Data Intake Report

Name: <G2M insight for Cab Investment firm(Must for all Specialization)>
Report date: <23 Sep 2022>
Internship Batch:<LISUM13>
Version:<1.0>
Data intake by:<Tatiana Moteu Ngoli>
Data intake reviewer:</>
Data storage location:
<https://github.com/TatianaMoteuN/Data-Glacier/tree/master/week2/DataSets-main>

Tabular data details:

Total number of observations	<359,392>
Total number of files	<4>
Total number of features	<14>
Base format of the file	<.csv>
Size of the data	< 359,392 B >

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:
- Data Understanding
- Forecasting profit and number of rides for each Cab type
- Finding the most profitable Cab company base on features
- Recommendations for investment
- Assumptions:
  - Outliers are present in Price_Charged features but due to unavailability of trip duration details, we are not treating this as outlier.
  - There is an equal distribution on the Age for both Cab company and we will also not use this feature 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 features of city datasets as number of Cab users in the city. We have then assumed that this can be other Cab users as well (including Yellow and Pink Cab)