Progress report

1.Project scope update

The project scope remains consistent with the original proposal: analyzing how restaurant and neighborhood characteristics influence Yelp ratings across Los Angeles. The main analytical goals are to identify correlations between restaurant success and factors like cuisine type, price level, and local demographics, and to build predictive models (Multiple Linear Regression and Random Forest Regressor) to estimate ratings.

2.Data sources

1) Restaurant data has been collected successfully from the Yelp Fusion API, including restaurant names, categories, ratings, review counts, price levels, and ZIP codes within the Los Angeles area. 200 restaurants data retrived successfully in the test runs. Preliminary analyses have been conducted on these records, including correlation heatmaps and feature importance plots.
2) API used:
    - Yelp Fusion API
    - Planned: Google API

3.Issues / difficulties

I encountered a Yelp API limit where requests fail (HTTP 400) once the offset exceeds about 200, so only ~200 restaurant records can be retrieved per query. To address this, I plan to split Los Angeles into smaller areas (e.g., by ZIP code) and merge results. The 50-record per request and rate-limit constraints also slow data collection, and some entries lack fields like price or ZIP code, requiring cleaning.