# Week 1 and Week 2

# Pre Processing

# Missing value

## Visualize Missing Value

## Impute missing value by mean or by median or by most frequent or MICE or KNN imputation

# Feature Engineering

# Encoder

# Scale data

# Outlier

## visualize

## Detect outlier

# Week 3 and Week 4

# Feature Engineering

* Write function to create binned column from existing numeric columns automatically
* Write function to create a numeric column from categorical column to show % distribution of each category

# Visualization

## SNS/panda

* Categorical scatter plot with hue
  + strip
  + Swarm
* Categorical distribution plots
  + Box
* Categorical estimate plots
  + Point
  + Bar
  + Count
* Joinplot
* Pairplot
* Line plot
* Numerical scatterplot
* Multiple relationships with facets

## Descriptive Statistics

## Correlation

# PCA

# Week 5 to Week 8

Write down top 10 observations from graphs that relates to business

Find optimal value of eps for DBScan programmatically

# Modelling

* Build model with different dataset like raw data, featured engineered data, with label encoding, with dummy columns, after removal of outlier etc
  + sklearn classifier,
    - Cross validation
    - Recursive Feature Elimination
    - Feature Importance
    - Hyper Parameter Tuning
    - Predict
    - Model metrics - R-square, Adjusted r-square, MAPE, MAE, RMSE, MSE.
    - Calculate Accuracy and compare models