Group 4 - Project 1

Michelin Star Restaurants

Elyssa Irizarry
Cody Gunter
Megan Adams
Markeis Williams

Introduction

Our exploratory data analysis is over Michelin Star Restaurants of 2018-2019, provided on Kaggle. Three datasets are given regarding one, two, and three star Michelin restaurants. The datasets provide information on restaurant location, including latitude, longitude, city, region, and zip code. The data also includes cuisine type, and price.

Michelin Stars are a global restaurant rating system created by Michelin Tire. The Michelin Star Guide was initially created to encourage travel, consequently, leading to an increase in tire sales. Over time, a Michelin Star rating became highly coveted and respected in the restaurant business. It is a difficult honor to achieve for any chef and restaurateur. Restaurants initially gain a one star rating. The rating can be increased to two and then three stars as cuisine, quality, and experience rose.

In order to analyze the data, we merged the data sets for one, two, and three star restaurants. From this data we addressed the following research questions.

- What is the density and location of Michelin Star Restaurants worldwide, from the data provided?
- What is the proportion of Michelin stars (one versus two versus three) by region, cuisine, and price?
- How do prices compare between location and star rating?
- What cuisine types are most popular at different star levels and price levels?

Data Approach

Six hundred and ninety five restaurants were analyzed through the information provided from Kaggle. Currently, there are over 2,800 Michelin Star restaurants in the world. The data analyzed is a representative sample. We estimate our sample is approximately twenty five percent of total Michelin Star restaurants worldwide. Two hundred and two restaurants are located within the United States.

Inspiration regarding the initial analysis of the data was located on Kaggle from previous analyses. A previous exploratory data analysis from Luna McBride provides ideas on dealing with null values, and incorporating map usage. Analysis from Thomas Konstantin included ideas on how to address missing price values.

Null values for zip code and inconsistencies in region labeling were addressed by focusing on latitude and longitude for analysis. The restaurant rating and price point were provided in string values and maintained as string values through the analysis. Null rating values were dropped since approximately twenty five percent of the global values were missing. This is a significant amount and alters the price analysis. Star count was altered to a float/integer for the price

analysis. Through exploration of the Michelin Star Guide, we learned many restaurant locations were missing from the data provided.

US Census data from 2019 was merged with the US Michelin Star data, specifically, population, median age, household income, per capita income, poverty count, unemployment count, and unemployment rate. US Census data was used for the logistical regression analysis.

Results

Location Analysis

Keeping in mind that the data from Kaggle is only a representative sample of all Michelin Star restaurants, the locations are still scattered on nearly every continent. All were mapped with markers for each restaurant with its name, as well as a heat map layer to determine the concentration area.

In the US, the main regions are New York City, Washington D.C., Chicago, San Francisco, and Los Angeles. All of the locations in California were put into the region "California," whereas the others outside of the state had the city name in the Region column. Until the locations were mapped, it was unknown in exactly what city they were.

In Europe, a great majority are in the UK, specifically London, according to the heat map layer. Scandinavia also has several restaurants in Sweden, Norway, Finland, and Denmark. Eastern Europe has approximately one per country, such as one in Hungary, Czechia, Poland, and Slovenia.

In Asia, the locations are more concentrated. Thailand has several in Bangkok, as does Singapore, Seoul, Taipei, and Hong Kong. (Appendix 5-11.2)

Proportion of Star Rating Analysis

Through the merged data of one, two, and three star restaurants it was determined that 79% of the 695 restaurants were one star recipients, or 549. Two star restaurants made up 15.8%, or 110, and three stars constituted 5.2%, or 36 restaurants. (Appendix 14) Since the majority of restaurants is one star, this skews all proportional data to one star rating. This is seen more thoroughly in the regression analysis.

Due to use of the US Census data, a further analysis was done on restaurants located within the United States. A total of 202 restaurants broke down to 155, or 76.7%, one star, 33, or 16.2% two star, and 14, or 16.9% three star. (Appendix 15) This sample reflects global data numbers.

A region by region breakdown comparing the average Michelin star rating is shown (Appendix 16) with the US regions in Red. Overall, the average Michelin star count per restaurant hovers around 1.25, meaning for every 4 Michelin star restaurants, a total of 5 stars will be the cumulative average sum. Of the countries that rank amongst the highest in volume, Hong Kong, New York City, and California exceeded that 1.25 mark.

Price Analysis

The average price rating of the 519 restaurants with a provided price rating was 3.4. A bar chart depicting each region ranked ascendingly in average price is shown (Appendix 19). Four regions with a smaller sample size had an average price point of 5, but the majority of the regions were within close proximity to the 3.4 mark. The US regions outside of Washington DC had greater price points than the mean, and further analysis of the US regions is shown (Appendix 20.1 and 20.2). The region that stands out the most when comparing average price and star count is Hong Kong. Hong Kong ranks as the third greatest region in average star count, while being fourth from the bottom in their price rate. To see the cause we had to analyze how cuisine affects price.

Appendix 20 displays the ten top cuisines by the number of restaurants with the average star count and price point. The cuisine that stands out with the lowest price rating and star count greater than the 1.25 average is Cantonese. Cantonese cuisine restaurants average a miniscule price rating of 2.2. In Hong Kong, Cantonese cuisine is the most prevalent cuisine type which explains Hong Kong's lower price rating with a greater Michelin star average.

When comparing Michelin star count against their respective average price, the results were not surprising (Appendix 17.1 and 17.2). Restaurants with two or three Michelin stars prove to be much more expensive than one star restaurants.

Cuisine Analysis

Of the 695 restaurants, seventy different cuisine types were represented. (Appendix 23) Many cuisines are represented only one or two times. Deeper analysis determined that the ten most popular cuisines in Michelin Star restaurants are (in order of popularity) modern cuisine, contemporary, Japanese, creative, Cantonese, modern British, French, innovative, Italian, and French contemporary. One hundred and eight restaurants are labeled as modern cuisine. (Appendix 24) According to the Michelin Star guide, restaurants label their own cuisine type. This is the likely reason for the variety of cuisine types listed and the duplication of cuisine styles, i.e. modern and modern British.

Analysis of cuisine type by Michelin Star count revealed that a majority portion of modern cuisine restaurants have received a one Michelin Star rating. The top ten cuisines which have earned one star ratings are modern, contemporary, Japanese, modern British, Cantonese, creative, innovative, French, Italian, and classic. It was also noted that Italian and classic cuisine in 2019

only received one star ratings. All of the sushi restaurants within this data received two star ratings. Asian, American, and Korean cuisines have all earned one star ratings. The top two star cuisines are modern, contemporary, Japanese, modern British, Cantonese, creative, innovative, French, French contemporary, and sushi. The top three star cuisines are modern, contemporary, Japanese, Cantonese, creative, French, French contemporary, Asian, American, and Korean. (Appendix 25)

A further analysis was completed on restaurants within the United States. The top ten cuisines altered, with contemporary being the most popular cuisine at 75 restaurants. The top ten cuisines within the US are contemporary, Japanese, Californian, American, Italian, French, Mexican, Korean, Seafood, and fusion. (Appendix 27) Analysis of cuisines based upon star rating show that contemporary cuisine is majority composed of one star rated restaurants. Fusion and Mediterranean restaurants have only received one star ratings. Scandinavian and Indian have been awarded two star ratings, and Asian is the only cuisine in the United States to solely receive three Michelin stars. The top one star cuisines in the US are contemporary, Japanese, Californian, American, Italian, French, Mexican, Korean, fusion, and Mediterranean. The top two star restaurant cuisines are contemporary, Japanese, Califorian, Italian, French, Mexican, Korean, Scandinavian, seafood, and Indian. The top three star cuisines are contemporary, Japanese, American, seafood, and Asian. (Appendix 28)

Logistic Regression

Traditionally, data analysis would require a linear regression. However, our data was much more suited to a logistic regression. That is, we determined which price point a restaurant would fall into. The worldwide options were one, two, three, four, and five dollar signs, shown as $, $$, $$$, $$$$, and $$$$$.

In order to get as accurate a result as possible, we brought in the following census data: Population, Median Age, Household Income, Per Capita Income, Poverty Count, and Unemployment Count. We added an additional column for Unemployment Rate, which was calculated by dividing Unemployment Count by the Population. We restricted the data to the United States only, due to including the census data. This reduced the price point options to only $$, $$$, and $$$$.

The logistic regression predicted overwhelmingly 4$ price points. This is because of the 180 restaurants in the US, 118 or 66% were at that price point. The rest were 45 3$ price points, and only 17 2$ price points. The algorithm determined that since the majority of the restaurants were 4$ restaurants, it would be correct the majority of the time if it predicted all were 4$ price points. A probability array was also run. It showed that some restaurants were quite close in the predictions - 26% chance of 2$, 34% chance of 3$, and 39% chance of a 4$ price point - but still overwhelmingly had 4$ price points at the highest probability.

The confusion matrix shows how inaccurate the regression is. Only one 2$ price point was accurately predicted, and zero 3$ price points were predicted correctly. A few were inaccurately predicted to be 2$ and 3$ as well, when they were actually 4$.

The correlation heatmap also confirms the lack of correlation with price point.. Correlation was seen within the census data. For example, poverty count and unemployment rate show a strong correlation. Michelin Star count and per capita income show a low correlation at 0.149. A scatter plot of per capita income versus Michelin Star rating was completed to look for trends. One star restaurants were spread through all income ranges with the highest concentration between $40,000 and $100,000. Comparing two and three star restaurants with one star restaurants show a slight trend towards higher per capita income where the restaurant is located.

The ANOVA tests the differences between the average regional Per Capita Income versus the Michelin Star count. This was pictured using a violin plot, as well as a boxplot. The boxplot showed there were no outliers for PCI based on Michelin Star count.

The average Per Capita Income is higher for 3-star restaurants. The average PCI for 1-star restaurants is $75,000, 2-star is around $83,000, and 3-star is nearly $100,000. Average Per Capita Income increases as star count increases. Another way to look at it, the higher the PCI, the more likely it is that the region has a 3-star restaurant.

Limitations and Future Work

Time was the biggest limitation to this project, as it is with most ventures, but we have built a foundation for potential future work. Some possible ideas would involve including Yelp reviews as part of our model, testing whether consumers rate two and three star Michelin star restaurants more favorably than one star restaurants, or whether certain cuisine out competes others in reviews. Another would be further analysis on non-US regions, doing a similar breakdown looking for patterns or outliers internationally. Lastly, analyzing a larger sample of restaurants would also be recommended.

Conclusion

In conclusion, the region with the most Michelin Star restaurants is the United Kingdom. Approximately, 79% of Michelin Star restaurants, from the data, are one star. Average price increases with star count and per capita income in the United States. Modern and contemporary cuisine types are the most popular worldwide and within the United States, respectfully. Higher star count also correlates to higher restaurant prices worldwide. Predictive analysis was flawed

due to the large percentage of one star and $$$$ priced restaurants in the sample. A future analysis of all current Michelin Star restaurants will confirm or refute the analysis of this sample.

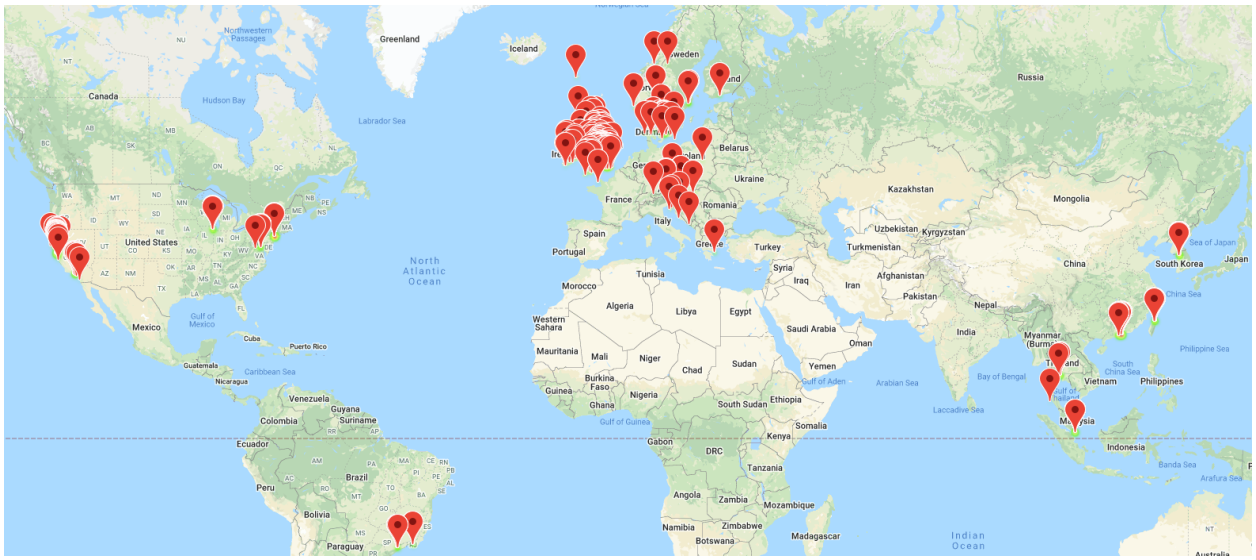
References

- https://www.kaggle.com/jackywang529/michelin-restaurants?select=one-star-michelin-restaurants.csv
- https://www.kaggle.com/lunamcbride24/michelin-star-exploration-classification
- https://www.kaggle.com/thomaskonstantin/michelin-restaurants-eda-missing-price-prediction
- https://www.census.gov/data/developers/data-sets.html
  https://guide.michelin.com/us/en/restaurants
- Colors reference image - #C9002A, #0C0948, #EDF2E8
    - Reference image (see Appendix 40.1)-
      https://michelinmedia.com/site/user/images/Michelin_Man_Chicago_Red_Guide.jpg

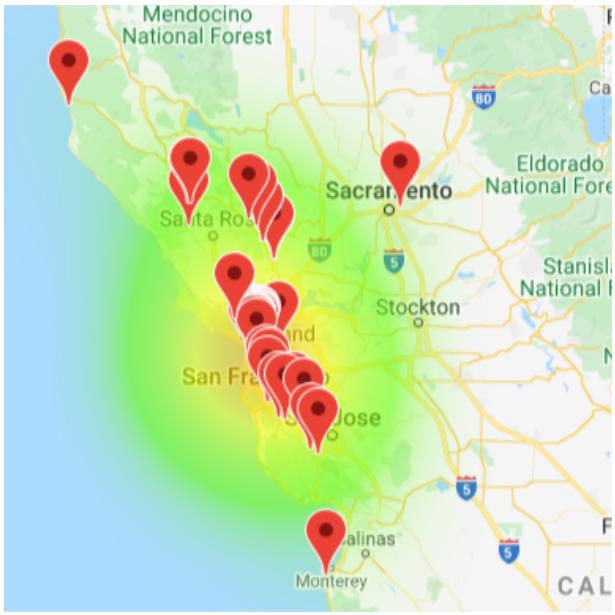
Appendix

Images labeled in correlation to their appearance in slideshow by slide number.

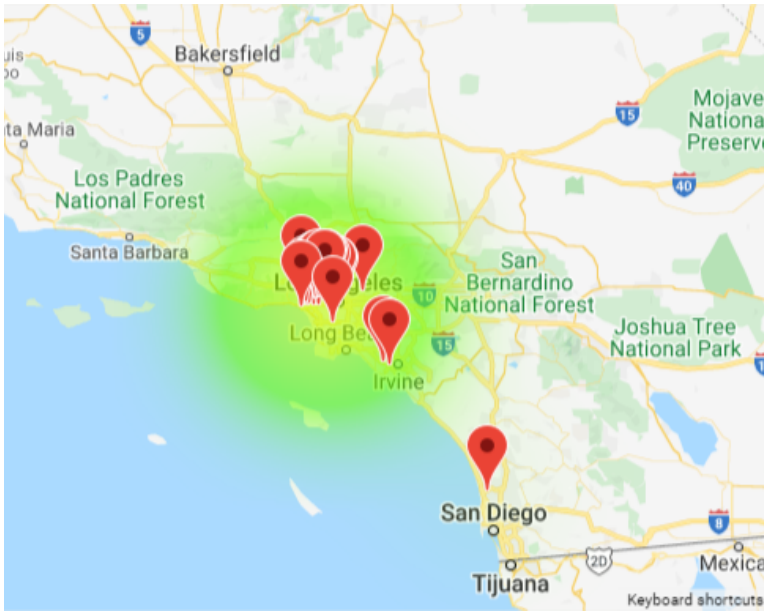
5



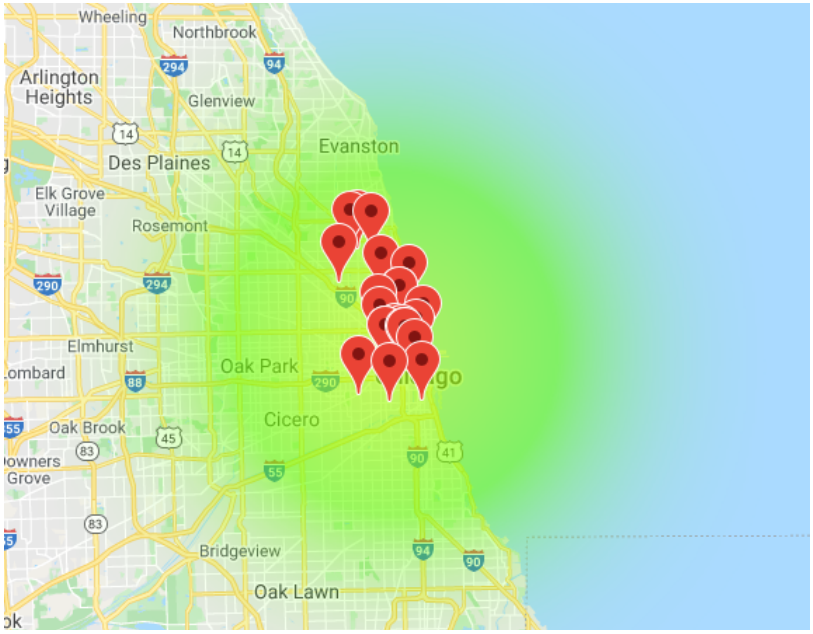
6.1



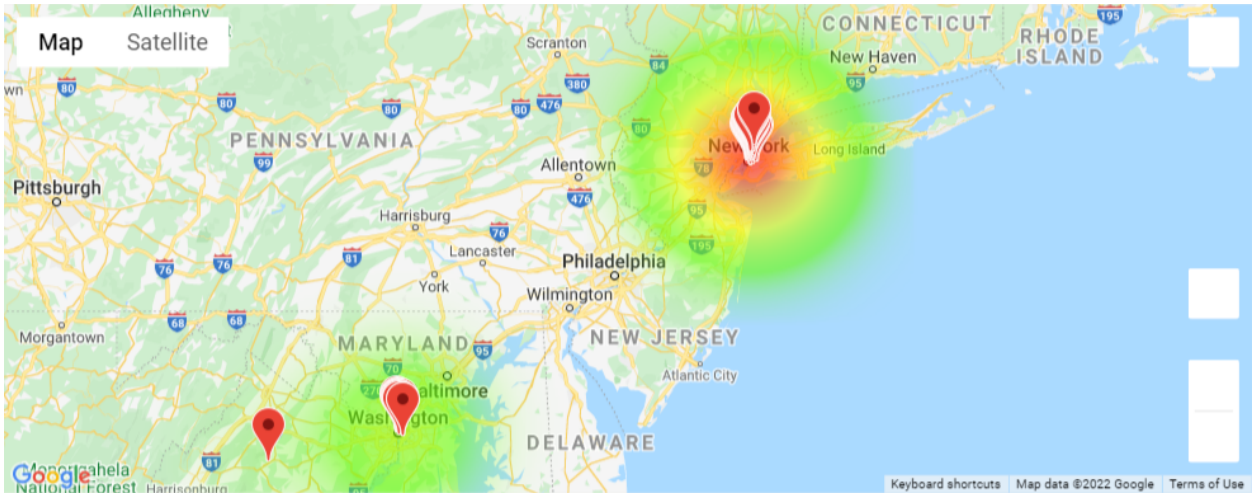
6.2


Bakersfield
Mojave National Preserve
Los Padres National Forest
Santa Barbara
San Bernardino National Forest
Joshua Tree National Park
Irvine
San Diego
Tijuana
Keyboard shortcuts

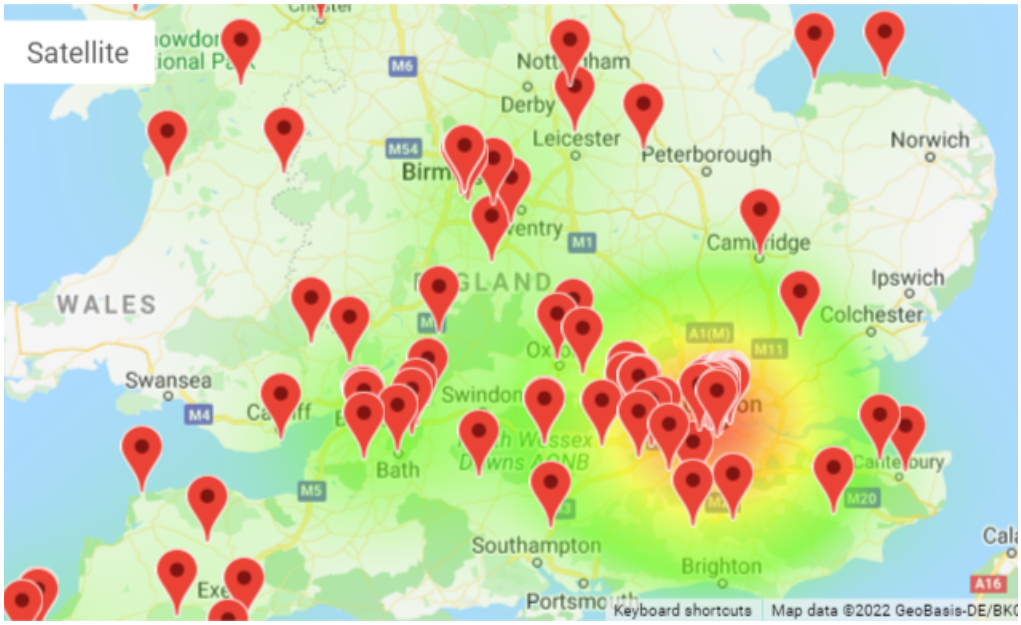
7


Wheeling
Northbrook
Arlington Heights
Glenview
Evanston
Des Plaines
Elk Grove Village
Rosemont
Elmhurst
Oak Park
Cicero
Oak Brook
Bridgeview
Oak Lawn

8


Map
Satellite
CONNECTICUT
RHODE ISLAND
New Haven
Scranton
PENNSYLVANIA
Allentown
Long Island
Pittsburgh
Harrisburg
Lancaster
Philadelphia
York
Wilmington
Morgantown
MARYLAND
NEW JERSEY
Atlantic City
DELAWARE
Keyboard shortcuts
Map data ©2022 Google
Terms of Use

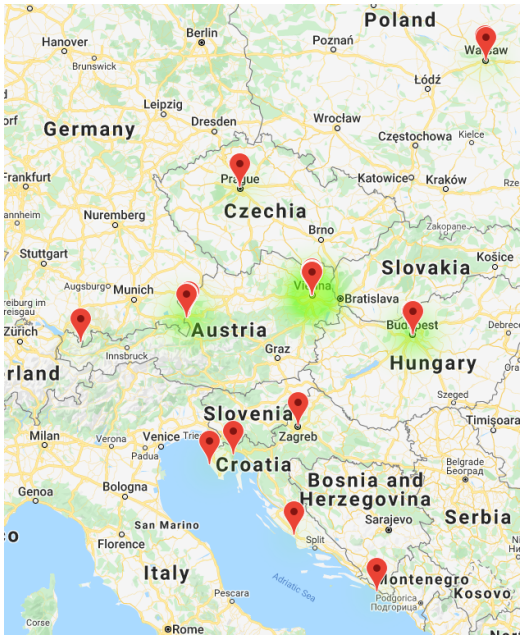
9


Satellite
Derby
Leicester
Peterborough
Norwich
Cambridge
Ipswich
Colchester
WALES
Swansea
Swindon
Bath
Southampton
Brighton
Keyboard shortcuts
Map data ©2022 GeoBasis-DE/BK

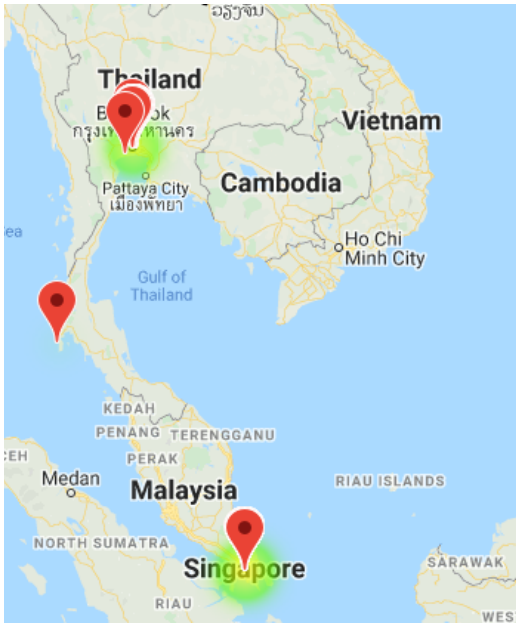
10.1


Sweden
Norway
Finland
Estonia
Latvia
Lithuania
Denmark
Tallinn
Riga
Vilnius
Kaunas
Gothenburg
Malmö
Stockholm
Helsinki
Turku
Tampere
Bergen
Oslo
Aalborg
Odense
Gdynia
Baltic Sea
Gulf of Bothnia

10.2


Poland
Germany
Czechia
Slovakia
Austria
Hungary
Slovenia
Croatia
Bosnia and Herzegovina
Serbia
Italy
Berlin
Hanover
Poznań
Łódź
Leipzig
Dresden
Wrocław
Częstochowa
Katowice
Kraków
Nuremberg
Stuttgart
Munich
Brno
Bratislava
Košice
Graz
Innsbruck
Zagreb
Timișoara
Milan
Venice
Padua
Verona
Belgrade
Sarajevo
Split
Genoa
Bologna
San Marino
Florence
Pescara
Rome
Adriatic Sea
Kosovo
Zakopane

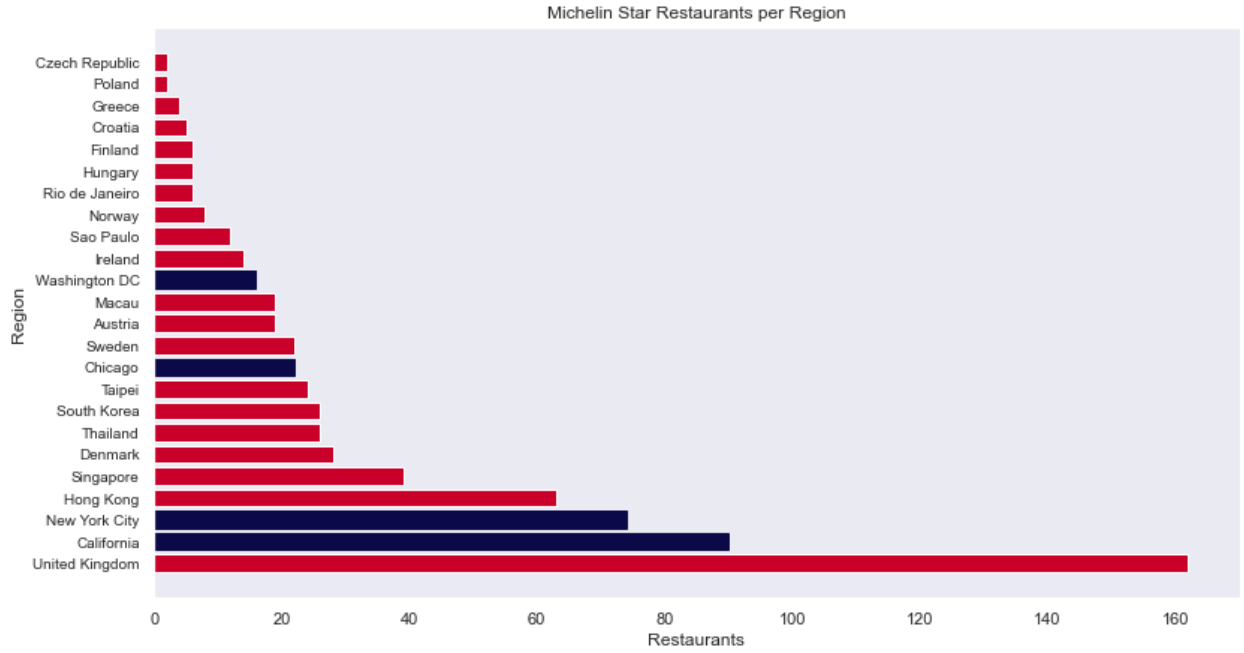
11.1


Thailand
Vietnam
Cambodia
Pattaya City
Ho Chi Minh City
Gulf of Thailand
KEDAH
PENANG
TERENGGANU
PERAK
Medan
Malaysia
RIAU ISLANDS
NORTH SUMATRA
Singapore
SARAWAK
RIAU

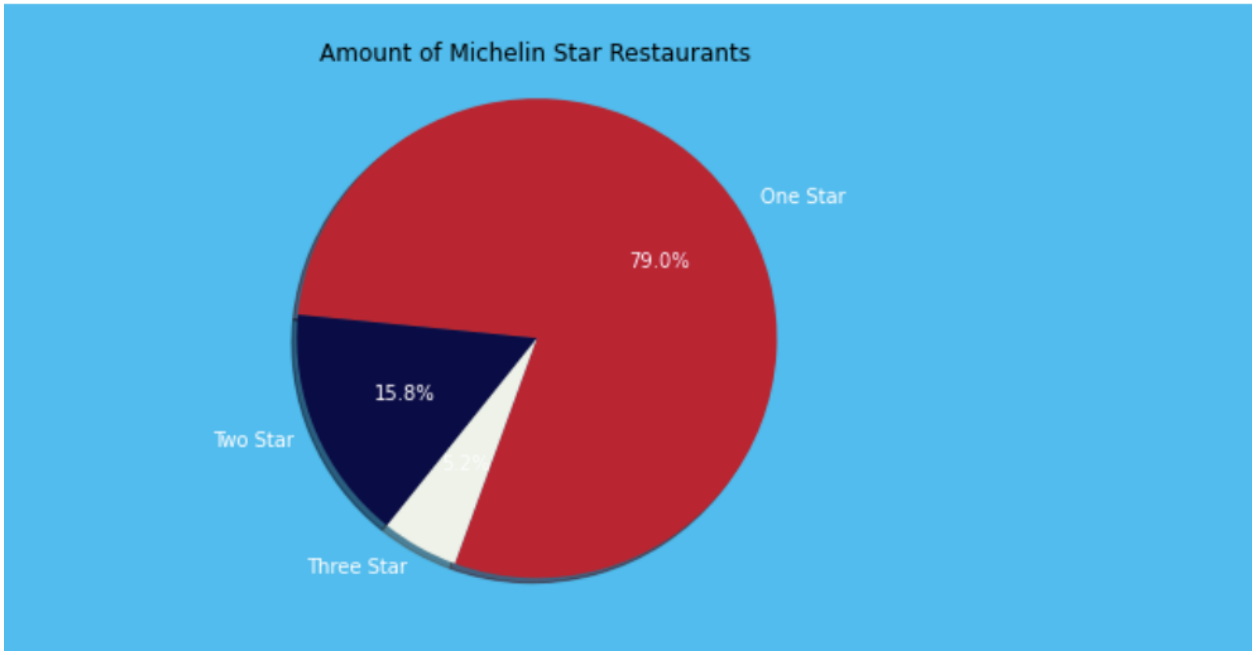
11.2


North Korea
Pyongyang
Seoul
South Korea
Yellow Sea
Busan
Jeju
HEBEI
SHANXI
SHANDONG
HENAN
JIANGSU
Shanghai
HUBEI
Wuhan
ANHUI
East China Sea
ZHEJIANG
HUNAN
JIANGXI
FUJIAN
Taiwan
Kaohsiung City
Hong Kong

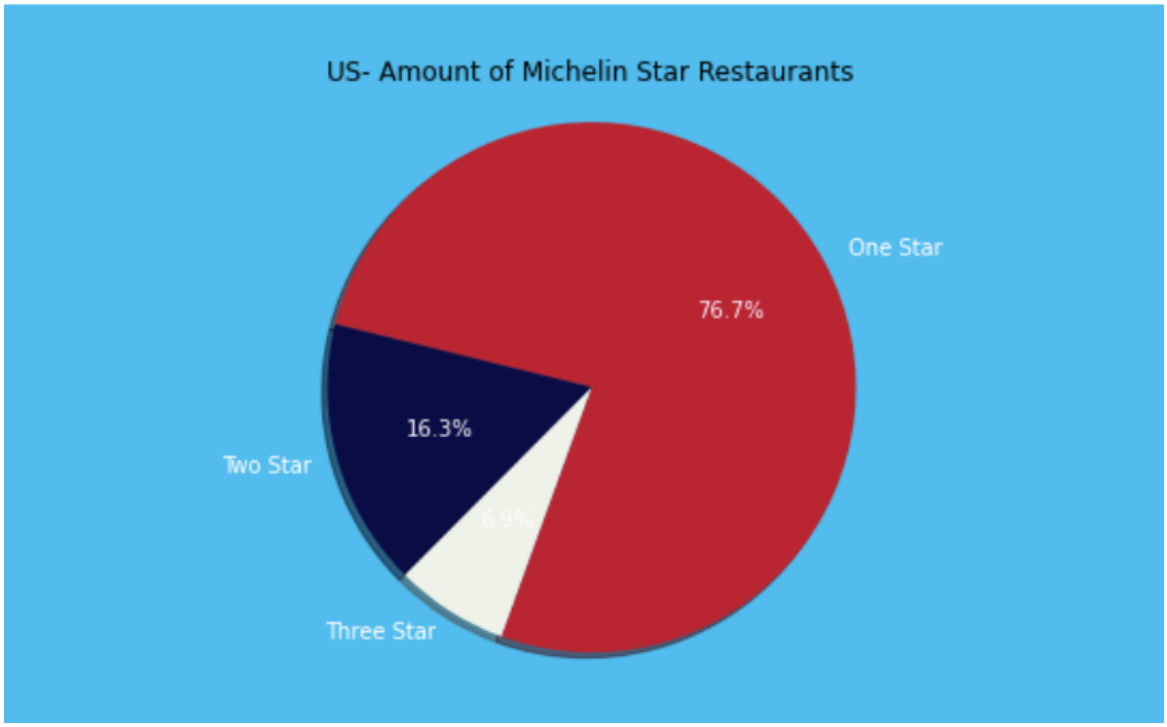
12


Michelin Star Restaurants per Region
Region
Czech Republic
Poland
Greece
Croatia
Finland
Hungary
Rio de Janeiro
Norway
Sao Paulo
Ireland
Washington DC
Macau
Austria
Sweden
Chicago
Taipei
South Korea
Thailand
Denmark
Singapore
Hong Kong
New York City
California
United Kingdom
0
20
40
60
80
100
120
140
160
Restaurants

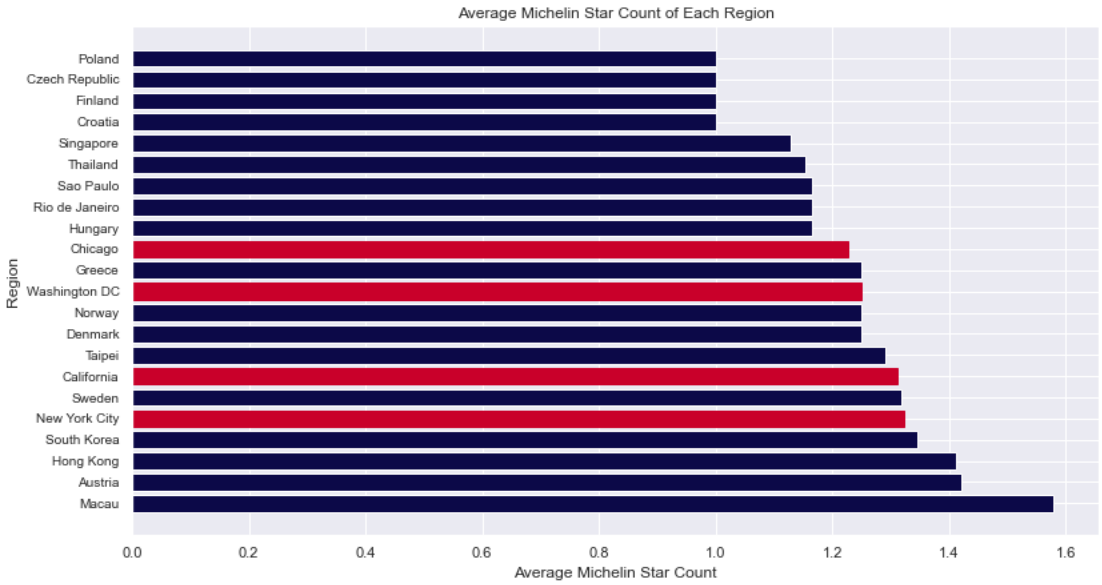
14


Amount of Michelin Star Restaurants
One Star
79.0%
15.8%
Two Star
Three Star

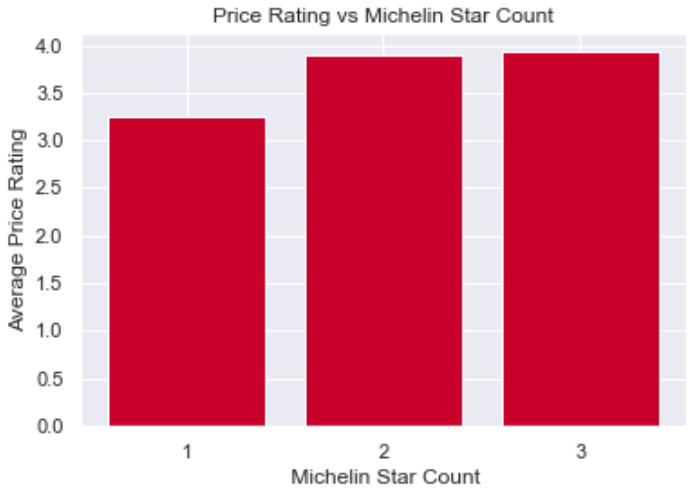
15


US- Amount of Michelin Star Restaurants
One Star
76.7%
16.3%
Two Star
6.9%
Three Star

16


Average Michelin Star Count of Each Region
Poland
Czech Republic
Finland
Croatia
Singapore
Thailand
Sao Paulo
Rio de Janeiro
Hungary
Chicago
Greece
Washington DC
Norway
Denmark
Taipei
California
Sweden
New York City
South Korea
Hong Kong
Austria
Macau
Region
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Average Michelin Star Count

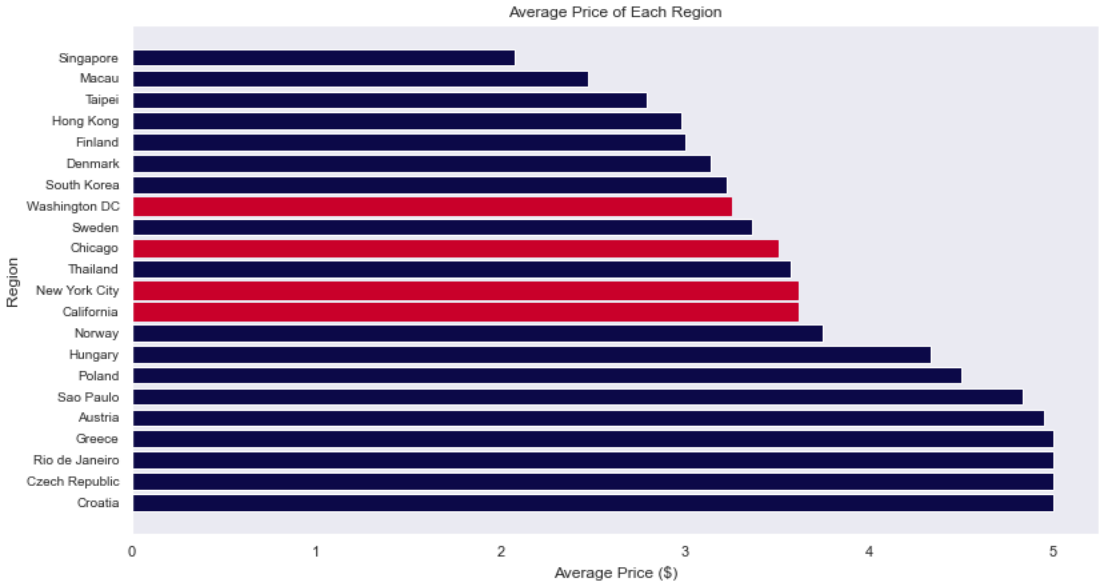
17.1


Price Rating vs Michelin Star Count
Average Price Rating
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
Michelin Star Count

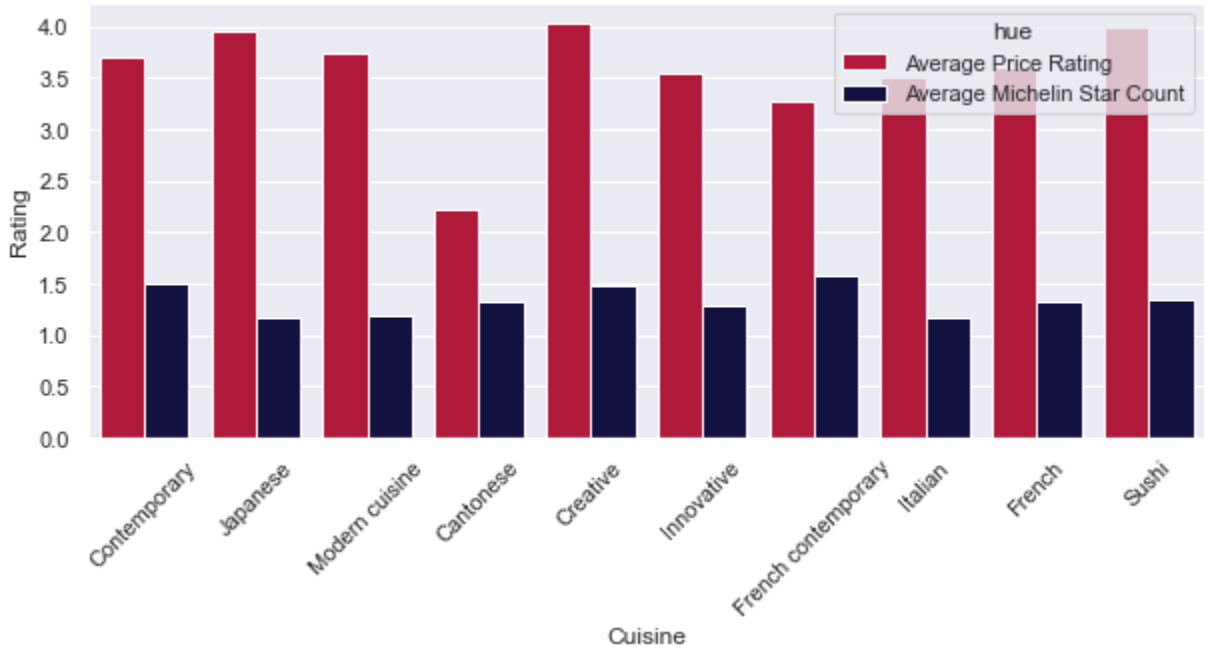
17.2


Price Rating Against Michelin Star Count
Average Michelin Star Count
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
Price Rating ($)

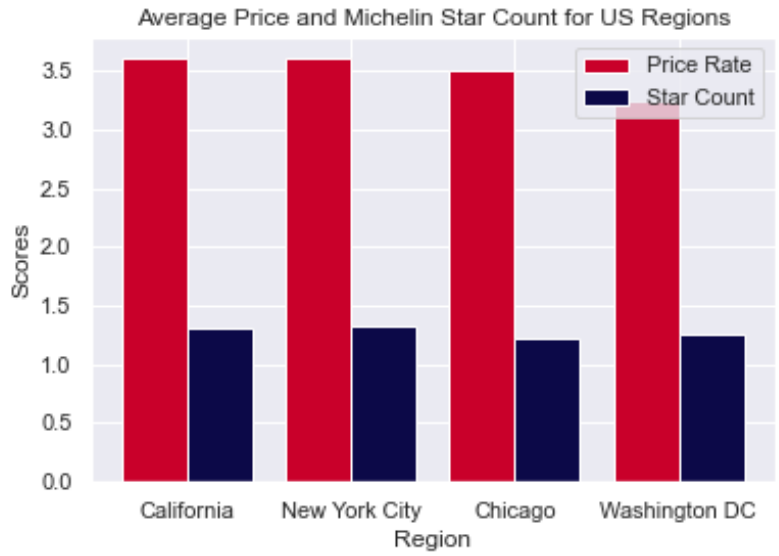
19


Average Price of Each Region
Singapore
Macau
Taipei
Hong Kong
Finland
Denmark
South Korea
Washington DC
Sweden
Chicago
Thailand
New York City
California
Norway
Hungary
Poland
Sao Paulo
Austria
Greece
Rio de Janeiro
Czech Republic
Croatia
Region
0
1
2
3
4
5
Average Price ($)

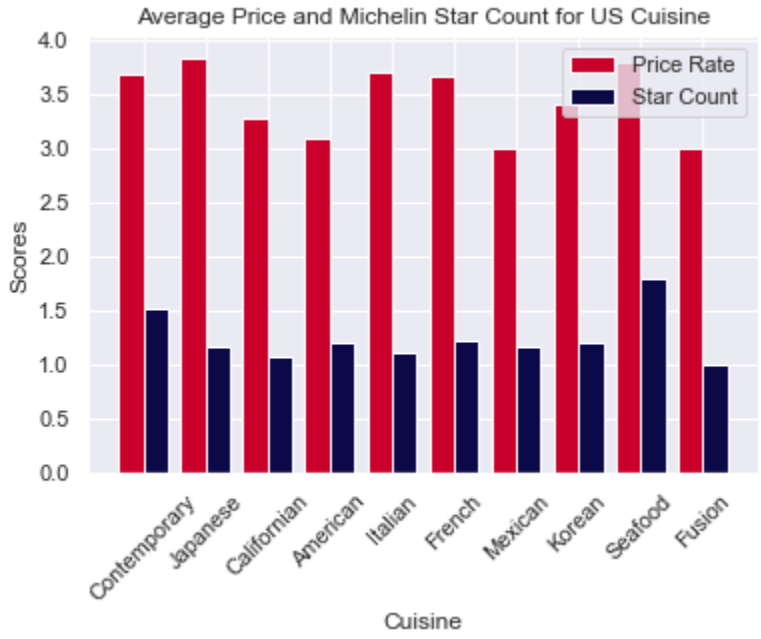
20


hue
Average Price Rating
Average Michelin Star Count
Rating
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Contemporary
Japanese
Modern cuisine
Cantonese
Creative
Innovative
French contemporary
Italian
French
Sushi
Cuisine

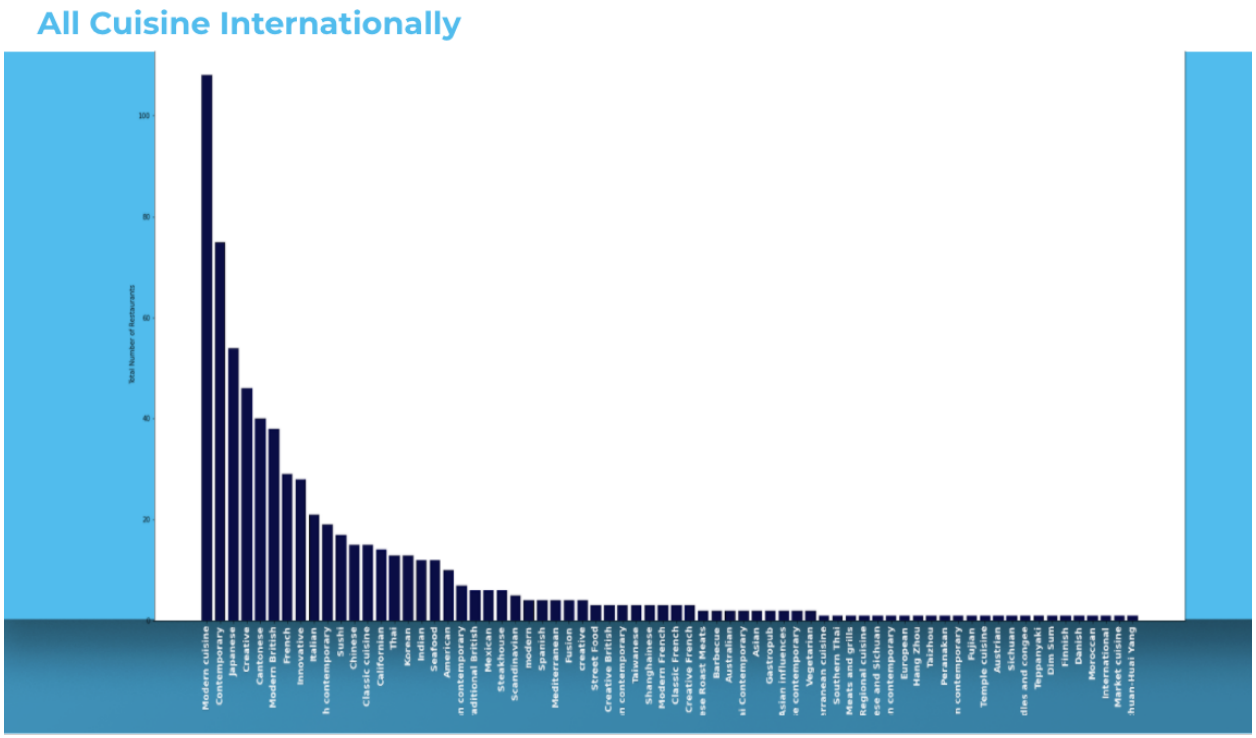
20.1


Average Price and Michelin Star Count for US Regions
Price Rate
Star Count
Scores
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
California
New York City
Chicago
Washington DC
Region

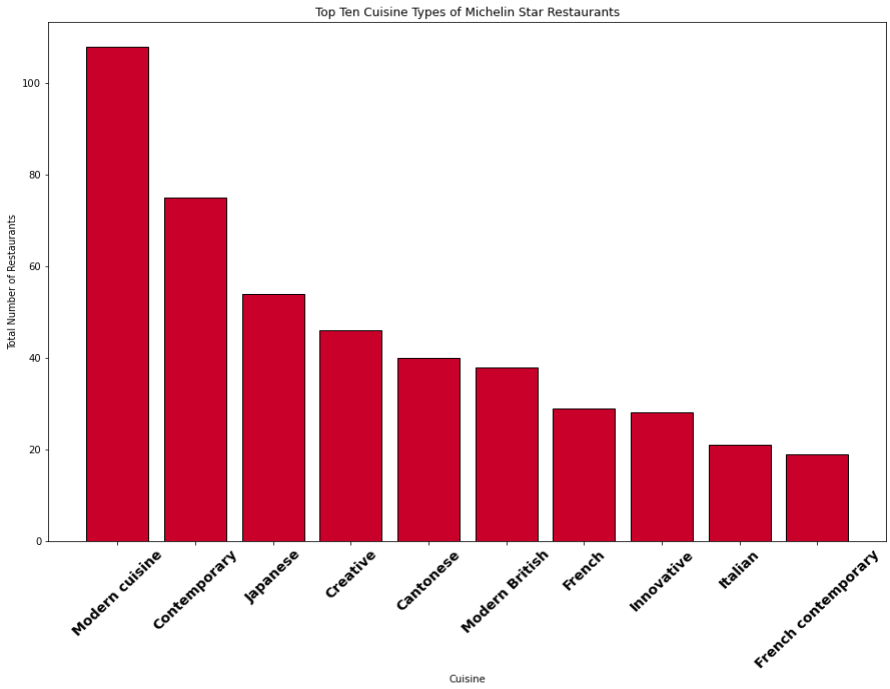
20.2


Average Price and Michelin Star Count for US Cuisine
Price Rate
Star Count
Scores
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Contemporary
Japanese
Californian
American
Italian
French
Mexican
Korean
Seafood
Fusion
Cuisine

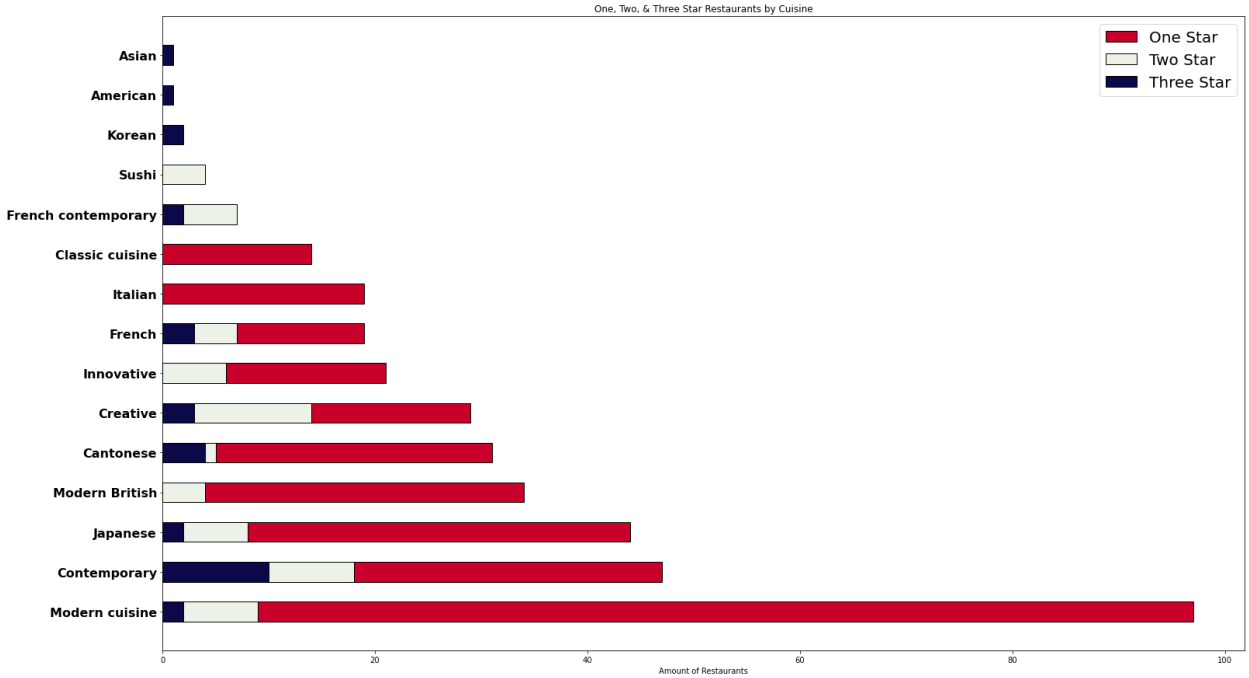
23


All Cuisine Internationally

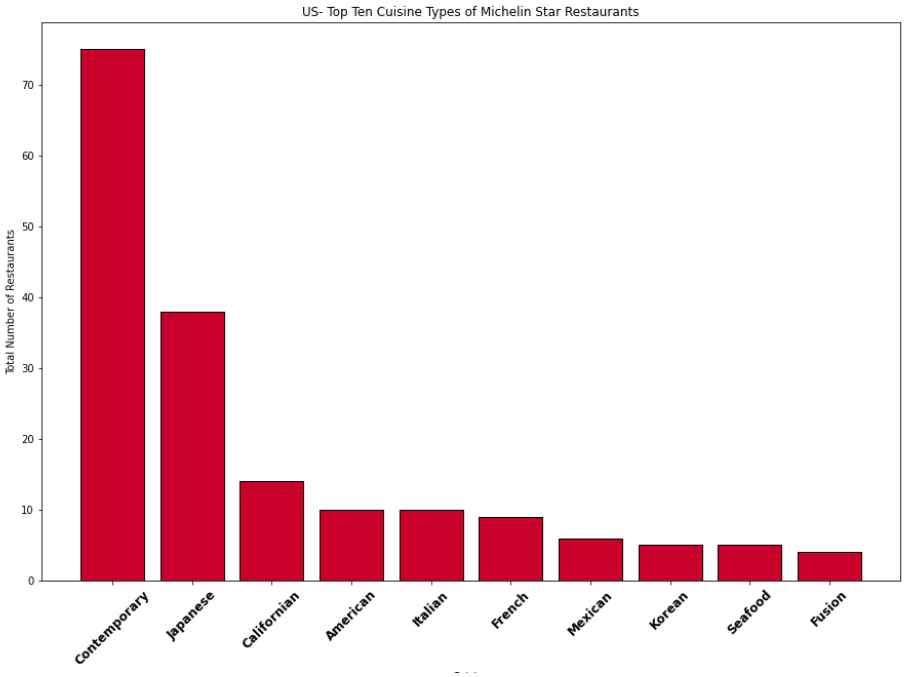
24


Top Ten Cuisine Types of Michelin Star Restaurants
Total Number of Restaurants
0
20
40
60
80
100
Modern cuisine
Contemporary
Japanese
Creative
Cantonese
Modern British
French
Innovative
Italian
French contemporary
Cuisine

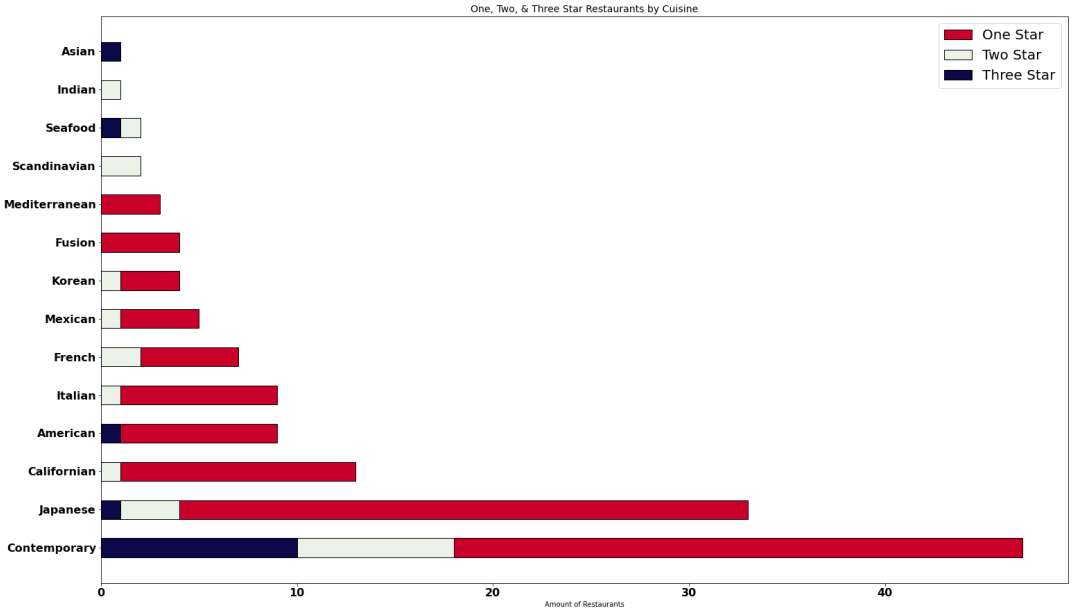
25


One, Two, & Three Star Restaurants by Cuisine
One Star
Two Star
Three Star
Asian
American
Korean
Sushi
French contemporary
Classic cuisine
Italian
French
Innovative
Creative
Cantonese
Modern British
Japanese
Contemporary
Modern cuisine
0
20
40
60
80
100
Amount of Restaurants

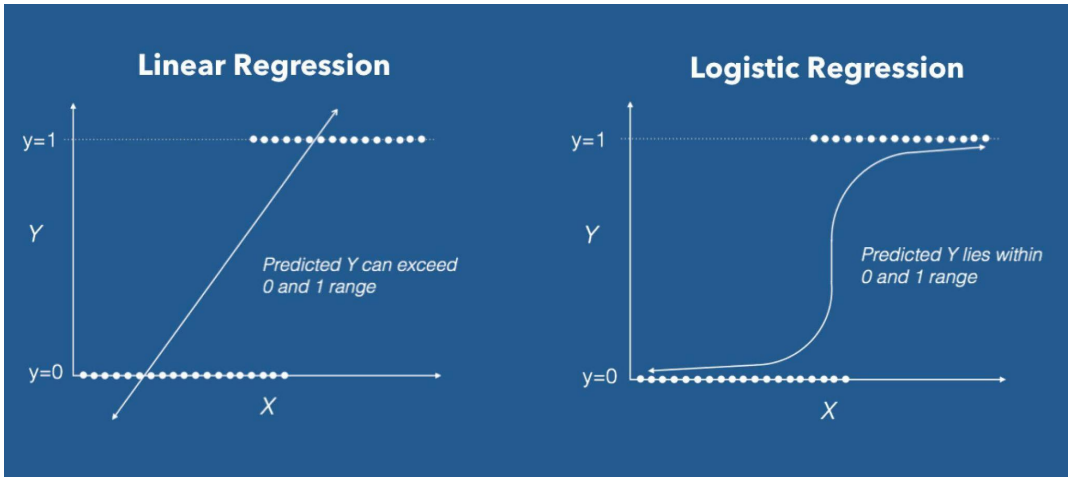
27


US- Top Ten Cuisine Types of Michelin Star Restaurants
Total Number of Restaurants
0
10
20
30
40
50
60
70
Contemporary
Japanese
Californian
American
Italian
French
Mexican
Korean
Seafood
Fusion

28


One, Two, & Three Star Restaurants by Cuisine
One Star
Two Star
Three Star
Asian
Indian
Seafood
Scandinavian
Mediterranean
Fusion
Korean
Mexican
French
Italian
American
Californian
Japanese
Contemporary
0
10
20
30
40
Amount of Restaurants

30


Linear Regression
y=1
Y
Predicted Y can exceed
0 and 1 range
y=0
X
Logistic Regression
y=1
Y
Predicted Y lies within
0 and 1 range
y=0
X

31

```
In [27]: predicted

Out[27]: array(['$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$', '$$', '$$',
                '$$', '$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$', '$$$$',
                '$$$$', '$$$$'], dtype=object)
```

32.1

```
In [28]: actual

Out[28]: 0      $$$$
         1       $$$
         2       $$$
         3       $$$
         4      $$$$
                ...
         175    $$$$
         176    $$$$
         177    $$$$
         178    $$$$
         179    $$$$
         Name: price, Length: 180, dtype: object
```

32.2

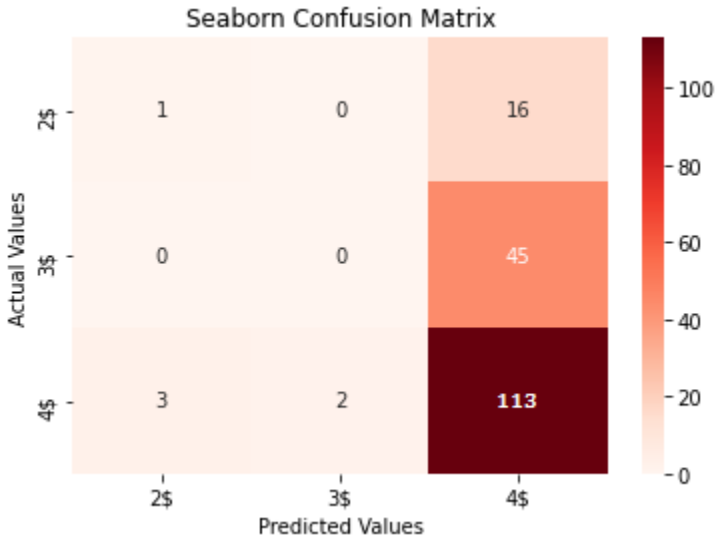
```
In [21]: mergedcensus_df.price.value_counts()

Out[21]: $$$$    118
         $$$      45
         $$       17
         Name: price, dtype: int64
```

33

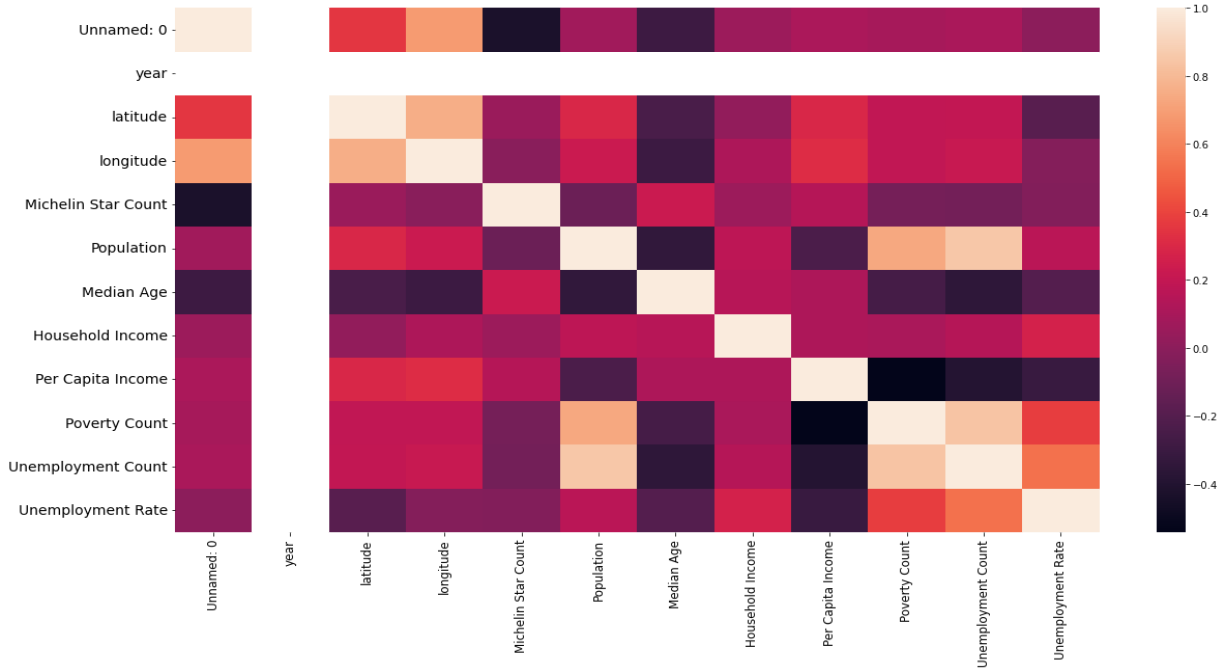
```
In [29]: probs
```

```
Out[29]: array([[0.07574935, 0.28126144, 0.6429892 ],
                [0.03103115, 0.20415998, 0.76480887],
                [0.13178075, 0.30858274, 0.55963651],
                [0.07642574, 0.25900651, 0.66456776],
                [0.09859288, 0.29648585, 0.60492127],
                [0.11192692, 0.30738191, 0.58069117],
                [0.2214439 , 0.35416454, 0.42439156],
                [0.11995433, 0.34186283, 0.53818285],
                [0.26237597, 0.34147989, 0.39614414],
                [0.26237597, 0.34147989, 0.39614414],
                [0.26237597, 0.34147989, 0.39614414],
                [0.26237557, 0.34147866, 0.39614577],
                [0.26237557, 0.34147866, 0.39614577],
                [0.06833885, 0.20077151, 0.73088964],
                [0.06833885, 0.20077151, 0.73088964],
                [0.06833885, 0.20077151, 0.73088964],
                [0.06833885, 0.20077151, 0.73088964],
                [0.01073561, 0.14847727, 0.84078712],
                [0.01073551, 0.14847533, 0.84078916],
```

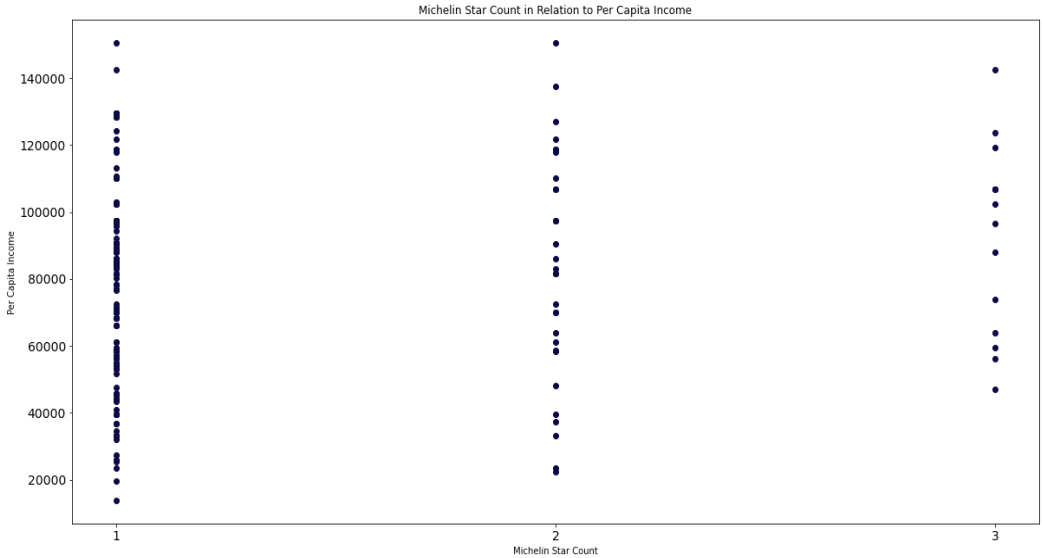
34



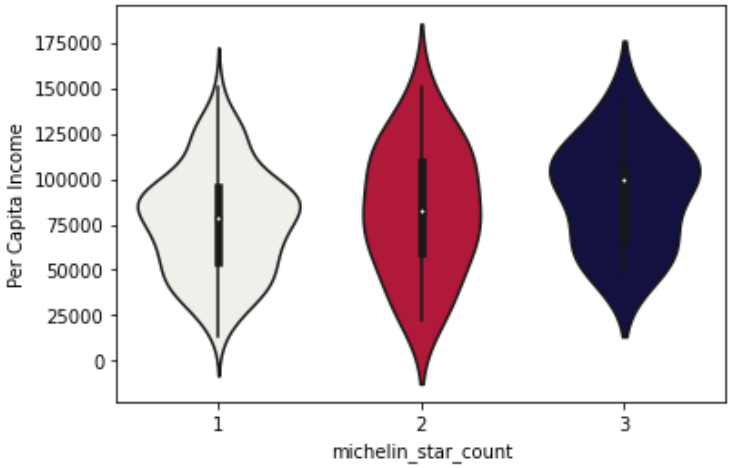
35


Unnamed: 0
year
latitude
longitude
Michelin Star Count
Population
Median Age
Household Income
Per Capita Income
Poverty Count
Unemployment Count
Unemployment Rate
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

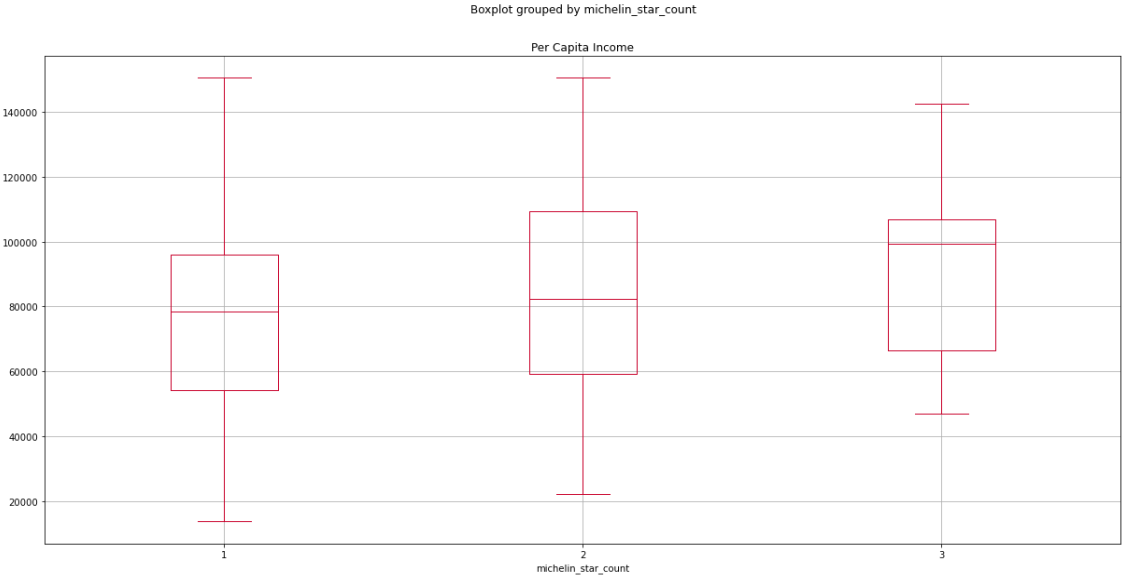
36


Michelin Star Count in Relation to Per Capita Income
Per Capita Income
140000
120000
100000
80000
60000
40000
20000
1
2
3
Michelin Star Count

37


175000
150000
125000
100000
75000
50000
25000
0
Per Capita Income
1
2
3
michelin_star_count

38


Boxplot grouped by michelin_star_count
Per Capita Income
140000
120000
100000
80000
60000
40000
20000
1
2
3
michelin_star_count

40.1


Chicago
2013