Data Science Lifecycle

- Data Collection (crawling)

- Data Mining

- Data Analysis - numpy, pandas

- Data Cleaning - pandas scikit

- Irrelevant data

- Handle missing values

- Duplicate values

- Unwanted Outliers

- Structural errors

- Data Preprocessing - scikit

- Feature Scaling

- Label Encoding

- One Hot Label Encoding

- Train/test split

- Statistical Analysis - sklearn

- Algebra, stats, matrix, calculus, probability

- Data visualisation - matplotlib ,seaborn, bokeh

- Predictive Modelling (Machine Learning)

- Optimisation

- Data visualization