GeakMinds
geeks in action

System Test - Orders & Delivery Time

This dataset is about product delivery timelines of Orders (something like e-commerce delivery). The dataset comprises of two csv files, one showing all details of order delivery and other showing the time it took to deliver the order.

The problem statement is: When a customer places an order, how long does it take to get delivered in days?

1. Write an **SQL Query** to join both datasets (assuming each csv file is an SQL Table) and obtain order delivery details and timelines in a single table.
2. Perform **Exploratory Analysis** on the data and identify at least 3 aspects which impact delivery time and how much does it impact.
3. Analyse the dataset and build a **Machine Learning model** to predict the "Delivery Time" for orders.
4. Present the results of EDA and Modelling in a Power Point Presentation/Google slides. (Slide Preparation can be done offline)

Data Dictionary

order-delivery.csv

- Order Number – A unique identifier for each order
- Product Name – Name of the Product
- Order Type – New or Additional, indicating if it was first time or not.
- Product Cost – Cost of the Product in $
- Cash on Delivery – Is it prepaid or not.
- Product Unavailable Flag – Is the product available in stock or needs to procure from manufacturer to deliver to customer.
- Multi-Mode Transport Flag – Does it requires more than one mode of transport in delivering (flight/rail/road/ship)
- Speed Delivery Flag – Has the customer requested for faster delivery.
- Remote Location – Is the customer location outside major cities.
- Multi Hop Delivery – Does this delivery involves multiple transit hubs.
- Product Size – Size of the product

order-delivery-time.csv

- Order Number – A unique identifier for each order
- Delivery Time – No of days taken to deliver the product to customer.