

Hospitality & Tourism

- Dina
- Nourah
- Ammar
- Maan
- Salman

Contents

- Story of the dataset
- Exploratory Data Analysis (EDA)
- Visualization
- Results
- Challenges
- Conclusion
- Q&A





The story behind the dataset

Airbnb Rome Dataset

- Airbnb is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.
- Founded in 2008. It started as a way to make extra income by renting a spare room, but today its a multi billion dollar industry.



Airbnb Rome Dataset

- We choose Rome because it's one of the most visited cities by tourists in Europe. Thanks to the rich history, amazing food, and the relatively cheaper price compared to other major european cities.



Data Munging



Data Munging

- First we checked for duplicated and null values.
- Then we dropped columns of no use to us.
- We also dropped columns with too many null values,
- Cleaning price, from `\$` sign, and the commas separator.

```
df2["price"] = df2["price"].apply(lambda x : float(x[1:].replace(",",""))))
```

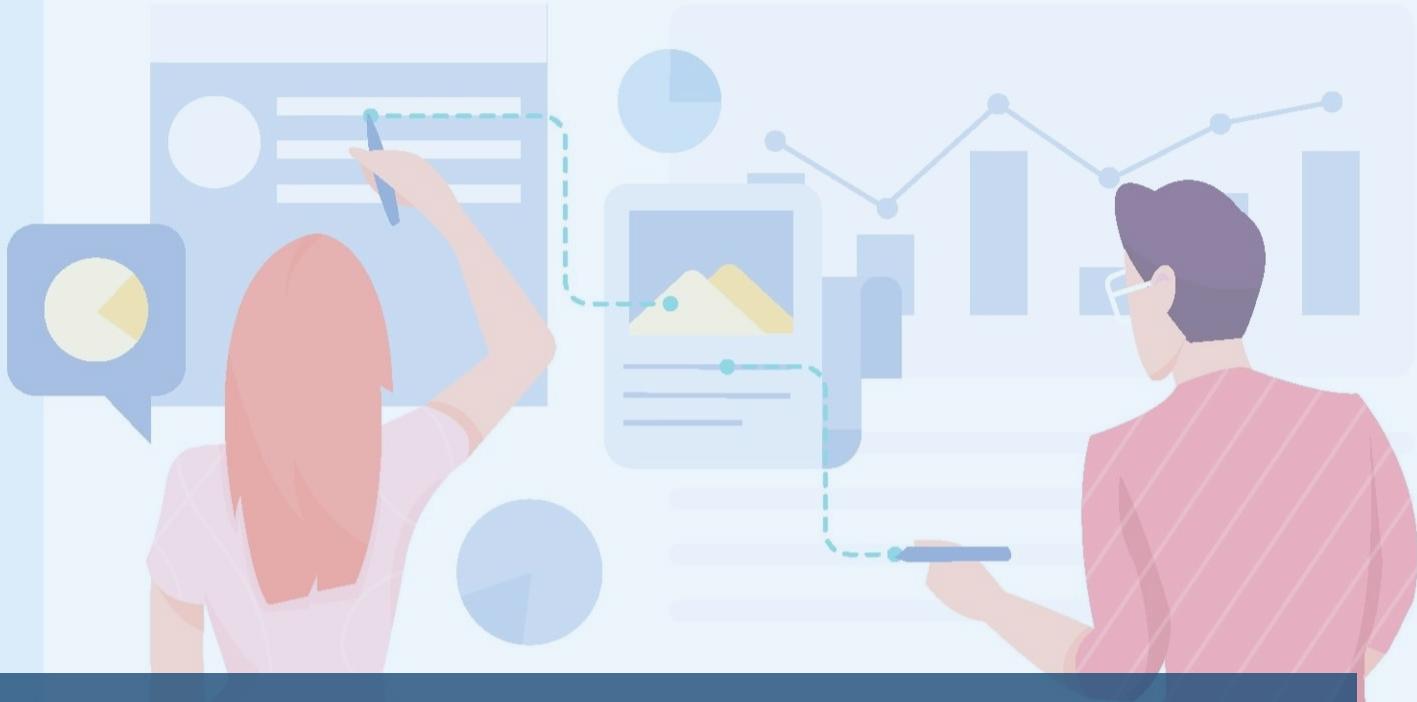
```
df.duplicated().sum()  
0  
  
df.isnull().sum()
```



Data Dictionary

Field	Description	Field	Description
host_name	Name of the host	minimum_nights	minimum number of night stay for the listing
neighbourhood	Name of the neighbourhood	number_of_reviews	The number of reviews the listing has
latitude	Used to make an interactive map	last_review	The date of the last/newest review
longitude	Used to make an interactive map	availability_90	The availability of the listing x days in the future.
room_type	Entire apt, private room, hotel room.	number_of_reviews_ltm	The number of reviews the listing has (in the last 12 months)
Price	Price in Euro	amenities	A list of amenities in the listing

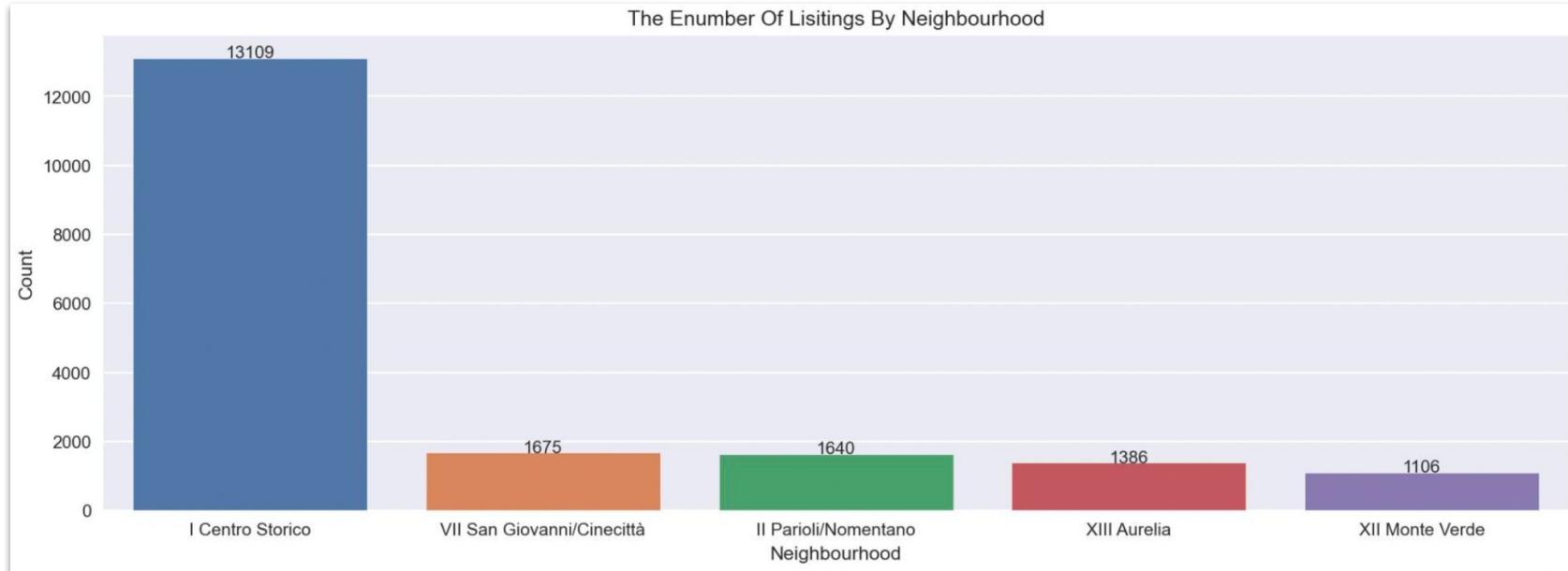




Exploratory Data Analysis (EDA)

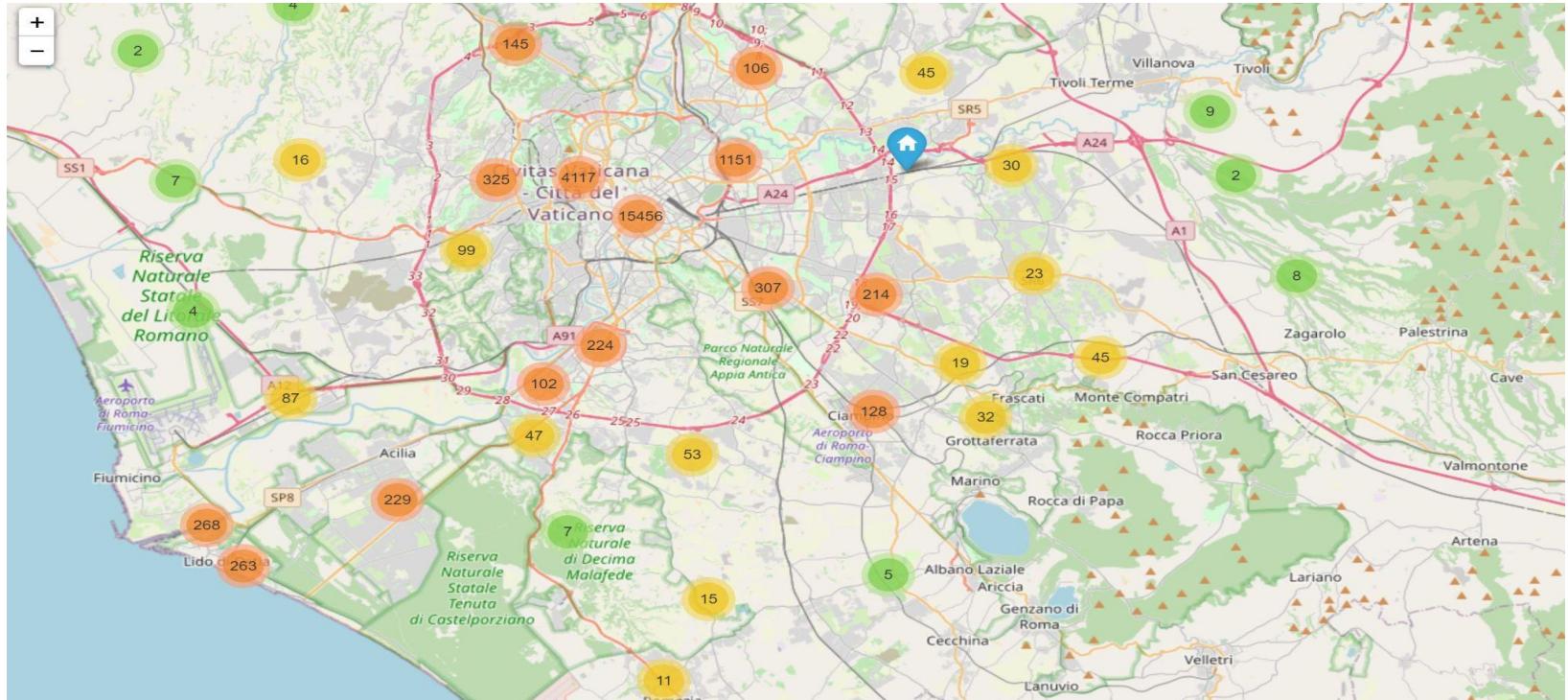
Listings by location

- We wanted to see the top 5 locations for listings in Rome



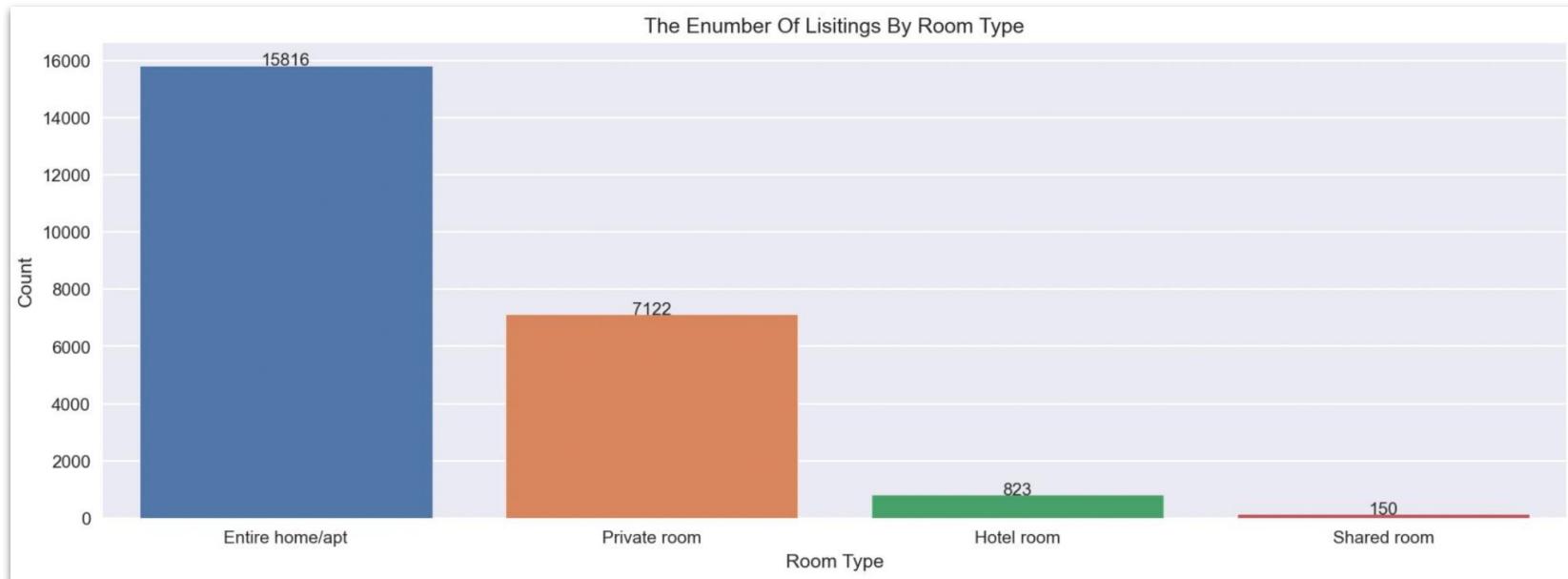
Listings by location

- We used folium to see where listings are located in Rome



Listings room type count

- We wanted to see what are the most common room types in Rome.
- Surprisingly, there are Hotel rooms here which is something I didn't know Airbnb offer.



What about income ?

How did we calculate income

- the dataset doesn't show the income directly so...
- We calculated the expected income over the next 3 months this way!
- Firstly, We calculated the booked nights over the next 3 months using this line

```
romeListings.apply(lambda x : 90-x["availability_90"],axis=1)
```

- Secondly , we simply multiplied it with the price

```
romeListings.apply(lambda x : x["expected_booked_nights_coming_3m"]*x["price"],axis=1)
```



How did we calculate income

- And we calculated the minimum income over the past 12 months this way!
- Using the number of reviews for the past 12 months , minimum nights we calculated the minimum booked nights over the past 12 months

```
romeListings.apply(lambda x : x["number_of_reviews_ltm"]*x["minimum_nights_avg_ntm"],axis=1)
```

- Then we also multiplied it with the price

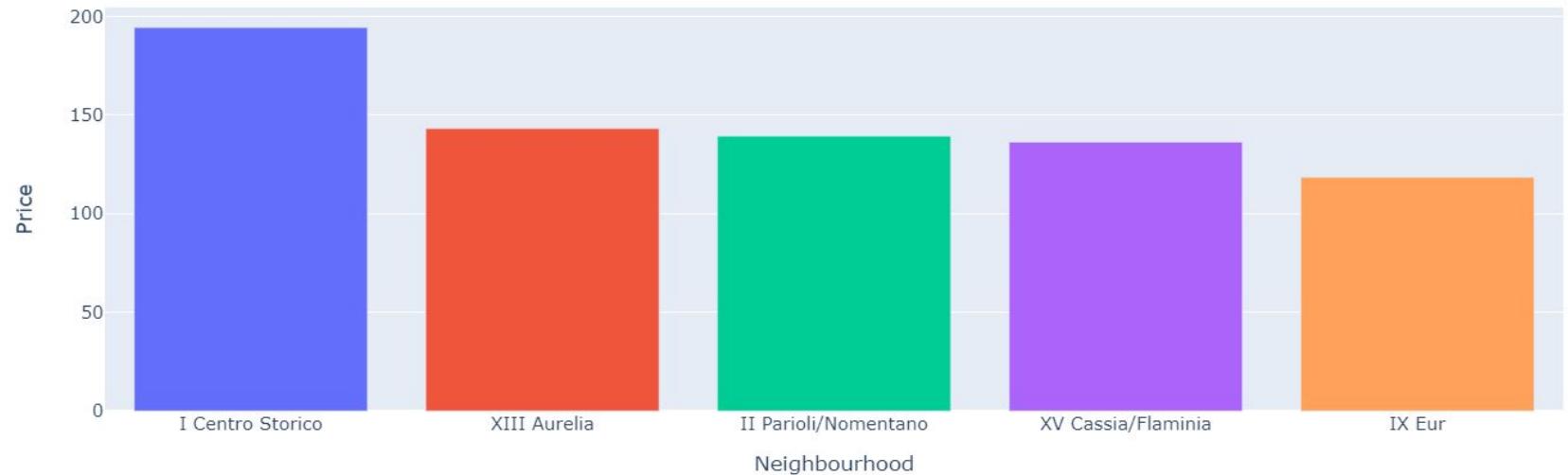
```
romeListings.apply(lambda x : x["min_booked_nights_past_12m"]*x["price"],axis=1)
```



Average price per night (Neighbourhood)

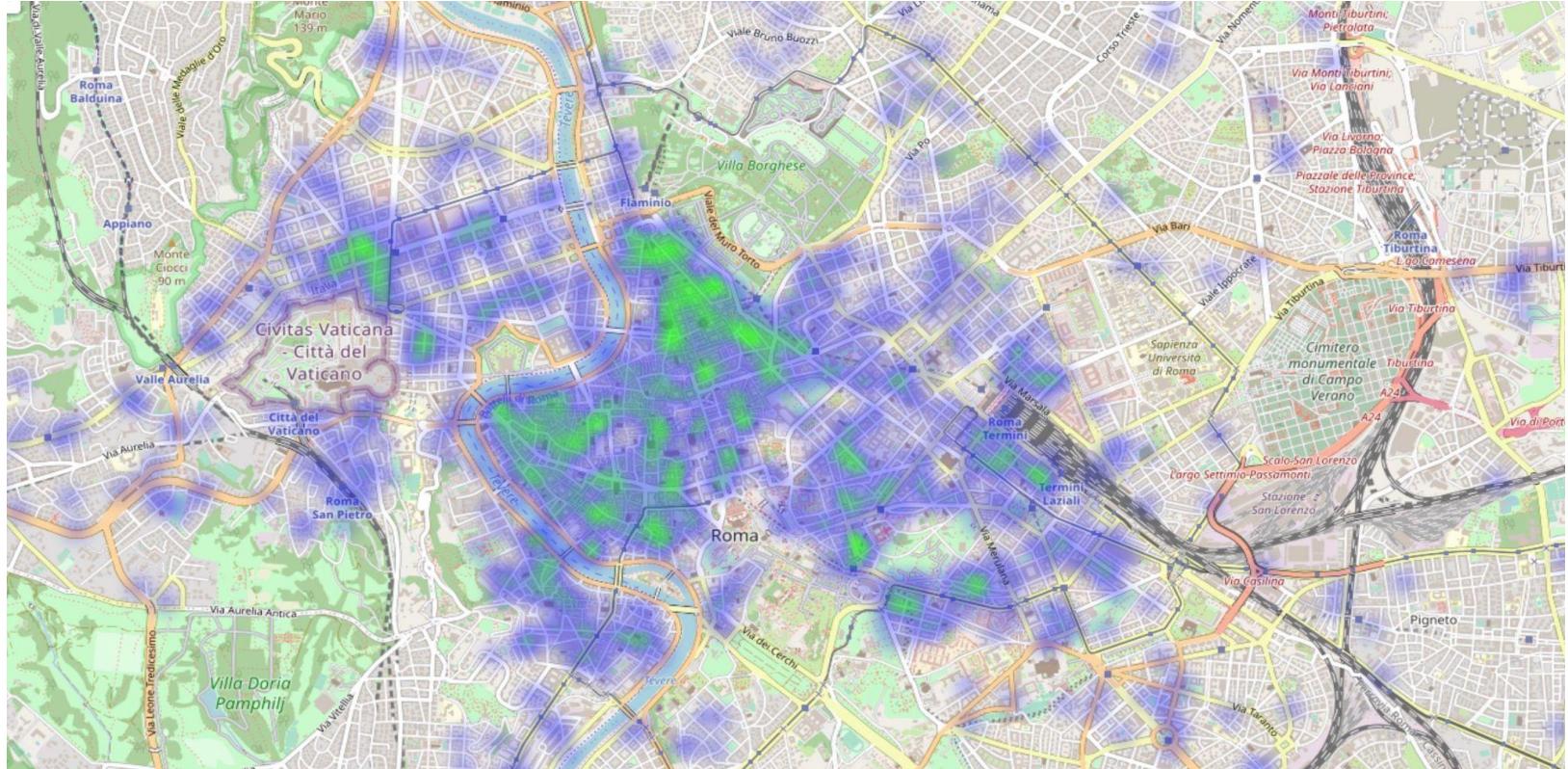
- Lets first take a look at the prices.
 - I Centro Storico looks interesting.

The highest price per night average of listings within a neighbourhood



Average price per night (Neighbourhood)

- Let's look at a heatmap for prices above 200



Average price per night (Room Type)

- Lets first take a look at the prices.
 - Hotel room sand entire homes/apts is pretty expensive unlike shared rooms, but do shared rooms get occupied more often than the others which might help them get back to the income game

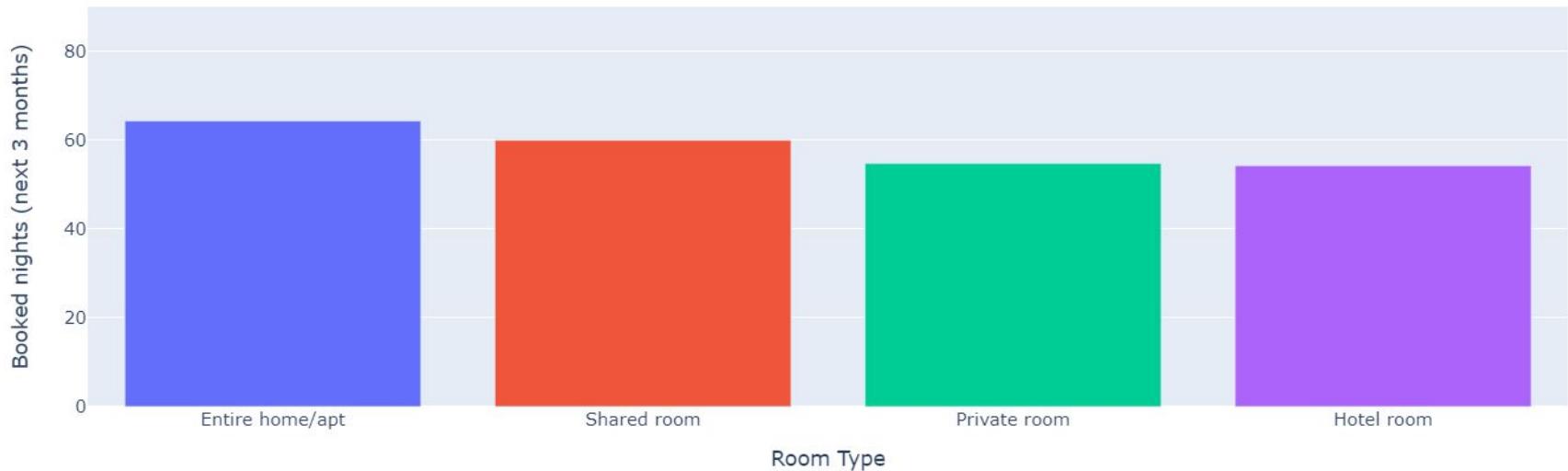
The highest price per night average of listings of a room type



Average booked nights for the coming 3 months (Room Type)

- Well, shared rooms get occupied more often than some of the other but not much more often.
- And entire home/apt is still interesting.

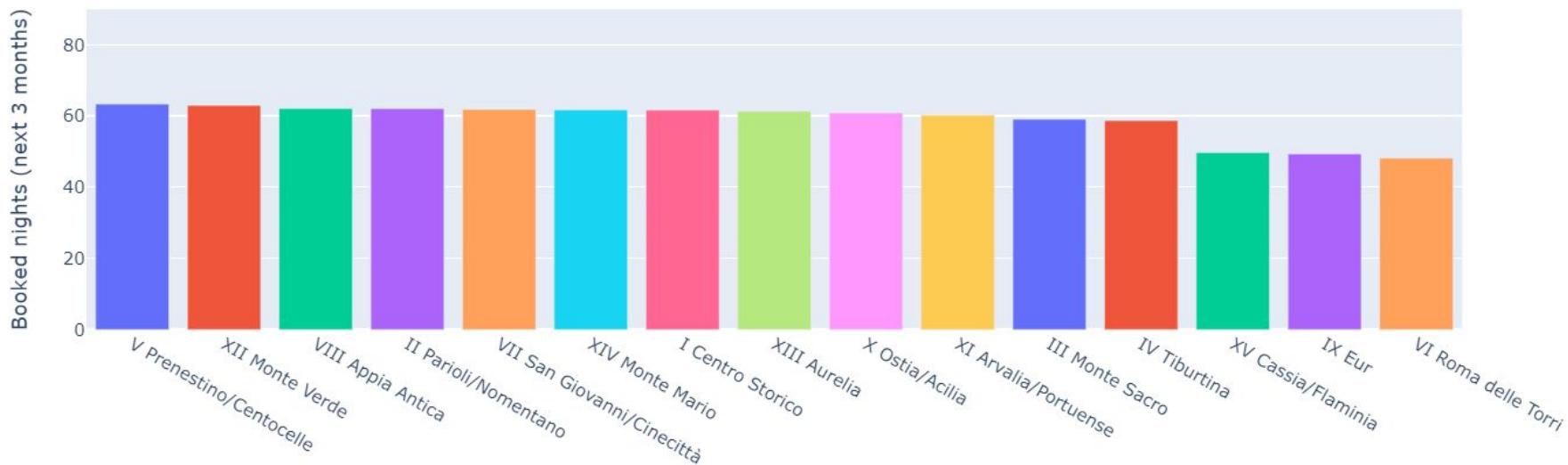
The highest booked nights average of listings of a room type



Average booked nights for the coming 3 months (Neighbourhood)

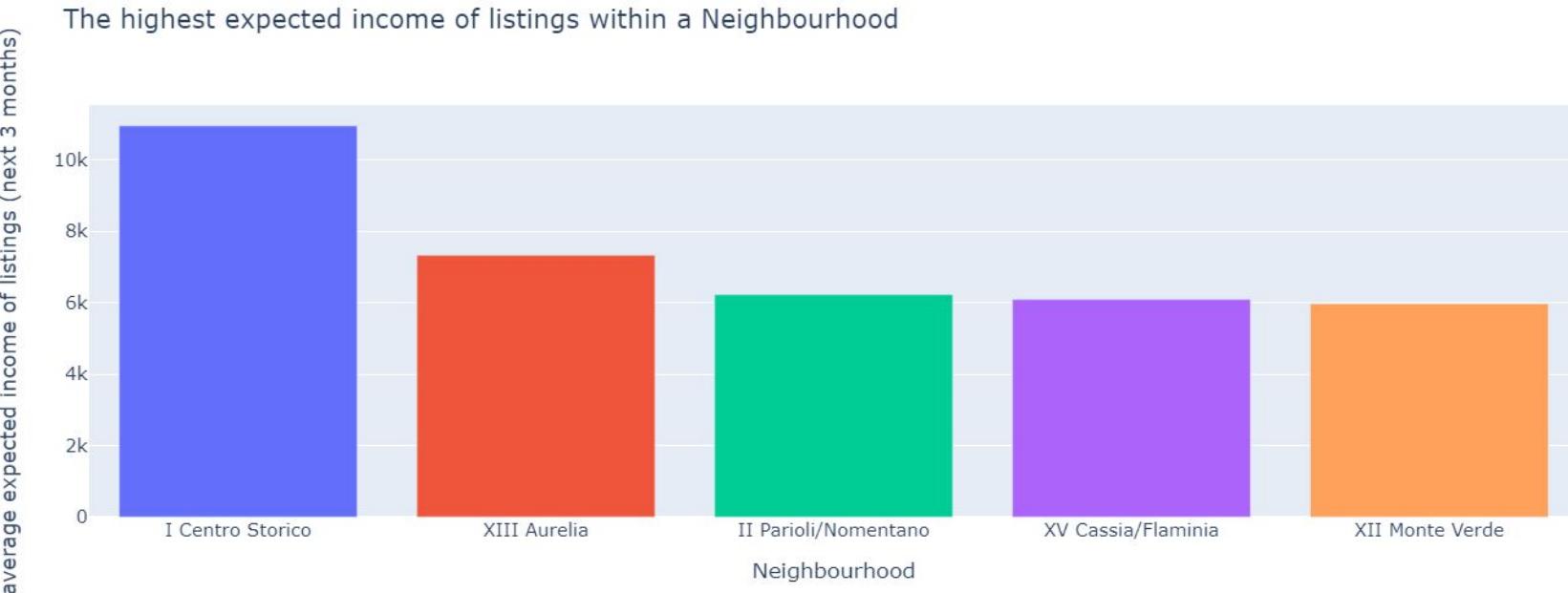
- Here we clearly don't see an outstanding neighbourhood in terms of occupancy

The highest booked nights average of listings within a neighbourhood



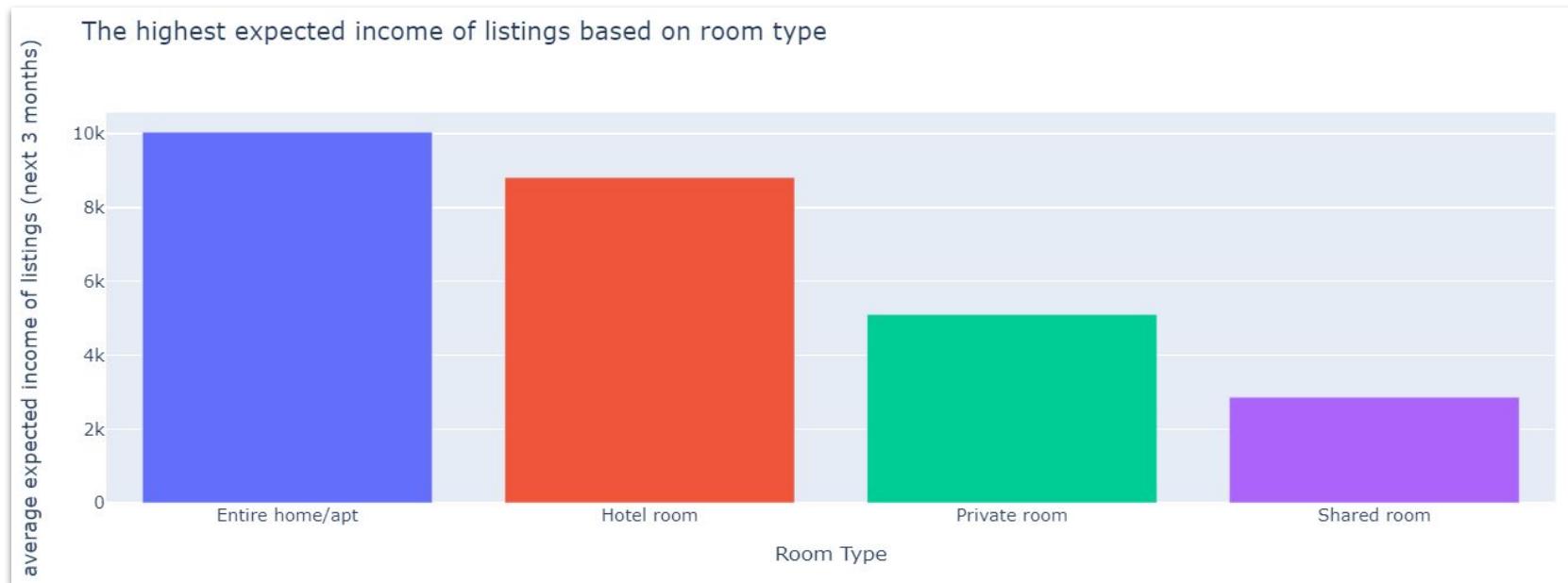
Average expected income for the next 3 months (Neighbourhood)

- What about income ?
 - I Centro Storico still interesting.



Average expected income for the next 3 months (Room Type)

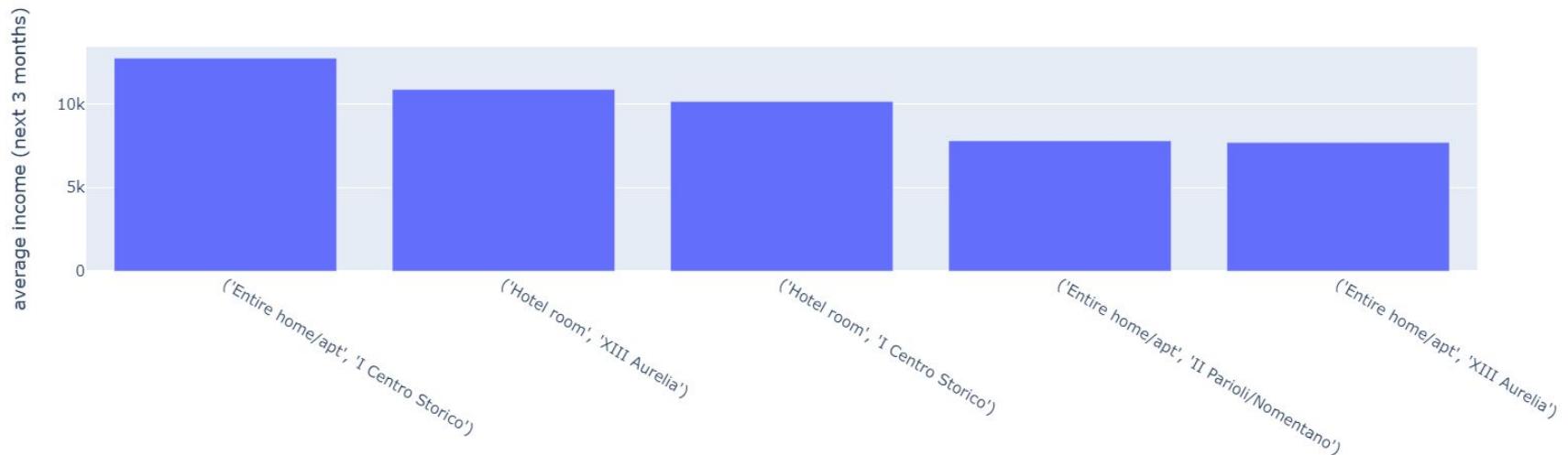
- What about income ?
 - And Entire home/apt is also still interesting.



Average expected income for the next 3 months

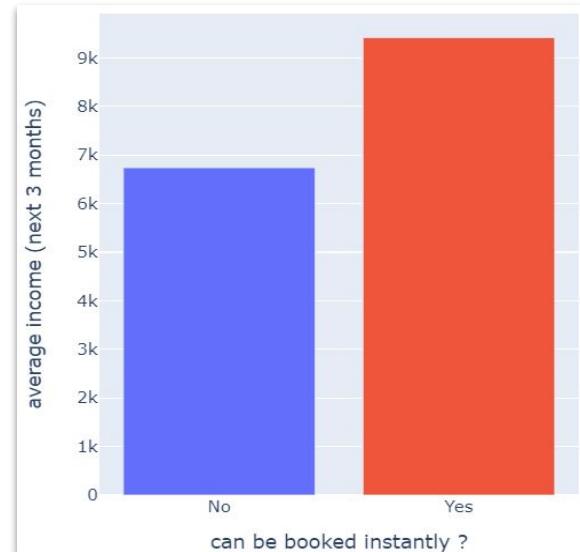
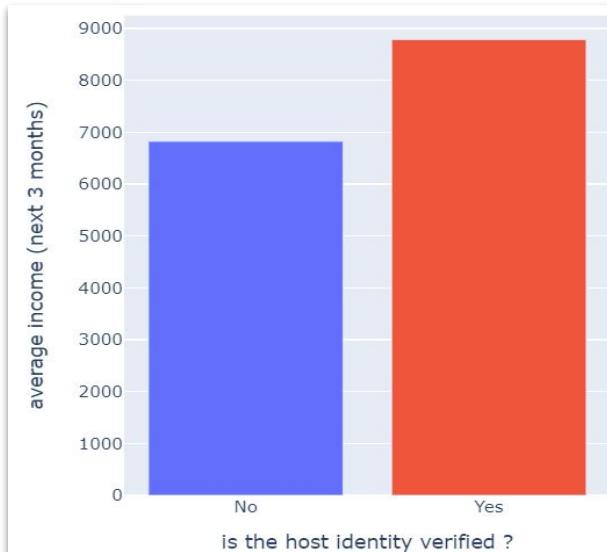
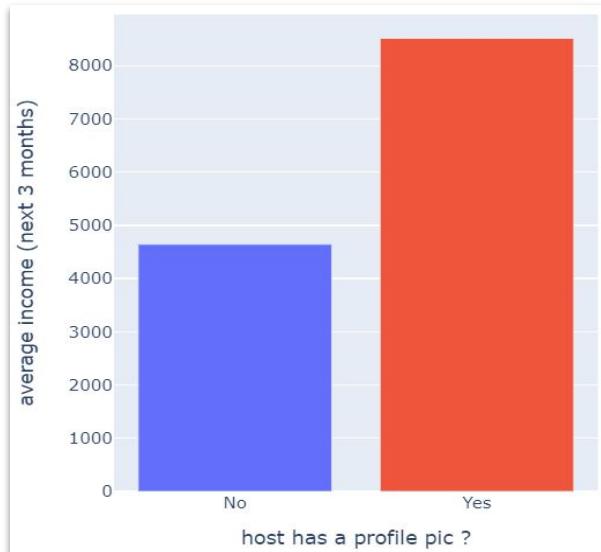
- What about income ?
 - Obviously Entire home/apt in I Centro Storico is the winner here.

The highest income average of listings for every (Room Type,Neighbourhood) combination



Airbnb Account Appearance

- The property attributes in not the only attributes that affect income.
- Here we clearly can see the accounts who are verified and have the booked instantly service and the attribute with highest correlation with the income is simply having a profile pic.



Amenities Analysis

- Cleaning the amenities column
- Performing one-hot encoding

```
df2.amenities[0]                                     Python
✓ 0.4s
['Patio or balcony", "Iron", "Wifi", "Long term stays
allowed", "Essentials", "Air conditioning", "Crib", "Lock on
bedroom door", "Private entrance", "Hair dryer", "Fire
extinguisher", "Fire pit", "Shampoo", "Dedicated workspace",
"BBQ grill", "Host greets you", "Heating", "Free parking on
premises", "Hangers", "Backyard"]'
```



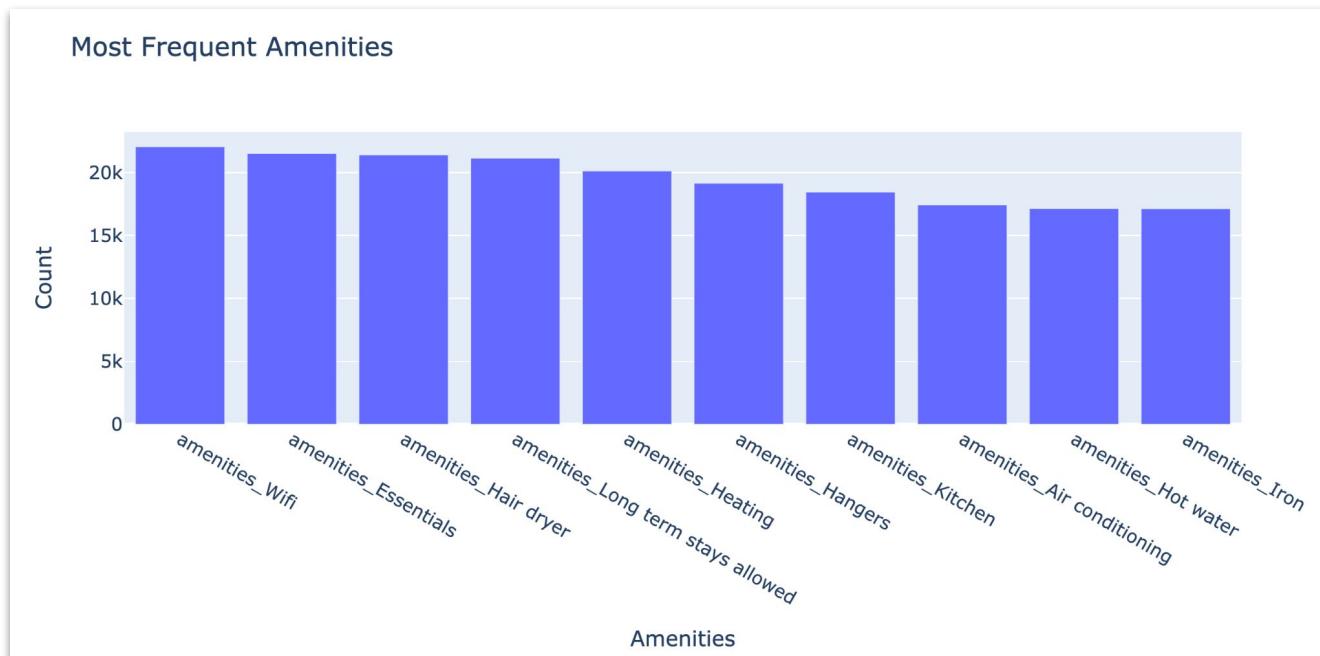
```
df_new[['amenities_Wifi']].head()
```

✓ 0.5s

	amenities_Wifi
id	
2737	1
11834	1
12398	1
19965	1
19967	1

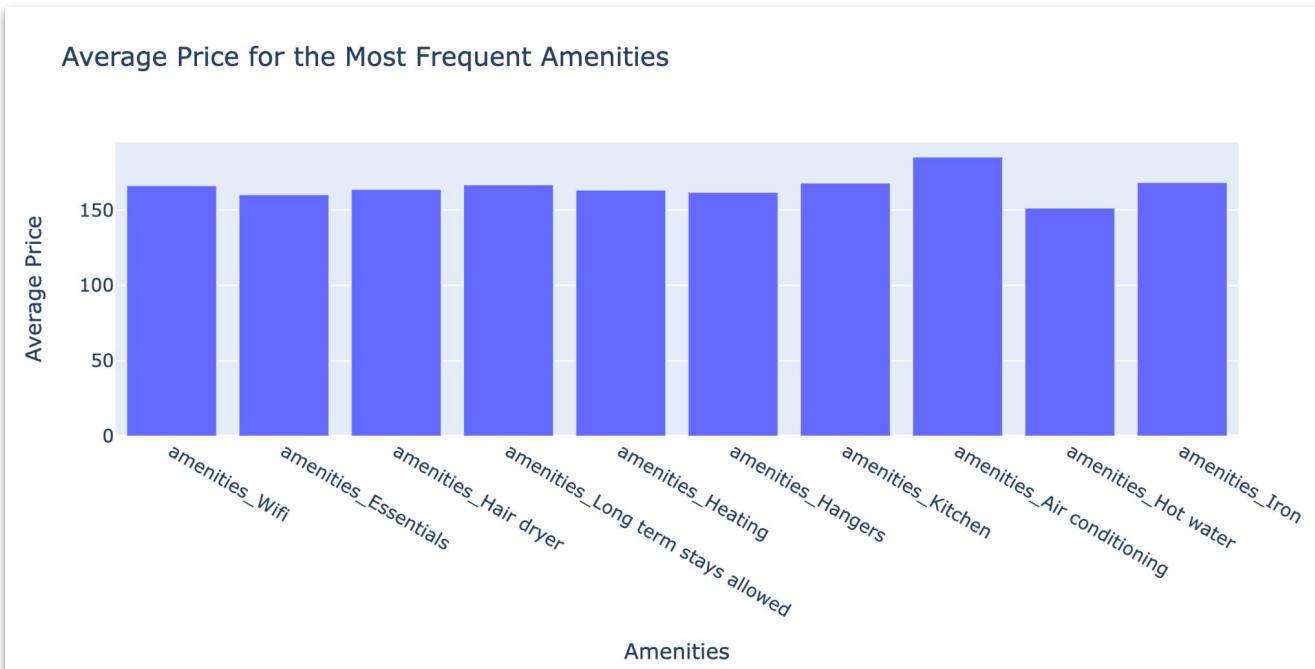
What are the most Frequent Amenities?

- Wifi
- Hair dryer
- Heating
- Kitchen
- Hot water
- Essentials
- Long term stay
- Hangers
- Air conditioning
- Ironing



Average Price for the Most Frequent Amenities

- Air Conditioning has the highest average price with 185.24



Creating a Categorical Column from Review Scores

- Filling NA's with the mean
- Using cut from pandas to create labels and ranges

review_scores_rating
4.390000
4.750000
4.520000
5.000000
4.673438



rating_category
Good
Good
Good
Great
Good

Location Review Scores Analysis

- Listings that have Great or Good rating on locations have a higher average minimum income for the past year.



Location Review Scores Analysis

- Listings that had Great, Good, or Okay rating on locations are expected to be booked more than others in the next three months



Number of Bedrooms Analysis

- Listings that are an entire home or an apartment with seven bedrooms

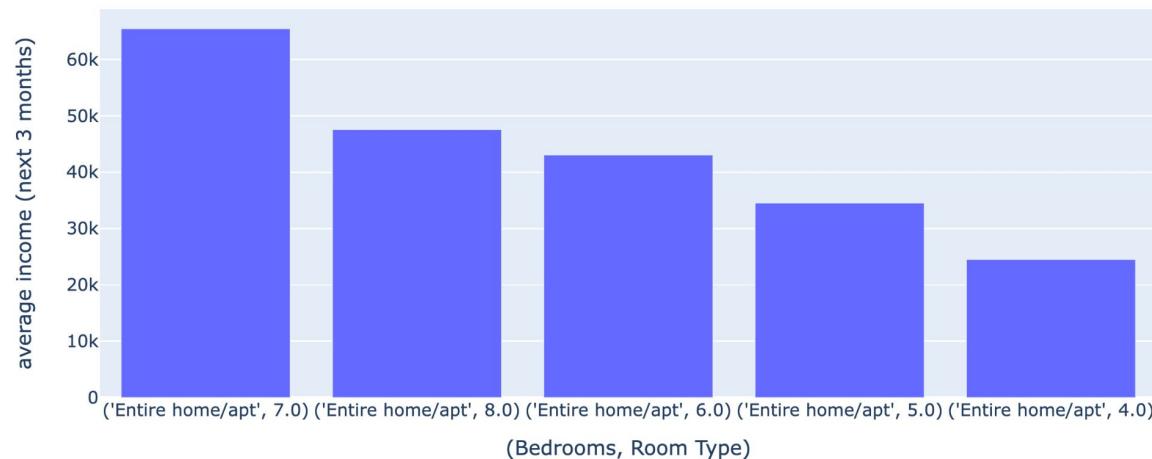
have the highest income average for the past year



Number of Bedrooms Analysis

- Listings that are an entire home or an apartment with seven bedrooms are expected to have the highest income average for the next three months

The highest Expected Income average of listings for every (Bedrooms, Room Type) combination



Models & Hypothesis Testing

- Review Score Value as predicted by Price

★ 4.92 · 12 reviews



Models & Hypothesis Testing

- Review Score Value as predicted by Price



Models & Hypothesis Testing

$$H_0 : \beta = 0 \quad H_1 : \beta \neq 0$$

- Review Score Value as predicted by Price
- $r = 0.03$
- The data showed statistically significant deviation from the null hypothesis of 0 at the 0.5% level.
- As a result we reject the null hypothesis, i.e., the mean is best to describe the `review_score_value`

OLS Regression Results									
Dep. Variable:	review_scores_value		R-squared:	0.001					
Model:	OLS	Adj. R-squared:	0.001						
Method:	Least Squares		F-statistic:	12.84					
Date:	Wed, 06 Jul 2022		Prob (F-statistic):	0.000340					
Time:	14:58:15		Log-Likelihood:	-9733.7					
No. Observations:	19841		AIC:	1.947e+04					
Df Residuals:	19839		BIC:	1.949e+04					
Df Model:	1								
Covariance Type:	nonrobust								
	coef	std err	t	P> t	[0.025	0.975]			
Intercept	4.6754	0.003	1457.006	0.000	4.669	4.682			
price	-3.908e-05	1.09e-05	-3.583	0.000	-6.05e-05	-1.77e-05			
Omnibus:	17281.362		Durbin-Watson:	1.968					
Prob(Omnibus):	0.000		Jarque-Bera (JB):	673666.432					
Skew:	-4.095		Prob(JB):	0.00					
Kurtosis:	30.346		Cond. No.	337.					

Models & Hypothesis Testing

- Review Score Value as predicted by Price



Challenges

- Choosing a dataset based on cities.
- A dataset with more than 70 columns, is not easy.
- Cleaning the `amenities` column, into discrete values.
- First time to use world map library.
- Language and namings difficulties.
- Estimating occupancy and income.

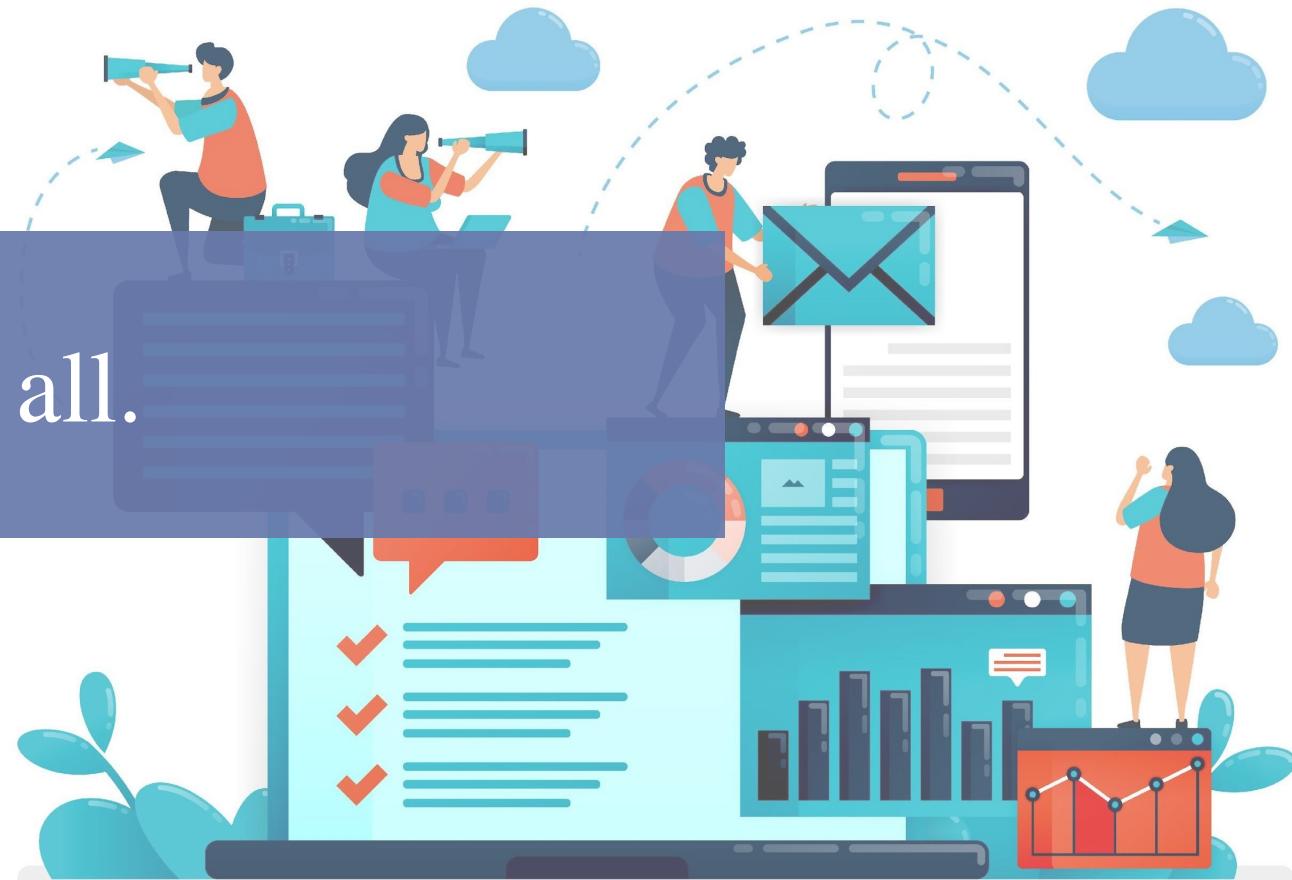


Conclusion

1. Invest in the room type: Entire Home/Apt in ‘II Centro Storico’, with 7 bedrooms
2. Try not to exceed the related average prices, it might lead to bad Scores Value.
3. Include the following amenities: WiFi, Hair Dryer, Heating, Kitchen, Hot Water
4. Try to response as quick as possible.
5. Provide an instant booking.
6. Have a verification mark.
7. Have a profile picture.



Thank you all.





Q&A

Any questions?