Lesson 2:

Analytic problem solving framework doesn't help us understand which methodology we will use.

Types of non-predictive data analysis.

Geospatial:

(1) use location based data to derive conclusions.

egs (1) identifying customers by geographic location

(2) Calculating distance store locations

(3) Creating a trade area based upon customer locations

Segmentation:

is the process of grouping data together.

Identify stores based on demographics of their customers.

Aggregation:

Calculating a value across a group or dimension.

for example : aggregating sales data for a sales person by month, ~~add~~.

~~or~~: sales by month per sales territory.

Aggregation is often done in reporting to be able to slice and dice information.

Description :

provide simple summaries of a data sample.

eg : Calculate avg temperature per hr / per day.

mean, median, mode, IR.

Classifying Business Problem

have enough data to solve the problem — data rich.

do not have enough data to solve the problem — data poor

data poor : do not have any data related to sales to predict an outcome.

Data . Poor Business Problem

A/B Test → If we don't have sufficient data, then we need to set up experiment to help us get data we need.

For example . we could introduce the new product at . all some stores to estimate the sales at all stores.

Data Rich Business Problem.

Numeric outcomes : predicting . the demand of electricity / regression model to predict numeric outcomes

Non- numeric outcomes :

Predicting whether. a customer will pay on time, pay late or default on payment → Classification models to predict non numeric outcomes.

Target variable : represent the outcome we are trying to predict . In order to select the right

3 Types of Numeric variable.

(1) Continuous : A continuous variable is a one that can take all values in a range.

(2) Time - based : A time based numeric variable is one where we are trying to predict what will happen over time.

Count variables are numbers that are discrete +ve/the integers.

Non-numeric models

(1) if there are only 2 possible outcomes then variable is classified as binary.

(2) If there are more than 2 possible outcomes, then variable is classified as non-binary.