

Исследование рынка заведений общественного питания Москвы

Инвесторы из фонда «Shut Up and Take My Money» решили попробовать себя в новой области и открыть заведение общественного питания в Москве. Заказчики ещё не знают, что это будет за место: кафе, ресторан, пиццерия, паб или бар, — и какими будут расположение, меню и цены.

Задача аналитика — подготовить исследование рынка Москвы, найти интересные особенности и презентовать полученные результаты, которые в будущем помогут в выборе подходящего инвесторам места.

Необходимо сделать презентацию информативной и лаконичной. Её структура и оформление сильно влияют на восприятие информации читателями вашего исследования. Выбирать инструменты (matplotlib, seaborn и другие) и типы визуализаций рекомендуется самостоятельно.

В распоряжении датасет с заведениями общественного питания Москвы, составленный на основе данных сервисов Яндекс Карты и Яндекс Бизнес на лето 2022 года. Информация, размещённая в сервисе Яндекс Бизнес, могла быть добавлена пользователями или найдена в общедоступных источниках. Она носит исключительно справочный характер.

Цель исследования:

1. Найти особенности рынка заведений общественного питания Москвы
2. Презентовать результаты исследования, которые помогут в выборе подходящего инвесторам места.

Ход исследования:

1. Обзор данных.
2. Предобработка данных.
3. Анализ данных.
4. Детализация исследования: открытие кофейни.
5. Общий вывод. Подготовка презентации.

Описание данных:

Файл `moscow_places.csv`:

- `name` — название заведения;
- `address` — адрес заведения;
- `category` — категория заведения, например «кафе», «пиццерия» или «кофейня»;
- `hours` — информация о днях и часах работы;
- `lat` — широта географической точки, в которой находится заведение;
- `long` — долгота географической точки, в которой находится заведение;
- `rating` — рейтинг заведения по оценкам пользователей в Яндекс Картах (высшая оценка — 5.0);
- `price` — категория цен в заведении, например «средние», «ниже среднего», «выше среднего» и так далее;
- `avg_bill` — строка, которая хранит среднюю стоимость заказа в виде диапазона, например:
 - «Средний счёт: 1000–1500 ₽»;
 - «Цена чашки капучино: 130–220 ₽»;
 - «Цена бокала пива: 400–600 ₽».
 - и так далее;

- `middle_avg_bill` — число с оценкой среднего чека, которое указано только для значений из столбца `avg_bill`, начинающихся с подстроки «Средний счёт»:
 - Если в строке указан ценовой диапазон из двух значений, в столбец войдёт медиана этих двух значений.
 - Если в строке указано одно число — цена без диапазона, то в столбец войдёт это число.
 - Если значения нет или оно не начинается с подстроки «Средний счёт», то в столбец ничего не войдёт.
- `middle_coffee_cup` — число с оценкой одной чашки капучино, которое указано только для значений из столбца `avg_bill`, начинающихся с подстроки «Цена одной чашки капучино»:
 - Если в строке указан ценовой диапазон из двух значений, в столбец войдёт медиана этих двух значений.
 - Если в строке указано одно число — цена без диапазона, то в столбец войдёт это число.
 - Если значения нет или оно не начинается с подстроки «Цена одной чашки капучино», то в столбец ничего не войдёт.
- `chain` — число, выраженное 0 или 1, которое показывает, является ли заведение сетевым (для маленьких сетей могут встречаться ошибки):
 - 0 — заведение не является сетевым
 - 1 — заведение является сетевым
- `district` — административный район, в котором находится заведение, например Центральный административный округ;
- `seats` — количество посадочных мест.

Обзор данных.

1. Откроем файл и изучим общую информацию.

```
In [1]: #Импортируем библиотеки
```

```
In [2]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import json
import folium
from folium import Map, Marker
from folium.plugins import MarkerCluster
from folium import Map, Choropleth
import plotly.express as px
from plotly import graph_objects as go
import warnings
warnings.filterwarnings('ignore')
```

```
In [3]: from matplotlib.axes import Axes
```

```
In [4]: #Прочитаем файл и сохраним его в переменную data
```

```
In [5]: try:
    data = pd.read_csv('/datasets/moscow_places.csv')
except:
    data = pd.read_csv('https://code.s3.yandex.net/datasets/moscow_places.csv')
```

```
In [6]: #Получим общую информацию о датасете, выведем первые 5 строк
```

```
In [7]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8406 entries, 0 to 8405
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   name              8406 non-null    object  
 1   category          8406 non-null    object  
 2   address            8406 non-null    object  
 3   district           8406 non-null    object  
 4   hours              7870 non-null    object  
 5   lat                8406 non-null    float64 
 6   lng                8406 non-null    float64 
 7   rating             8406 non-null    float64 
 8   price              3315 non-null    object  
 9   avg_bill           3816 non-null    object  
 10  middle_avg_bill   3149 non-null    float64 
 11  middle_coffee_cup 535 non-null     float64 
 12  chain              8406 non-null    int64  
 13  seats              4795 non-null    float64 
dtypes: float64(6), int64(1), object(7)
memory usage: 919.5+ KB
```

```
In [8]: data['district'].unique()
```

```
Out[8]: array(['Северный административный округ',
 'Северо-Восточный административный округ',
 'Северо-Западный административный округ',
 'Западный административный округ',
 'Центральный административный округ',
 'Восточный административный округ',
 'Юго-Восточный административный округ',
 'Южный административный округ',
 'Юго-Западный административный округ'], dtype=object)
```

```
In [9]: data.head()
```

```
Out[9]:   name  category      address  district        hours      lat      lng  rating  price  avg_bill  middle_avg_bill  middle_coffee_cup  chain  seats
 0   WoWфли    кафе    Москва, улица Дыбенко, 7/1  Северный административный округ  ежедневно, 10:00–22:00  55.878494  37.478860  5.0    NaN      NaN      NaN      NaN      0      NaN
 1  Четыре комнаты  ресторан  Москва, улица Дыбенко, 36, корп. 1  Северный административный округ  ежедневно, 10:00–22:00  55.875801  37.484479  4.5  выше среднего  Средний счёт:1500–1600 ₽  1550.0      NaN      0      4.0
 2      Хазри    кафе    Москва, Клязьминская улица, 15  Северный административный округ  пн-чт 11:00–02:00; пт,сб 11:00–05:00; вс 11:00...  55.889146  37.525901  4.6  средние  Средний счёт:от 1000 ₽  1000.0      NaN      0      45.0
 3  Dormouse Coffee Shop  кофейня  Москва, улица Маршала Федоренко, 12  Северный административный округ  ежедневно, 09:00–22:00  55.881608  37.488860  5.0    NaN  Цена чашки капучино:155–185 ₽  NaN      170.0      0      NaN
 4    Иль Марко  пиццерия  Москва, Правобережная улица, 1Б  Северный административный округ  ежедневно, 10:00–22:00  55.881166  37.449357  5.0  средние  Средний счёт:400–600 ₽  500.0      NaN      1     148.0
```

```
In [10]: data.describe()
```

Out[10]:

	lat	lng	rating	middle_avg_bill	middle_coffee_cup	chain	seats
count	8406.000000	8406.000000	8406.000000	3149.000000	535.000000	8406.000000	4795.000000
mean	55.750109	37.608570	4.229895	958.053668	174.721495	0.381275	108.421689
std	0.069658	0.098597	0.470348	1009.732845	88.951103	0.485729	122.833396
min	55.573942	37.355651	1.000000	0.000000	60.000000	0.000000	0.000000
25%	55.705155	37.538583	4.100000	375.000000	124.500000	0.000000	40.000000
50%	55.753425	37.605246	4.300000	750.000000	169.000000	0.000000	75.000000
75%	55.795041	37.664792	4.400000	1250.000000	225.000000	1.000000	140.000000
max	55.928943	37.874466	5.000000	35000.000000	1568.000000	1.000000	1288.000000

Вывод: в таблице 8406 строк и 14 столбцов. Некоторые столбцы содержат пропуски. В столбце `seats` тип данных `float64`, но логичнее чтобы тип данных был `int`, так как количество посадочных мест - целое число. Также в столбце `seats` минимальное количество посадочных мест составляет 0, это нелогично. В столбце `float64` медиана и среднее значительно отличаются, что означает выбросы в данных. В столбце `middle_avg_bill` минимальный средний чек составляет 0 руб, что выглядит неправдоподобно.

Предобработка данных.

In [11]: #Посчитаем количество пропусков в датафрейме

In [12]: `data.isna().sum()`

```
Out[12]: 
name          0
category      0
address       0
district      0
hours         536
lat           0
lng           0
rating        0
price         5091
avg_bill     4590
middle_avg_bill 5257
middle_coffee_cup 7871
chain         0
seats        3611
dtype: int64
```

In [13]: #Посчитаем количество пропусков в процентном соотношении

In [14]: `(data.isna().sum()/len(data)).round(4)*100`

```
Out[14]: name      0.00
category   0.00
address    0.00
district   0.00
hours      6.38
lat        0.00
lng        0.00
rating     0.00
price      60.56
avg_bill   54.60
middle_avg_bill 62.54
middle_coffee_cup 93.64
chain      0.00
seats      42.96
dtype: float64
```

Пропущены значения в столбцах `price`, `avg_bill`, `middle_avg_bill`, `middle_coffee_cup`, `seats`. Оставим пропуски как есть, поскольку любые методы заполнения будут не точны и исказят данные.
Пропуски в столбце `hours` могут означать, что заведение работает ежедневно и круглосуточно.

```
In [15]: #Проверим строки таблицы на наличие дубликатов:
```

```
In [16]: data.duplicated().sum()
```

```
Out[16]: 0
```

Явные убликаты отсутствуют.

```
In [17]: #Приведем столбцы с типом object к нижнему регистру.
```

```
In [18]: data['name'] = data['name'].str.lower()
```

```
In [19]: data['address'] = data['address'].str.lower()
```

```
In [20]: data['avg_bill'] = data['avg_bill'].str.lower()
```

```
In [21]: #Проверим данные на неявные дубликаты:
```

```
In [22]: data.duplicated(subset=['name', 'category', 'address']).sum()
```

```
Out[22]: 2
```

Найдено два неявных дубликата, посмотрим на эти строки поближе

```
In [23]: data.loc[data.duplicated(subset = ['name', 'category', 'address'], keep=False)]
```

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	
189	кафе	кафе	москва, парк ангарские пруды	Северный административный округ	ежедневно, 09:00–23:00	55.880327	37.530786	3.2	NaN	NaN	NaN	NaN	NaN	0	NaN
215	кафе	кафе	москва, парк ангарские пруды	Северный административный округ	ежедневно, 10:00–22:00	55.881438	37.531848	3.2	NaN	NaN	NaN	NaN	NaN	0	NaN
1430	more poke	ресторан	москва, волоколамское шоссе, 11, стр. 2	Северный административный округ	ежедневно, 09:00–21:00	55.806307	37.497566	4.2	NaN	NaN	NaN	NaN	NaN	0	188.0
1511	more poke	ресторан	москва, волоколамское шоссе, 11, стр. 2	Северный административный округ	пн-чт 09:00–18:00; пт,сб 09:00–21:00; вс 09:00...	55.806307	37.497566	4.2	NaN	NaN	NaN	NaN	NaN	1	188.0

Заведения с индексами 189 и 215 мы не будем относить к дубликатам, поскольку у них отличается геопозиция. Заведения с индексами 1430 и 1511 отнесем к дубликатам, поскольку у них все данные одинаковые, кроме часов работы. У

```
In [24]: #Удалим неявные дубликаты
```

```
In [25]: data = data.drop_duplicates(subset=['name', 'category', 'address', 'lat'], keep='first')
```

```
In [26]: #Создадим столбец street с названиями улиц из столбца с адресом
```

```
In [27]: def split_address (row):
    return row.split(',') [1]
```

```
In [28]: data['street_name'] = data['address'].apply(split_address)
```

```
In [29]: data.head()
```

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street_name
0	wowфли	кафе	москва, улица дыбенко, 7/1	Северный административный округ	ежедневно, 10:00–22:00	55.878494	37.478860	5.0	NaN	NaN	NaN	NaN	0	NaN	улица дыбенко
1	четыре комнаты	ресторан	москва, улица дыбенко, 36, корп. 1	Северный административный округ	ежедневно, 10:00–22:00	55.875801	37.484479	4.5	выше среднего	средний счёт:1500–1600 ₽	1550.0	NaN	0	4.0	улица дыбенко
2	хазри	кафе	москва, клязьминская улица, 15	Северный административный округ	пн-чт 11:00–02:00; пт,сб 11:00–05:00; вс 11:00...	55.889146	37.525901	4.6	средние	средний счёт:от 1000 ₽	1000.0	NaN	0	45.0	клязьминская улица
3	dormouse coffee shop	кофейня	москва, улица маршала федоренко, 12	Северный административный округ	ежедневно, 09:00–22:00	55.881608	37.488860	5.0	NaN	цена чашки капучино:155–185 ₽	NaN	170.0	0	NaN	улица маршала федоренко
4	иль марко	пиццерия	москва, правобережная улица, 16	Северный административный округ	ежедневно, 10:00–22:00	55.881166	37.449357	5.0	средние	средний счёт:400–600 ₽	500.0	NaN	1	148.0	правобережная улица

```
In [30]: #Создадим столбец is_24/7 с обозначением, что заведение работает ежедневно и круглосуточно (24/7):
```

```
#логическое значение True – если заведение работает ежедневно и круглосуточно;
```

```
#логическое значение False – в противоположном случае.
```

```
In [31]: def categorize(row):
    try:
        if 'ежедневно, круглосуточно' in row:
            return True
        else:
            return False
    except: pass
```

```
In [32]: data['is_24/7'] = data['hours'].apply(categorize)
```

```
In [33]: data.head()
```

Out[33]:	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street_name	is_24/7
0	wowфли	кафе	москва, улица дыбенко, 7/1	Северный административный округ	ежедневно, 10:00–22:00	55.878494	37.478860	5.0	NaN	NaN	NaN	NaN	0	NaN	улица дыбенко	False
1	четыре комнаты	ресторан	москва, улица дыбенко, 36, корп. 1	Северный административный округ	ежедневно, 10:00–22:00	55.875801	37.484479	4.5	выше среднего	средний счёт:1500–1600 ₽	1550.0	NaN	0	4.0	улица дыбенко	False
2	хазри	кафе	москва, клязьминская улица, 15	Северный административный округ	пн-чт 11:00–02:00; пт,сб 11:00–05:00; вс 11:00...	55.889146	37.525901	4.6	средние	средний счёт:от 1000 ₽	1000.0	NaN	0	45.0	клязьминская улица	False
3	dormouse coffee shop	кофейня	москва, улица маршала федоренко, 12	Северный административный округ	ежедневно, 09:00–22:00	55.881608	37.488860	5.0	NaN	цена чашки капучино:155–185 ₽	NaN	170.0	0	NaN	улица маршала федоренко	False
4	иль марко	пиццерия	москва, правобережная улица, 16	Северный административный округ	ежедневно, 10:00–22:00	55.881166	37.449357	5.0	средние	средний счёт:400–600 ₽	500.0	NaN	1	148.0	правобережная улица	False

Вывод: В ходе предобработки данных исследованы пропуски, дубликаты. Пропущены значения в столбцах price, avg_bill, middle_avg_bill, middle_coffee_cup, seats. Пропуски оставлены, поскольку любые методы заполнения будут не точны и исказят данные.

Анализ данных.

Исследование категорий объектов

Исследуем количество объектов общественного питания по категориям

```
In [34]: #Посмотрим на уникальные значения категорий
```

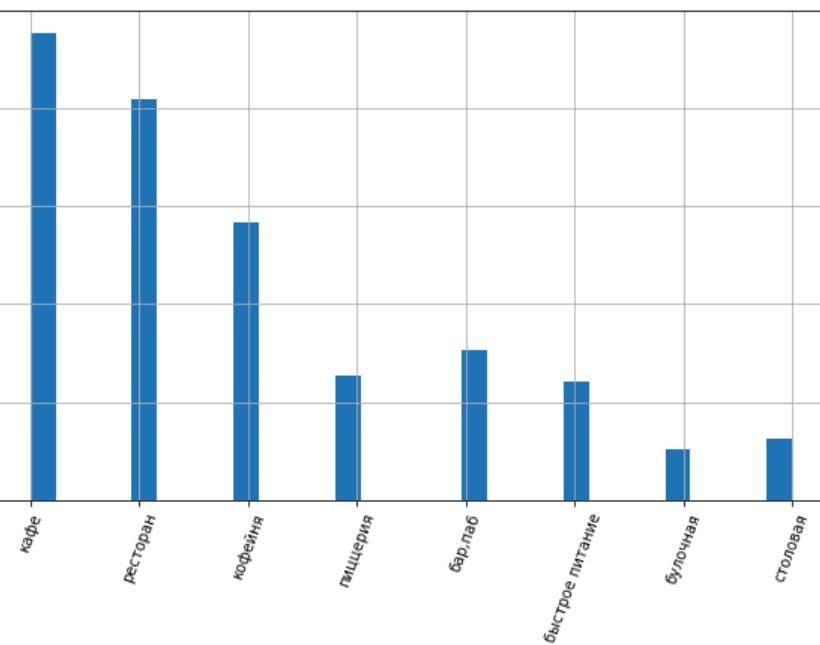
```
In [35]: data['category'].unique()
```

```
Out[35]: array(['кафе', 'ресторан', 'кофейня', 'пиццерия', 'бар,паб', 'быстрое питание', 'булочная', 'столовая'], dtype=object)
```

В данных представлены такие типы заведений, как кафе, ресторан, кофейня, пиццерия, бар,паб, быстрое питание, булочная, столовая.

```
In [36]: #Построим гистограмму по категориям заведений
```

```
In [37]: data['category'].hist(bins=30, figsize=(10,6))
plt.xticks(rotation=70)
#plt.figure(figsize=(16,10))
plt.show();
```



Чаще всего среди всех заведений общественного питания Москвы встречаются кафе, на втором месте рестораны, а менее популярные булочные и столовые. Посмотрим на численные данные поближе.

```
In [38]: category_count = data.groupby('category').agg({'name': 'count'}).sort_values(by='name', ascending=False).reset_index()
category_count
```

```
Out[38]:
```

category	name	count
кафе		2378
ресторан		2042
кофейня		1413
бар,паб		765
пиццерия		633
быстрое питание		603
столовая		315
булочная		256

```
In [39]: #Построим столбчатую диаграмму
```

```
In [40]: fig, ax = plt.subplots(figsize=(12, 6))
total = category_count['name'].sum()
width = 0.6 # Увеличиваем ширину столбцов
padding = 20 # Отступ между меткой и столбцом
bars = ax.bar(category_count['category'], category_count['name'], width=width)

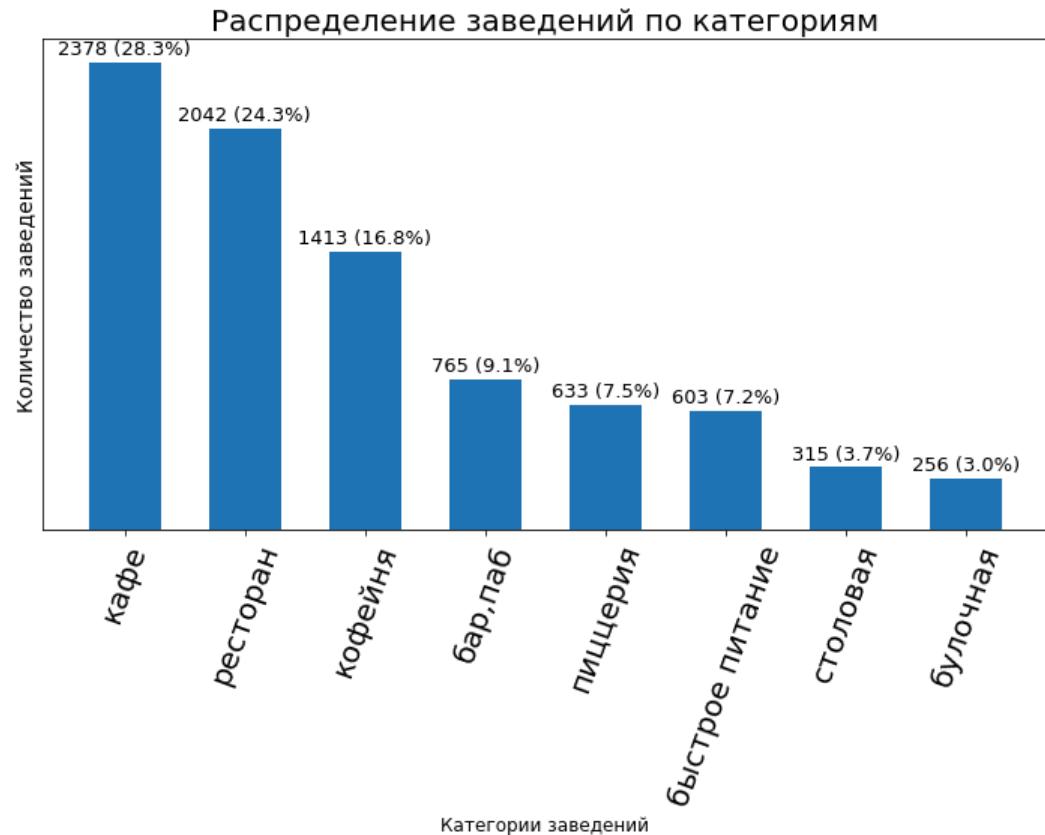
for bar in bars:
    height = bar.get_height()
    percentage = height / total * 100
    ax.text(bar.get_x() + bar.get_width() / 2,
            height + padding, f'{height} ({percentage:.1f}%)',
            ha='center',
```

```

va='bottom',
fontsize=13)
plt.xticks(fontsize=20)
plt.yticks([])
plt.xlabel('Категории заведений', fontsize=12)
plt.ylabel('Количество заведений', fontsize=14)
ax.set_xticklabels(category_count['category'], rotation=70)
plt.title('Распределение заведений по категориям', fontsize=20)

plt.show()

```



Вывод: самые распространенная категория заведений - это кафе. Кафе занимают долю в размере 28.3% от всех заведений общественного питания. С небольшим отрывом по популярности идут рестораны (24.3%). На третьем и четвертом месте находятся кофейни (16.8%) и пабы (9.1%). Наименее популярные столовые и булочные, доля которых составила 3.7% и 3%

Исследование количества посадочных мест по категориям объектов

In [41]: #Построим гистограмму столбца 'seats'

In [42]: data['seats'].hist(bins=100, figsize=(10,6));
plt.xlabel('Количество посадочных мест в заведении')
plt.ylabel('Количество заведений')
plt.title('Распределение количества посадочных мест для заведений Москвы');



```
In [43]: data['seats'].value_counts()
```

```
Out[43]:
```

seats	count
40.0	253
100.0	213
60.0	175
50.0	168
80.0	160
...	
600.0	1
268.0	1
61.0	1
236.0	1
33.0	1

Name: seats, Length: 229, dtype: int64

Мы видим, что в данных присутствуют заведения, с количеством посадочных мест, равных 0. Посмотрим, что это за строки, также посмотрим на их количество.

```
In [44]: data.loc[data['seats'] == 0]
```

Out[44]:		name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street_name	is_24/7
	84	meat doner kebab	булочная	москва, улица лескова, 22	Северо-Восточный административный округ	ежедневно, круглосуточно	55.896987	37.608126	4.5	NaN	средний счёт:300 ₽	300.0	NaN	0	0.0	улица лескова	True
	177	арамье	булочная	москва, улица 800-летия москвы, 22, корп. 2	Северный административный округ	ежедневно, 09:00–21:00	55.879392	37.556832	4.0	NaN	NaN	NaN	NaN	1	0.0	улица 800-летия москвы	False
	196	донер-шашлык	ресторан	москва, улица лескова, 22	Северо-Восточный административный округ	ежедневно, круглосуточно	55.896962	37.608300	4.5	NaN	NaN	NaN	NaN	0	0.0	улица лескова	True
	203	тандыр № 1	кафе	москва, улица лескова, 22г	Северо-Восточный административный округ	ежедневно, круглосуточно	55.895615	37.611049	4.0	NaN	NaN	NaN	NaN	1	0.0	улица лескова	True
	211	неаполитан пицца	кафе	москва, улица пришвина, 23	Северо-Восточный административный округ	ежедневно, 12:00–23:00	55.885416	37.604650	4.3	NaN	NaN	NaN	NaN	0	0.0	улица пришвина	False
	
	8118	бико	булочная	москва, улица симоновский вал, 13, стр. 3	Юго-Восточный административный округ	ежедневно, 08:00–20:00	55.723340	37.664252	1.3	NaN	NaN	NaN	NaN	0	0.0	улица симоновский вал	False
	8168	чайхана уч кудук	кофейня	москва, улица симоновский вал, 26, корп. 1	Южный административный округ	ежедневно, 09:00–00:00	55.718663	37.662887	4.3	NaN	NaN	NaN	NaN	0	0.0	улица симоновский вал	False
	8172	масса кофе	кофейня	москва, холодильный переулок, 2	Южный административный округ	пн–пт 08:00–20:00; сб, вс 09:00–20:00	55.709713	37.624388	4.3	средние	цена чашки капучино:100–210 ₽	NaN	155.0	0	0.0	холодильный переулок	False
	8336	сочная шаурма в кузьминках	быстрое питание	москва, волгоградский проспект, 102	Юго-Восточный административный округ	ежедневно, круглосуточно	55.703834	37.773831	3.9	NaN	средний счёт:120–130 ₽	125.0	NaN	0	0.0	волгоградский проспект	True
	8362	достор	кафе	москва, вешняковская улица, 43	Восточный административный округ	ежедневно, 09:00–00:00	55.716153	37.821948	4.1	средние	средний счёт:300–800 ₽	550.0	NaN	0	0.0	вешняковская улица	False

136 rows × 16 columns

136 заведений с нулевым количеством посадочных мест. По отфильтрованным данным очевидно, что некоторые типы заведения, такие, как булочная или кафе, точки быстрого питания могут работать без посадочных мест.

```
In [45]: #Посмотрим в процентном соотношении сколько таких заведений с количеством посадочных мест, #рабных 0
```

```
In [46]: len(data.loc[data['seats'] == 0])/data.shape[0]*100
```

Out[46]: 1.618084473527662

1.61% или 136 заведений с нулевыми значениями. 43% или 3611 заведений с пропусками. Предположим, это ошибочный ввод данных, или отсутствие данных для соответствующего заведения, либо булочные или точки быстрого питания.

Напишем функцию и разделим предприятия общественного питания по категориям в зависимости от количества посадочных мест:

- 0 п.м.
- 1 - 6 п.м.

- 7 - 20 п.м.
- 21 - 30 п.м.
- 31 - 60 п.м.
- 61 - 100 п.м.
- 101 - 200 п.м.
- 201 - 300 п.м.
- 301 и выше п.м.
- N/A

Для того, чтобы не потерялись данные с пропусками и в дальнейшем были отражены на графиках, заполним их заглушками со значением равным 2000 и проверим произведенные изменения.

```
In [47]: data['seats'] = data['seats'].fillna(2000)
```

```
In [48]: data.query('seats == 2000')
```

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street_name	is_24/7
0	wowфли	кафе	москва, улица дыбенко, 7/1	Северный административный округ	ежедневно, 10:00–22:00	55.878494	37.478860	5.0	NaN	NaN	NaN	NaN	0	2000.0	улица дыбенко	False
3	dormouse coffee shop	кофейня	москва, улица маршала федоренко, 12	Северный административный округ	ежедневно, 09:00–22:00	55.881608	37.488860	5.0	NaN	цена чашки капучино:155–185 ₽	NaN	170.0	0	2000.0	улица маршала федоренко	False
5	sergio pizza	пиццерия	москва, ижорская улица, вл86	Северный административный округ	ежедневно, 10:00–23:00	55.888010	37.509573	4.6	средние	NaN	NaN	NaN	0	2000.0	ижорская улица	False
11	шашлык шеф	кафе	москва, улица маршала федоренко, 10с1	Северный административный округ	ежедневно, 10:00–21:00	55.881770	37.492362	4.9	NaN	NaN	NaN	NaN	0	2000.0	улица маршала федоренко	False
12	заправка	кафе	москва, мкад, 80-й километр, 1	Северный административный округ	вт-сб 09:00–18:00	55.899938	37.517958	4.3	средние	средний счёт:330 ₽	330.0	NaN	0	2000.0	мкад	False
...
8387	pab&burg	бар,паб	михайлова, 22, корп. 4	Юго-Восточный административный округ	ежедневно, 12:00–21:30	55.726303	37.769842	4.2	средние	цена бокала пива:от 140 ₽	NaN	NaN	0	2000.0	улица михайлова	False
8389	assa	ресторан	москва, улица авиаконструктора миля, За	Юго-Восточный административный округ	пн-чт 09:00–22:00; пт,сб 09:00–23:00; вс 09:00...	55.686817	37.851225	4.9	NaN	NaN	NaN	NaN	0	2000.0	улица авиаконструктора миля	False
8392	касабланка	кафе	москва, большая косинская улица, 27	Восточный административный округ	пн-чт 08:00–17:00; пт 08:00–16:00	55.718514	37.857576	3.3	NaN	NaN	NaN	NaN	0	2000.0	большая косинская улица	False
8394	намангали	кафе	москва, ферганская улица, вл17-21	Юго-Восточный административный округ	ежедневно, круглосуточно	55.705332	37.819244	4.3	NaN	NaN	NaN	NaN	0	2000.0	ферганская улица	True
8395	истира запракфа	кафе	москва, юго-восточный административный округ, ...	Юго-Восточный административный округ	NaN	55.724686	37.710558	2.9	NaN	NaN	NaN	NaN	0	2000.0	юго-восточный административный округ	None

3611 rows x 16 columns

```
In [49]: #Напишем функцию, которая разделяет заведения на категории в зависимости от количества посадочных мест
```

```
In [50]: def categorize_seats (row):
```

```
    """
    Возвращает категории по количеству посадочных мест
    - '0 п.м.' если seats == 0;
    - '1 - 10 п.м.' если seats от 1 до 6;
    - '11 - 30 п.м.' если seats от 21 до 30;
    - '31 - 60 п.м.' если seats от 31 до 60;
    - '61 - 100 п.м.' если seats от 61 до 100;
    - '101 - 200 п.м.' если seats от 101 до 200;
    - '201 - 300 п.м.' если seats 201 до 300;
    - '301 и выше п.м.' если seats от 301 и выше;
    - 'N/A' если пропуски в seats.
    """

    if row == 0:
        return '0'
    if 1 <= row <= 10:
        return '1-10'
    if 11 <= row <= 30:
        return '11-30'
    if 31 <= row <= 60:
        return '31-60'
    if 61 <= row <= 100:
        return '61-100'
    if 101 <= row <= 200:
        return '101-200'
    if 201 <= row <= 300:
        return '201-300'
    if 1500 >= row >= 301:
        return '301 и выше'
    if row == 2000:
        return 'N/A'
```

```
In [51]: data['seats_category'] = data['seats'].apply(categorize_seats)
```

```
In [52]: data.head()
```

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street_name	is_24/7	seats_category
0	wowфли	кафе	москва, улица дыбенко, 7/1	Северный административный округ	ежедневно, 10:00-22:00	55.878494	37.478860	5.0	NaN	NaN	NaN	NaN	0	2000.0	улица дыбенко	False	N/A
1	четыре комнаты	ресторан	москва, улица дыбенко, 36, корп. 1	Северный административный округ	ежедневно, 10:00-22:00	55.875801	37.484479	4.5	выше среднего	средний счёт:1500-1600 ₽	1550.0	NaN	0	4.0	улица дыбенко	False	1-10
2	хазри	кафе	москва, клязьминская улица, 15	Северный административный округ	пн-чт 11:00-02:00; пт,сб 11:00-05:00; вс 11:00..	55.889146	37.525901	4.6	средние	средний счёт:от 1000 ₽	1000.0	NaN	0	45.0	клязьминская улица	False	31-60
3	dormouse coffee shop	кофейня	москва, улица маршала федоренко, 12	Северный административный округ	ежедневно, 09:00-22:00	55.881608	37.488860	5.0	NaN	цена чашки капучино:155-185 ₽	NaN	170.0	0	2000.0	улица маршала федоренко	False	N/A
4	ильмарко	пиццерия	москва, правобережная улица, 16	Северный административный округ	ежедневно, 10:00-22:00	55.881166	37.449357	5.0	средние	средний счёт:400-600 ₽	500.0	NaN	1	148.0	правобережная улица	False	101-200

```
In [53]: #сгруппируем данные по столбцам 'category', 'seats_category'  
#и посчитаем количество заведений для каждой группы
```

```
In [54]: seats_count = data.groupby(['category', 'seats_category']).agg(seats_count=('seats_category', 'count'))  
seats_count
```

```
Out[54]: seats_count
```

category	seats_category	seats_count
бар,паб	0	4
	1-10	4
	101-200	89
	11-30	38
	201-300	33

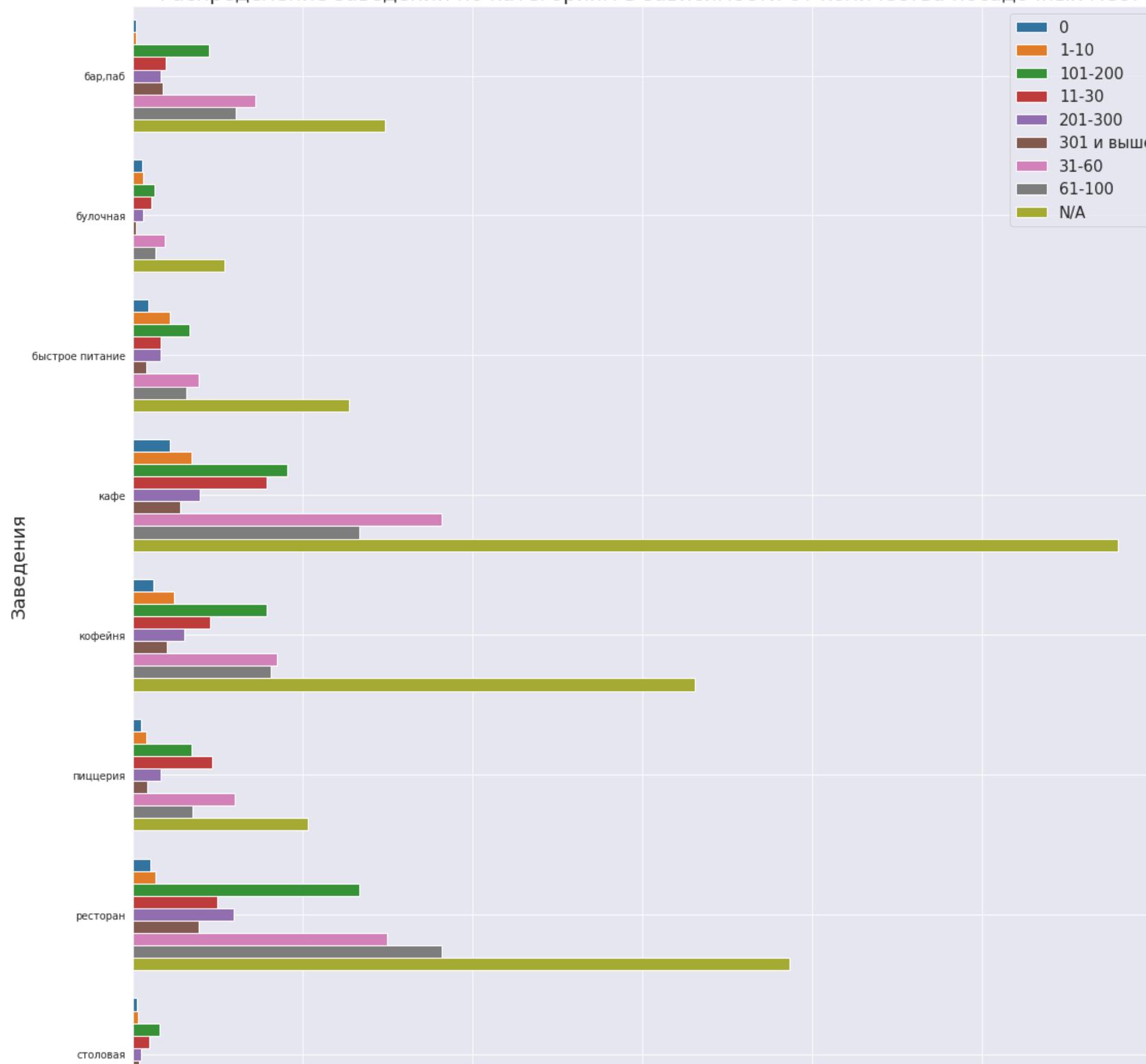
столовая	201-300	9
	301 и выше	7
	31-60	46
	61-100	41
	N/A	151

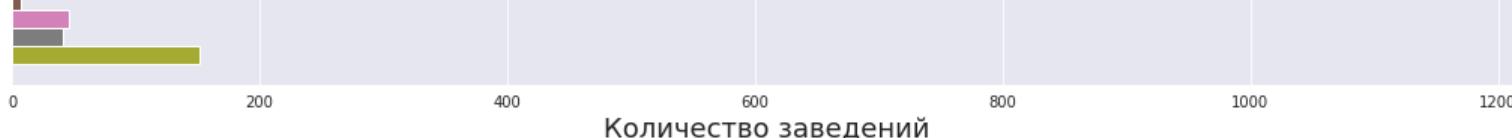
72 rows × 1 columns

```
In [55]: #визуализируем распределение заведений по категориям в зависимости от количества посадочных мест,  
#включая заведения с отсутствующими данными по количеству мест
```

```
In [56]: seats_category_na = data.groupby(['category', 'seats_category'], as_index = False).agg(seats_count=('seats_category', 'count'))  
sns.set_style('dark')  
plt.figure(figsize=(18, 20))  
  
sns.barplot(x='seats_count', y='category', data=seats_category_na, hue='seats_category')  
plt.title('Распределение заведений по категориям в зависимости от количества посадочных мест', fontsize=20)  
plt.xlabel('Количество заведений', fontsize=18)  
plt.ylabel('Заведения', fontsize=18)  
plt.legend(loc='upper right', fontsize=15)  
plt.grid()  
sns.set(font_scale=1)  
plt.show()
```

Распределение заведений по категориям в зависимости от количества посадочных мест





```
In [57]: #визуализируем распределение заведений по категориям в зависимости от количества посадочных мест,  
#не включая заведения с отсутствующими данными по количеству мест
```

```
In [58]: seats_category = data.query('seats_category!="N/A"').groupby(  
    ['category', 'seats_category'], as_index = False  
).agg(seats_count=('seats_category', 'count'))  
  
sns.set_style('dark')  
plt.figure(figsize=(20, 18))  
  
sns.barplot(x='seats_count', y='category', data=seats_category, hue='seats_category')  
plt.title(  
    'Распределение заведений по категориям в зависимости от кол-ва посадочных мест без учета отсутствующих данных', fontsize=20  
)  
plt.xlabel('Количество заведений', fontsize=18)  
plt.ylabel('Заведения', fontsize=18)  
plt.legend(loc='upper right', fontsize=15)  
plt.grid()  
sns.set(font_scale=1)  
plt.show()
```

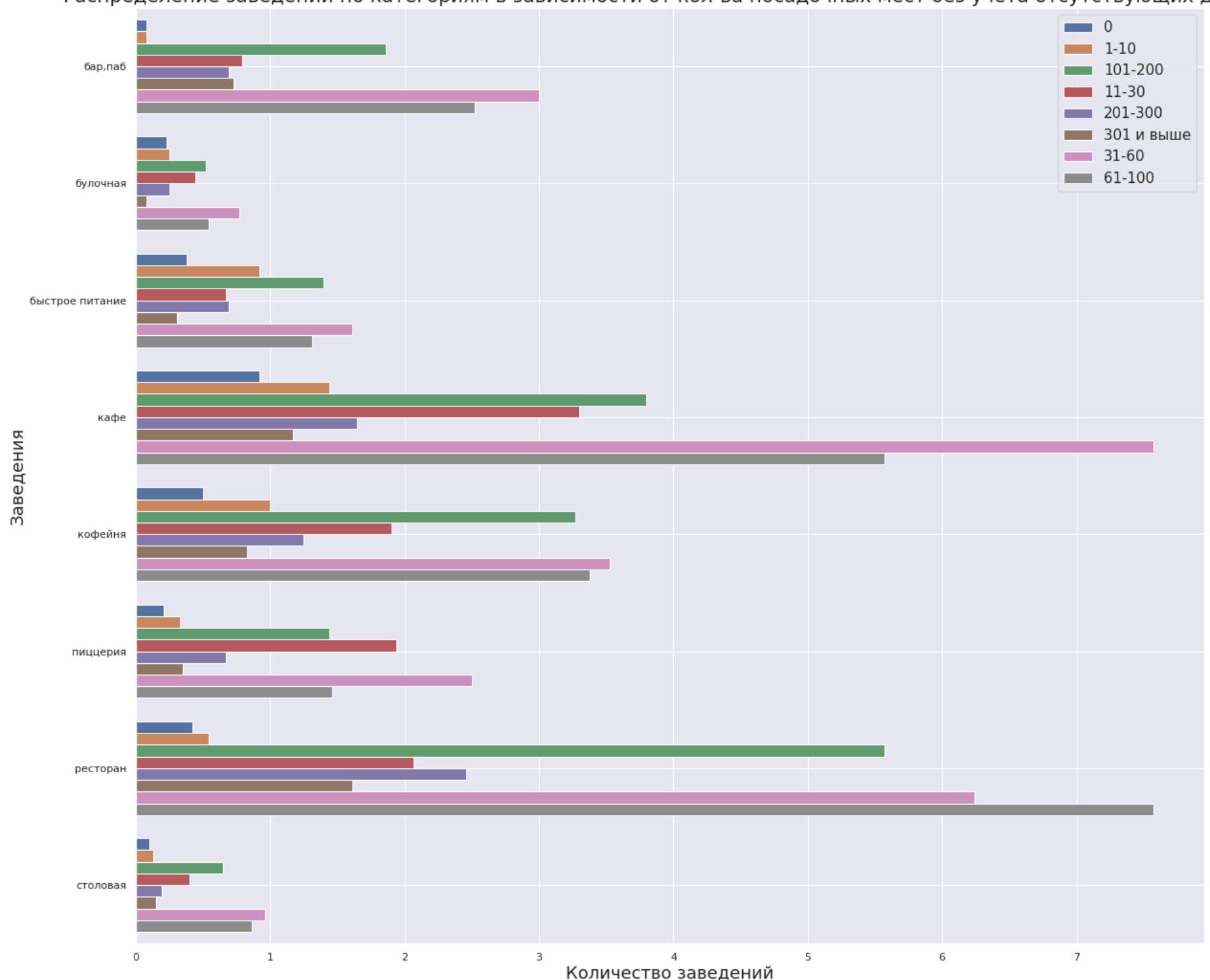
Распределение заведений по категориям в зависимости от кол-ва посадочных мест без учета отсутствующих данных



Визуализируем процентное соотношение заведений в зависимости от категорий по количеству посадочных мест.

```
In [59]: seats_category_perc = data.query('seats_category!="N/A"').groupby(['category', 'seats_category'], as_index = False).agg(seats_count=('seats_category', 'count'))  
seats_category_perc['perc'] = round(seats_category_perc['seats_count']/seats_category_perc['seats_count'].sum()*100, 2)  
  
sns.set_style('dark')  
plt.figure(figsize=(20, 18))  
  
sns.barplot(x='perc', y='category', data=seats_category_perc, hue='seats_category')  
plt.title(  
    'Распределение заведений по категориям в зависимости от кол-ва посадочных мест без учета отсутствующих данных', fontsize=20  
)  
plt.xlabel('Количество заведений', fontsize=18)  
plt.ylabel('Заведения', fontsize=18)  
plt.legend(loc='upper right', fontsize=15)  
plt.grid()  
sns.set(font_scale=1)  
plt.show()
```

Распределение заведений по категориям в зависимости от кол-ва посадочных мест без учета отсутствующих данных



Вывод: Для каждой категории заведений большую часть занимают заведения с отсутствующими данными по количеству посадочных мест. Наибольшая доля - 7.6% или 370 заведений приходится на рестораны с количеством посадочных мест от 61 до 100, и столько же - 7.6% или 370 заведений приходится на кафе с количеством посадочных мест от 31 до 60.

Также в Москве распространены кофейни примерно с равнозначной долей по количеству посадочных мест от 31 до 60 - 165 заведений или 3.5% от всех заведений города, от 61 до 100 - 160 заведений или 3.4% от всех заведений города, от 101 до 200 - 153 заведений от 3.2% от всех заведений города.

Исследование соотношения сетевых и несетевых заведений в Москве

```
In [60]: #Построим сводную таблицу, отражающую количество сетевых и несетевых заведений
```

```
In [61]: data_chain = data.pivot_table(index='chain', values='name', aggfunc='count').rename(index={0:'несетевые', 1:'сетевые'})  
data_chain
```

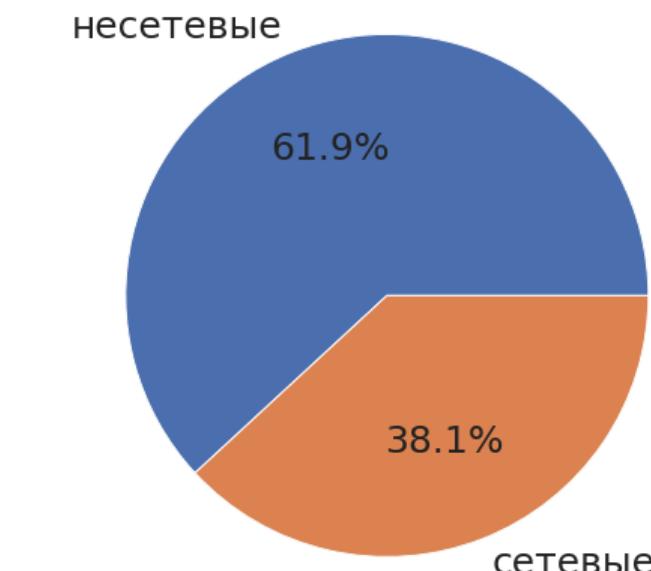
```
Out[61]:      name
```

chain	name
несетевые	5201
сетевые	3204

```
In [62]: #визуализируем соотношение сетевых и не сетевых заведений
```

```
In [63]: data_chain.plot(y='name', kind='pie', figsize=(8,8), autopct='%1.1f%%', textprops={'fontsize': 25});  
plt.ylabel('') # Удаляем подпись оси у  
plt.legend().remove()  
plt.title('Доли сетевых и несетевых заведений Москвы', fontsize=20) # Добавляем заголовок  
plt.show()
```

Доли сетевых и несетевых заведений Москвы



Вывод: преобладает доля несетевых заведений - 61.9%. Доля сетевых заведений составила 38.1%

Исследование сетевых заведений в разрезе категорий

Построим сводную таблицу, отражающую доли сетевых и несетевых заведений в разрезе категорий

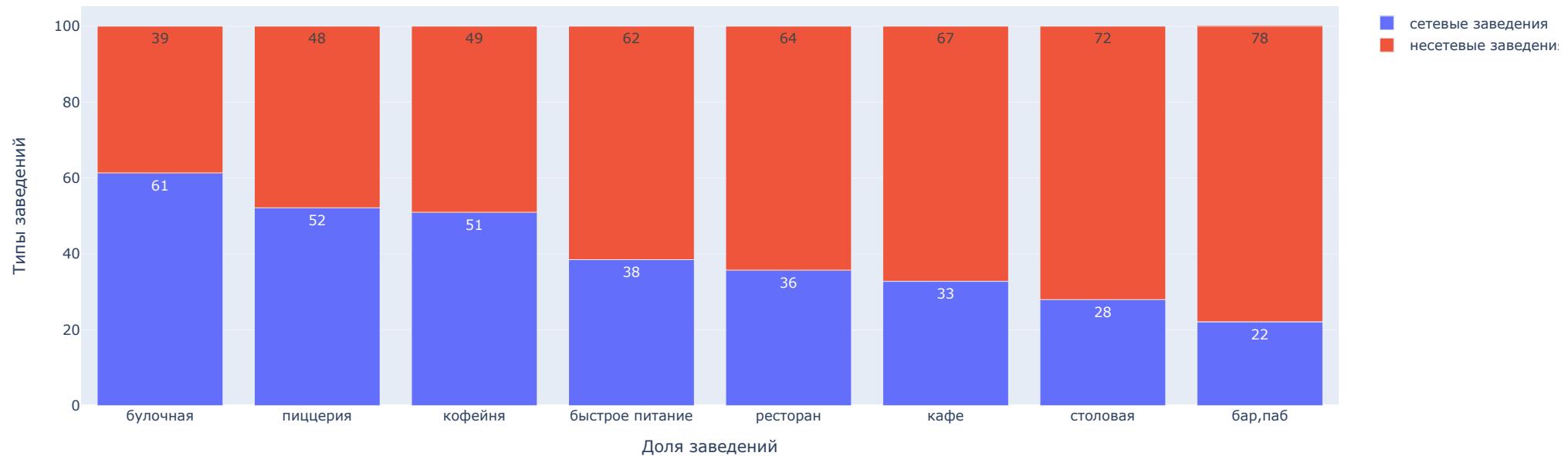
```
In [64]: chain = data.groupby('category').agg({'chain':'mean'})
chain['non_chain'] = 1 - chain['chain']
chain[['chain', 'non_chain']] = round(chain[['chain', 'non_chain']]*100, 2)
chain.columns = ['сетевые заведения', 'несетевые заведения']
chain = chain.sort_values(by='сетевые заведения', ascending=False)
chain
```

```
Out[64]:
```

category	сетевые заведения	несетевые заведения
булочная	61.33	38.67
пиццерия	52.13	47.87
кофейня	50.96	49.04
быстрое питание	38.47	61.53
ресторан	35.70	64.30
кафе	32.76	67.24
столовая	27.94	72.06
бар,паб	22.09	77.91

```
In [65]: fig = px.bar(chain,
                  title = 'Доля сетевых заведений в разрезе категорий')
fig.update_xaxes(title_text='Доля заведений', tickfont=dict(size=12))
fig.update_yaxes(title_text='Типы заведений')
fig.update_layout(legend_title_text='')
fig.update_traces(texttemplate='%{value:.2}')
fig.show()
```

Доля сетевых заведений в разрезе категорий



Вывод: доля сетевых заведений составляет 60% среди булочных. Для пиццерий и кофеин доля сетевых заведений составляет 50%. Среди заведений быстрого питания, ресторанов, кафе, столовых, баров преобладает доля несетевых заведений. Наиболее редко сетевые заведения встречаются среди баров и пабов в 22% случаев.

ТОП-15 популярных сетей в Москве

```
In [66]: #отсортируем сетевые заведения
```

```
In [67]: top_15 = data.loc[data['chain'] == 1]
```

```
In [68]: #Сгруппируем данные по названию и найдем ТОП-15 самых популярных сетей в Москве
```

```
In [69]: top_15 = top_15.groupby('name')['name'].count().sort_values(ascending=False)
len(top_15)
```

```
Out[69]: 747
```

Всего в датасете 747 сетевых заведений с уникальным названием.

```
In [70]: #Найдем ТОП-15 самых популярных сетей Москвы
```

```
In [71]: top_15 = top_15.head(15)
top_15
```

```
Out[71]:
```

name	
шоколадница	120
домино'с пицца	76
додо пицца	74
one price coffee	71
яндекс лавка	69
cofix	65
prime	50
хинкальная	44
кофепорт	42
кулинарная лавка братьев караваевых	39
теремок	38
чайхана	37
cofefest	32
буханка	32
му-му	27

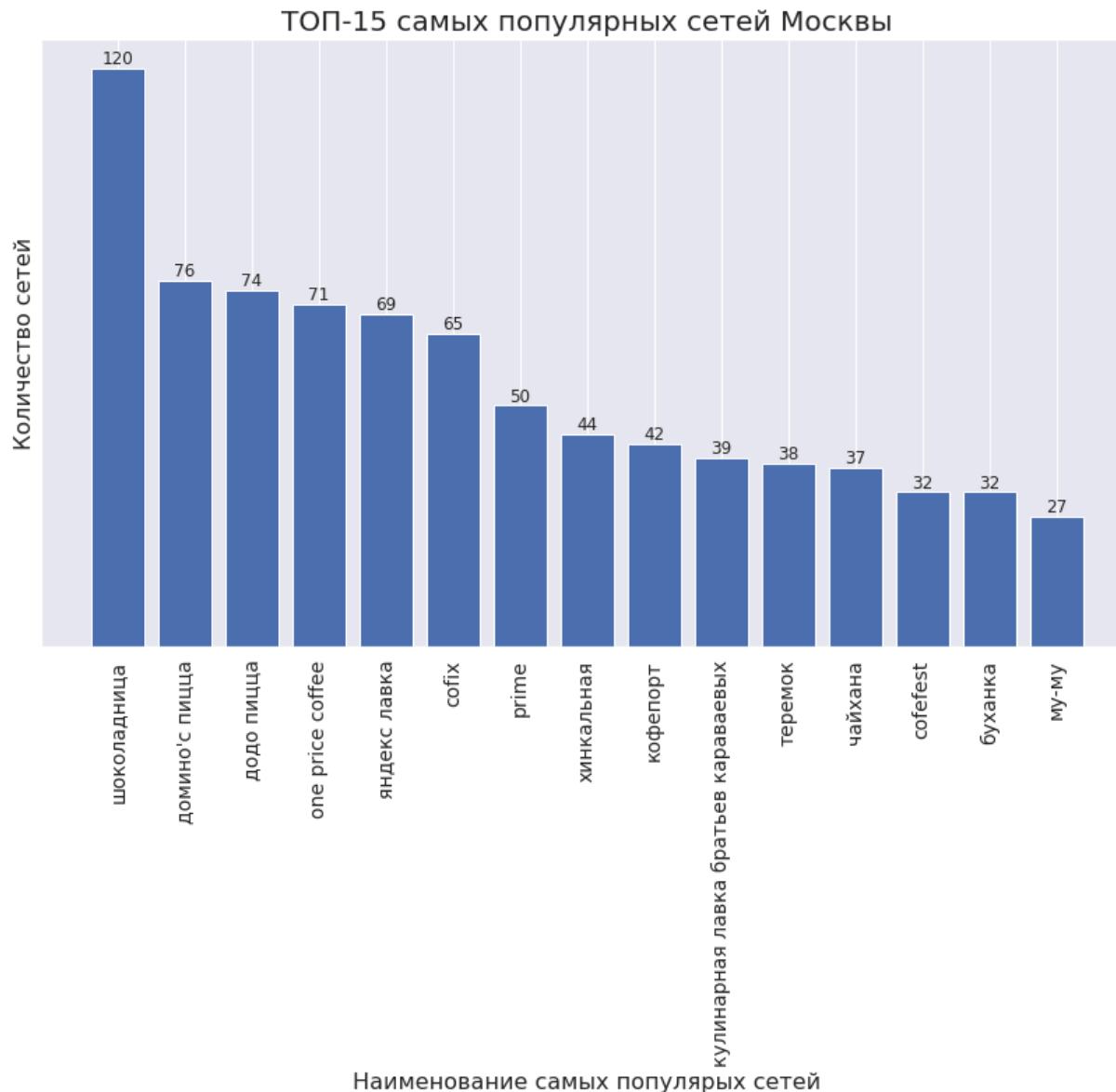
Name: name, dtype: int64

Самая популярная и распространенная сеть предприятий общественного питания - это Шоколадница, всего по Москве 120 точек Шоколадницы. На втором и третьем месте по популярности находятся пиццерии: домино'с пицца 76 и додо пицца 74 точки питания.

```
In [72]:
```

```
fig, ax = plt.subplots(figsize=(14, 8))
bars = ax.bar(top_15.index, top_15.values)

# добавление значений на столбы
for bar in bars:
    height = bar.get_height()
    ax.text(bar.get_x() + bar.get_width() / 2, height, height,
            ha='center', va='bottom')
plt.xticks(fontsize=14)
plt.yticks([])
plt.xlabel('Наименование самых популярных сетей', fontsize=16)
plt.ylabel('Количество сетей', fontsize=16)
ax.set_xticklabels(top_15.index, rotation=90)
plt.title('ТОП-15 самых популярных сетей Москвы', fontsize=20)
plt.show()
```



```
In [73]: # найдем категории заведений, к которым относятся ТОП-15 самых популярных сетей Москвы
```

```
In [74]: top_15_category = data.query('name in @top_15.index')
top_15_category = top_15_category[['name', 'category']]
top_15_category = top_15_category['category'].value_counts()
top_15_category
```

```
Out[74]:
```

кофейня	337
ресторан	186
пиццерия	152
кафе	100
булочная	25
быстрое питание	12
бар, паб	4
столовая	2

```
Name: category, dtype: int64
```

```
In [75]:
```

```
fig, ax = plt.subplots(figsize=(12, 6))
top_15_category.plot.bar()
for i, v in enumerate(top_15_category):
    ax.text(i, v, str(v), ha='center', va='bottom', fontsize=15)
plt.xticks(fontsize=12)
plt.yticks([])

plt.xlabel('Категории заведений', fontsize=14)
plt.ylabel('Количество заведений', fontsize=14)
plt.title('Распределение заведений по категориям по самым популярным сетям Москвы', fontsize=18);
```



```
In [76]:
```

```
#посмотрим, какая категория заведения соответствует каждому наименованию заведения из ТОП-15
```

```
In [77]:
```

```
top_15_1 = data.loc[data['chain'] == 1]
top_15_1 = top_15_1.groupby('name').agg({'name':'count', 'category':'first'})
top_15_1.columns = ['count', 'category']
top_15_1.sort_values(by='count', ascending=False).head(15)
```

Out[77]:

	count	category
name		
шоколадница	120	кофейня
домино'с пицца	76	пиццерия
додо пицца	74	пиццерия
one price coffee	71	кофейня
яндекс лавка	69	ресторан
cofix	65	кофейня
prime	50	ресторан
хинкальная	44	быстрое питание
кофепорт	42	кофейня
кулинарная лавка братьев караваевых	39	кафе
теремок	38	ресторан
чайхана	37	кафе
cofefest	32	кофейня
буханка	32	булочная
му-му	27	кафе

Вывод:

Всего в датасете 747 сетевых заведений с уникальным названием. Самая популярная и распространенная сеть предприятий общественного питания - это кофейня Шоколадница, всего по Москве 120 точек Шоколадницы. На втором и третьем месте по популярности находятся пиццерии: домино'с пицца и додо пицца 76 и 74 точки питания.

Среди самых популярных сетевых заведений Москвы наиболее распространены кофейни: 337 заведений, а также рестораны: 186 заведений. На третьем месте по распространенности находятся пиццерии: 152 заведения. В десятку также входят кафе, булочные, предприятия быстрого питания, бары и пабы, и столовые.

Исследование категорий заведений по административным районам

In [78]: *#С помощью сводной таблицы посчитаем по районам количество заведений каждой категории*In [79]: *#Построим столбчатую диаграмму, где по оси X будут указаны административные округи, а по оси Y будут указаны количество заведений определенной категории, а также общее количество заведений*In [80]: `district = data.groupby(['district', 'category']).agg({'name':'count'}).reset_index()`

Out[80]:

	district	category	name
0	Восточный административный округ	бар,паб	53
1	Восточный административный округ	булочная	25
2	Восточный административный округ	быстрое питание	71
3	Восточный административный округ	кафе	272
4	Восточный административный округ	кофейня	105
...
67	Южный административный округ	кафе	264
68	Южный административный округ	кофейня	131
69	Южный административный округ	пиццерия	73
70	Южный административный округ	ресторан	202
71	Южный административный округ	столовая	44

72 rows x 3 columns

In [81]: #Посмотрим, какие административные округа присутствуют в датасете

In [82]: district['district'].unique()

Out[82]: array(['Восточный административный округ',
 'Западный административный округ',
 'Северный административный округ',
 'Северо-Восточный административный округ',
 'Северо-Западный административный округ',
 'Центральный административный округ',
 'Юго-Восточный административный округ',
 'Юго-Западный административный округ',
 'Южный административный округ'], dtype=object)

In [83]: #Для удобства заменим длинные названия административных округов на более короткие аббревиатуры:

In [84]: district['district']= district['district'].replace({'Восточный административный округ':'ВАО',
 'Западный административный округ':'ЗАО',
 'Северный административный округ':'САО',
 'Северо-Восточный административный округ':'СВАО',
 'Северо-Западный административный округ':'СЗАО',
 'Центральный административный округ':'ЦАО',
 'Юго-Восточный административный округ':'ЮВАО',
 'Юго-Западный административный округ':'ЮЗАО',
 'Южный административный округ':'ЮАО'})

In [85]: district['district'].unique()

Out[85]: array(['ВАО', 'ЗАО', 'САО', 'СВАО', 'СЗАО', 'ЦАО', 'ЮВАО', 'ЮЗАО', 'ЮАО'], dtype=object)

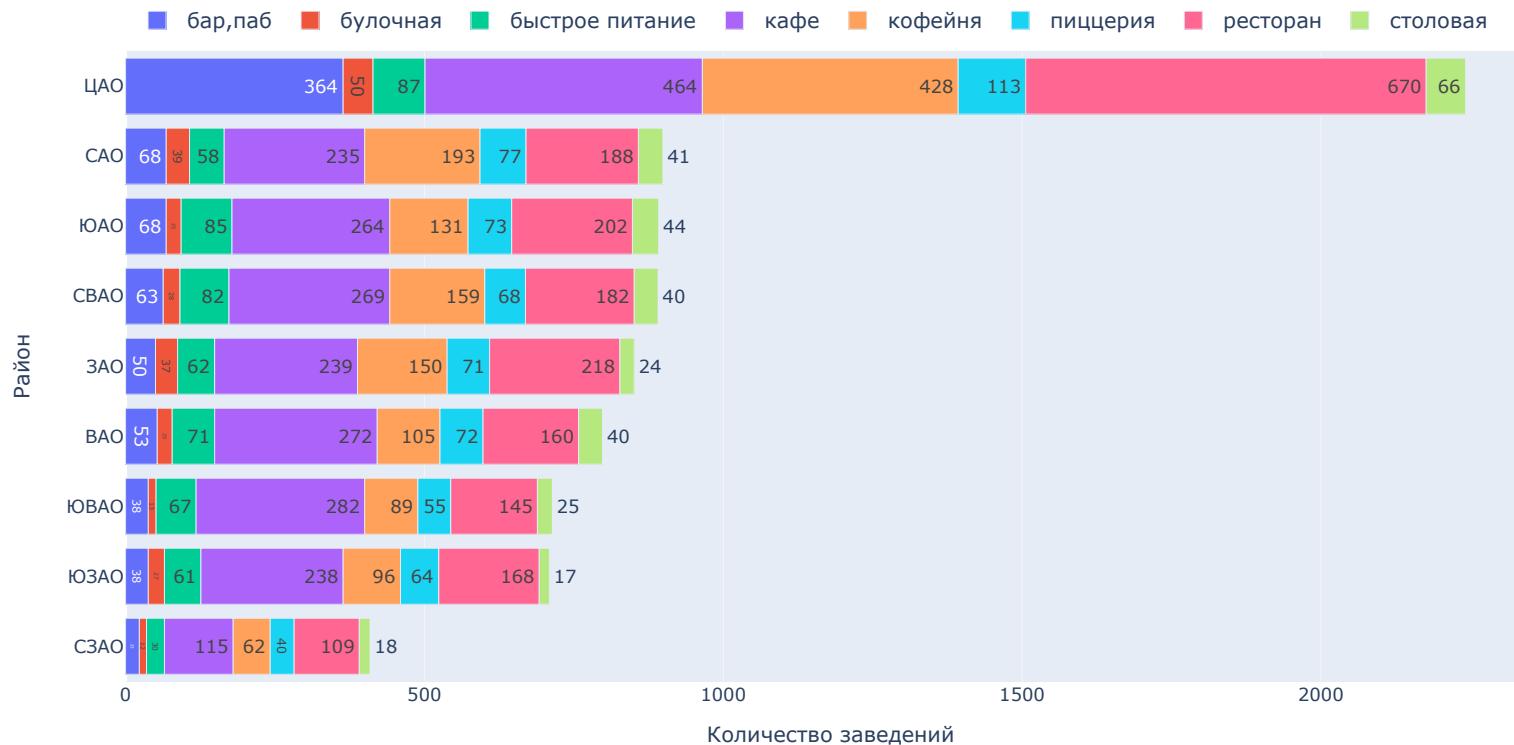
In [86]: fig = px.bar(
 district, y='district', x='name', color='category', title='Деление предприятий общественного питания по районам',
 text=district['name'].values
)
fig.update_yaxes(title_text='Район', categoryorder='total ascending')
fig.update_xaxes(title_text='Количество заведений')
fig.update_layout(
 width=1100,

```

height=600,
legend=dict(
    font=dict(
        size=14 # размер шрифта легенды
    )
),
legend_orientation='h', legend_yanchor='bottom', legend_y=1.01,
legend_xanchor='left', legend_x=.001, legend_title=None
)
fig.update_traces(hovertemplate='В <i>%{y}</i> %{x:.0f} заведений(-я, -е) <br> данного типа')
fig.show()

```

Деление предприятий общественного питания по районам



Вывод:

Больше всего предприятий общественного питания расположено в центральном административном округе (ЦАО): 2400 заведений, что более, чем на 100% больше, чем в любом другом округе Москвы. В ЦАО преобладают рестораны: 670 заведений данного типа, что составляет примерно 28% от всех заведений округа. На втором и третьем месте в ЦАО по популярности находятся кафе (464 заведения или 19%) и кофейни(428 заведений или 17%). Наиболее редки булочные (50 заведений или 2%) и столовые (66 заведений или 2.7%).

В остальных районах города на первом месте преобладают кафе, а на втором месте по численности находятся рестораны. Менее развиты булочные и столовые.

Меньше всего предприятий общественного питания находится в Северо-западном административном округе (всего около 400 заведений).

Исследование распределения средних рейтингов по категориям заведений

```
In [87]: #Создадим сводную таблицу, где каждой категории соответствует средний рейтинг
```

```
In [88]: data_rate = data.pivot_table(index='category', values='rating').sort_values(by='rating', ascending=False).reset_index()
data_rate['rating'] = round(data_rate['rating'], 2)
data_rate
```

```
Out[88]: category rating
0 бар.паб 4.39
1 пиццерия 4.30
2 ресторан 4.29
3 кофейня 4.28
4 булочная 4.27
5 столовая 4.21
6 кафе 4.12
7 быстрое питание 4.05
```

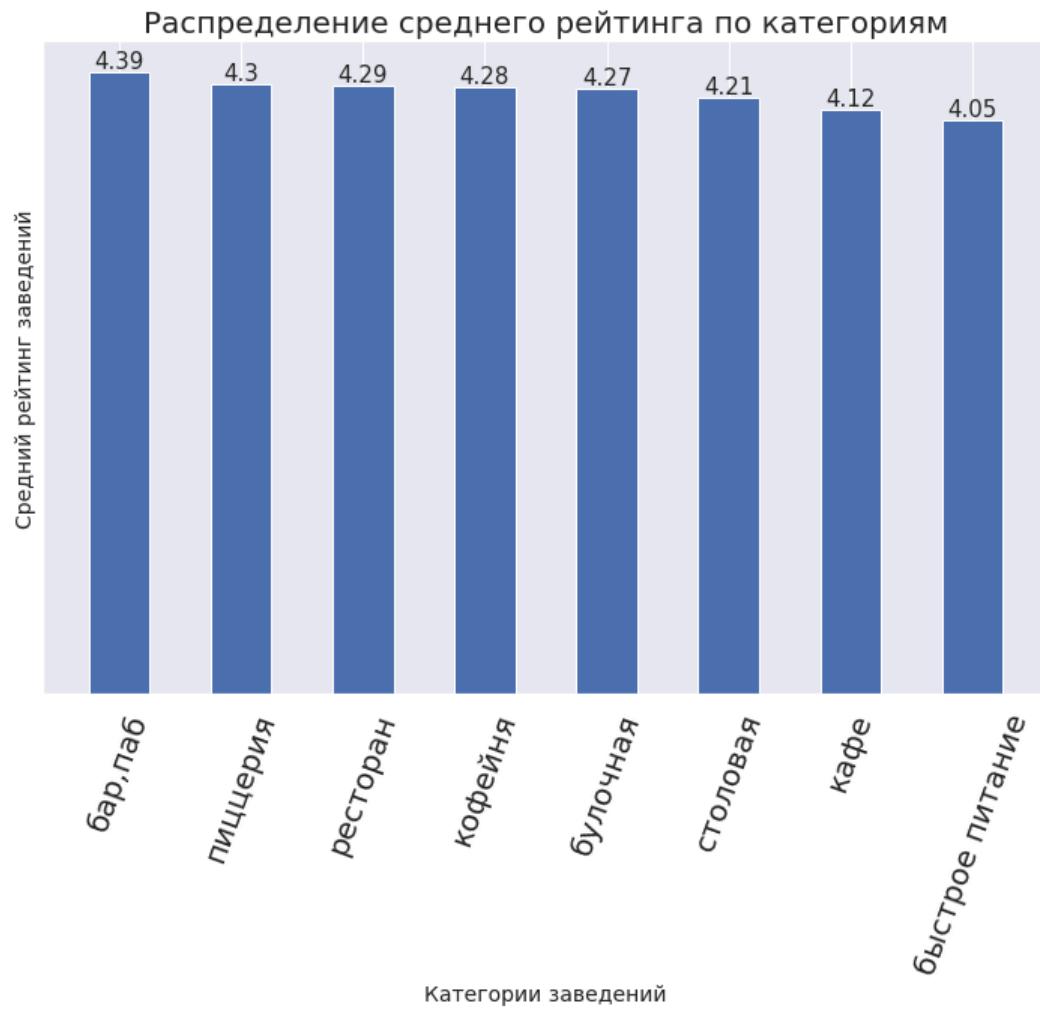
```
In [89]: #Визуализируем распределение рейтинга по категориям заведений и построим столбчатую диаграмму
```

```
In [90]: fig, ax = plt.subplots(figsize=(12, 8))
width = 0.5 # Увеличиваем ширину столбцов
padding = 50 # Отступ между меткой и столбцом
bars = ax.bar(data_rate['category'], data_rate['rating'], width=width)

for i, v in enumerate(data_rate['rating']):
    ax.text(i, v, str(v), ha='center', va='bottom', fontsize=15)

plt.xticks(fontsize=20)
plt.yticks([])
plt.xlabel('Категории заведений', fontsize=14)
plt.ylabel('Средний рейтинг заведений', fontsize=14)
ax.set_xticklabels(data_rate['category'], rotation=70)
plt.title('Распределение среднего рейтинга по категориям', fontsize=20)

plt.show()
```



```
In [91]: #Посчитаем общий средний рейтинг заведений Москвы:
```

```
In [92]: print(round(data['rating'].mean(), 2))
```

4.23

Вывод: средний рейтинг заведений всех заведений Москвы довольно высокий и составляет 4.23. Если смотреть в разрезе категорий, то самый высокий рейтинг у баров и пабов: 4.39; средний рейтинг пиццерий и ресторанов примерно одинаковый и составляет 4.3 и 4.29. Самый низкий рейтинг у предприятий быстрого питания: 4.05. Усредненные рейтинги заведений по типам питания различаются минимально.

Фоновая картограмма (хороплет) со средним рейтингом для заведений каждого района.

```
In [93]: #Выведем средний рейтинг по районам Москвы
```

```
In [94]: district_rate = data.pivot_table(  
    index='district', values='rating'  
)  
.reset_index().sort_values(by='rating', ascending=False)  
district_rate['rating'] = round(district_rate['rating'], 2)  
district_rate
```

Out[94]:

	district	rating
5	Центральный административный округ	4.38
2	Северный административный округ	4.24
4	Северо-Западный административный округ	4.21
8	Южный административный округ	4.18
1	Западный административный округ	4.18
0	Восточный административный округ	4.17
7	Юго-Западный административный округ	4.17
3	Северо-Восточный административный округ	4.15
6	Юго-Восточный административный округ	4.10

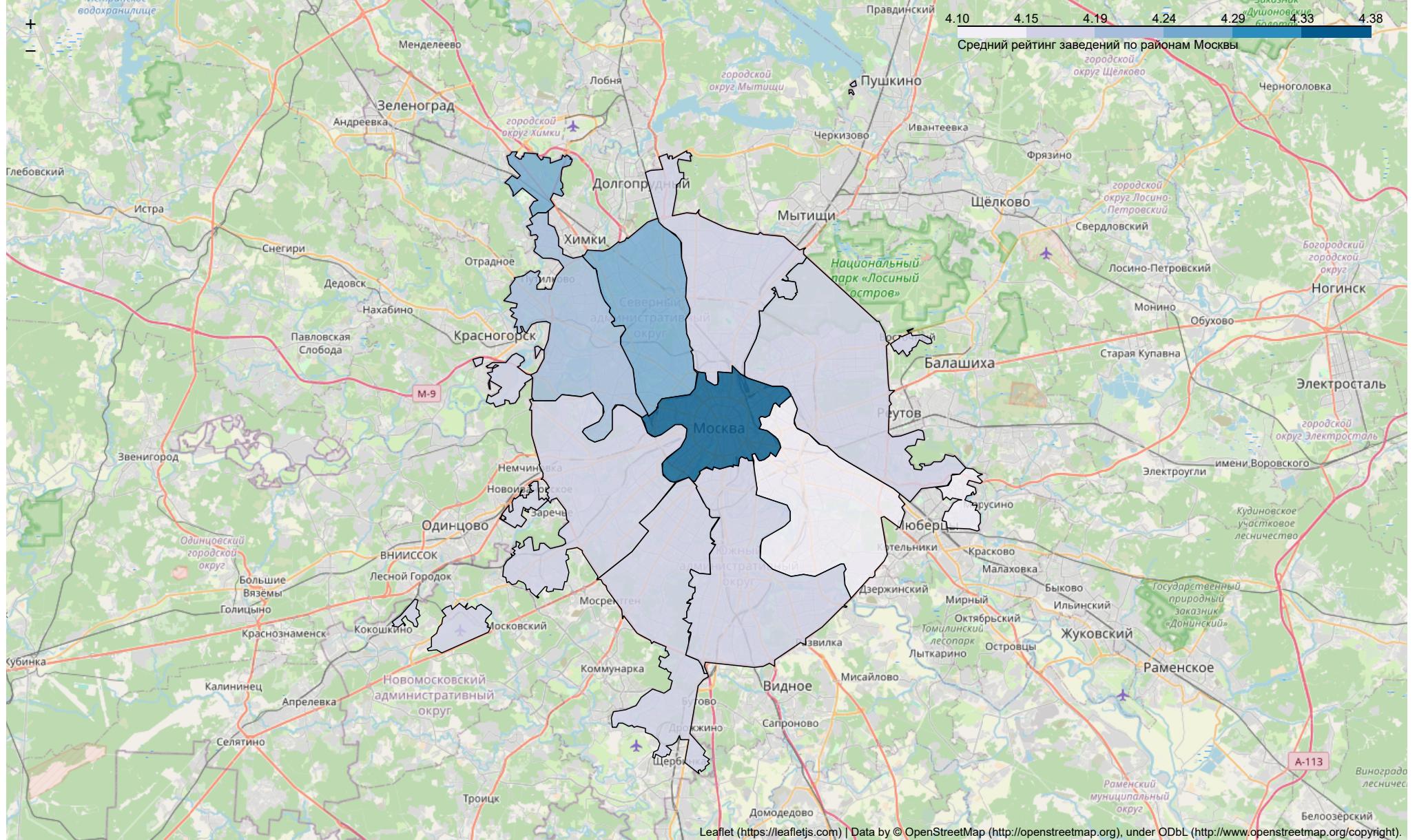
In [95]: #загрузим файл с границами районов

In [96]: with open('/datasets/admin_level_geomap.geojson', 'r', encoding='utf-8') as f:
 geo_json = json.load(f)In [97]: try:
 geo = '/datasets/admin_level_geomap.geojson'
except:
 geo = 'C:/Users/User/Downloads/admin_level_geomap.geojson_adm'In [98]: # moscow_lat - широта центра Москвы, moscow_lng - долгота центра Москвы
moscow_lat, moscow_lng = 55.751244, 37.618423

```
# создаём карту Москвы
m = Map(location=[moscow_lat, moscow_lng], zoom_start=10)
k=folium.Choropleth(
    geo_data=geo,
    data=district_rate,
    columns=['district','rating'],
    key_on='feature.name',
    fill_color='PuBu',
    fill_opacity=0.8,
    legend_name='Средний рейтинг заведений по районам Москвы',
    legend_font_size='30px'
).add_to(m)
```

```
#Добавим интерактивное отражение названий районов
k.geojson.add_child(folium.features.GeoJsonTooltip(['name'], labels=False))
folium.LayerControl().add_to(m)
# выводим карту
m
```

Out[98]:



Leaflet (<https://leafletjs.com>) | Data by © OpenStreetMap (<http://openstreetmap.org>), under ODbL (<http://www.openstreetmap.org/copyright>).

Вывод: Самый высокий средний рейтинг у Центрального административного района: 4.38, за ним следует Северный административный округ: средний рейтинг заведений 4.24, и Северо-Западный административный округ 4.21. Самый низкий рейтинг в Юго-Восточном административном округе: 4.10.

Визуализация заведений на карте с помощью кластеров.

In [99]:

```
moscow_lat, moscow_lng = 55.751244, 37.618423
# создаём карту Москвы
m_Moscow = Map(location=[moscow_lat, moscow_lng], zoom_start=10)
marker_cluster = MarkerCluster().add_to(m_Moscow)

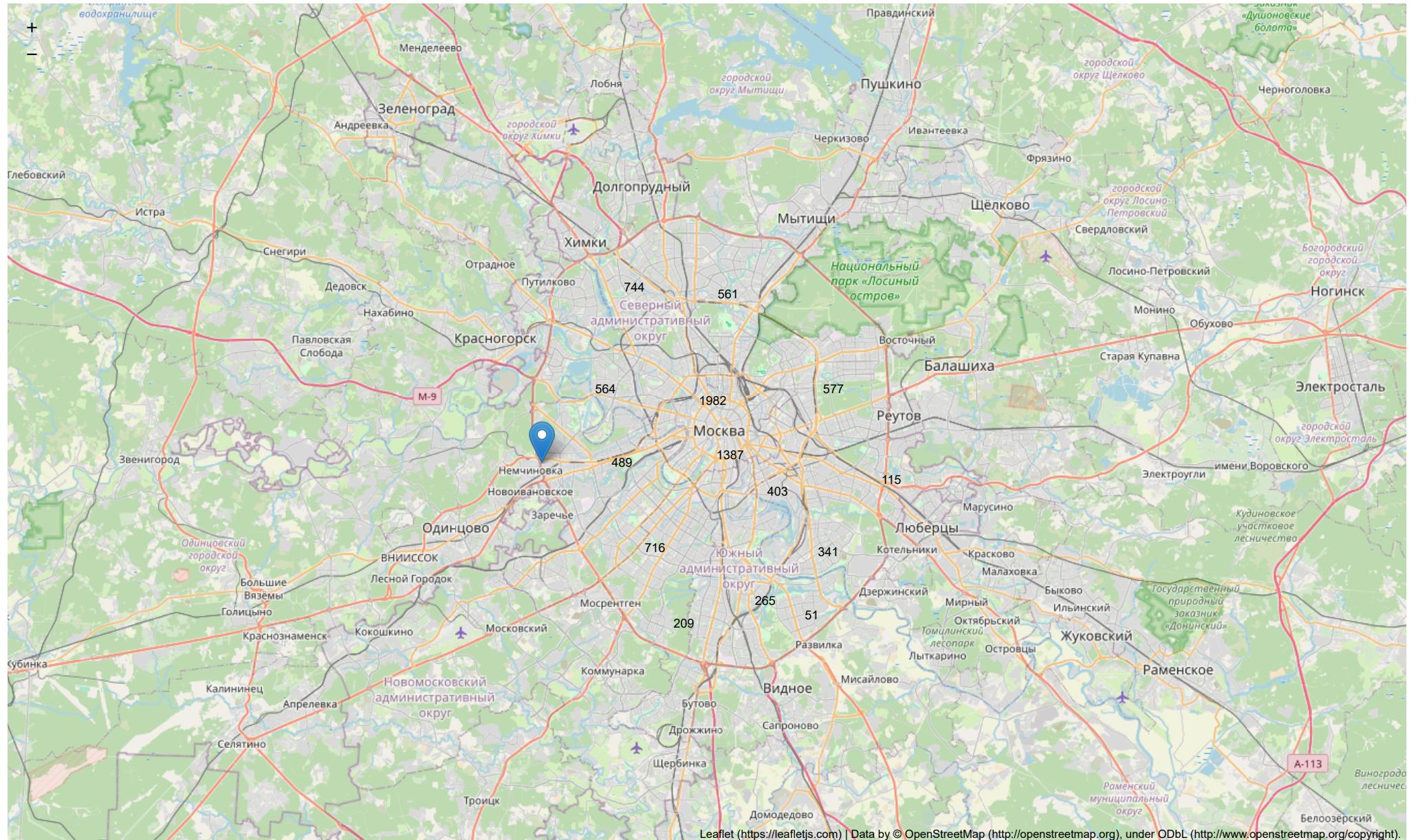
# пишем функцию, которая принимает строку датафрейма,
```

```
# создаёт маркер в текущей точке и добавляет его в кластер marker_cluster
def create_clusters(row):
    Marker(
        [row['lat'], row['lng']],
        popup=f'{row['name']} {row['rating']} {row['category']}'
    ).add_to(marker_cluster)
```

```
# применяем функцию create_clusters() к каждой строке датафрейма
data.apply(create_clusters, axis=1)
```

```
# выводим карту
m_Moscow
```

Out[99]:



Исследование и визуализация ТОП-15 улиц по количеству заведений в разрезе категорий

In [100...]

```
#Построим сводную таблицу и посчитаем общее количество заведений для каждой улицы,  
#выведем топ-15 улиц с наибольшим количеством заведений
```

In [101...]

```
top_streets = data.pivot_table(  
    index=['street_name'], values=['name'], aggfunc='count'  
) .sort_values(by='name', ascending=False).head(15)  
top_streets
```

Out[101]:

street_name	name
проспект мира	184
профсоюзная улица	122
проспект вернадского	108
ленинский проспект	107
ленинградский проспект	95
дмитровское шоссе	88
каширское шоссе	77
варшавское шоссе	76
ленинградское шоссе	70
мкад	65
люблинская улица	60
улица вавилова	55
кутузовский проспект	54
улица миклухо-маклая	49
пятницкая улица	48

In [102...]

```
top_streets_category = data.query('street_name in @top_streets.index').groupby(  
    ['street_name', 'category'])  
.agg(count=('category', 'count')).reset_index()
```

```
top_streets_category
```

Out[102]:

	street_name	category	count
0	варшавское шоссе	бар,паб	6
1	варшавское шоссе	быстрое питание	7
2	варшавское шоссе	кафе	18
3	варшавское шоссе	кофейня	14
4	варшавское шоссе	пиццерия	4
...
106	улица миклухо-маклая	быстрое питание	4
107	улица миклухо-маклая	кафе	21
108	улица миклухо-маклая	кофейня	4
109	улица миклухо-маклая	пиццерия	2
110	улица миклухо-маклая	ресторан	15

111 rows × 3 columns

In [103...]

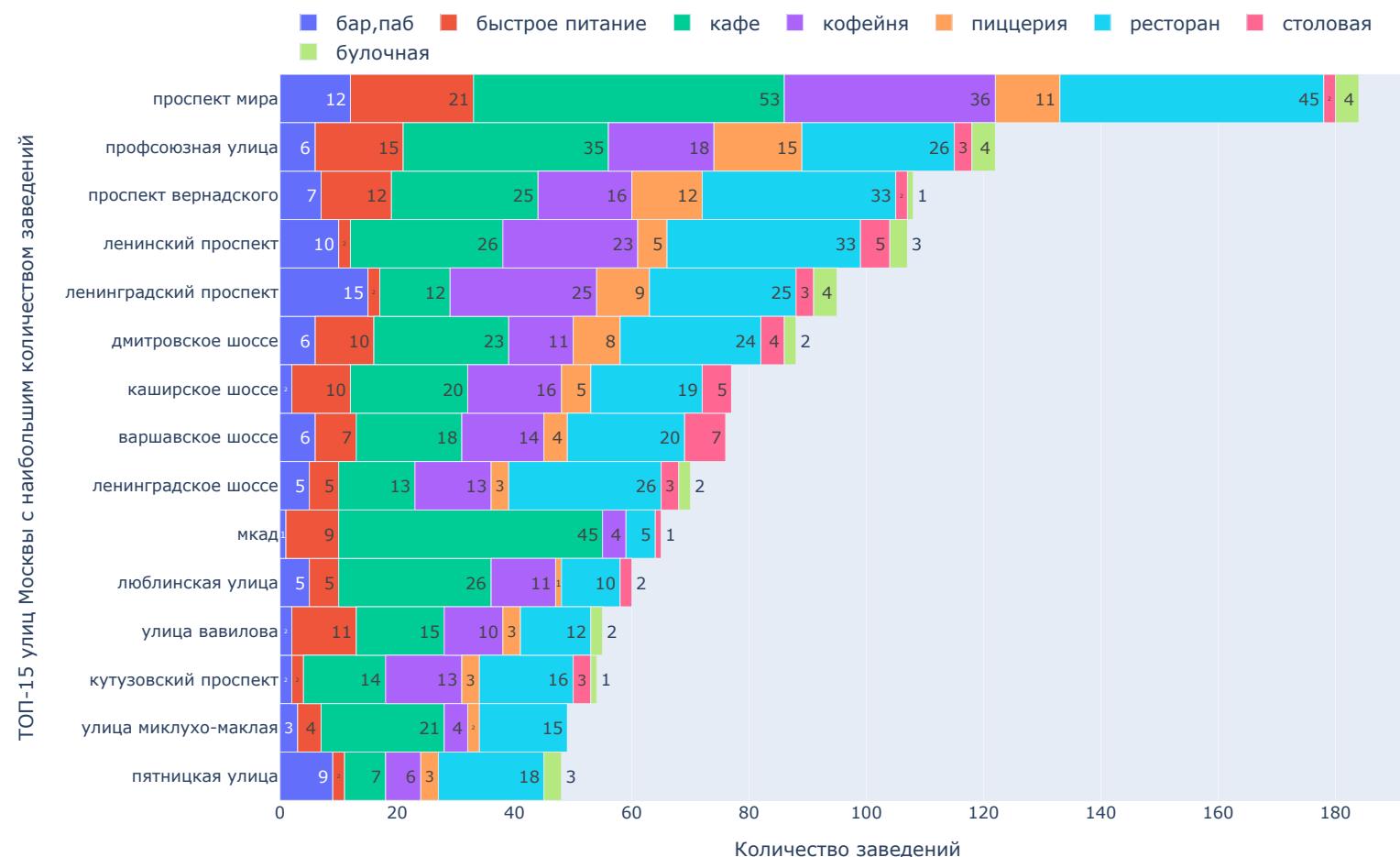
#Визуализируем распределение заведений в разрезе улиц

In [104...]

```
fig = px.bar(
    top_streets_category, y='street_name', x=top_streets_category['count'], color='category',
    title='Количество предприятий общественного питания в разрезе улиц',
    width=1000, height=1200,
    text=top_streets_category['count'].values
)
fig.update_yaxes(
    title_text='ТОП-15 улиц Москвы с наибольшим количеством заведений', categoryorder='total ascending'
)

fig.update_xaxes(title_text='Количество заведений')
fig.update_layout(
    width=1100,
    height=800,
    legend=dict(
        font=dict(
            size=14 # размер шрифта легенды
        )
    ),
    legend_orientation='h', legend_yanchor='bottom', legend_y=1.24,
    legend_xanchor='left', legend_x=0.001, legend_title=None
)
fig.update_traces(
    hovertemplate='В <i>%{y}</i> %{x:.0f} заведений(-я, -е) <br> данного типа',
    width=1
)
fig.show()
```

Количество предприятий общественного питания в разрезе улиц



Вывод:

В ТОП-15 улиц по количеству заведений вошли такие улицы, как:

- 1) проспект мира
- 2) профсоюзная улица
- 3) проспект вернадского
- 4) ленинский проспект
- 5) ленинградский проспект

6) дмитровское шоссе

7) каширское шоссе

8) варшавское шоссе

9) ленинградское шоссе

10) мкад

11) люблинская улица

12) улица вавилова

13) кутузовский проспект

14) улица миклухо-маклая

15) пятницкая улица

Больше всего заведений расположено на проспекте Мира: свыше 180 предприятий общественного питания, из них в тройке лидеров кафе(53 заведений), рестораны(45 заведений) и кофейни (36 заведений). По остальным улицам типы заведений распределились также пропорционально: большая часть предприятий приходится на рестораны и кафе, за исключением МКАДа: 70% предприятий общественного питания составляют кафе - 45 заведений.

Исследование улиц с одним объектом общественного питания

In [105...]

```
one_cafe = data.pivot_table(  
    index=['street_name'], values=['name'], aggfunc='count'  
)  
  
one_cafe = one_cafe.query('name == 1')  
one_cafe.columns = ['count']  
  
one_cafe
```

Out[105]:

	count
street_name	
1-й автозаводский проезд	1
1-й балтийский переулок	1
1-й варшавский проезд	1
1-й вешняковский проезд	1
1-й голутвинский переулок	1
...	...
электрический переулок	1
юго-западный административный округ	1
якиманский переулок	1
яконоапостольский переулок	1
№ 7	1

457 rows × 1 columns

In [106...]

```
one_cafe_filtered = data.query('street_name in @one_cafe.index')
```

Всего в данных 457 улиц с одним заведением общественного питания. Посмотрим на распределение по категориям.

In [107...]

```
one_cafe_filtered_data = one_cafe_filtered.groupby('category').agg({'name': 'count'}).sort_values(by='name', ascending=False)
one_cafe_filtered_data
```

Out[107]:

	name
category	
кафе	159
ресторан	93
кофейня	84
бар,лаб	39
столовая	36
быстрое питание	23
пиццерия	15
булочная	8

In [108...]

#Посмотрим, как распределились доли в зависимости от категории

In [109...]

```
one_cafe_filtered_total = one_cafe_filtered_data['name'].sum()

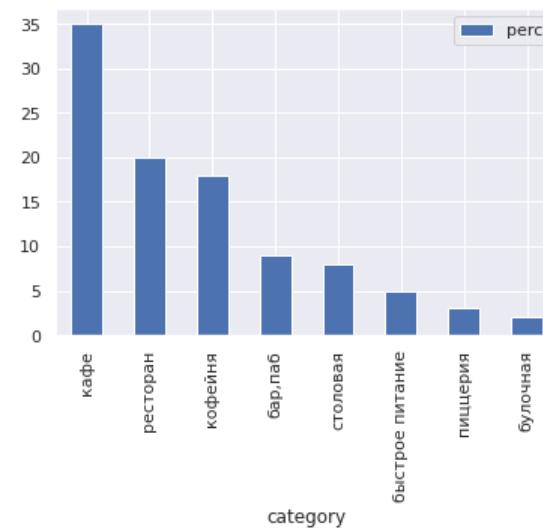
one_cafe_filtered_data['perc'] = round(one_cafe_filtered_data['name']/one_cafe_filtered_total, 2)*100
one_cafe_filtered_data
```

```
Out[109]:
```

category	name	perc
кафе	159	35.0
ресторан	93	20.0
кофейня	84	18.0
бар,паб	39	9.0
столовая	36	8.0
быстрое питание	23	5.0
пиццерия	15	3.0
булочная	8	2.0

```
In [110...:
```

```
one_cafe_filtered_data.plot(y='perc', kind='bar');
```



По большей части на 35% улицах одно единственное заведение - это кафе. Примерно в равных долях: 20% случаев и 17% случаев одно единственное заведение является рестораном или кофейней.

```
In [111...:
```

```
#Посмотрим по карте Москвы, где расположены исследуемые улицы
```

```
In [112...:
```

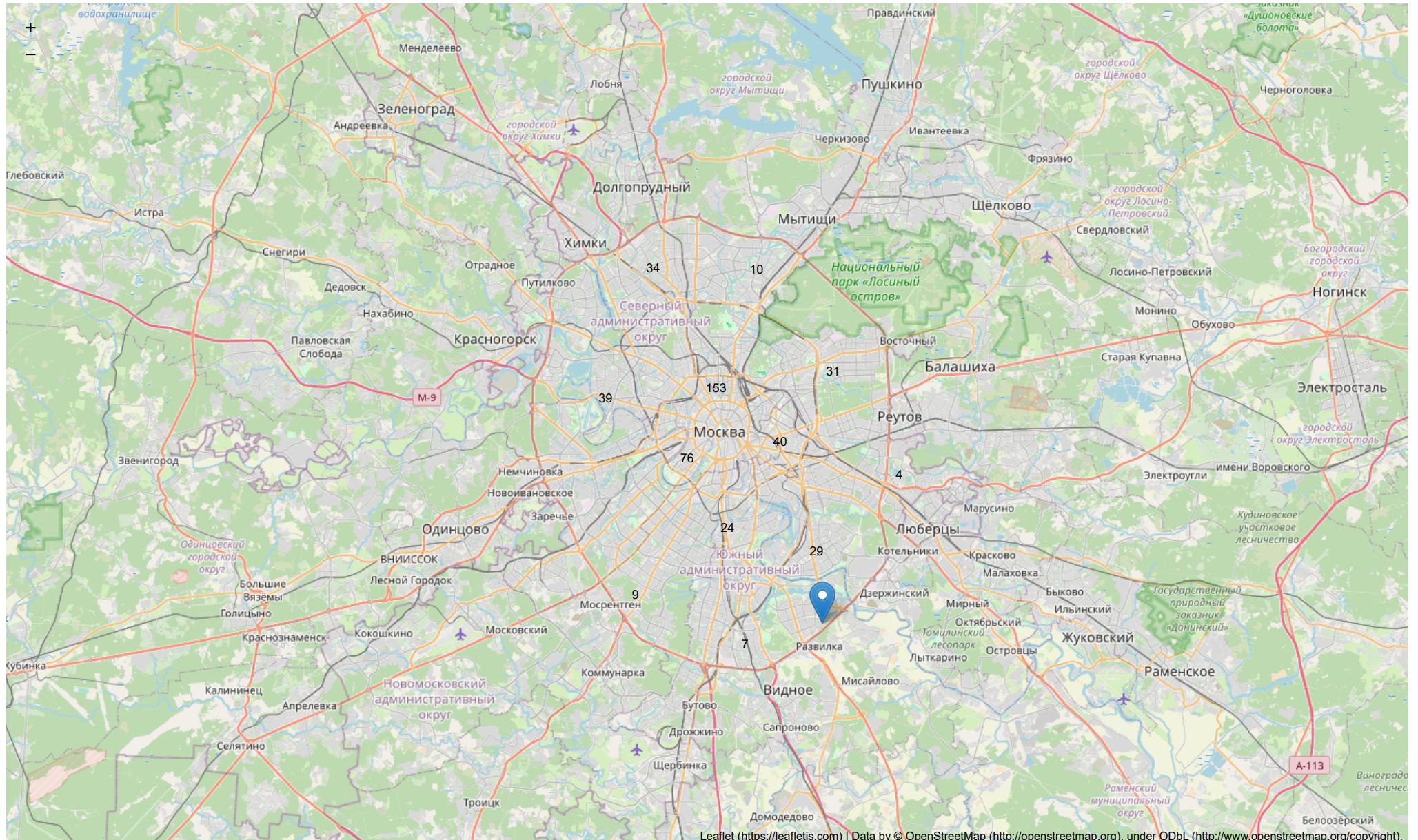
```
moscow_lat, moscow_lng = 55.751244, 37.618423
# создаём карту Москвы
m_Moscow = Map(location=[moscow_lat, moscow_lng], zoom_start=10)
marker_cluster = MarkerCluster().add_to(m_Moscow)

# пишем функцию, которая принимает строку датафрейма,
# создаёт маркер в текущей точке и добавляет его в кластер marker_cluster
def create_clusters(row):
    Marker(
        [row['lat'], row['lng']],
        popup=f'{row['street_name']}'
    ).add_to(marker_cluster)

# применяем функцию create_clusters() к каждой строке датафрейма
one_cafe_filtered.apply(create_clusters, axis=1)
```

выводим карту
m_Moscow

Out[112]:



Вывод: Улицы с одним заведением могут располагаться как в центре, так и по всем районам Москвы (в том числе на окраинах) равномерно, эти улички скорее являются достаточно маленькими, переулками или проездами, но также встречаются и на известных набережных, например таких, как Софийская набережная и Раушская набережная. По большей части на 35% улицах одно единственное заведение - это кафе. Примерно в равных долях: 20% случаев и 17% случаев одно единственное заведение является рестораном или кофейней. Всего в данных 457 улиц с одним заведением общественного питания.

Исследование среднего чека заведений в зависимости от района Москвы

```
In [113... #Посчитаем медианное значение столбца middle_avg_bill для каждого района
```

```
In [114... median_bill = data.pivot_table(index='district', values='middle_avg_bill', aggfunc='median') .reset_index().sort_values(by='middle_avg_bill', ascending=False) median_bill
```

```
Out[114]:
```

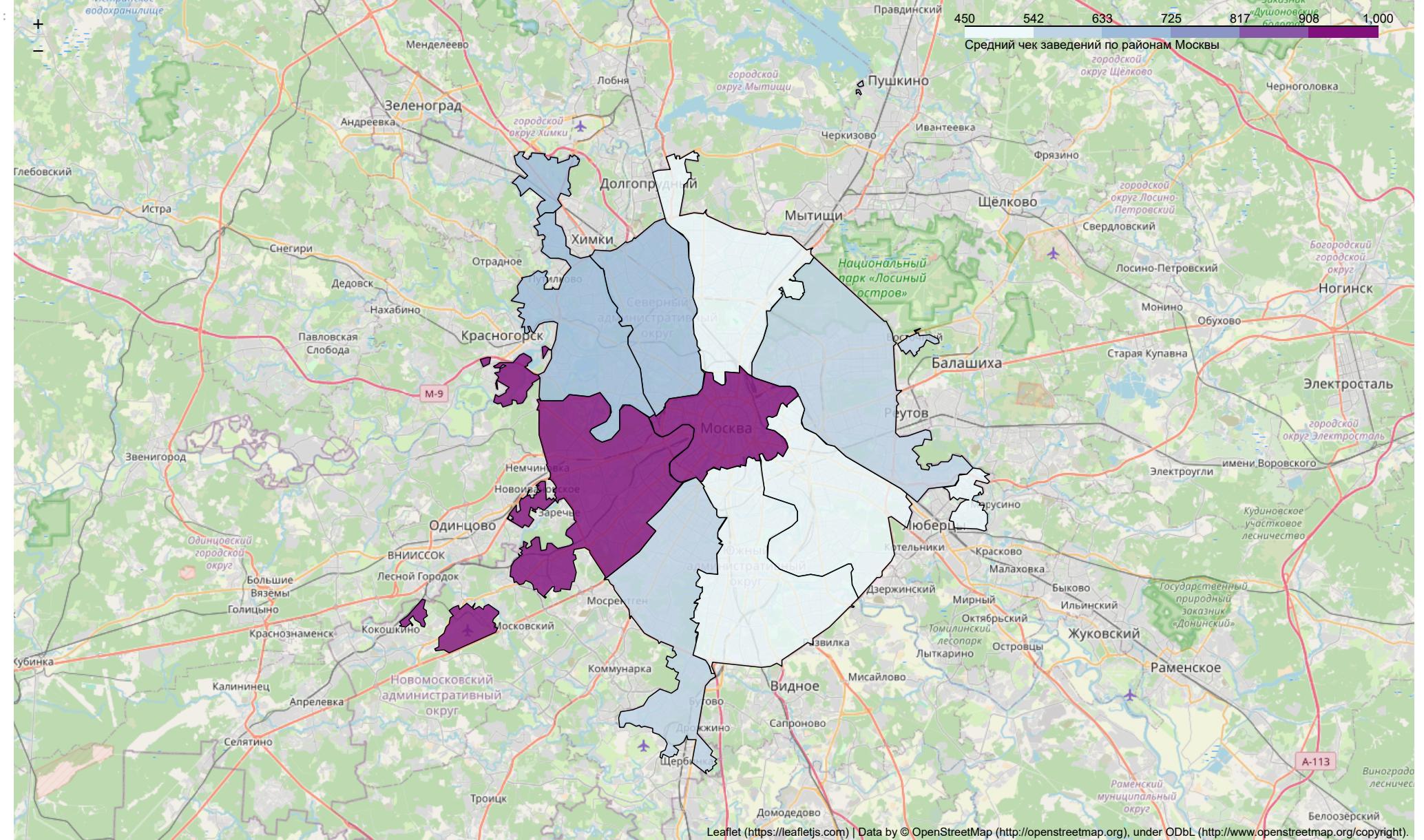
	district	middle_avg_bill
1	Западный административный округ	1000.0
5	Центральный административный округ	1000.0
4	Северо-Западный административный округ	700.0
2	Северный административный округ	650.0
7	Юго-Западный административный округ	600.0
0	Восточный административный округ	575.0
3	Северо-Восточный административный округ	500.0
8	Южный административный округ	500.0
6	Юго-Восточный административный округ	450.0

```
In [115... #Построим фоновую картограмму
```

```
In [116... # moscow_lat - широта центра Москвы, moscow_lng - долгота центра Москвы moscow_lat, moscow_lng = 55.751244, 37.618423
```

```
# создаём карту Москвы
m = Map(location=[moscow_lat, moscow_lng], zoom_start=10)
k=folium.Choropleth(
    geo_data=geo,
    data=median_bill,
    columns=['district','middle_avg_bill'],
    key_on='feature.name',
    fill_color='BuPu',
    fill_opacity=0.8,
    legend_name='Средний чек заведений по районам Москвы',
    legend_font_size='30px'
).add_to(m)
```

```
#Добавим интерактивное отражение названий районов
k.geojson.add_child(folium.features.GeoJsonTooltip(['name'], labels=False))
folium.LayerControl().add_to(m)
# выводим карту
m
```



Вывод:

Самый высокий средний чек в Центральном и Западном административных округах, и составляет 1000 руб., что обусловлено самыми высокими ценами для центра города и преобладающим количеством ресторанов, а Западный административный округ является одним из самых престижных для проживания. Чем дальше заведение находится от Центрального административного округа и Западного административного круга, тем средний чек ниже. Самый низкий средний чек в Южном административном округе, в Юго-Восточном административном округе, в Северо-Восточном административном округе и составляет 450-500 руб.

Исследование часов работы заведений, их зависимость от расположения и категории заведения.

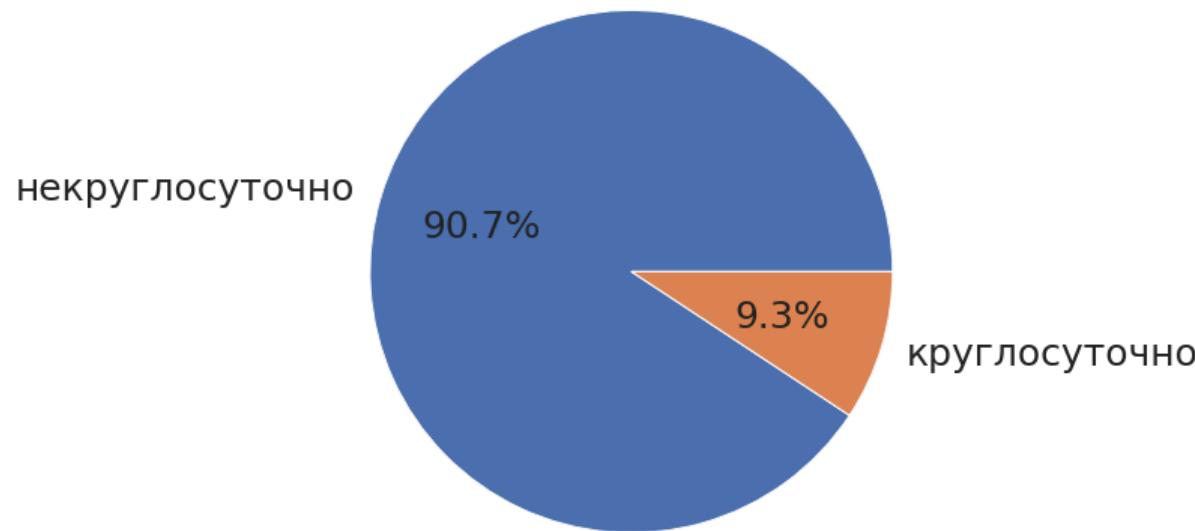
```
In [118]: hours = data.pivot_table(  
    index='is_24/7', values='name', aggfunc='count'  
) .rename(index={True:'круглосуточно', False:'некруглосуточно'})  
hours
```

```
Out[118]:  
      name  
is_24/7  
некруглосуточно    7139  
круглосуточно     730
```

```
In [119]: #визуализируем соотношение сетевых и несетевых заведений
```

```
In [120]: hours.plot(y='name', kind='pie', figsize=(8,8), autopct='%1.1f%%', textprops={'fontsize': 25});  
plt.ylabel('') # Удаляем подпись оси у  
plt.legend().remove()  
plt.title('Доли круглосуточных и некруглосуточных заведений Москвы', fontsize=20) # Добавляем заголовок  
plt.show()
```

Доли круглосуточных и некруглосуточных заведений Москвы



Доля заведений Москвы, которые работают круглосуточно незначительна и составляет всего 9.3%, в то время, как доля некруглосуточных заведений составляет 90.7%

```
In [121]: category_hours = data.pivot_table(  
    index='category', columns='is_24/7', values='name', aggfunc=['count'])  
)  
category_hours.columns=['некруглосуточные', 'круглосуточные']  
  
category_hours = category_hours.sort_values(by='некруглосуточные', ascending=False)  
category_hours['общее количество'] = category_hours['некруглосуточные'] + category_hours['круглосуточные']  
category_hours['некруглосуточные, %'] = round(category_hours['некруглосуточные']/category_hours['общее количество']*100)  
category_hours['круглосуточные, %'] = round(category_hours['круглосуточные']/category_hours['общее количество']*100)
```

```
category_hours_perc = category_hours.drop(columns=['некруглосуточные', 'круглосуточные', 'общее количество'])  
category_hours_perc
```

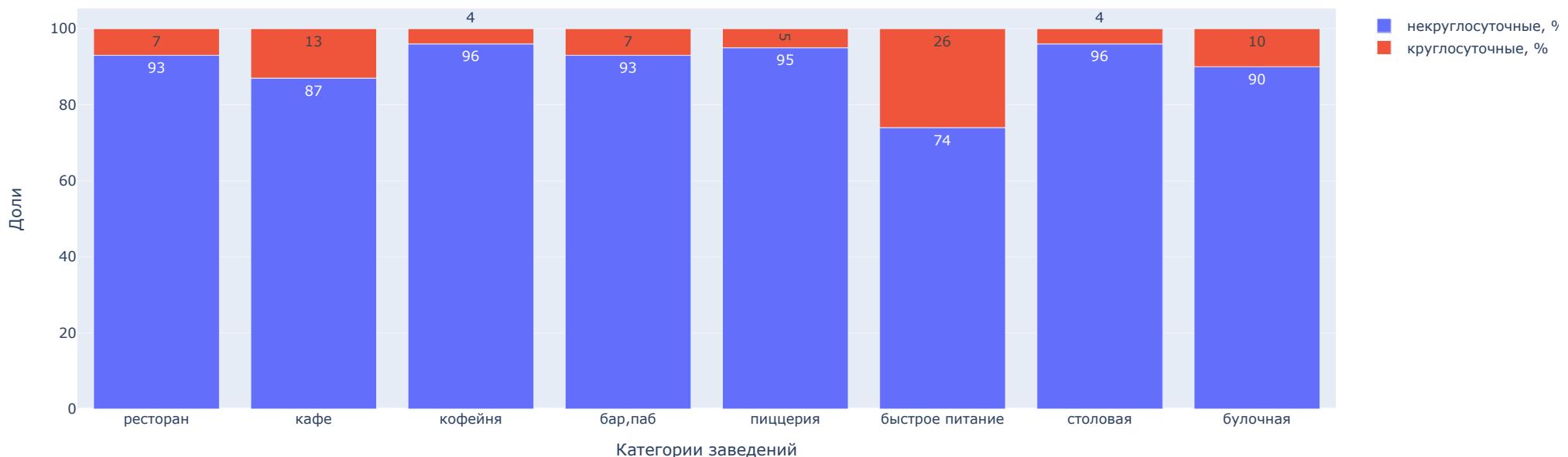
Out[121]:

некруглосуточные, % круглосуточные, %

category	некруглосуточные, %	круглосуточные, %
ресторан	93.0	7.0
кафе	87.0	13.0
кофейня	96.0	4.0
бар,паб	93.0	7.0
пиццерия	95.0	5.0
быстрое питание	74.0	26.0
столовая	96.0	4.0
булочная	90.0	10.0

```
In [122...]  
fig = px.bar(category_hours_perc,  
             title = 'Процентное соотношение круглосуточных и некруглосуточных заведений в Москве')  
  
fig.update_xaxes(title_text='Категории заведений', tickfont=dict(size=12))  
fig.update_yaxes(title_text='Доли')  
fig.update_layout(legend_title_text='')  
fig.update_traces(texttemplate='%{value:.2}')fig.show()
```

Процентное соотношение круглосуточных и некруглосуточных заведений в Москве



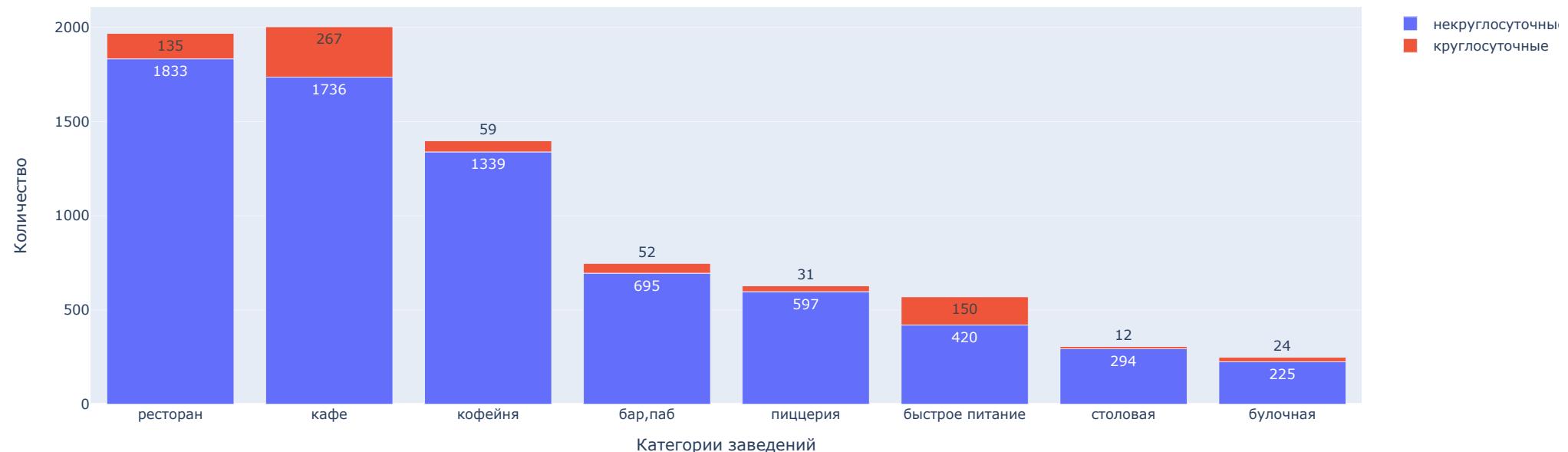
```
In [123...]: category_hours_count = category_hours.drop(  
    columns=['общее количество', 'некруглосуточные, %', 'круглосуточные, %'])  
)  
category_hours_count
```

```
Out[123]: некруглосуточные  круглосуточные
```

category	некруглосуточные	круглосуточные
ресторан	1833	135
кафе	1736	267
кофейня	1339	59
бар,паб	695	52
пиццерия	597	31
быстрое питание	420	150
столовая	294	12
булочная	225	24

```
In [124...]: fig = px.bar(category_hours_count,  
                    title = 'Количественное соотношение круглосуточных и некруглосуточных заведений в Москве')  
  
fig.update_xaxes(title_text='Категории заведений', tickfont=dict(size=12))  
fig.update_yaxes(title_text='Количество')  
fig.update_layout(legend_title_text='')  
fig.update_traces(texttemplate='%{y:.0f}')  
fig.show()
```

Количественное соотношение круглосуточных и некруглосуточных заведений в Москве



Вывод: В Москве на 90.7 % преобладают заведения, которые работают некруглосуточно, в то же время доля заведений, работающих круглосуточно составляет всего 9.3%. Наибольшее количество круглосуточных заведений приходится на кафе: 267 заведений, и на предприятия быстрого питания: 150 заведений. В процентном соотношении среди предприятий быстрого питания доля заведений, работающих круглосуточно наибольшая и составляет 26%. В тройку лидеров среди наибольшей доли круглосуточных заведений также вошли кафе: 13%, и круглосуточные булочные: 10%.

ВЫВОД ПО РАЗДЕЛУ "АНАЛИЗ ДАННЫХ": - Самые распространенная категория заведений Москвы - это кафе. Кафе занимают долю в размере 28.3% от всех предприятий общественного питания Москвы, представленных в данных. С небольшим отрывом по популярности идут рестораны : 24.3%. На третьем и четвертом месте находятся кофейни (16.8%) и пабы (9.1%). Наименее популярные в Москве столовые и булочные, доля которых составила 3.7% и 3%. - Для каждой категории заведений большую часть занимают заведения с отсутствующими данными по количеству посадочных мест. Наибольшая доля - 7.6% или 370 заведений приходится на рестораны с количеством посадочных мест от 61 до 100, и столько же - 7.6% или 370 заведений приходится на кафе с количеством посадочных мест от 31 до 60. Также в Москве распространены кофейни примерно с равнозначной долей по количеству посадочных мест от 31 до 60 - 165 заведений или 3.5% от всех заведений города, от 61 до 100 - 160 заведений или 3.4% от всех заведений города, от 101 до 200 - 153 заведений от 3.2% от всех заведений города. - В Москве преимущественно работают несетевые заведения, их доля составляет 61.9%. Доля сетевых заведений составила 38.1%. Наибольшая доля сетевых заведений среди булочных и составляет 60%. Для пиццерий и кофеин доля для сетевых и несетевых заведений разделена поровну и составляет 50%. Среди заведений быстрого питания, ресторанов, кафе, столовых, баров преобладает доля несетевых заведений. Наиболее редко сетевые заведения встречаются среди баров и пабов: в 22% случаев. - Всего в датасете 747 сетевых заведений с уникальным названием. Самая популярная и распространенная сеть предприятий общественного питания - это кофейня "Шоколадница", всего по Москве 120 точек Шоколадницы. На втором и третьем месте по популярности находятся пиццерии: "Домино'с пицца" и "Додо пицца": 76 и 74 точки питания. - Среди категорий сетевых заведений самые распространенные кофейни: 337 заведений, а также рестораны: 186 заведений. На третьем месте по распространенности находятся пиццерии: 152 заведения. В десятку также входят кафе, булочные, предприятия быстрого питания, бары и пабы, и столовые. - Больше всего предприятий общественного питания расположено в центральном административном округе (ЦАО): 2400 заведений, что более, чем на 100% больше, чем в любом другом округе Москвы. В ЦАО преобладают рестораны: 670 заведений данного типа, что составляет примерно 28% от всех заведений округа. На втором и третьем месте в ЦАО по популярности находятся кафе (464 заведения или 19%) и кофейни(428 заведений или 17%). Наиболее редки булочные (50 заведений или 2%) и столовые (66 заведений или 2.7%). В остальных районах города на первом месте преобладают кафе, а на втором месте по численности находятся рестораны. Менее развиты булочные и столовые. - Меньше всего предприятий общественного питания находится в Северо-западном административном округе (всего около 400 заведений). - Средний рейтинг заведений всех заведений Москвы довольно высокий и составляет 4.23. Если смотреть в разрезе категорий, то самый высокий рейтинг у баров и пабов: 4.39; средний рейтинг пиццерий и ресторанов примерно

одинаковый и составляет 4.3 и 4.29. Самый низкий рейтинг у предприятий быстрого питания: 4.05. Усредненные рейтинги заведений по типам питания различаются минимально Самый высокий средний рейтинг у Центрального административного района: 4.38, за ним следует Северный административный округ: средний рейтинг заведений 4.24, и Северо-Западный административный округ 4.21. Самый низкий рейтинг в Юго-Восточном административном округе: 4.10. - В ТОП-15 улиц по количеству заведений вошли такие улицы, как: 1) проспект мира 2) профсоюзная улица 3) проспект вернадского 4) ленинский проспект 5) ленинградский проспект 6) дмитровское шоссе 7) каширское шоссе 8) варшавское шоссе 9) ленинградское шоссе 10) мкад 11) люблинская улица 12) улица вавилова 13) кутузовский проспект 14) улица миклухо-маклая 15) пятницкая улица Больше всего заведений расположено на проспекте Мира: свыше 180 предприятий общественного питания, из них в тройке лидеров кафе(53 заведений), рестораны(45 заведений) и кофейни (36 заведений). По остальным улицам типы заведений распределились также пропорционально: большая часть предприятий приходится на рестораны и кафе, за исключением МКАДа: 70% предприятий общественного питания составляют кафе - 45 заведений. - Улицы с одним заведением могут располагаться как в центре, так и по всем районам Москвы (в том числе на окраинах) равномерно, эти улочки скорее являются достаточно маленькими, переулками или проездами, но также встречаются и на известных набережных, например таких, как Софийская набережная и Раушская набережная. По большей части на 35% улицах одно единственное заведение - это кафе. Примерно в равных долях: 20% случаев и 17% случаев одно единственное заведение является рестораном или кофейней. Всего в данных 457 улиц с одним заведением общественного питания. - Самый высокий средний чек в Центральном и Западном административных округах, и составляет 1000 руб., что обусловлено самыми высокими ценами для центра города и преобладающим количеством ресторанов, а Западный административный округ является одним из самых престижных для проживания. Чем дальше заведение находится от Центрального административного округа и Западного административного округа, тем средний чек ниже. Самый низкий средний чек в Южном административном округе, в Юго-Восточном административном округе, в Северо-Восточном административном округе и составляет 450-500 руб. - В Москве на 90.7 % преобладают заведения, которые работают некруглосуточно, в то же время доля заведений, работающих круглосуточно составляет всего 9.3%. Наибольшее количество круглосуточных заведений приходится на кафе: 267 заведений, и на предприятия быстрого питания: 150 заведений. В процентном соотношении среди предприятий быстрого питания доля заведений, работающих круглосуточно наибольшая и составляет 26%. В тройку лидеров среди наибольшей доли круглосуточных заведений также вошли кафе: 13%, и круглосуточные булочные: 10%.

Открытие кофейни

```
In [125...]: #Отфильтруем все кофейни и положим их в переменную coffee_house
```

```
In [126...]: coffee_house = data.loc[data['category'] == 'кофейня']
coffee_house.head()
```

		name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street_name	is_24/7	seats_category
3	dormouse coffee shop	кофейня	москва, улица маршала федоренко, 12	Северный административный округ	ежедневно, 09:00-22:00	55.881608	37.488860	5.0	Nan	цена чашки капучино:155-185 ₽	NaN	NaN	170.0	0	2000.0	улица маршала федоренко	False	N/A
25	в парке вкуснее	кофейня	москва, парк левобережный	Северный административный округ	ежедневно, 10:00-21:00	55.878453	37.460028	4.3	Nan	NaN	NaN	NaN	NaN	1	2000.0	парк левобережный	False	N/A
45	9 bar coffee	кофейня	москва, коровинское шоссе, 41, стр. 1	Северный административный округ	пн-пт 08:00-18:00	55.885837	37.513422	4.0	Nan	NaN	NaN	NaN	NaN	1	46.0	коровинское шоссе	False	31-60
46	cofefest	кофейня	москва, улица маршала федоренко, 6с1	Северный административный округ	пн-пт 09:00-19:00	55.879934	37.492522	3.6	Nan	NaN	NaN	NaN	NaN	1	2000.0	улица маршала федоренко	False	N/A
52	cofix	кофейня	москва, улица дыбенко, 7/1	Северный административный округ	ежедневно, 08:00-22:00	55.878531	37.479395	3.8	Nan	NaN	NaN	NaN	NaN	1	2000.0	улица дыбенко	False	N/A

```
In [127...]: #Посчитаем количество кофеен в датасете
```

```
In [128...]: total_coffee = len(coffee_house)
total_coffee
```

```
Out[128]: 1413
```

Всего в данных представлено 1413 кофейни

```
In [129...]: #Посмотрим на распределение кофеен по районам
```

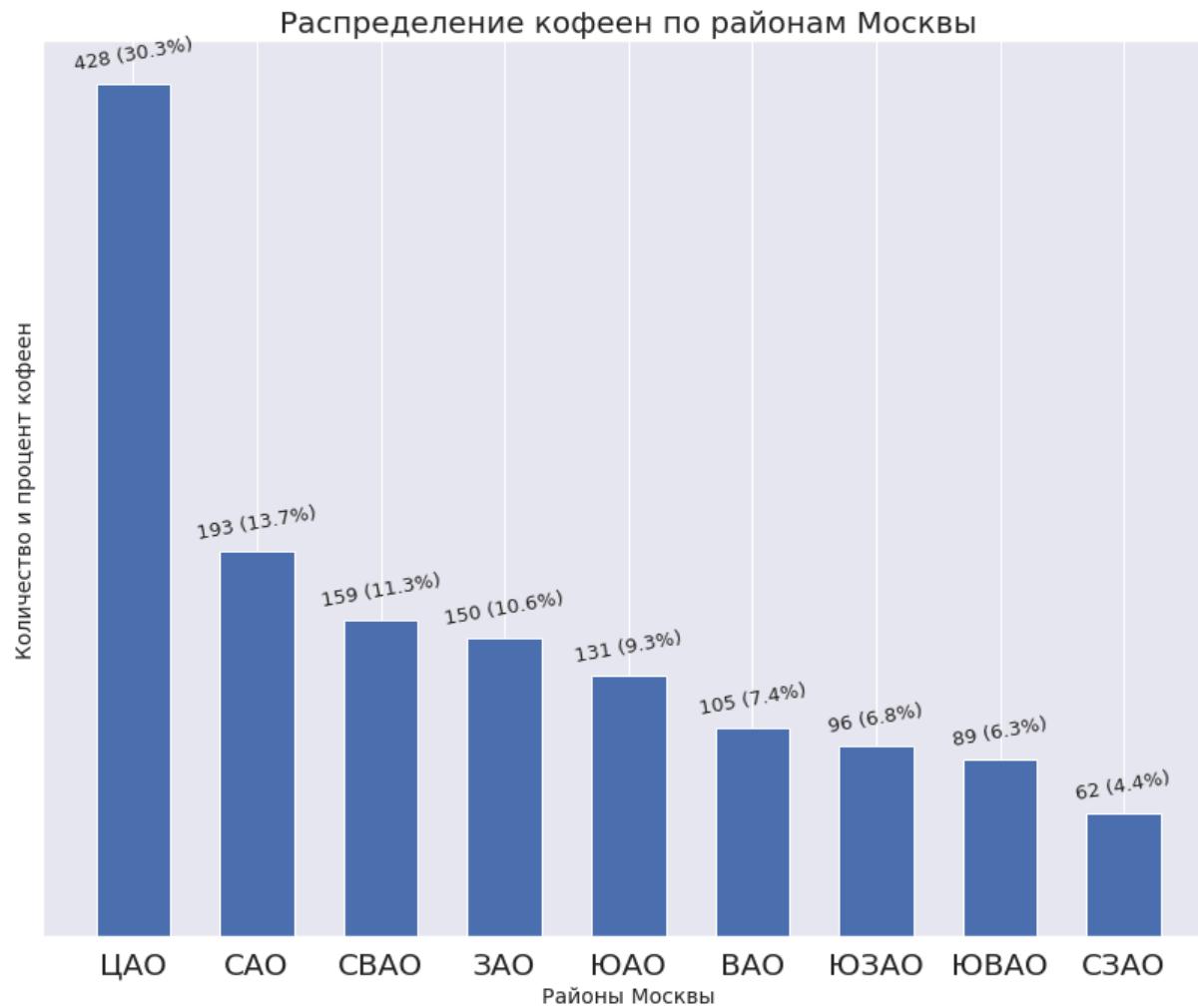
```
In [130...]: cofee_house_count = cofee_house.pivot_table(  
    index='district', values='name', aggfunc='count'  
)  
.reset_index()  
cofee_house_count.columns = ['Район Москвы', 'Количество кофеен']  
cofee_house_count = cofee_house_count.sort_values(by='Количество кофеен', ascending=False)  
cofee_house_count
```

Out[130]:

	Район Москвы	Количество кофеен
5	Центральный административный округ	428
2	Северный административный округ	193
3	Северо-Восточный административный округ	159
1	Западный административный округ	150
8	Южный административный округ	131
0	Восточный административный округ	105
7	Юго-Западный административный округ	96
6	Юго-Восточный административный округ	89
4	Северо-Западный административный округ	62

```
In [131...]: cofee_house_count['Район Москвы'] = cofee_house_count['Район Москвы'].replace(  
    {'Восточный административный округ':'ВАО',  
     'Западный административный округ':'ЗАО',  
     'Северный административный округ':'САО',  
     'Северо-Восточный административный округ':'СВАО',  
     'Северо-Западный административный округ':'СЗАО',  
     'Центральный административный округ':'ЦАО',  
     'Юго-Восточный административный округ':'ЮВАО',  
     'Юго-Западный административный округ':'ЮЗАО',  
     'Южный административный округ':'ЮАО'})
```

```
In [132...]: fig, ax = plt.subplots(figsize=(14, 11))  
  
width = 0.6 # Увеличиваем ширину столбцов  
padding = 5 # Отступ между меткой и столбцом  
bars = ax.bar(cofee_house_count['Район Москвы'], cofee_house_count['Количество кофеен'], width=width)  
  
for bar in bars:  
    height = bar.get_height()  
    percentage = height / total_coffee * 100  
    ax.text(bar.get_x() + bar.get_width() / 2,  
            height + padding, f'{height} ({percentage:.1f}%)',  
            ha='center',  
            va='bottom',  
            fontsize=13,  
            rotation=10)  
plt.xticks(fontsize=20)  
plt.yticks([])  
plt.xlabel('Районы Москвы', fontsize=14)  
plt.ylabel('Количество и процент кофеен', fontsize=14)  
ax.set_xticklabels(cofee_house_count['Район Москвы'])  
plt.title('Распределение кофеен по районам Москвы', fontsize=20)  
  
plt.show()
```



30% кофеен города расположились в Центральном административном округе, их количество составило 428 кофейни. Меньше всего кофеен расположено в Северо-Западном административном округе: на весь округ 62 кофейни, что составляет 4.4% всех кофеен города.

In [133...]

```
#Посчитаем, сколько в Москве круглосуточных и некруглосуточных кофеен,
#А также в каких районах города расположены круглосуточные кофейни
```

In [134...]

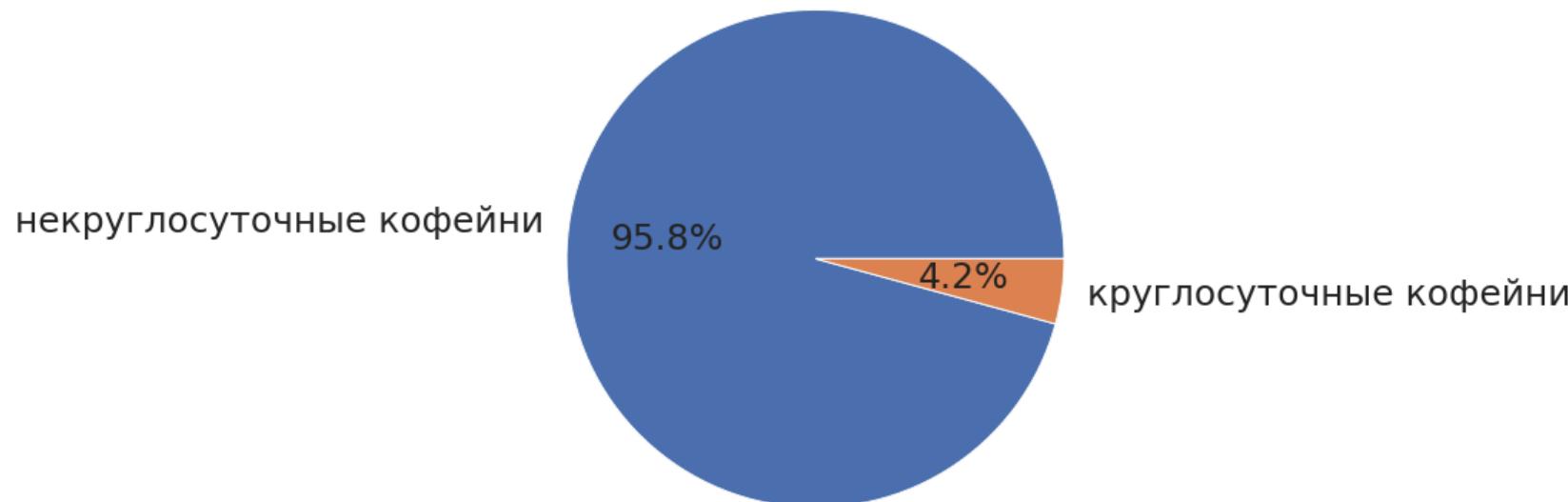
```
cofee_house_wt = cofee_house.pivot_table(
    index='is_24/7', values='name', aggfunc=['count']
)
cofee_house_wt.columns = ['count']
cofee_house_wt = cofee_house_wt.rename(index={True:'круглосуточные кофейни', False:'некруглосуточные кофейни'})
cofee_house_wt
```

Out[134]:

	count
is_24/7	
некруглосуточные кофейни	1339
круглосуточные кофейни	59

```
In [135...]  
cofee_house_wt.plot(y='count', kind='pie', figsize=(8,8), autopct='%.1f%%', textprops={'fontsize': 25});  
plt.ylabel('') # Удаляем подпись оси у  
plt.legend().remove()  
plt.title('Доли круглосуточных и некруглосуточных кофеен Москвы', fontsize=20) # Добавляем заголовок  
plt.show()
```

Доли круглосуточных и некруглосуточных кофеен Москвы



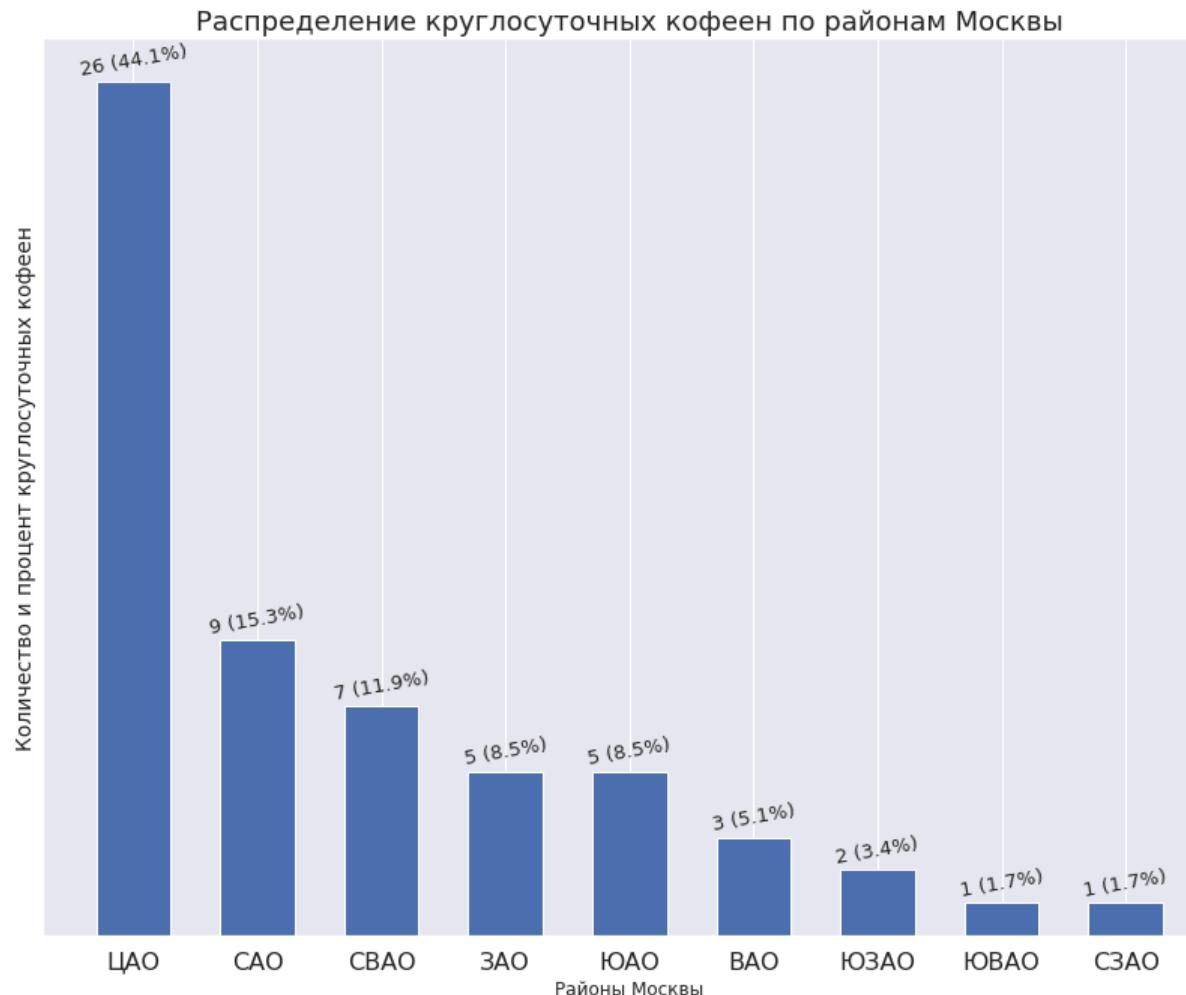
В Москве 95.8% или 1339 некруглосуточных кофеен, и всего 4.2% или 59 круглосуточных кофеен.

```
In [136...]  
#Посмотрим, в каких районах расположены круглосуточные кофейни
```

```
In [137...]  
cofee_house_district = cofee_house.loc[cofee_house['is_24/7'] == True]  
cofee_house_district = cofee_house_district.groupby('district').agg(count=('name','count')).reset_index()  
cofee_house_district = cofee_house_district.sort_values(by='count', ascending=False)
```

```
In [138...]  
fig, ax = plt.subplots(figsize=(14, 11))  
cofee_house_total = cofee_house_district['count'].sum()  
width = 0.6 # Увеличиваем ширину столбцов  
padding = 0.1 # Отступ между меткой и столбцом  
bars = ax.bar(cofee_house_district['district'], cofee_house_district['count'], width=width)  
  
for bar in bars:  
    height = bar.get_height()  
    percentage = height / cofee_house_total * 100  
    ax.text(bar.get_x() + bar.get_width() / 2,  
            height + padding, f'{height} ({percentage:.1f}%)',  
            ha='center',  
            va='bottom',  
            fontsize=13,  
            rotation=10)  
plt.xticks(fontsize=16)  
plt.yticks([])  
plt.xlabel('Районы Москвы', fontsize=12)  
plt.ylabel('Количество и процент круглосуточных кофеен', fontsize=14)
```

```
ax.set_xticklabels(cofee_house_count['Район Москвы'])
plt.title('Распределение круглосуточных кофеен по районам Москвы', fontsize=18)
plt.show()
```



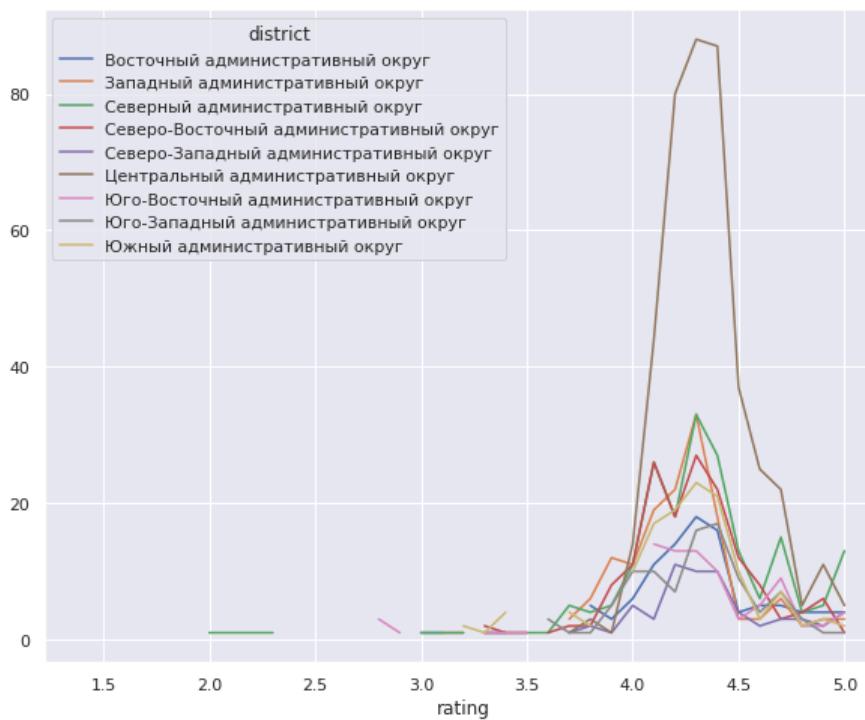
44% или 26 круглосуточных кофеен расположены в Центральном административном округе Москвы, что обусловлено высоким спросом на кофе в центре города в ночное время суток в связи с тем, что в центре ночью работают большое количество других круглосуточных заведений и предприятий. В Северном административном округе расположено всего 15% или 9 кофеен. В юго-восточном административном округе или северо-западном административном округе расположено всего по 1 круглосуточному кафе, что, вероятно обусловлено низким спросом на кофе ночью.

In [139...]

```
#Изучим рейтинги кофеен
#Визуализируем распределение рейтингов по районам Москвы
```

In [140...]

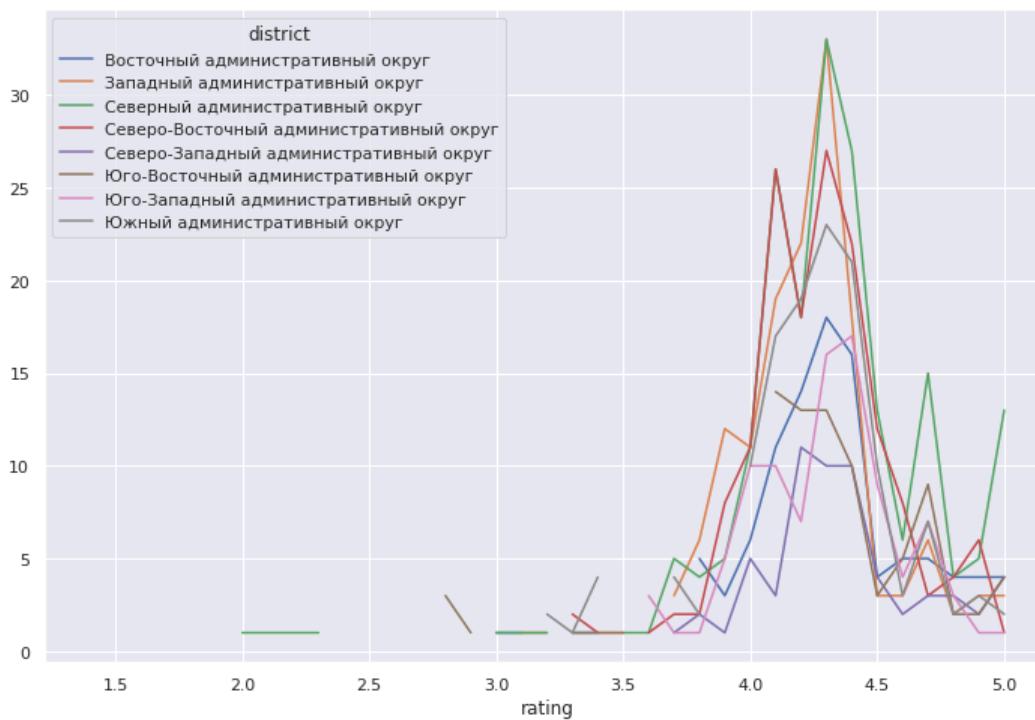
```
cofee_house.pivot_table(index='rating', columns='district', values='name', aggfunc='count').plot(figsize=(10,8));
```



В центральном административном округе чаще всего оценивают кофейни, что связано в принципе с самым большим числом кофеен для этого района Москвы. Для Центрального административного округа рейтинг большинства кофеен колеблется в пределах 4-4.7, хотя встречаются кофейни с маленьким рейтингом 3.7 и с высоким рейтингом 5, но таких кофеен не более 5 штук.

In [141...]: #Для наглядности, посмотрим на рейтинги кофеен без учета Центрального административного округа, удалив его с графика

In [142...]: cofee_house.query('district != "Центральный административный округ"').pivot_table(index='rating', columns='district', values='name', aggfunc='count').plot(figsize=(12,8));



В Северном административном округе большой разброс по рейтингам среди кофеен: встречаются кофейни с рейтингом 2.0 и 3.0, большинство кофеен округа имеют средний рейтинг от 4.0 до 4.5, и в тоже время среди всех других районов города в Северном административном округе больше всего кофеен с высоким рейтингом от 4.7 до 5.0. В других округах города количество кофеен с высоким рейтингом минимально. В остальных районах Москвы большинство кофеен с рейтингом от 3.7 до 4.5.

In [143]: #Составим сводную таблицу со средним рейтингом по районам

```
cofee_house_rate = coffee_house.groupby(['district']).agg({'rating':'mean'}).reset_index().sort_values(
    by='rating', ascending=False
)
cofee_house_rate['rating'] = round(cofee_house_rate['rating'], 1)
cofee_house_rate.columns = ['Административный округ', 'Рейтинг']
cofee_house_rate
```

Out[144]:

Административный округ Рейтинг

5	Центральный административный округ	4.3
4	Северо-Западный административный округ	4.3
2	Северный административный округ	4.3
7	Юго-Западный административный округ	4.3
0	Восточный административный округ	4.3
8	Южный административный округ	4.2
6	Юго-Восточный административный округ	4.2
3	Северо-Восточный административный округ	4.2
1	Западный административный округ	4.2

Что касается среднего рейтинга, то во всех районах он примерно одинаков и колеблется от 4.2 до 4.3.

```
In [145...]: #Посчитаем медианный чек на чашку капучино в зависимости от района Москвы  
  
In [146...]: median_coffee_cup = cofee_house.groupby('district').agg({'middle_coffee_cup' : 'median'}).reset_index()  
median_coffee_cup.columns = ['Район', 'Цена одной чашки капучино']  
median_coffee_cup = median_coffee_cup.sort_values(by='Цена одной чашки капучино', ascending=False)  
median_coffee_cup
```

Out[146]:

	Район	Цена одной чашки капучино
7	Юго-Западный административный округ	198.0
5	Центральный административный округ	190.0
1	Западный административный округ	189.0
4	Северо-Западный административный округ	165.0
3	Северо-Восточный административный округ	162.5
2	Северный административный округ	159.0
8	Южный административный округ	150.0
6	Юго-Восточный административный округ	147.5
0	Восточный административный округ	135.0

```
In [147...]: #Посчитаем медианную стоимость одной чашки капучино
```

```
In [148...]: print(  
    'Цена одной чашки капучино', median_coffee_cup['Цена одной чашки капучино'].median(), 'руб.'  
)
```

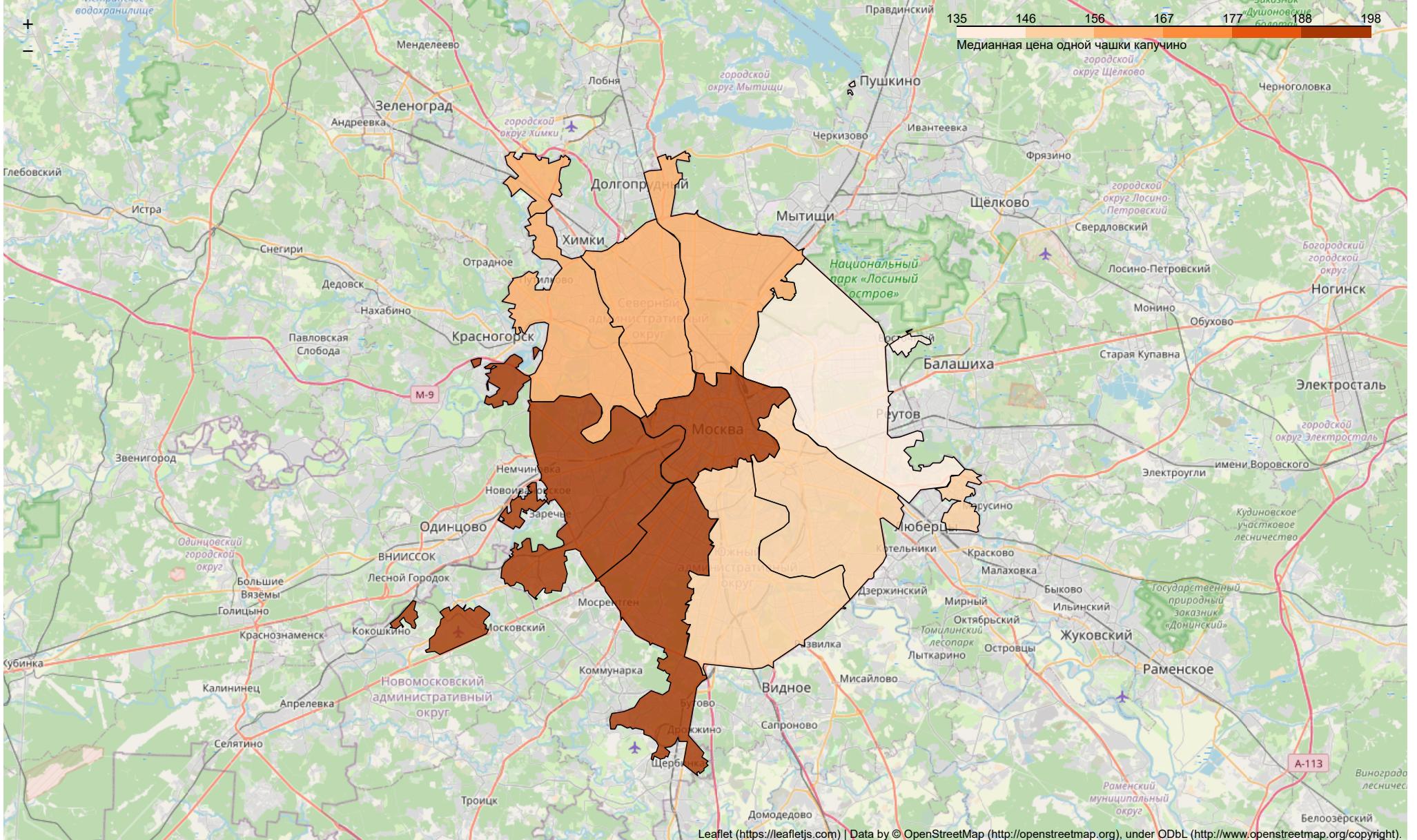
Цена одной чашки капучино 162.5 руб.

```
In [149...]: #Визуализируем на фоновой картошрамме распределение по медианному чеку на чашку капучино
```

```
In [150...]: # moscow_lat - широта центра Москвы, moscow_lng - долгота центра Москвы  
moscow_lat, moscow_lng = 55.751244, 37.618423
```

```
# создаем карту Москвы  
m = Map(location=[moscow_lat, moscow_lng], zoom_start=10)  
k=folium.Choropleth(  
    geo_data=geo,  
    data=median_coffee_cup,  
    columns=['Район', 'Цена одной чашки капучино'],  
    key_on='feature.name',  
    fill_color='Oranges',  
    fill_opacity=0.8,  
    legend_name='Медианная цена одной чашки капучино',  
    legend_font_size='30px'  
) .add_to(m)  
  
#Добавим интерактивное отражение названий районов  
k.geojson.add_child(folium.features.GeoJsonTooltip(['name'], labels=False))  
folium.LayerControl().add_to(m)  
# выводим карту  
m
```

Out[150]:



```
In [151...]: cofee_house['district'] = cofee_house['district'].replace({'Восточный административный округ': 'BAO',  
                                                               'Западный административный округ': 'ЗАО',  
                                                               'Северный административный округ': 'САО',  
                                                               'Северо-Восточный административный округ': 'СВАО',  
                                                               'Северо-Западный административный округ': 'СЗАО',  
                                                               'Центральный административный округ': 'ЦАО',  
                                                               'Юго-Восточный административный округ': 'ЮВАО',  
                                                               'Юго-Западный административный округ': 'ЮЗАО',  
                                                               'Южный административный округ': 'ЮАО'})
```

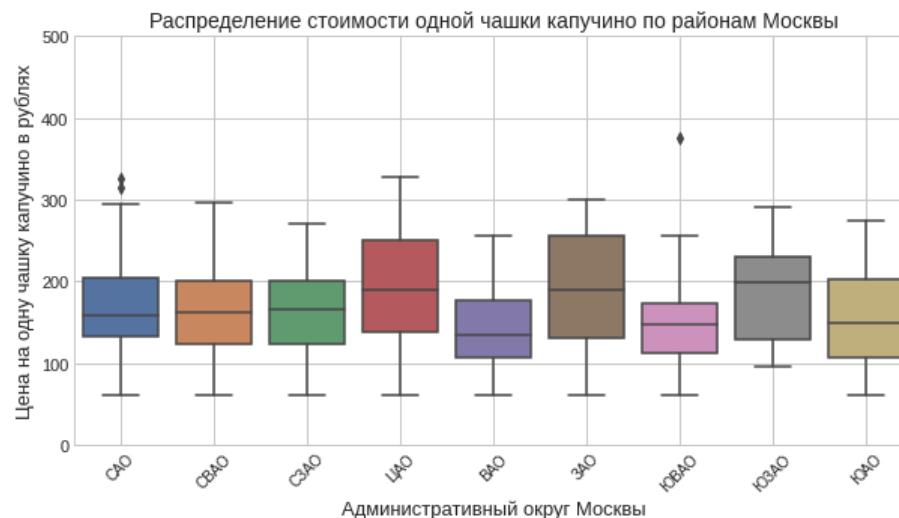
```
In [152...]: plt.style.use('seaborn-whitegrid')
sns.set_palette('deep')
plt.figure(figsize=(10, 5))
```

```

sns.boxplot(data=cofee_house.loc[:, ['district', 'middle_coffee_cup']], x='district', y='middle_coffee_cup')
plt.title('Распределение стоимости одной чашки капучино по районам Москвы', size=14)
plt.xlabel('Административный округ Москвы', size=13)
plt.ylabel('Цена на одну чашку капучино в рублях', size=13)

plt.xticks(rotation=45, size=10)
plt.ylim(0,500)
plt.grid(True)
plt.show()

```



Медианная цена одной чашки капучино по районам Москвы составляет 162.5 руб. Самая дорогая чашка капучино в Юго-Западном административном округе и составляет 198 руб., а также в Центральном административном округе и составляет 190 руб. Самая низкая медианная цена на чашку капучино в Восточном административном округе и составляет 135 руб.

In [153...]: #Визуализируем все кофейни из предоставленных данных на карте Москвы

In [154...]: cofee_house_high_rating = cofee_house.query('rating <= 3.0')

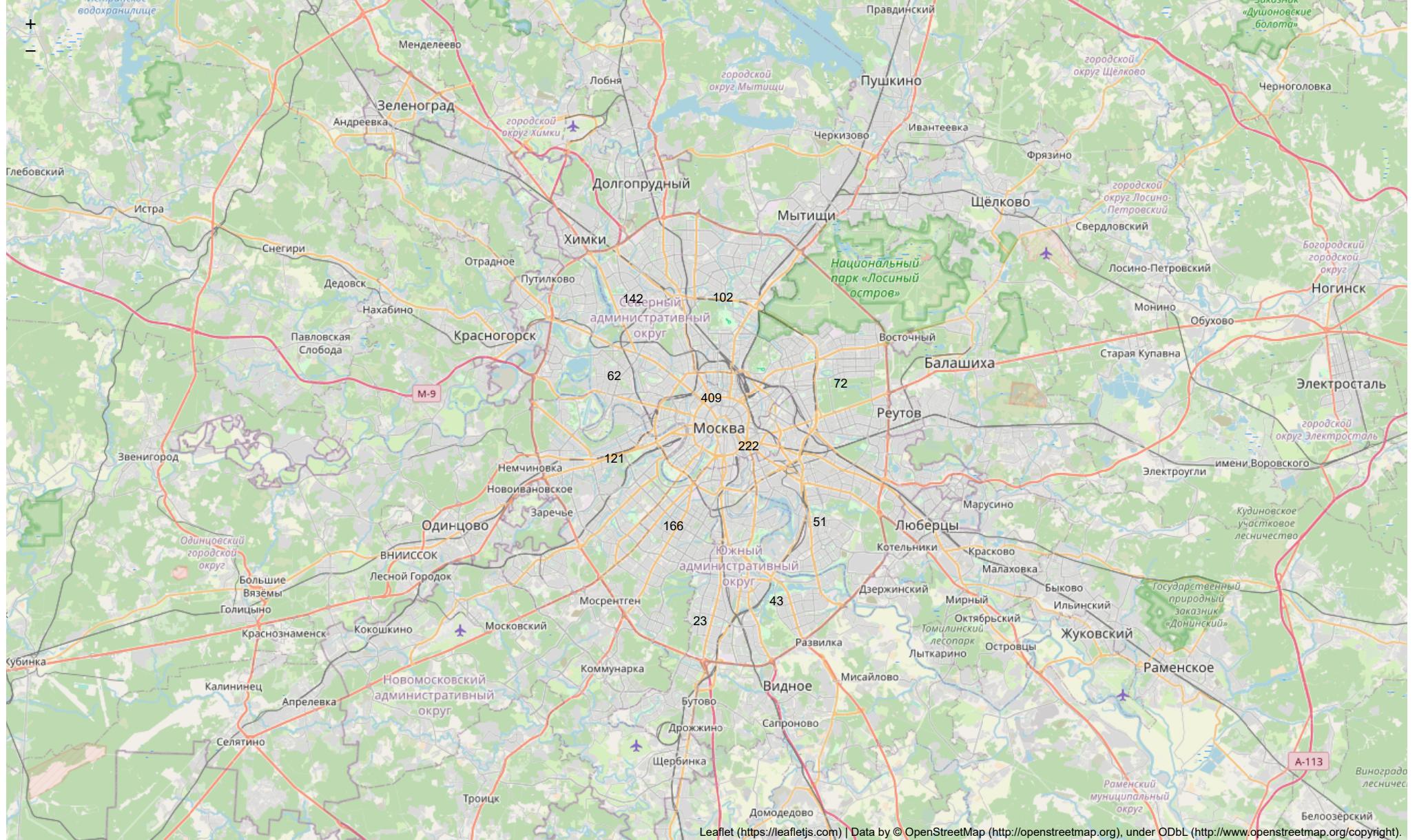
In [155...]: moscow_lat, moscow_lng = 55.751244, 37.618423
создаём карту Москвы
m_Moscow = Map(location=[moscow_lat, moscow_lng], zoom_start=10)
marker_cluster = MarkerCluster().add_to(m_Moscow)

пишем функцию, которая принимает строку датафрейма,
создаёт маркер в текущей точке и добавляем его в кластер marker_cluster
def create_clusters(row):
 Marker(
 [row['lat'], row['lng']],
 popup=f'{row['name']} {row['rating']} {row['middle_coffee_cup']}')
 .add_to(marker_cluster)

применяем функцию create_clusters() к каждой строке датафрейма
cofee_house.apply(create_clusters, axis=1)

выводим карту
m_Moscow

Out[155]:



In [156...]

```
#Выведем кофейни с низким рейтингом менее 3.0 на карте Москвы
```

In [157...]

```
cofee_house_low_rating = cofee_house.query('rating <= 3.0')
```

In [158...]

```
moscow_lat, moscow_lng = 55.751244, 37.618423
# создаём карту Москвы
m_Moscow = Map(location=[moscow_lat, moscow_lng], zoom_start=10)
marker_cluster = MarkerCluster().add_to(m_Moscow)
```

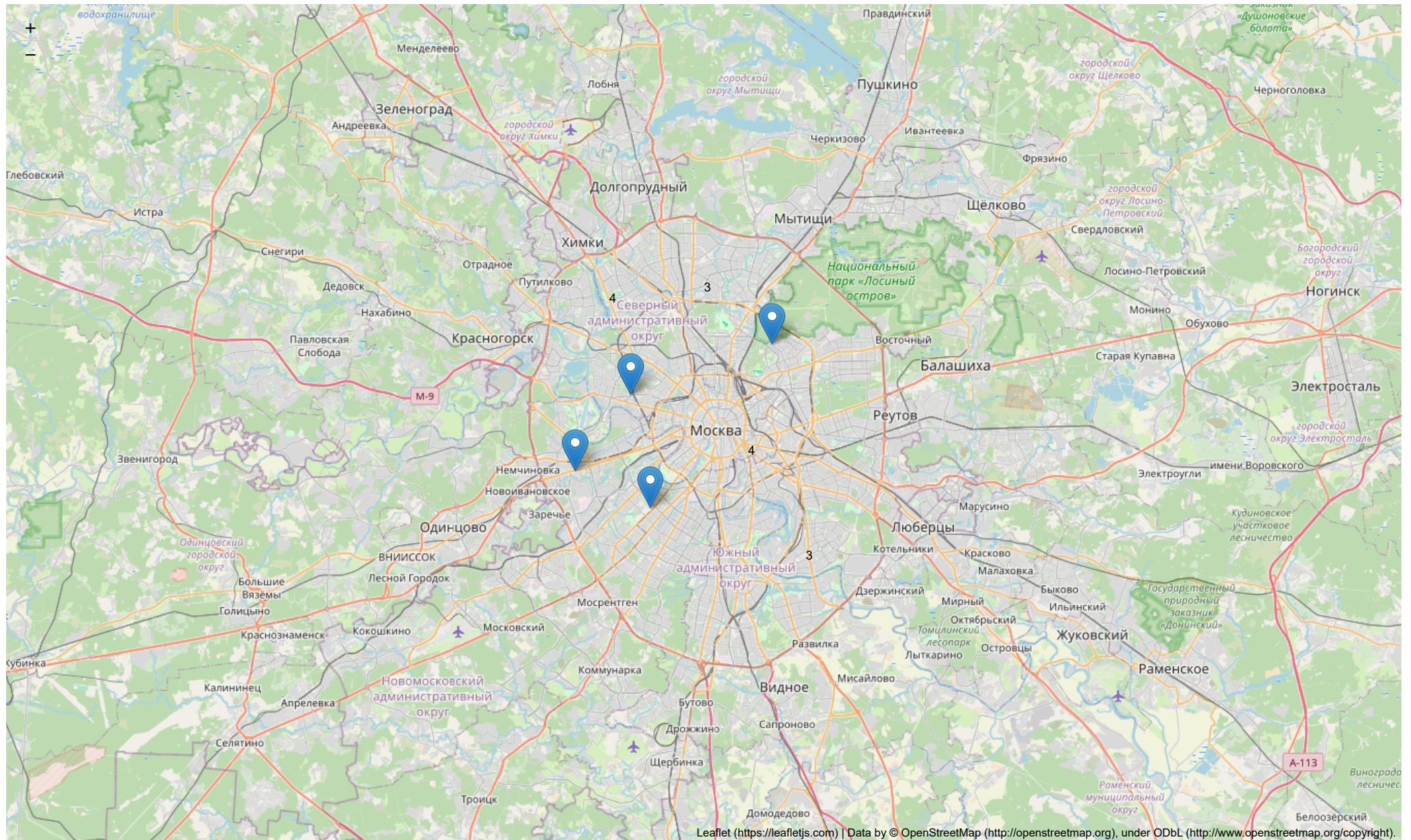
```
# пишем функцию, которая принимает строку датагрейма,
# создаёт маркер в текущей точке и добавляем его в кластер marker_cluster
def create_clusters(row):
```

```
Marker(  
    [row['lat'], row['lng']],  
    popup=f'{row['name']} {row['rating']} {row['middle_coffee_cup']}"  
) .add_to(marker_cluster)
```

```
# применяем функцию create_clusters() к каждой строке датафрейма  
coffee_house_low_rating.apply(create_clusters, axis=1)
```

```
# выводим карту  
m_Moscow
```

Out[158]:



Для открытия кофейни рекомендовано выбрать точку с большим проходным потоком людей, или вблизи функционирующих предприятий (больниц, бизнес-центров и других) поблизости с кофейнями с низким рейтингом в Центральном или Северном административном округах, так как в указанных районах присутствует спрос на кофе. К примеру, в центре Москвы, на ул. Гончарная, 23, или 5-й Котельничий

переулок, рядом с клиникой "Женское здоровье", рядом с клиникой "Остеон", рядом с Музеем Русского зарубежья, рядом с Московским институтом открытого образования и участковым пунктом полиции, и другими предприятиями. Стоимость чашки капучино установить от 160 до 200 руб. в зависимости от района города.

Вывод: Всего в данных представлено 1413 кофейни. 30% кофеен города расположились в Центральном административном округе, их количество составило 428 кофейни. Меньше всего кофеен расположено в Северо-Западном административном округе: на весь округ 62 кофейни, что составляет 4.4% всех кофеен города. В Москве 95.8% или 1339 некруглосуточных кофеен, и всего 4.2% или 59 круглосуточных кофеен.

44% или 26 круглосуточных кофеен расположены в Центральном административном округе Москвы, что обусловлено высоким спросом на кофе в центре города в ночное время суток в связи с тем, что в центре ночью работают большое количество других круглосуточных заведений и предприятий. В Северном административном округе расположено всего 15% или 9 кофеен. В юго-восточном административном округе или северо-западном административном округе расположено всего по 1 круглосуточному кафе, что, вероятно обусловлено низким спросом на кофе ночью.

В центральном административном округе чаще всего оценивают кофейни, что связано в принципе с самым большим числом кофеен для этого района Москвы. Для Центрального административного округа рейтинг большинства кофеен колеблется в пределах 4-4.7, хотя встречаются кофейни с маленьким рейтингом 3.7 и с высоким рейтингом 5, но таких кофеен не более 5 штук.

В Северном административном округе большой разброс по рейтингам среди кофеен: встречаются кофейни с рейтингом 2.0 и 3.0, большинство кофеен округа имеют средний рейтинг от 4.0 до 4.5, и в тоже время среди всех других районов города в Северном административном округе больше всего кофеен с высоким рейтингом от 4.7 до 5.0. В других округах города количество кофеен с высоким рейтингом минимально. В остальных районах Москвы большинство кофеен с рейтингом от 3.7 до 4.5.

Что касается среднего рейтинга, то во всех районах он примерно одинаков и колеблется от 4.2 до 4.3.

Медианная цена одной чашки капучино по районам Москвы составляет 162.5 руб. Самая дорогая чашка капучино в Юго-Западном административном округе и составлят 198 руб., а также в Центральном административном округе и составляет 190 руб. Самая низкая медианская цена на чашку капучино в Восточном административном округе и составляет 135 руб.

Рекомендации по открытию кофейни:

Для открытия кофейни рекомендовано выбрать точку с большим проходным потоком людей, или вблизи функционирующих предприятий (больниц, бизнес-центров и других) поблизости с кофейнями с низким рейтингом в Центральном или Северном административных округах, так как в указанных районах присутствует спрос на кофе. К примеру, в центре Москвы, на ул. Гончарная, 23, или 5-й Котельничий переулок, рядом с клиникой "Женское здоровье", рядом с клиникой "Остеон", рядом с Музеем Русского зарубежья, рядом с Московским институтом открытого образования и участковым пунктом полиции, и другими предприятиями: на пересечении указанных улиц функционирует всего одна кофейня с низким рейтингом, при этом логично предположить по количеству функционирующих рядом предприятий, что спрос на кофе присутствует. Стоимость чашки капучино установить от 160 до 200 руб. в зависимости от района города.

****ОБЩИЙ ВЫВОД:**** **1. Обзор данных.** В ходе обзора данных выявлено, что в таблице 8406 строк и 14 столбцов. Некоторые столбцы содержат пропуски. В столбце seats тип данных float64, но логичнее чтобы тип данных был int, так как количество посадочных мест - целое число. Также в столбце seatsминимальное количество посадочных мест составляет 0, это нелогично. В столбце float64 медиана и среднее значительно отличаются, что означает выбросы в данных. **2. Предобработка данных.** В ходе предобработки данных исследованы пропуски, дубликаты. Пропущены значения в столбцах price, avg_bill, middle_avg_bill, middle_coffee_cup, seats. Пропуски оставлены, поскольку любые методы заполнения будут не точны и исказят данные. **3. Анализ данных.** - Самые распространенная категория заведений Москвы - это кафе. Кафе занимают долю в размере 28.3% от всех предприятий общественного питания Москвы, представленных в данных. С небольшим отрывом по популярности идут рестораны : 24.3%. На третьем и четвертом месте находятся кофейни (16.8%) и пабы (9.1%). Наименее популярные в Москве столовые и булочные, доля которых составила 3.7% и 3%. - Для каждой категории заведений большую часть занимают заведения с отсутствующими данными по количеству посадочных мест. Наибольшая доля - 7.6% или 370 заведений приходится на рестораны с количеством посадочных мест от 61 до 100, и столько же - 7.6% или 370 заведений приходится на кафе с количеством посадочных мест от 31 до 60. Также в Москве распространены кофейни примерно с равнозначной долей по количеству посадочных мест от 31 до 60 - 165 заведений или 3.5% от всех заведений города, от 61 до 100 - 160 заведений или 3.4% от всех заведений города, от 101 до 200 - 153 заведений от 3.2% от всех заведений города. - В Москве преимущественно работают несетевые заведения, их доля составляет 61.9%. Доля сетевых заведений составила 38.1%. Наибольшая доля сетевых заведений среди булочных и составляет 60%. Для пиццерий и кофеин доля для сетевых и несетевых заведений разделена поровну и составляет 50%. Среди заведений быстрого питания, ресторанов, кафе, столовых, баров преобладает доля несетевых заведений. Наиболее редко сетевые заведения встречаются среди баров и пабов: в 22% случаев. - Всего в датасете 747 сетевых заведений с уникальным названием. Самая популярная и распространенная сеть предприятий общественного питания - это кофейня "Шоколадница", всего по Москве 120 точек Шоколадницы. На втором и третьем месте по популярности находятся пиццерии: "Домино'с пицца" и "Додо пицца": 76 и 74 точки питания. - Среди самых популярных сетевых заведений Москвы наиболее распространены кофейни: 337 заведений, а также рестораны: 186 заведений. На третьем месте по распространенности находятся пиццерии: 152 заведения. В десятку также входят кафе, булочные, предприятия быстрого питания, бары и пабы, и столовые. - Больше всего предприятий общественного питания расположено в центральном административном округе (ЦАО): 2400 заведений, что более, чем на 100% больше, чем в любом другом округе Москвы. В ЦАО преобладают рестораны: 670 заведений данного типа, что составляет примерно 28% от всех

заведений округа. На втором и третьем месте в ЦАО по популярности находятся кафе (464 заведения или 19%) и кофейни(428 заведений или 17%). Наиболее редки булочные (50 заведений или 2%) и столовые (66 заведений или 2.7%). В остальных районах города на первом месте преобладают кафе, а на втором месте по численности находятся рестораны. Менее развиты булочные и столовые. - Меньше всего предприятий общественного питания находится в Северо-западном административном округе (всего около 400 заведений). - Средний рейтинг заведений всех заведений Москвы довольно высокий и составляет 4.23. Если смотреть в разрезе категорий, то самый высокий рейтинг у баров и пабов: 4.39; средний рейтинг пиццерий и ресторанов примерно одинаковый и составляет 4.3 и 4.29. Самый низкий рейтинг у предприятий быстрого питания: 4.05. Усредненные рейтинги заведений по типам питания различаются минимально Самый высокий средний рейтинг у Центрального административного района: 4.38, за ним следует Северный административный округ: средний рейтинг заведений 4.24, и Северо-Западный административный округ 4.21. Самый низкий рейтинг в Юго-Восточном административном округе: 4.10. - В ТОП-15 улиц по количеству заведений вошли такие улицы, как: 1) проспект мира 2) профсоюзная улица 3) проспект вернадского 4) ленинский проспект 5) ленинградский проспект 6) дмитровское шоссе 7) каширское шоссе 8) варшавское шоссе 9) ленинградское шоссе 10) мкад 11) люблинская улица 12) улица вавилова 13) кутузовский проспект 14) улица миклухо-маклая 15) пятницкая улица Больше всего заведений расположено на проспекте Мира: свыше 180 предприятий общественного питания, из них в тройке лидеров кафе(53 заведений), рестораны(45 заведений) и кофейни (36 заведений). По остальным улицам типы заведений распределились также пропорционально: большая часть предприятий приходится на рестораны и кафе, за исключением МКАДа: 70% предприятий общественного питания составляют кафе - 45 заведений. - Улицы с одним заведением могут располагаться как в центре, так и по всем районам Москвы (в том числе на окраинах) равномерно, эти улочки скорее являются достаточно маленькими, переулками или проездами, но также встречаются и на известных набережных, например таких, как Софийская набережная и Раушская набережная. По большей части на 35% улицах одно единственное заведение - это кафе. Примерно в равных долях: 20% случаев и 17% случаев одно единственное заведение является рестораном или кофейней. Всего в данных 457 улиц с одним заведением общественного питания. - Самый высокий средний чек в Центральном и Западном административных округах, и составляет 1000 руб., что обусловлено самыми высокими ценами для центра города и преобладающим количеством ресторанов, а Западный административный округ является одним из самых престижных для проживания. Чем дальше заведение находится от Центрального административного округа и Западного административного округа, тем средний чек ниже. Самый низкий средний чек в Южном административном округе, в Юго-Восточном административном округе, в Северо-Восточном административном округе и составляет 450-500 руб. - В Москве на 90.7 % преобладают заведения, которые работают некруглосуточно, в то же время доля заведений, работающих круглосуточно составляет всего 9.3%. Наибольшее количество круглосуточных заведений приходится на кафе: 267 заведений, и на предприятия быстрого питания: 150 заведений. В процентном соотношении среди предприятий быстрого питания доля заведений, работающих круглосуточно наибольшая и составляет 26%. В тройку лидеров среди наибольшей доли круглосуточных заведений также вошли кафе: 13%, и круглосуточные булочные: 10%. **4. Открытие кофейни** Всего в данных представлено 1413 кофейни. 30% кофеен города расположились в Центральном административном округе, их количество составило 428 кофейни. Меньше всего кофеен расположено в Северо-Западном административном округе: на весь округ 62 кофейни, что составляет 4.4% всех кофеен города. В Москве 95.8% или 1339 некруглосуточных кофеен, и всего 4.2% или 59 круглосуточных кофеен. 44% или 26 круглосуточных кофеен расположены в Центральном административном округе Москвы, что обусловлено высоким спросом на кофе в центре города в ночное время суток в связи с тем, что в центре ночью работают большое количество других круглосуточных заведений и предприятий. В Северном административном округе расположено всего 15% или 9 кофеен. В юго-восточном административном округе или северо-западном административном округе расположено всего по 1 круглосуточному кафе, что, вероятно обусловлено низким спросом на кофе ночью. В центральном административном округе чаще всего оценивают кофейни, что связано в принципе с самым большим числом кофеен для этого района Москвы. Для Центрального административного округа рейтинг большинства кофеен колеблется в пределах 4-4.7, хотя встречаются кофейни с маленьким рейтингом 3.7 и с высоким рейтингом 5, но таких кофеен не более 5 штук. В Северном административном округе большой разброс по рейтингам среди кофеен: встречаются кофейни с рейтингом 2.0 и 3.0, большинство кофеен округа имеют средний рейтинг от 4.0 до 4.5, и в тоже время среди всех других районов города в Северном административном округе больше всего кофеен с высоким рейтингом от 4.7 до 5.0. В других округах города количество кофеен с высоким рейтингом минимально. В остальных районах Москвы большинство кофеен с рейтингом от 3.7 до 4.5. Что касается среднего рейтинга, то во всех районах он примерно одинаков и колеблется от 4.2 до 4.3. Медианная цена одной чашки капучино по районам Москвы составляет 162.5 руб. Самая дорогая чашка капучино в Юго-Западном административном округе и составляет 198 руб., а также в Центральном административном округе и составляет 190 руб. Самая низкая медианная цена на чашку капучино в Восточном административном округе и составляет 135 руб.

Рекомендации по открытию кофейни: Для открытия кофейни рекомендовано выбрать точку с большим проходным потоком людей, или вблизи функционирующих предприятий (больниц, бизнес-центров и других) поблизости с кофейнями с низким рейтингом в Центральном или Северном административном округах, так как в указанных районах присутствует спрос на кофе. К примеру, в центре Москвы, на ул. Гончарная, 23, или 5-й Котельничий переулок, рядом с клиникой "Женское здоровье", рядом с клиникой "Остеон", рядом с Музеем Русского зарубежья, рядом с Московским институтом открытого образования и участковым пунктом полиции, и другими предприятиями: на пересечении указанных улиц функционирует всего одна кофейня с низким рейтингом, при этом логично предположить по количеству функционирующих рядом предприятий, что спрос на кофе присутствует. При открытии кофейни ограничиться вместимостью 30-60 посадочных мест. Стоимость чашки капучино установить от 160 до 200 руб. в зависимости от района города. При выборе места для открытия кафе следует избегать улиц, входящих в ТОП-15 по количеству заведений общественного питания.