

Midterm Project CSS 328 07-p

Mukashev Sultan

Data collected from [Krisha.kz](#) and [Etaqi.kz](#)

Questions:

1. What city has highest price for apartments?

```
# 1.What city has highest price for apartments?
cities = res1.groupby('City')['Price'].apply(np.median).reset_index(name='avg_price')
cities[cities['avg_price'] == cities['avg_price'].max()]
```

	City	avg_price
4	Алматы	39900000.0

Firstly, we group all data by their cities, then find median of prices of each city. Then find maximum value and find city where price equals to it and show.

2. What city is more pfofitable in case of Year/Price ratio?

```
#2. What city is more pfofitable in case of Year/Price ratio?
cities1 = res.groupby('City')['Price'].apply(np.median).reset_index(name='avg_price')
cities2 = res.groupby('City')['Year'].apply(np.median).reset_index(name='Year')
cities = pd.merge(cities1,cities2, on = 'City')
cities.sort_values(by = ['Year','avg_price'],ascending = [False,True]).head(1)
```

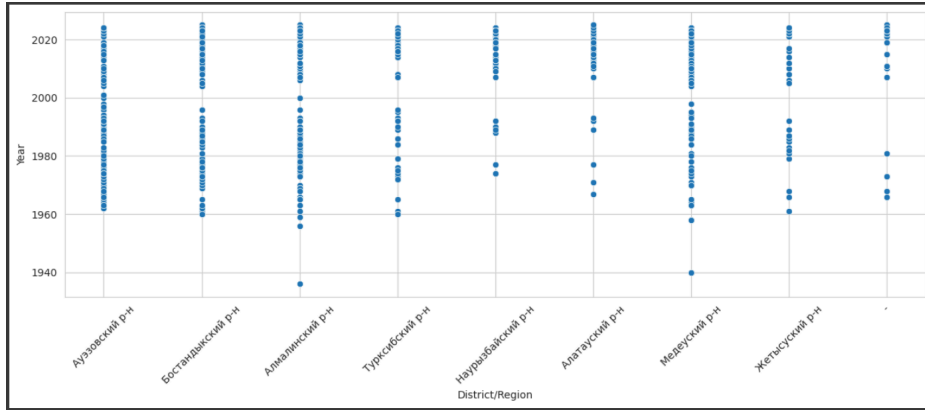
	City	avg_price	Year
33	Туркестан	19028800.0	2025.0

In this case, we also group data by cities and get their median price and year of building, then sort. Firstly by descending year, and then by ascending price, to find where we get newest houses with lowest price.

3. What districts in Almaty has most of new apartments?

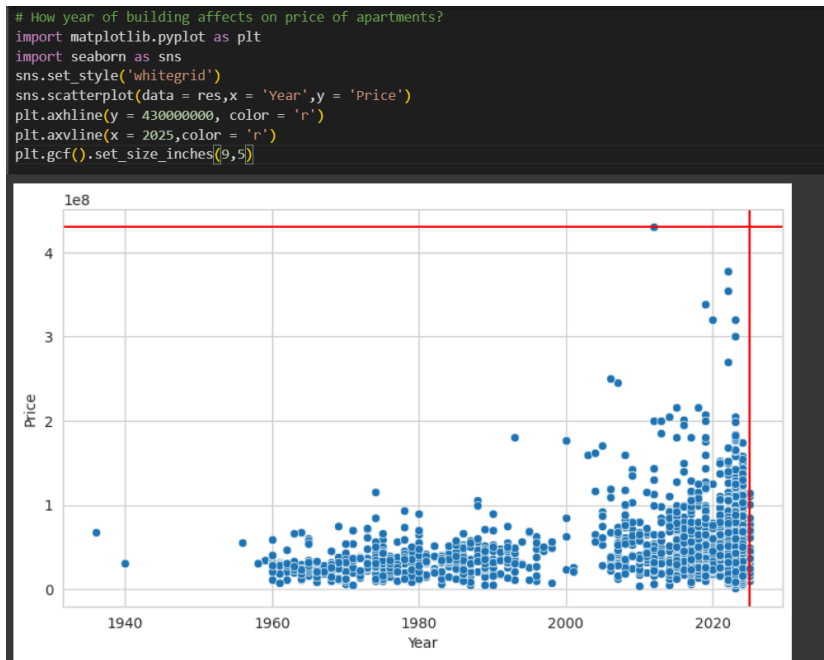
```
#3. What districts in Almaty has most of new apartments?
ala = res.query('City == "Алматы"')
dist = ala.groupby('District/Region')['Year'].apply(np.median).reset_index(name = 'Year')
dist.sort_values(by = 'Year',ascending=False)
```

	District/Region	Year
0	-	2023.0
1	Алатауский р-н	2022.0
4	Бостандыкский р-н	2022.0
5	Жетысуский р-н	2022.0
6	Медеуский р-н	2022.0
7	Наурызбайский р-н	2022.0
8	Турксибский р-н	2021.0
2	Алмалинский р-н	2014.5
3	Ауэзовский р-н	2006.0



According to table we can say that most of the districts has new apartments, built in 2022, and looking at scatterplot we can also conclude that 'Nauрызбай' and 'Alatau' districts has lowest percentage of old buildings which can make them winners in this question.

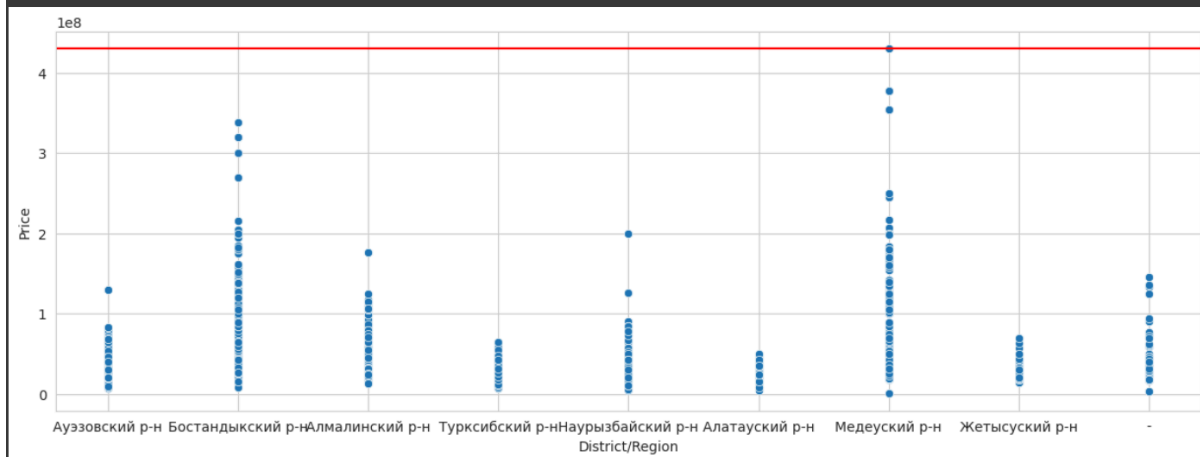
4. How year of building affects on price of apartments?



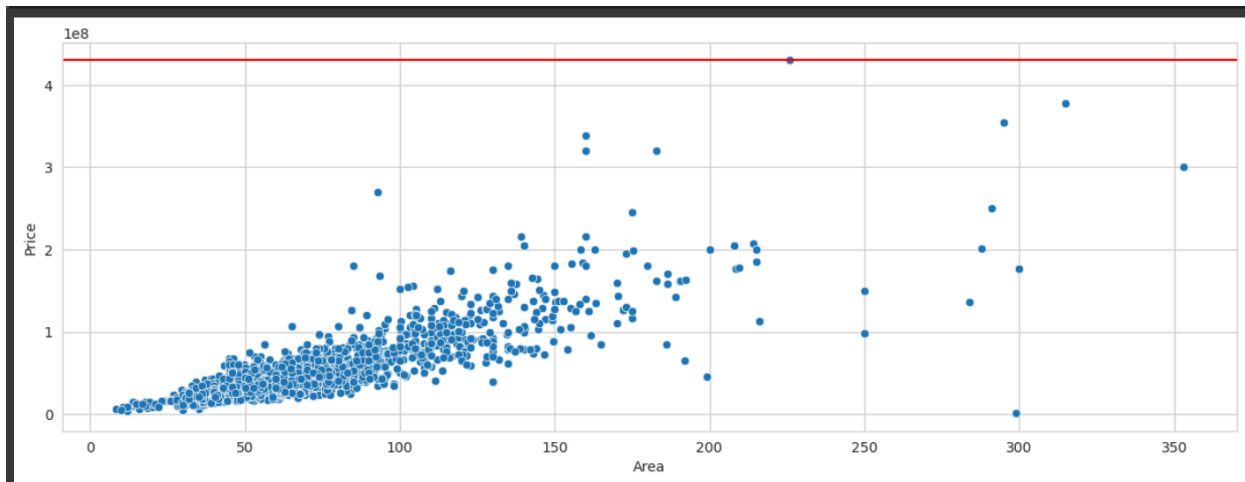
A: So we can see that newer apartments costs more, but generally, most of houses has similar prices. And we can see that older buildings don't have prices higher than 150 million tenge.

5. How district/area/quantity of rooms of building affects on price?

```
# How district/area/quantity of rooms of building affects on price?
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style('whitegrid')
sns.scatterplot(data = ala, x = 'District/Region', y = 'Price')
plt.axhline(y = 430000000, color = 'r')
plt.gcf().set_size_inches(15,5)
```

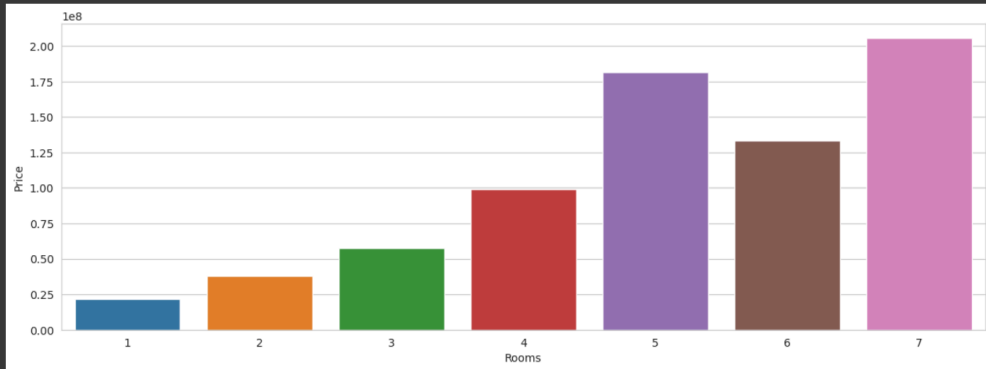


A1: The most expensive apartments are in 'Medeu' and 'Bostandyk' districts and lowest ones are in 'Alatau' and 'Turksib';

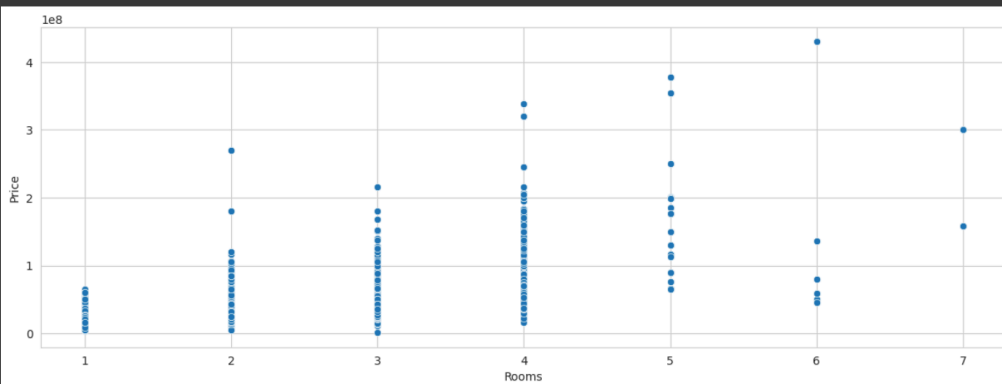


A2: Price is dependent on area almost linearly, the more is area, more the price. But there are also some outliers.

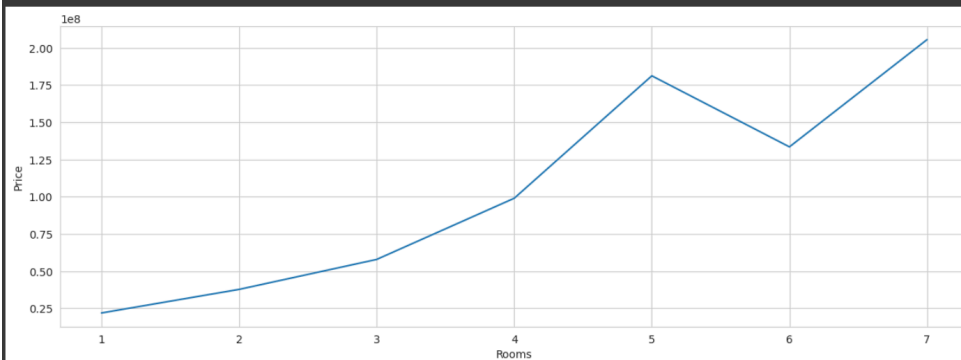
```
# How district/area/quantity of rooms of building affects on price?
import matplotlib.pyplot as plt
import seaborn as sns
rooms = res.groupby('Rooms')['Price'].apply(np.mean).reset_index(name = 'Price')
sns.set_style('whitegrid')
sns.barplot(data = rooms,x = 'Rooms',y = 'Price')
plt.gcf().set_size_inches(15,5)
```



```
# How district/area/quantity of rooms of building affects on price?
import matplotlib.pyplot as plt
import seaborn as sns
rooms = res.groupby('Rooms')['Price'].apply(np.mean).reset_index(name = 'Price')
sns.set_style('whitegrid')
sns.scatterplot(data = res,x = 'Rooms',y = 'Price')
plt.gcf().set_size_inches(15,5)
```



```
# How district/area/quantity of rooms of building affects on price?
import matplotlib.pyplot as plt
import seaborn as sns
rooms = res.groupby('Rooms')['Price'].apply(np.mean).reset_index(name = 'Price')
sns.set_style('whitegrid')
sns.lineplot(data = rooms,x = 'Rooms',y = 'Price')
plt.gcf().set_size_inches(15,5)
```



A3: Price changes almost linearly quantity of rooms also, but as we see in scatter plot, there are not a lot of data about apartments with more rooms, which can cause some unexpected results as avg price of 5 room apart. Is more than ones with 6 rooms.

6. What district in Almaty is most profitable in case of Year/Price ratio?

```
# What district in Almaty is most profitable in case of Year/Price ratio?
dist1 = ala.groupby('District/Region')['Year'].apply(np.median).reset_index(name = 'Year')
dist2 = ala.groupby('District/Region')['Price'].apply(np.median).reset_index(name = 'Price')
dist = pd.merge(dist1,dist2,on='District/Region')
dist.sort_values(by = ['Year','Price'],ascending = [False,True])
```

	District/Region	Year	Price
0	-	2023.0	34496788.0
1	Алатауский р-н	2022.0	25000000.0

In this case, we group data by districts, firstly taking only ones which are in Almaty and get their median price and year of building, then sort. Firstly by descending year, and then by ascending price, to find where we get newest houses with lowest price. Since we have some outliers with undefined district, we should consider second best as our result.

7. Apartments in which district are not worth of their cost considering their other features?

```
# Apartments in which district are not worth of their cost considering their other features?
dist1 = ala.groupby('District/Region')['Year'].apply(np.median).reset_index(name = 'Year')
dist2 = ala.groupby('District/Region')['Price'].apply(np.median).reset_index(name = 'Price')
dist = pd.merge(dist1,dist2,on='District/Region')
dist.sort_values(by = ['Year','Price'],ascending = [False,True]).tail(1)
```

	District/Region	Year	Price
3	Ауэзовский р-н	2006.0	38000000.0

Other features are not considered, because they are not dependent on districts, but effects prices, so we can say that 'Auezov' district has oldest houses but not lowest prices. That's why they are not worth it, considering only relation between year and price(things like location an etc. are not included in analysis).

8. What percentage of houses were built this year?

```
new = res.query('Year == 2023')
ans = len(new)*100/len(res)
ans
```

24.682957005876894

A:Firstly we take only data with year 2023, then check how many apartments are there and divide it by total amount of them. So, lmost 25% of houses were built in 2023

9. What percentage of houses are not finished yet?

```
# What percentage of houses are not finished yet?
new = res.query('Year > 2023')
ans = len(new)*100/len(res)
ans|

21.497061552737396
```

A: Same as in Q8, we use query to filter out data with year greater than current(2023), which means they are still not ready. Then divide its amount to total to get our ratio. As we can see, 21% of all apartments in those websites are not finished yet.

10. What are most expensive and the cheapest houses in each city?

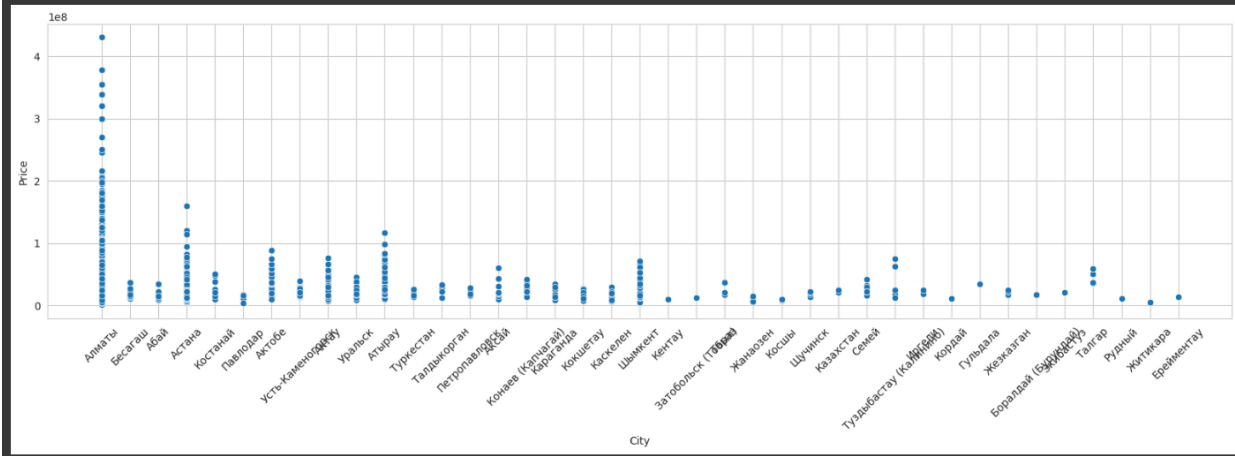
```
# What are most expensive and the cheapest houses in each city?
cmax = res.groupby('City')['Price'].apply(np.max).reset_index(name = 'MaxPrice')
cmin = res.groupby('City')['Price'].apply(np.min).reset_index(name = 'MinPrice')
cit = pd.merge(cmin,cmax,on='City')
cit|
```

	City	MinPrice	MaxPrice
0	Абай	9100000	35350000
1	Ақсай	10706500	60760000
2	Ақтау	7626000	76752400
3	Ақтобе	9779600	89150000
4	Алматы	1130000	430000000
5	Астана	5568000	159999999
6	Атырау	10437500	117483135
7	Бесағаш	11685000	37400000
8	Боралдай (Бурундай)	18000000	18000000
9	Гүлдала	35000000	35000000
10	Ерейментау	14000000	14000000
11	Жанаозен	7000000	14868000
12	Жезқазған	18000000	24500000
13	Жітіқара	5500000	5500000
14	Затобольск (Тобыл)	12500000	12500000

14	Затобольск (Тобыл)	12500000	12500000
15	Иргели	18800000	25000000
16	Казахстан	21386000	25466000
17	Караганда	9500000	34560000
18	Каскелен	6330000	30000000
19	Кентау	10500000	10500000
20	Кокшетау	8400000	27660000
21	Конаев (Капчагай)	13708200	42550000
22	Кордай	12000000	12000000
23	Костанай	10800000	50586800
24	Косшы	8415000	10500000
25	Павлодар	4700000	18000000
26	Петропавловск	16500000	29000000
27	Рудный	11700000	11700000
28	Семей	16000000	42173600
29	Талгар	36000000	59020000
30	Талдыкорган	12500000	33052800
31	Тараз	17500000	36914400
32	Туздыбастау (Калинино)	12287300	75825000
33	Туркестан	14333200	26283600
34	Уральск	9500000	46000000

35	Усть-Каменогорск	14022000	39692800
36	Шымкент	5000000	72000000
37	Щучинск	14001000	22545900
38	Экибастуз	21000000	21000000

```
import matplotlib.pyplot as plt
import seaborn as sns
# cities = res.groupby('City')['Price'].apply(np.mean).reset_index(name = 'Price')
sns.set_style('whitegrid')
sns.scatterplot(data = res, x = 'City', y = 'Price')
plt.xticks(rotation = 45)
plt.gcf().set_size_inches([20,5])
```



So we have to group all data by cities and get max and min prices for house in each city. First table shows us maximum and minimum prices, and in the second table we can visualize it and see range of prices in each city.