

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 13.06.2022

Internship Batch: LISUM10

Version:<1.0>

Data intake by: Kseniia Nosenko

Data intake reviewer:< >

Data storage location: <<https://github.com/KseniyaLem/Internship/Week2>>

Tabular data details:

Name of the file	<code>Cab_Data.csv</code>
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	20663 KB

Name of the file	<code>City.csv</code>
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	1 KB

Name of the file	<code>Customer_ID.csv</code>
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	1027KB

Name of the file	<code>Transaction_ID.csv</code>
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	8788 KB

Name of the file	Holidays.csv
Total number of observations	342
Total number of files	1
Total number of features	6
Base format of the file	.csv
Size of the data	16 KB

- The information was examined, no duplicates were found (a search for unique values in columns was applied: column Transaction_ID in files Transaction_ID and Cab_Data, Customer_ID in file Customer_ID).
- Transaction data that is in the transaction file, but not in the general file, was removed from the analysis. From the file with holidays, data was taken only for the period 01.01.2016-31.12.2018.