Shipment Cost Prediction for a Logistics Company

SwiftFreight is committed being the industry leader in logistic services through the continuous investment in technology solutions and talented individuals. SwiftFreight is the middleman between the customer and freight services. They will charge the customer and pay to the freight services and during the transaction they will make their profit/revenue. The challenge is, given a request from a customer for shipment they would like to decide the estimated cost for the carriers. This information will enable them to quote the best price for the customer and reduce the revenue leakages and increase their bottom line.

S.No	Variable	Description
1	ORDER_NBR	Order Number
2	EQUIPMENT_TYPE	Carrier used for the shipment
3	CUSTOMER_MILES	Distance in Miles
4	WEIGHT	Weight of the shipment
5	**ORDER_COST**	**Cost of Shipment**
6	FIRST_PICK_ZIP	Source City Zip Code / Pin Code
7	FIRST_PICK_EARLY_APPT	First Appointment Date to collect the shipment from Customer
8	FIRST_PICK_LATE_APPT	Late Appointment Date to collect the shipment from Customer
9	LAST_DELIVERY_ZIP	Destination City Zip Code / Pin Code
10	LAST_DELIVERY_EARLY_APPT	Shipment Delivery Date - Early
11	LAST_DELIVERY_LATE_APPT	Shipment Delivery Date - Late
12	IS_HAZARDOUS	Is the shipment hazardous or not
13	CREATED_DATE	Order Created Date

Feature Engineering

Study the problem carefully and identify the scope of creating new variables, feature engineering,

Data Visualizations and Hypothesis Testing

Choose suitable visualization to answer the following questions. Write at least 10 words about your observations or insights from the graphs. Use suitable tests to hypothesis

1. How is the shipment cost distributed?
2. Does the shipment cost depend on the distance?
3. Is the weight of the shipment an important predictor to decide the shipment cost?
4. Is the Shipment cost varying depending on the source and destinations?
5. How do you justify whether the cost varies depending on the whether the item is hazardous or not
    a. Write the hypothesis and use appropriate statistical test
6. Identify the most important variables from both linear and non-linear models

Model Building

1. What is the performance metric you choose and explain the reason for the choice
2. Build 5 different machine learning models and experiment with at least 3 parameters for hyper-parameter tuning
3. Feature engineering
    a. Carefully look at the features and based on your domain knowledge derive couple of new features from the data
    b. Identify couple of new features that could have helped you to improve the model performance
4. Does transforming the target variable improve the model performance?
    a. Take at least 3 different approaches to improve model performance
5. Consolidate the performance of all the models in a data frame and write your comments

Project Deliverables and Deadlines

Project Deliverables and Deadlines

Capstone Schedule with Dates (12th April 2023 – 2nd May 2023)		
Date	**Event/Deliverable**	**Type of file to submit through SLMS and timings**
21st April 2023	Visualizations and Hypothesis Testing results	Zip file with code in ipynb and HTML formats only
30th April 2023	Model Building results	Zip file with code in ipynb and HTML formats only
1st May 2023	PPT	PPT in PPTX format as per the guidelines
2nd May 2023	In-person 30 minutes viva	

Submission

1. Submission uploaded in SLMS only will be considered for the evaluation.
2. Please submit a Zip folder which includes all outputs in jupyter notebook (.ipynb) and HTML format only.
3. Zip folder name should follow the format as mentioned below

    < Banik_Project_01.zip >

Viva Metrics

Code Quality	Code Execution
	Code Structure and Completeness
	Commenting in Code
Presentation Skills	Quality of Presentation Material
	Overall, Clarity of Thought and Flow of Presentation
Coding Skills	Python
Modelling Skills	Data Handling
	Models
	Model Validation
Conceptual Skills	Statistics
	ML
	AI
	Data Visualizations
	Big Data and Data Engineering