**Data Visualization**

**MIS 6380.0U1**

***Group-01***

*Kannupriya*

*Heli Deepak Madani*

*Mayur Kumar Tikmani*

**Executive Summary:**

We are visualizing this dataset (Austin Code COVID19 complaint cases) to contemplate the time frame when the complaint was opened during COVID. The decrease in the complaints registered with time, along with the setting where, majority of complaints got registered, and the regulations that were neglected which led to the case getting registered are some of the other objectives we are going to visualize.

We want to envision as the number of complaints of social distancing, overcrowding and face covering increase the number of total COVID cases also increase in Austin, Texas following the increase in total deaths.

**INDEX****:**

Data Description

Data Cleaning

General Introduction

Insights and findings

Conclusion

**Data Description:**

* The dataset contains the following information: the type of complaint filed, the location where the complaint was filed, as well as the date and time. Each row represents a complaint, and there are 13 columns and 10,288 rows in all. Information on the number of deaths and COVID cases in Austin is provided in a supplementary dataset.
* We will pay close attention to the column Open date, which will track the number of complaints received in various months of 2020 and 2021. Our dataset's third column "Type of Business," gives insight into the environment in which COVID-19-related complaints were made. The dataset's fourth column Type of Complaints lists the rules that were breached and resulted in the case being recorded. Depending upon the severity of their violations, the Austin, Texas, government has specified several actions that will be taken against them. Address, council district, location, gives us a brief idea about the location where the complaints were registered.
* Primary dataset: Austin Code COVID-19 Complaint Cases

Link: <https://data.austintexas.gov/Public-Safety/Austin-Code-COVID-19-Complaint-Cases/4p54-9544/data>

* Secondary dataset: Daily Counts

Link: <https://atc-covid19data-austin.hub.arcgis.com/datasets/daily-counts-public-view/explore>

**Data** **Cleaning:**

1. For data cleaning, we removed the columns ConstructionReferral, CitationIssued, CouncilDistrict and ParcelID. These columns have a lot of empty records and they don’t align with the insights that we were trying to find.
2. We checked the interval between the OpenDate and CloseDate in majority of the cases it’s one day. There were 32 empty records in CloseDate so we imputed data with the date one day after the OpenDate.
3. The Outcome column had 39 blanks, so we replaced them with No Violation Found which formed the majority of the records in that column.
4. In the secondary dataset, we have 144 empty records in the TotalCases column the data is right skewed so we replaced the records with the median which was 75490.

**General Introduction:**

With the domestic and international spread of (COVID-19), much attention has been given to estimating the pandemic risk, fatalities, effect on business. We propose the novel application, to provide a visualization of the COVID-19 – related complaints filed in Austin, Texas for alleged violations of the CDCs’ STANDARD in YEAR 2020 TO 2022.

We want to envision as the number of complaints of social distancing, overcrowding and face covering increase the number of total COVID cases also increase in Austin, Texas following the increase in total deaths.

To visualize it we are using Tableau, which provides the features like cleaning, organizing, and visualizing the data. It is easier to create interactive visual analytics in the form of dashboards. This dashboard will make it easier for non-technical analysts and end users to convert the data into understandable ones. Tableau supports a wide range of data sources like Excel, Microsoft Access, and databases.

**Insights and findings:**

* Our first insight was that the maximum complaints werefiled in April 2020, thus our first hypothesis, which claimed that the most complaints were filed in March 2020, failed. To put the notion to the test, we built a computed variable called No. of complaints. In this representation, we've used a bar chart with years on the x-axis and no complaints on the y-axis. And the year 2020 had the most complaints when compared to the year 2021 and 2022.

Now, we are interested to know which month had the highest no. of complaints and for that we have utilized Line chart, which shows April 2020 had the highest no. Of cases registered, after which there was a progressive decline in COVID complaint instances.

* Second, we discovered that overall social distancing accounts for the greatest proportion of complaint files, demonstrating that our second hypothesis was correct.  To test this, we used a pie chart and grouped related type of complaint. As we can see, the greatest proportion was social distancing (37.18 percent), followed by face covering (36.28 percent). More than half of the complaints are about social distancing and face hiding.

However, in 2020, the number of complaints about social distancing outnumbered the number of complaints about face covering, but in 2021, the maximum number of complaints were about face covering and the number of complaints about social distancing reduced by 25.85 percent. We experienced an unexpected outcome in 2022, where neither face-covering nor social distancing was a major issue. The most common complaint was regarding eviction. It is apparent that when the number of complaints declined in 2022, the nature of the complaints changed dramatically.

* Third, we discovered that other types of businesses filed the most complaints, indicating that the hypothesis that residents filed the most complaints failed. To test our hypothesis, we have utilized line chart and animated the number of complaints made by different types of businesses to easily discover the components that substantially contribute to each category. Here in the animation, it is clear that there were major fluctuations in the no. of complaints filed by different businesses. However, the Highest no. of complaint recorded was in April 2020 by residents. But the category classified as OTHERS was the one who filed the maximum no. of total complaints over the period 2020-2022.

And to visualize that we have utilized a tree map which shows the %age of total no. of complaints which is 10287 filed by the type of business. It is notable that the category of others submitted the most complaints, accounting for (45.21 percent), followed by residents (24.73 percent). Because most of the complaints came from abrupt places, they cannot be classified in specific settings and are thus they are classified as Others.

* Fourth, we found that there is no consistent relation between the number of social distancing complaints and the number of covid instances. Although they did not move in the same direction in 2020, they did have a negative relationship after April 2020. As a result, Hypothesis 4 failed to demonstrate a positive association between the two.  To illustrate this, we utilized a line chart in which the number of covid instances and social distancing complaints are represented as lines over the month of 2020-2021. While the number of social distancing instances increased from March to April 2020, the overall number of cases fluctuated for total Covid Cases. However, as we approach May 2020, we can vividly observe the negative relationship between the two.

* Fifth, we discovered that no violation was detected in the majority of the complaints. In this situation, we used the Packed bubble chart, and we can see that the max. outcome of the complaints, no violation was found which accounts for 72.70 percent, followed by the outcome of delivering education linked to COVID-19, which accounts for 15.38 percent.

Now, we've attempted to use a bar chart, with types of complaints on the x-axis and each bar classified by kind of outcome. Though we know that in the majority of cases, no violation was detected, and the most common form of complaint was face hiding and social distancing. No violation accounts 31.30% in face covering and for social distancing no violation accounts to 26.03%. As per the data suggested it is evident that no major action was taken against those who violated the social distancing and face covering,  

Graphical user interface, table

Description automatically generated

Conclusion:

* As the Covid-19 cases increases in 2020, the highest cases were recorded in April 2020, and with time the cases decrease.
* We were more interested to know if there is a positive relationship with covid-19 complaints about social distancing, face covering etc, with no. of covid cases recorded, as with more violations of CDC rules, the cases should increase, but while the number of incidents of social distance grew from March to December 2020, the total number of Covid Cases varied. However, as we approach March 2021, we can clearly see the two's unfavorable relationship.
* There is no direct proportional relationship between the two and there must be other factors, we should introduce to know what CDC rules should consider to corelate with increase of COVID-19 cases like population density, covid testing frequency etc.