INTRODUCTION:

CRISP-DM stands for cross-industry process for data mining. The CRISP-DM methodology provides a structured approach to planning a data mining project. This model is an idealized sequence of events. In practice many of the tasks can be performed in a different order and it will often be necessary to backtrack to previous tasks and repeat certain actions. The model does not try to capture all possible routes through the data mining process.

Phases:

1.Business understanding

2.Data understanding

3.Data preparation

4.Modeling

5.Evaluation

6.Deployment

1.**BUSINESS UNDERSTANDING**

This is all about to understand what you want to achieve. These can be achieved by looking at the data given. According to this project there are some few questions to answer these include:

1.Is population directly affecting the grand electors? (relationship between population and grand electors)

2.what is the relationship between population and state?

3.relationship between state and grand electors?

N/B-in this phase what is very key is to understand the problem and the objective you want to achieve.

2.DATA UNDERSTANDING.

This is much about understanding the type of **data that required**, format, **processes to acquire** and method to integrate if they are many data types.

There are 3 reports need to be submitted.

1.initial data report-source

2.data description report. -format, quantity

3.data exploration report-first findings of the data.

For the case of our data:

1.initial data report-source. -the data was downloaded from the website

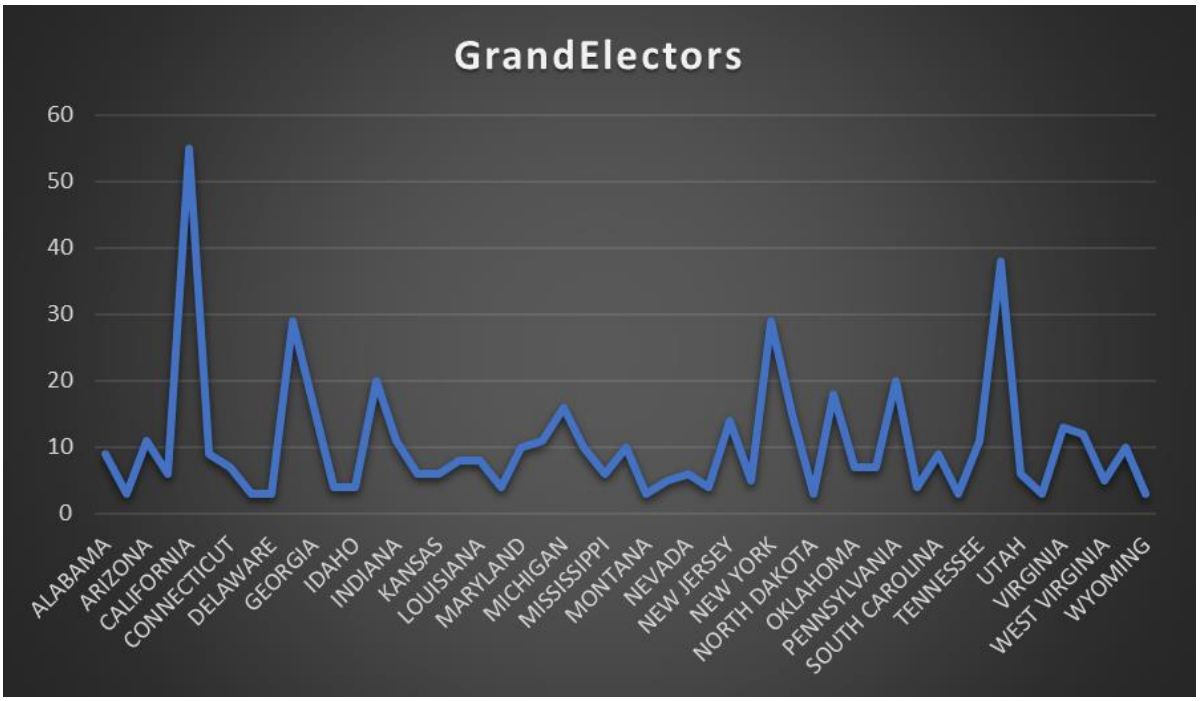
2.data description report. -it's in csv format and two column datasets

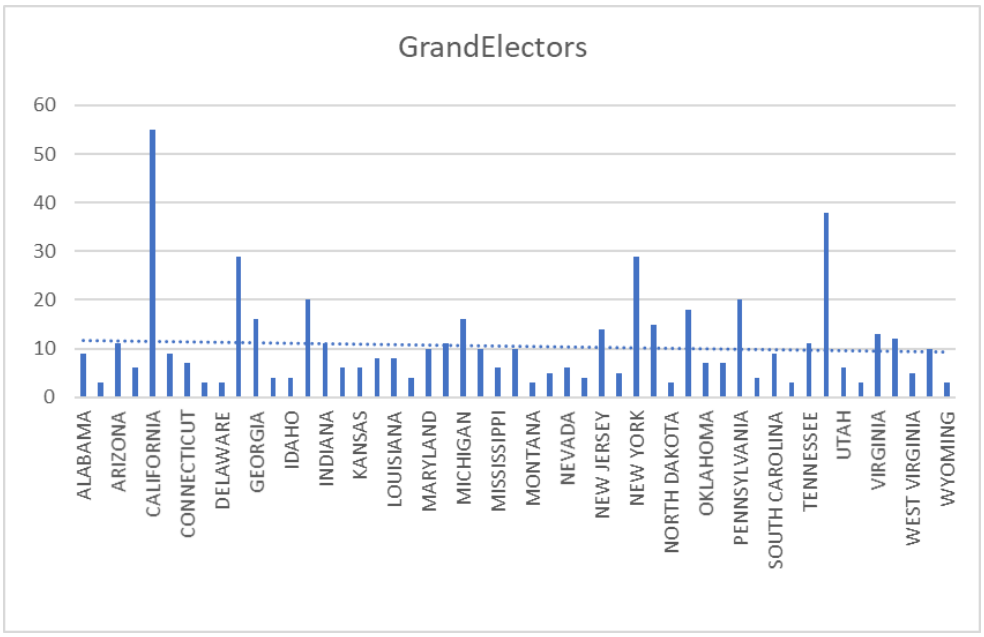
3.data exploration report-explore the data and verify data quality using data quality formats and produce a data quality report.

3.DATA PREPARATION.

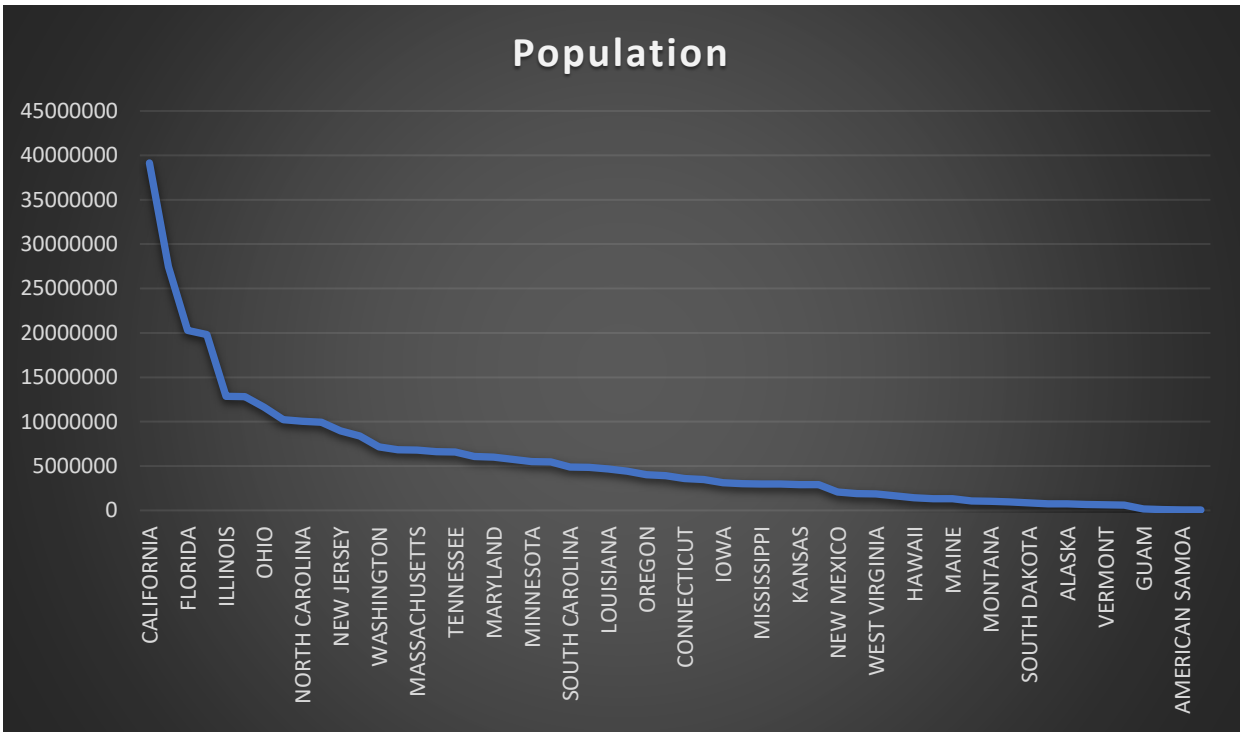
In this stage the analysis is used to come up with list of data to be excluded and in included, after which we clean our data using python where we check if there are columns which are empty. After which we construct the required data and then we integrate all of them. This is a stage where analysis starts.

The first two chats show the grand Electors versus state:





Population against state.



Point to Note:

Population is directly proportional to the Grand Electors.

4. MODELLING STAGE:

Before we build a mode here, we generate a mechanism to test the validity and Quality. after testing the validity and quality we can now generate a model to perform certain tasks.

This stage can be further divided into 4 stages:

1.seelecting the model techniques.

2.Desighning the test.

3.Building the model

4.Assessing the model.

Examples.

Using the C5.0 for decision three building or using back propagation in neural network generation, segmentation and clustering.

In segmentation: the dataset contains all the value no field is missing. After segmentation we perform an EDA where we check if the STATE, POPULATION and GRAND ELECTORS features are multimodal distributions (each having two or three heels) and using the plots we check the dense regions After EDA we are suppose to perform the clustering analysis.

5.EVALUATION

During this step we will assesses the degree to which the model meets this project objectives and seek to determine if there is some reason why this model is deficient. Another option we will test the model(s) on test applications in the real application, if time and budget constraints permit. The evaluation phase also involves assessing any other data mining results you've generated. Data mining results involve models that are necessarily related to the original business objectives and all other findings that are not necessarily related to the original business objectives, but might also unveil additional challenges, information, or hints for future directions.

6.DEPLOYMENT

In the deployment stage you'll take your evaluation results and determine a strategy for their deployment. If a general procedure has been identified to create the relevant model(s), this procedure is documented here for later deployment. It makes sense to consider the ways and means

of deployment during the business understanding phase as well, because deployment is absolutely crucial to the success of the project.