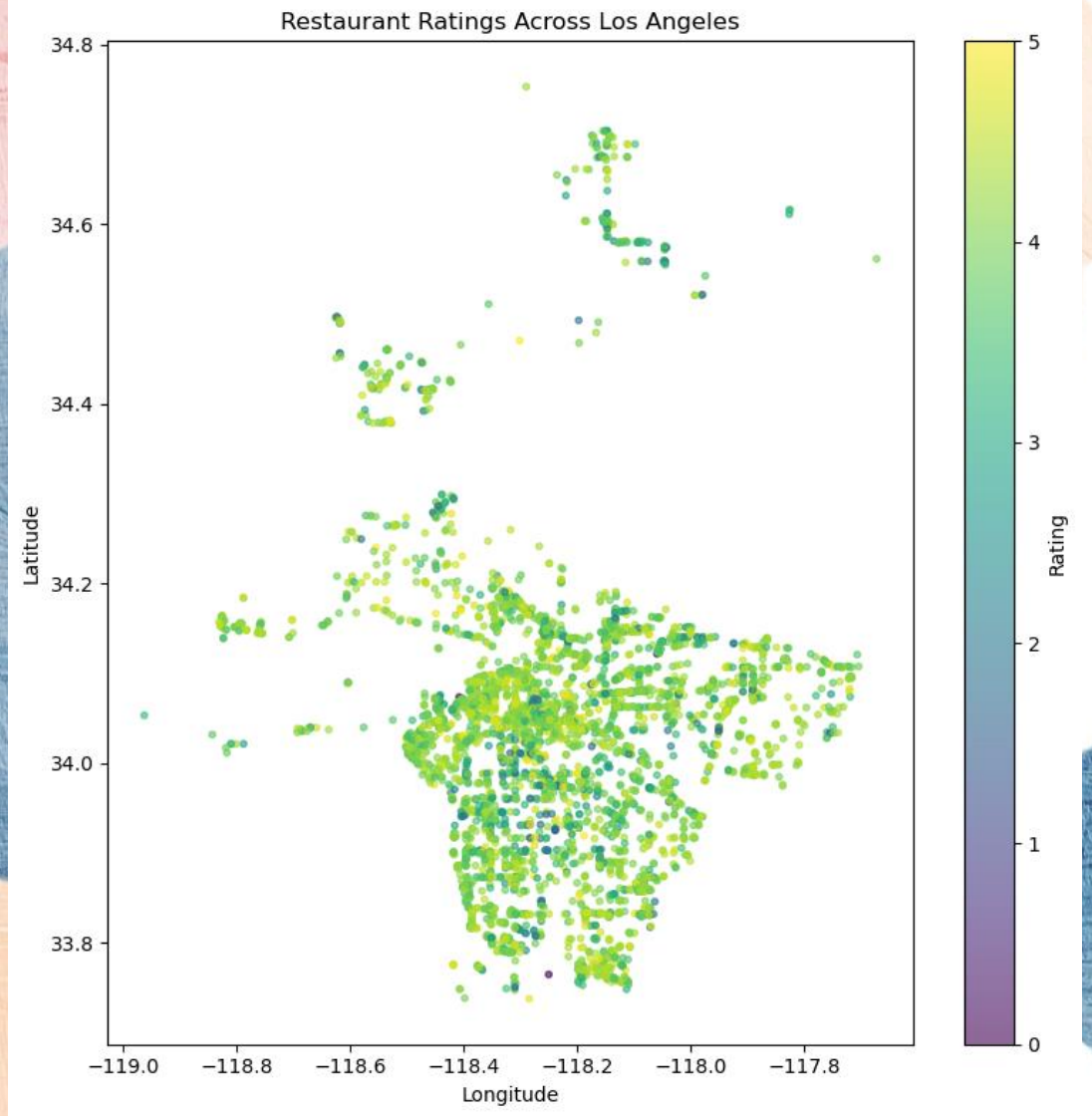
ANALYZING THE FACTORS INFLUENCING RESTAURANT RATINGS IN LOS ANGELES



Lauren Yu

INTRODUCTION

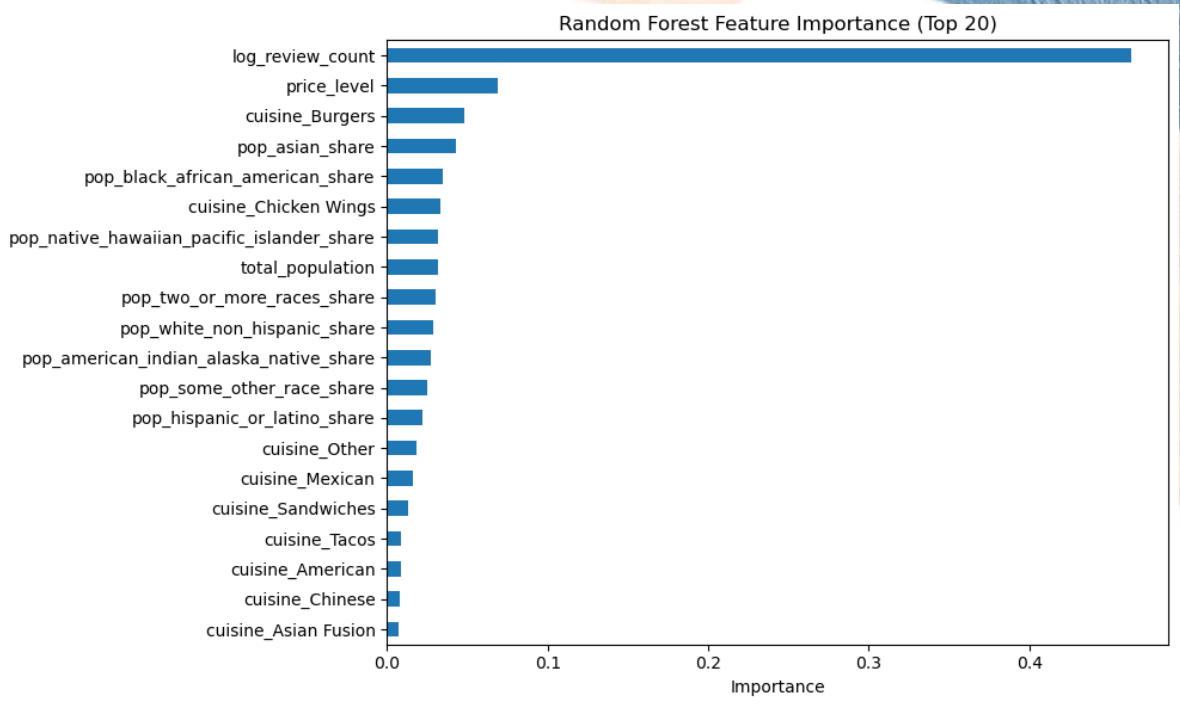
- Goal: **Identify what factors influence restaurant ratings across Los Angeles**.
- Methods:
  - Cleaned + merged using **coordinates and ZIP codes**
  - **Exploratory analysis** of ratings by cuisine, neighborhood, and demographics
  - **Correlation + feature importance** to find key predictors
- **Models:** Multiple Linear Regression & Random Forest using price level, review volume, cuisine type, and demographic factors
- Purpose: Understand how **community characteristics** and **restaurant attributes** drive rating patterns across LA.

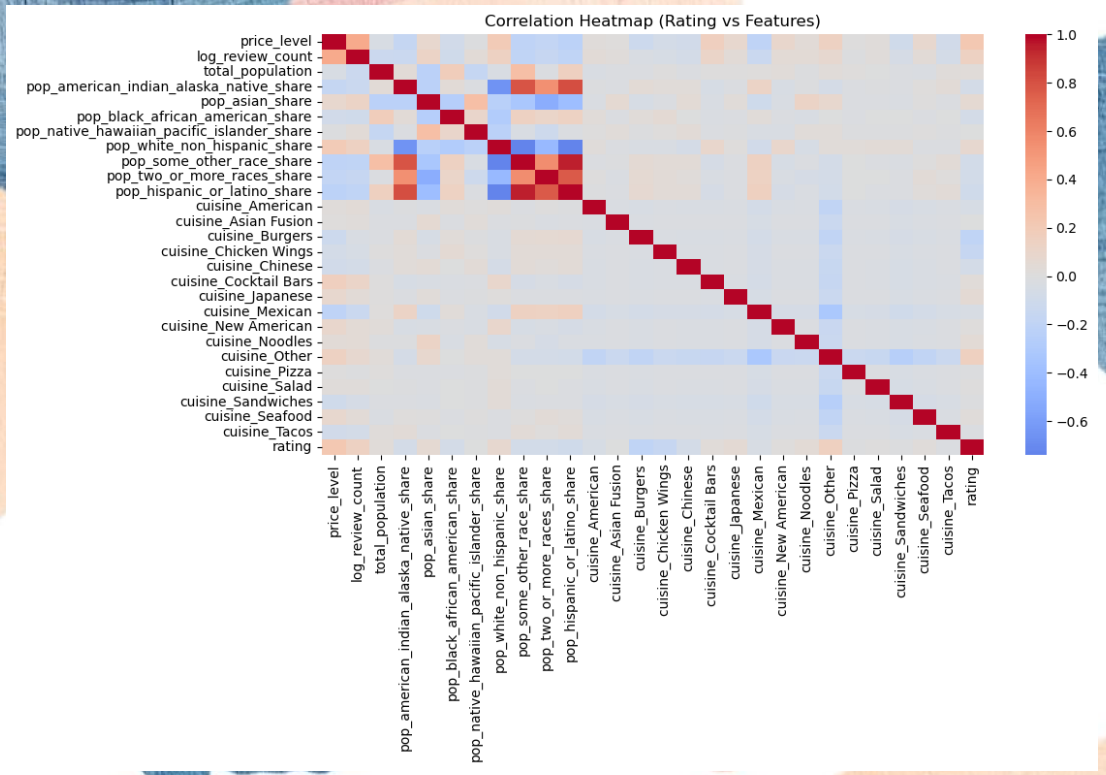
DATA SOURCES

Data	Name	Type	Format	data size
1	Yelp Fusion API – Los Angeles Restaurants	API	JSON / CSV	13,000+
2	LA cities and zipcodes	Web	HTML → CSV	529
3	Racial/Ethnic Composition Cities & Unincorporated Communities Los Angeles County	Web	HTML → CSV	144 cities and communities

KEY DRIVERS OF RESTAURANT RATINGS

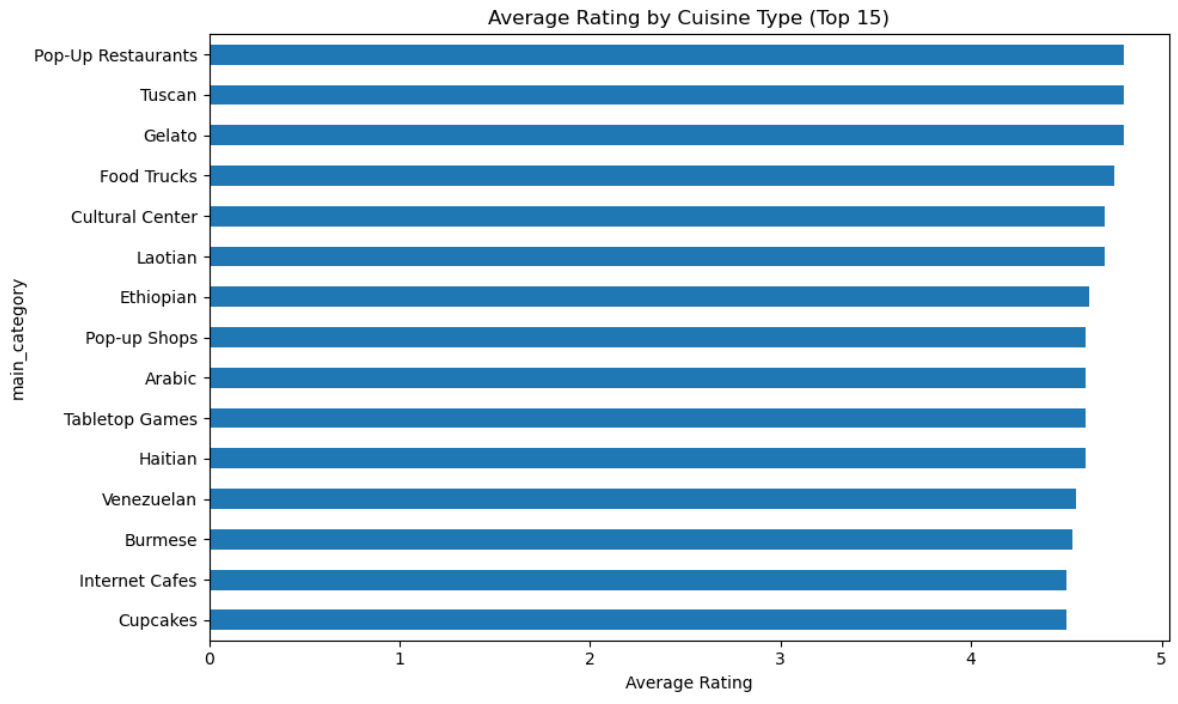
- Restaurant-level factors are the **strongest predictors** of Yelp ratings.
- Higher **price level** → higher average ratings
- Higher **review count** → higher ratings (popularity/visibility effect)





INFLUENCE OF CUISINE TYPE

- Top cuisines like *Pop-Up Restaurants, Tuscan, Gelato, Laotian, Ethiopian* consistently have ratings between 4.5–5
- Several cuisine dummies are among the highest positive or negative coefficients and cuisine-driven effects remain strong even after controlling for price & visibility
- Customer preferences vary significantly by cuisine—certain cuisines systematically receive higher ratings across LA

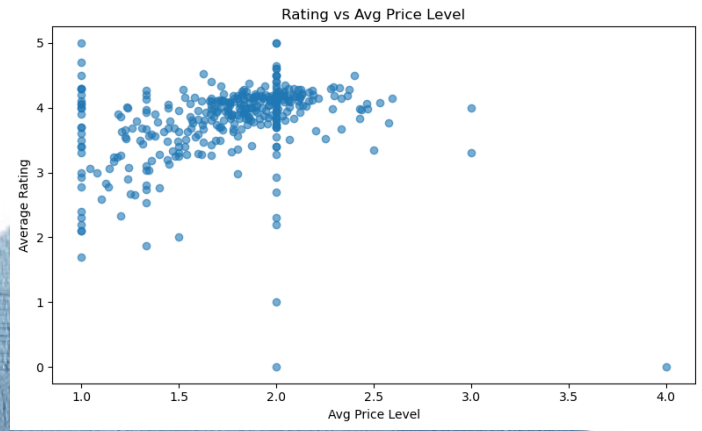


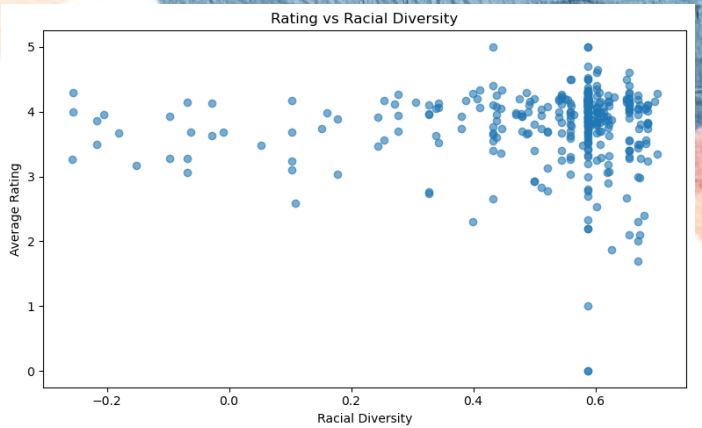

```
Top Linear Coefficients (positive influence on rating):
pop_american_indian_alaska_native_share       0.914188
pop_native_hawaiian_pacific_islander_share    0.243865
cuisine_Japanese                              0.209829
cuisine_Noodles                               0.190886
cuisine_Other                                 0.161798
price_level                                   0.160643
cuisine_Sandwiches                            0.104860
cuisine_Seafood                               0.099155
cuisine_Mexican                               0.097942
cuisine_Tacos                                 0.063741
cuisine_Salad                                 0.062297
cuisine_Cocktail Bars                         0.062135
cuisine_Asian Fusion                          0.054077
cuisine_New American                          0.049132
pop_some_other_race_share                     0.042172
```

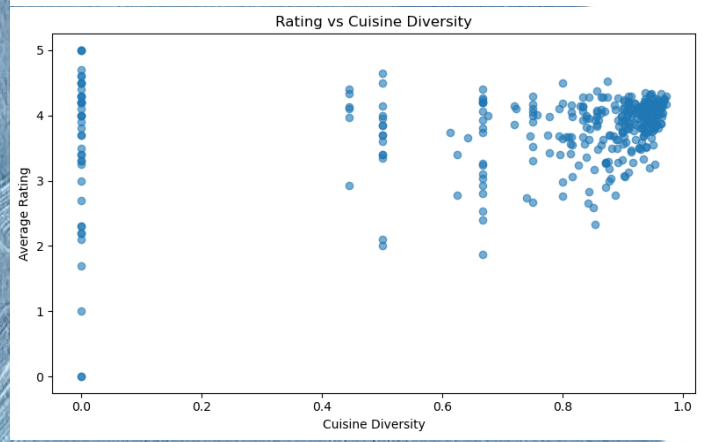
ZIP-LEVEL AGGREGATION REVEALS STRONGER PATTERNS

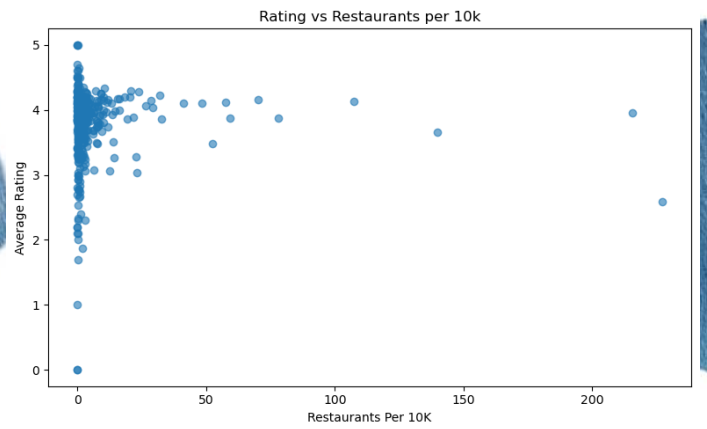
```
ZIP-level dataset built: 360 ZIP codes
=== ZIP-level Linear Regression ===
R^2:  0.3762
RMSE: 0.5200

=== ZIP-level Random Forest Regressor ===
R^2:  0.4278
RMSE: 0.4980
```

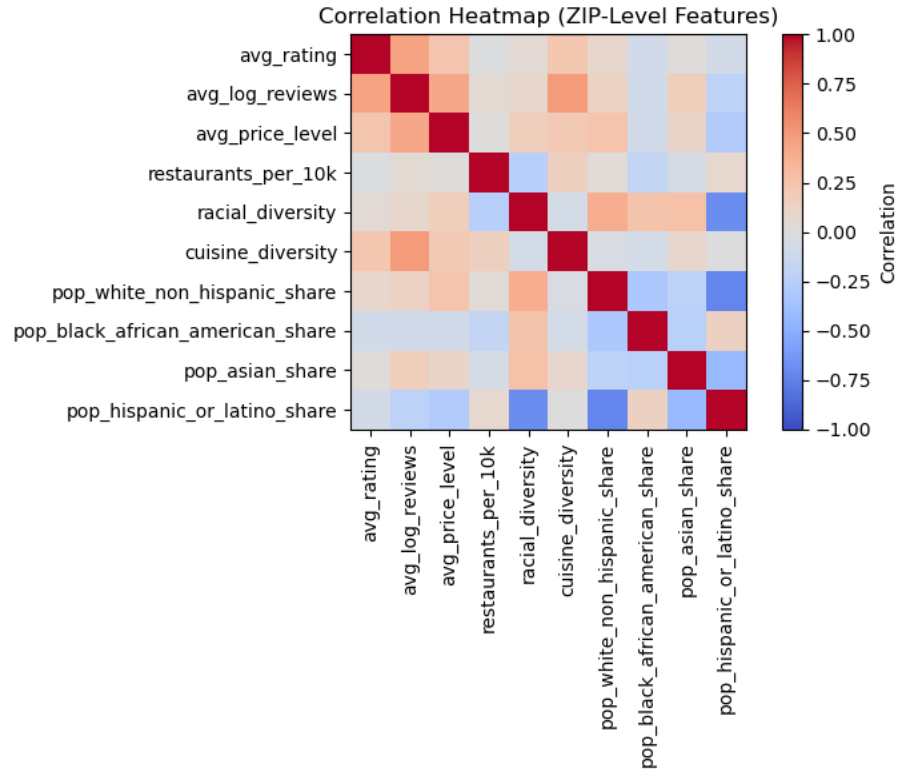
avg_log_reviews	0.324628
avg_price_level	0.256951
racial_diversity	0.106906
cuisine_diversity	0.092705
restaurants_per_10k	0.085259
pop_asian_share	0.043423
pop_black_african_american_share	0.037443
pop_white_non_hispanic_share	0.026578
pop_hispanic_or_latino_share	0.026107





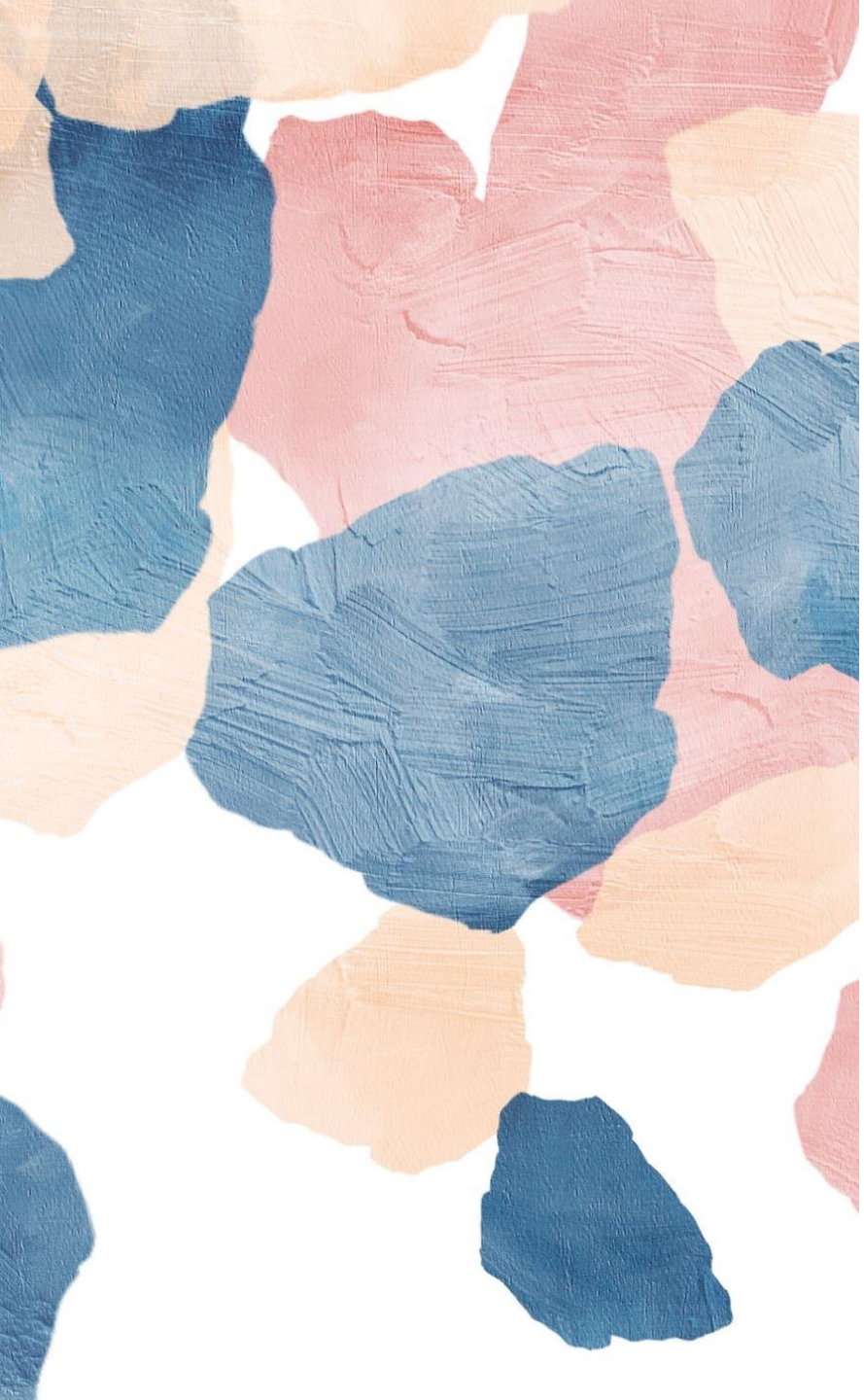






SUMMARY

Restaurant ratings in Los Angeles are primarily driven by restaurant-level factors such as price level, popularity, and cuisine type, as demonstrated by feature importance plots and correlation matrices. Neighborhood demographic variables show minimal influence at the individual restaurant level, but become meaningful at the ZIP-code level, where diversity and restaurant density moderately predict higher average ratings. Geographic visualizations reveal clear spatial clustering, with several LA regions consistently outperforming others in dining quality. Overall, the analysis shows that both micro-level restaurant characteristics and macro-level neighborhood environments contribute to restaurant performance.



CHALLENAGES

- Yelp categories needed extensive cleaning.
- City name inconsistencies -> NaN issues during merging.
- Spatial reasoning required for ZIP/city analysis.
- Large dataset required careful preprocessing.


THANK YOU!