# Are Influencers Actually Influencing Economy?

## 1. Project One

### 1.1 Introduction

With the development of the Internet, online business platforms such as Yelp have become widespread in our lives and have becomes an increasingly important source for us when deciding what restaurant to go. Specialists said that USD1.5billion food online sales in 2020 and it is expected to reach USD2.5billion in 2025 with a compound annual growth rate (CAGR) of 10.8%. Moreover, Yelp's IPO Quarterly Report showed that in quarter 1 of 2021, there were 1.7 million new users on Yelp and 14.5 million consumers visited Yelp. Yelp's Chief Economist Michael Luca suggested that having a high star rating on Yelp is crucial for increasing revenue for businesses (well-prepared, restaurant, factors influencing online restaurant ratings, Michael Luca, Harvard Business School, 2021).

Because Yelp's crucial for consumers' decision-making, it is important to understand what factors influence the number of users and reviews. Reviews to a large number of influencers, called influencers, have been considered more practical and valuable, which helps an enterprise to increase its influence and sales. Yelp's Elite Squad is a group of Yelp's most active users who are invited to attend exclusive events and parties, and they are also given a badge to be recognized by other users.

Prior research has examined influencers' purpose. Zhu, Yin and Hu (2019) and Li, Mai and Liu (2014) have analyzed the relationship between the number of friends, the number of reviews, and the number of votes. The results show that the number of friends and the number of reviews have a positive impact on the number of votes. Although the number of friends and the number of reviews have a positive impact on the number of votes, the results suggests that promoting prolificacy through influencers is a great choice.

Although the Elite Squad is a group of influencers, the Elite Squad's purpose is to provide a more significant impact on the overall ratings of a restaurant. Although the Elite Squad's purpose is to provide a more significant impact on the overall ratings of a restaurant, the Elite Squad's business and roles is fleeting. Furthermore, the algorithm that calculates the overall star rating is more likely to be designed so that recent reviews have a greater impact than older reviews. Although the Elite Squad's purpose is to provide a more significant impact on the overall rating, it is unknown if the Elite Squad's users' attributes successfully encouraging users to follow using Yelp.

Although the Elite Squad's purpose is to provide a more significant impact on the overall rating of a restaurant, Xie, Zhang and Zhang (2014) stated that consumers have important doubts in a customer's positive review rating. Therefore, it is also important to investigate the impact of influencers' star ratings on a restaurant's overall star rating to increase its profitability.

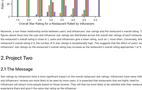
Although the Elite Squad's purpose is to provide a more significant impact on the overall rating of a restaurant, it is unknown if Yelp's Elite Squad actually attracted the attention of users and if the Elite Squad's star rating influence the number of reviews. Yelp has expanded since its founding in 2004. The Elite Squad that influencer star ratings significantly influences restaurant's overall star rating more than non-influencer star ratings, and influencer star ratings have a greater impact than non-influencer star ratings on the number of reviews. Therefore, it is worthwhile to collect data on reviews, the Elite Squad, and other factors that may influence restaurants' overall star rating. Yelp has expanded from its original business of providing reviews of restaurants to other businesses, including doctors, beauty salons, and repair services. Thus, it is worthwhile to collect data on other factors that may influence businesses' overall star rating, with the restaurant's overall star rating being known when the influencer's star rating is 1 and the rating being higher when the influencer star rating is higher.

### 1.2 Data Cleaning/Loading

#### 1.2.1 Rating and Report

Details

- Yelp Dataset: We obtained the Yelp Dataset in Kaggle. There are several subsets in the data, and we used a dataset on businesses, a dataset on reviews, and a dataset on users. First, we extracted data on businesses that belonged to Restaurants and Food. Second, we extracted data on reviews that were written for the restaurants in the business dataset. Third, we extracted data on users who wrote the reviews. We then combined the data with the Elite variable, including star rating for the restaurant, and the date of review. We also created a variable that indicates the number of years a user has been in the Elite Squad, the number of friends, the number of reviews, which proxies for LOEWUP observations. These reviews range from 2004 through 2021.
- Yelp's Elite Squad: Yelp's Elite Squad was created by Michael Luca and collected data on attributes of users who applied to the Elite Squad and who were accepted to the Elite Squad. Michael Luca's data includes the number of years a user has been in the Elite Squad and the number of years remaining in the Elite Squad. We added this data to the above dataset on the Elite.
- Yelp's Elite Squad Attributes: Yelp's Elite Squad attributes are collected by the Yelp's Elite Squad Recruiting Committee. Program uses collected data to determine which users are invited to the Elite Squad. Yelp's Elite Squad Attributes are used to calculate a "Quality" score for each user, which is used to LOEWUP observations. These reviews range from 2004 through 2012.
- Yelp's Elite Squad Attributes: Yelp's Elite Squad attributes are collected from an internal source per Yelp. The Elite Squad Attributes are used to calculate a "Quality" score for each user, which is used to LOEWUP observations. The Elite Squad Attributes are updated every three years




Moreover, a non-linear relationship exists between users and influencer ratings and the restaurant's overall rating. The two charts above show that influencers tend to rate restaurants higher than the average user, and that users tend to rate restaurants higher than influencers. Restaurant overall ratings tend to cluster around 3.5 stars, but users tend to cluster rating, such as influencers, around whole stars, while the restaurant's overall rating is 3.5, the number of 3 star ratings is exceedingly high. This suggests that the effect of each and influencer's rating on the restaurant's overall rating is dependent on the restaurant's overall rating expectation.

## 2. Project Two

### 2.1 The Message

Star ratings by influencers have a more significant impact on the overall restaurant star rating. Influencers have many followers, and influencer reviews are more likely to users by more users. It is expected that restaurants that are highly rated by influencers will attract more people based on those reviews. They will then be more likely to be satisfied with their restaurant experience and rate it higher than the average user.

Yelp's mission is to connect people to the best local businesses. Yelp's star rating is intended to be a fair index that uses the average of star scores which all users in an overall star rating for the restaurant. Influencers' ratings of restaurants tend to be higher than the average user, and influencers have more followers than the average user. Influencers' ratings of restaurants are more likely to attract users, and because the rate of star ratings of restaurants by influencers is higher. Therefore, as expected the influencers' ratings of restaurants have a more significant impact on the overall star rating of restaurants than the average user. Influencers' ratings of restaurants which are more reliable than other users, are reflected in the restaurant's overall rating.

### 2.2 Maps and Infographics

```
gd_data2 = df[df.Date > '2018-01-01']
gd_data2 = gd_data2[gd_data2.Influencers_Rating.isin(['Yes', 'No'])]
gd_data2 = gd_data2[gd_data2.Overall_Rating.isin(['1', '2', '3', '4', '5'])]
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.astype('category')
gd_data2.Overall_Rating = gd_data2.Overall_Rating.astype('int')
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.cat.rename_categories(['Non-Influencers', 'Influencers'])
gd_data2 = gd_data2[['Influencers_Rating', 'Overall_Rating']]
gd_data2 = gd_data2.groupby(['Influencers_Rating', 'Overall_Rating'])
gd_data2 = gd_data2.size().reset_index(name='Influencers_Count')
gd_data2 = gd_data2.sort_values(['Influencers_Rating', 'Overall_Rating'])
gd_data2.head(10)
```

```
  Influencers_Rating  Overall_Rating  Influencers_Count
0       Influencers               1                 16
1       Influencers               2                 19
2       Influencers               3                 71
3       Influencers               4                 71
4       Influencers               5                 23
5   Non-Influencers               1                 16
6   Non-Influencers               2                 19
7   Non-Influencers               3                 71
8   Non-Influencers               4                 71
9   Non-Influencers               5                 23
```

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```

Difference in Number of Ratings by User for a Restaurant Rated

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Influencers_Count", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Influencers Count")
plt.xlabel("Influencers_Rating")
plt.ylabel("Influencers_Count")
```

Difference in Number of Ratings by User for a Restaurant Rated

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```

Star Rating by Influencers

Rate of Star Ratings by Non-Influencers to Each Star Rating by Influencers

```
gd_data2 = gd_data2[gd_data2.Influencers_Rating == 'Non-Influencers']
gd_data2 = gd_data2[['Overall_Rating', 'Influencers_Count']]
gd_data2 = gd_data2.groupby(['Overall_Rating'])
gd_data2 = gd_data2.size().reset_index(name='Influencers_Count')
gd_data2 = gd_data2.sort_values(['Overall_Rating'])
gd_data2.head(10)
```

```
   Overall_Rating  Influencers_Count
0               1                 16
1               2                 19
2               3                 71
3               4                 71
4               5                 23
```

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Overall_Rating", y="Influencers_Count", data=gd_data2, palette="Paired")
plt.title("Rate of Star Ratings by Non-Influencers to Each Star Rating by Influencers")
plt.xlabel("Overall_Rating")
plt.ylabel("Influencers_Count")
```

Rate of Star Ratings by Non-Influencers to Each Star Rating by Influencers

```
gd_data2 = df[df.Influencers_Rating == 'Influencers']
gd_data2 = gd_data2[gd_data2.Influencers_Rating.isin(['Yes', 'No'])]
gd_data2 = gd_data2[gd_data2.Overall_Rating.isin(['1', '2', '3', '4', '5'])]
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.astype('category')
gd_data2.Overall_Rating = gd_data2.Overall_Rating.astype('int')
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.cat.rename_categories(['Non-Influencers', 'Influencers'])
gd_data2 = gd_data2[['Influencers_Rating', 'Overall_Rating']]
gd_data2 = gd_data2.groupby(['Influencers_Rating', 'Overall_Rating'])
gd_data2 = gd_data2.size().reset_index(name='Influencers_Count')
gd_data2 = gd_data2.sort_values(['Influencers_Rating', 'Overall_Rating'])
gd_data2.head(10)
```

```
  Influencers_Rating  Overall_Rating  Influencers_Count
0       Influencers               1                 16
1       Influencers               2                 19
2       Influencers               3                 71
3       Influencers               4                 71
4       Influencers               5                 23
```

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Overall_Rating", y="Influencers_Count", data=gd_data2, palette="Paired")
plt.title("Rate of Star Ratings by Influencers to Each Star Rating by Influencers")
plt.xlabel("Overall_Rating")
plt.ylabel("Influencers_Count")
```

Rate of Star Ratings by Influencers to Each Star Rating by Influencers

```
gd_data2 = df[df.Influencers_Rating == 'Influencers']
gd_data2 = gd_data2[gd_data2.Influencers_Rating.isin(['Yes', 'No'])]
gd_data2 = gd_data2[gd_data2.Overall_Rating.isin(['1', '2', '3', '4', '5'])]
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.astype('category')
gd_data2.Overall_Rating = gd_data2.Overall_Rating.astype('int')
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.cat.rename_categories(['Non-Influencers', 'Influencers'])
gd_data2 = gd_data2[['Influencers_Rating', 'Overall_Rating']]
gd_data2 = gd_data2.groupby(['Influencers_Rating', 'Overall_Rating'])
gd_data2 = gd_data2.size().reset_index(name='Influencers_Count')
gd_data2 = gd_data2.sort_values(['Influencers_Rating', 'Overall_Rating'])
gd_data2.head(10)
```

```
  Influencers_Rating  Overall_Rating  Influencers_Count
0       Influencers               1                 16
1       Influencers               2                 19
2       Influencers               3                 71
3       Influencers               4                 71
4       Influencers               5                 23
```

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```

Difference in Relationships of Ratings by User for a Restaurant Rated

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Influencers_Count", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Influencers Count")
plt.xlabel("Influencers_Rating")
plt.ylabel("Influencers_Count")
```

Difference in Relationships of Ratings by User for a Restaurant Rated

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```

Most Common Star Rating for a Restaurant

```
gd_data2 = df[df.Date > '2018-01-01']
gd_data2 = gd_data2[gd_data2.Influencers_Rating.isin(['Yes', 'No'])]
gd_data2 = gd_data2[gd_data2.Overall_Rating.isin(['1', '2', '3', '4', '5'])]
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.astype('category')
gd_data2.Overall_Rating = gd_data2.Overall_Rating.astype('int')
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.cat.rename_categories(['Non-Influencers', 'Influencers'])
gd_data2 = gd_data2[['Influencers_Rating', 'Overall_Rating']]
gd_data2 = gd_data2.groupby(['Influencers_Rating', 'Overall_Rating'])
gd_data2 = gd_data2.size().reset_index(name='Influencers_Count')
gd_data2 = gd_data2.sort_values(['Influencers_Rating', 'Overall_Rating'])
gd_data2.head(10)
```

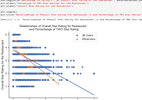
```
  Influencers_Rating  Overall_Rating  Influencers_Count
0       Influencers               1                 16
1       Influencers               2                 19
2       Influencers               3                 71
3       Influencers               4                 71
4       Influencers               5                 23
```

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```

Most Common Star Rating for a Restaurant

Whether the Restaurant Received 1 Star Rating by Influencers

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Influencers_Count", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Influencers Count")
plt.xlabel("Influencers_Rating")
plt.ylabel("Influencers_Count")
```




Most Common Star Rating for a Restaurant

Whether the Restaurant Received 1 Star Rating by Influencers

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```

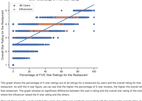
Most Common Star Rating for a Restaurant

Whether the Restaurant Received 1 Star Rating by Influencers

```
gd_data2 = df[df.Date > '2018-01-01']
gd_data2 = gd_data2[gd_data2.Influencers_Rating.isin(['Yes', 'No'])]
gd_data2 = gd_data2[gd_data2.Overall_Rating.isin(['1', '2', '3', '4', '5'])]
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.astype('category')
gd_data2.Overall_Rating = gd_data2.Overall_Rating.astype('int')
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.cat.rename_categories(['Non-Influencers', 'Influencers'])
gd_data2 = gd_data2[['Influencers_Rating', 'Overall_Rating']]
gd_data2 = gd_data2.groupby(['Influencers_Rating', 'Overall_Rating'])
gd_data2 = gd_data2.size().reset_index(name='Influencers_Count')
gd_data2 = gd_data2.sort_values(['Influencers_Rating', 'Overall_Rating'])
gd_data2.head(10)
```

```
  Influencers_Rating  Overall_Rating  Influencers_Count
0       Influencers               1                 16
1       Influencers               2                 19
2       Influencers               3                 71
3       Influencers               4                 71
4       Influencers               5                 23
```

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```




Most Common Star Rating for a Restaurant

Whether the Restaurant Received 1 Star Rating by Influencers

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Influencers_Count", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Influencers Count")
plt.xlabel("Influencers_Rating")
plt.ylabel("Influencers_Count")
```

Most Common Star Rating for a Restaurant

Whether the Restaurant Received 1 Star Rating by Influencers

Relationship between Overall Star Rating of Restaurant and Star Ratings by Users

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```




Relationship between Overall Star Rating of Restaurant and Star Ratings by Users

The chart above shows the relationship between a restaurant's overall star rating and the highest star rating the restaurant received from users. The chart shows that restaurants with a 1 star rating from influencers are more likely to have a 2.5 or 3 star rating from users. The chart also shows that restaurants with a 2 star rating from influencers are more likely to have a 3 or 4 star rating from users. The chart also shows that restaurants with a 3 star rating from influencers are more likely to have a 3 or 4 star rating from users. The chart also shows that restaurants with a 4 star rating from influencers are more likely to have a 4 or 5 star rating from users. The chart also shows that restaurants with a 5 star rating from influencers are more likely to have a 4 or 5 star rating from users.

```
gd_data2 = df[df.Date > '2018-01-01']
gd_data2 = gd_data2[gd_data2.Influencers_Rating.isin(['Yes', 'No'])]
gd_data2 = gd_data2[gd_data2.Overall_Rating.isin(['1', '2', '3', '4', '5'])]
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.astype('category')
gd_data2.Overall_Rating = gd_data2.Overall_Rating.astype('int')
gd_data2.Influencers_Rating = gd_data2.Influencers_Rating.cat.rename_categories(['Non-Influencers', 'Influencers'])
gd_data2 = gd_data2[['Influencers_Rating', 'Overall_Rating']]
gd_data2 = gd_data2.groupby(['Influencers_Rating', 'Overall_Rating'])
gd_data2 = gd_data2.size().reset_index(name='Influencers_Count')
gd_data2 = gd_data2.sort_values(['Influencers_Rating', 'Overall_Rating'])
gd_data2.head(10)
```

```
  Influencers_Rating  Overall_Rating  Influencers_Count
0       Influencers               1                 16
1       Influencers               2                 19
2       Influencers               3                 71
3       Influencers               4                 71
4       Influencers               5                 23
```

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```

Most Common Star Rating for a Restaurant

Whether the Restaurant Received 1 Star Rating by Influencers

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Influencers_Count", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Influencers Count")
plt.xlabel("Influencers_Rating")
plt.ylabel("Influencers_Count")
```




Most Common Star Rating for a Restaurant

Whether the Restaurant Received 1 Star Rating by Influencers

Relationship between Overall Star Rating of Restaurant and Star Ratings by Users

```
%matplotlib inline
sns.set(style="ticks")
sns.boxplot(x="Influencers_Rating", y="Overall_Rating", data=gd_data2, palette="Paired")
plt.title("Influencers Rating Vs Overall Rating")
plt.xlabel("Influencers_Rating")
plt.ylabel("Overall_Rating")
```




Relationship between Overall Star Rating of Restaurant and Star Ratings by Users

```
# Overall Star Ratings for Restaurants
# Defined by Major Highway and Secondary Highway
# Overall Star Ratings for Restaurants in Las Vegas
# Defined by Major Highway and Secondary Highway
# Average of Overall Star Ratings by Users in Each Zipcode
# Defined by Major Highway and Secondary Highway
# Average of Star Ratings by Users in Each Zipcode
# Defined by Major Highway and Secondary Highway
# Average of Star Ratings by Influencers in Each Zipcode
# Defined by Major Highway and Secondary Highway
# Overall Star Ratings for Restaurants in Charlotte
# Defined by Major Highway and Secondary Highway
# Average of Overall Star Ratings in Each Zipcode
# Defined by Major Highway and Secondary Highway
# Average of Star Ratings by Users in Each Zipcode
# Defined by Major Highway and Secondary Highway
# Average of Star Ratings by Influencers in Each Zipcode
# Defined by Major Highway and Secondary Highway
```

## 3. Project Three

### 3.1 Potential Data to Scrape

Since Yelp's API is limited, and Foursquare's API is limited in scope, we can go to a restaurant without having had data about going there, or a restaurant. Yelp's API limits us to 5 reviews per restaurant, and Foursquare's API only allows for restaurant information to be scraped, and not their evaluation of the restaurant and its fares.

The Yelp's data is limited to restaurants in the United States in one. Therefore, we'll be limited to the data that influences the quality of a restaurant in the United States. Yelp's API limits us to 5 reviews per restaurant, and Foursquare's API only allows for information to be scraped for a restaurant's location and not its evaluation or a review of the number of opening spaces.

The website should be a dashboard of the overall star rating by restaurants and per compiled UR of patrons' star ratings because we want to see how the star rating changes based on the number of people who have visited the restaurant. We want to see if there is a difference between those who actually go to eat and those who have the restaurant to critical to get good, and the number of opening spaces could influence the star rating because people

```
Sinaloa                    24803
Durango                    24659
Guerrero                   24314
Morelos                    23659
Jalisco                    23522
Nuevo Leon                 22851
Baja California            22808
Baja California Sur        21548
Sonora                     21486
Yucatan                    21453
Campeche                   21353
Q.R. Chetumal              21352
Tlaxcala                   21251
Nayarit                    21000
Quintana Roo               20900
Northern Mariana Islands   20416
Colima                     20000
Guam                       19900
```

A dataset of average per capita income was successfully created by the code above.

### 3.4 Merging the Scraped Data

The scraped dataset is going to be merged to the dataset of restaurants rated at Yelp on some of the states below.

```
restaurantsincome_businessdesp_merge
```

```
(0) 0
(1) 1
```

```
restaurantsincome_businessdesp_merge.shape(0)
```

```
(1) 15000
```

```
restaurantsincome_businessdesp_merge.sample(10).head()
```

```
(2)   businessid_business_desp_x  latitude  longitude  ...  \
0                 0x8d0ef0ef0ef0  29.6516  -82.3248  ...
1                 0x8d0ef0ef0ef1  29.6516  -82.3248  ...
2                 0x8d0ef0ef0ef2  29.6516  -82.3248  ...
3                 0x8d0ef0ef0ef3  29.6516  -82.3248  ...
4                 0x8d0ef0ef0ef4  29.6516  -82.3248  ...

   businessid_business_desp_y  \
0                         NaN
1                         NaN
2                         NaN
3                         NaN
4                         NaN

   business_name_business_desp  \
0                          NaN
1                          NaN
2                          NaN
3                          NaN
4                          NaN

   business_address_business_desp  \
0                              NaN
1                              NaN
2                              NaN
3                              NaN
4                              NaN

   business_city_business_desp  \
0                          NaN
1                          NaN
2                          NaN
3                          NaN
4                          NaN

   business_state_business_desp  \
0                           NaN
1                           NaN
2                           NaN
3                           NaN
4                           NaN

   business_postal_code_business_desp  \
0                                 NaN
1                                 NaN
2                                 NaN
3                                 NaN
4                                 NaN

   business_latitude_business_desp  \
0                              NaN
1                              NaN
2                              NaN
3                              NaN
4                              NaN

   business_longitude_business_desp  \
0                               NaN
1                               NaN
2                               NaN
3                               NaN
4                               NaN

   business_stars_business_desp  \
0                           NaN
1                           NaN
2                           NaN
3                           NaN
4                           NaN

   business_review_count_business_desp  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN

   business_is_open_business_desp  \
0                             NaN
1                             NaN
2                             NaN
3                             NaN
4                             NaN

   business_attributes_business_desp  \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN

   business_categories_business_desp  \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN

   business_hours_business_desp  \
0                           NaN
1                           NaN
2                           NaN
3                           NaN
4                           NaN

   business_type_business_desp  \
0                          NaN
1                          NaN
2                          NaN
3                          NaN
4                          NaN

   businessid_business_desp_y.1  \
0                           NaN
1                           NaN
2                           NaN
3                           NaN
4                           NaN

   business_name_business_desp.1  \
0                            NaN
1                            NaN
2                            NaN
3                            NaN
4                            NaN

   business_address_business_desp.1  \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN

   business_city_business_desp.1  \
0                            NaN
1                            NaN
2                            NaN
3                            NaN
4                            NaN

   business_state_business_desp.1  \
0                             NaN
1                             NaN
2                             NaN
3                             NaN
4                             NaN

   business_postal_code_business_desp.1  \
0                                   NaN
1                                   NaN
2                                   NaN
3                                   NaN
4                                   NaN

   business_latitude_business_desp.1  \
0                                NaN
1                                NaN
2                                NaN
3                                NaN
4                                NaN

   business_longitude_business_desp.1  \
0                                 NaN
1                                 NaN
2                                 NaN
3                                 NaN
4                                 NaN

   business_stars_business_desp.1  \
0                             NaN
1                             NaN
2                             NaN
3                             NaN
4                             NaN

   business_review_count_business_desp.1  \
0                                    NaN
1                                    NaN
2                                    NaN
3                                    NaN
4                                    NaN

   business_is_open_business_desp.1  \
0                               NaN
1                               NaN
2                               NaN
3                               NaN
4                               NaN

   business_attributes_business_desp.1  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN

   business_categories_business_desp.1  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN

   business_hours_business_desp.1  \
0                             NaN
1                             NaN
2                             NaN
3                             NaN
4                             NaN

   business_type_business_desp.1  \
0                            NaN
1                            NaN
2                            NaN
3                            NaN
4                            NaN

   zipcode_income_income_merge  \
0                        32601
1                        32601
2                        32601
3                        32601
4                        32601

   average_income_income_merge  \
0                      24659
1                      24659
2                      24659
3                      24659
4                      24659

   state_income_income_merge  \
0                         FL
1                         FL
2                         FL
3                         FL
4                         FL

   _merge_income_income_merge  \
0                        both
1                        both
2                        both
3                        both
4                        both

   businessid_restaurantsincome_businessdesp_merge  \
0                                     0x8d0ef0ef0ef0
1                                     0x8d0ef0ef0ef1
2                                     0x8d0ef0ef0ef2
3                                     0x8d0ef0ef0ef3
4                                     0x8d0ef0ef0ef4

   businessid_restaurantsincome_businessdesp_merge.1  \
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN

   _merge_restaurantsincome_businessdesp_merge
0                                        both
1                                        both
2                                        both
3                                        both
4                                        both
```

```
restaurantsincome_businessdesp_merge.columns.tolist()
```

```
(3) ['businessid_business_desp_x',
 'latitude',
 'longitude',
 'zipcode',
 'city',
 'state',
 'stars',
 'review_count',
 'is_open',
 'attributes',
 'categories',
 'hours',
 'type',
 'business_id',
 'name',
 'address',
 'city',
 'state',
 'postal_code',
 'latitude',
 'longitude',
 'stars',
 'review_count',
 'is_open',
 'attributes',
 'categories',
 'hours',
 'type',
'businessid_business_desp_y',
'business_name_business_desp',
'business_address_business_desp',
'business_city_business_desp',
'business_state_business_desp',
'business_postal_code_business_desp',
'business_latitude_business_desp',
'business_longitude_business_desp',
'business_stars_business_desp',
'business_review_count_business_desp',
'business_is_open_business_desp',
'business_attributes_business_desp',
'business_categories_business_desp',
'business_hours_business_desp',
'business_type_business_desp',
'businessid_business_desp_y.1',
'business_name_business_desp.1',
'business_address_business_desp.1',
'business_city_business_desp.1',
'business_state_business_desp.1',
'business_postal_code_business_desp.1',
'business_latitude_business_desp.1',
'business_longitude_business_desp.1',
'business_stars_business_desp.1',
'business_review_count_business_desp.1',
'business_is_open_business_desp.1',
'business_attributes_business_desp.1',
'business_categories_business_desp.1',
'business_hours_business_desp.1',
'business_type_business_desp.1',
'zipcode_income_income_merge',
'average_income_income_merge',
'state_income_income_merge',
'_merge_income_income_merge',
'businessid_restaurantsincome_businessdesp_merge',
'businessid_restaurantsincome_businessdesp_merge.1',
'_merge_restaurantsincome_businessdesp_merge']
```

```
restaurantsincome_businessdesp_merge.isna().sum()
```

```
(4) business_id                                                  0
 name                                                        1500
 address                                                     1500
 city                                                        1500
 state                                                       1500
 postal_code                                                 1500
 latitude                                                    1500
 longitude                                                   1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 business_id                                                 1500
 name                                                        1500
 address                                                     1500
 city                                                        1500
 state                                                       1500
 postal_code                                                 1500
 latitude                                                    1500
 longitude                                                   1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 businessid_business_desp_x                                  1500
 latitude                                                    1500
 longitude                                                   1500
 zipcode                                                     1500
 city                                                        1500
 state                                                       1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 businessid_business_desp_y                                  1500
 business_name_business_desp                                 1500
 business_address_business_desp                              1500
 business_city_business_desp                                 1500
 business_state_business_desp                                1500
 business_postal_code_business_desp                          1500
 business_latitude_business_desp                             1500
 business_longitude_business_desp                            1500
 business_stars_business_desp                                1500
 business_review_count_business_desp                         1500
 business_is_open_business_desp                              1500
 business_attributes_business_desp                           1500
 business_categories_business_desp                           1500
 business_hours_business_desp                                1500
 business_type_business_desp                                 1500
 businessid_business_desp_y.1                                1500
 business_name_business_desp.1                               1500
 business_address_business_desp.1                            1500
 business_city_business_desp.1                               1500
 business_state_business_desp.1                              1500
 business_postal_code_business_desp.1                        1500
 business_latitude_business_desp.1                           1500
 business_longitude_business_desp.1                          1500
 business_stars_business_desp.1                              1500
 business_review_count_business_desp.1                       1500
 business_is_open_business_desp.1                            1500
 business_attributes_business_desp.1                         1500
 business_categories_business_desp.1                         1500
 business_hours_business_desp.1                              1500
 business_type_business_desp.1                               1500
 zipcode_income_income_merge                                    0
 average_income_income_merge                                    0
 state_income_income_merge                                      0
 _merge_income_income_merge                                     0
 businessid_restaurantsincome_businessdesp_merge                0
 businessid_restaurantsincome_businessdesp_merge.1           1500
 _merge_restaurantsincome_businessdesp_merge                    0
 dtype: int64
```

```
restaurantsincome_businessdesp_merge.dropna(subset=['average_income_income_merge'].sample(10).head()
```

```
(5)   business_id  name  address  city  state  postal_code  latitude  \
0                NaN   NaN      NaN   NaN    NaN          NaN       NaN
1                NaN   NaN      NaN   NaN    NaN          NaN       NaN
2                NaN   NaN      NaN   NaN    NaN          NaN       NaN
3                NaN   NaN      NaN   NaN    NaN          NaN       NaN
4                NaN   NaN      NaN   NaN    NaN          NaN       NaN

   longitude  stars  review_count  is_open  attributes  categories  \
0        NaN    NaN           NaN      NaN         NaN         NaN
1        NaN    NaN           NaN      NaN         NaN         NaN
2        NaN    NaN           NaN      NaN         NaN         NaN
3        NaN    NaN           NaN      NaN         NaN         NaN
4        NaN    NaN           NaN      NaN         NaN         NaN

   hours  type  business_id  name  address  city  state  postal_code  \
0    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
1    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
2    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
3    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
4    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN

   latitude  longitude  stars  review_count  is_open  attributes  \
0       NaN        NaN    NaN           NaN      NaN         NaN
1       NaN        NaN    NaN           NaN      NaN         NaN
2       NaN        NaN    NaN           NaN      NaN         NaN
3       NaN        NaN    NaN           NaN      NaN         NaN
4       NaN        NaN    NaN           NaN      NaN         NaN

   categories  hours  type  businessid_business_desp_x  latitude  \
0         NaN    NaN   NaN                 0x8d0ef0ef0ef0  29.6516
1         NaN    NaN   NaN                 0x8d0ef0ef0ef1  29.6516
2         NaN    NaN   NaN                 0x8d0ef0ef0ef2  29.6516
3         NaN    NaN   NaN                 0x8d0ef0ef0ef3  29.6516
4         NaN    NaN   NaN                 0x8d0ef0ef0ef4  29.6516

   longitude  zipcode  city  state  stars  review_count  is_open  \
0  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
1  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
2  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
3  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
4  -82.3248    32601   NaN    NaN    NaN           NaN      NaN

   attributes  categories  hours  type  businessid_business_desp_y  \
0         NaN         NaN    NaN   NaN                         NaN
1         NaN         NaN    NaN   NaN                         NaN
2         NaN         NaN    NaN   NaN                         NaN
3         NaN         NaN    NaN   NaN                         NaN
4         NaN         NaN    NaN   NaN                         NaN

   business_name_business_desp  business_address_business_desp  \
0                          NaN                             NaN
1                          NaN                             NaN
2                          NaN                             NaN
3                          NaN                             NaN
4                          NaN                             NaN

   business_city_business_desp  business_state_business_desp  \
0                          NaN                           NaN
1                          NaN                           NaN
2                          NaN                           NaN
3                          NaN                           NaN
4                          NaN                           NaN

   business_postal_code_business_desp  business_latitude_business_desp  \
0                                 NaN                              NaN
1                                 NaN                              NaN
2                                 NaN                              NaN
3                                 NaN                              NaN
4                                 NaN                              NaN

   business_longitude_business_desp  business_stars_business_desp  \
0                               NaN                           NaN
1                               NaN                           NaN
2                               NaN                           NaN
3                               NaN                           NaN
4                               NaN                           NaN

   business_review_count_business_desp  business_is_open_business_desp  \
0                                  NaN                             NaN
1                                  NaN                             NaN
2                                  NaN                             NaN
3                                  NaN                             NaN
4                                  NaN                             NaN

   business_attributes_business_desp  business_categories_business_desp  \
0                                NaN                                NaN
1                                NaN                                NaN
2                                NaN                                NaN
3                                NaN                                NaN
4                                NaN                                NaN

   business_hours_business_desp  business_type_business_desp  \
0                           NaN                          NaN
1                           NaN                          NaN
2                           NaN                          NaN
3                           NaN                          NaN
4                           NaN                          NaN

   businessid_business_desp_y.1  business_name_business_desp.1  \
0                           NaN                            NaN
1                           NaN                            NaN
2                           NaN                            NaN
3                           NaN                            NaN
4                           NaN                            NaN

   business_address_business_desp.1  business_city_business_desp.1  \
0                               NaN                            NaN
1                               NaN                            NaN
2                               NaN                            NaN
3                               NaN                            NaN
4                               NaN                            NaN

   business_state_business_desp.1  business_postal_code_business_desp.1  \
0                             NaN                                   NaN
1                             NaN                                   NaN
2                             NaN                                   NaN
3                             NaN                                   NaN
4                             NaN                                   NaN

   business_latitude_business_desp.1  business_longitude_business_desp.1  \
0                               NaN                                NaN
1                               NaN                                NaN
2                               NaN                                NaN
3                               NaN                                NaN
4                               NaN                                NaN

   business_stars_business_desp.1  business_review_count_business_desp.1  \
0                             NaN                                    NaN
1                             NaN                                    NaN
2                             NaN                                    NaN
3                             NaN                                    NaN
4                             NaN                                    NaN

   business_is_open_business_desp.1  business_attributes_business_desp.1  \
0                               NaN                                  NaN
1                               NaN                                  NaN
2                               NaN                                  NaN
3                               NaN                                  NaN
4                               NaN                                  NaN

   business_categories_business_desp.1  business_hours_business_desp.1  \
0                                  NaN                             NaN
1                                  NaN                             NaN
2                                  NaN                             NaN
3                                  NaN                             NaN
4                                  NaN                             NaN

   business_type_business_desp.1  zipcode_income_income_merge  \
0                            NaN                        32601
1                            NaN                        32601
2                            NaN                        32601
3                            NaN                        32601
4                            NaN                        32601

   average_income_income_merge  state_income_income_merge  \
0                      24659                         FL
1                      24659                         FL
2                      24659                         FL
3                      24659                         FL
4                      24659                         FL

   _merge_income_income_merge  \
0                        both
1                        both
2                        both
3                        both
4                        both

   businessid_restaurantsincome_businessdesp_merge  \
0                                     0x8d0ef0ef0ef0
1                                     0x8d0ef0ef0ef1
2                                     0x8d0ef0ef0ef2
3                                     0x8d0ef0ef0ef3
4                                     0x8d0ef0ef0ef4

   businessid_restaurantsincome_businessdesp_merge.1  \
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN

   _merge_restaurantsincome_businessdesp_merge
0                                        both
1                                        both
2                                        both
3                                        both
4                                        both
```

```
restaurantsincome_businessdesp_merge.columns.tolist()
```

```
(6) ['business_id',
 'name',
 'address',
 'city',
 'state',
 'postal_code',
 'latitude',
 'longitude',
 'stars',
 'review_count',
 'is_open',
 'attributes',
 'categories',
 'hours',
 'type',
 'business_id',
 'name',
 'address',
 'city',
 'state',
 'postal_code',
 'latitude',
 'longitude',
 'stars',
 'review_count',
 'is_open',
 'attributes',
 'categories',
 'hours',
 'type',
'businessid_business_desp_x',
'latitude',
'longitude',
'zipcode',
'city',
'state',
'stars',
'review_count',
'is_open',
'attributes',
'categories',
'hours',
'type',
'businessid_business_desp_y',
'business_name_business_desp',
'business_address_business_desp',
'business_city_business_desp',
'business_state_business_desp',
'business_postal_code_business_desp',
'business_latitude_business_desp',
'business_longitude_business_desp',
'business_stars_business_desp',
'business_review_count_business_desp',
'business_is_open_business_desp',
'business_attributes_business_desp',
'business_categories_business_desp',
'business_hours_business_desp',
'business_type_business_desp',
'businessid_business_desp_y.1',
'business_name_business_desp.1',
'business_address_business_desp.1',
'business_city_business_desp.1',
'business_state_business_desp.1',
'business_postal_code_business_desp.1',
'business_latitude_business_desp.1',
'business_longitude_business_desp.1',
'business_stars_business_desp.1',
'business_review_count_business_desp.1',
'business_is_open_business_desp.1',
'business_attributes_business_desp.1',
'business_categories_business_desp.1',
'business_hours_business_desp.1',
'business_type_business_desp.1',
'zipcode_income_income_merge',
'average_income_income_merge',
'state_income_income_merge',
'_merge_income_income_merge',
'businessid_restaurantsincome_businessdesp_merge',
'businessid_restaurantsincome_businessdesp_merge.1',
'_merge_restaurantsincome_businessdesp_merge']
```

```
restaurantsincome_businessdesp_merge.isna().sum()
```

```
(7) business_id                                                  0
 name                                                        1500
 address                                                     1500
 city                                                        1500
 state                                                       1500
 postal_code                                                 1500
 latitude                                                    1500
 longitude                                                   1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 business_id                                                 1500
 name                                                        1500
 address                                                     1500
 city                                                        1500
 state                                                       1500
 postal_code                                                 1500
 latitude                                                    1500
 longitude                                                   1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 businessid_business_desp_x                                  1500
 latitude                                                    1500
 longitude                                                   1500
 zipcode                                                     1500
 city                                                        1500
 state                                                       1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 businessid_business_desp_y                                  1500
 business_name_business_desp                                 1500
 business_address_business_desp                              1500
 business_city_business_desp                                 1500
 business_state_business_desp                                1500
 business_postal_code_business_desp                          1500
 business_latitude_business_desp                             1500
 business_longitude_business_desp                            1500
 business_stars_business_desp                                1500
 business_review_count_business_desp                         1500
 business_is_open_business_desp                              1500
 business_attributes_business_desp                           1500
 business_categories_business_desp                           1500
 business_hours_business_desp                                1500
 business_type_business_desp                                 1500
 businessid_business_desp_y.1                                1500
 business_name_business_desp.1                               1500
 business_address_business_desp.1                            1500
 business_city_business_desp.1                               1500
 business_state_business_desp.1                              1500
 business_postal_code_business_desp.1                        1500
 business_latitude_business_desp.1                           1500
 business_longitude_business_desp.1                          1500
 business_stars_business_desp.1                              1500
 business_review_count_business_desp.1                       1500
 business_is_open_business_desp.1                            1500
 business_attributes_business_desp.1                         1500
 business_categories_business_desp.1                         1500
 business_hours_business_desp.1                              1500
 business_type_business_desp.1                               1500
 zipcode_income_income_merge                                    0
 average_income_income_merge                                    0
 state_income_income_merge                                      0
 _merge_income_income_merge                                     0
 businessid_restaurantsincome_businessdesp_merge                0
 businessid_restaurantsincome_businessdesp_merge.1           1500
 _merge_restaurantsincome_businessdesp_merge                    0
 dtype: int64
```

```
restaurantsincome_businessdesp_merge.dropna(subset=['average_income_income_merge'].sample(10).head()
```

```
(8)   business_id  name  address  city  state  postal_code  latitude  \
0                NaN   NaN      NaN   NaN    NaN          NaN       NaN
1                NaN   NaN      NaN   NaN    NaN          NaN       NaN
2                NaN   NaN      NaN   NaN    NaN          NaN       NaN
3                NaN   NaN      NaN   NaN    NaN          NaN       NaN
4                NaN   NaN      NaN   NaN    NaN          NaN       NaN

   longitude  stars  review_count  is_open  attributes  categories  \
0        NaN    NaN           NaN      NaN         NaN         NaN
1        NaN    NaN           NaN      NaN         NaN         NaN
2        NaN    NaN           NaN      NaN         NaN         NaN
3        NaN    NaN           NaN      NaN         NaN         NaN
4        NaN    NaN           NaN      NaN         NaN         NaN

   hours  type  business_id  name  address  city  state  postal_code  \
0    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
1    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
2    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
3    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
4    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN

   latitude  longitude  stars  review_count  is_open  attributes  \
0       NaN        NaN    NaN           NaN      NaN         NaN
1       NaN        NaN    NaN           NaN      NaN         NaN
2       NaN        NaN    NaN           NaN      NaN         NaN
3       NaN        NaN    NaN           NaN      NaN         NaN
4       NaN        NaN    NaN           NaN      NaN         NaN

   categories  hours  type  businessid_business_desp_x  latitude  \
0         NaN    NaN   NaN                 0x8d0ef0ef0ef0  29.6516
1         NaN    NaN   NaN                 0x8d0ef0ef0ef1  29.6516
2         NaN    NaN   NaN                 0x8d0ef0ef0ef2  29.6516
3         NaN    NaN   NaN                 0x8d0ef0ef0ef3  29.6516
4         NaN    NaN   NaN                 0x8d0ef0ef0ef4  29.6516

   longitude  zipcode  city  state  stars  review_count  is_open  \
0  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
1  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
2  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
3  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
4  -82.3248    32601   NaN    NaN    NaN           NaN      NaN

   attributes  categories  hours  type  businessid_business_desp_y  \
0         NaN         NaN    NaN   NaN                         NaN
1         NaN         NaN    NaN   NaN                         NaN
2         NaN         NaN    NaN   NaN                         NaN
3         NaN         NaN    NaN   NaN                         NaN
4         NaN         NaN    NaN   NaN                         NaN

   business_name_business_desp  business_address_business_desp  \
0                          NaN                             NaN
1                          NaN                             NaN
2                          NaN                             NaN
3                          NaN                             NaN
4                          NaN                             NaN

   business_city_business_desp  business_state_business_desp  \
0                          NaN                           NaN
1                          NaN                           NaN
2                          NaN                           NaN
3                          NaN                           NaN
4                          NaN                           NaN

   business_postal_code_business_desp  business_latitude_business_desp  \
0                                 NaN                              NaN
1                                 NaN                              NaN
2                                 NaN                              NaN
3                                 NaN                              NaN
4                                 NaN                              NaN

   business_longitude_business_desp  business_stars_business_desp  \
0                               NaN                           NaN
1                               NaN                           NaN
2                               NaN                           NaN
3                               NaN                           NaN
4                               NaN                           NaN

   business_review_count_business_desp  business_is_open_business_desp  \
0                                  NaN                             NaN
1                                  NaN                             NaN
2                                  NaN                             NaN
3                                  NaN                             NaN
4                                  NaN                             NaN

   business_attributes_business_desp  business_categories_business_desp  \
0                                NaN                                NaN
1                                NaN                                NaN
2                                NaN                                NaN
3                                NaN                                NaN
4                                NaN                                NaN

   business_hours_business_desp  business_type_business_desp  \
0                           NaN                          NaN
1                           NaN                          NaN
2                           NaN                          NaN
3                           NaN                          NaN
4                           NaN                          NaN

   businessid_business_desp_y.1  business_name_business_desp.1  \
0                           NaN                            NaN
1                           NaN                            NaN
2                           NaN                            NaN
3                           NaN                            NaN
4                           NaN                            NaN

   business_address_business_desp.1  business_city_business_desp.1  \
0                               NaN                            NaN
1                               NaN                            NaN
2                               NaN                            NaN
3                               NaN                            NaN
4                               NaN                            NaN

   business_state_business_desp.1  business_postal_code_business_desp.1  \
0                             NaN                                   NaN
1                             NaN                                   NaN
2                             NaN                                   NaN
3                             NaN                                   NaN
4                             NaN                                   NaN

   business_latitude_business_desp.1  business_longitude_business_desp.1  \
0                               NaN                                NaN
1                               NaN                                NaN
2                               NaN                                NaN
3                               NaN                                NaN
4                               NaN                                NaN

   business_stars_business_desp.1  business_review_count_business_desp.1  \
0                             NaN                                    NaN
1                             NaN                                    NaN
2                             NaN                                    NaN
3                             NaN                                    NaN
4                             NaN                                    NaN

   business_is_open_business_desp.1  business_attributes_business_desp.1  \
0                               NaN                                  NaN
1                               NaN                                  NaN
2                               NaN                                  NaN
3                               NaN                                  NaN
4                               NaN                                  NaN

   business_categories_business_desp.1  business_hours_business_desp.1  \
0                                  NaN                             NaN
1                                  NaN                             NaN
2                                  NaN                             NaN
3                                  NaN                             NaN
4                                  NaN                             NaN

   business_type_business_desp.1  zipcode_income_income_merge  \
0                            NaN                        32601
1                            NaN                        32601
2                            NaN                        32601
3                            NaN                        32601
4                            NaN                        32601

   average_income_income_merge  state_income_income_merge  \
0                      24659                         FL
1                      24659                         FL
2                      24659                         FL
3                      24659                         FL
4                      24659                         FL

   _merge_income_income_merge  \
0                        both
1                        both
2                        both
3                        both
4                        both

   businessid_restaurantsincome_businessdesp_merge  \
0                                     0x8d0ef0ef0ef0
1                                     0x8d0ef0ef0ef1
2                                     0x8d0ef0ef0ef2
3                                     0x8d0ef0ef0ef3
4                                     0x8d0ef0ef0ef4

   businessid_restaurantsincome_businessdesp_merge.1  \
0                                                NaN
1                                                NaN
2                                                NaN
3                                                NaN
4                                                NaN

   _merge_restaurantsincome_businessdesp_merge
0                                        both
1                                        both
2                                        both
3                                        both
4                                        both
```

```
restaurantsincome_businessdesp_merge.columns.tolist()
```

```
(9) ['business_id',
 'name',
 'address',
 'city',
 'state',
 'postal_code',
 'latitude',
 'longitude',
 'stars',
 'review_count',
 'is_open',
 'attributes',
 'categories',
 'hours',
 'type',
 'business_id',
 'name',
 'address',
 'city',
 'state',
 'postal_code',
 'latitude',
 'longitude',
 'stars',
 'review_count',
 'is_open',
 'attributes',
 'categories',
 'hours',
 'type',
'businessid_business_desp_x',
'latitude',
'longitude',
'zipcode',
'city',
'state',
'stars',
'review_count',
'is_open',
'attributes',
'categories',
'hours',
'type',
'businessid_business_desp_y',
'business_name_business_desp',
'business_address_business_desp',
'business_city_business_desp',
'business_state_business_desp',
'business_postal_code_business_desp',
'business_latitude_business_desp',
'business_longitude_business_desp',
'business_stars_business_desp',
'business_review_count_business_desp',
'business_is_open_business_desp',
'business_attributes_business_desp',
'business_categories_business_desp',
'business_hours_business_desp',
'business_type_business_desp',
'businessid_business_desp_y.1',
'business_name_business_desp.1',
'business_address_business_desp.1',
'business_city_business_desp.1',
'business_state_business_desp.1',
'business_postal_code_business_desp.1',
'business_latitude_business_desp.1',
'business_longitude_business_desp.1',
'business_stars_business_desp.1',
'business_review_count_business_desp.1',
'business_is_open_business_desp.1',
'business_attributes_business_desp.1',
'business_categories_business_desp.1',
'business_hours_business_desp.1',
'business_type_business_desp.1',
'zipcode_income_income_merge',
'average_income_income_merge',
'state_income_income_merge',
'_merge_income_income_merge',
'businessid_restaurantsincome_businessdesp_merge',
'businessid_restaurantsincome_businessdesp_merge.1',
'_merge_restaurantsincome_businessdesp_merge']
```

```
restaurantsincome_businessdesp_merge.isna().sum()
```

```
(10) business_id                                                  0
 name                                                        1500
 address                                                     1500
 city                                                        1500
 state                                                       1500
 postal_code                                                 1500
 latitude                                                    1500
 longitude                                                   1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 business_id                                                 1500
 name                                                        1500
 address                                                     1500
 city                                                        1500
 state                                                       1500
 postal_code                                                 1500
 latitude                                                    1500
 longitude                                                   1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 businessid_business_desp_x                                  1500
 latitude                                                    1500
 longitude                                                   1500
 zipcode                                                     1500
 city                                                        1500
 state                                                       1500
 stars                                                       1500
 review_count                                                1500
 is_open                                                     1500
 attributes                                                  1500
 categories                                                  1500
 hours                                                       1500
 type                                                        1500
 businessid_business_desp_y                                  1500
 business_name_business_desp                                 1500
 business_address_business_desp                              1500
 business_city_business_desp                                 1500
 business_state_business_desp                                1500
 business_postal_code_business_desp                          1500
 business_latitude_business_desp                             1500
 business_longitude_business_desp                            1500
 business_stars_business_desp                                1500
 business_review_count_business_desp                         1500
 business_is_open_business_desp                              1500
 business_attributes_business_desp                           1500
 business_categories_business_desp                           1500
 business_hours_business_desp                                1500
 business_type_business_desp                                 1500
 businessid_business_desp_y.1                                1500
 business_name_business_desp.1                               1500
 business_address_business_desp.1                            1500
 business_city_business_desp.1                               1500
 business_state_business_desp.1                              1500
 business_postal_code_business_desp.1                        1500
 business_latitude_business_desp.1                           1500
 business_longitude_business_desp.1                          1500
 business_stars_business_desp.1                              1500
 business_review_count_business_desp.1                       1500
 business_is_open_business_desp.1                            1500
 business_attributes_business_desp.1                         1500
 business_categories_business_desp.1                         1500
 business_hours_business_desp.1                              1500
 business_type_business_desp.1                               1500
 zipcode_income_income_merge                                    0
 average_income_income_merge                                    0
 state_income_income_merge                                      0
 _merge_income_income_merge                                     0
 businessid_restaurantsincome_businessdesp_merge                0
 businessid_restaurantsincome_businessdesp_merge.1           1500
 _merge_restaurantsincome_businessdesp_merge                    0
 dtype: int64
```

```
restaurantsincome_businessdesp_merge.dropna(subset=['average_income_income_merge'].sample(10).head()
```

```
(11)   business_id  name  address  city  state  postal_code  latitude  \
0                NaN   NaN      NaN   NaN    NaN          NaN       NaN
1                NaN   NaN      NaN   NaN    NaN          NaN       NaN
2                NaN   NaN      NaN   NaN    NaN          NaN       NaN
3                NaN   NaN      NaN   NaN    NaN          NaN       NaN
4                NaN   NaN      NaN   NaN    NaN          NaN       NaN

   longitude  stars  review_count  is_open  attributes  categories  \
0        NaN    NaN           NaN      NaN         NaN         NaN
1        NaN    NaN           NaN      NaN         NaN         NaN
2        NaN    NaN           NaN      NaN         NaN         NaN
3        NaN    NaN           NaN      NaN         NaN         NaN
4        NaN    NaN           NaN      NaN         NaN         NaN

   hours  type  business_id  name  address  city  state  postal_code  \
0    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
1    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
2    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
3    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN
4    NaN   NaN          NaN   NaN      NaN   NaN    NaN          NaN

   latitude  longitude  stars  review_count  is_open  attributes  \
0       NaN        NaN    NaN           NaN      NaN         NaN
1       NaN        NaN    NaN           NaN      NaN         NaN
2       NaN        NaN    NaN           NaN      NaN         NaN
3       NaN        NaN    NaN           NaN      NaN         NaN
4       NaN        NaN    NaN           NaN      NaN         NaN

   categories  hours  type  businessid_business_desp_x  latitude  \
0         NaN    NaN   NaN                 0x8d0ef0ef0ef0  29.6516
1         NaN    NaN   NaN                 0x8d0ef0ef0ef1  29.6516
2         NaN    NaN   NaN                 0x8d0ef0ef0ef2  29.6516
3         NaN    NaN   NaN                 0x8d0ef0ef0ef3  29.6516
4         NaN    NaN   NaN                 0x8d0ef0ef0ef4  29.6516

   longitude  zipcode  city  state  stars  review_count  is_open  \
0  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
1  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
2  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
3  -82.3248    32601   NaN    NaN    NaN           NaN      NaN
4  -82.3248    32601   NaN    NaN    NaN           NaN      NaN

   attributes  categories  hours  type  businessid_business_desp_y  \
0         NaN         NaN    NaN   NaN                         NaN
1         NaN         NaN    NaN   NaN                         NaN
2         NaN         NaN    NaN   NaN                         NaN
3         NaN         NaN    NaN   NaN                         NaN
4         NaN         NaN    NaN   NaN                         NaN

   business_name_business_desp  business_address_business_desp  \
0                          NaN                             NaN
1                          NaN                             NaN
2                          NaN                             NaN
3                          NaN                             NaN
4                          NaN                             NaN

   business_city_business_desp  business_state_business_desp  \
0                          NaN                           NaN
1                          NaN                           NaN
2                          NaN                           NaN
3                          NaN                           NaN
4                          NaN                           NaN

   business_postal_code_business_desp  business_latitude_business_desp  \
0                                 NaN                              NaN
1                                 NaN                              NaN
2                                 NaN                              NaN
3                                 NaN                              NaN
4                                 NaN                              NaN

   business_longitude_business_desp  business_stars_business_desp  \
0                               NaN                           NaN
1                               NaN                           NaN
2                               NaN                           NaN
3                               NaN                           NaN
4                               NaN                           NaN

   business_review_count_business_desp  business_is_open_business_desp  \
0                                  NaN                             NaN
1                                  NaN                             NaN
2                                  NaN                             NaN
3                                  NaN                             NaN
4                                  NaN                             NaN

   business_attributes_business_desp  business_categories_business_desp  \
0                                NaN                                NaN
1                                NaN                                NaN
2                                NaN                                NaN
3                                NaN                                NaN
4                                NaN                                NaN

   business_hours_business_desp  business_type_business_desp  \
0                           NaN                          NaN
1                           NaN                          NaN
2                           NaN                          NaN
3                           NaN                          NaN
4                           NaN                          NaN

   businessid_business_desp_y.1  business_name_business_desp.1  \
0                           NaN                            NaN
1                           NaN                            NaN
2                           NaN                            NaN
3                           NaN                            NaN
4                           NaN                            NaN

   business_address_business_desp.1  business_city_business_desp.1  \
0                               NaN                            NaN
1                               NaN                            NaN
2
```

The graph shows the relationship between crime rates in a state and the percentage of social user ratings. It can be seen that the higher the crime rate, the more negative reviews people leave. Conversely, the lower the crime rate, the more positive reviews. Conversely, can see that the percentage of high-rated restaurants increases in the safer regions.

From these maps and graphs, it is clear that there is a relationship between the percentage of restaurants with more stars and crime rate. The more dangerous the area, the the percentage of restaurants with lower star ratings is always with lower stars.

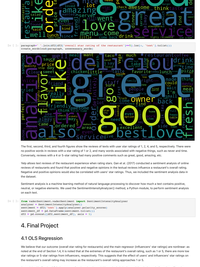
We could not obtain parking spaces data using Google API, but we did get data from Google that included information on parking availability in the area of 1 km. Parking Bars, was updated 2017. Therefore, we will change that data with the data from the survey.

```
parking = pd.read_csv('./Parking_Bars.csv')
parking['Bike_parking'] = parking['Bike_parking'].replace(False, 'No')
parking['Bike_parking'] = parking['Bike_parking'].replace(True, 'Yes')
parking = parking.groupby(['City','City', 'Park_Price', 'Bike_parking', 'Lot_Type', 'Lot_Name', 'Lot_Capacity', 'Lot_Occupied', 'Lot_Open', 'Lot_Closed', 'Lot_Updated', 'Lot_Latitude', 'Lot_Longitude'])['Lot_Name'].count().reset_index()
restaurant_parking = pd.read_csv('./restaurant_parking.csv')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(False, 'No')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(True, 'Yes')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(nan, 'No')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(u'free', 'Yes')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(u'paid', 'No')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(u'street', 'Yes')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(u'validated', 'Yes')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(u'valet', 'Yes')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(u'lot', 'Yes')
restaurant_parking['restaurant_parking'] = restaurant_parking['restaurant_parking'].replace(u'garage', 'Yes')
```

In addition, we analyzed the review data, which was recorded in Pickles in Russian language according, meaningless units and to determine which words appeared more regularly in the text in each star's reviews.

```
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer(ngram_range=(1,1))
vectorizer.fit(review_df['text'])
print('Vocabulary size: {}'.format(len(vectorizer.vocabulary_)))
vocabulary = vectorizer.vocabulary_
def count_vectors(dataset, text, ngrams=1):
    vectorizer = CountVectorizer(ngram_range=(1, ngrams))
    bow = vectorizer.fit_transform(dataset[text])
    for ngram, index in vectorizer.vocabulary_.items():
        print(ngram, bow[:, index].sum())
    return bow
df1_ngrams = count_vectors(review_df[review_df['stars'] == 1], 'text', 1)
df2_ngrams = count_vectors(review_df[review_df['stars'] == 2], 'text', 1)
df3_ngrams = count_vectors(review_df[review_df['stars'] == 3], 'text', 1)
df4_ngrams = count_vectors(review_df[review_df['stars'] == 4], 'text', 1)
df5_ngrams = count_vectors(review_df[review_df['stars'] == 5], 'text', 1)
```

```
Vocabulary size: 32
```


Vocabulary size: 32
manager
minute
chicken
goodness
people
location
pizza
even
like
order
time
ordered
employee
went
asked
told
service
pizza
mcdonald
drive-thru
drive-thru
manager
minute
staff
time
people
like
order
good
ordered
food
place
place
like
like
order
like
like
good
like
good
like
good