**QTW-Week 4 Presession Submission-Imputation**

**This week was all about Imputation**

Dataset can be incomplete in 3 forms:

* Incomplete
* Incorrect
* Missing data

Missing data patterns can be of 3 types:

* Missing completely at random (MCAR)
* Missing at random (MAR)
* Not at random

Imputation can be done in following ways:

* Ignore the missing data
* Use a method based on summary statistics
* Replacing the data is based on guess. We can use mean, median, mode, fit and sample.

There are several ways to fill the missing values:

• If the data is normally distributed, try the mean (as in replace all missing values with the mean).

• If the data is categorical, try the mode.

• If the data is non-normal, try the median.

• If multiple variables are correlated, either drop the column with missing data or perform a fit to predict missing data.

Best way of imputation depends on the problem.

If the data is missing at random then,

* A second variable can show a pattern for missing data.
* Once you identify the pattern, the missing data is completely at random.

Domain knowledge allows us deeper insights into solving the problem at hand.

NAN is the representation of missing data in python.