

Достану данные

```
In [1]: import pandas as pd
import numpy as np

f = "Задача 4 про географию и цены на недвижимость.xlsx"

a = pd.read_excel(f, 'Квартиры по городам')
c = pd.read_excel(f, 'Города')
m = pd.read_excel(f, 'Матрица времени пути', header=None)
```

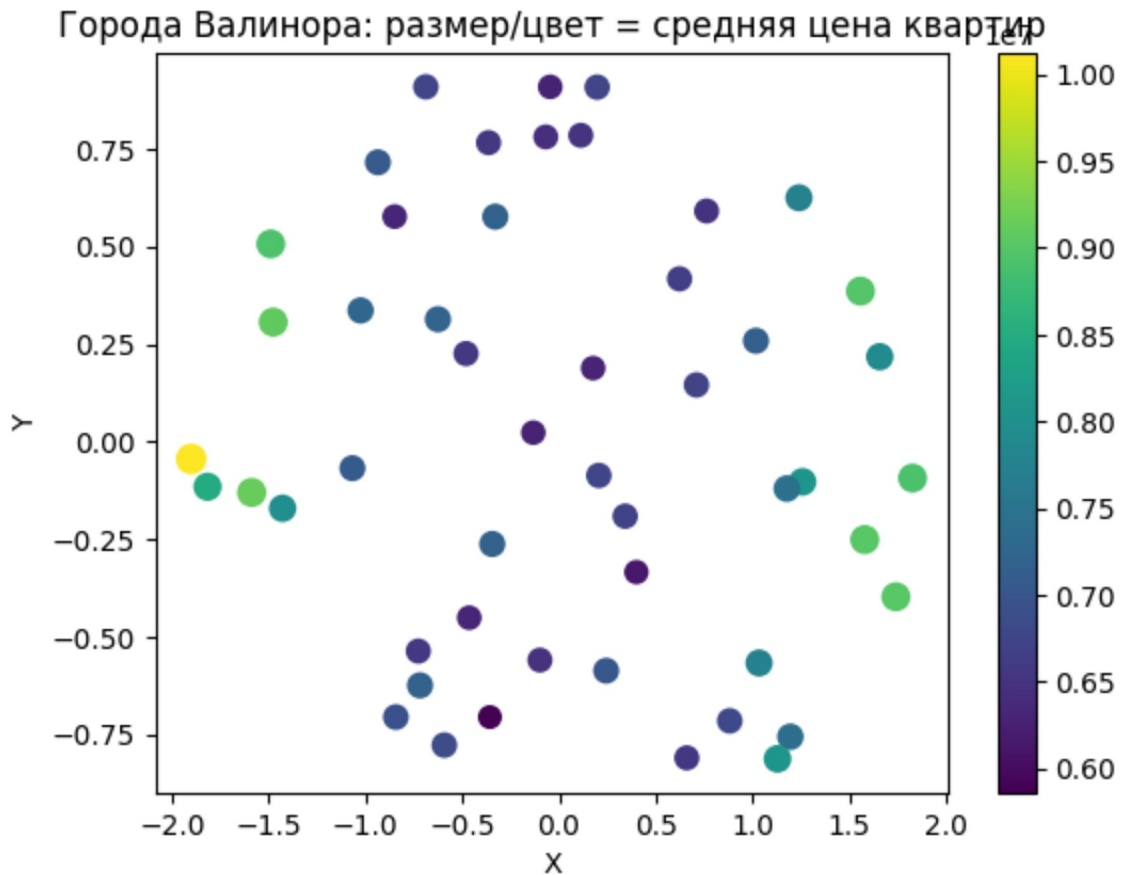
1. Средняя цена по городам и график для визуализации

на плоскости X–Y рисуем точки-города (x, y), а среднюю цену квартиры показываем размером и цветом точки

```
In [5]: g = a.groupby('city_id', as_index=False)['price'].mean()
g = g.merge(c, on='city_id')

import matplotlib.pyplot as plt

plt.scatter(g['x'], g['y'], s=g['price']/100000, c=g['price'])
plt.xlabel('X')
plt.ylabel('Y')
plt.title('Города Валинора: размер/цвет = средняя цена квартир')
plt.colorbar()
plt.show()
```



2. Корреляция: расстояние от Старого города и средняя цена

```
In [6]: o = c.loc[c['city_name']=='Старый Город','city_id'].iloc[0]
cx = c.set_index('city_id')
x0 = cx.loc[o,'x']
y0 = cx.loc[o,'y']

g['dist'] = np.sqrt((g['x']-x0)**2 + (g['y']-y0)**2)

b = g[g['city_id']!=o]
r2 = np.corrcoef(b['dist'], b['price'])[0,1]
print(r2)
```

0.11913752033518694

То есть слабая положительная корреляция: по всей стране в среднем более дальние (по прямой) от Старого города города чуть дороже, но связь очень слабая

3. Корреляция: расстояние от Старого города и цена для городов слева

```
In [7]: b2 = g[(g['city_id']!=o) & (g['x']<0)]
r3 = np.corrcoef(b2['dist'], b2['price'])[0,1]
print(r3, len(b2))
```

-0.8405102319050122 25

Здесь уже сильная отрицательная корреляция: чем дальше город от Старого города (по карте), тем дешевле там в среднем жильё

#### 4. Корреляция: время пути от Старого города и цена

```
In [8]: t = m.iloc[o-1,:].values
g['time'] = g['city_id'].apply(lambda z: t[z-1])

b3 = g[g['city_id']!=o]
r4 = np.corrcoef(b3['time'], b3['price'])[0,1]
print(r4)
```

-0.8529437785868481

Сильная отрицательная корреляция: города, до которых по времени добраться ближе к Старому городу, имеют более высокие средние цены на квартиры

Ответы на кач вопросы(5)

A. В городах, которые по реальной доступности (по времени пути) ближе к Старому городу, жильё заметно дороже. Простое расстояние по карте даёт слабую связь по всей стране, но сильную внутри его половины ( $X < 0$ ).

B. Старый город выступает главным экономическим центром: рядом больше работы, услуг, инфраструктуры, поэтому люди готовы переплачивать за жильё там и в городах, откуда до него удобно добираться

C. Разные меры дистанции по-разному учитывают реальность: геометрия не знает про горы и дороги; ограничение  $X < 0$  убирает барьер гор; время пути учитывает и горы, и сеть дорог, и телепорт, поэтому даёт самую чистую зависимость

D. Ещё один разумный вариант дистанции: величина, обратная пассажиропотоку между городами, например  $D(i,j) = 1 / F(i,j)$ , где  $F(i,j)$  — среднее число поездок между городами. Часто ездят друг к другу  $\rightarrow F$  большое  $\rightarrow D$  малая  $\rightarrow$  города близки в социально-экономическом смысле.