

```
In [137]: M import numpy as np
import pandas as pd
import plotly
import requests
import json
from plotly.graph_objs import Figure, Histogram, Layout, Scatter, Bar, Pie
plotly.tools.set_credentials_file(username="RAQH", api_key="s611ovOHT3qmMF1dLbii")
import scipy as stats
import seaborn as sns
plotly.offline.init_notebook_mode(connected=True)
```

Read "Open Table Data" scraped from "Open Table" site

```
In [138]: M openitable = pd.read_csv('openitable_data.csv') #import
```

```
In [139]: M openitable.head()
```

Out[139]:

	restaurant	location	cuisine	price	food_rating	service_rating	ambience_rating	value_rating	num_reviews	recommendation_percentage
0	Gabriel Kreuther	Midtown West	Contemporary French / American	\$50 and over	4.8	4.8	4.8	4.3	1269.0	91% K
1	Davo's Northern Italian Steakhouse - Manhattan	Midtown East	Steakhouse	31 to 50	4.5	4.5	4.4	4.2	947.0	93% D Chef
2	Empellon Taqueria	West Village	Mexican	31 to 50	4.5	4.6	4.4	4.1	1581.0	81% Alex
3	Añejo - Tribeca	TriBeCa - Downtown	Mexican	31 to 50	4.5	4.3	4.3	4.1	422.0	82% C
4	Yopparai	Lower East Side	Izakaya	\$50 and over	4.8	4.8	4.7	4.4	310.0	95%

Drop duplicate restaurant and covert it to lower case

```
In [140]: M openitable = openitable.drop_duplicates(subset=['restaurant'])
openitable['restaurant'] = openitable['restaurant'].str.lower()
openitable.head()
```

Out[140]:

	restaurant	location	cuisine	price	food_rating	service_rating	ambience_rating	value_rating	num_reviews	recommendation_percentage
0	gabriel kreuther	Midtown West	Contemporary French / American	\$50 and over	4.8	4.8	4.8	4.3	1269.0	91% K
1	davo's northern italian steakhouse - manhattan	Midtown East	Steakhouse	31 to 50	4.5	4.5	4.4	4.2	947.0	93% D Chef
2	empellon taqueria	West Village	Mexican	31 to 50	4.5	4.6	4.4	4.1	1581.0	81% Alex
3	añejo tribeca	TriBeCa - Downtown	Mexican	31 to 50	4.5	4.3	4.3	4.1	422.0	82% C
4	yopparai	Lower East Side	Izakaya	\$50 and over	4.8	4.8	4.7	4.4	310.0	95%

Second Data - Read restaurant violation data from NYC data

- Rename "dba" column to "restaurant".
- Convert "restaurant" column to lower case

```
In [141]: M from sodapy import Socrata
client = Socrata("data.cityofnewyork.us", None)
results = client.get("43nn-pn8j", limit=381633)
WARNING:root:Requests made without an app_token will be subject to strict throttling limits.
```

```
In [170]: M results_df_RAW = pd.DataFrame.from_records(results)
results_df_RAW.to_csv('RAWpractice_data.csv', index=True)
results_df_RAW = results_df_RAW.rename(columns = {"dba" : "restaurant"})
results_df_RAW['restaurant'] = results_df_RAW['restaurant'].str.lower()
results_df_RAW = results_df_RAW[results_df_RAW.boro == 'MANHATTAN']
results_df_RAW.head()
```

Out[170]:

	action	boro	building	camis	critical_flag	cuisine_description	restaurant	grade	grade_date	inspection_date	inspection_type
33	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Critical		irish dj reynolds pub and restaurant	A	2019-06-06T00:00:00.000	2019-06-06T00:00:00.000	Cycle Inspection / Initial Inspection 2122
34	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Not Critical		irish dj reynolds pub and restaurant	A	2019-06-06T00:00:00.000	2019-06-06T00:00:00.000	Cycle Inspection / Initial Inspection 2122
35	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Critical		irish dj reynolds pub and restaurant	A	2018-05-16T00:00:00.000	2018-05-16T00:00:00.000	Cycle Inspection / Initial Inspection 2122
36	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Not Critical		irish dj reynolds pub and restaurant	A	2018-05-16T00:00:00.000	2018-05-16T00:00:00.000	Cycle Inspection / Initial Inspection 2122
37	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Not Critical		irish dj reynolds pub and restaurant	A	2018-05-16T00:00:00.000	2018-05-16T00:00:00.000	Cycle Inspection / Initial Inspection 2122

```
In [172]: M results_df = results_df_RAW[results_df_RAW.zipcode != 'N/A']
```

```
results_df.head(2)
```

```
Out[172]:
```

	action	boro	building	camis	critical_flag	cuisine_description	restaurant	grade	grade_date	inspection_date	inspection_type
33	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Critical	Irish dj reynolds pub and restaurant	A	2019-06-06T00:00:00.000	2019-06-06T00:00:00.000	Cycle Inspection / Initial Inspection	2122
34	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Not Critical	Irish dj reynolds pub and restaurant	A	2019-06-06T00:00:00.000	2019-06-06T00:00:00.000	Cycle Inspection / Initial Inspection	2122

Drop duplicate restaurant and load to another DF

```
In [174]: results_df = results_df.drop_duplicates(subset=['zipcode','restaurant','street'])  
results_df.head()
```

```
Out[174]:
```

	action	boro	building	camis	critical_flag	cuisine_description	restaurant	grade	grade_date	inspection_date	inspection_type	
33	Violations were cited in the following area(s).	MANHATTAN	351	30191841	Critical	Irish dj reynolds pub and restaurant	A	2019-06-06T00:00:00.000	2019-06-06T00:00:00.000	Cycle Inspection / Initial Inspection	2122	
80	Violations were cited in the following area(s).	MANHATTAN	1	40359480	Critical	American 1 east 6th street kitchen	A	2018-09-28T00:00:00.000	2018-09-28T00:00:00.000	Cycle Inspection / Initial Inspection	2122	
164	Violations were cited in the following area(s).	MANHATTAN	730	40362264	Critical	American p & s deli grocery	A	2019-05-13T00:00:00.000	2019-05-13T00:00:00.000	Cycle Inspection / Initial Inspection	2122	
191	Violations were cited in the following area(s).	MANHATTAN	18	40362274	Critical	American angelika film center	A	2019-01-02T00:00:00.000	2019-01-02T00:00:00.000	Cycle Inspection / Initial Inspection	2122	
228	Violations were cited in the following area(s).	MANHATTAN	60	40362715	Critical	Sandwiches/Salads/Mixed Buffet	the country cafe	Z	2019-06-07T00:00:00.000	2019-06-07T00:00:00.000	Cycle Inspection / Re-inspection	2122

Find the count of "Critical and Not Critical Violation per restaurant"

```
In [180]: results_df_Criticalraw = results_df_RAW[results_df_RAW.critical_flag=='Critical']  
results_df_Critical = results_df_Criticalraw.groupby(['restaurant','zipcode']).agg({'critical_flag':'count'})  
results_df_Critical = results_df_Critical.rename(columns = {"critical_flag" : "CriticalViolationCount"})  
results_df_Critical = results_df_Critical.reset_index(level=['restaurant','zipcode'])  
results_df_Critical.head()
```

```
Out[180]:
```

	restaurant	zipcode	CriticalViolationCount
0	#1 chinese restaurant	10033	13
1	\$1 pizza	10001	4
2	& pizza	10001	6
3	& pizza	10003	6
4	'cesca	10023	9

```
In [181]: results_df_NotCriticalraw = results_df_RAW[results_df_RAW.critical_flag=='Not Critical']  
results_df_NotCritical = results_df_NotCriticalraw.groupby(['restaurant','zipcode']).agg({'critical_flag':'count'})  
results_df_NotCritical = results_df_NotCritical.rename(columns = {"critical_flag" : "NotCriticalViolationCount"})  
results_df_NotCritical = results_df_NotCritical.reset_index(level=['restaurant','zipcode'])  
results_df_NotCritical.head()
```

```
Out[181]:
```

	restaurant	zipcode	NotCriticalViolationCount
0	#1 chinese restaurant	10033	9
1	& pizza	10001	4
2	& pizza	10003	5
3	'cesca	10023	2
4	'essen	10017	3

Merge Critical and Not Critical Violation Dataframes

```
In [185]: results_df_Criticalall = pd.merge(results_df_Critical,results_df_NotCritical ,  
how='left',left_on=['restaurant','zipcode'],right_on=['restaurant','zipcode'])  
results_df_Criticalall['NotCriticalViolationCount'] = results_df_Criticalall['NotCriticalViolationCount'].fillna(0)  
results_df_Criticalall.head()
```

```
Out[185]:
```

	restaurant	zipcode	CriticalViolationCount	NotCriticalViolationCount
0	#1 chinese restaurant	10033	13	9.0
1	\$1 pizza	10001	4	0.0
2	& pizza	10001	6	4.0
3	& pizza	10003	6	5.0
4	'cesca	10023	9	2.0

Merge Open table Data and Restaurant Violation Data (without duplicates)

```
In [182]: Openable_ResulttableMerge = pd.merge(openable[['restaurant','cuisine','price', 'food_rating','dining_style','dress_code']],  
results_df[['restaurant','grade','zipcode']],  
how='left',left_on=['restaurant'],right_on=['restaurant'])  
Openable_ResulttableMerge.head()
```

```
Out[182]:
```

	restaurant	cuisine	price	food_rating	dining_style	dress_code	grade	zipcode
0	gabriel kreuther	Contemporary French / American	\$50 and over	4.8	Fine Dining	Business Casual	A	10036
1	davio's northern italian steakhouse - manhattan	Steakhouse	\$110+0	4.5	Fine Dining	Smart Casual	NaN	NaN
2	empillon taqueria	Mexican	\$110+0	4.5	Casual Dining	Smart Casual	NaN	NaN
3	a'zlejo tribeca	Mexican	\$110+0	4.5	Casual Dining	Smart Casual	NaN	NaN
4	yopparai	Izakaya	\$50 and over	4.8	Casual Elegant	Casual Dress	A	10002

Merge above table with critical violation count to do analysis

```
In [188]: M analysis_df = pd.merge(OpenTable_ResultSet, results_df_Criticalall,
                                how='left', left_on=['restaurant', 'zipcode'], right_on=['restaurant', 'zipcode'])
```

Out[188]:

	restaurant	cuisine	price	food_rating	dining_style	dress_code	grade	zipcode	CriticalViolationCount	NotCriticalViolationCount
0	gabriel kreuther	Contemporary French / American	\$50 and over	4.8	Fine Dining	Business Casual	A	10036	2.0	4.0
1	davio's northern italian steakhouse - manhattan	Steakhouse	\$110+0	4.5	Fine Dining	Smart Casual	NaN	NaN	NaN	NaN
2	empellon taqueria	Mexican	\$110+0	4.5	Casual Dining	Smart Casual	NaN	NaN	NaN	NaN
3	a\zejo tribeca	Mexican	\$110+0	4.5	Casual Dining	Smart Casual	NaN	NaN	NaN	NaN
4	yopparai	Izakaya	\$50 and over	4.8	Casual Elegant	Casual Dress	A	10002	13.0	4.0

Drop rows which does not have data(nan) and rename column names to be lower case to load to SQL

```
In [190]: M analysis_df = analysis_df.dropna()
```

Out[190]:

	restaurant	cuisine	price	food_rating	dining_style	dress_code	grade	zipcode	CriticalViolationCount	NotCriticalViolationCount
0	gabriel kreuther	Contemporary French / American	\$50 and over	4.8	Fine Dining	Business Casual	A	10036	2.0	4.0
4	yopparai	Izakaya	\$50 and over	4.8	Casual Elegant	Casual Dress	A	10002	13.0	4.0
5	db bistro moderne	French	\$110+0	4.5	Casual Dining	Smart Casual	A	10036	5.0	9.0
10	patsy's pizzeria	Pizzeria	\$30 and under	4.2	Casual Dining	Casual Dress	A	10017	9.0	9.0
11	patsy's pizzeria	Pizzeria	\$30 and under	4.2	Casual Dining	Casual Dress	A	10065	12.0	10.0

```
In [191]: M analysis_df_final = analysis_df.reset_index()
del analysis_df_final['index']
analysis_df_final = analysis_df_final.rename (columns = {"CriticalViolationCount" : "criticalviolationcount",
                                                "NotCriticalViolationCount" : "notcriticalviolationcount"})
analysis_df_final.head()
```

Out[191]:

	restaurant	cuisine	price	food_rating	dining_style	dress_code	grade	zipcode	criticalviolationcount	notcriticalviolationcount
0	gabriel kreuther	Contemporary French / American	\$50 and over	4.8	Fine Dining	Business Casual	A	10036	2.0	4.0
1	yopparai	Izakaya	\$50 and over	4.8	Casual Elegant	Casual Dress	A	10002	13.0	4.0
2	db bistro moderne	French	\$110+0	4.5	Casual Dining	Smart Casual	A	10036	5.0	9.0
3	patsy's pizzeria	Pizzeria	\$30 and under	4.2	Casual Dining	Casual Dress	A	10017	9.0	9.0
4	patsy's pizzeria	Pizzeria	\$30 and under	4.2	Casual Dining	Casual Dress	A	10065	12.0	10.0

Connect to local database

```
In [192]: M from sqlalchemy import create_engine
rds_connection_string = "postgres:Password@localhost:5432/restaurant_db"
engine = create_engine(f'postgresql:///{rds_connection_string}')
```

Use pandas to load csv converted DataFrame into database

```
In [193]: M analysis_df_final.to_sql(name='restaurant', con=engine, if_exists='append', index=False)
```

Confirm data has been added by querying the customer_location table

```
In [194]: M pd.read_sql_query('select * from restaurant fetch first 5 rows only', con=engine)
```

Out[194]:

	restaurant	cuisine	price	food_rating	dining_style	dress_code	grade	zipcode	criticalviolationcount	notcriticalviolationcount
0	gabriel kreuther	Contemporary French / American	\$50 and over	5	Fine Dining	Business Casual	A	10036	2	4
1	yopparai	Izakaya	\$50 and over	5	Casual Elegant	Casual Dress	A	10002	13	4
2	db bistro moderne	French	\$110+0	5	Casual Dining	Smart Casual	A	10036	5	9
3	patsy's pizzeria	Pizzeria	\$30 and under	4	Casual Dining	Casual Dress	A	10017	9	9
4	patsy's pizzeria	Pizzeria	\$30 and under	4	Casual Dining	Casual Dress	A	10065	12	10

Analysis

New table can be used to

1. we can analyse the rating of the restaurant along with the number of critical/non-critical violations to understand any significant pattern. Ratings are given by people, whereas Violations are recorded by health-inspectors. We can analyse to see if the two parameters have any mutual relationship.
2. we can further analyse the grade of a restaurant outlet with the rating of the chain to uncover any pattern associated with these two parameters.
3. In the same table, we can view the dining style for the restaurant. This data can be viewed in conjunction with grade to see if a particular dining style is associated with a particular grade violations to see if a particular dining style is associated with a more violations - ex. pickup joints tend to have more violations as compared to fine dining.
4. we can repeat the same analysis as mentioned in point 3 in conjunction to cuisine instead of dining style as well.

In []: