






 Felix-87 / Phase3_DS_project



 Code  Issues  Pull requests  Actions  Projects  Wiki  Security  Insights  Settings



★ 0 stars  0 forks  1 watching  Branches  Activity
 Tags

 Public repository

 1 Branch  0 Tags  













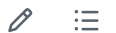


 Felix-87 Merge branch 'main' of https://github.com/Felix-87/Phase3_DS_project 

a11de02 · 5 minutes ago 

 .gitignore	updates on notebooks and dataset	10 hours ago
 README.md	Update README.md	45 minutes ago
 Train.csv	updates on notebooks and dataset	10 hours ago
 draft e_commerce.ipynb	updates on notebooks and dataset	10 hours ago
 index.ipynb	final update commits	12 minutes ago
 presentation.pdf	final update commits	12 minutes ago

 README



E-Commerce Shipping Data Analysis

1.0 Project Overview

This project involves analysis of shipping data to draw insights and build a predictive model on future deliveries if they will be delivered 'On-time' or 'Delayed'. The research follows cross industry standard procedures (CRISP-DM) methodology for the shipping industry (product supply chain).

2.0 Business Understanding

The modern technological advancements has elevated demand for convenience in human day to day life. Every business establishment is keen in tapping to this new norm and to remain afloat, they must understand customer needs and their historical performances in order to improve. The shipping industry is at the heart of supply chain. Orders are placed, received, processed and feedback on processing progress done on online platform. Analysis of data sitting this industry's database will enable identification of the essential factors affecting product shipping. Customers are keen metrics like delivery timelines, costs and even feedback during the whole process. Business entities in the industry with such understanding, will optimize their operations towards satisfying customer needs and will ultimately be rated highly by the same customers.

2.1 Business Objective

Company involved in trade of electronic items in the international market is interested in the analysis of its data in customer database to draw insights and apply machine learning techniques for predictive analytics. This aimed at optimizing its resources towards improving operational efficiency in meeting customer demands.

Specific Objectives

1. *Analyze and optimize shipping performance:* Identify factors influencing shipping delays.
2. *Improved customer satisfaction:* Establishment of areas for improvement regarding customer satisfaction, for example, to make the delivery faster or provide more reliable delivery windows.
3. *Reduce cost of shipping:* Explore data patterns to identify opportunities for reducing shipping costs without compromising delivery speed or customer satisfaction. Discounting to incentivize sales is a cost to the company.
4. *Predict delivery times:* Develop predictive models to accurately forecast delivery times based on historical shipping data.

3.0 Data Sources and Understanding

The Data

The dataset for analysis 'E-Commerce Shipping Data' was drawn from; <https://www.kaggle.com/datasets/prachi13/customer-analytics> on Product Shipment Delivery to Meet E-Commerce Customer Demand.

Content

The dataset used for model building contained 10999 observations of 12 variables.

The data contains the following information:

ID: ID Number of Customers.

Warehouse block: The company has big Warehouse which is divided into block; A,B,C,D,E.

Mode of shipment: The company delivers products either through Ship, Flight or Road.

Customer care calls: The number of the customer enquiry calls.

Customer rating: Customer rating on the company delivery. 1 is the lowest (Worst), 5 is the highest (Best).

Cost of the product: Cost of the product in US Dollars.

Prior purchases: The number of prior purchases.

Product importance: Product categorization into low, medium, high.

Gender: Male and Female.

Discount offered: Discount offered on that specific product.

Weight in gms: It is the weight in grams.

Reached on time: It is the target variable, where 0 Indicates that the product has NOT reached on time and 1 indicates it has reached on time.

Key questions to guide data analysis

1. What are the key determinants of shipping delays?
 - What variables (e.g., source location, shipping method, order volumes, etc.) have the largest impact on the timeliness of deliveries?
2. How do customer demographics and product preferences influence the delivery experience?
 - What various customer ratings and their product preferences affect delivery times?
3. How can we minimize our shipping cost without compromising any service level?
 - Which is the most cost-effective shipment method and route?
4. How do we forecast the delivery times accurately?
 - Develop a predictive model and evaluate based on the shipping data.

4.0 Data Preparation

This is the process of cleaning data in readiness for analysis and it involves the following:

1. Validity check; Checking for irrelevant features and removing them or selecting the relevant features
2. Data completeness; Checking for missing values and treating them.
3. Data accuracy; Checking for outlier values in the data that distort its accuracy.
4. Data consistency; Consistency is achieved through removal of duplicates in the dataframe.
5. Data Uniformity; Involves feature engineering

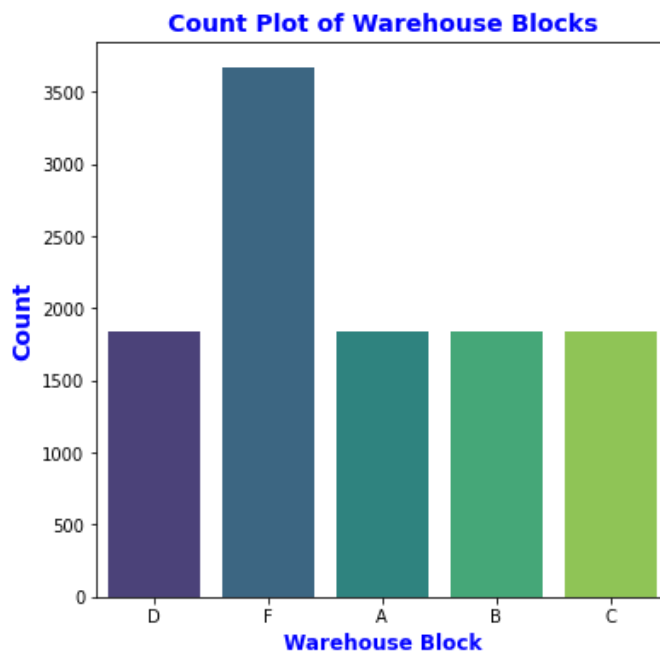
5.0 Analysis and Modelling

Analysis Approach

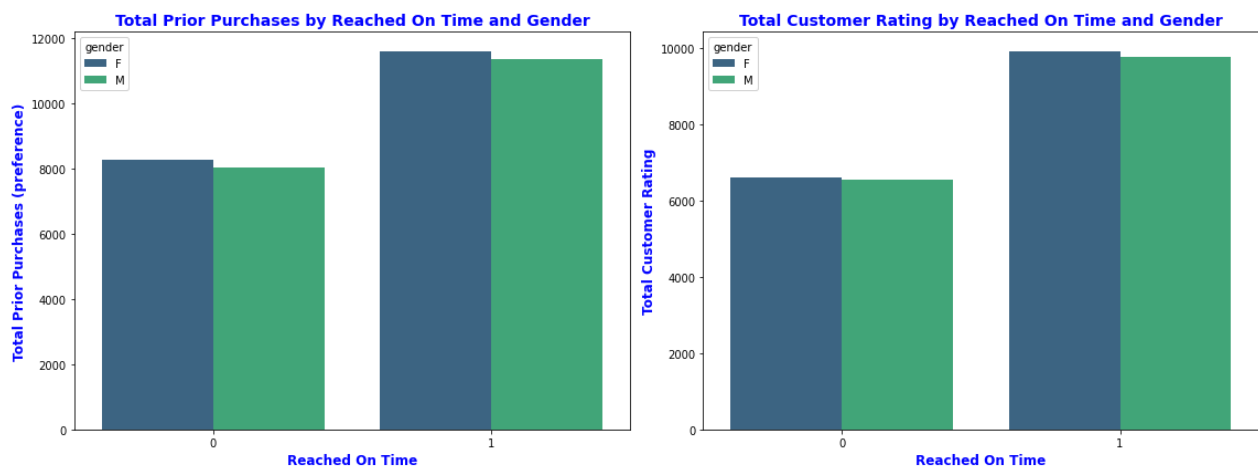
1. Univariate Analysis: Understanding single variables distribution.
2. Bivariate Analysis: Relationships between gender and ratings/preferences.
3. Multivariate Analysis: Combining factors (warehouse_block, cost_per_order and mode of shipment).
4. Develop predictive classification models, evaluate performances and pick the best performing.

5.1 Data Visualizations

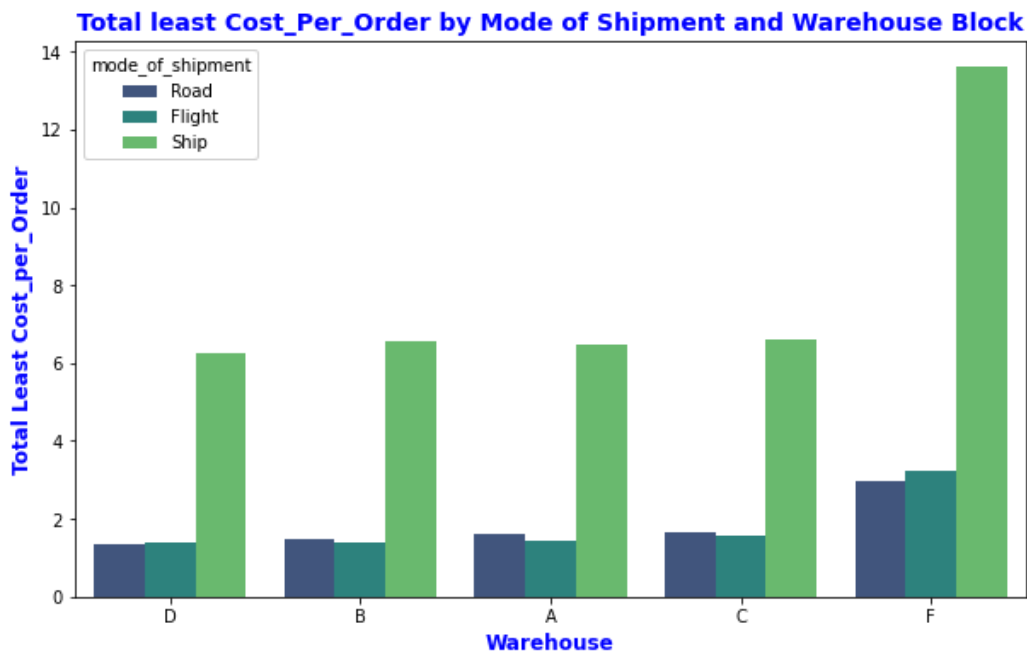
- 1.Count plot for single variable distribution (Warehouse Block)



2. Visualization of prior purchases, customer ratings by reached on time and gender variables



3. Visualization of total least Cost_Per_Order by Mode of Shipment and Warehouse Block



Findings

- *Warehouse block*, majority of the dispatches that ended up being delayed originated from block F
- *Mode of shipment*, movement of shipment by sea(ship) resulted in delays; possibly due to speed of ships which is lower.
- *Product importance*, product categories of low and medium shipped/ordered delayed
- The higher the customer ratings and prior purchases(product preferences) for both genders led to delivery on time.
- Higher customer preference and ratings possibly is as a result of being satisfied with delivery service.
- Shipment by Road and Flight are generally the cost effective modes of shipment where the company incurred the least rate in incentivizing sale.
- Cost effective routes is by shipment through all warehouse blocks except block F
- The top three cost effective routes is shipping through block D by Road, block B by Flight and block D by Flight respectively.
- Decision Tree classification model is the best performing

Use Case

- Clone this repository:
- bash
- Copy code
- git clone https://github.com/Felix-87/Phase3_DS_project.git
- Run financial_analysis.ipynb in Jupyter Notebook or any other compatible IDE.

6.0 Conclusions and Recommendations

6.1 Conclusions

Main factors contributing to shipping delays

- *Warehouse block*, majority of the dispatches that ended up being delayed originated from block F
- *Mode of shipment*, movement of shipment by sea(ship) resulted in delays; possibly due to speed of ships which is lower.
- *Product importance*, product categories of low and medium shipped/ordered delayed

Customer demographics and product preferences influencing delivery experiences

- Both genders exhibit same trend on levels of prior purchases and ratings, though females seem to have made slightly more purchases than males.
- Higher customer preference and ratings possibly is as a result of being satisfied with delivery service.

Cost effective shipping mode and routes

- Shipment by Road and Flight are generally the cost effective modes of shipment where the company incurred the least rate in incentivizing sale.
- Cost effective routes is by shipment through all warehouse blocks except block F
- The top three cost effective routes is shipping through block D by Road, block B by Flight and block D by Flight.



Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

- Jupyter Notebook 100.0%