Data Visualization (FOR EDA Practice)

# Importance of Data Visualization:

It is difficult to find the importance of underlying data patterns only through raw data which becomes much clearer when visualized through appropriate plots.

2 types of data:

1. Facts
2. Dimensions.

Facts – Numerical data

Dimensions - Metadata, that explains the additional information associated with the factual variables.

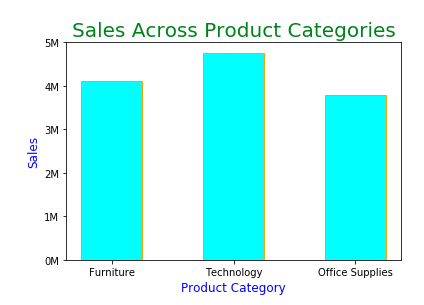
# Plots

1. Bar Graph – helpful when you need to visualize a numeric feature (Fact) across multiple categories. Better in numeric-categorical analysis.

*plt.bar(x\_component,y\_component)*

*plt.show()*

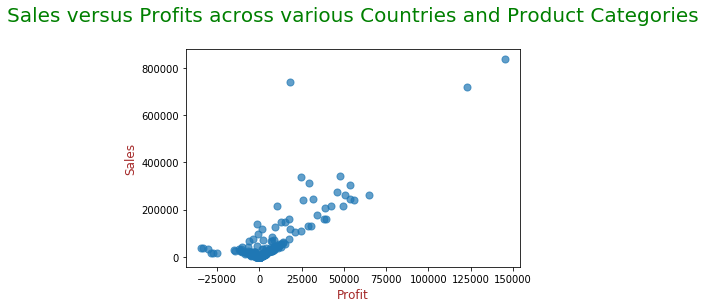
example. Sales(numeric) v/s Product category (Furniture,technology,office supplies)



1. Scatterplot – Used to identify a relationship or pattern between 2 quantitative variables and presence of outlier between them. Displays how the variables are spread across the range considered.

*plt.scatter(x\_axis,y\_axis)*

example, sales v/s profit. It will show us outlier between these two.



1. Line Graph - A line graph is used to present continuous time-dependent data. It accurately depicts the trend of a variable over a specified time period.

*plt.plot(x\_axis,y\_axis)*

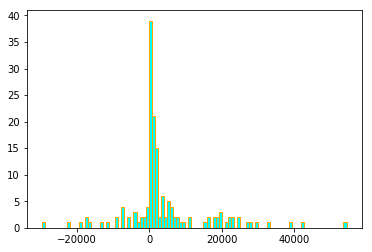
example: Sales in different months.



1. Histogram – Useful when you want to understand the distribution of a given series. It is a frequency chart that records the number of occurences of an entry or element in dataset.

*plt.hist(profit,bins=100,edgecolor=’Orange’,color=’cyan’)*

example: profit frequency histogram



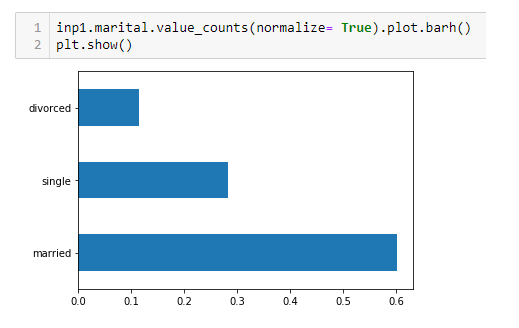
1. Box Plot – Quite effective in summarizing the spread of a large data set. They use percentiles to divide the data range.

plt.boxplot([list1,list2])



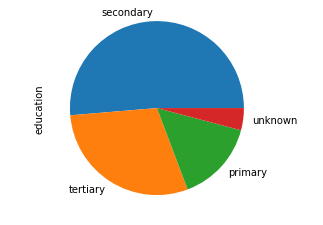
UNIVARIATE ANALYSIS

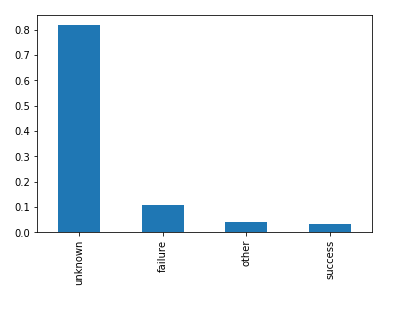
1. Categorical unordered variables - Bar graph/Pie chart. Use value\_counts() and plot to find insights.



1. Categorical ordered variables – Bar graph/Pie chart

(Same as above. Find value\_counts(), normalize true and plot bar/pie chart.)

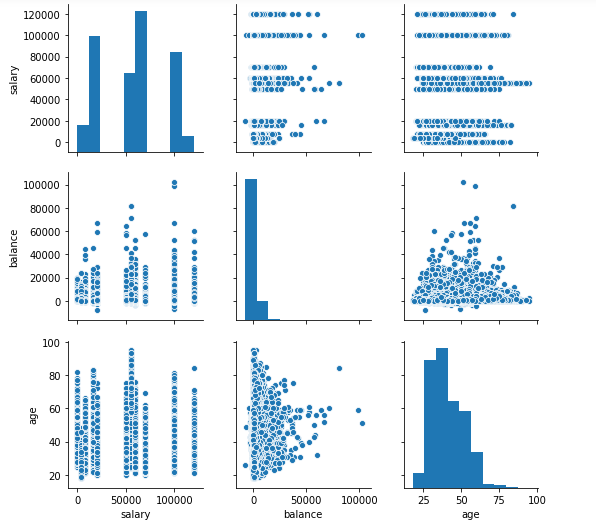




1. Statistics on Numeric fetures – Boxplot, describe(), IQR, Quantiles

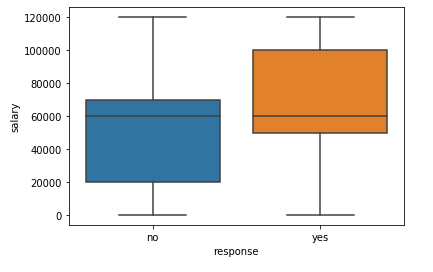
BIVARIATE AND MULTIVARIATE ANALYSIS

1. Numeric-Numeric analysis
2. Correlation matrix (relation b/w 2 numbers)
3. Scatter plot (x\_axis,y\_axis, mark every record on graph. We can see patterns) – easily visible.
4. Pair plot



1. Categorical – Numerical analysis
2. Boxplot

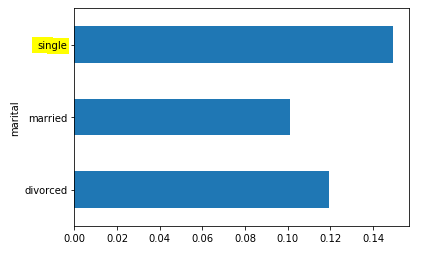
Sns.boxplot(data=inp1,x=’response’,y=’salary’)



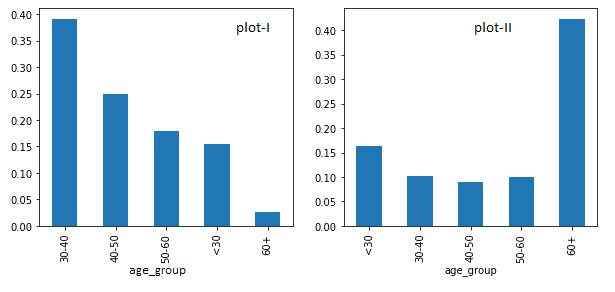
ii. Look at mean-median-quantile together

inp1.groupby(‘response’)[‘balance’].aggregate([‘mean’,’median’]).plot.bar()

1. Categorical – Categorical analysis (one would be our target variable).
   1. Using groupby and finding insights using mean, median -> then plot it. (barplot)
   2. Using countplot and setting second column as hue.



* 1. Create groups and then find relation with target variable.



1. MultiVariate analysis
   1. More than 2 variable
      1. Create pivot\_table and then plot heatmap.
      2. Use a center of reference for heat map (cmap = RdYlGn)
      3. 