Data Intake Report

Name: G2M Strategy
Report date: 12-08-2023
Internship Batch: <u>LISUM23: 30</u>
Version:1.0
Data intake by: Sri Saikamal Priyank Samireddi
Data intake reviewer:
Data storage location:

Tabular data details:

Total number of observations	359392
Total number of files	4
Total number of features	14
Base format of the file	.csv
Size of the data	30.479 MB

• 15 Features (including 5 features engineered)
• Timeframe of the data: 2016-01-31 to 2018-12-31
• Total data points: 359392 rows

Assumptions:

•The Price_Charged feature contains potential outliers, but without trip duration details available, these higher-priced rides cannot definitively be classified as anomalous data points. For this analysis, no adjustments were made to exclude potential outlier fare amounts.

•To simplify profit calculations at the ride level, assumptions were made to keep other cost factors constant. Profit per ride is calculated directly from the Price_Charged and Cost_of_Trip features, not accounting for additional variable costs. This focuses the profitability analysis specifically on the revenue versus direct costs per ride.

In summary:

•Outliers exist in the fare prices but could not be reliably identified without trip duration data, so no outlier removal was done.

•Direct profit per ride was calculated using only the price charged and trip cost data, holding other factors constant. This simplifies ride profit analysis.