AutoML EDA Libraryes

October 3, 2021

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
# D-Tale
df=pd.read_csv('train.csv')
```

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

[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

1 D-Tale

[6]:
```
#using d-tale use
import dtale
```

[ ]:
```
#Report:

1.The dtale library generates a report having:
2.An overview of the dataset Custom filters
3.Correlation, Charts, and Heatmaps
4.Highlight datatypes, missing values, ranges Code export
```

[7]:
```
dtale.show(df)
```

```
http://LAPTOP-1MT6GSPA:40000/dtale/main/1
It looks like this data may have already been loaded to D-Tale based on shape
and column names. Here is URL of the data that seems to match it:

None

If you still want to load this data please use the following command:

dtale.show(df, ignore_duplicate=True)
```

[8]:
```
dtale.show(df, ignore_duplicate=True)
```

```
<IPython.lib.display.IFrame at 0x1b62c1a2908>
```

[8]:

2 pandas profiling

[9]:
```
from pandas_profiling import ProfileReport
```

[ ]:
```
#Report:

The pandas-profiling library generates a report having: An overview of the␣
 ↪dataset
1.Variable properties
2.Interaction of variables Correlation of variables
3.Missing values
4.Sample data
```

[10]:
```
ProfileReport(df)
```

```
HBox(children=(FloatProgress(value=0.0, description='Summarize dataset', max=26.0, style=Progr
```

```
HBox(children=(FloatProgress(value=0.0, description='Generate report structure', max=1.0, style
```

```
HBox(children=(FloatProgress(value=0.0, description='Render HTML', max=1.0, style=ProgressStyle
```

```
<IPython.core.display.HTML object>
```

[10]:

3 sweetviz

[ ]:
```
pip install sweetviz
```

[ ]:
```
#Report:

The Sweetviz library generates a report having:
An overview of the dataset Variable properties Categorical associations␣
 ↪Numerical associations
Most frequent,
smallest, l
argest values for numerical features
```

[ ]:
```
import sweetviz as sv

my_report = sv.analyze(df)
my_report.show_html() # Default arguments will generate to "SWEETVIZ_REPORT.
 ↪html"
```

4 autoviz

[ ]:
```
from autoviz.AutoViz_Class import AutoViz_Class

AV = AutoViz_Class(d
```

```
Executing shutdown due to inactivity...
```

```
2021-10-03 17:25:30,741 - INFO     - Executing shutdown due to inactivity...
```

```
Executing shutdown...
```

```
2021-10-03 17:25:34,909 - INFO     - Executing shutdown...
```

[ ]:
```
df  =  AV.AutoViz('df)
```

[ ]:
```
#Report:
The Autoviz library generates a report having:
An overview of the dataset
Pairwise scatter plot of continuous variables Distribution of categorical ␣
 ↪variables Heatmaps of continuous variables
Average numerical variable by each categorical variable
```