

Memoria del Análisis Exploratorio de Datos

ÍNDICE

1. Introducción
2. Descripción del Proyecto
3. Fuentes de Datos
4. Técnicas Utilizadas
5. Exploración de Datos Descriptiva
6. Análisis Bivariado
7. Visualización de Datos
8. Resultados y Conclusiones
9. Recomendaciones
10. Referencias

Introducción

En esta memoria se documenta el proceso y los resultados del Análisis Exploratorio de Datos (EDA, por sus siglas en inglés) realizado en el proyecto **Optimizando la Eficiencia en Transacciones Portuarias: Un Enfoque Basado en Análisis de Datos**. El EDA es una etapa crucial en el proceso de análisis de datos, que permite comprender mejor el conjunto de datos, identificar patrones, tendencias y anomalías, y orientar el enfoque del análisis posterior.

Descripción del Proyecto

En base al histórico proporcionado por Exnova internacional sobre todas las transacciones portuarias en USA se pretende analizar los retrasos de las mercancías y encontrar patrones donde puedan darnos diferentes hipótesis para poder aplicar soluciones a dichos problemas y mejorar la eficiencia.

Fuentes de Datos

Las fuentes de datos ha sido la misma empresa Exnova internacional proporcionandonos todas las transacciones en un .csv, además de haber extraído información de la web searates para extraer todas las coordenadas y los puertos de USA mediante web scraping y regex y tradingview para extraer una gráfica del precio del petróleo desde el 2020 hasta el 2024.

Técnicas Utilizadas

Python, Pandas, Numpy, Regex Web scraping con la librería BeautifulSoup, matplotlib seaborn y pyplot para realizar gráficas, Folium para la creación de mapas además de la creación de un Dashboard con Looker Studio.

Exploración de Datos Descriptiva

- Resumen estadístico de las variables clave. Las variables más relevantes para la realización del Análisis han sido muchas, como loading_port, unloading_port, weight, estimate_arrival_date y complete_date. Con estos 2 últimos he creado una variable que me permite ver el tiempo total de retrasos en cada transacción creando una nueva columna llamada delay_time.
- Visualización de distribuciones de datos, como histogramas y gráficos de densidad.
- Identificación y manejo de valores atípicos y datos faltantes.

Análisis Bivariado

- Relación entre columnas para el análisis:
delay_time/weight, 'loading_port')[delay_time', shipper_name/delay_time, consignee_name/delay_time

Visualización de Datos

- Uso de gráficos de dispersión, barras para representar diferentes aspectos de los datos.
- Mapas para visualizar datos geoespaciales o categóricos.

Resultados y Conclusiones

Resultados:

Tras el cruce de datos y el análisis del mercado, nos da como resultado que USA tiene dificultades para recibir mercancía portuaria en el oeste del país dado que los puertos Long beach y los Ángeles no tienen la infraestructura necesaria para abarcar el mercado asiático. Además de ver la influencia del precio del petróleo dado que es directamente proporcional a los retrasos de la mercancía. También se ha podido visualizar las complicaciones en los contenedores más pesados y cabe destacar que antes de que pase un mes justo en el día 30 se realizan la mayoría de recepciones.

Conclusiones limitaciones y desafíos:

En conclusión el precio del petróleo es complejo de predecir, pero sí que hay tendencias, se podría realizar predicciones en base a la época del año, carga de la mercancía y precio del petróleo, a esto hay que añadir datos meteorológicos.

La geopolítica probablemente sea lo más complejo de predecir, pero es importante tener en cuenta que en este mercado afecta mucho.

Recomendaciones

Es importante cuidar este tipo de datos dado que cuanto más haya y mejor tratados están, la información puede ser mucho más relevante para poder realizar predicciones o análisis.

Añadir los datos climáticos sería un muy buen dato para poder mejorar el análisis. Además de un seguimiento del precio del petróleo.

También se podrían generar rutas alternativas a los puertos en base a una predicción de retraso.

Referencias

Datos proporcionados por Exnova Internacional, recopilación de coordenadas (https://www.searates.com/es/maritime/united_states) y datos del precio del precio del petróleo en Tradingview