Ben Postell

Dr. Su Dong

Data Wrangling

May 10, 2023

Business Analytics Project

After being presented with data revolving around COVID-19, I performed an analysis to understand key questions that are expanded upon below. In my opinion, females that are middle-aged are at the most risk for COVID-19 during the beginning months of the year.

**Insights of VAERS data**

*Data Comprehension; Most common location, age, and vaccine/symptoms*

With 217,337 people being diagnosed with COVID-19 in the year 2022, it is important to understand demographics revolving around that statistic. Finding the most common location, age and vaccine administered are trends that can help predict at-risk people. Using SQL to perform the analysis, the top ages are between 60 to 70 years old. This shows that middle-aged individuals are more susceptible to COVID-19 and should take more precautions. The top three states with COVID-19 are CA, MI, and CO. There is no correlation between states and at-risk individuals since the average ages with COVID-19 in those states are outside of the 60 to 70 year old range. Overall, the most popular vaccine administered is PFIZER/BIONTECH which is also the dominant vaccine for middle-aged individuals. The tables and information can be used to create a plan for allocation of resources to locations with the most cases as well as help people understand if they are at risk or not. Knowing common symptoms can help identify COVID-19 and prevent it from going undetected. Looking at the symptoms listed, an analysis of the most common symptoms was able to be done. Words that are not symptoms but were used when describing how the individual felt or simply too vague, were filtered out by skipping to the next most common symptom. The top three symptoms were blood, fatigue, and headache.

*Trends of COVID-19 Over Each Month for The Year 2022*

**Icon

Description automatically generated with medium confidence***Chart, line chart

Description automatically generated*The seasonality of COVID-19 was a necessary analysis to test external influencers that were not in the dataset provided. In the figures below, it is obvious that as the months progress, COVID-19 decreases. This could be from many contributing factors, such as the months become warmer, people are able to spend time outside which in turn, limits the spread of COVID-19.

*Chart, histogram

Description automatically generated*With major holidays circa the time COVID-19 cases are at the summit, it is expected to see a high influx. With the holiday’s past, there is a decline of travel and exposure which reflects in the tables above. Using seaborn to gain another perspective, the graphs in figure 3 reinforce the idea that cases decline after the holidays. External factors play a large role in influencing the data. These variables are hard to measure but assumptions can be made to complete an analysis.

*Correlation between Gender and Age Groups*

Diving deeper into the correlation between Male and Female, figure 4 and 5 both show that females are more likely to have COVID-19. The points are separated by four age groups: Adult, Middle\_Age, Old and Young. Young represents all individuals under the age of 18, Adult is between the ages of 18 to 40, Middle\_Age is between 40 to 70 years and Old is 70 years and up.



While there are many different questions that can still be answered, this analysis answered key questions that can be used to protect and plan for the future. Based off this data from 2022, females that are middle aged should take the most precautions during the beginning months of the year to reduce the spread of COVID-19.