

By Navyadhara Gogineni (UIN: 660392086)

- | Restaurants | 0.01 | 0.019 | 0.09 | 0.058 | 0.011 | 0.021 | 0.019 | 0.004 | 0.0018 | 0.001 | 0.075 | 0.11 | 0.19 | 0.0077 | 0.2 | 0.018 | 0.038 | 0.066 | 0.055 | 0.005 | 0.011 | 0.14 | | |
|----------------------------|-----------|-------|-------|-------|--------|--------|-------|-------|--------|-------|-------|-------|--------|--------|-------|--------|--------|-------|--------|--------|--------|-------|-------|------|
| Outstanding | 0.037 | 0.1 | 0.11 | 0.2 | 0.18 | 0.016 | 0.2 | 0.18 | 0.11 | 0.01 | 0.018 | 0.21 | 0.27 | 0.21 | 0.19 | 0.018 | 0.066 | 0.055 | 0.007 | 0.014 | 0.012 | 0.13 | | |
| Business/clients/Corporate | 0.09 | 0.11 | 0.1 | 0.21 | 0.0022 | 0.19 | 0.16 | 0.21 | 0.086 | 0.018 | 0.077 | 0.1 | 0.24 | 0.11 | 0.051 | 0.027 | 0.039 | 0.042 | 0.0020 | 0.0066 | 0.0026 | 0.016 | | |
| Black/White | 0.05 | 0.2 | 0.23 | 0.11 | 0.15 | 0.0077 | 0.22 | 0.23 | 0.077 | 0.031 | 0.16 | 0.12 | 0.11 | 0.14 | 0.021 | 0.021 | 0.039 | 0.065 | 0.0048 | 0.071 | 0.001 | 0.05 | | |
| Referrals/Referrals | 0.055 | 0.1 | 0.1 | 0.2 | 0.18 | 0.016 | 0.2 | 0.18 | 0.11 | 0.01 | 0.018 | 0.21 | 0.27 | 0.21 | 0.19 | 0.018 | 0.066 | 0.055 | 0.007 | 0.014 | 0.012 | 0.13 | | |
| Referrals/Referrals | 0.021 | 0.016 | 0.022 | 0.017 | 0.02 | 0.0055 | 0.022 | 0.018 | 0.18 | 0.003 | 0.012 | 0.013 | 0.013 | 0.0051 | 0.012 | 0.002 | 0.0051 | 0.012 | 0.001 | 0.0051 | 0.006 | 0.029 | | |
| Caters | 0.022 | 0.016 | 0.022 | 0.017 | 0.02 | 0.0055 | 0.022 | 0.018 | 0.18 | 0.003 | 0.012 | 0.013 | 0.013 | 0.0051 | 0.012 | 0.002 | 0.0051 | 0.012 | 0.001 | 0.0051 | 0.006 | 0.029 | | |
| Cherish/Cherish | 0.092 | 0.18 | 0.16 | 0.23 | 0.085 | 0.023 | 0.16 | 0.1 | 0.14 | 0.15 | 0.09 | 0.31 | 0.17 | 0.037 | 0.032 | 0.025 | 0.056 | 0.043 | 0.18 | 0.19 | 0.014 | 0.012 | 0.13 | |
| Good/Good | 0.019 | 0.1 | 0.1 | 0.16 | 0.0055 | 0.023 | 0.018 | 0.18 | 0.003 | 0.012 | 0.013 | 0.013 | 0.0051 | 0.012 | 0.002 | 0.0051 | 0.012 | 0.001 | 0.0051 | 0.006 | 0.029 | 0.016 | 0.05 | |
| Restaurants/Restaurants | 0.042 | 0.06 | 0.078 | 0.077 | 0.15 | 0.018 | 0.2 | 0.1 | 0.12 | 0.048 | 0.037 | 0.23 | 0.23 | 0.11 | 0.077 | 0.039 | 0.051 | 0.115 | 0.011 | 0.006 | 0.022 | 0.064 | 0.05 | |
| Good/Good | -0.000105 | 0.053 | 0.018 | 0.031 | 0.0013 | 0.014 | 0.035 | 0.009 | 0.019 | 0.019 | 0.11 | 0.019 | 0.11 | 0.015 | 0.038 | 0.14 | -0.018 | 0.035 | 0.1 | 0.007 | 0.015 | 0.001 | 0.076 | 0.19 |
| Days/Days | 0.019 | 0.18 | 0.077 | 0.16 | 0.034 | 0.054 | 0.092 | 0.31 | -0.016 | 0.048 | 0.019 | 0.1 | 0.12 | 0.048 | 0.037 | 0.23 | 0.056 | 0.043 | 0.099 | 0.014 | 0.009 | 0.039 | 0.016 | 0.05 |
| Restaurants/Restaurants | 0.075 | 0.21 | 0.16 | 0.16 | 0.0055 | 0.023 | 0.018 | 0.18 | 0.003 | 0.012 | 0.013 | 0.013 | 0.0051 | 0.012 | 0.002 | 0.0051 | 0.012 | 0.001 | 0.0051 | 0.006 | 0.029 | 0.016 | 0.05 | |
| Restaurants/Restaurants | 0.011 | 0.27 | 0.22 | 0.14 | 0.17 | 0.019 | 0.024 | 0.17 | 0.019 | 0.019 | 0.11 | 0.019 | 0.11 | 0.015 | 0.038 | 0.14 | -0.018 | 0.035 | 0.1 | 0.007 | 0.015 | 0.001 | 0.076 | 0.19 |
| Hot/Hot | 0.018 | 0.1 | 0.15 | 0.1 | 0.3 | 0.033 | 0.2 | 0.032 | 0.42 | 0.023 | 0.033 | 0.27 | 0.5 | 0.1 | 0.26 | 0.048 | 0.027 | 0.076 | 0.023 | 0.052 | 0.02 | 0.01 | 0.013 | 0.1 |
| Hot/Hot | 0.0037 | 0.28 | 0.11 | 0.14 | 0.087 | 0.045 | 0.12 | 0.023 | 0.32 | 0.14 | 0.16 | 0.17 | 0.26 | 0.26 | 0.11 | 0.091 | 0.022 | | | | | | | |

3. Feature Selection:

- a. After working through the above steps, we select the following features for regression analysis:
 - i. average_stars, average_good_review, average_funny, average_cool, score from review data
 - ii. average_review_count, average_fans, average_good_user, average_elite_count, average_friend_count from user-average data.

4. Train-Test Split: Imade a 80-20 train test split

5. Training/Modelling:

- a. Three prediction algorithms have been implemented as stated above

d. Testing Stage:

- i. The performance of the prediction models has been tested using R2 score and MSE.
- ii. The following are the metrics obtained for the 3 models:

Model	R2 Score	MSE
Linear Regression	0.967	0.024
Decision Tree Regressor	0.985	0.011
SVM Regressor	0.24	0.657

- e. **Inference:** Predicted rating for a local restaurant using google reviews.

```
Rating for Jarabe Mexican Street Food is: 4.0
Rating for Stax Cafe. is: 4.24
Rating for Siri Indian Restaurant is: 3.8
```

ii)

I. **Goal:** Cluster all restaurants of a city

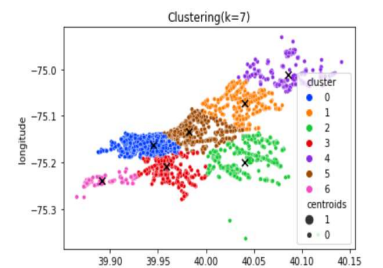
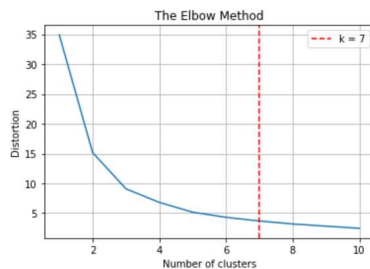
II. **Algorithms implemented:** K-Means clustering

III. **Procedure:**

○ Data Preparation:

- i. From the business data, first we select a city. “Philadelphia” has been used as city here as it had higher number of samples.
- ii. We fetch the latitude and longitude data from the separated data

○ Optimal K and Training:



○

