**Elizabeth Josephine**

**Data report on the data science lifecycle for the elections of the United states**

Having been asked to pinpoint the states that need to be prioritized so as to help a contesting candidate win the US elections, I will use the following techniques to determine the outcome of the election while ensuring a win for one of the candidates. Using the CRISP-DM methodology to analyse the situation at hand, below are some of the processes that will be in play during my analysis of the research question at hand concerning the elections of the US. an election with two prospects.

1. Understanding of the objectives of the exercise

This first step will help me determine what it is that is expected of me at the end of this research. The accomplishments from this sect allows for the uncovering of significant factors; that have the ability to affect the results of the whole election. With this understanding, it is easier to understand what outputs are desired from this election while assessing the situation at hand. Understanding of these objectives will also help me know what data mining goals are expected of me along with helping me come up with a comprehensive plan.

From the information provided, it is clear that only one of the candidates will win, which explains the need for a data scientist to determine which states are to be targeted during the campaign periods.

1. Understanding of the provided data which in this case is the population, grand electors and the states’ data

In this phase, I will describe, explore, verify the quality of data and finally make a quality report of the data. This is because this stage requires me to have the resources and especially the data in this project. While the provided data sets contain the population count by state and the number of grand electors in every state, it is crucial to know that the grand electors are always attributed at the state level. This is to say that, in each and every state out of the possible 51 states, there is a select count of the grand electors needed for a candidate to win. This also means that the presidential candidate who receives the highest number of local votes gets to win all the grand electors in a particular state.

1. Data preparation

In this phase, I will use the provided datasets to prepare the data for evaluation. This is only possible through data selection and cleaning. Other than that, i will have to construct the data required of me, even if it means having to join the tables for comprehensive data. The cleaning will determine the quality of data I use for analysis which is why it has to be thorough. When all this is done then i have to integrate the data from multiple databases and tables which in this case are just two.

1. Modelling

In this stage, I will select a modelling technique, generate a test design for my data along with building a model. However building a model is not enough because the models have to be assessed. This process involves a summary of the results of the research task, listing my models and doing a quality/accuracy check.

1. Evaluation

At this phase of the project, I will determine if the results did meet the objectives of the research question. This stage allows me to evaluate the results, review the processes used in the project so as to know that which has been overlooked. Depending on the results of the assessment, this phase will determine whether I continue or whether I have to fall back and make a newer plan of action.

1. Deployment.

In this phase, I will take the results from my evaluation and determine a deployment strategy from them. This will include how I plan on performing these steps, which in this case are the things to be done by the presidential candidates if they expect a win. With everything set, a maintenance and monitoring plan in place is appropriate along with the production of a final report. This report is to be presented by me to the task handler after a comprehensive view of the whole project in question.