Food Safety and Inspection Analysis of Restaurants located in San Francisco

[1]:
```
# Initialize OK
from client.api.notebook import Notebook
ok = Notebook('proj1b.ok')
```

```
=====================================================================
Assignment: proj1b
OK, version v1.13.11
=====================================================================
```

[2]:
```
import pickle
import matplotlib
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
sns.set()
plt.style.use('fivethirtyeight')
```

Load the cleaned data

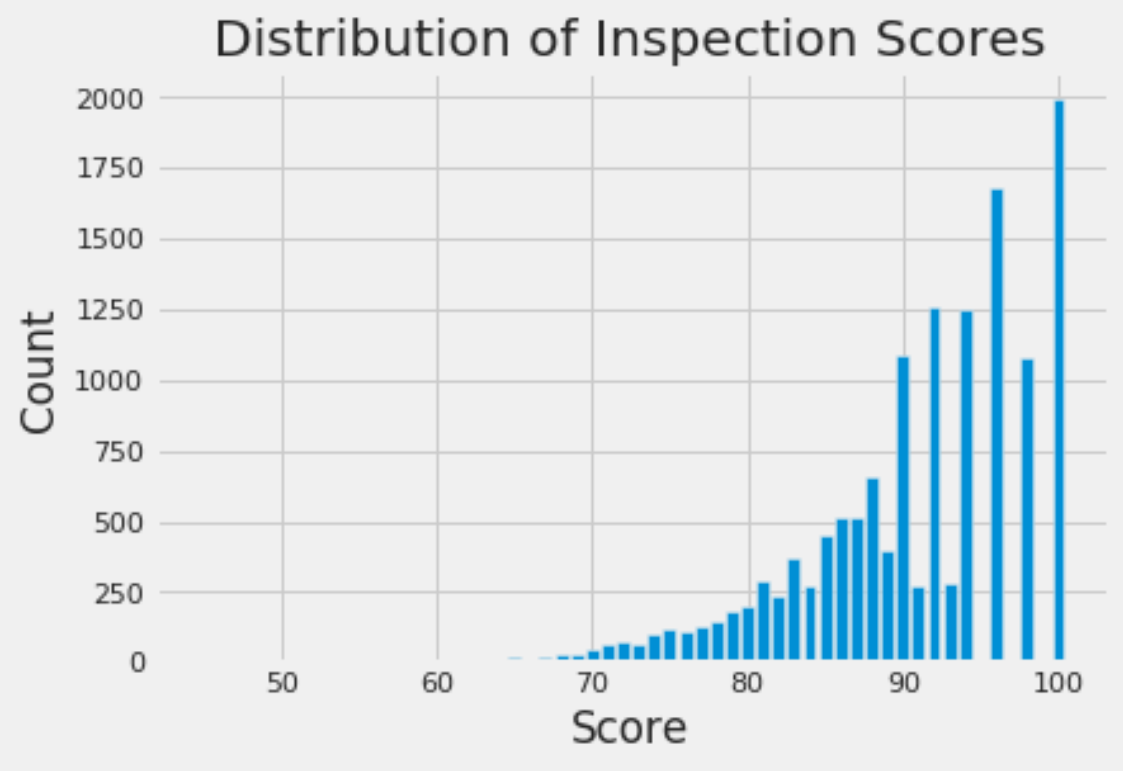
[3]:
```
ins = pickle.load(open('./data/ins.p', 'rb'))
vio = pickle.load(open('./data/vio.p', 'rb'))
ins2vio = pickle.load(open('./data/ins2vio.p', 'rb'))
bus = pickle.load(open('./data/bus.p', 'rb'))
```

Look at the distribution of inspection scores

[4]:
```

```

[4]:
```
Text(0.5, 1.0, 'Distribution of Inspection Scores')
```



Make a dataframe called scores_pairs_by_business indexed by business_id that contains the field score_pair consisting of the score pairs ordered chronologically [first_score, second_score].

[5]:

[5]:

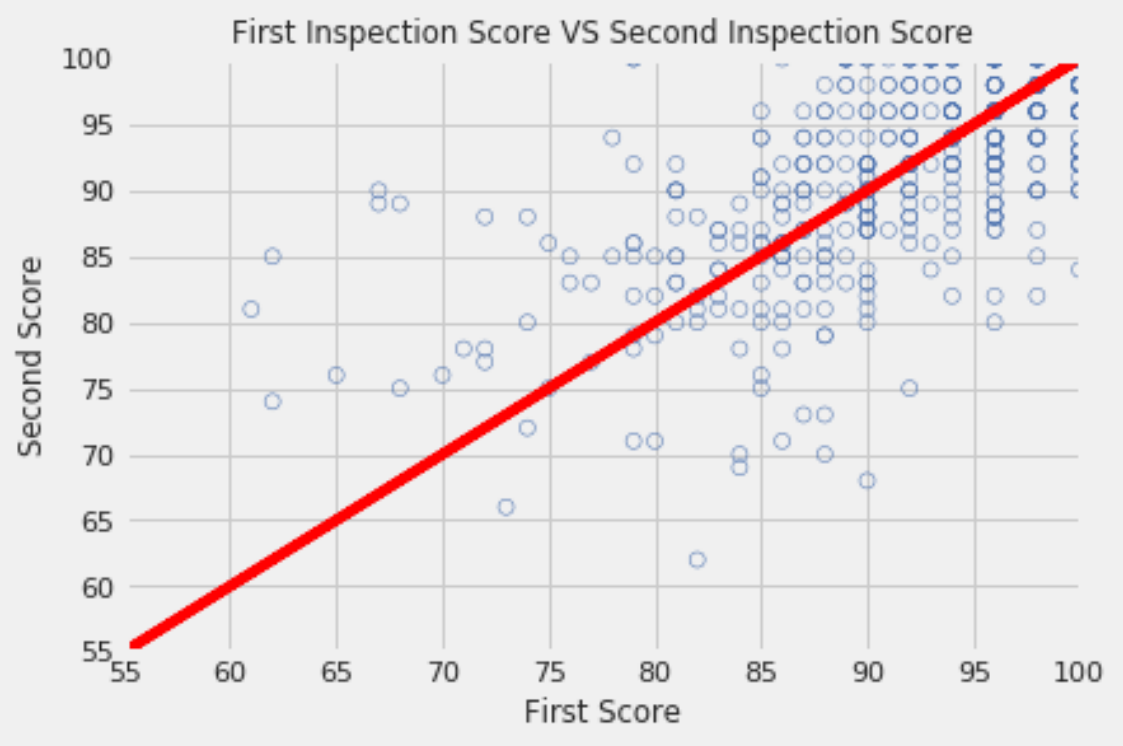
bid	score_pair
48	[94, 87]
66	[98, 98]
146	[81, 90]
184	[90, 96]
273	[83, 84]
…	…
95621	[100, 100]
95628	[75, 75]
95674	[100, 96]
95761	[91, 87]
95764	[100, 92]

[535 rows x 1 columns]

Create a scatter plot

[6]:

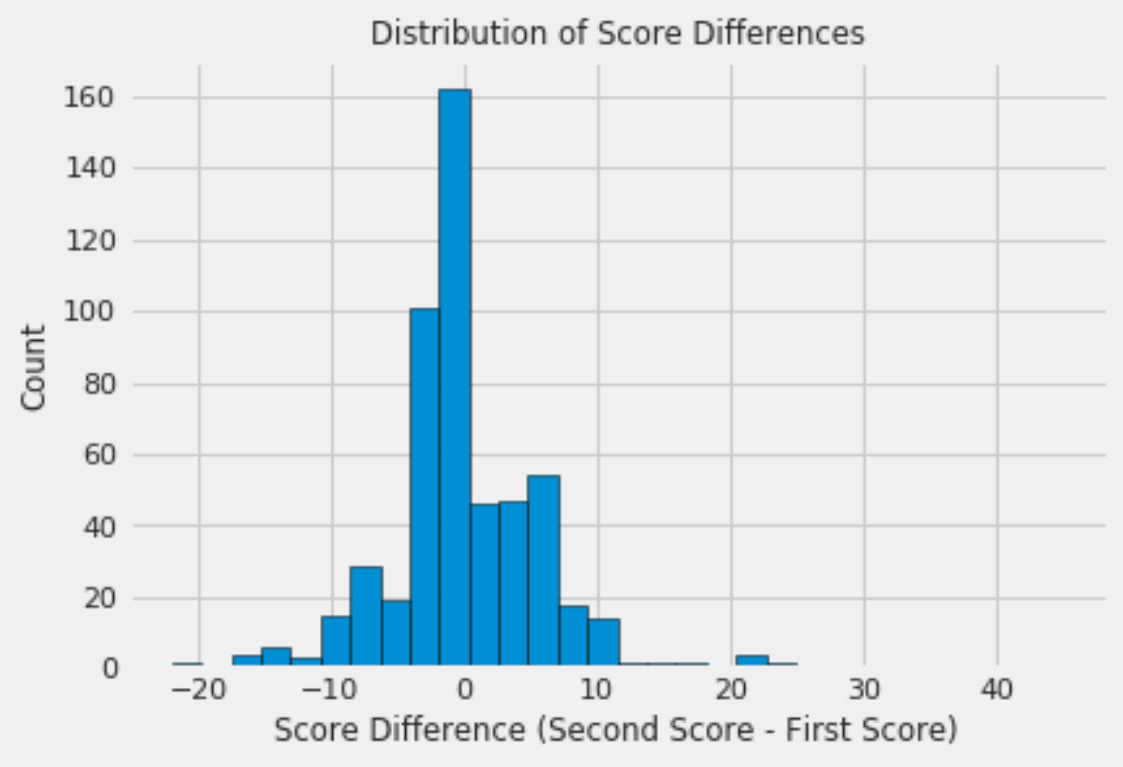
[6]: Text(0.5, 1.0, 'First Inspection Score VS Second Inspection Score')



Compare the scores from the two inspections: subtract the first score from the second in scores_pairs_by_business. Make a histogram of these differences in the scores

[7]:

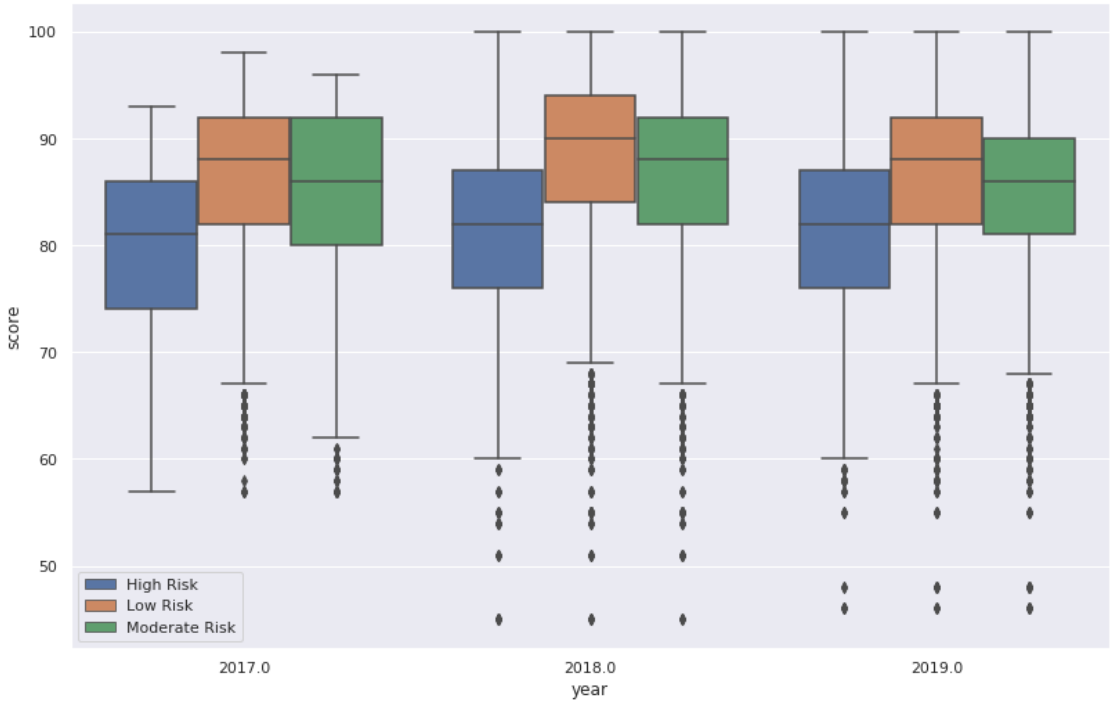
[7]: Text(0.5, 1.0, 'Distribution of Score Differences')



Looking at the distribution of restaurant scores over time. Create a side-by-side boxplot that shows the distribution of these scores for each different risk category from 2017 to 2019

[8]:

[8]: `<matplotlib.legend.Legend at 0x7f2ac8717cf8>`



[ ]:

In the context of restaurant ratings, we can choose our x[i], y[i], c[i] values to be the longitude, latitude, and inspection score for each restaurant in San Francisco respectively. First, create a DataFrame rated_geo that includes the longitude, latitude, and score for each restaurant.

[9]:

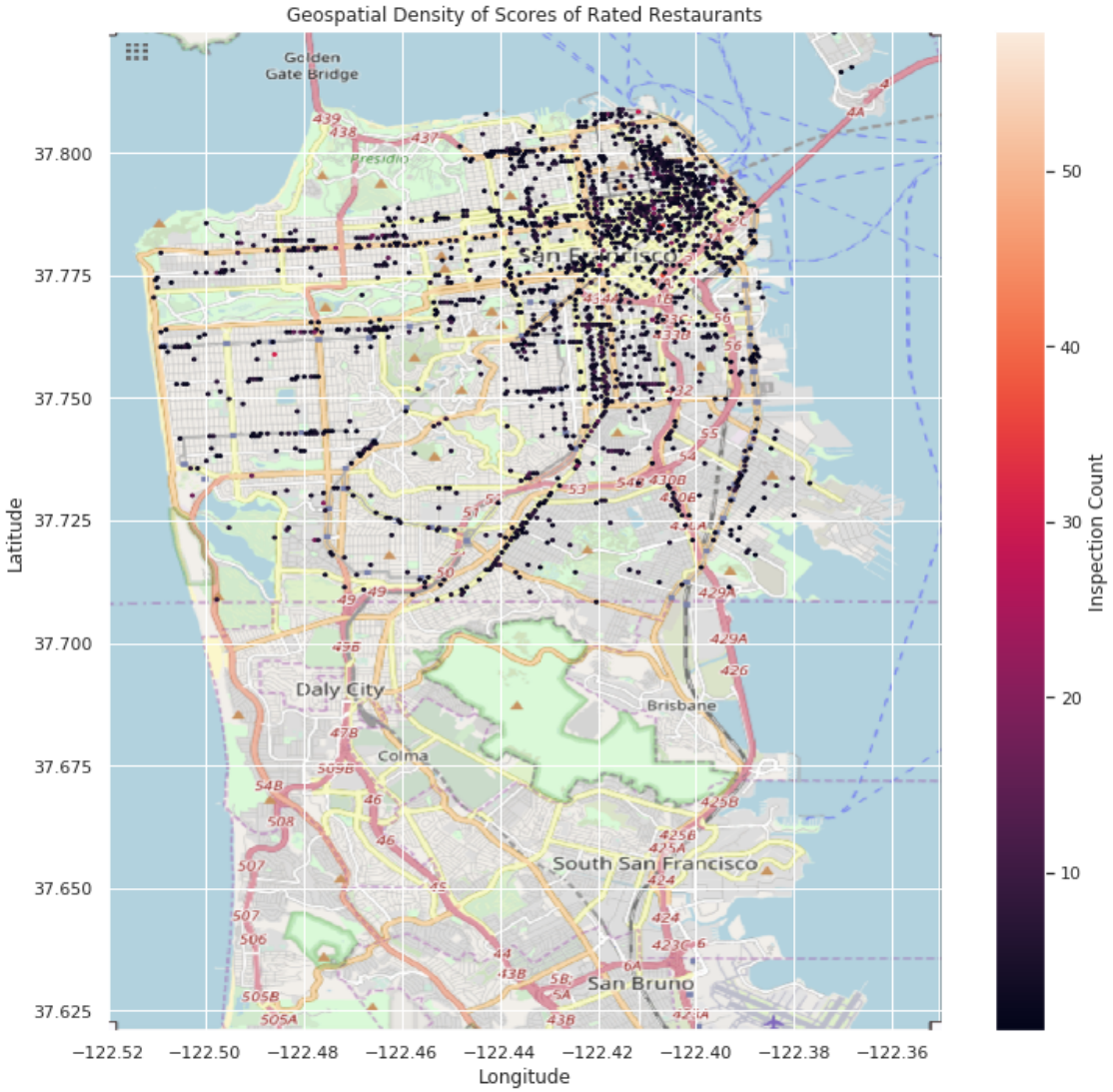
[9]:

	score	latitude	longitude
16	74	37.755282	-122.420493
17	76	37.755282	-122.420493
18	72	37.755282	-122.420493
36	85	37.752158	-122.420362
37	90	37.752158	-122.420362
…	…	…	…
14026	77	37.756997	-122.420534
14027	80	37.756997	-122.420534
14028	80	37.756997	-122.420534
14029	82	37.794293	-122.405967
14030	84	37.794293	-122.405967

[7390 rows x 3 columns]

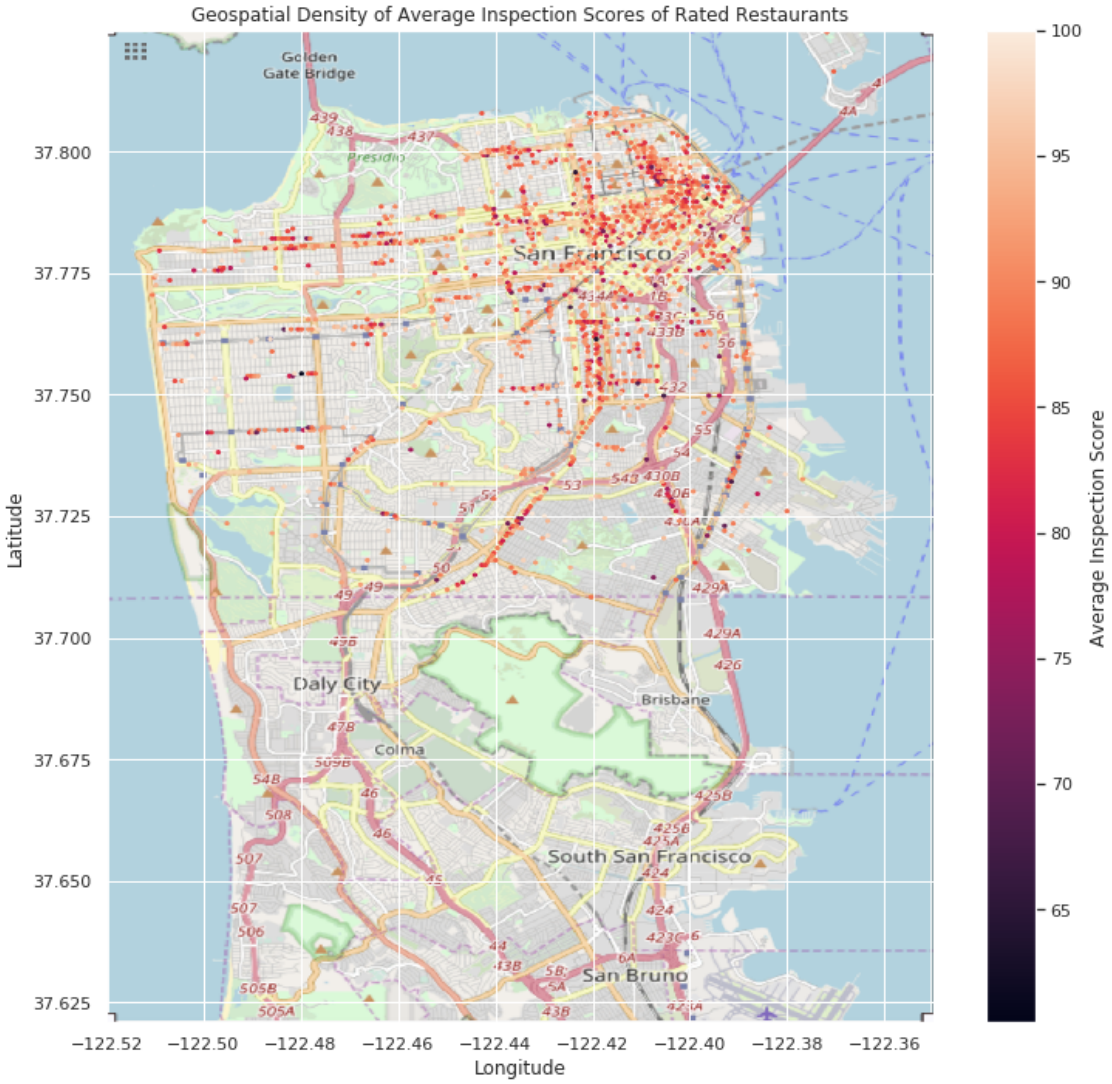
Create a geospatial hexbin plot that includes the inspection count for all restaurant locations in San Francisco

[11]:



Create another hexbin plot that visualizes the average inspection scores for restaurants in San Francisco.

[13]:


Geospatial Density of Average Inspection Scores of Rated Restaurants
Golden Gate Bridge
San Francisco
Daly City
Brisbane
Colma
South San Francisco
San Bruno
Latitude
Longitude
37.800
37.775
37.750
37.725
37.700
37.675
37.650
37.625
−122.52
−122.50
−122.48
−122.46
−122.44
−122.42
−122.40
−122.38
−122.36
Average Inspection Score
100
95
90
85
80
75
70
65

[ ]: