ZeroDown Hackathon

General Instructions:

- Create a private github repository and add collaborator access for zerodown2024@gmail.com as soon as you start the hackathon
- You can choose any one of the following problem statements.
- Completion of any 1 problem is mandatory, if time permits you may proceed with the next question.
- Ensure that you have a decent, stable internet connection.
- Push commits for every substantial progress in each problem as frequently as possible.
- Usage of the database to ingest and processing data is mandatory.
- Development of very basic UI, integrated with backend and hosting it publicly is an added advantage.
- Keep your code organized, modular and use comments.
- Make use of data dumps if necessary from the given link.

Problem Statement 1

PriceProbe: Predicting Property Values

In the dynamic realm of US real estate, accurate pricing stands as the cornerstone of successful transactions. With the market constantly evolving and property values fluctuating, the ability to determine fair and competitive prices is paramount. Using the partial raw market data provided, your task is to predict home prices for properties listed for sale, by progressing through the following milestones.

Milestones:

1. **ERD**: Add entity relationship diagram based on DDL statements provided.
2. **EDA**:
    - Range of attributes
    - Geographical spread
    - Temporal spread
    - Identify outlier homes and homes with incorrect data
3. **Homes Deduplication**: Devise an scalable algorithm to identify duplicate homes. Duplicates can be classified into 2 types, identify both separately.
    - Absolute duplicate - same home, listed in the market at almost same time.
    - Pseudo duplicate - same home, listed at different points in time.
4. **Home Comparables**: Given a home id, devise an algorithm to provide a list of similar homes.
5. **Price Estimation**: Given home attributes(bed, bath, city/zipcode etc...) estimate price based on sold homes.

Note:

1. For comparables & price estimation, bed, bath, city/zipcode will be mandatory inputs. Other attributes like finished_sqft, lot_size_sqft, home_type, etc... can be optional inputs.
2. Preferably, use PostgreSQL database.

Data:

home_info.sql
market.sql
market_geom.sql

Problem Statement 2

Urban Insights Explorer: A Dashboard for City and Zip Code Demographics

Create a visual dashboard that provides insights into the demographics of cities and zip codes. The dashboard should allow users to compare various demographic attributes across different cities and zip codes, enabling them to gain valuable insights into population characteristics, occupations, median income, and more.

Data Collection:

- Utilize the U.S. Census Bureau's website (https://www.census.gov/) to access the required demographic data.
- Specifically, retrieve the '2017-2021 ACS 5-year Estimates' dataset.
- Focus on gathering data related to demographics at the city and zip code levels (Consider Washington and California states), including but not limited to population size, age distribution, racial composition, educational attainment, employment status, median household income, median home value, etc.

Data Processing:

- Clean and preprocess the acquired data to ensure consistency and accuracy.
- Handle missing values, outliers, and any inconsistencies in the data.
- Perform any necessary transformations or aggregations to prepare the data for analysis.

Development:

- Develop a visual dashboard using a suitable technology stack such as Python, React, or any other web-based visualization framework.
- Design an intuitive user interface that allows users to interact with the dashboard easily. Implement interactive charts, graphs, and tables to present the demographic insights effectively.
- Include features for selecting cities/zip codes, choosing demographic attributes for comparison, and customizing the visualization settings.

Insights Generation:

- Analyze the demographic data to identify meaningful insights and trends. Provide statistical summaries, trend analyses, and visual representations to highlight key findings. Encourage exploration and discovery by allowing users to drill down into specific demographics and geographical regions.