

Data Intake Report

Name: G2M Insight for Cab Investment Firm

Report date: 6/13/2023

Internship Batch: LISUM22

Version: 1.0

Data intake by: Ayah Ibrahim

Data intake reviewer: -

Data storage location: -

Tabular data details:

Cab_Data

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 MB

Using isnull(): Are there any null/missing values in this dataset? No

Using duplicated(): Are there any duplicates in this dataset? No

Using dtypes(): What are the datatypes in this dataset?

2 integer columns, 3 float columns, and 2 object columns:

Transaction ID: int64

Date of Travel: int64

Company: object

City: object

KM Travelled: float64

Price Charged: float64

Cost of Trip: float64

Using nunique(): What columns are unique? Returns the count of distinct values present in the column.

Transaction ID: 359,392 (All values are unique)

Date of Travel: 1,095 (There are 1,095 unique dates)

Company: 2 (There are 2 unique company values)

City: 19 (There are 19 unique cities)

KM Travelled: 874 (There are 874 unique values for kilometers travelled)
Price Charged: 99,176 (There are 99,176 unique values for price charged)
Cost of Trip: 16,291 (There are 16,291 unique values for the cost of the trip)

From the above list, I can conclude that the 'Transaction ID' column is the only column that has all unique values, indicating that each row has a distinct transaction ID.

Tabular data details:

City

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	4 KB

Using isnull(): Are there any null/missing values in this dataset? No

Using duplicated(): Are there any duplicates in this dataset? No

Using dtypes(): What are the datatypes in this dataset?

3 object columns

City: object

Population: object

Users: object

Using nunique(): What columns are unique? Returns the count of distinct values present in the column.

City: 20

Population: 20

Users: 20

Based on the above list, all the values in each of the 3 columns are unique.

Tabular data details:

Customer_ID

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1 MB

Using isnull(): Are there any null/missing values in this dataset? No

Using duplicated(): Are there any duplicates in this dataset? No

Using dtypes(): What are the datatypes in this dataset?

3 integer columns and 1 object columns:

Customer ID: int64

Gender: object

Age: int64

Income (USD/Month): int64

Using nunique(): What columns are unique? Returns the count of distinct values present in the column.

Customer ID: 49,171 (All values are unique)

Gender: 2 (There are 2 unique gender values)

Age: 48 (There are 48 unique age values)

Income (USD/Month): 23,341 (There are 23,341 unique income values)

Based on the list above, the only column with all unique values is Customer ID.

Tabular data details:

Transaction_ID

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	8.58 MB

Using isnull(): Are there any null/missing values in this dataset? No

Using duplicated(): Are there any duplicates in this dataset? No

Using dtypes(): What are the datatypes in this dataset?

2 integer columns and 1 object column:

Transaction ID: int64

Customer ID: int64

Payment_Mode: object

Using nunique(): What columns are unique? Returns the count of distinct values present in the column.

Transaction ID: 440,098 (All values are unique)

Customer ID: 49,171 (There are 49,171 unique customer IDs)

Payment_Mode: 2 (There are 2 unique payment modes)

Based on the list above, the only column with all unique values is Transaction ID.