# Automated EDA - AutoViz

September 30, 2022

[1]:
```
#pip install autoviz
```

[2]:
```
import seaborn as sns
```

[3]:
```
df = sns.load_dataset('titanic')
df.head()
```

[3]:

| | survived | pclass | sex | age | sibsp | parch | fare | embarked | class | \ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | |

| | who | adult_male | deck | embark_town | alive | alone |
|---|---|---|---|---|---|---|
| 0 | man | True | NaN | Southampton | no | False |
| 1 | woman | False | C | Cherbourg | yes | False |
| 2 | woman | False | NaN | Southampton | yes | True |
| 3 | woman | False | C | Southampton | yes | False |
| 4 | man | True | NaN | Southampton | no | True |

[4]:
```
df.to_csv('titanic.csv')
```

[5]:
```
from autoviz.AutoViz_Class import AutoViz_Class
```

```
Imported v0.1.55. After importing, execute '%matplotlib inline' to display
charts in Jupyter.
    AV = AutoViz_Class()
    dfte = AV.AutoViz(filename, sep=',', depVar='', dfte=None, header=0,
verbose=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 also
displays charts.
        verbose=2 does not display charts but saves them in AutoViz_Plots folder
in local machine.
        chart_format='bokeh' displays charts in your local Jupyter notebook.
```

```
        chart_format='server' displays charts in your browser: one tab for each
chart type
        chart_format='html' silently saves interactive HTML files in your local
machine
```

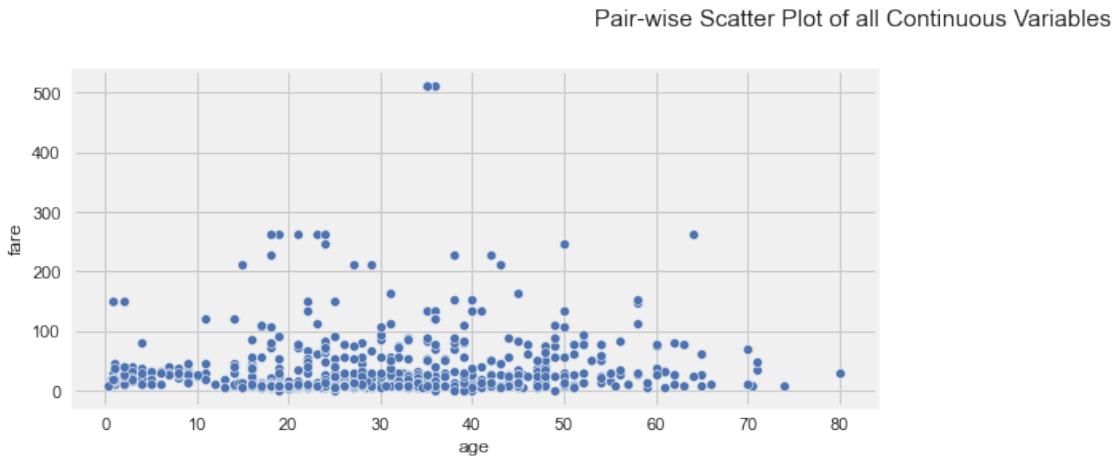
[6]:
```
av = AutoViz_Class()
```

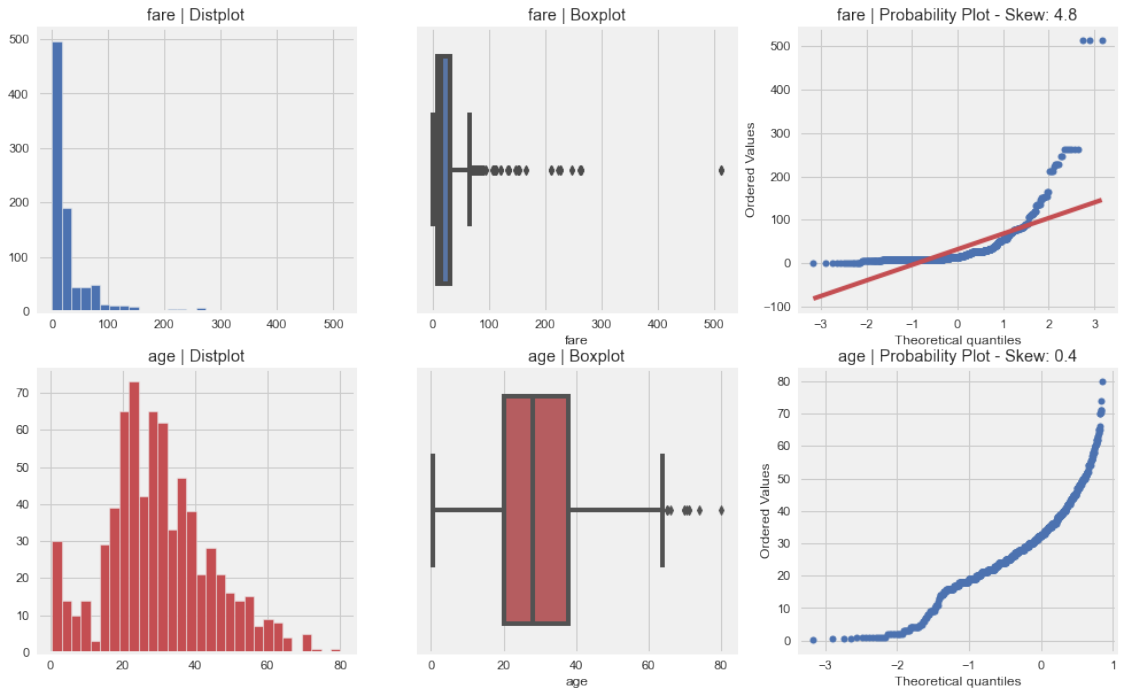
[7]:
```
import matplotlib.pyplot as plt
%matplotlib inline
filename = "titanic.csv"
dft = av.AutoViz(
    filename
)
```

```
Shape of your Data Set loaded: (891, 16)
#############################################################################
#######
######################## C L A S S I F Y I N G  V A R I A B L E S
####################
#############################################################################
#######
Classifying variables in data set…
```

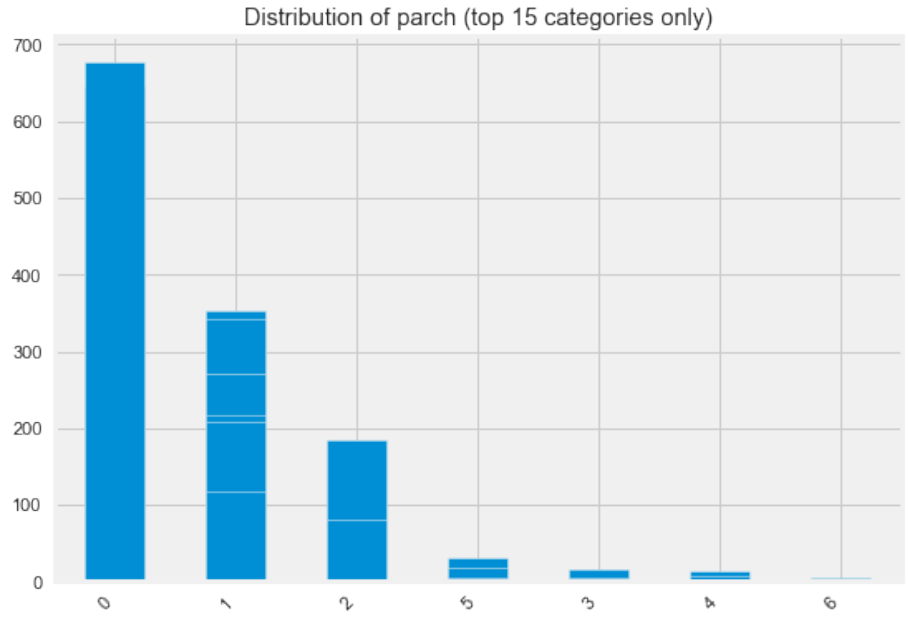
```
<pandas.io.formats.style.Styler at 0x1ffe7c1f5e0>
```

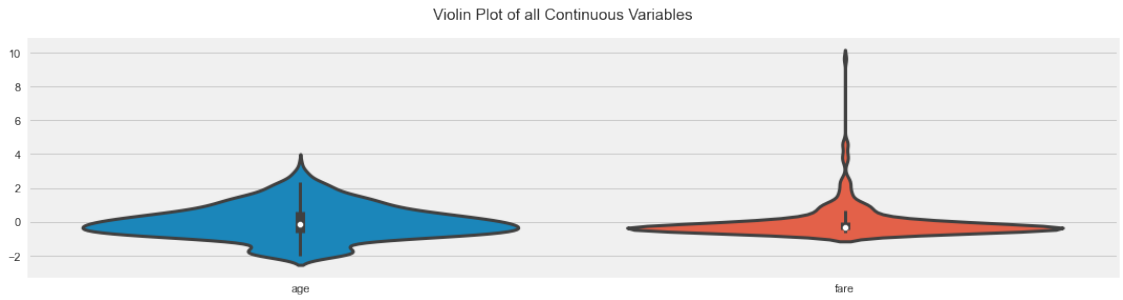
```
    16 Predictors classified…
        1 variables removed since they were ID or low-information variables
        List of variables removed: ['Unnamed: 0']
Number of All Scatter Plots = 3
```

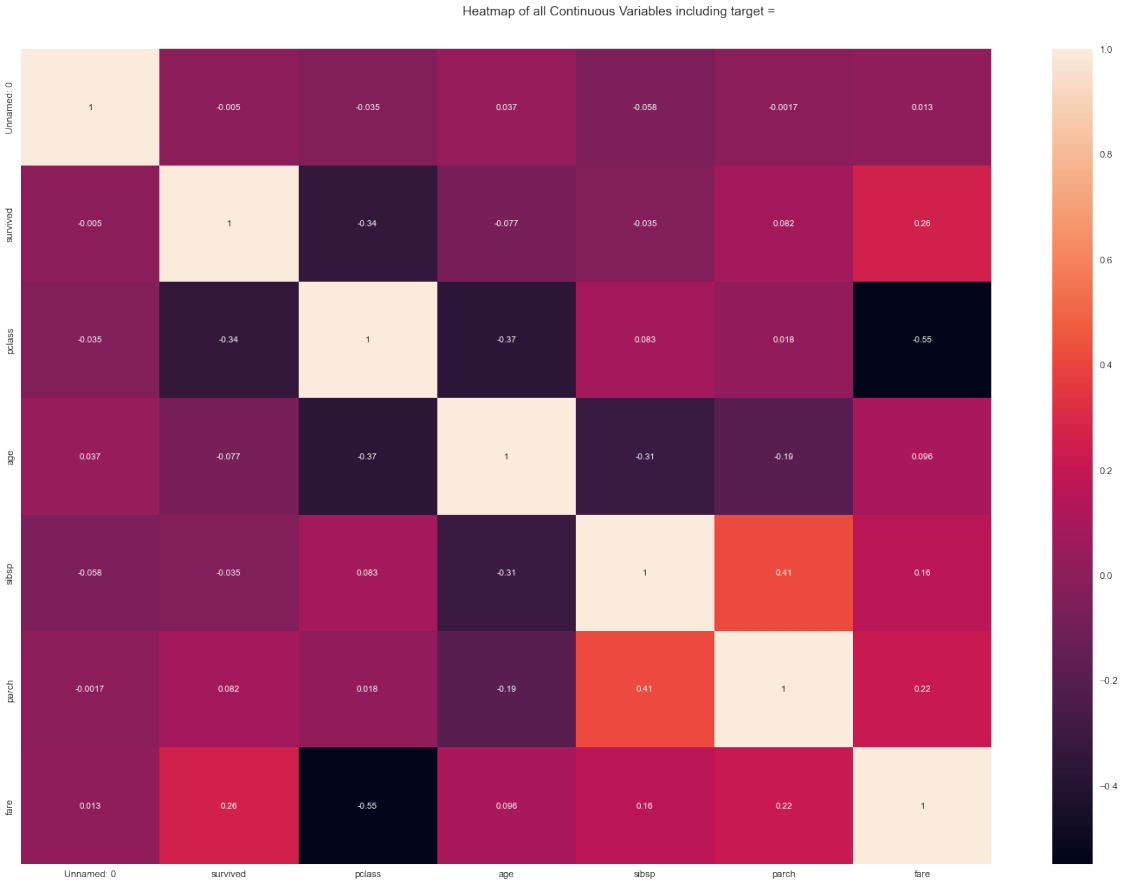




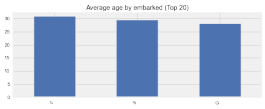

Histograms (KDE plots) of all Continuous Variables

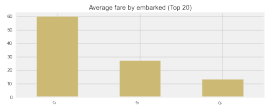


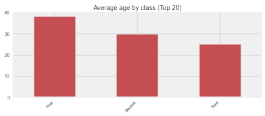

Violin Plot of all Continuous Variables
10
8
6
4
2
0
−2
age
fare

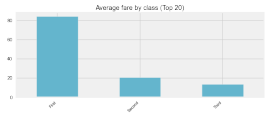

Heatmap of all Continuous Variables including target =
Unnamed: 0
survived
pclass
age
sibsp
parch
fare
1
-0.005
-0.035
0.037
-0.058
-0.0017
0.013
-0.005
1
-0.34
-0.077
-0.035
0.082
0.26
-0.035
-0.34
1
-0.37
0.083
0.018
-0.55
0.037
-0.077
-0.37
1
-0.31
-0.19
0.096
-0.058
-0.035
0.083
-0.31
1
0.41
0.16
-0.0017
0.082
0.018
-0.19
0.41
1
0.22
0.013
0.26
-0.55
0.096
0.16
0.22
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

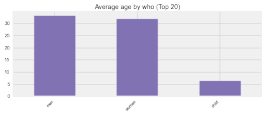
Bar plots for each Continuous by each Categorical variable

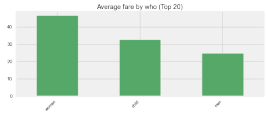


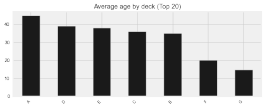



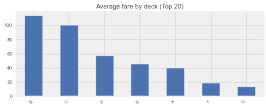



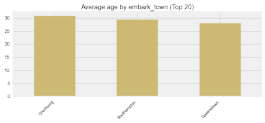



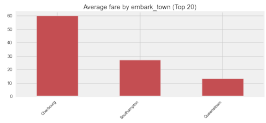



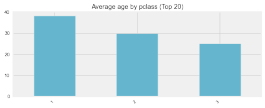



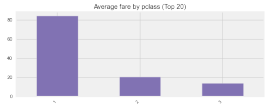



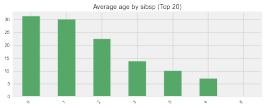



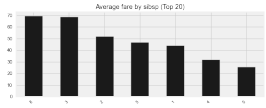



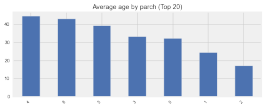



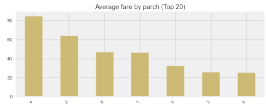



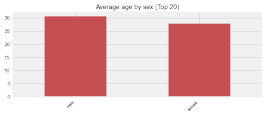



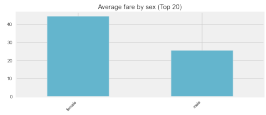



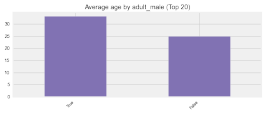



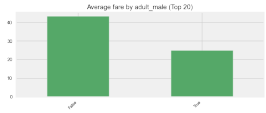



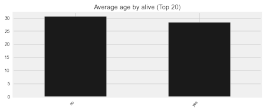



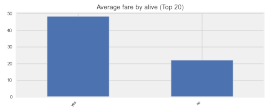



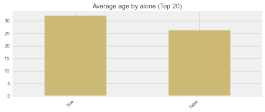



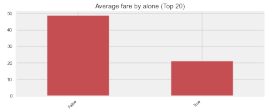



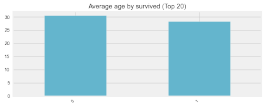



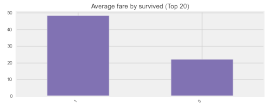



```
All Plots done
Time to run AutoViz = 4 seconds

 ###################### AUTO VISUALIZATION Completed ########################
```

[ ]: