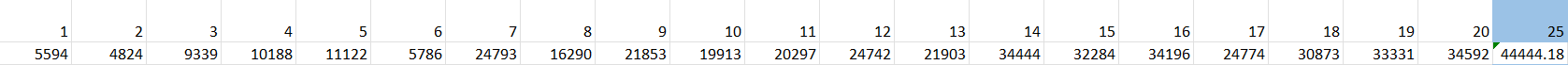
Given the attached dataset, the following tasks are to be performed:

1. Identify which classification methodology can be used to build a prediction model which can predict the expected new infected and new death by Coronavirus in the future.

*Linear regression is the perfect classification methodology to predict a numeric prediction, it uses x = “Days” and y=” New Cases” or “New deaths” to build the model, after that we get to predict new cases or new deaths for a given day, for example:*

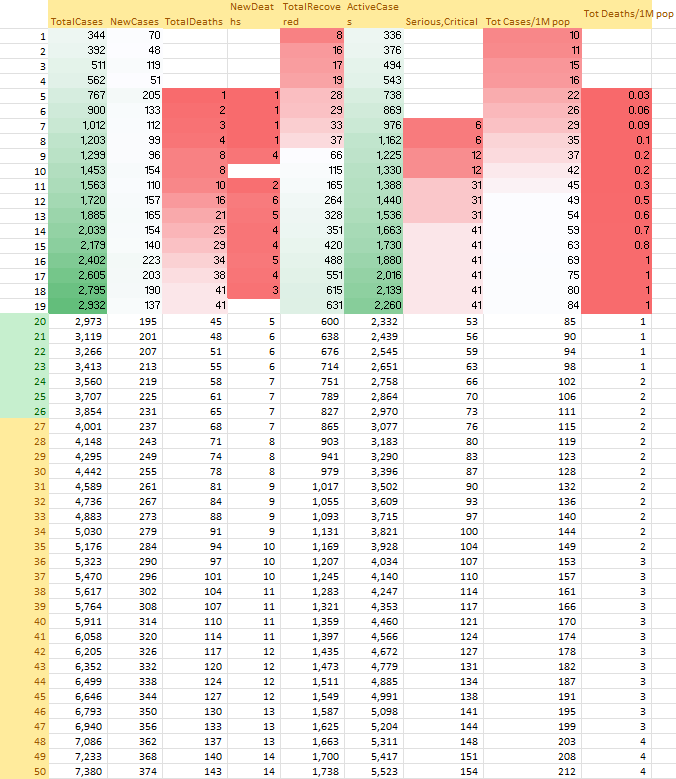
* *This is a prediction of new cases for USA in the 25th day of the virus:*

1. Try to use the attached data to build the model from a different perspective (Saudi Arabia in one the week, Saudi Arabia in all days, North Americas countries (US, Canada) for one-week, North Americas countries (US, Canada) for all days, all world in one week, all world in all days)

*Saudi Arabia:*

*One-week: Green*

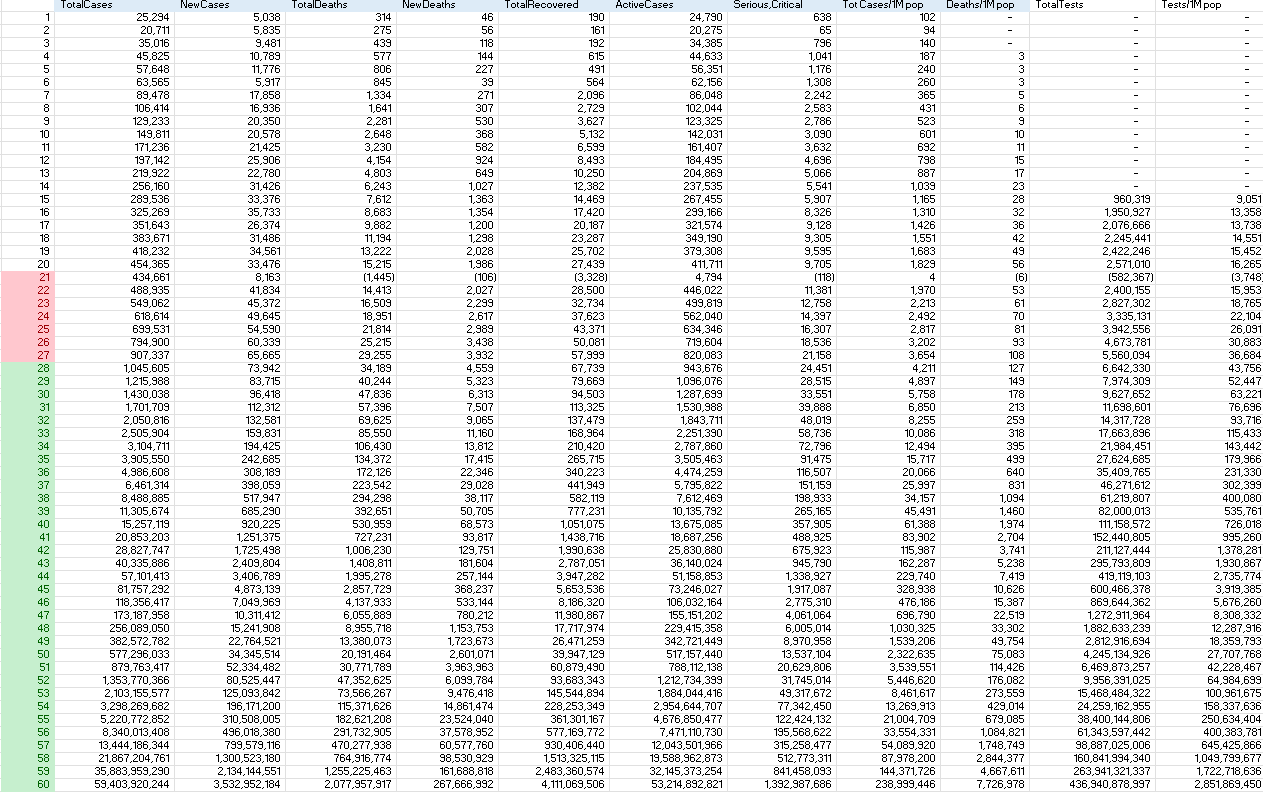
*All days: Yellow*



*North America:*

*One-week: Red*

*All-Days: Green*



1. Explain in detail your finding and justify your method.

We find that according to our prediction, all increases with time in both North America and Saudi Arabia with the exception of the total death per 1m pop in Saudi Arabia being relatively stable and controlled.

1. If you can use more than one model, choosing the best model, and justify your chosen model using the validation techniques (hold out or 10x cross validation)

**The validation techniques lack a class for application.**

1. Measure the accuracy of the used model.

*Total Cases: 43%*

*New Cases: 56%*

*Total Deaths: 12%*

*New Deaths: 17%*

*Total Recovered: 26%*

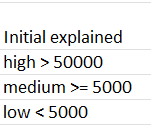
*Active Cases: 48%*

*Serious: 32%*

1. Using the clustering methodologies, grouped the county into three levels of risk (High, Medium, Low) using the mean for one round only,

*We used the k-means clustering method to group the countries into three groups (high, medium, low), the grouping was placed on the ground of “Total cases” of the last day “8th” of the given data:*

* *Our initial clustering explained:*

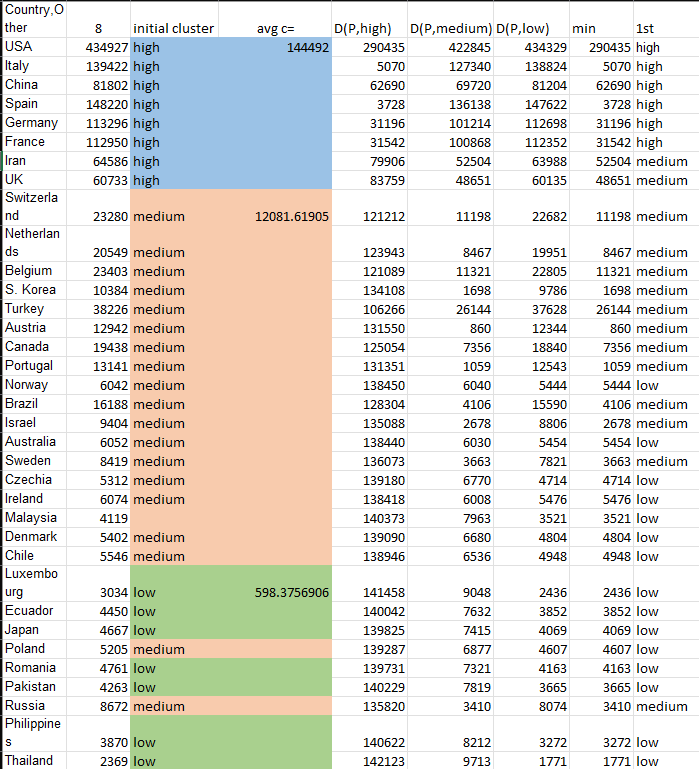


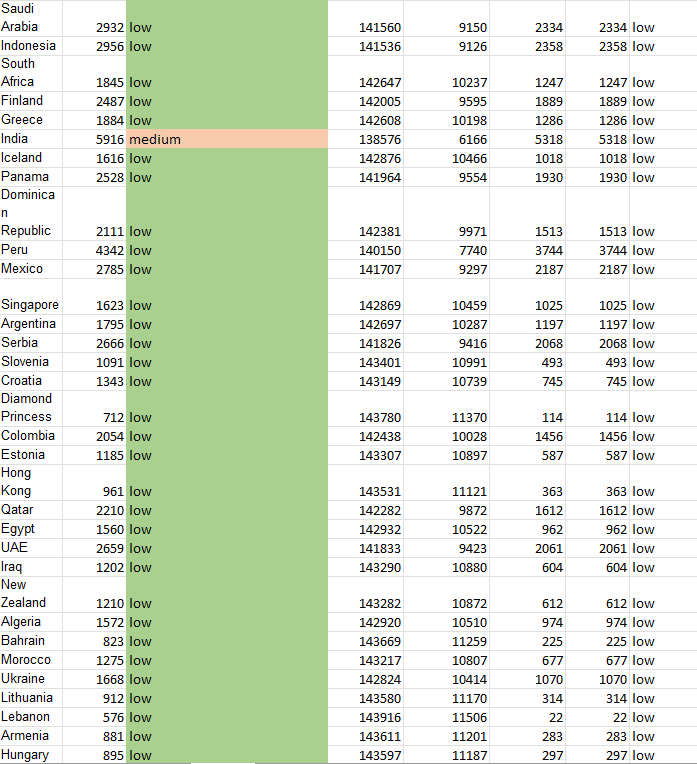
* *Note:*

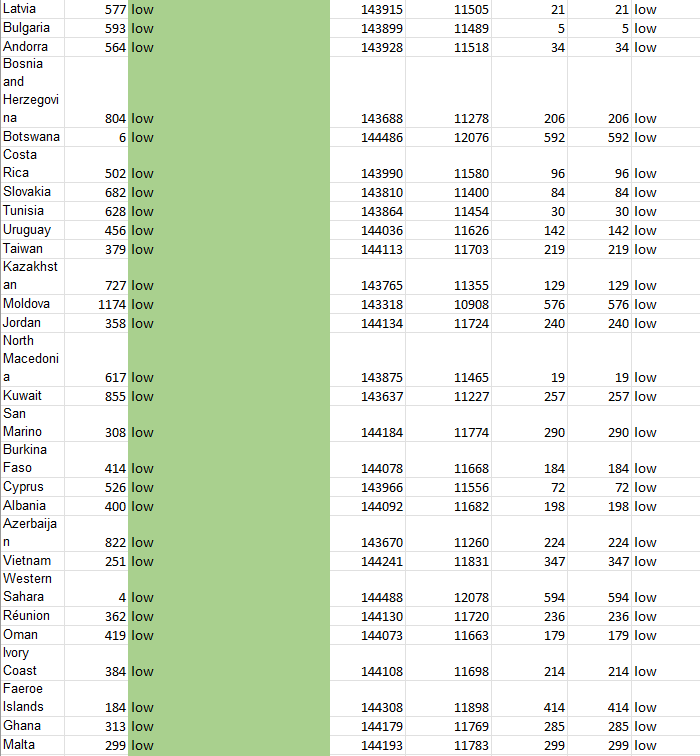
Graphical user interface, application

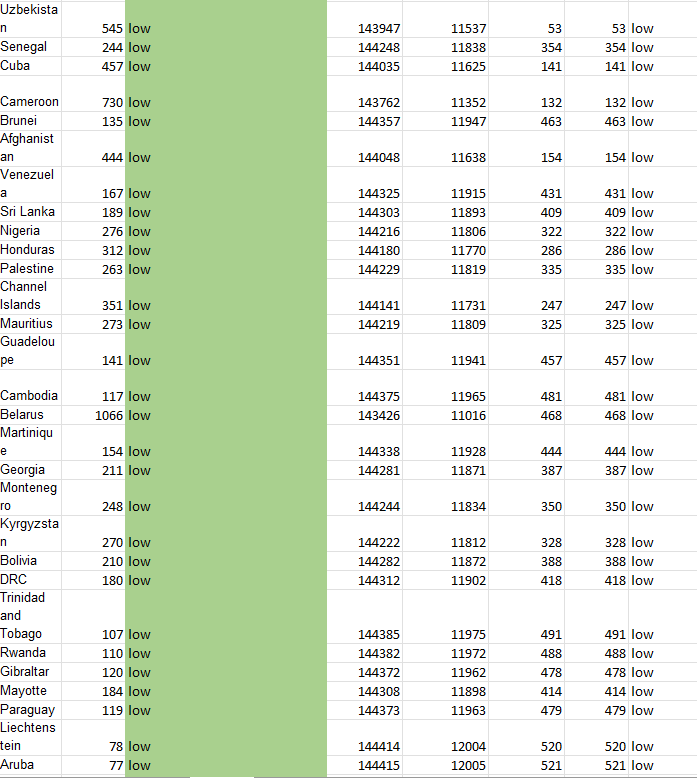
Description automatically generated

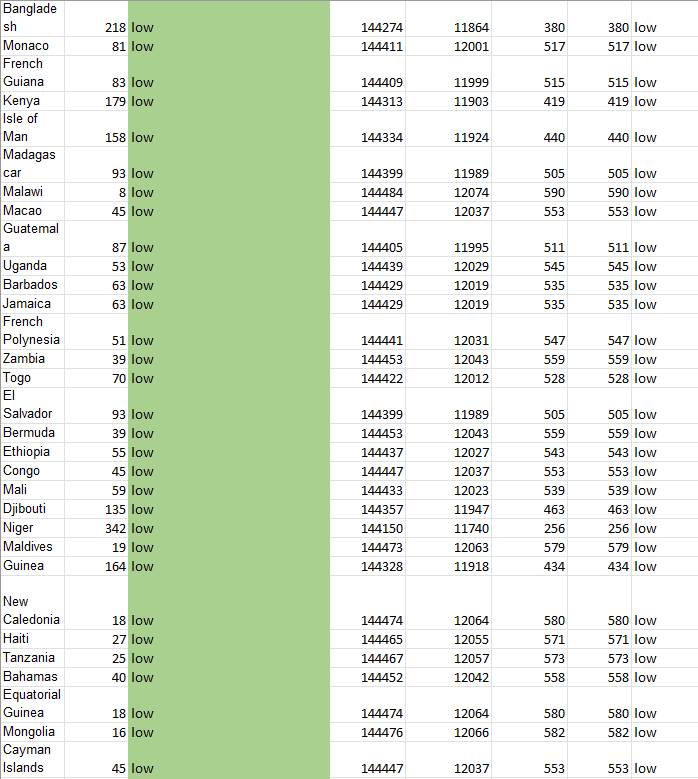
* *The grouping using k-means in full details:*

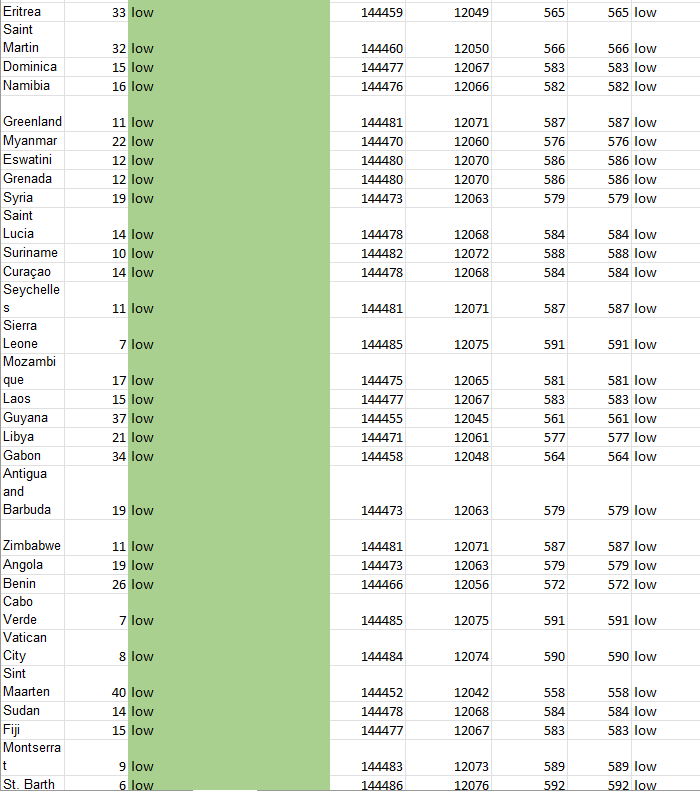


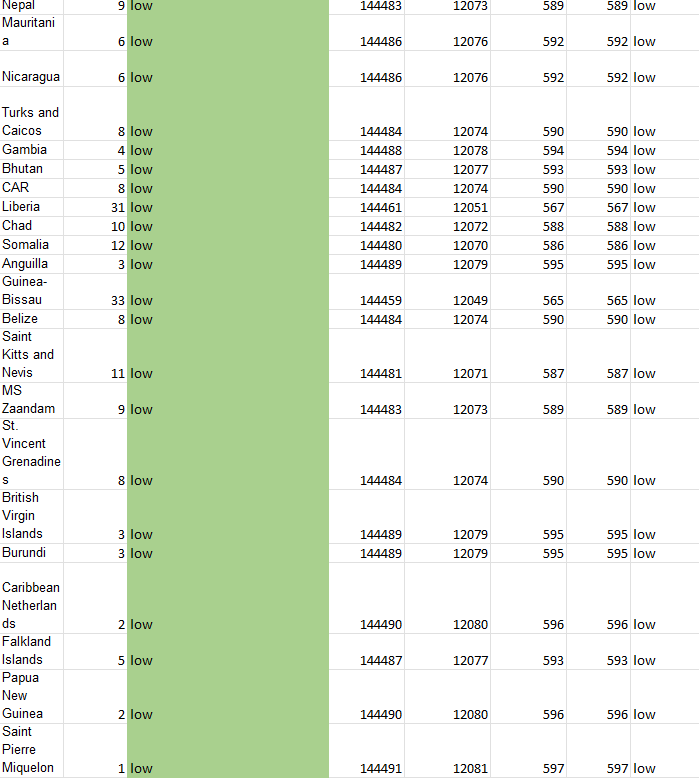


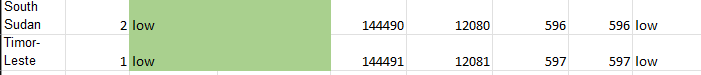












1. After you cluster the data, chooses two different countries from each level, and build a dendrogram to cluster the country into groups using a single link.

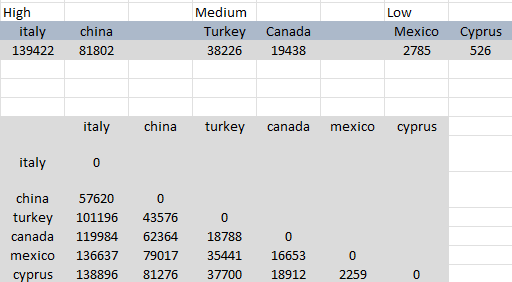
* *Countries chosen from each group for dendrogram are as follows:*

*High: Italy, China*

*Medium: Turkey, Canada*

*Low: Mexico, Cyprus*

* *Total cases for each country, and distance matrix:*

**

* *Dendrogram:*

