Business Case: Delhivery - Feature Engineering

Problem definition.

The company wants to understand and process the data coming out of data engineering pipelines:

• Clean, sanitize and manipulate data to get useful features out of raw fields

• Make sense out of the raw data and help the data science team to build forecasting models on it

Colab Link - Answer

Insights and recommendations : -

1. Distributions of all numerical variable columns exhibit right skewness. Non-linear transformations were applied during hypothesis testing to ensure a normal distribution in the data.
2. The busiest corridor is from Chandigarh_Mehmdpur_H (Punjab) to Gurgaon_Bilaspur_HB (Haryana). Exploring optimization strategies can significantly increase system efficiency.
3. The data primarily contains data from the month of September, potentially missing seasonal trends for future analysis.
4. The highest number of orders originates from Maharashtra, signifying a significant hub for business operations.
5. The notable difference observed between the numerical values of OSRM and actual metrics suggests the existence of optimization opportunities, leaving room for potential improvements in operations.
6. The absence of a discernible pattern in daily order counts, consistently around 300 throughout the September and October dataset, indicates that relying on this behavior may not be a feasible approach for future projections.
7. The removal of outliers using IQR method significantly reduced the number of columns in the dataset, potentially impacting the richness of available information and insights.