Personalized Apartment Recommendations

https://github.com/ansellim/apartments.git

Team 153 – Ansel, Daosheng, Keith, Key

Motivation and Introduction

Aspiring homeowners in Singapore are constrained by biases in multiple information sources, as the content creators on many platforms tend to be realtors, who have skin in the game.

Moreover, the current available tools for researching the property market involve filtering among certain attributes such as location, price and size of the home. The thought process in purchasing a home often involves a much greater variety of factors, such as proximity to various amenities (e.g. food & beverage, schools, recreational facilities). Each individual would have his/her own preferences and priorities, hence it is difficult to have a one-size-fits-all approach in determining the best place to live.

We have developed a tool which aspiring homebuyers in Singapore can use to analyse and visualise multiple unbiased sources of information, as well as provide recommendations to the users based on their specified preferences and constraints.

Approach

Obtaining Data

First and foremost, unbiased sources of data had to be identified. The Government of Singapore maintains and publishes publicly available datasets (www.data.gov.sg & www.ura.gov.sg/maps/api/), and a comprehensive map with an open API (www.onemap.gov.sg), both of which are extremely useful for our tool. Through these sources, we were able to download and scrape details of residential projects (e.g. sale transactions, addresses, map coordinates) and locations of various amenities (e.g. parks, educational institutions, food and beverage establishments). We further enriched the information with data from Google Places API (https://developers.google.com/maps), adding details such as review ratings and number of reviews.

All in all, we retrieved 11,488 rows of property transaction data and 32,695 rows across 15 different types of amenities data.

Data Massaging

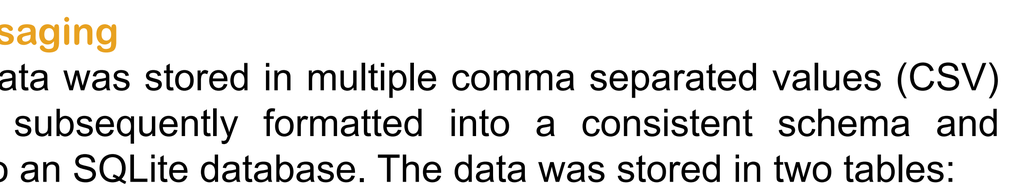
The raw data was stored in multiple comma separated values (CSV) files, and subsequently formatted into a consistent schema and loaded into an SQLite database. The data was stored in two tables:

1. "Properties" table, containing details of the various residential housing projects such as name, district, location, average transacted price.
2. "Features" table, containing details of the different amenities such as name, category, location, review ratings.

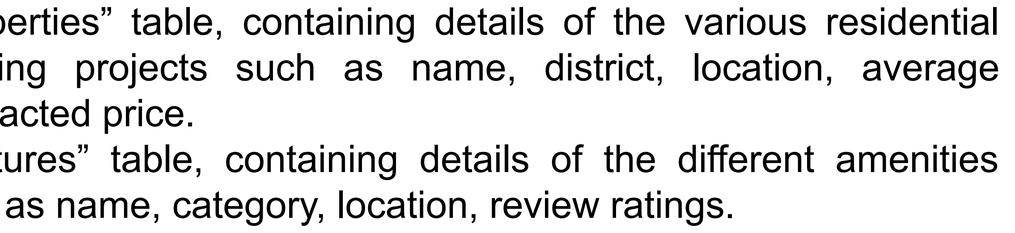
User Input

To take into account the user requirements, we created a web questionnaire for the user to indicate his/her constraints and preferences. Constraints, such as price and floor area, are applied as filters to the data. Preferences are indicated via slider bars, where a user indicates how important each feature is to him/her.

Constraints:



Preferences:



Algorithm

Each residential project in the "Properties" table is enriched with a series of scores, by taking the count and weighted ratings of all amenities from the "Features" table within a 1 kilometer radius of the residential project. This is a processing-intensive task, which involves iterating through thousands of housing projects and calculating the scores.

Approach (continued)

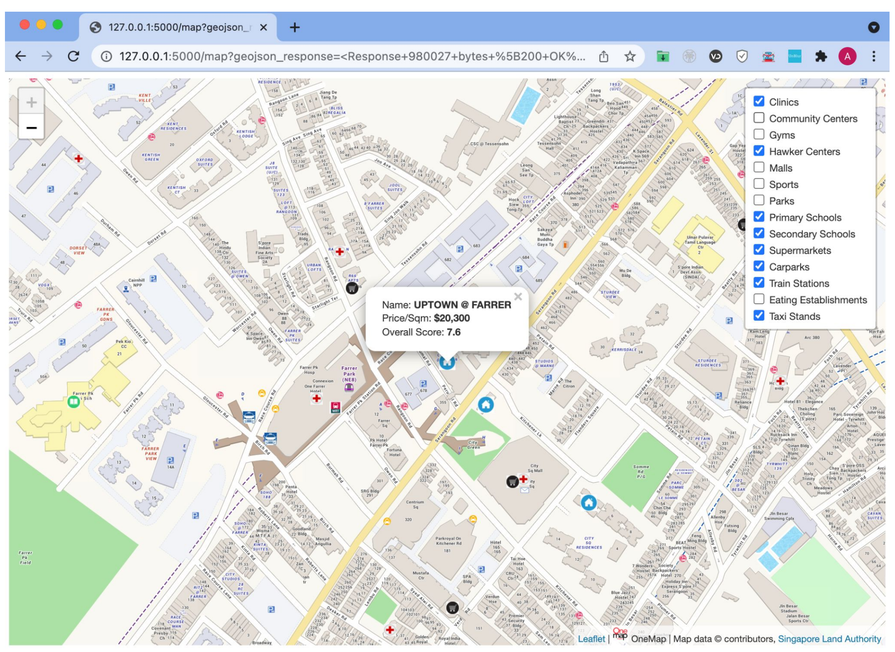
With the constraints and preference weights indicated by the user via the questionnaire, we then filter relevant housing projects and calculate the sum-product of indicated weights with the projects' feature scores to obtain an overall score for each housing project. In a simplified example with only two features of schools and transport links, the overall score of a housing project can be calculated as

$$\text{Overall score} = (weight_{school} * score_{school}) + (weight_{transport} * score_{transport})$$

By sorting the overall score in descending order, we can provide users with most relevant recommendations as per their constraints and preferences.

Interactive Map Interface

Once we obtain the list of recommendations, the results are shown in an interactive map. To avoid cluttering the interface, the top 5 residential projects are shown as individual pins on the map. Clicking on each pin allows the user to view the details of a particular housing project. The user can also check or uncheck options in a side menu to display or hide various amenities in the vicinity.



Novelty of approach

Our tool is able to take into account user requirements and a large variety of features to provide relevant recommendations. Our scoring method is novel as it incorporates quantity and quality aspects of a feature, where quality takes into account the raw rating and number of votes. Moreover, each feature has a different score assignment mechanism based on its distribution. Last but not least, we provide a good user experience for users to input their requirements and explore the recommended residential projects in a map.

Experiments and results

For our experiments, we tested the tool on several "user personas", and checked if the constraints defined have been met, and the recommended results included features within the top quartile of residential projects across Singapore. We also evaluated the time taken for the tool to compute the results, as this would significantly affect the user experience.

Persona #1 - Young parents, below average income
Fixed constraints: Price limit of $5,000 per sq m, focus on 2 districts.
Preferred features: Schools, healthcare, hawker centres, MRT stations.

Persona #2 - Single, young adult, average income.
Fixed constraints: Price limit of $8,000 per sq m, no district restriction.
Preferred features: Recreational facilities, food & beverage outlets, transport links.

Persona #3 - Large multi-generational family, wealthy.
Fixed constraints: Price range between $20,000 to $30,000 per sq m, focus on 3 upscale districts.
Preferred features: Schools, healthcare, recreational facilities, malls.

The tool performed well and achieved the intended outcomes. For all of the tested personas, we noted that the top 5 results returned top quartile results for most of the preferred features. Furthermore, it took under 2 seconds from questionnaire submission till the results are displayed, which is excellent from the perspective of user experience.