RIYADH COFFEE ANALYSIS

Hessah Alotaibi, Leen Alqahtani, Lamya Alnahdi, Norah Alfaheed

Supervised by: Dr. Mashael Aldayel

introduction

Cafés are a vital part of modern urban life, serving as both coffee spots and social hubs. In Riyadh, rapid growth in the café industry has led to intense competition. Despite offering similar products, some cafés achieve higher ratings and greater customer loyalty. This project aims to identify the key factors behind the success of top-rated cafés by analyzing customer reviews, pricing, atmosphere, and service quality. The findings will help café owners enhance customer satisfaction and stay competitive in a dynamic market where small differences can determine success.

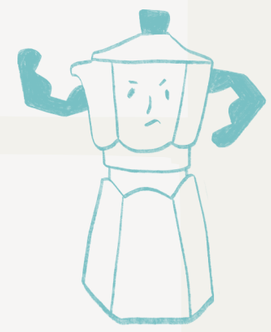
Objectives

- How does the number of reviews (rating count) correlate with the rating of a coffee shop?
- What is the overall sentiment of customer reviews for coffee shops in Riyadh?
- Which factors—service or atmosphere—have the strongest influence on customer satisfaction?
- Is there a relationship between the average price range of a coffee shop and its rating?
- How does time affect crowding and popularity in cafes?

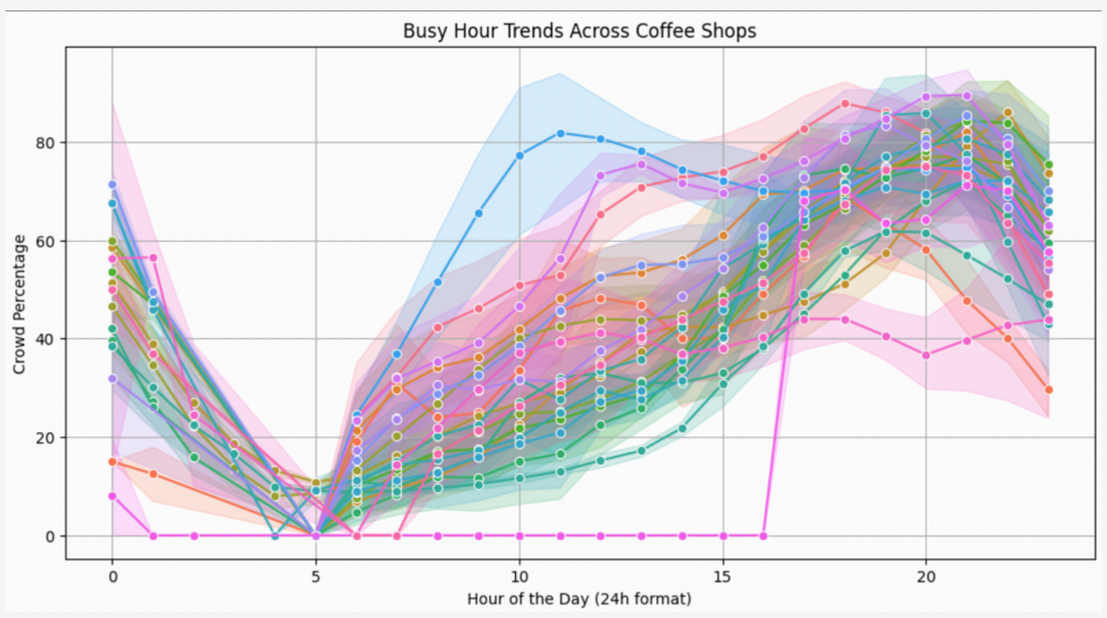
Data Collection

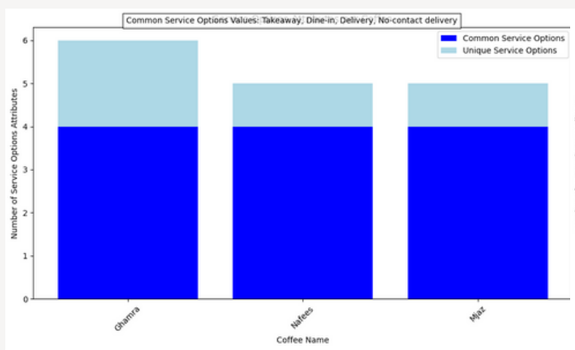
The dataset was collected through web scraping from Google Maps using the Instant Data Scraper extension, due to restrictions on traditional scraping tools like BeautifulSoup.

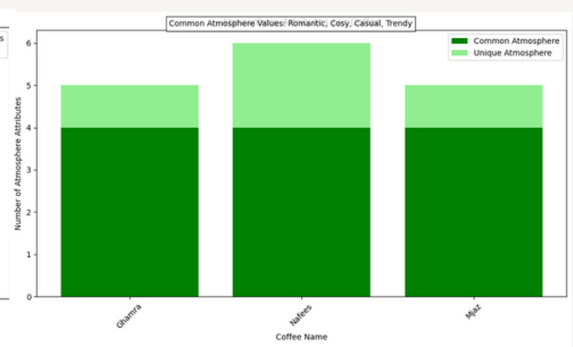
- Reviews, ratings, addresses, and average prices for cafés in Riyadh were manually extracted between February 1–4, 2025. Additionally,
- A custom script was later used to collect busy hours data. The final dataset includes features such as café names, customer comments, ratings, number of reviews, average price ranges, addresses, and busy hours.

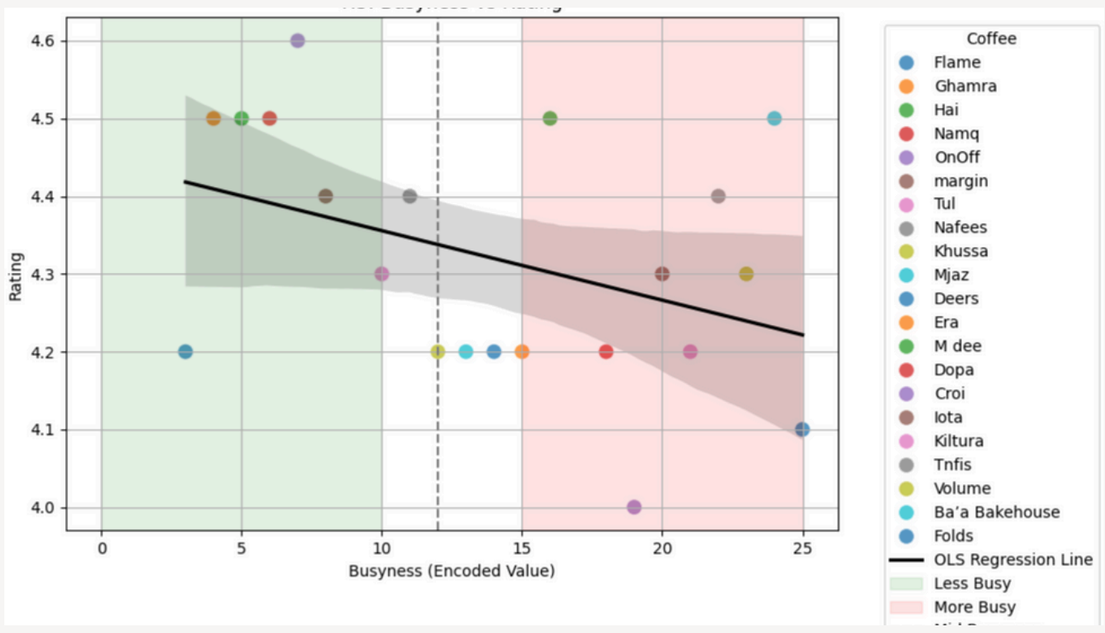


DATA Analysis









MODELS AND FINDINGS

Findings:

- All models struggled to predict customer ratings accurately. Linear Regression performed slightly better than the baseline, but overall, the models failed to capture meaningful relationships between the features and the ratings.

Models Developed:

- Dummy Regressor (Baseline)
- Linear Regression
- Random Forest Regressor

Model Evaluation:

- Linear Regression: Best model among those tested, but still unable to explain rating variations well. Showed slight improvement over the baseline, but results indicate weak predictive ability.
- Random Forest Regressor: Performed worse than Linear Regression, with higher error rates and less stability.
- Dummy Regressor: As expected, showed the poorest performance, simply predicting the average rating without learning any patterns.

Conclusion:

- Current data features (e.g., atmosphere, service options, customer comments) are insufficient for building reliable predictive models. Future work should include collecting richer data to improve model accuracy.

CONCLUSION

Among all tested hypotheses, negative comments were the only statistically significant factor that negatively affected customer ratings.

Other factors — including price, busy hours, atmosphere, service options, and rating count — showed weak or non-significant impacts on ratings.

The results of our hypotheses were not strongly supported because there was bias in the data collection process, which limited the accuracy and generalizability of our models.

Future improvements should focus on reducing data bias and exploring additional features to improve prediction quality.