

Desarrollo de Modelo de Predicción de Costos para Mineras

Por

Matias Montenegro Reyes

Proyecto de pasantía
presentado a la Universidad Adolfo Ibáñez
en cumplimiento del requisito de
proyecto de pasantía para la obtención del título de
Ingeniero Civil Industrial

Santiago, Región Metropolitana, Chile, 2023

Declaración del Autor

Por la presente declaro que soy el único autor de este proyecto de pasantía. Esta es una copia fiel del proyecto de pasantía, incluidas las revisiones finales requeridas, tal y como fue aceptada por mis examinadores.

Entiendo que mi tesis puede ponerse a disposición del público por medios electrónicos.

Resumen Ejecutivo

Este documento presenta una propuesta de proyecto de pasantía destinado a la empresa “HPR Soluciones Informáticas LTDA”, especializada en la venta de servicios tecnológicos dirigidos a la industria minera. Sus servicios abarcan desde el soporte en plataformas existentes, hasta el desarrollo de nuevos productos diseñados para abordar las necesidades de toma de decisiones administrativas a través de la inteligencia de negocios.

El proyecto que se propone tiene como objetivo la expansión de los servicios ofrecidos hacia la toma de decisiones relacionadas con los costos de estas operaciones, un campo en el que la empresa aún no ha incursionado. Para lograrlo, se plantea el desarrollo de un sistema capaz de generar proyecciones de costos con un nivel de certeza del 90 %, basado en la comparación con registros históricos. Estas proyecciones estarán destinadas a una o varias plantas mineras y deberán estar disponibles antes de enero de 2024.

La solución desarrollada considera un modelo de regresión lineal implementado con Python donde se entrena utilizando variables ampliamente contabilizadas en la minería, tales como los tonelajes de material extraído y los costos totales, los que ayudarán a predecir los costos operacionales y su fluctuación. Para el usuario de esta solución se proporciona una visualización de los datos proyectados en PDF y Power Bi, además de un archivo Excel que contiene todos los valores proyectados para su utilización.

Finalmente, las proyecciones realizadas con datos normalizados obtenidos de la mina “Los Bronces” de Anglo American consiguen un porcentaje de error de un 1,253 % para el caso del tonelaje y de un 2,758 % para el caso de los costos, lo que cumple con el objetivo general de una certeza mayor al 90 %.

Executive summary

This document presents a proposal for an internship project intended for the company “HPR Soluciones Informáticas LTDA”, specializing in the provision of technological services to the mining industry. Their services range from supporting existing platforms to developing new products designed to address administrative decision-making needs through business intelligence.

The proposed project aims to expand the services offered towards decision-making related to the costs of mining operations, an area in which the company has not yet ventured. To achieve this, the development of a system capable of generating cost projections with a 90% level of certainty, based on a comparison with historical records, is proposed. These projections will be intended for one or more mining plants and must be available before January 2024.

The developed solution considers a linear regression model implemented with Python, where it is trained using variables widely accounted for in mining, such as extracted material tonnages, and total costs, which will help predict operational costs and their fluctuations. For the user of this solution, a visualization of the projected data is provided in PDF and Power BI, in addition to an Excel file containing all the projected values for their use.

Finally, the projections made with normalized data obtained from the “Los Bronces” mine of Anglo American achieve an error percentage of 1.253% for the tonnage case and 2.758% for the cost case, fulfilling the overall objective of accuracy greater than 90%.

Tabla de Contenidos

Declaración del Autor	I
Resumen Ejecutivo	II
Executive summary	III
1. Introducción	1
1.1. Contexto de la Empresa	1
1.2. Área de Desarrollo del Proyecto	3
1.3. Oportunidad	3
2. Objetivos	5
2.1. Objetivo General	5
2.2. Objetivos Específicos	5
2.3. Métricas	6
3. Estado del Arte	7

4. Metodología y Solución	11
4.1. Metodología	11
4.2. Solución	13
5. Desarrollo y Evaluación Económica	17
5.1. Desarrollo	17
5.2. Análisis de Riesgo	19
5.3. Evaluación Económica	22
6. Resultados y Conclusiones	23
6.1. Resultados	23
6.2. Evaluación de Métricas	27
6.3. Conclusiones	28

Capítulo 1

Introducción

1.1. Contexto de la Empresa

HPR Soluciones Informáticas LTDA. es una empresa chilena fundada en el año 2006, siendo su área de especialización el campo de la consultoría informática, la optimización de procesos y la inteligencia de negocios. A lo largo de su historia, la empresa ha centrado su atención en brindar sus servicios a la industria minera, siendo “Anglo American” su cliente principal en este sector, manejando contratos con ellos desde hace más de 15 años, desarrollando proyectos en los cuatro yacimientos presentes en Chile. Además de Anglo American, cuenta con otros contratos de proyectos en minería, sin embargo, estos representan menos del 10 % de los ingresos de la empresa.

Para llevar a cabo sus operaciones y ofrecer sus servicios a sus clientes, HPR cuenta con un equipo de 22 empleados que se distribuyen en cinco departamentos distintos, estos son:

- Soporte TI: Este departamento se encarga de brindar soporte técnico en tecnologías

de la información para garantizar que los sistemas y las redes informáticas de los clientes funcionen de manera eficiente y segura.

- **Administración de Proyectos:** Gestiona la planificación y ejecución de los proyectos de la empresa, asegurando que se entreguen en tiempo y forma, y que cumplan con los objetivos y requerimientos establecidos.
- **Desarrollo:** Este equipo se enfoca en crear soluciones informáticas a medida para los clientes, adaptándose a sus necesidades específicas y garantizando un funcionamiento óptimo.
- **Reportabilidad:** Departamento encargado de la generación de informes y reportes, que son fundamentales para la toma de decisiones informadas en el ámbito de la inteligencia de negocios.
- **Recursos Humanos:** La gestión de recursos humanos es crucial para el reclutamiento, capacitación y desarrollo de los empleados de la empresa, contribuyendo a un equipo competente y motivado.

Además de estos departamentos, la empresa cuenta con la figura de un Gerente de Servicios y Proyectos, y un Gerente General. Su operación consiste en el desarrollo y/o mantenimiento de proyectos para áreas específicas de las minas, lo que va desde generar reportes diarios para la toma de decisiones en terreno a nuevas formas de calcular los gases emanados por las plantas, por ejemplo.

1.2. Área de Desarrollo del Proyecto

El proyecto se sitúa en el área de Reportabilidad de la empresa. Como se detalla en la sección 1.1, este departamento despliega sus operaciones en la generación de informes y reportes, estableciendo una conexión directa con las bases de datos de los clientes para recopilar los registros necesarios. El propósito es generar análisis, modelos y gráficos que relacionen de manera numérica la situación de la empresa.

Este sector opera con tres herramientas fundamentales: SQL Server, Python y Power BI, concentrándose en todo lo relacionado con la visualización para la inteligencia de negocios. En la actualidad, los proyectos se enfocan principalmente en actividades directamente vinculadas con la extracción y procesamiento de materiales excavados. En otras palabras, se desarrollan proyectos centrados exclusivamente en el monitoreo, extracción y procesamiento de materiales en una mina específica. Un ejemplo reciente aborda la visualización de los materiales extraídos en distintos polígonos, su flujo y porcentaje de cobre, la operación de camiones y excavadoras, la cantidad movida y la comparación de cada uno de estos valores con su plan para evaluar el desempeño en relación a las expectativas. En resumen, los proyectos hasta ahora se centran en la producción exclusivamente, pero podrían ser complementados con la vertiente financiera de los proyectos mineros.

1.3. Oportunidad

En vista de esto, la oportunidad que se presenta es la de emprender un proyecto que amplíe los contratos de la empresa hacia el aspecto administrativo de la minería. Existe la posibilidad de crear un nuevo sistema orientado a la toma de decisiones, teniendo en cuenta los costos incurridos en las operaciones mineras. Esta propuesta ayudaría a la empresa a

ampliar su cartera de productos para sus clientes, mejorando así la confianza en la toma de decisiones a nivel de las plantas mediante una expansión en la Inteligencia Empresarial.

La propuesta concreta implica desarrollar un modelo capaz de generar proyecciones de costos mensuales para diversos proyectos mineros. Este producto sería presentado como una herramienta valiosa para la toma de decisiones dentro de las minas, contribuyendo así a mejorar las operaciones. La implementación exitosa de este modelo podría ser percibida como un valor añadido por parte de la empresa, ofreciendo a los clientes una herramienta efectiva para la planificación y gestión de sus proyectos mineros.

Esta oportunidad tendría un impacto directo en los ingresos de la empresa, lo que resultaría en un aumento correspondiente a las horas de trabajo invertidas en el contrato, la duración de la adaptación del producto para la planta respectiva, los roles necesarios para la implementación, multiplicados por la cantidad de plantas que contratan el producto y sus requisitos específicos.

Capítulo 2

Objetivos

2.1. Objetivo General

El objetivo general de este proyecto es el de desarrollar un sistema de proyecciones de costos con una certeza del 90 %, por medio de comparación con registros anteriores, para una planta minera, antes de enero del 2024.

2.2. Objetivos Específicos

- Extraer y modificar datos utilizados actualmente en proyectos mineros con SQL y almacenarlos en nuevas bases de datos con fin exclusivo de análisis.
- Crear un modelo capaz de predecir los movimientos y costos de un mes futuro con los datos almacenados.
- Analizar y comparar las proyecciones con datos de los meses previos.

- Corregir el modelo hasta obtener el porcentaje de certeza objetivo.

2.3. Métricas

Para la medición de la eficiencia de la solución generada, se han propuesto las siguientes métricas:

- **Tiempo de extracción y modificación de los datos:** Al contabilizar el tiempo que toma esta tarea, se puede medir cuánto del tiempo total de desarrollo de la solución se lleva a cabo en esta etapa.
- **Error Porcentual Absoluto Medio (MAPE):** Al comparar los valores proyectados con los valores reales, podemos obtener un promedio de cercanía de los resultados para medir la eficacia del modelo.
- **Análisis de tendencias:** Observar gráficamente cómo se comporta la proyección en relación con los valores reales, con tal de evaluar su posible ajuste.
- **Conteo de iteraciones de ajustes:** Al contabilizar la cantidad de iteraciones que se realizan para ajustar el modelo al porcentaje deseado, se puede analizar la dificultad de proyección de los datos trabajados.

Capítulo 3

Estado del Arte

Dentro del mundo de la proyección de costos en la industria minera, fueron exploradas diversas soluciones que han sido implementadas a lo largo del tiempo. Este análisis crítico del estado del arte es fundamental para comprender la evolución y las limitaciones de las metodologías existentes en la proyección de costos operacionales. A lo largo de esta sección, se exploran tres enfoques principales: modelos basados en Machine Learning, regresiones lineales y modelos matemáticos complejos. Cada uno de ellos ofrece perspectivas únicas y desafíos específicos. A medida que se desentrañan las complejidades de estas soluciones, es esencial destacar cómo estas metodologías se alinean con las necesidades particulares de este proyecto en HPR Soluciones Informáticas LTDA. para mejorar la proyección de costos en la minería.

En primer lugar, se encuentran las soluciones basadas en “Machine Learning”; en segundo lugar, aquellas vinculadas directamente a regresiones lineales; y, finalmente, un tercer grupo relacionado con diversas formulaciones de resoluciones de problemas similares que abordan costos de operaciones. No obstante, la mayoría de estas proyecciones se centran

en cálculos previos al inicio de las operaciones mineras, siendo principalmente generadas para la evaluación de proyectos. La documentación específica orientada a la proyección de costos mensuales o anuales es limitada.

En la primera categoría, destaca la propuesta de Guo y Nguyen (2021), quienes crean un modelo de Machine Learning, particularmente con la creación de redes neuronales para la predicción de los costos totales de un proyecto minero, es decir, cuánto será el costo final desde que se retira la primera baldada hasta el cierre de la mina. Este enfoque, al ingresar un conjunto específico de variables con datos y combinaciones determinadas en la red neuronal, logra predecir los costos de operaciones de un proyecto minero con un margen absoluto de error del 7,770 % [Guo et al., 2021]. Lo importante de este modelo es que es posible de implementar en cualquier tipo de proyecto minero, tanto para minas subterráneas como a rajo abierto, siendo altamente adaptable con las variables a implementar y periodos a proyectar, sin embargo, aunque este modelo se posiciona como el más preciso hasta la fecha, su aplicación es costosa en términos de recursos humanos, tiempo de desarrollo e implementación, lo que lo hace poco viable para proyecciones mensuales, principalmente porque el desarrollo de las redes neuronales para un proyecto de esta envergadura, si bien podría llegar a tener un nivel bajo de error, aumenta considerablemente el tiempo de desarrollo y dificultad de implementación a corto plazo.

En contraste, las soluciones basadas en regresiones lineales son las más comúnmente empleadas en la industria minera, contando con modelos que han sido implementados durante más de 50 años. Ejemplificando este enfoque, el modelo propuesto por Nourali y Osanloo (2019) utiliza regresiones lineales múltiples, incorporando diversas variables previamente almacenadas en la mina para generar predicciones de la demanda y los costos tanto mensuales como anuales. Estas variables incluyen la producción anual del mineral, la cantidad extraída desechada anual, la producción anual a molienda, el grado de concentración, el

mineral en inventario y los gastos de capital [Nourali and Osanloo, 2019]. La aplicación de este método implica generar regresiones lineales a partir de los datos de cada variable y luego multiplicarlas por sus costos proyectados, generando así una proyección acorde a la certeza y cantidad de datos introducidos. Este acercamiento de solución permite desarrollar un modelo con la menor cantidad de información posible, llegando a tener resultados menores al 10 % de margen de error. Utiliza además variables que se encuentran dentro del alcance del proyecto, sin embargo, para asegurar una proyección mensual más acertada se debería aumentar la granularidad de los datos de anual a mensual.

La última categoría de soluciones involucra modelos matemáticos complejos, que utilizan una amplia gama de variables junto con diferentes formas de proyección. Un ejemplo relacionado es la propuesta en la tesis de magíster de Parra (2011), que plantea una función de costos operacionales basada en una función translogarítmica modificada [Parra Pizarro, 2011] para la proyección de un grupo de datos que se determinan como los principales para calcular los costos totales de una mina. Este modelo incorpora una mayor cantidad de variables que su contraparte original, proporcionando una proyección más precisa, pero a la vez más extensa. A diferencia de los enfoques anteriores, este modelo no se limita a costos operacionales, sino que abarca todos los costos presentes en una mina desde su base, lo que si bien podría servir para un enfoque posterior a la realización de este modelo, actualmente se busca realizar proyección exclusivamente a los costos operacionales. En conjunto con este ejemplo existen variados modelos matemáticos para la proyección de costos dentro de la literatura, sin embargo, su especificidad hace que la dificultad de desarrollo aumente, llegando a tener resultados similares a, por ejemplo, las regresiones lineales.

En resumen, los modelos implementados, según la revisión del estado del arte, tienden a carecer de robustez, siendo la regresión lineal múltiple el método más utilizado. Esto se debe

a que, independientemente del costo proyectado, las operaciones mineras tienden a mantener su forma de producción. Estos sistemas son particularmente útiles para la contabilidad en el eventual cierre de operaciones mineras, ya que la gerencia suele mostrar escaso interés en los costos mes a mes o año a año, considerándolos asumidos y llegando a promediar costos elevados en un 22 %, según lo proyectado entre 1980 y 2001 [Gypton, 2002].

En conclusión, la revisión exhaustiva del estado del arte ha proporcionado una visión integral de las soluciones existentes para la proyección de costos en la industria minera. Se observa que, si bien modelos basados en Machine Learning ofrecen precisión, su implementación puede ser costosa en términos de recursos y tiempo, lo que plantea desafíos para proyecciones mensuales. Por otro lado, los modelos basados en regresiones lineales, como el propuesto por Nourali y Osanloo, demuestran estabilidad y eficiencia en la proyección de costos operacionales, aunque requieren una mayor granularidad de datos para proyecciones mensuales. Además, los modelos matemáticos complejos, como el presentado por Parra, ofrecen una perspectiva completa pero pueden ser demasiado extensos para el alcance actual del proyecto.

Capítulo 4

Metodología y Solución

4.1. Metodología

La solución que se implementará se basará en la segunda categoría previamente descrita, enfocada en regresiones lineales. La elección de este enfoque se debe a que, debido a varios costos de implementación, resulta ser la opción menos engorrosa, tanto en tiempo como en recursos materiales. No obstante, se adaptará a las particularidades de cada caso, ya que el modelo que se desarrollará para HPR será de naturaleza general. Esto permitirá una rápida adaptabilidad para su aplicación en diversas minas.

La regresión lineal opera al analizar el comportamiento de un conjunto de datos, examinando si estos datos presentan algún tipo de secuencia, repetición, tendencia o estacionalidad. Una vez que estos datos son analizados, se busca proyectar información hacia el futuro, teniendo en cuenta todas las variables previamente descritas. La precisión de los valores proyectados tiende a mejorar a medida que se cuenta con más datos y se pueden interpretar patrones de manera más precisa.

En el caso de esta solución, las variables seleccionadas son: movimiento total (toneladas) y costos (\$). El nivel de detalle con el que HPR puede obtener información varía según cada mina y contrato. Sin embargo, en el caso de las toneladas, se puede disponer de una base de datos que registre la extracción tonelada por tonelada, lo que proporciona miles de datos en un mes.

En contraste, los datos relacionados con los costos que ha manejado HPR suelen ser totales mensuales. La granularidad de estos datos es mensual, lo que implica que la cantidad de datos disponibles para una regresión lineal es menor. No obstante, es posible proyectarlos de manera efectiva. Una interpretación alternativa de estas variables es la relación entre los costos totales mensuales y las toneladas totales mensuales (\$/ton), es decir, el costo promedio de extraer una tonelada en un mes específico. Se optará por esta última interpretación en la solución, ya que, al proyectar las toneladas por día, al final del ejercicio para obtener el costo total operacional del mes proyectado, se debe realizar la siguiente operación:

$$C_t = T_p * CT_p$$

Donde:

- C_t : Costo total del mes proyectado en [\$].
- T_p : Toneladas totales del mes proyectado en [ton].
- CT_p : Costo por tonelada del mes proyectado en [\$/ton].

Trabajar con variables y llevar a cabo las limpiezas necesarias para obtener datos con la menor cantidad de ruido posible contribuye a obtener una regresión lineal más precisa. No obstante, es importante señalar que prever desviaciones considerables en la producción

debido a factores externos escapa del alcance del proyecto. La razón radica en la implicación de numerosas variables adicionales, lo que llevó al descarte de un modelo basado en redes neuronales debido a la complejidad adicional que involucraría su implementación.

En relación a la proyección de datos, es esencial destacar que esta se realiza en función de una serie temporal. En otras palabras, no se comparan entre sí, sino que representan extracciones de valores a lo largo del tiempo. Por ejemplo, al tener una base de datos con 1.000 registros de tonelaje, equivalente a un registro diario durante 2,7 años, se realiza un entrenamiento con un porcentaje de los valores totales. En este caso, se podría entrenar con el 80 % de los datos, es decir, 800 valores unitarios con sus respectivas series temporales. Estos valores son consecutivos, desde el día 1 hasta el día 800, y no se eligen de forma aleatoria.

Una vez entrenado el modelo con este porcentaje, se procede a proyectar una cantidad específica de valores, como por ejemplo, 200. Para evaluar su porcentaje de error, se comparan con los 200 valores restantes que no se utilizaron en el entrenamiento del modelo, es decir, el 20 % restante. De este modo, se determina la precisión del modelo entrenado. Aunque se espera que el porcentaje de error sea menor al 10 % según el objetivo general, un modelo de este tipo permite ajustar los valores significativamente hasta obtener un porcentaje que se acerque o sea inferior a este umbral.

4.2. Solución

La solución en sí funcionará mediante código Python, que se encargará de recopilar los datos, manipularlos, generar la regresión lineal, realizar la proyección y entregar los datos en un formato extraíble, en este caso, un archivo .xlsx que contendrá tanto los datos originales como los proyectados para su visualización. La segunda parte de la solución consiste en un

archivo en Power BI que facilitará la visualización de la proyección. Además, se creará un informe en formato PDF con las mismas características que el archivo .pbix, en caso de que este último falle.

El código operará de la siguiente manera: en primera instancia, se importarán las bibliotecas necesarias para cada etapa. A grandes rasgos, estas bibliotecas incluirán pyodbc para la conexión con los servidores, pandas para el manejo y almacenamiento de datos, sktime para la regresión lineal y proyección de datos, fpdf para la creación del informe en formato .pdf, y matplotlib para la creación de gráficos. Las dos últimas bibliotecas mencionadas han sido ampliamente utilizadas a lo largo de la carrera.

A continuación, se lleva a cabo el almacenamiento de las variables, las cuales consisten principalmente en las consultas a realizar en SQL para la obtención de los datos, junto con la información necesaria para la conexión con el servidor mediante la librería pyodbc. Con estas variables creadas, se procede a generar la conexión con el servidor. Posteriormente, se realiza la consulta mediante las funciones cursor, execute y fetchall de pyodbc. Los valores obtenidos son luego separados por columnas en distintas variables, teniendo así una variable para las fechas, otra para los tonelajes, y otra para el porcentaje de cobre.

Estas variables se combinan en un “Dataframe” que contiene la fecha y una de las variables mencionadas, ya sea toneladas o el porcentaje de cobre. En este punto, los datos son analizados y transformados para eliminar valores anómalos y ceros, que no son proyectados correctamente por la librería sktime. Con los datos ya transformados, se crea un conjunto de entrenamiento del modelo para evaluar la eficacia de la proyección, definiendo porcentajes específicos para el entrenamiento y la prueba del modelo. Estos porcentajes dependerán de cada variable, la cantidad de valores disponibles y la cantidad de valores que se necesitan proyectar.

Con el modelo ya entrenado, se realiza la proyección, que posteriormente se compara con los valores de prueba para analizar la eficacia del modelo. Se utiliza la función “mean_absolute_percentage_error” (MAPE) para obtener un porcentaje promedio absoluto de error entre los datos reales y los datos de prueba. Si este valor es menor al 10 %, los datos proyectados cumplen con el objetivo principal. En caso contrario, se debe seguir iterando y manipulando los datos hasta alcanzar el porcentaje deseado.

Este proceso se repite para cada una de las variables a analizar, lo que resulta en un modelo muy robusto para generar proyecciones con la especificidad que se desee. Por ejemplo, aunque actualmente se esté utilizando exclusivamente para el tonelaje total de operaciones de una mina, dado que no se está desarrollando para una mina en específico, este enfoque puede ser desglosado para analizar el tonelaje separado por flota de camiones que mueve las toneladas, o por las excavadoras que extraen estos tonelajes. Se pueden proyectar variables como las velocidades, las distancias recorridas, el tiempo de uso de los camiones, entre muchos otros. Esto podría eventualmente servir para ajustar los modelos a los planes específicos de cada mina.

Una vez analizadas todas las variables necesarias para modelar los costos proyectados, se almacenan en un archivo Excel, separando las tablas por hojas mediante la función de Pandas, “ExcelWriter”. Luego, se crean y guardan las visualizaciones necesarias para incorporarlas en el informe en PDF con Matplotlib, ajustando la estética mediante diversas funciones presentes en la librería, pudiendo así ser representadas gráficamente a gusto del cliente. Finalmente, todo se incorpora dentro del informe.

En última instancia, los valores almacenados en el archivo .xlsx son extraídos por un archivo .pbix previamente diseñado. Este archivo proyectará visualmente el costo total mensual proyectado en comparación con el del mes anterior real, junto con gráficas que muestran la evolución de los tonelajes proyectados a lo largo de los días del mes. Adicional-

mente, incluye una barra de una variable móvil que abarca un rango probable de valores de dólares y/o libras esterlinas para la evaluación del cambio con la moneda seleccionada.

Desde el punto de vista del usuario, en caso de ser controlado por el cliente, solo deberá ejecutar el código cuando desee realizar una proyección. Este código reemplazará todos los archivos creados con los valores actualizados, proporcionando así una proyección actualizada cada vez que sea necesario.

Capítulo 5

Desarrollo y Evaluación Económica

5.1. Desarrollo

Para el desarrollo de la solución, al no contar con una base de datos (DB) disponible en un proyecto que requiera esta solución, se crea una en un servidor local con el fin de tener una respuesta más rápida y mayor privacidad en el uso de estos datos. Los datos dentro de esta nueva DB corresponden a información extraída de las operaciones de la mina “Los Bronces” de Anglo American, siendo normalizados para eliminar la posible fuga de información. Dentro de esta DB, se crean dos tablas: una que almacena la información diaria de tonelaje y porcentaje de cobre extraído, con registros desde el año 2019 (aproximadamente 1.400 registros), y otra que contiene los costos por tonelada mensual, también desde el año 2019. Sin embargo, estos últimos tienen una granularidad mensual, no diaria como en el caso de las toneladas, debido al alcance de los registros originales, siendo aproximadamente 48 filas de datos.

Una vez creada la DB, se procede a desarrollar el código Python, la base de la solución.

La razón por la cual se crea primero la DB antes que el código Python es para asegurar que la extracción de los datos con el código sea funcional. Es necesario tener una DB de prueba para comenzar a realizar las extracciones de datos y posteriormente las proyecciones.

Dentro del código, se ajustan los parámetros para la conexión y se analizan las mejores maneras de extraer y almacenar los datos. Para la extracción de los datos, al ser un servidor creado con “SQL Service Management Studio”, las consultas integradas dentro del código Python para la extracción de los datos deben estar escritas en el lenguaje de SQL Server. Estas consultas se utilizan en conjunto con la librería pyodbc, que genera una conexión con el servidor, ejecuta las consultas y recupera los datos, los cuales luego se almacenan por columna en distintas variables.

Luego, se procede al proceso de limpieza de los datos. Para ello, se analiza la distribución de los datos gráficamente utilizando un gráfico de dispersión y un gráfico de caja y bigotes. Estos gráficos permiten observar la cantidad de datos anómalos y en qué rangos se encuentran, los cuales serán posteriormente eliminados del modelo.

Una vez que los datos han sido procesados, se define el rango de datos de prueba y los datos que se utilizarán para entrenar el modelo. Este rango dependerá de la cantidad de datos disponibles y de la cantidad de datos que se requiere proyectar. Los datos proyectados se comparan con los datos reales utilizando la métrica MAPE, que proporciona información sobre el error de los datos.

En esta sección, se completa el desarrollo de la base para una de las variables y se inicia la iteración del proceso para ajustar el modelo y obtener mejores resultados. Sin embargo, al tener la base desarrollada, se repite el mismo procedimiento para todas las variables que sean necesarias.

Una vez proyectadas todas las variables necesarias, se procede a desarrollar el código

encargado de exportar esta información a un archivo Excel. Cada hoja del archivo contendrá una tabla distinta. Es crucial asegurarse de que el formato en que se están extrayendo los datos sea el deseado. En caso contrario, se deben realizar modificaciones en los Datasets donde se encuentran almacenadas las tablas hasta lograr el formato requerido.

Una vez completado este proceso, se procede a crear las visualizaciones que estarán presentes en el reporte en PDF. Estas gráficas deben ser almacenadas en la carpeta donde se encuentra el código, para luego ser llamadas por la función encargada de generar el reporte. En esta etapa, las gráficas son desarrolladas utilizando la librería matplotlib, aprovechando sus diversas funciones para ajustar la visualización según las preferencias del cliente.

Como se mencionó, el siguiente paso consiste en incorporar las gráficas creadas en el reporte PDF, mientras que los datos generados en Excel serán utilizados en el reporte de Power BI.

Con los datos de Excel ya creados, estos se importarán a Power BI. Dentro de Power BI, se generarán visualizaciones dinámicas que incluirán tanto los costos por tonelada como el costo total proyectado. Además, se presentarán las toneladas del mes anterior y las toneladas proyectadas para el mes siguiente. Se creará una variable dinámica que permitirá ajustar el valor del dólar o de la libra esterlina según la preferencia del usuario, facilitando la visualización de los costos convertidos a la moneda de elección, considerando las expectativas del cliente sobre las variaciones monetarias.

5.2. Análisis de Riesgo

Gracias a la naturaleza de la solución, los principales riesgos se centran en el comportamiento del código y los archivos asociados. Se realizó un desglose de los riesgos primarios

que podrían surgir durante la implementación de la solución, categorizándolos en una escala del 1 al 10 en tres variables:

- Posibilidad de ocurrir
- Impacto
- Dificultad de detección

Finalmente, se identifica el momento en que podría ocurrir el evento de riesgo. La siguiente matriz de riesgo muestra el detalle de cada punto.

Evento de Riesgo	Posibilidad de Ocurrir	Impacto	Dificultad de Detección	Cuando Podría Ocurrir
Problemas en el código	7	8	2	Al ejecutar el código
Archivos tipo .txt dañados	6	8	2	Al ejecutar el código
Problemas en reporte .pbix	3	6	1	Al abrir el archivo .pbix o al modificar los datos
Falla de conexión	4	2	4	Al ejecutar el código
Proyección fuera de rango	6	9	3	Al analizar los datos obtenidos
Usuario no capacitado para su uso	2	9	2	Al conversar con el cliente
Rechazo del cliente	5	10	7	Al ser presentado al cliente

Esta información finalmente ayuda a dilucidar que los problemas pueden ser evitables, siempre y cuando el producto final presentado a los clientes abarque todas estas posibles complicaciones, con el fin de reducir el riesgo al mínimo.

No obstante, el posible rechazo del cliente, o más bien, que ningún cliente adquiera este servicio, sí afectaría negativamente el proyecto. En tal caso, sería necesario buscar un enfoque completamente distinto para asegurar que el cliente quede satisfecho y que el servicio cumpla con sus expectativas.

5.3. Evaluación Económica

De acuerdo con los precios de venta de la empresa, siendo un producto desarrollado por un “Analista”, el valor de la hora hombre es de 1,1 UF. Se calcula un total de desarrollo de la solución en 400 horas, lo que se traduciría en 440 UF, equivalente a \$16.000.000, pesos chilenos, aproximadamente.

El plan de pago para este producto, así como su mantenimiento, será propuesto por la empresa y estará sujeto a negociación con el cliente.

Cantidad de Horas de Desarrollo	400
Precio HH Analista	1,1 UF
Total UF	440
Total CLP	16.060.299

Adicionalmente, al tratarse de un modelo novedoso con un enfoque hacia el área administrativa de las minas, HPR podría abrirse a un nuevo mercado que se centre en el análisis de datos relacionados con costos, aspectos financieros o, como en este caso, en análisis predictivos. Esto contribuiría a ampliar su gama de productos y fortalecer su posicionamiento en la industria de empresas que ofrecen servicios tecnológicos.

Capítulo 6

Resultados y Conclusiones

6.1. Resultados

El código final desarrollado se puede encontrar en el [6.3](#) y tiene el mismo comportamiento descrito anteriormente. Para la finalización de este se realizaron un total de 25 iteraciones con el objetivo de mejorar la calidad de los datos para su proyección, empleando un tiempo total de 30 horas dedicadas exclusivamente a esta fase, como se detalla en las métricas contabilizadas. A pesar de la inversión de tiempo, los resultados de las iteraciones resultaron beneficiosos. Inicialmente, la distribución de los valores del tonelaje se veía de la siguiente manera:

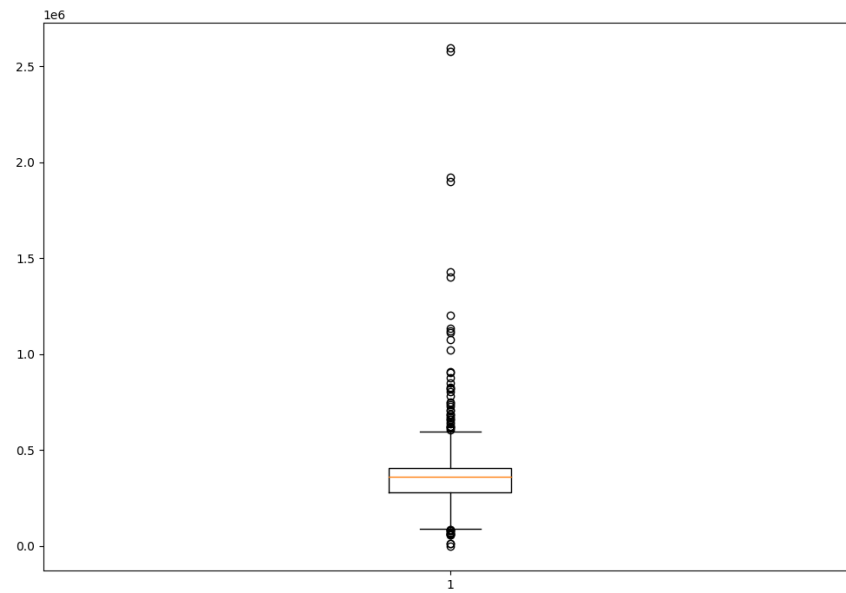


Figura 6.1: Diagrama de caja de valores originales del tonelaje.

Sin embargo, después de las modificaciones realizadas, que consistieron principalmente en cambiar la granularidad de los datos de mensuales a diarios y en extraer los datos atípicos, que se pueden apreciar en la figura 6.1, se observa la nueva distribución:

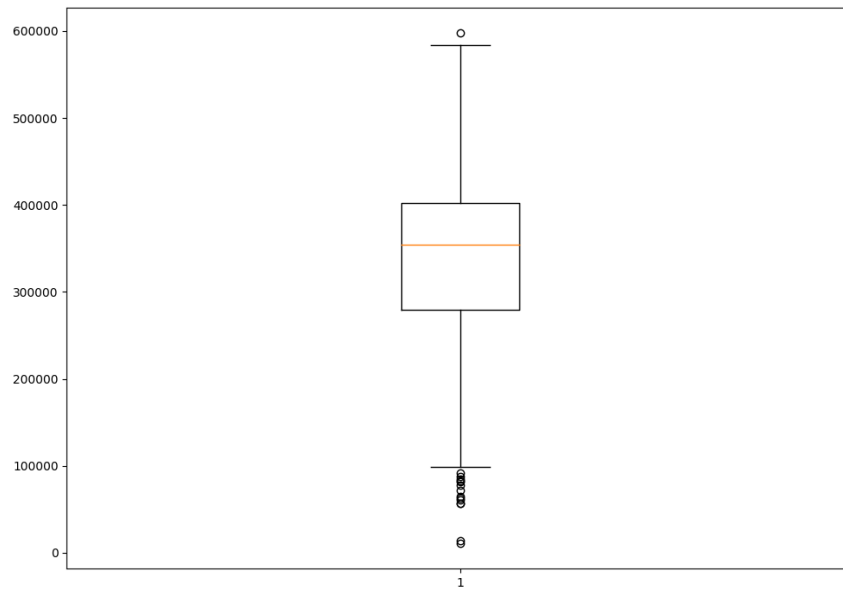


Figura 6.2: Diagrama de caja de valores originales del tonelaje.

El primer modelo proyectado, utilizando los valores de la figura 6.1, inicialmente mostró un porcentaje de error (MAPE) cercano al 300 %. Sin embargo, después de realizar modificaciones significativas en los datos, como se ilustra en la figura 6.2, este error se redujo notablemente a un 1,253 %. Este valor refleja la comparación entre los datos proyectados y los datos reales del mismo mes, específicamente para los tonelajes.


```

sum_real = tonnage_df["Tonnage"].sum()
sum_proy = tonnage_df_pred["Tonnage"].sum()
print("Promedio Tonelaje Mes Proyectado Real: ", "{:.2f}".format(sum_real))
print("Promedio Tonelaje Mes Proyectado Proyectado: ", "{:.2f}".format(sum_proy))
mape = (abs(sum_proy / sum_real) - 1)*100
print("MAPE: ", "{:.3f}".format(mape), "%")

```

[276] ✓ 0.0s Python

```

... Promedio Tonelaje Mes Proyectado Real: 6613088.73
Promedio Tonelaje Mes Proyectado Proyectado: 6695953.30
MAPE: 1.253 %

```

Figura 6.3: Captura de pantalla de resultados MAPE tonelaje.

Cuando se comparan los valores proyectados con los originales, como se muestra en la figura 6.3, se observa que el MAPE acumulado de la producción mensual es del 1,253 %, siendo este el valor de interés. Esto se debe a que, al analizarlo en una gráfica de tiempo, como se muestra en la figura 6.4, se percibe que la proyección diaria no es muy precisa. Sin embargo, para el alcance de este proyecto, esta imprecisión no resulta significativa.

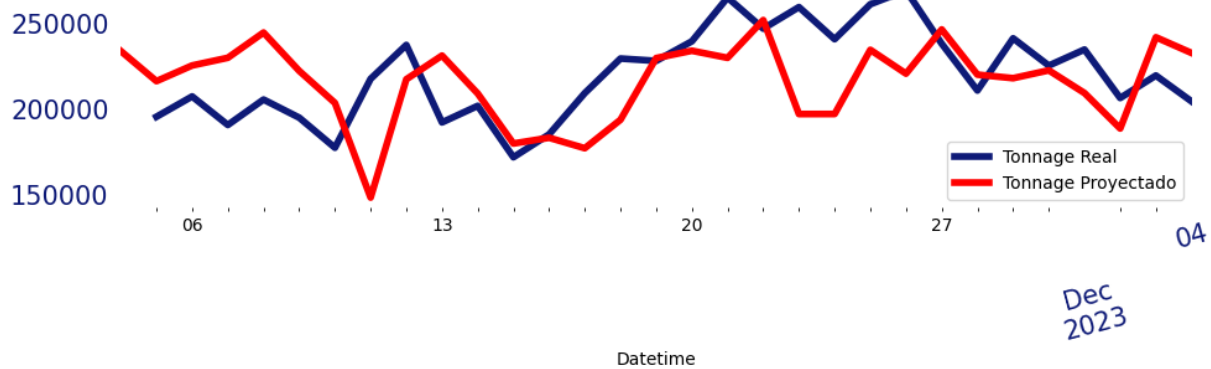


Figura 6.4: Gráfica de comparación tonelaje original con proyectado.

Ahora bien, como se mencionó anteriormente, estos resultados se refieren específicamente a las proyecciones de los tonelajes. En cuanto a los costos, se obtuvieron resultados similares, alcanzando un MAPE del 2,758 %, como se detalla en el Anexo 6.3, el cual muestra una comparación de los costos en el mismo mes y su MAPE correspondiente. Estos valores, conforme a la fórmula presentada en 4.1, son luego multiplicados para observar los costos totales proyectados, que representan la suma de sus MAPE. El MAPE total de la solución es entonces de 4,011 %, lo que corresponde a un margen de error menor al 10 % propuesto en el objetivo principal de este proyecto, logrando así a cabalidad los resultados esperados. Este porcentaje de error indica que, en un mes proyectado, si la proyección fue de \$100.000, por ejemplo, se puede esperar una variación de un $\pm 4,011$ %, lo que se traduce en $\pm \$4.011$ en el valor real.

Para la visualización de estos resultados de parte del cliente, se desarrolló un reporte simple en Power BI el cual contiene gráficas similares y los resultados esperados, el cuál se puede observar en el Anexo 6.3 en formato de captura de pantalla del reporte, pudiendo así observar los detalles presentes en la visualización.

6.2. Evaluación de Métricas

Como se mencionaba, el porcentaje de certeza esperado era de un 90 %, sin embargo, luego de la realización del proyecto, este valor llegó a ser de un 95,989 %, cumpliendo así el objetivo general de este proyecto. Luego, al ir comparando las métricas una a una obtenemos lo siguiente:

- **Tiempo de extracción y modificación de los datos:** El tiempo total dedicado a esta tarea fue de 30 horas, lo que representa una proporción baja en relación al

tiempo total de desarrollo del proyecto. Esta métrica sugiere que el proyecto es viable, ya que no consume una parte significativa del tiempo total.

- **Error Porcentual Absoluto Medio (MAPE):** En el caso de los tonelajes, se obtuvo un MAPE del 1,253 %, mientras que para los costos fue del 2,758 %. Ambos valores son bajos, lo cual es positivo. En la solución final, estos porcentajes se suman para obtener un MAPE total de 4,011 %, considerablemente por debajo del 10 % originalmente propuesto.
- **Análisis de tendencias:** El análisis de tendencias mediante la observación de los gráficos, tanto de diagramas de caja como de series de tiempo, permitió identificar la distribución de los datos. Gracias a esta observación, se realizaron modificaciones y ajustes que contribuyeron a reducir el porcentaje de error de aproximadamente un 300 % en las primeras iteraciones a un 1,253 %.
- **Conteo de iteraciones de ajustes:** Se llevaron a cabo un total de 25 iteraciones desde la primera proyección. Estas iteraciones fueron fundamentales para la reducción significativa del porcentaje de error, ya que en cada iteración se pudo identificar y aplicar ajustes mejorados al modelo.

6.3. Conclusiones

Para concluir, la realización de proyecciones de datos para la minería se revela como un proceso complejo que exige amplios conocimientos y dedicación. El cumplimiento del objetivo general de alcanzar un porcentaje de certeza superior al 90 % se atribuye en gran medida a la naturaleza estática de los datos utilizados. El próximo reto crucial implica la automatización del proceso de limpieza de datos, eliminando la necesidad de ajustes

manuales y asegurando la adaptabilidad del modelo a medida que se incorporan nuevos datos a las bases de datos.

Se reconoce la necesidad de desarrollar un modelo dinámico para mantener un margen de error consistente mes a mes, permitiendo que la empresa evolucione y genere proyecciones precisas con la continua adición de nuevos registros. Este enfoque se postula como esencial para garantizar la efectividad a largo plazo de la solución. Se podría lograr incorporando una mayor cantidad de datos a proyectar, como el grado de cobre y la proyección de movimiento por equipo, tanto de excavadoras como de camiones. Esto permitiría evaluar también la incorporación de nueva maquinaria en la mina y ajustar el modelo a medida que la operación se expande y se introduce mayor complejidad.

Sin duda, una mejora continua en la solución propuesta podría proporcionar herramientas valiosas para la toma de decisiones en proyectos mineros. Permitiría mejorar la planificación de la extracción, generar planes más realistas y aumentar la certeza en su cumplimiento. Para HPR, contar con un producto de este tipo podría mejorar su posición en la industria, atrayendo nuevos contratos y diversificando su cartera de clientes. La capacidad de ofrecer proyecciones precisas y ajustadas a la realidad puede ser un diferenciador significativo en un sector tan estratégico como la minería.

El aprendizaje adquirido durante el desarrollo del proyecto proporciona perspectivas valiosas sobre métodos adicionales para mejorar continuamente el rendimiento del modelo. La presentación del proyecto fue bien recibida por la empresa, subrayando su capacidad de adaptación. Aunque concebida como una solución generalizada para una oportunidad específica, se destaca su versatilidad para ser adaptada y presentada a los clientes, permitiendo ajustes en las proyecciones según las necesidades particulares en el momento requerido.

Bibliografía

- [Guo et al., 2021] Guo, H., Nguyen, H., Vu, D.-A., and Bui, X.-N. (2021). Forecasting mining capital cost for open-pit mining projects based on artificial neural network approach. *Resources Policy*, 74:101474.
- [Gypton, 2002] Gypton, C. (2002). How have we done? *Engineering and Mining Journal*, 203(1):40.
- [Nourali and Osanloo, 2019] Nourali, H. and Osanloo, M. (2019). Mining capital cost estimation using support vector regression (svr). *Resources policy*, 62:527–540.
- [Parra Pizarro, 2011] Parra Pizarro, A. L. (2011). Construcción de una función de costos operacionales para producción de cobre.

Anexos

Anexo A. Sección de Código

```
# Creación de Variables
consulta_tonnage = "SELECT * FROM Dump_Old"
consulta_costos = "SELECT Cast(Concat([Date], '-', '01')
as Date) as Date, CostPerTon FROM Costs"

conn_str = "DRIVER={ODBC Driver 17 for SQL Server};
SERVER= (localdb)\localdb;
Database=Storage; Trusted_Connection=yes;"

forecaster2 = Prophet()
horizon = 1 # Dias

# Conexión con el servidor y obtención de datos
try:
    connection = pyodbc.connect(conn_str)
    print("Conexión exitosa.")
    cursor = connection.cursor()
    cursor.execute(consulta_tonnage)
    records = cursor.fetchall()

# Almacenaje de datos en variables
```

```
Date = [record[0] for record in records]
Tonnage = [record[2] for record in records]
Distance = [record[3] for record in records]
LostTime = [record[4] for record in records]
except Exception as ex:
    print("Error durante la conexión: {}".format(ex))
finally:
    connection.close() # Se cierra la conexión a la BD.
    print("La conexión ha finalizado.")

# Creación del Dataframe y guardado de valores originales en .csv
tonnage_df = pd.DataFrame(list(zip(Date, Tonnage)),
columns=["Datetime", "Tonnage"])

tonnage_df["Datetime"] = pd.to_datetime(
    tonnage_df["Datetime"]
) # Cambio de formato de columna Datetime del DF a datetime

tonnage_df = tonnage_df.set_index(
    pd.DatetimeIndex(tonnage_df["Datetime"])
) # Definir la columna Datetime como Index

tonnage_df["Tonnage"] = tonnage_df["Tonnage"].astype(
```

```
float
) # cambio de formato de columna Tonnage

tonnage_df = tonnage_df.drop(
    ["Datetime"], axis=1
) # Eliminación de columna Datetime del df, queda solamente
como Index

tonnage_df = tonnage_df.resample(
    rule="D"
).sum() # Agrupación de los datos en "D" (días), siendo sumados
en la agrupación

tonnage_df["Tonnage"] = tonnage_df["Tonnage"].replace(
    0, tonnage_df["Tonnage"].mean()
) # Manipulación de datos, se cambian los datos 0, que no se
pueden procesar en la regresión, por el promedio
de la columna Tonnage

tonnage_df.where(
    tonnage_df <= 600000, tonnage_df["Tonnage"].mean(),
    inplace=True
) # Manipulación de datos, los datos más grandes de lo
permitido (errores) sean reemplazados por el promedio de la
columna Tonnage
```

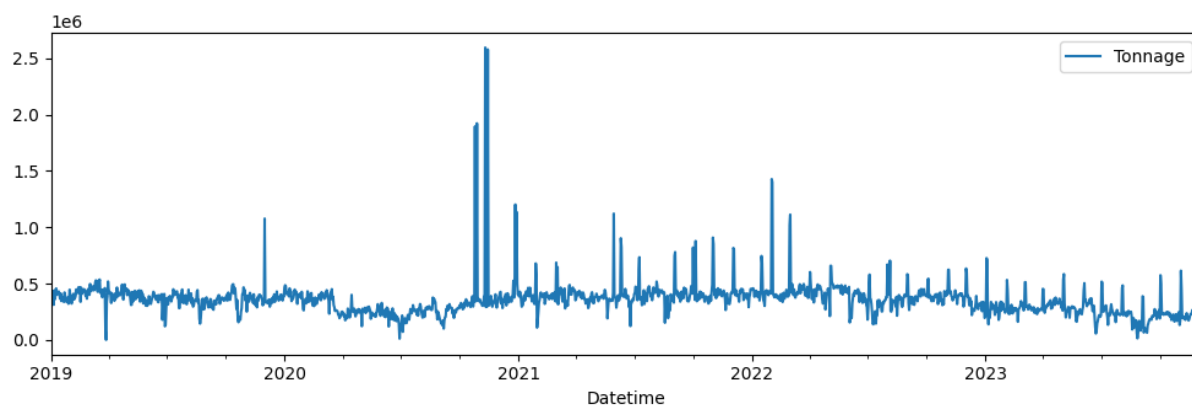


```
tonnage_df_train , tonnage_df_test = temporal_train_test_split(  
    tonnage_df, train_size=0.983  
) # Definición de las variables de prueba y de entrenamiento  
con su margen de entrenamiento  
fh = ForecastingHorizon(  
    tonnage_df_test.index, is_relative=False  
) # Se define el Horizonte de proyección  
  
forecaster = ThetaForecaster(  
    sp=720  
) # Se crea esta variable en base a la función ThetaForecaster ,  
con un largo definido de acuerdo a la  
cantidad de datos a procesar  
  
forecaster.fit(tonnage_df_train) # Se entrena la variable  
  
tonnage_df_pred = forecaster.predict(  
    fh  
) # Se predicen los datos con los valores entrenados  
  
print(  
    "Porcentaje de error absoluto tonelaje: ",  
    mean_absolute_percentage_error(tonnage_df_test ,  
    tonnage_df_pred) * 100,
```

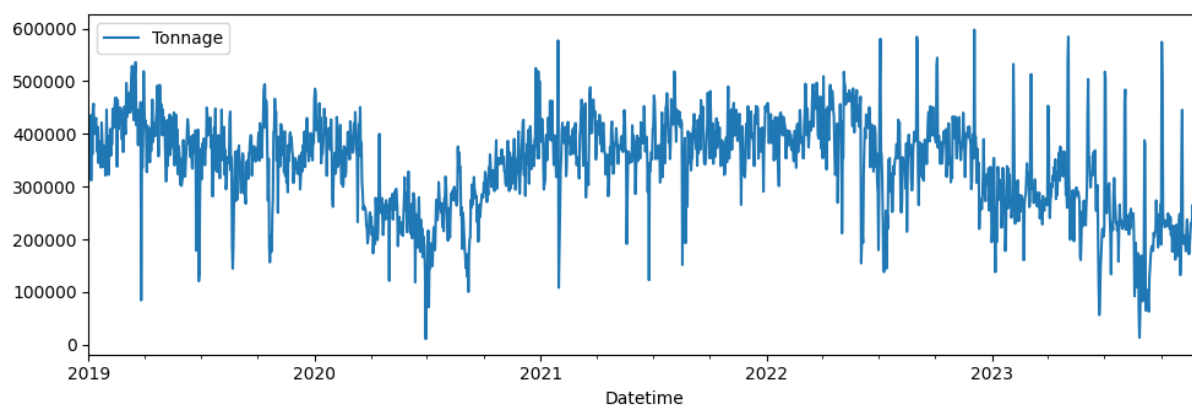
” 0%”,

)

Anexo B. Gráfica Tonelajes Originales



Anexo C. Gráfica Tonelajes Originales Modificados



Anexo D. Captura Pantalla Resultados MAPE Costos

```
cost_df = cost_df.tail(1)
print(cost_df)
print(cost_df_pred)
print("MAPE Costos: ", "{:.3f}".format(mape_cost), "%")
```

[277] ✓ 0.0s Python

...	Datetime	Costs
58	2023-11-30	3486.42499
	Date	Costs
0	2023-11-30	3582.588199
MAPE Costos:		2.758 %

Anexo E. Reporte Base Power Bi

