Instructor’s mail: jerry@vempower.org

Detecting Outliers

Missing value: mainly depends on the data. Use median, mean, mode, knn (k-nearest neighbor) is most effective which helps 90% to detect  
ask questions to stakeholders. Why is the data missing, what is the pipeline? Always ask the right questions.

Project:

Create master table – load cleaned data to master table using ETL (postgreSQL)

Left, right join is important but full outer join is a must do

3 things are to submit in a combined pdf

Deadline: 8th sept