**Project Planning:**

**Outline:**

# Understand the nature of the data (numeric, categorical, etc.)

* *.info()*
* *.describe()*

# Histogram and boxplots

* *discover any trends in the data*
* *outliers*

# Value counts

* *for categorical, we can’t use histogram.*

# Missing values

* *To remove the missing data or*
* *Impute the missing data*

# Correlation between the metrics

* *Heatmap*
  + *For regression problem, we need to know what are the related metrics*
  + *For classification, it’s just yes/no type results*

# Explore interesting features

* *Brainstorm if some of the features contribute to the outcome such age, gender, location*

# Feature engineering

* *Maybe create new variable if necessary to reduce the dimensionality*

# pre-process data together or use a transformer

# Use label for train & test

# Scaling

* *Model building may require scaling*

# Model baseline

# Model comparison

**Conclusion:**

Model tuning may provide better results, however better data can provide far better results. So, it is important to analyze and prepare the data prior to modelling.