Covid -19 Vaccine distribution:

Rajesh kumar singh

16.05.2020

1. **Introduction**
   1. **Background**

**Coronavirus disease 2019** (**COVID-19**) is a communicable disease caused by [severe acute respiratory syndrome coronavirus 2](https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome_coronavirus_2) (SARS-CoV-2). It was first identified in December 2019 in [Wuhan](https://en.wikipedia.org/wiki/Wuhan), China, and has since spread globally, resulting in [an ongoing pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic).

Common symptoms include [fever](https://en.wikipedia.org/wiki/Fever), [cough](https://en.wikipedia.org/wiki/Cough), [fatigue](https://en.wikipedia.org/wiki/Fatigue), [shortness of breath](https://en.wikipedia.org/wiki/Shortness_of_breath), and [loss of smell](https://en.wikipedia.org/wiki/Anosmia) and [taste](https://en.wikipedia.org/wiki/Ageusia). While the majority of cases result in mild symptoms, some progress to [acute respiratory distress syndrome](https://en.wikipedia.org/wiki/Acute_respiratory_distress_syndrome) (ARDS), [multi-organ failure](https://en.wikipedia.org/wiki/Multiple_organ_dysfunction_syndrome), [septic shock](https://en.wikipedia.org/wiki/Septic_shock), and [blood clots](https://en.wikipedia.org/wiki/Blood_clot).

The effect of corona virus is different in old and younger people. The death rate of older people is more as compared to people of age in between 0 to 10 years.

As the doses for adult and children are different, we need to supply the vaccine as soon as possible and in the correct volume. For this we need to know the density of adult and minors population in the region.

* 1. **Problem**

How to make this COVID19 vaccine available to the people as soon as possible. The region of NY has around 8.3 million population, out of which 40 % of population is of people aged 50 and above and 60 % of are below 50 years.

We need to provide them the desirable amount of vaccines.

* 1. **Interest**

The US government is very interested in accurate prediction of vaccine requirements. Others who are interested are the hospitals or the medical centres where the vaccine supply would happen.

1. **Data acquisition and cleaning**
   1. **Data sources**

The population of NY city data can be obtained from the municipal corporation. But here, I have used the customer data from one of the link provided in the earlier chapter.

<https://cocl.us/customer_dataset>

* 1. **Data cleaning**

We get data from different sources. We need to combine these dataset into one . After that we need to remove the duplicates from it, we need to fill the empty cells with some suitable data like mean values , we need to convert the datatype of the columns to into to perform some mathematics into it.

* 1. **Feature section**

Once the data cleaning is done we need to find out the number of columns and rows in the dataset, that we can do by applying the shape function in the data frame. I have taken only the required column from the data set like age of the customer, address of the customer.

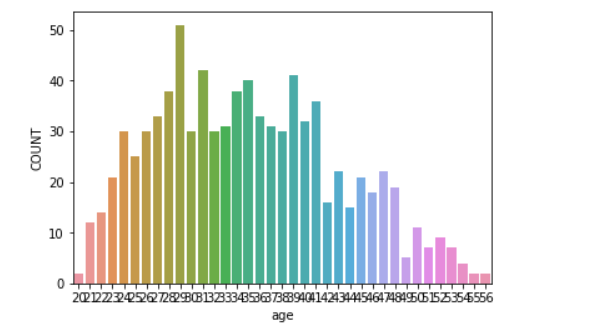
I have applied the group by clause to obtain the desired results.

The library I have used to connect to the database is ibm\_db . An on cloud db2 database is used . I created a db2 hosed service and under that I loaded the customer\_segmention data on customer table and did some modification on the data types, like converted small int column to integer.

Then, I set the credentials for ssl and without ssl connections. Then I used the library pandas, and ibm\_db\_dbi. I created a data frame to obtain the result in a form of table.

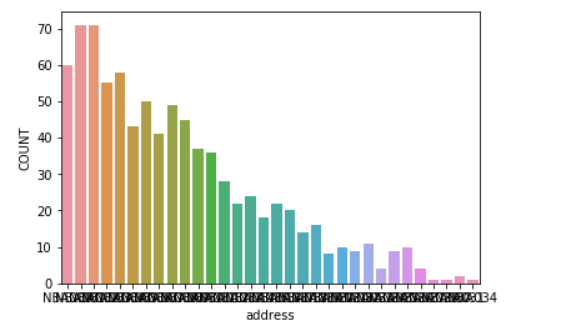
1. **Exploratory Data Analysis**

* 1. **Calculation of age count using group by clause**



We need to work on only the columns which we need. In my case I will need the customer id, customer age and the customer address. Customer ID is the unique column, customer age will give me the age of every customer and customer address will tell me about their area position in NY. Once we know the population of older and the younger people in each area , it will become easier for us to distribute vaccine to this region

* 1. **Calculating the population based on the address/location.**



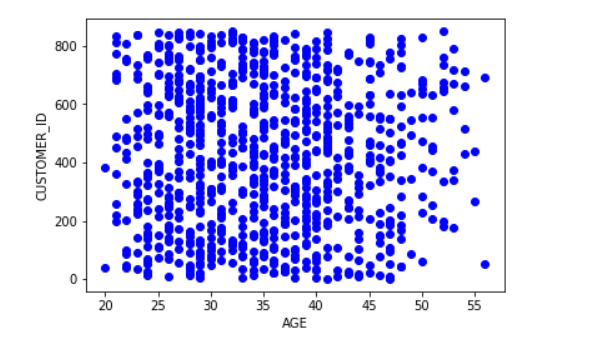
The above chat shows the count of population at different location. This can be obtained by applying the group by clause on population of different age and different location.

1. **Predictive Modelling**

There are two types of supervised learning, regression and classification, that can he used to predict that how fast the virus is going to spread at different parts of the region.

Regression is a process of predicting continues values. Regression model would provide additional information on the amount of people getting affected , while classification is the process of predicting discrete class labels or categories. Classification model focus on the probability of a person might affect.

Plotting the linear regression: lets plot the linear regression between customer\_id and there age.



I used the group by function with count to find out the population of aged and younger people at any particular part of NY region.

1. **Future directions**

* Creating more robust module .
* Frequent update of the new born babies in the dataset.
* Would visualize the population density in a map using Folium .
* Would get the location of hospitals in each location.

1. **Conclusion**

In this study I analysed relation ship between the population and location. How densely an area in NY city is populate. How can we distribute the vaccines and the medicines depending upon the affected area. What is the movement of people at any area. At what rate the virus spread could happen .