

Basic Machine Learning: Data Visualization

Oleh: Muhammad Angga Muttaqien | Founder & Mentor

Goal

Goal

To understand how important data visualization is in term of building machine learning model

1. Answer questions
2. Generate hypothesis for further analysis
3. Prepare data for machine learning modeling process



About Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

“A picture is worth a thousand words”



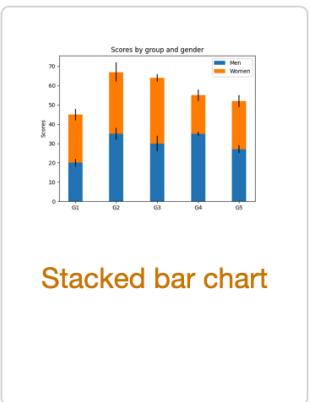
About Matplotlib

Gallery

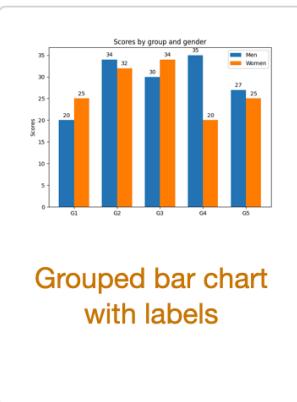
This gallery contains examples of the many things you can do with Matplotlib. Click on any image to see the full image and source code.

For longer tutorials, see our [tutorials page](#). You can also find [external resources](#) and a [FAQ](#) in our [user guide](#).

Lines, bars and markers



Stacked bar chart



Grouped bar chart
with labels

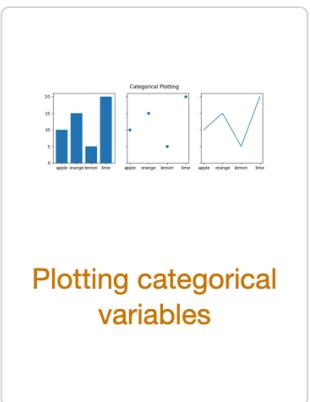


Horizontal bar chart

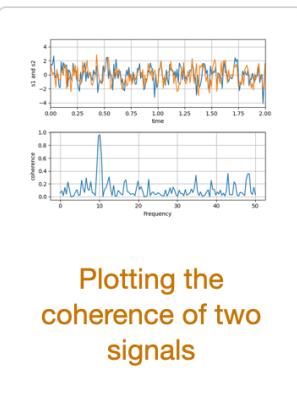


Broken Barh

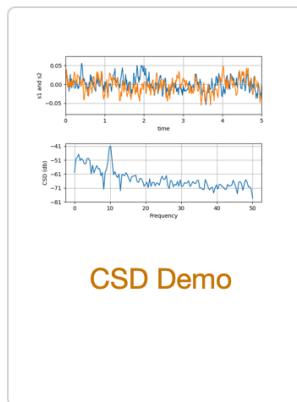
[Matplotlib Galery](#)



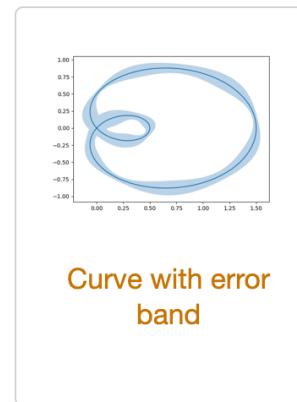
Plotting categorical
variables



Plotting the
coherence of two
signals



CSD Demo



Curve with error
band

About Matplotlib

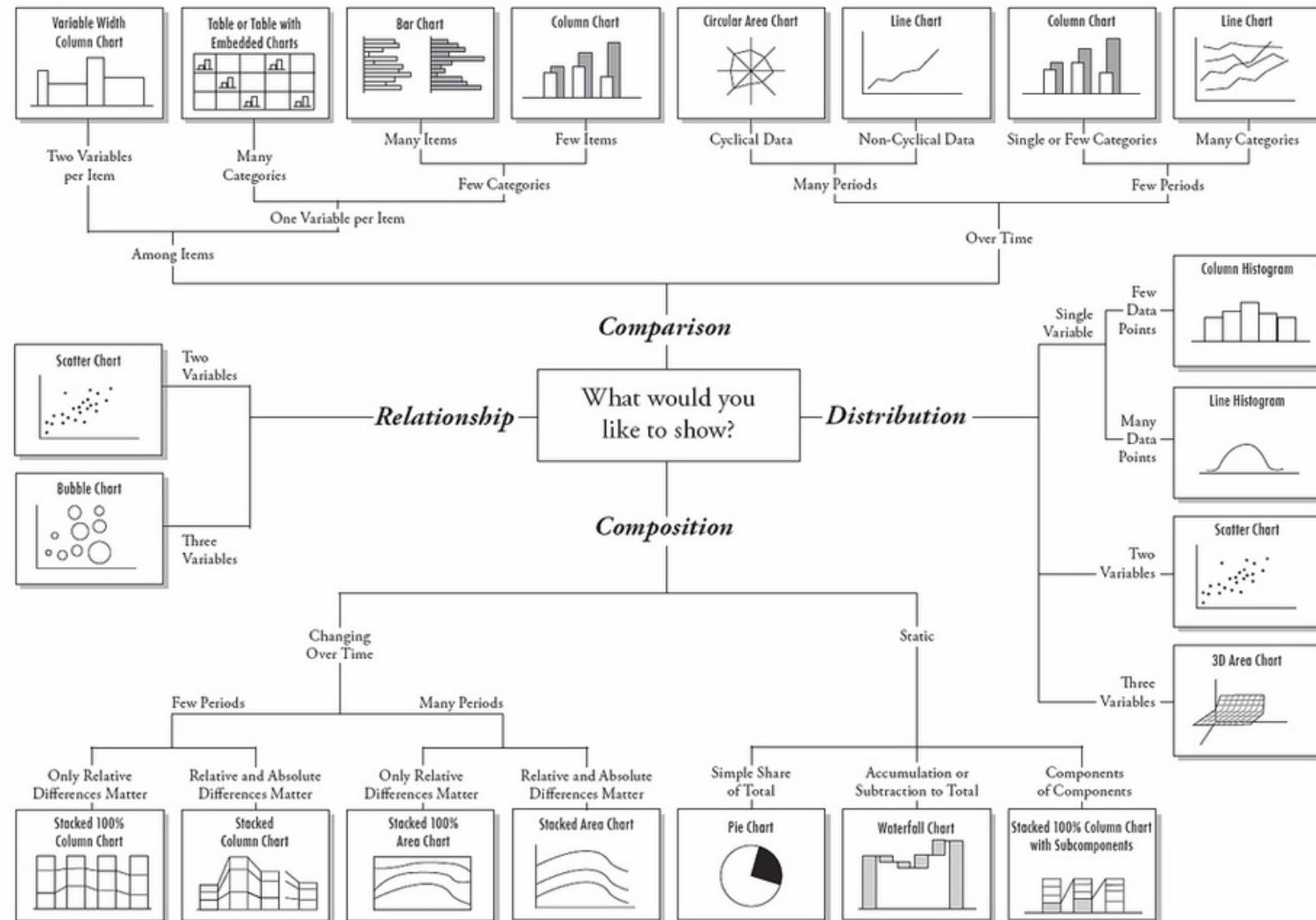
We can divide the chart messages in four categories namely:

1. Distribution
2. Comparison
3. Relationship
4. Composition



About Matplotlib

Chart Suggestions—A Thought-Starter



About Matplotlib

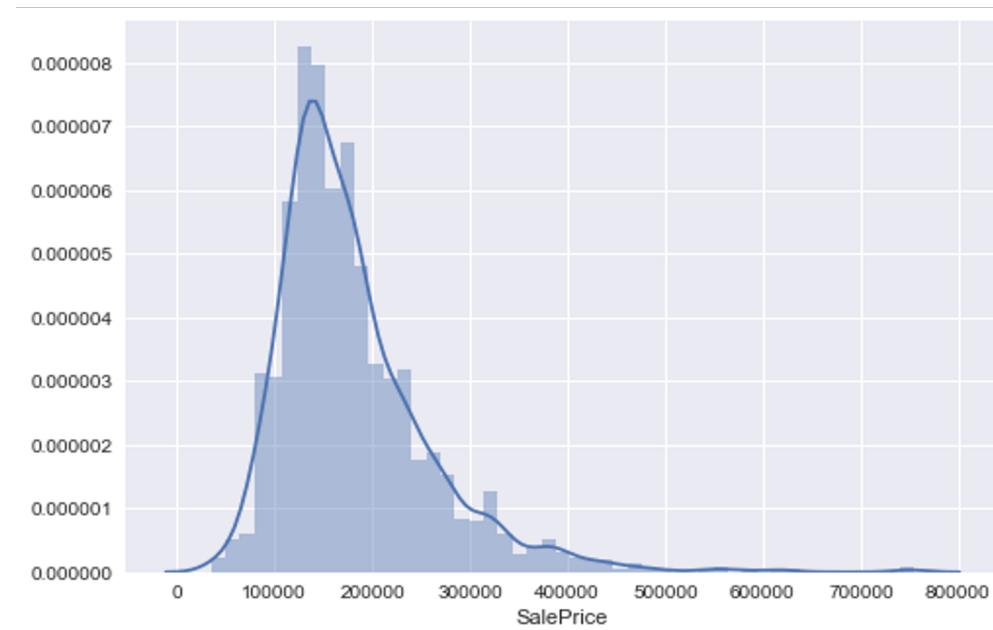
We can also divide the chart messages in three categories namely

1. Single Variable Visualization
2. Two Variable Visualization
3. Multi Variable Visualization



Single Variable – Displot

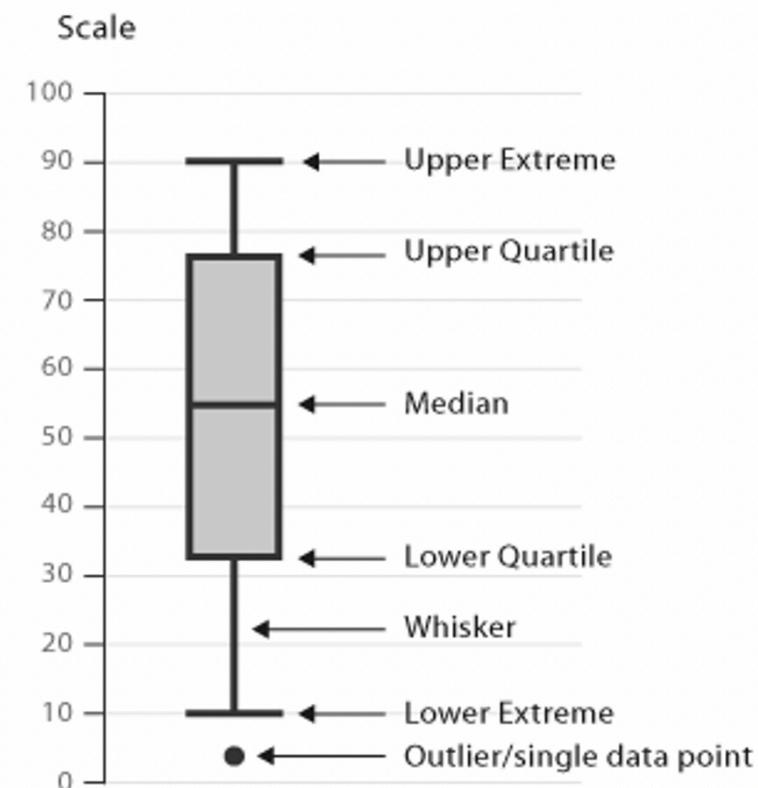
- Summarize the distribution of a single variable dataset
- Show centre, spread, skewness, outliers, multiple modes



Single Variable – Box Plot

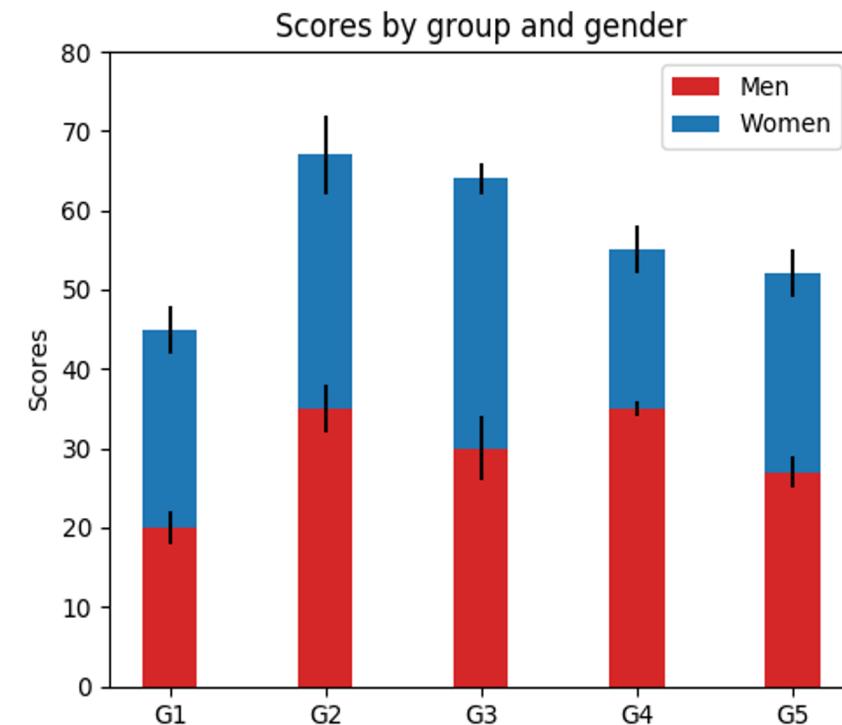
- Depicting groups of numerical data through their quartiles
- It shows a lot of information about a variable in one plot

1. Median
2. Inter Quartile Range (IQR)
3. Outliers
4. Range



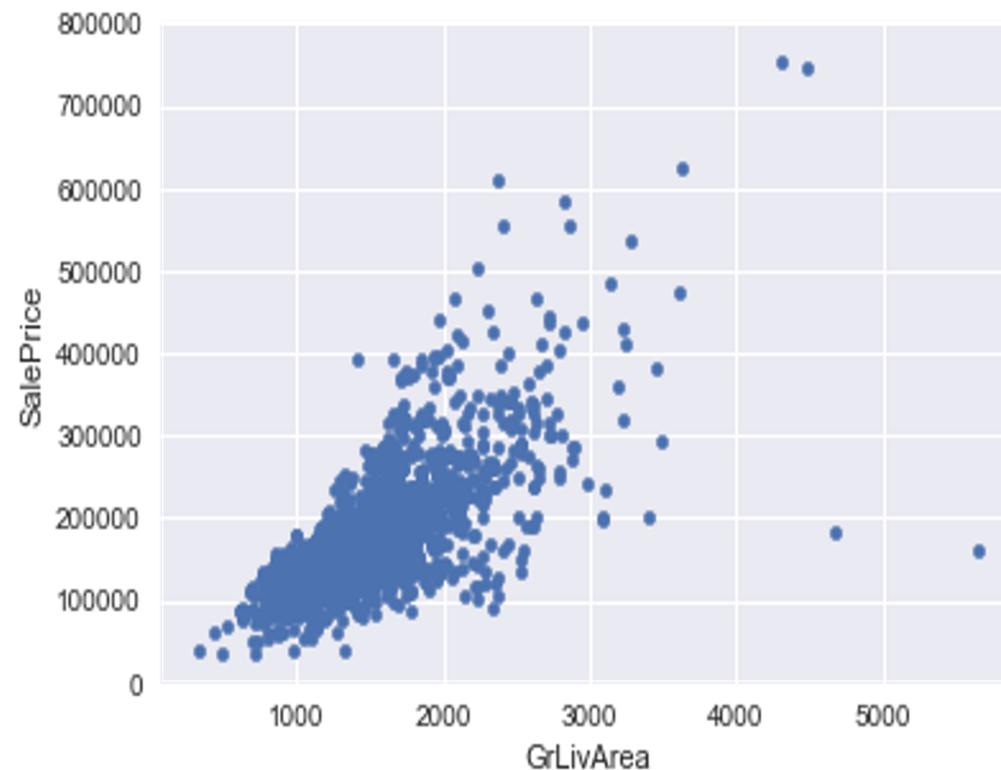
Single Variable – Stacked Bar

- Show comparisons between categories of data, but with ability to break down and compare parts of a whole

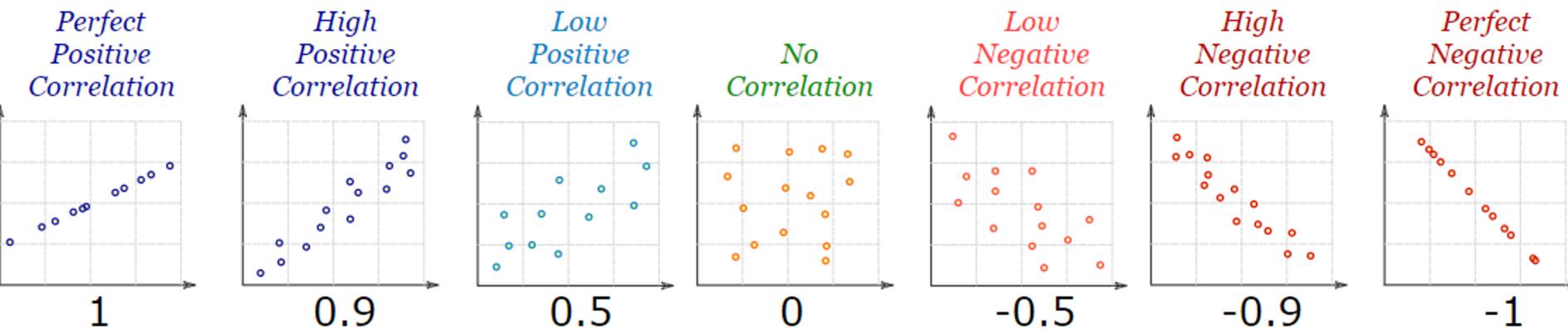


Two Variable – Scatter Plot

- For two numeric variables
- Use cartesian coordinates to display values for typically two variables for a set of data
- This sample plot reveals a linear relationship between the two variables

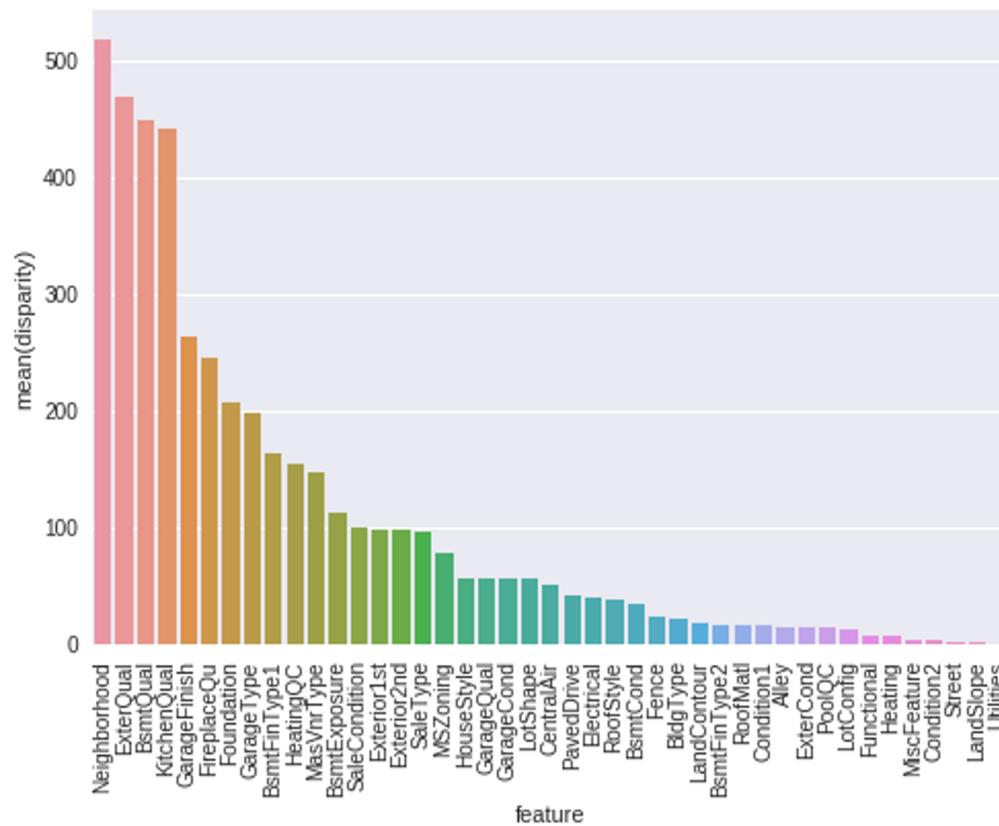


Two Variable – Scatter Plot



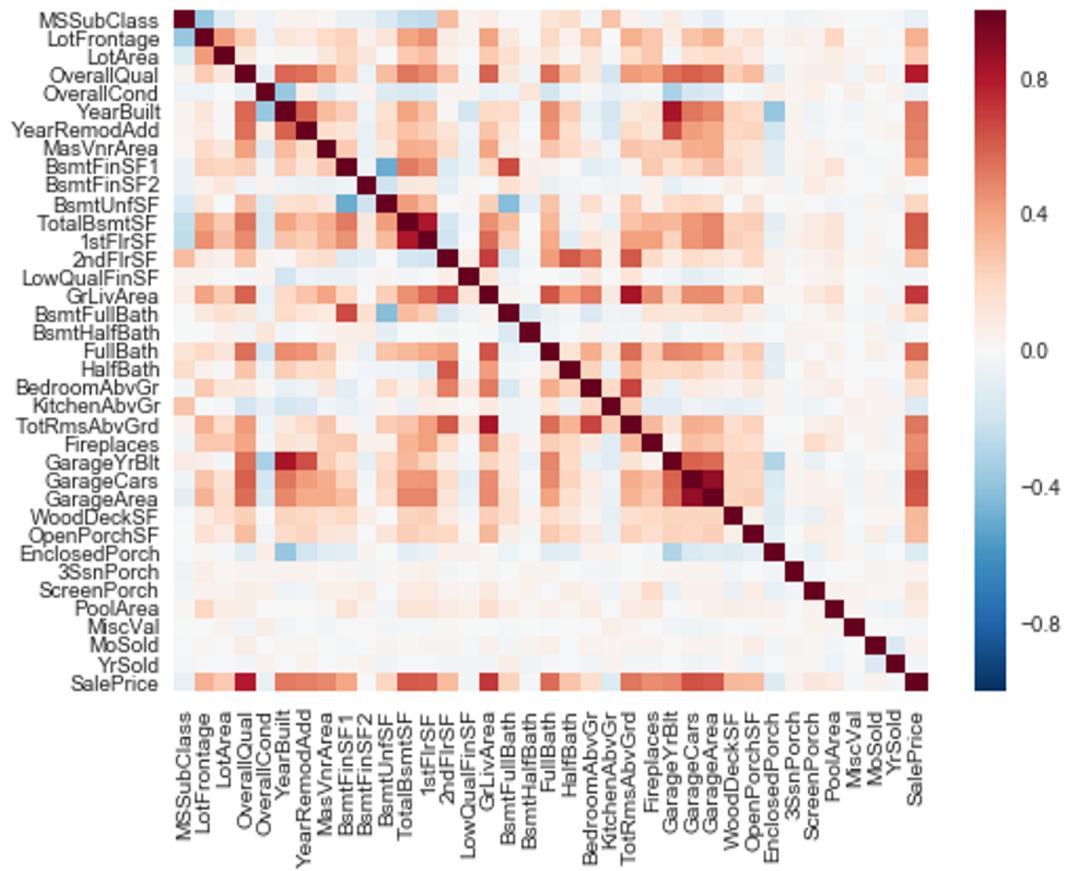
Two Variable – Bar Plot

- A bar chart or bar graph is a chart or graph that presents grouped data with rectangular bars with lengths proportional to the values that they represent



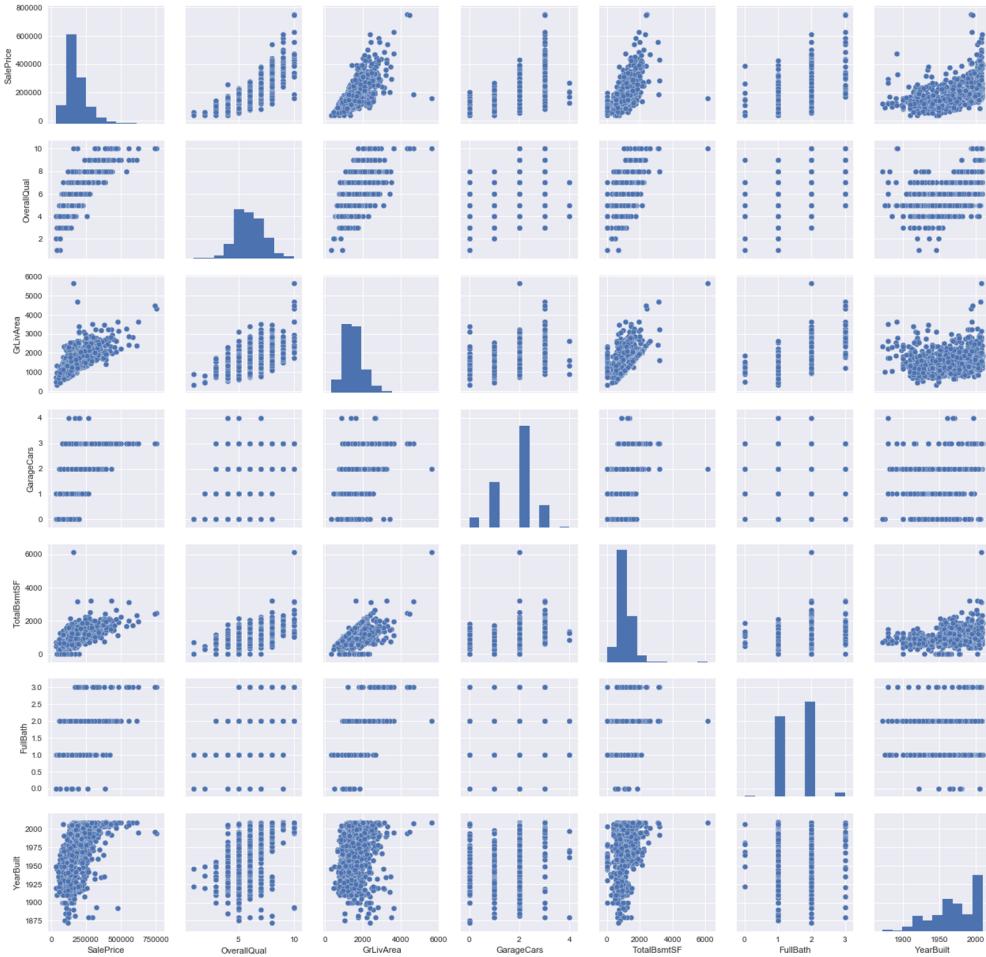
Multi Variable – Heat Map

- Graphical representation of data where the individual values contained in a matrix are represented as colours
- Helpful to represent the correlation of data



Multi Variable – Pair Plot

- Another useful seaborn plot is the pair plot which shows the bivariate relation
- Show two across all feature combinations



Assignment 1

- Lakukan Data Visualization menggunakan datasets *titanic.csv* untuk menganalisa informasi *Single Variable*, *Two Variable* dan *Multi Variable!*
- Akan dipilih dua mentee untuk mempresentasikan hasilnya pada pertemuan berikutnya!



Thanks!