

UCSanDiegoX-s-Data-Science-MicroMasters

1.1 Course 1-Python For Data Science

Part-1 of some of the important notes that I took from the week-1 lecture of my Micro-Master in Data Science by UC san diego.

- Learning to understand "BIG PICTURE" of data science through case studies
- Understanding the statistics of the datasets
- Learning to summarize how data cleaning and correlations are applied to the datasets
- Reciting the data visualization techniques applied in the case study
- Importance of how clustering of similar groups and plotting these clusters helps in any case study.
- Recalling what was used in drawing conclusion based on data analysis.

DIMENSIONALITY REDUCTION: Enables you to reduce feature while preserving variance.

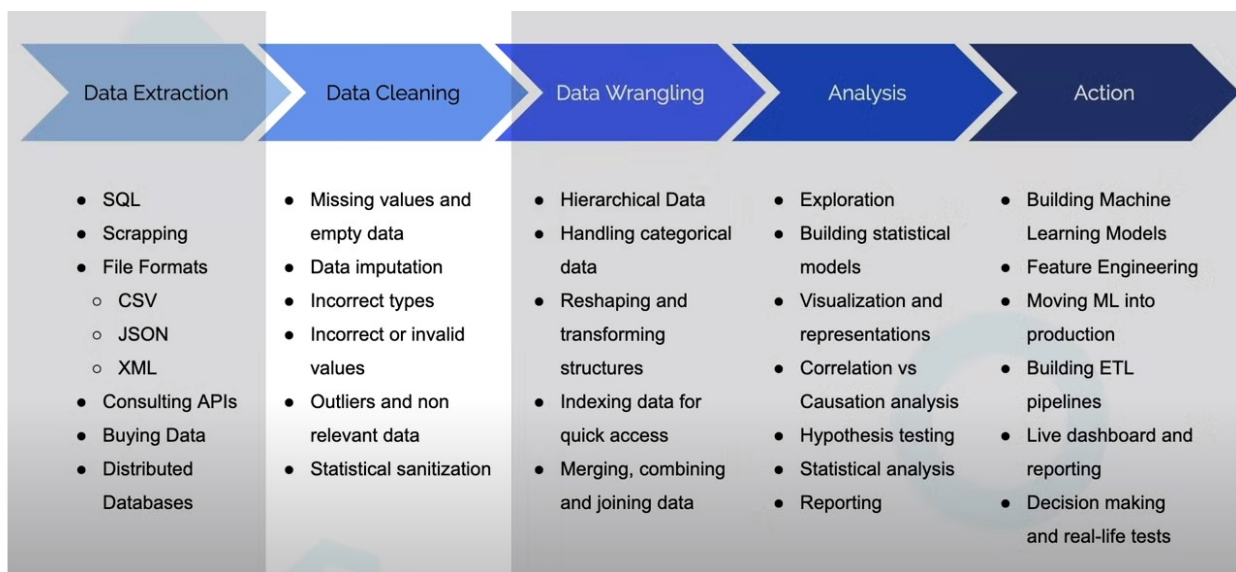
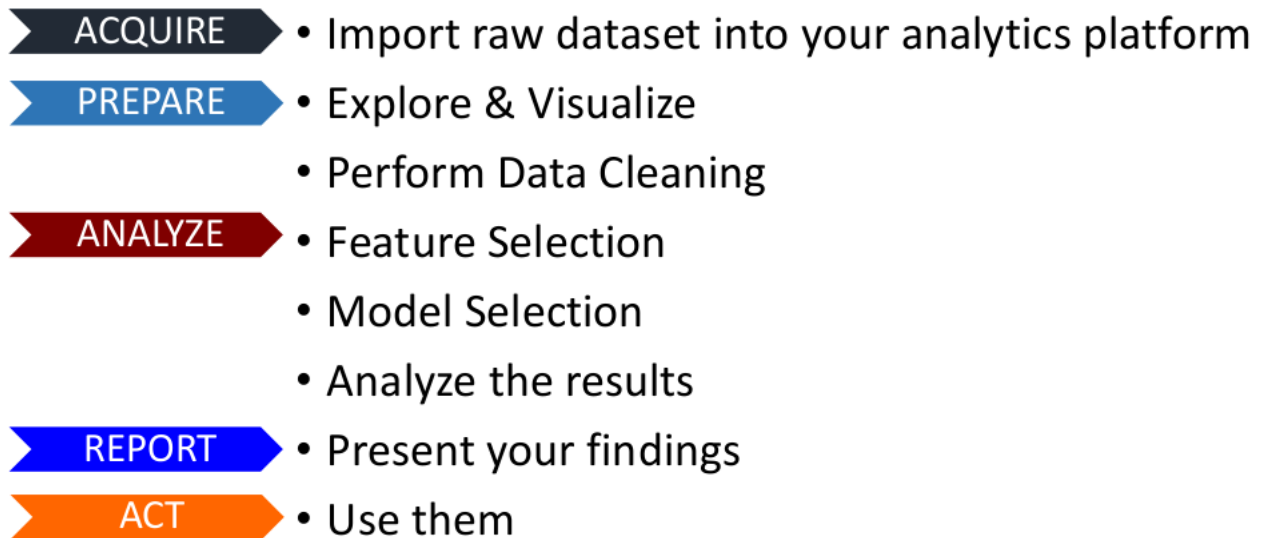
SCIKIT LEARN: Scikit learn has abilities for:

- Principle Component Analysis
- Singular Value Composition
- Factor Analysis
- Independent component Analysis
- Matrix Factorization
- Latent Dirichlet Allocation

DATA SCIENCE TECHNIQUES: An important aspect of data science is choosing correct model . And selection of model depends on the types of problem statement. It could be supervised,semi supervised ,unsupervised or reinforcement learning.

BASIC STEPS IN DATA SCIENCE

Basic Steps in a Data Science Project



IMPORTANCE OF STATISTICS IN DATA SCIENCE: Summary statistics are quantities that capture various characteristics of a set of values with a single number or a small set of numbers like mean median MOde standard deviation Mean and median are measures of the location of specific values . Mode is the value that occurs most frequently in your data set. And range and standard deviation are measures of spread in your data.