

Университет ИТМО

## Практическая работа №2

по дисциплине «Визуализация и моделирование»

**Автор:** Бичук Екатерина Дмитриевна

**Поток:** ВИМ 1.2

**Группа:** K3221

**Факультет:** ИКТ

**Преподаватель:** Чернышева А.В.

Санкт-Петербург, 2021 г.

В датасете хранятся данные об объявлениях о сдаче жилья в Нью-Йорке с сайта airbnb за 2019 год. Также информация об самом жилье и арендодателях.

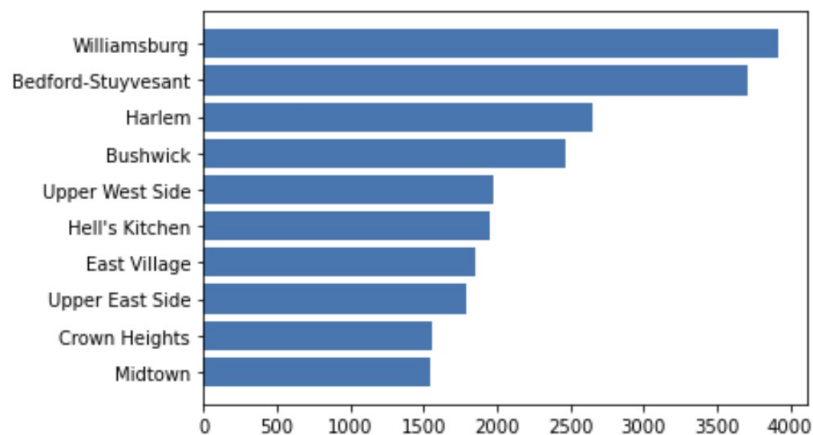
Таблица 1: Информация хранящаяся в столбцах

Название столбца в датасете	Данные, хранящиеся в столбце	Тип данных	Шкала
id	Индивидуальный номер объявления	int	Номинальная
name	Название объявления	string	Номинальная
host_id	Индивидуальный номер арендодателя	int	Номинальная
host_name	Имя арендодателя	string	Номинальная
neighbourhood_group	Боро( единица административного деления Нью-Йорка)	string	Номинальная
neighbourhood	Название района	string	Номинальная
latitude	Координаты широты	point	Номинальная
longitude	Координаты долготы	point	Номинальная
room_type	Тип жилья	string	Номинальная
price	Цена за сутки в долларах	int	Относительная
minimum_nights	Минимальное количество ночей	int	Относительная
number_of_reviews	Количество оценок на объявлении	int	Относительная
last_review	Дата последней оценки	data	Интервальная
reviews_per_month	Количество оценок в месяц	double	Относительная
calculated_host_list	Количество объявлений у одного арендодателя	int	Относительная
availability_365	Количество дней в году для съема жилья	int	Относительная

Описательная статистика и визуализация:

1. Какие районы чаще всего встречаются в датасете?

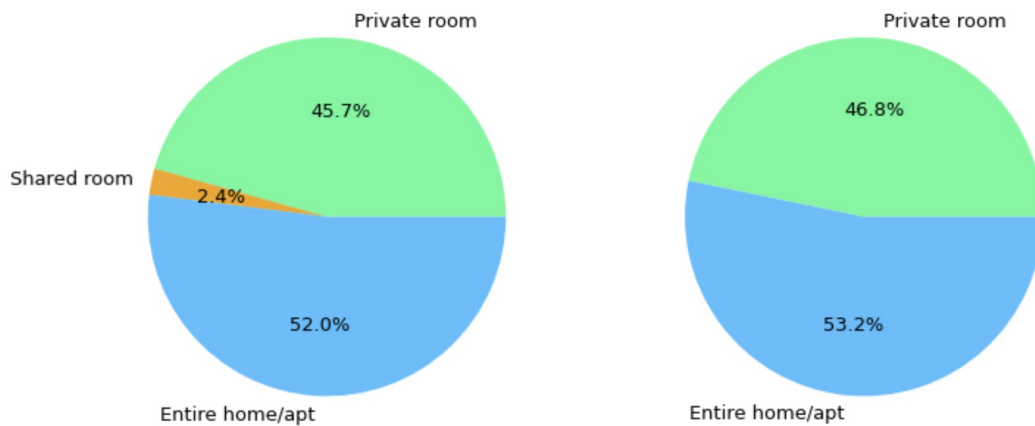
Рис. 1: Районы с наибольшим количеством объявлений



Получилось, что всего у 11 из 221 районов больше 1500 объявлений.

## 2. Какой тип жилья преобладает в целом по городу?

Рис. 2: Сравнение количества объявлений с определенным типом жилья



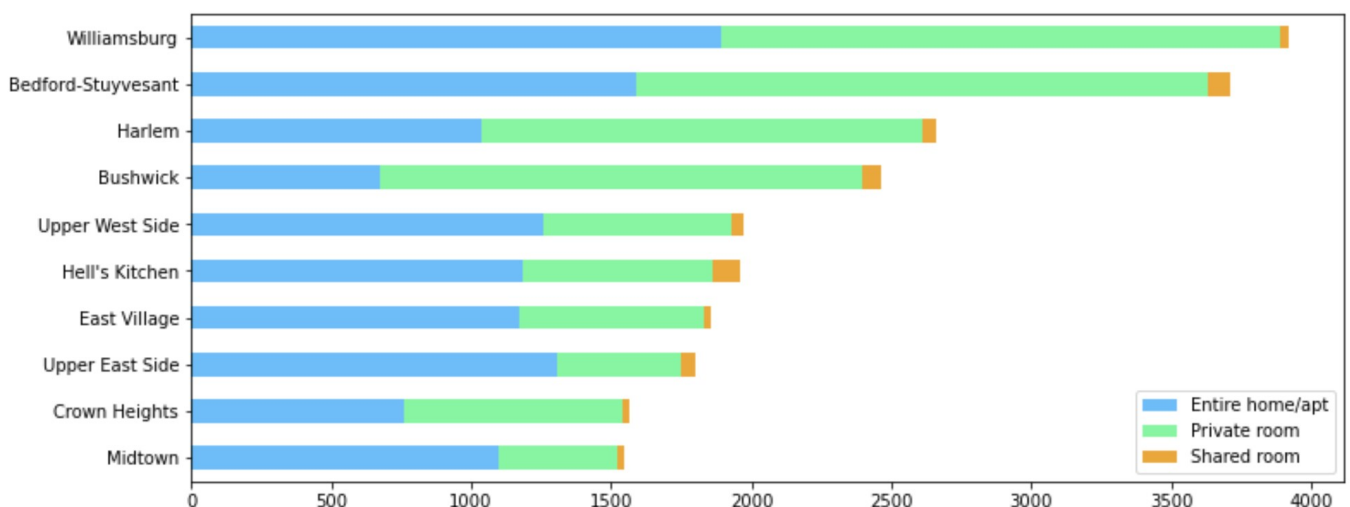
```
# тип жилья по всем районам
room_type_dt = {room_type: df["room_type"].to_list().count(room_type) for room_type in set(df["room_type"])}
print(room_type_dt)
# получилось три варианта. Построим две круговые диаграммы.

{'Private room': 22326, 'Shared room': 1160, 'Entire home/apt': 25409}
```

Представлены всего три типа жилья: апартаменты, отдельная и общая комната. Сравнивая диаграммы, можно сделать вывод, что распределение не сильно меняется, если убрать "shared room".

## 3. Остается ли такое распределение по типу жилья в районах с большим количеством объявлений?

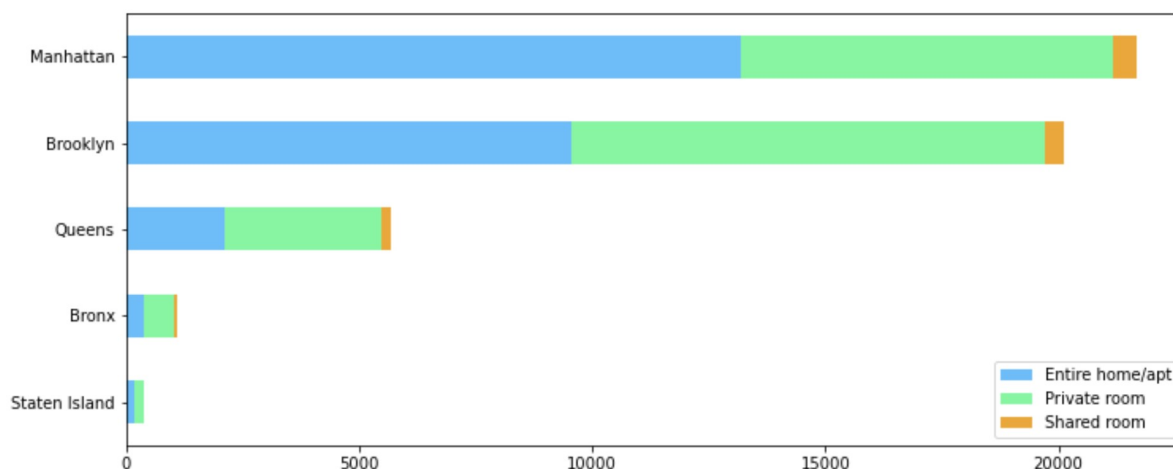
Рис. 3: Распределение объявлений по типу жилья в найденных ранее районах



В районах, где меньше 2000, но больше 1500 объявлений распределение остается, так как тип жилья "entire home/apt" занимает около половины от всех объявлений. Однако, в районах свыше 2000 объявлений преобладает "private room" отдельная комната.

4. Остается ли такое распределение по типу жилья в разных боро Нью-Йорка?

Рис. 4: Распределение объявлений по типу жилья в разных боро

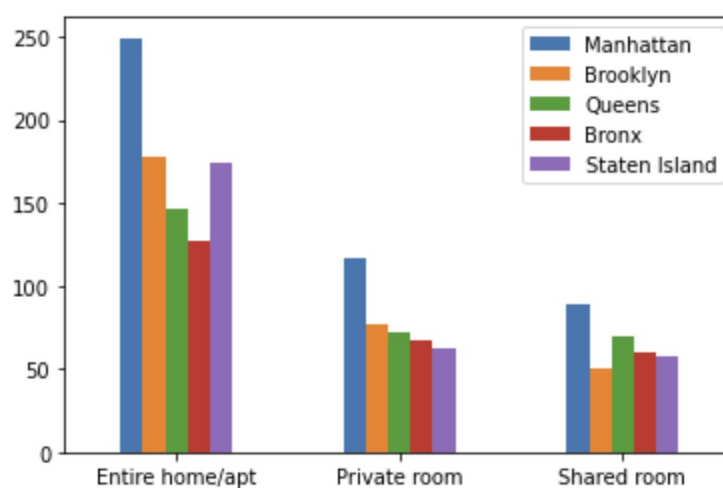


	Region	Total	Entire home/apt	Private room	Shared room
0	Staten Island	373	176	188	9
1	Bronx	1091	379	652	60
2	Queens	5666	2096	3372	198
3	Brooklyn	20104	9559	10132	413
4	Manhattan	21661	13199	7982	480

Получилось, что распределение из пункта 2, соблюдается только в Манхэттене. В остальных боро преобладают объявления "private room".

5. В каком боро средняя цена самая высокая?

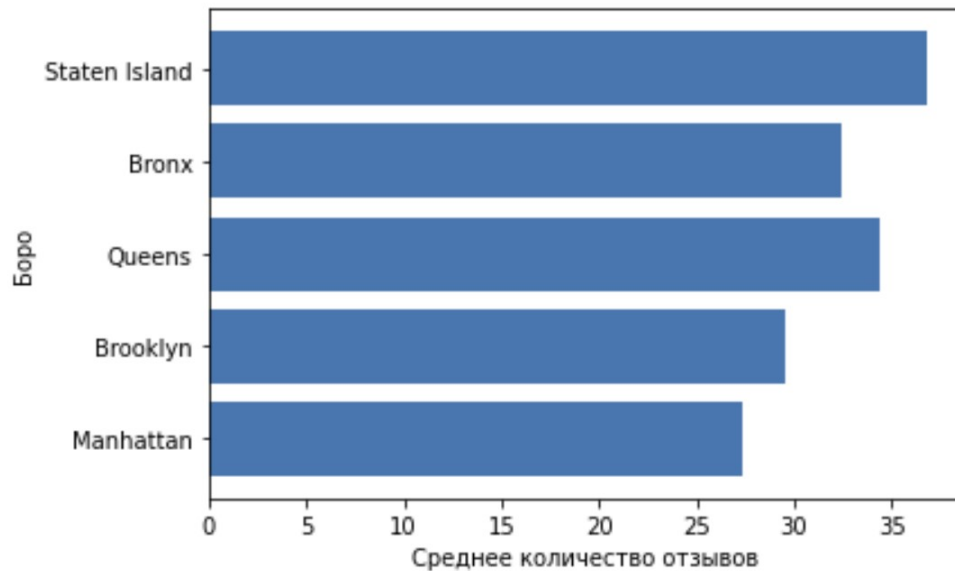
Рис. 5: Цена разного типа жилья во всех боро



Из гистограммы видно, что в Манхэттене средняя цена на много выше чем в других боро.

6. В каком боро в среднем, не учитывая объявления без оценок, оставляют больше отзывов?

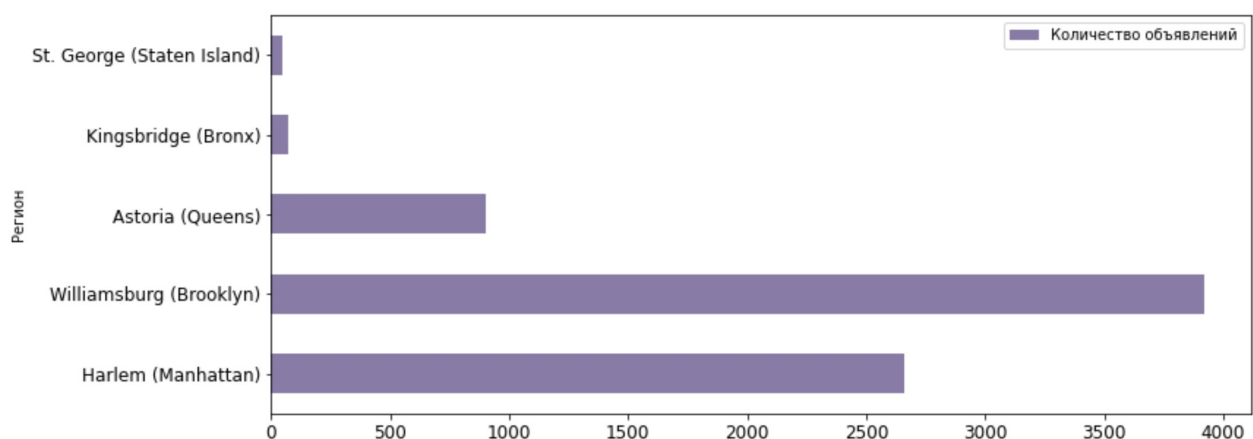
Рис. 6: Среднее количество оценок в боро



Из гистограммы видно, что в самом дорогом боро, Манхэттене, оставляют в среднем меньше оценок, чем во всех остальных. Больше всего оценивают в Статен Айленде.

7. Визуализируем наиболее популярные для сдачи районы в каждом из боро

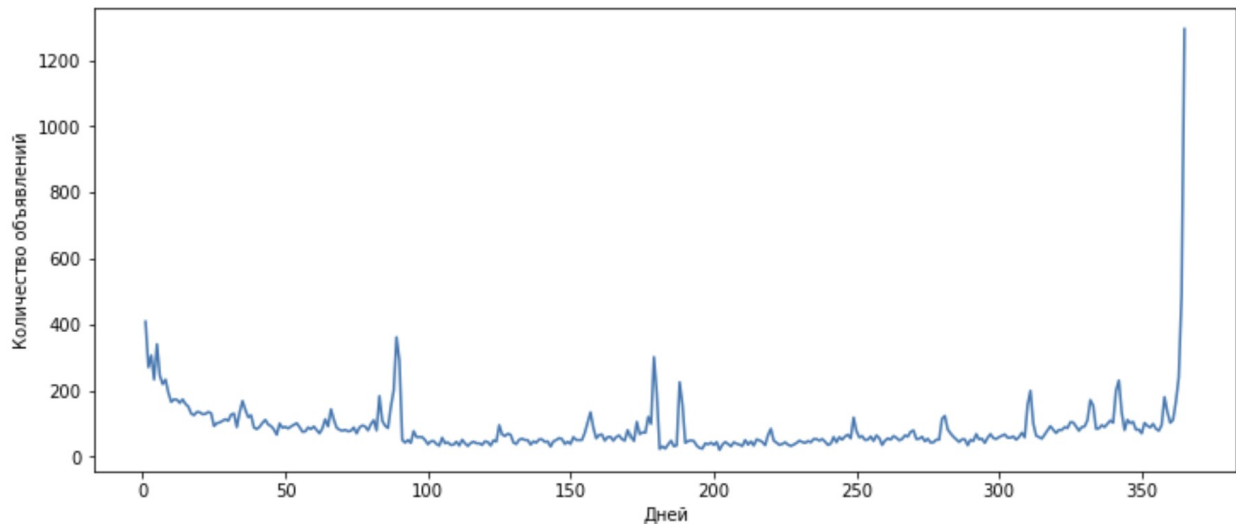
Рис. 7: Популярные районы для сдачи в каждом боро



Из гистограммы можно сделать вывод, что наиболее популярный район в Бруклине обходит по количеству объявлений остальные популярные районы из других боро. Хотя общее количество объявлений в Манхэттене больше, чем в Бруклине.

8. Визуализируем доступность объявлений для съема за год

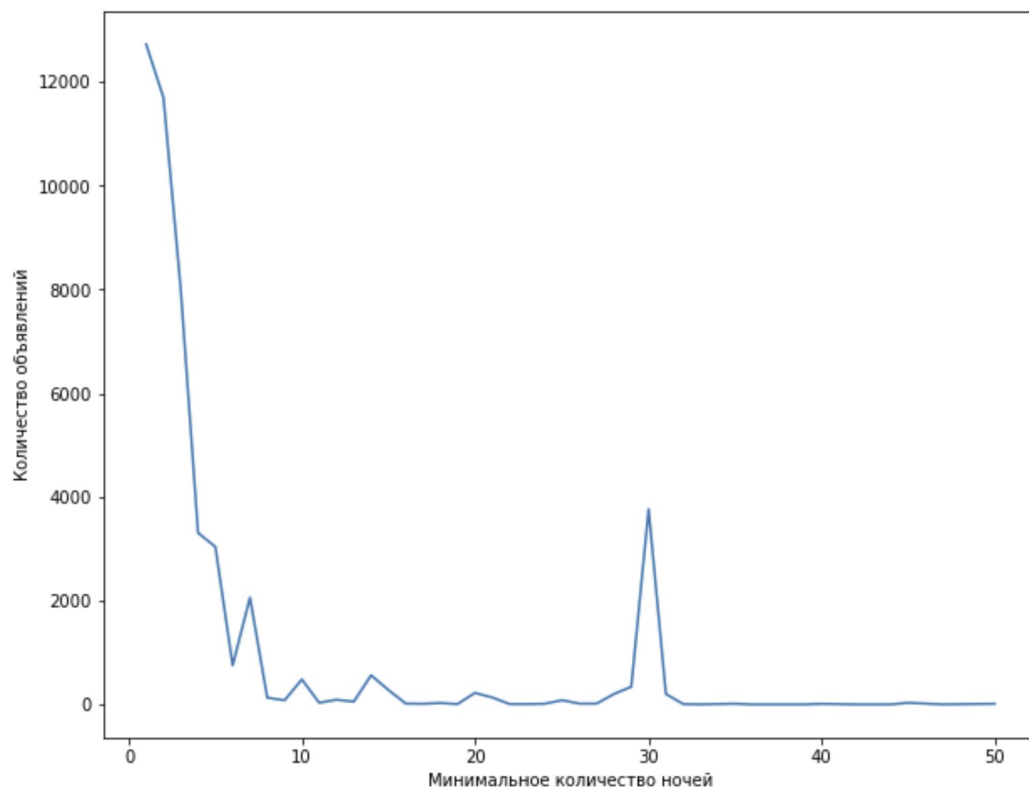
Рис. 8: Количество дней для съема жилья в году



Из графика видно что более 1200 объявлений доступны целый год, также видны скачки на 92 и 182-183 днях, т.е. 3 месяцев и полгода.

9. Есть ли какой-то разброс в параметре минимальное количество ночей?

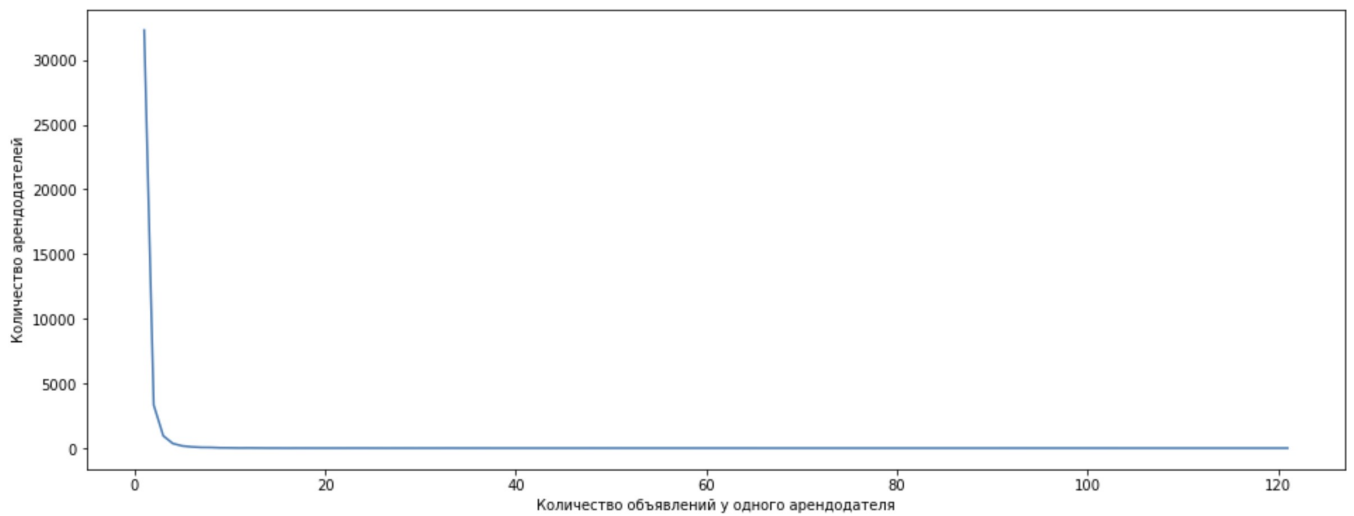
Рис. 9: Минимальное количество ночей



Из графика видно, что от 1 до 3 минимальных ночей имеют более 12000 объявлений. Также скачек виден на 7 и 31 ночи, т.е. преимущественно объявления имеют минимальный съем жилья от одного до трех дней, неделю и 1 месяц.

10. Визуализируем, сколько арендодателей имеют одинаковое количество объявлений.

```
fig, ax = plt.subplots()
ax.plot( regions1, cases1)
ax.set_ylabel('Количество арендодателей')
ax.set_xlabel('Количество объявлений у одного арендодателя')
fig.set_figwidth(16)
fig.set_figheight(6)
```



Из графика видно, что большинство арендодателей имеют по одному объявлению. Также на графике можно заметить, что самое большое количество объявлений 121 у одного арендодателя.