EDA Implementation with the help of Automted EDA Tools:- 28-09-2022

In [1]:
```python
# Importing csv Dataset with a help of Pandas Library:-
import pandas as pd


df=pd.read_csv(r'C:\Users\Mukul\Data Kaggle\Downloads\Datasets\Dataset\Dataset\da
```

In [2]:
```python
# df used to provide all data from datasets
df
```

Out[2]:

	User_ID	Product_ID	Gender	Age	Occupation	City_Category	Stay_In_Current_City_Years
0	1000001	P00069042	F	0-17	10	A	2
1	1000001	P00248942	F	0-17	10	A	2
2	1000001	P00087842	F	0-17	10	A	2
3	1000001	P00085442	F	0-17	10	A	2
4	1000002	P00285442	M	55+	16	C	4+
...	...	...	...	...	...	...	...
550063	1006033	P00372445	M	51-55	13	B	1
550064	1006035	P00375436	F	26-35	1	C	3
550065	1006036	P00375436	F	26-35	15	B	4+
550066	1006038	P00375436	F	55+	1	C	2
550067	1006039	P00371644	F	46-50	0	B	4+

550068 rows × 12 columns

In [3]: ```
# head() used to privide top 5 data from dataset
df.head()
```

Out[3]:

| | User_ID | Product_ID | Gender | Age | Occupation | City_Category | Stay_In_Current_City_Years | Mar |
|---|---|---|---|---|---|---|---|---|
| 0 | 1000001 | P00069042 | F | 0-17 | 10 | A | 2 | |
| 1 | 1000001 | P00248942 | F | 0-17 | 10 | A | 2 | |
| 2 | 1000001 | P00087842 | F | 0-17 | 10 | A | 2 | |
| 3 | 1000001 | P00085442 | F | 0-17 | 10 | A | 2 | |
| 4 | 1000002 | P00285442 | M | 55+ | 16 | C | 4+ | |

In [4]: ```
# tail() used to provide last  5 data from dataset
df.tail()
```

Out[4]:

	User_ID	Product_ID	Gender	Age	Occupation	City_Category	Stay_In_Current_City_Years
550063	1006033	P00372445	M	51-55	13	B	1
550064	1006035	P00375436	F	26-35	1	C	3
550065	1006036	P00375436	F	26-35	15	B	4+
550066	1006038	P00375436	F	55+	1	C	2
550067	1006039	P00371644	F	46-50	0	B	4+

Performing EDA with the help of Pandas - Profiling Automated EDA Tools:-

In [5]:
```python
# Performing EDA with the help of Pandas -Profiling Automated EDA Tools

import pandas as pd
from pandas_profiling import ProfileReport
profile = ProfileReport(df,title ='Report')
profile
```

Summarize dataset: 0%| | 0/5 [00:00<?, ?it/s]

Generate report structure: 0%| | 0/1 [00:00<?, ?it/s]

Render HTML: 0%| | 0/1 [00:00<?, ?it/s]

Overview

Dataset statistics

Number of variables	12
Number of observations	550068
Missing cells	556885
Missing cells (%)	8.4%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	50.4 MiB
Average record size in memory	96.0 B

Variable types

Numeric	6
Categorical	6

Alerts

`Product_ID` has a high cardinality: 3631 distinct values	High cardinality
`Product_Category_2` is highly correlated with `Product_Category_1` and 2 other fields (Product_Category_1, Product_Category_3, Purchase)	High correlation

Out[5]:

Performing EDA with the help of Sweetviz Automated EDA Tools:-

In [6]: *# Performing EDA with the help of Sweet viz Automated EDA Tools:*

In [7]:
```
pip install sweetviz
```

Requirement already satisfied: sweetviz in c:\users\mukul\anaconda3\lib\site-pa
ckages (2.1.4)Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3 in c:\use
rs\mukul\anaconda3\lib\site-packages (from sweetviz) (1.4.2)
Requirement already satisfied: scipy>=1.3.2 in c:\users\mukul\anaconda3\lib\sit
e-packages (from sweetviz) (1.7.3)
Requirement already satisfied: importlib-resources>=1.2.0 in c:\users\mukul\ana
conda3\lib\site-packages (from sweetviz) (5.9.0)
Requirement already satisfied: numpy>=1.16.0 in c:\users\mukul\anaconda3\lib\si
te-packages (from sweetviz) (1.21.5)
Requirement already satisfied: tqdm>=4.43.0 in c:\users\mukul\anaconda3\lib\sit
e-packages (from sweetviz) (4.64.0)
Requirement already satisfied: jinja2>=2.11.1 in c:\users\mukul\anaconda3\lib\s
ite-packages (from sweetviz) (2.11.3)
Requirement already satisfied: matplotlib>=3.1.3 in c:\users\mukul\anaconda3\li
b\site-packages (from sweetviz) (3.5.1)
Requirement already satisfied: zipp>=3.1.0 in c:\users\mukul\anaconda3\lib\site
-packages (from importlib-resources>=1.2.0->sweetviz) (3.7.0)

Requirement already satisfied: MarkupSafe>=0.23 in c:\users\mukul\anaconda3\lib
\site-packages (from jinja2>=2.11.1->sweetviz) (2.0.1)
Requirement already satisfied: packaging>=20.0 in c:\users\mukul\anaconda3\lib
\site-packages (from matplotlib>=3.1.3->sweetviz) (21.3)
Requirement already satisfied: pillow>=6.2.0 in c:\users\mukul\anaconda3\lib\si
te-packages (from matplotlib>=3.1.3->sweetviz) (9.0.1)
Requirement already satisfied: cycler>=0.10 in c:\users\mukul\anaconda3\lib\sit
e-packages (from matplotlib>=3.1.3->sweetviz) (0.11.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\mukul\anaconda3
\lib\site-packages (from matplotlib>=3.1.3->sweetviz) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\mukul\anaconda3\li
b\site-packages (from matplotlib>=3.1.3->sweetviz) (1.3.2)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\mukul\anaconda3\li
b\site-packages (from matplotlib>=3.1.3->sweetviz) (4.25.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\mukul\anaconda3\lib
\site-packages (from matplotlib>=3.1.3->sweetviz) (3.0.4)
Requirement already satisfied: pytz>=2020.1 in c:\users\mukul\anaconda3\lib\sit
e-packages (from pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3->sweetviz) (2021.3)
Requirement already satisfied: six>=1.5 in c:\users\mukul\anaconda3\lib\site-pa
ckages (from python-dateutil>=2.7->matplotlib>=3.1.3->sweetviz) (1.16.0)
Requirement already satisfied: colorama in c:\users\mukul\anaconda3\lib\site-pa
ckages (from tqdm>=4.43.0->sweetviz) (0.4.4)

In [8]:
```python
import sweetviz as sv
analyze_report = sv.analyze(df)
analyze_report.show_notebook()
```

| [0%]   00:00 ->…

Sweetviz
2.1.4
Get updates, docs & report issues here
Created & maintained by Francois Bertrand
Graphic design by Jean-Francois Hains

DataFr

ASSOCIATIONS

DataFram

1 User_ID

VALUES:	550,068 (100%)	MAX	1,006,040	RANGE
MISSING:	---	95%	1,005,747	IQR
		Q3	1,004,478	STD
DISTINCT:	5,891 (1%)	MEDIAN	1,003,077	VAR
		AVG	1,003,029	
ZEROES:	---	Q1	1,001,516	KURT.
		5%	1,000,329	SKEW
		MIN	1,000,001	SUM

2 Product_ID

VALUES:	550,068 (100%)	1,880	<1%	P00265242
MISSING:	---	1,615	<1%	P00025442
		1,612	<1%	P00110742
DISTINCT:	3,631 (<1%)	1,562	<1%	P00112142
		1,470	<1%	P00057642
		1,440	<1%	P00184942
		1,438	<1%	P00046742
		539,051	98%	(Other)

3 Gender

VALUES:	550,068 (100%)
MISSING:	---
DISTINCT:	2 (<1%)

4 Age

VALUES:	550,068 (100%)

Performing EDA with help of Autoviz Automated EDA Tools:-

In [9]:
```python
import matplotlib.pyplot as plt
```

In [10]:
```python
pip install autoviz
```

Requirement already satisfied: autoviz in c:\users\mukul\anaconda3\lib\site-p
ackages (0.1.55)Note: you may need to restart the kernel to use updated packa
ges.
Requirement already satisfied: panel~=0.12.6 in c:\users\mukul\anaconda3\lib
\site-packages (from autoviz) (0.12.7)
Requirement already satisfied: fsspec>=0.8.3 in c:\users\mukul\anaconda3\lib
\site-packages (from autoviz) (2022.2.0)
Requirement already satisfied: xlrd in c:\users\mukul\anaconda3\lib\site-pack
ages (from autoviz) (2.0.1)
Requirement already satisfied: pandas in c:\users\mukul\anaconda3\lib\site-pa
ckages (from autoviz) (1.4.2)
Requirement already satisfied: wordcloud in c:\users\mukul\anaconda3\lib\site
-packages (from autoviz) (1.8.2.2)
Requirement already satisfied: emoji in c:\users\mukul\anaconda3\lib\site-pac
kages (from autoviz) (2.1.0)
Requirement already satisfied: matplotlib>=3.3.3 in c:\users\mukul\anaconda3
\lib\site-packages (from autoviz) (3.5.1)
Requirement already satisfied: hvplot>=0.7.3 in c:\users\mukul\anaconda3\lib
\site-packages (from autoviz) (0.7.3)

In [11]:
```python
from autoviz.AutoViz_Class import AutoViz_Class
AV = AutoViz_Class()
df_av = AV.AutoViz(r"C:\Users\Mukul\Data Kaggle\Downloads\Datasets\Dataset\Datase
```

Imported v0.1.55. After importing, execute '%matplotlib inline' to display ch
arts in Jupyter.
    AV = AutoViz_Class()
    dfte = AV.AutoViz(filename, sep=',', depVar='', dfte=None, header=0, verb
ose=1, lowess=False,
                chart_format='svg',max_rows_analyzed=150000,max_cols_analyzed=
30, save_plot_dir=None)
Update: verbose=0 displays charts in your local Jupyter notebook.
        verbose=1 additionally provides EDA data cleaning suggestions. It als
o displays charts.
        verbose=2 does not display charts but saves them in AutoViz_Plots fol
der in local machine.
        chart_format='bokeh' displays charts in your local Jupyter notebook.
        chart_format='server' displays charts in your browser: one tab for ea
ch chart type
        chart_format='html' silently saves interactive HTML files in your loc
al machine

In [ ]: