Presented by:

Romeo

Sarthak

Zahra

Feature Recognition as a consultant based on Airbnb data

Business Scope:

- Assuming ourselves as an Airbnb consultant, we are preparing the list of features in order to maximize the profit for hosts. Also, we are preparing models to advise them the type of houses in which customers are more interested. If any hosts possess more than one houses listed on Airbnb, we're expecting to give them insights on which house generates more profit for them.

Outline

I. About our dataset

II. Data cleaning

III. Advice on how to benefit from your Airbnb experience as a host

   a) 1- Guest Preference and seasonality
      2 - Price and seasonality
   b) Price and relationship with other variables in our dataset

IV. Conclusion

I. About our dataset

- Dataset that is used in this project is 'The Berlin Airbnb data'.
- Despite 6 tables in the dataset, we've used 3 tables as per the information.

  We're trying to deduce the tables as follows:

  - Calendar_summary :(8231480, 9)
  - Listings: (22552, 16)
  - Listings_summary: (22552, 96)

I. About our dataset

	listing_id	date	available	price
0	2015	2019-09-10	f	NaN
1	2015	2019-09-09	f	NaN
2	2015	2019-09-08	f	NaN
3	2015	2019-09-07	f	NaN
4	2015	2019-09-06	f	NaN

Calendar_summary table

Deeper insights in Data

I. About our dataset

	id	listing_url	scrape_id	last_scraped	name	summary	space	description	experiences_offered	neighborl
0	2015	https://www.airbnb.com/rooms/2015	20181107122246	2018-11-07	Berlin-Mitte Value! Quiet courtyard/very central	Great location! 30 of 75 sq meters. This wood...	A+++ location! This „Einliegerwohnung“ is an e...	Great location! 30 of 75 sq meters. This wood...	none	It is locat Eas
1	2695	https://www.airbnb.com/rooms/2695	20181107122246	2018-11-07	Prenzlauer Berg close to Mauerpark	NaN	In the summertime we are spending most of our ...	In the summertime we are spending most of our ...	none	
2	3176	https://www.airbnb.com/rooms/3176	20181107122246	2018-11-07	Fabulous Flat in great Location	This beautiful first floor apartment is situa...	1st floor (68m2) apartment on Kollwitzplatz/ P...	This beautiful first floor apartment is situa...	none	The n famous fc
3	3309	https://www.airbnb.com/rooms/3309	20181107122246	2018-11-07	BerlinSpot Schöneberg near KaDeWe	First of all: I prefer short-notice bookings. ...	Your room is really big and has 26 sqm, is ver...	First of all: I prefer short-notice bookings. ...	none	My flat is Wes
4	7071	https://www.airbnb.com/rooms/7071	20181107122246	2018-11-07	BrightRoom with sunny greenview!	Cozy and large room in the beautiful district ...	The BrightRoom is an approx. 20 sqm (215ft²), ...	Cozy and large room in the beautiful district ...	none	Great ne plenty o

5 rows × 96 columns

Listings_summary table

- Includes different features of house along with price

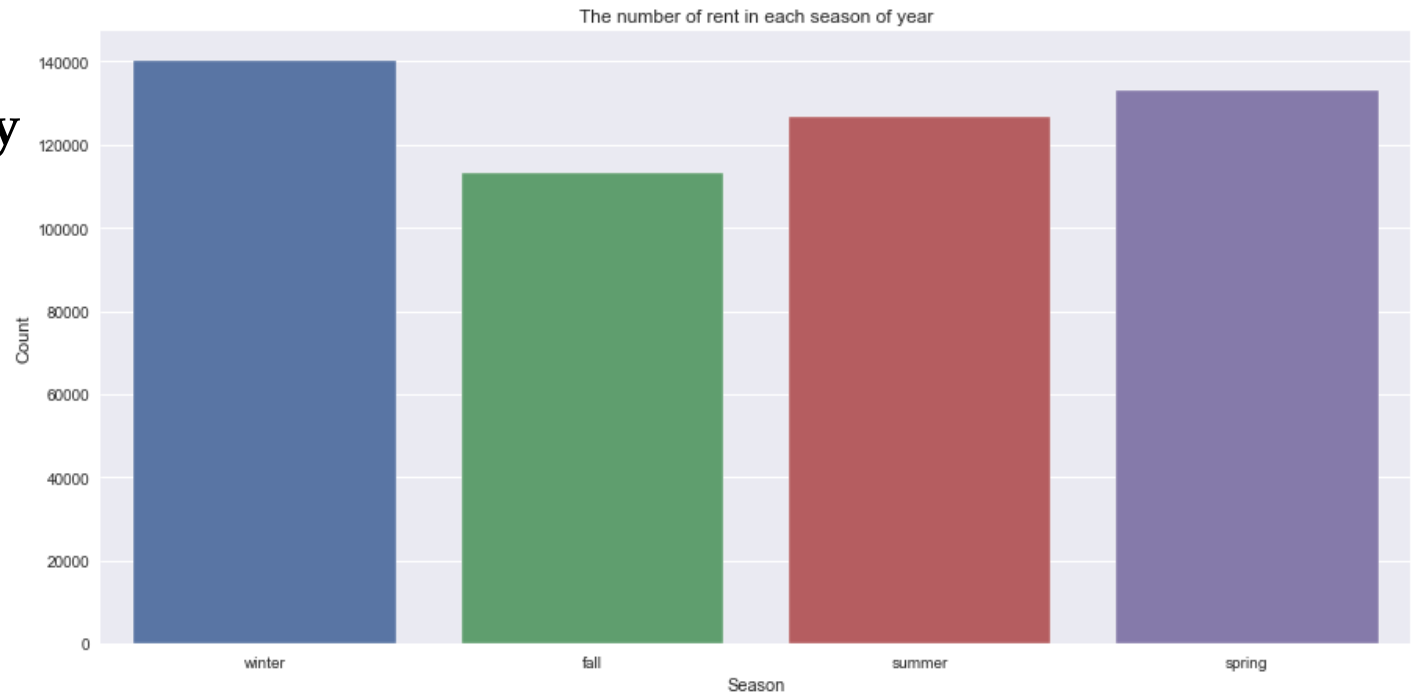
II. Data Cleaning

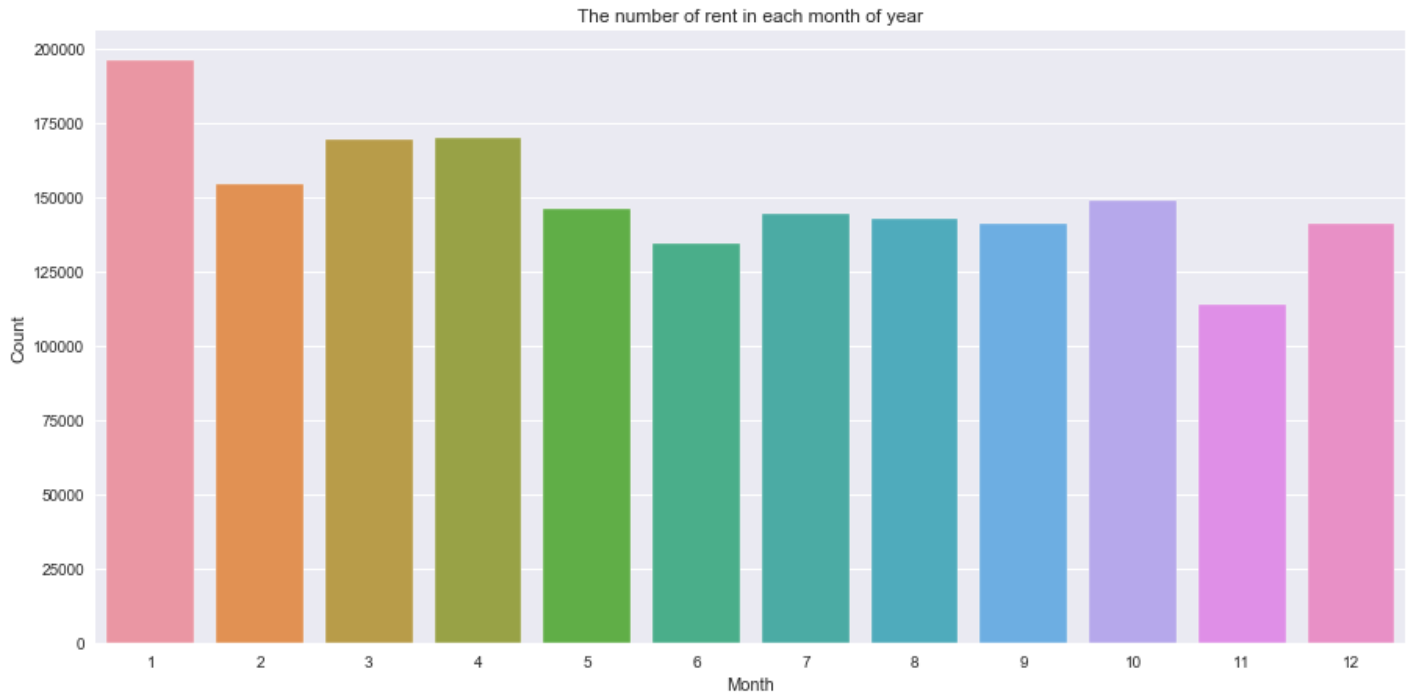
- We chose following as our main tables:
  - The Calendar_summary and
  - Listings_summary
- We reduce the columns in Listings_summary to what we needed in our analysis
- We've selected the price lower than 300 based on outliers detection analysis.

III a) 1 - Guest Preference and seasonality

- By checking seasonality, we found that the **fall has the lowest interest for guests**

- In accordance to month, January is the most preferred month by Airbnb customers.

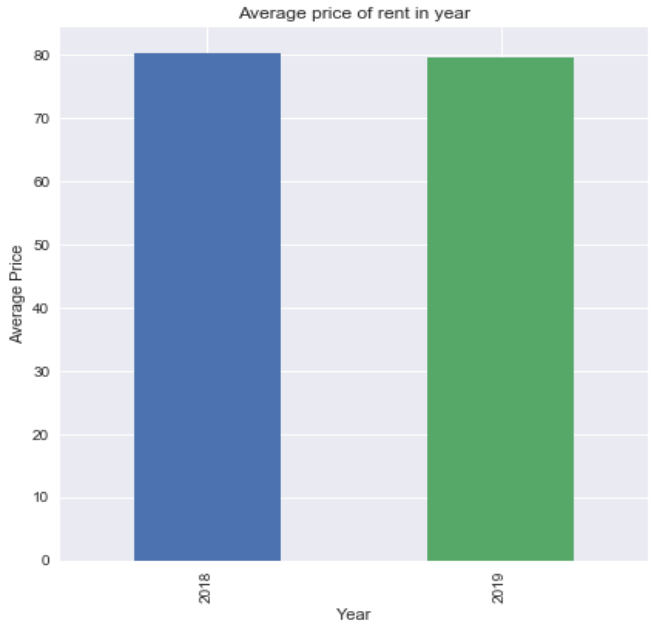


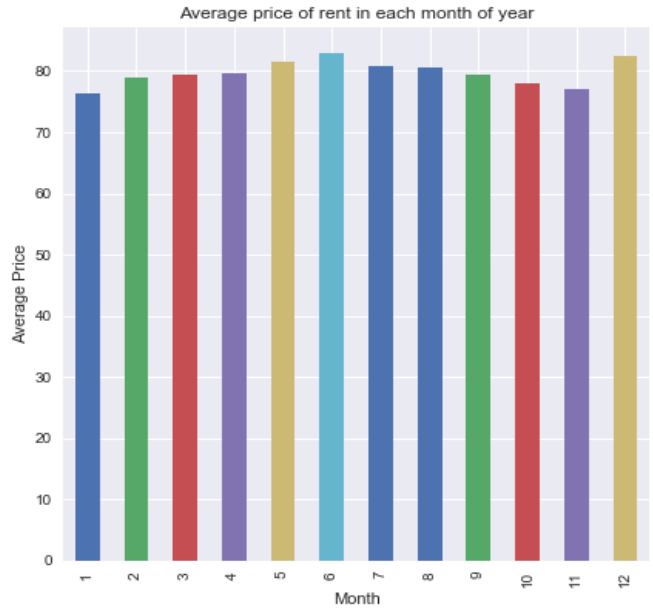


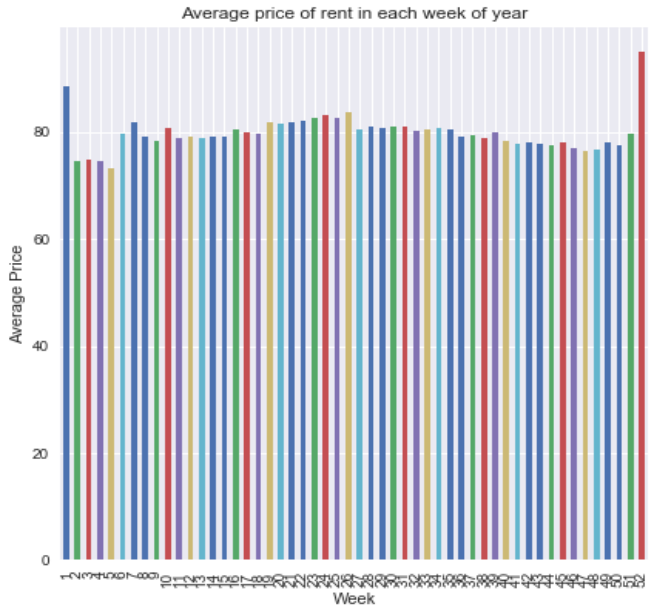
III a) 2- Price and seasonality

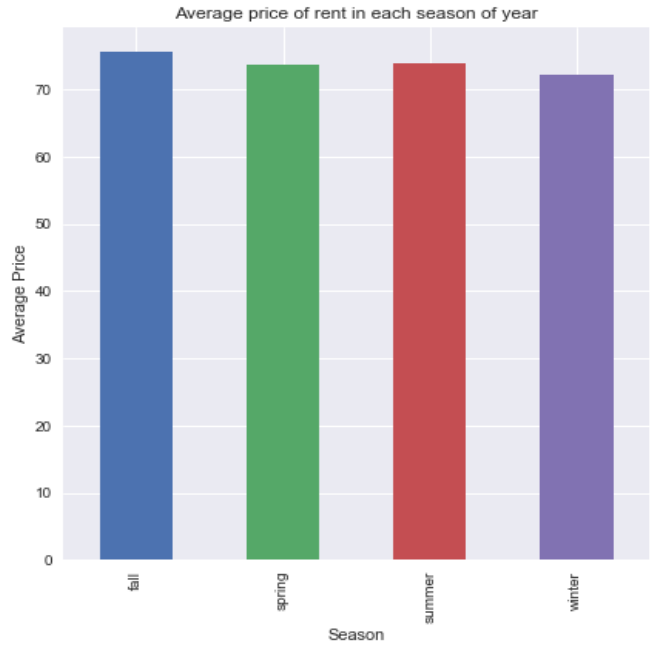
Seasonality and behavior of data during different time of the year, month, week based on average price. all season.

- It shows December and July have the highest average price in months.
- Fall has the highest average price in all season.









Gold List:

1. December and July in months
2. fall and summer as season

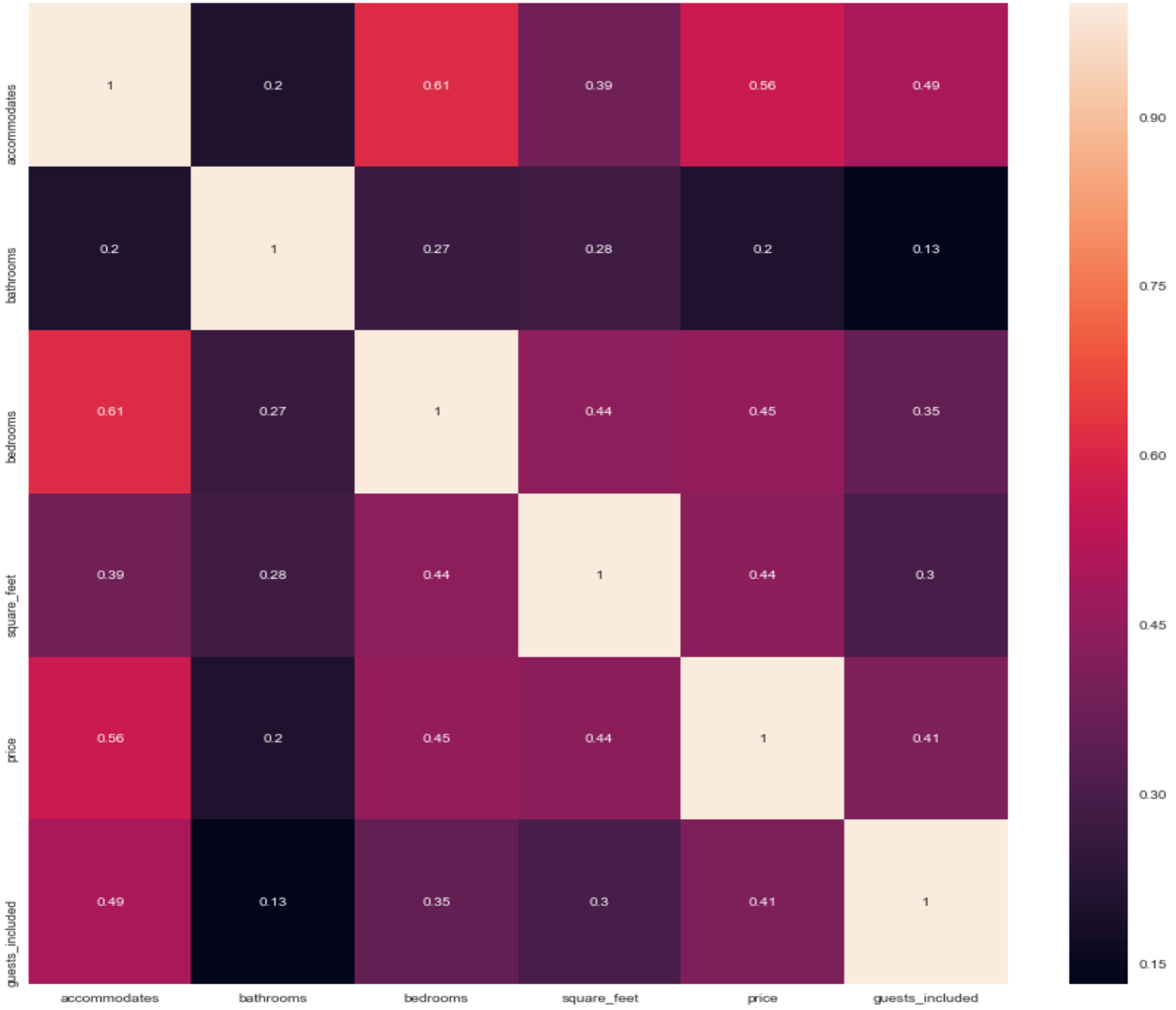
III b) Price and relationship with other variables in our dataset

Gold List:

1. Seasonality
2. Accommodates
3. Bedrooms
4. Area (Square feet)
5. Guest included
6. Neighborhood

III a) Price and relationship with other variables in our dataset

- From this heatmap, the highest correlations to price are:
  - Accomodates
  - Bedrooms
  - Square feet
  - Guests included



III b) Price and relationship with other variables in our dataset

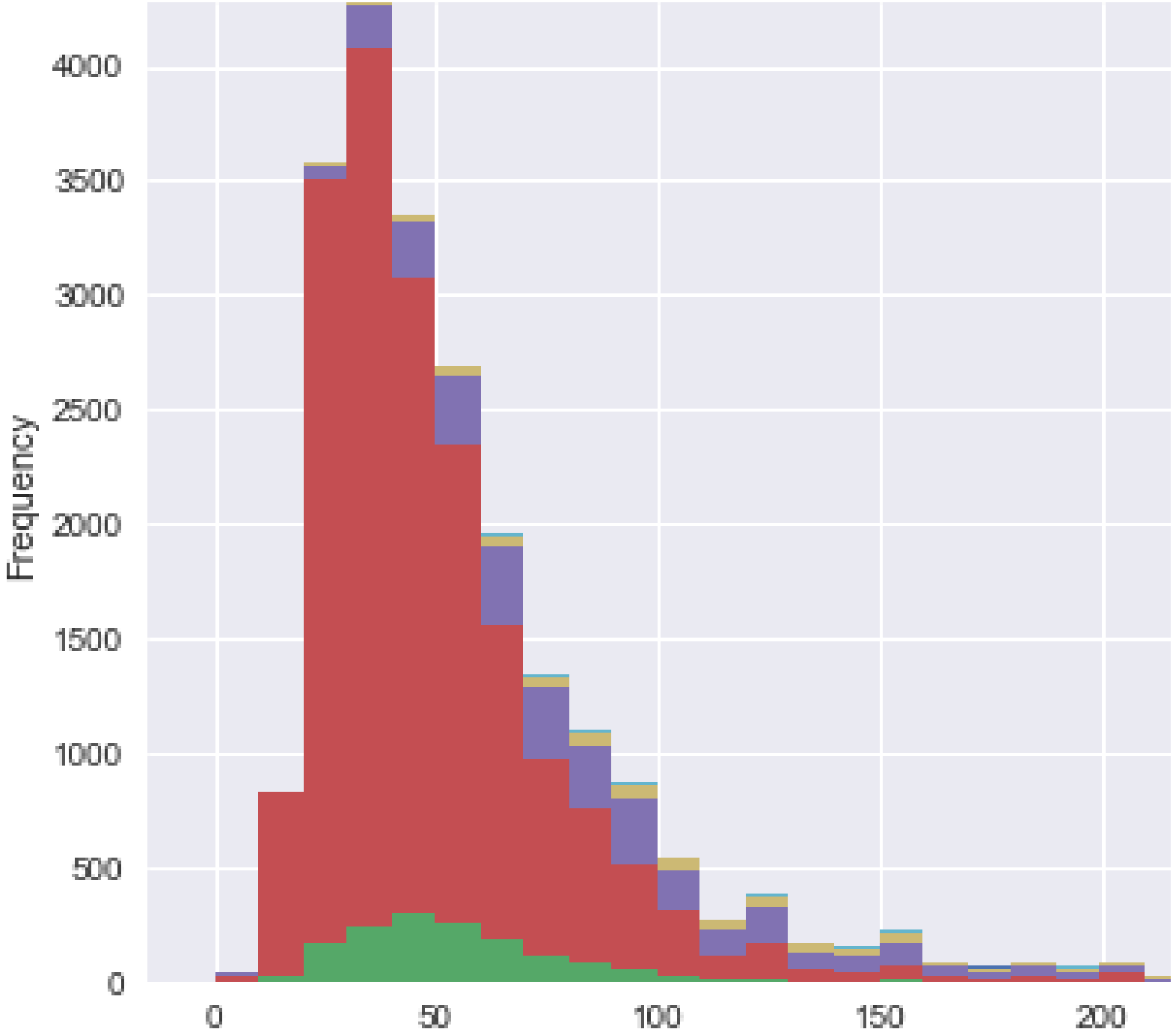
Gold List:

1. Accommodates
2. Bedrooms
3. Area (Square feet)
4. Guest included
5. Neighborhood
6. Seasonality

III **b) Price and relationship with other variables in our dataset**

Distribution of bedrooms

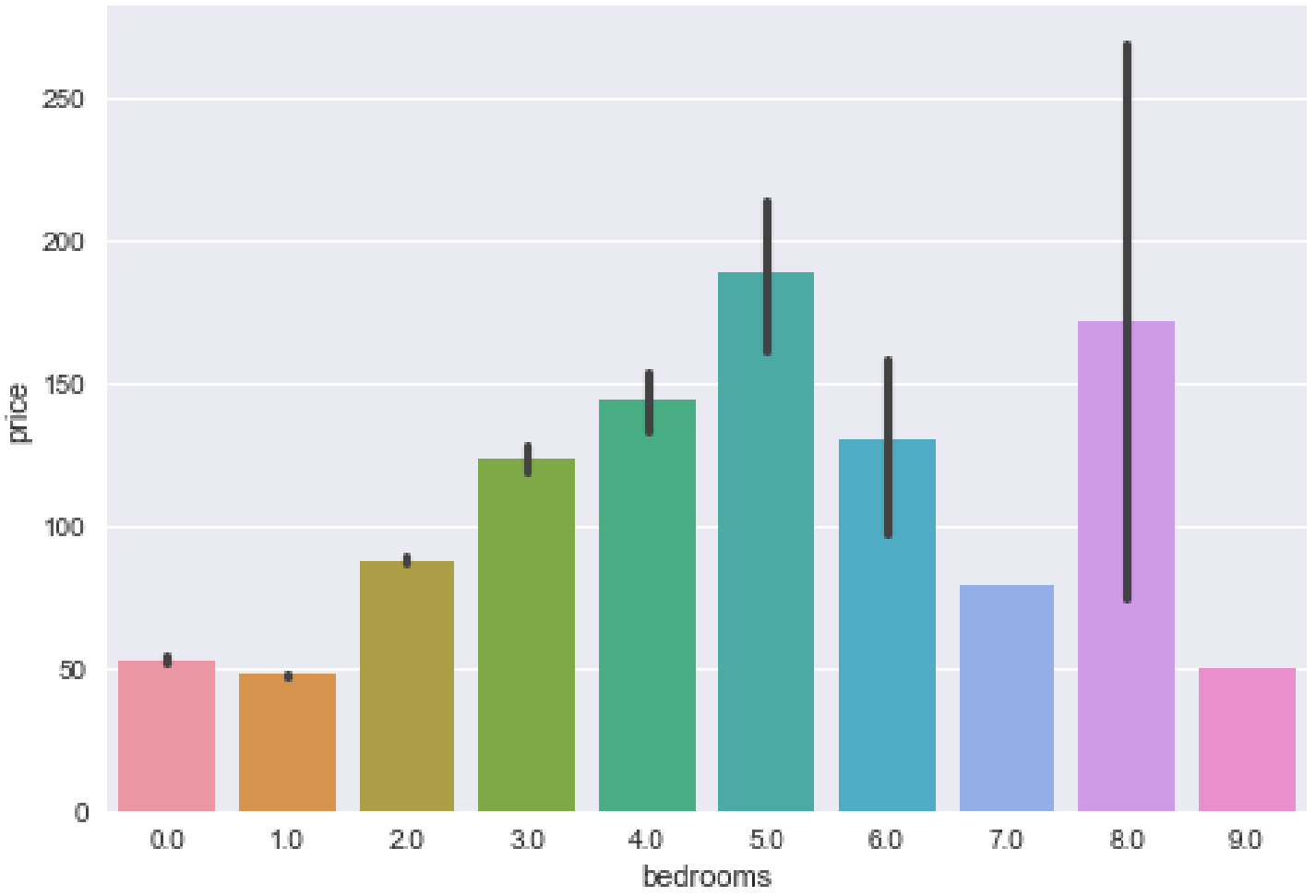
- It shows the most interesting and common rooms are with 0 and 1 bedrooms.



III b) Price and relationship with other variables in our dataset

Relation of price with the number of rooms

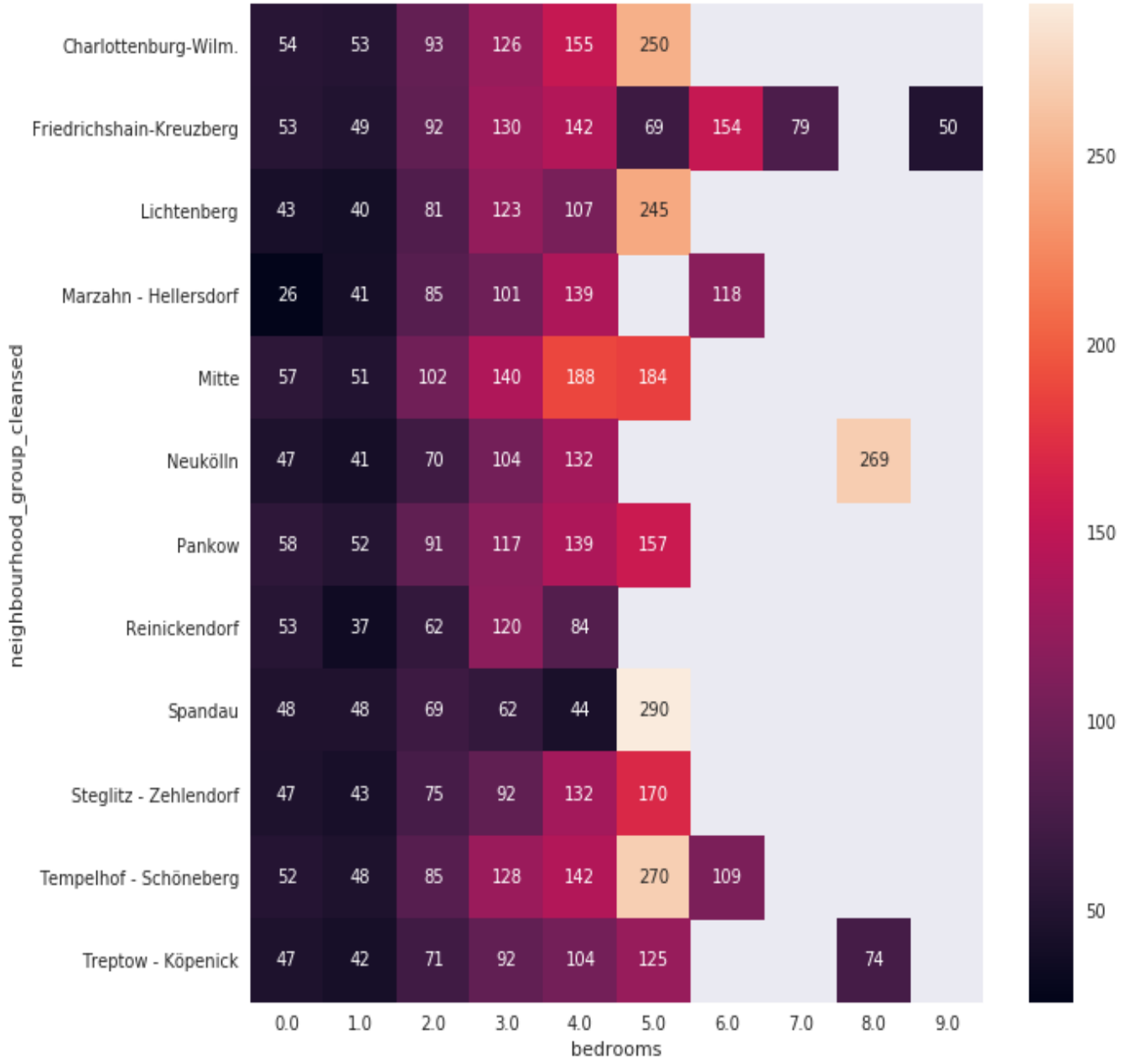
The number of rooms has direct effect on price till 5 bedrooms and has and after that we can see the fluctuation price for upper than 5.



III b) Price and relationship with other variables in our dataset

Average price by number of bedrooms in each neighborhood area

- We can see again the rooms with 5 bedrooms has the highest average price in this plot.



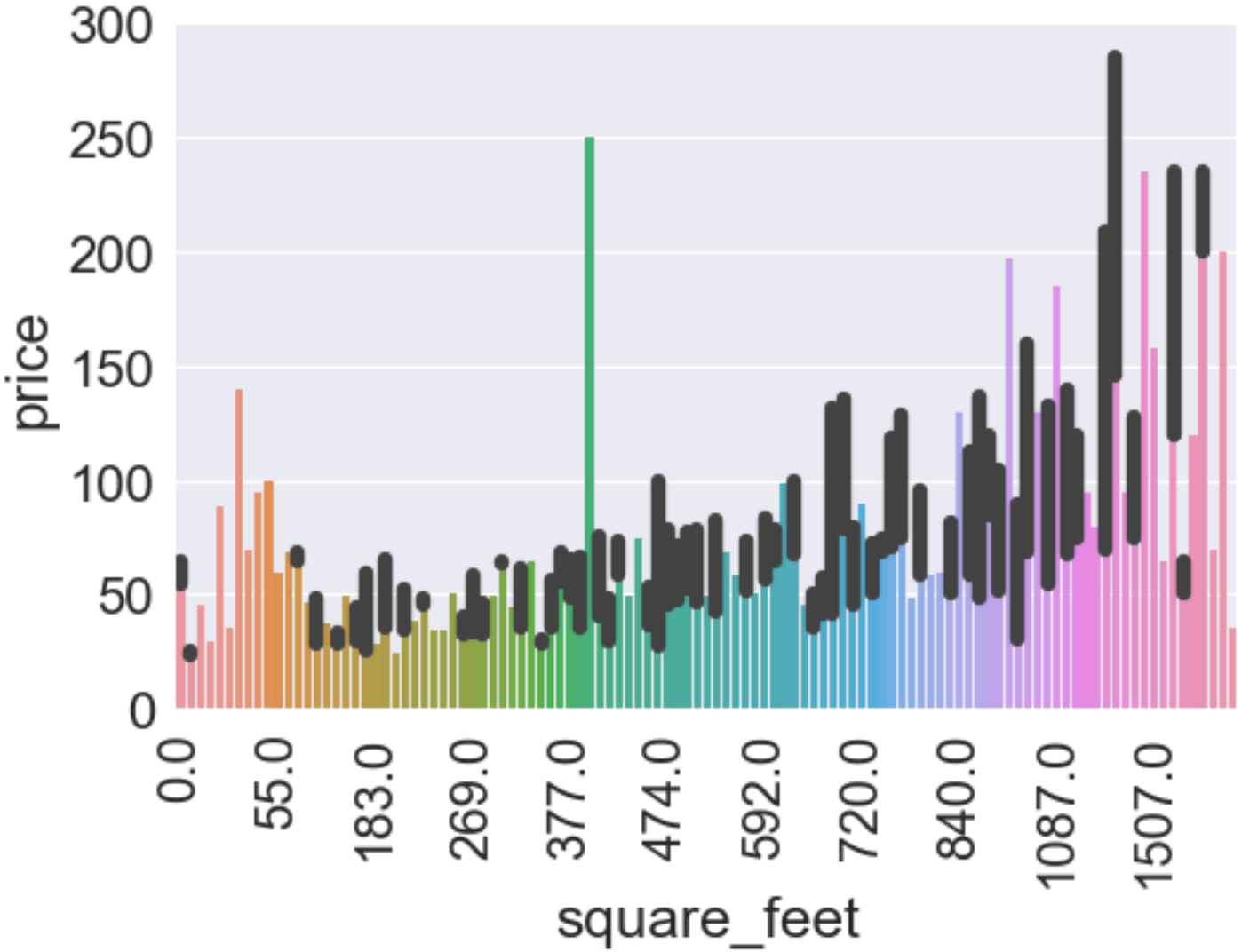
III b) Price and relationship with other variables in our dataset

Gold List:

1. Accommodates
2. Bedrooms
3. Area (Square feet)
4. Guest included
5. Neighborhood
6. Seasonality

III b) Price and relationship with other variables in our dataset

The plot support what heatmap claimed before and the price going up if square feet increase



III b) Price and relationship with other variables in our dataset

Gold List:

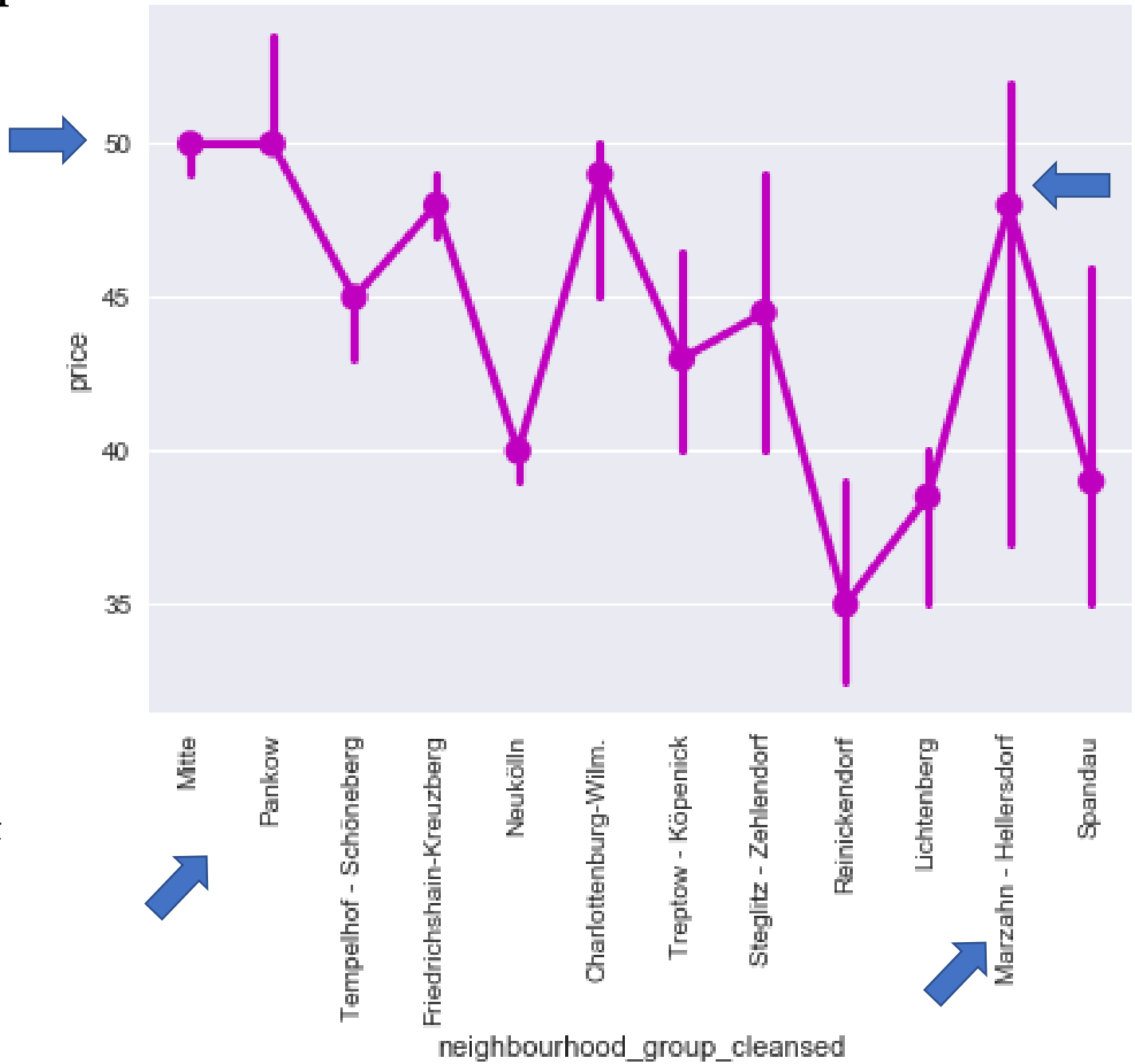
1. Accommodates
2. Bedrooms
3. Area (Square feet)
4. Guest included
5. Neighborhood
6. Seasonality

III **b) Price and relationship with other variables in our dataset**

Behaviour of price in Neighborhood

As we can see in the figure:

- Pankow and Mitte are more consistent neighborhood with relatively greater median than other neighborhoods.
- Marzahn-Hellersdorf also has greater price input, however, it also has a large fraction below the median price.



III **b) Price and relationship with other variables in our dataset**

Gold List:

1. Accommodates
2. Bedrooms
3. Area (Square feet)
4. Guest included
5. Neighborhood
6. Seasonality

More about factors Airbnb hosts should consider

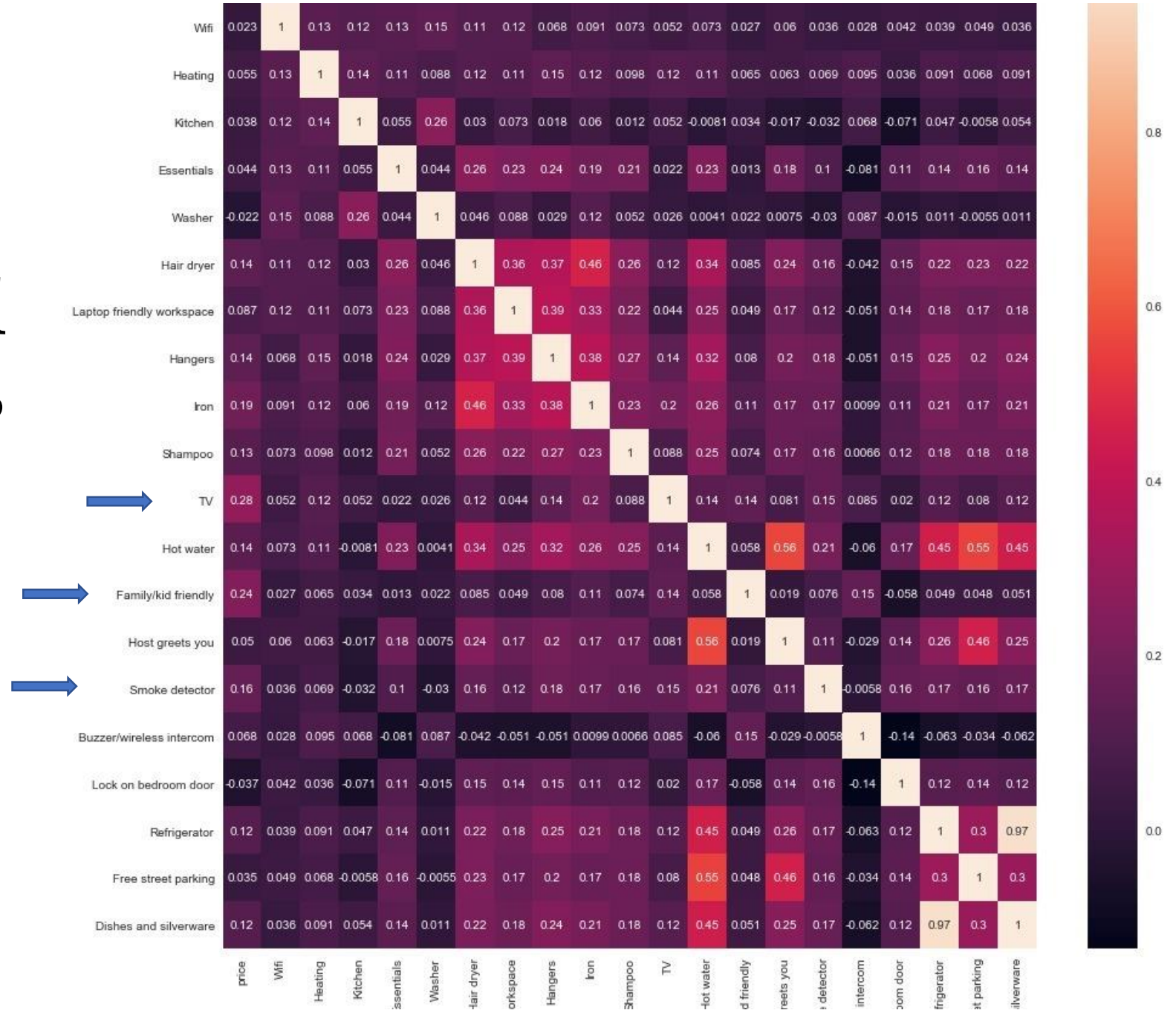
- We dwelled into details in the dataset and found out about the features that can be added in the house to make more profit.

- We are going to focus on 'Amenities' columns in Listings_summary table.

III **c) Lessons learned from our data**

Correlation between price and amenities features

- Most relevant:
- 1- TV
- 2- Family/kid friendly
- 3- Smoke detector
- … and so on



III **c) Lessons learned from our data**

The most relevant amenities with price:

1- TV
2- Family/kid friendly
3- Smoke detector
4-Iron
5- hair dryer
6- Hangers
7- Hot water
8- Shampoo
9- Refrigerator
10- Dishes and silverware

III **c) Lessons learned from our data**

Final Golden List:

Best time to rent your house:

1- December and July in months

2- fall and summer as season

House feature:

1- Accommodates
2- Bedrooms
3- Square feet
4- Guest included
5- Neighborhood

Amenities:

1- TV
2- Family/kid friendly
3- Smoke detector
4-Iron
5- hair dryer
6- Hangers
7- Hot water
8- Shampoo
9- Refrigerator
10- Dishes and silverware

Conclusion

- Key features for maximum profit are Accommodates, Bedrooms, Bathrooms, Area of the house and so on.
- Apart from these features, the host can focus on some amenities as TV, kid-friendly environment, Smoke detector, iron and so on to maximize profit.
- Keeping these features into account, hosts can turn his Airbnb business stand out among the crowd.

Questions?

Thank you!