

Python operations on (Q = 9 10 11 12)

- a. Create data subsets for type of post
- b. Merge two subsets
- c. Sort Data on Page total likes
- d. Transposing Data
- e. Melting Data to long format
- f. Casting data to wide format

Facebook metrics

Iris data sets

Movie data sets

Adult data sets

=====

Model Building

1) Heart Deseases

- a. Data cleaning(Remove NA, ?, Negative values etc.)
- b. Error correcting(Outlier detection and removal)
- c. Data transformation
- d. Build Data model using regression and kNN methods and compare accuracy of heart disease prediction.

2) Iris data sets

- a. Data cleaning(Remove NA, ?, Negative values etc.)
- b. Error correcting(Outlier detection and removal)
- c. Data transformation
- d. Build Data model using regression and Naïve Bayes methods and compare accuracy of Iris Species Prediction

4) Breast Cancer data sets

- a. Data cleaning(Remove NA, ?, Negative values etc.)
- b. Error correcting(Outlier detection and removal)
- c. Data transformation
- d. Build Data model using regression and Naïve Bayes methods and compare accuracy of benign and malignant tumors in Breast Cancer Dataset

- 5) Adult data sets
- Data cleaning(Remove NA, ?, Negative values etc.)
 - Error correcting(Outlier detection and removal)
 - Data transformation
 - Build Data model using regression and Naïve Bayes methods for prediction of income category ($\geq 50k$ or $\leq 50k$) and compare accuracy Prediction.

=====

Visulization:

Visualize the Heart disease dataset by plotting the following graphs using Python. (Define objective for every graph)

- Histograms
- Dot Plots
- Bar Plots
- Line Charts
- Add Histogram and Scatter plot to box plot.

- Histograms
- Pie Charts
- Box Plots
- Scatter Plots
- Add boxplots to a scatterplot