



**Informe Final**  
ING500: Pasantía Full-Time  
Ingeniería Industrial

Tomás Ignacio Feliú Jarpa [tfeliu@alumnos.uai.cl](mailto:tfeliu@alumnos.uai.cl)

10 de Diciembre de 2023

## Índice

1. Executive Summary:	3
2. Resumen ejecutivo	4
3. Contexto	5
4. Objetivos y Metodologías	6
4.1. Umbral de normalidad	6
4.2. Clasificación licitación	7
4.3. Sistema de alertas	7
4.4. Objetivo principal	8
5. Estado del arte y propuestas de solución	9
5.1. Regresión simple y multiple	11
5.2. CNN	12
5.3. Análisis estadístico	12
5.4. LSTM/Long short term memory	12
5.5. ARIMA/SARIMA	13
5.6. MLP	15
6. Plan implementación y Metodologia	16
7. Evaluación económica y medidas de desempeño	17
8. Análisis de riesgo	17
9. Desarrollo proyecto	19
10. Resultados	22
11. Conclusiones	25

## 1. Executive Summary:

This report provides a comprehensive overview of WherEX, identifying an area for improvement related to errors in offer prices or offered quantities in a small percentage of tenders. The project focuses on minimizing these errors and improving the accuracy of critical calculations for decision-making.

The current state of outlier detection in WherEX, using a z-score-based model, is discussed. However, limitations are acknowledged when certain requirements are not met, leading to the exploration of various solutions such as regression, CNN, statistical analysis, LSTM, ARIMA, and MLP. The potential of ARIMA and MLP is highlighted, with plans to enhance their implementation and explore combinations.

The economic evaluation reveals that the project does not seek direct economic benefits but will impact crucial financial areas such as commissions, Gross Transaction Value (GTV), and Gross Merchandise Value (GMV). The goal is to correct tenders awarded with incorrect data to minimize economic impacts in these three key areas.

Risk analysis identifies two main risks: project implementation and the proper functioning of the model. The risk of non-implementation is linked to supplier trust, while the risk of incorrect functioning is assessed through confusion matrices for atypical offers and real prices.

The chosen solution involves performing a robust z-score for each type of potential outlier, classifying them based on their characteristics. Initial results are positive, supporting the project's continuation and ongoing model improvement.

In conclusion, the report emphasizes the importance of WherEX leveraging the valuable historical dataset of the company. The project demonstrates a promising approach to enhancing outlier detection and precision in key metrics, suggesting a path towards continuous improvement and more informed decision-making in the future.

## 2. Resumen ejecutivo

El presente informe ofrece una visión completa de WherEX, identificando un área de mejora relacionada con errores en el precio de oferta o la cantidad ofertada en un pequeño porcentaje de licitaciones. El proyecto se centra en minimizar estos errores y mejorar la precisión de los cálculos cruciales para la toma de decisiones.

Se aborda el estado actual de la detección de outliers en WherEX, que utiliza un modelo basado en z-score. Sin embargo, se reconocen limitaciones cuando no se cumplen ciertos requisitos, lo que lleva a explorar varias soluciones, como regresión, CNN, análisis estadístico, LSTM, ARIMA y MLP. Se destaca el potencial de ARIMA y MLP, con planes para mejorar su implementación y explorar combinaciones.

La evaluación económica revela que el proyecto no busca beneficios económicos directos, pero afectará áreas financieras cruciales como comisiones, Gross Transaction Value (GTV) y Gross Merchandise Value (GMV). El objetivo es corregir licitaciones adjudicadas con datos incorrectos para minimizar impactos económicos en estos tres aspectos clave.

El análisis de riesgo identifica dos riesgos principales: la implementación del proyecto y el correcto funcionamiento del modelo. El riesgo de no implementar se vincula con la confianza de los proveedores, mientras que el riesgo de un funcionamiento incorrecto se evalúa mediante matrices de confusión para ofertas atípicas y precios reales.

La solución seleccionada consiste en realizar un z-score robusto específico para cada tipo de posible caso de outliers, clasificándolos según sus características. Los resultados iniciales son positivos, respaldando la continuación del proyecto y la mejora constante del modelo.

En conclusión, se destaca la importancia de que WherEX aproveche el valioso conjunto de datos históricos de la empresa. El proyecto muestra un enfoque prometedor para mejorar la detección de outliers y la precisión en las medidas clave, lo que sugiere un camino hacia la mejora continua y una toma de decisiones más informada en el futuro.

### 3. Contexto

WherEX, fundada en 2015 bajo el nombre de AQmarket es un Marketplace B2B, el cual se propuso como misión facilitar el acceso de grandes empresas de la industria acuícola (a las que llamaremos "compradores") a un mayor número de cotizaciones de las que solían recibir de un número limitado de empresas (denominadas "proveedores"). Su objetivo era fortalecer la conexión en el mercado. WherEX generó una competencia más saludable, permitiendo que los proveedores ofrecieran precios más competitivos frente a los productos demandados por los compradores, lo que resultó en una mayor visibilidad para las empresas compradoras que antes desconocían su existencia.

Este aumento de competencia, cotizaciones, precios más bajos y una mayor diversidad de proveedores permitió a las empresas compradoras reducir costos y obtener una comprensión más completa del mercado en relación con los productos necesarios para su funcionamiento. En resumen, AQmarket (ahora WherEX) abrió las puertas para que proveedores y compradores se conectaran por medio de su plataforma, brindando visibilidad a los primeros y ofreciendo ofertas más asequibles a los últimos debido al aumento de la competencia y las cotizaciones, en contraste con la limitación previa a 3-4 cotizaciones con los mismos proveedores de siempre.

Después del éxito en la industria acuícola, WherEX identificó la escalabilidad del modelo y decidió expandir su Marketplace B2B a otras industrias del mercado. En ese momento, se llevó a cabo una reforma de la imagen de la empresa, cambiando su nombre a WherEX y extendiendo su presencia a prácticamente todas las grandes industrias. Actualmente, WherEX cuenta con más de 150 clientes compradores en 4 países y cerca de 14,000 empresas proveedoras.

Después de la incorporación de la empresa a mediados del mes de julio, se identificó un problema dentro de WherEX, específicamente en un pequeño porcentaje de licitaciones cerradas. El propósito fundamental de este proyecto es reducir al máximo los errores presentes en las licitaciones de WherEX, ya que estos desencadenan una serie de cálculos inexactos que, en última instancia, llegan a los encargados de la toma de decisiones en la empresa, generando resultados que no son 100% exactos.

El problema identificado se concentra en inexactitudes en el precio de oferta o la cantidad ofertada en un pequeño porcentaje de las licitaciones de WherEX. Actualmente, estos errores, categorizados por un modelo de detección de valores atípicos, no son completamente identificados debido a la falta de utilización de información histórica. Las consecuencias de estos errores son significativas, ya que generan desfases entre lo solicitado, lo ofertado y lo adjudicado, afectando los cálculos de comisiones, GTV, GMV y los informes gerenciales.

El presente proyecto tiene como objetivo principal minimizar estos errores, mejorando la precisión de los cálculos críticos para la toma de decisiones. El informe detalla de manera exhaustiva los objetivos específicos y

la metodología implementada, que incluye el uso de Python y las bases de datos de la empresa. Estos objetivos conducen a la definición clara de un objetivo central. Además, se presentan medidas de desempeño relevantes para evaluar el éxito o fracaso del proyecto. La planificación a lo largo del semestre se expone en un diagrama de Gantt, destacando hitos importantes en el desarrollo del proyecto.

## **4. Objetivos y Metodologías**

El objetivo de este proyecto es crear un sistema de gestión de decisión que permita a wherEX involucrarse en aquellas licitaciones que presenten ofertas o datos atípicos de tal nivel que ameriten la acción por parte de los CSM (Customer success managers) para aclarar la licitación y la oferta en específico, corrigiendo el potencial error que pueda estar presente. A continuación, se pondrán en evidencia los distintos objetivos específicos que se necesitan lograr para concretar el proyecto declarado anteriormente.

### **4.1. Umbral de normalidad**

El primer objetivo se centra en realizar una proyección del precio que debería corresponder a un determinado producto, y a esta proyección la denominaremos el umbral "normal", considerando el producto, la categoría y la cantidad. Para calcular este umbral, se empleará un código en Python que utilizará un método de aprendizaje automático. Este código será capaz de analizar una extensa base de datos histórica de todas las licitaciones publicadas y las ofertas presentadas, almacenada en la base de datos denominada "sql", la cual cuenta con más de 3 millones de filas y aproximadamente 126 columnas.

Las ofertas se clasificarán por producto, donde se emplearan métodos predictivos que puedan entregar una proyección correcta y acertada del supuesto precio unitario de oferta que corresponde a la realidad. Es importante destacar la importancia de realizar una predicción que tenga en cuenta la variable del tiempo en las ofertas históricas, ya que un producto ofertado en 2015 probablemente tenga un precio muy diferente al de uno ofertado en 2023, por lo que ambos sirvan como información para la proyección, pero no tiene la misma relevancia o peso que una oferta mas reciente.

Aún no se ha determinado el método específico de aprendizaje automático que mejor se adapte a este proyecto en particular. Se buscará la ayuda de profesores de la universidad para recibir orientación sobre el tipo de código a crear y qué método de proyección proporcionará un umbral más realista dado el producto y su cantidad.

## 4.2. Clasificación licitación

La correcta clasificación de las ofertas que se desvían significativamente del umbral es un paso crítico para garantizar la eficacia y relevancia del sistema implementado. Destacaré la importancia de este proceso en el siguiente párrafo:

Una vez que el umbral "normal" ha sido definido, se revela la trascendencia de clasificar con precisión aquellas licitaciones que se desvían considerablemente de este umbral. Este paso no solo sirve como filtro para determinar cuáles ofertas deben ser sometidas a un análisis más profundo, sino que también asegura la integridad y validez de las alertas generadas. La clasificación precisa es esencial para evitar la generación innecesaria de alertas o, en el peor de los casos, la omisión de casos críticos.

La diferencia considerable respecto al umbral, especialmente en relación con la categoría del producto, sirve como indicador clave para la necesidad de un análisis más detallado. Esta fase es fundamental para comprender las razones detrás de la desviación y determinar si la oferta se considera errónea. La evaluación cuidadosa de estas ofertas fuera del rango establecido es crucial para asegurar que el sistema de alertas solo notifique a los Customer Success Managers (CSM) sobre situaciones que requieran su intervención y atención inmediata.

En resumen, la correcta clasificación de las ofertas que se alejan significativamente del umbral es un componente vital para la eficiencia operativa del sistema, permitiendo una gestión focalizada y efectiva de las alertas por parte de los CSM, quienes podrán abordar los casos críticos de manera oportuna y precisa.

## 4.3. Sistema de alertas

Como último objetivo específico, se implementará un sistema de alertas en Slack, que es el medio de comunicación utilizado por la empresa. Se creará un canal específico destinado a los Customer Success Managers (CSM). En este canal, se enviarán alertas únicamente para aquellos casos que realmente lo ameriten, con el fin de evitar generar un canal saturado de información innecesaria que podría dificultar la identificación de casos relevantes. Si esto llegara a suceder, no se cumpliría el propósito de notificar solo los casos de importancia.

En este canal de alertas, se espera destacar el elemento específico donde ocurre el potencial error (puede ser la cantidad, el precio, el tiempo, etc.), la empresa compradora involucrada y la empresa proveedora asociada. La responsabilidad de gestionar el caso o distribuirlo entre los CSM recaerá en ellos. Dependiendo del contenido de la alerta, podrán decidir tomar el caso por completo o asignarlo a la empresa proveedora correspondiente que requiere un seguimiento constante.

Es crucial destacar la importancia de los objetivos anteriores en este proceso: lograr un umbral correcto y, sobre todo, clasificar adecuadamente las licitaciones como correctas o incorrectas. Esto es esencial para evitar

alertar innecesariamente a los CSM o, en su defecto, garantizar que se les notifique cuando sea pertinente. La implementación de este sistema de alertas en Slack busca optimizar la eficiencia y la relevancia de la comunicación, asegurando que los CSM se centren únicamente en los casos críticos y pertinentes.

#### 4.4. Objetivo principal

Una vez identificados los objetivos específicos, la consolidación del objetivo principal es esencial para el éxito del proyecto. Como se mencionó anteriormente, el objetivo principal es establecer un sistema de gestión de decisiones capaz de analizar, identificar e informar sobre casos inusuales y alejados de la realidad en las licitaciones de wherEX. Este sistema busca obtener datos precisos y verídicos de cada licitación y oferta, garantizando así un cobro justo y correcto de comisiones a los proveedores. La importancia radica en evitar perjuicios tanto para los proveedores como para la confiabilidad de wherEX, ya que un cobro indebido podría afectar la relación con los proveedores o los ingresos de la plataforma.

Los puntos esenciales de este objetivo incluyen la proyección precisa, la identificación correcta y la notificación oportuna de casos relevantes. A continuación, se describe el objetivo general al estilo SMART:

1. **Specific** : Corregir licitaciones y ofertas erróneas presentes en la plataforma wherEX, mejorando el cálculo de distintas medidas como GMW, GTV y, en especial, la comisión real que corresponde cobrar.

2. **Measurable** : Actualmente, no se tiene claridad sobre el porcentaje de error existente ni en cuántos millones de pesos asciende. Este proyecto tiene como objetivo identificar y cuantificar estos errores, facilitando una medición precisa de la mejora realizada.

3. **Achievable** : Se logrará mediante el desarrollo de un código en Python que genere un sistema de gestión de decisiones capaz de identificar y alertar a los Customer Success Managers (CSM) sobre los casos realmente relevantes.

4. **Relevant** : La importancia de este proyecto radica en la obtención de datos más precisos y alineados con la realidad, ya que estos datos influyen en el cálculo de numerosas medidas que impactan directamente a los tomadores de decisiones. Además, garantiza el cobro justo y adecuado de comisiones a los proveedores.

5. **Time**: Se espera lograr el objetivo y completar el proyecto en aproximadamente 2 meses y medio, asegurando así una implementación oportuna de las mejoras planificadas.

En resumen, el objetivo principal del proyecto SMART es corregir, medir, alcanzar, relevante y con un plazo determinado, asegurando la eficacia y el impacto positivo en la gestión de decisiones en la plataforma wherEX.



## 5. Estado del arte y propuestas de solución

La comprensión detallada de la situación actual en cuanto a la detección de ofertas atípicas en wherEX es crucial antes de abordar el análisis del estado del arte en regresiones temporales y predicciones. La base de datos "sql" proporciona información vital sobre todas las ofertas realizadas en la historia de wherEX, siendo las columnas más relevantes para este proyecto las siguientes:

- **Licid:** Identificador de la licitación.
- **Licstatus:** Estado de la licitación, que puede ser Desierta, Adjudicada, En evaluación o Publicada.
- **Licfechacierre:** Fecha de cierre de la licitación.
- **Tipo:** Indica si el producto es un bien o servicio.
- **Product\_id:** Identificador único del producto para cada comprador.
- **Categoriaid:** Identificador de la categoría del producto.
- **Cantidadsolicitada:** Cantidad solicitada por el comprador del producto.
- **Cantidadofertada:** Cantidad ofertada por el proveedor del producto.
- **Cantidadadjudicada:** Cantidad adjudicada al proveedor del producto.
- **Preciounitarioof:** Precio unitario de la oferta. Para servicios, se divide en cantidadadjudicada \* duración.
- **Checkz:** Determina si la oferta se considerará en los cálculos, siendo -1 si no y 0 si sí. Las condiciones para considerarla son: cantidad solicitada = ofertada y el total ofertado no es un outlier.
- **Mazbien:** Monto adjudicado para bienes, correspondiente al total pagado.
- **Mazservicio:** Monto adjudicado para servicios, correspondiente al total pagado.

Es crucial comprender especialmente la información en la columna `product\_id`, que representa el id único del producto para cada comprador. Cada comprador tiene su propio conjunto de ids, y no se repiten entre distintos compradores. La columna `categoriaid` muestra el id de la categoría a la que pertenece el producto, siendo este id transversal para los compradores.

La columna más relevante es `Checkz`, que actúa como un clasificador mostrando si la oferta fue identificada como atípica o no. Se clasifica como atípica cuando la cantidad solicitada es diferente a la ofertada con `Checkz=-1`. También se clasifica como oferta atípica cuando el total ofertado es un outlier, situación que se detallará más adelante.

WherEX emplea un modelo de detección de ofertas atípicas en dos etapas. La primera utiliza un método robusto de z-score, donde se utilizan las ofertas de cada producto en específico de dicha licitación, se detectan outliers si el valor z es  $<-1.75$  o  $>1.96$ . La segunda etapa se basa en datos históricos, utilizando las últimas tres licitaciones adjudicadas anteriores para obtener un valor histórico, si el precio unitario es tres veces mayor o

menor al valor histórico, es clasificado como outliers. Este método es preciso, pero presenta limitaciones cuando no se cumplen ciertos requisitos, como la falta de licitaciones anteriores para un producto específico o cuando las ofertas actuales son insuficientes para el modelo z, para poder realizar un modelo Z se necesitan a lo menos 3 ofertas por producto. En resumen si es la primera vez que se licita un producto o presenta menos de tres ofertas no ingresa al modelo ya existente y no se analiza si es una oferta atípica o no.

Este es el punto de intervención del proyecto actual. Busca mejorar estas situaciones mediante el uso de datos históricos basados en `product\_id` o `categoriaid` cuando no se cumplen los requisitos anteriores. Aunque la última columna puede no ser tan precisa debido a la diversidad de productos y precios en cada categoría, proporciona una idea del rango de precios esperado para las ofertas en una licitación no analizable mediante el modelo tradicional. La mejora de este aspecto es esencial para optimizar la detección de outliers en condiciones menos ideales y garantizar la integridad de los procesos de wherEX.

Una vez entendida la situación actual de la empresa, se estudiaron soluciones similares en el campo, poniendo especial énfasis en regresiones con distintos metodos. A continuación, se explorarán los modelos considerados para mejorar la detección de outliers en situaciones donde el modelo tradicional no puede aplicarse. Para poder analizar y comparar cada modelo se utilizaron los datos de la categoría "Boas", ID 625, con un total de 542 ofertas, 99 adjudicaciones, 14 tipos de productos y un mega outlier de 11.05.487\$ CLP. Las imágenes siguientes ilustrarán la ya mencionada variación de precios unitarios de oferta dentro de esta categoría lo cual dificulta la implementación de un modelo predictivo. A continuación se mostraran dos imágenes en las cuales se puede ver la distribución de precios a lo largo del tiempo y la gran variedad que existe dentro de las categorías.

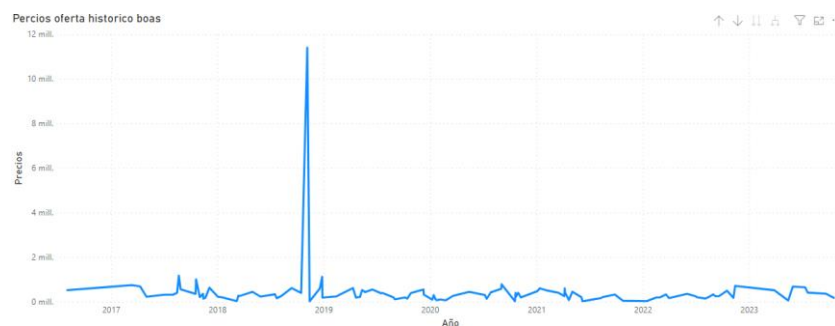


Figura 1. Distribución de precios dentro de la categoría boas con mega outlier a lo largo del tiempo

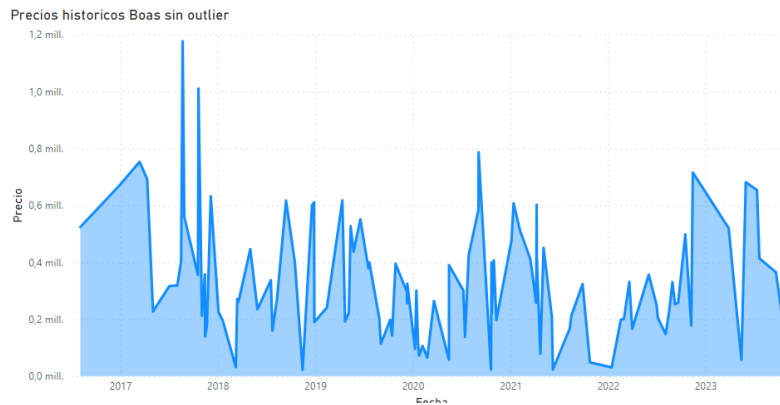


Figura 2. Distribución de precios dentro de la categoría boas con mega outlier a lo largo del tiempo

Debido a la gran variedad de productos y precios dentro de las categorías, como se puede apreciar en las imágenes, es necesario realizar un análisis estandarizado de estos datos dentro de cada modelo se utilizaron 3 tipos:

- **MinMax scaler:** El cual no es recomendable por verse muy afectado por outliers grandes en su máximo y mínimo

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- **Z-Score:** El cual es el método mas clásico de escalado tomando la media y desviación estándar

$$z = \frac{x - \mu}{\sigma}$$

$$\begin{aligned} \mu &= \text{Mean} \\ \sigma &= \text{Standard Deviation} \end{aligned}$$

- **Escalamiento robusto/Z-score robusto:** el cual es el mas recomendado en este caso debido a que es la misma formula que el Z-score clásico solo que utiliza mediana y el rango intercuartílico

A continuación se abordara uno por uno los modelos estudiados que son soluciones potenciales para solucionar el problema detectado.

### 5.1. Regresión simple y multiple

Este tipo de modelado se centra principalmente en datos que experimentan cambios medibles y estudiabiles a lo largo del tiempo. En este caso, a pesar de contar con datos desde 2015, no siguen una tendencia de comportamiento que permita una regresión adecuada. Esto se debe a que los datos se analizan por categoría, y la diversidad dentro de las categorías es tan amplia que impide un modelado preciso.

## 5.2. CNN

Este modelo predictivo posibilita la adecuada interpretación de datos y la predicción de resultados mediante una serie de capas neuronales. Sin embargo, tras estudiar su funcionamiento, se descartó de inmediato como una posible solución para el presente proyecto. Esto se debe a que opera con datos bidimensionales, como imágenes, los cuales no son relevantes ni útiles para el alcance de este proyecto.

## 5.3. Análisis estadístico

La creación de un modelo mediante el análisis estadístico nos brindará varios valores, como la media, mediana, desviación y varianza. Estos datos nos posibilitarán establecer un umbral junto con sus percentiles. Este umbral será utilizado para identificar las ofertas en la licitación analizada que se desvían de lo normal, de manera similar al modelo ya existente en WherEX.

En otras palabras, la implementación de esta solución implica realizar un nuevo método de análisis de outliers, pero esta vez utilizando la información histórica disponible dentro de cada categoría. Esto permitirá un análisis más preciso de las licitaciones que no cumplen con los requisitos del comprador que haya licitado previamente ese producto específico y que exista una cantidad mínima de ofertas en la licitación que se está analizando.

## 5.4. LSTM/Long short term memory

Este modelo, diseñado para predecir a largo plazo, no se ve afectado por datos más recientes, lo que teóricamente lo hace adecuado para nuestro análisis. Para evaluar su idoneidad, se estandarizaron sus datos utilizando los tres métodos mencionados previamente. Los resultados iniciales sugieren que el modelo no es altamente predictivo, mostrando una línea recta que podría representar un cálculo promedio de los datos. Como se ve en la siguiente imagen.

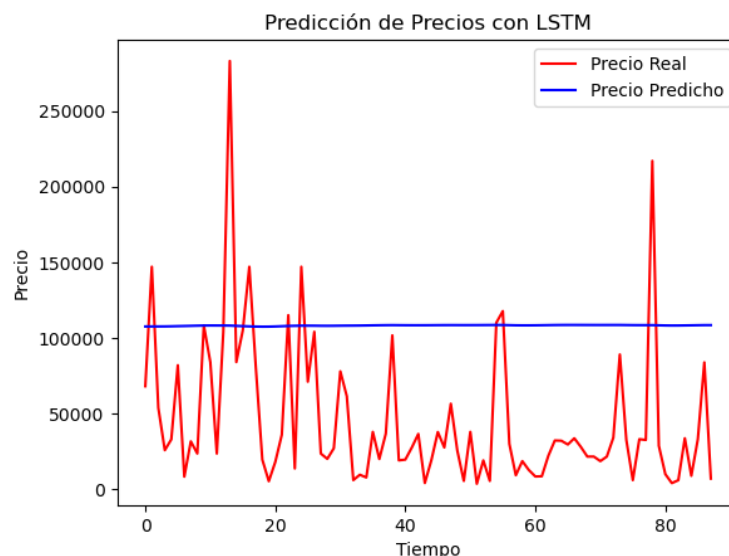


Figura 3. Distribución de precios dentro de la categoría boas junto a su predicción

Posteriormente, se aplicaron los tres métodos de estandarización, y se destaca el que utilizó el método robusto z, ya que mostró resultados más favorables como se ve en la figura 4. Aunque la predicción inicial fue similar a los datos no estandarizados, se observó un aumento en los precios. Es importante señalar que este análisis fue preliminar, no enfocado en obtener resultados óptimos, sino en comprender la interacción del modelo con los datos y determinar su utilidad para este tipo de información y para futuras escalas de datos de la empresa.

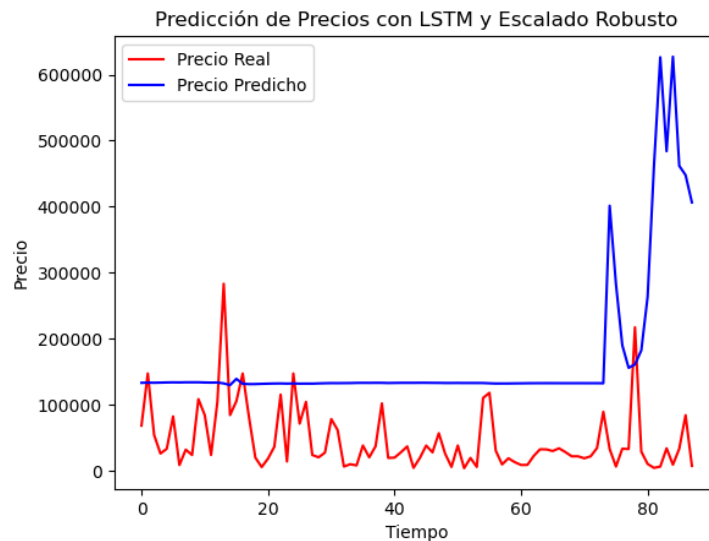


Figura 4. Distribución de precios (estandarizados) dentro de la categoría boas junto a su predicción

## 5.5. ARIMA/SARIMA

El modelo ARIMA, ampliamente reconocido en el ámbito financiero, se utiliza para predecir movimientos y fluctuaciones en acciones y diversos productos. Este modelo es relevante en nuestro caso, ya que buscamos analizar la variación de precios a lo largo del tiempo y los factores asociados a cada licitación. El ARIMA facilita el análisis y predicción de series temporales, y su variante, conocida como SARIMA, agrega la capacidad de identificar estacionalidad en los datos, revelando patrones estacionales en las ventas.

En el análisis del ARIMA, se estandarizaron los datos mediante los tres métodos previamente expuestos, al igual que en otros modelos revisados anteriormente. Se realizaron dos iteraciones diferentes, una considerando un outlier de más de 11 millones de pesos y otra sin considerarlo, totalizando seis iteraciones. Los resultados fueron satisfactorios, especialmente cuando no se consideró el outlier. A continuación, se presentarán dos imágenes que ilustran la estandarización mediante el puntaje z y los datos con y sin la presencia del mega outlier.

En la primera imagen, los resultados sin el outlier revelan un umbral de 5,345 pesos a 186,974 pesos. Estos resultados son fundamentales para identificar las ofertas que se desvían de este umbral, generando alertas inicialmente.

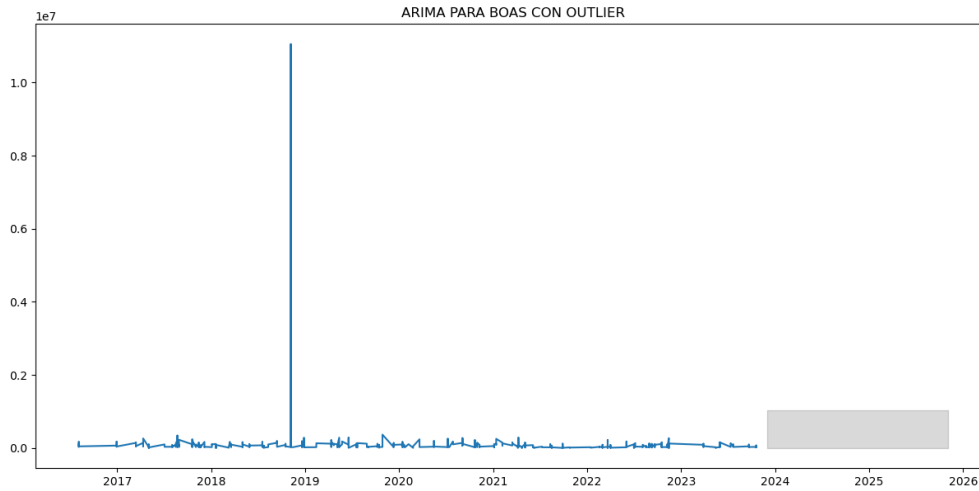


Figura 5. Distribución de precios dentro de la categoría boas con outlier junto a la predicción de ARIMA

La segunda imagen muestra un cambio significativo al considerar el outlier de 11 millones. La detección preliminar de outliers disminuye, y el umbral resultante es de 8,390 pesos a 1,021,325 pesos. En este caso, el umbral identifica una oferta "incorrecta" de 11 millones de pesos.

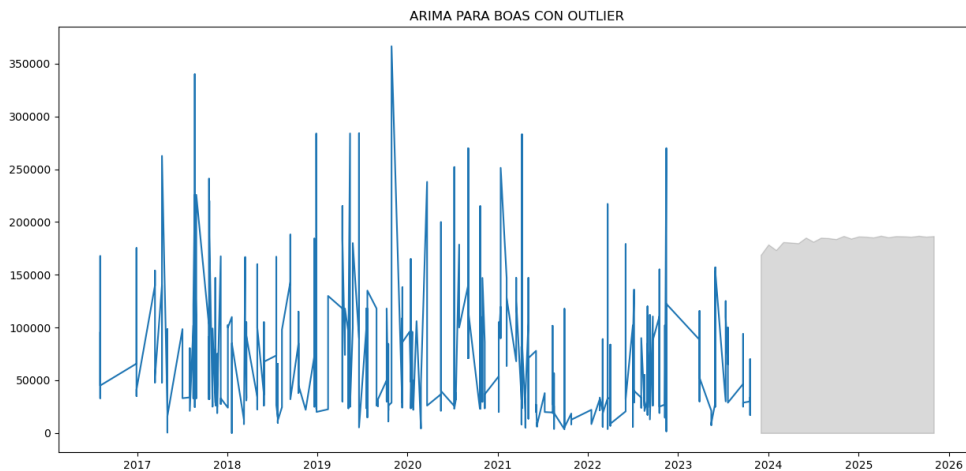


Figura 6. Distribución de precios dentro de la categoría boas sin outlier junto a la predicción de ARIMA

En resumen, el modelo ARIMA demuestra su eficacia en su versión original. Sin embargo, la versión SARIMA no es aplicable, ya que no se observa estacionalidad en los datos. Se buscará perfeccionar este modelo para hacerlo más específico y preciso, con un umbral adaptable a cada categoría y sus distintas ofertas en futuras etapas.

## 5.6. MLP

El modelo Multilayer Perceptron (MLP) es capaz de clasificar y predecir datos mediante la modelación de relaciones no lineales, una característica valiosa para la predicción de datos complejos. Este modelo busca minimizar la diferencia entre las predicciones y los datos reales, aprendiendo de manera autónoma a medida que se le proporciona más interés y datos.

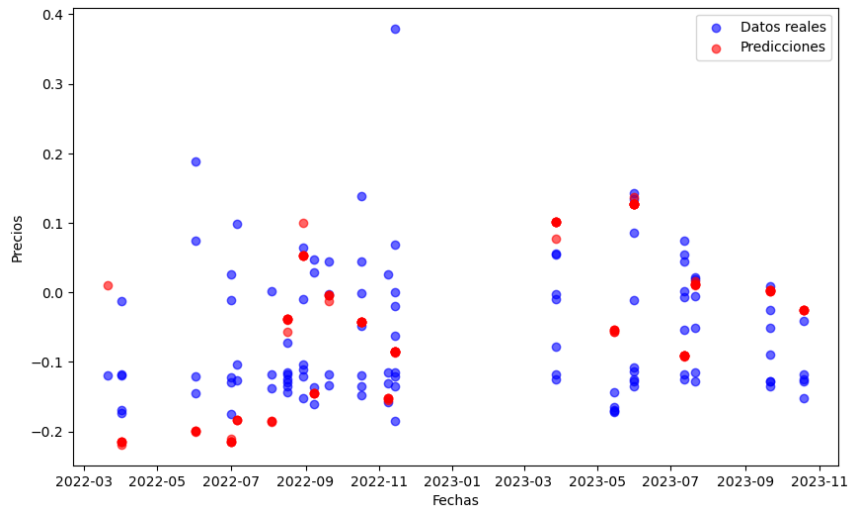


Figura 7. Predicción específica modelo MLP

Las dos imágenes presentadas muestran las predicciones del modelo y la visualización de datos estandarizados mediante el z score robusto. En la primera imagen, los puntos rojos representan las predicciones del modelo, mientras que los puntos azules representan las ofertas reales en una licitación específica. Se observa que las predicciones no son precisas y muestran una discrepancia considerable entre lo esperado y lo real.

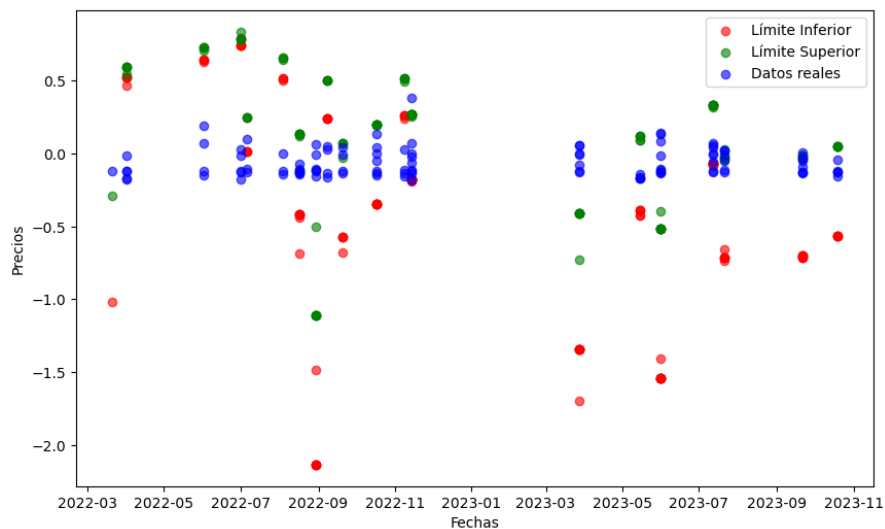


Figura 8. Predicción umbral modelo MLP

En la segunda imagen, se intenta mostrar un umbral predictivo para cada licitación. Los puntos verdes y rojos representan los límites superior e inferior, respectivamente, mientras que los puntos azules indican las ofertas reales. Los resultados no cumplen completamente las expectativas, mostrando aciertos en algunos casos pero alejándose de la realidad en otros. Aunque el modelo MLP, al igual que ARIMA, tiene potencial, se sugiere mejorar las predicciones ajustando factores y valores antes del modelado. Además, sería interesante no solo predecir para una licitación específica, sino crear un umbral consistente y con pocos cambios para el conjunto de licitaciones futuras, similar al enfoque utilizado en el modelo ARIMA.

## 6. Plan implementación y Metodología

En la fase de implementación, tras exhaustivos análisis de los resultados de las diversas soluciones propuestas, se ha tomado la decisión de mejorar y expandir el actual modelo de detección de outliers. Se incorporarán dos nuevas etapas de z-score, que se explicarán detalladamente más adelante, junto con la introducción de un tercer paso que utilizará el modelo ARIMA. Estas etapas analizarán los distintos posibles casos de outliers dependiendo de su origen y características propias de cada oferta haciendo una correcta limpieza y filtrado inicial de datos.

La implementación se llevará a cabo de manera gradual por medio de un código en Python para cada etapa. Inicialmente, se realizará una evaluación de los resultados al analizar las ofertas desde principios de noviembre hasta la fecha actual. Este análisis se centrará en los indicadores clave como GTV, GMV y comisiones. Posteriormente, se escalará el modelo, adaptándolo para su utilización diaria. Se analizarán las ofertas de las licitaciones cerradas en el mismo día, permitiendo una alerta cercana a la fecha de cierre de la licitación.

El último paso del plan consistirá en establecer un canal de alerta, preferiblemente a través de Slack, dirigido a los Customer Success Managers (CSM) responsables de cada empresa compradora. Este canal proporcionará la capacidad de alertar sobre posibles casos con datos incorrectos que requieran corrección. La comunicación efectiva y oportuna con los CSM asegurará una respuesta rápida y precisa ante situaciones críticas, mejorando así la calidad y confiabilidad de las transacciones en wherEX.



## 7. Evaluación económica y medidas de desempeño

El propósito central de este proyecto no radica en la generación directa de beneficios económicos ni en la reducción de costos, pero sí se anticipa un impacto en áreas financieras clave de la empresa. A continuación, se detallan los elementos más relevantes afectados por el proyecto:

1. **Comisiones:** La corrección de licitaciones adjudicadas con datos incorrectos influirá directamente en la cantidad de comisiones generadas, ya que los montos adjudicados podrían variar respecto a lo inicialmente registrado.
2. **Cálculos de GTV y GMV:** La corrección de montos adjudicados y transados tendrá repercusiones en los cálculos de Gross Transaction Value (GTV) y Gross Merchandise Value (GMV). Estos indicadores son fundamentales en diversas proyecciones y decisiones estratégicas.
3. **Metas de los CSM:** Las metas de los Customer Success Managers (CSM) se verán afectadas, ya que la cantidad de montos transados y adjudicados en las carteras de compradores asignados a cada CSM experimentará cambios.

En cuanto a las medidas de desempeño que se utilizarán, se destacan las siguientes:

1. **GTV corregido:** Esta medida reflejará la modificación, ya sea positiva o negativa, que se logre al corregir los montos de las transacciones adjudicadas.
2. **GMV corregido:** Similar al GTV corregido, esta medida cuantificará la modificación, en términos positivos o negativos, que se pueda realizar en los cálculos de Gross Merchandise Value.
3. **Cantidad de outliers detectados que no ingresaron al modelo existente:** Esta métrica permitirá visualizar la eficacia del proyecto al alertar sobre ofertas consideradas atípicas que no habrían sido identificadas por el modelo de detección de outliers existente..

Estas medidas de desempeño proporcionarán una evaluación cuantitativa y cualitativa del impacto del proyecto, permitiendo una valoración precisa de su efectividad y contribución a la mejora de los procesos financieros y operativos en wherEX.

## 8. Análisis de riesgo

En cuanto a los riesgos asociados al proyecto, se han identificado dos aspectos críticos. El primero se relaciona con la decisión de implementar o no el proyecto, es decir, resolver o no el problema identificado. Este riesgo no es considerado muy alto para WherEX, dado que la empresa ya cuenta con un modelo robusto para detectar outliers. Sin embargo, como se evidenció en este informe, el modelo actual no logra detectar el 100% de los errores, lo que constituye un riesgo inherente.

El riesgo de no implementar el proyecto es más complejo de evaluar, ya que su impacto será específico para cada licitación, afectando diversas métricas y comisiones. Evaluar este riesgo como una gran pérdida económica o una diferencia sustancial en los cálculos con datos reales resulta complicado. El mayor riesgo recae en la percepción y confianza de los compradores y proveedores en la plataforma. Dado que situaciones atípicas y erróneas podrían influir en las adjudicaciones, existe la posibilidad de cobrar de más o de menos a los proveedores, afectando la credibilidad y confianza de la plataforma.

En este contexto, no se presentará una matriz con valores numéricos para cuantificar el riesgo de no implementar el proyecto, debido a la dificultad inherente de predecir cuánto serán los montos corregidos.

		Datos Reales	
		Es Outlier	NO Es Outlier
Resultados Filtro	Lo Filtramos	True Positive 5	False Positive 1
	No lo Filtramos	False Negative 8	True Negative 1

Figura 9. Matriz de riesgo sobre Outliers

Por otro lado, en relación a los riesgos asociados a la implementación del proyecto, se realizaron dos matrices de confusión. La primera aborda si la oferta es o no un outlet y si se filtra correctamente. En esta matriz, se observa que el filtrado incorrecto afecta, pero no de manera significativa.

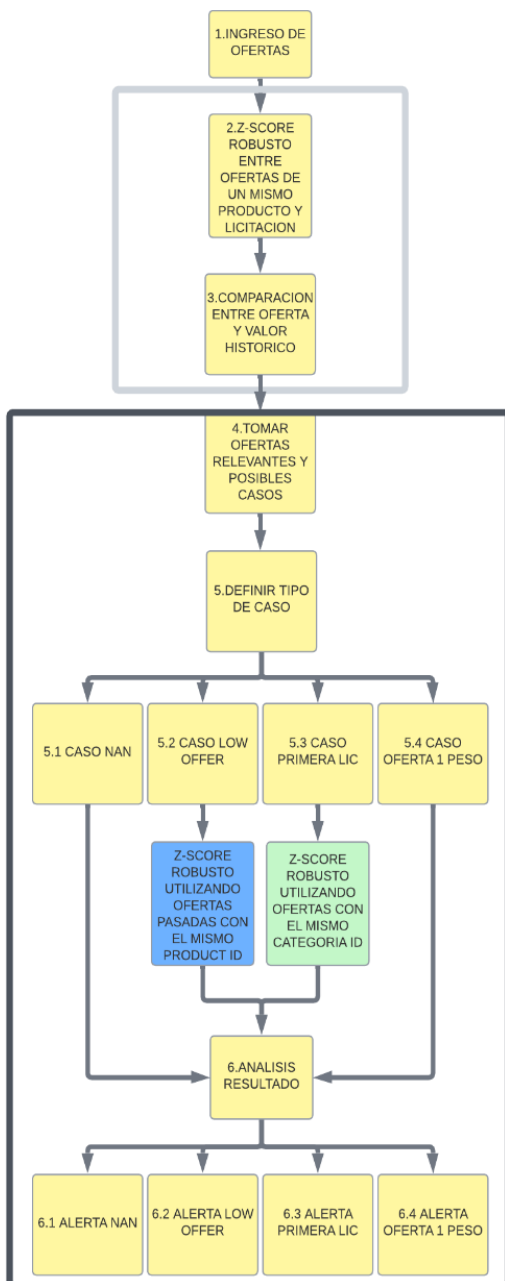
		Datos Reales	
		Es Precio Real	No es Precio Real
Resultados Filtro	Lo dejamos pasar	True Positive 10	False Positive 1
	No lo dejamos pasar	False Negative 20	True Negative 1

La segunda matriz se centra en el precio, determinando si este es real o no, y si se filtra correctamente. En este caso, los errores en el filtrado afectarán más notoriamente, ya que se trata de precios reales que impactan directamente a los proveedores. Si se filtra incorrectamente un precio real, demasiado bajo o alto, podría perjudicar al proveedor, generando desequilibrios en la libre competencia. Es crucial destacar que, en ambos casos, los efectos negativos son limitados, ya que no se realiza un filtrado directo, sino que se alerta a los Customer Success Managers (CSM), quienes tomarán decisiones humanas antes de corregir la oferta específica. Sin

embargo, se subraya la importancia de ser cautelosos al notificar a los proveedores, ya que un aviso incorrecto podría infringir normativas de libre competencia.

## 9. Desarrollo proyecto

Como se mencionó anteriormente, wherEX ya cuenta con un sistema de detección de outliers. Sin embargo, el problema que existe dentro de este sistema radica en ciertos casos en los cuales las ofertas no son identificadas. Estos casos específicos se presentan cuando hay 2 o menos ofertas por producto en cada licitación o cuando es la primera vez que se oferta ese producto en particular, es decir, cuando un comprador está licitando ese producto por primera vez. A continuación, se presenta un diagrama que ilustra los pasos del sistema de detección existente, encerrado en un cuadro de color gris, y los pasos que se agregarán con el nuevo sistema de detección de este proyecto, encerrado en un cuadro gris oscuro.



### Paso 1: Ingreso de Ofertas

- Todas las ofertas existentes son ingresadas en el sistema.

### Paso 2: Z-score Robusto

- Se realiza un z-score robusto con las ofertas de cada producto de cada licitación, identificando posibles outliers.

### Paso 3: Valor Histórico

- Se obtiene el valor histórico a partir de las últimas 3 licitaciones adjudicadas para ese producto y se compara si es tres veces mayor o menor.

### Inicio del Modelo del Proyecto

### Paso 4: Selección de Ofertas Relevantes

- Se seleccionan las ofertas relevantes a analizar, considerando condiciones como licstatus='Adjudicada', checkz=0 y licfechacierre posterior al 01-11-2023.

### Paso 5: Definición de Casos

- Se definen los posibles casos, clasificándolos en:
  - Ofertas con Precio NaN: Ofertas con precio oferta igual a NaN.
  - Primera Licitación: Primera licitación de ese producto en la historia.
  - Low Offer: Total de ofertas de la licitación  $\leq 2$ .
  - Ofertas con Precio de \$1: Ofertas con preciooferta=1.

**Paso 5.1 y 5.4: Casos Especiales**

- Estos casos requieren un análisis manual y supervisado, ya que son situaciones permitidas por wherEX pero que deben ser notificadas

**Paso 6: Z-score para Casos 5.2 y 5.3**

- Low Offer: Con datos históricos de licitaciones pasadas de ese producto
- Primer licitacion: Con datos históricos de licitaciones adjudicadas de la categoría del producto.

**Paso 6: Resultados**

- Se analizan los resultados de los z-score. Si los valores son  $<-1.75$  o  $>1.96$ , se consideran outliers y son alertados.

**Paso 7: Base de Datos de Casos Pertinentes**

- Todos los casos que ahora pasaron a ser 'Alerta' pasan a una base de datos con la información relevante de la oferta.

**Paso 8: Alerta a los CSM**

- Se generan alertas a los Customer Success Managers (CSM) para su revisión y posible corrección.

*Los pasos 7 y 8 no se evidencian en el diagrama ya que aun no han sido implementados dentro de la empresa*

Este modelo detallado describe el flujo de procesos desde la entrada de ofertas hasta la identificación y alerta de posibles irregularidades, asegurando una gestión efectiva de las transacciones en el entorno de wherEX.

Cada uno de estos pasos esta programado en Python, en un futuro se busca que el código se corra una vez por día y analice las ofertas del día en especifico, en este caso en particular se analizaron ofertas de aproximadamente 5 semanas para poder evaluar su efectividad. A continuación se evidenciaran algunas tablas que permitirán visualizar el alcance del proyecto.

Esta primera tabla e imagen se evidencia la cantidad total de ofertas de licitaciones adjudicadas con valor  $checkz=0$  y que se encontraban dentro del rango de fechas. En la segunda fila de la tabla, podemos observar la cantidad de posibles casos que no ingresaron al sistema de detección existente en wherEX. Es decir, un total de 20,019 ofertas tienen menos de 3 ofertas por producto por licitación, son la primera vez que se licitan o tienen precio unitario =1 o precio unitario=NaN.

### Ofertas totales vs posibles casos

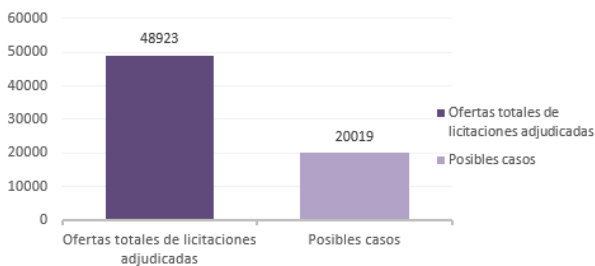


Figura 11. Ofertas totales vs Posibles casos

Ofertas totales de licitaciones adjudicadas	48.923
Posibles Casos	20.019

A continuación se evidenciará una tabla con dos imágenes donde se clasificará en lo en el tipo de caso que podría pertenecer cada oferta es decir que se clasifican los posibles casos en cada condición particular. Se puede visualizar que la mayoría de posibles casos pertenecen a los casos de primera licitación es decir que son productos Por Primera Vez están siendo licitados por el comprador.

### Posibles casos por tipo

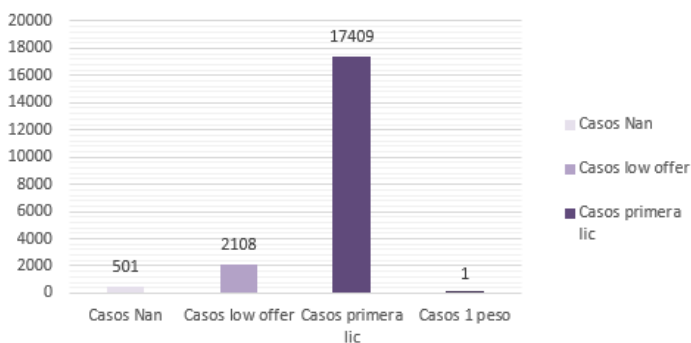


Figura 12. Clasificación posibles casos

Tipos de Posibles Casos	Cantidad
Casos Nan	501
Casos low offer	2.108
Casos primera lic	17.409
Casos 1 peso	1

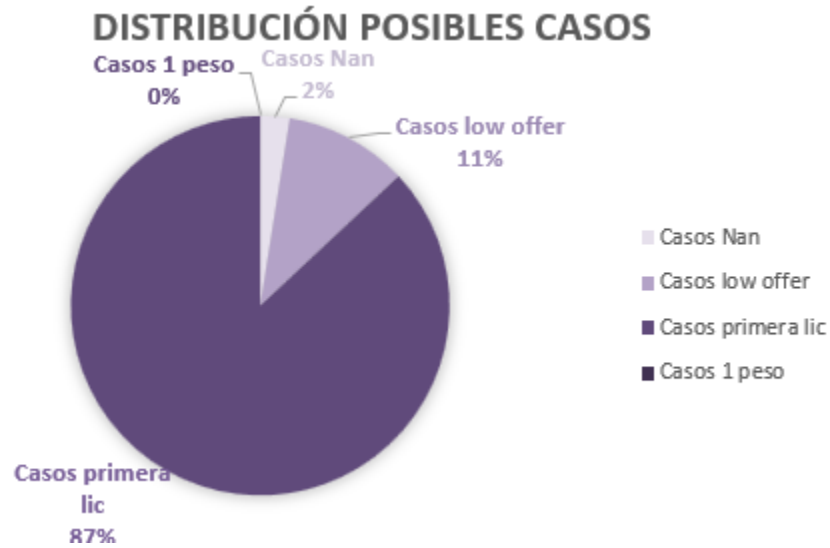


Figura 13. Distribucion de posibles Casos

## 10. Resultados

En los resultados, podemos observar una imagen y varias tablas que semostraran los resultados. En primer lugar, tenemos la cantidad de alertas resultantes para cada tipo de caso. Con respecto a las alertas "NAN", es importante señalar que estas requieren un análisis más detallado y la intervención de los CSM, ya que no se pueden clasificar automáticamente como alertas o permitir estas ofertas. En cuanto a las alertas "Low Offer", se logró detectar un total de 48 ofertas, lo que representa un 2.28% del total de posibles casos en esta categoría. De estas 48 ofertas, 13 son outliers inferiores y 35 son outliers superiores. En la categoría de "Primera licitación", se lograron detectar 63 outliers, correspondientes al 0.36% del total de posibles casos. De estos, 62 pertenecen a outliers inferiores y solo uno pertenece a outliers superiores. Finalmente, en el único caso de oferta a un peso, se confirmó que era un outlier inferior. Todo lo anterior se puede visualizar en la figura 14 y en la tabla a continuación.

Tipos de Alertas	Cantidad Alertas	Porcentaje Alertas/Posibles casos	Alertas outliers inferiores	Alertas Outliers superiores
Alertas Nan	0	0%	0	0
Alertas low offer	48	2,28%	13	35
Alertas primera lic	63	0,36%	62	1
Alertas 1 peso	1	100%	1	0

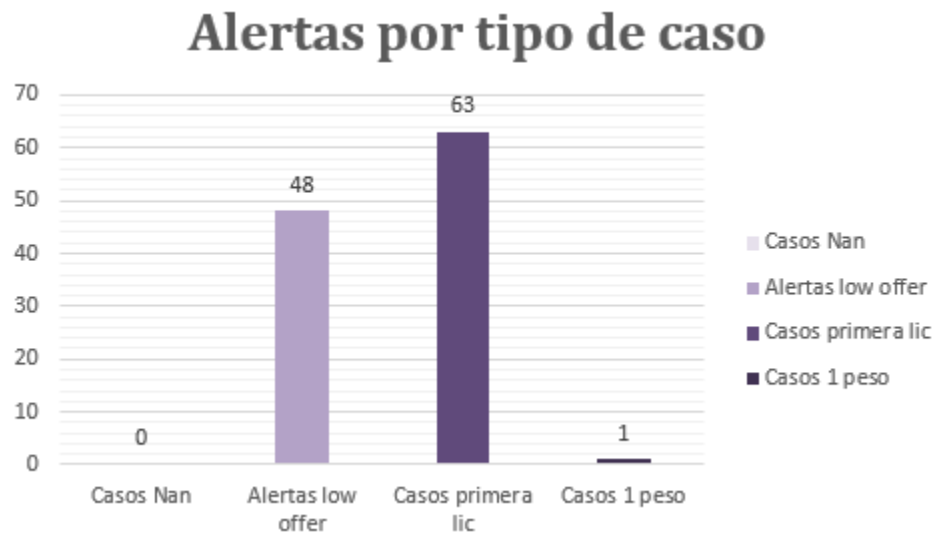


Figura 14. Alertas por tipo de Caso

Respecto de los efectos en la medida de desempeño que se han declarado en el proyecto se pondrán a prueba 2 medidas el GTV el GMV la comisión de agraciadamente no va a poder ser analizada ya que ésta varía muy en específico respecto a la cantidad que se está adjudicando el precio efectivo en el que se va a adjudicar el tipo de producto y otras variables que desgraciadamente tiene son analizadas muy posteriormente al cierre de una licitación por lo cual sus efectos en esta medida de desempeño se verán en un futuro pero no se podrán visualizar inmediatamente. Cabe destacar que ahora se asumirán que todas las alertas fueron correctamente indentificadas (es decir que si son outliers) y su precio de oferta será cambiado por la mediana del z-score que fueron puesotos a prueba

Comenzando con los efectos en GTV, se pueden visualizar en la tabla a continuación, que muestra cuatro columnas. En relación a las alertas "Low Offer", la primera columna muestra el GTV antiguo antes del proyecto, y la segunda columna presenta el GTV nuevo corregido, asumiendo que todas las alertas estaban correctamente identificadas y necesitaban ser corregidas. En este caso, el precio unitario de oferta fue cambiado por la mediana del z-score correspondiente.

Tipo de Alerta	GTV OLD	GTV NEW	Corregido hacia abajo	Corregido hacia arriba
Alertas Low offer	\$17.819.688	\$7.614.215	\$-11.081.167	\$875.693
Alertas primera Lic	\$16.633.324	\$51.721.931	\$-9.840.680	\$45.929.286

En relación al análisis de las correcciones de alertas en el Gross Transaction Value (GTV), específicamente en las categorías de alertas "Low Offer" y "Primera Lic", se observan ajustes significativos en el valor total transado. En el caso de las alertas "Low Offer", se destaca una disminución marcada del GTV después de la corrección, indicando la identificación de ofertas con precios originalmente más altos de lo esperado. La mayor parte de la corrección fue hacia abajo, representando un cambio negativo de \$11.081.167. Además, se observa una corrección hacia arriba de \$875.693, sugiriendo la revisión y corrección de algunas ofertas inicialmente consideradas outliers bajos.

Por otro lado, las alertas "Primera Lic" revelan un cambio sustancial en el GTV después de la corrección, con un ajuste notable hacia arriba de \$45.929.286. Esto sugiere que estas alertas estaban relacionadas principalmente con ofertas que originalmente tenían un precio más bajo de lo esperado. Es fundamental señalar que estos resultados deben interpretarse con precaución debido a la gran variabilidad de precios dentro de cada categoría.

En resumen, las correcciones de alertas "Low Offer" y "Primera Lic" han influido significativamente en el GTV, identificando y corrigiendo ofertas con precios que se desviaban de las expectativas. Estos ajustes son esenciales para mejorar la precisión y confiabilidad de los datos, proporcionando así una representación más precisa del valor total transado en el mercado. Estas correcciones respaldan la obtención de insights más precisos y fundamentan la toma de decisiones informadas en el negocio.

El análisis de las correcciones de alertas en el Gross Merchandise Value (GMV), específicamente en las categorías de alertas "Low Offer" y "Primera Lic", adjuntos en la tabla a continuación

Tipo de Alerta	GMV OLD	GMV NEW	Corregido hacia abajo	Corregido hacia arriba
Alertas Low offer	\$ 7.229.392	\$ 4.016.603	\$3.596.308	\$ 383.519
Alertas primera Lic	\$ 1.495.994	\$22.319.394	0	\$20.823.399

Los resultados revelan impactos significativos en el valor transado. En el caso de las alertas "Low Offer", se observa una marcada disminución del GMV después de la corrección, indicando la identificación de ofertas con precios originalmente más altos de lo esperado. La mayor parte de la corrección (\$3.596.308) fue hacia abajo, pero se aprecia también una corrección hacia arriba de \$383.519, sugiriendo la revisión y corrección de algunas ofertas inicialmente consideradas outliers bajos.



Por otro lado, las alertas "Primera Lic" muestran un cambio sustancial en el GMV después de la corrección, con un ajuste considerable hacia arriba de \$20.823.399. Esto sugiere que estas alertas estaban relacionadas principalmente con ofertas que originalmente tenían un precio más bajo de lo esperado. Es crucial destacar que estos resultados deben interpretarse con precaución debido a la gran variabilidad de precios dentro de cada categoría.

En resumen, tanto las correcciones de alertas "Low Offer" como "Primera Lic" influyeron notablemente en el GMV, identificando y corrigiendo ofertas con precios que se desviaban de las expectativas. Estas correcciones contribuyen significativamente a mejorar la precisión y confiabilidad de los datos, proporcionando así una representación más fiel del valor transado en el mercado. Estos ajustes son esenciales para obtener insights más precisos y respaldar la toma de decisiones informadas en el negocio.

## 11. Conclusiones

En conclusión, el proyecto ha cumplido con éxito su objetivo principal, proporcionando a wherEX un valioso conocimiento sobre la utilización de sus datos históricos. La inclusión de la información desde el año 2015 ha demostrado ser esencial para fortalecer y escalar el modelo de detección de outliers existente. La utilización de datos históricos permite que wherEX avance hacia la obtención de valores más fidedignos y alineados con la realidad, evitando ser afectado por posibles errores humanos que podrían comprometer la precisión de las medidas obtenidas.

El proyecto refuerza el modelo de detección de outliers existente al abordar casos particulares que previamente no eran identificados y se clasificaban directamente como ofertas normales. Los beneficios obtenidos permiten que wherEX continúe escalando y robusteciendo su capacidad de detección, brindando una mayor meticulosidad en la identificación de outliers.

En términos de correcciones, se ha logrado detectar y corregir el Gross Transaction Value (GTV) y el Gross Merchandise Value (GMV) hacia valores más realistas, proporcionando así una representación más precisa del valor total transado. Este enfoque en la corrección de valores apegados a la realidad es crucial para evitar suposiciones y mejorar la confiabilidad de las medidas obtenidas.

Como recomendación final, se insta a wherEX a continuar con el proceso de detección de ofertas atípicas, poniendo especial énfasis en evaluar los efectos de estas correcciones en las comisiones cobradas a los proveedores. Esto garantizará que las comisiones sean calculadas de manera precisa, evitando pérdidas financieras para la empresa y manteniendo la lealtad de los proveedores en el proceso de cobro de comisiones por cada licitación adjudicada.





