MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING

Analyze dining-out behavior of people during pre and post Covid-19 pandemic

By Department of Data Science

Het Bhatt-D003
Taher Gari - D007
Sneha Pradeep - D012
Fahad Syed - D016

Abstract

Purpose: This paper aims to analyze changes in the behavior of people, before and after the onset of coronavirus disease 2019 (COVID-19) by evaluating the dine-out reviews.The outbreak of COVID-19 in late 2019 has had a huge impact on people's daily life. Many restaurant businesses have been greatly affected by it. Consumer's preferences for the restaurant industry in US counties have changed, such as environmental hygiene, variety of dishes, and service methods.
Design/methodology/approach: Restaurant reviews data was acquired for the U.S. counties: pre and post-pandemic.
Originality: This study represents a pioneering attempt to investigate the impact of COVID-19 on restaurant businesses.
Practical implications: These results characterize the regional restaurant industry's resilience to COVID-19 and identify particularly vulnerable areas that may require supplementary assistance to recover.

Introduction

The outbreak of coronavirus disease 2019 (COVID-19) has shaken the world in an unprecedented way. This global catastrophe has thus far claimed the U.S. as its largest victim. The latency of the virus and a relative lack of countermeasures in the early stages of transmission resulted in more than 700,000 deaths and 4.73 cr cases in the U.S. as of late August 2021. COVID-19 is highly transmissible between humans. Its spread can be most effectively controlled by reducing interpersonal contact through physical distancing. In the U.S., physical distancing has been implemented on two authoritative levels: the federal government and state/local governments. COVID-19 was declared a national emergency on March 1, 2020, when the domestic outbreak was still in incubation. Then, as infections began to escalate, state and local governments started to implement preventive community measures such as school closures, shutdowns of non-essential businesses, and stay-at-home orders.

COVID-19 has dramatically impacted the restaurant industry nationwide. Although outbreak severity varies by U.S. region and community, nearly every state and local government has enforced physical distancing orders by banning restaurants' dine-in services. While these intervention efforts have minimized personal interaction and alleviated the virus's spread, they have greatly threatened the restaurant industry's survival. Directory and review site Yelp in July announced that 60% of restaurants that closed down completely during the pandemic had marked those closures as permanent for a total of nearly 16,000 restaurants by July 24. The New York Times on March 20 reported that industry analysts were predicting that two thirds of restaurants would not survive, and as many as 75 percent of independents. It was estimated that restaurant sales had decreased 47% nationwide from March 1-22nd. Downloads for grocery apps spiked during this period as more consumers were choosing to cook their own food instead of ordering out. Downloads for Instacart had spiked 215% from February 14th-March 15th. Forbes on March 19 estimated the job losses in the restaurant industry to be in the millions. The National Restaurant Association estimated probable job losses to be five to seven million. Industry experts on March 18 forecast that $225 billion in direct losses and a full economic impact of $675 billion because of additional revenues generated elsewhere in the economy when customers patronize restaurants.

Alternative off-premise models, such as drive-thru and food delivery, were implemented by restaurants to offset the impact. However, these service models offered an edge for fast food restaurants that had already equipped with digital infrastructure and drive-thru windows; full-service restaurants, however, were not able to quickly adapt to the change, and consequently, took the biggest financial hit.
In the hospitality literature, scholars have explored numerous factors associated with restaurant demand, such as food quality, price, location, and online reviews. Among these factors, crisis events represent an uncommon yet critical force, some of which have brought grave consequences to the restaurant industry. To the best of our knowledge, no empirical research has considered the analysis on dining-out behavior of people in US restaurants during the pandemic period.

To fill this research gap, we gathered restaurant reviews data from two big perspectives: preCovid and postCovid. An analysis is then shown which is used to investigate the impacts of the pandemic and reviews on restaurant businesses.First and foremost, this study represents a pioneering effort to unveil the consequences of COVID-19, an unforeseen global pandemic, in the restaurant industry.
We will be comparing the trends PreCovid and PostCovid so that we can analyze to which extent the restaurant industry was affected. We plan to find out the dining-out trends all over the US and also calculate the YOY impact in the restaurant's business remaining open.

Research Questions

From this project we aim to analyse better the following problems and come up with possible explanations:

1. How has the customer's preferences for the restaurant industry in the US counties changed?
2. What are the particularly vulnerable areas that restaurants may require supplementary assistance to recover from the aftermath of the pandemic?

Objectives

1. To analyze changes in the restaurant industry in the U.S.A. before and after the onset of coronavirus disease 2019 (COVID-19) by evaluating the dine-out reviews.
2. To analyze the number of reviews statewise with respect to the number of cases and deaths in the U.S.A.
3. To analyze review count based on the pre-Covid and post-Covid timeline.
4. To find out how many restaurants are still open compared to the pre-covid time period.

Literature Review

Customer review analysis is particularly important for the restaurant industry, given the perishability of related products and services. Predicting trends and exploring the characteristics that influence consumer demand for restaurants can inform restaurant management, strategic planning, and marketing. Such information can also facilitate foodservice related policy making and evaluation among governments and other public sectors.

The restaurant industry is particularly vulnerable to epidemic crises, as it relies on human interaction and gatherings. Several factors are playing a crucial role to which most of the businesses are getting affected, factors like perceived personal susceptibility to infectious diseases; perceived seriousness of the consequences of infection; a perceived sense of control, referring to one's belief in whether they can take effective actions to combat a health risk; personal motivation to maintain good health; and personal risk aversion, referring to one's tendency to avoid uncertainty or risky choices . Moreover, individuals' risk perceptions of an epidemic can be shaped by risk communication via mass media and authorities such as the World Health Organization (WHO) and the federal government. These sources may miscommunicate health-related information such as the confirmed infections, facts about a disease, and whether effective therapeutic or countermeasures exist . Risk communication can, therefore, influence the perceived severity and controllability of health risks, even leading to widespread public panic in some cases. In the context of COVID-19, restaurant dining could be associated with high perceived personal susceptibility due to the highly infectious nature of human interactions in a relatively enclosed, less ventilated space.

Methodology

Data details:

Review dataset

Sort fields: Data source order | Show aliases | Show hidden fields | 1,000 rows

Busi...	Name	State	City	Postal...	Latit...	Longit...	Stars	Revie...	Is Open	Categories	Review Id	User Id	Custo...	Us...	Fu...	Cool	Text	Dat
2WRCc...	Yook Ko...	British ...	Vanc...	V5N 5E4	49.2627	-123.0564	4.000...	74	1	Restaurants...	t2I2CLHd7...	iBRhL3g...	1	0	0	0	Decen...	1
oug5bL...	Gourdo...	Texas	Austin	78704	30.2454	-97.7804	4.000...	1,534	1	Karaoke, Pu...	gcJK-cQHFr...	qcpg7m...	4	1	0	0	I've b...	3
imUIM...	Donut E...	Massa...	Med...	02052	42.1937	-71.2903	5.000...	42	1	Food, Bakeri...	wvtTSr4GG...	os_bbA...	2	0	0	0	The la...	2
mP1EdI...	Barcelo...	Massa...	Boston	02116	42.3449	-71.0705	4.500...	1,097	1	Spanish, Ba...	twUhcgH6...	cIEAaBN...	4	1	0	0	I cam...	1
4UjU7F...	Shan-A-...	Massa...	Broo...	02446	42.3480	-71.1296	4.000...	426	1	Nightlife, In...	vPFrIWYmx...	evqTtzt...	1	1	0	1	Ate th...	3
pKvT7...	Virgils ...	Georgia	Colle...	30337	33.6537	-84.4499	4.000...	252	1	Soul Food, C...	rJ0cbBwdX...	yG8mW...	1	6	0	0	Very v...	1
v1UzkU...	Gus's ...	Texas	Austin	78701	30.2635	-97.7417	4.500...	2,490	1	Soul Food, S...	dJXzdV468...	4ghiDth...	4	0	0	0	One o...	1
VaZPH...	Norwo...	Massa...	Norw...	02062	42.1935	-71.2015	4.500...	19	1	Restaurants...	bObYIm3T4...	mcWU_...	5	0	0	0	The d...	1
J8Ha6y...	Barley ...	Texas	Austin	78757	30.3413	-97.7384	4.500...	1,086	1	Restaurants...	cdz56-rOcu...	IVgyib2...	5	1	0	1	Alway...	3

Us Covid dataset

Sort fields: Data source order | Show aliases | Show hidden fields | 1,000 rows

Date	County	State	Fips	Cases	Deaths
02-06-2020	Ashtabula	Ohio	39,007	340	35
02-06-2020	Athens	Ohio	39,009	18	1
02-06-2020	Auglaize	Ohio	39,011	75	3
02-06-2020	Belmont	Ohio	39,013	438	13
02-06-2020	Brown	Ohio	39,015	28	1
02-06-2020	Butler	Ohio	39,017	919	32
02-06-2020	Carroll	Ohio	39,019	31	3

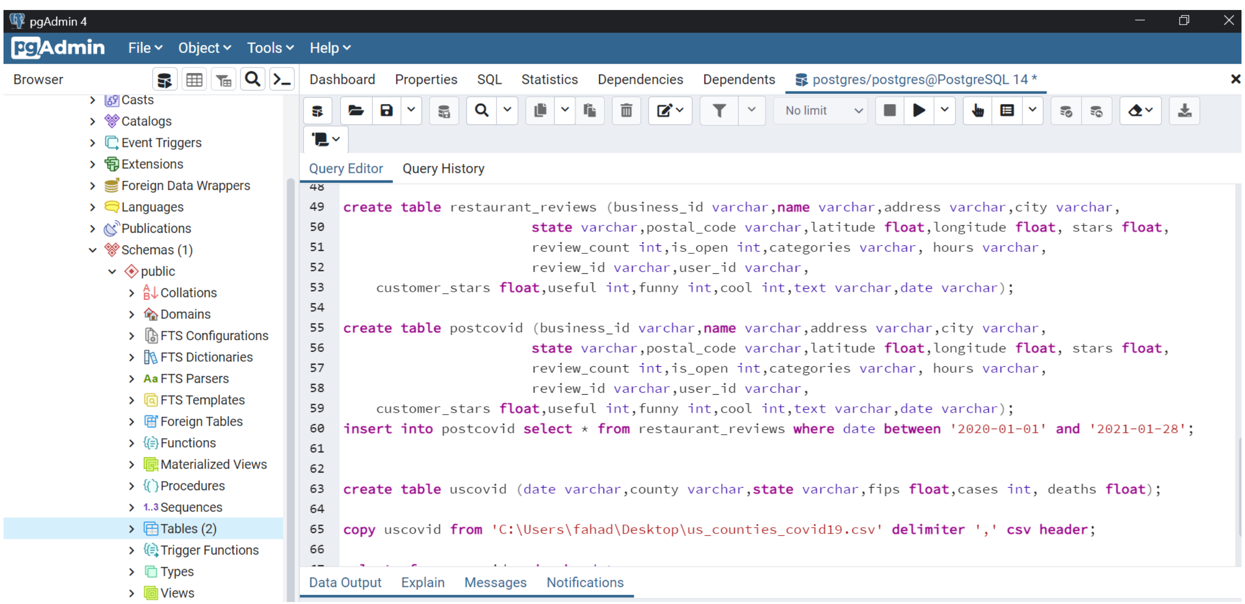
Collecting/Preparing Data:

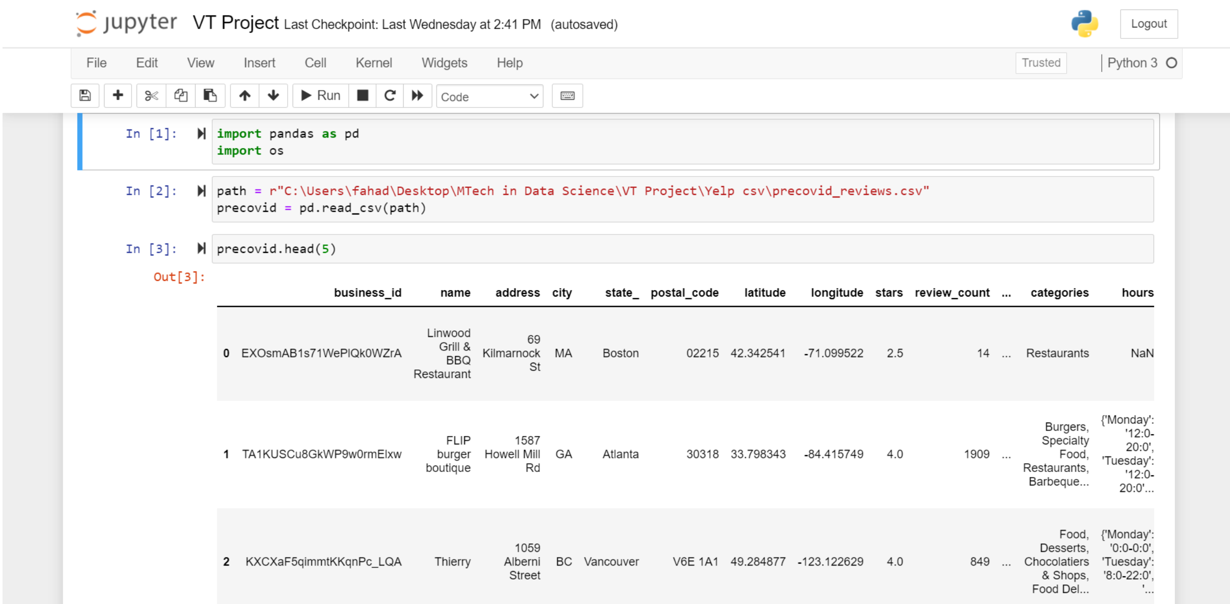
We Separated the data into two main parts named pre-COVID and post-COVID Data. Since covid19 was discovered in the US during the early month of January 2020, we have the Restaurant customer reviews generated from 2020-01-01 to 2021-01-31. For the pre-COVID data, the time period that we selected is of the past 24 years of data which is before the pandemic started.

The following steps are carried out in order to fetch out the relevant data for this study.

1. The first step was to convert the json type file into csv file form and was done using python on jupyter notebook using pandas library.
2. Postgresql (pgadmin 4) used for categorizing data based on restaurants and to make a new table to get the restaurants and reviews as a single output table keeping Business ID as primary key.
3. During this process of forming a Restaurant reviews data table inner join, create table, like, between & copy format was used. The difficulty faced during this procedure was to access the csv file and the error which occurred while importing the data.

Data that we have after carrying out the task, for pre-COVID analysis is *51,72,981* and for pandemic duration i.e. post-COVID is *4,00,295*. We also have *8,00,437* observations of US covid data mentioning the number of cases and deaths based on the country which will be used for further analysis.





jupyter VT Project Last Checkpoint: Last Wednesday at 2:41 PM (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [1]:
```
import pandas as pd
import os
```

In [2]:
```
path = r"C:\Users\fahad\Desktop\MTech in Data Science\VT Project\Yelp csv\precovid_reviews.csv"
precovid = pd.read_csv(path)
```

In [3]:
```
precovid.head(5)
```

Out[3]:

	business_id	name	address	city	state_	postal_code	latitude	longitude	stars	review_count	...	categories	hours
0	EXOsmAB1s71WePlQk0WZrA	Linwood Grill & BBQ Restaurant	69 Kilmarnock St	MA	Boston	02215	42.342541	-71.099522	2.5	14	...	Restaurants	NaN
1	TA1KUSCu8GkWP9w0rmElxw	FLIP burger boutique	1587 Howell Mill Rd	GA	Atlanta	30318	33.798343	-84.415749	4.0	1909	...	Burgers, Specialty Food, Restaurants, Barbeque...	{'Monday': '12:0-20:0', 'Tuesday': '12:0-20:0'...
2	KXCXaF5qimmtKKqnPc_LQA	Thierry	1059 Alberni Street	BC	Vancouver	V6E 1A1	49.284877	-123.122629	4.0	849	...	Food, Desserts, Chocolatiers & Shops, Food Del...	{'Monday': '0:0-0:0', 'Tuesday': '8:0-22:0', '...

Data Cleaning:

We started with the US covid dataset, in this we again use pgAdmin since the review dataset focussed on 12 states in the USA we list out all the relevant states needed for the analysis from the uscovid dataset using a in function in sql. Now the dataset reads *2,52,857* observations. Further checking the dataset we found out that there were no null values.

Now to the post-COVID review dataset we found out that there were *13,224* null values.

These null values were in the column of restaurants address (*992*) and restaurants opening and closing hours (*12,232*). Since we didn't require these columns for analysis we used the drop function to omit them from the dataset.

Further on looking at the dataset we found out that some of the cities were in UpperCase while others were in LowerCase. To overcome this we first used a unique function to check the unique values of the city in the dataset (*394 cities*) and later on used the str.title() function to make them distinct (*361 cities*). Also, state abbreviation was converted into full name using replace function with dictionary. All this was important for Tableau Analysis so as to read the state and city column for mapping.

There were *2,29,910* null values in the pre-COVID dataset.

These null values were in the column of restaurants address (*13,409*) and restaurants opening and closing hours (*2,16,501*). The cleaning procedure is the same as what we did in the post-COVID dataset.

Jupyter Data Cleaning Last Checkpoint: 10/02/2021 (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

400294	dmkDZKPsK8ImwFuLiFQ0Zw	Planted PDX	4390 SW Garden Home Rd	OR	Portland	97219	45.465756	-122.721888	4.5	24	...	Food Trucks Restaurants Food, Vega

400295 rows × 21 columns

```
In [4]:  post_covid['address'].isnull().sum()
Out[4]:  992

In [5]:  post_covid['postal_code'].isnull().sum()
Out[5]:  4

In [6]:  post_covid['address'].isnull().values.any()
Out[6]:  True

In [7]:  post_covid['hours'].isnull().sum()
Out[7]:  12232

In [8]:  post_covid = post_covid.drop(columns = ['address','hours'])

In [9]:  post_covid
Out[9]:
```

	business_id	name	state_	city	postal_code	latitude	longitude	stars	review_count	is_open	categories	
0	2WRCcQATOe_Em0k61T6kvQ	Yook Korean Grilled BBQ & Bistro	BC	Vancouver	V5N 5E4	49.262732	-123.056371	4.0	74	1	Restaurants, Barbeque, Korean	t2

Jupyter Data Cleaning Last Checkpoint: 10/02/2021 (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

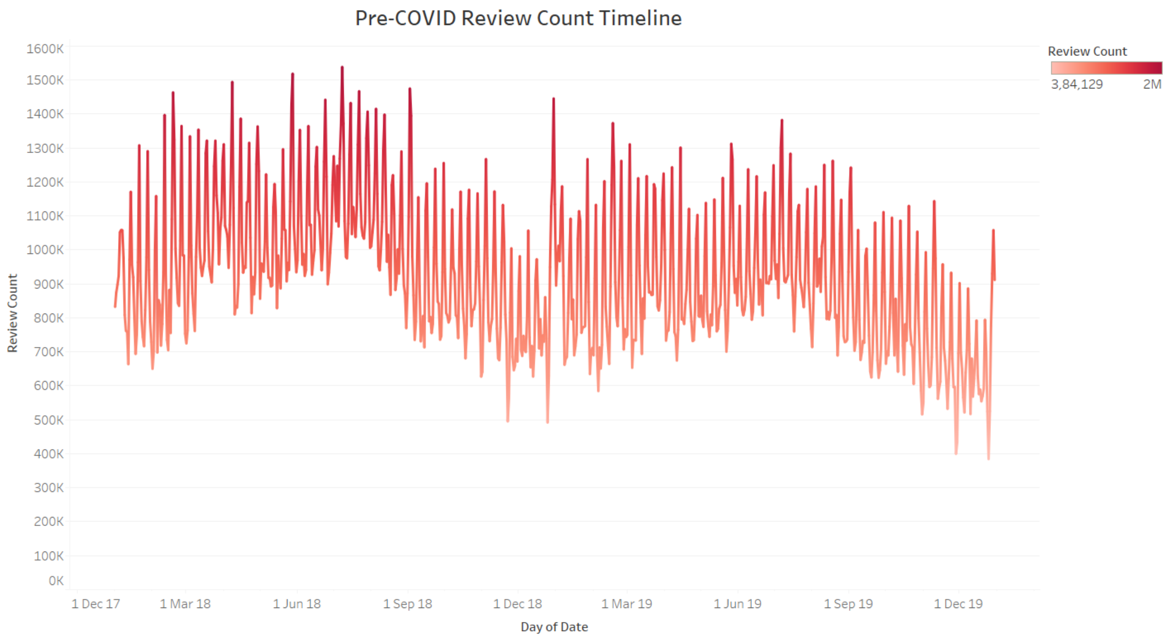
```
        'Franklin', 'Ladner', 'Columbus ', 'Bee Caves', 'Lake Monroe',
        'St Johns', 'Union Park', 'Newton Corner', 'Chuluota',
        'Urbancrest', 'Beaverton Or', 'Holb', 'Berkshire', 'Waban',
        'Gunbarrel', 'Vinings', 'Wollaston', 'Liberty', 'Hollbrook',
        'Aloma', 'Lago Vista', 'Arvada', 'Marble Cliff', 'Fulton',
        'College Park Ga', 'Austin Texas', 'San Marcos'], dtype=object)

In [14]:  post_covid.city.nunique()
Out[14]:  361

In [15]:  post_covid.dropna(subset=["postal_code"], axis = 0, inplace = True)

In [38]:  post_covid.isnull().sum().sum()
Out[38]:  0

In [17]:  post_covid.state_.unique()
Out[17]:  array(['BC', 'TX', 'MA', 'GA', 'FL', 'OH', 'OR', 'WA', 'CO', 'NH', 'WY',
                 'VA', 'ABE', 'KY'], dtype=object)

In [35]:  post_covid['state_'] = post_covid['state_'].str.replace('KY','Kentucky')

In [82]:  post_covid.state_.unique()
Out[82]:  array(['British Columbia', 'Texas', 'Massachusetts', 'Georgia', 'Florida',
                 'Ohio', 'Oregon', 'Washington', 'Colorado', 'New Hampshire',
                 'Wyoming', 'Virginia', 'ABE', 'Kentucky'], dtype=object)

In [36]:  post_covid.state_.unique()
Out[36]:  array(['British Columbia', 'Texas', 'Massachusetts', 'Georgia', 'Florida',
                 'Ohio', 'Oregon', 'Washington', 'Colorado', 'New Hampshire',
                 'Wyoming', 'Virginia', 'Kentucky'], dtype=object)

In [37]:  post_covid
Out[37]:
```

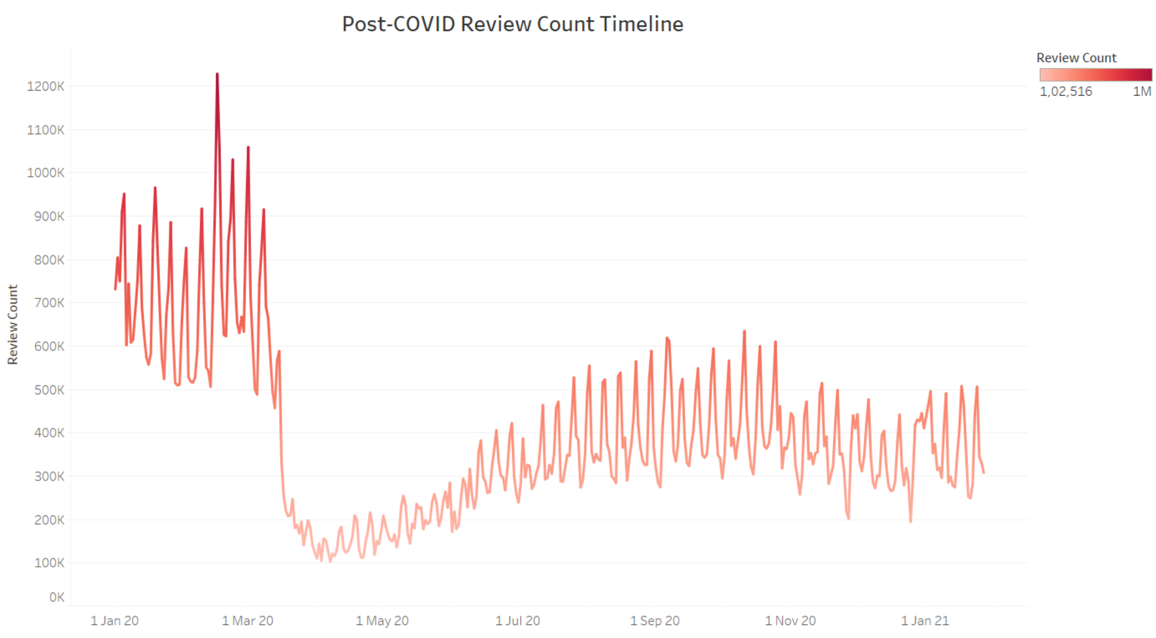
	business_id	name	state_	city	postal_code	latitude	longitude	stars	review_count	is_open	categorie
0	2WRCcQATOe_Em0k61T6kvQ	Yook Korean Grilled BBQ	British Columbia	Vancouver	V5N 5E4	49.262732	-123.056371	4.0	74	1	Restaurant Barbequ

Data Analysis:

1. Change in Review Count

We started with the analysis of the review count timeline. This was done on Tableau by showing the change in the review count by number of days during the pre-COVID and post-COVID duration. We can clearly analyze that the effect of Covid was seen in the USA during the mid of March 2020 as there is a big decline in the daily number of review count when compared to the pre covid timeline.





2. Correlation

Second analysis was to check the correlation between review count, cases and deaths. For this we combined the dataset of post-COVID and USA covid using groupby function in python state wise and formed a new dataset having column review count, cases and deaths.

For this we used the Pearson correlation.

After analysis

a)Review Count and Cases

The Pearson Correlation Coefficient is 0.7006474748704289 with a P-value of P = 0.011149912451599887

b)Review Count and Deaths

The Pearson Correlation Coefficient is 0.7699278263748179 with a P-value of P = 0.0033990330082125137

We can clearly see that both the correlation result shows a moderate positive correlation with a very significant p value.

Even though we see a positive correlation we cannot conclude that there is a relationship between how cases and deaths affect the review count. In order to check this we used SLR and MLR for further analysis since correlation doesn’t imply causation.

3. SLR vs MLR

Simple linear regression vs Multiple linear regression

For SLR we used review count as dependent variable and cases as independent variable. Such that when we perform linear regression in python using OLS or sklearn package we get the value of R and adj R along with the p-value.

Now for MLR we used review count as a dependent variable and cases, deaths and is_open (restaurants that were open post-COVID) as independent variables. We followed the same process for this as well and saw a significant change in the value of R and adj R along with the p-value.

Further on we have also shown a distribution plot for SLR and MLR showing the difference between both the models. These models help us to understand the change in the distribution of graphs by comparing the actual value to the fitted value.

Results after analysis.

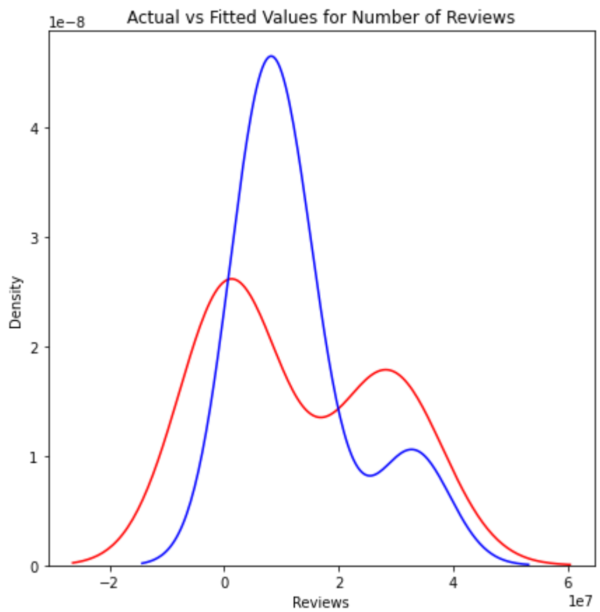
Left graph shows the distribution graph for SLR.

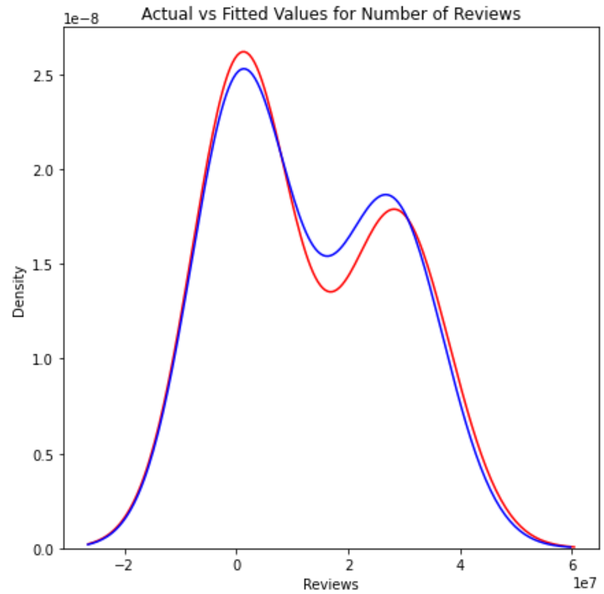
Right graph shows the distribution graph for MLR.

Hence we can conclude that here MLR provides us with much better insights as compared to SLR due to an additional independent variable (is_open) in MLR.

As R square shows the goodness of fit in MLR better and also we have a statistically significant p value for is_open we can say that there is a relationship between the number of review counts going down due to the fact of restaurants getting closed instead of the number of cases and deaths.

However we can also see that there is also a relationship between number of cases and the number of review counts in SLR but we have a moderate goodness of fit.





- **SLR:**

```
==============================================================================
Dep. Variable:           review_count   R-squared:                       0.491
Model:                            OLS   Adj. R-squared:                  0.440
Method:                 Least Squares   F-statistic:                     9.643
Date:                Sun, 07 Nov 2021   Prob (F-statistic):             0.0111
Time:                        00:28:24   Log-Likelihood:                -210.31
No. Observations:                  12   AIC:                             424.6
Df Residuals:                      10   BIC:                             425.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       3.802e+06   4.23e+06      0.899      0.390   -5.62e+06    1.32e+07
cases          0.2580      0.083      3.105      0.011       0.073       0.443
------------------------------------------------------------------------------
```

- **MLR:**

```
==============================================================================
Dep. Variable:           review_count   R-squared:                       0.945
Model:                            OLS   Adj. R-squared:                  0.924
Method:                 Least Squares   F-statistic:                     45.52
Date:                Sun, 07 Nov 2021   Prob (F-statistic):           2.26e-05
Time:                        00:55:35   Log-Likelihood:                -197.00
No. Observations:                  12   AIC:                             402.0
Df Residuals:                       8   BIC:                             403.9
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -7.061e+05   1.71e+06     -0.413      0.690   -4.65e+06    3.23e+06
cases          0.0705      0.066      1.065      0.318      -0.082       0.223
deaths        -2.9587      3.796     -0.779      0.458     -11.712       5.794
is_open      438.9523     61.695      7.115      0.000     296.684     581.220
==============================================================================
```

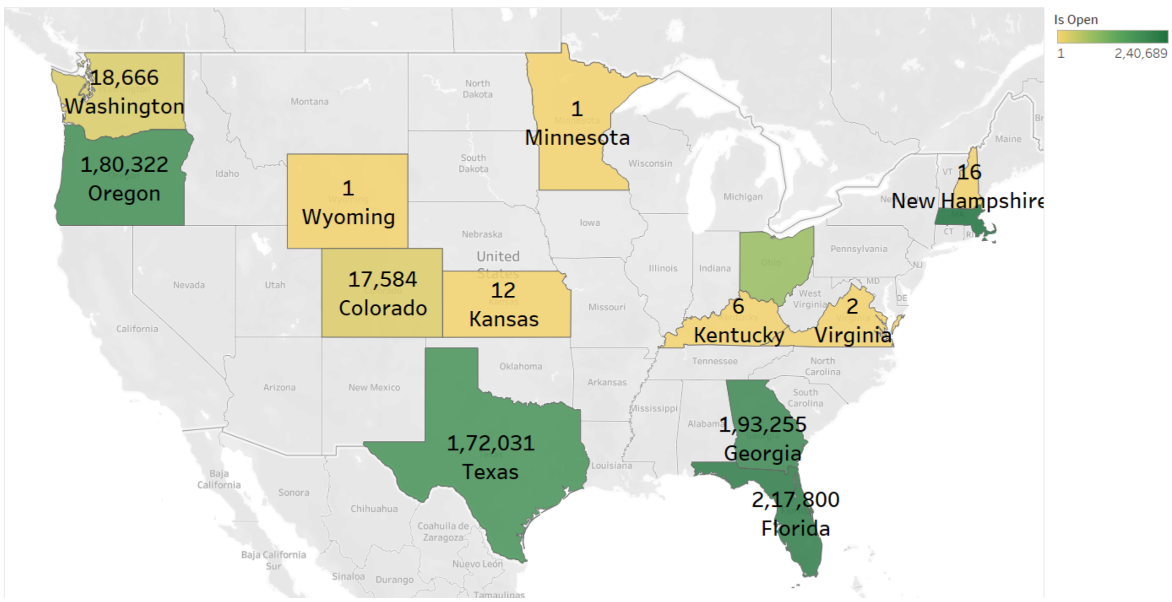
4. Restaurants getting closed post-COVID

In this we used the column *is_open* from the dataset where it shows the counts of restaurants that were open in the USA based on year. We used the map chart for understanding the state wise restaurants getting closed in pandemic as we compare it before the pandemic took place.

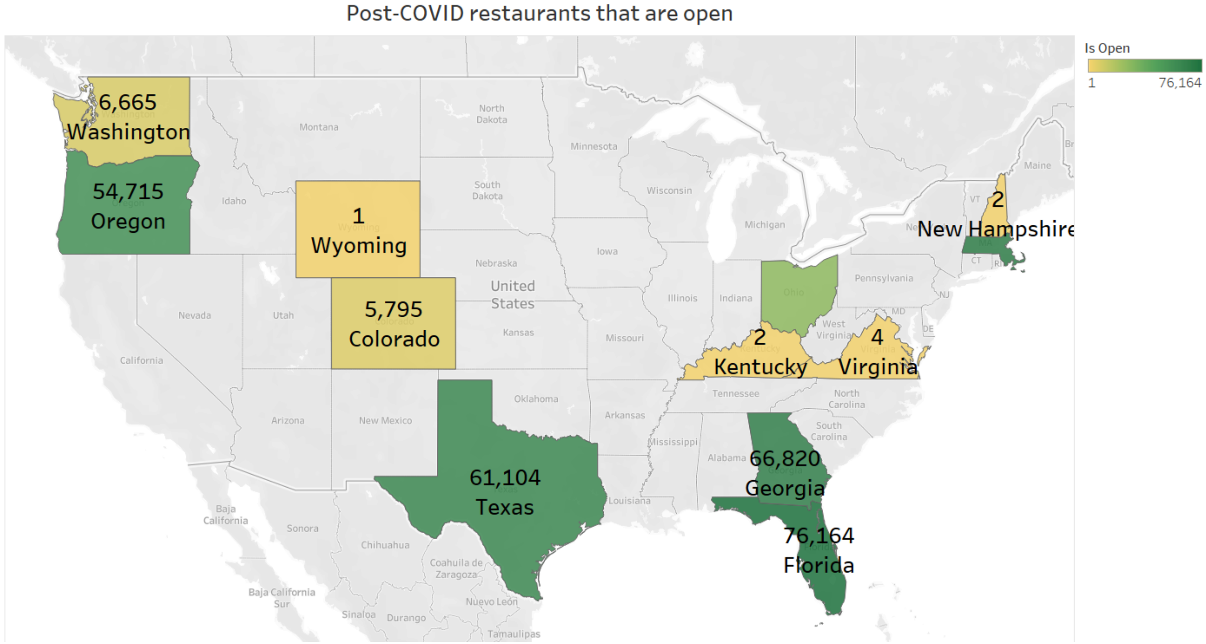
Through this visualization we can clearly say that the restaurants business were greatly affected by the COVID situation and hence can relate to the decline in the review count of the restaurants.

We can clearly see that Georgia, Oregon, Florida and Texas were hugely affected by the COVID pandemic situation. Also, there were no restaurants that were open in Kansas and Minnesota post pandemic.

Pre-COVID restaurants that are open



Post-COVID restaurants that are open

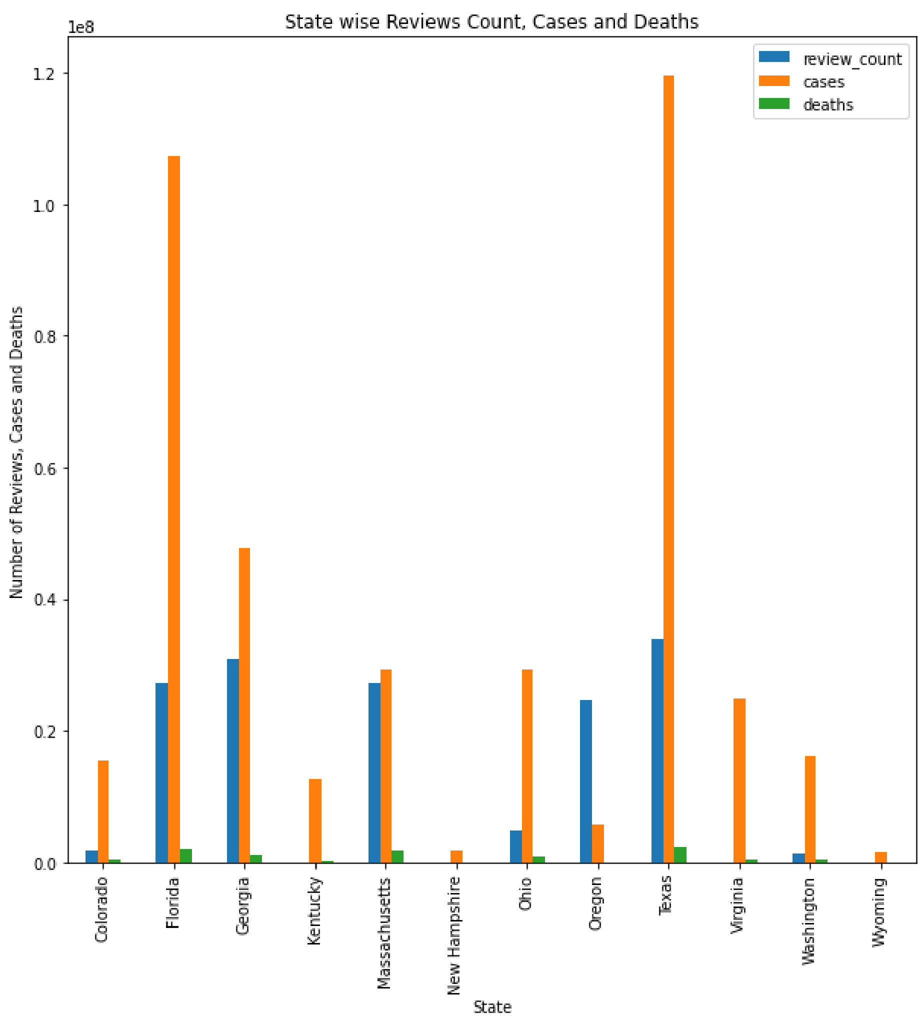


5. State wise review count, cases and deaths

As we can see that the most number of cases was found in the state Texas followed by Florida and Georgia.

We can say that the review count in Georgia and Massachusetts were not affected by the pandemic crisis as compared to other states.

Wyoming and New Hampshire are the only states with zero number of deaths during the post-COVID period. Which we can truly see as the review count in these states are the least as well as the number of cases..



References

1. Lasek, A., Cercone, N. and Saunders, J. , "Smart Restaurants: Survey on customer demand and sales forecasting", in Obaidat, M. S. and Nicopolitidis, P. (Eds.) Smart cities and homes: : Key enabling technologies, Elsevier, Cambridge, MA, pp. 361-386.
2. Kim, E.-J. and Geistfeld, L. V. , "Consumers' restaurant choice behavior and the impact of socio-economic and demographic factors", Journal of Foodservice Business Research, Vol. 6 No. 1, pp. 3-24.
3. Beck, M. J., Fifer, S. and Rose, J. M. , "Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty", Transportation Research Part B: Methodological, Vol. 89, pp. 149-167
4. Champion, V. L. and Skinner, C. S. , "The health belief model", Health behavior and health education: Theory, research, and practice, 4th ed, Jossey-Bass, San Francisco, CA, US, pp. 45-65.