

Data Intake Report

Name: <G2M insight for Cab Investment firm City>

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/City.csv>>

Tabular data details:

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	<4.0 KB>

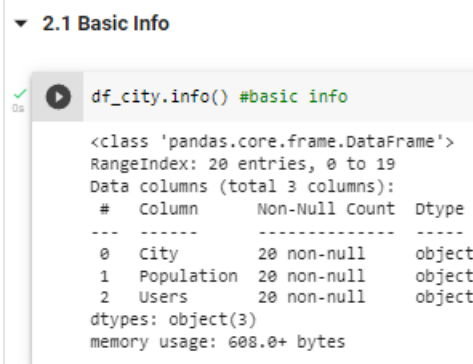
Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

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

1. Pandas code : df_city. Info

Sample screenshot



```
2.1 Basic Info

df_city.info() #basic info

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0    City        20 non-null    object
1    Population  20 non-null    object
2    Users       20 non-null    object
dtypes: object(3)
memory usage: 608.0+ bytes
```

2. Pandas Profiling: widget command of pandas profiling module

Sample screenshot

▼ 2.2 Interactive Profile Report

```
profile2 = ProfileReport(df_city, title = "Profile Report of dataset City")
profile2.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
    "Ipywidgets is not yet fully supported on Google Colab (https://github.com/googlecolab/colabtools/issues/60)."
```

Summarize dataset: 17/? [00:00<00:00, 15.69it/s, Completed]

Generate report structure: 100% 1/1 [00:01<00:00, 1.11s/it]

Overview	Variables	Interactions	Correlations	Missing values	Sample
Overview	Reproduction	Warnings (0)			
Number of variables	3		CAT		3
Number of observations	20				
Missing cells	0				
Missing cells (%)	0.0%				
Duplicate rows	0				
Duplicate rows (%)	0.0%				
Total size in memory	4.0 KiB				
Average record size in memory	205.2 B				

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