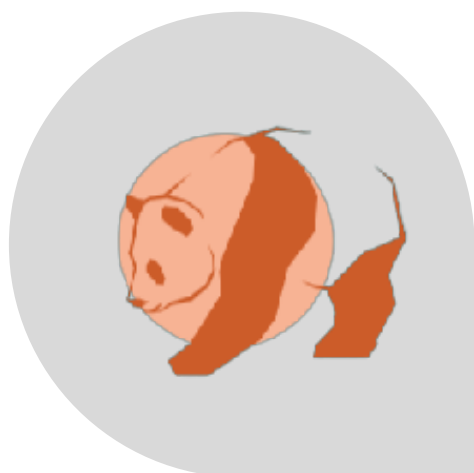


Automated EDA TOOLS

1



Pandas Profiling

It is an open-source library written in python and it generated interactive HTML reports and describes various aspects of the dataset. Key functionalities include handling missing values, statistics of dataset like mean, mode, median, skewness, standard deviation etc, charts like histograms and correlations as well.

2



Dtale

It is a library that allows us to visualize pandas' data frame easily. It has many features which are very handy for exploratory data analysis. It is made using the flask backend and reacts frontend. It supports interactive plots, 3d plots, heat maps, the correlation between features, builds custom columns, and many more.

3



Sweetviz

It is an open-source python library that used to get visualizations which is useful in exploratory data analysis with just a few lines of codes. The library can be used to visualize the variables and comparing the dataset.

4



Autoviz

It stands for Automatically Visualize. Visualization is possible with any size of the dataset with a few lines of code.

Installation

```
pip install dtale
pip install pandas-profiling
pip install sweetviz
pip install autoviz
```

Pandas Profiling

```
import dtale
import pandas as pd

df = pd.read_csv('data.csv')
d = dtale.show(df)

d.open_browser()
```

Sweetviz

```
import sweetviz
import pandas as pd

df = pd.read_csv('sample.csv')
my_report = sweetviz.analyze([df, 'Train'], target_feat='SalePrice')
my_report.show_html('FinalReport.html')
```

Autoviz

```
from autoviz.AutoViz_Class import AutoViz_Class

AV = AutoViz_Class()
df = AV.AutoViz('sample.csv')
```