Strategy: to tackle a data science problem.

1) Explortary Analysis: (Data Exploration)

Tools & methods to to be used to explore the data.

a) <u>COFOUNDING REASONS</u>:

develop cofounding reasons to build and justify which key features should be used to harness the results. These selected features (co-founding reasons) will steer an ideal learning curve where there is a short gap between training and testing data. The ideal set will contain minimum feature which will give the mentioned result.

b) perform statical analysis on the data.

* measure MEAN, Median, Mode.
* Take care of Out liers so mean $\bar{X}$ is not bias.

* Make sure the mean $\bar{X}$ lies between IQR.
  Inter quartile range is the range between 1^{ST} Qtr → 3^{rd} Qtr.
* If the mean does not lie between IQR. Then there are Out liers

* Identify the Outliers, Isolate the data from the outliers and compare the $\bar{X}$, mean and the model before and after the $\bar{X}$, mean is corrected.
* If we have incorret model due to outliers, we will introduce an in incorect mean to fill in null value. , which will skew the prediction.
* when the model is skewed, the recall and precision of the predictions will be incorret, since our data is skewed -
* when we are training the data on the offsets will outliers introduced our training and test splits will have contradicti values since the existing data was correct but the data which was imputed with an incorrect mean $\bar{X}$, due to outliers will have drifted values leaning towards the outlier. Therefore, the Recall & Precision of the Test predictions will will reduce and the results will contain poor accuracy.