26/05/2023- Final Presentation

Project Deliverables and Goal achieved

PropertyAI: A one-stop solution for real estate data powered by AI

Problem Statement & Project Goal

- **Bangladesh's property market is important for the country's money matters, but it has problems with sharing information and being open.**
- **This makes it hard for people to make smart choices.**
- **The lack of good data encourages dishonest behaviour and makes people lose trust.**
- **To solve this, we need a system that can show easy-to-understand, correct property data. Current websites give basic services, but they don't have advanced data analysis or complete data, so they can't serve people's needs fully.**

PropertyAI: A one-stop solution for real estate data powered by AI

- Project Duration March **28/03/2023 - 28/05/2023**
- Initially Registered Collaborators: 143
- Final number of collaborators : **104**
- Tasks - Data Collection, Data Pre-processing , Exploratory Data Analysis, Feature Engineering and Model Development, Model Deployment
- We also have knowledge gathering and sharing task working in the background throughout the project.
- Each task was lead by a **Task-Leads and co-leads**

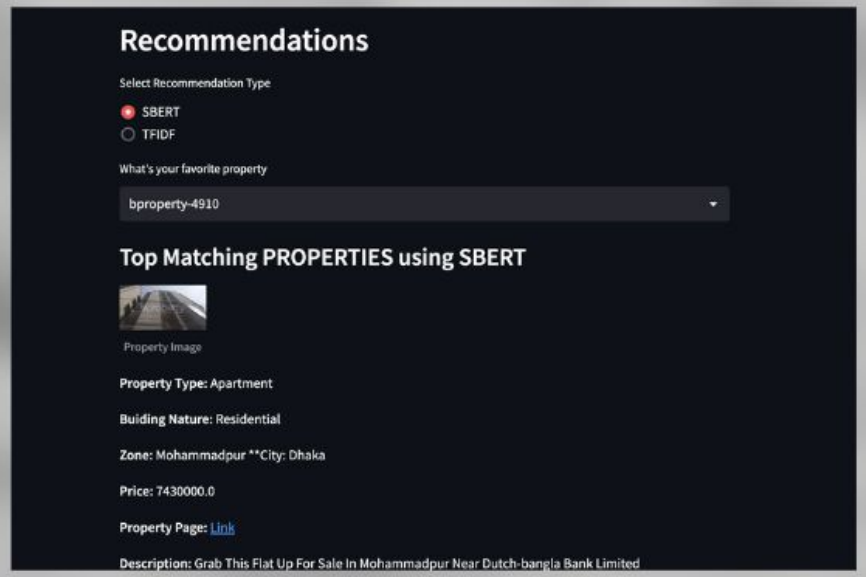
The Platform we used

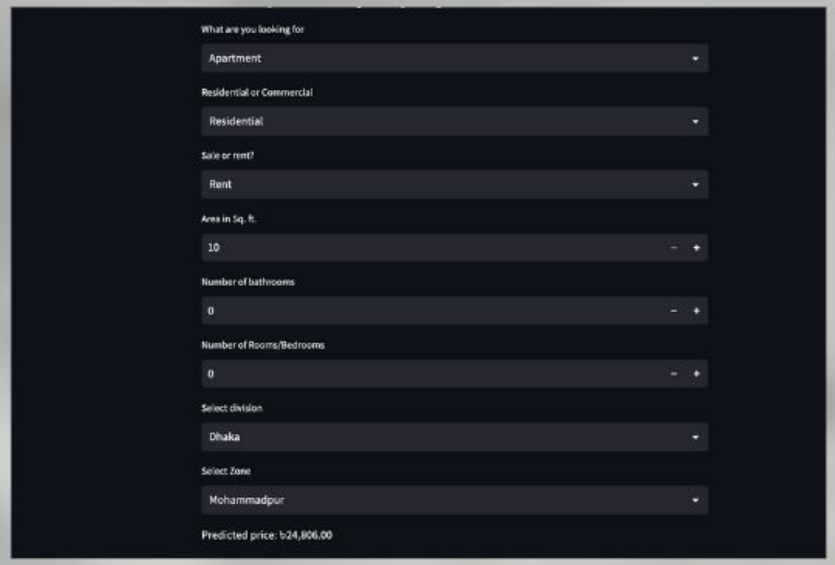
- Slack for Communication
- Github Private Repository
- AWS for Web App deployment
- Notion and Google docs for Project management and document sharing


Interact with APIs
Understand how APIs of the sites that hold the data work and how to extract useful information from it.
Web scrape
Learn different scraping techniques and determine which is better for our case.
EDA & Visualization
Delivery of Insights
Finalize the project outcome and represent the findings in an interactive Web-App.
mdena

RESULTS

Interactive Web-App on streamlit





mdena
LOCAL CHAPTERS

Task 00 - Knowledge

Objective

Definition of the problem

The present project is going to be developed with the purpose of generating recommendations for rental real estate properties based on their attributes or characteristics.

Problem to be solved

Enhancing the user's experience and assist them in finding real estate products that align with their search queries

Project structure

It will be designed as an end-to-end project, encompassing all stages from data collection to project deployment.

Summary

1. Creating the project tracker to facilitate easy monitoring and management of project milestones
2. Updating participants with recent developments in ML
3. Generating a blog post summarizing knowledge base and key findings

Work in Progress - Next Steps

- Relocating documents to notion

Task 01 - Data collection

Problem Statement

Addressing the Lack of Transparency and Information Asymmetry in the Real Estate Sector in Bangladesh

:

Work Pipeline

Step 1

- Research and list various websites and other sources from where to scrape the data.
- Make a final list of the identified sources that are relevant to the project

Step 2

- Out of various scrapers present, finalize scraper or combination of scrapers that will help in getting the data in the desired format
- Finalize scraping script and generalize if possible to be used by everyone

Step 3

QC the scraped data and rework if required

Output

Put the data in a collated sheet or repository

Workflow:

- The task was primarily divided between the Co-leads under whom the collaborators were mapped
- Each co-lead then divided the task further between the collaborators
- Knowledge sharing sessions were conducted in order to enable the collaborators to contribute confidently to the scraping activity

Work Process

- 19 collaborators actively participated in this task
- A total of 132 websites were identified for scraping
- 37 websites were found to be relevant after final review and research
- 20+ websites have been completed and pushed to Git
- Nearly 60k rows of data scraped
- Leveraged on tools like Scrapy and Selenium

Co-Leads : Al Momin Faruk, Vijay Mamilla, Sunitha L V

Work Process - Learnings

- Data was scattered over numerous regional websites
- Some data were outdated as websites did not update them
- Vernacular website content made it a challenge for translation
- Most of the features were hidden in the ‘property description’
- Most listings were biased towards a few locations like Dhaka etc.
- Different websites had different dynamics and needed customised scripts

Results & Visualizations

Current progress(In Progress, On Review, Done)	COUNT
Done	20
In Progress (incomplete)	5
Coil not scrape	12
Grand Total	**37**

Task 02 - Data preprocessing

Problem Statement

- *Clean the data, and put it in the right format for the work in Tasks 03 and 04*

:

Work Pipeline

Step 1

Explore the features of the data from Task 01 to understand their structure

Step 2

- Assess the data to locate cleaning work that needed to be done
- Enumerate cleaning tasks

Step 3

Forming sub-teams to work on cleaning tasks

Output

Data in the right format for Tasks 03 and 04

Work process

- Each Collaborator worked on cleaning csv files he chose
- Cleaned csv files were put together in one final and global csv file
- Final csv file where further cleaned
- Feature engineering were performed on some features of the final csv file

Results

- Delivered two intermediary/initial csv files (in order to allow some works to be started in [Task 03 - EDA]
- Delivered a final, clean and tidy dataset:
  - 24 features by 33710 samples

Task 03 - EDA

Problem Statement

- Build an intuition about the real-estate industry in Bangladesh
- Providing informations that can orient model development (recalling that we have two models: properties recommender engine and properties prices recommender on rent/sale)

:

Work Pipeline

Step 1

Check if the data from Task 2 is in the required format

Step 2

1. Framing questions
2. Univariate Visualisations
3. Bivariate/Multivariate Visualisations

Step 3

Carry out descriptive statistical analysis of the patterns found in step-2

Output

- List of features that influence a property price
- Other insights generated through visual and statistical explorations

Week 03 (Task 02 week)

- **First 02 days**: work in tandem with collaborators of Task 02 and give inputs on their work, while getting familiar with the features in the dataset
- **Following 03 days**: form team of 02-03 members; each team suggest 10-15 questions to be explored
- **Following day**: Tidy the questions together, and eliminate duplicates or those that are irrelevants
- **Last day**: quick-off meeting for Task 03, each team choose questions to explore

Week 04 (Task 03 week)

- **First day**: check the data (refer to step 01 above in the work pipeline)
- **Following 02-03 days**: each team explore its questions (refer to points 02 and 03 from step 02 above in the work pipeline)
- **Following day**: step 03 (in the above work pipeline)
- **Following day**: generate a report based on the finding from all team members
- **Last day**: present our work report (through a meeting)

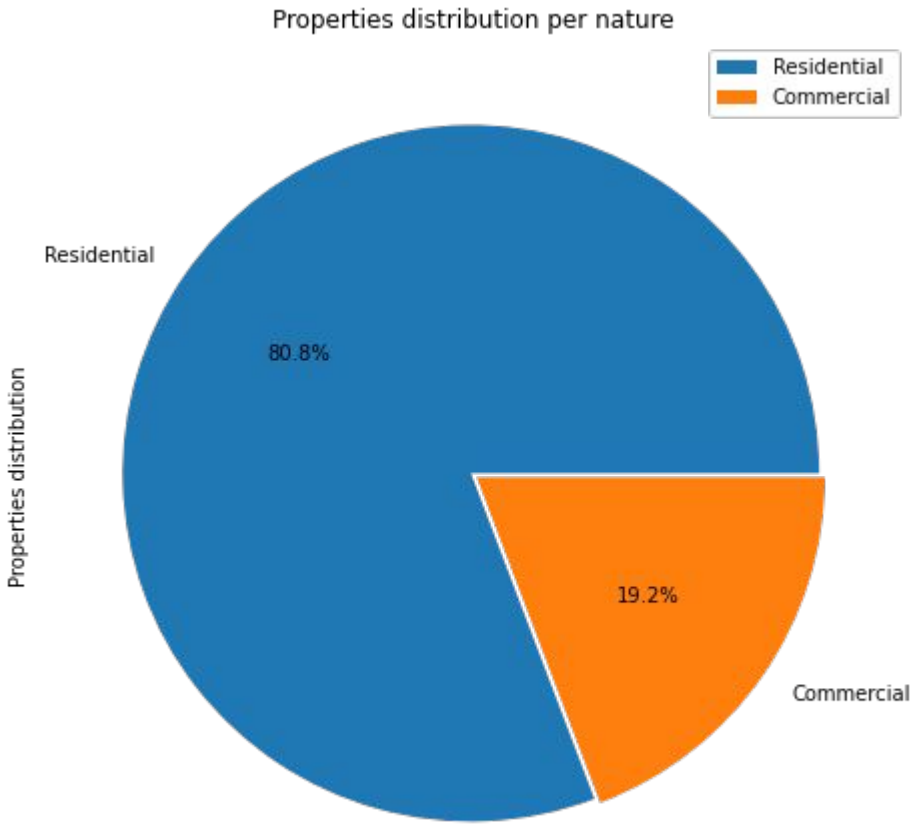
EDA - process

- Got answers from our (collected) data
  - Work done in Teams of 03 Collaborators
  - Each Team got some questions to work on
- Got answers from (real-estate) industry experts

EDA - Findings and Industry Expert inputs

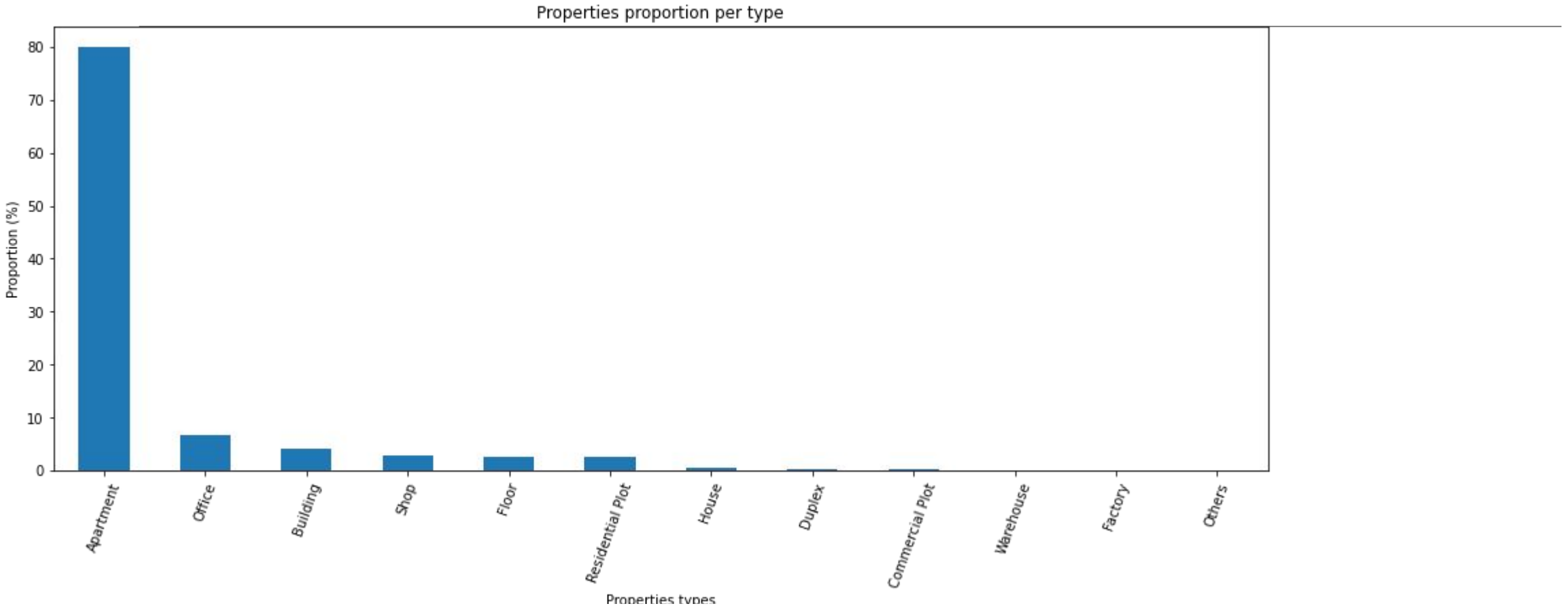
[please refer to next slides]

Properties distribution per nature



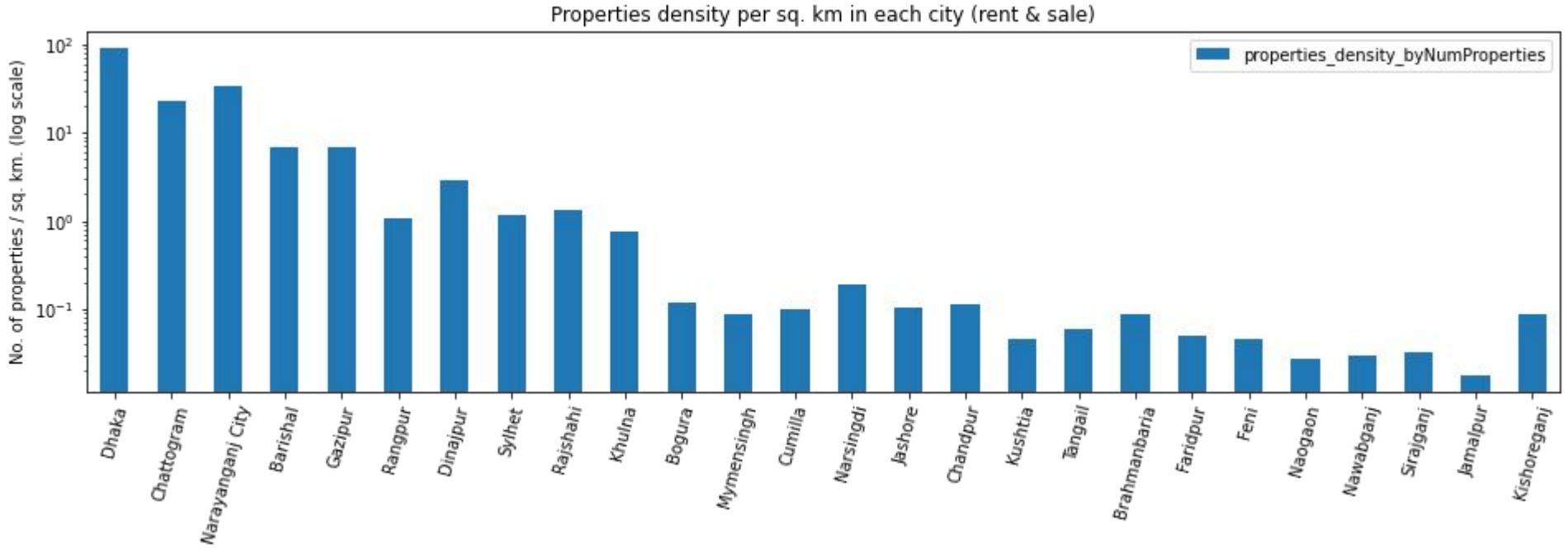
81% of our properties are `Residential`, while the rest are `Commercial`.

Properties distribution per type



1. Nearly 80% of our properties are `Apartment`, for a total of nearly 27000 samples. We also have some `Office`, `Building`, `Shop`, `Floor`, `Residential Plot`, whose number are under 10% of the total dataset; that is to say, their numbers are under 2500.
2. There are other types of properties, in a very negligible number.

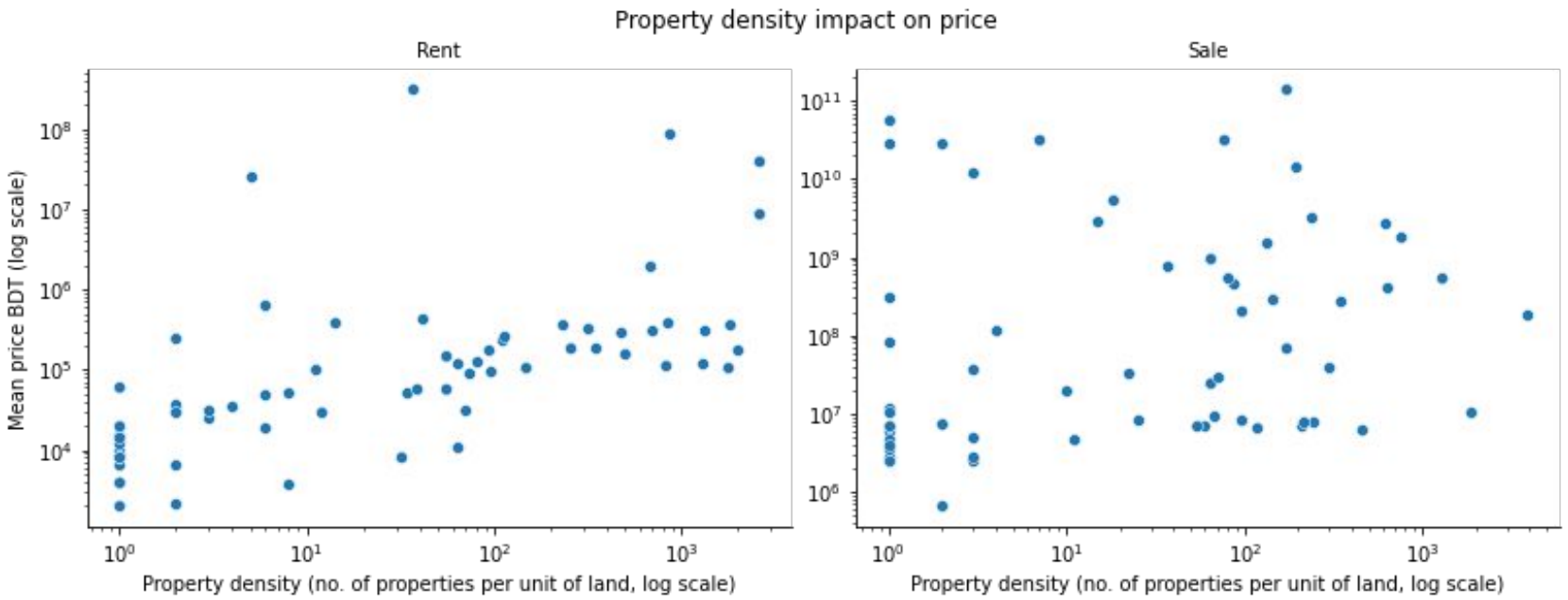
Properties density per sq. km in each city

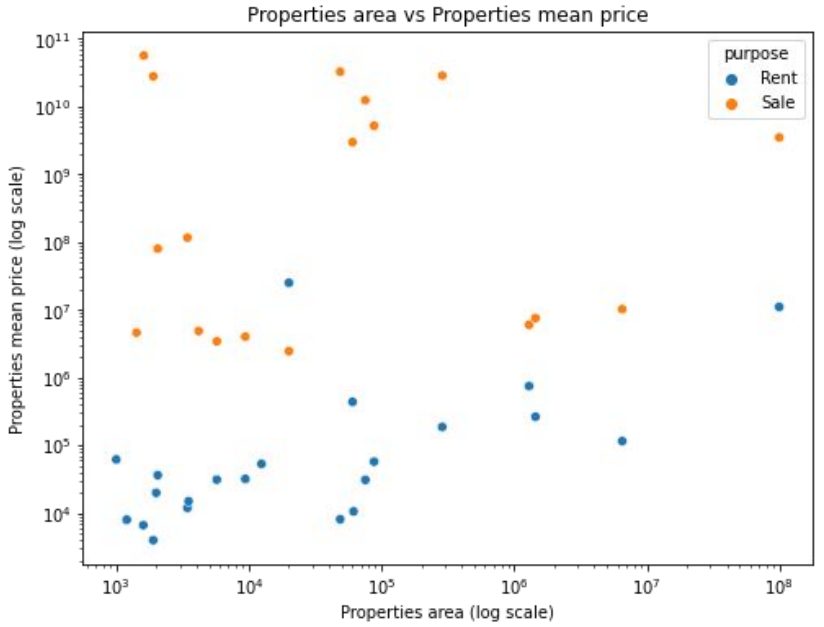


Properties density is a measure of the number of properties we find per unit area of land (here, km).

- `Dhaka`, `Chattogram`, `Narayanganj` are the cities with the highest properties density.
- The density is concentrated toward those 03 cities, specially `Dhaka`
- We are expecting high properties density to mean high population density, which in turn will mean higher properties prices

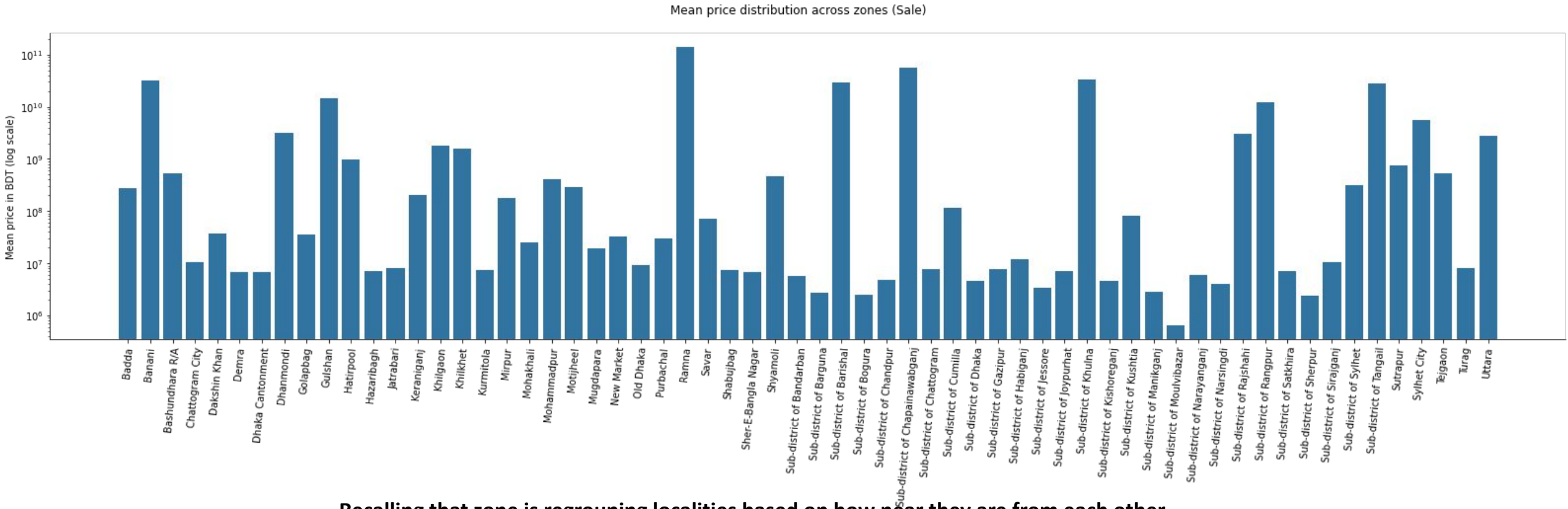
Properties density’ and area’ impact on prices





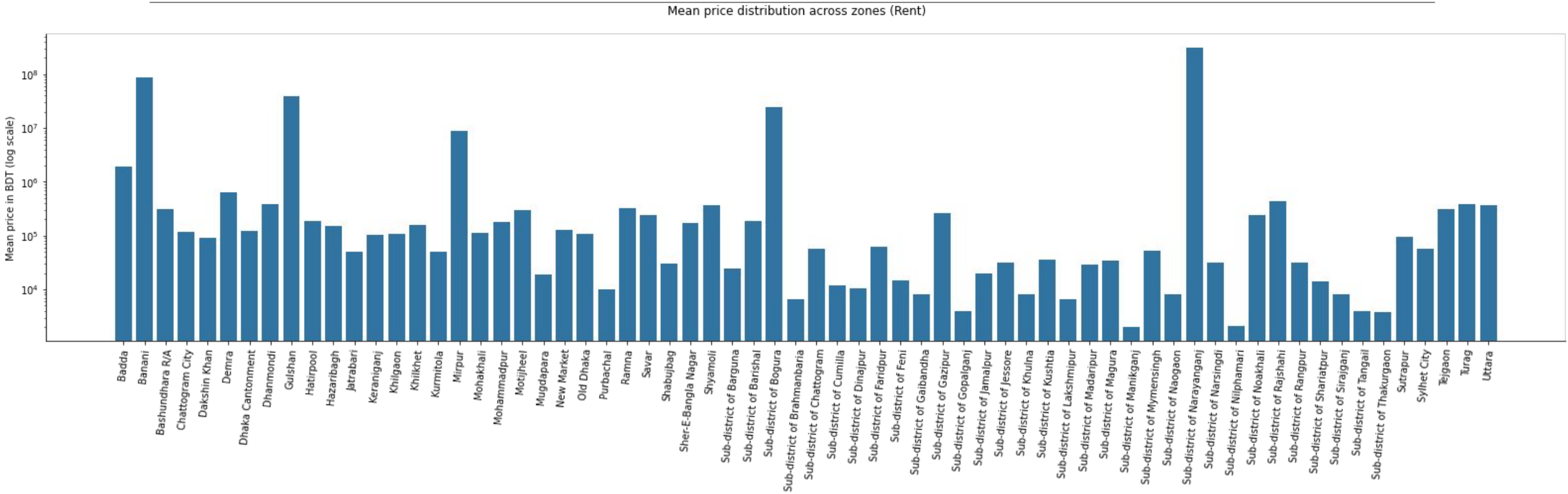
We can confirm that properties density and area are impacting the rent’ and sale’ prices

Geographical situation impact on sale prices



- **Recalling that zone is regrouping localities based on how near they are from each other**
- **Sale prices are impacted by the geographical location of the properties**

Geographical situation impact on rent prices

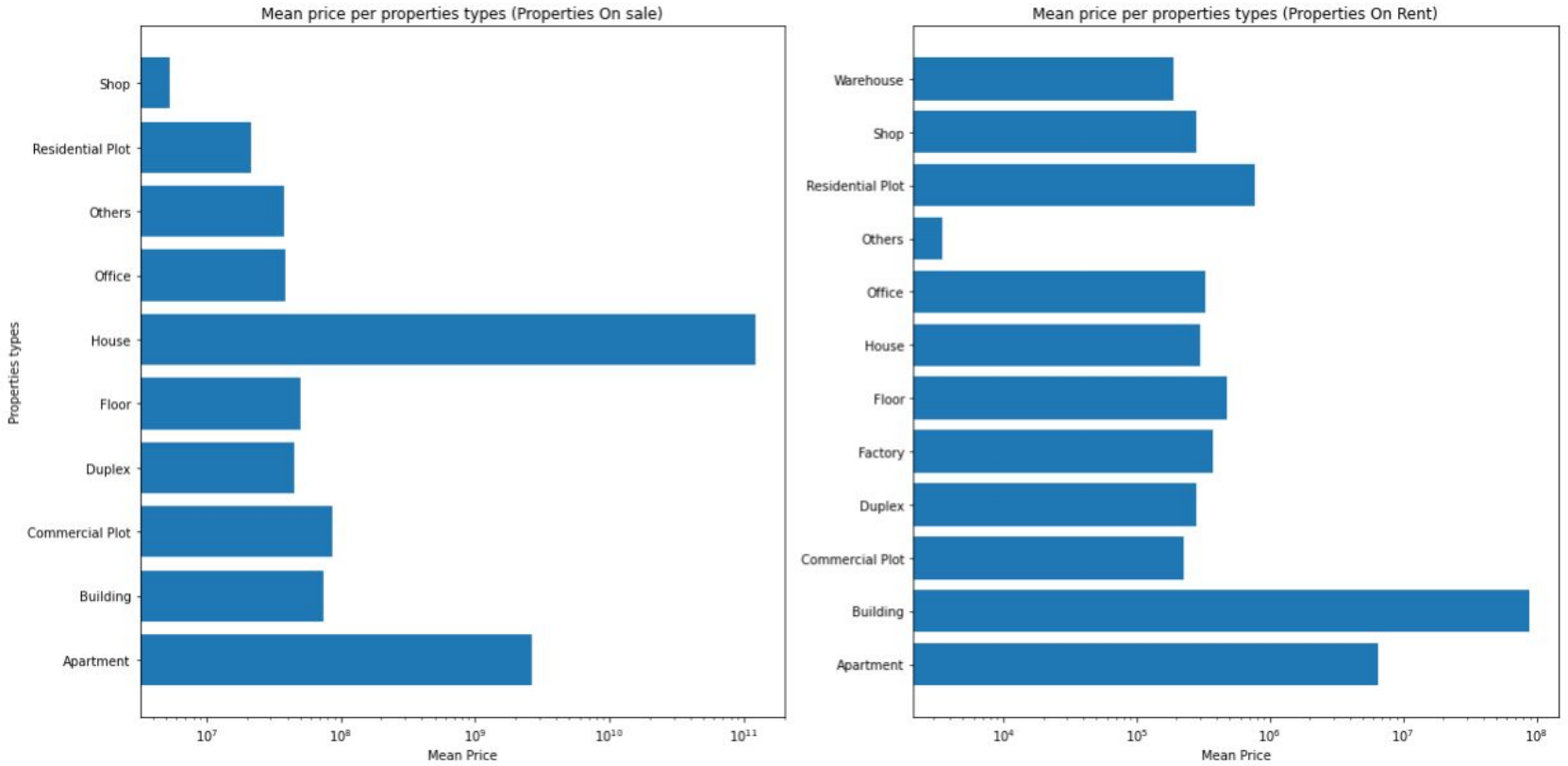


- **Recalling that zone is regrouping localities based on how near they are from each other**
- **Rent prices are also impacted by the geographical location of the properties**

Geographical location’ impact on prices
[What do Industry Experts say]

- “I think this is the basic factor in price differences among properties”
-

Mean price for each property type

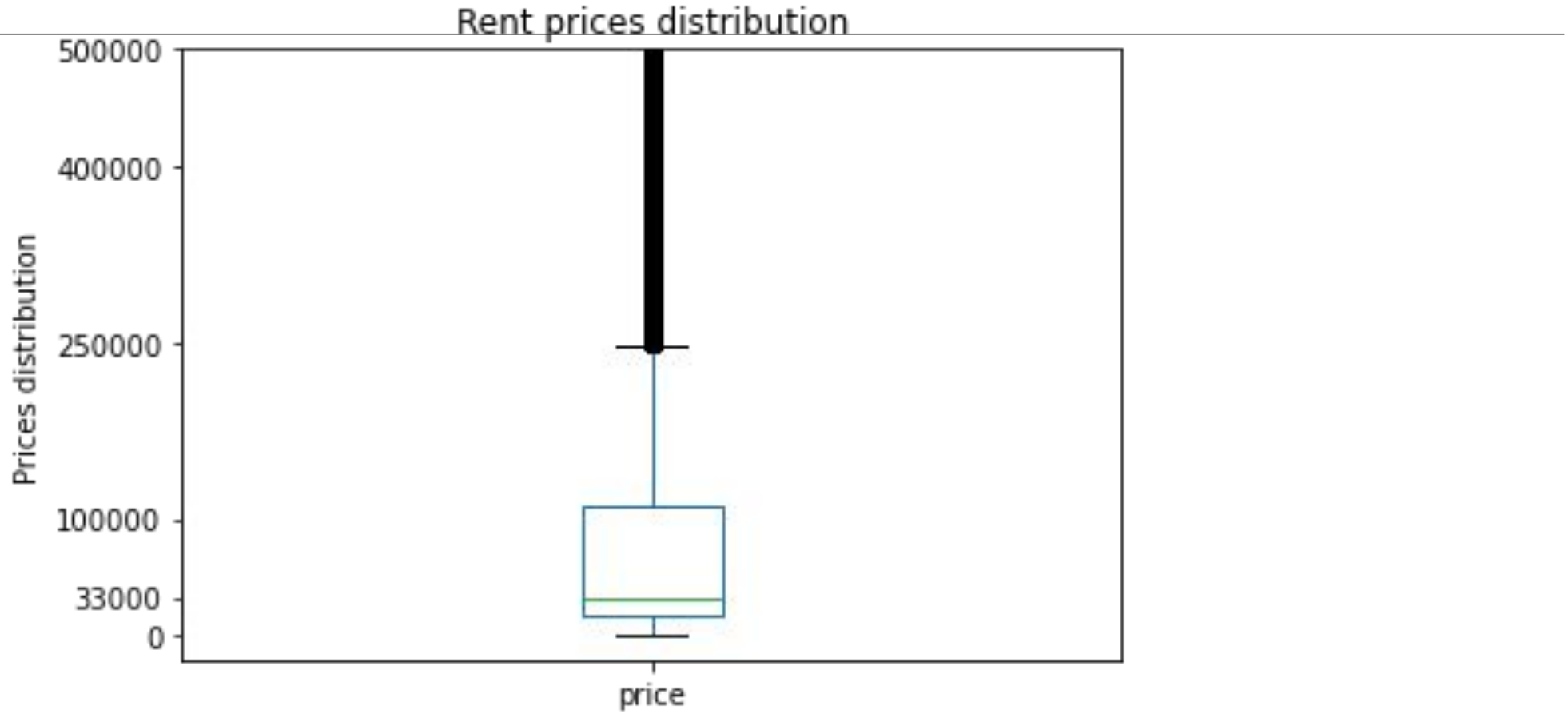


Please take note of the fact that the prices are not on the same scale over the two plots.

`House` and `Apartment` on sale are very expensive compared to other types on sale. The `Shop` is the cheapet of all. Other property types on sale tend to have nearly the same prices.

`Building` and `Apartment` on rent are the most expensive types of property. `Residential plot` is third in line in term of price. As for the other types of property, their rent prices are nearly the same.

Properties on rent' prices distribution

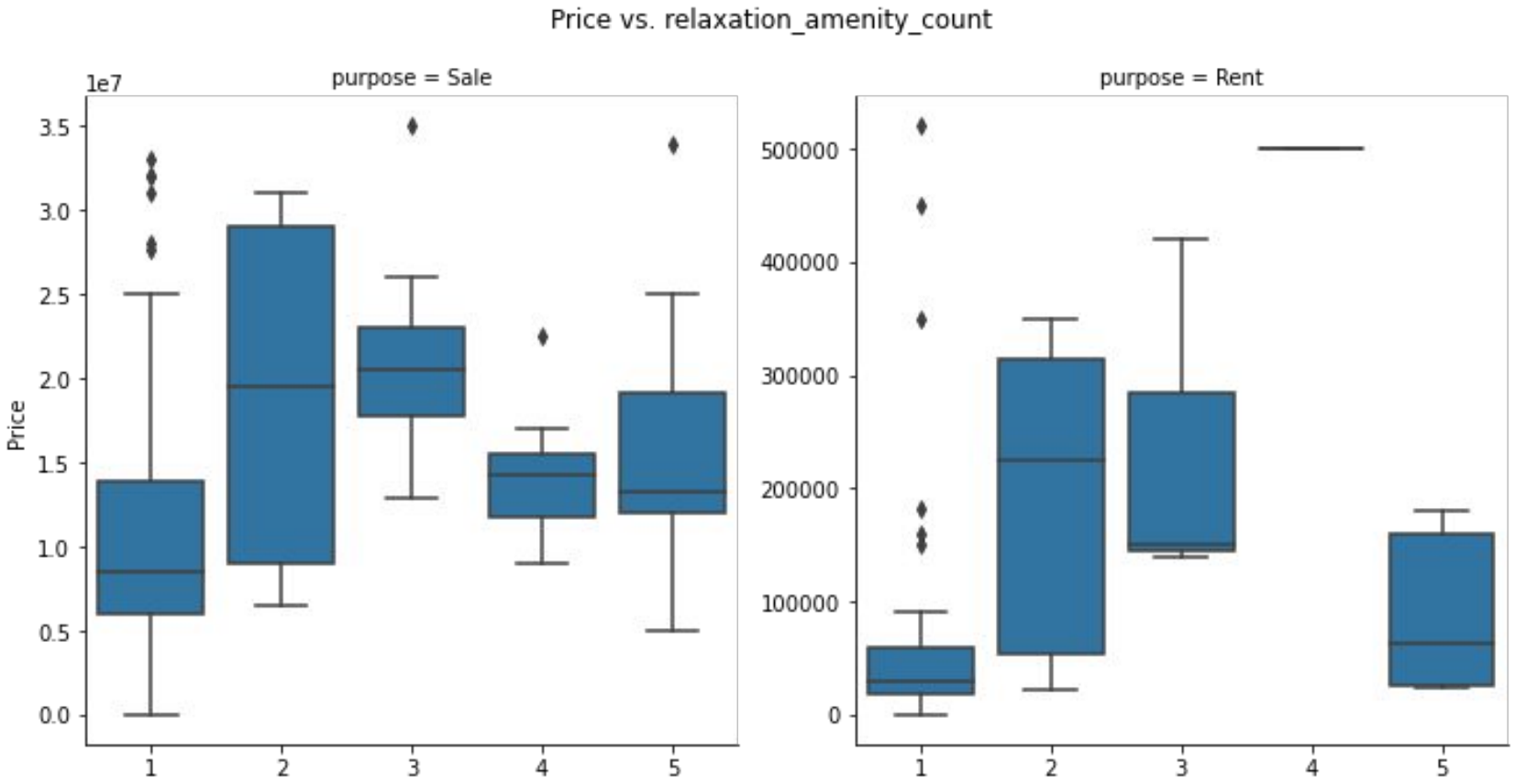


Most rented properties tend to cost 33000 BTD. We can also see that the majority of rented properties cost between 33,000 and 100,000 BTD. Some rented properties have prices situated between 100,000 and 250,000 BTD.
As for properties rented below 33,000 BTD, they are very few in number.

Properties on sale’ prices distribution

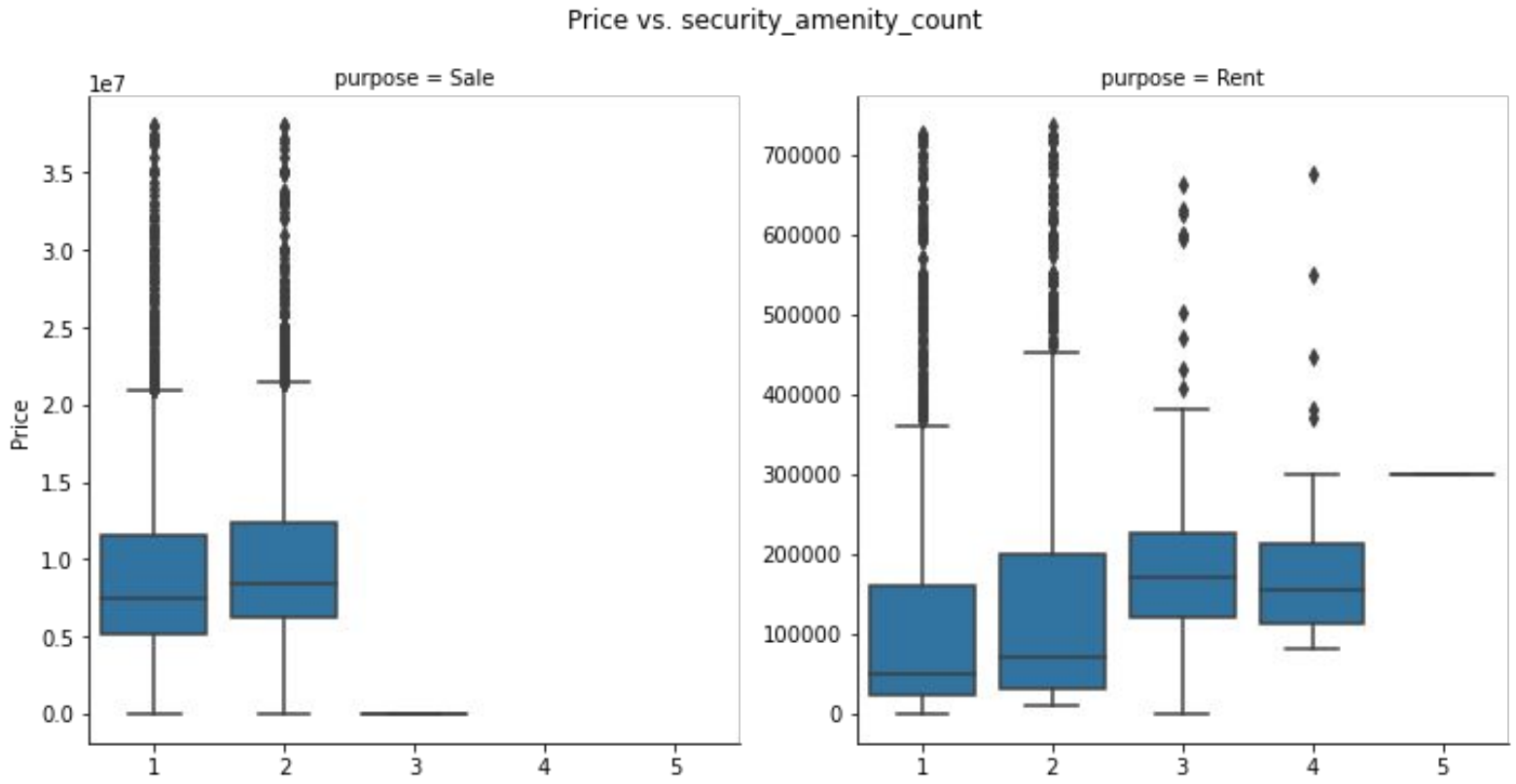
- Most properties are sold at 7,350,000 BTD.
- 02 third of the sale prices are comprised between 7,350,000 and 11,500,000 BTD.
- A third of the sold properties are between 5,000,000 and 7,350,000 BTD.

Relaxation amenities' impact on the price



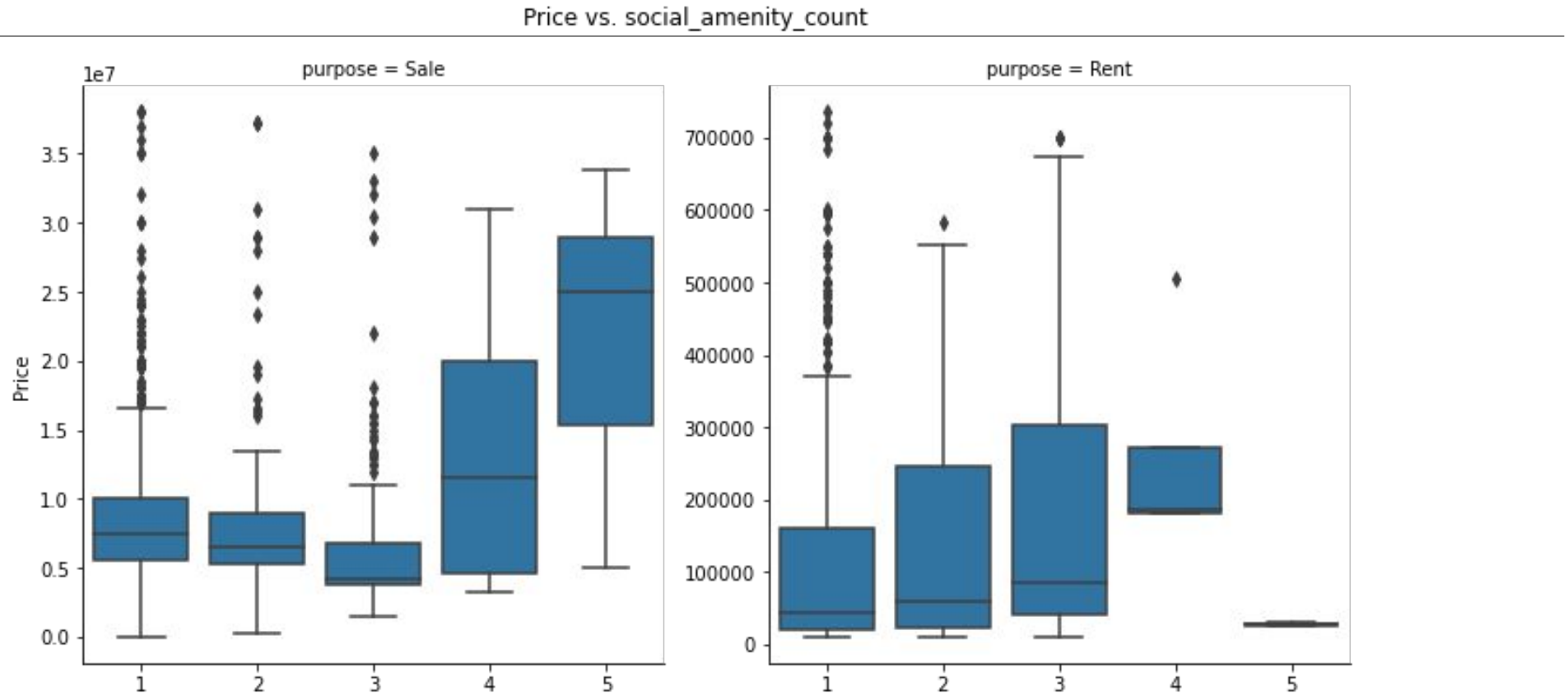
- The number of relaxation amenities a property has has an impact on its price. Let's note that when the property is on sale, the impact on the prices range vary for properties having 01, 02, and 05 amenities, but stay nearly for properties having 03 and 04 amenities (though they prices values are different)

Security amenities' impact on the price



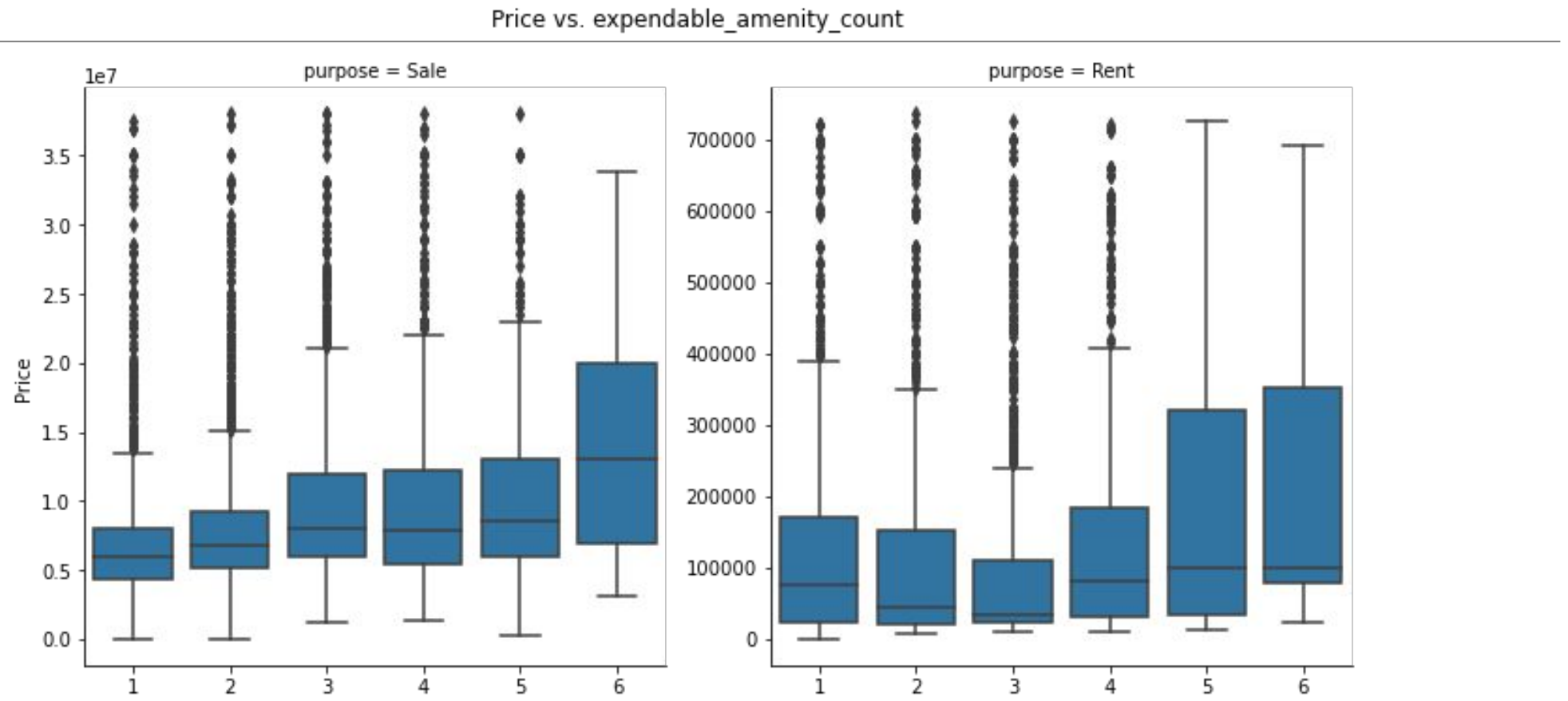
- Properties on sale tend to have 01 to 02 security amenities, and their prices ranges as well as their mean prices are nearly the same.
- However for properties on rent, there can be up to 04 security amenities. The mean prices as well as prices range for properties with 01 or 02 security amenities are basically the same. The mean prices as well as prices range for properties with 03 or 04 security amenities are basically the same.

Social amenities' impact on the price



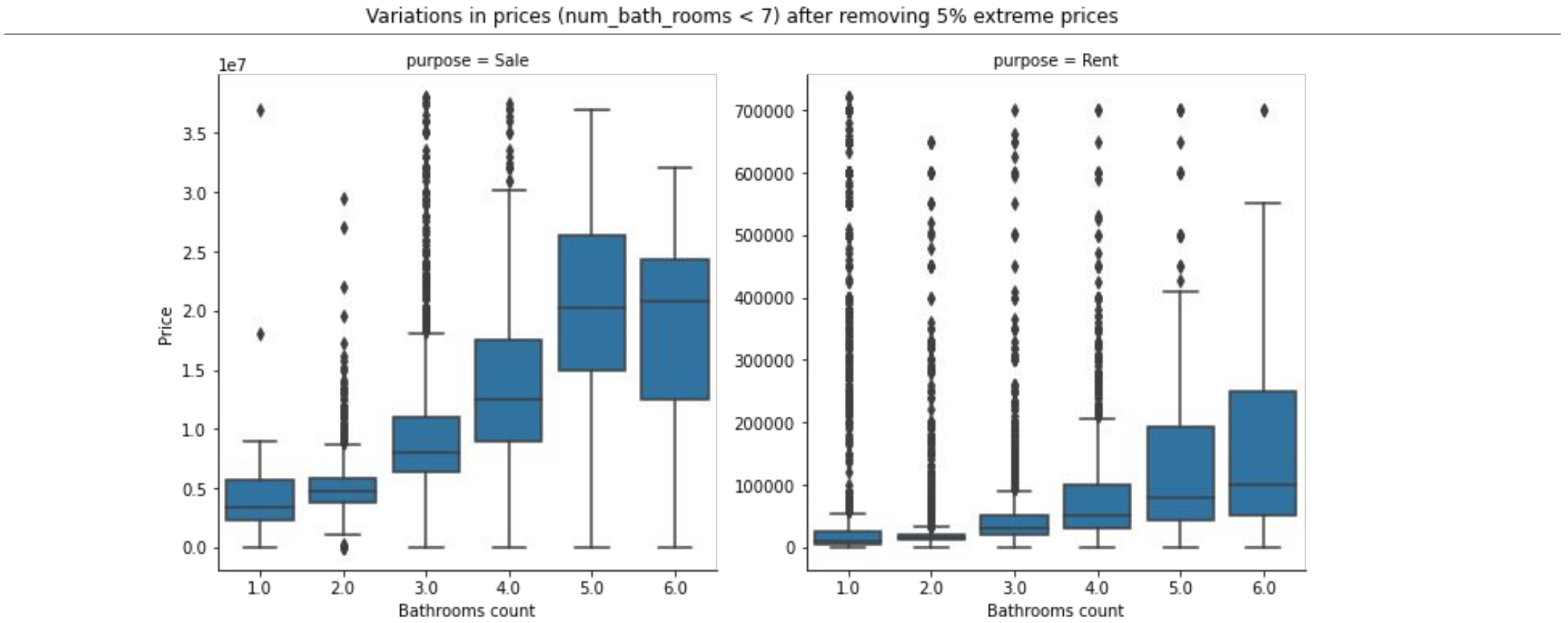
- For properties on sale, the mean prices tend to decrease from 01 to 03 social social amenities, which is astonishing since we are expecting and increase in the prices. The expected increase only appear from 04 to 06 social amenities.
- Properties on rent see their mean prices increase the most social amenities they have. The exception is the properties with 05 social amenities; looking at the prices range of its boxplot, we are expecting this to be an outlier

Expendable amenities' impact on the price



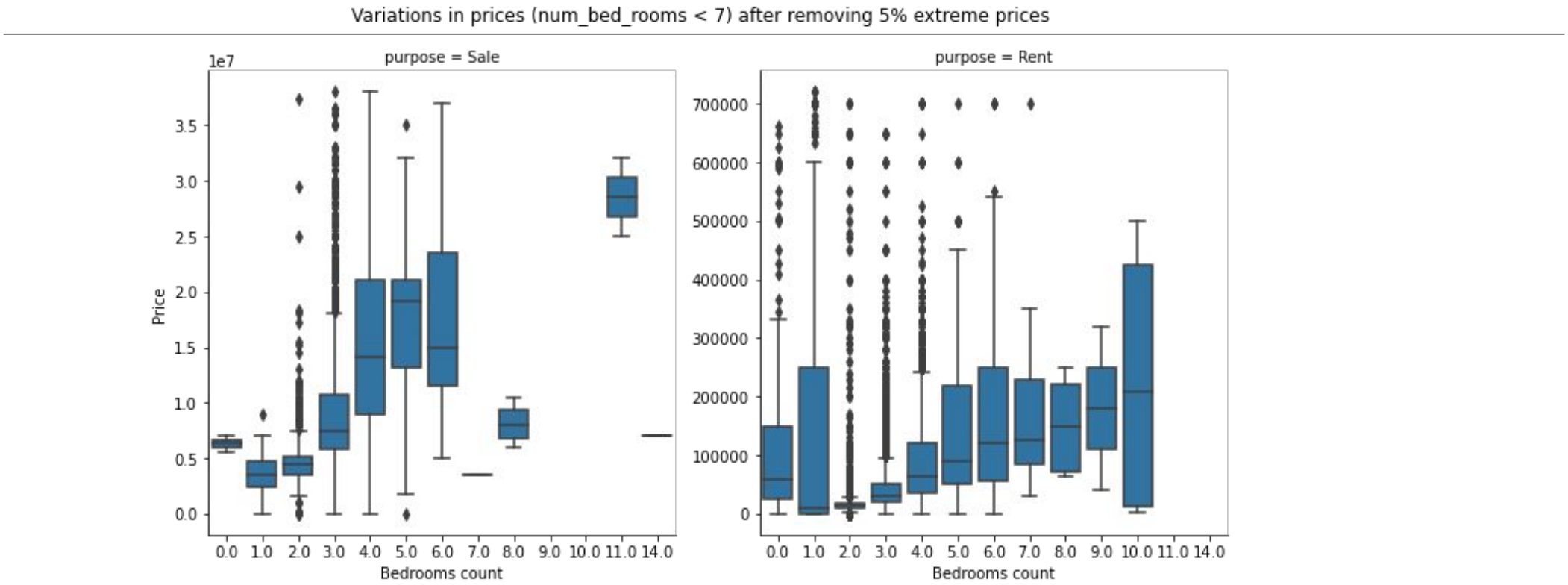
- The mean price of the properties on sale are increasing with the number of expendable amenities they have. The prices ranges are also increasing.
- For the properties on rent, their mean price as well as prices ranges are decreasing from 01 to 03 expendable amenities. The expected increase only appear from 04 to 06 expendable amenities.

Bathrooms' impact on the price



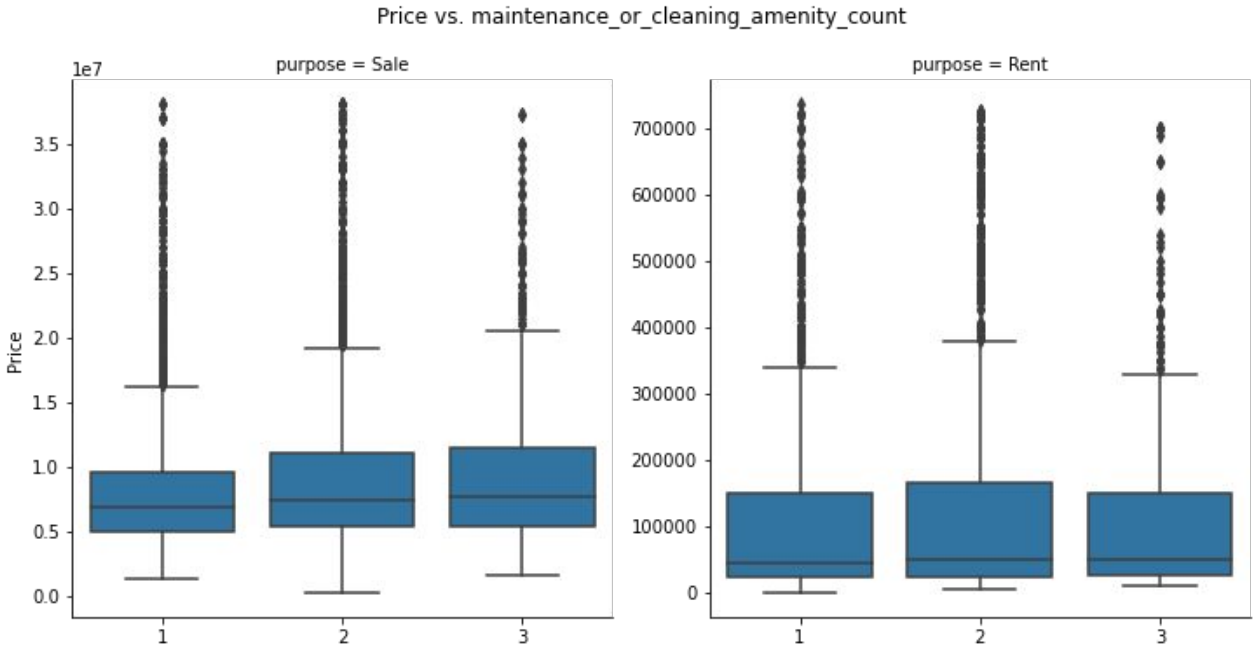
- Please note that the two charts are not on the same scale
- The number of bathrooms is an indicator of properties prices, either for sale or rent
- Properties with 01 and 02 bathrooms tend to have sale prices not far from each other, though their prices ranges are different
- Properties with 05 and 06 bathrooms are sold nearly at the same prices, though their prices ranges are different. Surprisingly, properties with 05 bathrooms have a range of prices superior to the range of prices for 06 bathrooms' properties

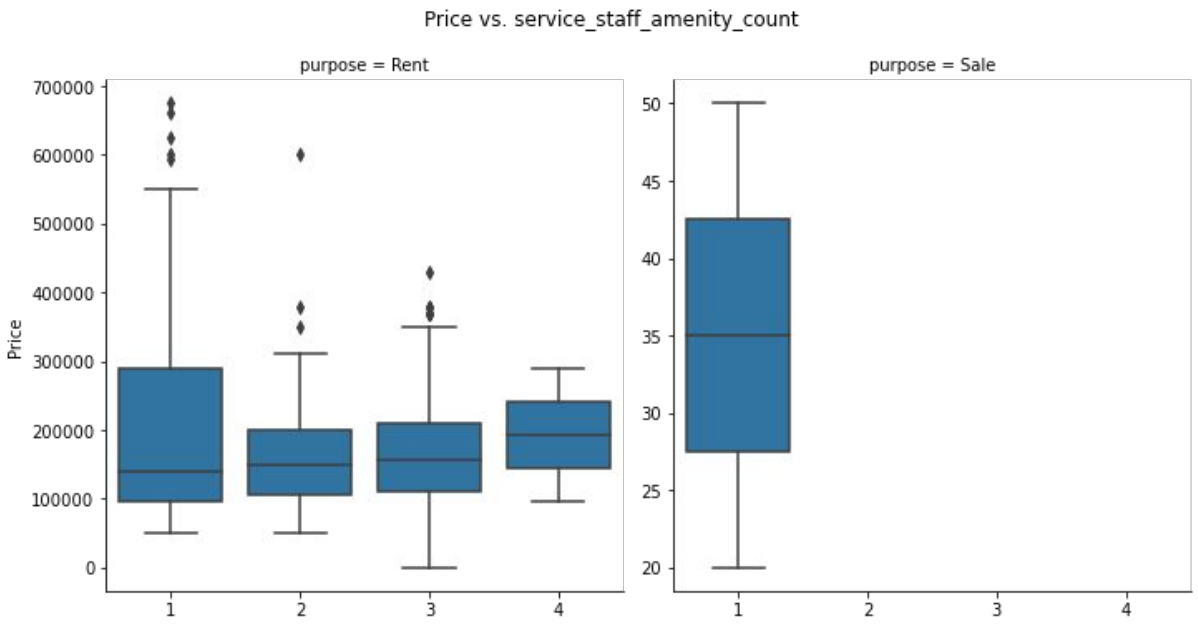
Bedrooms' impact on the price



- Please note that the two charts are not on the same scale
- The number of bedrooms is also an indicator of properties prices, either for sale or rent
- Properties with 00 bedrooms should be Commercial. We can see that properties on sale with 00 bedroom are on a very small range of price, which mean that it should be easy for Commercial properties on rent to be quoted.
- Generally, the most bedrooms a (residential) proprety has, the higher its rent/sale price; that make sens since it should mean more area available in the property. The exception is the properties on sale with 06 bedrooms, which have a fall in mean price, though they have a chance to be quoted higher (their prices range is bigger).

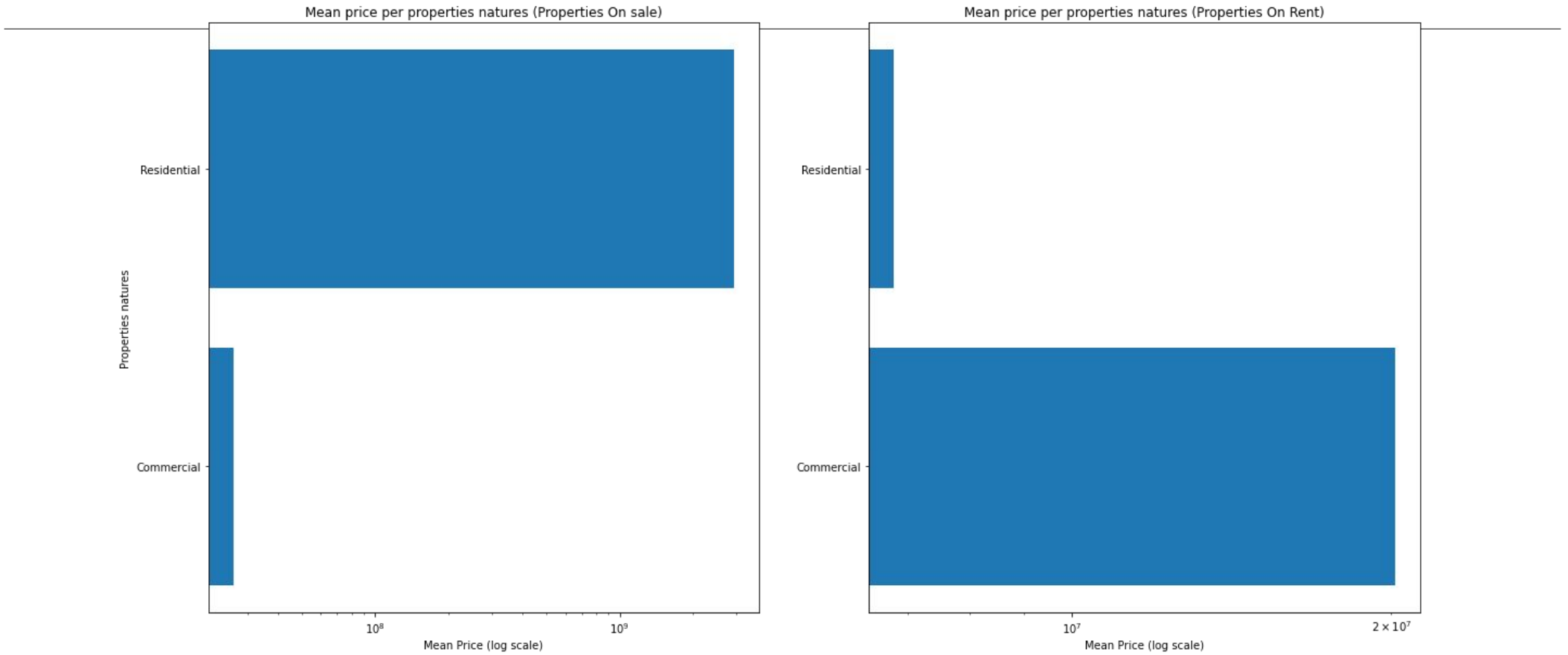
Amenities with little impact on the price





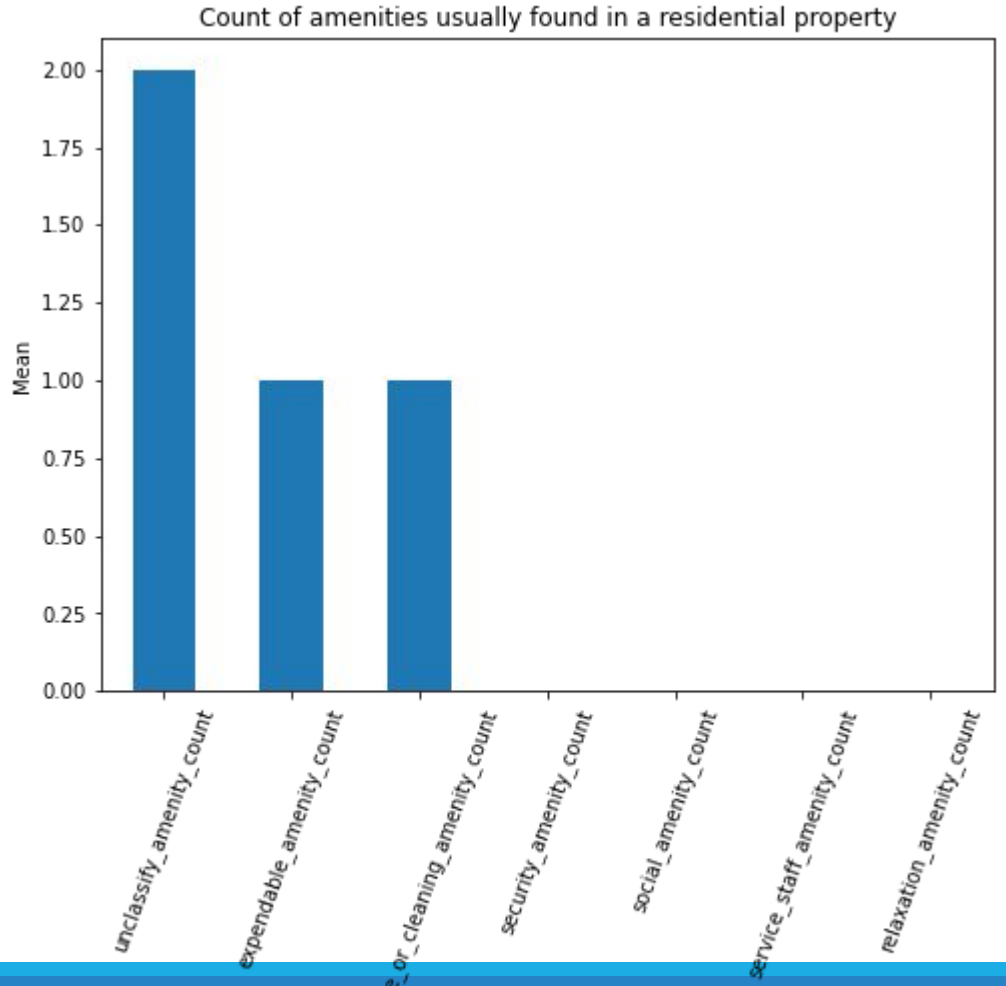
- **Maintenance or cleaning amenities, as well as service staff amenities don't make the price change that much**
- **We can deduce that they are expected/standard goods in a property, or that buyer/renter doesn't care too much about their presence when getting a property**

Prices variation depending on property nature



- **Residential properties on sale are much more expensive compared to Commercial properties on sale**
- **Rented Commercial properties tend to cost much more than rented Residential properties**

Usual amenities in residential properties



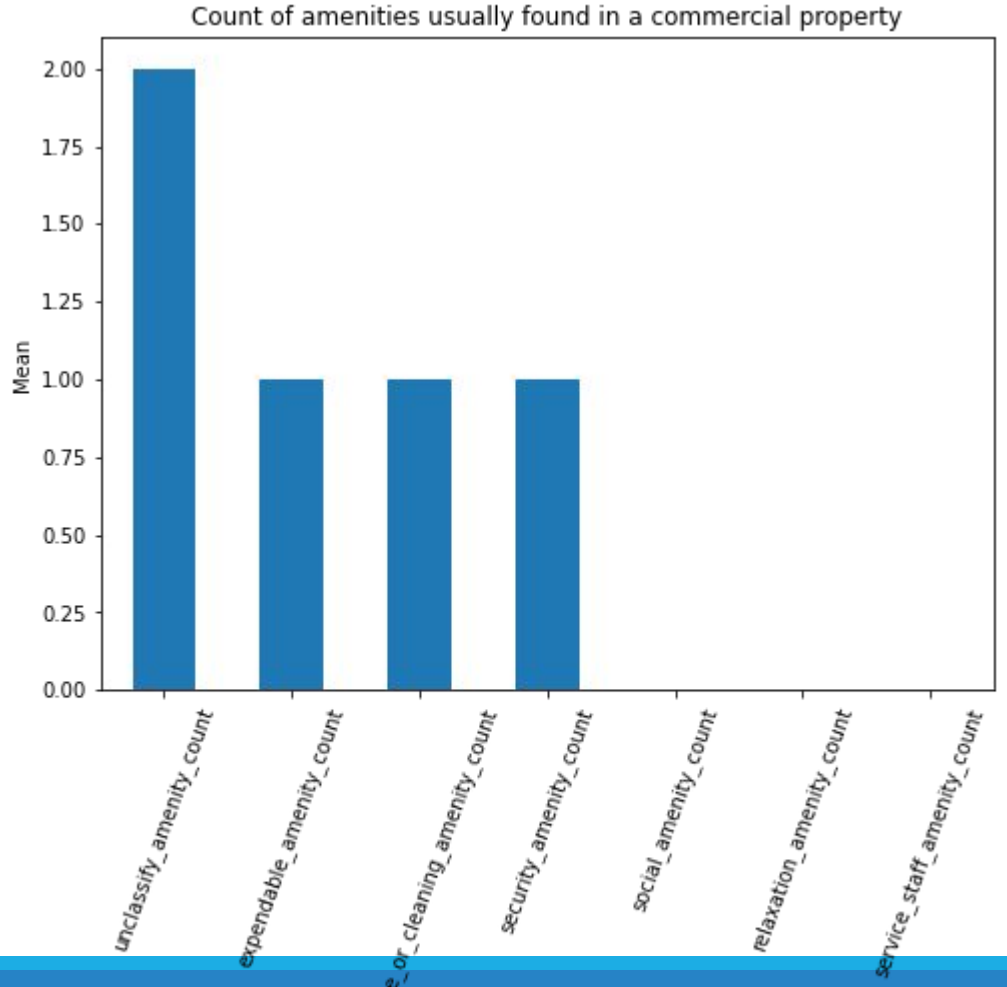
In residential properties, we usually found 01 expendable amenity and 01 maintenance/cleaning amenity.

As for other type of amenities, they are rarely found.

Usual amenities in residential properties
[What do Industry Experts say]

- Parking, security guard, lift, generator, intercom, check-in counter, CCTV, building manager, open roof, praying area, community hall/room, and gardening area.
- Very few have those: playground, swimming pool, automation in the building, room

Usual amenities in commercial properties



In commercial properties, we usually found 01 expendable amenity, 01 maintenance/cleaning amenity, and 01 security amenity.

Other types of amenities are rarely found.

Usual amenities in commercial properties
[What do Industry Experts say]

- Parking, security guard, lift, generator, check-in counter, CCTV, building manager, praying area
- Very few have those: automation in the building, room

Properties density’ and count’ impact on prices
[What does Industry Experts say]

- High price areas have more property density and govt. regulation plays a factor here too
- Does the density of available properties influence the price?
  - Yes. If the customers have options they tend to negotiate more on the price. Less option less negotiation
-

Amenities impact on prices
[What do Industry Experts say]

- Having more amenities means this is a premium building and a premium building charges more. If you follow the property companies, the more amenities they include in the building the higher they charge.
- How do they influence the price ?
  - Stepwise increase in relation to the amenities it offers but it will follow linearly price increase when we consider price hikes or inflation.
- How do property prices fluctuate for same / similar amenities?
  - will be same.. For example two flat on the same floor and same side will have same price.
- Which amenities have the more impact on price ?
  - The common amenities are a must every building has them. The price changes with the rare amenities or if the common amenities are present in a higher number.

Amenities impact on prices
[What do Industry Experts say] (cont)

- Which amenities have the less impact on price ?
  - swimming pool, gymnasium, or rooftop garden those are only present in premium buildings doesn't effect the mass market
- Impact of the basic facilities (e.g. num_bath_room, num_bed_room) on the property prices
  - It has a higher effect on price.
  - The size and number of bedrooms and bathrooms, the amount of comfort it provides, the openness, and how bright the room is, access to fresh air plays a significant role here.
-

Other inputs from Industry Experts

- Relationship between government regulations and sale prices
  - Property registration costs, Zoning and building codes, interest rate can affect sales prices.

- Impact of facilities (malls, educational institutions, business hubs, hospitals, ...) in the vicinity on the prices
  - Usually, people are willing to pay more to have those facilities near them. The better the institute is, the more effect it will have on price.

- How do property prices fluctuate for same property size?
  - If all the other factors are exactly the same almost no fluctuation.

Other inputs from Industry Experts (cont)

- How near should a property be to influence the price of another property ?
  - It will vary from area to area.
  - In some areas, it has to be close to having an impact but in some areas, the influence can span maybe kilometers depending on the securities and other unique traits the area has to offer.
  - Ultimately it is important to consider the local market factors.

- How do nearby properties prices influence a given property prices ?
  - Unless the other property has any significant factor there will be significant influence.
  - The more factors there are and the more important those factors are to the buyer, the more influence there will be.

Other inputs from Industry Experts (cont 2)

- How has the real estate market in Bangladesh evolved over time? What are the main drivers of this demand?
  - GDP growth, urbanization, population growth, raising income, and economic growth have evolved this sector significantly.
  - People having high-income cap now wants to invest because they have more free money and upgrade their lifestyle. Having a home gives them a sense of stability and high-status sense.

- How could the mean/median income of an area relate to prices in the area.
  - The higher the income cap the bigger and more luxury they want in their home hence the high price.
  - After a certain point, price matters little while service and relationship matters most.

Conclusion

Recommendations for Task 04

- The data is skewed towards “Apartments” (property’ type) and certain localities (and zones)
- We could only consider properties in Dhaka and Chattogram divisions/cities when building our future models, since they hold 90% of our data
- Properties type (apartment, shop, floor,...) column should be kept while Properties nature (residential, commercial) column could dropped, since the first one allow to infer the second one
- The price increase is non-linear: it increases exponentially with area, number of bathrooms, and bedroom
- Maintenance or cleaning amenities, as well as Service staff amenities surprisingly don’t influence the price of properties

Conclusion

Results

- Built an understanding of the real-estate market in Bangladesh
- Delivered key insights for [Task 04 - Model development]

Task 04 - Model Development

Problem Statement

- *Develop a model that will propose price for a given property*
- *Develop a model that will recommend a property to a given user, based on the search parameters he will input*

:

Work Pipeline (Fair Price model)

-

Data from Task 02

Step 1

Making sure data from Task 02 is correctly cleaned

Step 2

Remove outliers

Step 3

Choosing modeling method (Machine Learning / Deep Learning / both)

Step 4

Train, test, and compare different models according to Step 3, to choose the best one

Step 5

Provide the best trained model

Work Pipeline (property recommender)

-

Data from Task 02

Step 1

Making sure data from Task 02 is correctly cleaned

Step 2

Remove outliers

Step 3

Choosing modeling method (Machine Learning / Deep Learning / both)

Step 4

Train, test, and compare different models according to Step 3, to choose the best one

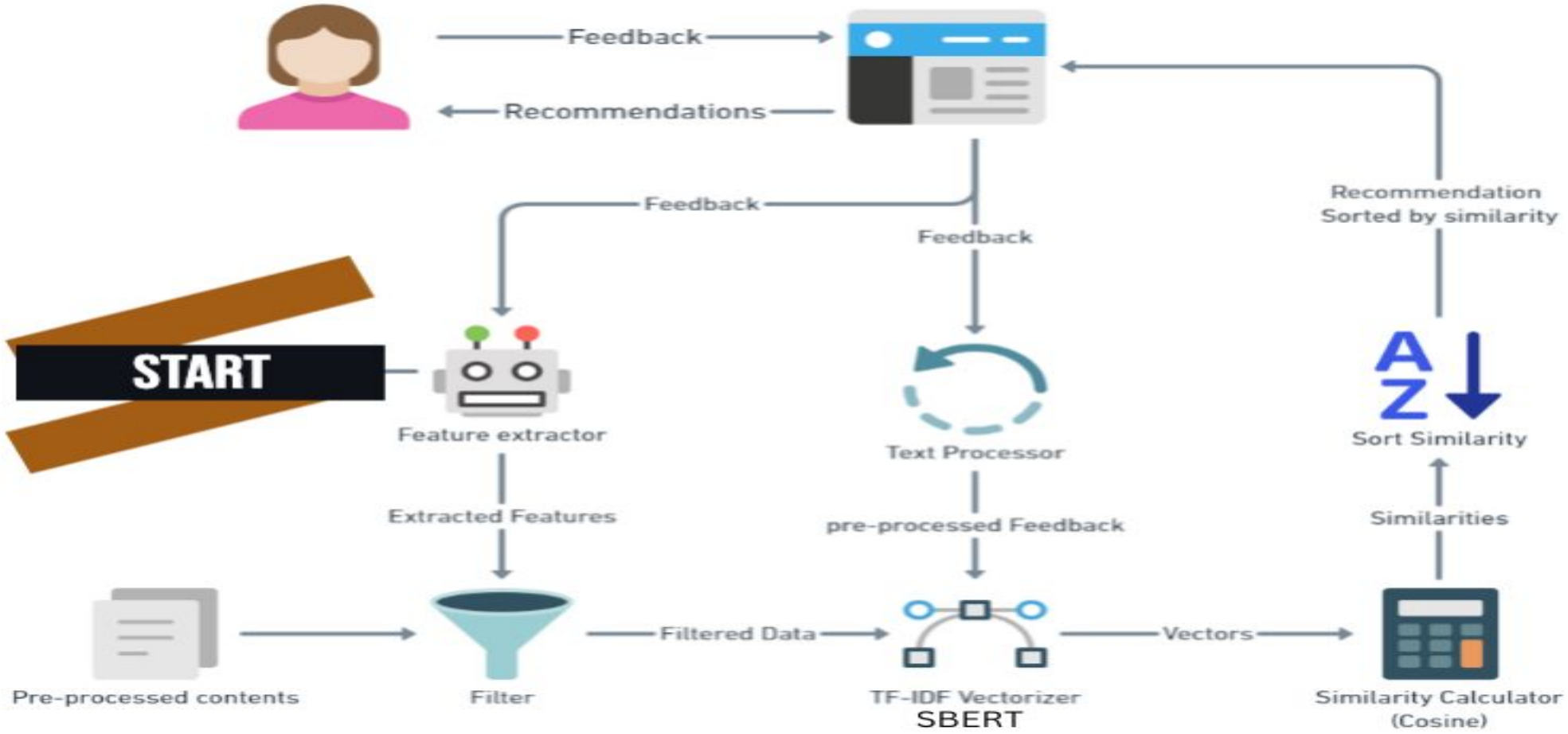
Step 5

Different trained models should be provided in order to select the best through A/B tests

Regression Model

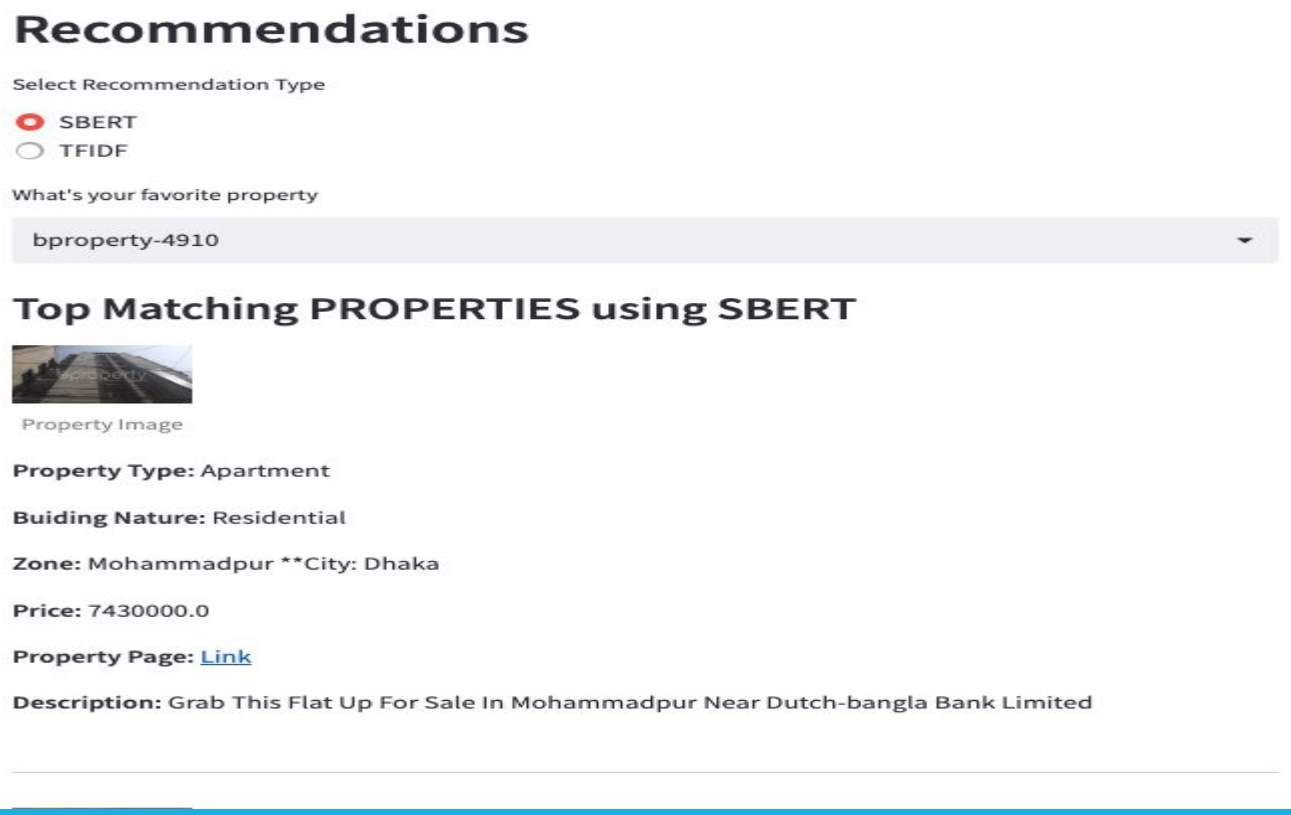
- *Developed price predicting model using Scikit Learn*

Recommendation Model Flow Diagram



Recommendation Model

- *Developed recommendation model using TD-IDF and Hagging Face framework.*



Task 05 - Model Deployment

Problem Statement

- *Develop a model that will propose price for a given property*
- *Develop a model that will recommend a property to a given user, based on the search parameters he will input*

:

Work Pipeline (Model Deployment)

-

Data from Task 02

Step 1

Making sure data from Task 02 is correctly cleaned

Step 2

Remove outliers

Step 3

Choosing modeling method (Machine Learning / Deep Learning / both)

Step 4

Train, test, and compare different models according to Step 3, to choose the best one

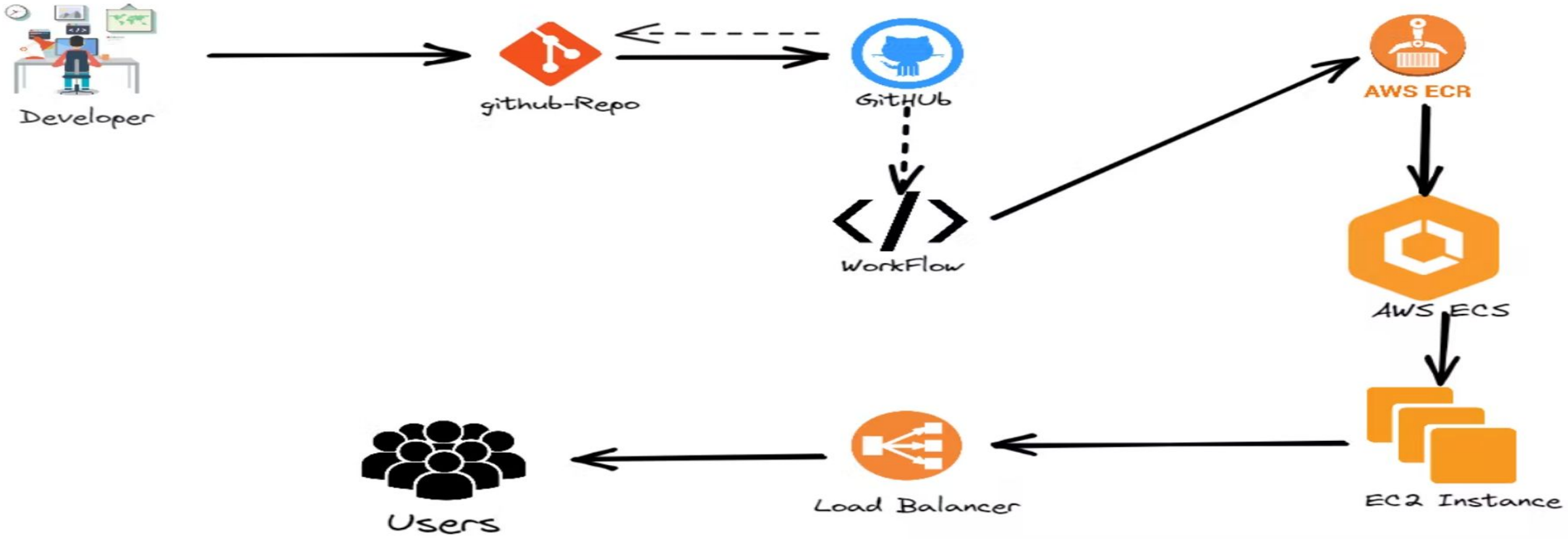
Step 5

Provide the best trained model

Model Deployment - Task 5

- *Developed the Streamlit application and converted into Docker image.*
- *Deployed the Streamlit app in the AWS ec2 instance and application is dockerized.*
- *Application is up and running.*

Model Deployment - Task 5



Work in Progress - Next Steps

- *Define clear working plan : which technologies to choose ? should there be a session to explain recommender engines creation to task members ? etc.*
- *Define working technologies for each of the two models: Deep Learning ? Machine Learning ? both ?*

What's Next after this Project closure ?

- *We will handover the WebApp to Omdena.*
- *Provide Certificates to the contributors after activity analysis.*
- *We will handover the project report to Omdena .*

IT’S SELFIE TIME! PLEASE TURN ON YOUR VIDEO !

Thank you

Follow Us:

LinkedIN : https://www.linkedin.com/company/omdena-bd-chapter/

Facebook : https://www.facebook.com/profile.php?id=100089549531493

Any Questions?