Facilitate surge price using predictive analysis

Student ID: 33059047

Project Brief

- The aim of the project here is to analyse the uber data to make a predictive analysis of when and where to increase the uber ride prices i.e. a predictive analysis of the surge pricing for uber.
- As a result of this research, we can identify the busiest times when more customers are available to implement surge pricing. – Time analysis
- Analysis would be used to identify which regions would be the busiest in order to activate surge pricing there. This would increase the number of drivers on the road and enable surge pricing. - Location Analysis

Uber Advantages using this model

- Add more drivers to the road and implement surge pricing where it is suitable.
- Implement surge pricing on particular days and times by identifying the peak hours and days and proactively tell drivers to move within this time and day and position the supply chain accordingly.
- Achieve their running target and thereby meeting revenue generation.
- Predicting which weather conditions or economic conditions are suitable for which surge multiplier.

Value adds and surge price

- Real-time traffic monitoring by these algorithms allows prices to be changed as demand for rides changes and increases the likelihood that journeys will take longer. This encourages more drivers to drive when required and to stay at home when there is little demand for their services and thereby increasing revenue generation.
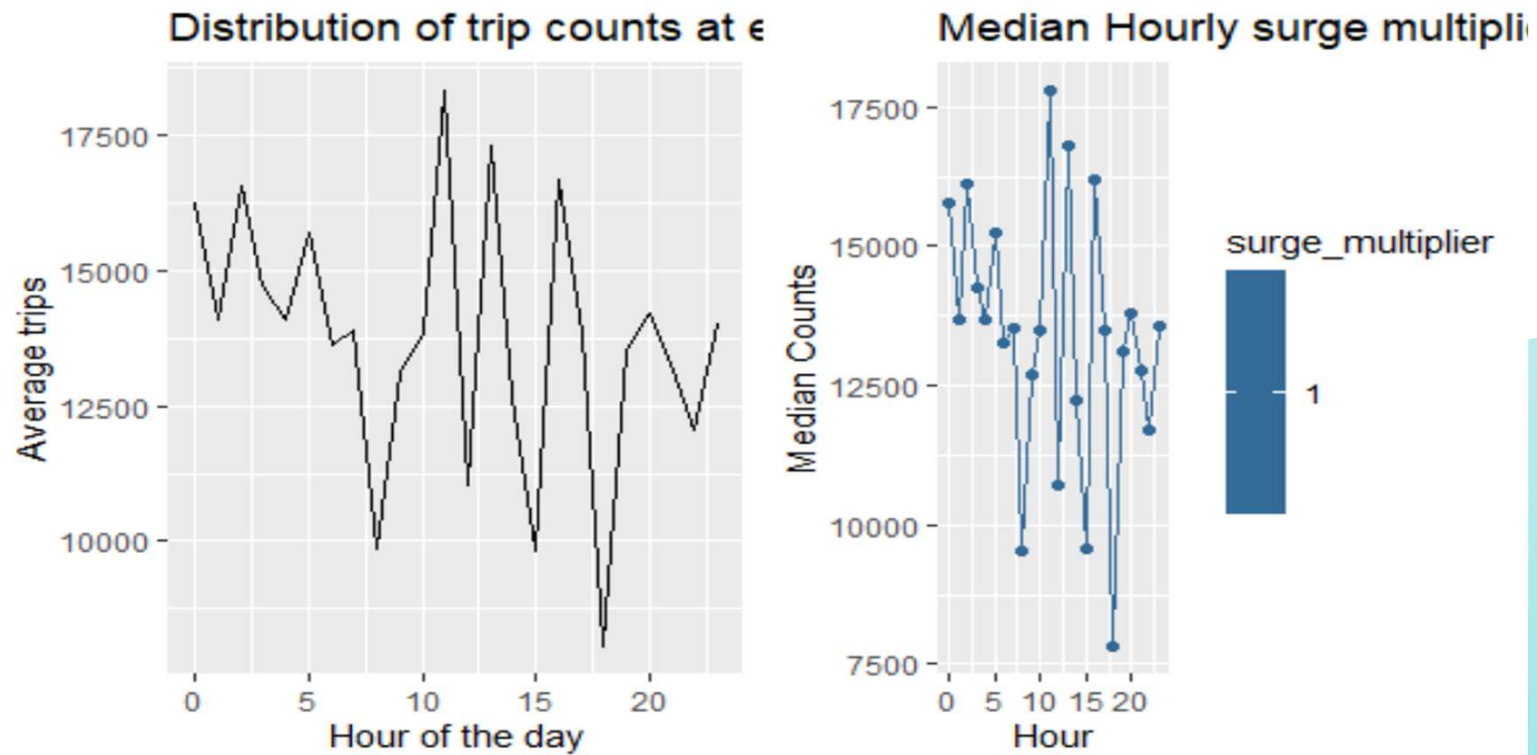
Data preprocessing

- Large amounts of structured data can be kept in a reliable enterprise data warehouse which can also act as the organization's central point of access or Hive.

- Real-time data analysis and presentation are required to support surge pricing and location hotspot forecasting. To achieve this we can use real – time data streaming tools

- stream processing of big data at uber

- Patterns can be found, outcomes can be examined, different levels of emphasis can be examined, and looking at data from several streams at once is also simple. Detecting patterns across time and time series data easily fit with stream processing. Stream – processing also enables delivery of analytical results fast.

- Apache Spark, an open-source stream processing platform that offers in-memory data processing. HDFS, as well as OpenStack Swift and Apache Cassandra, are compatible with Spark. You may process structured, unstructured, and semi-structured data using distributed computing with Spark.
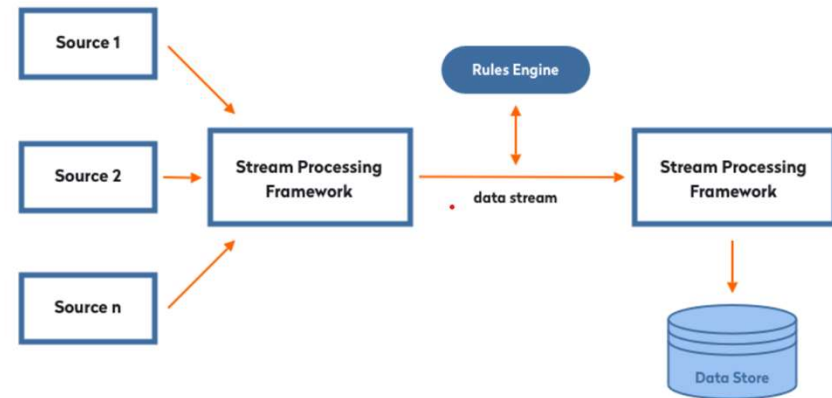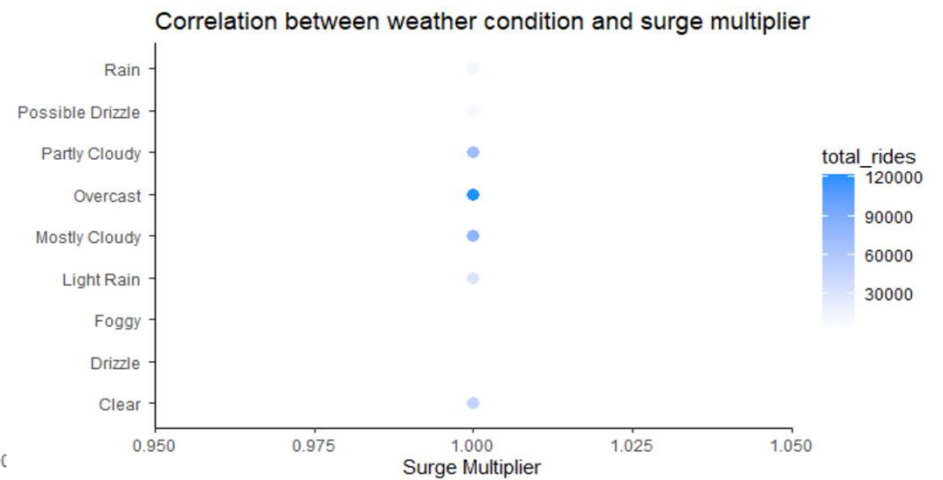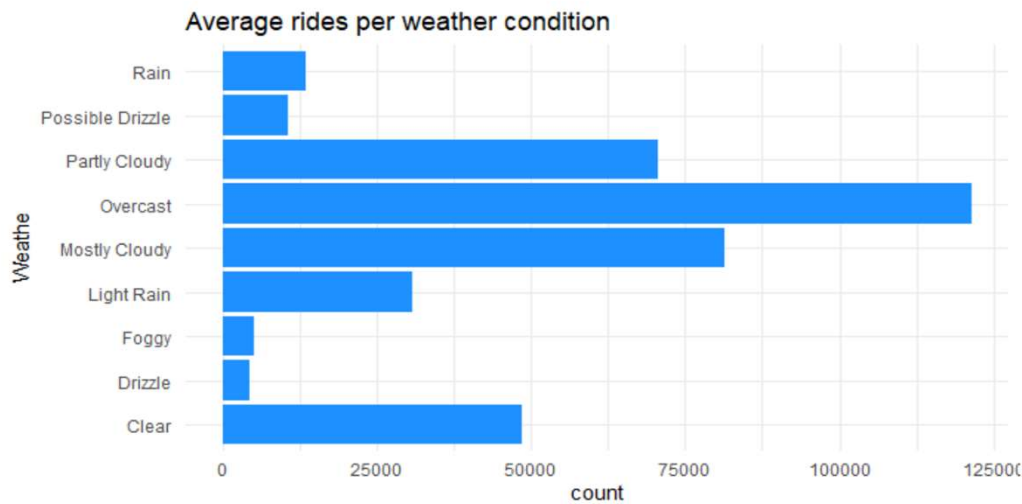
EDA: Correlation of busiest days & surge multiplier
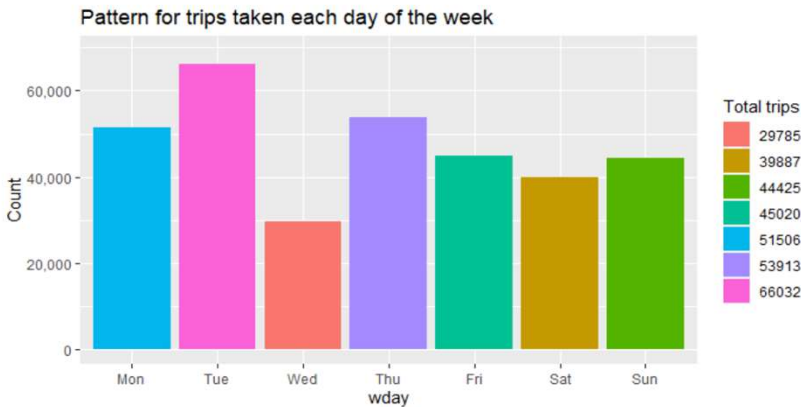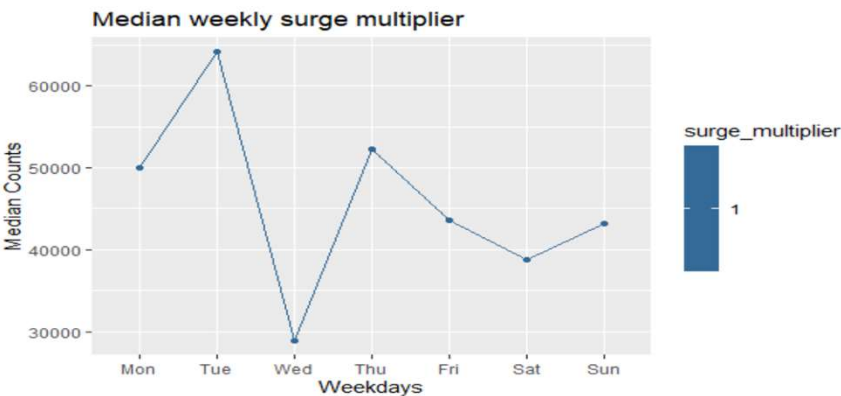
EDA: Correlation of busiest hour vs surge multiplier

EDA: Correlation of weather vs surge multiplier

Data Governance & Management framework 1/2

- **Data privacy & security**: Authentication: The process of confirming a person's identity before allowing them access to data using techniques including passwords, PINs, security tokens, and biometrics.

- **Archiving: Backups & disaster recovery**: like cloud archiving services as historical data at uber can be used for predictions.

- **Data Quality checks**: The data needs to be checked before it is committed to the database to retain a clear, standardised, and deduplicated picture of it always this can be implemented by incorporating data quality checks or deploying a centra data quality firewall. This is an important step as the results of the model depends directly on the data quality.

- **Metadata Management**: The quality of your data is directly impacted by your metadata and hence it is very necessary for managing metadata .

- **Master & Reference Data management**: Uber can improve its data management procedures and cut down on the time and effort needed to maintain and update data by centralising and standardising master and reference data.

Data Governance & Management framework 2/2

- **Information lifecycle governance:** One of the most important benefits is by doing this uber may lessen the amount of data it maintains and the accompanying expenses by putting in place procedures for data retention and disposal.