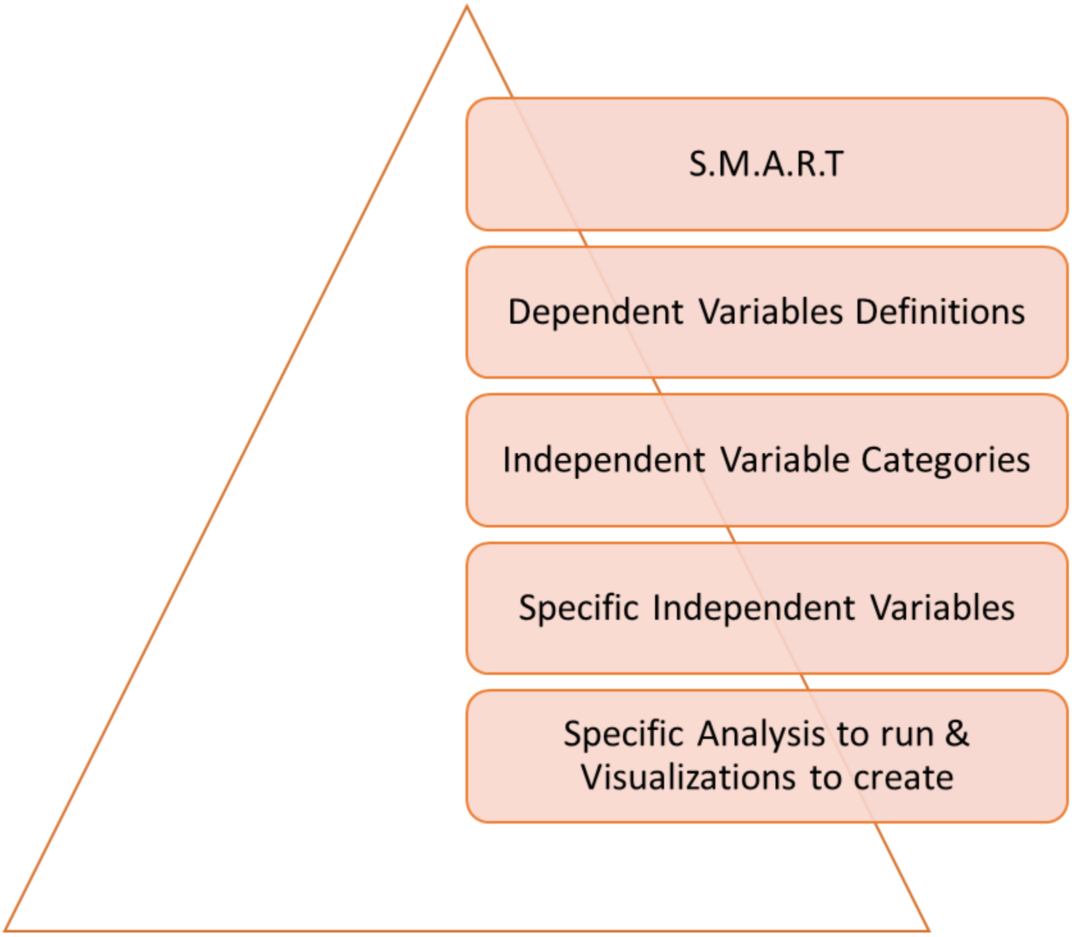
STRUCTURED PYRAMID ANALYSIS



S.M.A.R.T

Dependent Variables Definitions

Independent Variable Categories

Specific Independent Variables

Specific Analysis to run & Visualizations to create

MULTIVARIATE ANALYSIS

- Under the guidance of
  Prof. Raunak Parik

- Team members:
  Trishala Ghone
  Aakash Borhade
  Siddhesh Inamdar

S.M.A.R.T

SPECIFIC

- We are a group of three and our project is based on Multivariate Analysis.
- This dataset we have selected is from National Institute of Diabetes and Digestive and Kidney Diseases.
- The aim of our project is to diagnose if a particular patient has diabetes based on various medical parameters in the dataset.
- We will accomplish this by using machine learning techniques Technologies used: R and Rstudio

MEASURABLE

- Perform operations on the dataset depending upon the nature of the dataset and the questions asked.
- The performance can be measured based on the completion of the assigned task and also on the results from the techniques performed.

ACHIEVABLE

- Operations performed on the pima diabetes dataset:
    - Understood the nature of dataset.
    - Asked relevant questions.
    - Cleaned the data
    - Data Visualization
    - Performed Statistical tests

RELEVANT

- The power of machine learning in diagnosing disease and in sorting and classifying health data will empower physicians and speed-up decision making in the clinic.
- Using different analytic techniques and algorithms can provide better information to doctors at the point of patient care.

TIME-BOUND

- The long term goal is to predict if a patient has diabetes or not based on a few medical factors.
- Research and perform the best algorithms which will help us in accurate prediction.

Dependent Variables Definitions

- Here, the 'Outcome' variable is the Dependent variable. This variable depends on the variation of the other independent variables.
- Each independent variable has a different impact on the output variable.

Independent Variable Categories

The following are the independent variables

- Pregnancies – The number of times the women is pregnant
- GlucosePlasma - Glucose concentration in an oral glucose tolerance test
- BloodPressure - Diastolic blood pressure (mm Hg)
- SkinThicknessTriceps - skin fold thickness (mm) #Insulin2- Hour serum insulin (mu U/ml)
- BMIBody mass index- (weight in kg/(height in m)^2)
- DiabetesPedigreeFunctionDiabetes -pedigree function
- Age - (years)
- OutcomeClass variable – Dependent Variable
- All are numeric variables except for the outcome variable (yes/no)

Specific Analysis to run & Visualizations to create

- Questions raised on our data set
  - What exactly do I want to find out?
  - What does the data 'look' like? Does it follow any known probability distributions?
  - How are the various measurements related?
  - Number of Pregnancies has an impact over diabetes outcome?
- Statistical tests performed:
  - T-test
  - Hotelling
  - F-test
  - Levene test
- Data Visualisations
  - Missingmap
  - Stripchart
  - Correlation matrix
  - Correlation Plot
  - Density Plot