**AutoViz**

Visualization is a technique that is used to visualize the data using different graphs and plots. In data science, we generally use data visualization techniques to understand the dataset and find the relation between the Dependent data and independent data. Visualization can also help in finding the pattern in the dataset which is used for further analysis.

There are many techniques and libraries in python which are used for Data Visualization like Matplotlib, Seaborn, Plotly, etc. But while using all these libraries we need to define the type of graph we want to visualize and the arguments which we need to visualize.

**AutoViz** which can automate the whole process of Data Visualization in just a single line of code.

AutoViz performs automatic visualization of any dataset with just one line of code. AutoViz can find the most important features and plot impactful visualizations only using those automatically selected features. Also, AutoViz is incredibly fast so it creates visualization within seconds.

If the number of observations is large, AutoViz will take a random sample; likewise, if the number of variables is large (which you can decide) AutoViz can find the most important features and plot impactful visualizations only using those automatically selected features. The user can set the sample number of rows and the maximum number of features to visualize by simply passing a parameter to AutoViz. AutoViz is capable of adapting to any number of different data contexts such as regression, classification, or even time-series data. It also delivers output incredibly quickly.

**AutoViz can be implemented in 4 simple steps:**

1. Install using "pip install autoviz"
2. Import with "from autoviz.AutoViz\_Class import AutoViz\_Class"
3. Instantiate a class "AV = AutoViz\_Class()"
4. Run an experiment in the following line with our data set:

**Notes:**

* **AutoViz** will visualize any sized file using a statistically valid sample.
* **COMMA** is assumed as default separator in file. But you can change it.
* Assumes first row as header in file but you can change it.
* **verbose** option
* if 0, display minimal information but displays charts on your notebook
* if 1, print extra information on the notebook and also display charts
* if 2, will not display any charts, it will simply save them in your local machine under AutoViz\_Plots directory

**Important Arguments**

**filename** - Make sure that you give filename as empty string ("") if there is no filename associated with this data and you want to use a dataframe, then use dfte to give the name of the dataframe. Otherwise, fill in the file name and leave dfte as empty string. Only one of these two is needed to load the data set.

**sep** - this is the separator in the file. It can be comma, semi-colon or tab or any value that you see in your file that separates each column.

**depVar** - target variable in your dataset. You can leave it as empty string if you don't have a target variable in your data.

**dfte** - this is the input data frame in case you want to load a pandas data frame to plot charts. In that case, leave filename as an empty string.

**header** - the row number of the header row in your file. If it is the first row, then this must be zero.

**verbose** - it has 3 acceptable values: 0, 1 or 2. With zero, you get all charts but limited info. With 1 you get all charts and more info. With 2, you will not see any charts but they will be quietly generated and save in your local current directory under the AutoViz\_Plots directory which will be created. Make sure you delete this folder periodically, otherwise, you will have lots of charts saved here if you used verbose=2 option a lot.

**lowess** - this option is very nice for small datasets where you can see regression lines for each pair of continuous variable against the target variable. Don't use this for large data sets (that is over 100,000 rows)

**chart\_format -** this can be SVG, PNG or JPG. You will get charts generated and saved in this format if you used verbose=2 option. Very useful for generating charts and using them later.

**max\_rows\_analyzed -** limits the max number of rows that is used to display charts. If you have a very large data set with millions of rows, then use this option to limit the amount of time it takes to generate charts. We will take a statistically valid sample.

**max\_cols\_analyzed** - limits the number of continuous vars that can be analyzed

Reference: <https://towardsdatascience.com/autoviz-a-new-tool-for-automated-visualization-ec9c1744a6ad>

<https://github.com/AutoViML/AutoViz>

<https://github.com/danroth-nyt/autoviz_test>