YYDS Group Project

House Price of Beijing Shanghai and Gaungzhou

Reason topic was selected

All of our team members are from China and are ready to graduate, after graduation, we may face the problem of buying a house for us. So, we are curious about the house price in China, and also the trend for it.

With our analysis, we know what kind of house is cheaper and we can predict the house price of any kind of house in Beijing, Shanghai and Guangzhou.

Description of the source of data

We selected some data from Kaggle

https://www.kaggle.com/ruiqurm/lianjia

All Data in Kaggle is from the following website

https://bj.lianjia.com/chengjiao.

but thats not enough, to make our analysis more complex and meaningful, we did a lot of web scraping and got House Price in Shanghai and Guangzhou as well.

Moreover, we not only considered houses that already been sold, but we also considered houses that are currently on sell.

Questions the team hopes to answer with the data

1. Which city's houses are more expensive.

2. Which district's houses are more expensive.

3. Create interactive tables and charts for customers and our audience to select their dreaming house.

4. Predict house price for customer's dreaming house. If houses with these features do not exist yet.

Description of the data exploration phase of the project

As I discribed in the source of data, first we did a lot of web scraping to make our data more complex and meaningful, the challenge in this part is that it take a long time for us to actually scrap and store those data.

Then we do cleaning, we drop those outliers and uncommon rows, split the Chinese characters and English characters. And then make sure the data types for each column is exactly what we suppose. And we also dropped useless columns.

Finally, we created a database, we need to database not only for our project requirment, but also for team communication. We transformed our data in different forms such as js, json and sqlite. So that the data can also be used in creating maps and dashboards.

Description of the analysis phase of the project

We created dashboards to present the data to our audience, include interactive tables and map with marks.

Also we used some machine learning technics to predict house price with given features. During this progress we selected models which performs best with the highest R2 score and lowest mean square error.

DashBoard

Now let's have a look at our beautiful, gorgeous, remarkable, splendid, amazing dashboard :)))

We aim to create a dashbaord that can link to different pages.

- We will have home page, about us, team, page information for different cities, mapping, etc.
- We use javascript to make the dashboard
- We plan to add what we learn from the module (searching bar, mapping, graphing)