Introduction

In statistics, exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task. Exploratory data analysis was promoted by John Tukey to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. EDA is different from initial data analysis (IDA), which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA.

Content

There is a list of over 10,000 fast food restaurants provided by Datafiniti's Business Database. The dataset includes the restaurant's address, city, latitude and longitude coordinates, name, and more.

What You Can Do with This Data

You can use this data to rank cities with the most and least fast food restaurants across the U.S.. E.g.:
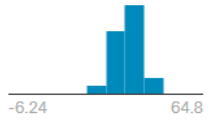
- Cities with the most and least McDonald's per capita
- Fast food restaurants per capita for all states
- Fast food restaurants with the most locations nationally
- Major cities with the most and least fast food restaurants per capita
- Small cities with the most fast food restaurants per capita
- States with the most and least fast food restaurants per capita
- The number of fast food restaurants per capita

Columns used from the file

- City
- Country
- Keys
- Latitude
- Longitude
- Name
- Postal Code

- Province
- Websites

Data From FastfoodRestaurant.CSV File

	address	city	country	keys	# latitude	# longi
	9934 unique values	Cincinnati 1% Las Vegas 1% Other (2773) 98%	US 100%	10000 unique values	-6.24 — 64.8	-159.38
1	324 Main St	Massena	US	us/ny/massena/324mainst/-1161002137	44.9213	
2	530 Clinton Ave	Washington Court House	US	us/oh/washingtoncourthouse/530clintonave/-791445730	39.53255	
3	408 Market Square Dr	Maysville	US	us/ky/maysville/408marketsquaredr/1051460804	38.62736	
4	6098 State Highway 37	Massena	US	us/ny/massena/6098statehighway37/-11610021 37	44.95008	
5	139 Columbus Rd	Athens	US	us/oh/athens/139columbusrd/990890980	39.35155	
6	4182 Tonya Trl	Hamilton	US	us/oh/hamilton/4182tonyatrl/-1055723171	39.4176	

Run Info

Run Info

Succeeded	True	Run Time	9.8 seconds
Exit Code	0	Queue Time	0 seconds
Docker Image Name	/python(Dockerfile)	Output Size	0
Timeout Exceeded	False	Used All Space	False
Failure Message			

Importing Libraries & Reading CSV File

```python
import numpy as np
import pandas as pd
import string
from plotly.offline import init_notebook_mode, iplot
from plotly.graph_objs import *
init_notebook_mode(connected=True)
print(os.listdir("../input"))
df = pd.read_csv('../input/FastFoodRestaurants.csv')
```

```python
df.head()
```

	address	city	country	keys	latitude	longitude	name	postalCode
0	324 Main St	Massena	US	us/ny/massena/324mainst/-1161002137	44.92130	-74.89021	McDonald's	13662
1	530 Clinton Ave	Washington Court House	US	us/oh/washingtoncourthouse/530clintonave/-7914...	39.53255	-83.44526	Wendy's	43160
2	408 Market Square Dr	Maysville	US	us/ky/maysville/408marketsquaredr/1051460804	38.62736	-83.79141	Frisch's Big Boy	41056
3	6098 State Highway 37	Massena	US	us/ny/massena/6098statehighway37/-1161002137	44.95008	-74.84553	McDonald's	13662
4	139 Columbus Rd	Athens	US	us/oh/athens/139columbusrd/990890980	39.35155	-82.09728	OMG! Rotisserie	45701

Data Exploration

```python
print('Total no of columns:',df.shape[1])
print('Total no of rows:',df.shape[0])
```

```
Total no of columns: 10
Total no of rows: 10000
```

```python
print('countries are:',df.country.unique())
print('city are:',df.city.unique())
```

```
countries are: ['US']
city are: ['Massena' 'Washington Court House' 'Maysville' ... 'Chatham' 'Miles City'
 'East Chicago']
```

```python
df.name.value_counts()[:10]
```

```
McDonald's            1886
Burger King           1154
Taco Bell              873
Wendy's                731
Arby's                 518
KFC                    421
Domino's Pizza         345
Subway                 322
SONIC Drive In         226
Hardee's               192
Name: name, dtype: int64
```

```python
df.name=df.name.apply(lambda x: x.lower())
df.name=df.name.apply(lambda x:''.join([i for i in x
                      if i not in string.punctuation]))

df.name.value_counts()[:10]
```

```
mcdonalds            2105
burger king          1154
taco bell             873
wendys                734
arbys                 520
kfc                   441
subway                404
dominos pizza         359
sonic drive in        248
jack in the box       201
Name: name, dtype: int64
```
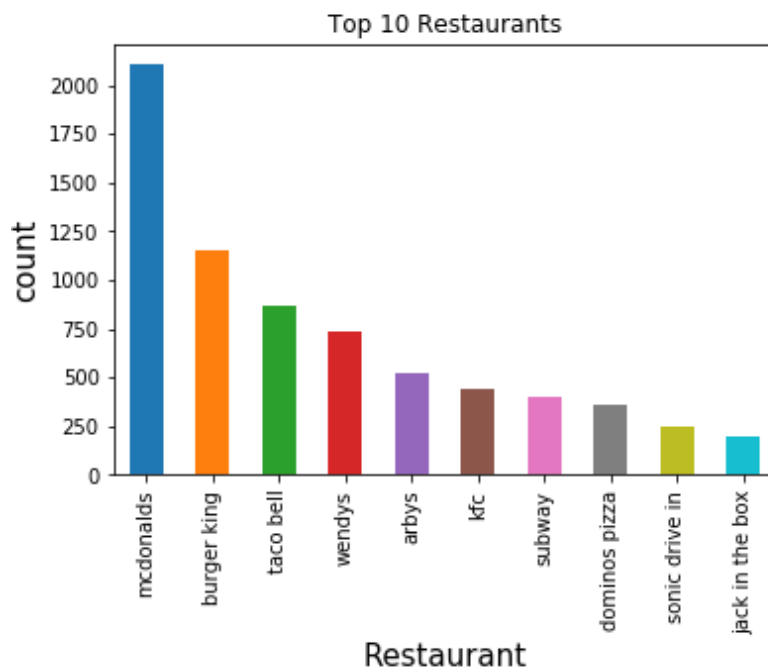
Data Visualization

Histograms

1.Top 10 Restaurants using EDA

```
nameplt=df.name.value_counts()[:10].plot.bar(title="Top 10 Restaurants")
nameplt.set_xlabel('Restaurant',size=15)
nameplt.set_ylabel('count',size=15)
```
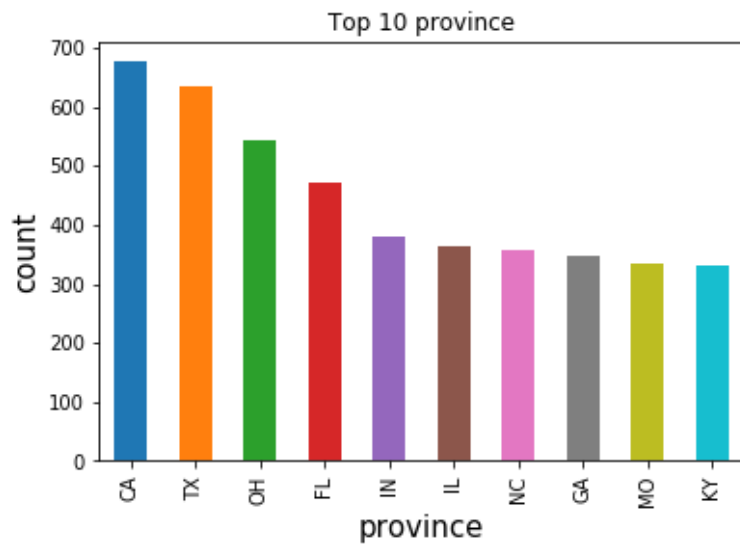
```
Text(0,0.5,'count')
```


2.Top 10 Restaurants using EDA

```
provplt=df.province.value_counts()[:10].plot.bar(title='Top 10 province')
provplt.set_xlabel('province',size=15)
provplt.set_ylabel('count',size=15)
```
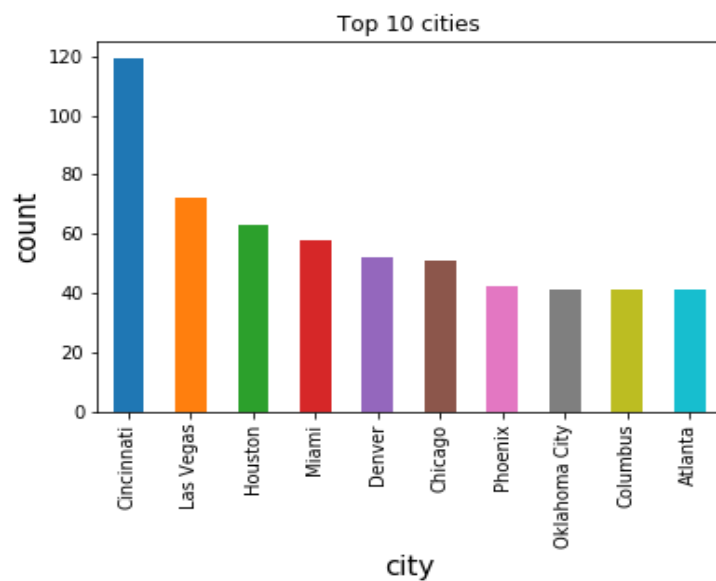
```
Text(0,0.5,'count')
```

3.Histogram of Top 10 Cities using EDA

```
cityplt=df.city.value_counts()[:10].plot.bar(title='Top 10 cities')
cityplt.set_xlabel('city',size=15)
cityplt.set_ylabel('count',size=15)
```

```
Text(0,0.5,'count')
```

Plotting Restaurant Locations on U.S Map

1.McDonalds Restaurants on Map

```python
mcd=df[df.name =='mcdonalds']
mcd_lat = mcd.latitude
mcd_lon = mcd.longitude
mcd_city = mcd.city

data = [
    go.Scattermapbox(
        lat=mcd_lat,
        lon=mcd_lon,
        mode='markers',
        marker=dict(
            size=5,
            color='rgb(255, 0, 0)',
            opacity=0.3
        ))]
layout = go.Layout(
    title='Mcdonalds Restaurants',
    autosize=True,
    hovermode='closest',
    showlegend=False,
    mapbox=dict(
        accesstoken=mapbox_access_token,
        bearing=0,
        center=dict(
            lat=38,
            lon=-94
        ),
```

```python
        ),
        pitch=0,
        zoom=3,
        style='light'
    ),
)

fig = dict(data=data, layout=layout)
iplot(fig, filename='Mcdonalds restaurants')
```

Mcdonalds Restaurants

2.Burger King Restaurants on Map

```python
bg=df[df.name =='burger king']
bg_lat = bg.latitude
bg_lon = bg.longitude
bg_city = bg.city

data = [
    go.Scattermapbox(
        lat=bg_lat,
        lon=bg_lon,
        mode='markers',
        marker=dict(
            size=5,
            color='rgb(0,255, 0)',
            opacity=0.8
        ))]
layout = go.Layout(
    title='Burgerking Restaurants',
```

```python
    autosize=True,
    hovermode='closest',
    showlegend=False,
    mapbox=dict(
        accesstoken=mapbox_access_token,
        bearing=0,
        center=dict(
            lat=38,
            lon=-94
        ),
        pitch=0,
        zoom=3,
        style='light'
    ),
)

fig = dict(data=data, layout=layout)
iplot(fig, filename='burgerking restaurants')
```
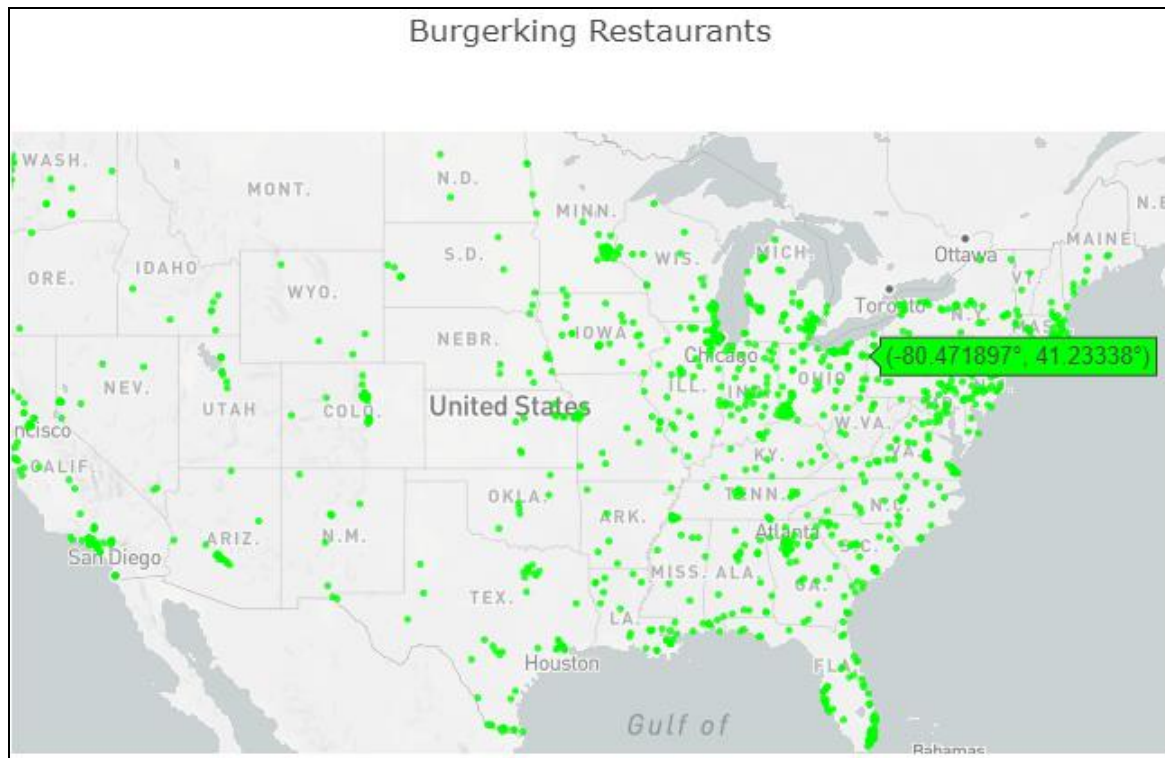

Burgerking Restaurants

3.Tacobell Restaurants on Map

```python
tb=df[df.name =='taco bell']
tb_lat = tb.latitude
tb_lon = tb.longitude
tb_city = tb.city

data = [
    go.Scattermapbox(
        lat=tb_lat,
        lon=tb_lon,
        mode='markers',
        marker=dict(
            size=5,
            color='rgb(0,0,255)',
            opacity=0.8
        ))]
layout = go.Layout(
    title='Tacobell Restaurants',
    autosize=True,
    hovermode='closest',
    showlegend=False,
    mapbox=dict(
        accesstoken=mapbox_access_token,
        bearing=0,
        center=dict(
            lat=38,
            lon=-94
        ),
        pitch=0,
        zoom=3,
        style='light'
    ),
)

fig = dict(data=data, layout=layout)
iplot(fig, filename='Tacobell restaurants')
```

Tacobell Restaurants

(-97.14379°, 34.18602°)
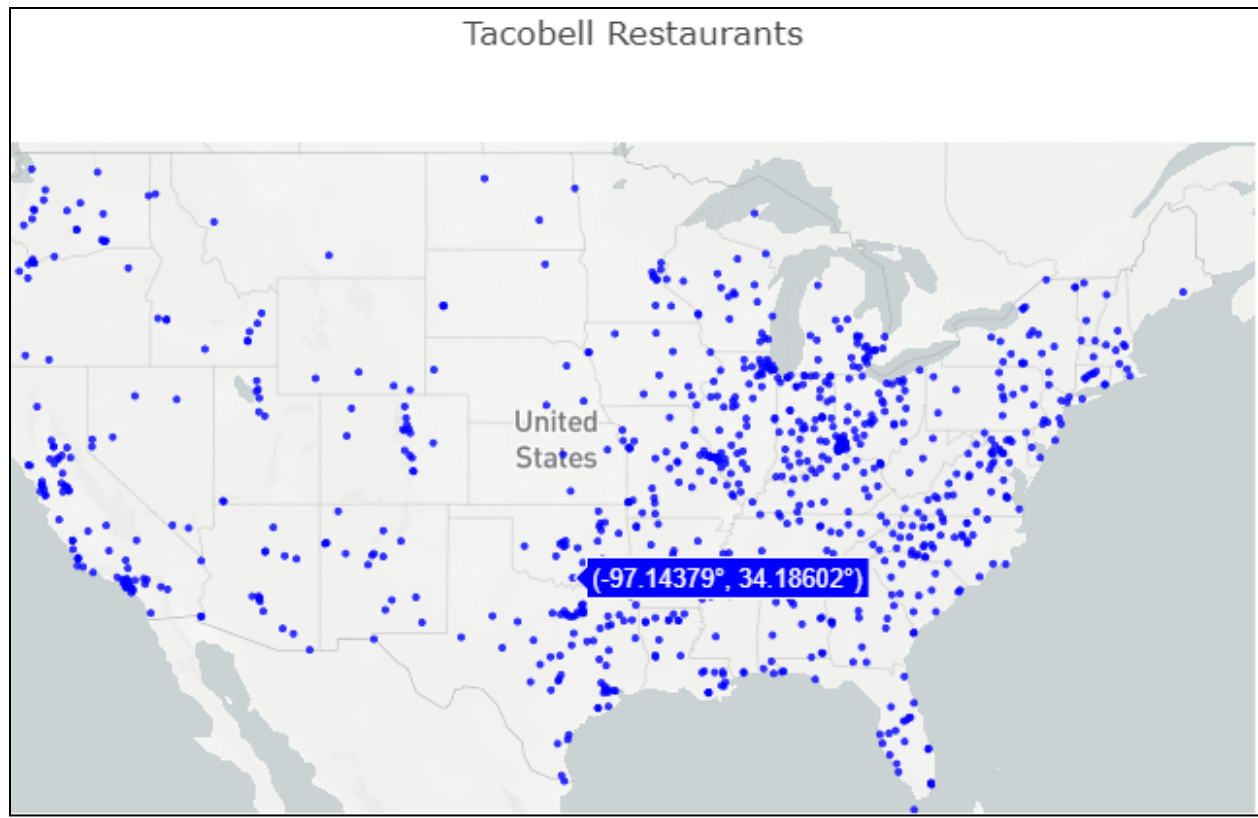
Conclusion

After doing this project we conclude that by using this project on Exploratory Data Analysis of Restaurants , we are now able to understand the locations of various restaurants present in U.S. Also we can do an analysis on the data from FastfoodRestaurant.csv file to see the number of cities or number of provinces having how many restaurants of fastfood.