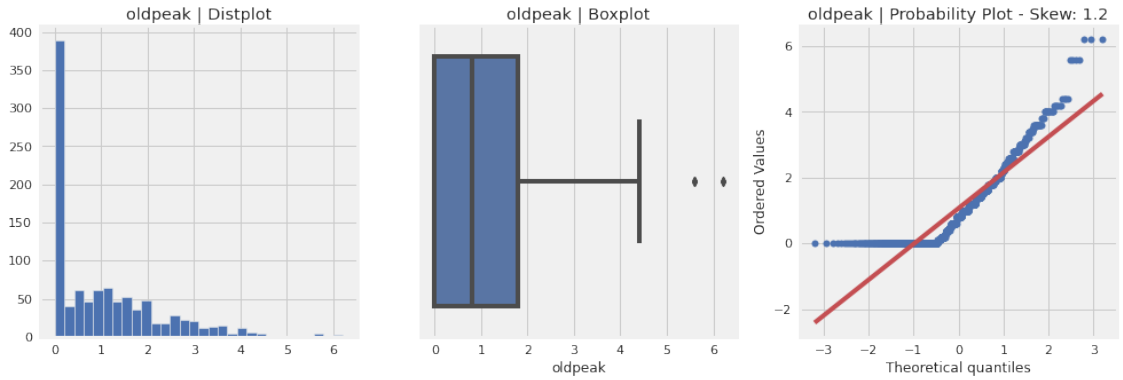
Autovizmain

December 10, 2022

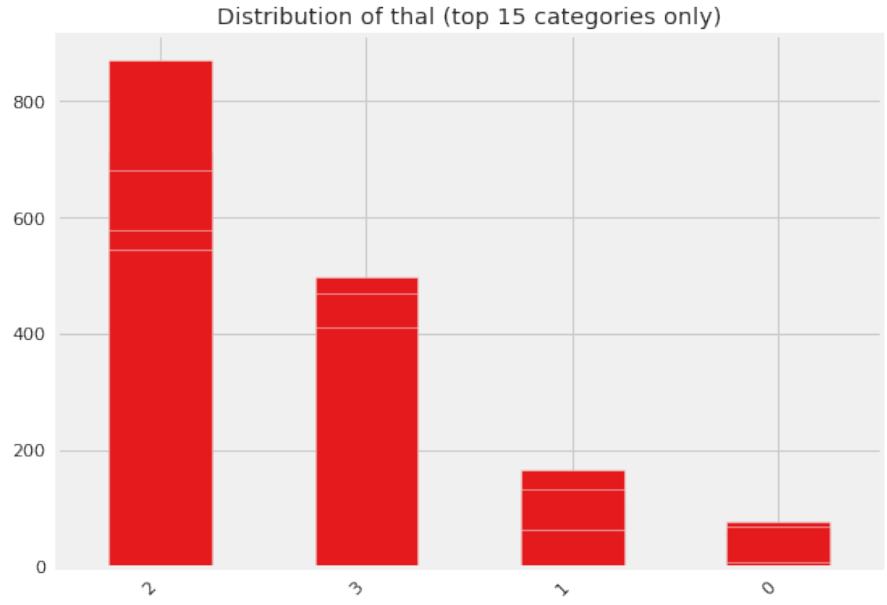
[6]:
```
import pandas as pd
import matplotlib as plt
import numpy as np
import seaborn as sns
from autoviz import data_cleaning_suggestions
from autoviz.AutoViz_Class import AutoViz_Class
AV= AutoViz_Class()
%matplotlib inline
import jovian
```

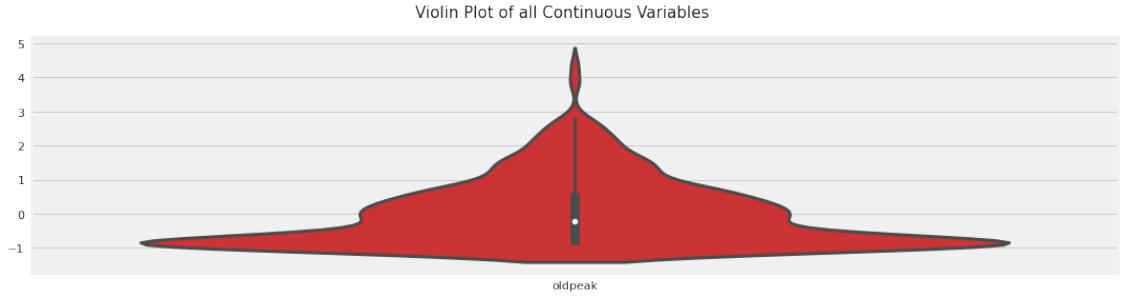
[2]:
```
df=AV.AutoViz('/home/miracle/Downloads/heart.csv')
#dfte = AV.AutoViz(filename, sep=',', depVar='', dfte=None, header=0,␣
 ↪verbose=1, lowess=False,
                ␣
 ↪#chart_format='svg',max_rows_analyzed=150000,max_cols_analyzed=30,␣
 ↪save_plot_dir=None)
```

```
Shape of your Data Set loaded: (1025, 14)
#############################################################################
#######
######################## 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…
Data cleaning improvement suggestions. Complete them before proceeding to ML
modeling.

<pandas.io.formats.style.Styler at 0x7f4ea5c36dc0>

    14 Predictors classified…
        No variables removed since no ID or low-information variables found in
data set
```

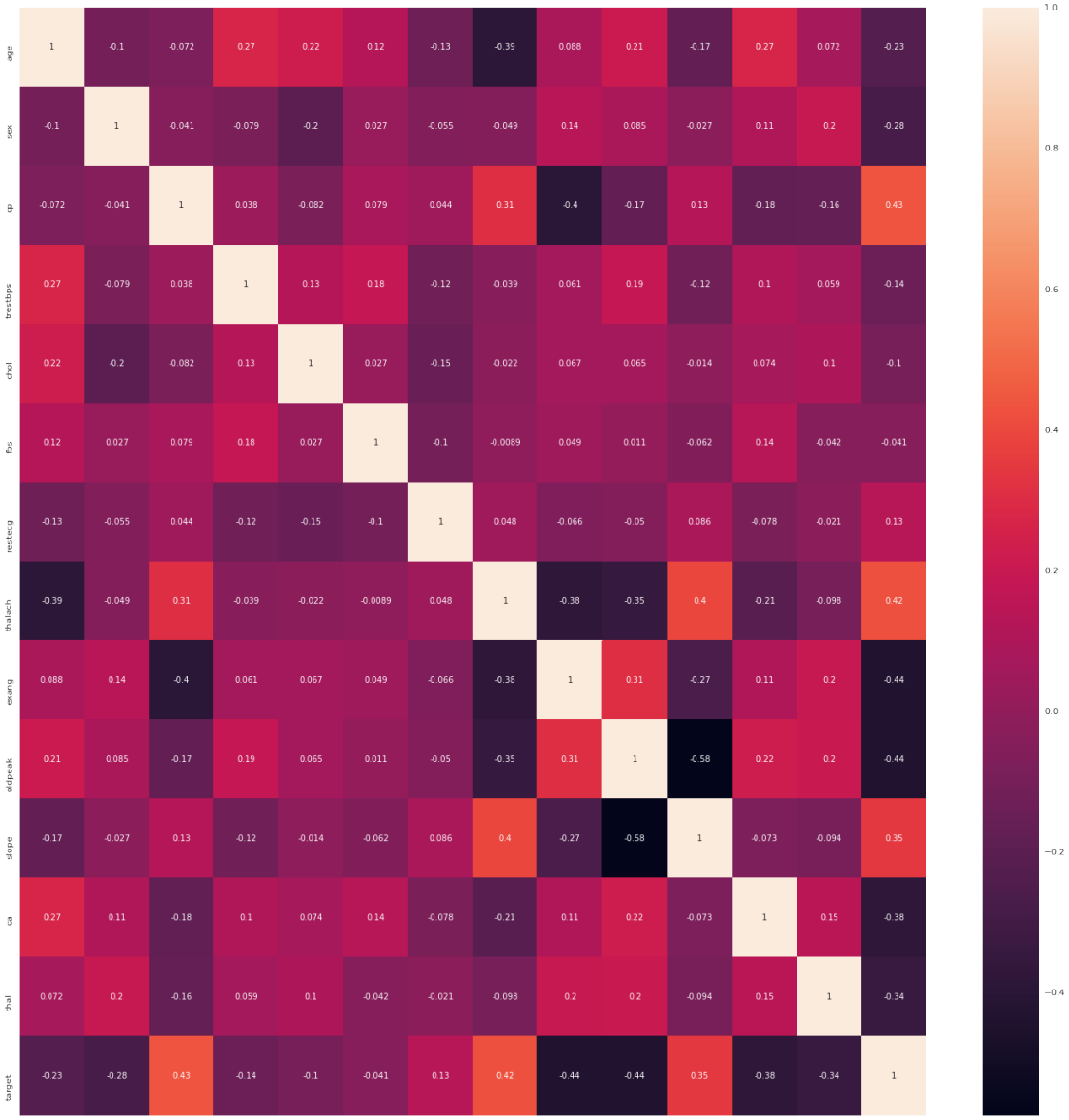


Histograms (KDE plots) of all Continuous Variables

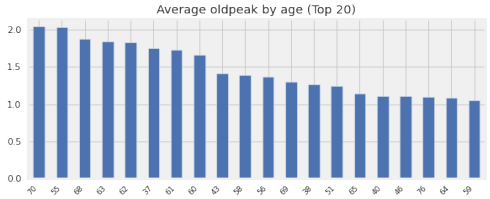


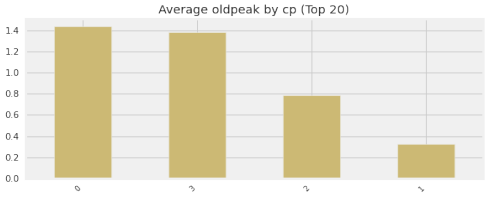


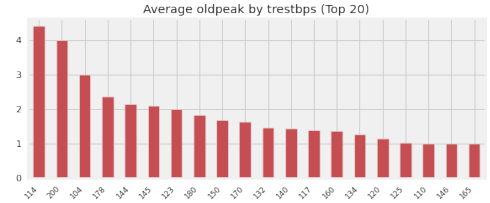
Heatmap of all Continuous Variables including target =

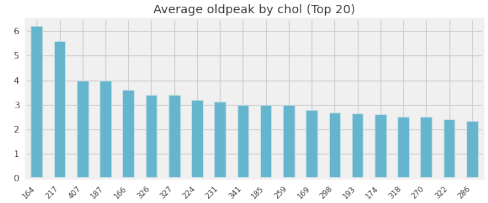


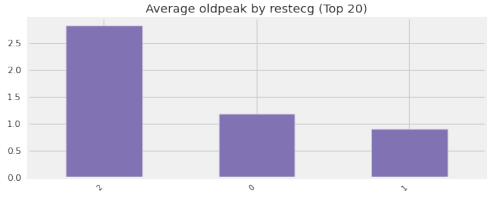
Bar plots for each Continuous by each Categorical variable

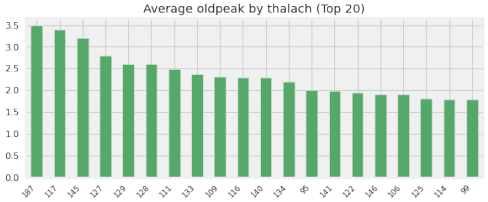


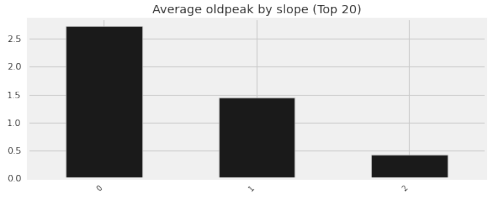


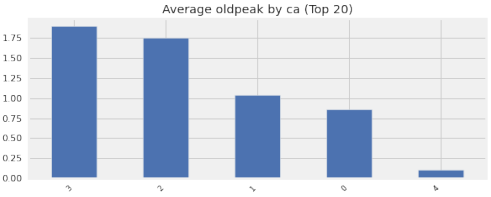


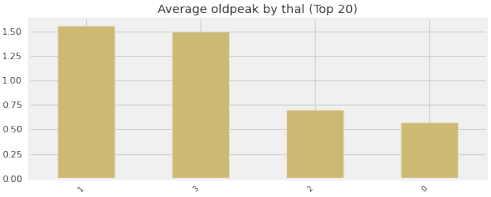


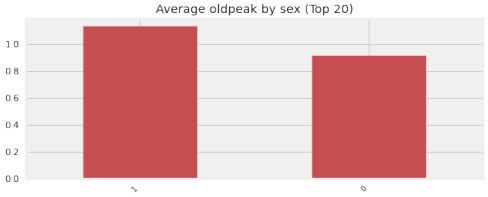


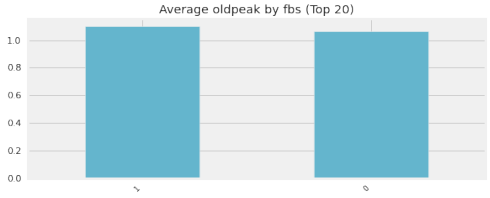


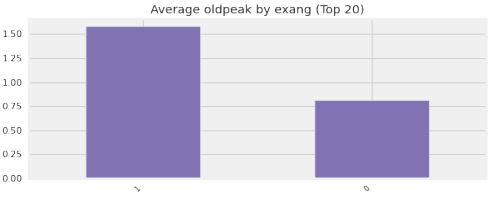


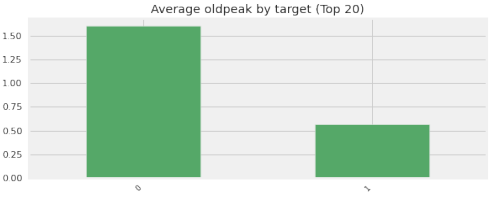












```
All Plots done
Time to run AutoViz = 25 seconds

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

[3]:
```
df= pd.read_csv('/home/miracle/Downloads/heart.csv')
df.head(2)
```

[3]:

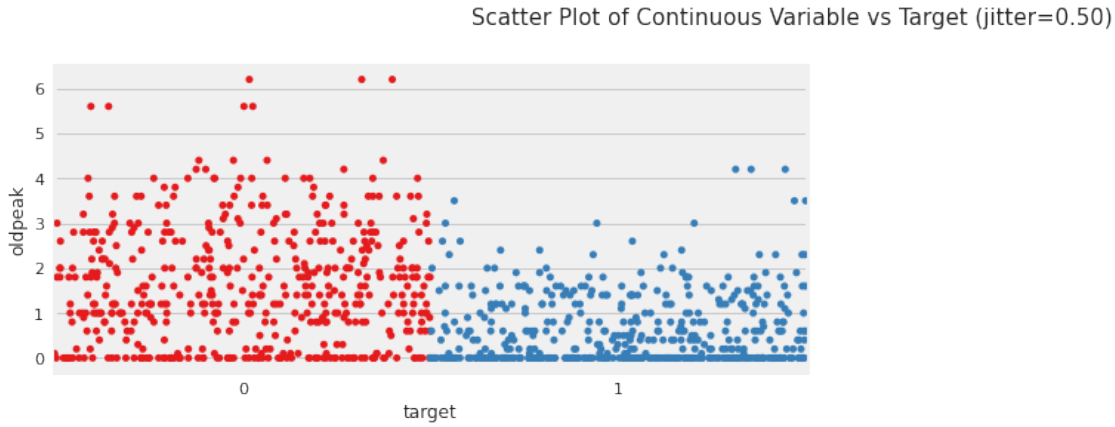
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0

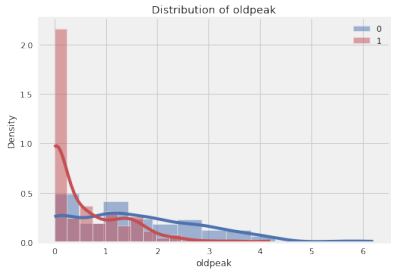
[4]:
```
data_cleaning_suggestions(df)
```

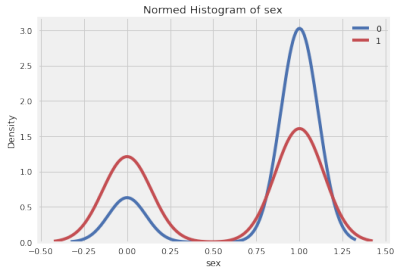
```
Data cleaning improvement suggestions. Complete them before proceeding to ML
modeling.

<pandas.io.formats.style.Styler at 0x7f4ea45bb550>
```

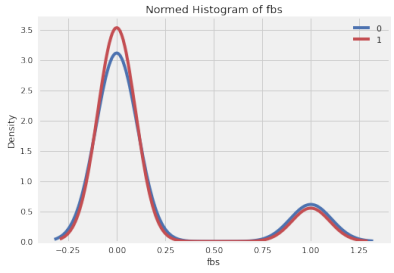
[5]:
```
df=AV.AutoViz('/home/miracle/Downloads/heart.csv',depVar='target')
```

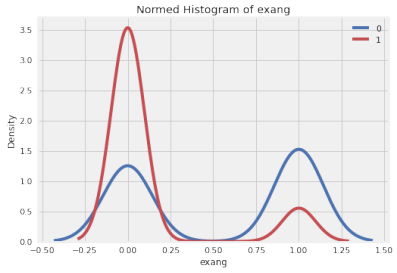
```
Shape of your Data Set loaded: (1025, 14)
#############################################################################
#######
######################## C L A S S I F Y I N G  V A R I A B L E S
####################
#############################################################################
#######
Classifying variables in data set…
Data cleaning improvement suggestions. Complete them before proceeding to ML
modeling.

<pandas.io.formats.style.Styler at 0x7f4ea4f8f850>

    13 Predictors classified…
        No variables removed since no ID or low-information variables found in
data set

################ Binary_Classification problem ######################
```

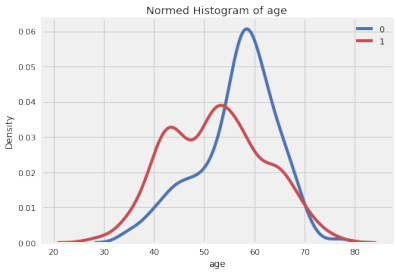

Scatter Plot of Continuous Variable vs Target (jitter=0.50)
oldpeak
target
0
1
2
3
4
5
6

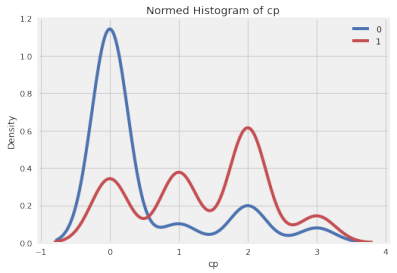

Distribution of oldpeak
Density
oldpeak
0
1

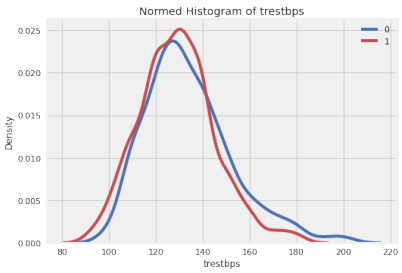

Normed Histogram of sex
Density
sex
0
1

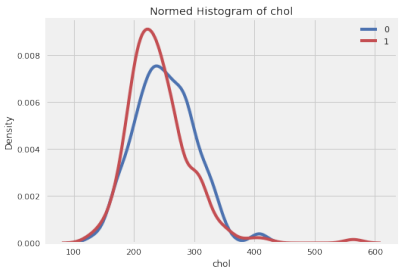

Normed Histogram of fbs
Density
fbs
0
1

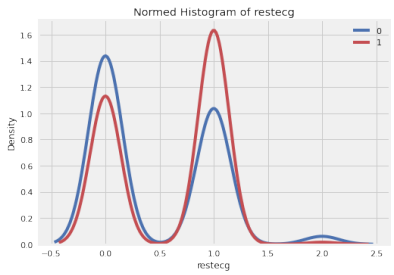

Normed Histogram of exang
Density
exang
0
1

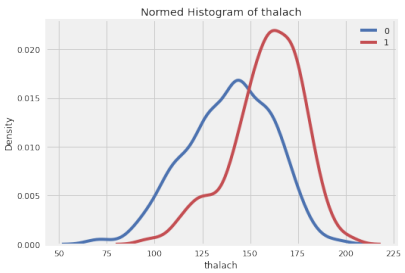

Normed Histogram of age
Density
age
0
1

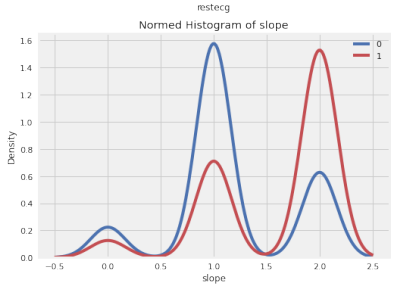

Normed Histogram of cp
Density
cp
0
1

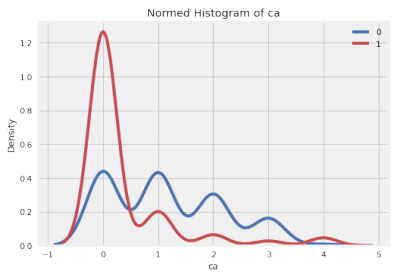

Normed Histogram of trestbps
Density
trestbps
0
1

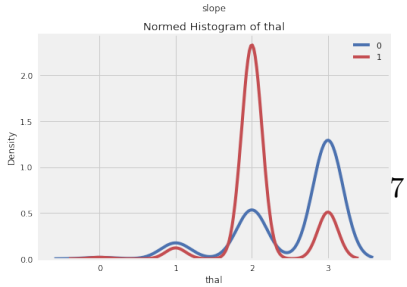

Normed Histogram of chol
Density
chol
0
1


Normed Histogram of restecg
Density
restecg
0
1

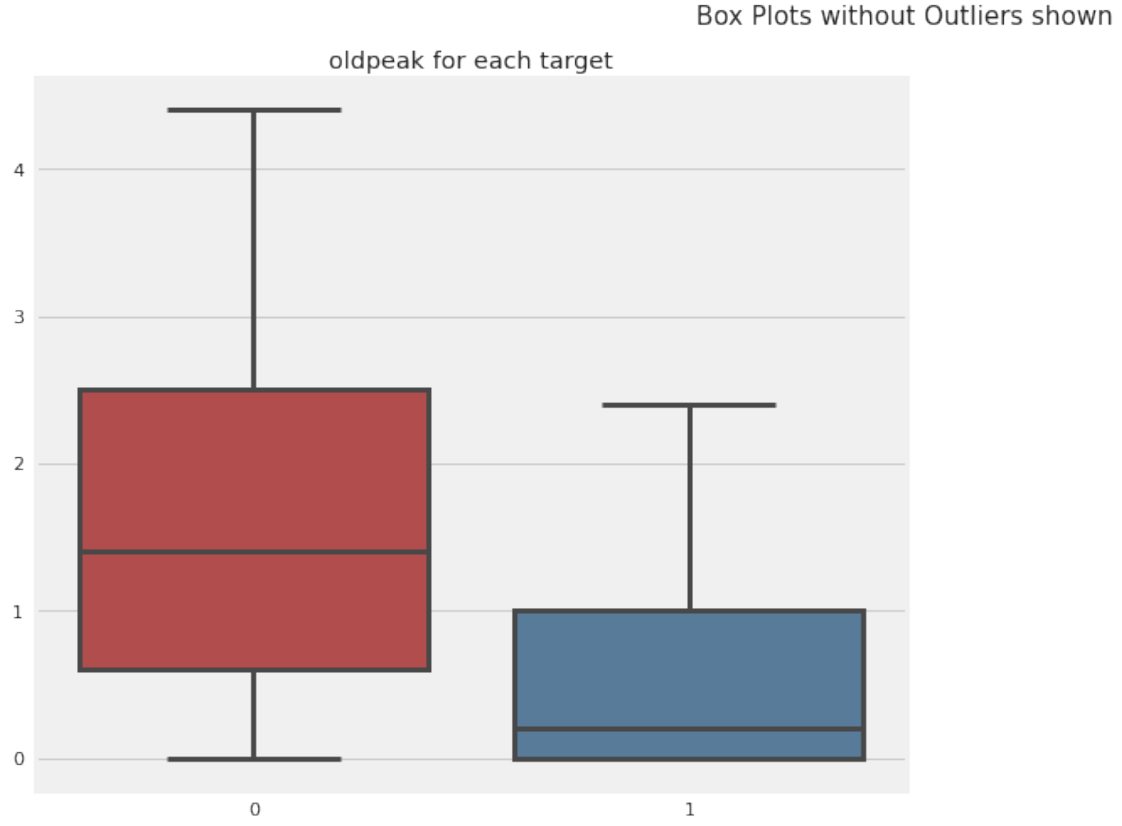

Normed Histogram of thalach
Density
thalach
0
1

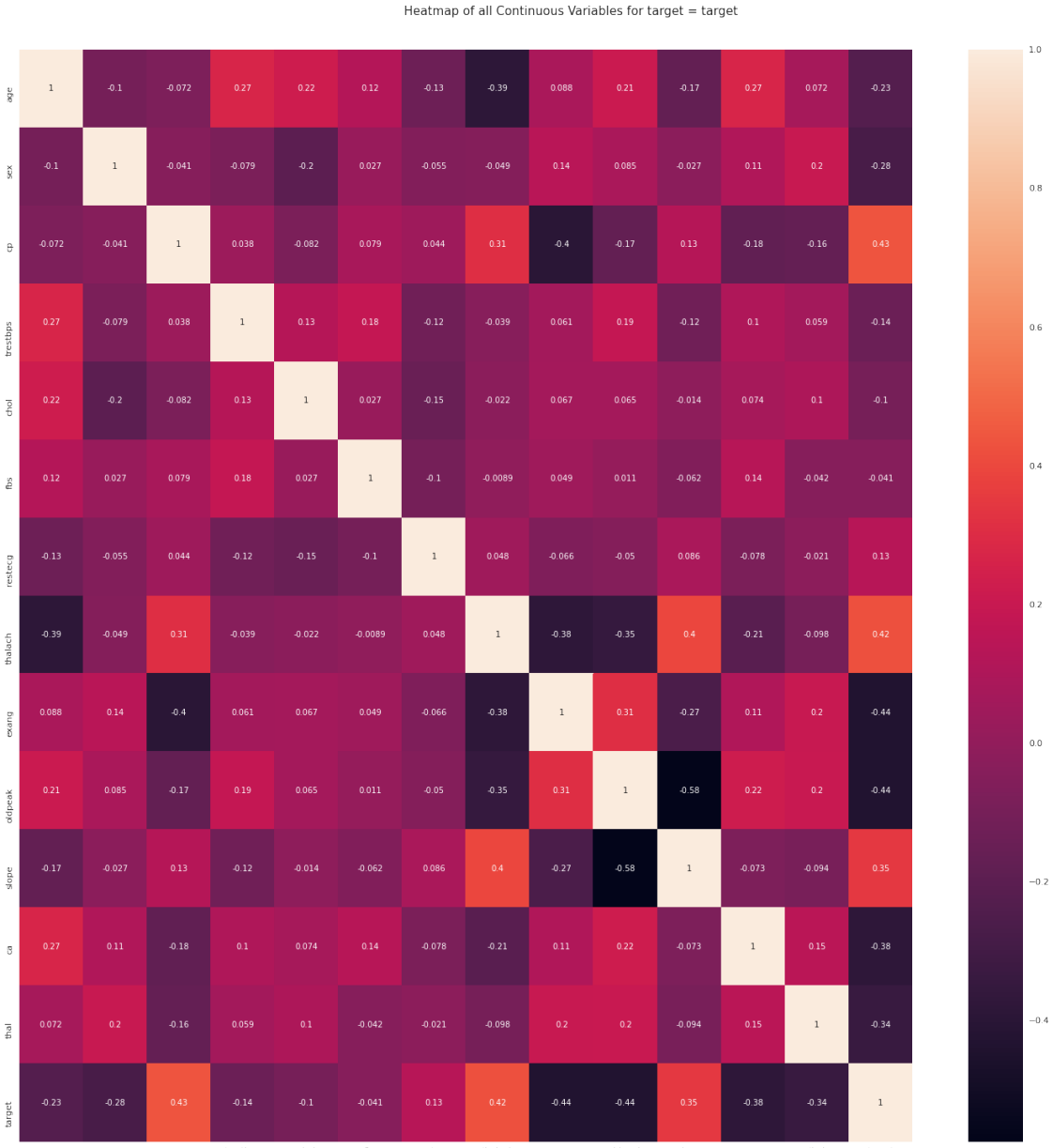

Normed Histogram of slope
Density
slope
0
1

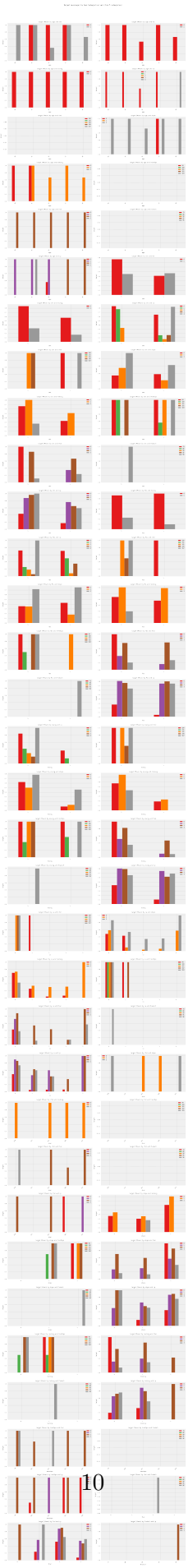

Normed Histogram of ca
Density
ca
0
1


Normed Histogram of thal
Density
thal
0
1

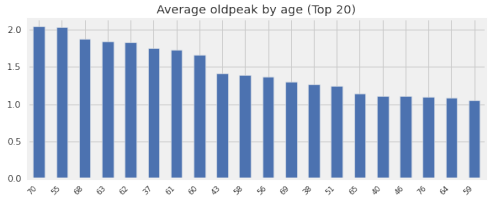

target : Distribution of Target Variable
Percentage Distribution of Target = target
0.51
0.49
0.5
0.4
0.3
0.2
0.1
0.0
0
1
Freq Distribution of Target Variable = target
526
499
500
400
300
200
100
0
0
1

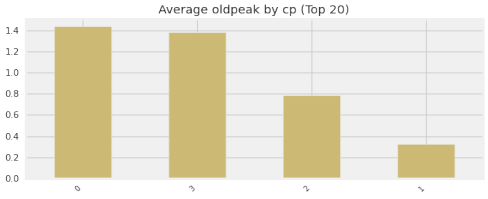

Box Plots without Outliers shown
oldpeak for each target
4
3
2
1
0
0
1

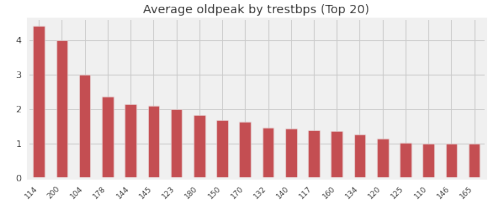

Heatmap of all Continuous Variables for target = target
age 1 -0.1 -0.072 0.27 0.22 0.12 -0.13 -0.39 0.088 0.21 -0.17 0.27 0.072 -0.23
sex -0.1 1 -0.041 -0.079 -0.2 0.027 -0.055 -0.049 0.14 0.085 -0.027 0.11 0.2 -0.28
cp -0.072 -0.041 1 0.038 -0.082 0.079 0.044 0.31 -0.4 -0.17 0.13 -0.18 -0.16 0.43
trestbps 0.27 -0.079 0.038 1 0.13 0.18 -0.12 -0.039 0.061 0.19 -0.12 0.1 0.059 -0.14
chol 0.22 -0.2 -0.082 0.13 1 0.027 -0.15 -0.022 0.067 0.065 -0.014 0.074 0.1 -0.1
fbs 0.12 0.027 0.079 0.18 0.027 1 -0.1 -0.0089 0.049 0.011 -0.062 0.14 -0.042 -0.041
restecg -0.13 -0.055 0.044 -0.12 -0.15 -0.1 1 0.048 -0.066 -0.05 0.086 -0.078 -0.021 0.13
thalach -0.39 -0.049 0.31 -0.039 -0.022 -0.0089 0.048 1 -0.38 -0.35 0.4 -0.21 -0.098 0.42
exang 0.088 0.14 -0.4 0.061 0.067 0.049 -0.066 -0.38 1 0.31 -0.27 0.11 0.2 -0.44
oldpeak 0.21 0.085 -0.17 0.19 0.065 0.011 -0.05 -0.35 0.31 1 -0.58 0.22 0.2 -0.44
slope -0.17 -0.027 0.13 -0.12 -0.014 -0.062 0.086 0.4 -0.27 -0.58 1 -0.073 -0.094 0.35
ca 0.27 0.11 -0.18 0.1 0.074 0.14 -0.078 -0.21 0.11 0.22 -0.073 1 0.15 -0.38
thal 0.072 0.2 -0.16 0.059 0.1 -0.042 -0.021 -0.098 0.2 0.2 -0.094 0.15 1 -0.34
target -0.23 -0.28 0.43 -0.14 -0.1 -0.041 0.13 0.42 -0.44 -0.44 0.35 -0.38 -0.34 1
age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4

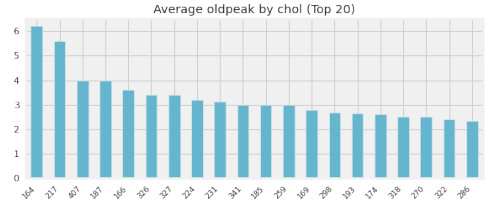


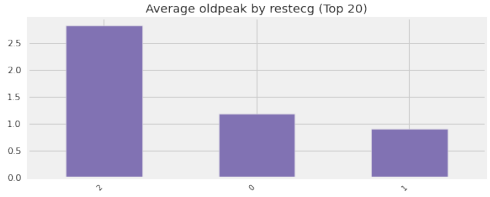
Bar plots for each Continuous by each Categorical variable

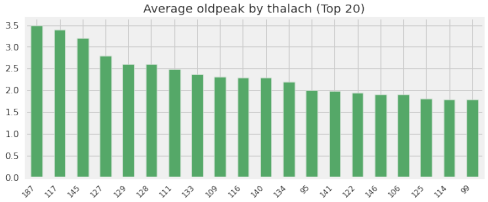


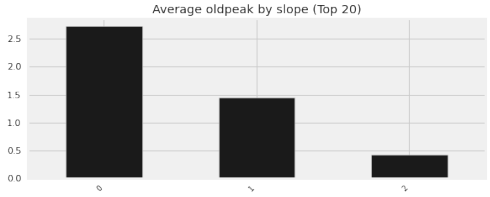


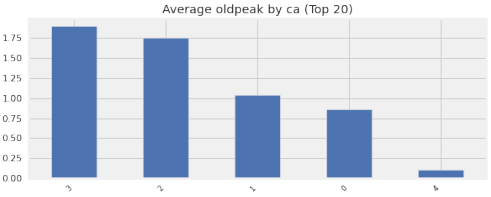


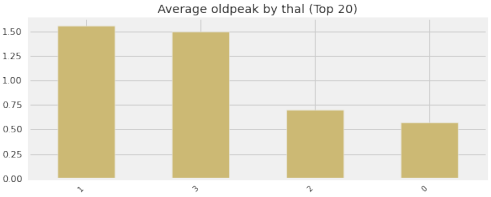


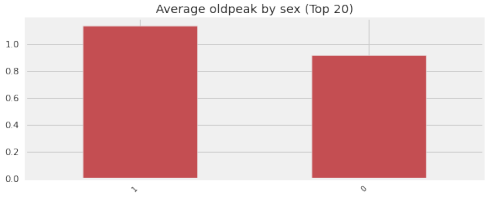


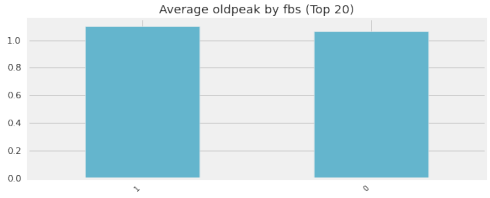


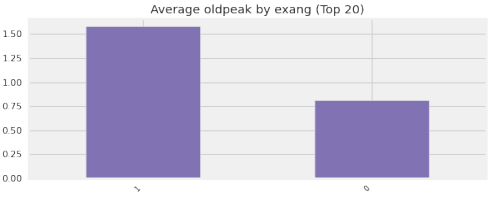


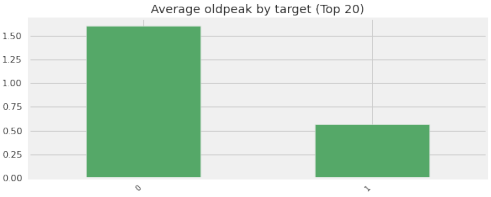












```
All Plots done
Time to run AutoViz = 101 seconds

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

[8]:
```
pd.show_versions(as_json=False)
```

```
INSTALLED VERSIONS
------------------
commit           : 8dab54d6573f7186ff0c3b6364d5e4dd635ff3e7
python           : 3.8.8.final.0
python-bits      : 64
OS               : Linux
OS-release       : 5.11.0-41-generic
Version          : #45~20.04.1-Ubuntu SMP Wed Nov 10 10:20:10 UTC 2021
machine          : x86_64
processor        : x86_64
byteorder        : little
LC_ALL           : None
LANG             : en_US.UTF-8
LOCALE           : en_US.UTF-8

pandas           : 1.5.2
numpy            : 1.23.5
pytz             : 2021.1
dateutil         : 2.8.1
setuptools       : 52.0.0.post20210125
pip              : 21.0.1
Cython           : 0.29.23
pytest           : 6.2.3
hypothesis       : None
sphinx           : 4.0.1
blosc            : None
feather          : None
xlsxwriter       : 1.3.8
lxml.etree       : 4.6.3
html5lib         : 1.1
pymysql          : None
psycopg2         : None
jinja2           : 2.11.3
IPython          : 7.22.0
pandas_datareader: None
bs4              : 4.9.3
bottleneck       : 1.3.2
```

```
brotli          :
fastparquet     : None
fsspec          : 0.9.0
gcsfs           : None
matplotlib      : 3.3.4
numba           : 0.53.1
numexpr         : 2.7.3
odfpy           : None
openpyxl        : 3.0.7
pandas_gbq      : None
pyarrow         : 9.0.0
pyreadstat      : None
pyxlsb          : None
s3fs            : None
scipy           : 1.6.2
snappy          : None
sqlalchemy      : 1.4.15
tables          : 3.6.1
tabulate        : None
xarray          : None
xlrd            : 2.0.1
xlwt            : 1.3.0
zstandard       : None
tzdata          : 2022.5
```

[7]:
```
project_name='autoviz'
jovian.commit(project= project_name)
```

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

```
[jovian] Creating a new project "oguejiofor-mbah/autoviz"
[jovian] Committed successfully! https://jovian.ai/oguejiofor-mbah/autoviz
```

[7]:
```
'https://jovian.ai/oguejiofor-mbah/autoviz'
```

[ ]:
```

```