

## Motivation

Yelp is a website and mobile app that connects people with local businesses. The Yelp community is made up of engaged locals who connect on-line and off-line to share their opinions about local businesses. Millions of users connect on this application to, rate and give reviews for businesses.

In this project, I have analyzed various factors to be considered for establishing a new restaurant business in San Francisco area.

## Data Collection

Using Web scrapping I have collected 990 restaurants information for San Francisco. I have collected Name, Cuisine served, delivery availability for a restaurant, ratings & number of reviews for analysis.

## Analysis

### Pre-Work:

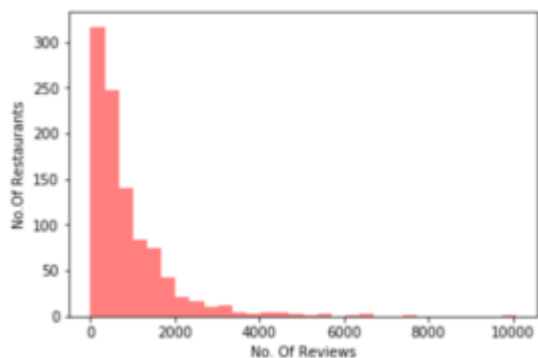
Before analysis of various factors recommending for a new restaurant business, let's consider to review the following:

#### 1. Missing Data

```
#Finding out the percentage of missing data  
yelp.isna().mean().round(4) * 100
```

```
Category      0.00  
Overallrating 0.00  
Name          0.00  
Delivery      2.42  
NumberReviews 0.00  
PriceRange    0.00  
dtype: float64
```

#### 2. Invalid Data

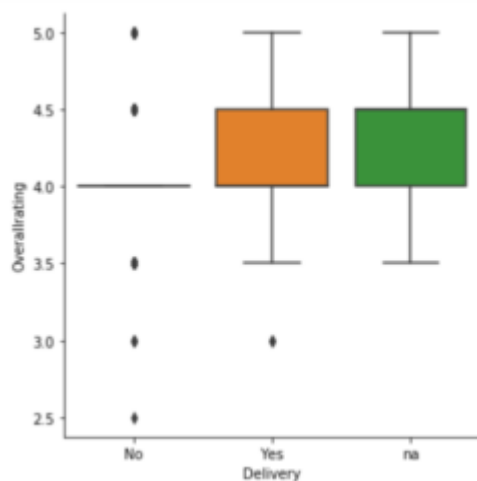


From the above observations:

1. we have only 2.4% of missing data for Delivery column which is replaced by NA and can be ignored for further analysis.
2. Considering restaurants having more number of Reviews[1000] - Based on the above distribution we can't ignore any rows for column "Number of Reviews" as major percentage of restaurants are having less than 100 reviews.

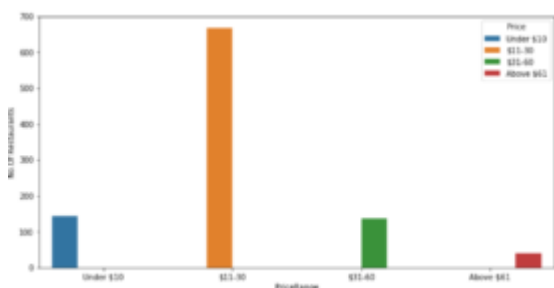
Now Let's consider the various factors influencing for a new restaurant business:

1. **Delivery** - Analyzing if delivery have an impact on restaurant business. The following graph shows the user rates whether the restaurant has delivery service or not based on Yelp reviews.



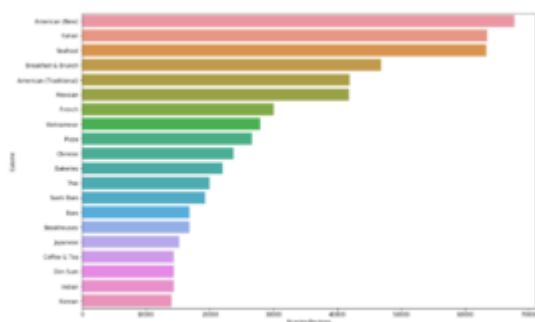
From above analysis - Restaurants with Delivery have average of more ratings compared to restaurants without delivery.

2. **Price Range** - To Analyze if Price Range have an impact on restaurant business, I have to replace the dollar value for the column PriceRange of the data-set with relevant categories of Price ranges.



I have identified that restaurants in price range of Under \$10 and Above \$61 have higher ratings but the count of restaurants with price range \$11-\$30 are way higher than any other category with 4.0 rating.

**3. Cuisine** - Identifying the best Cuisines, Cuisine plays a major role in the restaurant business. I have aggregated the data by cuisine and shown by the most number of reviews which provides the information of famous cuisines around SFO.



## Conclusion:

The above analysis shows that a new restaurant in SFO will have a better scope in business by considering Delivery service as one of the key factors. The second factor to be considered is Price Range i.e \$11-\$30, as restaurant business is very competitive price range should be matching to the city along with good food served. The last factor I have considered is Cuisine as different regions have different taste of cuisine.

Further analysis can be done by considering data from different major cities and websites like yelp. More information about my analysis is found in [GitHub](#).