v) For more complex analysis, you may want to walk the audience through the analytical problem solving process.

vi) Always reference the data sources used.

vii) Make sure your analysis supports the decision that need to be made.

3/1/21 **Lesson 2 :- Selecting an Analytical Frame work**

1) **Selecting an Analytical Methodology :-**

- **Methodology Map :-** The analytical Problem Solving frame work helps systomatically work through a business problem; however, it doesn't help us to understand which methodology to use.

  The Methodology Map will help us as we decide which methodology to use to solve a business problem, and is meant to be used as we work our way through the Problem solving frame work.

▷ <u>Non - Predictive Business Problem</u> :-

The non- predictive analysis are divided into four categories :

i) <u>Geospatial Analysis</u> :- This type of analysis uses location based data to help drive your conclusions. Example includes identifying customers by a geographic region, calculating the distance store location or creating a trade area based upon customer locations.

ii) <u>Segmentation Analysis</u> :- Segmentation is the process of grouping data together. Groups can be simple, such as customers who have purchased different items, to more complex segmentation techniques where you identify stores that are similar based upon the demographic of their customers.

iii) <u>Aggregation Analysis</u> :- This methodology simply means calculating a value across a group or dimensions and is commonly used in data analysis. for example :- we may want to aggregate sales data for a salesperson by month - adding all of the sales closed for each ~~other~~ month. Then

you may want to aggregate across dimensions, such as Sales by month per Sales territory. Aggregation is often done in reporting to be able to "slice and dice" information to help managers make decisions and view performance.

iv) **Descriptive Analysis :-** Descriptive Statistics provide simple summaries of a data Sample. Example could be calculating average GPA for application to a school and etc. Some of Commonly use descriptive Statistics are mean, median, mode, Standard deviation and. Inter quartile range. ~~for normal distribution~~ :-

$$Mean = \mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$Variance = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 = \sigma^2$$

$$Standard \ deviation = \sigma = \sqrt{Variance}$$

Median = 1st sort them then.

$$\begin{cases} n+1/2 & n = odd. \\ Average \ of \ 2 & n = even. \end{cases}$$

Mode = Most often appears

$$IQR = Q_3 - Q_1 = Q_{50\%}$$

 75th 25th

 upper lower quartiles.

▷ **Predictive Business Problem :-**

— **Data Rich vs Data Poor :-** Do you have data on what you are trying to predict? If so, you can proceed down the data rich Path, otherwise the data poor path is only option.

▷ **Data Poor Business Problem :-**

— **A/B testing :-** If there is not sufficient usable data to solve the problem, then we need to setup an experiment to help us get the data we need. An experiment in a business context is usually referred to as an A/B Test.

▷ **Data Rich Business Problems :-**

— **Numeric vs Non-Numeric Analysis :-**

Assuming we have enough data to proceed with the

analysis.

i) <u>Regression Model</u> :- Numeric outcome are those where the outcome is simply a number. Predicting the demand for electricity or the hourly temperature are both numeric outcomes. Models predicting numeric data are called regression models.

ii) <u>Classification Model</u> :- Non numeric outcomes are those where we're trying to predict the category into which a cases (eg. customer) falls, such as wheather a customer will pay on-time, pay late or default on a payment. Models predicting non-numeric data are called classification models.

▷ <u>Introduction to Numeric Models</u> :-

— <u>Target Variable</u> :- Target variable represent the outcome we are trying to predict. 1st determine whether the target variable is numeric or non-numeric.

- ## Type of Numeric Variables :-

i) **Continaous** :- A continuous variable is one that can take on all values in a range.
eg :- Height.

ii) **Time based** :- A time based numeric variable is one where you are trying to predict what will happen over time.
eg :- forecasting.

iii) **Count** :- Count variable are numbers that are discrete, positive integers. They're called count numbers because they're used to analyze variable that you can count.

D **Introduction to Non Numeric Models** :-

- **Non numeric Variable** :- A non numeric variable is often called categorical, because the value of the variable take on a discrete number of possible values or categories.
eg :- i) Electronic device will fail before £k hrs or not
ii) wheather a customer will pay on time, pay late, or default on payment.

iii) whether a store is classified as large, medium or small.

Classification Model.

Binary	Non Binary
When modeling categorical variables, the no. of possible outcomes is an important factor. If there are only two possible categorical outcome such as Yes or NO or True or False. then they are called binary.	If there are more than two possible categorical outcomes, such as small, medium or large.