Neil Bai, Colin Leung
DSCI 551 Project Proposal

Project Option: SQL

Datasets:

1) https://data.lacounty.gov/datasets/19b6607ac82c4512b10811870975dbdc/about
    a) Los Angeles County restaurant health inspection grades
2) LA County Demographic Data by ZIP Code(Tentative)
    a) Source a second dataset with meaningful information to join with our primary data set. Currently not set on what that will be, but for now we are thinking of potentially using demographic data by zip code.

Implementation Summary:

For this project, we will implement a custom CSV parser to load real-world datasets containing Los Angeles County restaurant data into memory as a simplified DataFrame structure. This structure will be built using Python dictionaries and NumPy arrays, supporting core SQL-like operations including filtering, projection, group by, aggregation, and join. We will develop modular functions for each operation to enable flexible search and recommendation capabilities, allowing users to explore restaurants by ZIP code, inspection grade, and other criteria.

Functions/Features

- Search inspections by name/ZIP/date window (filtering/group by)
- Compute avg inspection score by different categories(ex: avg inspection score in a specific zip code)

Timeline:

9/19 - Project Proposal

9/26 - Finalize and clean the two datasets. Design the class structure for our DataFrame and parser.

10/3 - Complete the implementation of the custom CSV parser.

10/10 - Implement and test the filter operations, prepare for midterm progress report.

10/17 - Midterm Progress Report

10/24 - Implement and test group_by and aggregation/analysis functions.

10/31 - Polish functions and start implementation on user-facing applications and interface.

11/7 - Integrate all backend functions with user-facing interface and conduct thorough testing for edge cases and bugs.

11/14 - Finalize project and error testing. Begin working on final presentation.

11/21 - Complete final presentation and practice presenting.

11/23 - Project Implementation

11/24 - Class Demo

12/5 - Complete final report first draft.

12/12 - Finalize final report.

12/15 - Project Final Report

Division of Labor:

Code

- Neil: data cleaning, aggregation, join, user interface
- Colin: data cleaning, CSV parser, filtering, projection, group by

Presentation + Final Report

During the in-class demo, each member will present the portion of the project for which they were responsible for. Similarly, for the final written report, each member will write the technical sections that describe the implementation of their respective components. We will then collaborate on the general parts of the presentation and report, such as the introduction, application description, and conclusion to ensure a cohesive project.