

# Data Intake Report

**Name:** Insight for Cab Investment firm

**Report date:** February 26, 2021

**Internship Batch:** LISP01

**Version:** 1.0

**Data intake by:** Swapnil Vishwakarma

**Data intake reviewer:**

**Data storage location:** Github

## Tabular data details:

<b>Total number of observations</b>	359392
<b>Total number of files</b>	4
<b>Total number of features</b>	22
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	60.3+ MB

## Proposed Approach:

- Merged Cab and City data on 'City' column.
- Merged Customer and Transaction data on 'Transaction\_ID'.
- Finally merged the above two data on 'Transaction\_ID'.
- Convert 'Date of Travel' column into pandas datetime column and set it as the index
- Created new columns to better analyze the trend.
- Find null values in the dataset and duplicates.
- Outliers are present in "Price Charged" feature but due to unavailability of trip duration details, we are not treating this as outlier.
- Profit of rides are calculated keeping other factors constant and only "Price Charged" and "Cost of Trip" features used to calculate profit.
- Users feature of city dataset is treated as number of cab users in the city.
- Perform univariate, bivariate and time series analysis.
- Hypothesis Testing.