Data exploration & preparation — B. Lagya Srivalli 2300032262

HA-1

Q1) Outlier detection & feature scaling

A) Highly skewed dataset can negatively impact statistical measures & predictive models. Outlier detection methods such as IQR methods & Z-score analysis are used to identify extreme values that distort data distribution. Once detected, outliers can be removed or capped to reduce their influence.

These preprocessing steps improve data integrity & reduce model bias.

Q2 Traffic congestion forecasting using data preparation

A) Raw traffic sensor data often contains noise, missing values & inconsistencies. The first step in data preparation involves cleaning the data by removing errors & handling missing values. The data is then aggregated based on time intervals & locations. Feature engineering is applied to derive meaningful variables such as traffic volume, average speed & peak hour indicators.

EDA is performed using trend analysis & correlation studies to identify congestion patterns & traffic behaviour. These insights help forecast congestion.