***Zomato Bangalore Restaurants***

***Business problem:***

· Ratings and reviews play a very important role in attracting and retaining new customers.

· Our target would be **improving ratings** based on the insights and based on these factors **predict ratings** for a prospect restaurant.

· Understand what people like the most in a highly rated restaurant, in a particular locality, which are related to ratings for a prospect restaurant.

· Have an **insight of approx\_cost** which is based on many factors like neighborhood, restaurant type which can be related to ratings.

· Given a locality, a prospect restaurant can have an insight of the factors to get the best rating.

· Marketing strategies like personalized notifications, discounts etc. can be set up.

***Clients:***

· Potential clients would be **existing Zomato restaurant owners** and **prospects restaurants.**

· Having insights on factors ratings could help the decision makers take actions which would eventually increase the ratings and clients.

***Data extraction:***

· The data is extracted from the Zomato website using the Python package ‘**Beautiful Soup’**.

· This data is for all restaurants of Bangalore city which is around 12k-13k of records, pulled as of January 2020.

· Following are the fields

1. restaurant\_link: Link for the restaurant

2. restaurant\_ID: Unique restaurant id

3. restaurant\_name: Name of the restaurant

4. locality: neighbourhood of the restaurant

5. restaurant\_category: Category of restaurant based on what food they serve, like dining or quick bites, etc.

6. zomato\_gold: Whether the restaurant provides zomato gold benefits

7. discounts: Discounts offered by the restaurant

8. photos\_taken: Number of photos taken at the restaurant

9. rating: Zomato rating

10. votes: Votes for the ratings or reviews

11. cuisines: Type of cuisines served

12. approx.\_cost\_for\_2: Approx cost for 2 people

13. opening timings: Opening and closing timings of the restaurant

14. address: Detailed address of the restaurant

15. latitude: Latitude of restaurant

16. longitude :Longitude of the restaurant

17. more\_info: main features or services provided by the restaurant like delivery, outside seating, etc

18. featured\_in: Featured in which categories of Zomato collections

19. most\_liked\_food: Most liked or famous for in food items and rating

20. most\_liked\_service: Most liked service of the restaurant and rating

21. most\_liked\_look&field: Most liked, look and feel of the restaurant and rating

22. reviews: Reviews available on the first page of the restaurant along with time of review posted and sentiments.

Reference:<https://github.com/Anandpatil412/DSC/tree/master/CapstoneProject1/DataExtraction(WebZomato)>

***Data cleaning:***

· Most of the data is cleaned while scraping.

· Some columns are manipulated to tuples.

· Opening and closing timings are transformed to datetime formats.

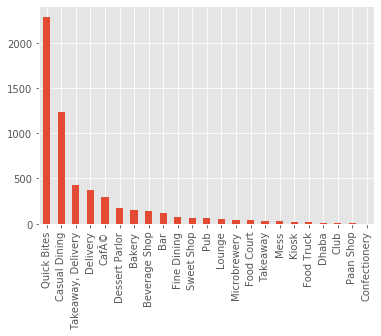
· Missing values are transformed to np.NAN

· Duplicates rows, if any, are removed based on the restaurant\_id.

· No outliers.

***Findings from EDA:***

**Different restaurant categories**

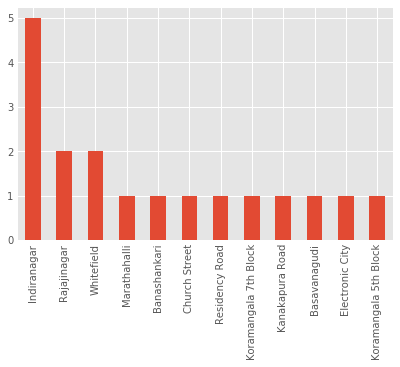
****

· **Quick bites** and **Casual dining** are the most common restaurant categories.

**Unique cuisines**

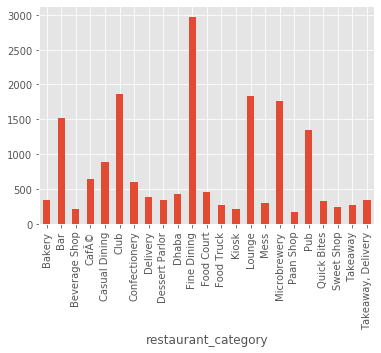
· **Raw meat** and **Nepalese** are the most unique cuisines found.

**Locality having maximum restaurants**



· **Indiranagar** has the maximum number of restaurants.

**Restaurant category having highest cost for 2**

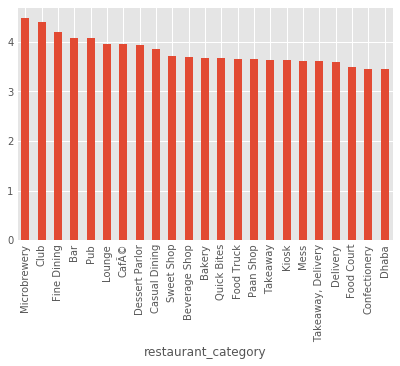


· **Fine Dining** category has the highest cost for 2.

**Locality having highest average rating**

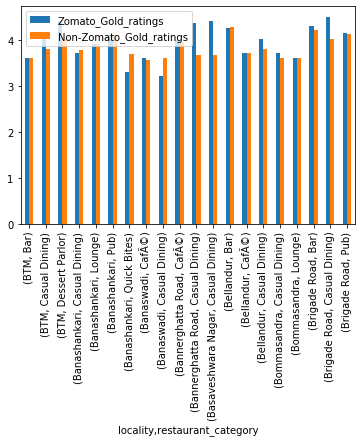
· **Sankey road** and **Lavelle road** have the highest average restaurant ratings.

**Restaurant category having highest and lowest average rating**



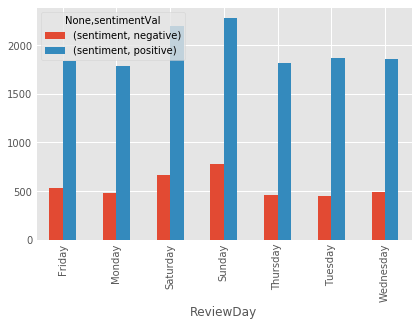
· **Microbrewery** has the highest while Dhaba has the lowest average restaurant rating.

**Comparison of avg ratings of Zomato Gold and Non Zomato Gold restaurant locality wise**



· Average ratings of Zomato Gold restaurants seem higher than the Non Zomato Gold restaurants.

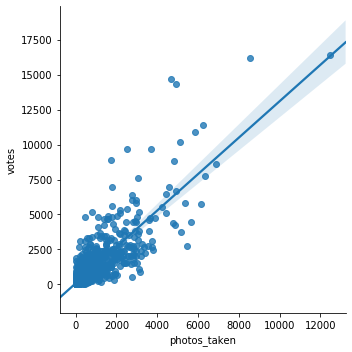
**Comparison of Positive and Negative Sentiments for days of a week**



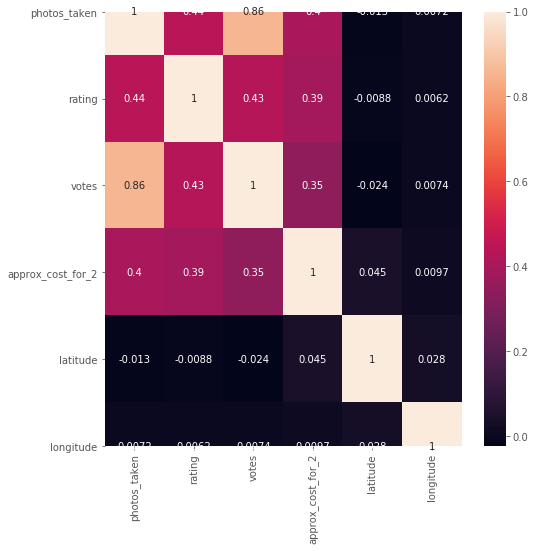
· There seem more positive sentiments than negative,weekends being the highest.

Reference:<https://github.com/Anandpatil412/DSC/tree/master/CapstoneProject1/DataWrangling>

***Correlations and conclusions:***

****

· There is a high positive correlation between photos\_taken and votes, which indicates there is more voting from people who take photos.So factors like ambience, food presentation etc play a very important role in contributing votes.



· Null hypothesis tests on correlation between approx\_cost\_for\_2 and ratings come to a conclusion that there is definitely a positive correlation between the two,which indicates people highly rated restaurants are expensive.

· Null hypothesis tests on comparing zomato gold ratings over non zomato ratings come to a conclusion that restaurants having Zomato gold have higher ratings than the one which don't have.