

Data Intake Report

Name: <G2M insight for Cab Investment firm Customer id>

Report date: <14/07/2022>

Internship Batch:<LISUM11>

Version:<1.0>

Data intake by:<Md Khalid Siddiqui>

Data intake reviewer:<intern who reviewed the report>

Data storage location: < https://raw.githubusercontent.com/DataGlacier/DataSets/main/Customer_ID.csv>

Tabular data details:

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	<4.0 MB>

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

Proposed Approach:

- Two methods were used to explore initial characteristics (univariate analysis)

1. Pandas code : df_dataframe. Info

Sample screenshot

3.1 Basic Info

```
df_customer_id.info() #basic info

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49171 entries, 0 to 49170
Data columns (total 4 columns):
 #   Column              Non-Null Count  Dtype  
---  -
 0   Customer ID         49171 non-null  int64  
 1   Gender              49171 non-null  object  
 2   Age                 49171 non-null  int64  
 3   Income (USD/Month)  49171 non-null  int64  
dtypes: int64(3), object(1)
memory usage: 1.5+ MB
```

2. Pandas Profiling: widget command of pandas profiling module

Sample screenshot

▼ 3.2 Interactive Profile Report

23s

▶

```
profile3 = ProfileReport(df_customer_id, title = "Profile Report of dataset customer_id")
profile3.to_widgets() #interactive widget form
```

⏏

/usr/local/lib/python3.7/dist-packages/pandas_profiling/profile_report.py:361: UserWarning: Ipywidgets is not yet fully supported on Google Colab (https://github.com/googlecolab/colabtools/issues/68).
"Ipywidgets is not yet fully supported on Google Colab (https://github.com/googlecolab/colabtools/issues/68)."
Summarize dataset: ██████████ 18/? [00:04<00:00, 2.47it/s, Completed]
Generate report structure: 100% ██████████ 1/1 [00:02<00:00, 2.23s/it]

Overview

Variables

Interactions

Correlations

Missing values

Sample

Overview

Reproduction

Warnings (1)

Number of variables	4	NUM	3
Number of observations	49171	CAT	1
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	4.0 MiB		
Average record size in memory	85.9 B		

Report generated with [pandas-profiling](#).