

```
EDA-using-libraries (/github/AnkitDubey040/EDA-using-libraries/tree/main)
/
EDA using inbuilt libraries.ipynb (/github/AnkitDubey040/EDA-using-libraries/tree/main/EDA using inbuilt libraries.ipynb)
```

..... EDA using Inbuilt pandas libraries

Using pandas-profiling

```
In [1]: # Firstly we import the data on which we are going to perform the EDA on using pandas as pd
```

```
In [2]: df = pd.read_csv("E:\\DATA_SCIENCE\\Dataset\\data7\\train.csv") #enter your own
```

```
In [3]: df.head()
```

```
Out[3]:
```

	Item_Identifier	Item_Weight	Item_Visibility	Item_MRP	Outlet_Identifier	Item_Outlet_Sales	Outlet_Type
0	FDA15	9.30	0.016047	249.8092	OUT049	3735.1380	1
1	DRC01	5.92	0.019278	48.2692	OUT018	443.4228	2
2	FDN15	17.50	0.016760	141.6180	OUT049	2097.2700	3
3	FDX07	19.20	0.017834	182.0950	OUT010	732.3800	4
4	NCD19	8.93	0.009780	53.8614	OUT013	994.7052	

5 rows × 33 columns

We will first install the pandas-profiling library

In [6]:

```
pip install pandas-profiling
```

```
Collecting pandas-profiling
  Downloading pandas_profiling-3.3.0-py2.py3-none-any.whl (268 kB)
Collecting phik<0.13,>=0.11.1
  Downloading phik-0.12.2-cp39-cp39-win_amd64.whl (685 kB)
Requirement already satisfied: statsmodels<0.14,>=0.13.2 in c:\users\hp\anaconda3\lib\site-packages\statsmodels\__init__.py
Requirement already satisfied: scipy<1.10,>=1.4.1 in c:\users\hp\anaconda3\lib\site-packages\scipy\__init__.py
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in c:\users\hp\anaconda3\lib\site-packages\PyYAML\__init__.py
Requirement already satisfied: requests<2.29,>=2.24.0 in c:\users\hp\anaconda3\lib\site-packages\requests\__init__.py
Requirement already satisfied: matplotlib<3.6,>=3.2 in c:\users\hp\anaconda3\lib\site-packages\matplotlib\__init__.py
Requirement already satisfied: pandas!=1.4.0,<1.5,>1.1 in c:\users\hp\anaconda3\lib\site-packages\pandas\__init__.py
Collecting missingno<0.6,>=0.4.2
  Downloading missingno-0.5.1-py3-none-any.whl (8.7 kB)
Requirement already satisfied: tqdm<4.65,>=4.48.2 in c:\users\hp\anaconda3\lib\site-packages\tqdm\__init__.py
Requirement already satisfied: seaborn<0.12,>=0.10.1 in c:\users\hp\anaconda3\lib\site-packages\seaborn\__init__.py
Collecting multimethod<1.9,>=1.4
  Downloading multimethod-1.8-py3-none-any.whl (9.8 kB)
Requirement already satisfied: numpy<1.24,>=1.16.0 in c:\users\hp\anaconda3\lib\site-packages\numpy\__init__.py
Requirement already satisfied: joblib~=1.1.0 in c:\users\hp\anaconda3\lib\site-packages\joblib\__init__.py
Requirement already satisfied: jinja2<3.2,>=2.11.1 in c:\users\hp\anaconda3\lib\site-packages\jinja2\__init__.py
Collecting htmlmin==0.1.12
  Downloading htmlmin-0.1.12.tar.gz (19 kB)
Collecting visions[type_image_path]==0.7.5
  Downloading visions-0.7.5-py3-none-any.whl (102 kB)
Collecting tangled-up-in-unicode==0.2.0
  Downloading tangled_up_in_unicode-0.2.0-py3-none-any.whl (4.7 MB)
Collecting pydantic<1.10,>=1.8.1
  Downloading pydantic-1.9.2-cp39-cp39-win_amd64.whl (2.0 MB)
Requirement already satisfied: networkx>=2.4 in c:\users\hp\anaconda3\lib\site-packages\networkx\__init__.py
Requirement already satisfied: attrs>=19.3.0 in c:\users\hp\anaconda3\lib\site-packages\attrs\__init__.py
Requirement already satisfied: Pillow in c:\users\hp\anaconda3\lib\site-packages\Pillow\__init__.py
Collecting imagehash
  Downloading ImageHash-4.3.1-py2.py3-none-any.whl (296 kB)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\hp\anaconda3\lib\site-packages\MarkupSafe\__init__.py
Requirement already satisfied: packaging>=20.0 in c:\users\hp\anaconda3\lib\site-packages\packaging\__init__.py
Requirement already satisfied: cycler>=0.10 in c:\users\hp\anaconda3\lib\site-packages\cycler\__init__.py
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hp\anaconda3\lib\site-packages\fonttools\__init__.py
Requirement already satisfied: python-dateutil>=2.7 in c:\users\hp\anaconda3\lib\site-packages\python_dateutil\__init__.py
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\hp\anaconda3\lib\site-packages\pyparsing\__init__.py
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp\anaconda3\lib\site-packages\kiwisolver\__init__.py
Requirement already satisfied: pytz>=2020.1 in c:\users\hp\anaconda3\lib\site-packages\pytz\__init__.py
Requirement already satisfied: typing-extensions>=3.7.4.3 in c:\users\hp\anaconda3\lib\site-packages\typing_extensions\__init__.py
Requirement already satisfied: six>=1.5 in c:\users\hp\anaconda3\lib\site-packages\six\__init__.py
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\hp\anaconda3\lib\site-packages\charset_normalizer\__init__.py
Requirement already satisfied: certifi>=2017.4.17 in c:\users\hp\anaconda3\lib\site-packages\certifi\__init__.py
Requirement already satisfied: idna<4,>=2.5 in c:\users\hp\anaconda3\lib\site-packages\idna\__init__.py
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\hp\anaconda3\lib\site-packages\urllib3\__init__.py
Requirement already satisfied: patsy>=0.5.2 in c:\users\hp\anaconda3\lib\site-packages\patsy\__init__.py
Requirement already satisfied: colorama in c:\users\hp\anaconda3\lib\site-packages\colorama\__init__.py
Requirement already satisfied: PyWavelets in c:\users\hp\anaconda3\lib\site-packages\PyWavelets\__init__.py
Building wheels for collected packages: htmlmin
  Building wheel for htmlmin (setup.py): started
  Building wheel for htmlmin (setup.py): finished with status 'done'
  Created wheel for htmlmin: filename=htmlmin-0.1.12-py3-none-any.whl size=27096
  Stored in directory: c:\users\hp\appdata\local\pip\cache\wheels\1d\05\04\c6d7c
Successfully built htmlmin
```

```
Installing collected packages: tangled-up-in-unicode, multimethod, visions, imagehash, htmlmin, missingno, multimethod, vislens
Successfully installed htmlmin-0.1.12 imagehash-4.3.1 missingno-0.5.1 multimethod-0.0.8 vislens-0.1.0
```

In [4]:

```
# importing other important Libraries as well
import numpy as np
import pandas as pd
from pandas_profiling import ProfileReport
```

ProfileReport tool of the pandas_profiling library is used to create an eda report of the dataframe

After this we create a data_profile variable that contains the report and we give title to it.

In [5]:

```
data_profile = ProfileReport(df,title = "Pandas profiling report train dataset")
```

In [6]:

```
# Now we can view the created report:  
data_profile
```

Overview

Dataset statistics

Number of variables	33
Number of observations	8523
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	2.1 MiB
Average record size in memory	264.0 B

Variable types

Categorical	28
Numeric	5

Alerts

Item_Identifier has a high cardinality: 1559 distinct values	High cardinality
Item_MRP is highly correlated with Item_Outlet_Sales	High correlation
Item_Outlet_Sales is highly correlated with Item_MRP and 2 other fields (Item_MRP, Outlet_Identifier, Outlet_Type_0)	High correlation

Out[6]:

As we can see this library automatically detects all the variables their types and do a full analysis of each feature to give a detailed report about each variable.

1. In the top we can see the title of the report **Pandas profiling report train dataset** , along with that we see many tabs i.e. Overview , variables , Intercations , corerlation , missing values and sample which can be used to directly jump to different sections of the report.
2. First we see an **Overview** section which gives us summary of the whole data , and gives us basic idea about the dataframe like no of variables , type of variables , missing cells , duplicate cells etc. Along with this it also gives us an **Alerts** and **Reproduction** tabs.
3. Then we have **Vatriables** section hwic is one the most important as it gives a detailed and seperate analysis of each feature or each columnm of the table or dataframe individually as well as with respect to other columns/variables.
Also for each variable we have a **Google Details** button in the lower right side where we get to see a whole statistical analysis of the data, mean , median , mode, sd , corr , max, min , and also we can view histograms , common values, Q1, Q3, IQR , 95th percetile etc.
4. The third section is **Interactions** section which as the name suggests gives a visualazion of intercations between all the variables with each other.
5. The fourth section is **Correlation** which give us corr. between variables , ypu can view spearmen's rank corr, pearson's corr, Cramer's corr. and kendell's corr.
6. The fifth section is **Missing Values** which shows us missing valus(if any) , its's effect on the data , and gives us idea as to how we can handle it efficiently.
7. The last section is **Sample** where we can view the sampople data of first few rows and last few rows.

Using SweetVIZ

Sweetviz is an open-source Python library that generates beautiful, high-density visualizations to kickstart EDA (Exploratory Data Analysis) with just two lines of code. Output is a fully self-contained HTML application. The system is built around quickly visualizing target values and comparing datasets.

First we intall the sweetviz library into our notebook

In [11]:

```
pip install sweetviz
```

```
Note: you may need to restart the kernel to use updated packages. Collecting sweetviz
  Downloading sweetviz-2.1.4-py3-none-any.whl (15.1 MB)
Requirement already satisfied: scipy>=1.3.2 in c:\users\hp\anaconda3\lib\site-packages\scipy\__init__.py
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3 in c:\users\hp\anaconda3\lib\site-packages\pandas\__init__.py
Collecting importlib-resources>=1.2.0
  Downloading importlib_resources-5.10.0-py3-none-any.whl (34 kB)
Requirement already satisfied: matplotlib>=3.1.3 in c:\users\hp\anaconda3\lib\site-packages\matplotlib\__init__.py
Requirement already satisfied: numpy>=1.16.0 in c:\users\hp\anaconda3\lib\site-packages\numpy\__init__.py
Requirement already satisfied: jinja2>=2.11.1 in c:\users\hp\anaconda3\lib\site-packages\jinja2\__init__.py
Requirement already satisfied: tqdm>=4.43.0 in c:\users\hp\anaconda3\lib\site-packages\tqdm\__init__.py
Requirement already satisfied: zipp>=3.1.0 in c:\users\hp\anaconda3\lib\site-packages\zipp\__init__.py
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\hp\anaconda3\lib\site-packages\MarkupSafe\__init__.py
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp\anaconda3\lib\site-packages\kiwisolver\__init__.py
Requirement already satisfied: cycler>=0.10 in c:\users\hp\anaconda3\lib\site-packages\cyclер\__init__.py
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hp\anaconda3\lib\site-packages\fontTools\__init__.py
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\hp\anaconda3\lib\site-packages\pyparsing\__init__.py
Requirement already satisfied: packaging>=20.0 in c:\users\hp\anaconda3\lib\site-packages\packaging\__init__.py
Requirement already satisfied: pillow>=6.2.0 in c:\users\hp\anaconda3\lib\site-packages\PIL\__init__.py
Requirement already satisfied: python-dateutil>=2.7 in c:\users\hp\anaconda3\lib\site-packages\python_dateutil\__init__.py
Requirement already satisfied: pytz>=2020.1 in c:\users\hp\anaconda3\lib\site-packages\pytz\__init__.py
Requirement already satisfied: six>=1.5 in c:\users\hp\anaconda3\lib\site-packages\six\__init__.py
Requirement already satisfied: colorama in c:\users\hp\anaconda3\lib\site-packages\colorama\__init__.py
Installing collected packages: importlib-resources, sweetviz
Successfully installed importlib-resources-5.10.0 sweetviz-2.1.4
```

After installation we now import the library and create a report with name **data_profile2** on the dataframe **df**.

sv.analyze(dataframe_name) automatically does an analysis of the data and creates an html report the same.

In [7]:

```
import sweetviz as sv
data_profile2 = sv.analyze(df)
```

report_name.show_notebook() is used to view the report in notebook itself, if we wanted to view it in html in a web browser we can do that using **report_name.show_html()**.

In [8]:

```
data_profile2.show_notebook()
```



We can see that it creates a report which is similar in pattern to that of pandas profiling , the main difference is effects and visualization which is different . Here also we get an **Overview** on the top of the report which gives us a summary of the data , also there is an **Association** tab where we can see association between all the variables.

Then there is a list of all the features and on clicking on any feature , the section expands to give us all the information about the data , The visualization is **better** than pandas profiling as we get to see all the details very clearly and with nicely formed histograms and charts .

Using D-Tale

D-Tale is the combination of a Flask back-end and a React front-end to bring you an easy way to view & analyze Pandas data structures. It integrates seamlessly with ipython notebooks & python/ipython terminals. Currently this tool supports such Pandas objects as DataFrame, Series, MultiIndex, DatetimeIndex & RangeIndex.

First we install the **dtale** library

In [1]:

```
pip install dtale
```

```
Collecting dtale
  Note: you may need to restart the kernel to use updated packages
    Using cached dtale-2.8.1-py2.py3-none-any.whl (12.8 MB)
Requirement already satisfied: seaborn in c:\users\hp\anaconda3\lib\site-packages
Collecting squarify
  Using cached squarify-0.4.3-py3-none-any.whl (4.3 kB)
Requirement already satisfied: networkx in c:\users\hp\anaconda3\lib\site-packages
Requirement already satisfied: xarray in c:\users\hp\anaconda3\lib\site-packages
Requirement already satisfied: six in c:\users\hp\anaconda3\lib\site-packages (f
Requirement already satisfied: itsdangerous in c:\users\hp\anaconda3\lib\site-pa
Collecting flask-ngrok
  Using cached flask_ngrok-0.0.25-py3-none-any.whl (3.1 kB)
Requirement already satisfied: scikit-learn in c:\users\hp\anaconda3\lib\site-pa
Requirement already satisfied: et-xmlfile in c:\users\hp\anaconda3\lib\site-pac
Requirement already satisfied: plotly>=5.0.0 in c:\users\hp\anaconda3\lib\site-p
Requirement already satisfied: statsmodels in c:\users\hp\anaconda3\lib\site-pac
Collecting Flask-Compress
  Using cached Flask_Compress-1.13-py3-none-any.whl (7.9 kB)

Collecting dash-daq
  Using cached dash_daq-0.5.0.tar.gz (642 kB)
Requirement already satisfied: requests in c:\users\hp\anaconda3\lib\site-pac
Requirement already satisfied: openpyxl in c:\users\hp\anaconda3\lib\site-pac
Collecting missingno<=0.4.2
  Using cached missingno-0.4.2-py3-none-any.whl (9.7 kB)
Requirement already satisfied: matplotlib in c:\users\hp\anaconda3\lib\site-pac
Requirement already satisfied: Flask in c:\users\hp\anaconda3\lib\site-packages
Collecting dash>=2.0.0
  Using cached dash-2.6.2-py3-none-any.whl (9.8 MB)
Collecting dash-colorscales
  Using cached dash_colorscales-0.0.4.tar.gz (62 kB)
Collecting kaleido
  Downloading kaleido-0.2.1-py2.py3-none-win_amd64.whl (65.9 MB)
Requirement already satisfied: pandas in c:\users\hp\anaconda3\lib\site-packages
Requirement already satisfied: certifi in c:\users\hp\anaconda3\lib\site-package
Requirement already satisfied: numpy in c:\users\hp\anaconda3\lib\site-packages
Collecting strsimpy
  Using cached strsimpy-0.2.1-py3-none-any.whl (45 kB)
Requirement already satisfied: cycler in c:\users\hp\anaconda3\lib\site-packages
Requirement already satisfied: future>=0.14.0 in c:\users\hp\anaconda3\lib\site-
Collecting lz4
  Using cached lz4-4.0.2-cp39-cp39-win_amd64.whl (98 kB)
Collecting dash-bootstrap-components
  Using cached dash_bootstrap_components-1.2.1-py3-none-any.whl (216 kB)
Requirement already satisfied: scipy in c:\users\hp\anaconda3\lib\site-packages
Requirement already satisfied: xlrd in c:\users\hp\anaconda3\lib\site-packages (
Collecting dash-html-components==2.0.0
  Downloading dash_html_components-2.0.0-py3-none-any.whl (4.1 kB)
Collecting dash-table==5.0.0
  Downloading dash_table-5.0.0-py3-none-any.whl (3.9 kB)
Collecting dash-core-components==2.0.0
  Downloading dash_core_components-2.0.0-py3-none-any.whl (3.8 kB)
Requirement already satisfied: Werkzeug>=2.2.2 in c:\users\hp\anaconda3\lib\site
Requirement already satisfied: click>=8.0 in c:\users\hp\anaconda3\lib\site-pac
Requirement already satisfied: Jinja2>=3.0 in c:\users\hp\anaconda3\lib\site-pac
Requirement already satisfied: importlib-metadata>=3.6.0 in c:\users\hp\anacond
Requirement already satisfied: colorama in c:\users\hp\anaconda3\lib\site-pac
Requirement already satisfied: zipp>=0.5 in c:\users\hp\anaconda3\lib\site-pac
```

```
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\hp\anaconda3\lib\site-packages\MarkupSafe-2.0-py3.9.egg\__init__.py
Requirement already satisfied: tenacity>=6.2.0 in c:\users\hp\anaconda3\lib\site-packages\tencore-1.1.0-py3.9.egg\__init__.py
Collecting brotli
  Downloading Brotli-1.0.9-cp39-cp39-win_amd64.whl (383 kB)
Requirement already satisfied: pillow>=6.2.0 in c:\users\hp\anaconda3\lib\site-packages\Pillow-8.2.0-py3.9.egg\__init__.py
Requirement already satisfied: packaging>=20.0 in c:\users\hp\anaconda3\lib\site-packages\packaging-20.4-py3.9.egg\__init__.py
Requirement already satisfied: python-dateutil>=2.7 in c:\users\hp\anaconda3\lib\site-packages\python_dateutil-2.8.1-py3.9.egg\__init__.py
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\hp\anaconda3\lib\site-packages\pyparsing-2.4.7-py3.9.egg\__init__.py
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\hp\anaconda3\lib\site-packages\kiwisolver-1.4.0-py3.9.egg\__init__.py
Requirement already satisfied: fonttools>=4.22.0 in c:\users\hp\anaconda3\lib\site-packages\fontTools-4.22.0-py3.9.egg\__init__.py
Requirement already satisfied: pytz>=2020.1 in c:\users\hp\anaconda3\lib\site-packages\pytz-2020.1-py3.9.egg\__init__.py
Requirement already satisfied: idna<4,>=2.5 in c:\users\hp\anaconda3\lib\site-packages\idna-2.10-py3.9.egg\__init__.py
Requirement already satisfied: charset-normalizer~2.0.0 in c:\users\hp\anaconda3\lib\site-packages\charset_normalizer-2.0.0-py3.9.egg\__init__.py
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\hp\anaconda3\lib\site-packages\urllib3-1.26.12-py3.9.egg\__init__.py
Requirement already satisfied: joblib>=0.11 in c:\users\hp\anaconda3\lib\site-packages\joblib-1.0.1-py3.9.egg\__init__.py
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\hp\anaconda3\lib\site-packages\threadpoolctl-2.2.0-py3.9.egg\__init__.py
Requirement already satisfied: patsy>=0.5.2 in c:\users\hp\anaconda3\lib\site-packages\patsy-0.5.2-py3.9.egg\__init__.py
Building wheels for collected packages: dash-colorscales, dash-daq
  Building wheel for dash-colorscales (setup.py): started
  Building wheel for dash-colorscales (setup.py): finished with status 'done'
  Created wheel for dash-colorscales: filename=dash_colorscale-0.0.4-py3-none-any.whl size=6697
  Stored in directory: c:\users\hp\appdata\local\pip\cache\wheels\ec\32\ec\dc8a1
  Building wheel for dash-daq (setup.py): started
  Building wheel for dash-daq (setup.py): finished with status 'done'
  Created wheel for dash-daq: filename=dash_daq-0.5.0-py3-none-any.whl size=6697
  Stored in directory: c:\users\hp\appdata\local\pip\cache\wheels\53\73\ea\8a422
Successfully built dash-colorscales dash-daq
Installing collected packages: brotli, Flask-Compress, dash-table, dash-html-compat, dash
  Attempting uninstall: missingno
    Found existing installation: missingno 0.5.1
    Uninstalling missingno-0.5.1:
      Successfully uninstalled missingno-0.5.1
Successfully installed Flask-Compress-1.13 brotli-1.0.9 dash-2.6.2 dash-bootstr
```

In [9]: `import dtale`

```
In [10]: import pandas as pd  
df = pd.read_csv("E:\\DATA SCIENCE\\Dataset\\data7\\train.csv")
```

Now after importing the dtale lib we can view our dataframe in dtale table .

```
In [11]: df_table = dtale.show(df)
```

Unlike other libraries dtale does not give us direct analysis but instead gives us a format like excel with powerful and direct tools to analyze the data, using which we can directly do analysis of data and also visualize it using different plots.

In [12]:

```
d_table
```

Out[12]:

In the top we see many options and on choosing any one of the option we can directly open a new window in web browser where we can do manipulation on the data and also use various Visualization features

Using AutoViz

In [1]:

```
pip install autoviz --user --quiet --upgrade
```

Note: you may need to restart the kernel to use updated packages.

```
WARNING: The script panel.exe is installed in 'C:\Users\HP\AppData\Roaming\Pyt
Consider adding this directory to PATH or, if you prefer to suppress this warr
ERROR: pip's dependency resolver does not currently take into account all the pa
spyder 5.1.5 requires pyqt5<5.13, which is not installed.
spyder 5.1.5 requires pyqtwebengine<5.13, which is not installed.
conda-repo-cli 1.0.4 requires pathlib, which is not installed.
anaconda-project 0.10.2 requires ruamel-yaml, which is not installed.
distributed 2022.2.1 requires dask==2022.02.1, but you have dask 2022.10.0 which
```

In [1]:

```
from autoviz.AutoViz_Class import AutoViz_Class  
%matplotlib inline
```



Imported v0.1.58. After importing, execute '%matplotlib inline' to display chart
AV = AutoViz_Class()
dfte = AV.AutoViz(filename, sep=',', depVar='', dfte=None, header=0, verbose
chart_format='svg',max_rows_analyzed=150000,max_cols_analyzed=30,
Update: verbose=0 displays charts in your local Jupyter notebook.
verbose=1 additionally provides EDA data cleaning suggestions. It also c
verbose=2 does not display charts but saves them in AutoViz_Plots folder
chart_format='bokeh' displays charts in your local Jupyter notebook.
chart_format='server' displays charts in your browser: one tab for each
chart_format='html' silently saves interactive HTML files in your local

In [2]:

```
av = AutoViz_Class()
```

In []:

```
df = av.AutoViz(r'E:\DATA_SCIENCE\Dataset\data7\train.csv', save_plot_dir=r'EDA
```

Shape of your Data Set loaded: (8523, 33)

CLASSIFYING VARIABLES

Classifying variables in data set...

Data cleaning improvement suggestions. Complete them before proceeding to ML mod

	Nuniques	dtype	Nulls	Nullpercent	NuniquePercent	Value counts Min	Data cleaning improvement suggestions
Item_Visibility	8325	float64	0	0.000000	97.676874	0	skewed: ca or drc outlie
Item_MRP	5938	float64	0	0.000000	69.670304	0	
Item_Outlet_Sales	3493	float64	0	0.000000	40.983222	0	skewed: ca or drc outlie
Item_Identifier	1559	object	0	0.000000	18.291681	1	combine rare categories
Item_Weight	555	float64	0	0.000000	6.511792	0	
Outlet_Identifier	10	object	0	0.000000	0.117330	528	
Outlet_Years	9	int64	0	0.000000	0.105597	0	
Outlet_3	2	int64	0	0.000000	0.023466	0	
Item_Type_Combined_2	2	int64	0	0.000000	0.023466	0	
Outlet_0	2	int64	0	0.000000	0.023466	0	
Outlet_1	2	int64	0	0.000000	0.023466	0	
Outlet_2	2	int64	0	0.000000	0.023466	0	
Outlet_6	2	int64	0	0.000000	0.023466	0	
Outlet_4	2	int64	0	0.000000	0.023466	0	
Outlet_5	2	int64	0	0.000000	0.023466	0	
Item_Type_Combined_0	2	int64	0	0.000000	0.023466	0	
Outlet_7	2	int64	0	0.000000	0.023466	0	
Outlet_8	2	int64	0	0.000000	0.023466	0	
Item_Type_Combined_1	2	int64	0	0.000000	0.023466	0	
Outlet_Type_0	2	int64	0	0.000000	0.023466	0	
Outlet_Type_3	2	int64	0	0.000000	0.023466	0	
Outlet_Type_2	2	int64	0	0.000000	0.023466	0	
Outlet_Type_1	2	int64	0	0.000000	0.023466	0	

	Nuniques	dtype	Nulls	Nullpercent	NuniquePercent	Value counts	Min	Data cleaning improvement suggestions
Outlet_Size_3	2	int64	0	0.000000	0.023466	0	0	
Outlet_Size_2	2	int64	0	0.000000	0.023466	0	0	
Outlet_Size_1	2	int64	0	0.000000	0.023466	0	0	
Outlet_Size_0	2	int64	0	0.000000	0.023466	0	0	
Outlet_Location_Type_2	2	int64	0	0.000000	0.023466	0	0	
Outlet_Location_Type_1	2	int64	0	0.000000	0.023466	0	0	
Outlet_Location_Type_0	2	int64	0	0.000000	0.023466	0	0	
Item_Fat_Content_1	2	int64	0	0.000000	0.023466	0	0	
Item_Fat_Content_0	2	int64	0	0.000000	0.023466	0	0	
Outlet_9	2	int64	0	0.000000	0.023466	0	0	

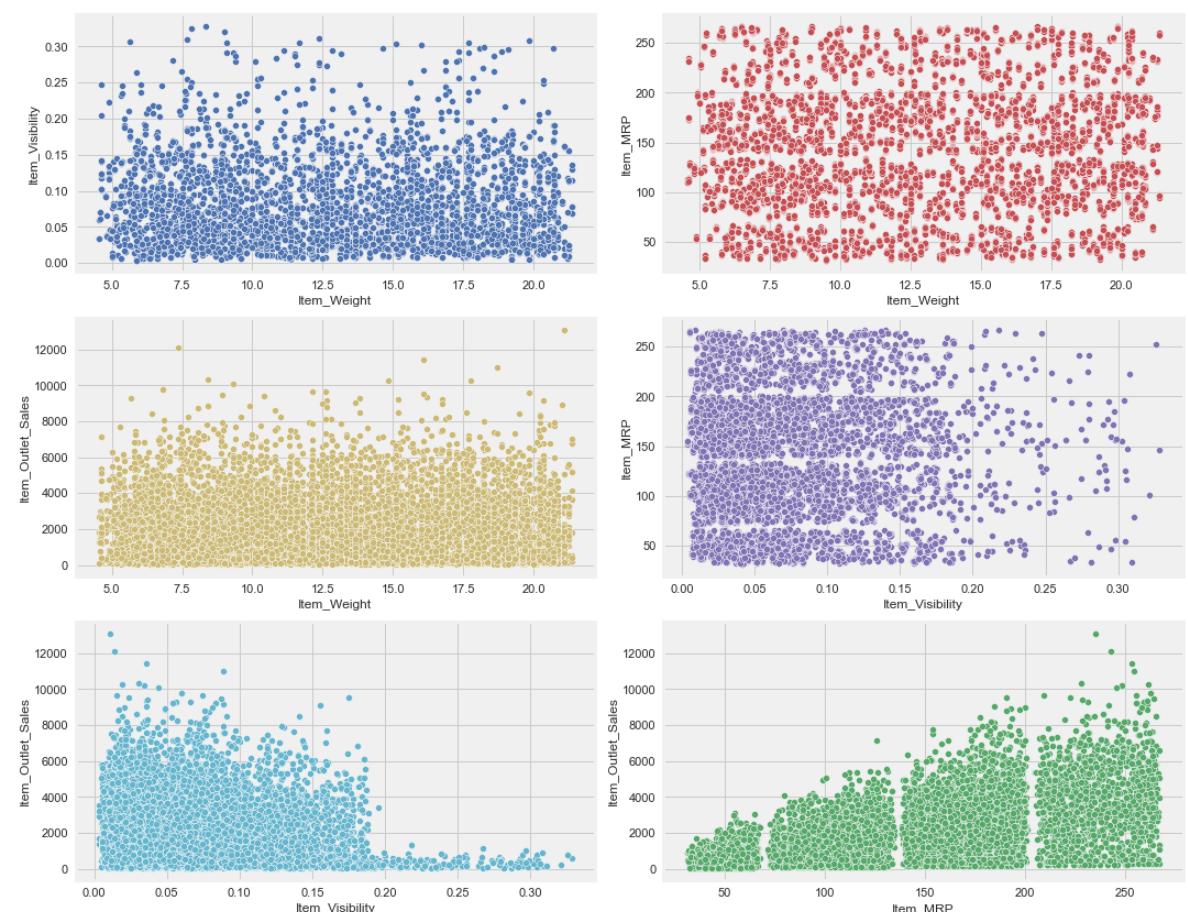
33 Predictors classified...

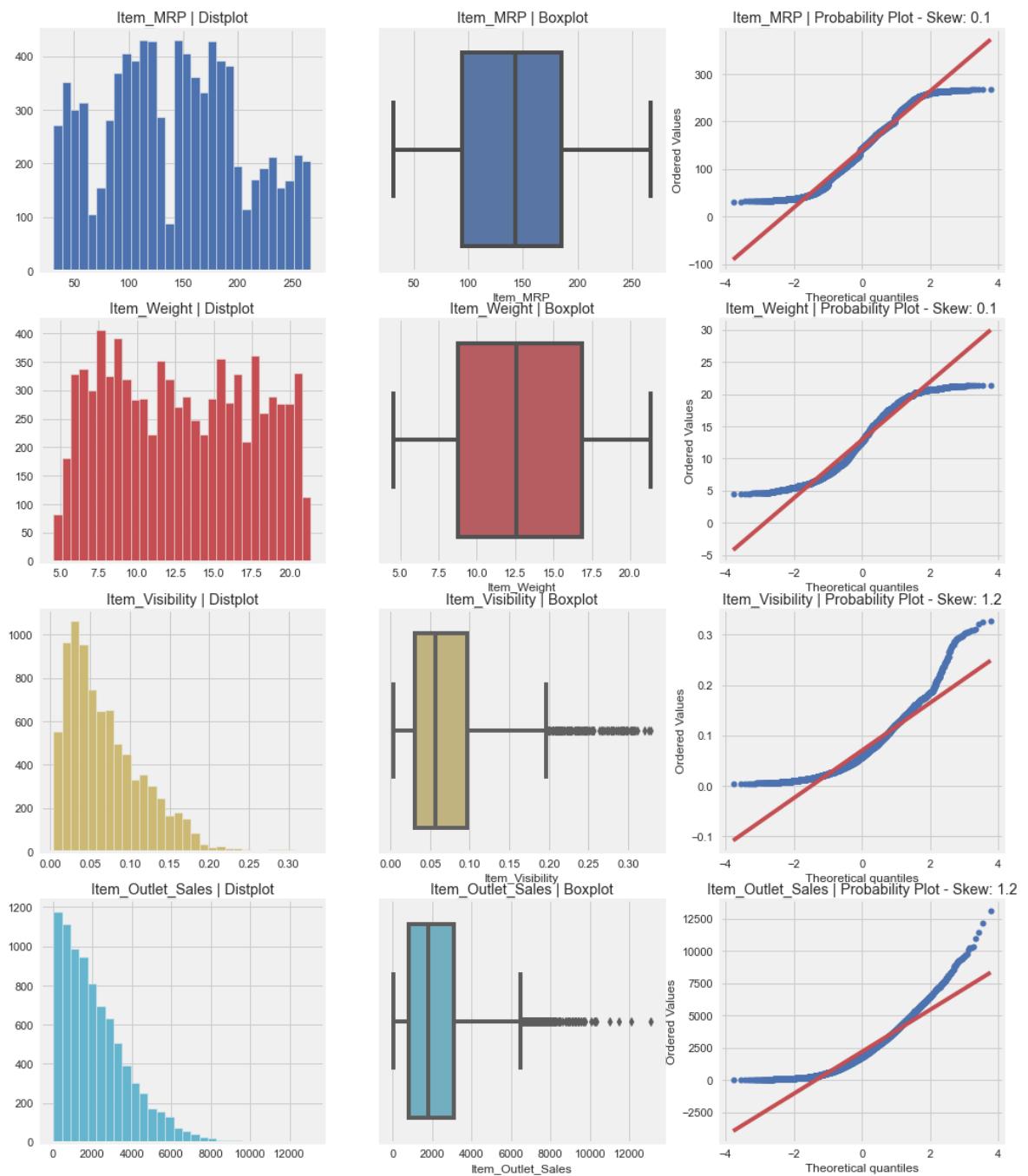
No variables removed since no ID or low-information variables found in column
4 numeric variables in data exceeds limit, taking top 30 variables

List of variables selected: ['Item_Weight', 'Item_Visibility', 'Item_MRP', 'Item_Outlet_Sales']
Total columns > 30, too numerous to print.

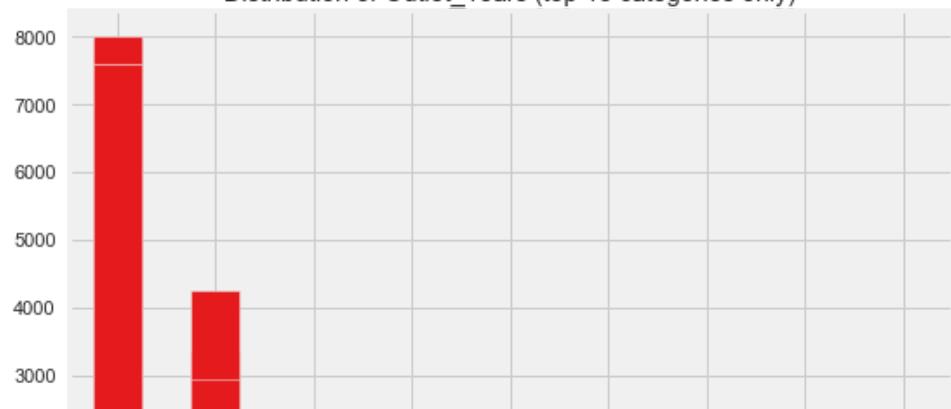
Number of All Scatter Plots = 10

Pair-wise Scatter Plot of all Continuous Variables



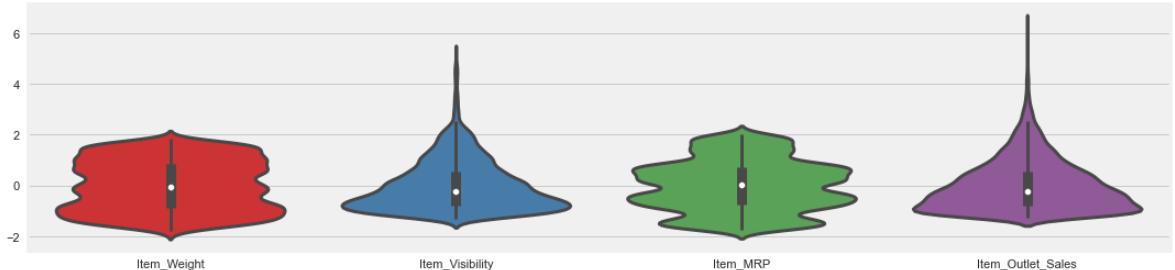


Distribution of Outlet_Years (top 15 categories only)

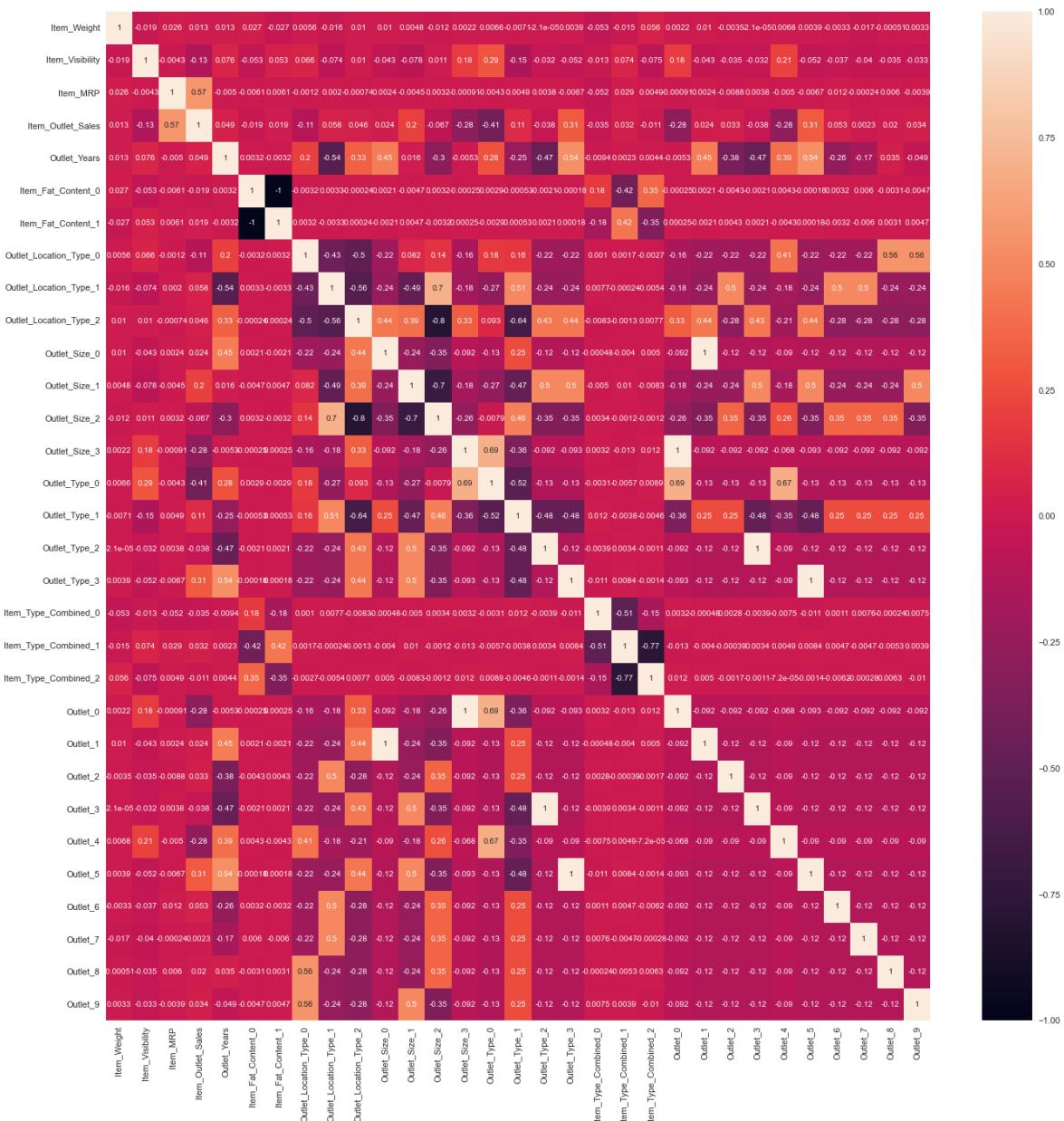




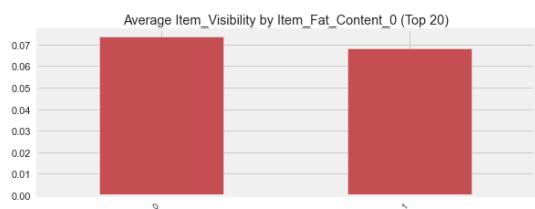
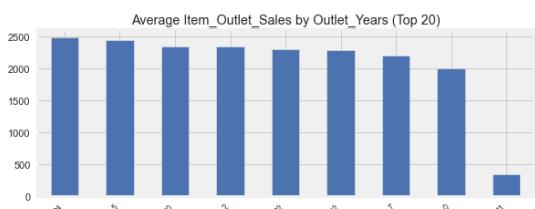
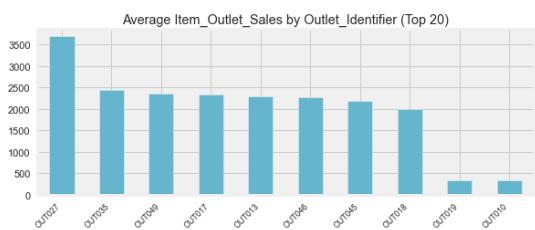
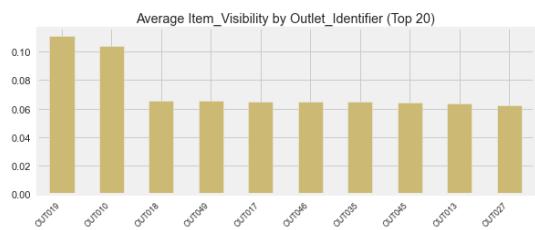
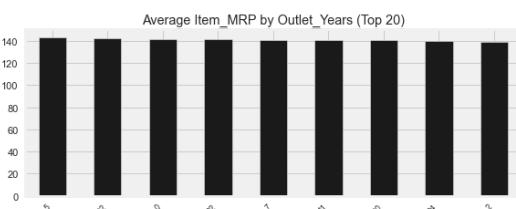
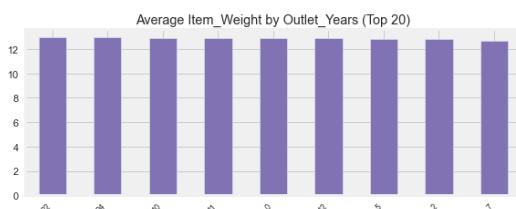
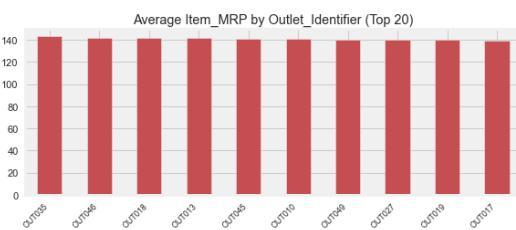
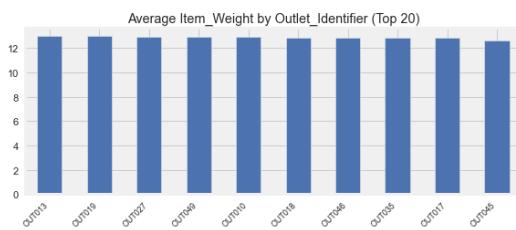
Violin Plot of all Continuous Variables



Heatmap of all Continuous Variables including target =



Bar plots for each Continuous by each Categorical variable

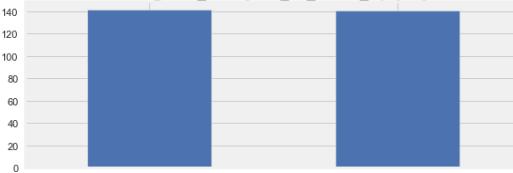




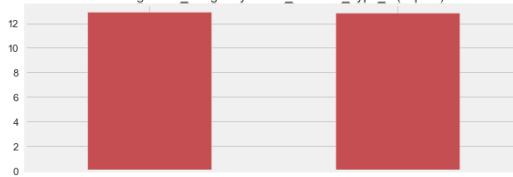
Average Item_Weight by Item_Fat_Content_1 (Top 20)



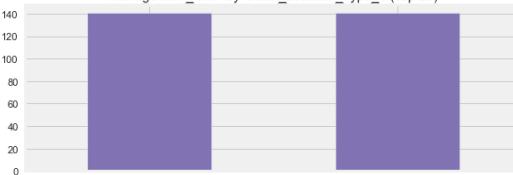
Average Item_MRP by Item_Fat_Content_1 (Top 20)



Average Item_Weight by Outlet_Location_Type_0 (Top 20)



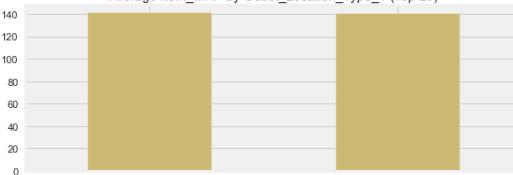
Average Item_MRP by Outlet_Location_Type_0 (Top 20)



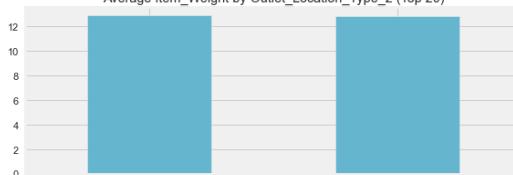
Average Item_Weight by Outlet_Location_Type_1 (Top 20)



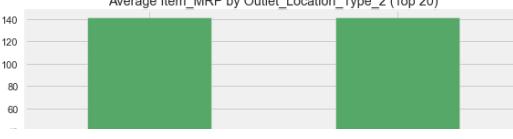
Average Item_MRP by Outlet_Location_Type_1 (Top 20)



Average Item_Weight by Outlet_Location_Type_2 (Top 20)



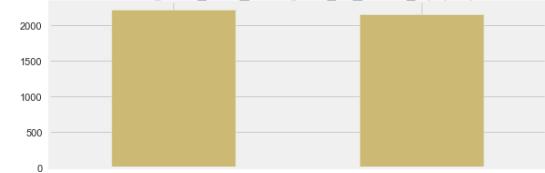
Average Item_MRP by Outlet_Location_Type_2 (Top 20)



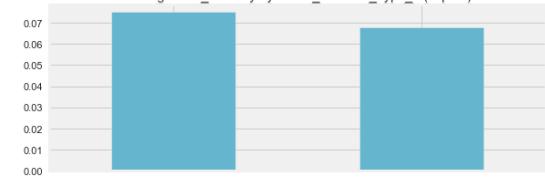
Average Item_Visibility by Item_Fat_Content_1 (Top 20)



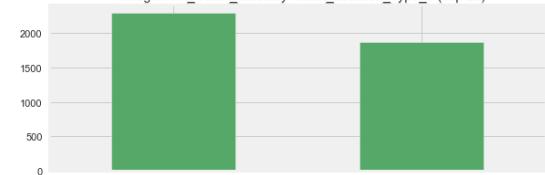
Average Item_Outlet_Sales by Item_Fat_Content_1 (Top 20)



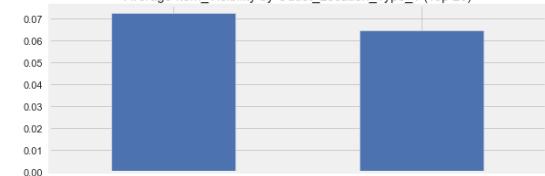
Average Item_Visibility by Outlet_Location_Type_0 (Top 20)



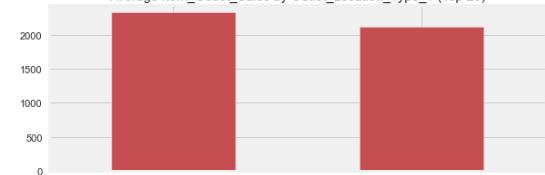
Average Item_Outlet_Sales by Outlet_Location_Type_0 (Top 20)



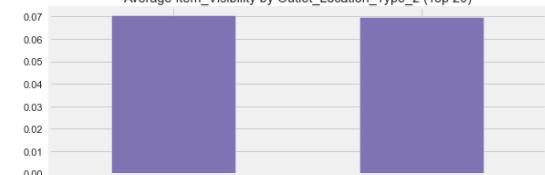
Average Item_Visibility by Outlet_Location_Type_1 (Top 20)



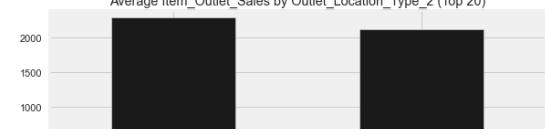
Average Item_Outlet_Sales by Outlet_Location_Type_1 (Top 20)

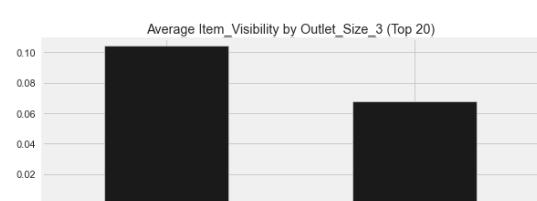
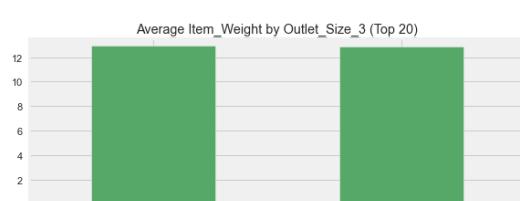
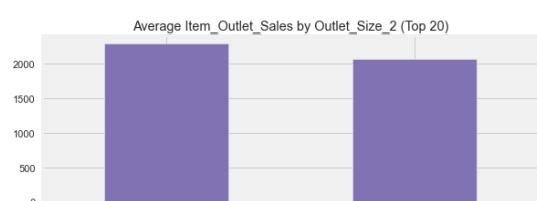
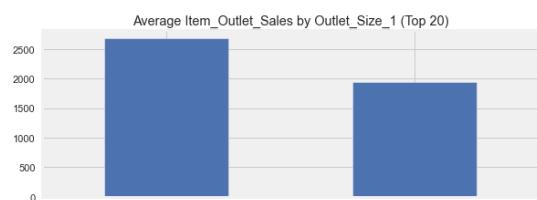
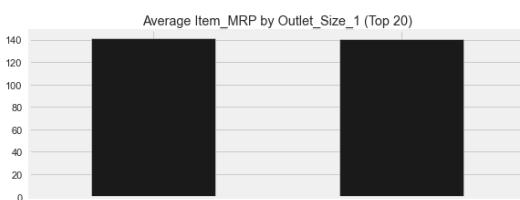
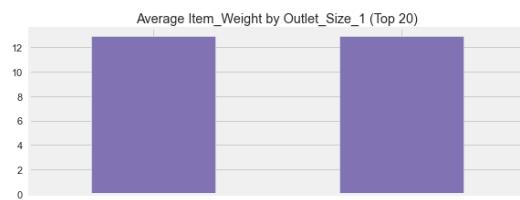
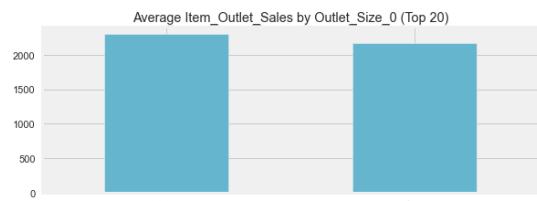
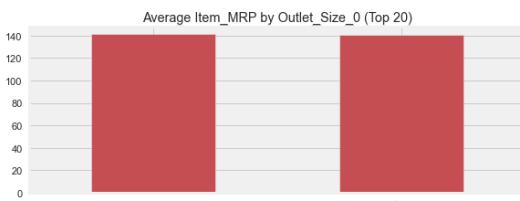
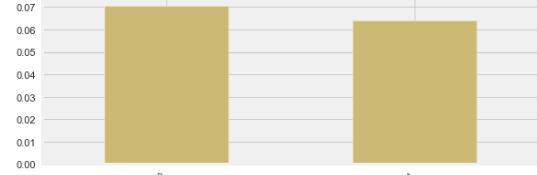


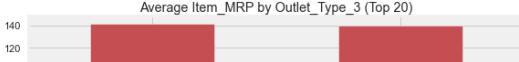
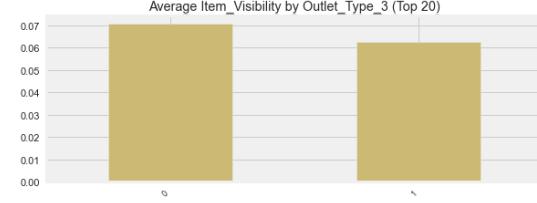
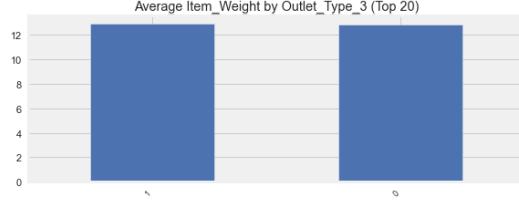
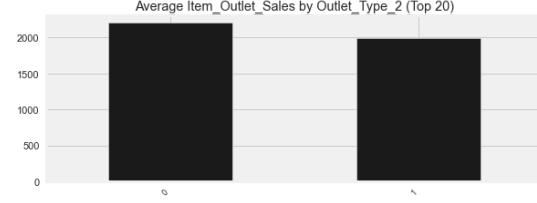
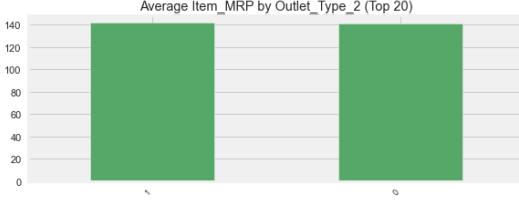
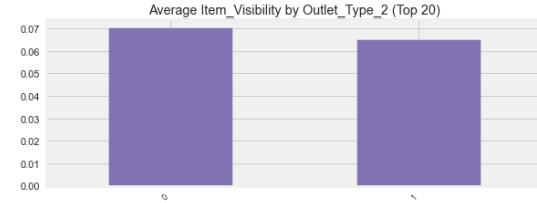
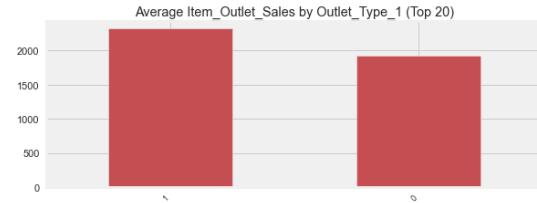
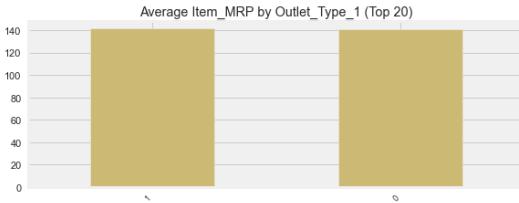
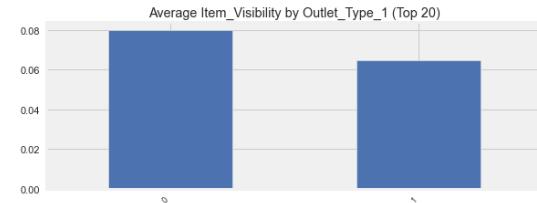
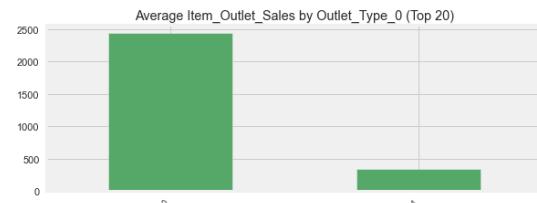
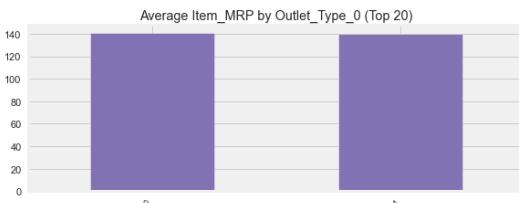
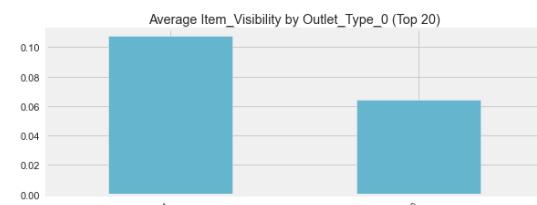
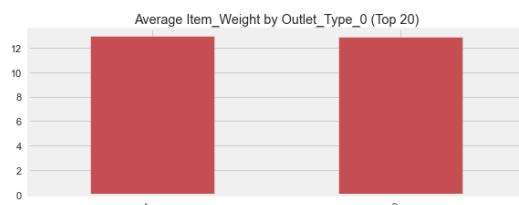
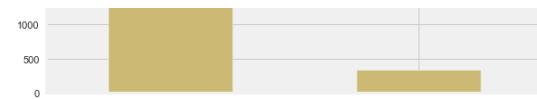
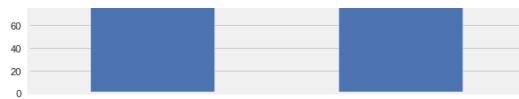
Average Item_Visibility by Outlet_Location_Type_2 (Top 20)

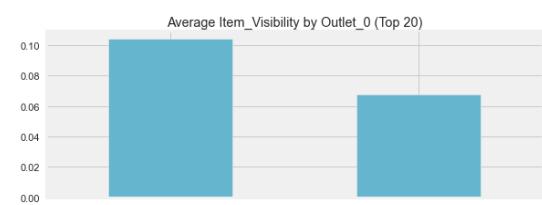
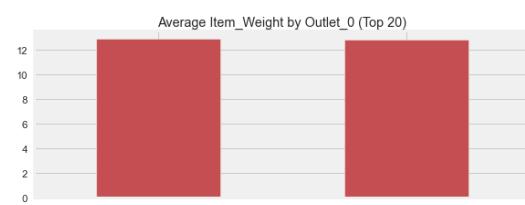
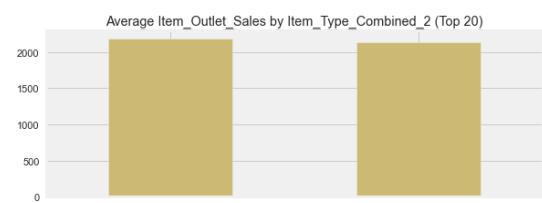
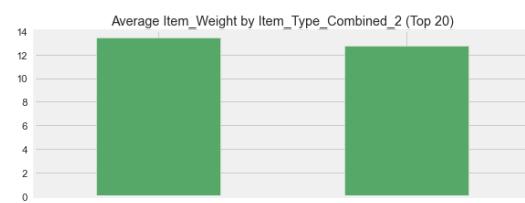
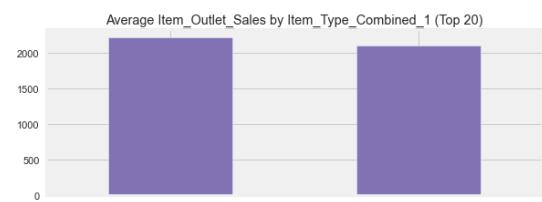
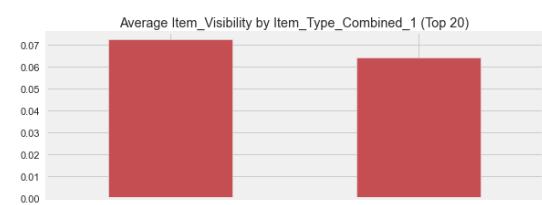
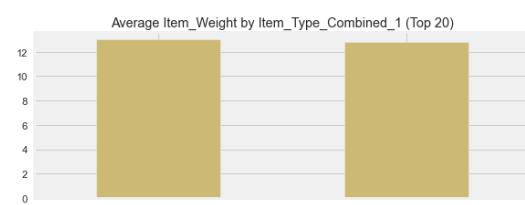
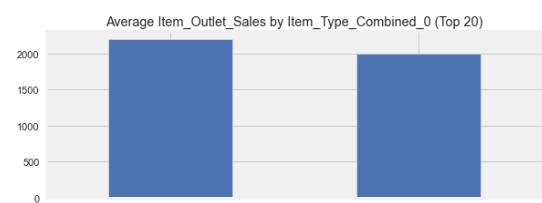
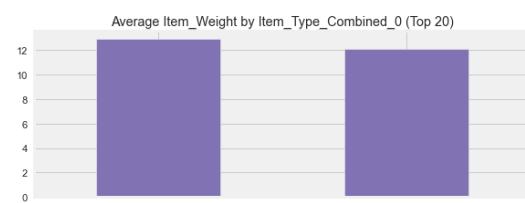
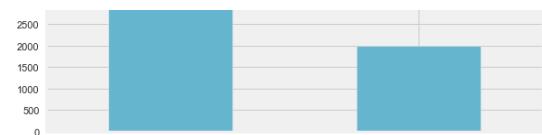
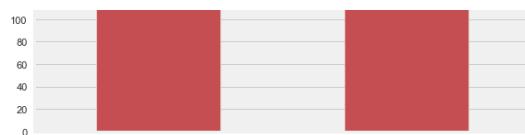


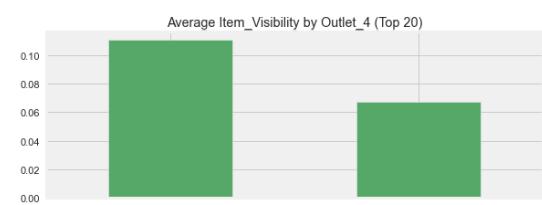
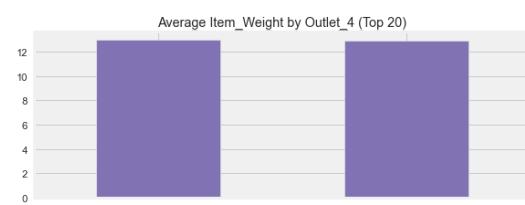
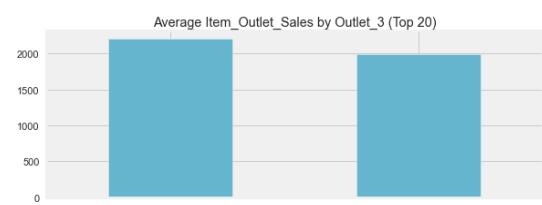
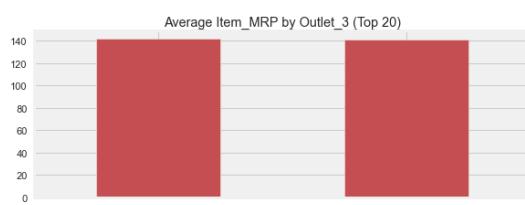
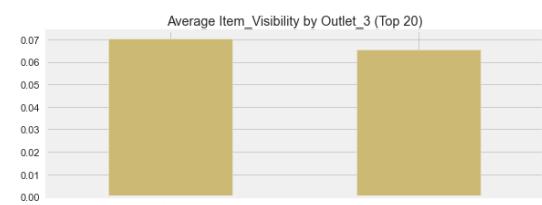
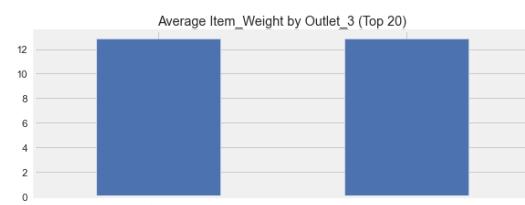
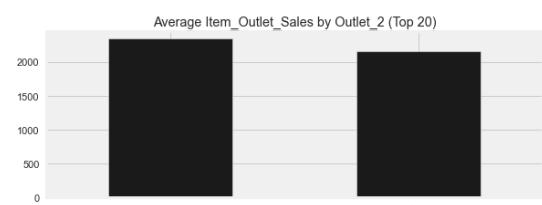
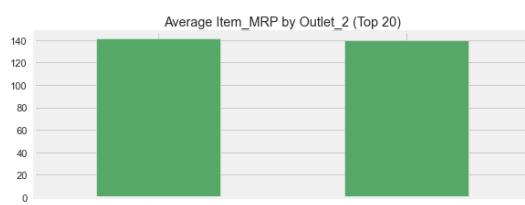
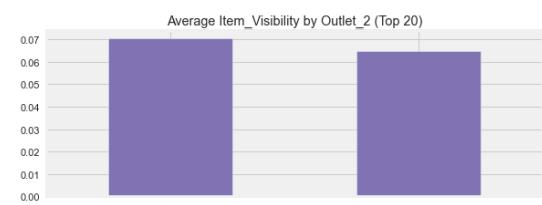
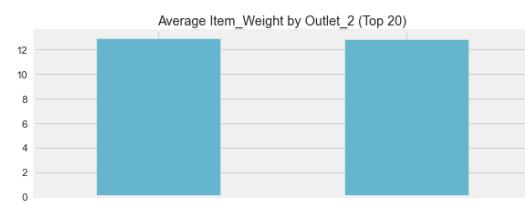
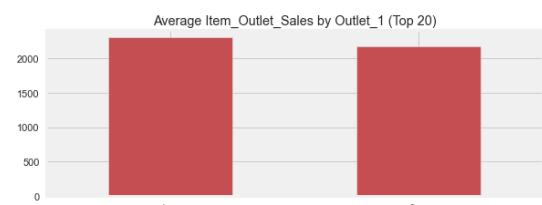
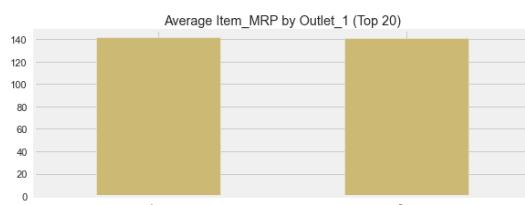
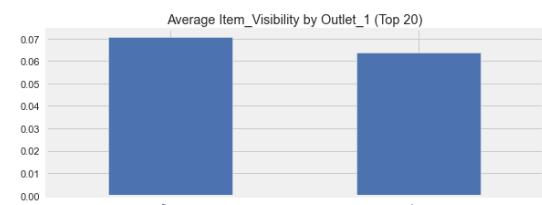
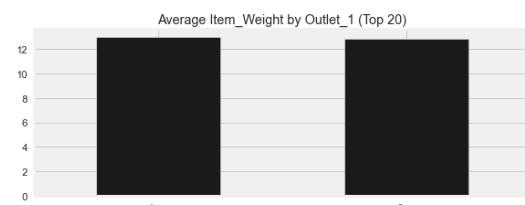
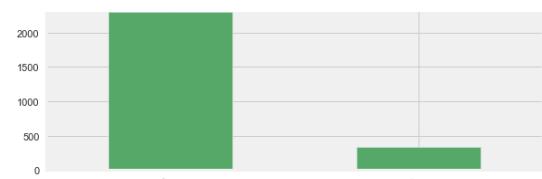
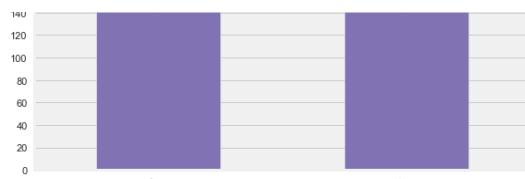
Average Item_Outlet_Sales by Outlet_Location_Type_2 (Top 20)

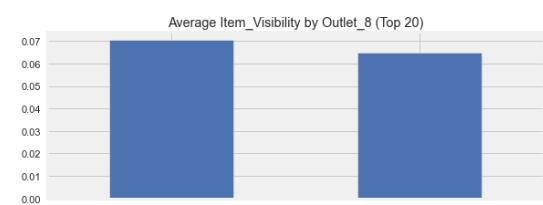
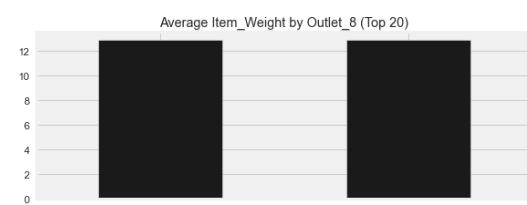
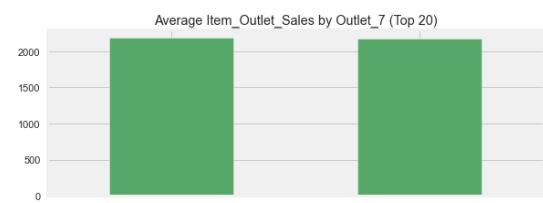
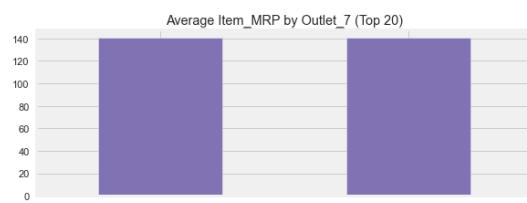
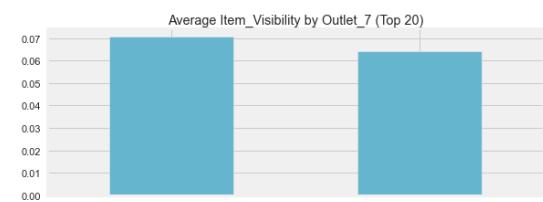
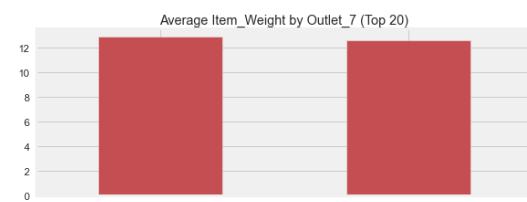
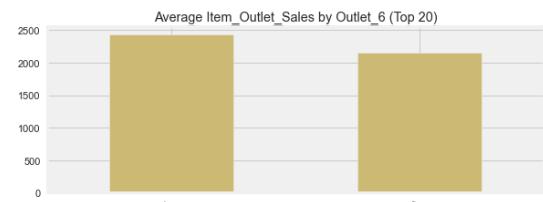
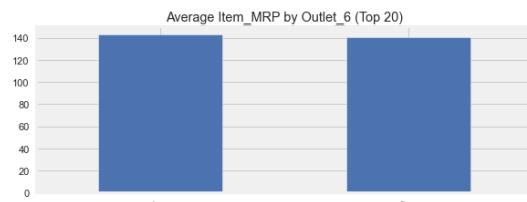
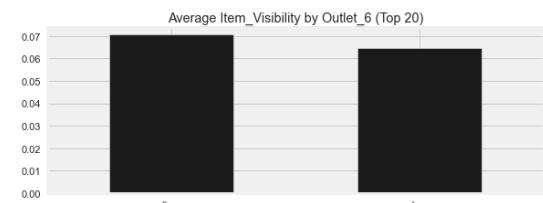
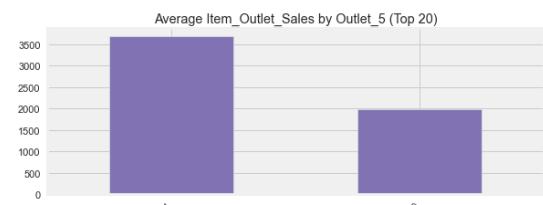
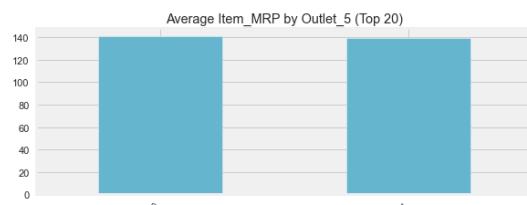
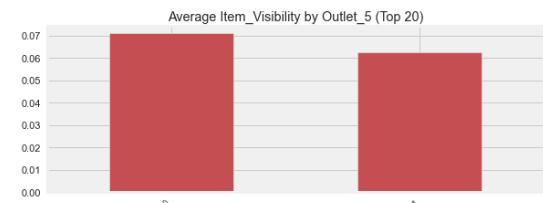
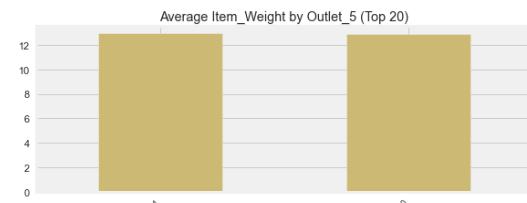
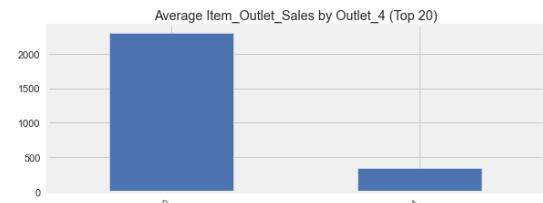
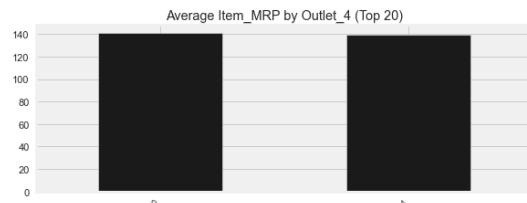


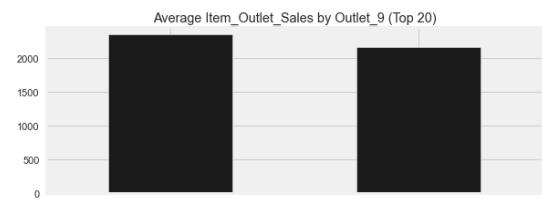
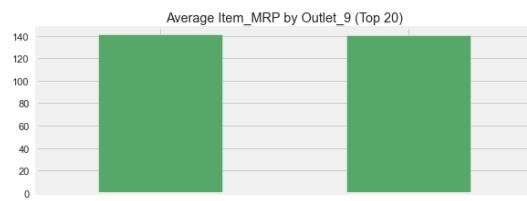
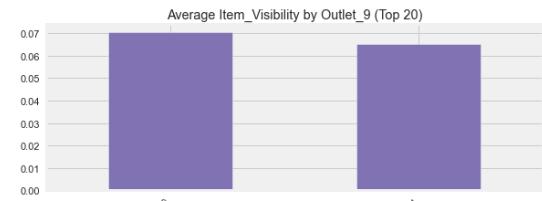
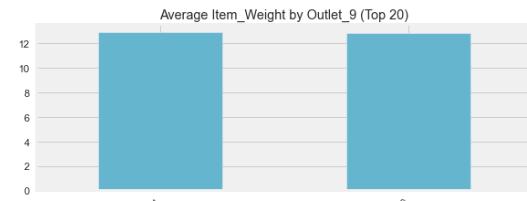
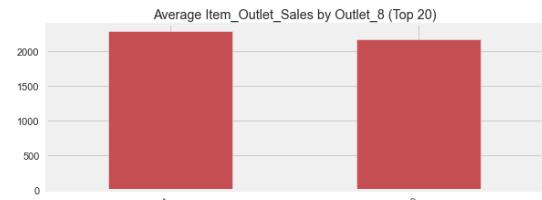
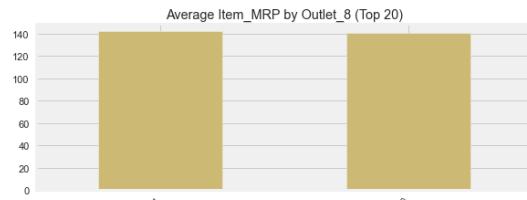












```
[nltk_data] Downloading collection 'popular'
[nltk_data] |   Downloading package cmudict to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\cmudict.zip.
[nltk_data] |   Downloading package gazetteers to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\gazetteers.zip.
[nltk_data] |   Downloading package genesis to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\genesis.zip.
[nltk_data] |   Downloading package gutenberg to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\gutenberg.zip.
[nltk_data] |   Downloading package inaugural to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\inaugural.zip.
[nltk_data] |   Downloading package movie_reviews to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\movie_reviews.zip.
[nltk_data] |   Downloading package names to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\names.zip.
[nltk_data] |   Downloading package shakespeare to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\shakespeare.zip.
[nltk_data] |   Downloading package stopwords to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\stopwords.zip.
[nltk_data] |   Downloading package treebank to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\treebank.zip.
[nltk_data] |   Downloading package twitter_samples to
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] |       Unzipping corpora\twitter_samples.zip.
[nltk_data] |   Downloading package omw to
```

```
[nltk_data] | Downloading package emoji to  
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Downloading package omw-1.4 to  
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...  
[nltk_data] | Downloading package wordnet to  
[nltk_data] |     C:\Users\HP\AppData\Roaming\nltk_data...
```

ON using Autoviz we can see that it gives all the essential details about the data in a tabular form highlighting all the important aspects of each feature. Also if we keep scrolling down we find that the visualization of data does automatically and all types of charts, graphs , bar , pie , violin , histo , distribution etc . all are available with comprising different features .

In []: