**Automobile Analysis Project**

**Introduction:**

This analysis delves into comprehensive analysis on finding origin of car from three different countries Europe, America, Japan from features mpg, cylinders, displacement, horsepower, weight, acceleration, model year.

**Hardware and Software Requirements:**

Hardware Requirements:

1. CPU: A contemporary processor that can handle the computational demands of developing machine learning models. For speedier processing, multi-core CPUs are preferred.
2. RAM: The dataset and model training procedure will fit in enough RAM. Depending on the size and complexity of the dataset, a different quantity may be needed, although 8GB or more is typically advised.
3. Storage: Enough room to keep the dataset, the programme, and any intermediate files produced throughout the project.

Software requirements:

1. Python is a popular programming language for machine learning tasks. Install the most recent version of Python on your machine (for example, Python 3.9 or higher).
2. Development environment integrated (IDE): Select a Python IDE based on your personal tastes. Popular choices that offer a convenient environment for code development, execution, and data analysis include PyCharm, Jupyter Notebook, or Anaconda.
3. Libraries for machine learning: Install the necessary Python machine learning libraries, including:

**NumPy:** For manipulating arrays and doing numerical calculations.

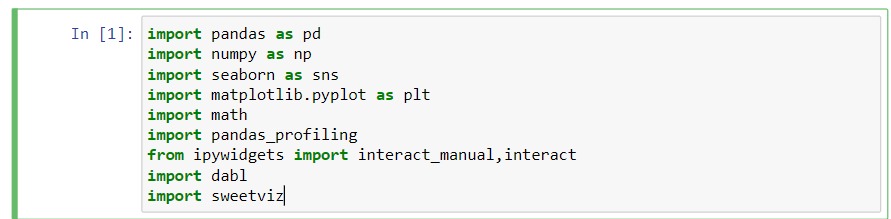
**Pandas:** Used for data analysis and manipulation.

A thorough machine learning library with a variety of tools and algorithms through **Scikit-learn**.

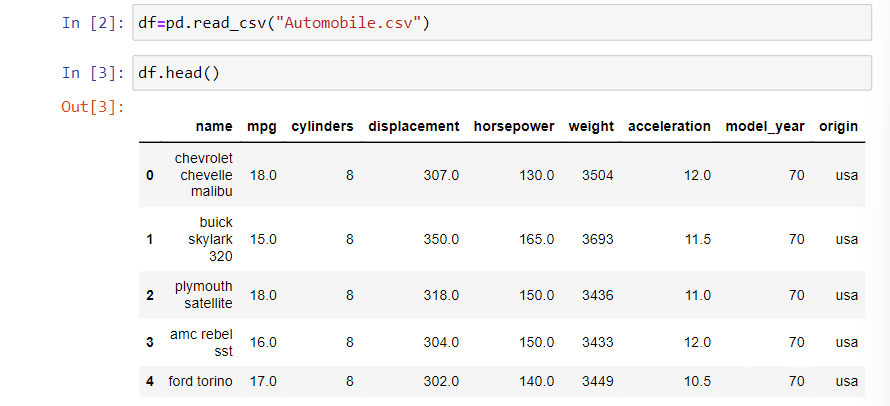
To visualise data and produce charts for data analysis, use the data visualisation libraries **Matplotlib or Seaborn**.

1. Jupyter Notebooks (Optional): If you prefer an interactive environment for developing and documenting your code, install Jupyter Notebooks or JupyterLab.

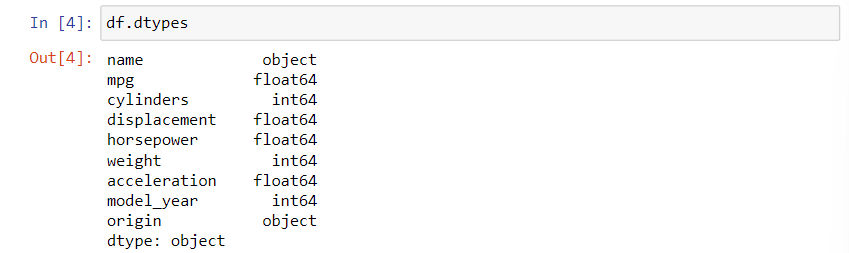
**Experimental Investigation:**

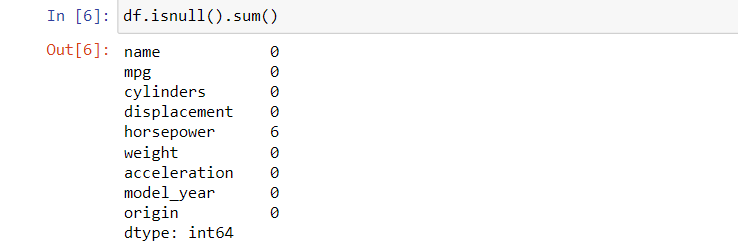
****

Importing required packages for the data analysis.

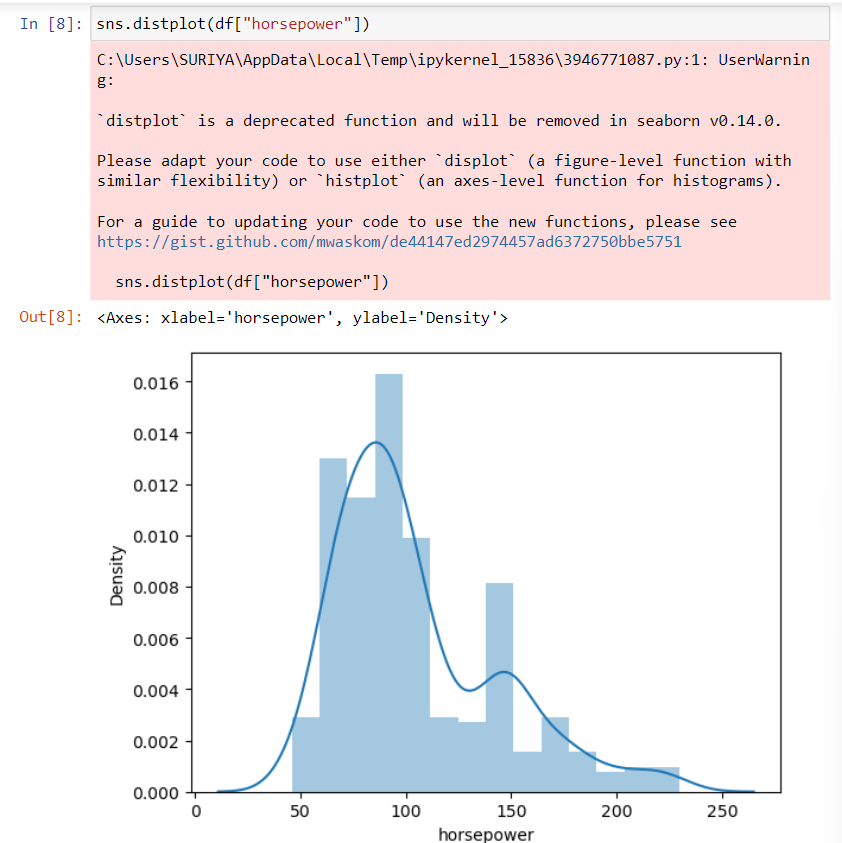


Reading the csv file using read\_csv() function.

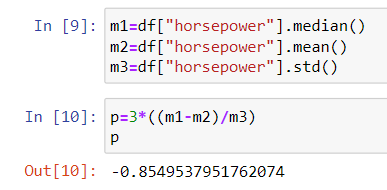
The dataset has data types object, float and int.



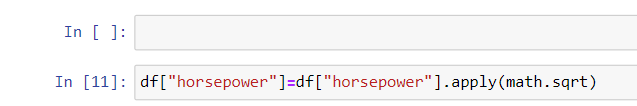
The column horsepower has six null values.



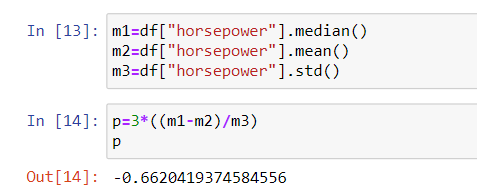
Finding how the column horsepower is distributed to fill the null values based on median or mean values.



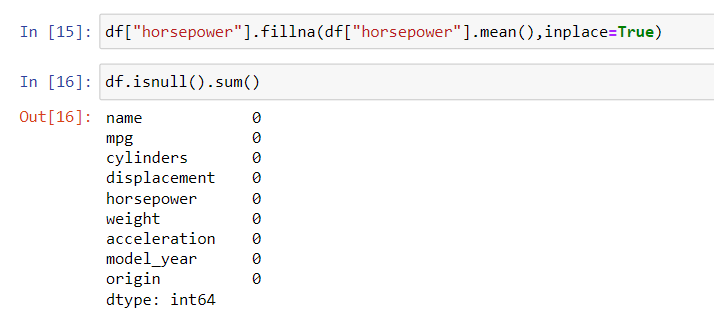
Measuring the skewness and found the column is negatively skewed.



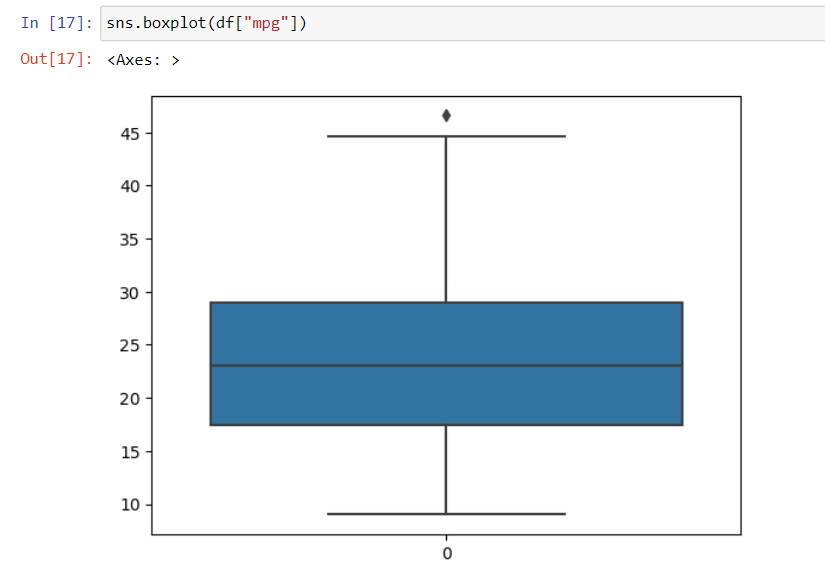
Reducing the skewness by applying square root function.



The skew is reduced now and now we can apply median or mean for the null values present in the column so we go by mean.

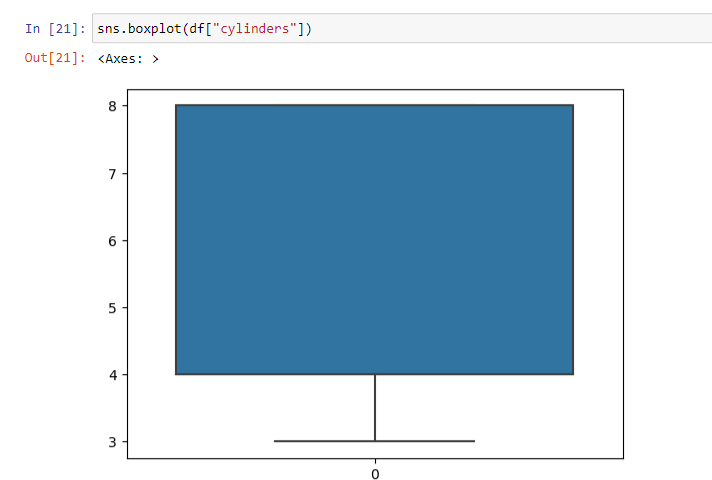


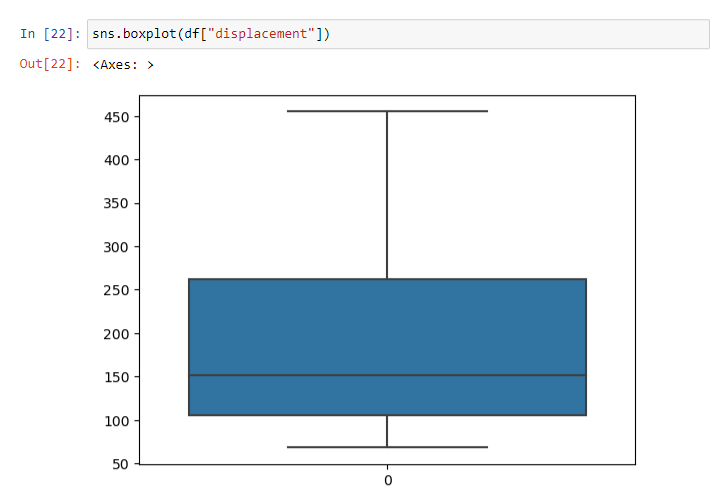
Outlier analysis:

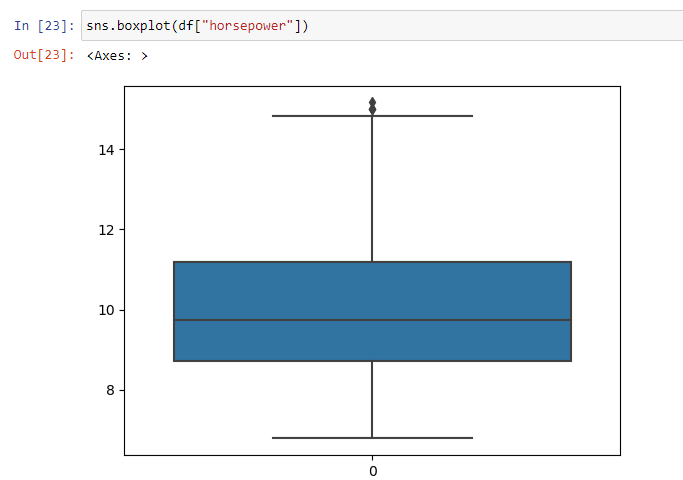


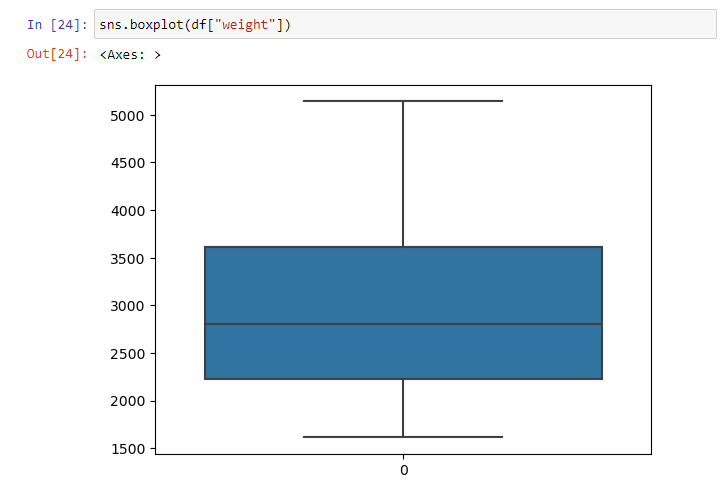
The mpg column has outlier and im removing it.



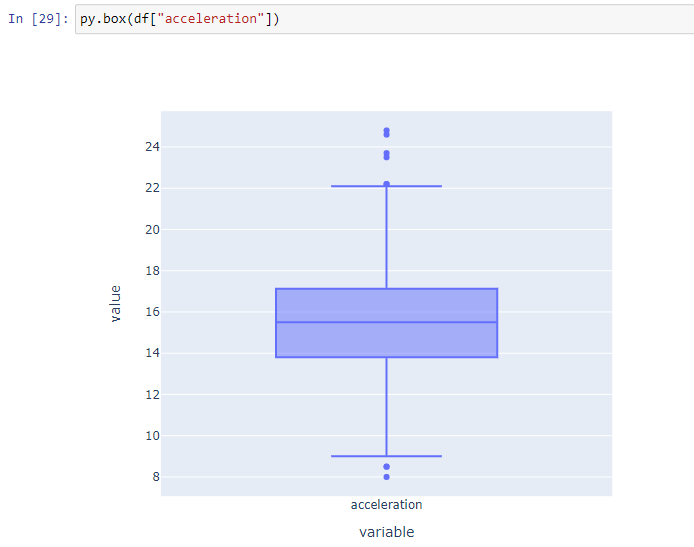




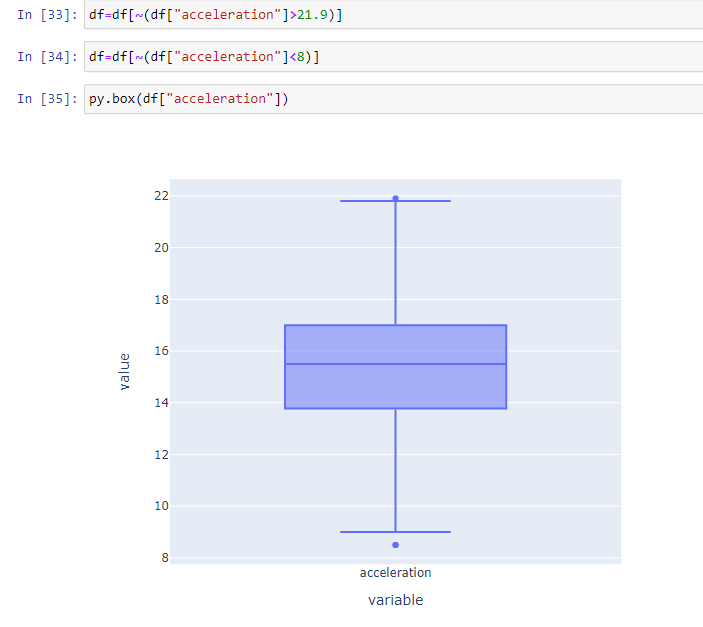


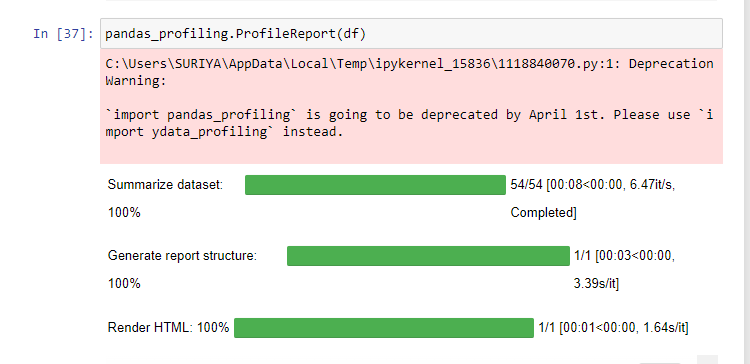


Columns weight, horsepower, displacement, cylinders has no outlier.

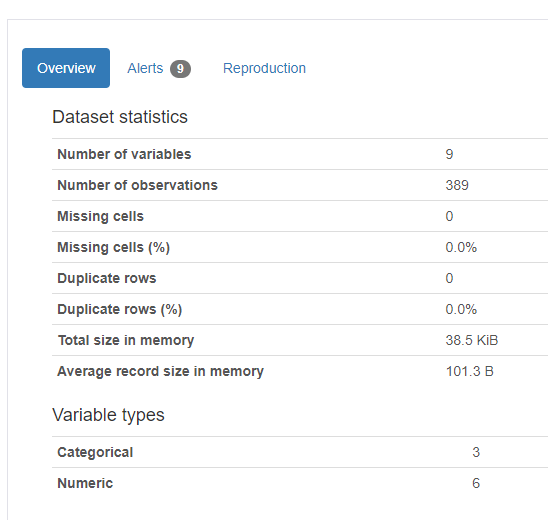


Removing outlier on acceleration column:

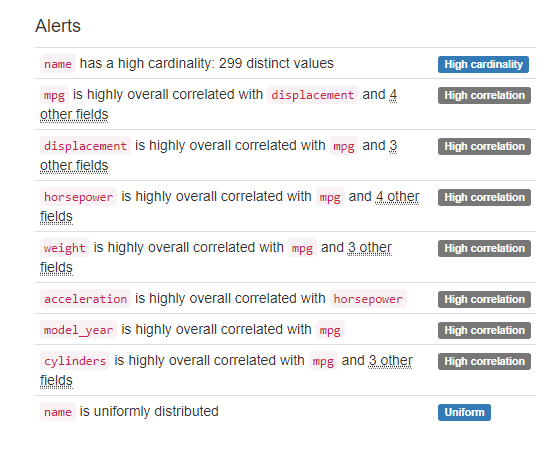




Pandas Profiling the dataset:



The overview we have 9 columns 389 rows no missing cell no duplicate columns with 3 categorical and 6 numeric columns.



All columns are highly correlated.

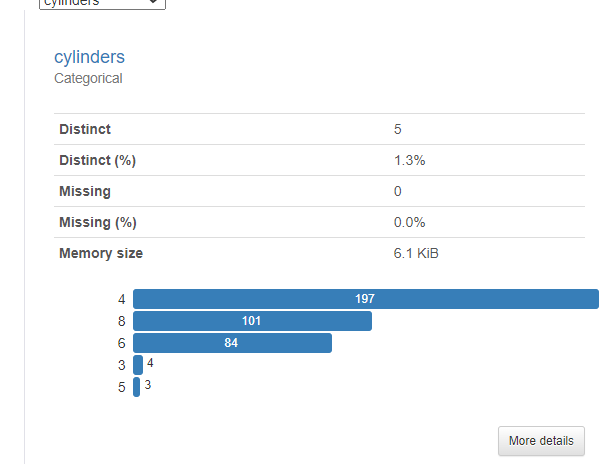
MPG column analysis:

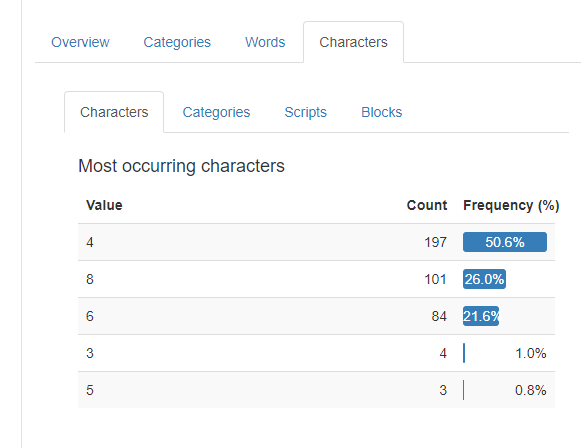


MPG column analysis:

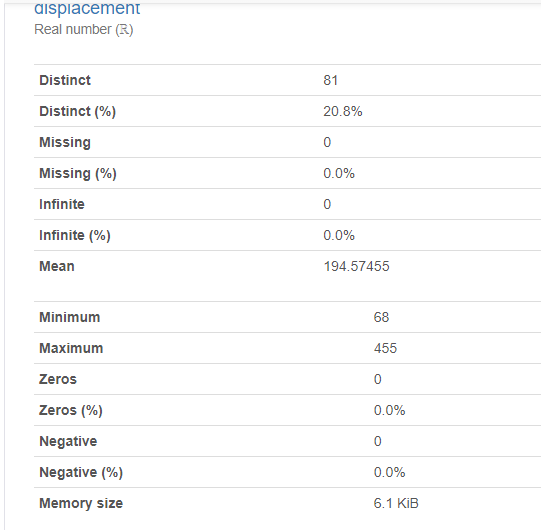


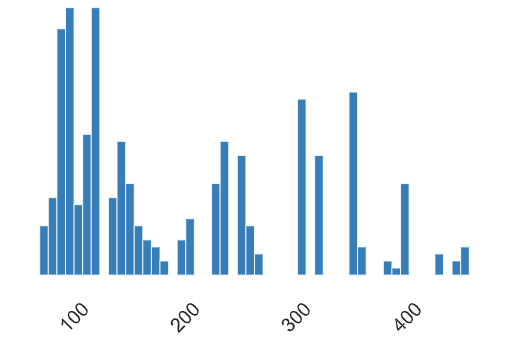
Cylinders column analysis:

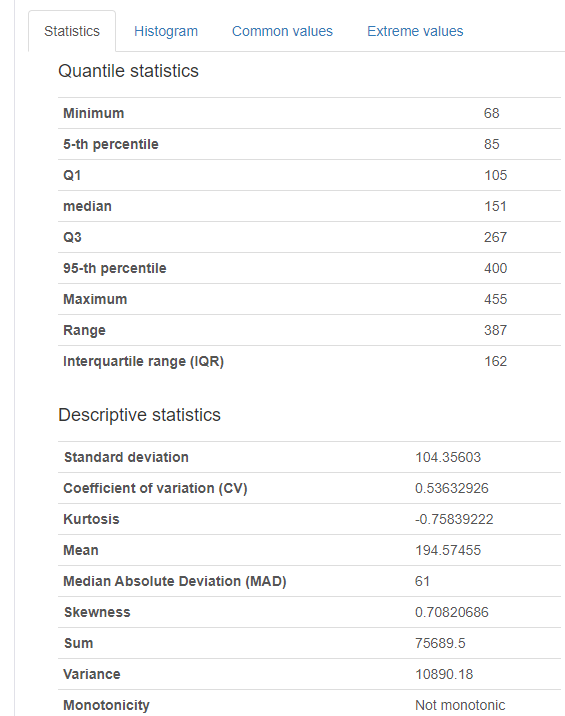


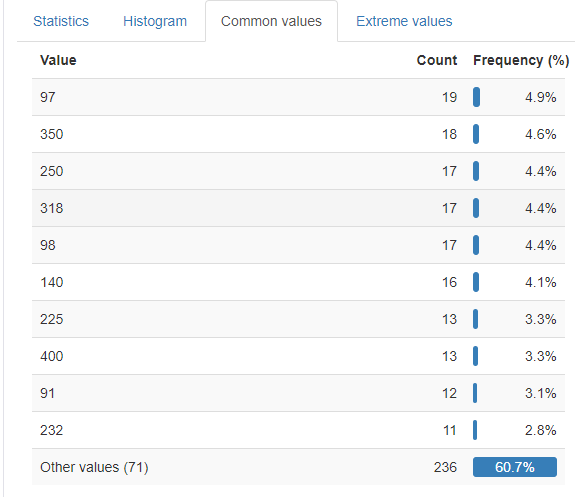


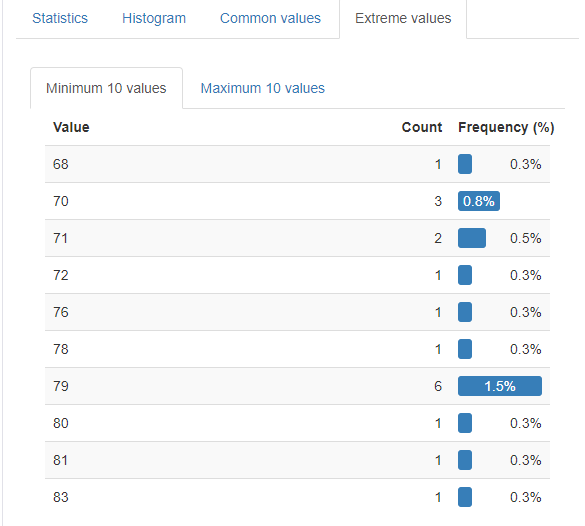
Displacement column analysis:



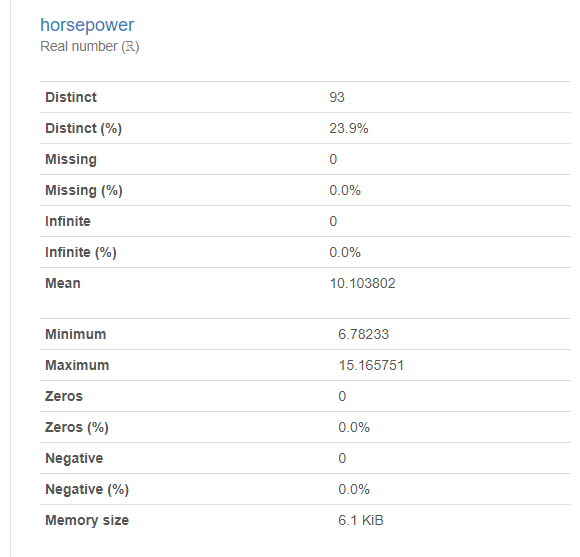


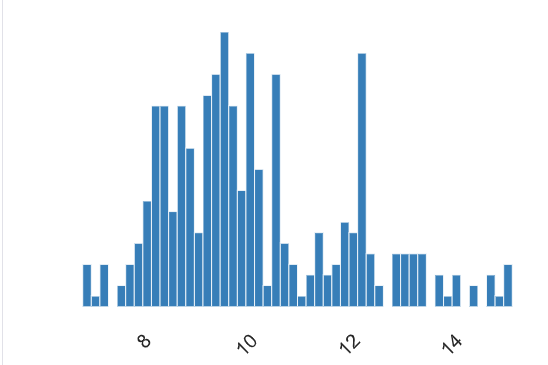


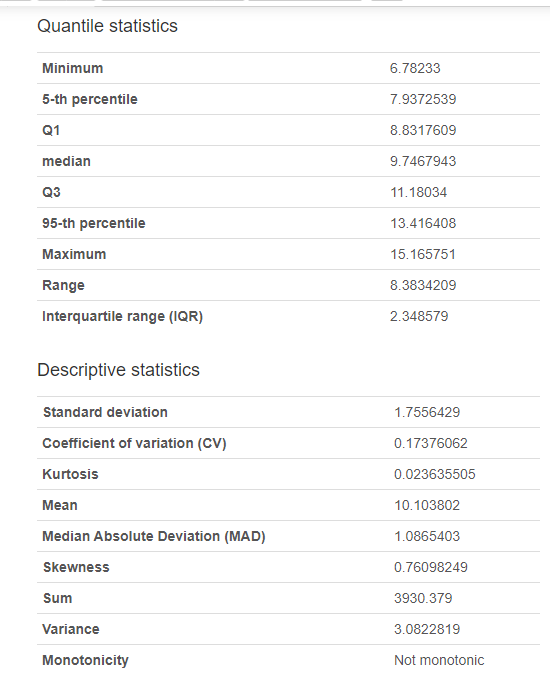


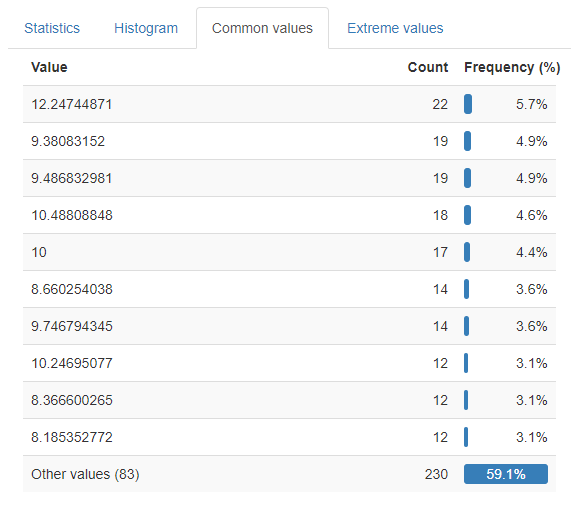


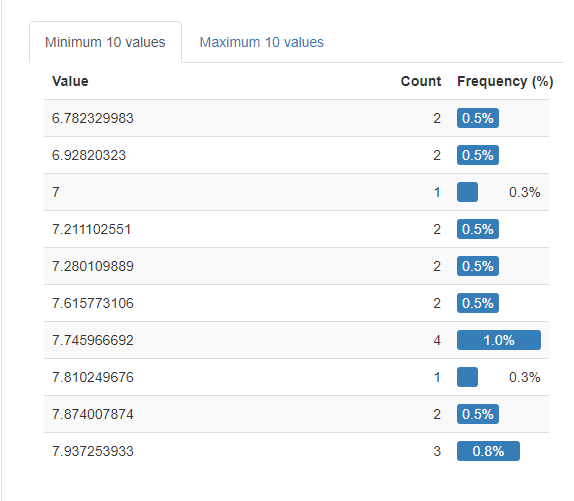
Horsepower column analysis:



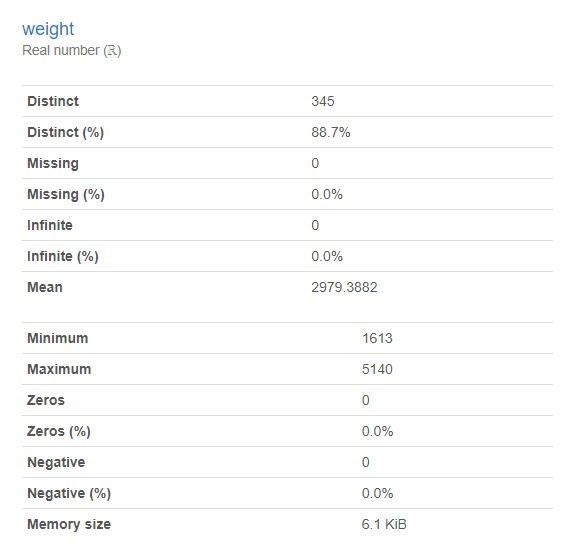


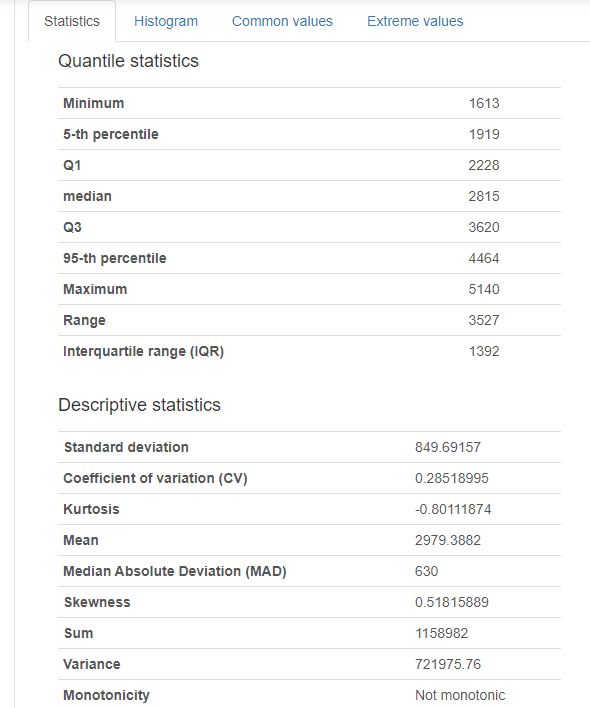


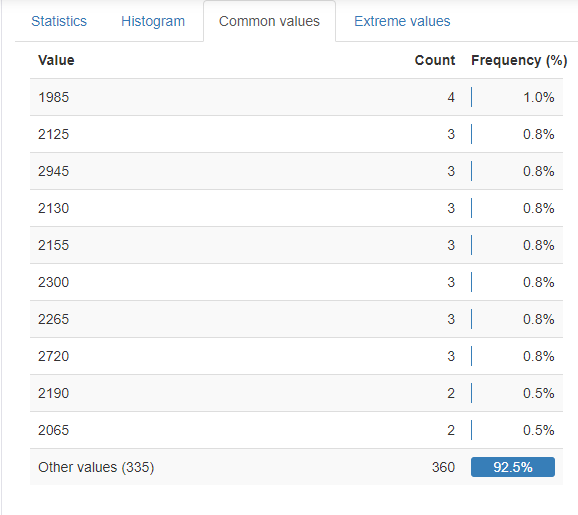


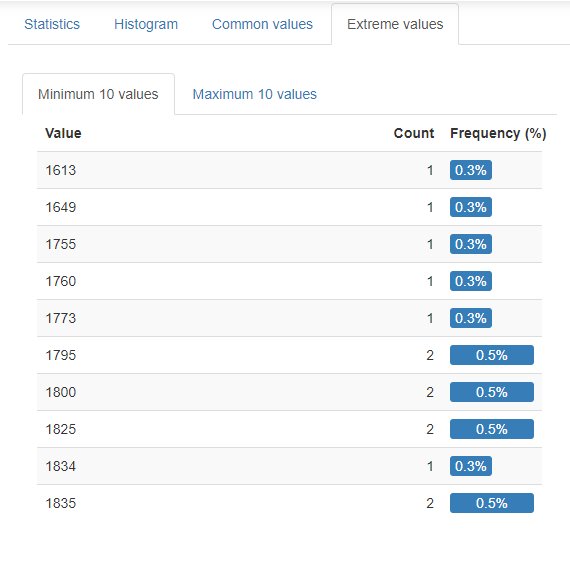


Weight column analysis:

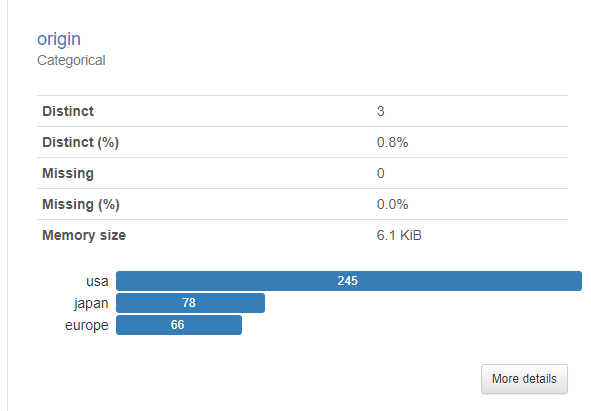


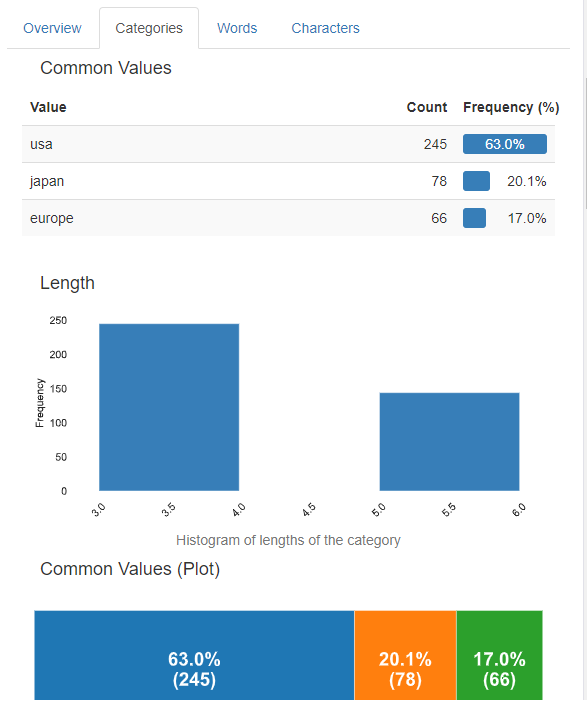


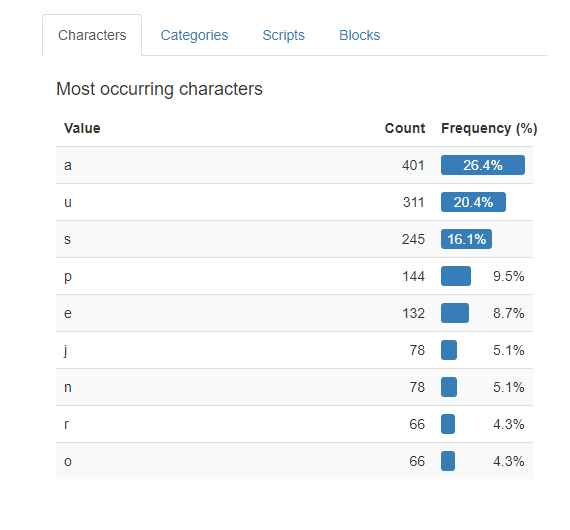


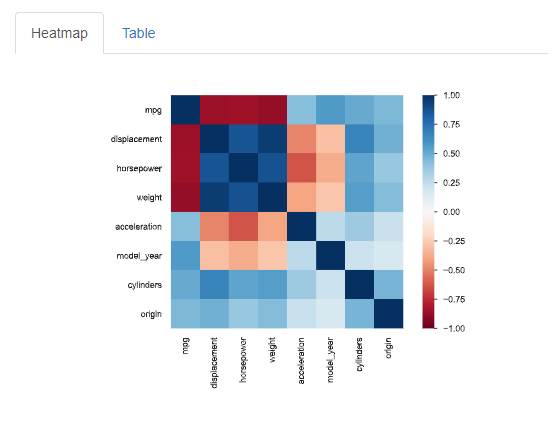


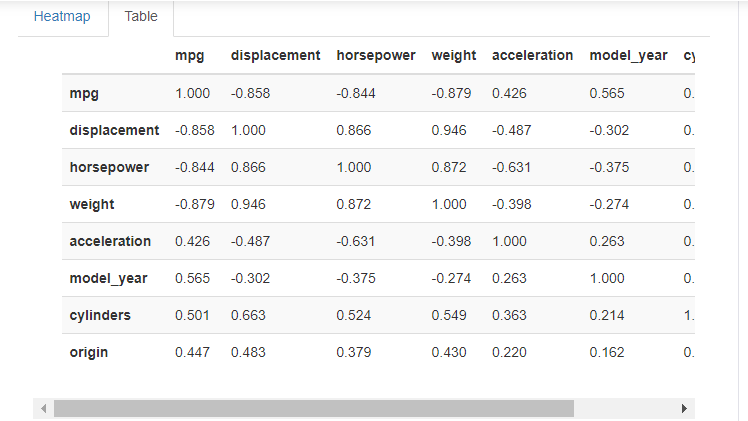
Origin column analysis:





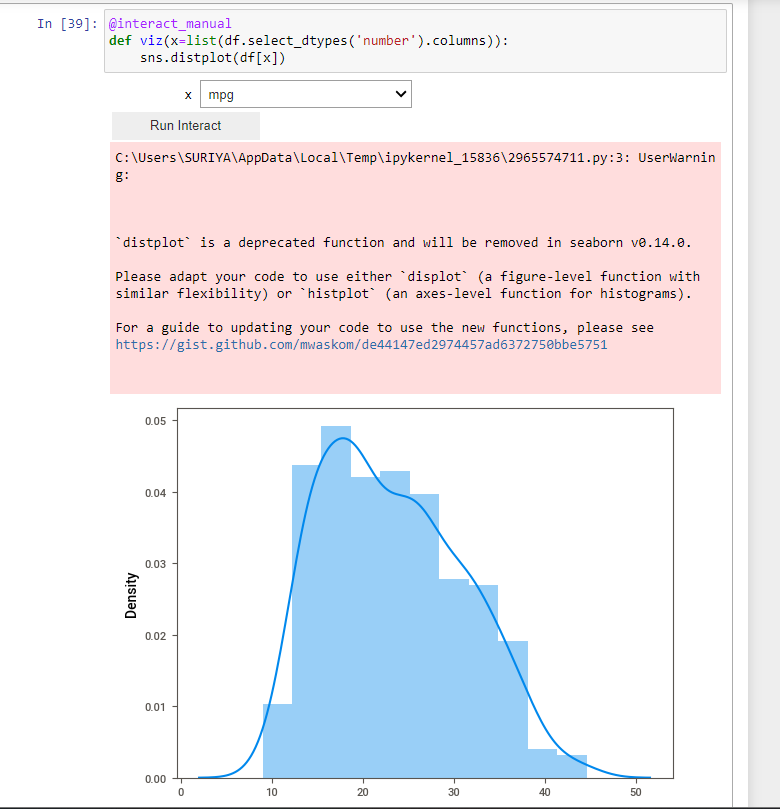


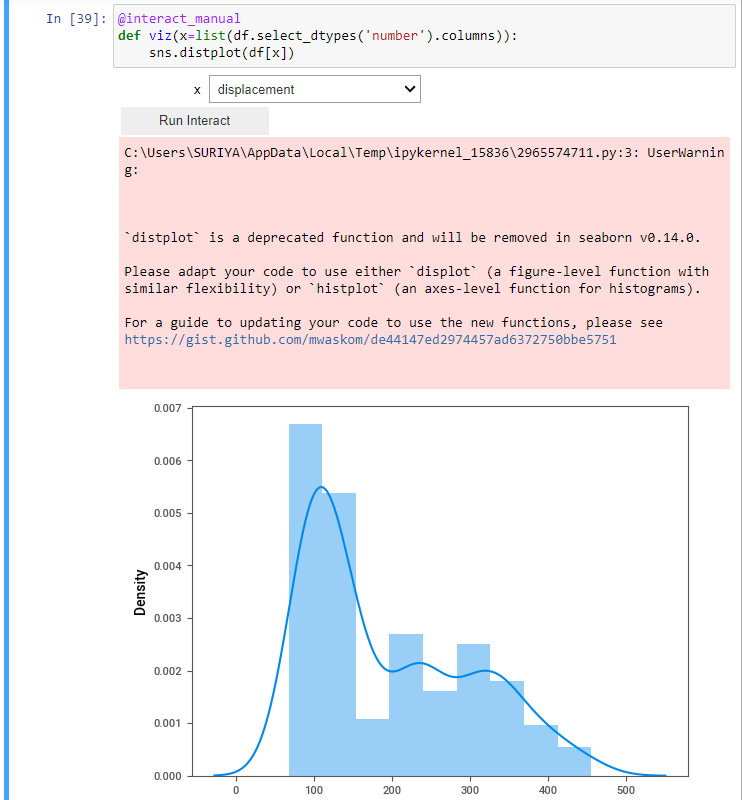


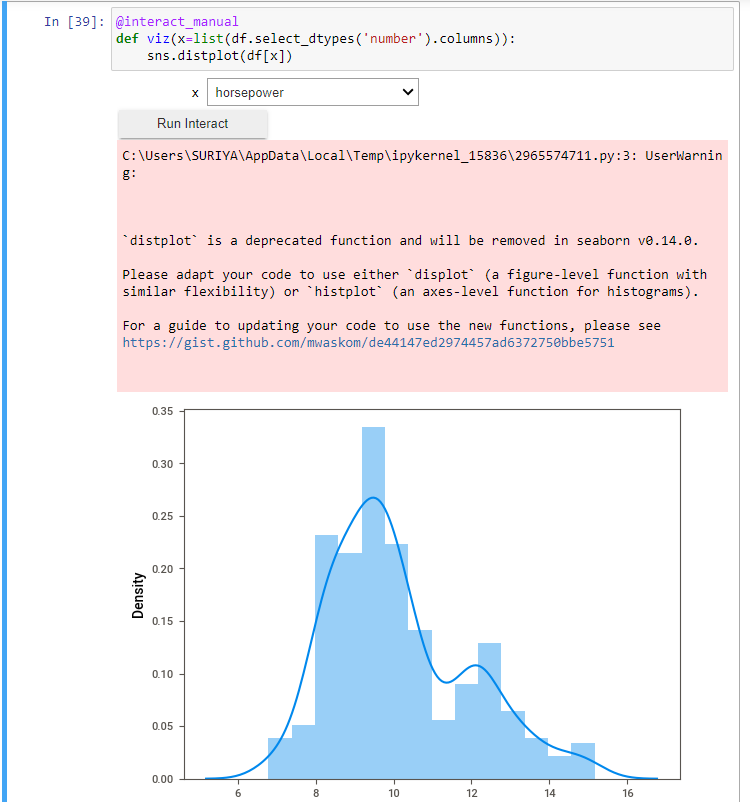


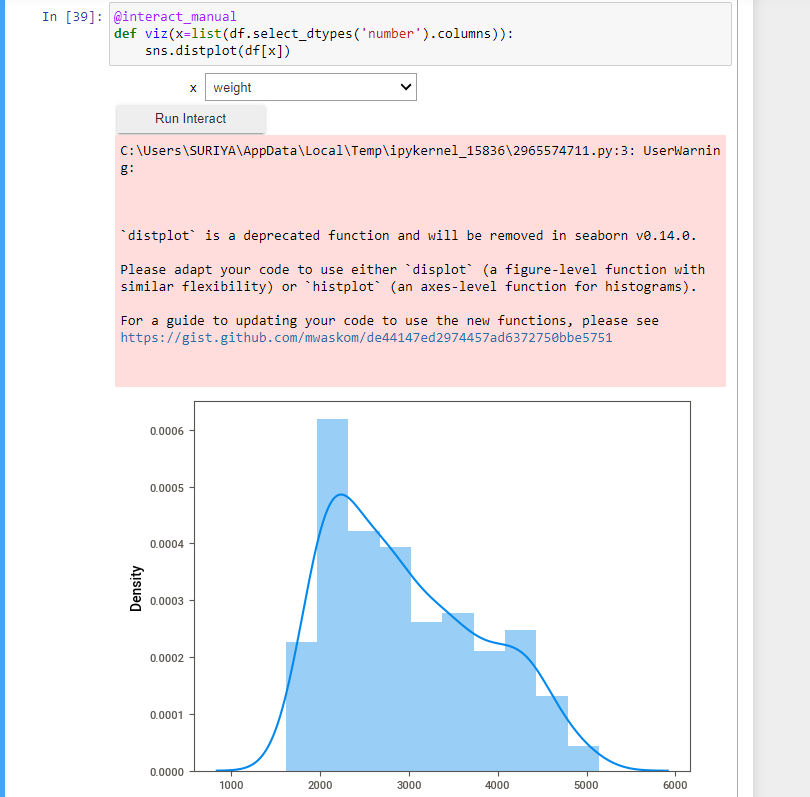
**Visualizations:**

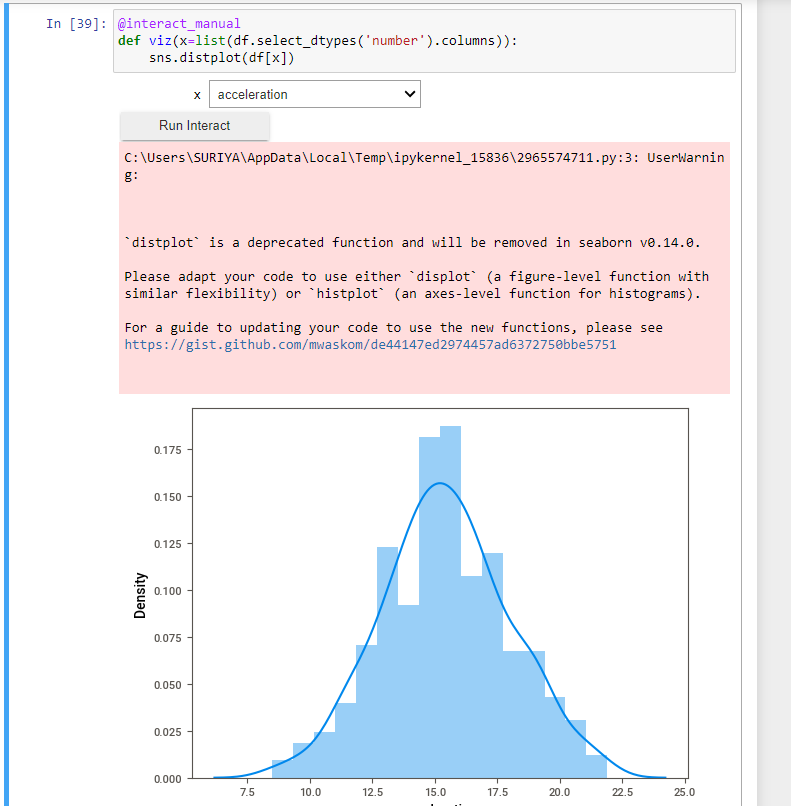
**Distribution plot:**

****

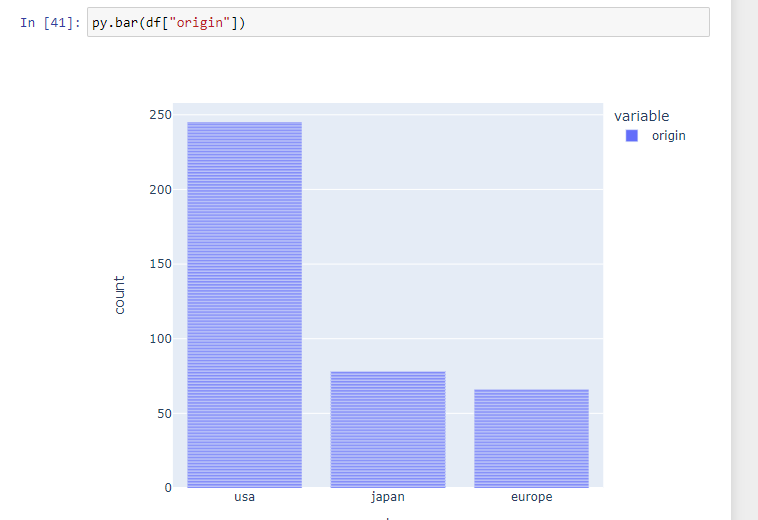
****

****

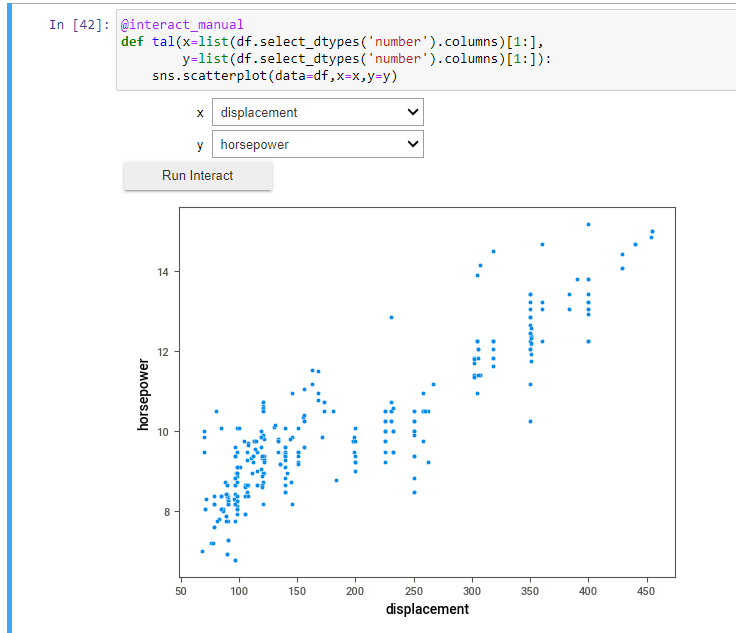
****

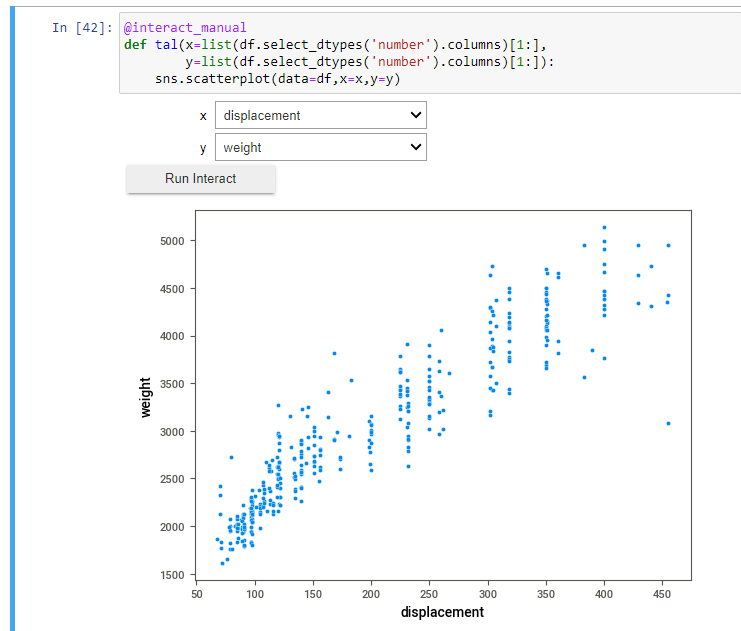
****

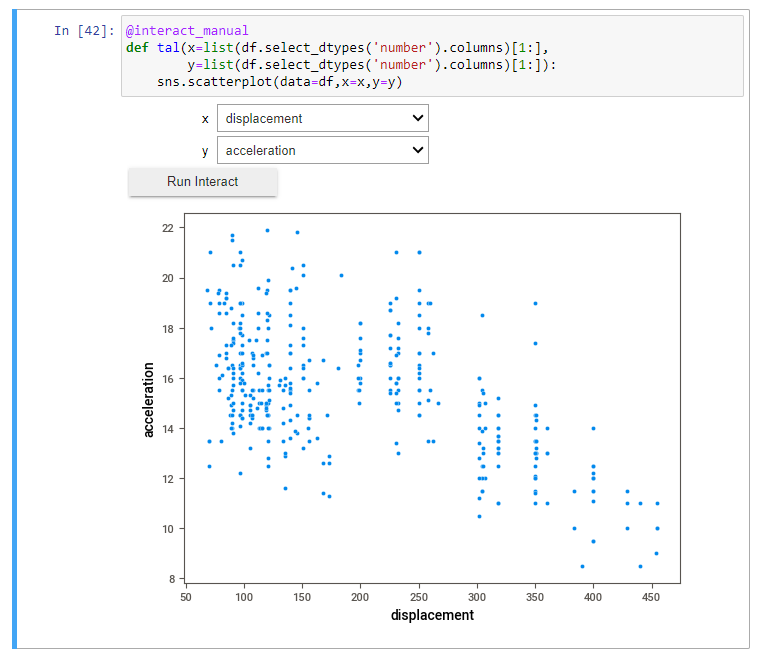
**Barplot:**

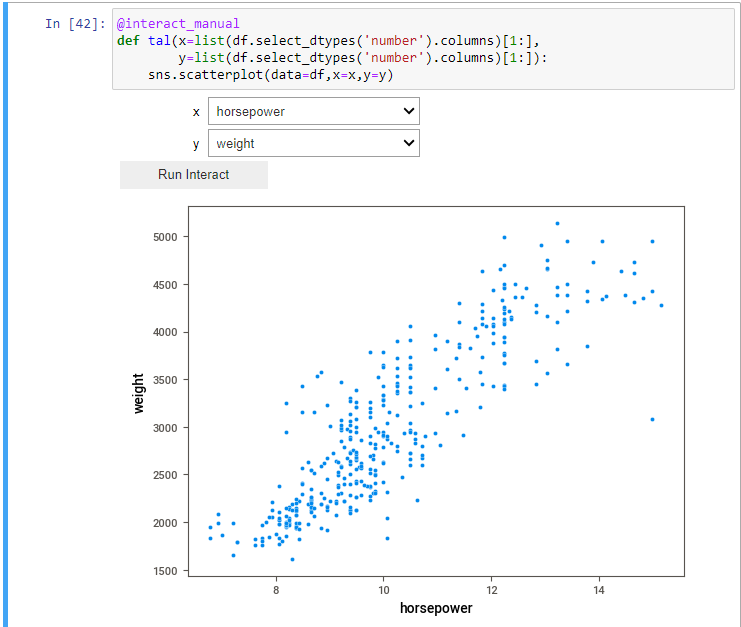
****

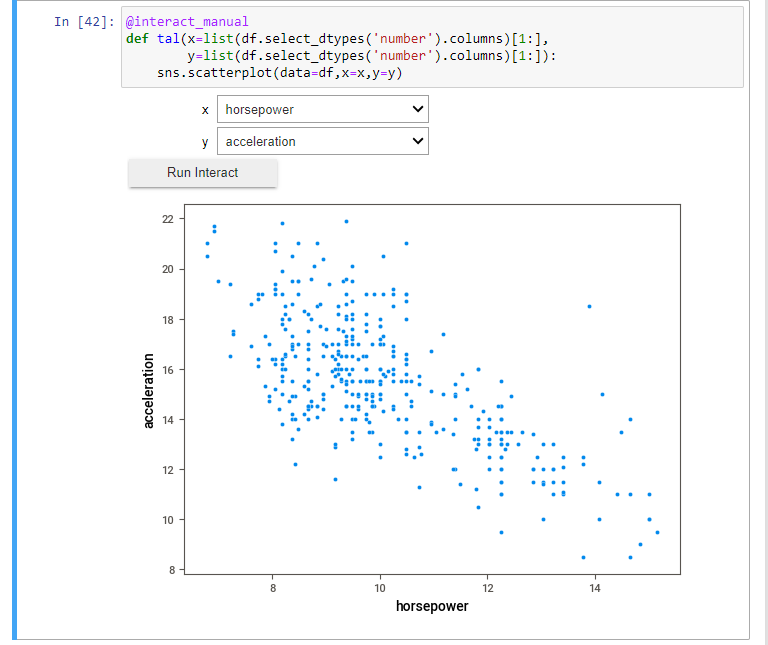
**Scatterplot:**

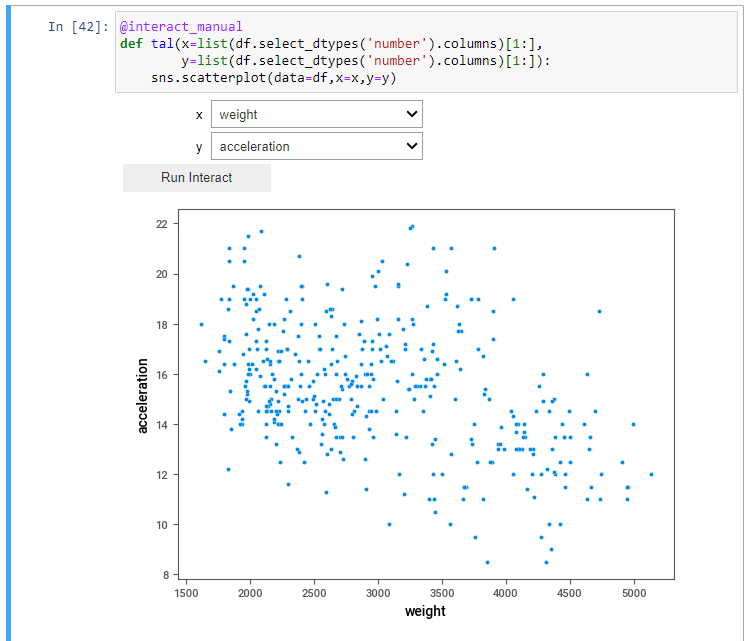
****

****

****

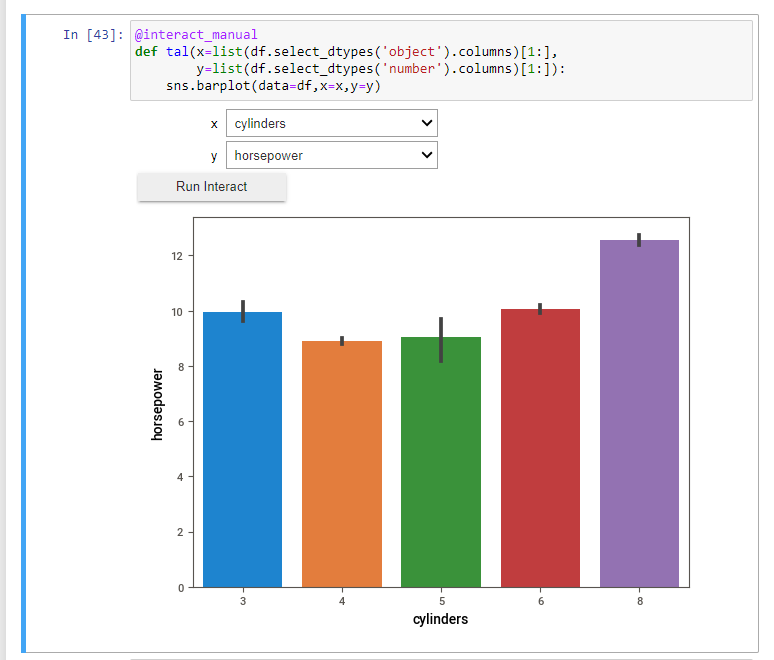
****

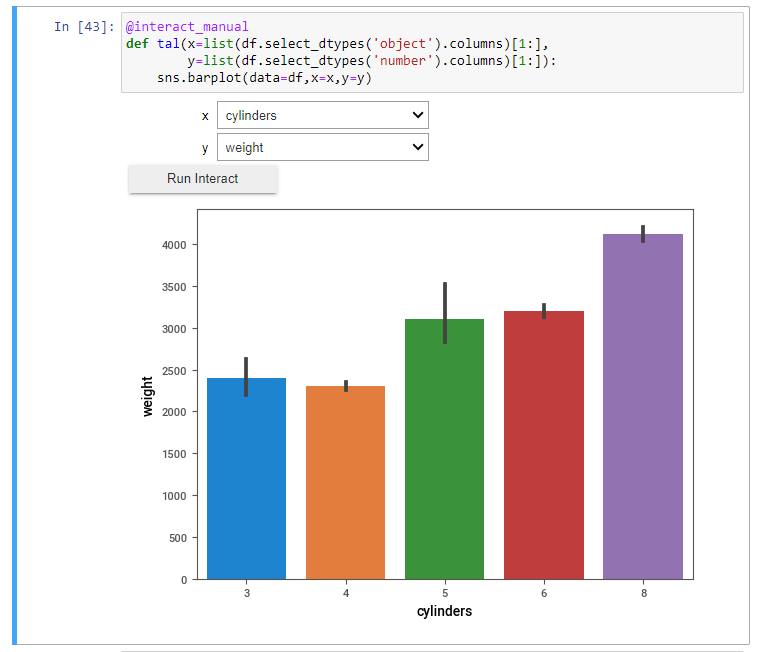
****



**Barplot:**

****

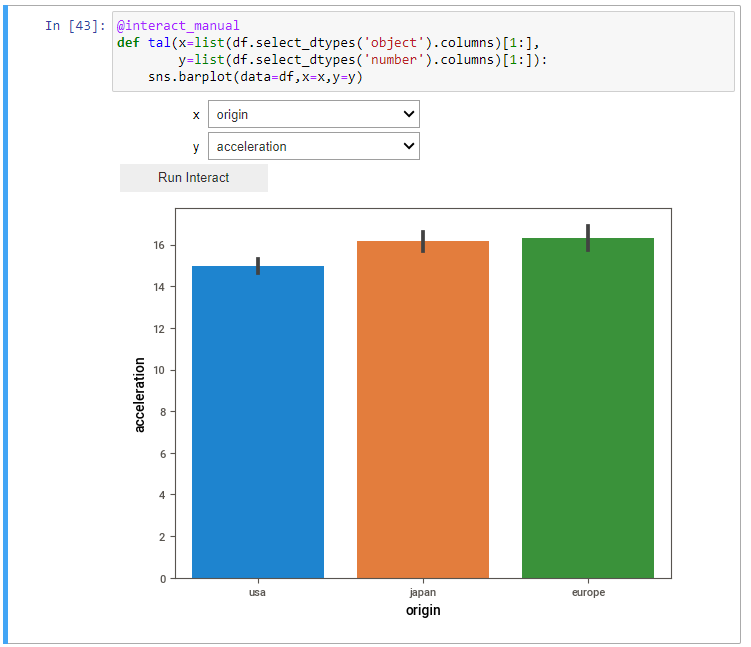
****

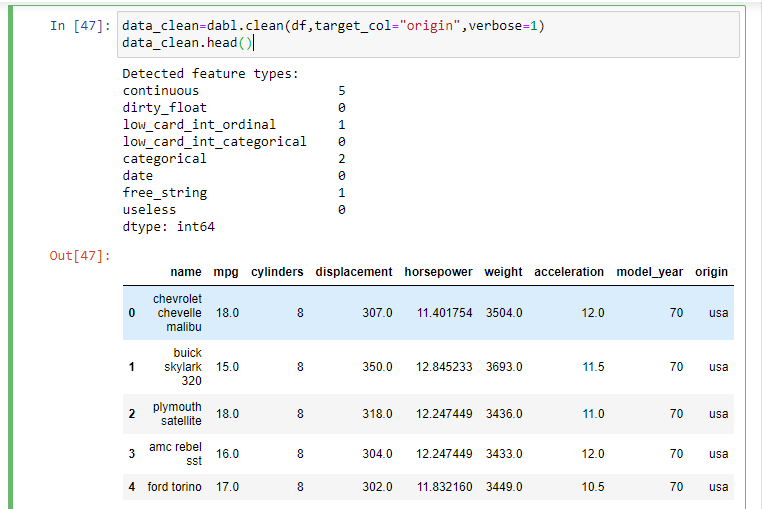
****

****

****

****

****

****