

Predicción de costo de Pick up para Importaciones FCL-EXW



Alumno: Rodrigo Valenzuela Luppá

Universidad Adolfo Ibáñez

Carrera: Ingeniería Civil Industrial

Empresa/Área: Shippter/Pricing and Revenue Management

Resumen Ejecutivo:

El proyecto de pasantía desarrollado en Shippter, una empresa líder en asesoría de importaciones, representa un esfuerzo significativo para mejorar la eficiencia y competitividad en el mercado de importaciones FCL-EXW, particularmente con enfoque en las operaciones desde Estados Unidos. Frente a los retos de variabilidad en los costos de pick-up y la ausencia de un sistema de cotizaciones EXW automatizado, el proyecto introduce un modelo predictivo avanzado, fundamentado en técnicas de machine learning como la regresión cuantílica y los bosques aleatorios. Esta innovación busca ofrecer estimaciones de costos más precisas y rápidas.

En el núcleo del proyecto, se plantea el objetivo de **reducir significativamente la variabilidad del margen de pick-up en origen, concretamente una disminución de al menos un 30% en los próximos cuatro meses. Esto implica pasar de una variabilidad actual del 20% a aproximadamente un 14%.** Al alcanzar este objetivo, se espera no solo minimizar las pérdidas operativas sino también mejorar la velocidad en la entrega de tarifas a los clientes, un factor crucial en el competitivo mercado de importaciones.

La implementación del modelo predictivo está diseñada para integrarse sin problemas con la infraestructura existente de Shippter, proporcionando una solución eficaz y escalable. Los resultados preliminares indican un impacto positivo en la precisión de las cotizaciones y una notable reducción en la variabilidad de los márgenes. Este avance no solo mejora la eficiencia operativa de Shippter, sino que también fortalece su posición en el mercado, ofreciendo un servicio más rápido y confiable a sus clientes. Aunque el modelo requiere iteraciones futuras para su perfeccionamiento, su implementación ya señala un cambio significativo en la estrategia comercial y operativa de la empresa.

Abstract:

The internship project at Shippter, a leading company in import advisory, represents a significant effort to enhance efficiency and competitiveness in the FCL-EXW import market, focusing particularly on operations from the United States. Faced with challenges such as variability in pick-up costs and the lack of an automated EXW quotation system, the project introduces an advanced predictive model, based on machine learning techniques like quantile regression and random forests. This innovation aims to provide more accurate and faster cost estimates.

At the core of the project is the **goal to significantly reduce the variability of the pick-up margin at origin, specifically aiming for a reduction of at least 30% over the next four months. This means reducing the current variability from 20% to approximately 14%.** Achieving this objective is expected not only to minimize operational losses but also to enhance the speed of delivering rates to clients, a crucial factor in the competitive import market.

The implementation of the predictive model is designed to integrate seamlessly with Shippter's existing infrastructure, providing an effective and scalable solution. Preliminary results indicate a positive impact on the accuracy of quotations and a noticeable reduction in margin variability. This advancement not only improves Shippter's operational efficiency but also strengthens its market position by offering faster and more reliable service to its clients. Although the model requires further iterations for refinement, its implementation already signifies a significant shift in the company's business and operational strategy.

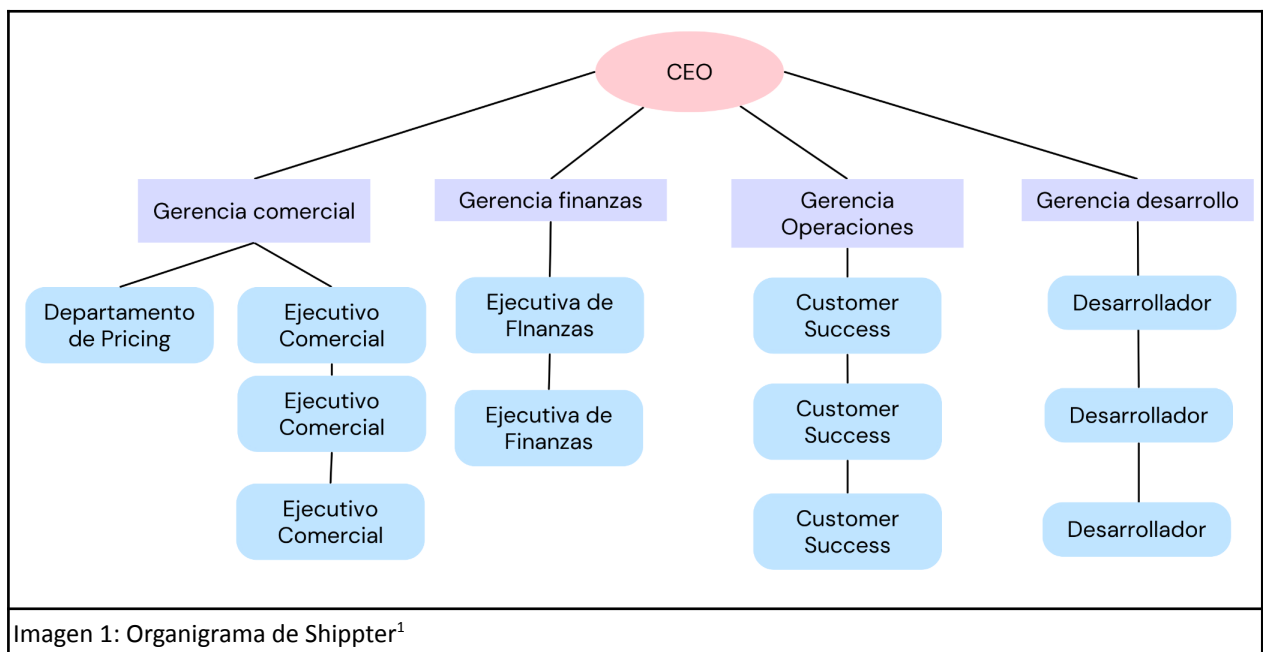
Índice

Contexto de la empresa:	5
Planteamiento del Problema:	6
Antecedente:	6
Síntomas:	7
Objetivos:	9
General:	10
Específicos:	10
Estado del arte:	10
Propuestas de solución:	12
Solución escogida:	12
Plan de implementación:	13
Análisis Riesgo:	14
Medidas de desempeño:	16
Desarrollo del proyecto:	17
Base de datos:	17
Limpieza y cambios en Base de datos:	17
Modelos Predictivos:	21
Regresión Cuantílica:	22
Modelo de Bosques aleatorios:	24
Resultados:	26
Evaluación Económica:	27
Conclusiones:	29
Referencias:	30
Anexos:	31

Contexto de la empresa:

Shippter es un Forwarder digital e integral, el cual tiene como objetivo principal el asesorar las importaciones de sus clientes desde la ubicación del manufacturer en el país de origen hasta las puertas de su fábrica aquí en Chile. La propuesta de valor que ofrece la empresa es la centralización de esta asesoría, dado que el importar algún embarque requiere de varios pasos y coordinaciones entre los distintos individuos y Shippter logra operar como único coordinador de la importación.

Shippter lleva aproximadamente 5 años en el mercado, sin embargo ha tenido un alto crecimiento los últimos dos años. La empresa actualmente consta de 30 trabajadores, estos se distribuyen en 4 áreas principales: Comercial, Operaciones, Finanzas y Desarrollo. A continuación el organigrama que representa a la empresa:



El desarrollo del trabajo se realizó en el departamento de Pricing and Revenue Management el cual pertenece al área comercial. El objetivo del departamento es el análisis y fijación de las tarifas de todos los conceptos en origen y algunos en destino, es decir, tratar de conseguir la tarifa all-in² de la importación (según el término de compra que desee el cliente) lo antes posible, como política de

¹ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

² Tarifa All-in: Tarifa total de la importación que incluye pick-up, gastos documentales en origen, flete internacional y gasto documental en destino.

empresa, de no estar automatizada por la plataforma, el departamento de pricing es el encargado de conseguirla en las próximas 24 hrs hábiles.

Planteamiento del Problema:

Antecedente:

Como se mencionó en el contexto de la empresa, uno de los lineamientos principales de Shippter es presentar al cliente una tarifa lo antes posible, por lo cual presenta tarifas automatizadas por plataforma. Esto dado que la industria del forwarding es muy competitiva y el entregar tarifa a cargas urgentes se traduce en cierre de más operaciones y el afianzar a más clientes.

Shippter ofrece a sus clientes 3 tipos de importaciones, Aéreas, FCL (Full Container Load) y LCL (Less Than Full Container Load). Además, el cliente puede elegir el incoterm que desea comprar/contratar con el forwarder según lo acordado con su proveedor en el país de origen. Esta puede ir desde FOB el cual el forwarder solo gestiona y tienen responsabilidad de la carga desde que la importación sale del puerto de origen hasta la fábrica del cliente, o EXW el cual la responsabilidad y gestión la toma el forwarder desde que la carga sale de la fábrica del manufacturer, por lo que tiene que gestionar el flete local con documentación en origen, el flete internacional y la documentación y flete local en el país de destino. A continuación una imagen que logra visualizar la diferencia entre incoterms, siendo en rojo la responsabilidad que tiene Shippter y en azul el embarcador en origen:

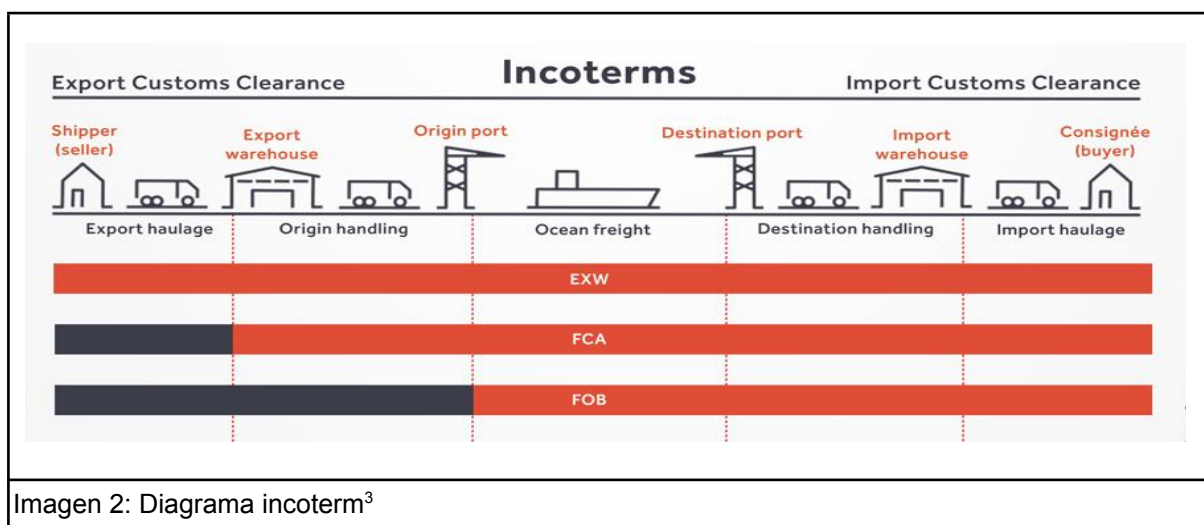


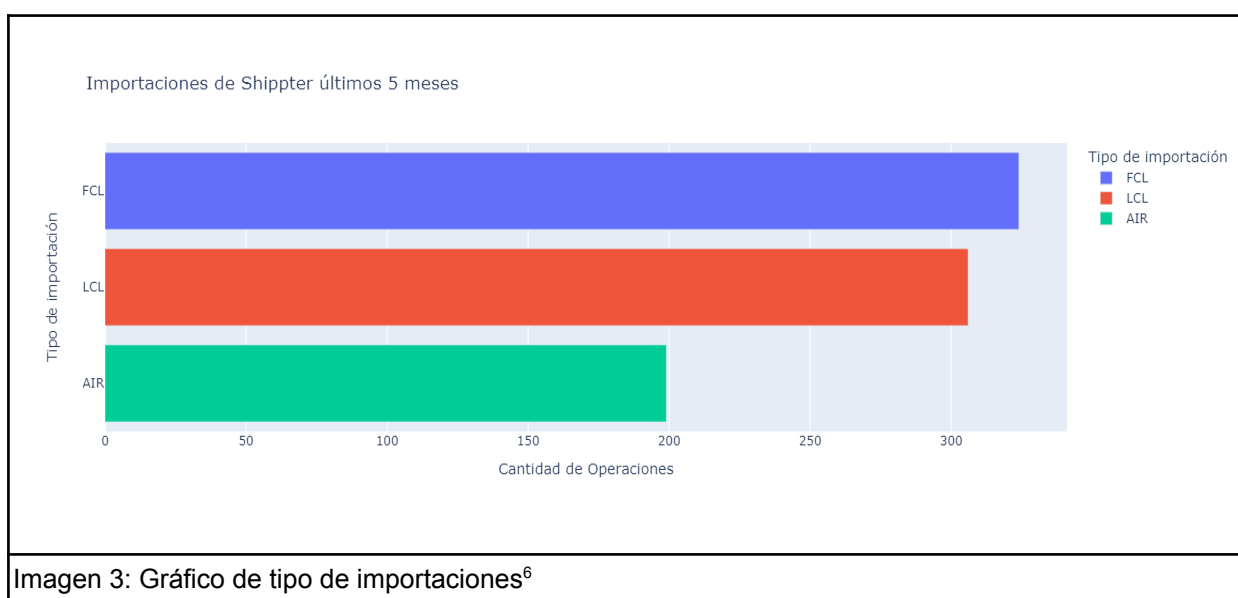
Imagen 2: Diagrama incoterm³

³ Incoterm. (2019). descripción de incoterms [Página Web]. Transporteca.
<https://transporteca.es/fob-o-exw/>

Síntomas:

En esta industria es interesante captar las importaciones tipo EXW, dado que da la posibilidad al forwarder de marginar con los conceptos de pick-up⁴ y gastos locales⁵ en origen, además de facilidades de coordinación para el forwarder en destino en este caso Shippter, dado que tiene un mayor control de la importación.

La empresa tiene como objetivo estratégico el aumentar el nivel de importaciones tipo FCL, en la cual últimamente ha ido en aumento en sus distintos tipos de importaciones (FCL, LCL y aéreo). La proporción de las importaciones de Shippter se pueden ver en el siguiente gráfico (datos hasta octubre 2023):

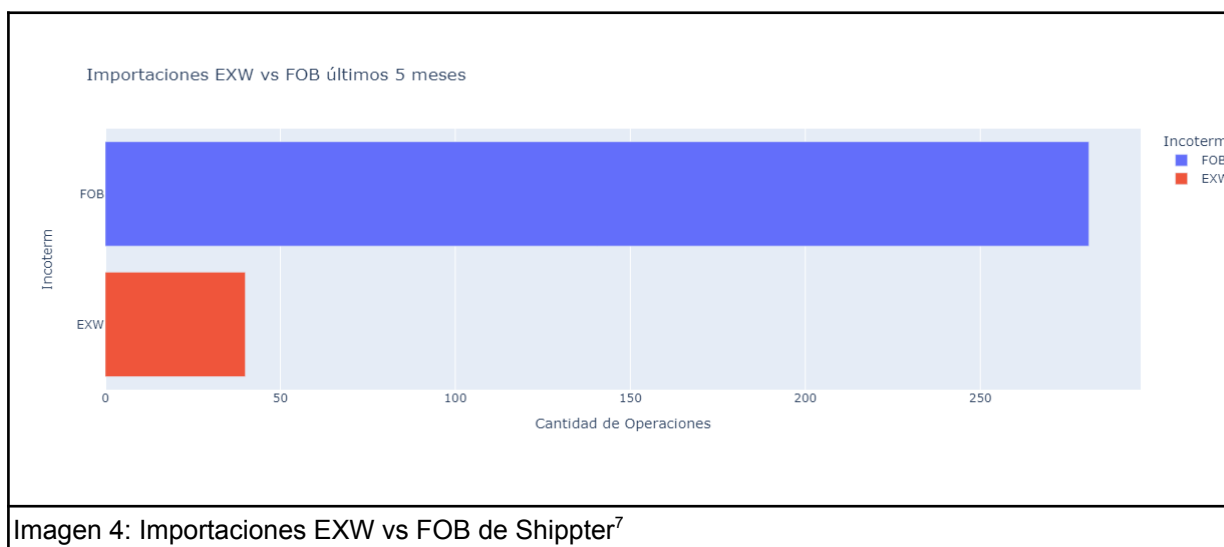


Con el gráfico anterior se puede apreciar que en los últimos meses la importación FCL ha tenido un gran flujo en la empresa, Sin embargo de las importaciones FCL la mayoría son de tipo FOB, la cual quita a Shippter la posibilidad de marginar por conceptos en origen ya anteriormente mencionados por la importación tipo EXW. como podemos ver en el siguiente gráfico que nos muestra la proporción EXW vs FOB de Shippter:

⁴ Pick-up: Costo del flete en el país de origen

⁵ Gastos Locales: Gastos documentales en el puerto, usualmente no varían

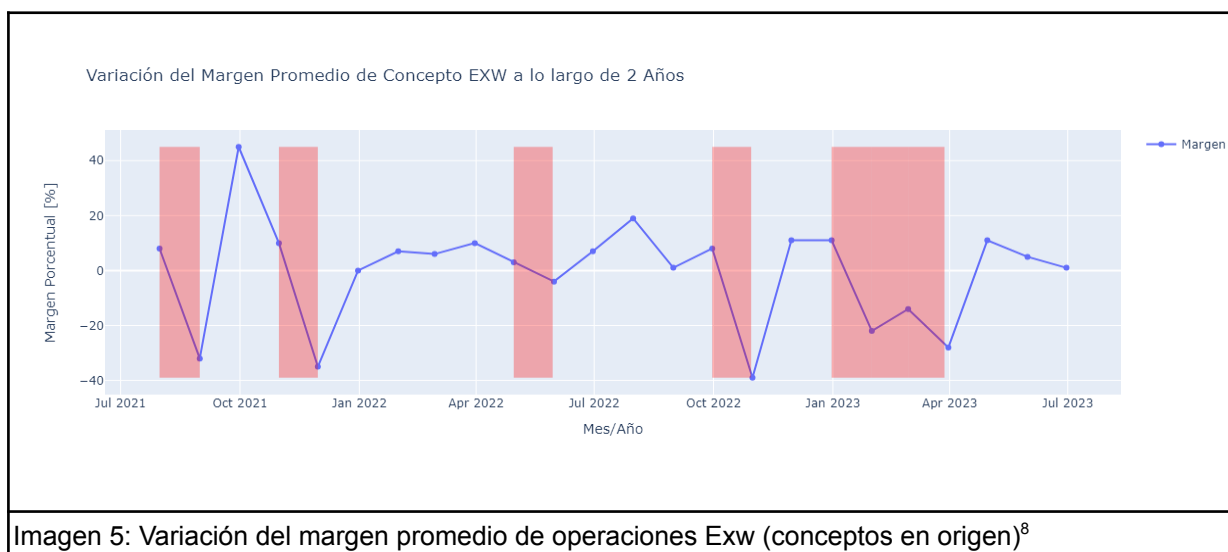
⁶ Rodrigo Valenzuela, 2023. Desarrollo del proyecto



Lo anterior dado que las importaciones de categoría FOB tienen tarifarios mensuales, por ende los clientes tienen la tarifa de la cotización de manera inmediata. En caso contrario, las tarifas EXW se cotizan de manera manual dado que el concepto de Pick up en origen varía según la dirección que escoja el cliente, lo que significa que puede tardar un mínimo de 1 día, sin embargo en promedio tardan 2 días por el contacto de Shippter con el proveedor y el proveedor con camioneros.

Anteriormente Shippter tenía automatizado el cotizador de EXW para importaciones FCL, sin embargo, en junio de 2023 se eliminó de plataforma, dado que no se lograba llegar a los márgenes deseados en los conceptos en origen. El rendimiento anterior se puede apreciar en el siguiente gráfico siendo la línea azul el margen promedio del mes y los meses que llegan a margen promedio negativos contienen una columna roja:

⁷ Rodrigo Valenzuela, 2023. Desarrollo del proyecto



Se puede apreciar que presenta una alta variabilidad del margen llegando a una desviación estándar de 20% en el periodo analizado, además el desempeño de los primeros 4 meses de 2023 del margen fue nocivo para la empresa, a tal manera que la gerencia de Shippter tomó la decisión de solamente cotizar las importaciones EXW de manera manual.

Esto supone un desperdicio de oportunidad, dado que muchos clientes no pueden comprar en términos FOB por limitaciones de sus proveedores en los países de origen, por lo que se pierden potenciales clientes, y más aún, se pierde la proyección de la propuesta de valor de la empresa.

Objetivos:

Los objetivos que se presentarán a continuación fueron decididos en conjunto con el departamento de pricing y la gerencia comercial, esto en razón de la necesidad de la empresa, alineamientos estratégicos, capacidades ingenieriles y el alcance del proyecto. Por último mencionar que se acota el proyecto únicamente para importaciones desde Estados Unidos por petición del jefe del departamento.

⁸ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

General:

En los próximos 4 meses, reducir la variación del margen de pick-up en origen en al menos un 30% (de un 20% a aproximadamente un 14% de variabilidad) con el fin de disminuir las pérdidas operativas y acelerar la determinación de tarifas para el cliente.

Específicos:

1. Aumentar y mantener el margen mensual operacional en promedio del 10% para operaciones EXW-FCL
2. Obtener el costo del Pick-up en menos de 24 horas, para que el cliente la tenga lo antes posible y seguir lineamientos de la empresa.
3. Aumentar el flujo de operaciones EXW un 20% para tipo de importación FCL

Estado del arte:

Para el estudio del estado del arte se encontró una tesis a máster el cual utiliza un contexto similar para la explicación de modelo predictivo que supera a la regresión lineal múltiple el cual es interesante revisar, puesto que logra realizar una predicción de costos.

Un caso de éxito de empresa papelería que utiliza modelo predictivo y machine learning para poder reducir sus costos y optimizar la producción de la empresa.

Y por último un caso en que utilizan subasta inversa para reducir costos y obtener la mejor tarifa posible, lo cual se traduce en un precio competitivo.

Ejemplo 1: Javier D. Martín⁹ ingeniero perteneciente a la facultad de ciencias económicas de la Universidad de Córdoba, realizó una investigación con un contexto similar al presente en este trabajo, en este analiza el transporte local dentro de estados unidos realizando una mejora de ruteo logística y, para lo interesante de este trabajo, una predicción de costo utilizando regresión cuantílica. Fundamentando que este tipo de regresión es mejor para variables que presentan datos atípicos, heterocedasticidad o cambios estructurales, lo cual la típica fórmula de de la regresión lineal múltiple

⁹ Javier D. Martín. (2015). Estimación de tarifas de fletes a través del uso del método de la Regresión Cuantílica (Tesis de maestría, Universidad nacional de Córdoba). rdu.unc.edu.ar

no puede explicar su relación para una predicción precisa; El único defecto de este método es la alta complejidad que resulta implementarlo. El funcionamiento de la regresión cuantílica es prácticamente similar a una regresión lineal múltiple, sin embargo, en vez de utilizar la media como la regresión tradicional, utiliza distintos cuartiles (o cualquier partición deseada) para el set de datos, para cada nivel de cuartil habrá una regresión en específico que logra estimar un valor de la variable dependiente. Como resultado obtuve un valor en promedio 20% más cercano al valor real que el de una regresión lineal. Sin embargo, como desventaja necesito una gran cantidad de datos para lograr resultados significativos

Ejemplo 2:

En el artículo¹⁰ encontrado, se propone una problemática similar a la nuestra, ¿qué pasa si una empresa tiene una alta variabilidad en el margen operacional, o que en este caso la calidad de producción no llegaba a satisfacer a los clientes?, lo cual era un daño comercial para la empresa. La empresa ProcessMiner Inc. presentaba este problema. Tomó como solución el implementar un análisis predictivo a su sistema de sensores, además al modelo predictivo lo alimentaron de la base de datos histórica y se sigue actualizando en tiempo real. Con esto lograron reducir en gran manera la variabilidad, además de aumentar la satisfacción de sus clientes, esto sin hacer un cambio gigantesco a la maquinaria de la empresa.

Ejemplo 3:

Para el tercer caso se tomará de ejemplo a la empresa Super Procure¹¹, esta empresa plantea la falta de transparencia y eficiencia de costo en el uso de transportistas locales. Dado que se presenta mucha variabilidad. Por ende ellos implementaron un software de subasta inversa, lo cual significa que el cliente establece una tarifa, y posteriormente los transportistas realizan una puja para poder quedarse con la operación. La operación se cierra luego de varias rondas de puja de los proveedores. Con esto aumenta drásticamente la competencia entre los proveedores, lo cual trae grandes beneficios económicos al cliente, además de que eficacia el proceso de negociación entre las partes.

¹⁰ ManufacturingTomorrow. (2021, 31 de agosto). Variability Reduction: Why Important To Manufacturers? ManufacturingTomorrow. <https://www.manufacturingtomorrow.com/article/2021/08/variability-reduction-why-important-to-manufacturers/18768/>

¹¹ SuperProcure. (Fecha no especificada). Digital reverse auction empowers freight procurement strategy. SuperProcure. Recuperado de <https://www.superprocure.com/blog/digital-reverse-auction-empowers-freight-procurement-strategy/>

Según lo observado en la investigación, hay una tendencia de solucionar problemas similares con modelos predictivos o machine learning de tipo supervisado. Esto en vista de que estos tipos de soluciones siempre pueden aprender y son de un leve costo modificar para ajustar las predicciones.

Propuestas de solución:

En razón de lo investigado en el estado del arte, y en conjunto de la jefatura del departamento de pricing, se llegaron a 3 posibles soluciones a la problemática, esto alineado con el objetivo propuesto anteriormente y los objetivos específicos.

1. Implementación de Subasta inversa para proveedores. Shippter cuenta con 3 proveedores importantes: MSL, pluscargo y Traffic Tech. Al ser un número acotado nos permitirá lanzar un tipo de plataforma que les permita competir en costos, lo cual nos permitirá lograr una reducción significativa en costo y tener poder de negociación con ellos para reducirlos, así controlar de mejor manera la variabilidad de estos.
2. Implementar un agente de Shippter en Estados Unidos, para poder negociar de primera mano con los transportistas y así tener un control o previsión de los costos en origen para mantener la variabilidad. Esta solución supondría un cambio en la cadena de suministro, caso similar que se aplica con las cargas de China.
3. Modelar un estimador predictivo de costos, según últimas cotizaciones y data histórica. Con esto podemos marginar según la predicción y además entregar de manera rápida una tarifa a los clientes de Shippter. Esto sería empleando algún tipo de algoritmo de machine learning.

Solución escogida:

Para la selección de solución se realizó una categorización de las soluciones presentadas previamente. Se tomó la decisión en base a los siguientes criterios:

- **Impacto:** La solución escogida Impacta Ingenierilmente de forma positiva a la empresa. (30%)
- **Alcance:** La solución es posible realizar dentro de los plazos del proyecto (25%)
- **Costo de implementación:** La solución propuesta tiene bajos costo de implementación, además que incurra a bajo costo operativos en el tiempo. (20%)

- **Rapidez de tarifa:** Gracias a la solución propuesta se pueda entregar una tarifa rápida al cliente el cual es un requerimiento de la gerencia comercial de Shippter. (25%)

La solución será escogida mediante una ponderación de los criterios anteriormente descritos con una escala del 1 al 5, siendo que 1 no logra cumplir el criterio y 5 cumple con el criterio.

A continuación la tabla de los resultados de selección con las ponderación obtenidas:

Solución	Impacto (30%)	Alcance (25%)	Costo de Implementación (20%)	Rapidez de tarifa (25%)	Total Ponderado
1	4	3	3	2	3,05
2	3	1	2	3	2,3
3	4	4	4	5	4,25

Imagen 6: Resultados de selección¹²

La solución con mayor puntaje es la 3, generar un modelo predictivo de los costos. Esta propuesta toma distancia de las demás con un puntaje de 4,25. Esto debido a que cumple de muy buena manera los criterios expuestos. Llama la atención al ser la que mejor destaca en el criterio de rapidez de tarifa, dado que predecir los costos supondría la posibilidad de enviar la tarifa de venta al cliente de manera inmediata, lo cual es un factor sumamente valorado por Shippter

Plan de implementación:

Para la implementación de la solución se requieren 3 pasos importantes:

1. Predicción paralela con operaciones actuales: Al ser el predictor un mvp, este puede que tenga severas desviaciones respecto a la realidad, a medida que Shippter cierre alguna operación de Estados Unidos y obtenga los costos reales de la operación, el modelo en manera paralela hará una predicción con la distancia recorrida por el contenedor, de esta manera ver la diferencia entre el estimador y lo real.

¹² Rodrigo Valenzuela, 2023. Desarrollo del proyecto

2. Ajuste del modelo: Según la diferencias con las nuevas operaciones, ajustar el modelo y entrenarlo nuevamente, además de agregar los datos de operación a la base de datos del predictor, para así pueda seguir aprendiendo y mejorando. Considerando si los datos de la información son relevantes para el modelo.
3. Implementación en Back-end de plataforma online: Después de hacer las correcciones habilitar la sección de cotizar EXW en plataforma con el estimador de costo integrado. Mantenerlo arriba por dos semanas para posteriormente realizar un análisis de indicadores.
4. Repetir paso 2, posteriormente el encargado de pricing deberá hacer el reajuste cada mes para mantener actualizado el estimador.

Es importante tener en consideración que este plan de implementación debería ser iterativo a largo plazo, ya que pueden generarse fluctuaciones que no estén consideradas en la base de datos original del proyecto.

Analisis Riesgo:

Dado que ya tenemos una solución seleccionada, amerita realizar un análisis de riesgo respecto a esta solución, con el fin de poder ver posibles casos de riesgo y eventualmente generar mitigaciones para reducir la incertidumbre del proyecto.

Para esto existen 3 variables importantes a considerar: el evento de riesgo a analizar, la probabilidad de que este evento ocurra y el impacto que este evento tendría para la empresa. A continuación podemos observar una visualización de la matriz de riesgo que representa la relación entre el impacto y la probabilidad:

Probabilidad	Impacto				
	Mínimo	Moderado	Serio	Elevado	Grave
Casi Seguro	Aceptable	Moderado	Alto	Muy alto	Muy alto
Probable	Aceptable	Aceptable	Moderado	Alto	Muy alto
moderado	Bajo	Aceptable	Aceptable	Moderado	Alto
Poco Probable	Muy bajo	Bajo	Aceptable	Aceptable	Moderado
Raro	Muy bajo	Muy bajo	Bajo	Aceptable	Aceptable

Imagen 7: Matriz de riesgo¹³

Se puede apreciar según la matriz que existen 6 niveles de riesgo según la ponderación que lleve entre la probabilidad y el impacto del evento.

Eventos de riesgo:

A continuación se presentarán los eventos de riesgo que estarían presentes en el proyecto:

1. Baja cotización: Esto se produciría que al implementar el estimador, los cliente no coticen en plataforma
2. Error de Codeo: Que al momento de implementar se presenta error de código, ya sea con la compatibilidad de plataforma o con estimaciones en específico.
3. Desviación de tarifa: Que la tarifa de la predicción esté muy alejada del costo real de la operación
4. Tarifas Desactualizadas: Que las tarifa de predicción de costo sean muy altas, en este caso dejarían de ser tarifas competitivas
5. Cambio Organizacional: El cotizador es una herramienta de “combate” para los ejecutivos de venta, este evento ocurriría si no utilizan el estimador a implementar.

Luego una vez planteado los eventos de riesgo al aplicarlos en la matriz de riesgo quedaría de la siguiente forma:

Evento	Probabilidad	Impacto	Riesgo
Baja cotización	Probable	Serio	Moderado
Error de Codeo	Raro	Serio	Bajo
Desviación de Tarifa	Probable	Elevado	Alto
Tarifas Desactualizada	Moderado	Serio	Aceptable
Cambio Organizacional	Poco Probable	Moderado	Bajo

Imagen 8: Matriz de riesgo¹⁴

Como se puede apreciar en la tabla anterior, hay eventos los cuales hay que tomar acciones de mitigación para bajar la incertidumbre del proyecto, serían los que tienen riesgo moderado hasta alto. Las mitigaciones serían las siguientes:

Mitigación “Baja cotización”: Realizar una campaña de ventas enfocado a promocionar la rapidez nuevamente del cotizador automático, con el fin de que los clientes se incentiven en cotizar y en consecuencia a más operaciones cerradas.

¹³ Rodrigo Valenzuela,2023. Desarrollo del proyecto

¹⁴ Rodrigo Valenzuela,2023. Desarrollo del proyecto

Mitigación “Desviación de Tarifa”: Este evento ocurriría si hay una desviación excesiva del estimador, en este caso en el plan de implementación habría que reducir el reajuste del modelo a cada semana para poder verificar cual es el error que lo provoca.

Medidas de desempeño:

Los principales KPI que utilizaremos para medir el desempeño del proyecto son los siguientes:

Reducción en variabilidad: Consiste en el porcentaje que cambió la variabilidad del margen promedio sobre los costos EXW en comparación entre los datos iniciales y el mes de prueba. Tiene la siguiente fórmula:

$$\text{Reducción en Variabilidad (\%)} = \left(\frac{\text{Variabilidad Inicial} - \text{Variabilidad Post Implementación}}{\text{Variabilidad Inicial}} \right) \times 100$$

Certeza: Consiste en el porcentaje de certeza que fue el predictor respecto al costo real de la operación. Tiene la siguiente fórmula:

$$\text{Certeza (\%)} = \left(\frac{\text{Costo Estimado} - \text{Costo real}}{\text{Costo Estimado}} \right) \times 100$$

Operaciones EXW: Consiste en la cantidad de operaciones FCL fueron de tipo EXW en el periodo de prueba, para verificar si el proyecto está realizando un impacto en la estrategia de ventas. Tiene la siguiente fórmula:

$$\text{Operaciones EXW (\%)} = \left(\frac{\text{Operaciones EXW}}{\text{Total Operaciones}} \right) \times 100$$

Tiempo de cotización: Consiste en el tiempo que se demora en tener disponible la tarifa para el cliente desde el momento en que se recibe los datos de la solicitud de cotización.

$$\text{Tiempo de cotización (hrs)} = \text{tiempo de tarifa enviada} - \text{tiempo de cotización recibida}$$

Desarrollo del proyecto:

Para este punto se planteó, dado los plazos, el alcance datos e información del proyecto, la elaboración de un prototipo mínimo viable que logre cumplir el mínimo del objetivo general y de objetivos específicos.

Base de datos:

Para la recolección de datos se utilizaron datos de costos pick-ups solicitados a proveedores de Shippter. Por contrato únicamente se puede solicitar esta tarifa cuando un cliente de Shippter realiza una solicitud de cotización marítima. La base de datos en la cual se realizó el análisis tiene 40 observaciones recopiladas entre los meses de septiembre, octubre y noviembre. Esto debido al poco flujo que solicitan los clientes de este tipo de tarifas. Las variables disponibles son las siguientes:

Cotización: Identificador de la solicitud de cotización enviada al proveedor.

Origen: Lugar de origen y código postal de la dirección de recogida para el pick-up

Puerto: Puerto al cual se dirige el contenedor dentro de la cotización

Volumen: Volumen de contenedor de la cotización, estos pueden ser de 20' o 40' pies.

Costo/Contenedor: El costo por contenedor de trasladarlo desde el punto de origen al puerto correspondiente. Se utilizará como divisa el dólar estadounidense (USD), esta sería nuestra variable más importante dado que es la que queremos predecir.

Fecha: Fecha en donde se recibió la tarifa solicitada al proveedor.

Proveedor: Cuál fue el proveedor que nos facilitó la tarifa.

West or East: Variable que identifica si el movimiento se hizo en la zona este u oeste de Estados Unidos.

Al ser pocas observaciones el modelo predictivo podría no llegar a una exactitud ideal, sin embargo, obtener esta herramienta le será de suma utilidad a la empresa, más aún al equipo comercial para agilizar el envío de cotizaciones.

Limpieza y cambios en Base de datos

Para el manejo de los datos y del proyecto se utilizara el lenguaje de programación python, las principales librerías a utilizar serán pandas y numpy para la limpieza de datos. La base de datos inicial se puede ver a continuación:

```
import numpy as np
import pandas as pd
✓ 0.0s
```

```
coti = pd.read_csv("https://docs.google.com/spreadsheets/d/1Hw64pkkQrA5lxf5SnYT-Og1xdc7296A8rIwDR3u4Y/export?gid=0&format=csv")
coti.head()
✓ 1.1s
```

	cotizacion	Origen	Puerto	Volumen	costo/contenedor	fecha	Proveedor	West o East
0	MFEXW5707	Evans City, PA (16033)	New york	1x40	1150	23/11/2023	MSL	E
1	MFEXW5665	Monroe Township, NJ (08831	New york	1x20	990	14/11/2023	MSL	E
2	MFEXW5643	Houston, TX (77041)	Houston	4x40	950	10/11/2023	MSL	E
3	MFEXW5643	Houston, TX 77041	Houston	4x40	970	10/11/2023	TT	E
4	MFEXW5586	Sumner, WA (98390)	Seattle	1x20	1090	31/10/2023	MSL	W

Imagen 9: BBDD Inicial¹⁵

Posteriormente se da origen a la variable “distancia”, dado que los camioneros en Estados Unidos cambian sus costos según la distancia recorrida a los puertos; este dato no es facilitado en la BBDD original. Se utilizo la libreria Geopy para calcular la distancias utilizando los datos de “origen” y “puerto” de la BBDD original como se muestra a continuación:

```
from geopy.distance import geodesic
from geopy.geocoders import Nominatim

app = Nominatim(user_agent="Exw-distance")
def distance( origen, destino):

    localizacion1 = origen

    location1 = app.geocode(localizacion1).raw
    latitude1 = location1["lat"]
    longitude1 = location1["lon"]

    localizacion2 = destino

    location2 = app.geocode(localizacion2).raw
    latitude2 = location2["lat"]
    longitude2 = location2["lon"]

    distancia = geodesic((latitude1, longitude1), (latitude2, longitude2))

    return(distancia)
✓ 0.0s
```

```
coti["distancia"]=np.zeros(shape=(37,1))
for i in range(len(coti)):
    coti["distancia"][i] = distance(coti["Origen"][i],coti["Puerto"][i]+" Port,USA")
```

Imagen 10: Origen Variable Distancia¹⁶

¹⁵ Rodrigo Valenzuela,2023. Desarrollo del proyecto

¹⁶ Rodrigo Valenzuela,2023. Desarrollo del proyecto

Por último se separó la variable Volumen en dos variables, una con el mismo nombre que indica la cantidad de contenedores de la cotización, y una segunda variable de nombre “tipo” la cual indica el tipo de contenedor de la cotización. Se puede apreciar el resultado en la siguiente imagen:

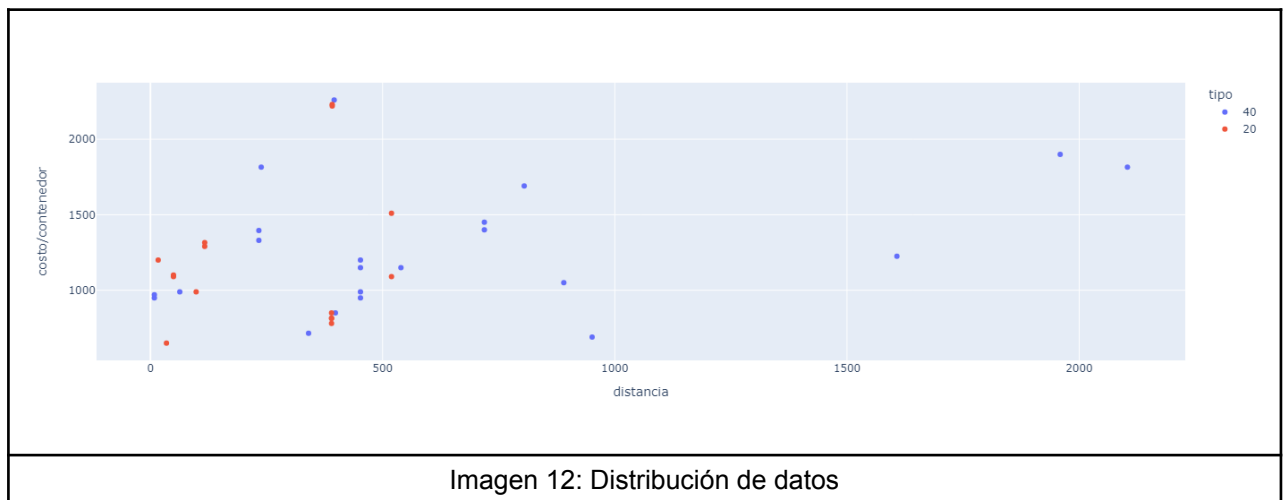
```
coti.head()
```

✓ 0.0s

	cotizacion	Origen	Puerto	Volumen	costo/contenedor	fecha	Proveedor	West o East	distancia	tipo
0	MFEXW5707	Evans City, PA (16033)	New york	1	1150	23/11/2023	MSL	E	539.45	40
1	MFEXW5665	Monroe Township, NJ (08831)	New york	1	990	14/11/2023	MSL	E	98.41	20
2	MFEXW5643	Houston, TX (77041)	Houston	4	950	10/11/2023	MSL	E	8.45	40
3	MFEXW5643	Houston, TX 77041	Houston	4	970	10/11/2023	TT	E	8.45	40
4	MFEXW5586	Sumner, WA (98390)	Seattle	1	1090	31/10/2023	MSL	W	49.64	20

Imagen 11: BBDD¹⁷

Ahora bien, puesto que para tener una noción de los datos atípicos se realizó un gráfico de dispersión entre las variables cuantitativas “costo/contenedor” ,“distancia” y con la variable cualitativa “tipo” para comprobar la existencia de estos outliers y si es necesario eliminarlos:



Podemos ver en el gráfico anterior que existen algunos puntos atípicos, estos pueden distorsionar seriamente la predicción de nuestro estimador por lo que se procederá a eliminarlos. A continuación la visualización del procedimiento de identificación/eliminación de outliers y el gráfico de caja de ambas variables (“costo/contenedor” y “distancia”) que justifica dicha acción:

¹⁷ Rodrigo Valenzuela,2023. Desarrollo del proyecto

```
# Función para calcular los límites para identificar outliers
def calculate_outlier_bounds(column):
    Q1 = column.quantile(0.25)
    Q3 = column.quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    return lower_bound, upper_bound

# Calculando límites para 'costo/contenedor' y 'distancia's
lower_bound_costo, upper_bound_costo = calculate_outlier_bounds(data['costo/contenedor'])
lower_bound_distancia, upper_bound_distancia = calculate_outlier_bounds(data['distancia'])

# Filtrando los outliers
filtered_data = data[(data['costo/contenedor'] >= lower_bound_costo) &
                    (data['costo/contenedor'] <= upper_bound_costo) &
                    (data['distancia'] >= lower_bound_distancia) &
                    (data['distancia'] <= upper_bound_distancia)]

# Comparando tamaños antes y después de la eliminación de outliers
print("Tamaño original:", data.shape[0])
print("Tamaño después de filtrar outliers:", filtered_data.shape[0])
```

Imagen 13: Borrado de outliers¹⁸

A continuación gráfico de caja que visualiza los outliers:

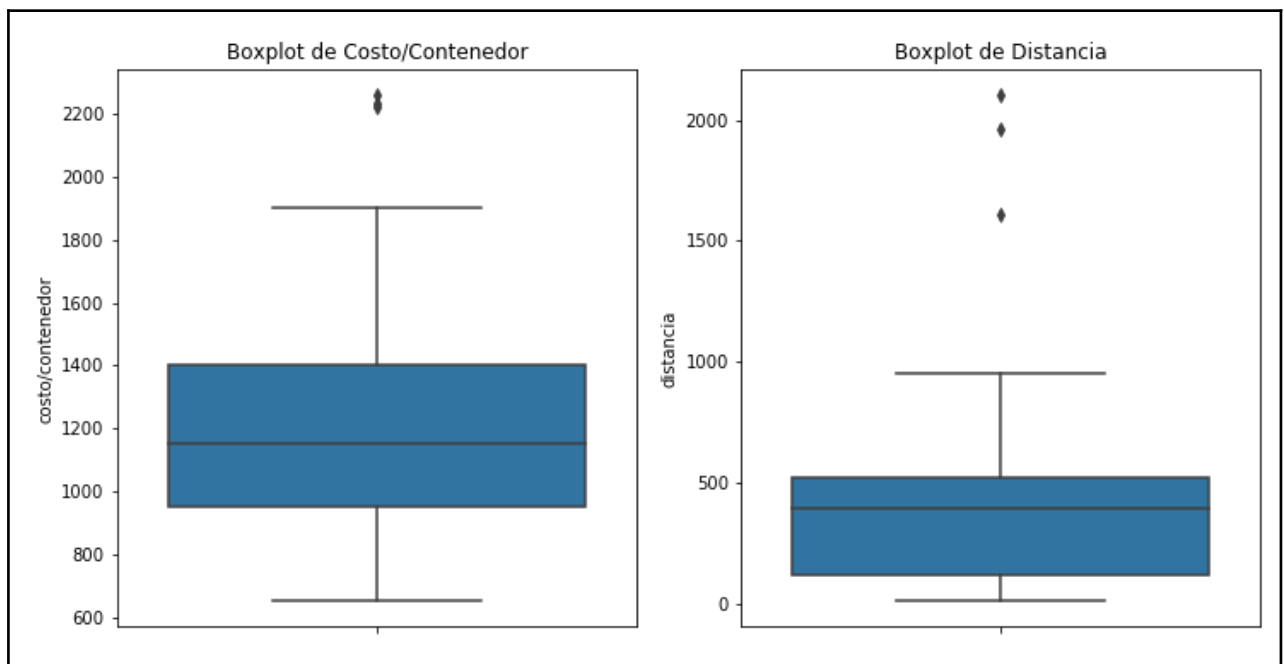


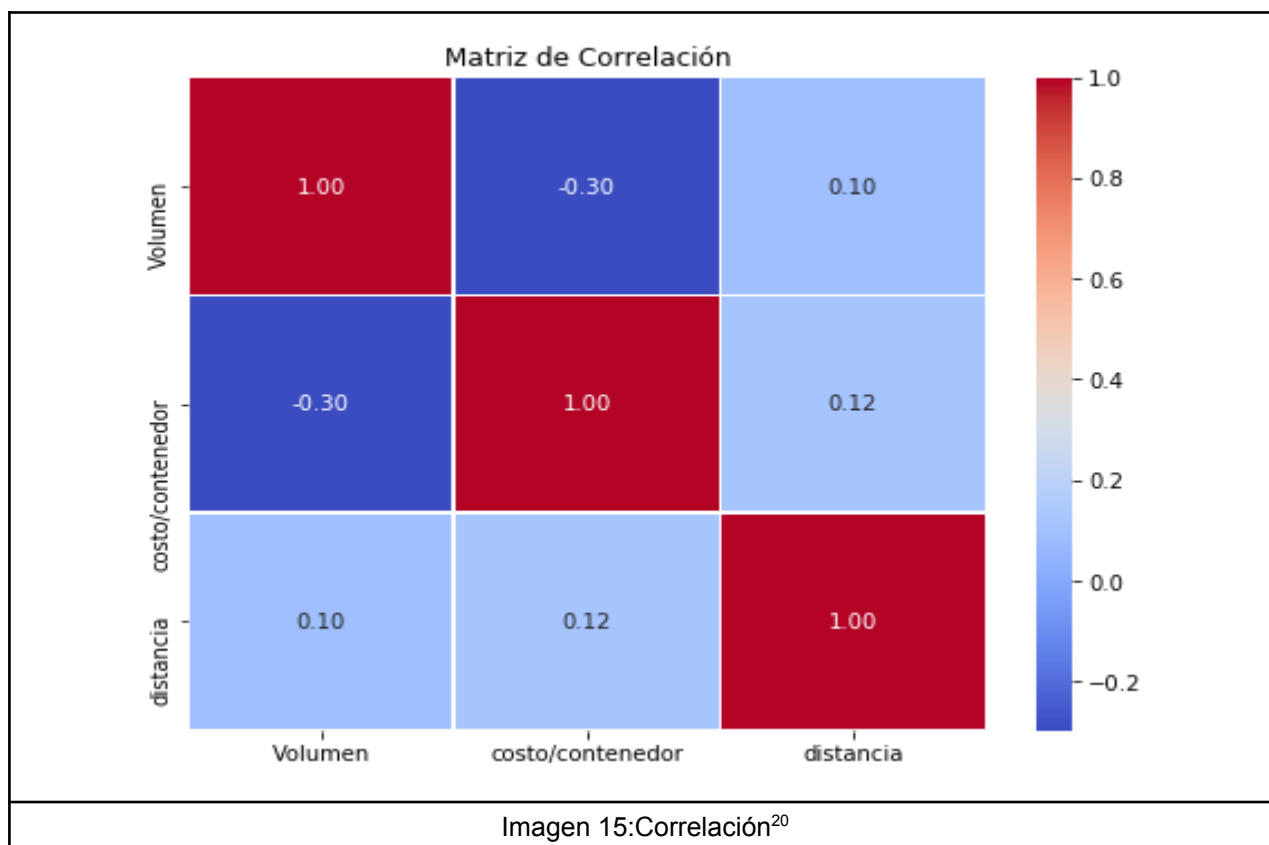
Imagen 14: Diagrama de caja de “Costo/contenedor” y “Distancia”¹⁹

Como se puede ver del gráfico anterior los datos atípicos estaban muy distantes de la mayoría de los datos este factor se ve claramente en la variable de distancia (gráfico de la derecha). Ya que

¹⁸ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

¹⁹ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

poseemos pocas observaciones al tener tal cantidad de datos atípicos afectará de manera importante el modelo predictivo. Ahora bien como último de análisis de la base de datos se realizará un gráfico de la correlación de las variables cuantitativas:



Se puede ver como afecta la variable “volumen” y “distancia” a nuestra variable dependiente, siendo la primera con una correlación negativa y la segunda con una positiva. Esto tiene sentido con la naturaleza de los datos en razón de que a un mayor volumen de contenedores se espera que el embarcador haga un descuento de costo por contenedor de la cotización; también que a mayor distancia el camión tenga que recorrer, el costo suba de una manera leve.

Modelos Predictivos:

Para este punto se evaluaron dos modelos predictivos, primero se replicará el modelo de regresión cuantílica para 3 cuartiles y un algoritmo de aprendizaje supervisado: el modelo de bosques aleatorios, dado que este modelo es versátil para pequeños conjuntos de datos.

²⁰ Rodrigo Valenzuela,2023. Desarrollo del proyecto

Las variables que utilizaremos para los modelos predictivos son: “Tipo”, “Distancia”, “Volumen”, “West or East” y la variable dependiente “costo/contenedor”

Primero para poder evaluar los modelos cambiamos las variable “West or East” a una variable binaria de nombre “West_o_East_encoded” y “tipo” a “tipo_encoded”, siendo 0 para “East” en la primera variable y 20 para la segunda variable.

```
from sklearn.preprocessing import LabelEncoder

# Preparación de los datos
# 1. Convertir la variable 'West o East' en un formato numérico
label_encoder = LabelEncoder()
filtered_data['West_o_East_encoded'] = label_encoder.fit_transform(filtered_data['West o East'])
filtered_data['tipo_encoded'] = label_encoder.fit_transform(filtered_data['tipo'])
```

Imagen 16: Variables Binarias ²¹

Ahora que las variables son numéricas podemos continuar entrenando los modelos.

Regresión Cuantílica:

Para la regresión cuantílica se entrenó el modelo para 3 cuantiles (25%, 50% o mediana y 75%) a continuación el desarrollo de la regresión:

```
import statsmodels.formula.api as smf

# Definiendo los cuantiles a analizar
quantiles = [0.25, 0.50, 0.75]

# Creando un modelo de regresión cuantílica para cada cuantil
models = {}
for qt in quantiles:
    # La fórmula especifica la relación entre las variables
    model = smf.quantreg('y ~ distancia + Volumen + West_o_East_encoded + tipo_encoded', filtered_data)
    results = model.fit(q=qt)
    models[qt] = results

# Mostrando los resultados de los modelos
model_summaries = {qt: models[qt].summary() for qt in quantiles}
model_summaries[0.25], model_summaries[0.50], model_summaries[0.75]
```

Imagen 17: Código Regresión Cuantílica²²

²¹ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

²² Rodrigo Valenzuela, 2023. Desarrollo del proyecto

Para comparar las distintas curvas de la regresión se utiliza el pseudo R^2 el cual cumple la misma función que el R^2 de la regresión lineal, analizar qué tanto explica el modelo de la variable dependiente. Primero veremos el resultado del primer cuartil a analizar:

QuantReg Regression Results						
Dep. Variable:	y	Pseudo R-squared:	0.1786			
Model:	QuantReg	Bandwidth:	485.8			
Method:	Least Squares	Sparsity:	985.5			
Date:	Mon, 11 Dec 2023	No. Observations:	31			
Time:	20:26:04	Df Residuals:	26			
		Df Model:	4			
	coef	std err	t	P> t	[0.025	0.975]
Intercept	930.0000	209.926	4.430	0.000	498.491	1361.509
distancia	1.901e-09	0.334	5.69e-09	1.000	-0.686	0.686
Volumen	-115.0000	91.512	-1.257	0.220	-303.105	73.105
West_o_East_encoded	275.0000	267.857	1.027	0.314	-275.588	825.588
tipo_encoded	175.0000	206.671	0.847	0.405	-249.818	599.818

Imagen 18: Primer cuartil regresión cuantílica²³

Podemos apreciar del resultado que ninguna variable independiente es significativa en el primer cuartil para poder predecir el costo, sin embargo, el pseudo R^2 es sumamente bajo para poder ayudar a una predicción de tarifa. Continuemos con el siguiente cuartil:

QuantReg Regression Results						
Dep. Variable:	y	Pseudo R-squared:	0.1441			
Model:	QuantReg	Bandwidth:	414.3			
Method:	Least Squares	Sparsity:	796.1			
Date:	Mon, 11 Dec 2023	No. Observations:	31			
Time:	20:26:04	Df Residuals:	26			
		Df Model:	4			
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1105.1102	146.747	7.531	0.000	803.468	1406.753
distancia	-0.1759	0.291	-0.605	0.550	-0.774	0.422
Volumen	-97.8003	70.186	-1.393	0.175	-242.070	46.469
West_o_East_encoded	101.4215	227.013	0.447	0.659	-365.210	568.054
tipo_encoded	237.5773	163.456	1.453	0.158	-98.411	573.566

Imagen 19: Segundo cuartil regresión cuantílica²⁴

²³ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

²⁴ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

Para el segundo cuartil ocurre una situación similar al primero, ninguna variable es significativa para el modelo, sin embargo a este nivel ya se acercan las variables de “Volumen” y “tipo” a ser significativa, lo cual podemos inferir que a un nivel de más alto de costo importa el volumen y tipo de contenedor que se quiere traer, el pseudo R^2 sigue siendo una limitante para el modelo. El último cuartil presenta lo siguiente:

QuantReg Regression Results						
Dep. Variable:	y	Pseudo R-squared:	0.1673			
Model:	QuantReg	Bandwidth:	485.8			
Method:	Least Squares	Sparsity:	1001.			
Date:	Mon, 11 Dec 2023	No. Observations:	31			
Time:	20:26:04	Df Residuals:	26			
		Df Model:	4			
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1340.7221	142.855	9.385	0.000	1047.080	1634.365
distancia	0.0103	0.341	0.030	0.976	-0.692	0.712
Volumen	-140.8944	81.235	-1.734	0.095	-307.875	26.087
West_o_East_encoded	41.0135	281.386	0.146	0.885	-537.384	619.411
tipo_encoded	192.7683	186.252	1.035	0.310	-190.078	575.614

Imagen 20: Último cuartil regresión cuantílica²⁵

El pseudo R^2 sigue siendo bajo para el modelo, a pesar de que varias variables se acercan a la significancia, esto ocurre principalmente por dos factores, la cantidad de datos es muy baja para que la predicción del modelo sea fiable y la variabilidad de la muestra es muy alta. Cabe destacar que este modelo a largo plazo logrará tener mayor relevancia, dado por su robustez a los datos atípicos y la heterocedasticidad, por lo que sería interesante volver a entrenar el modelo cuando se tengan más de 200 observaciones.

Modelo de Bosques aleatorios:

Utilizaremos las mismas variables independientes para este modelo, la fortaleza de este algoritmo supervisado es la robustez que tiene frente a la dispersión, datos atípicos y a una cantidad reducida de datos. El desarrollo del modelo fue el siguiente:

²⁵ Rodrigo Valenzuela, 2023. Desarrollo del proyecto


```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
import seaborn as sns

# Carga tus datos
# Asegúrate de cambiar 'tu_archivo.csv' por la ruta de tu archivo de datos
data = filtered_data

# Preparando los datos
# Convierte 'West o East' a una variable numérica
data['West_o_East_encoded'] = pd.get_dummies(data['West o East'], drop_first=True)
X = data[['distancia', 'West_o_East_encoded', 'tipo_encoded', "Volumen"]] # Variables predictoras
y = data['costo/contenedor'] # Variable objetivo

# Dividiendo los datos en conjunto de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Ajustando un modelo de árbol de decisión
tree_model = DecisionTreeRegressor(random_state=42)
tree_model.fit(X_train, y_train)

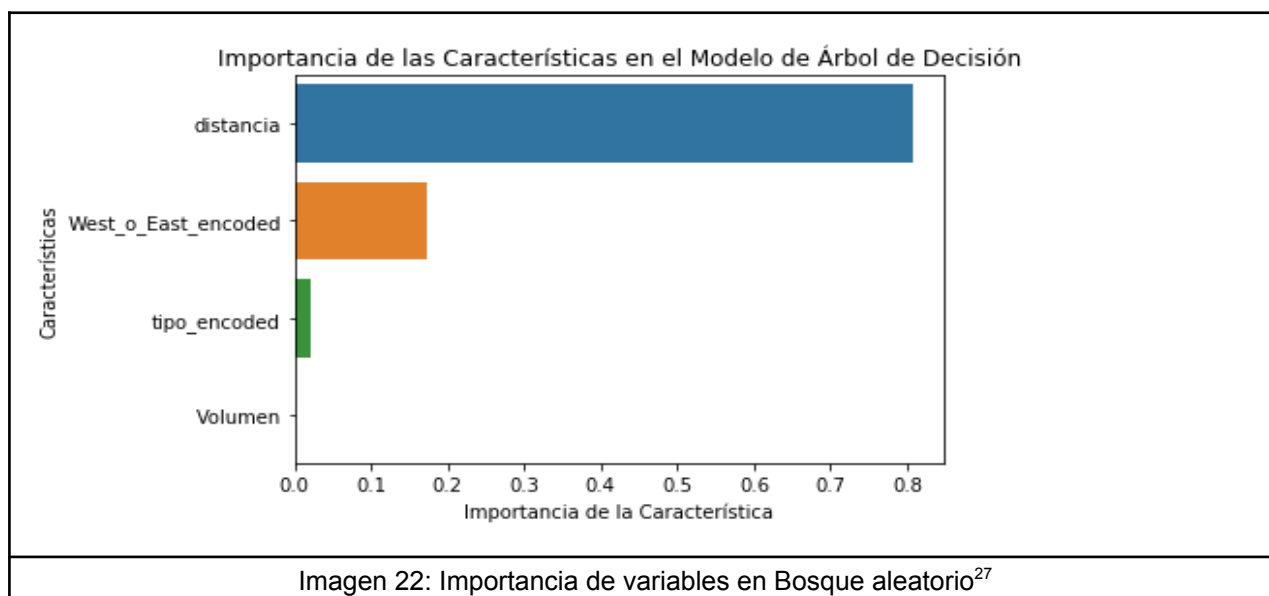
# Prediciendo en el conjunto de prueba
y_pred = tree_model.predict(X_test)

# Evaluando el modelo
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

Imagen 21: Código Modelo de bosque aleatorios²⁶

Como se puede ver, se utilizó un 30% de la base de datos para el testeado del modelo utilizando las mismas variables cuantitativa y cualitativas que en el modelo anterior; como resultado el modelo de árbol aleatorio nos da un $R^2 = 0.35$. Por lo tanto nos dice que este modelo nos ayudará de mejor manera que la regresión cuantílica para la predicción de costos. Además podemos ver que tan influyente son las variables independientes en el siguiente gráfico:

²⁶ Rodrigo Valenzuela, 2023. Desarrollo del proyecto



Con esto podemos ver que distancia es la variable más importante al realizar predicciones, lo cual es lo más adecuado dado al contexto del problema. Con todo lo anterior podemos ver que de las herramientas propuestas la más adecuada es la de árboles aleatorios para nuestro problema.

Resultados:

Para los resultados, se utilizó la información de solicitud de cotización de noviembre, se remitió la tarifa a los clientes con la predicción del modelo y se esperó la respuesta del proveedor para comparar las diferencias que estás podrías obtener. Se obtuvieron cerca de 8 observaciones.

De manera paralela se persuadió al equipo comercial que promocionan las importaciones EXW para poder estimar el modelo. A continuación una tabla con los resultados según los KPI's estipulados en medidas de desempeño:

KPI	Valor Actual	Valor Esperado	Resultados
Reducción en Variabilidad	20%	14%	18%
Certeza	-	10%	6%
Operaciones EXW	10%	20%	12%
Tiempo de cotización	36 Hrs	0	0

Imagen 23: Resultados²⁸

²⁷ Rodrigo Valenzuela,2023. Desarrollo del proyecto

²⁸ Rodrigo Valenzuela,2023. Desarrollo del proyecto

Podemos ver que para la reducción de variabilidad se logró reducir, sin embargo, no al nivel deseado anteriormente expuesto, esto debido al bajo nivel de observaciones que se lograron obtener en el periodo de prueba. Cabe resaltar que se mejoró el nivel antes de la presentación del proyecto.

Para el KPI certeza, no se tenían datos pre-proyecto, dado que no se utilizaba un predictor para las importaciones estipuladas para este trabajo. Se puede apreciar que es de los KPI que tuvieron peor desempeño. Lo más probable que sea por la leve cantidad de datos que se utilizaron para el entrenamiento del modelo.

El KPI Operaciones EXW logró tener un mejor desempeño que el kpi anterior, dado que se logró mejorar la situación actual de la empresa incrementando levemente la cantidad de importación EXW, no se logra al nivel deseado debido a estacionalidades del mercado, ya que usualmente en el periodo que se realiza el testeó las importaciones disminuyen a nivel global.

Por último el tiempo de cotización es el indicador con más éxito, dado que es una predicción realizada a los costos, la tarifa puede ser entregada inmediatamente a los clientes, por lo que el tiempo resultante de espera es 0.

Evaluación Económica:

Para la evaluación económica según los datos facilitados por Shippter este sería el flujo de caja para lo que alborada el proyecto (Importaciones FCL) considerando la implementación de este:

	0	1	2	3	4
Ingresos		\$35.543.786	\$35.987.986	\$36.547.347	\$36.434.354
Costos Variables		-\$21.326.272	-\$21.592.792	-\$21.928.408	-\$21.860.612
Costos Fijos		-\$7.000.000	-\$7.000.000	-\$7.000.000	-\$7.000.000
Flujo Operacional		\$7.217.514	\$7.395.194	\$7.618.939	\$7.573.742
Inversión Fija	-\$6.200.0000				

Imagen 24:Flujo de Caja²⁹

²⁹ Rodrigo Valenzuela,2023. Desarrollo del proyecto

La inversión Inicial considera los siguiente costo que se utilizan solamente para la implementación, considerando que posteriormente los trabajadores del área de pricing puedan mantener en ejecución el proyecto:

Activo	Costo aproximado
Analista de datos	1.500.000
Desarrollador	2.300.000
Servidor	1.000.000
Ingeniero	1.400.000
Imagen 25: Costo inversión ³⁰	

Ahora bien, en conjunto con lo anterior es importante realizar un análisis de rentabilidad del proyecto. Para ello se investigaron los indicadores de riesgos y libre de riesgo presentes en Chile como se puede ver en la siguiente tabla:

Tasa de mercado (Chile)	4,53%
Forwarding	1,1
Premio por riesgo (Chile)	-1,07%
Tasa libre de riesgo	5,60%
Imagen 26: Tasas	

Para la tasa libre de riesgo³¹ y la tasa de mercado³² se obtuvieron de Investing y para el tasa de Forwarding se utilizó el beta histórico de Maersk³³ el cual es una empresa presente en la misma industria. El cálculo del CAPM el cual da como resultado el siguiente:

CAPM	4,42%
Imagen 27: Cálculo CAPM	

Con esto puede calcular el VAN y TIR los cuales son los siguientes:

³⁰ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

³¹ Investing, 2023. Rentabilidad del bono Chile 10 años
<https://es.investing.com/rates-bonds/chile-10-year-bond-yield>.

³² Investing, 2023. S&P CLX IPSA (SPIPSA)
<https://es.investing.com/rates-bonds/chile-10-year-bond-yield>.

³³ Infrotanalytics, 2023, A.P. Moller - Maersk A/S
<https://www.infrontanalytics.com/fe-EN/301645D/A-P-Moller-Maersk-A-S/Beta>

TIR	95%
VAN	\$19.554.828
Imagen 28: Van y TIR ³⁴	

Dado que el VAN y TIR son positivos, el proyecto es rentable, este debido principalmente a la baja inversión Inicial.

Conclusiones:

Es posible rescatar que la solución planteada logra de manera ajustada, completar algunos de los objetivos impuestos, a pesar de la leve información disponible para el desarrollo del proyecto. Además, el proyecto surge como una nueva “herramienta” para que shippter pueda incrementar su valor comercial y seguir el alineamiento de su propuesta de valor de tarifas rápidas y competitivas.

Dada a la naturaleza de los modelos predictivos este proyecto es un punto inicial para que la empresa pueda seguir escalando y afinando el modelo. Así lograr aumentar los R cuadrado de los modelos y obtener modelos más consistentes de los presentados en este trabajo. Se recomienda que para desarrollar de manera apresurada el modelo predictivo, se logren alianzas con proveedores para ampliar la BBDD, dado que al momento de desarrollo de este proyecto no se logró aquello y fue una limitante importante para la calidad de los modelos predictivos.

Finalmente se puede concluir que el proyecto logra satisfacer de manera ajustada la necesidad principal de la empresa. Logrando alguno de los objetivos específicos propuestos

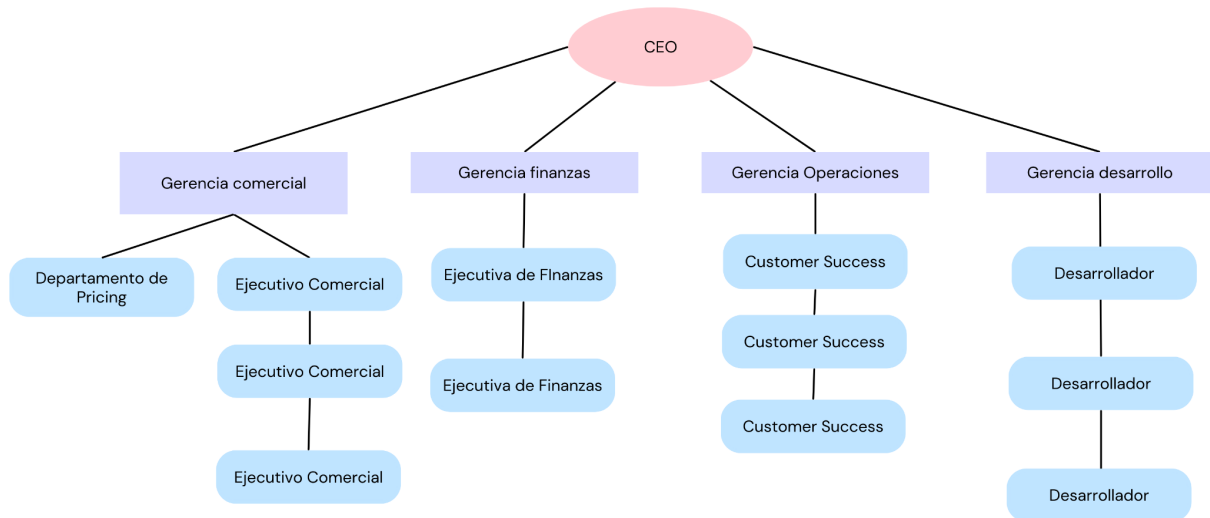
³⁴ Rodrigo Valenzuela, 2023. Desarrollo del proyecto

Referencias:

1. Javier D. Martin. (2015). Estimación de tarifas de fletes a través del uso del método de la Regresión Cuantílica (Tesis de maestría, Universidad nacional de córdoba). rdu.unc.edu.ar
2. ManufacturingTomorrow. (2021, 31 de agosto). Variability Reduction: Why Important To Manufacturers?
ManufacturingTomorrow. <https://www.manufacturingtomorrow.com/article/2021/08/variability-reduction-why-important-to-manufacturers/18768/>
3. SuperProcure. (Fecha no especificada). Digital reverse auction empowers freight procurement strategy. SuperProcure. Recuperado de <https://www.superprocure.com/blog/digital-reverse-auction-empowers-freight-procurement-strategy/>
4. Investing, 2023. Rentabilidad del bono Chile 10 año, <https://es.investing.com/rates-bonds/chile-10-year-bond-yield>.
5. Infrotanalytics, 2023, A.P. Moller - Maersk A/S, <https://www.infrontanalytics.com/fe-EN/30164SD/A-P-Moller-Maersk-A-S/Beta>
6. Investing, 2023.S&P CLX IPSA (SPIPSA), <https://es.investing.com/rates-bonds/chile-10-year-bond-yield>.

Anexos:

1.- Organigrama de Shippter



2.-

imagen

incoterms

