

CIS 450 Project Milestone 2

Kush Pandey, Aakash Jajoo, Yathu Nadanapathan, Karan Jaisingh

1. Motivation for the idea/description of the problem the application solves

Airbnb is a rapidly growing company that provides users with the opportunity to rent houses/apartments on short notice. Due to the size of the company, users have several options on what place to rent. It is a challenge to determine which place is the best for a specific user: certain users may look for cheaper spots, others may look for highly eloquent spots, while others may simply care about the rating of the host. With the large amount of data regarding Airbnbs, we place on creating a website that allows users to filter Airbnb listings to find their perfect match.

In addition to providing general information about Airbnb listings across several popular cities in the US, we plan to specifically analyze the correlation between noise complaints in NYC with Airbnbs. If someone is looking to have a fun time in NYC, we may suggest an Airbnb near a bar, which is the source of many noise complaints. Meanwhile, if someone is looking for a relaxing, tranquil trip in NYC, we will suggest an Airbnb in a quiet neighborhood.

2. List of features you will definitely implement in the application

- Creating a search bar (for the specific city location) and multiple filters such as price, rating, purpose of visit (optional, but includes things like leisure, business etc.), number of reviews, room type etc. that users can change as they need to.
- If the user chooses to visit New York City, then we will also include features like making recommendations based on the number of noise complaints in the neighborhood or the number of bars in the neighborhood if they are visiting for tourism or leisure. In essence, there will be added search filters if the user is visiting NYC.
- Number of reviews, price and rating filters will be implemented using a slide bar with a scale depending on the quantity (rating from 1 to 5, reviews from 0 to the max in the dataset, price from 0 to the maximum in the dataset).

3. List of features you might implement in the application, given enough time.

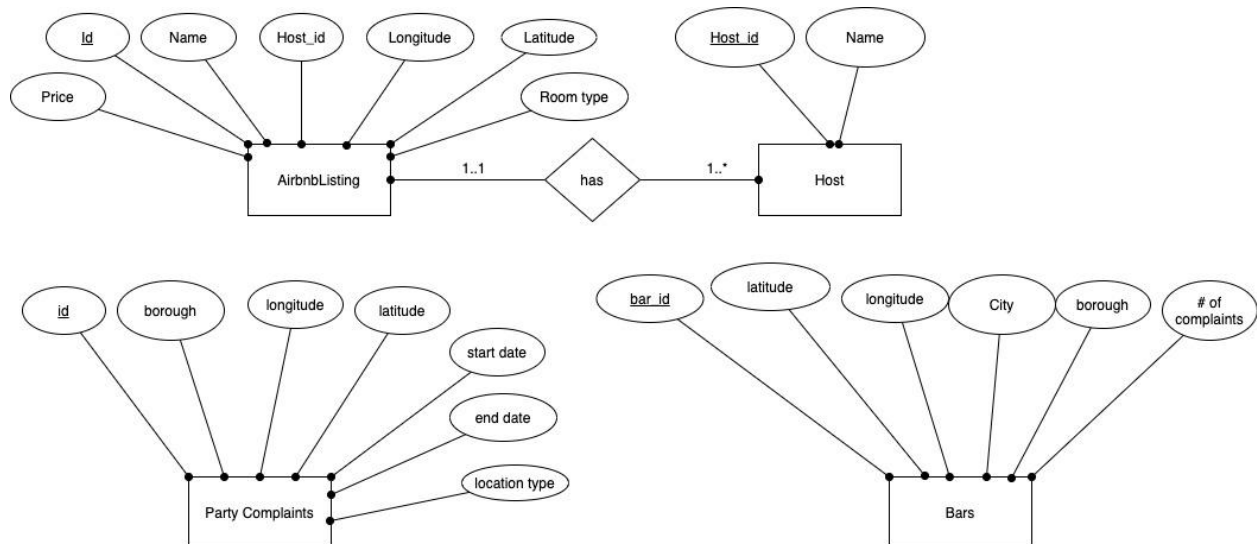
- Create a profile for users and give them personalised recommendations based on their previous stays
- Similar to NYC, have data on different neighbourhoods across the country so that the recommendations can be more comprehensive
- Get more data on factors that determine quality of life in a particular neighbourhood, including crime rates, quality of restaurants etc, and allow users to have more ways to filter neighbourhoods/cities
- An interactive map that will allow users to select a specific neighborhood.

4. List of pages the application will have and a 1-2 sentence description of each page. We expect that the functionality of each page will be meaningfully different than the functionality of the other pages.

- Home Page
 - Display random listings

- Provide a summary of the rest of the website
- Find My Airbnb
 - A web page that allows users to filter their Airbnb listings
 - Filter By:
 - Location
 - Price
 - Rating
 - Number of Reviews
 - Room Type
- Explore New York City
 - Provide some slider/scale from 1-10
 - 1 corresponds to very tranquil
 - 10 being very noise / party behavior
 - The scale allows people to filter airbnbs depending on how fun they want their trip to NYC to be
 - Once a person choose the level of fun, we will provide them with the following options:
 - Choose a borough
 - Choose price/rating/room type
 - Suggest some local bars
 - Using data on parties, we can recommend popular boroughs / time of the year

5. Relational schema as an ER diagram.



There is no direct relationship between the latitude and longitude of AirBNB listings and that of the noise complaints, though a formula will be utilised to match the two (as long as the distance between the two of them falls under a particular threshold).

6. SQL DDL for creating the database.

AirbnbListing(id, name, host_id, longitude, latitude, price, room type)

Host(host_id, name)

- Host_id foreign key references AirbnbListing(host_id)

Party_Complaints(party_id, borough, longitude, latitude, start_date, end_date, city, location_type)

Bars(bar_id, latitude, longitude, city, borough, number_of_complaints)

7. Explanation of how you will clean and pre-process the data. This tutorial demonstrates how to do simple pre-processing in Python.

We will be doing the preprocessing using Python. This may involve using libraries such as Pandas, Numpy and Matplotlib (for visualisation purposes). Simple CSV operations using Microsoft Excel may also be helpful.

Our preprocessing will consist of simple operations, such as removing entries in the database that have characters that cannot be recognised and removing null rows. We will also remove rows and columns that have impossible values like negative price, negative number of reviews etc.

8. List of technologies you will use. You must use some kind of SQL database. We recommend using MySQL specifically because you will use MySQL in HW2, and we will provide guidance for setting up a MySQL database.

We will be implementing our platform using a stack similar to what we used in Homework 2 - the frontend will be done using HTML, CSS, Javascript and React, and we will use Javascript and SQL (via MySQL) for the backend.

9. Description of what each group member will be responsible for.

- Yathu: Responsible for cleaning the data and writing SQL queries.
- Aakash: Responsible for setting up the website (Frontend + UI + React).
- Karan: Responsible for setting up the website (Frontend + UI + React).
- Kush: Responsible for cleaning the data and writing SQL queries.