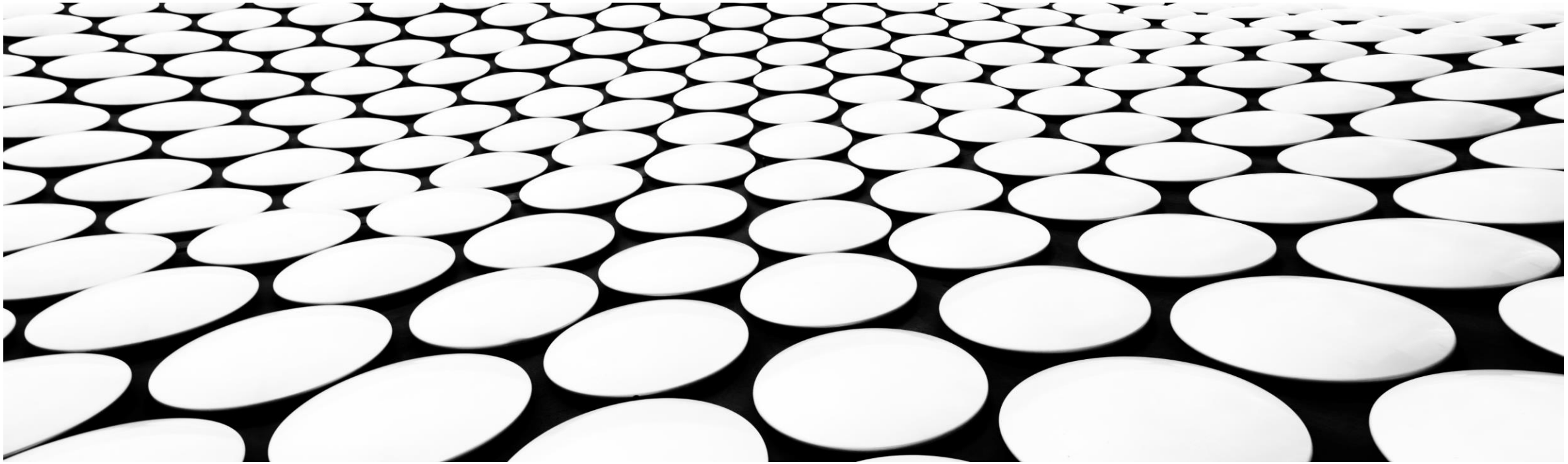
BRAZIL HOUSING

JABASTIN MICHAEL PRIYA F



AGENDA

- INTRODUCTION
- TOOLS
- VISUALIZATION
- RECOMMENDATION

INTRODUCTION

- Work From Home (WFH) has become the new normal for workers around the world. Several companies are allowing at least 40% of their employees to work from home permanently.

OBJECTIVE:

- Rio De Janeiro and Sao Paulo are among the most expensive cities to live in Brazil.
- Castro Brazila Inc., a top real estate management firm with a nationwide presence in Brazil, wishes to help people choose an alternate city to relocate.
- As a data analyst, help the firm figure out suitable cities for relocation for bachelors, for mid-sized families, and for large families

TOOLS AND TECHNIQUES

- EXPLORATORY DATA ANALYSIS – JUPYTER NOTEBOOK
- USER DEFINED FUNCTION
- IF CONDITION
- FOR LOOP
- Pandas
- Numpy

INFORMATION ABOUT DATASETS

INFORMATION

- It has 10692 rows and 13 columns
- Column are separated by Numerical and Categorical variable
- By using .info(), Find that there is no Null values and their datatypes
- There is No Missing Values in the Data Set.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10692 entries, 0 to 10691
Data columns (total 13 columns):
# Column Non-Null Count Dtype
 --- ------ -------------- -----
 0 city 10692 non-null object
 1 area 10692 non-null int64
2 rooms 10692 non-null int64
3 bathroom 10692 non-null int64
4 parking spaces 10692 non-null int64
5 floor 10692 non-null int64
6 animal 10692 non-null object
7 furniture 10692 non-null object
8 hoa (R$) 10692 non-null int64
9 rent amount (R$) 10692 non-null int64
10 property tax (R$) 10692 non-null
int64
11 fire insurance (R$) 10692 non-null
int64
12 total (R$) 10692 non-null int64
dtypes: int64(10), object(3)
memory usage: 1.1+ MB
```

DEFINING A CATEGORICAL AND NUMERICAL VARIABLE

```
def separate(brazil_data):
    categorical=[]
    numerical=[]
    for columns in brazil_data.columns:
        if brazil_data[columns].nunique() < 40:
            categorical.append(columns)
        else:
            numerical.append(columns)
    return categorical, numerical

categorical, numerical=separate(brazil_data)

from tabulate import tabulate
table=[categorical, numerical]
print(tabulate({"Categorical Columns":categorical,
                "numerical Columns":numerical}, headers=['categorical','numerical']))
```

CATEGORICAL AND NUMERICAL VARIABLE COLUMNS

CATEGORICAL

- CITY
- ROOMS
- BATHROOMS
- PARKING SPACES
- FLOOR
- ANIMAL
- FURNITURE

NUMERICAL

- AREA
- HOA
- RENT AMOUNT (R$)
- PROPERTY TAX (R$)
- FIRE INSURANCE (R$)
- TOTAL (R$)

DEFINING A FUNCTION FOR CATEGORICAL COLUMNS

```
def fun_cat(col):
    if brazil_data[col].isnull().sum()>0:
        print(f"\n There are null values in {col} columns")
    print(f"{col}: Mode of {col} are {brazil_data[col].mode()[0]}")
    print(f"{col}: Unique Values in {col} are: {brazil_data[col].unique()}")
    print(f"{col}: Number of Missing values in {col} is {brazil_data[col].isnull().sum()}")
```

DEFINING A FUNCTION FOR NUMERICAL COLUMNS

```
def fun_num(col):
    if brazil_data[col].isnull().sum()>0:
        brazil_data[col].fillna(brazil_data[col].mean(),
inplace=True)
    print(f"{col}: Mean is {brazil_data[col].mean()}")
    print(f"{col}: Median is {brazil_data[col].median()}")
```

MEAN, MEDIAN AND MODE OF CATEGORICAL AND NUMERICAL COLUMNS

CATEGORICAL COLUMNS

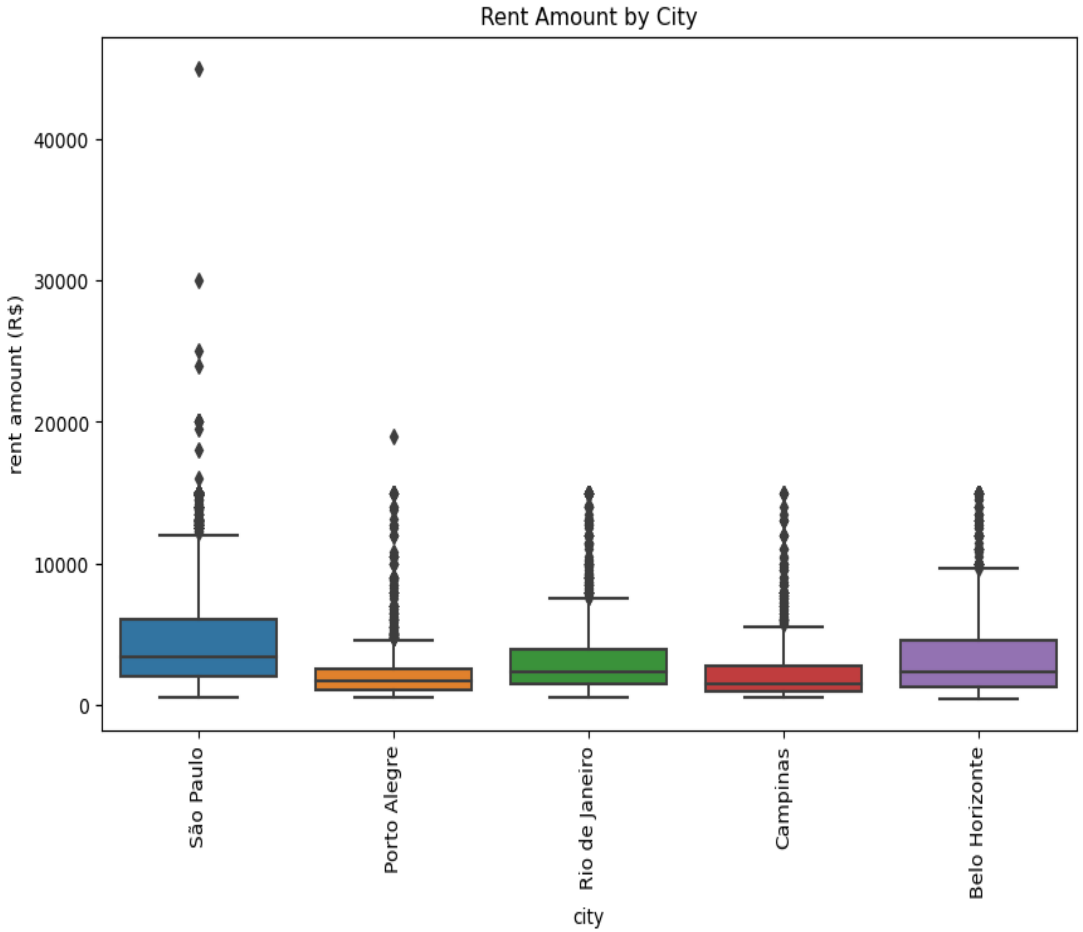
- MODE OF CITY IS SAO PAULO
- MODE OF ROOMS IS 3
- MODE OF BATHROOM IS 1
- MODE OF PARKING SPACES IS 1
- MODE OF FURNITURE IS NOT FURNISHED
- MODE OF PETS IS ACCEPT
- MODE OF FLOOR IS 0

NUMERICAL COLUMNS

- MEAN OF AREA IS 137.97
- MEDIAN IS 90
- MEAN OF HOA IS 1174.02
- MEDIAN OF HOA IS 560
- MEAN OF RENT AMOUNT IS 3896.24
- MEDIAN OF RENT AMOUNT IS 2661.0
- MEAN OF PROPERT TAX IS 366.70
- MEDIAN OF PEOPERTY TAX IS 125.0
- MEAN OF TOTAL COST IS 5228.78
- MEDIAN IS 5357.0
- MEAN OF FIRE INSURANCE IS 53.30
- MEDIAN OF FIRE INSURANCE IS 36.0

CITY VS RENT AMOUNT

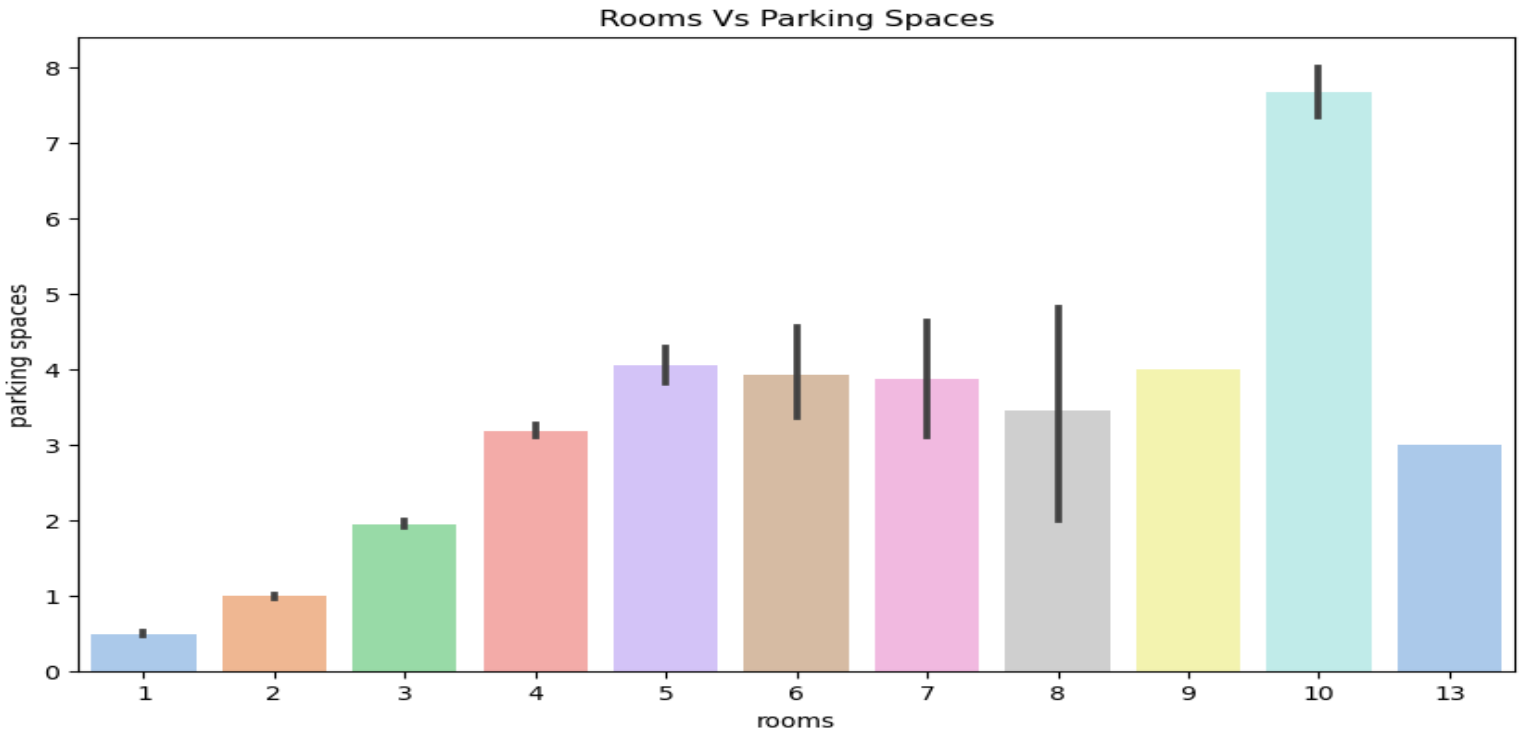
- plt.figure(figsize=(10,6))
- sb.boxplot(data=brazil_data, x='city', y='rent amount (R$)')
- plt.xticks(rotation=90)
- plt.xlabel('city')
- plt.ylabel('rent amount (R$)')
- plt.title('Rent Amount by City')
- plt.show()



RECOMMENDATION: **Campinas and Porto Alegre are the cheapest city with less rent amount**

ROOMS VS PARKING SPACES

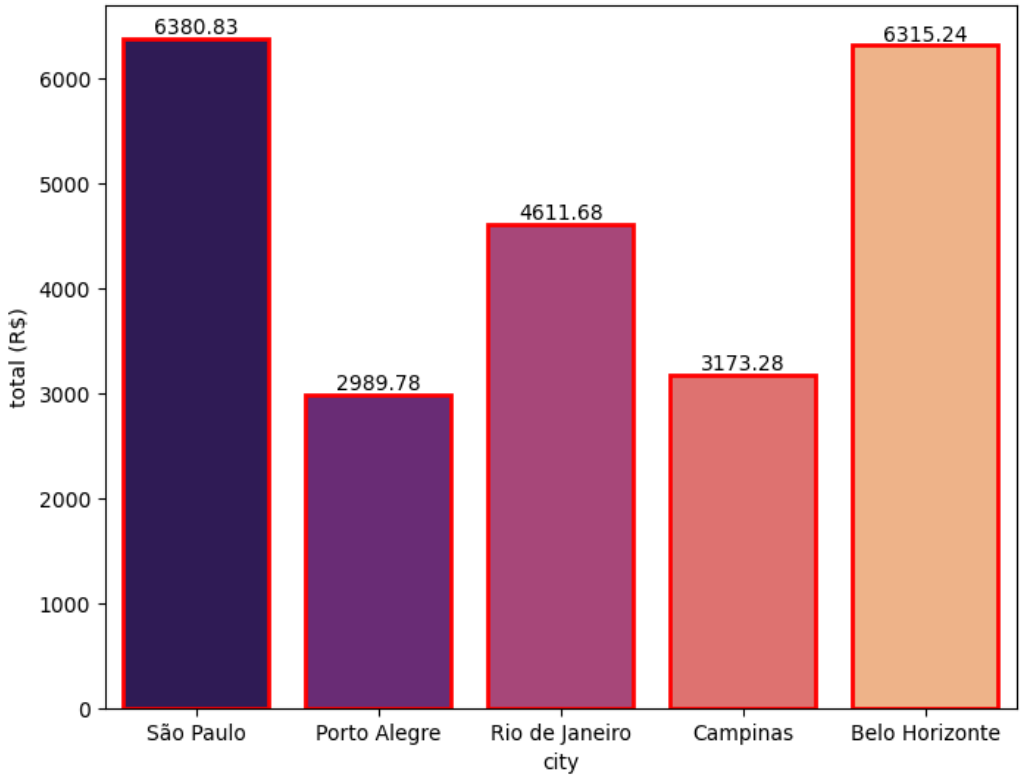
- plt.figure(figsize=(10,6))
- sb.barplot(x='rooms', y='parking spaces'
  data=brazil_data, palette='pastel')
- plt.show()



RECOMMENDATION: **The Parking space is higher when number of rooms increases on increase of rooms and that's true considering bigger Family would have more cars and need more parking spaces.**

TOTAL COST FOR RENT IN DIFFERENT CITY

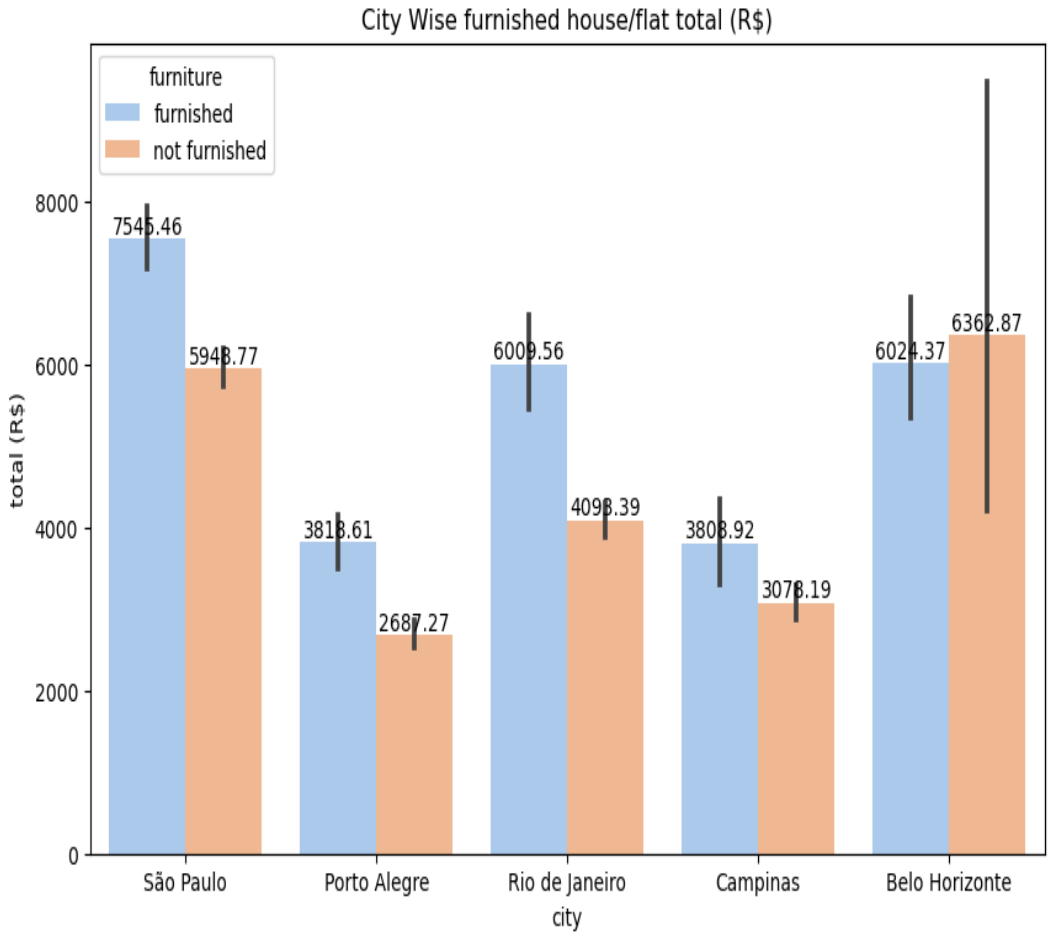
```
plt.figure(figsize=(8,6))
ax=sb.barplot(x='city', y='total (R$)', hue='rent amount
(R$)', data=brazil_data, palette='magma', edgecolor='red'
linewidth=2)
for bar in ax.patches:
    percentage = f"{round(bar.get_height() , 2)}"

    x=bar.get_x() + bar.get_width()/2
    y=bar.get_height()
    ax.annotate(percentage, (x,y), va='bottom',
ha='center')
plt.show()
```



RECOMMENDATION: THE TOTAL COST FOR RENT IN CITY OF PORTO ALEGRE IS CHEAPER THAN OTHER CITIES

CITY, TOTAL COST , FURNITURE

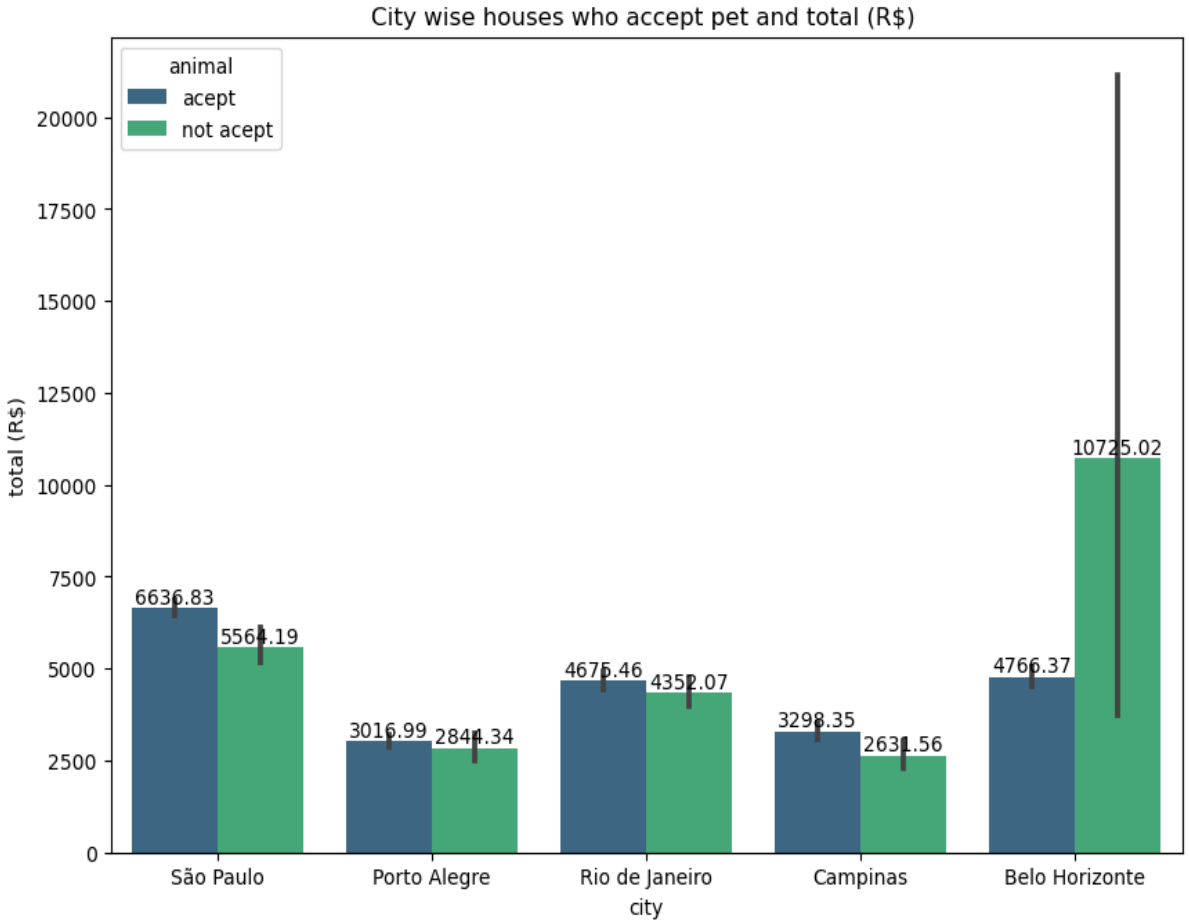
- fig,ax=plt.subplots(figsize=(10,6))
- ax=sb.barplot(x='city', y='total (R$)',data=brazil_data, hue='furniture', palette='pastel', linewidth=2, linestyle='--')
- plt.title('City Wise furnished house/flat total (R$)')
- for bar in ax.patches:
- percentage = f"{round(bar.get_height() , 2)}"
- x=bar.get_x() + bar.get_width()/2
- y=bar.get_height()
- ax.annotate(percentage, (x,y), va='bottom', ha='center')
- plt.show()



RECOMMENDATION: **If someone looking for furnished house and that should be cheap is again Porto Alegre**

CITY, TOTAL VS PET

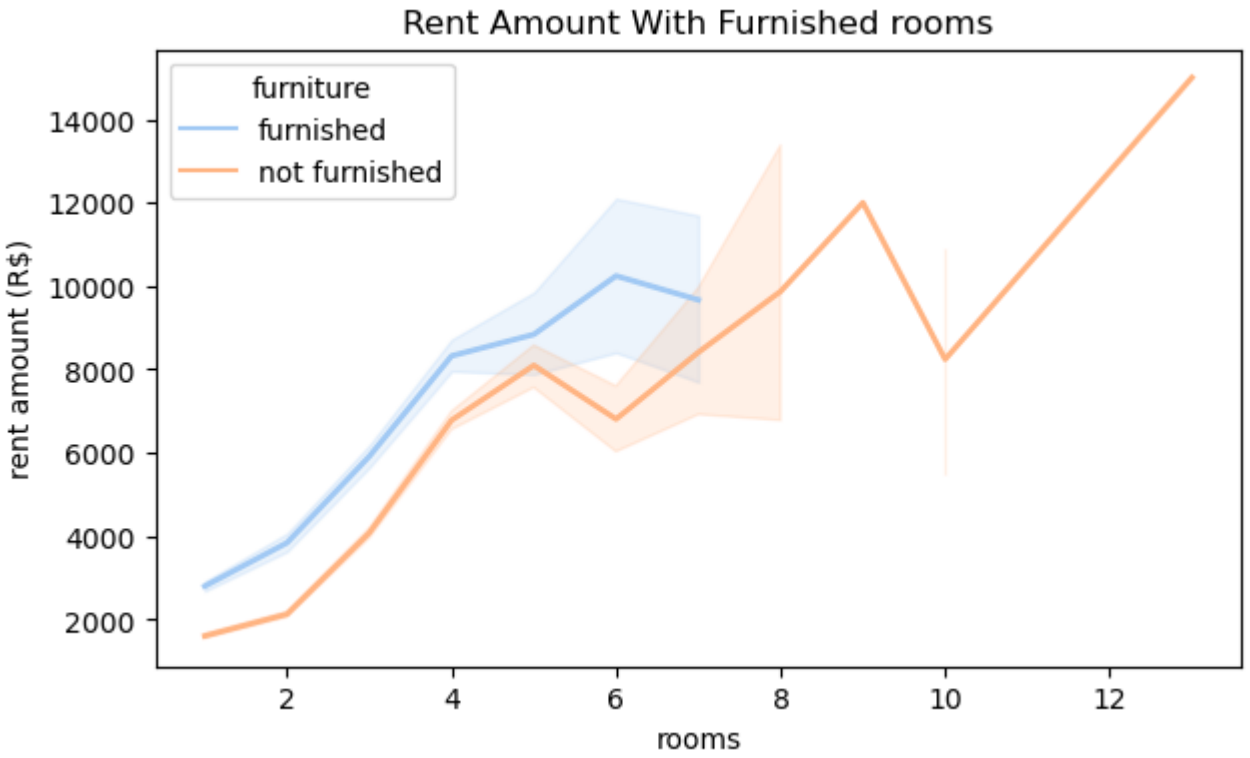
- fig,ax=plt.subplots(figsize=(10,7))
- ax=sb.barplot(x='city', y='total (R$)',data=brazil_data, hue='animal', palette='viridis', linewidth=2, linestyle='--')
- plt.title('City wise houses who accept pet and total (R$) ')
- for bar in ax.patches:
- percentage = f"{round(bar.get_height() , 2)}"
-
- x=bar.get_x() + bar.get_width()/2
- y=bar.get_height()
- ax.annotate(percentage, (x,y), va='bottom', ha='center')
- plt.show()



RECOMMENDATIONS: **Cities Porto Alegre and Campinas are almost accepts pets 50% and total cost is also not expensive**

ROOMS, RENT AMOUNT VS FURNITURE

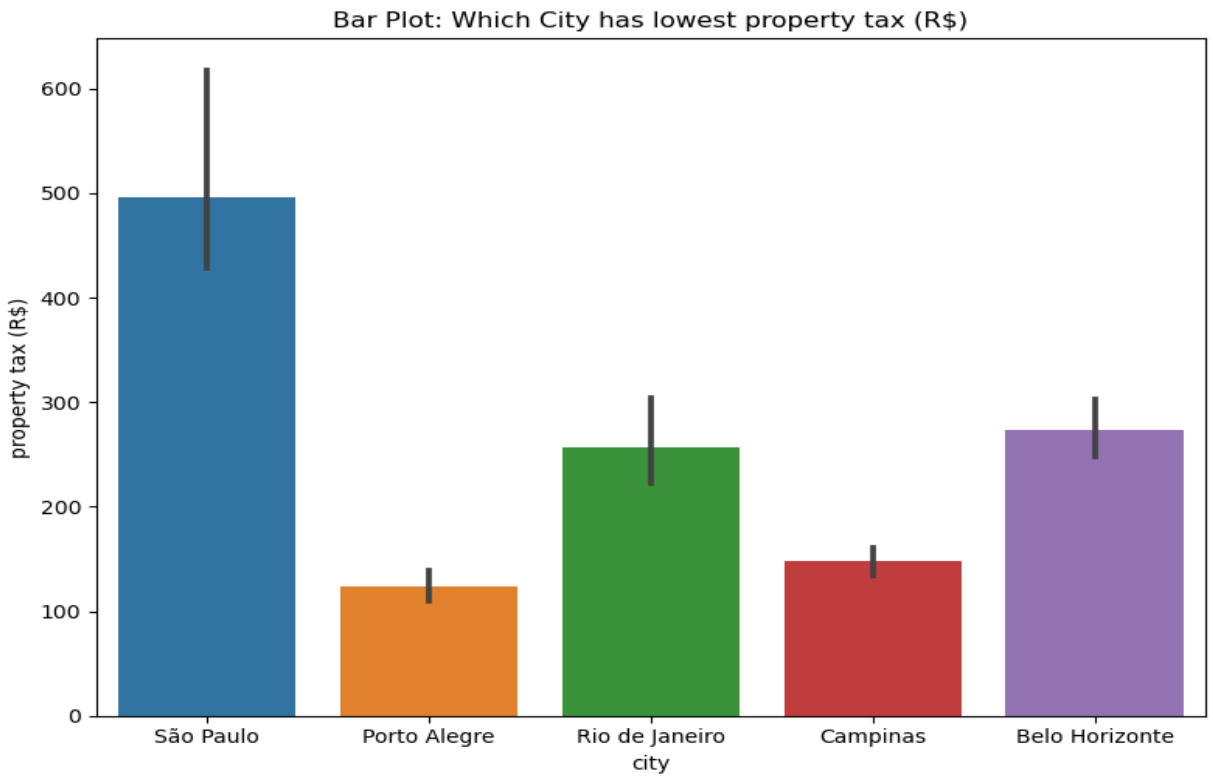
- plt.figure(figsize=(7,4))
- sb.lineplot(data=brazil_data, x='rooms', y='rent amount (R$)', hue='furniture', palette='pastel', linewidth=2)
- plt.xlabel('rooms')
- plt.ylabel('rent amount (R$)')
- plt.title('Rent Amount With Furnished rooms')
- plt.show()



RECOMMENDATIONS: **The Two room set are really affordable which are furnished, otherwise the furnished homes are really expensive.**

WHICH CITY HAS THE LOWEST PROPERTY TAX

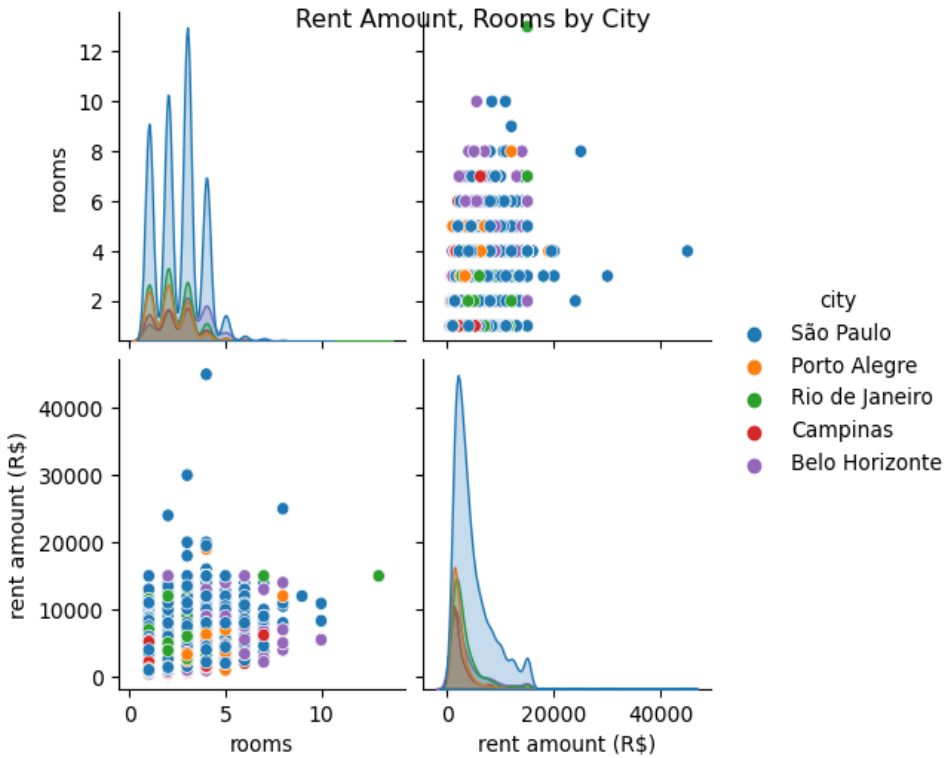
- plt.figure(figsize=(8,6))
- sb.barplot(data=brazil_data, x='city', y='property tax (R$)')
- plt.title('Bar Plot: City has a property tax (R$)')
- plt.tight_layout()
- plt.show()



RECOMMENDATION: Porto Alegre has lowest Property tax and Campinus also have less property Tax

ROOMS, RENT AMOUNT AND CITY

- plt.figure(figsize=(10, 6))
- sb.pairplot(data=brazil_data, vars=(['rooms','rent amount (R$)']), hue='city')
- plt.xlabel('Number of Rooms')
- plt.ylabel('Rent Amount')
- plt.suptitle('Rent Amount, Rooms by City')
- plt.show()



RECOMMENDATION: If we take room wise Sao Paulo can be considered. If we take rent amount (R$) wise let consider Campinas

RECOMMENDATION

- **Campinas and Porto Alegre are the cheapest city with less rent amount**
- **Porto Alegre has lowest Property tax and Campinus also have less property Tax**
- **The Total cost for rent in city of Porto Alegre is cheaper than other cities.**
- **If someone looking for furnished house and that should be cheap is again Porto Alegre**
- **Cities Porto Alegre and Campinas are almost accepts pets 50% and total cost is also not expensive**
- **The Two room set are really affordable which are furnished, otherwise the furnished homes are really expensive.**
- **If we take room wise Sao Paulo can be considered. If we take rent amount (R$) wise let consider Campinas**
- **The Parking space is higher when number of rooms increases on increase of rooms and that true considering bigger Family would have more cars and need more parking spaces.**