



Maestría en

SISTEMAS DE INFORMACIÓN



Mención Inteligencia de Negocios y Analítica de Datos Masivos.

Tesis previa a la obtención del título de Magíster en Sistemas de Información mención Inteligencia de Negocios y Analítica de Datos Masivos.

AUTOR: Aizprua Barrios Jaris Surya
Campoverde Pilco Ivonne Geovanna
Hidalgo Guamán Lesly Maribel
Parra Rosero Mario Andrés

TUTOR: Reyes Chacón Iván Galo

Análisis de datos sintéticos de ventas de productos de redes tecnológicas y comparación de dos Modelos de Series Temporales que permitan predecir su demanda

APROBACIÓN DEL TUTOR

Yo, Iván Reyes, certifico que conozco los autores/as del presente trabajo siendo los responsables exclusivos tanto de su originalidad y autenticidad, como de su contenido.

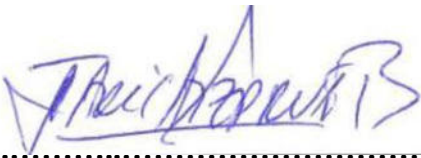
A handwritten signature in blue ink, appearing to read 'Iván Reyes', is positioned above a horizontal line.

Ing. Iván Reyes Mgtr.
DIRECTOR DE TESIS

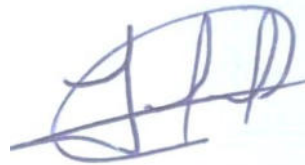
CERTIFICACIÓN DE AUTORÍA

Nosotros, Aizprua Barrios Jaris Surya, Campoverde Pilco Ivonne Geovanna, Hidalgo Guamán Lesly Maribel, Parra Rosero Mario Andrés, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador, para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



.....
Aizprua Barrios Jaris Surya



.....
Campoverde Pilco Ivonne Geovanna



.....
Hidalgo Guamán Lesly Maribel



.....
Parra Rosero Mario Andrés

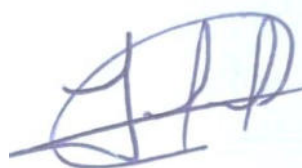
AUTORIZACIÓN DE DERECHOS DE PROPIEDAD INTELECTUAL

Nosotros, Aizprua Barrios Jaris Surya, Campoverde Pilco Ivonne Geovanna, Hidalgo Guamán Lesly Maribel, Parra Rosero Mario Andrés, en calidad de autores del trabajo de investigación titulado Análisis de datos sintéticos de ventas de productos de redes tecnológicas y comparación de dos Modelos de Series Temporales que permitan predecir su demanda autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, Octubre 2024



Aizprua Barrios Jaris Surya



Campoverde Pilco Ivonne Geovanna



Hidalgo Guamán Lesly Maribel



Parra Rosero Mario Andrés

DEDICATORIAS

A mi mamá, ella nos dejó recientemente pero siempre nos motivó a superarnos y seguir creciendo.

Aizprua Barrios Jaris Surya

El presente proyecto está dedicado a mis padres Geovana Pilco y Marco Campoverde, que han sido mi principal pilar a lo largo de toda mi vida y quienes en cada momento me han motivado a alcanzar cada uno de los objetivos que me he planteado.

A mis hermanos Marco Xavier Campoverde, Santiago Campoverde y a mi sobrino Nicolas Campoverde, que han estado siempre para mí y quienes me motivan siempre a mejorar.

A mi novio Jimmy Puente, porque fue la persona que me motivo a seguir la maestría y quien me contagia las ganas de aprender, investigar y superar los retos que se presenten.

Campoverde Pilco Ivonne Geovanna

A mis padres y hermano, por creer en mí y apoyarme en cada paso de mi vida. Gracias por su paciencia y por enseñarme el valor del esfuerzo y la perseverancia.

A mis amigos, por estar en los momentos difíciles y compartir alegrías y retos. Su compañía y consejos hicieron que este proceso fuera más llevadero.

Este proyecto es el resultado del apoyo que he recibido de todos ustedes.

Hidalgo Guamán Lesly Maribel

A Dios, por darme vida y salud, por permitirme tener y disfrutar a mi hermosa familia, por guiarme en cada etapa de mi vida permitiéndome crecer y ser una persona de bien en todo aspecto posible.

A mi esposa, por su amor, compañía y apoyo incondicional en cada momento y decisión tomada

A mis padres y hermanos, por su comprensión y ayuda en cada paso de mi vida.

A mis abuelitas, por sus sabios y largos consejos de vida, y a mis abuelitos, que sé que me acompañan, ya que siento su presencia en todo momento.

Parra Rosero Mario Andrés

AGRADECIMIENTOS

Gracias mamá por todo lo que hizo por nosotros.

Aizprua Barrios Jaris Surya

Agradezco a DIOS, por darme la oportunidad de vivir este momento y porque es el motor principal de toda acción en mi vida.

A mis abuelitos tanto maternos como paternos, que ya sea en vida o desde el cielo, han cuidado de mí y me han bendecido en todos los procesos de mi vida.

A mis padres Geovana Pilco y Marco Campoverde, por haberme guiado en cada paso y apoyarme en cada decisión.

A mis dos hermanos y mi sobrino, porque son mi fuente de motivación, por el apoyo que me brindan y por estar siempre en todos los momentos buenos y malos.

A mi novio, por haberme motivado a alcanzar este logro, y por todo el apoyo que me ha dado en este proceso de aprendizaje.

A los docentes de la maestría por los conocimientos impartidos, por su paciencia y disposición para ayudar en todo momento.

Campoverde Pilco Ivonne Geovanna

Quiero expresar mi más sincero agradecimiento a todas aquellas personas que hicieron posible la realización de este proyecto. A mis padres, hermano, amigos por su apoyo incondicional y su confianza en mis capacidades. Su motivación constante me ha ayudado a superar los desafíos a lo largo de este proceso, a no rendirme y seguir adelante con determinación.

A todos, gracias por ser parte de este logro.

Hidalgo Guamán Lesly Maribel

En primer lugar, agradezco a Dios por darme cada vez un día más de vida y permitirme llegar a este momento tan importante dentro de mi crecimiento profesional y pudiendo compartirlo con las personas que amo.

A mi esposa, que siempre está a mi lado en todo momento, sin importar la situación que se presente, llenando mi vida de felicidad, le amo.

A mis padres y hermanos, que siempre están dispuestos a brindarme su ayuda y apoyo, los amo.

Parra Rosero Mario Andrés

RESUMEN

En un mundo cada vez más impulsado por la conectividad, la demanda de productos de tecnología de redes, como routers, ha experimentado un crecimiento significativo, especialmente tras el aumento del trabajo remoto, la educación en línea y el entretenimiento digital durante la pandemia de COVID-19. Predecir con precisión la demanda de estos productos es crucial para que las empresas mantengan niveles óptimos de inventario, reduzcan costos y mejoren la eficiencia operativa. Este proyecto se centra en la aplicación de modelos de series temporales, específicamente SARIMA y Prophet, para pronosticar la demanda de productos tecnológicos utilizando datos de ventas históricos y sintéticos. El objetivo es identificar un modelo predictivo robusto capaz de adaptarse a las dinámicas del mercado y optimizar la gestión de inventarios.

El estudio subraya la importancia de una predicción precisa de la demanda para optimizar la producción, la gestión de la cadena de suministro y la satisfacción del cliente. Además, destaca la creciente relevancia del uso de datos sintéticos en el análisis, lo que permite modelar sin comprometer la privacidad. Se evaluó el desempeño de los modelos SARIMA y Prophet, siendo Prophet el que arrojó mejores resultados gracias a su capacidad para manejar irregularidades y datos faltantes. Las conclusiones de este análisis ofrecen valiosos insights para las empresas del sector tecnológico, especialmente en redes, permitiéndoles mejorar sus estrategias de predicción de demanda y obtener una ventaja competitiva.

Palabras Claves

Predicción de demanda, series temporales, SARIMA, Prophet, datos sintéticos, gestión de inventarios, productos tecnológicos, tecnología de redes, análisis de ventas, modelos predictivos.

ABSTRACT

In a world increasingly driven by connectivity, the demand for network technology products, such as routers, has seen a significant rise, especially following the surge of remote work, online education, and digital entertainment during the COVID-19 pandemic. Accurately forecasting demand for these products is critical for companies to maintain optimal inventory levels, reduce costs, and improve operational efficiency. This project focuses on the application of time series models, specifically SARIMA and Prophet, to predict the demand for technological products using both historical and synthetic sales data. The goal is to identify a robust predictive model that can adapt to market dynamics and enhance inventory management.

The study emphasizes the importance of accurate demand forecasting in optimizing production, supply chain management, and customer satisfaction. Furthermore, it highlights the growing relevance of synthetic data for analysis, ensuring data privacy while allowing for comprehensive modeling. The performance of SARIMA and Prophet models is evaluated, with Prophet showing superior results due to its ability to handle irregularities and missing data more effectively. The conclusions drawn from this analysis offer valuable insights for companies in the tech industry, particularly in network technologies, enabling them to improve their demand forecasting strategies and gain a competitive edge.

Keywords

Demand forecasting, time series, SARIMA, Prophet, synthetic data, inventory management, technological products, network technology, sales analysis, predictive models.

LISTA DE CONTENIDOS

CERTIFICACIÓN DE AUTORÍA	I
AUTORIZACIÓN DE DERECHOS DE PROPIEDAD INTELECTUAL	II
APROBACIÓN DE DIRECCIÓN Y COORDINACIÓN DEL PROGRAMA.....	III
DEDICATORIAS	IV
AGRADECIMIENTOS.....	VI
RESUMEN.....	VIII
ABSTRACT	IX
LISTA DE CONTENIDOS	X
LISTA DE TABLAS.....	XII
LISTA DE FIGURAS	XIII
CAPÍTULO I.....	1
INTRODUCCIÓN	1
1.1 Introducción.....	1
1.2 Justificación del proyecto	2
1.3 Necesidades.....	3
1.4 Objetivos.....	4
1.4.1 Objetivo General.....	4
1.4.2 Objetivos Específicos.....	4
1.5 Alcance del proyecto.....	4
1.5.1 Análisis Pestel.....	4
1.5.2 Alcance del proyecto.....	17
1.6 Stakeholders y áreas del negocio involucradas.....	17
1.7 Impacto de negocio	18
1.8 Fuentes de información.....	19
CAPITULO II	20
METODOLOGÍA	20
2.1 Series Temporales.....	20
2.2.1 Introducción.....	20
2.2.2 Sarima	22
2.2.3 Prophet	25
2.2 Arquitectura	28
2.2.1 Arquitectura de Delta Lake.....	28
2.2.2 Fuentes de datos.....	29

2.2.3	Orquestación de datos	30
2.2.4	Procesamiento de datos	31
2.2.5	Análisis e informes	32
2.2.6	Costo de la arquitectura	33
2.3	Cronograma de Ejecución de Tareas	34
2.4	Planificación de Recursos	35
2.5	Planteamiento Metodología Agile	37
2.6	Seguridad de la Información	38
2.6.1	Anonimización de datos	39
2.6.2	Datos Sintéticos	39
CAPITULO III.....		40
DESARROLLO.....		40
3.1	Aprovisionamiento de la arquitectura	40
3.2	Ingesta y almacenamiento de la información	40
3.3	Análisis exploratorio de los datos	44
3.4	Tratamiento de los datos	46
3.5	Aplicación de modelos	50
CAPITULO IV		57
RESULTADOS		57
4.1	Informe de Visualización (Dashboard)	57
4.2	Análisis de Resultados (Mejor modelo)	57
4.3	KPI y Métricas de Valor	60
4.2.3	Indicadores Clave	60
4.2.4	Métricas de Valor	61
4.4	Implementación en Producción y Aplicabilidad Entorno Real	63
CAPITULO V.....		65
CONCLUSIONES Y RECOMENDACIONES.....		65
5.1	Conclusiones	65
5.2	Recomendaciones	66
REFERENCIAS BIBLIOGRÁFICAS.....		68
APÉNDICES		70

LISTA DE TABLAS

Tabla 1	<i>Costos por Atributo de Azure DataLake</i>	30
Tabla 2	<i>Costos por Atributo de Azure Factory</i>	31
Tabla 3	<i>Costos por Atributo de Azure Databricks</i>	32
Tabla 4	<i>Costos por Atributo de Power BI</i>	33
Tabla 5	<i>Costos por Atributo de Microsof Entra ID</i>	34
Tabla 6	<i>Cálculo costo mensual y anual de la arquitectura propuesta</i>	34
Tabla 7	<i>Diagrama de Gantt</i>	34
Tabla 8	<i>Planificación de recursos por hora y perfil</i>	36

LISTA DE FIGURAS

Figura 1 <i>Diagrama de arquitectura Microsoft Azure</i>	28
Figura 2 <i>Metodología Scrum</i>	37
Figura 3 <i>Archivo fuente de datos cargado en la cuenta de almacenamiento</i>	40
Figura 4 <i>Código creación tabla delta Bronze</i>	41
Figura 4 <i>Tabla delta Bronze.AST_Ventas en el catálogo</i>	41
Figura 5 <i>Pipeline de Ingesta de datos - Origen</i>	42
Figura 6 <i>Pipeline de Ingesta de datos - Destino</i>	42
Figura 7 <i>Código verificación datos tabla delta Bronze.AST_Ventas</i>	43
Figura 8 <i>Código revisión dimensión del dataframe</i>	44
Figura 9 <i>Código revisión primeras filas del dataframe</i>	44
Figura 10 <i>Código revisión últimas filas del dataframe</i>	44
Figura 11 <i>Código visualización resumen del dataframe</i>	45
Figura 12 <i>Código visualización estadísticas del dataframe</i>	45
Figura 13 <i>Código eliminación duplicados del dataframe</i>	46
Figura 14 <i>Código filtrado ventas del remitente seleccionado</i>	46
Figura 15 <i>Código filtrado marcas correctas del remitente</i>	47
Figura 16 <i>Código filtrado inicial de columnas necesarias para curación datos</i>	47
Figura 17 <i>Código lectura de tabla delta Bronze.AST_Diccionario con diccionario de modelos</i>	47
Figura 18 <i>Código estandarización y curación de modelos routers con diccionario</i>	48
Figura 19 <i>Código generación dataframe para series temporales</i>	48
Figura 20 <i>Código generación dataframe tranpuesto para unidades por producto</i>	49
Figura 21 <i>Código reemplazo de valores Nan</i>	49
Figura 22 <i>Código generación tabla delta Silver.AST_VentasProcesadas</i>	49

Figura 23	<i>Código verificación datos tabla delta Silver.AST_VentasProcesadas</i>	50
Figura 24	<i>Código librerías necesarias para aplicación de modelos series temporales</i>	51
Figura 25	<i>Código aplicación prueba ADF</i>	51
Figura 26	<i>Código visualización resultados prueba ADF</i>	52
Figura 27	<i>Código filtro columnas no estacionarias</i>	52
Figura 28	<i>Código división conjuntos entrenamiento y pruebas</i>	53
Figura 29	<i>Código aplicación modelo SARIMA</i>	53
Figura 30	<i>Código visualización predicciones modelo SARIMA</i>	54
Figura 31	<i>Código aplicación modelo Prophet</i>	55
Figura 32	<i>Código visualización predicciones modelo Prophet</i>	55
Figura 33	<i>Código unificación predicciones modelos SARIMA y PROPHET</i>	56
Figura 34	<i>Código generación tabla delta Gold.AST_VentasPredicciones</i>	56
Figura 35	<i>Código verificación datos tabla delta Gold.AST_VentasPredicciones</i>	56
Figura 36	<i>Dashboard Predicción Venta de Routers</i>	57
Figura 37	<i>Proyección ventas de un producto sin ventas actuales</i>	58
Figura 38	<i>Proyección ventas de un producto con histórico y ventas actuales</i>	59
Figura 39	<i>Proyección ventas de un producto “nuevo”</i>	59
Figura 40	<i>Fórmula de cálculo de indicador Total Unidades Pronosticadas vs Vendidas</i>	60
Figura 41	<i>Fórmula de cálculo de indicador Unidades Pronosticadas vs Vendidas</i>	60
Figura 42	<i>Fórmula de cálculo de indicador Porcentaje de Unidades por tecnología Wifi</i>	61
Figura 43	<i>Fórmula de cálculo de indicador Porcentaje de Unidades más vendidas</i>	61

CAPÍTULO I

INTRODUCCIÓN

1.1 Introducción

En un mundo cada vez más interconectado, la demanda de productos de tecnología que faciliten la conectividad a Internet ha mostrado una tendencia creciente. Esta demanda es impulsada por la expansión del teletrabajo, la educación en línea y el entretenimiento digital, fenómenos que se aceleraron significativamente a partir de la pandemia de COVID-19 (Smith, 2021). Dada la naturaleza fluctuante de la demanda de estos productos, resulta crucial para las empresas del sector desarrollar modelos predictivos eficientes que les permitan ajustar su producción y gestión de inventarios de manera proactiva.

En la actualidad, la gestión eficiente del inventario es un factor crítico para el éxito de las empresas, especialmente en la industria tecnológica, donde los ciclos de vida de los productos son cortos y la demanda es altamente volátil. La capacidad de predecir con precisión la demanda futura de productos tecnológicos puede significar la diferencia entre mantener niveles óptimos de stock y enfrentar pérdidas significativas por exceso o falta de inventario.

Según Hans Franses et al. (2014) el análisis de series temporales emerge como una herramienta poderosa en este contexto, permitiendo a los analistas y a las empresas anticipar tendencias y patrones futuros basándose en datos históricos. La aplicación de modelos avanzados de series temporales, como los modelos SARIMA y los modelos de aprendizaje profundo, ha demostrado ser particularmente efectiva en la predicción de ventas en diversos sectores económicos (Lee, 2020).

El siguiente proyecto se centra en la aplicación de modelos de series temporales para el pronóstico de ventas de productos de tecnología de redes y comunicaciones. A través de un

preprocesamiento y análisis exhaustivo de datos históricos junto a datos generados de manera sintética se buscará identificar un modelo robusto que no solo ofrezca predicciones precisas, sino que también sea capaz de adaptarse a las dinámicas cambiantes del mercado y permita a las empresas optimizar su capacidad de respuesta ante la demanda del mercado, sus estrategias de marketing, ventas y robustecer su posición competitiva.

1.2 Justificación del proyecto

Importancia de la Predicción de la Demanda: La predicción precisa de la demanda es crucial para la planificación de la producción, la gestión de inventarios y la satisfacción del cliente. Según un estudio de Chopra et al. (2013), "una predicción precisa de la demanda permite a las empresas optimizar su cadena de suministro, reduciendo costos y mejorando la eficiencia operativa".

Relevancia de los Datos Sintéticos: El uso de datos sintéticos en análisis de ventas es cada vez más común debido a la creciente necesidad de proteger la privacidad de los datos y al acceso limitado a datos reales en ciertas circunstancias. Como señala Bowser (2019), "los datos sintéticos ofrecen una alternativa viable para el análisis y modelado de datos sin comprometer la privacidad de los individuos".

Eficiencia de los Modelos de Series Temporales: Los modelos de series temporales, como SARIMA y Prophet, son herramientas poderosas para la predicción de la demanda debido a su capacidad para capturar patrones temporales y estacionalidades. Hyndman et al. (2018) explican que "los modelos de series temporales son ampliamente utilizados en pronósticos debido a su capacidad para manejar datos dependientes del tiempo y proporcionar predicciones precisas".

Comparación de Modelos para Mejorar la Precisión: Comparar diferentes modelos de predicción permite identificar el más adecuado para un conjunto específico de

datos, mejorando así la precisión de las predicciones. Según Makridakis et al. (2018), "la comparación de diferentes enfoques de modelado es esencial para seleccionar el modelo más eficaz y garantizar predicciones confiables".

Impacto en la Industria de Tecnologías de Red: La industria de tecnologías de red, incluyendo routers, es altamente competitiva y sujeta a rápidos cambios tecnológicos. La capacidad de predecir la demanda en este sector es vital para mantener una ventaja competitiva. Un informe de MarketsandMarkets (2021) destaca que "las predicciones precisas de la demanda son fundamentales para las empresas de tecnologías de red, permitiéndoles responder rápidamente a las fluctuaciones del mercado y a las innovaciones tecnológicas".

1.3 Necesidades

Predecir y gestionar el inventario de los productos dependiendo la temporalidad: Predecir con precisión la demanda de productos proporciona a las empresas la oportunidad de optimizar sus niveles de inventario, reducir el exceso de stock y el riesgo de desabastecimiento. Esto mejora la eficiencia operativa y reduce costos asociados con el almacenamiento y la gestión de inventarios.

Optimizar la cadena de suministro, coordinar actividades de producción, distribución y logística: Las predicciones precisas de la demanda, permiten a las empresas coordinar sus actividades e incrementar la capacidad de respuesta ante cambios en la demanda del mercado, mejorando la flexibilidad y eficiencia de la cadena de suministro.

Apoyar en la toma de decisiones estratégicas en diferentes áreas: La información obtenida de la aplicación de modelos de predicción permite tomar decisiones estratégicas de

mercado, marketing, entre otras. Estas decisiones ayudan al crecimiento del negocio y a la obtención de ventajas competitivas.

Reducir riesgos por medio de las predicciones de demanda y la gestión eficiente de recursos y control de la información: Evita las inversiones y gastos innecesarios, facilita el cumplimiento de regulaciones de privacidad de datos y protege la información sensible.

1.4 Objetivos

1.4.1 Objetivo General

Evaluar los resultados de dos modelos de series temporales aplicados a datos históricos sintéticos de ventas que ayuden en la predicción de stock de productos tecnológicos y permitan mejorar la predicción y optimización de la gestión de inventarios.

1.4.2 Objetivos Específicos

Realizar el análisis exploratorio del conjunto de datos para entender las características, distribuciones y posibles relaciones entre diferentes variables que puedan influir en la predicción del stock de productos.

Aplicar y evaluar dos modelos de series temporales diferentes SARIMA y Prophet para predecir la demanda de stock.

Diseñar indicadores y métricas de valor que permitan optimizar la gestión de inventarios de manera proactiva de acuerdo con el nivel de demanda.

1.5 Alcance del proyecto

1.5.1 Análisis Pestel

Político

En cuanto al factor político, Ecuador tiene regulaciones específicas en telecomunicaciones y tecnología que las empresas dedicadas a esta comercialización deben cumplir, esto incluye normativas de importación y estándares técnicos. La estabilidad política es otra variable que influye en las operaciones de las empresas dedicadas a las ventas de producto tecnológico ya que dependen de cambios en el gobierno o en las políticas.

Regulaciones de Telecomunicaciones

La Agencia de Regulación y Control de las Telecomunicaciones (ARCOTEL), es el organismo encargado de regular, controlar y supervisar el sector de las telecomunicaciones en Ecuador, como dictan sus artículos. (ARCOTEL, 2015):

Derechos de los Usuarios (Art. 22):

- **Privacidad y Protección de Datos:** Los usuarios tienen derecho a que sus datos personales sean protegidos por los proveedores de servicios de telecomunicaciones.
- **Información Clara y en español:** Los usuarios deben recibir información precisa, gratuita y no engañosa sobre los servicios y tarifas en su idioma predominante.
- **Acceso a Servicios de Emergencia:** Los usuarios tiene en derecho a servicios gratuitos de llamadas de emergencia e información relevante.

Obligaciones de los Prestadores de Servicios (Art. 24):

- **Protección de Datos:** Los proveedores deben adoptar medidas para proteger los datos personales de los usuarios.

- **Seguridad de Redes:** Deben garantizar la seguridad y calidad de las redes de telecomunicaciones.
- **Manejo de Infraestructura y Desechos:** Deben cumplir con normas ambientales y de infraestructura, como el soterramiento y manejo de desechos tecnológicos.
- **Planes de Contingencia:** Deben tener planes para asegurar la continuidad del servicio durante emergencias.

Regulación de Mercados (Art. 31):

- **Evaluación de Competencia:** La autoridad regula y revisa periódicamente los mercados relevantes de telecomunicaciones para asegurar la competencia justa.

Homologación de Equipos (Art. 86-87):

- **Obligatoriedad de Homologación:** Equipos de telecomunicaciones deben ser homologados y certificados para garantizar la seguridad y evitar interferencias.
- **Prohibiciones:** Está prohibido el uso y comercialización de equipos no homologados, que generen interferencias, o sean incompatibles con las normativas.

Estabilidad política

La estabilidad política en Ecuador es crucial para las empresas comercializadoras de tecnología, sus operaciones y estrategias comerciales, dependen de: Cambio de Gobiernos, Políticas Gubernamentales, Seguridad y Orden Público, Relaciones Internacionales, Políticas Económicas.

En Ecuador la estabilidad política se ha visto afectada por protestas que han paralizado las actividades económicas del país, afectando el PIB, como lo indica el Banco Central, el valor de las pérdidas representa el 1% del Producto Interno Bruto (PIB) corriente (USD 116.010 millones) de acuerdo con la previsión macroeconómica del BCE publicada el 15 de septiembre de 2022, además lo que ocasiona poco interés en inversión extranjera. (Villarreal, 2022)

La inseguridad por la que atraviesa el país es otro factor que afecta al desarrollo económico, como lo mencionan los medios de comunicación. La inseguridad impacta negativamente en la economía. Se paralizan los procesos de producción, comercio y consumo, y por ende se reduce el empleo, el ingreso y el bienestar.

La inversión, que ya estaba contraída al alcanzar los USD 796 por persona en 2022 frente a USD 957 de 2019 (17% menos) y a USD 1.162 de 2014 (32% menos), se deprimirá aún más por el contexto de inseguridad que desincentiva a potenciales inversionistas nacionales e internacionales. (Mideros Mora, 2024)

Económico

Condiciones Económicas Generales:

Debido a las fluctuaciones de varios factores como los estragos de la pandemia de COVID-19, el aspecto tecnológico y el uso del internet ha aumentado considerablemente y se ha impulsado la digitalización y la adopción de nuevas tecnologías en varios sectores. Antes del año 2020, la penetración del consumo de internet era del 46%, durante la pandemia creció en un 40% y posterior a la pandemia, la penetración del consumo de internet a crecido un 65%. (La Republica, 2021)

El estado de la economía en el 2019, debido a la caída de los precios del petróleo y los problemas políticos, influyeron en que en ese año el crecimiento del PIB sea de 0.1%, lo que generó un enfoque limitado en la transformación digital. En el año 2020, la pandemia del COVID-19 provocó una caída del PIB de alrededor de un 7.8% y la adopción de la tecnología tuvo que acelerarse. A partir del 2021 el PIB tuvo un crecimiento del 4%, lo cual fue posible debido a que varios sectores como la educación, el comercio, la banca e incluso las PYMES adoptaron tecnología para acoplarse a la nueva era a la cual no enfrentamos, dando como resultado un aumento del 30% en el uso e implementación de tecnología. (World Bank Group, 2021)

La tasa de desempleo durante el año 2019 creció en un 6%, sin embargo, con la adaptación y uso de la tecnología en el año 2023 disminuyó en un 3.55%. (CEIC, 2024)

Las empresas buscan modernizar su infraestructura de TI, sistemas y aplicación de modelos para mejorar la eficiencia y competitividad, lo que abre oportunidades para empresas que optimicen sus operaciones mediante análisis predictivos. Hasta el año 2019 el uso y aplicación de Inteligencia artificial y modelos de análisis predictivos era menor al 5%, a partir del año 2021 esto aumentó en un 10 al 15%, actualmente estos modelos y la IA se han consolidado como sectores claves creciendo entre un 20 y 25%. (Pew Research Center, 2021) (Banco Interamericano de Desarrollo (BID), n.d.)

Tasas de Interés y Acceso a Financiamiento

En el Ecuador se han implementado diferentes políticas para facilitar el acceso a créditos y reducciones de tasas de interés, con el objetivo de estimular la inversión en diversos sectores como el tecnológico, estas medidas fueron cruciales para la reactivación post COVID-19, aumentando en un 25% con respecto a la situación del país antes del 2019. (World Bank Group, 2021)

Uno de los programas lanzados en el año 2021, que aumento la inversión extranjera fue el “1X30”, que genero un aumento del 15% en inversión en sectores energéticos, mineros y de tecnología. (El Telégrafo, 2023) (BanEcuador, s.f.)

Para el año 2023 se ha observado un aumento del 20% en el financiamiento de proyectos tecnológicos y a partir del 2021, la cartera de créditos del BNF creció en un 30%, reflejando un mayor acceso al financiamiento para inversiones en estos sectores. (Pew Research Center, 2021)

Costo de Tecnologías y Herramientas:

Ecuador depende en gran medida de la importación de tecnologías avanzadas y software necesarios para proyectos de análisis de datos y modelado de series temporales. Los costos pueden ser elevados debido a impuestos y aranceles sobre productos tecnológicos importados. A partir del 2021 el incremento de importación de equipos de telecomunicaciones creció en un 29.3%, para el año 2023 se reportó un crecimiento del 20%, estas cifras indican la aceleración de la adopción de tecnologías y refleja una estabilización del mercado conforme las empresas y hogares se adaptaron a la nueva normalidad tecnológica. (MINTEL, s.f.)

Social

“POLÍTICA PARA LA TRANSFORMACIÓN DIGITAL DEL ECUADOR 2022-2025, Ministerio de Telecomunicaciones.

Infraestructura, Penetración de Internet, Equipamiento del hogar, Porcentaje de hogares con acceso a internet a nivel nacional 53.20%.

Los cambios en preferencias y comportamientos de los consumidores afectan directamente en la demanda de routers. Como se observó en la pandemia el aumento de

teletrabajo y la educación en línea incrementó la necesidad de una conexión internet rápida y confiable. Los consumidores buscan routers de alto rendimiento para garantizar una experiencia fluida en videoconferencias y transmisión de datos, lo que ha aumentado la demanda de routers con tecnologías avanzadas en las empresas distribuidoras.

Tecnológico

Los avances en tecnología de redes como Wi-Fi 6 y acceso de fibra óptica mediante GPON, pueden crear nuevas demandas de routers con capacidades mejoradas. La empresa comercializadora de routers debe mantenerse a la vanguardia de los avances tecnológicos, garantizando conectividad eficiente y segura en un mundo cada vez más digitalizado. Su liderazgo en innovación permite a consumidores y negocios mantenerse competitivos y preparados para el futuro.

Infraestructura: En Ecuador el “PLAN DE SERVICIO UNIVERSAL 2022-2025” del Ministerio de Telecomunicaciones y de la Sociedad de la Información establece mecanismos para el despliegue o migración a redes de alta velocidad con el fin de contribuir al mejoramiento de calidad del servicio de acceso a internet en los hogares ecuatorianos.

Uno de los mecanismos indica el despliegue de infraestructura de fibra óptica en el país, en las redes de transporte y última milla; masificando la oferta del servicio de internet fijo por fibra óptica, en especial en las zonas priorizadas. Para lo mencionado, el Indicador No 7 “Porcentaje de hogares con acceso al servicio de Internet fijo a través de enlaces de fibra óptica” tiene como meta al 2024 un 38,05% y para 2025 un 42,09%. Estos porcentajes proyectados y la calidad de la infraestructura de telecomunicaciones en Ecuador afectará directamente al rendimiento y demanda de routers.

Ecológico

En Ecuador el manejo de desechos electrónicos está regulado por varias leyes y normativas que buscan garantizar la gestión adecuada de estos residuos, con el objetivo de proteger el medio ambiente y la salud pública. Algunas de las principales leyes y normativas son:

- **Ley de Gestión Ambiental (1999):** Establece el marco general para la gestión ambiental en el país. Incluye disposiciones sobre la gestión de residuos peligrosos, dentro de los cuales se consideran los desechos electrónicos.
- **Código Orgánico del Ambiente (2017):** Este es uno de los marcos legales más recientes y amplios en materia ambiental en Ecuador. En él se establece la obligación de gestionar de manera adecuada los desechos electrónicos, promoviendo la reducción, reutilización y reciclaje. Se consideran a estos desechos como residuos peligrosos.
- **Acuerdo Ministerial 161 (2008):** Este acuerdo establece las directrices para la gestión de desechos electrónicos en el país, incluyendo las responsabilidades de productores, importadores, distribuidores y comercializadores de aparatos electrónicos y eléctricos. Obliga a las empresas a recolectar y gestionar adecuadamente estos residuos.
- **Política Nacional de Residuos Sólidos (2010):** Define las políticas nacionales para la gestión integral de residuos sólidos, incluyendo los desechos electrónicos, promoviendo la minimización de generación de residuos y fomentando su adecuada disposición final.
- **Resolución No. 067 (2020):** Esta resolución emitida por el Ministerio del Ambiente y Agua regula la gestión de residuos especiales, que incluyen a los desechos

electrónicos, promoviendo la recolección y reciclaje a través de sistemas de responsabilidad extendida del productor (REP).

Adicionalmente por parte del Ministerio del Ambiente, Agua, y Transición Ecológica, se desarrolló el documento “Guía informativa sobre la Residuos de Aparatos Eléctricos y Electrónicos RAEE en Ecuador” el cual tiene como propósito brindar información accesible sobre la gestión de residuos electrónicos en el territorio ecuatoriano, dicho documento fue publicado en el mes de mayo del 2024.

En el caso de productos electrónicos tipo routers, los fabricantes suelen mencionar la clasificación del equipo y sugerencias de reciclaje en el manual o guía de usuario que viene en la caja del equipo; por ejemplo, para equipos del fabricante TP-Link podemos encontrar lo siguiente:

- **Símbolos de clasificación selectiva:** Los productos vienen marcados con símbolos que indican que se clasifica bajo la categoría de RAEE (Residuos de Aparatos Eléctricos y Electrónicos), dicha clasificación engloba a todos los dispositivos eléctricos y electrónicos que al final de su vida útil, se convierten en residuos y por lo tanto deben ser entregados en un punto de recolección autorizado para el reciclaje o desecharlo siguiendo las normativas vigentes en el país o región de uso. Estos símbolos forman parte del esfuerzo de la Unión Europea para asegurar que los desechos eléctricos y electrónicos se reciclen de manera adecuada.
- **Directiva Europea 2012/19/UE:** Esta directiva proporciona el marco legal para la recolección, tratamiento y reciclaje de residuos electrónicos dentro de la Unión Europea. Su objetivo es minimizar el impacto ambiental de los residuos electrónicos, garantizando que estos productos se manejen correctamente al final de su ciclo de vida.

- **Reciclaje o desmontaje:** La directiva enfatiza la importancia de reciclar o desmontar el producto para reducir su impacto ambiental. Este proceso asegura que se puedan recuperar materiales valiosos y gestionar adecuadamente las sustancias peligrosas.

En Ecuador existen varias organizaciones que se dedican al reciclaje y destrucción de desechos electrónicos cumpliendo con las normativas ambientales vigentes, de esa manera promueven la reducción del impacto ambiental; a continuación, se mencionan algunas de las más relevantes:

- **Recicla Ecuador:** Se especializa en la recolección, tratamiento, y reciclaje de residuos electrónicos. Ofrecen servicios tanto a empresas como a individuos, y garantizan una gestión adecuada y segura de los residuos.
- **Remar Ecuador:** Aunque su enfoque principal es el apoyo social, Remar Ecuador tiene un programa de reciclaje de desechos electrónicos. Recolectan estos desechos, los reciclan y destinan los fondos recaudados a proyectos sociales.
- **SERTECPET S.A.:** Se dedican a la gestión integral de residuos peligrosos, incluyendo residuos electrónicos. Ofrecen servicios de reciclaje y destrucción de estos residuos conforme a la normativa ecuatoriana.
- **Reciclaje Electrónico ECOLTEC:** Se dedica a la gestión de residuos tecnológicos, recolectando, clasificando y reciclando equipos electrónicos en desuso. También ofrecen servicios de destrucción segura de información almacenada en dispositivos electrónicos.

- **Fundación Hermano Miguel:** Cuenta con un programa de reciclaje de residuos electrónicos llamado "Reciclaje Inclusivo". Este programa busca no solo el reciclaje de electrónicos, sino también la inclusión social de personas con discapacidad a través del trabajo en el reciclaje.

Legal

Las regulaciones de telecomunicaciones en Ecuador son fundamentales para las empresas comercializadoras de productos tecnológicos, ya que afectan directamente su capacidad para operar y comercializar sus productos en el país. Además de la protección de datos como responsabilidad de cada empresa con los consumidores de sus productos.

- **Regulaciones de telecomunicaciones**
 - Agencia de Regulación y Control de las Telecomunicaciones (ARCOTEL)
 - **Licencias y Autorizaciones:** Las empresas debe obtener las licencias y autorizaciones necesarias para operar y comercializar sus productos en Ecuador. Esto incluye la certificación de conformidad con las normativas técnicas establecidas por ARCOTEL.
 - **Registro de Equipos:** Todos los dispositivos de telecomunicaciones que hagan uso del espectro inalámbrico, tales como routers WiFi, deben estar registrados y aprobados por ARCOTEL antes de su comercialización en el país.
- **Normativas Técnicas**
 - **Cumplimiento de Especificaciones Técnicas:** Los productos tecnológicos deben cumplir con las especificaciones técnicas exigidas por ARCOTEL para garantizar

su funcionamiento adecuado y seguro. Esto incluye normas relacionadas con la compatibilidad electromagnética, seguridad eléctrica y eficiencia espectral.

- **Pruebas y Certificaciones:** las empresas deben someter sus productos a pruebas de conformidad y obtener las certificaciones necesarias para demostrar el cumplimiento de las normativas técnicas.
- **Regulaciones de Importación y Comercialización**
 - **Aranceles y Tarifas:** Los productos importados están sujetos a aranceles y tarifas aduaneras. La empresa debe asegurarse de cumplir con todas las regulaciones aduaneras para evitar retrasos y costos adicionales.
 - **Etiquetado y Manuales:** Los productos deben cumplir con las normativas de etiquetado y deben incluir los manuales respectivos que expliquen la instalación de los equipos.
- **Protección del Consumidor**
 - **Garantías y Servicio al Cliente:** la empresa debe ofrecer garantías y atención al cliente que cumplan con las leyes de protección del consumidor en Ecuador. Esto incluye el reemplazo o reembolso en productos defectuosos.
 - **Transparencia en la Información:** La empresa debe proporcionar información clara y precisa sobre las características y capacidades de sus productos, evitando publicidad engañosa.
- **Ciberseguridad y Protección de Datos**
 - **Normativas de Ciberseguridad:** la empresa debe asegurarse de que sus productos cumplan con las normativas locales para proteger contra ciberataques y garantizar la seguridad de la red.

- **Protección de Datos:** Las empresas deben cumplir con la Ley Orgánica de Protección de Datos Personales en Ecuador, asegurando que los datos de los usuarios se manejen de manera segura y respetuosa con su privacidad.
- **Competencia y Monopolio**
 - Regulación de la Competencia: la empresa debe operar de manera justa y evitar prácticas monopólicas que puedan ser sancionadas por la Superintendencia de Control del Poder de Mercado (SCPM).

Protección de datos

- **Ley Orgánica de Protección de Datos Personales (LOPDP)**
 - Publicada en mayo de 2021, esta ley establece un marco legal para la protección de datos personales en Ecuador.
 - Tratamiento Lícito y Transparente: Los datos deben ser manejados de manera justa y clara, recopilándolos solo para propósitos específicos y legítimos.
 - Consentimiento: La empresa debe obtener el consentimiento explícito e informado de los usuarios antes de recopilar y procesar sus datos personales.
 - Derechos de los Titulares de Datos: Los usuarios tienen derechos sobre sus datos, como acceso, corrección, eliminación, oposición y portabilidad.
 - Responsabilidad y Transparencia: las empresas deben mostrar cumplimiento con la normativa de protección de datos, manteniendo registros detallados de las actividades de tratamiento.
- **Transferencia Internacional de Datos**

- Transferencias Seguras: las empresas deben garantizar que cualquier transferencia de datos personales fuera de Ecuador se realice de manera segura y cumpla con las normativas locales.

1.5.2 Alcance del proyecto

El proyecto se centrará en el análisis de datos sintéticos generados de ventas de routers. Lo que se busca es predecir la demanda futura de routers mediante el uso de modelos de series temporales. La predicción de la demanda es esencial para la gestión efectiva de la cadena de suministro y la planificación de inventarios.

Se compararán dos modelos de series temporales, como SARIMA y Prophet, para evaluar el desempeño en la predicción de la demanda. Se realizará un análisis exploratorio de los datos sintéticos que permitan identificar patrones, tendencias y estacionalidades que puedan influir en la demanda de routers. Este análisis exploratorio es un paso crucial que ayuda a comprender los datos, definir las variables objetivas y aplicar los modelos predictivos.

Finalmente se compararán los resultados de la aplicación de los modelos, que permitirá la selección del método más adecuado en base al contexto del negocio y las necesidades de gestionar el inventario disponible.

1.6 Stakeholders y áreas del negocio involucradas

Los stakeholders/áreas que participan en el proyecto son:

Internos

- **Equipo de Ciencia de Datos:** Científicos, analistas e ingenieros de datos responsables del manejo de la información.
- **Equipo de Producto o Marketing:** Gerentes de Producto, analistas de mercado interesados en los insights de ventas que permitan entender la demanda futura.
- **Equipo de Ventas:** Representantes de ventas interesados en las predicciones para establecer objetivos o ajustar los enfoques de captación.
- **Equipo de TI:** Encargados de la implementación tecnológica, aseguramiento de la información.
- **Equipo de Logística:** Requieren predicciones de demanda para optimizar gestión de inventarios y la cadena de suministro.

Externos

- **Clientes:** Aunque no estén directamente involucrados, sus comportamientos y demanda son el foco del análisis.
- **Distribuidores o Partners:** Podrían estar interesados en las predicciones para planificar su propio inventario y estrategias de distribución.

1.7 Impacto de negocio

Optimización de Inventarios y mejora la Eficiencia Operativa: Al predecir con precisión la demanda de routers, la organización puede mantener niveles de inventario óptimos, reduciendo tanto el exceso de stock como el riesgo de desabastecimiento. Esto disminuye costos de almacenamiento y mejora la eficiencia operativa. La capacidad de anticipar la demanda permite a la organización ajustar sus procesos operativos, mejorando la eficiencia en la producción y distribución.

Mejora en la Planificación de la Producción: La predicción precisa de la demanda permite una mejor planificación de la producción, asegurando que la capacidad de producción se ajuste adecuadamente a las necesidades del mercado.

Reducción de Riesgos Financieros: Una mejor predicción de la demanda ayuda a evitar la sobreproducción y el exceso de inventarios, lo que reduce los riesgos financieros asociados con el almacenamiento y la depreciación de productos.

Soporte a la Toma de Decisiones Estratégicas: La información precisa y oportuna sobre la demanda futura permite a los directivos tomar decisiones estratégicas más informadas en áreas como el lanzamiento de nuevos productos, la entrada a nuevos mercados y la planificación de promociones.

1.8 Fuentes de información

Como fuente de información se utilizará los datos correspondientes a las estadísticas de comercio exterior a través de la empresa Datos Ecuador (<https://www.datosecuador.com/>). Aproximadamente se dispone de novecientos mil registros de ventas históricas desde Enero de 2016 hasta Julio de 2024.

La base de información presenta datos cualitativos y cuantitativos relacionados a:

- Información de la empresa
- Información del producto (Marca, Descripción, Modelo, Origen, Cantidad)
- Información Financiera (Costo, Seguro)
- Información de tiempo (Fecha Venta)

CAPITULO II

METODOLOGÍA

2.1 Series Temporales

2.2.1 *Introducción*

Las series temporales son una rama de la estadística que se enfoca en analizar datos recolectados a través del tiempo. Este análisis es fundamental en múltiples disciplinas como la economía, la meteorología, la medicina y la ingeniería, donde es crucial no solo comprender la evolución histórica de los datos, sino también prever su comportamiento futuro. A diferencia de los datos aleatorios o independientes, los datos de series temporales tienen la característica distintiva de estar correlacionados temporalmente, lo que implica que las observaciones en puntos temporales cercanos tienden a estar más relacionadas entre sí que aquellas separadas por intervalos de tiempo más largos (Wei, 2006).

El análisis de series temporales implica varios pasos clave, entre los que destacan la identificación de patrones, la modelización, y la predicción. Los patrones comunes incluyen tendencias (cambios a largo plazo), estacionalidad (fluctuaciones periódicas) y ciclos, que son variaciones que ocurren en intervalos irregulares y se diferencian de la estacionalidad. La identificación y comprensión de estos patrones son esenciales para construir modelos precisos y útiles (George, Gwilym, Gregory, & Greta, 2015).

En la práctica, el análisis de series temporales enfrenta varios desafíos, como la estacionalidad, las tendencias no lineales, y la presencia de datos atípicos. Además, la precisión de las predicciones puede verse afectada por eventos externos o cambios estructurales en los datos, lo que requiere una modelización cuidadosa y, en muchos casos, el uso de técnicas avanzadas como el modelado de funciones de transferencia o la descomposición de series temporales (Chatfield, 2004).

Entre los modelos más utilizados en el análisis de series temporales se encuentra el modelo SARIMA (Seasonal Autoregressive Integrated Moving Average). SARIMA es un modelo paramétrico que combina los componentes principales: la estacionalidad (S), la autorregresión (AR), la integración (I), y la media móvil (MA). El componente S representa la estacionalidad que se refiere a patrones repetitivos en los datos, AR captura la dependencia lineal entre una observación y sus valores pasados, la integración (I) es utilizada para transformar la serie temporal en estacionaria, y el componente MA modela la relación entre una observación y los errores residuales de observaciones anteriores (Hyndman & Athanasopoulos, 2018). El proceso de especificación de un modelo SARIMA implica la identificación del orden de cada componente, lo que generalmente se hace mediante un enfoque iterativo que incluye pruebas de diagnóstico como los correlogramas de autocorrelación y autocorrelación parcial (Lütkepohl, 2005).

Otro modelo destacado es Prophet, desarrollado por Facebook, que se ha convertido en una herramienta popular debido a su capacidad para manejar datos con fuertes efectos de estacionalidad y descomponer las series temporales en componentes interpretables. Prophet es particularmente útil para series temporales que contienen múltiples ciclos estacionales y cambios en tendencias, y es conocido por su facilidad de uso, incluso en escenarios donde los datos presentan irregularidades o faltan algunos valores. El enfoque de Prophet combina modelos aditivos para capturar tendencia, estacionalidad y efectos festivos o de eventos, lo que lo convierte en una opción robusta para aplicaciones comerciales y de investigación (Taylor, 2018).

En síntesis, el análisis de series temporales se basa en la identificación de patrones temporales y la construcción de modelos que puedan capturar estas estructuras para predecir futuros valores. Modelos como SARIMA y Prophet son herramientas fundamentales en este campo, cada uno con sus fortalezas para diferentes tipos de datos y problemas de predicción.

La elección del modelo adecuado y su correcta implementación son críticos para el éxito en la predicción de series temporales (Shumway, 2017).

2.2.2 *Sarima*

El modelo SARIMA (Seasonal Autoregressive Integrated Moving Average) es una técnica fundamental para el análisis y pronóstico de series temporales. Se utiliza para modelar datos que pueden ser estacionarios o que requieren diferenciación para alcanzar la estacionariedad que se estudian desde un punto de vista estocástico o moderno y que permiten utilizar otros métodos para series más largas. (Mauricio, 2007)

A continuación, se describen sus componentes y el proceso de modelado.

- **AR (AutoRegresivo):** Este componente modela la dependencia entre una observación y un número determinado de rezagos (valores anteriores). Es decir, el valor de la serie en el tiempo t depende linealmente de los valores anteriores en $t-1$, $t-2$, etc.
- **I (Integrado):** Este componente se refiere al número de diferencias que se deben aplicar para que la serie temporal se vuelva estacionaria. Una serie es estacionaria cuando sus propiedades estadísticas (como la media y la varianza) no cambian con el tiempo.
- **MA (Media Móvil):** Este componente modela la dependencia entre una observación y un error residual de un modelo de media móvil aplicado a las observaciones anteriores.
- **Ordenes Estacionales:** En el modelo SARIMA, se especifican dos conjuntos de órdenes: (Morales Oñate, 2024)(Martin Mateos, 2020)
 - **(p,d,q):** Para los componentes no estacionales.

- **(P,D,Q,s):** Para los componentes estacionales, donde S corresponde a la periodicidad de la estacionalidad.

Un modelo SARIMA se denota de la siguiente manera:

$$P, D, Q$$

Donde:

P o p: Orden de la parte autorregresiva (AR), que representa la relación entre una observación y un número de observaciones anteriores.

D o d: Orden de diferenciación (I), que indica cuántas veces se debe diferenciar la serie para que sea estacionaria.

Q o q: Orden de la parte de media móvil (MA), que representa la relación entre una observación y un número de errores de predicción anteriores. (Ríos, 2008)

S: Ventana de tiempo del patrón estacional. (Morales Oñate, 2024)

Proceso de Modelado SARIMA

- **Identificación:** Se analiza la serie temporal para determinar si es estacionaria. Si no lo es, se seleccionan los órdenes correspondientes mediante análisis de autocorrelación y autocorrelación parcial.
- **Estimación:** Se ajusta el modelo SARIMA a los datos utilizando técnicas de estadísticas que minimicen el error en las predicciones.
- **Diagnóstico:** Se evalúan los residuos del modelo para asegurarse de que se comporten como ruido blanco, lo que indica que el modelo ha capturado la estructura de la serie.

- **Pronóstico:** Una vez validado el modelo, se utiliza para realizar pronósticos futuros. (Jenkins & Pelham Box)

Ventajas del Modelo SARIMA

- **Captura de Estacionalidad:** Permite modelar tanto componentes estacionales como no estacionales, lo cual lo hace eficaz al momento de analizar patrones repetitivos.
- **Flexibilidad:** Puede modelar una amplia variedad de patrones en series temporales.
- **Precisión del Pronóstico:** Proporciona pronósticos más precisos en comparación de modelos ARIMA estándar.
- **Aplicabilidad:** Se pueden aplicar a diferentes contextos o áreas en donde la estacionalidad es un factor crítico.
- **Manejo de Dependencias a Largo Plazo:** Captura dependencias a corto plazo a través del componente MA y a largo plazo a través del componente AR, lo que permite que el análisis sea más completo. (Técnicas de Trading, 2024)

Limitaciones del Modelo SARIMA

- **Dificultad en la Estimación**
 - En series temporales que tienen bastantes valores faltantes puede tener dificultad.

- Requiere una cuidadosa selección de parámetros caso contrario puede generar un mal ajuste.
- **Limitaciones de Cómputo**
 - En series temporales robustas o largas puede tomar bastante tiempo computacional de procesamiento.
- **Dificultad en Series Volátiles**
 - Si existe cambios abruptos que generan un dinamismo en las series temporales no es lo suficientemente flexible para ajustarse.

2.2.3 *Prophet*

Prophet es un modelo de series temporales de código abierto desarrollado por Facebook. Su enfoque se basa en un modelo aditivo que permite descomponer las series temporales en componentes de tendencia, estacionalidad y efectos de días festivos. Esta herramienta es particularmente útil para conjuntos de datos que presentan estacionalidades marcadas y que pueden tener datos faltantes o valores atípicos. (AWS, 2024) (Alvarez, 2024)

Entre las ventajas que proporciona este modelo, se encuentran:

- **Tendencia:** Permite modelar la tendencia como un crecimiento lineal que puede cambiar en diferentes segmentos del tiempo y ajustarse a los cambios.
- **Estacionalidad:** Al utilizar series de Fourier permite modelar estacionalidades que se ajusten a los ciclos.
- **Pronostico:** Es eficaz en contextos de pronóstico de datos con patrones estacionales complejos ya que tiene la capacidad de relacionar automáticamente los patrones sin una formación matemática profunda.

- **Robustez:** Está diseñado para manejar datos faltantes y valores atípicos de manera efectiva. Lo que es útil en situaciones donde los datos pueden ser incompletos o contener errores, lo que puede afectar a otros modelos de pronóstico.
- **Facilidad de Uso:** Su facilidad de implementación y lo intuitivo que es, permite que su uso no dependa de experiencia avanzada en estadística para realizar pronósticos precisos. El modelo requiere poca configuración y permite a los usuarios centrarse en los datos en lugar de en la complejidad del modelo.
- **Visualización de resultados:** Prophet incluye herramientas para visualizar las tendencias y estacionalidades subyacentes, lo que ayuda a los analistas a interpretar los resultados y comunicar sus hallazgos de manera efectiva. (Meta, 2017)

Componentes de Tendencia en Prophet

Prophet utiliza un modelo de tendencia que puede ser lineal o logístico.

- **Lineal:** Se representa como una línea recta
- **Logística:** Incluye un término de saturación que limita el crecimiento, evitando que la tendencia crezca indefinidamente.

La tendencia se modela como una función continua que puede cambiar en varios puntos a lo largo del tiempo, conocidos como puntos de cambio. Estos puntos permiten que la pendiente de la tendencia se ajuste, reflejando cambios abruptos en la dirección de la serie temporal. (AWS, 2024)

- **Detección Automática de Puntos de Cambio:** Prophet tiene la capacidad de detectar automáticamente estos puntos de cambio, lo que le permite adaptarse a

cambios inesperados en la tendencia. Esto es especialmente útil en series temporales donde los patrones de comportamiento pueden variar significativamente a lo largo del tiempo.

- **Descomposición Aditiva:** La descomposición de la serie temporal se realiza de manera aditiva, donde la serie se expresa como la suma de la tendencia, la estacionalidad y el error. Esto permite a los analistas entender cómo cada componente contribuye al comportamiento general de la serie temporal.
- **Inferencia Bayesiana:** Prophet utiliza inferencia bayesiana para estimar los parámetros del modelo de tendencia. Esto implica el uso del teorema de Bayes para determinar las distribuciones más probables de los parámetros basándose en los datos observados.

Estructura del Modelo

El modelo de Prophet se puede expresar como:

$$y(t) = G(t) + S(t) + H(t) + \epsilon(t)$$

Donde:

- $y(t)$ es el valor observado en el tiempo
- $G(t)$ es el componente de tendencia.
- $S(t)$ es el componente de estacionalidad.
- $H(t)$ es el efecto de días festivos.
- $\epsilon(t)$ representa el error o residual.

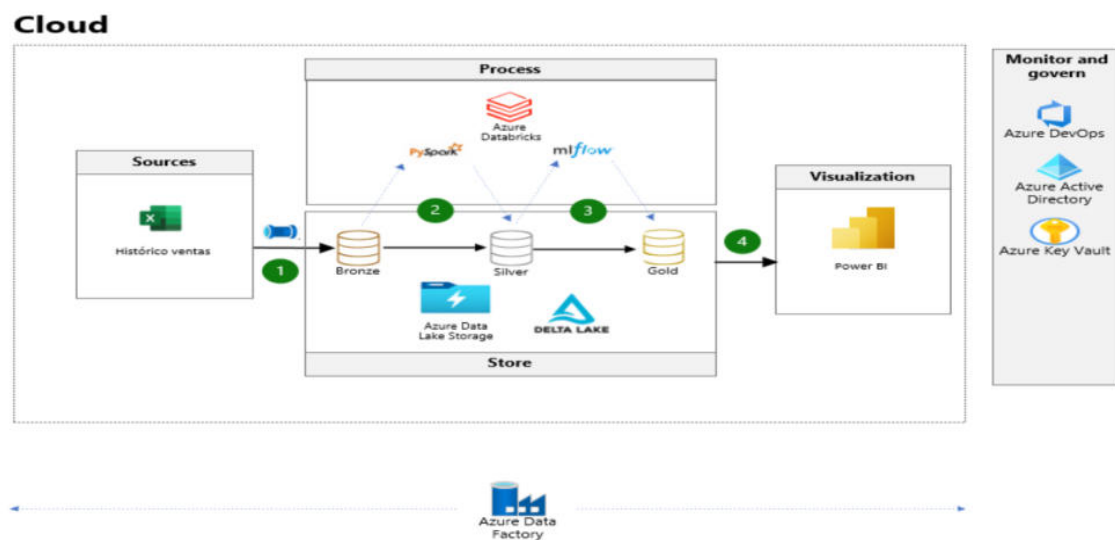
Esta estructura permite a Prophet realizar pronósticos precisos al descomponer los datos en sus componentes fundamentales, facilitando así el análisis y la interpretación de las series temporales. (Mirete Blanco, 2019)

2.2 Arquitectura

La arquitectura propuesta está diseñada para soportar una gran carga de datos y usuarios sin comprometer su rendimiento, estabilidad o eficiencia. En este contexto para el proyecto se ha decidido utilizar una arquitectura Delta Lake basada en Microsoft Azure. En la siguiente figura se encuentra el ecosistema definido:

Figura 1

Diagrama de arquitectura Microsoft Azure



Nota: El gráfico representa las herramientas a utilizar en la arquitectura propuesta. Autoría propia.

2.2.1 Arquitectura de Delta Lake

Delta Lake es una capa de datos optimizada entre los datos brutos y los motores de procesamiento. La base de estas tablas está colocada sobre un lago de datos, extendiéndolas como archivos de datos Parquet. (Microsoft, 2024)

Esta arquitectura es una evolución de una arquitectura lambda, permite un alto rendimiento sobre los datos, ya que construye índices que entregan un tiempo de recuperación más rápido. Su diseño está constituido sobre 3 capas para el flujo de datos:

Capa Bronce: Contiene información bruta sin procesar, ingerida de múltiples fuentes

Capa Plata: Contiene información con cierto tratamiento y procesos de ingeniería.

Capa Oro: Contiene información resultante lista para usuarios finales de informes.

2.2.2 Fuentes de datos

Los datos de origen están compuestos por archivo/s plano/s en formato Excel que almacenan toda la información de ventas de los fabricantes de routers hacia los distribuidores. La herramienta utilizada para esta etapa es Azure DataLake del proveedor Microsoft.

Azure DataLake: También conocido como Lago de datos, es un repositorio centralizado que puede almacenar todo tipo de datos, tanto estructurados o no estructurados. Es decir, permite a las organizaciones centralizar la información en su formato original sin necesidad de adaptarlos a una estructura. (Microsoft, 2024)

Sus principales características son:

- Están contruidos sobre Azure Blob Storage, lo que permite escalar horizontalmente para manejar hasta exabytes de datos, es decir puede manejar cantidades enormes de datos sin preocuparse de los límites de capacidad.
- Permite a la organización escalar sus operaciones hacia arriba o hacia abajo dependiendo la demanda sin interrumpir el servicio.
- Utiliza tecnologías Apache Spark que permiten el procesamiento en paralelo de grandes volúmenes de datos, facilitando su análisis y procesamiento.

Para el cálculo del costo se considera las características propias de cada herramienta que el proveedor factura y las necesidades del proyecto. dando un total de \$1,22 mensuales de acuerdo con el detalle presentado a continuación:

Tabla 1

Costos por Atributo de Azure DataLake

Azure DataLake		
Atributo	Detalle	Costo Mensual
Capacidad	Hasta 51,200 GB/mes	\$ 0,019
Operaciones Escritura	Cada 4 MB, por 10 000	\$ 0,650
Operaciones Lectura	Cada 4 MB, por 10 000	\$ 0,520
Otras operaciones	Cada 4 MB, por 10 000	\$ 0,030
	Total	\$ 1,219

2.2.3 Orquestación de datos

Para conseguir un completo ciclo de vida del dato, debe existir una automatización que permita el flujo de información a través de las plataformas analíticas. La herramienta utilizada para esta etapa es Azure DataFactory del proveedor Microsoft.

Azure DataFactory: Es una herramienta que permite la automatización del procesamiento de datos, tiene compatibilidad con una gran gama de orígenes de datos.

(Microsoft, 2024)

Sus principales características son:

- Permite escalar horizontalmente sin preocuparse por las limitaciones físicas de los recursos locales.
- Permite la creación de pipelines elásticos que pueden escalar según sea necesario para manejar grandes volúmenes de datos.

- Los pipelines pueden ejecutarse en paralelo y en múltiples entornos lo que aumenta la eficiencia y la capacidad de procesamiento.

Para el cálculo del costo se considera las características propias de cada herramienta que el proveedor factura y las necesidades del proyecto, dando un total de \$8,65 mensuales de acuerdo con el detalle presentado a continuación:

Tabla 2

Costos por Atributo de Azure Factory

Azure DataFactory		
Atributo	Detalle	Costo Mensual
Orquestación	Por 1000 ejecuciones	\$ 1,00
Actividad movimiento datos	Por 30 horas	\$ 7,50
Actividad canalización	Por 30 horas	\$ 0,15
	Total	\$ 8,65

2.2.4 Procesamiento de datos

Para extraer el valor de los datos es esencial un tratamiento de la información, el cual mantenga una correcta exploración, limpieza y modelado de los datos. La herramienta utilizada para esta etapa es Azure Databricks del proveedor Microsoft.

Azure Databricks: Es una herramienta que permite crear, implementar y mantener soluciones de analítica de datos que impulsan las decisiones empresariales. (Microsoft, 2024)

Sus principales características son:

- Está construido sobre Apache Spark, lo que permite un procesamiento paralelo distribuido de grandes volúmenes de datos.

- Permite el ajuste de los clústeres de computación en función de la carga de trabajo, es decir puede escalar horizontalmente agregando más nodos cuando aumenta la carga o reduciéndolos cuando esta disminuye.
- Tiene una optimización de rendimiento como la ejecución de memoria y el uso de técnicas avanzadas de gestión de recursos, reduciendo tiempos de procesamiento y mejorando el manejo de grandes conjuntos de datos.

Para el cálculo del costo se considera las características propias de cada herramienta que el proveedor factura y las necesidades del proyecto, dando un total de \$252,00 mensuales de acuerdo con el detalle presentado a continuación:

Tabla 3

Costos por Atributo de Azure Databricks

Azure Databricks		
Atributo	Detalle	Costo Mensual
Proceso multiuso	Por 60 horas	\$ 24,00
Proceso trabajos	Por 60 horas	\$ 9,00
Instancia virtual de proceso	Por mes	\$ 219,00
	Total	\$ 252,00

2.2.5 Análisis e informes

El objetivo de la presentación de resultados es trasladar el conocimiento a los implicados en el proyecto. La visualización debe diseñarse cuidadosamente ya que no toda la audiencia tiene el conocimiento teórico o técnico para comprender lo que se ha realizado. La herramienta utilizada para esta etapa es Power BI del proveedor Microsoft.

Power BI: Es una herramienta que combina servicios y conectores para convertir orígenes de datos en información relevante, interactiva de manera visual. (Microsoft, 2024)

Sus principales características son:

- Permite la creación de informes que pueden compartirse y usarse empresarialmente para la toma de decisiones.
- Utiliza un motor de análisis VertiPaq, que es altamente optimizado para consultas rápidas y procesamiento en memoria, el cual le permite manejar conjuntos de datos grandes y complejos de manera eficiente.

Para el cálculo del costo se considera las características propias de cada herramienta que el proveedor factura y las necesidades del proyecto, dando un total de \$10,00 mensuales de acuerdo con el detalle presentado a continuación:

Tabla 4

Costos por Atributo de Power BI

Power BI		
Atributo	Detalle	Costo Mensual
Licencia	Licencia Pro por usuario	\$ 10,00
	Total	\$ 10,00

2.2.6 Costo de la arquitectura

Las herramientas detalladas anteriormente componen la arquitectura propuesta para cumplir las necesidades analíticas del proyecto, es importante considerar que para dar acceso inicialmente a los recursos encargados del diseño e implementación es necesario contar con un servicio adicional que gestiona usuarios y la seguridad en las herramientas, en este caso al utilizar recursos Microsoft el servicio se llama Microsoft Entra ID. A continuación, se detalla su costo:

Tabla 5*Costos por Atributo de Microsoft Entra ID*

Microsoft Entra ID		
Atributo	Detalle	Costo Mensual
Usuarios Premium	4 usuarios	\$ 24,00
	Total	\$ 24,00

Una vez detallado las herramientas necesarias para abastecer la arquitectura propuesta con Microsoft Azure, se puede concluir que el costo anual que tendría su implementación es de \$3350,44 de acuerdo con el detalle presentado a continuación:

Tabla 6*Cálculo costo mensual y anual de la arquitectura propuesta*

Herramienta	Costo Mensual	Costo Anual
Azure DataLake	\$ 1,22	\$ 14,64
Azure DataFactory	\$ 8,65	\$ 103,80
Azure Databricks	\$ 252,00	\$ 3024,00
Power BI	\$ 10,00	\$ 120,00
Microsoft Entra ID	\$ 24,00	\$ 288,00
	Total	\$ 3550,44

2.3 Cronograma de Ejecución de Tareas

En la tabla a continuación se presenta el cronograma macro de realización de las etapas establecidas para la ejecución del proyecto y su tiempo estimado en semanas:

Tabla 7*Diagrama de Gantt*

No	Etapas	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
-----------	---------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	------------	------------	------------	------------	------------	------------

1	Levantamiento del proyecto															
2	Diseño de la arquitectura															
3	Recopilación y procesamiento															
4	Aplicación de modelos															
5	Presentación de resultados															

2.4 Planificación de Recursos

Para el diseño del proyecto es necesario contar con los siguientes perfiles:

- **Ingeniero de Datos:** responsable de la extracción y preparación de los datos, que posteriormente serán tratados.
- **Analista de Datos:** responsable de procesar y analizar datos generando insights para la toma de decisiones.
- **Científicos de Datos:** responsable de explorar y analizar los datos aplicando técnicas de machine learning, minería de datos y modelos estadísticos para obtener predicciones y recomendaciones basadas en los datos.
- **Arquitecto de Datos:** responsable de diseñar el ecosistema analítico, definiendo como los datos se almacenarán, accederán o moverán a través de las diferentes herramientas y asegurando que la arquitectura sea segura, escalable y eficiente.

A continuación, se presenta la planificación de recursos por hora y perfil en base a las diferentes etapas planteadas para el proyecto:

Tabla 8*Planificación de recursos por hora y perfil*

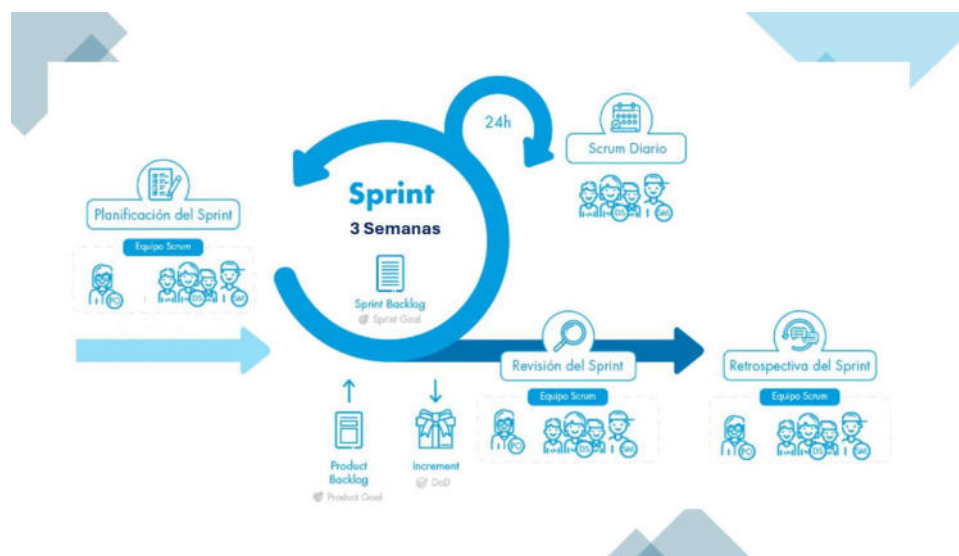
		Estimación	Ingeniero Datos	Científico Datos	Analista Datos	Arquitecto Datos
		Total	Horas	Horas	Horas	Horas
Diseño		22				
	Arquitectura					4
	Recopilación datos	4	4			
	Carga datos	2	2			
	Análisis, preprocesamiento	3	3			
	Selección modelos	2		2		
	Evaluación modelos	4		4		
	Presentación resultados	3			3	
Implementación		82				
	Arquitectura	10				10
	Recopilación datos	12	12			
	Carga datos	4	4			
	Análisis, preprocesamiento	20	20			
	Selección modelos	8		8		
	Evaluación modelos	16		16		
	Presentación resultados	12			12	
Implantación		85				
	Pruebas	40	10	20	10	
	Documentación	45	15	15	15	
TOTAL		189	70	65	40	14

2.5 Planteamiento Metodología Agile

Para la ejecución del proyecto se propone hacer uso de la metodología ágil de scrum, esto debido a que la aplicación y análisis de los modelos conlleva un proceso de revisión y mejora continua de los modelos empleados.

Figura 2

Metodología Scrum



Nota: El gráfico representa las fases de Scrum. Tomado de netmind, por Rodríguez M, 2020, Scrum: el pasado y el futuro.

Debido a esto se plantea la posibilidad de trabajar en 5 sprints de tres semanas cada uno, proyectados de la siguiente manera:

- **Sprint 1 Levantamiento del proyecto:** Definición de objetivos y planteamiento preliminar.
- **Sprint 2 Diseño de la arquitectura:** Establecimiento del ecosistema y herramientas analíticas que serán utilizadas.

- **Sprint 3 Recopilación y preprocesamiento:** Identificación de las fuentes de datos, carga de información y limpieza de datos.
- **Sprint 4 Aplicación de modelos:** Selección y evaluación de modelos de series temporales
- **Sprint 5 Presentación de resultados:** Análisis de los resultados obtenidos y elaboración del informe empresarial.

Se emplearán varias herramientas, artefactos y flujos de trabajo correspondientes a scrum y que se adaptan a la naturaleza del proyecto propuesto.

- **Planificación de Sprint:** Se realizará una reunión entre los involucrados para plantear los objetivos y planificar los ítems a atender
- **Daily:** Se mantendrá una comunicación diaria de los temas e impedimentos por resolver.
- **Review:** Se revisará al final de cada sprint los entregables obtenidos y se irán refinando en base a las observaciones
- **Retrospectiva:** Se evaluarán las fortalezas, oportunidades de mejora, debilidades y amenazas que puedan afectar al proyecto.
- Se mantendrá un control del backlog de tareas empleando un tablero Kanban que permita al equipo tener claridad del trabajo que debe realizarse.

2.6 Seguridad de la Información

Para el aseguramiento de la información adquirida, se resguardó campos sensibles como la razón social de las empresas propietarias de las ventas de los productos tecnológicos, con la finalidad de proteger la información y garantizar el manejo de los datos según la

LOPDP. Adicional se implementó técnicas de generación de datos sintéticos para completar información correspondiente al segundo semestre del año 2024.

2.6.1 Anonimización de datos

El proceso utilizado para proteger la razón social de las empresas es la anonimización de los datos, que se define como el “dato que no puede asociarse a una persona identificada o identificable por haberse destruido el nexo con toda información que identifique al sujeto, o porque dicha asociación exige un esfuerzo no razonable, entendiendo por tal el empleo de una cantidad de tiempo, gastos y trabajo desproporcionados”. (Martinez Velencoso, 2021)

La técnica de anonimización empleada para la protección de los datos es la de reemplazo/generalización, que utiliza etiquetas genéricas, por ejemplo: “Razón social de la empresa” por “Empresa1”. En esta técnica los nombres específicos son reemplazados por términos generales que mantienen la estructura y utilidad para el análisis, pero no exponen la identidad real.

2.6.2 Datos Sintéticos

Para obtener los datos sintéticos se utilizaron ciertas técnicas como:

- Transformaciones geométricas: Es decir se aplicaron ciertas rotaciones o recortes en los datos existentes para obtener nuevas variantes.
- Reglas basadas en conocimientos: Es decir se crearon datos sintéticos aplicando reglas específicas en base a un diccionario de modelos de routers, obteniendo nuevos registros pertenecientes a los modelos de routers a utilizar.

CAPITULO III

DESARROLLO

3.1 Aprovisionamiento de la arquitectura

Para el aprovisionamiento de la arquitectura se ha diseñado un Manual de Usuario con el paso a paso para la creación y configuración de cada recurso dentro de Microsoft Azure. El documento se encuentra como parte de los apéndices.

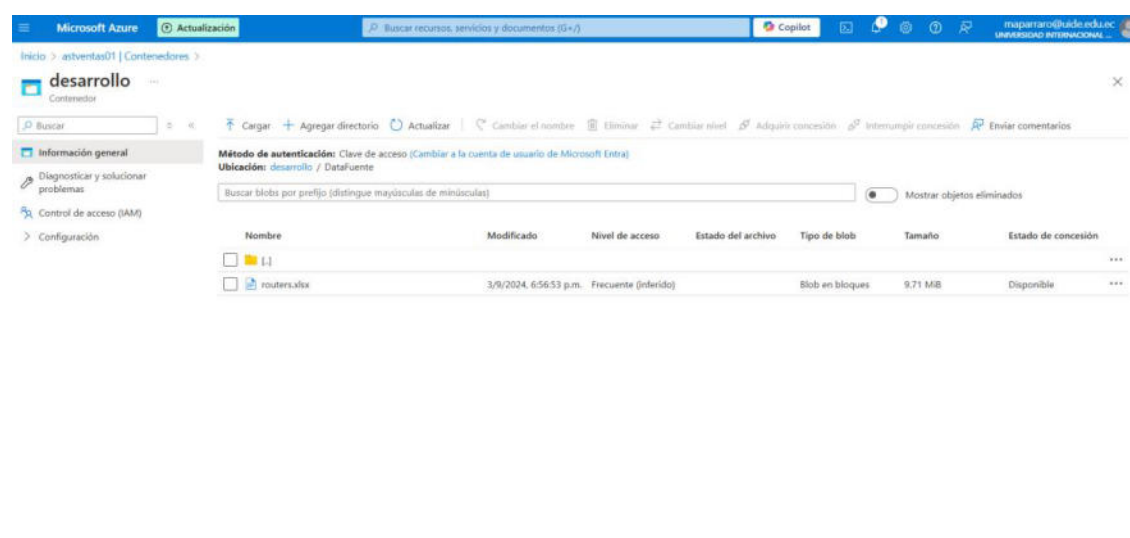
3.2 Ingesta y almacenamiento de la información

Para la carga de información se depositó el archivo histórico de datos en la cuenta de almacenamiento en el contenedor “desarrollo”, identificando como su ruta de acceso la siguiente:

desarrollo/DataFuente/datos_fabricantes.xlsx

Figura 3

Archivo fuente de datos cargado en la cuenta de almacenamiento



Previo a la carga de información se creó la tabla delta en la capa Bronze para su extracción.

Figura 4

Código creación tabla delta Bronze

```

CREATE TABLE spark_catalog.Bronze.ast_ventas (
  AÑO BIGINT,
  MES BIGINT,
  CAPITULO BIGINT,
  SUBCAPITULO BIGINT,
  RUC STRING,
  RAZON_SOCIAL STRING,
  RAZON_SOCIAL_DIRECCION STRING,
  REFRENDO STRING,
  NUME_SERIE BIGINT,
  TIPO_AFORO STRING,
  COD_REGIMEN BIGINT,
  REGIMEN STRING,
  DISTRITO STRING,
  AGENTE_AFIANZADO STRING,
  REMITENTE STRING,
  NOTIFY STRING,
  EMBARCADOR_CONSIGNEE STRING,
  EMBARCADOR_CONSIGNEE_ADDRESS STRING,
  AGENCIA STRING,
  LINEA STRING,
  MANIFIESTO STRING,
  FECH_EMBAR TIMESTAMP,
  FECH_LLEGA TIMESTAMP,
  FECH_INGRESO TIMESTAMP,
  FECH_PAGO TIMESTAMP,
  FECH_SALIDA TIMESTAMP,
  FACTURA STRING,
  BL STRING,
  NAVE STRING,
  VIAJE STRING,
  ALMACEN STRING,
  DEP_COMERCIAL STRING,
  PARTIDA BIGINT,
  TMAN BIGINT,
  TASA_ADVALOREM DOUBLE,
  DESC_ARAN STRING,

```

Figura 4

Tabla delta Bronze.AST_Ventas en el catálogo

ast_ventas

Overview Sample Data Details History

Filter columns...

Column	Type	Comment
AÑO	bigint	
MES	bigint	
CAPITULO	bigint	
SUBCAPITULO	bigint	
RUC	string	
RAZON_SOCIAL	string	
RAZON_SOCIAL_DIRECCION	string	
REFRENDO	string	
NUME_SERIE	bigint	
TIPO_AFORO	string	
COD_REGIMEN	bigint	
REGIMEN	string	
DISTRITO	string	
AGENTE_AFIANZADO	string	
REMITENTE	string	
NOTIFY	string	
EMBARCADOR_CONSIGNEE	string	
EMBARCADOR_CONSIGNEE_ADDRESS	string	
AGENCIA	string	
LINEA	string	
MANIFIESTO	string	
FECH_EMBAR	timestamp	
FECH_LLEGA	timestamp	
FECH_INGRESO	timestamp	
FECH_PAGO	timestamp	
FECH_SALIDA	timestamp	
FACTURA	string	
BL	string	
NAVE	string	
VIAJE	string	
ALMACEN	string	
DEP_COMERCIAL	string	
PARTIDA	bigint	
TMAN	bigint	
TASA_ADVALOREM	double	
DESC_ARAN	string	

About this table

Data source: [View](#)

Last updated: [View](#)

Size: 3.1MB, 4 files

En la plataforma ETL para realizar la ingesta de información se utilizó el data set configurado hacia Azure Databricks Delta Lake mediante una tarea de Copia de Datos, configurando como Origen el contenedor de almacenamiento y como Destino la tabla Delta en capa Bronze.

Figura 5

Pipeline de Ingesta de datos - Origen

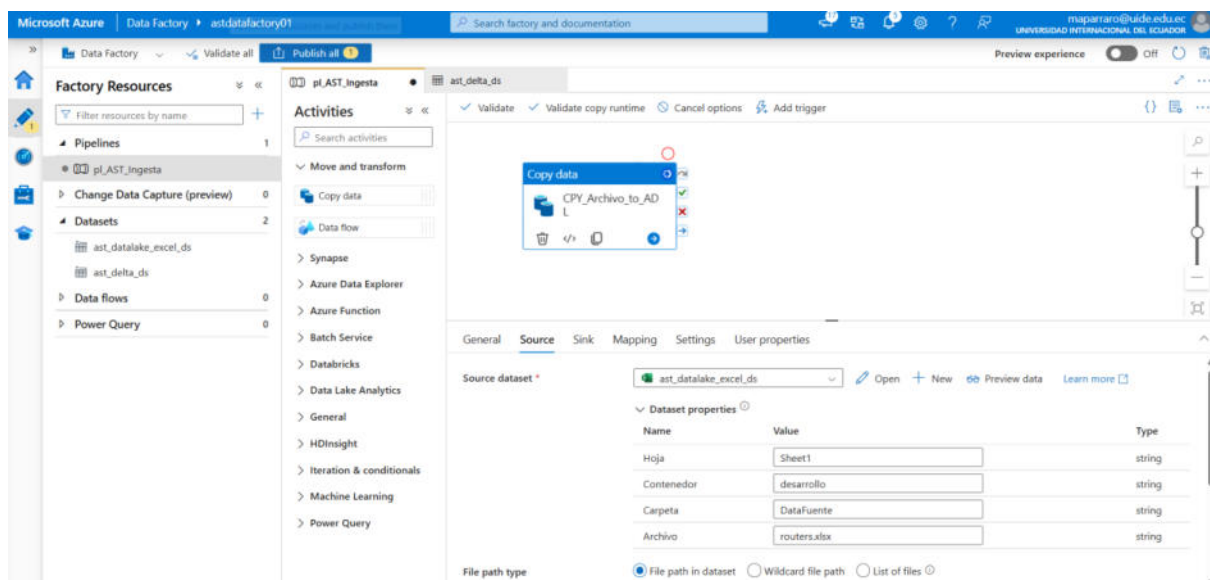
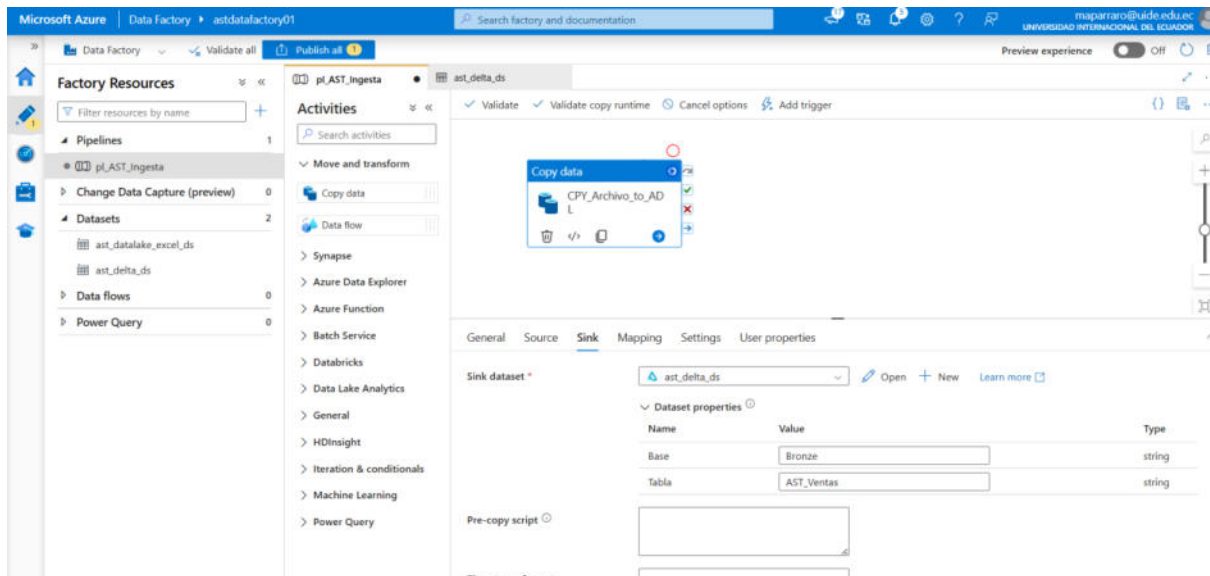


Figura 6

Pipeline de Ingesta de datos - Destino



Dado que se está utilizando una arquitectura Delta Lake la información cruda cargada se almacenó en una tabla Delta de la capa Bronze, para su posterior tratamiento.

Figura 7

Código verificación datos tabla delta Bronze.AST_Ventas

The screenshot shows a Jupyter Notebook with a SQL query executed. The result is displayed as a table with 14 rows and 7 columns. The columns are: AÑO, MES, CAPITULO, SUBCAPITULO, RUC, RAZON_SOCIAL, and RAZON_SOCIAL_DIRECCION. The data is as follows:

	AÑO	MES	CAPITULO	SUBCAPITULO	RUC	RAZON_SOCIAL	RAZON_SOCIAL_DIRECCION
1	2019	1	85		09927177500...	SMART TECHNOLOGY S.A. TECHSMART	OLMEDO 112 Y GUAYAQUIL
2	2019	1	85		09927177500...	SMART TECHNOLOGY S.A. TECHSMART	OLMEDO 112 Y GUAYAQUIL
3	2019	1	85		17922916660...	NOVISOLUTIONS CIA. LTDA.	AV. LOS SHYRIS N44-26 Y AV. RIO COCA
4	2019	1	85		17922916660...	NOVISOLUTIONS CIA. LTDA.	AV. LOS SHYRIS N44-26 Y AV. RIO COCA
5	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
6	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
7	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
8	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
9	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
10	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
11	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
12	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
13	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA
14	2019	1	85		17917431480...	INTCOMEX DEL ECUADOR S.A.	YANEZ PINZON 26-67 Y LA NIÑA

3.3 Análisis exploratorio de los datos

El análisis exploratorio de los datos es una fase crítica cuyo objetivo es comprender los datos, detectar anomalías o identificar patrones. En esta fase se aplicaron diferentes acciones detalladas a continuación:

Se realizó una inspección general del tamaño y de las últimas o primeras filas del conjunto de datos para tener una visión general.

Figura 8

Código revisión dimensión del dataframe

```
df.shape
(23031, 79)
```

Figura 9

Código revisión primeras filas del dataframe

```
df.head()
```

	AÑO	MES	CAPITULO	SUBCAPITULO	RUC	RAZON_SOCIAL	RAZON_SOCIAL_DIRECCION	REFERENDO	NUME_SERIE	TIPO_AFORO	COD_REGIMEN	REGIMEN	DISTRITO	AGENTE_AFIANZ
0	2022	3	85	8517	1091787616001	NEGOCIOS CESCH CIA LTDA.	IMBABURA ANTONIO ANTE ATUNTAQUI AVENIDA LEO...	028-2022- 10- 00273176	320	A	10	IMPORTACION A CONSUMO	GUAYAQUIL - MARITIMO	NOVACO SERVI ADUANE NOVACOMEXSA
1	2022	3	85	8517	1091787616001	NEGOCIOS CESCH CIA LTDA.	IMBABURA ANTONIO ANTE ATUNTAQUI AVENIDA LEO...	028-2022- 10- 00273176	330	A	10	IMPORTACION A CONSUMO	GUAYAQUIL - MARITIMO	NOVACO SERVI ADUANE NOVACOMEXSA
2	2022	3	85	8517	1091787616001	NEGOCIOS CESCH CIA LTDA.	IMBABURA ANTONIO ANTE ATUNTAQUI AVENIDA LEO...	028-2022- 10- 00273176	331	A	10	IMPORTACION A CONSUMO	GUAYAQUIL - MARITIMO	NOVACO SERVI ADUANE NOVACOMEXSA
3	2022	3	85	8517	1091787616001	NEGOCIOS CESCH CIA LTDA.	IMBABURA ANTONIO ANTE ATUNTAQUI AVENIDA LEO...	028-2022- 10- 00273176	332	A	10	IMPORTACION A CONSUMO	GUAYAQUIL - MARITIMO	NOVACO SERVI ADUANE NOVACOMEXSA
4	2022	3	85	8517	1091787616001	NEGOCIOS CESCH CIA LTDA.	IMBABURA ANTONIO ANTE ATUNTAQUI AVENIDA LEO...	028-2022- 10- 00273176	54	A	10	IMPORTACION A CONSUMO	GUAYAQUIL - MARITIMO	NOVACO SERVI ADUANE NOVACOMEXSA

Figura 10

Código revisión últimas filas del dataframe

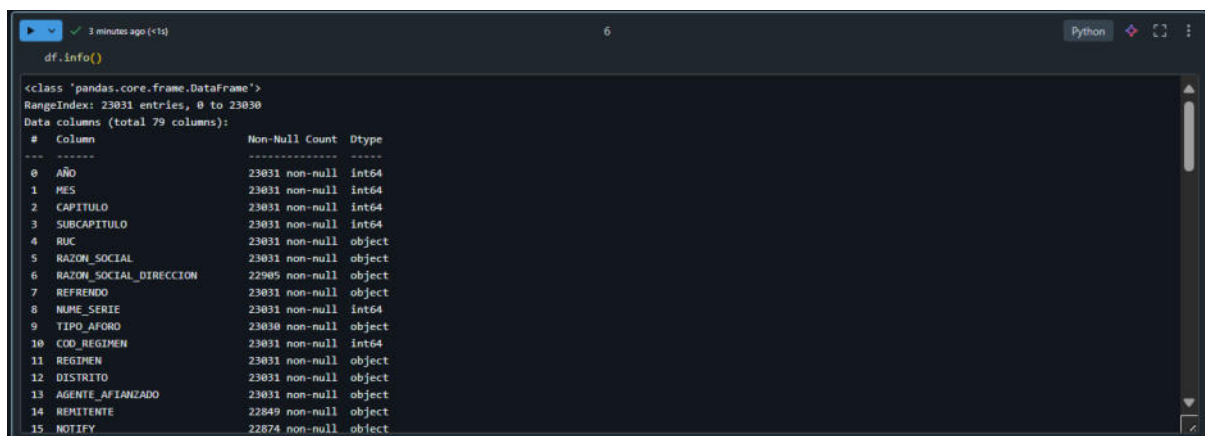


AÑO	MES	CAPITULO	SUBCAPITULO	RUC	RAZON_SOCIAL	RAZON_SOCIAL_DIRECCION	REFRENDO	NUME_SERIE	TIPO_AFORO	COD_REGIMEN	REGIMEN	DISTRITO	AGENTE_AFIA
23026	2020	9	85	8517 0993177725001	ODERTEK-TECHNOLOGIES S.A.	KM 5 VIA DURAN TAMBO INTERSECCION CON KM5	055-2020-10-00595277	22	A	10	IMPORTACION A CONSUMO	QUITO	ASTET AGEN ADUANA CIF
23027	2020	9	85	8517 1191714268001	NECUSOFT CIA. LTDA.	SUCRE 209 13 Y AZUAY LOJA ECUADOR	055-2020-10-00597295	1	A	10	IMPORTACION A CONSUMO	QUITO	ASTET AGEN ADUANA CIF
23028	2020	9	85	8517 1191714268001	NECUSOFT CIA. LTDA.	SUCRE 209 13 Y AZUAY LOJA ECUADOR	055-2020-10-00597295	17	A	10	IMPORTACION A CONSUMO	QUITO	ASTET AGEN ADUANA CIF
23029	2020	9	85	8517 1191714268001	NECUSOFT CIA. LTDA.	SUCRE 209 13 Y AZUAY LOJA ECUADOR	055-2020-10-00597295	18	A	10	IMPORTACION A CONSUMO	QUITO	ASTET AGEN ADUANA CIF
23030	2020	9	85	8517 1191714268001	NECUSOFT CIA. LTDA.	SUCRE 209 13 Y AZUAY LOJA ECUADOR	055-2020-10-00597295	19	A	10	IMPORTACION A CONSUMO	QUITO	ASTET AGEN ADUANA CIF

Se obtuvo información del conjunto de datos como el número de filas, columnas y tipos de datos.

Figura 11

Código visualización resumen del dataframe



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23031 entries, 0 to 23030
Data columns (total 79 columns):
#   Column                Non-Null Count  Dtype
---  -
0   AÑO                    23031 non-null  int64
1   MES                    23031 non-null  int64
2   CAPITULO               23031 non-null  int64
3   SUBCAPITULO            23031 non-null  int64
4   RUC                    23031 non-null  object
5   RAZON_SOCIAL            23031 non-null  object
6   RAZON_SOCIAL_DIRECCION  22905 non-null  object
7   REFRENDO                23031 non-null  object
8   NUME_SERIE             23031 non-null  int64
9   TIPO_AFORO              23030 non-null  object
10  COD_REGIMEN             23031 non-null  int64
11  REGIMEN                 23031 non-null  object
12  DISTRITO                 23031 non-null  object
13  AGENTE_AFIANZADO        23031 non-null  object
14  REMITENTE               22849 non-null  object
15  NOTIFY                  22874 non-null  object
```

Se obtuvo estadísticas descriptivas del del conjunto de datos como la mediana, media y desviación estándar.

Figura 12

Código visualización estadísticas del dataframe

3 minutos ago (<1s) 7

`df.describe()`

	AÑO	MES	CAPITULO	SUBCAPITULO	NUME_SERIE	COD_REGIMEN	PARTIDA	TNAN	TASA_ADVALOREM	TIPO_CARGA	UNIDADES	KILOS_NETO	FOB	FLETE
count	23031.000000	23031.000000	23031.0	23031.0	23031.000000	23031.000000	2.303100e+04	23031.0	0.0	0.0	23031.000000	23031.000000	2.303100e+04	23031.000000
mean	2019.744258	6.506969	85.0	8517.0	28.945986	27.361426	8.517627e+09	0.0	NaN	NaN	171.791810	120.630183	7.072750e+03	171.620165
std	2.336998	3.366953	0.0	0.0	59.389802	32.995556	1.898093e+04	0.0	NaN	NaN	909.691401	665.575242	4.461433e+04	1053.200720
min	2016.000000	1.000000	85.0	8517.0	0.000000	10.000000	8.517110e+09	0.0	NaN	NaN	1.000000	0.000000	1.000000e+02	0.000000
25%	2018.000000	4.000000	85.0	8517.0	1.000000	10.000000	8.517622e+09	0.0	NaN	NaN	2.000000	3.201000	3.224750e+02	4.060000
50%	2020.000000	7.000000	85.0	8517.0	6.000000	10.000000	8.517622e+09	0.0	NaN	NaN	8.000000	12.303000	1.060550e+03	17.550000
75%	2022.000000	9.000000	85.0	8517.0	28.000000	10.000000	8.517629e+09	0.0	NaN	NaN	50.000000	50.343000	3.600000e+03	70.986000
max	2024.000000	12.000000	85.0	8517.0	1507.000000	96.000000	8.517790e+09	0.0	NaN	NaN	50000.000000	39349.638000	2.180394e+06	53600.000000

3.4 Tratamiento de los datos

El tratamiento de datos es clave para asegurar que los datos sean adecuados para el análisis y modelado. En esta fase se aplicaron diferentes acciones detalladas a continuación:

Se identificó y eliminó registros duplicados en el conjunto de datos.

Figura 13

Código eliminación duplicados del dataframe

```

#Duplicados
duplicate_count = df.duplicated().sum()
print(f"Se encontraron {duplicate_count} filas duplicadas.")

Se encontraron 11 filas duplicadas.

df_duplicados=df.drop_duplicates(keep=False)
duplicate_count = df_duplicados.duplicated().sum()
print(f"Se encontraron {duplicate_count} filas duplicadas.")

Se encontraron 0 filas duplicadas.

```

Se filtró las ventas de un Remitente para el posterior análisis.

Figura 14

Código filtrado ventas del remitente seleccionado

```

#Filtrar información de remitente TPLINK
remitentes_unicos = df_duplicados['REMITENTE'].unique()
remitentes_unicos= np.sort(remitentes_unicos.astype(str))
palabras_clave_remitente = ['TP LINK', 'TP-LINK', 'TPLINK']
expresion_regular_remitente = '|'.join(palabras_clave_remitente)
df_filtro_router = df[df['REMITENTE'].str.contains(expresion_regular_remitente, case=False, na=False)]

```

Se filtró las marcas vendidas por el Remitente, excluyendo posibles datos incorrectos.

Figura 15*Código filtrado marcas correctas del remitente*

```

# Filtrar las marcas vendidas por TPLINK
marcas_unicas = df_filtro_router["MARCA_COMERCIAL"].unique()
marcas_unicas
marcas_correctas = ['TP-LINK', 'TP-LINK', 'TPLINK', 'T P LINK', 'TP-Link']
expresion_regular_marcas_correctas = '|'.join(marcas_correctas)
df_filtro_router["MARCA_COMERCIAL"] = df_filtro_router["MARCA_COMERCIAL"].fillna('')
df_filtro_router_marca = df_filtro_router[df_filtro_router["MARCA_COMERCIAL"].str.contains(expresion_regular_marcas_correctas, regex=True)]
df_filtro_router_marca["MARCA_COMERCIAL"].unique()

/root/.ipykernel/2610/command-2647293233563141-3702256701:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df_filtro_router["MARCA_COMERCIAL"] = df_filtro_router["MARCA_COMERCIAL"].fillna('')
array(['TP-LINK', 'TP LINK', 'TPLINK', 'TP-Link', 'T P LINK'],
      dtype=object)

```

Se realizó un primer filtro de columnas, necesarias para curar los nombres de los modelos de routers y eliminar los incorrectos.

Figura 16*Código filtrado inicial de columnas necesarias para curación datos*

```

df_filtro_columnas = df_filtro_router_marca[['RAZON_SOCIAL', 'FECH_INGRESO', 'UNIDADES', 'CIF', 'CARACTERISTICA', 'PRODUCTO', 'MODELO_MERCADERIA']]
df_filtro_columnas.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1378 entries, 7 to 23010
Data columns (total 7 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   RAZON_SOCIAL           1378 non-null   object
 1   FECH_INGRESO           1378 non-null   datetime64[ns]
 2   UNIDADES               1378 non-null   float64
 3   CIF                   1378 non-null   float64
 4   CARACTERISTICA         1378 non-null   object
 5   PRODUCTO              1378 non-null   object
 6   MODELO_MERCADERIA      1378 non-null   object
dtypes: datetime64[ns](1), float64(2), object(4)
memory usage: 86.1+ KB

```

Se cargó otra tabla con un diccionario de modelos de routers para estandarizar los nombres de los modelos y puedan agruparse correctamente en el análisis.

Figura 17*Código lectura de tabla delta Bronze.AST_Diccionario con diccionario de modelos*

```

#Crear diccionario, para estandarizar los nombres modelos y eliminar modelos que no son routers
spark_df_dicc = spark.read.format("delta").table("Bronze.AST_Diccionario")
df_dicc = spark_df_dicc.toPandas()

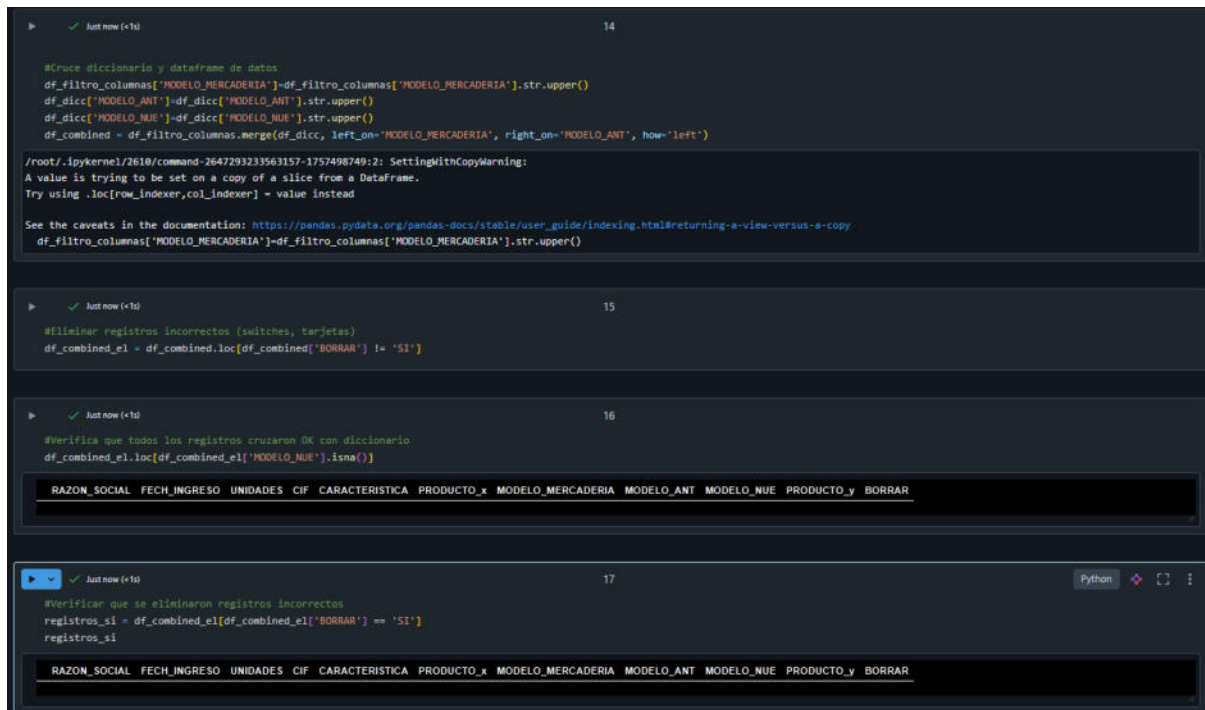
(2) Spark jobs
spark_df_dicc: pyspark.sql.dataframe.DataFrame = [MODELO_ANT: string, MODELO_NUE: string ... 2 more fields]

```

Se realizó un cruce de datos entre el diccionario y las ventas para generar una nueva columna con los nombres de los modelos de routers estandarizados, una vez realizado también se eliminaron productos que no son routers.

Figura 18

Código estandarización y curación de modelos routers con diccionario



```

#Cruce diccionario y dataframe de datos
df_filtro_columnas["MODELO_MERCADERIA"] = df_filtro_columnas["MODELO_MERCADERIA"].str.upper()
df_dicc["MODELO_ANT"] = df_dicc["MODELO_ANT"].str.upper()
df_dicc["MODELO_NUE"] = df_dicc["MODELO_NUE"].str.upper()
df_combined = df_filtro_columnas.merge(df_dicc, left_on="MODELO_MERCADERIA", right_on="MODELO_ANT", how="left")

/root/.ipykernel/2610/command-2647293233563157-1757498749:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df_filtro_columnas["MODELO_MERCADERIA"] = df_filtro_columnas["MODELO_MERCADERIA"].str.upper()

#Eliminar registros incorrectos (switches, tarjetas)
df_combined_el = df_combined.loc[df_combined["BORRAR"] != "SI"]

#Verifica que todos los registros cruzaron OK con diccionario
df_combined_el.loc[df_combined_el["MODELO_NUE"].isna()]

RAZON_SOCIAL  FECH_INGRESO  UNIDADES  CIF  CARACTERISTICA  PRODUCTO_x  MODELO_MERCADERIA  MODELO_ANT  MODELO_NUE  PRODUCTO_y  BORRAR

```

```

#Verificar que se eliminaron registros incorrectos
registros_si = df_combined_el[df_combined_el["BORRAR"] == "SI"]
registros_si

RAZON_SOCIAL  FECH_INGRESO  UNIDADES  CIF  CARACTERISTICA  PRODUCTO_x  MODELO_MERCADERIA  MODELO_ANT  MODELO_NUE  PRODUCTO_y  BORRAR

```

Una vez finalizada la limpieza se generó el dataframe más pequeño con las columnas necesarias para los modelos de series temporales.

Figura 19

Código generación dataframe para series temporales



```

# Se genera dataframe más pequeño solo columnas necesarias para series
df_series = df_combined_el[["FECH_INGRESO", "MODELO_NUE", "UNIDADES"]]

# Se cambia nombre columnas
df_series.columns = ["FechaVenta", "Producto", "Unidades"]

df_series.head()

FechaVenta  Producto  Unidades
0  2022-01-04  XC220-G3V    700.0
1  2022-01-04  TL-WR940N    800.0
2  2022-01-04  ARCHER C50   2000.0
3  2022-01-04  EC220-G5    1500.0
4  2022-01-04  TL-WR840N    2000.0

```


Debido que los modelos de series temporales necesitan únicamente como columnas de entrada la Fecha y los Valores, se generó un dataframe transpuesto para ubicar las unidades vendidas por producto a nivel de columna.

Figura 20

Código generación dataframe tranpuesto para unidades por producto



```
# Transponer el DataFrame al formato para series
df_series_traspuesto = df_series.pivot_table(index='FechaVenta', columns='Producto', values='Unidades', aggfunc='sum')
# Reiniciar el índice para que 'Fecha' vuelva a ser una columna
df_series_traspuesto.reset_index(inplace=True)
df_series_traspuesto.head()
```

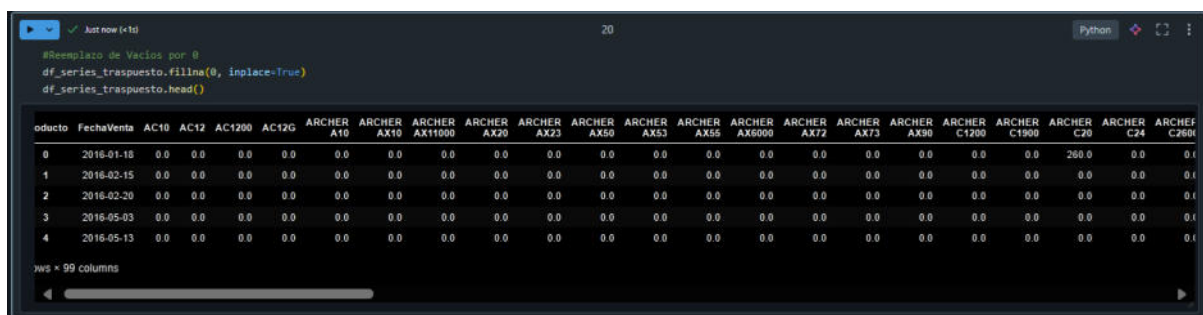
Producto	FechaVenta	AC10	AC12	AC1200	AC12G	ARCHER A10	ARCHER AX10	ARCHER AX11000	ARCHER AX20	ARCHER AX23	ARCHER AX50	ARCHER AX53	ARCHER AX55	ARCHER AX6000	ARCHER AX72	ARCHER AX73	ARCHER AX90	ARCHER C1200	ARCHER C1900	ARCHER C20	ARCHER C24	ARCHER C2600
0	2016-01-16	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	260.0	NaN	NaN
1	2016-02-15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	2016-02-20	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	2016-05-03	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	2016-05-13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows x 99 columns

Como último paso, se rellenó las fechas que no poseen unidades vendidas cada producto con valor de 0 y se almacenó el conjunto de datos procesado en la capa Silver.

Figura 21

Código reemplazo de valores Nan



```
#Reemplazo de Valores por 0
df_series_traspuesto.fillna(0, inplace=True)
df_series_traspuesto.head()
```

Producto	FechaVenta	AC10	AC12	AC1200	AC12G	ARCHER A10	ARCHER AX10	ARCHER AX11000	ARCHER AX20	ARCHER AX23	ARCHER AX50	ARCHER AX53	ARCHER AX55	ARCHER AX6000	ARCHER AX72	ARCHER AX73	ARCHER AX90	ARCHER C1200	ARCHER C1900	ARCHER C20	ARCHER C24	ARCHER C2600
0	2016-01-16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	260.0	0.0	0.0
1	2016-02-15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	2016-02-20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	2016-05-03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	2016-05-13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows x 99 columns

Figura 22

Código generación tabla delta Silver.AST_VentasProcesadas

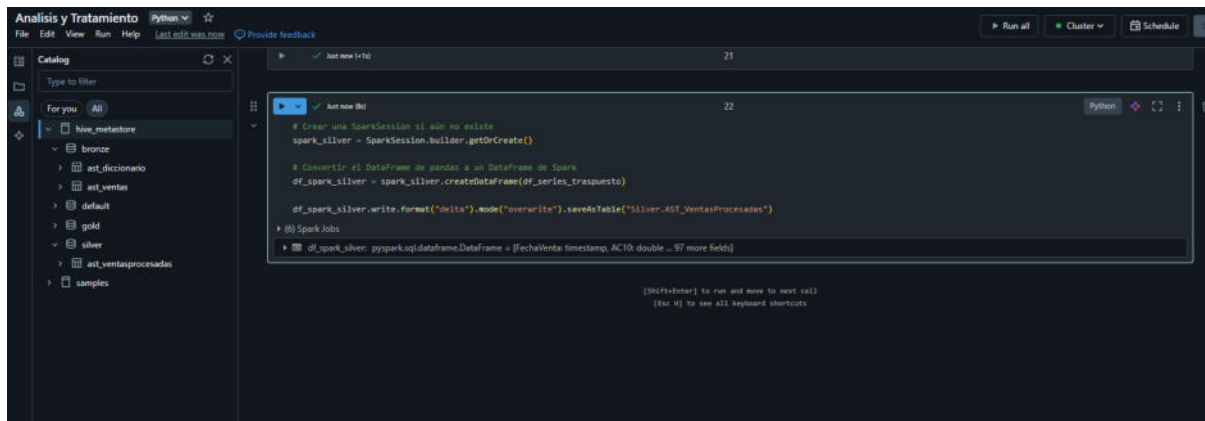


Figura 23

Código verificación datos tabla delta Silver.AST_VentasProcesadas

The screenshot shows a Databricks notebook interface with a SQL query executed. The query is:

```
select * from Silver.AST_VentasProcesadas
```

The result is displayed as a table with 142 rows and 11 columns. The columns are: FechaVenta, 1.2 AC10, 1.2 AC12, 1.2 AC1200, 1.2 AC12G, 1.2 ARCHER_A10, 1.2 ARCHER_AX10, 1.2 ARCHER_AX1000, 1.2 ARCHER_AX20, and 1.2 AF. The first 14 rows are shown in the screenshot.

	FechaVenta	1.2 AC10	1.2 AC12	1.2 AC1200	1.2 AC12G	1.2 ARCHER_A10	1.2 ARCHER_AX10	1.2 ARCHER_AX1000	1.2 ARCHER_AX20	1.2 AF
1	2016-01-18T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
2	2016-02-15T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
3	2016-02-20T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
4	2016-05-03T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
5	2016-05-13T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
6	2016-07-25T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
7	2016-08-18T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
8	2016-08-19T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
9	2016-08-30T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
10	2016-09-09T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
11	2016-10-21T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
12	2016-12-07T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
13	2016-12-16T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0
14	2016-12-27T00:00:00.000+00:00	0	0	0	0	0	0	0	0	0

At the bottom of the table, it says '142 rows | 1.90 seconds runtime' and 'Refreshed now'. A note at the very bottom states: 'This result is stored as _sqldf and can be used in other Python cells.'

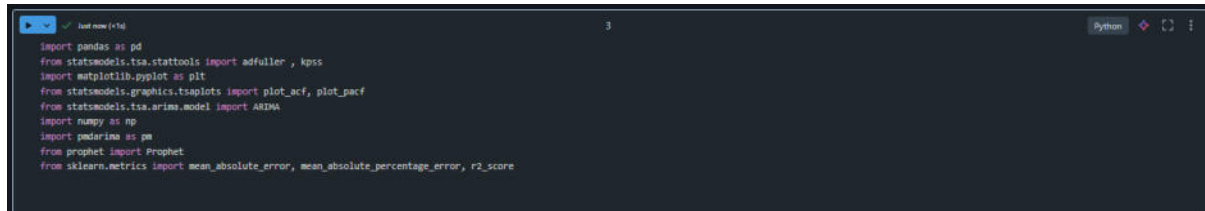
3.5 Aplicación de modelos

La fase de modelos de aprendizaje es el momento en que los patrones y relaciones entre los datos se hacen visibles. Los datos que han sido tratados, son formalizados y utilizados para realizar predicciones, clasificaciones, o tomar decisiones automatizadas. En esta fase de aplicaron diferentes acciones detalladas a continuación:

Primero se realizó la importación de las librerías necesarias para la aplicación de los modelos.

Figura 24

Código librerías necesarias para aplicación de modelos series temporales



```
import pandas as pd
from statsmodels.tsa.stattools import adfuller, kpss
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima.model import ARIMA
import numpy as np
import paderina as pa
from prophet import Prophet
from sklearn.metrics import mean_absolute_error, mean_absolute_percentage_error, r2_score
```

Se aplicó la prueba ADF para determinar si la serie temporal es estacionaria o no, obteniendo como resultado las modelos de routers que no cumplen las condiciones estacionarias.

Figura 25

Código aplicación prueba ADF

```

# Lista para almacenar los resultados
results = []

# Función para realizar la prueba ADF
def adf_test(series, column_name):
    # Verificar si la serie tiene valores constantes
    if series.max() == series.min():
        results.append({'Columna': column_name, 'Estacionaria': 'Error: La serie temporal es constante'})
        return
    try:
        result = adfuller(series)
        adf_statistic = result[0]
        p_value = result[1]
        critical_values = result[4]
        # Determinar si la serie es estacionaria
        if p_value < 0.05:
            estacionaria = 'Si'
        else:
            estacionaria = 'No'
        # Almacenar los resultados
        results.append({
            'Columna': column_name,
            'Estacionaria': estacionaria,
            'Estadístico ADF': adf_statistic,
            'Valor p': p_value,
            'Valores críticos': critical_values
        })
    except Exception as e:
        results.append({'Columna': column_name, 'Estacionaria': f'Error: {e}'})

# Iterar sobre las columnas del DataFrame
for col in df_simplificado.columns:
    adf_test(df_simplificado[col], col)

# Convertir la lista de resultados en un DataFrame
df_resultados = pd.DataFrame(results)

# Mostrar el DataFrame de resultados
print(df_resultados)

```

Figura 26

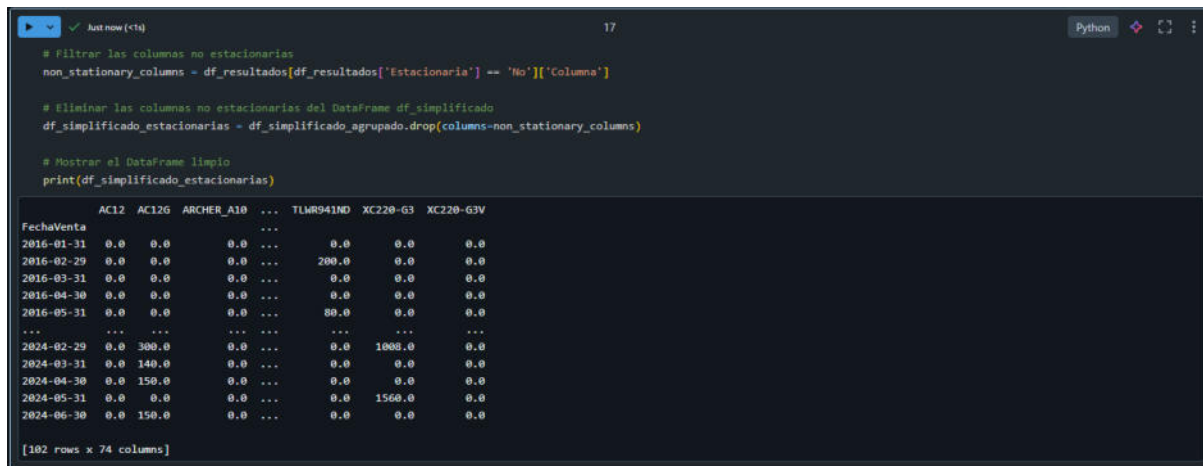
Código visualización resultados prueba ADF

df_resultados.head()

	Columna	Estacionaria	Estadístico ADF	Valor p	Valores críticos
0	AC10	No	-1.819976	3.705901e-01	[1%: -3.5019123847798657, 5%: -2.892815255...
1	AC12	Si	-3.001362	3.476739e-02	[1%: -3.4996365338407074, 5%: -2.891830773...
2	AC1200	No	-1.056306	7.321785e-01	[1%: -3.497501033, 5%: -2.89090644, 10%:...
3	AC12G	Si	-8.949459	8.804069e-15	[1%: -3.4968181663902103, 5%: -2.890610751...
4	ARCHER_A10	Si	-10.100807	1.060763e-17	[1%: -3.4968181663902103, 5%: -2.890610751...

Figura 27

Código filtro columnas no estacionarias



```

Just now (11s) 17 Python
# Filtrar las columnas no estacionarias
non_stationary_columns = df_resultados[df_resultados['Estacionaria'] == 'No']['Columna']

# Eliminar las columnas no estacionarias del DataFrame df_simplificado
df_simplificado_estacionarias = df_simplificado_agrupado.drop(columns=non_stationary_columns)

# Mostrar el DataFrame limpio
print(df_simplificado_estacionarias)

```

FechaVenta	AC12	AC12G	ARCHER_A10	...	TLMR941ND	XC220-G3	XC220-G3V
2016-01-31	0.0	0.0	0.0	...	0.0	0.0	0.0
2016-02-29	0.0	0.0	0.0	...	280.0	0.0	0.0
2016-03-31	0.0	0.0	0.0	...	0.0	0.0	0.0
2016-04-30	0.0	0.0	0.0	...	0.0	0.0	0.0
2016-05-31	0.0	0.0	0.0	...	80.0	0.0	0.0
...	...	...	...	...	...	...	...
2024-02-29	0.0	300.0	0.0	...	0.0	1808.0	0.0
2024-03-31	0.0	140.0	0.0	...	0.0	0.0	0.0
2024-04-30	0.0	150.0	0.0	...	0.0	0.0	0.0
2024-05-31	0.0	0.0	0.0	...	0.0	1560.0	0.0
2024-06-30	0.0	150.0	0.0	...	0.0	0.0	0.0

[102 rows x 74 columns]

Previo al entrenamiento es necesario dividir el conjunto de datos en entrenamiento y test.

Figura 28

Código división conjuntos entrenamiento y pruebas



```

Just now (11s) 18 Python
#Separa los dataframe de entrenamiento y test
fecha_corte = '2022-10-31'
df_train=df_simplificado_estacionarias[df_simplificado_estacionarias.index <= fecha_corte]
df_test=df_simplificado_estacionarias[df_simplificado_estacionarias.index > fecha_corte]

```

Se aplicó el modelo SARIMA ajustado por cada modelo router obteniendo su predicción

Figura 29

Código aplicación modelo SARIMA

```

# Crear un diccionario para almacenar los modelos, predicciones y métricas
predicciones_sarima_ajustado = {}

metricas_sarima = pd.DataFrame(columns=['Producto', 'SAR-MAE', 'SAR-RMSE', 'SAR-MAPE', 'SAR-R2'])

# Iterar sobre las columnas del DataFrame
for col in df_train.columns:

    # Ajustar el modelo usando auto_arima para encontrar los mejores parámetros
    model_auto_arima = pw.auto_arima(df_train[col],
                                     seasonal=False,
                                     n_num_meses_predecir,
                                     trace=True,
                                     stepwise=True,
                                     suppress_warnings=True)

    # Ajustar el modelo ARIMA con los parámetros encontrados
    p, d, q = model_auto_arima.order
    P, D, Q, s = model_auto_arima.seasonal_order

    model_sarima = SARIMAX(df_train[col], order=(p, d, q), seasonal_order=(P, D, Q, s))
    model_sarima_fit = model_sarima.fit()

    # Hacer predicciones para el número de meses
    forecast = model_sarima_fit.forecast(steps=num_meses_predecir)

    # Almacenar las predicciones en el diccionario
    predicciones_sarima_ajustado[col] = forecast

    # Calcular las métricas
    actual = df_test[col].values[num_meses_predecir:] # Asegúrate de que tengas suficientes datos en df_test
    mae = mean_absolute_error(actual, forecast)
    rmse = np.sqrt(mean_squared_error(actual, forecast))
    mape = np.mean(np.abs(actual - forecast) / actual) * 100
    r2 = 1 - (np.sum((actual - forecast) ** 2) / np.sum((actual - np.mean(actual)) ** 2))

    # Crear un DataFrame temporal con las métricas para este producto
    metricas_tempora = pd.DataFrame({
        'Producto': [col],
        'SAR-MAE': [mae],
        'SAR-RMSE': [rmse],
        'SAR-MAPE': [mape],
        'SAR-R2': [r2]
    })

    # Usar pd.concat() para agregar las métricas al DataFrame general
    metricas_sarima = pd.concat([metricas_sarima, metricas_tempora], ignore_index=True)

```

Figura 30

Código visualización predicciones modelo SARIMA

FechaVenta	AC12	AC120	ARCHER_A10	ARCHER_AX10	ARCHER_AX20	ARCHER_AX50	ARCHER_AX55	ARCHER_AX6000	ARCHER_AX72	ARCHER_C1200	ARCHER_C1900	ARCHER_C20	ARCHER_C2600	ARCHER_C315	ARCHER_C3150	ARC
0	2022-11-30	0.0	24.958733	0.0	11.996854	3.483415	6.087485	-1.601811	0.0	0.0	0.0	868.880858	0.0	0.0	0.0	
1	2022-12-31	0.0	29.533499	0.0	12.716301	3.268624	5.528987	0.840363	0.0	0.0	0.0	803.612187	0.0	0.0	0.0	
2	2023-01-31	0.0	30.372023	0.0	10.802523	4.400180	7.363061	0.251547	0.0	0.0	0.0	278.345026	0.0	0.0	0.0	
3	2023-02-28	0.0	30.525718	0.0	15.828598	4.261451	7.104232	-0.537095	0.0	0.0	0.0	341.446245	0.0	0.0	0.0	
4	2023-03-31	0.0	30.553860	0.0	16.828880	4.833279	7.808038	0.149133	0.0	0.0	0.0	1903.937979	0.0	0.0	0.0	
5	2023-04-30	0.0	30.559053	0.0	15.296230	4.585815	7.497794	0.194867	0.0	0.0	0.0	900.031909	0.0	0.0	0.0	
6	2023-05-31	0.0	30.560000	0.0	17.273300	4.689346	7.846781	-0.145720	0.0	0.0	0.0	1022.595643	0.0	0.0	0.0	
7	2023-06-30	0.0	30.560173	0.0	18.022448	4.680070	7.804769	-0.002192	0.0	0.0	0.0	813.852344	0.0	0.0	0.0	
8	2023-07-31	0.0	30.560205	0.0	17.990350	4.701534	7.849537	0.085400	0.0	0.0	0.0	560.537322	0.0	0.0	0.0	
9	2023-08-31	0.0	30.560211	0.0	17.783338	4.689578	7.834432	-0.030717	0.0	0.0	0.0	565.534195	0.0	0.0	0.0	
10	2023-09-30	0.0	30.560212	0.0	18.297859	4.703627	7.848071	-0.013993	0.0	0.0	0.0	1067.335378	0.0	0.0	0.0	
11	2023-10-31	0.0	30.560212	0.0	17.812486	4.686822	7.842825	0.020342	0.0	0.0	0.0	862.786157	0.0	0.0	0.0	
12	2023-11-30	0.0	30.560212	0.0	18.007033	4.703723	7.847029	-0.003577	0.0	0.0	0.0	978.349159	0.0	0.0	0.0	
13	2023-12-31	0.0	30.560212	0.0	18.319530	4.701916	7.845247	-0.007200	0.0	0.0	0.0	876.107786	0.0	0.0	0.0	
14	2024-01-31	0.0	30.560212	0.0	18.065311	4.703590	7.846556	0.005112	0.0	0.0	0.0	727.772621	0.0	0.0	0.0	
15	2024-02-29	0.0	30.560212	0.0	18.118536	4.702887	7.845958	0.000702	0.0	0.0	0.0	703.968207	0.0	0.0	0.0	
16	2024-03-31	0.0	30.560212	0.0	18.291687	4.703460	7.846370	-0.002597	0.0	0.0	0.0	886.747889	0.0	0.0	0.0	
17	2024-04-30	0.0	30.560212	0.0	18.200195	4.703205	7.846172	0.000948	0.0	0.0	0.0	832.790761	0.0	0.0	0.0	
18	2024-05-31	0.0	30.560212	0.0	18.177973	4.703405	7.845302	0.000967	0.0	0.0	0.0	913.382894	0.0	0.0	0.0	
19	2024-06-30	0.0	30.560212	0.0	18.286725	4.703311	7.846237	-0.000752	0.0	0.0	0.0	878.830049	0.0	0.0	0.0	

Se aplicó el modelo Prophet por cada modelo router obteniendo su predicción

Figura 31

Código aplicación modelo Prophet

```

# Crear un diccionario para almacenar los modelos y las predicciones
predicciones_prophet = pd.DataFrame()

# Crear un DataFrame para almacenar las métricas de evaluación
metricas_prophet = pd.DataFrame(columns=['Producto', 'PRO-MAE', 'PRO-RMSE', 'PRO-MAPE', 'PRO-R2'])

# Ajustar el modelo Prophet para cada columna y hacer predicciones
for col in df_train.columns:
    # Crear un DataFrame compatible con Prophet
    df_prophet = df_train[col].reset_index()
    df_prophet = df_prophet.rename(columns={'fecha_venta': 'ds', 'col': 'y'})

    # Ajustar el modelo Prophet
    model_prophet = Prophet()
    model_prophet.add_seasonality(
        name='monthly',
        period=30.5,
        fourier_order=5 # ajusta este valor según la complejidad de tus datos
    )
    model_prophet_fit = model_prophet.fit(df_prophet)

    # Hacer predicciones para el número de meses
    future = model_prophet_fit.make_future_dataframe(periods=(num_meses_predicir-1), freq='M')
    forecast = model_prophet_fit.predict(future)
    # Reemplazar valores negativos por cero
    forecast['yhat'] = forecast['yhat'].apply(lambda x: max(x, 0))

    # Almacenar las predicciones en el diccionario
    predicciones_prophet['col'] = forecast['yhat']

    # Extraer las predicciones solo para el periodo de test
    pred_test = forecast['yhat'][-num_meses_predicir:]

    # Calcular las métricas de evaluación
    mae = mean_absolute_error(df_test[col], pred_test)
    rmse = np.sqrt(mean_squared_error(df_test[col], pred_test))
    mape = np.mean(np.abs(df_test[col] - pred_test) / df_test[col]) * 100
    r2 = r2_score(df_test[col], pred_test)

    # Crear un DataFrame temporal con las métricas para este producto
    metricas_temp = pd.DataFrame({
        'Producto': [col],
        'PRO-MAE': [mae],
        'PRO-RMSE': [rmse],
        'PRO-MAPE': [mape],
        'PRO-R2': [r2]
    })

    # Usar pd.concat() para agregar las métricas al DataFrame general
    metricas_prophet = pd.concat([metricas_prophet, metricas_temp], ignore_index=True)

```

Figura 32

Código visualización predicciones modelo Prophet

✓ Just now (1:16)

predicciones_prophet_filtrado_df

37

Python

	FechaVenta	AC12	AC120	ARCHER_A10	ARCHER_AX10	ARCHER_AX20	ARCHER_AX50	ARCHER_AX55	ARCHER_AX6000	ARCHER_AX72	ARCHER_C1200	ARCHER_C1900	ARCHER_C20	ARCHER_C200	ARCHER_C315	ARCHER_C3150
0	2022-11-30	11.755041	73.005400	0.000000	51.126611	12.272065	15.609187	4.922066	2.092067	12.717121	0.000000	0.0	958.845878	0.000000	0.0000	0.884035
1	2023-12-31	0.000000	34.350341	0.000000	15.420155	7.040021	0.000000	0.000000	0.448688	5.580534	0.000000	0.0	653.503688	0.000000	0.0000	2.450747
2	2023-01-31	1.517466	17.845849	1.759621	6.758277	18.492834	0.000000	7.189305	6.001440	0.616704	0.000000	0.0	491.806206	0.000000	0.0000	2.501808
3	2023-02-28	0.000000	208.858397	0.000000	324.439301	0.000000	269.552298	238.303478	26.913470	2.841499	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
4	2023-03-31	0.000000	170.743835	0.419268	152.010893	0.000000	121.818442	104.207906	12.117919	9.987353	0.000000	0.0	159.898490	0.000000	0.0000	0.000000
5	2023-04-30	0.000000	365.289002	0.520406	278.632719	0.000000	269.423679	212.810906	22.513458	0.000000	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
6	2023-05-31	0.000000	189.531936	0.000000	185.812780	0.000000	121.943596	105.209137	13.980984	35.198024	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
7	2023-06-30	0.000000	228.548962	1.147372	301.899615	0.000000	245.410995	225.455775	22.612164	0.000000	0.000000	0.0	0.000000	0.000000	0.0000	16.908532
8	2023-07-31	2.448727	106.320844	1.233128	157.006791	0.000000	124.188090	104.898648	12.174401	11.925889	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
9	2023-08-31	6.969245	104.887350	0.289449	81.107967	0.000000	49.581824	38.278834	5.002845	17.044970	0.000000	0.0	1027.854773	1.130008	0.0000	0.000000
10	2023-09-30	6.743571	185.782870	0.000000	156.990082	0.000000	125.803704	104.062491	12.216263	12.861555	10.673882	0.0	191.489501	0.000000	8.4274	0.000000
11	2023-10-31	0.000000	125.825912	0.000000	79.725714	0.000000	50.463200	37.357022	5.047588	17.384488	0.000000	0.0	746.345344	0.144713	0.0000	0.359316
12	2023-11-30	14.765520	227.742063	0.000000	177.824886	0.237946	132.613304	104.831369	13.345057	15.362896	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
13	2023-12-31	0.000000	89.504020	0.000000	81.263551	0.000000	39.750350	35.987507	5.181177	18.002899	0.000000	0.0	887.124874	0.000000	0.0000	0.000000
14	2024-01-31	3.807378	72.805570	2.200010	35.723655	13.517278	11.706644	12.093825	9.988408	11.867678	0.000000	0.0	1167.276884	0.000000	0.0000	1.038893
15	2024-02-29	7.038797	312.107529	0.828480	301.262825	0.000000	265.189897	226.095156	25.095892	0.000000	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
16	2024-03-31	21.619988	223.898821	0.884304	166.526687	0.000000	131.725487	106.400154	12.888497	15.782119	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
17	2024-04-30	7.654042	337.811182	0.000000	313.244183	0.000000	267.748230	223.721306	24.820993	0.000000	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
18	2024-05-31	55.780299	238.382109	10.070143	166.525954	0.000000	131.553783	105.806866	13.001912	15.614361	0.000000	0.0	0.000000	0.000000	0.0000	0.000000
19	2024-06-30	3.229686	315.285960	0.000000	295.780642	0.000000	263.321527	224.258408	24.812584	0.000000	0.000000	0.0	0.000000	0.000000	0.0000	0.000000

Como último paso, se cruzo los 2 dataframes de las predicciones y se almacenó el conjunto de datos en la capa Gold.

Figura 33

Código unificación predicciones modelos SARIMA y PROPHET

```

# Agrega una columna que indique el nombre del DataFrame
predicciones_sarima_df['Origen'] = 'SARIMA'
predicciones_prophet_filtrado_df['Origen'] = 'PROPHET'

/root/.ipykernel/914/command-4079658072762835-3523836063:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
predicciones_prophet_filtrado_df['Origen'] = 'PROPHET'

# Alinear las columnas para asegurarse de que coincidan
predicciones_sarima_df = predicciones_sarima_df[predicciones_prophet_filtrado_df.columns]

# Unir los DataFrames
df_unido = pd.concat([predicciones_sarima_df, predicciones_prophet_filtrado_df], ignore_index=True)

```

Figura 34

Código generación tabla delta Gold.AST_VentasPredicciones

```

# Crear una SparkSession si aún no existe
spark_gold = SparkSession.builder.getOrCreate()

# Convertir el DataFrame de pandas a un DataFrame de Spark
df_spark_silver = spark_gold.createDataFrame(df_unido)

df_spark_silver.write.format("delta").mode("overwrite").saveAsTable("Gold.AST_VentasPrediccionesArimaProphet")

# (6) Spark Job
df_spark_silver: pyspark.sql.dataframe.DataFrame = [FechaVenta: timestamp, AC12: double ... 74 more fields]

```

Figura 35

Código verificación datos tabla delta Gold.AST_VentasPredicciones

SQL

```

select * from Gold.AST_VentasPredicciones

```

(2) Spark Jobs

_sqlid: pyspark.sql.dataframe.DataFrame = [FechaVenta: timestamp, AC1200: double ... 75 more fields]

	1.2 TL-WRB40N	1.2 TL-WRB41HP	1.2 TL-WRB41N	1.2 TL-WRB41ND	1.2 TL-WRB42N	1.2 TL-WRB43N	1.2 TL-WRB50N	1.2 TL-WRB41ND	1.2 TL-WRB41ND	1.2 KC220-GJV	Origen
25	1276.328257803245	58.54833911356052	187.1681552587007	0.6648791218563541	0.7079595989364178	38.31862500968214	195.86751527162488	0.5309684493844395	2.3247799380002087	62.36392218316157	ARIMA
26	1276.328257803245	57.96962788762738	187.1681552587007	0.6775765181968723	0.7079595989364178	38.31862500968214	195.86751527162488	0.5309684493844395	2.324779924898371	62.496347258711354	ARIMA
27	1276.328257803245	58.4760590763083	187.1681552587007	0.687238610010665	0.7079595989364178	38.31862500968214	195.86751527162488	0.5309684493844395	2.3247800305580346	62.609028096543405	ARIMA
28	1276.328257803245	58.03079373210356	187.1681552587007	0.6919466225585412	0.7079595989364178	38.31862500968214	195.86751527162488	0.5309684493844395	2.324780304373894	62.70490848739384	ARIMA
29	1276.328257803245	58.4207075885018	187.1681552587007	0.6916704382574151	0.7079595989364178	38.31862500968214	195.86751527162488	0.5309684493844395	2.3247802039738945	62.78649334549708	ARIMA
30	993.3660318098007	202.16825535170387	798.6106927228542	18.65464908896488	0.9944550077013861	-1.4338597788323006	-249.95867801130021	0.6667439944308597	2.640472559645733	-149.824969716287...	PROPHET
31	1923.8185511067933	139.573434330035	422.40115041172055	96.00811330092283	1.3439237780943187	28.086819787620588	-180.1530035447538	1.6014548622571851	3.444503825885185	150.18439210563514	PROPHET
32	1072.1049229978878	12.865834801718767	40.454885497533945	-0.0245701157608	-0.0628995996093707	-14.46875082959535	-71.52350488097231	-0.1739674589781596	200.1368838891554	116.2539596645706	PROPHET
33	341.1260924354551	143.47700624466234	242.43257102181207	26.7887964330856	1.4468099654947246	48.484999780971485	-157.837879764773	12.86730409748365	6.817574536450941	-84.11539679815127	PROPHET
34	73.8232848289174	109.46204526516519	37.98472709605289	43.04296620960764	4.8449134970002365	-6.2253082825653365	-102.94613707867353	4.327320529003874	14.52863124634123	-3.021107052243785	PROPHET
35	1035.2120059383771	153.8463337009715	430.20636238629957	25.34791038207142	1.5855802215321213	13.892942485241533	-62.16074303859118	1.1072835376900546	2.360248322322274	-54.03410053932404	PROPHET
36	2121.1598767378664	228.0530429804392	375.94251235042896	4.9129801958048435	0.91238169362573	11.466684748016146	-101.37630223550613	1.967443351550554	1.4081594832986688	-9.64227518132941	PROPHET
37	1917.7083659860388	132.59022023106038	222.49544428839118	25.43520085678155	3.469114119526691	19.900547977165386	-101.98893176041866	1.8020542922562	3.719586666411804	-4.044800187603185	PROPHET
38	2122.036048456816	274.97910213886786	439.11108295382416	-3.8735301246358276	-0.5893842892868517	5.391260340994492	-258.72165416000104	1.5389949234407352	-0.4141728064644923	277.4068375910364	PROPHET

200 rows | 2.68 seconds runtime

This result is stored as `_sqlid` and can be used in other `Python` cells.

CAPITULO IV

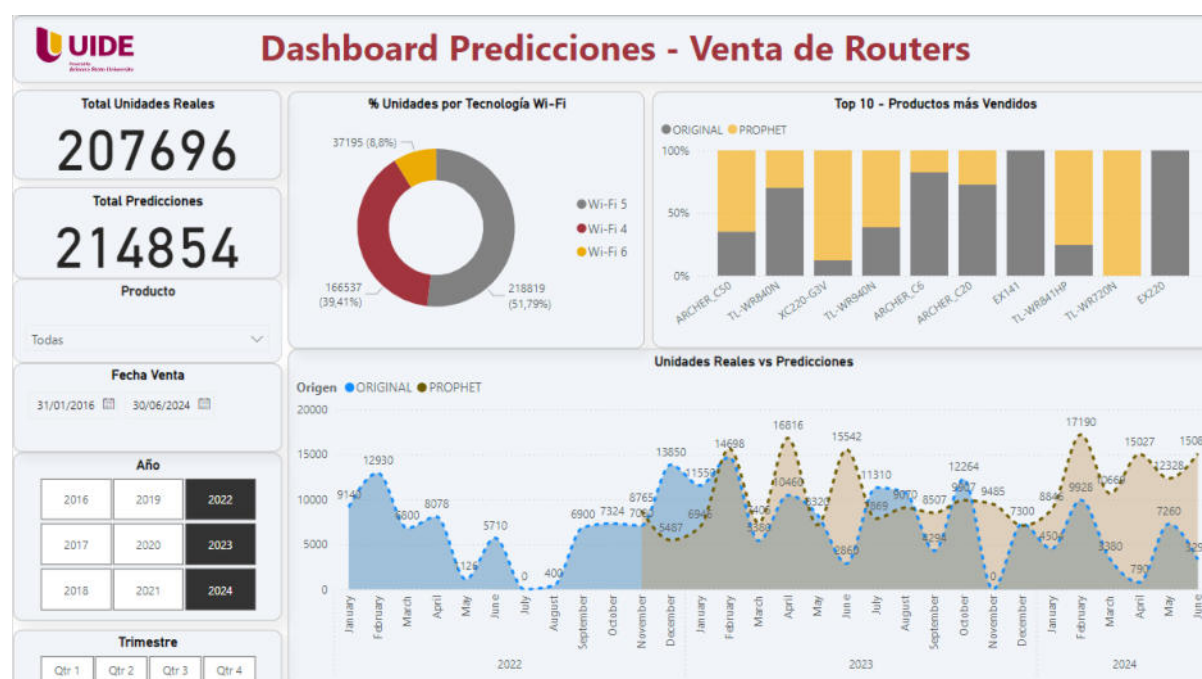
RESULTADOS

4.1 Informe de Visualización (Dashboard)

Se propuso el diseño de un tablero utilizando la herramienta Power BI como se detalló en la arquitectura planteada, que entregará un valor agregado ya que transmite diferentes indicadores clave que permiten evaluar correctamente la demanda y obtener una adecuada planificación de inventarios y su stock.

Figura 36

Dashboard Predicción Venta de Routers



4.2 Análisis de Resultados (Mejor modelo)

Para seleccionar el mejor modelo se aplicó las métricas de evaluación a los 74 productos que fueron aplicados los dos modelos de series temporales. Al evaluar producto

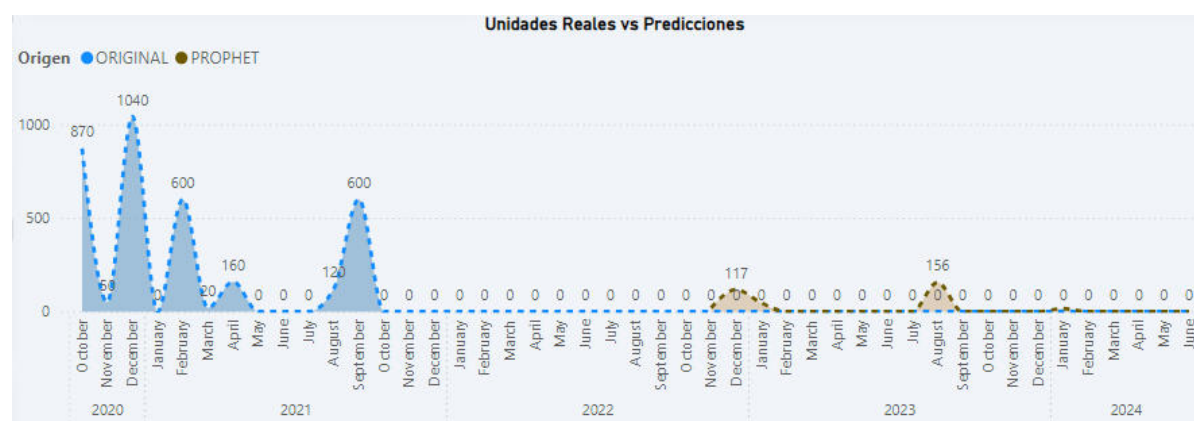
por producto visualmente y junto a sus métricas, se pudo obtener que el modelo Prophet tuvo mejores métricas en el 80% de los productos, seleccionándose como el mejor modelo.

Dentro de los productos evaluados existieron diferentes tipos de datos históricos en las series temporales, entre los principales se pueden detallar los siguientes y el desempeño de Prophet con un ejemplo visual:

Productos que mantenían ventas históricas y dejaron de venderse posiblemente por renovación o cambio en las tecnologías. Para estos escenarios es evidente que el modelo identificó en su entrenamiento que ya no existían ventas para ese producto pero a pesar de eso proyectó una pequeña cantidad de ventas que en la realidad nunca llegaron a darse.

Figura 37

Proyección ventas de un producto sin ventas actuales



Productos que llevan algunos años en el mercado y mantienen ventas históricas y actuales. Para estos escenarios el modelo tuvo un desempeño mayor con una proyección más acertada.

Figura 38

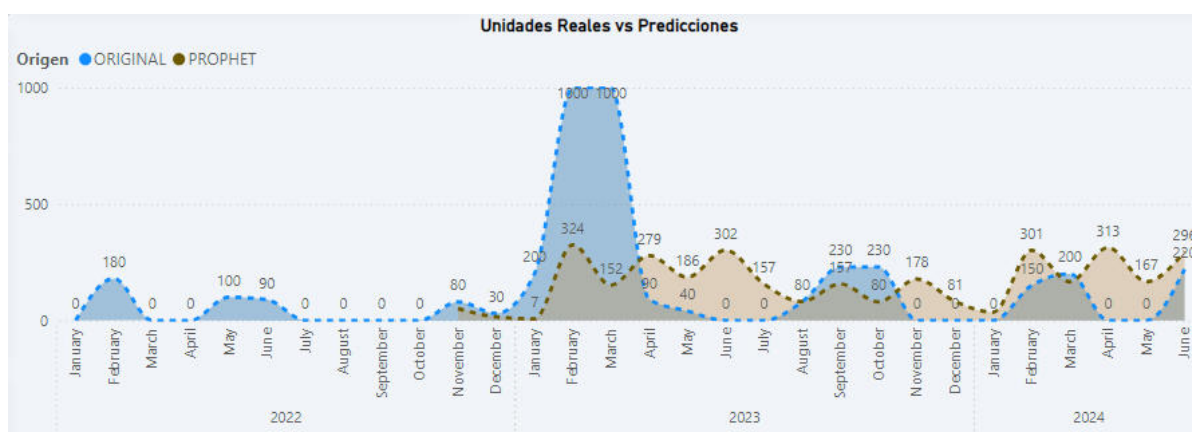
Proyección ventas de un producto con histórico y ventas actuales



Productos “nuevos” que empezaron sus ventas en un corto plazo. Para estos escenarios el modelo tiene cierto porcentaje de acierto pero posiblemente pueda ser mejor contando con un historial más grande de información.

Figura 39

Proyección ventas de un producto “nuevo”



La teoría indica que Prophet es más robusto y flexible ante datos faltantes o valores atípicos, en la práctica se comprobó que este modelo si tiene el desempeño mencionado, pero necesita de ciertos factores para cumplirlo. Adicional en un entorno real no se evaluarían productos que fueron discontinuados, ya no cuentan con ventas en el mercado o tienen un

historial muy corto de información, debido que un modelo de aprendizaje automático es más acertado cuando cuenta con bastante información para su entrenamiento.

4.3 KPI y Métricas de Valor

Con la fuente de datos utilizada se han determinado los indicadores que permiten alcanzar el objetivo de optimizar la gestión de inventarios de manera proactiva de acuerdo con el nivel de demanda.

Estos indicadores permitirán tomar decisiones acertadas al fabricante respecto al volumen de unidades, tomando en cuenta el producto o su categoría.

4.2.3 Indicadores Clave

Total unidades pronosticadas vs vendidas: Mide la precisión de la predicción de demanda en términos de las unidades de producto, evaluando cuan acertado fue el pronóstico en comparación con las unidades vendidas realmente.

Figura 40

Fórmula de cálculo de indicador Total Unidades Pronosticadas vs Vendidas

$$\text{Total Unidades Pronosticadas vs Vendidas} = \frac{\text{Unidades Pronosticadas}}{\text{Unidades Vendidas}}$$

Cantidad unidades pronosticadas vs vendidas por tiempo: Es la comparación de las previsiones de demanda con las ventas reales a nivel del producto, brindando una visión detallada de cuan acertado fue el pronóstico en el periodo determinado.

Figura 41

Fórmula de cálculo de indicador Unidades Pronosticadas vs Vendidas

$$\text{Unidades Pronosticadas vs Vendidas} = \frac{\text{Unidades Pronosticadas del Producto}}{\text{Unidades Vendidas del Producto}}$$

Porcentaje de Unidades por tecnología Wifi: Mide la participación de las unidades vendidas de la categoría de producto en relación con el número de unidades. Permite analizar como se distribuye la demanda entre las diferentes categorías e identificar las más relevantes.

Figura 42

Fórmula de cálculo de indicador Porcentaje de Unidades por tecnología Wifi

$$\text{Porcentaje de Unidades Vendidas por Categoría} = \frac{\text{Total Unidades Vendidas de la Categoría}}{\text{Total Unidades Vendidas Global}} \times 100$$

Productos más vendidos: Este análisis identifica los productos con el mayor volumen de ventas en un periodo determinado. Es clave para comprender que productos están generando mayor demanda.

Figura 43

Fórmula de cálculo de indicador Porcentaje de Unidades más vendidas

$$\text{Porcentaje de Unidades Vendidas} = \frac{\text{Unidades Vendidas del Producto}}{\text{Total de Unidades Vendidas}} \times 100$$

4.2.4 Métricas de Valor

Detección de patrones de crecimiento o descenso: En base a los patrones de crecimiento o descenso se puede identificar si las ventas de un producto muestran una tendencia creciente, decreciente o estables en el tiempo, lo que aporta a la toma de decisiones en cuanto a inversiones en campañas de marketing y promociones.

Estos patrones pueden ser calculados por medio de regresión lineal que modele la relación que existe entre el valor de unidades vendidas y el tiempo (segmentos de tiempo específicos).

Para este pronóstico, se debe aplicar los siguientes cálculos

$$\sum X : \text{Suma de los valores de tiempo}$$

$$\sum Y : \text{Suma de las unidades vendidas}$$

$$\sum XY : \text{Suma de los productos de tiempo y ventas}$$

$$\sum X^2 : \text{Suma de los cuadrados de los números de tiempo}$$

$$\text{Pendientes} = \frac{n \cdot \text{Suma de productos} - \text{Suma de tiempos} \cdot \text{Suma de ventas}}{n \cdot \text{Suma de cuadrados de tiempos} - (\text{Suma de tiempos})^2}$$

Donde:

- n es el número total de meses (12 en este caso)

Interpretación:

- Pendiente Positiva: Las ventas aumentan - crecimiento.
- Pendiente Negativa: Las ventas disminuyen - reducción.
- Pendiente Cero: Las ventas son estables – se mantienen.

Optimización de inventarios – Modelo punto de reorden: El modelo de punto de reorden permite determinar el nivel de inventario en el cual se debe adquirir nuevo stock, basada en el número de unidades indica como mantener un flujo constante de ventas.

Por lo cual aporta a:

- Optimizar la Disponibilidad de Productos: Mantener suficientes unidades para satisfacer la demanda.

- Buen servicio al Cliente: Evitar la falta de stock que podrían llevar a la pérdida de ventas por clientes insatisfechos.
- Aumentar y asegurar las Ventas: Asegurar que siempre haya productos disponibles para la venta.

Formulas:

$$\text{Punto de reorden} = \text{Demanda} \cdot \text{Tiempo de reabastecimiento}$$

$$\text{Punto de reorden} = \text{Unidades vendidas} \cdot \text{Tiempo de reabastecimiento (Frecuencia)}$$

4.4 Implementación en Producción y Aplicabilidad Entorno Real

Para implementar en producción el modelo de predicción de ventas generalmente se pueden seguir los pasos a continuación:

- **Despliegue del modelo:** Utilizando herramientas de control de versionamiento de código fuente se puede subir a producción el modelo para su consumo.
- **Automatización:** En la actualidad todas las decisiones se basan en los indicadores generados por la data, por lo cual contar con un proceso automatizado que considere tanto los datos históricos y los datos en tiempo real para alimentar el modelo periódicamente es fundamental. La herramienta planteada en este proyecto para la carga de información (Azure Data Factory) también tiene la función de orquestación de Datos, que permitiría programar tareas y actualizar continuamente todo el flujo de información.
- **Monitoreo del modelo:** Supervisar periódicamente el rendimiento del modelo una vez implementado es clave para asegurar que las predicciones se mantengan precisas en el tiempo y se conozca cuando es necesario un ajuste.

- **Reentrenamiento:** En el mundo de la tecnología las condiciones del mercado y demanda cambian constantemente, reentrenar periódicamente al modelo con nuevos datos resulta fundamental para estar a la vanguardia.

La implementación de un modelo de predicción puede ser clave para lograr una ventaja competitiva en el mercado, permitiendo a la empresa:

- **Optimizar los inventarios:** Con predicciones más precisas, es posible optimizar los niveles de inventario, minimizando el exceso o la falta de stock. Generando una mejor rotación de inventario y reducción de costos asociados con el almacenamiento.
- **Reducir los desperdicios:** En el caso de empresas de productos tecnológicos que tienen tendencia hacia la alta obsolescencia, una predicción de demanda precisa evitaría que productos se queden sin vender o descontinuados.
- **Mejorar la planificación estratégica:** Las predicciones ayudan en la toma de decisiones estratégicas ya que pueden incurrir en la creación de nuevas sucursales, una mejor planificación con proveedores o incluso una expansión de líneas de productos.
- **Tomar decisiones basadas en datos:** Los equipos campañas pueden ajustar sus estrategias de promoción según las tendencias de demanda esperadas, aumentando la efectividad de las acciones comerciales.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

La comparación entre los modelos SARIMA y Prophet reveló resultados contradictorios en cuanto a la precisión de sus predicciones. Mientras que SARIMA demostró que necesita una estacionalidad conocida y regular, Prophet se destacó por su manejo automático de estacionalidad, similar caso en temas de tendencia el primer modelo requiere una diferenciación manual mientras que el segundo es más automático y flexible.

El análisis de la estacionariedad mediante pruebas como ACF y PACF revelan la importancia de elegir modelos adecuados en función de la naturaleza de las series temporales. La identificación de series no estacionarias y la adaptación de los modelos empleados a las características específicas de las series son cruciales para obtener pronósticos fiables. Para las empresas que hagan uso de estos modelos, esto se traduce en una mejor logística, planificación de inventarios, optimización de estrategias y ventajas en cuanto a la toma de decisiones.

El uso de gráficas de modelamiento es una herramienta clave para comprender el comportamiento de las series temporales, ya que ofrecen una visión clara de la estructura subyacente de las variables y permiten identificar patrones estacionales y tendencias. La correcta interpretación de estas gráficas facilita el ajuste preciso de los parámetros del modelo, lo que optimiza las predicciones.

El modelo Prophet demostró ser significativamente más robusto y preciso que SARIMA en series temporales con datos faltantes. Mientras que SARIMA requiere datos completos y a intervalos regulares para obtener predicciones precisas, Prophet fue capaz de manejar la falta de datos sin la necesidad de realizar interpolaciones o preprocesamiento

adicional. Esto permitió una mejor captura de tendencias y estacionalidades en las series temporales evaluadas.

5.2 Recomendaciones

Es crucial implementar un modelo predictivo que se ajuste continuamente a las condiciones del mercado, esto implica monitorear factores externos, como lanzamientos de nuevos dispositivos, cambios en los costos y avances tecnológicos, para ajustar las estrategias de ventas y distribución, maximizando el impacto de las transiciones tecnológicas. Además, profundizar en el análisis de la estacionalidad que permitirá comprender mejor la demanda, incorporando variables que afectan la temporalidad de las ventas, lo que mejora la precisión de los modelos y el control del inventario.

A partir de este estudio, se recomienda que para futuros análisis de series temporales donde los datos faltantes sean un problema común, se utilice Prophet como el modelo de preferencia. Sin embargo, en casos donde los datos sean completos y la estacionalidad sea regular, SARIMA sigue siendo una alternativa viable. Además, se sugiere explorar métodos híbridos que combinen las ventajas de ambos modelos para mejorar aún más la precisión de las predicciones.

Es fundamental seguir cargando nuevos datos a una serie temporal para que el modelo pueda captar patrones más recientes, ya que estos reflejan cambios en tendencias, estacionalidad o eventos imprevistos. Además, con más datos, el modelo puede ajustarse mejor y mejorar su capacidad de predicción al reducir errores y sesgos causados por datos insuficientes o desactualizados.

Para llevar a producción una arquitectura basada en estos modelos, es esencial considerar aspectos como la escalabilidad, seguridad y resiliencia de la plataforma, en función del tamaño de la información y su proyección a futuro. Dado que estas plataformas

suelen ser PaaS (Plataforma como Servicio), es igualmente crucial definir un presupuesto claro que permita sostener la operación de manera eficiente y alineada con los objetivos del negocio.

REFERENCIAS BIBLIOGRÁFICAS

- Alvarez, D. (11 de mayo de 2024). *Blog Datlas*. Obtenido de Predicciones de series temporales con PROPHET de facebook (META) – Columna de Investigación DATLAS:
<https://blogdatlas.wordpress.com/2024/05/11/predicciones-de-series-temporales-con-prophet-de-facebook-meta-columna-de-investigacion-datlas/>
- ARCOTEL. (18 de febrero de 2015). *LEY ORGANICA DE TELECOMUNICACIONES*. Obtenido de
<https://www.arcotel.gob.ec/wp-content/uploads/2015/10/ley-organica-de-telecomunicaciones.pdf>
- AWS. (2024). *Algoritmo Prophet*.
- Banco Interamericano de Desarrollo(BID). (s.f.). *nformes sobre la adopción de tecnología y transformación digital en América Latina*. Obtenido de <https://www.iadb.org/es>
- BanEcuador. (s.f.). *Credito 1x30 continua reactivando a sectores productivos del país*. Obtenido de
<https://www.banecuador.fin.ec/2023/11/11/credito-1x30-continua-reactivando-a-sectores-productivos-del-pais/>
- Bowser, A. (2019). Synthetic Data: An Alternative to Real Data. *Data Privacy Journal*.
- Box, G. E. (2008). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- CEIC. (2024). *Ecuador Unemployment Rate*. Obtenido de
<https://www.ceicdata.com/en/indicator/ecuador/unemployment-rate>
- Chatfield, C. (2004). *The Analysis of Time Series: An Introduction*. CRC Press.
- Chopra, S., & Meindl, P. (2013). *Supply Chain Management: Strategy, Planning, and Operation*. Pearson.
- El Telégrafo. (17 de febrero de 2023). BanEcuador amplía su crédito 1x30 y llega al sector pesquero. págs. <https://www.eltelegrafo.com.ec/noticias/nacionales/44/banecuador-amplia-su-credito-1x30-y-llega-al-sector-pesquero>.
- George, B., Gwilym, J., Gregory, R., & Greta, L. (2015). *Time Series Analysis: Forecasting and Control, 5th Edition*. New Jersey: John Wiley and Sons Inc., Hoboken.
- González Casimiro, M. P. (2009). *Análisis de series temporales*.
- Hans Franses, P., van Dijk, D., & Opschoor, A. (2014). *Time Series Models for Business and Economic Forecasting*. Cambridge University Press.
- Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
- Jenkins, G. M., & Pelham Box, G. E. (s.f.). *Series Temporales, Modelo ARIMA*.
- Joseph, M. (2022). *Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning*. Packy Publishing.
- La Republica. (11 de diciembre de 2021). *Consumo de internet en el mundo aumentó 19,5% durante la pandemia de covid-19*. Obtenido de <https://www.larepublica.co/consumo/consumo-de-internet-en-el-mundo-aumento-19-5-durante-la-pandemia-de-covid-19-3274945>
- Lee, K. (2020). *Deep Learning for Time Series Forecasting*. Springer Nature.

- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 802-808.
- MarketsandMarkets. (2021). *Network Equipment Market by Component (Routers, Switches, & Gateways)*. MarketsandMarkets.
- Martinez Velencoso, L. (2021). *La protección jurídica de la persona en el ámbito de la Biotecnología y del Big Data*. Dykinson.
- Mauricio, J. A. (2007). *Análisis de Series Temporales*.
- Meta. (23 de febrero de 2017). *Prophet: forecasting at scale*. Obtenido de <https://research.facebook.com/blog/2017/2/prophet-forecasting-at-scale/>
- Microsoft. (2024). *Microsoft Learn*. Obtenido de <https://learn.microsoft.com/>
- Mideros Mora, A. (16 de enero de 2024). *PRIMICIAS*. Obtenido de Los desafíos económicos frente a la crisis de inseguridad: <https://www.primicias.ec/noticias/firmas/desafios-economicos-crisis-seguridad-ecuador/>
- Mifsud, A. G. (2019). *Automatic forecasting y sus aplicaciones*.
- MINTEL. (s.f.). *Informes sobre importaciones y costos de tecnología*.
- Mirete Blanco, H. (2019). *Extracción y predicción de datos de series*.
- Pew Research Center. (1 de septiembre de 2021). *The Internet and the Pandemic*. Obtenido de <https://www.pewresearch.org/internet/2021/09/01/the-internet-and-the-pandemic/>
- Ríos, G. (2008). *Series de Tiempo*. Chile.
- Rodriguez, M. (2020). *Scrum: el pasado y el futuro*. Obtenido de Netmind: <https://netmind.net/scrum-el-pasado-y-el-futuro/>
- Shumway, R. H. (2017). *Time Series Analysis and Its Applications: With R Examples*. Springer.
- Smith, J. (2021). *Technology and Connectivity: A New Era Post-COVID*. Academic Press.
- Taylor, S. J. (2018). *Prophet: Forecasting at Scale*. *The American Statistician*.
- Villarreal, F. (01 de octubre de 2022). *Banco Central del Ecuador*. Obtenido de Estimación de pérdidas y daños ocasionados por el paro nacional de junio de 2022: <https://www.bce.fin.ec/publicaciones/editoriales/estimacion-de-perdidas-y-danos-ocasionados-por-el-paro-nacional-de-junio-de-2022>
- Wei, W. W. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson.
- World Bank Group. (18 de octubre de 2021). *Ecuador, the Country that Vanquished the Nightmare Pandemic in 100 Day*. Obtenido de <https://www.worldbank.org/en/news/feature/2021/10/18/ecuador-the-country-that-vanquished-the-nightmare-pandemic-in-100-days>

APÉNDICES

MANUAL TÉCNICO: ANÁLISIS DE DATOS SINTÉTICOS DE VENTAS DE PRODUCTOS DE REDES TECNOLÓGICAS Y COMPARACIÓN DE DOS MODELOS DE SERIES TEMPORALES QUE PERMITAN PREDECIR SU DEMANDA

Fecha	Autor	Descripción	Versión
2024-08-27	Grupo 4	Creación del documento.	1.0

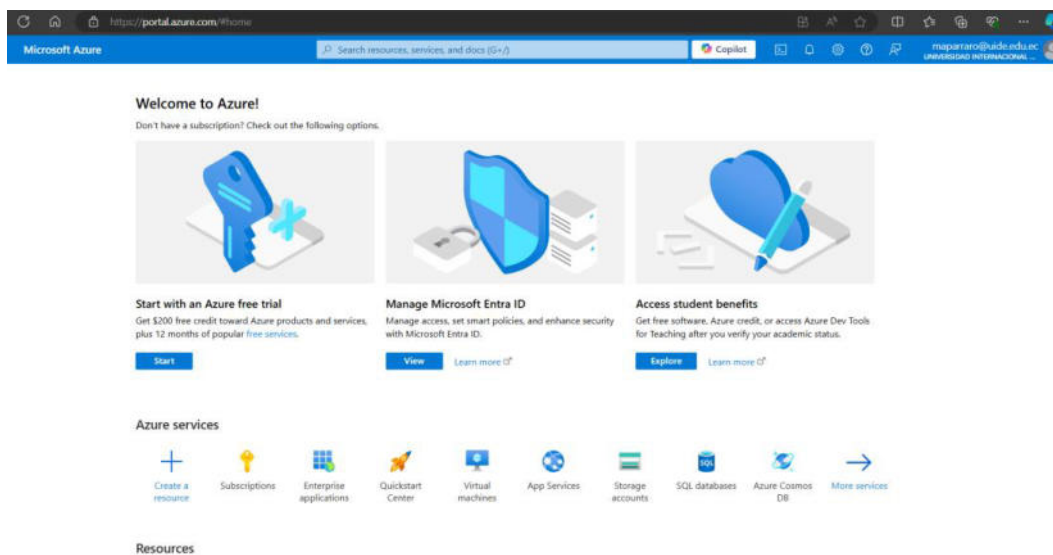
Objetivo

El objetivo de este documento es brindar una guía técnica para el aprovisionamiento de la arquitectura Microsoft Azure.

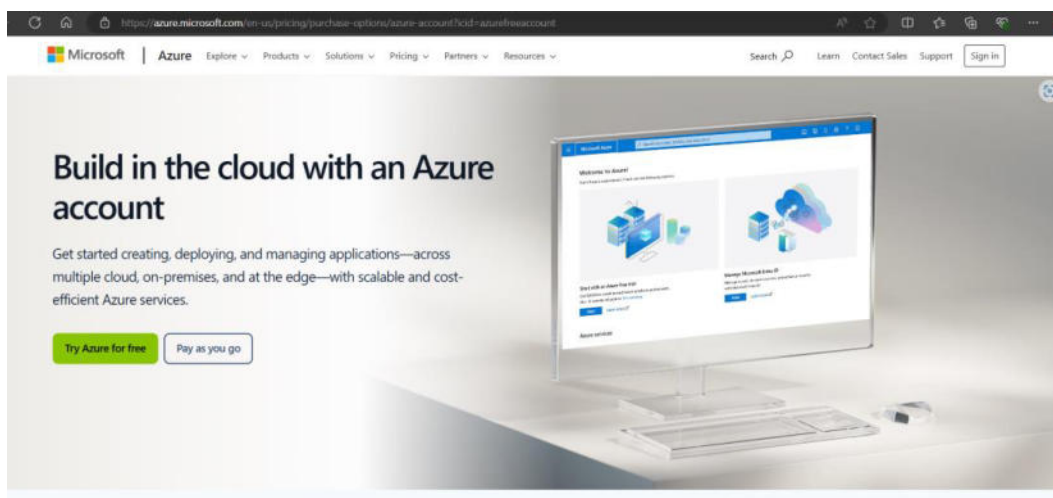
Aprovisionamiento arquitectura

Creación usuario Microsoft Azure

Es necesario contar con una cuenta correo Outlook, Microsoft Office o Office 365 para crear la cuenta en el portal de Azure. <https://portal.azure.com/>

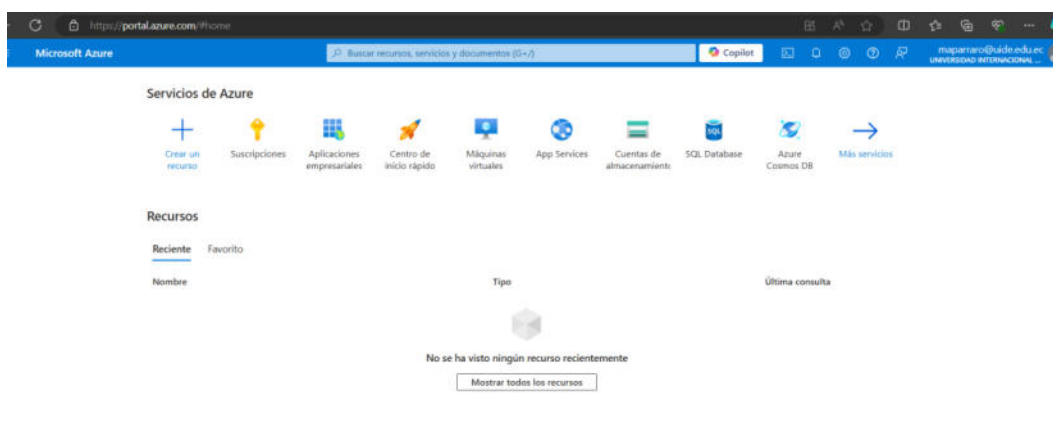


Para la vinculación del usuario existen 2 métodos: prueba gratis por 12 meses o pago por uso.

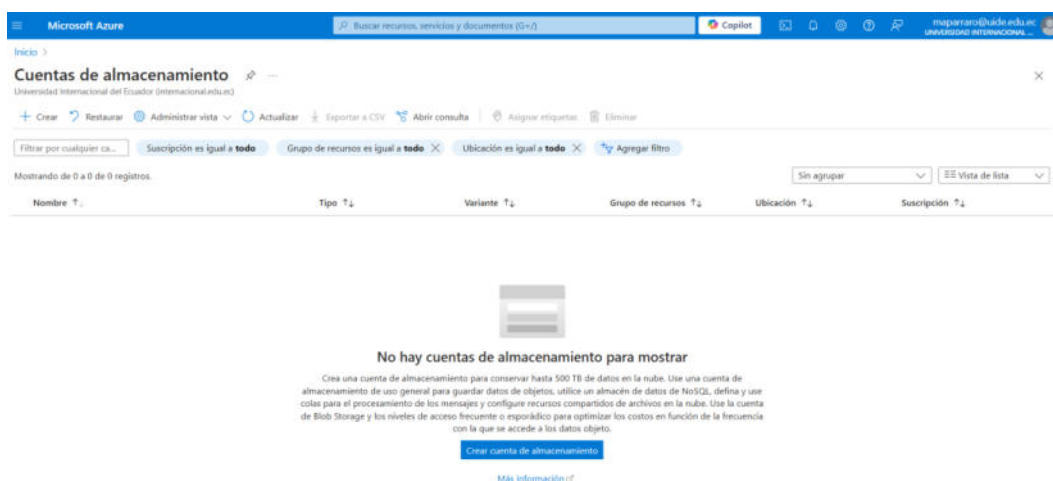


Creación cuenta almacenamiento

Seleccionamos cuentas de almacenamiento.



Seleccionamos Crear una nueva cuenta del almacenamiento.



Completamos los datos básicos solicitados en el formulario de creación.

Microsoft Azure | Buscar recursos, servicios y documentos (0+0) | Copilot

Inicio > Cuentas de almacenamiento >

Crear una cuenta de almacenamiento

Datos básicos | Avanzado | Redes | Protección de datos | Cifrado | Etiquetas | Revisar y crear

Azure Storage es un servicio administrado por Microsoft que proporciona almacenamiento en la nube altamente disponible, seguro, duradero, escalable y redundante. Azure Storage incluye Azure Blob (objetos), Azure Data Lake Storage Gen2, Azure Files, Azure Queues y Azure Tables. El costo de una cuenta de Storage depende del uso y de las opciones que elija a continuación. [Más información sobre las cuentas de almacenamiento de Azure](#)

Detalles del proyecto
 Seleccione la suscripción en la que se creará la nueva cuenta de almacenamiento. Elija un grupo de recursos nuevo o uno ya existente para organizar y administrar la cuenta de almacenamiento junto con otros recursos.

Suscripción * | Azure subscription 1 |
 Grupo de recursos * | (Nuevo) GR_ALMACENAMIENTO |
[Crear nuevo](#)

Detalles de la instancia

Nombre de la cuenta de almacenamiento * | asventas |
 Región * | (US) East US |
[Implementación en una zona extendida de Azure](#)
 Servicio principal | Azure Blob Storage o Azure Data Lake Storage Gen2 |

[Anterior](#) | [Siguiente](#) | [Revisar y crear](#) | [Enviar comentarios](#)

Microsoft Azure | Buscar recursos, servicios y documentos (0+0) | Copilot

Inicio > Cuentas de almacenamiento >

Crear una cuenta de almacenamiento

Detalles de la instancia

Nombre de la cuenta de almacenamiento * | asventas |
 Región * | (US) East US |
[Implementación en una zona extendida de Azure](#)
 Servicio principal | Azure Blob Storage o Azure Data Lake Storage Gen2 |
 Carga de trabajo principal | Seleccionar... |
 Personalice la cuenta de almacenamiento para la carga de trabajo o con un valor preestablecido recomendado modificable en función del procedimiento recomendado. Consulte la tabla de resumen para ver los valores preestablecidos recomendados.
 Rendimiento * | ☒ Estándar: Opción recomendada para la mayoría de los escenarios (cuenta de uso general v2) |
☐ Premium: Se recomienda para escenarios que requieren una latencia baja.
 Redundancia * | Almacenamiento con redundancia geográfica (GRS) |
☐ Habilite el acceso de lectura a los datos en el caso de que la región no esté disponible.

[Anterior](#) | [Siguiente](#) | [Revisar y crear](#) | [Enviar comentarios](#)

Una vez validados todos los campos seleccionamos Crear el recurso.

Microsoft Azure | Buscar recursos, servicios y documentos (0+0) | Copilot

Inicio > Cuentas de almacenamiento >

Crear una cuenta de almacenamiento

Datos básicos | Avanzado | Redes | Protección de datos | Cifrado | Etiquetas | **Revisar y crear**

[Ver plantilla de automatización](#)

Datos básicos

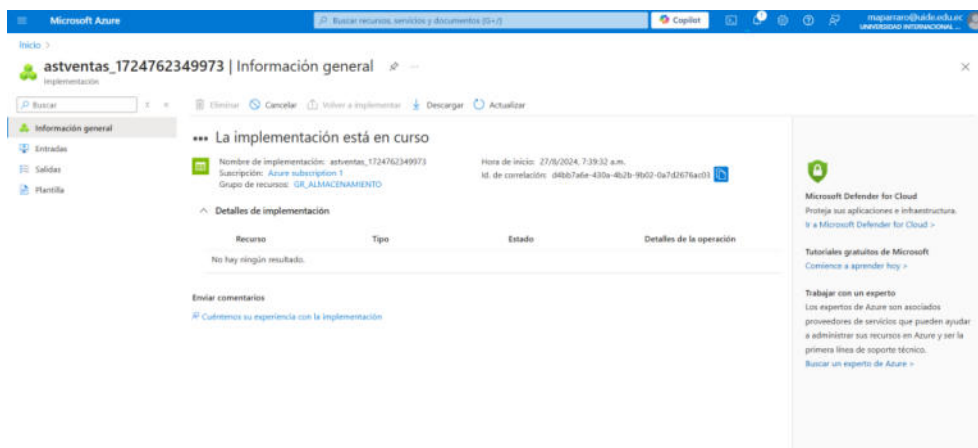
Suscripción	Azure subscription 1
Grupo de recursos	GR_ALMACENAMIENTO
Ubicación	East US
Nombre de la cuenta de almacenamiento	asventas
Servicio principal	Azure Blob Storage o Azure Data Lake Storage Gen2
Carga de trabajo principal	
Rendimiento	Estándar
Replicación	Almacenamiento con redundancia geográfica (GRS)

Opciones avanzadas

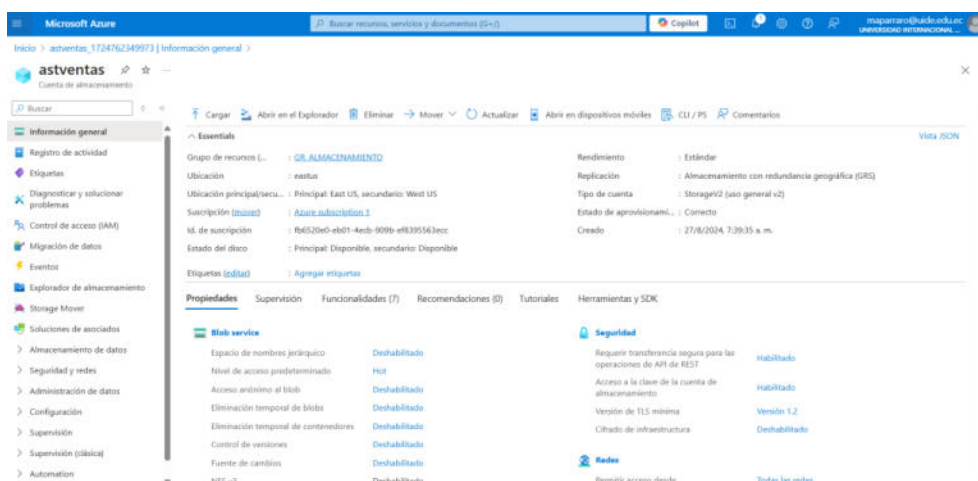
Habilitar el espacio de nombres jerárquico	Deshabilitado
Habilitar SFTP	Deshabilitado
Habilitar el sistema de archivos de red v3	Deshabilitado
Permitir replicación entre espacios empresariales	Deshabilitado

[Anterior](#) | [Siguiente](#) | [Crear](#) | [Enviar comentarios](#)

Se visualizará la pantalla de espera mientras el recurso es aprovisionado.

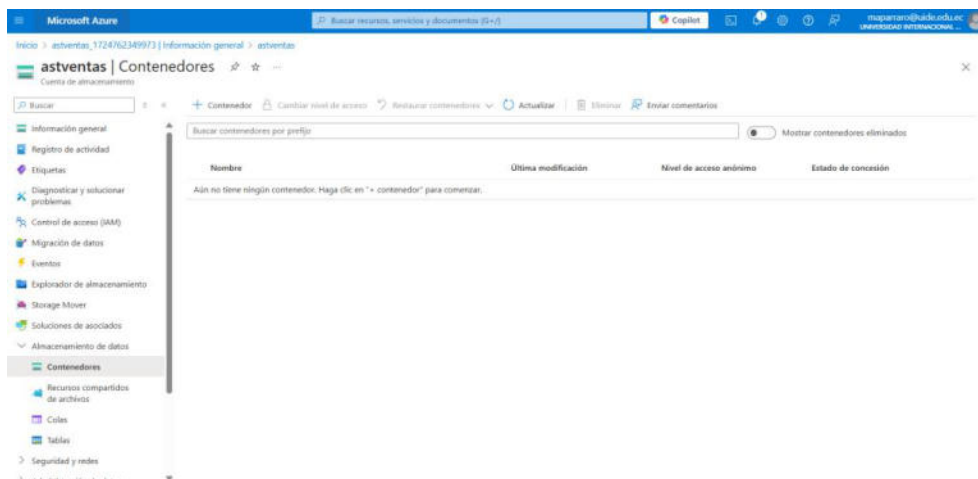


Completada la creación ingresamos al recurso.

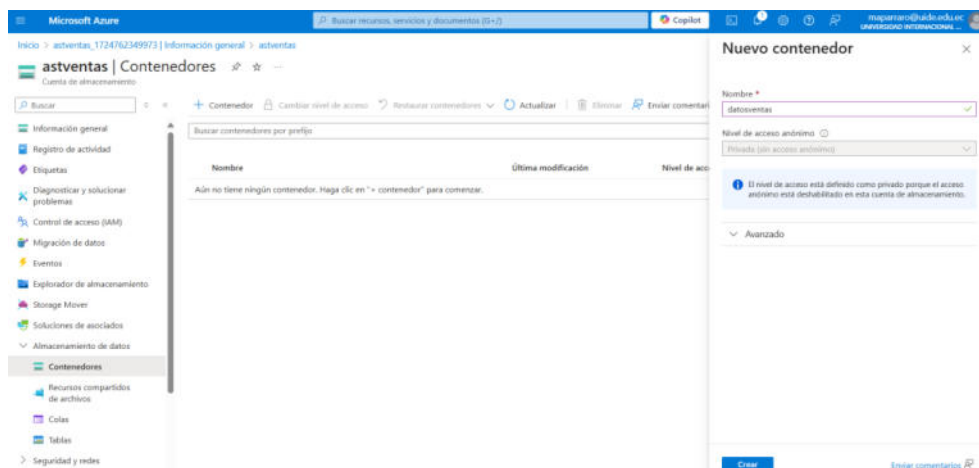


Creación del contenedor en la cuenta almacenamiento

En el panel izquierdo seleccionamos Contenedores.

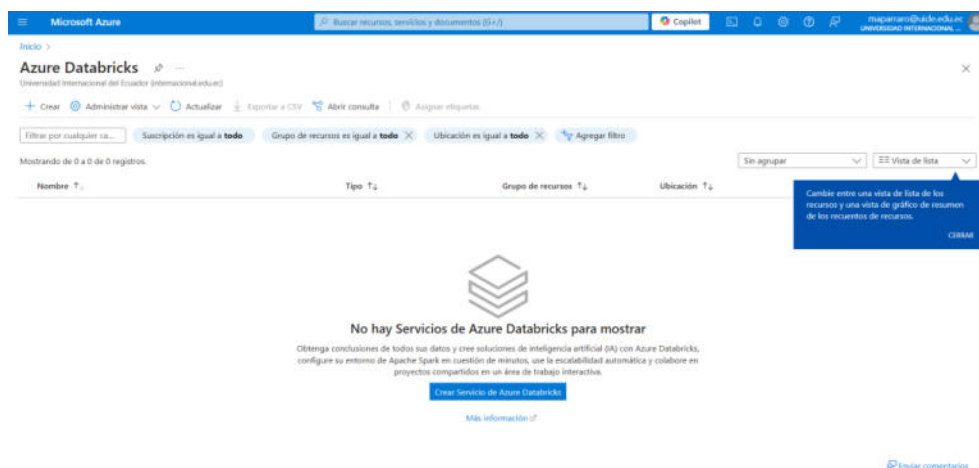


Seleccionamos Nuevo Contenedor y completamos el nombre deseado.

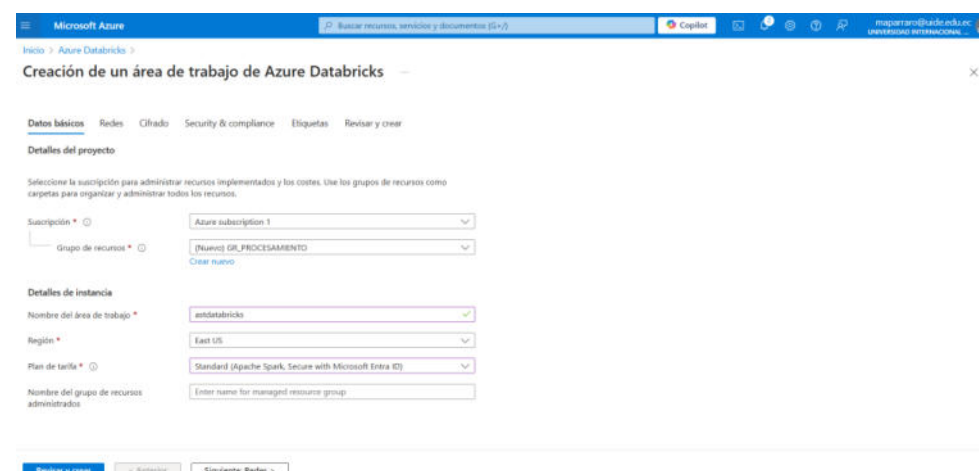


Creación Plataforma analítica Azure Databricks

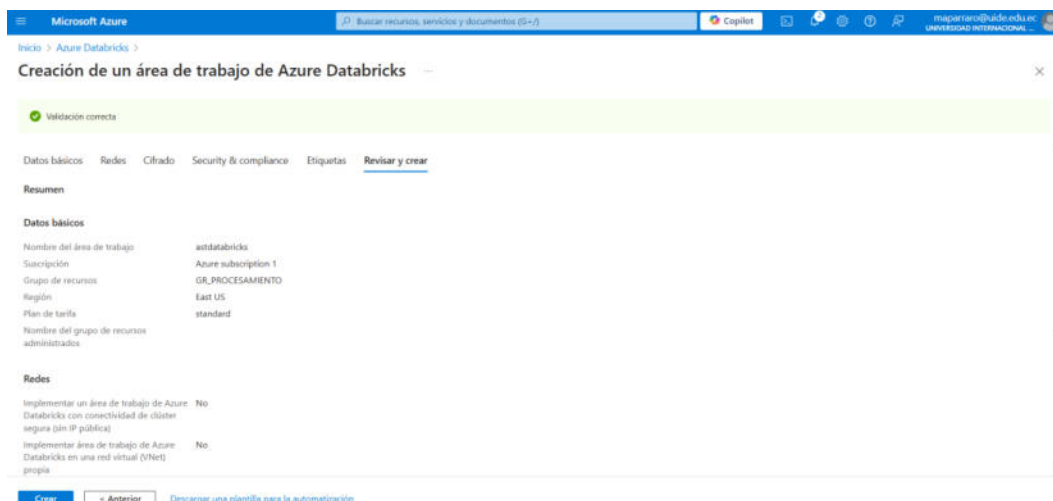
Seleccionamos Azure Databricks y posterior Crear un nuevo Azure Databricks.



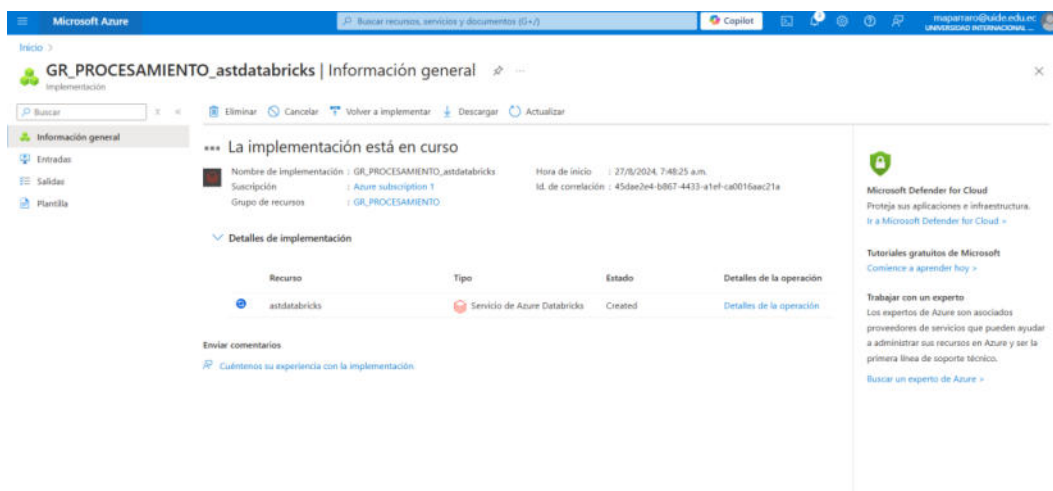
Completamos los datos básicos solicitados en el formulario de creación.



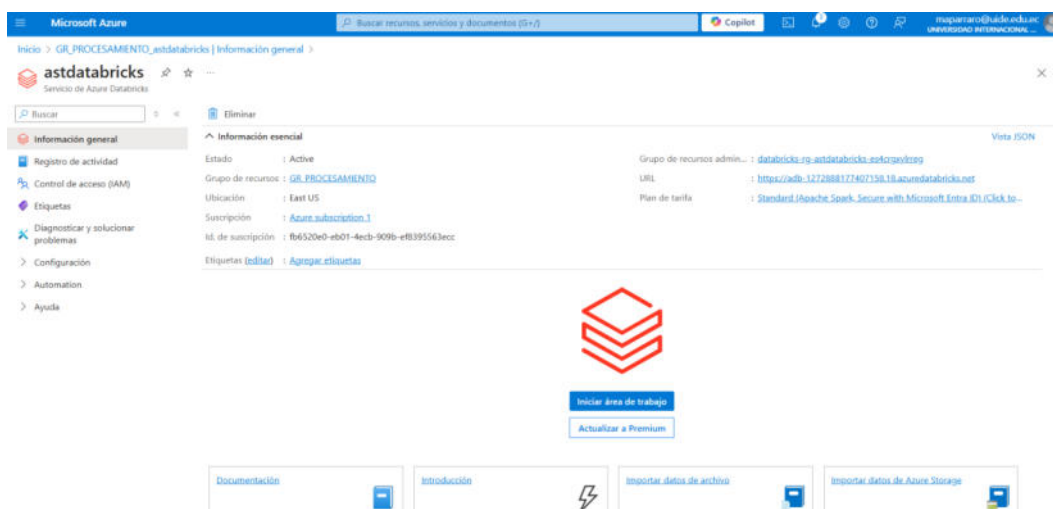
Una vez validados todos los campos seleccionamos Crear el recurso.



Se visualizará la pantalla de espera mientras el recurso es aprovisionado

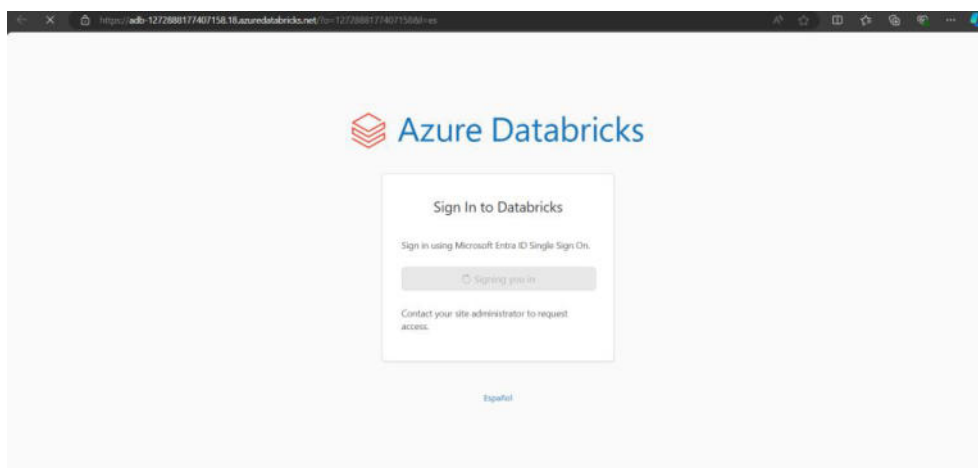


Completada la creación ingresamos al recurso

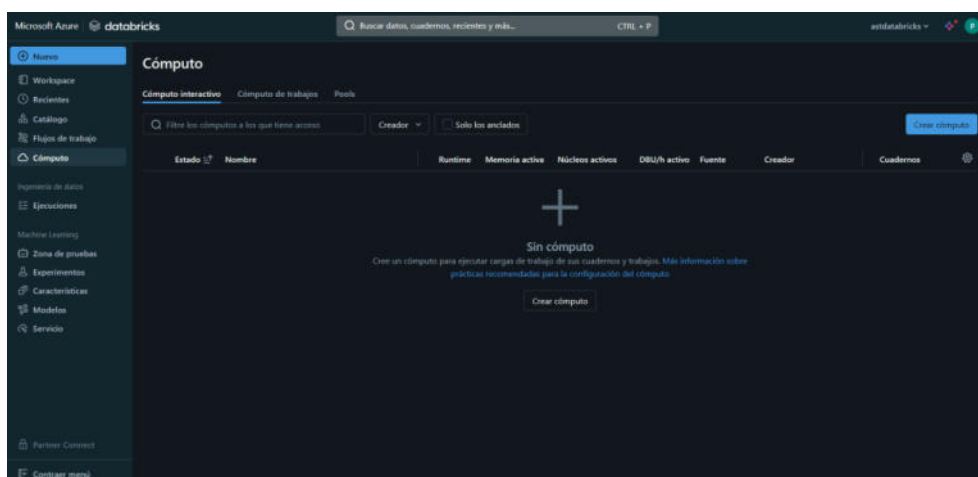


Creación cluster de trabajo en Azure Databricks

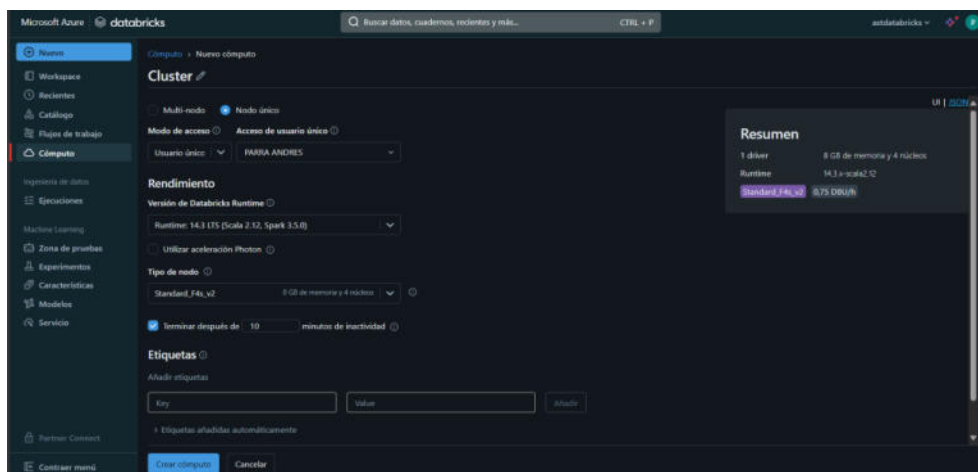
Ingresamos a la cuenta de Azure Databricks.



Seleccionamos Cómputo y Crear Cómputo

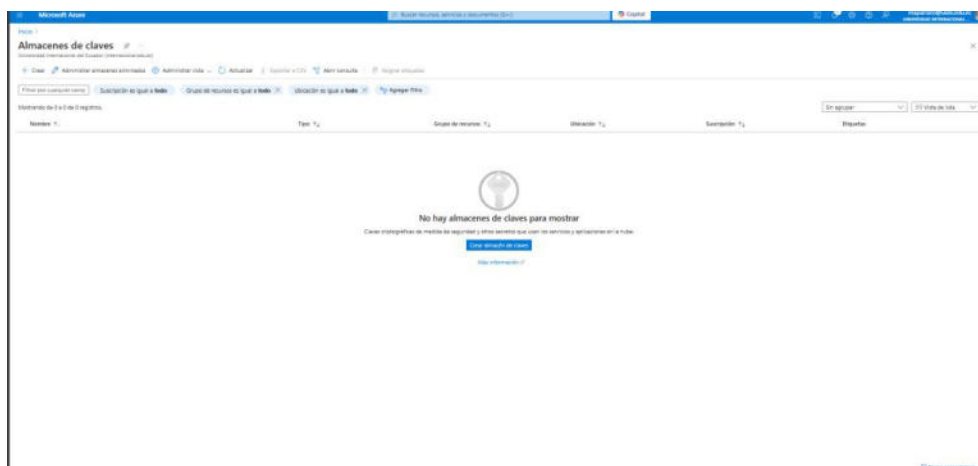


Completamos las características del cluster y seleccionamos Crear Cómputo.

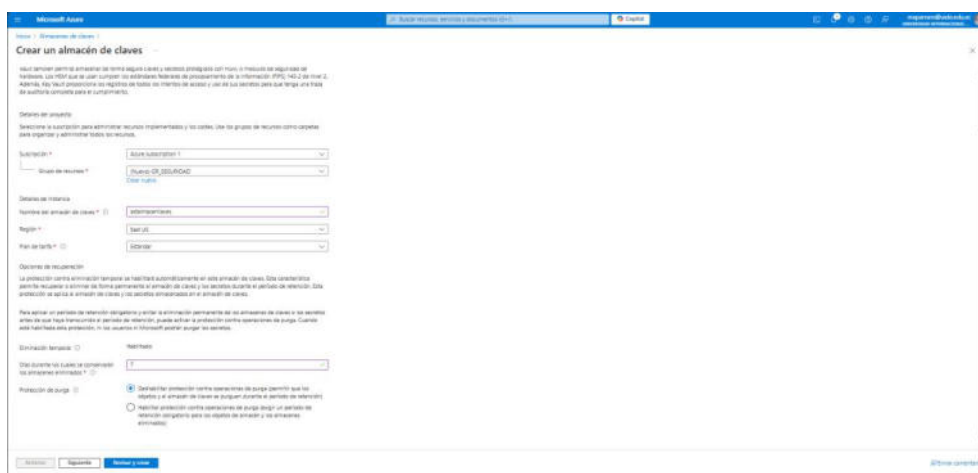


Creación almacén de llaves

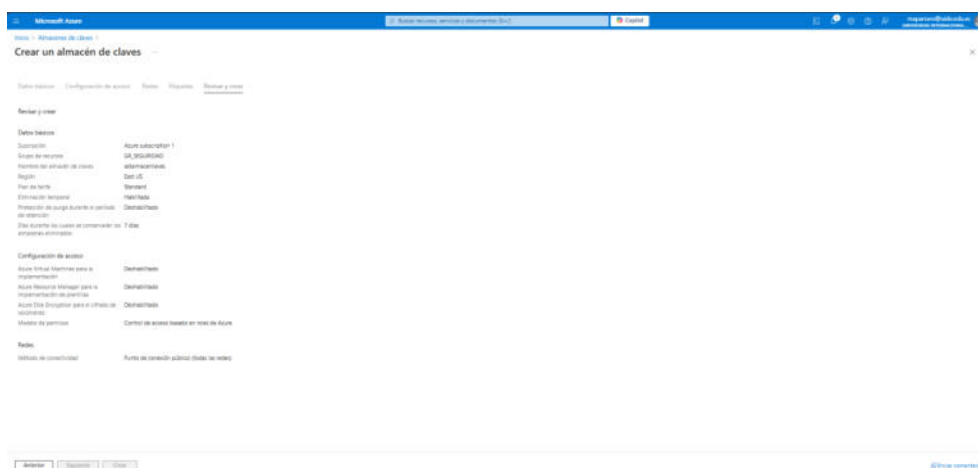
Seleccionamos Almacenes de llaves y posterior Crear.



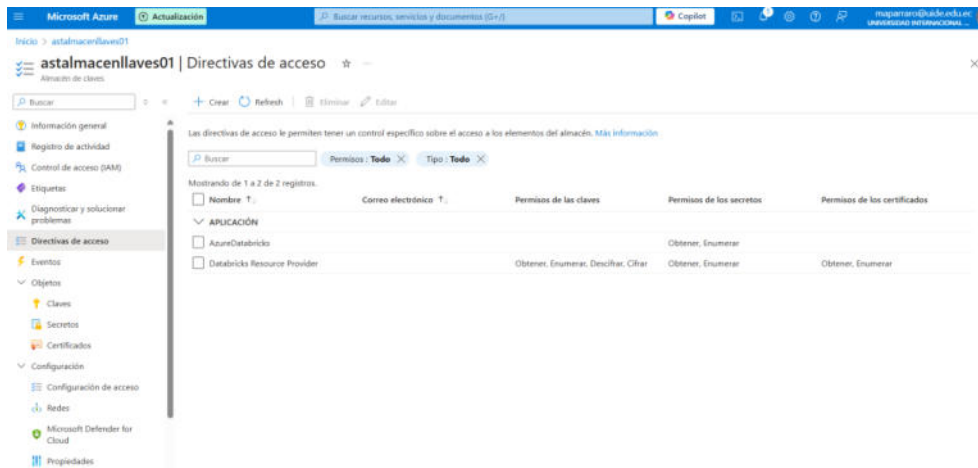
Completamos los datos básicos solicitados en el formulario de creación



Una vez validados todos los campos seleccionamos Crear el recurso.

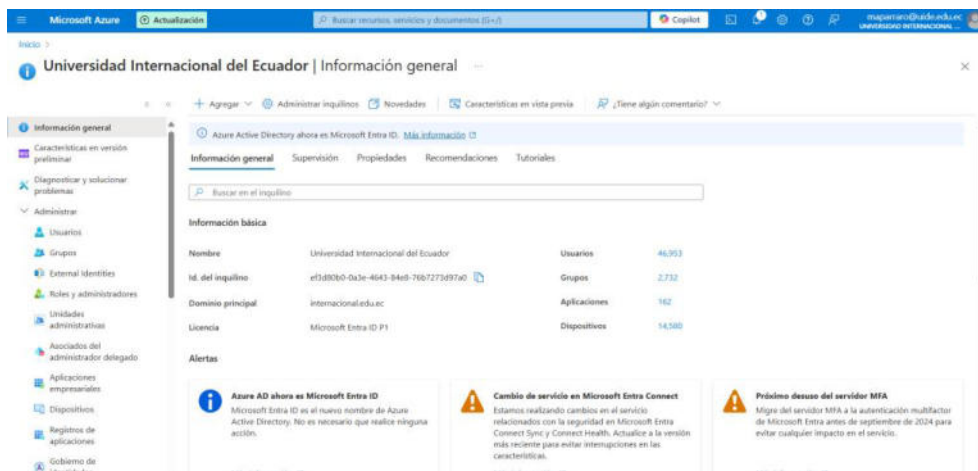


Asignamos los permisos de Obtener y Enumerar en las directivas de acceso del almacén de llaves al recurso de Azure Databricks. Necesario para que la plataforma tenga acceso a la llave.



Creación entidad de servicio

Ingresamos al Entra ID y seleccionamos Registro de Aplicaciones.



Seleccionamos Nuevo Registro y completamos el nombre de la Entidad de Servicio.

Microsoft Azure Actualización Buscar recursos, servicios y documentos (5+0) Copilot mapamw@univ.edu.ec UNIVERSIDAD INTERNACIONAL

Inicio > Universidad Internacional del Ecuador | Registro de aplicaciones >

Registrar una aplicación

* Nombre

Nombre para mostrar accesible por los usuarios de esta aplicación. Se puede cambiar posteriormente.

AST-APP-DATABRICKS

Tipos de cuenta compatibles

¿Quién puede usar esta aplicación o acceder a esta API?

☒ Solo cuentas de este directorio organizativo (solo de Universidad Internacional del Ecuador: inquilino único)

☐ Cuentas en cualquier directorio organizativo (cualquier inquilino de id. de Microsoft Entra - multinquilino)

☐ Cuentas en cualquier directorio organizativo (cualquier inquilino del id. de Microsoft Entra - multinquilino) y cuentas personales de Microsoft (por ejemplo, Skype, Xbox)

☐ Solo cuentas personales de Microsoft

[Ayúdame a elegir...](#)

URI de redirección (opcional)

Devolveremos la respuesta de autenticación a esta dirección URI después de autenticar correctamente al usuario. Este dato es opcional y se puede cambiar más tarde, pero se necesita un valor para la mayoría de los escenarios de autenticación.

[Al continuar, acepta las directivas de la plataforma Microsoft.](#)

[Registrar](#)

Copiamos los valores de ID aplicación, ID directorio que nos servirán mas adelante para la creación del punto de montura.

Microsoft Azure Actualización Buscar recursos, servicios y documentos (5+0) Copilot mapamw@univ.edu.ec UNIVERSIDAD INTERNACIONAL

Inicio > Universidad Internacional del Ecuador | Registro de aplicaciones >

AST-APP-DATABRICKS

Buscar Eliminar Puntos de conexión Características en versión preliminar

Información general Inicio rápido Asistente para integración Diagnosticar y solucionar problemas Administrar Soporte técnico y solución de problemas

Información esencial

Nombre para mostrar: [AST-APP-DATABRICKS](#)

id. de aplicación (id. de cliente): [\[Redacted\]](#)

Identificador de objeto: [\[Redacted\]](#)

id. de directorio (inquilino): [\[Redacted\]](#)

Tipos de cuenta compatibles: [Solo id. de organización](#)

Credenciales de cliente

id. de cliente: [\[Redacted\]](#)

URI de redirección: [Agregar un URI de redirección](#)

URI de id. de aplicación: [Agregar un URI de id. de aplicación](#)

Aplicación administrada...: [AST-APP-DATABRICKS](#)

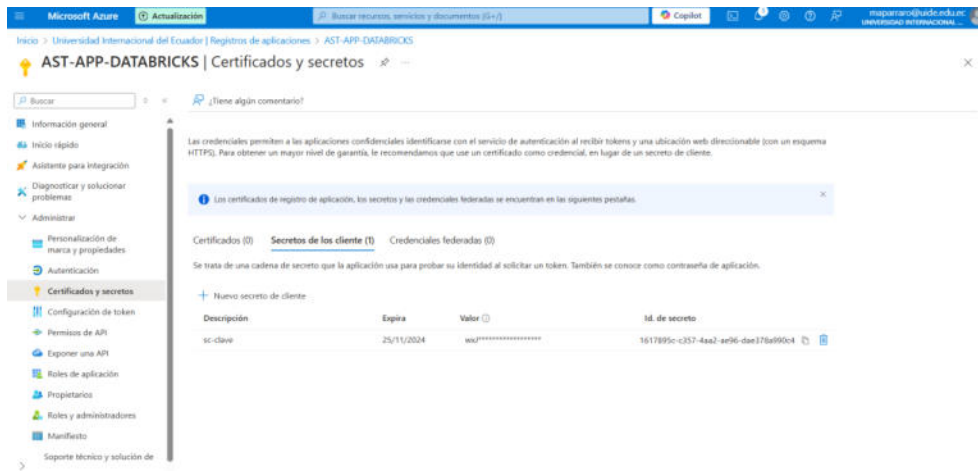
A partir del 30 de junio de 2020, ya no agregaremos ninguna característica nueva a la Biblioteca de autenticación de Azure Active Directory (AAD) y a Azure Active Directory Graph. Continuaremos brindando soporte técnico y actualizaciones de seguridad, pero ya no brindaremos actualizaciones de funciones. Las aplicaciones deberán actualizarse a la Biblioteca de autenticación de Microsoft (MSAL) y Microsoft Graph. [Más información](#)

Introducción Documentación

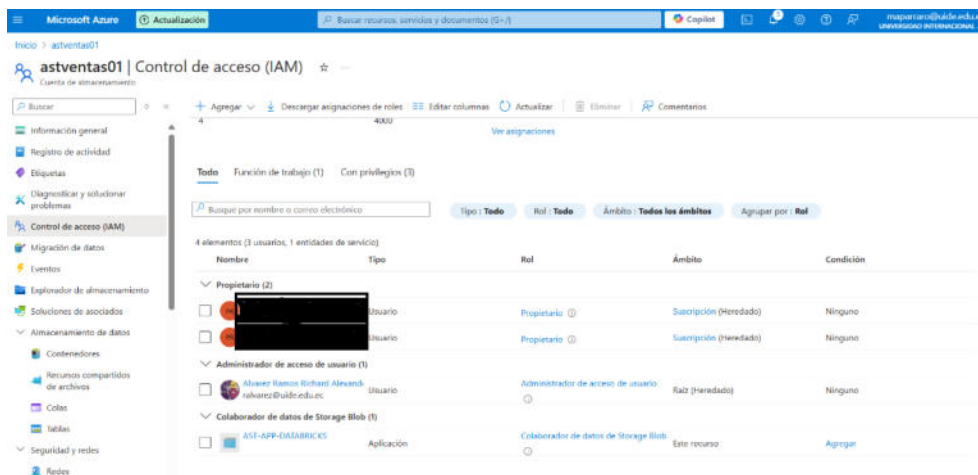
Compilación de la aplicación con la Plataforma de identidad de Microsoft

La Plataforma de identidad de Microsoft consta de un servicio de autenticación, bibliotecas de código abierto y herramientas de administración de aplicaciones. Puede crear soluciones de autenticación modernas basadas en estándares, obtener acceso a las API y protegerlas, y agregar información de inicio de sesión para sus usuarios y clientes. [Más información](#)

En credenciales de cliente seleccionamos crear un nuevo secreto y lo copiamos para configurar el secreto del almacén de llaves.

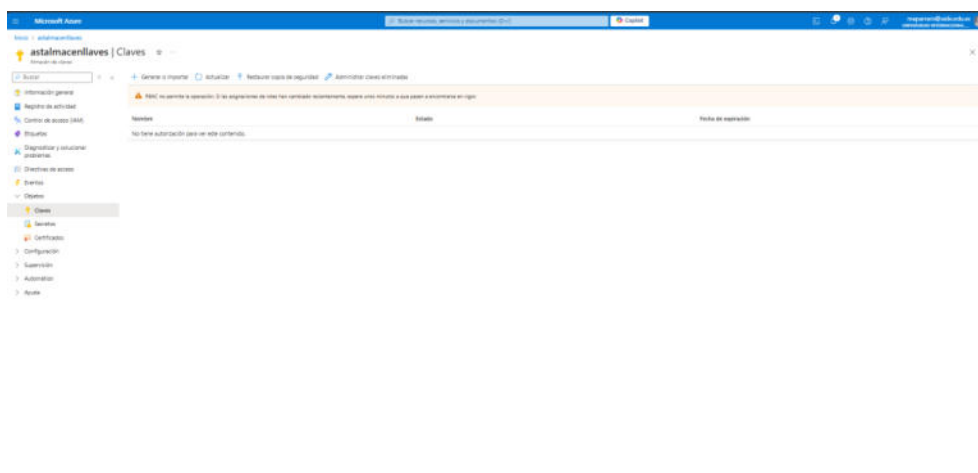


Asignamos permiso de Colaborador de blob de datos a la entidad de Servicio en el Control de Acceso IAM de la cuenta del almacenamiento.



Creación secreto en el almacén de llaves

Ingresamos a la cuenta de almacén de llaves y seleccionamos Secretos.



Completamos un nombre y pegamos el valor del secreto de la Entidad de Servicio.

Microsoft Azure

Crear un secreto

Nombre *

Valor secreto *

Tipo de contenido (opcional)

Establecer la fecha de expiración *

Establecer fecha de expiración *

Mostrar

Crear Cancelar

Creación scope en Azure Databricks

Ingresamos en Azure Databricks y en la url ingresamos a la creación de secretos con la siguiente URL.

<https://<databricks-instance>#secrets/createScope>

Microsoft Azure

databricks

Buscar datos, cuadros, recientes y más...

CTRL + P

autodatabricks01

HomePage / Crear un secret scope

Create Secret Scope

Cancel Create

Learn more

Scope Name *

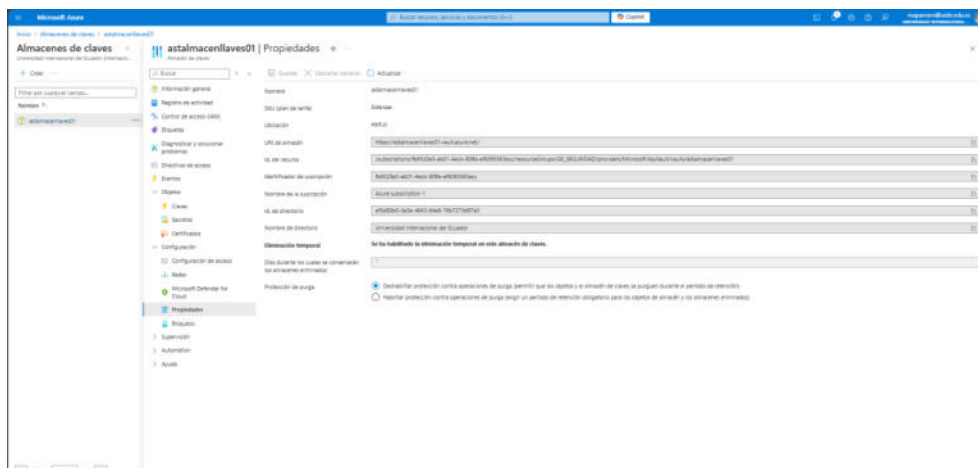
Manager Principal *

Azure Key Vault *

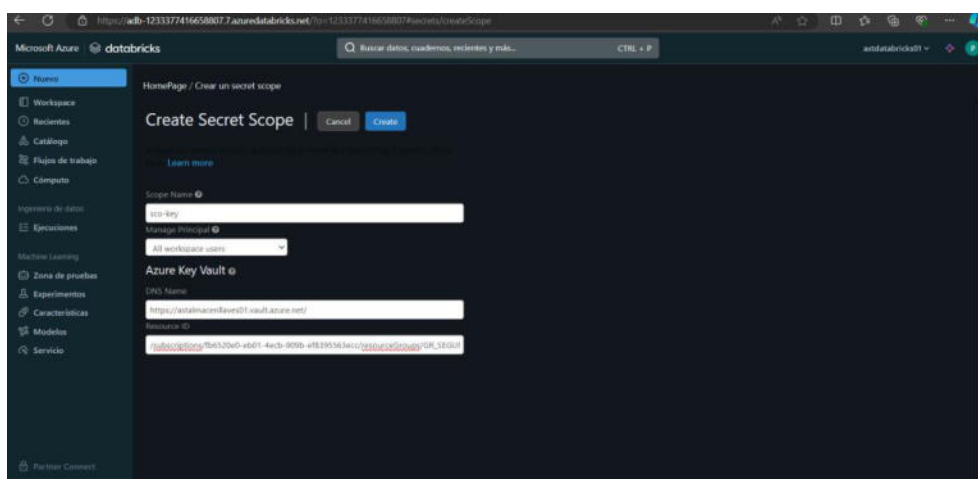
DNS Name *

Resource ID *

Ingresamos al Almacén de llaves, seleccionamos Propiedades y copiamos los valores de la URL de almacén e ID del recurso.



Pegamos los valores copiados en el respectivo campo de la creación del Scope.

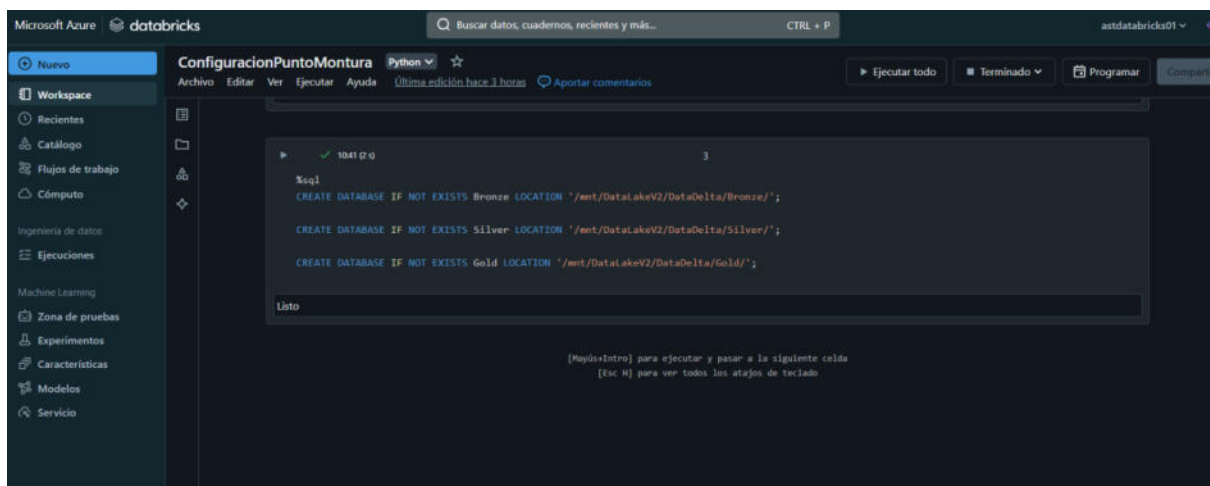


Creación punto de montura hacia la cuenta de almacenamiento en Azure Databricks

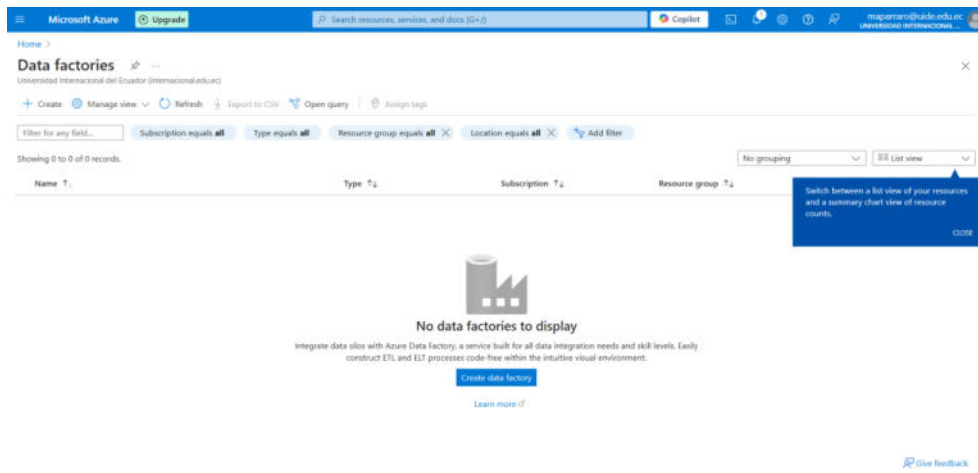
Una vez configurado el scope, encendemos el cluster y utilizamos el código a continuación para la creación del punto de montaje. En el mismo se debe reemplazar los valores de:

- ClientId y ClientEndpoint de la Entidad de Servicio
- Scope y Secreto del Almacén de llaves
- Nombre contenedor de la Cuenta de Almacenamiento

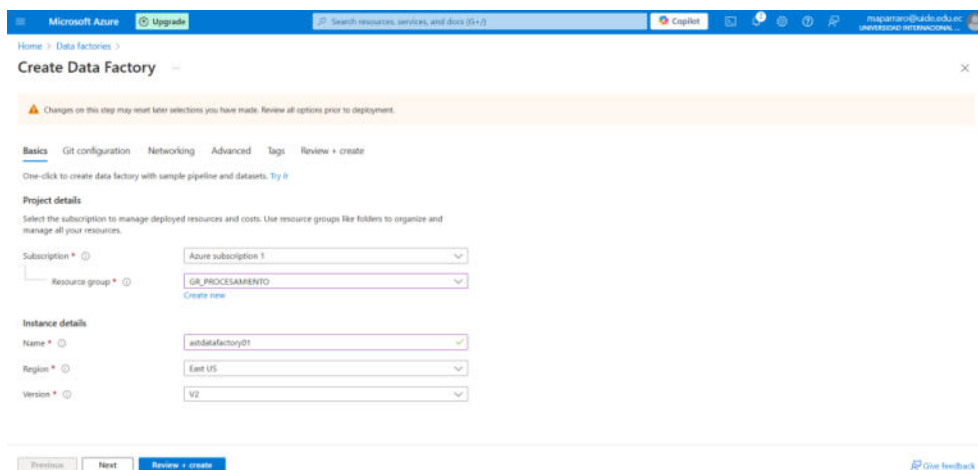
Una vez creado el punto de montura utilizamos el código a continuación para la creación de las bases Delta Lake (Bronce, Plata y Oro).



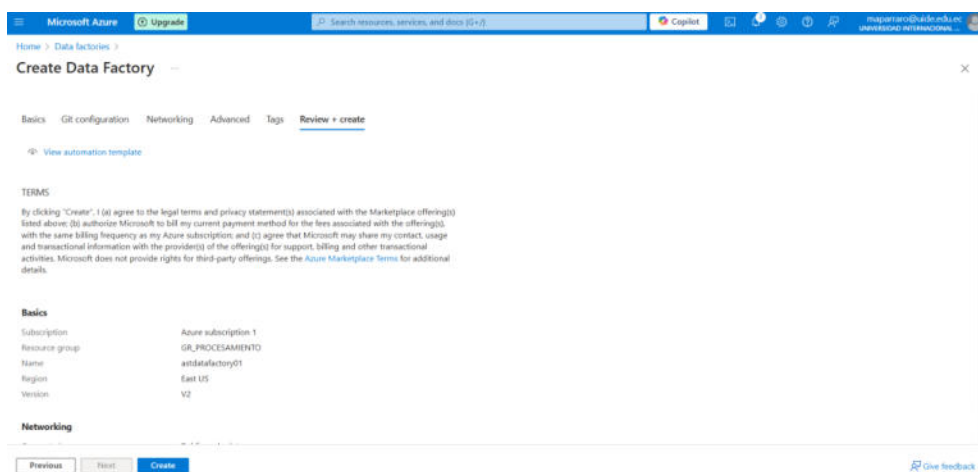
Seleccionamos Azure DataFactory y posterior Crear un nuevo Azure DataFactory.



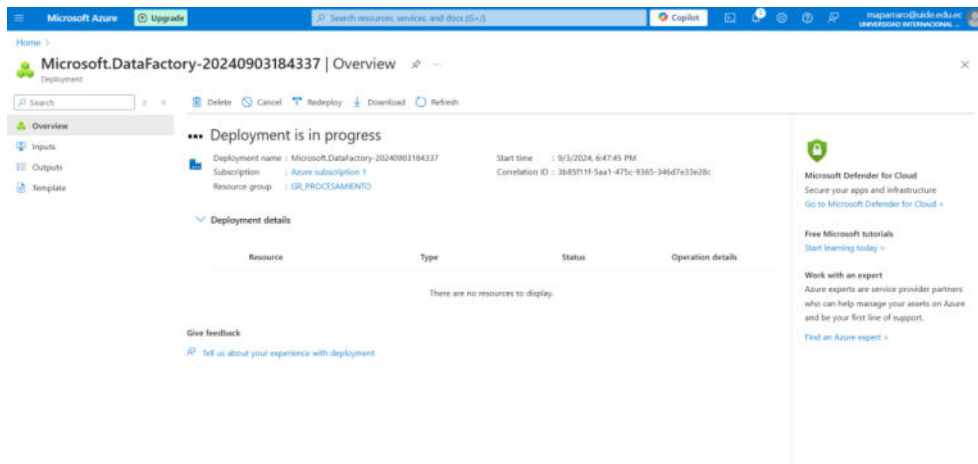
Completamos los datos básicos solicitados en el formulario de creación.



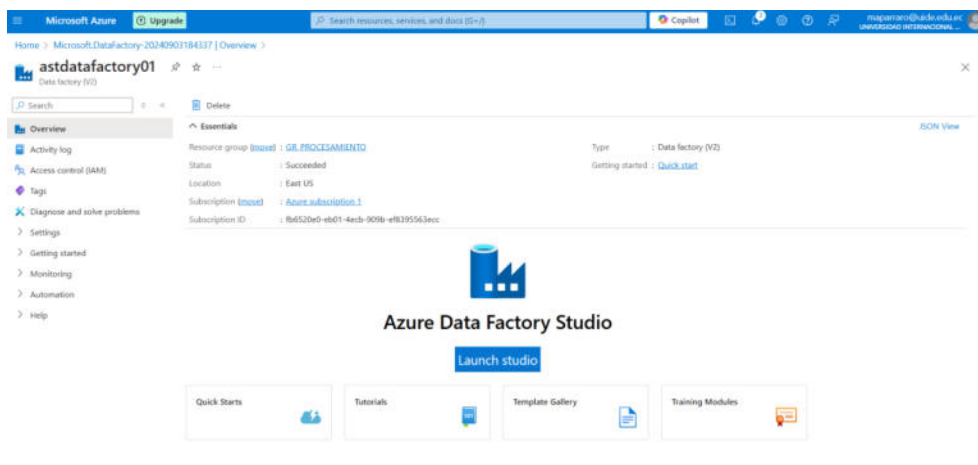
Una vez validados todos los campos seleccionamos Crear el recurso.



Se visualizará la pantalla de espera mientras el recurso es aprovisionado

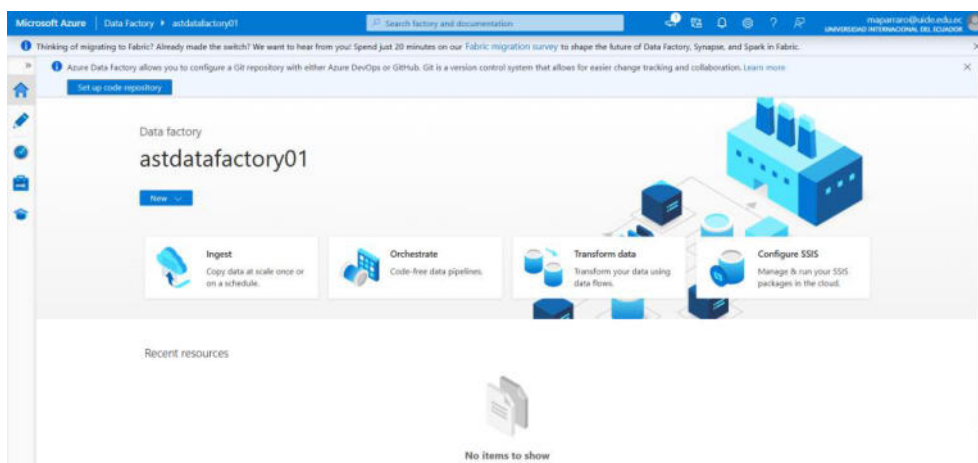


Completada la creación ingresamos al recurso

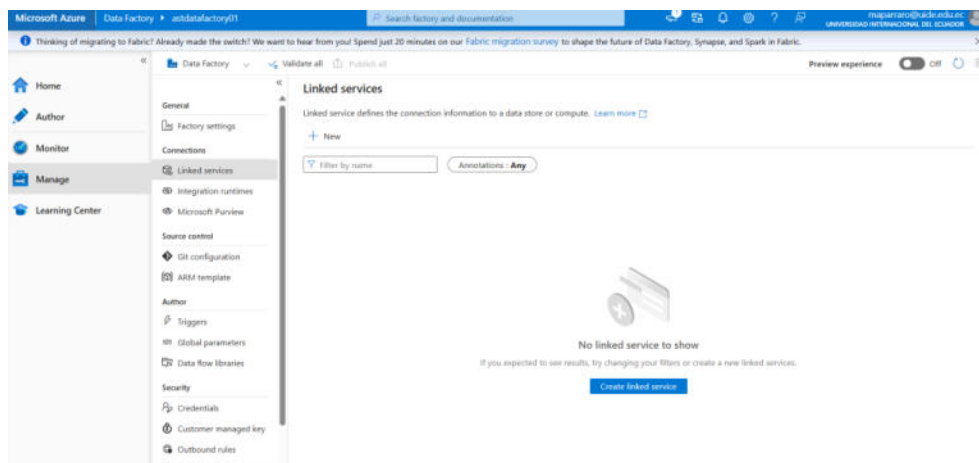


Creación DataSources en Azure DataFactory

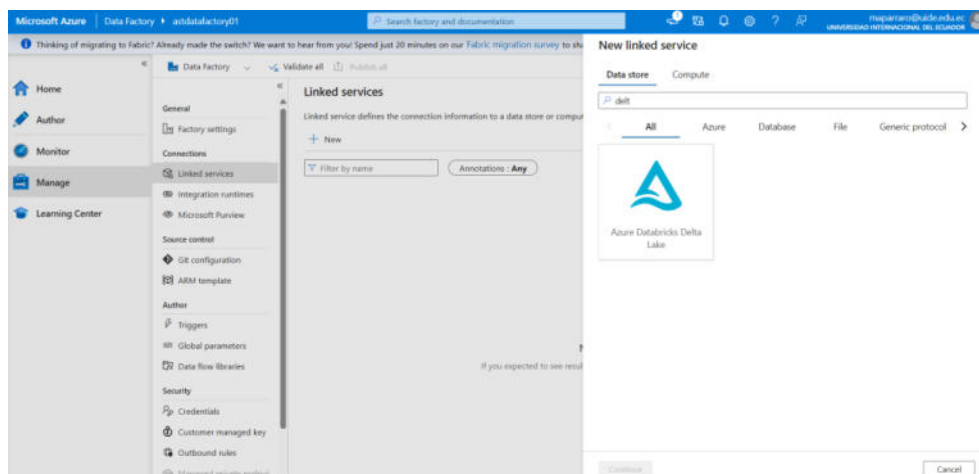
Ingresamos a la cuenta de Azure DataFactory.



Seleccionamos Manage, Linked Services, New



Para crear la conexión hacia tablas delta buscamos Azure Databricks Delta Lake y lo seleccionamos.



Completamos los datos del formulario, nombre del recurso databricks, id del cluster, y el token de acceso que debe ser generado desde la consola databricks.

Microsoft Azure | Data Factory | adlfactory01

Thinking of migrating to Fabric? Already made the switch? We want to hear from you! Spend just 20 minutes on our Fabric migration survey to all...

Home | Author | Monitor | Manage | Learning Center

General | Factory settings | Connections | Linked services | Integration runtimes | Microsoft Purview | Source control | Git configuration | ARM template | Author | Triggers | Global parameters | Data flow libraries | Security | Credentials | Customer managed key | Outbound rules | Managed private endpoint

Linked services

Linked service defines the connection information to a data store or computer.

+ New

Filter by name | Annotations: Any

New linked service

Azure Databricks Delta Lake | Learn more

Name * | adl_delta

Description

Connect via integration runtime * | AutolinkedIntegrationRuntime

Authentication method * | Access token

Account selection method | From Azure subscription | Enter manually

Domain * | https://adb-1233377418658007.7.azure.databricks.net/

Existing cluster ID * | D9D3-234227-48857a2

Access token | Access Key Vault

Access token * | *****

Connection successful

Create | Back | Test connection | Cancel

Para crear una conexión hacia Databricks para procesamiento batch de notebooks buscamos Azure Databricks.

Microsoft Azure | Data Factory | adlfactory01

Thinking of migrating to Fabric? Already made the switch? We want to hear from you! Spend just 20 minutes on our Fabric migration survey to all...

Home | Author | Monitor | Manage | Learning Center

General | Factory settings | Connections | Linked services | Integration runtimes | Microsoft Purview | Source control | Git configuration | ARM template | Author | Triggers | Global parameters | Data flow libraries | Security | Credentials | Customer managed key | Outbound rules | Managed private endpoint

Linked services

Linked service defines the connection information to a data store or computer.

+ New

Filter by name | Annotations: Any

Showing 1 - 1 of 1 items

Name	Type
adl_delta	Azure Databricks Delta Lake

New linked service

Data store: Compute

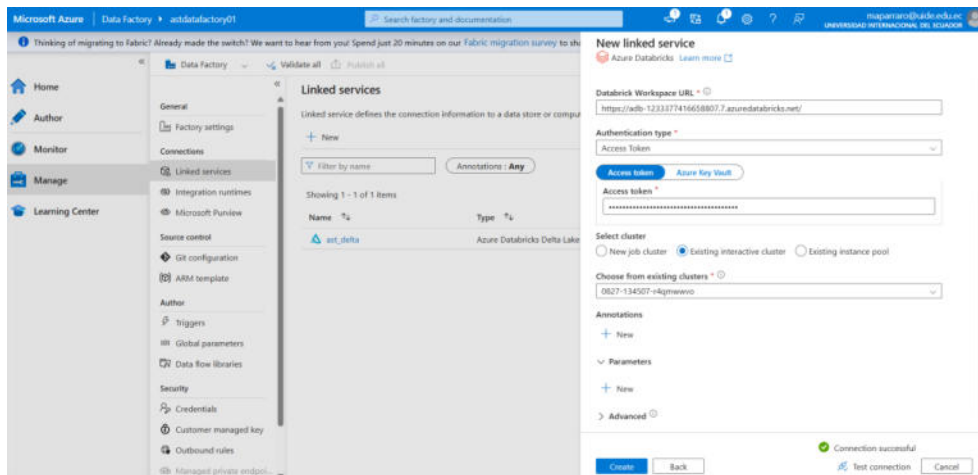
azure databricks

All | Azure | Compute

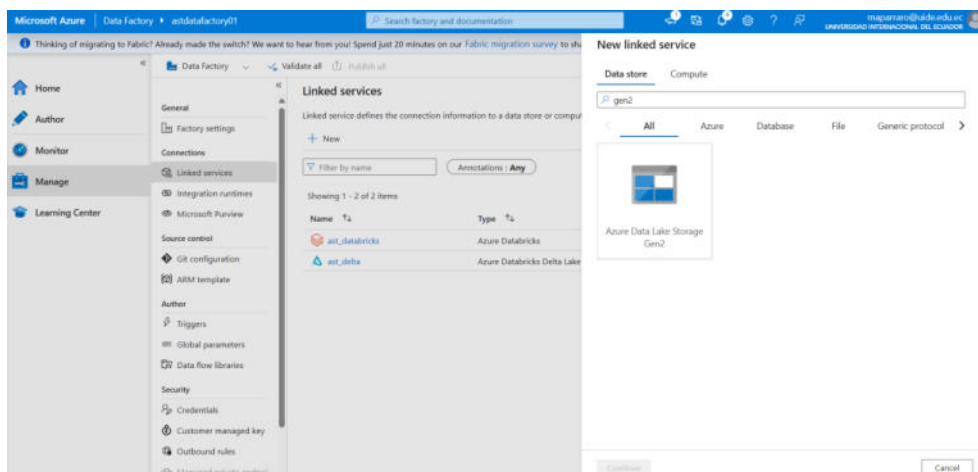
Azure Databricks

Cancel

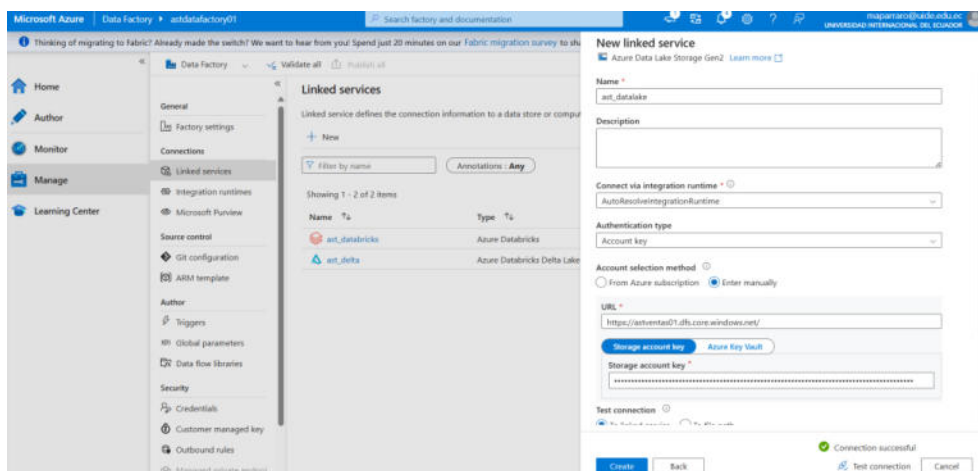
Completamos los datos del formulario, nombre del recurso databricks, id del cluster, y el token de acceso que debe ser generado desde la consola databricks.



Para crear una conexión hacia Azure DataLake para leer archivos planos buscando una conexión Azure Data Lake Gen2

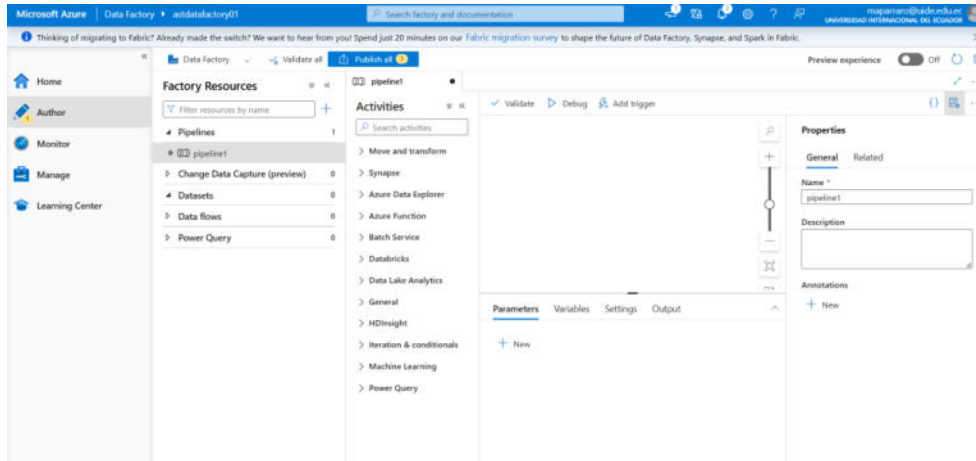


Completamos los datos del formulario, nombre del recurso cuenta almacenamiento y su clave de cuenta

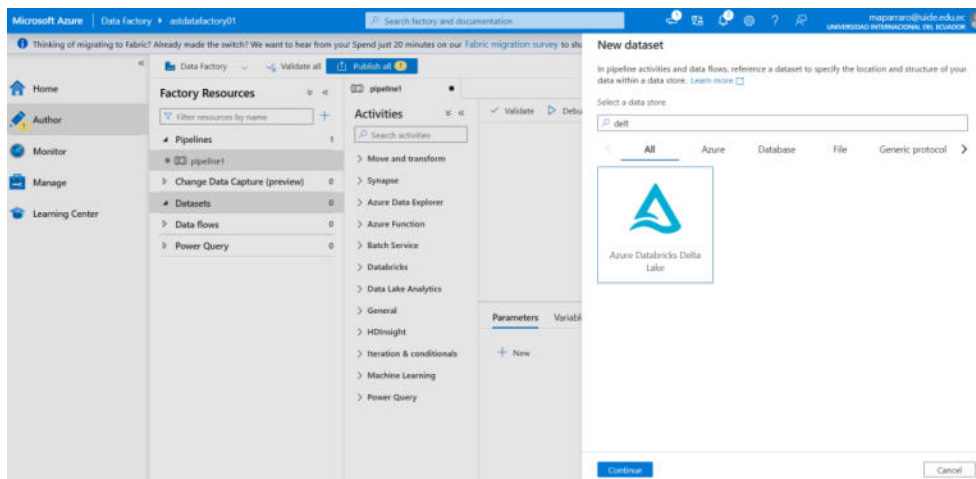


Creación DataSets en Azure DataFactory

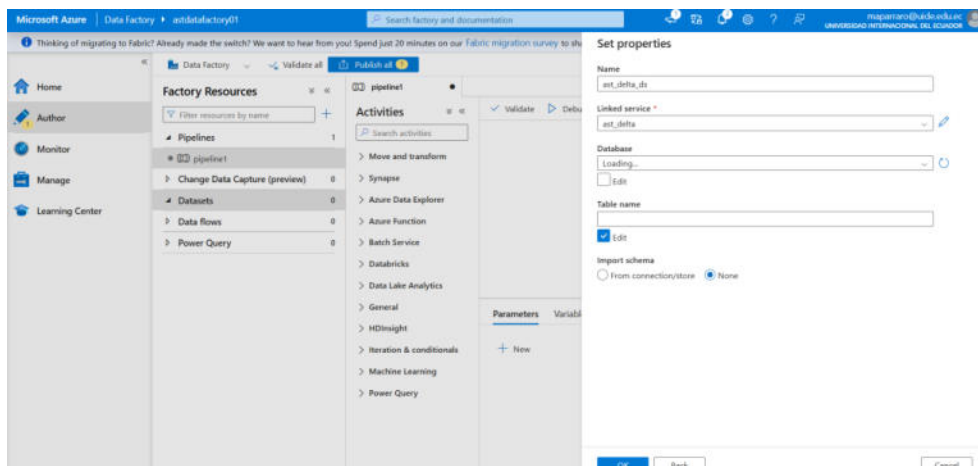
Creamos un nuevo pipeline y se nos habilitará la sección de Datasets



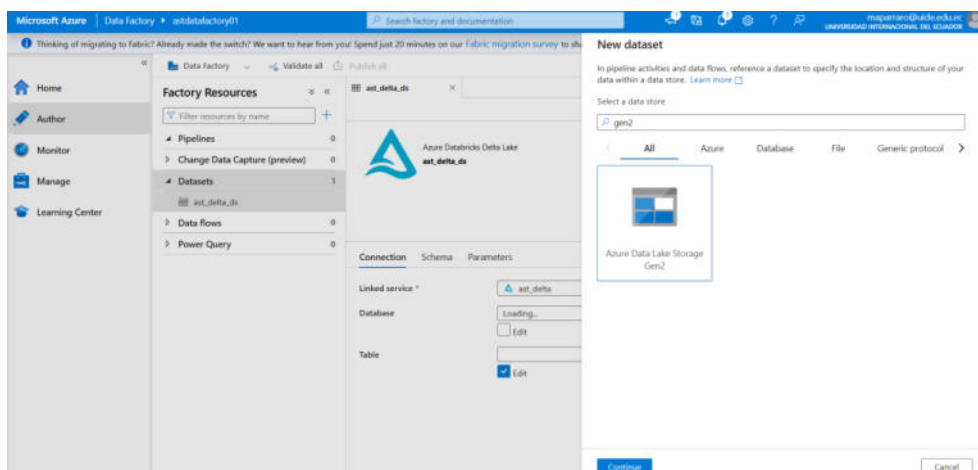
Para crear un dataset hacia delta buscamos Azure Databricks Delta Lake.



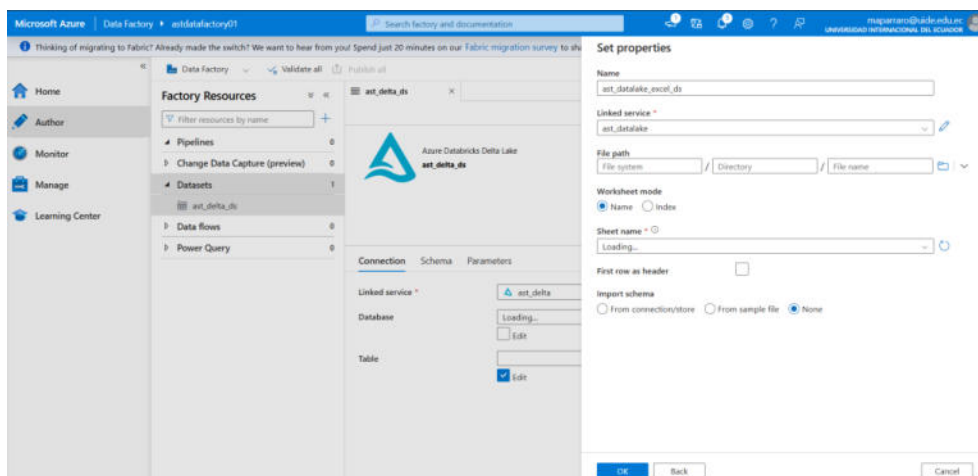
Damos un nombre y seleccionamos el DataSource para delta creado anteriormente.



Para crear un dataset hacia cuenta almacenamiento buscamos Azure Data Lake Gen2



Damos un nombre y seleccionamos el DataSource para datalake creado anteriormente



REPOSITORIO GITHUB

El código e insumos generados para el proyecto se encuentran en el repositorio de GitHub en la siguiente dirección:

[GitHub - BonecitaCP/SeriesTemporales_Grupo4_UIDE: Análisis de datos sintéticos de ventas de productos de redes tecnológicas y comparación de dos Modelos de Series Temporales que permitan predecir su demanda](#)