

# Data Intake Report

Name: G2M Strategy

Report date: 12-08-2023

Internship Batch: [LISUM23: 30](#)

Version: 1.0

Data intake by: Sri Saikamal Priyank Samireddi

Data intake reviewer:

Data storage location: <https://github.com/DataGlacier/DataSets>

## Tabular data details:

<b>Total number of observations</b>	359392
<b>Total number of files</b>	4
<b>Total number of features</b>	14
<b>Base format of the file</b>	.csv
<b>Size of the data</b>	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.