

Анализ рынка заведений общественного питания Москвы

Описание проекта: 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.

Цель проекта: Необходимо подготовить исследование рынка Москвы, найти интересные особенности и презентовать полученные результаты, которые в будущем помогут в выборе подходящего инвесторам места.

Описание данных: Доступен датасет с заведениями общественного питания Москвы, составленный на основе данных сервисов Яндекс Карты и Яндекс Бизнес на лето 2022 года. 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;
- `lng` — 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 Р»;
 - «Цена чашки капучино: 130–220 Р»;
 - «Цена бокала пива: 400–600 Р».

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.

Импортируем библиотеки

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import plotly.express as px
from plotly import graph_objects as go
import re
import json
from folium import Map, Choropleth
from folium import Map, Marker
from folium.plugins import MarkerCluster
from folium.features import CustomIcon

# подключаем модуль для работы с JSON-форматом
import json
# читаем файл и сохраняем в переменной
with open('C:/Users/dimch/OneDrive/Рабочий стол/Яндекс Практикум/Datasets/admin_level_geomap.geojson', 'r') as f:
    geo_json = json.load(f)
```

Шаг. Загрузите данные и изучите общую информацию

```
In [2]: # Откроем файл
path = 'C:/Users/dimch/OneDrive/Рабочий стол/Яндекс Практикум/Datasets/'
data = pd.read_csv(path + 'moscow_places.csv')
# Выведем первые 20 строк
data.head(20)
```

Out[2]:		name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats
		0	WoWfli	кафе	Москва, улица Дыбенко, 7/1	Северный административный округ	ежедневно, 10:00–22:00	55.878494	37.478860	5.0	NaN	NaN	NaN	NaN	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 Р	1550.0	NaN	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 Р	1000.0	NaN	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 Р	NaN	170.0	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 Р	500.0	NaN	148.0
		5	Sergio Pizza	пиццерия	Москва, Ижорская улица, вл8Б	Северный административный округ	ежедневно, 10:00–23:00	55.888010	37.509573	4.6	средние	NaN	NaN	NaN	NaN
		6	Огни города	бар,паб	Москва, Клязьминская улица, 9, стр. 3	Северный административный округ	пн 15:00–04:00; вт-вс 15:00–05:00	55.890752	37.524653	4.4	средние	Средний счёт:199 Р	199.0	NaN	45.0
		7	Mr. Уголёк	быстрое питание	Москва, Клязьминская улица, 9, стр. 3	Северный административный округ	пн-чт 10:00–22:00; пт,сб 10:00–23:00; вс 10:00...	55.890636	37.524303	4.7	средние	Средний счёт:200–300 Р	250.0	NaN	45.0
		8	Donna Maria	ресторан	Москва, Дмитровское шоссе, 107, корп. 4	Северный административный округ	ежедневно, 10:00–22:00	55.880045	37.539006	4.8	средние	Средний счёт:от 500 Р	500.0	NaN	79.0
		9	Готика	кафе	Москва, Ангарская улица, 39	Северный административный округ	ежедневно, 12:00–00:00	55.879038	37.524487	4.3	средние	Средний счёт:1000–1200 Р	1100.0	NaN	65.0
		10	Great Room Bar	бар,паб	Москва, Левобережная улица, 12	Северный административный округ	ежедневно, круглосуточно	55.877832	37.469171	4.5	средние	Цена бокала пива:250–350 Р	NaN	NaN	102.0
		11	Шашлык Шефф	кафе	Москва, улица Маршала Федоренко, 10с1	Северный административный округ	ежедневно, 10:00–21:00	55.881770	37.492362	4.9	NaN	NaN	NaN	NaN	NaN
		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 Р	330.0	NaN	NaN
		13	Буханка	булочная	Москва, Базовская улица, 15, корп. 1	Северный административный округ	ежедневно, 08:00–22:00	55.877007	37.504980	4.8	NaN	NaN	NaN	NaN	180.0
		14	У Сильвы	бар,паб	Москва, Ангарская улица, 42с1	Северный административный округ	ежедневно, 13:00–00:00	55.885528	37.528371	4.2	выше среднего	Средний счёт:1500 Р	1500.0	NaN	NaN
		15	Дом обеда	столовая	Москва, улица Бусиновская Горка, 2	Северный административный округ	пн-пт 08:30–18:30; сб 10:00–20:00	55.885890	37.493264	4.1	средние	Средний счёт:300–500 Р	400.0	NaN	180.0
		16	База Стритфуд	кафе	Москва, Базовская улица, 15, корп. 8	Северный административный округ	ежедневно, 10:00–23:00	55.877859	37.507754	4.2	средние	Средний счёт:140–350 Р	245.0	NaN	NaN
		17	Чайхана Беш-Бармак	ресторан	Москва, Ленинградское шоссе, 71Б, стр. 2	Северный административный округ	ежедневно, круглосуточно	55.876908	37.449876	4.4	средние	Средний счёт:350–500 Р	425.0	NaN	96.0
		18	Час-Пик	столовая	Москва, Коровинское шоссе, 30А	Северный административный округ	ежедневно, 09:00–21:00	55.884651	37.517482	4.3	средние	Средний счёт:200–300 Р	250.0	NaN	25.0
		19	Пекарня	булочная	Москва, Ижорский проезд, 5	Северный административный округ	ежедневно, круглосуточно	55.887969	37.515688	4.4	NaN	NaN	NaN	NaN	NaN

```
In [3]: # Выведем информацию о данных
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 [4]: # Посчитаем количество пропусков
data.isna().sum()
```

```
Out[4]: 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 [5]: # Вывожу список характерных значений для столбцов
data.describe()
```

	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:

- Перед нами датасет, который представляет собой список заведений москвы с рейтингами, координатами, информацией о ценах и прочими данными;
- Всего в датасете 8406 строк, каждая из которых содержит информацию о заведении общественного питания Москвы, после проверки на дубликаты можно будет точнее посчитать их количество;
- Данные типа object содержатся в столбцах 'name', 'category', 'address', 'district', 'hours', 'price', 'avg_bill';
- Данные типа float содержатся в столбцах 'lat', 'lng', 'address', 'rating', 'middle_avg_bill' и 'middle_coffee_cup';
- В столбце chain содержатся данные типа int, где 1 соответствует сетевому заведению, а 0 несетевому;
- В некоторых столбцах встречаются пропуски. Больше всего пропусков в столбцах middle_coffee_cup (7871), price (5091) и middle_avg_bill (5257).

Шаг. Выполните преобработку данных

Обработка дубликатов

```
In [6]: # Посчитаем дубликаты
print('Количество явных дубликатов:', data.duplicated().sum())
print('Количество дубликатов по названию заведения:',data.duplicated(subset = 'name').sum())
```

Количество явных дубликатов: 0
Количество дубликатов по названию заведения: 2792

```
In [7]: # Посчитаем дубликаты по названию и координатам заведения
data[data.duplicated(subset = ['name', 'lat', 'lng'])]
```

Out[7]:

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats
--	------	----------	---------	----------	-------	-----	-----	--------	-------	----------	-----------------	-------------------	-------	-------

```
In [8]: # Посчитаем дубликаты по названию и адресу заведения
data[data.duplicated(subset = ['name', 'address'])]
```

Out[8]:

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats
--	------	----------	---------	----------	-------	-----	-----	--------	-------	----------	-----------------	-------------------	-------	-------

Информации о дубликатах нет, попробуем привести данные к единой форме и посчитать снова.

```
In [9]: data['name'] = data['name'].replace(" ", " ")
data['address'] = data['address'].replace(" ", " ")
data['name'] = data['name'].str.replace('ё', 'е')
data['address'] = data['address'].str.replace('ё', 'е')
data.name = data['name'].str.lower()
data['address'] = data['address'].str.lower()
```

```
In [10]: # Снова посчитаем дубликаты по названию и адресу заведения
data[data.duplicated(subset = ['name', 'address'], keep = False)]
```

Out[10]:

	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	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	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	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	1	188.0
2211	раковарня клешни и хвосты	ресторан	москва, проспект мира, 118	Северо-Восточный административный округ	ежедневно, 12:00–00:00	55.810553	37.638161	4.4	NaN	NaN	NaN	NaN	0	150.0
2420	раковарня клешни и хвосты	бар,лаб	москва, проспект мира, 118	Северо-Восточный административный округ	пн-чт 12:00–00:00; пт,сб 12:00–01:00; вс 12:00...	55.810677	37.638379	4.4	NaN	NaN	NaN	NaN	1	150.0
3091	хлеб да выпечка	булочная	москва, ярцевская улица, 19	Западный административный округ	ежедневно, 09:00–22:00	55.738886	37.411648	4.1	NaN	NaN	NaN	NaN	1	276.0
3109	хлеб да выпечка	кафе	москва, ярцевская улица, 19	Западный административный округ	NaN	55.738449	37.410937	4.1	NaN	NaN	NaN	NaN	0	276.0

У заведений 189 и 215 совпадают названия и адреса, но координаты и режим работы разные, по карте это два разных заведения, 189-get&fly, а 215- это Гриль у пруда & кофейня. Остальные пункты просто дублируют друг друга. Переименую первые два, а дубликаты удалю по индексам.

```
In [11]: data.loc[189, 'name'] = data.loc[189, 'name'].replace('кафе','get&fly')
data.loc[189, 'name'] = data.loc[215, 'name'].replace('кафе','Гриль у пруда & кофейня')
data = data.drop(index= [1430, 2211 ,3109])
data.reset_index(drop=True)
data[data.duplicated(subset = ['name', 'address'], keep = False)]
```

Out[11]:

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats
--	------	----------	---------	----------	-------	-----	-----	--------	-------	----------	-----------------	-------------------	-------	-------

```
In [12]: # Снова посчитаем количество заведений
data['name'].count()
```

Out[12]: 8403

Дубликаты обработаны, по итогу у нас остаётся 8403 заведения.

Работа с пропусками

```
In [13]: data.isna().sum()

Out[13]: name                0
category                0
address                 0
district                0
hours                  535
lat                    0
lng                    0
rating                 0
price                  5088
avg_bill               4587
middle_avg_bill        5254
middle_coffee_cup      7868
chain                  0
seats                  3611
dtype: int64
```

Пропусков обнаружено достаточно много. Заполнять пропуски будет проблематично и трудозатратно, так как даже два соседних заведения могут очень сильно различатся, заполнение медианными значениями будет некорректным, поэтому пропуски оставлю без изменений.

Создание столбца с названиями улиц

```
In [14]: # Создадим столбец street
data['street'] = data['address'].str.split(',').str[1]
data.head(10)
```

Out[14]:

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street
0	wowfli	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 Р	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 Р	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 Р	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 Р	500.0	NaN	1	148.0	правобережная улица
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	NaN	ижорская улица
6	огни города	бар,паб	москва, клязьминская улица, 9, стр. 3	Северный административный округ	пн 15:00–04:00; вт-вс 15:00–05:00	55.890752	37.524653	4.4	средние	Средний счёт:199 Р	199.0	NaN	0	45.0	клязьминская улица
7	mr. уголек	быстрое питание	москва, клязьминская улица, 9, стр. 3	Северный административный округ	пн-чт 10:00–22:00; пт,сб 10:00–23:00; вс 10:00...	55.890636	37.524303	4.7	средние	Средний счёт:200–300 Р	250.0	NaN	0	45.0	клязьминская улица
8	donna maria	ресторан	москва, дмитровское шоссе, 107, корп. 4	Северный административный округ	ежедневно, 10:00–22:00	55.880045	37.539006	4.8	средние	Средний счёт:от 500 Р	500.0	NaN	0	79.0	дмитровское шоссе
9	готика	кафе	москва, ангарская улица, 39	Северный административный округ	ежедневно, 12:00–00:00	55.879038	37.524487	4.3	средние	Средний счёт:1000–1200 Р	1100.0	NaN	0	65.0	ангарская улица

Создание столбца с обозначением, что заведение работает ежедневно и круглосуточно (24/7)

In [15]:

```
# Создадим столбец is_24/7 с обозначением, что заведение работает ежедневно и круглосуточно
data['is_24/7'] = data['hours'].str.contains('24/7|круглосуточно', case=False, na=False)
data[data['is_24/7'] == True]
```

Out[15]:

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street	is_24/7
10	great room bar	бар,паб	москва, левобережная улица, 12	Северный административный округ	ежедневно, круглосуточно	55.877832	37.469171	4.5	средние	Цена бокала пива:250–350 Р	NaN	NaN	0	102.0	левобережная улица	True
17	чайхана беш-бармак	ресторан	москва, ленинградское шоссе, 716, стр. 2	Северный административный округ	ежедневно, круглосуточно	55.876908	37.449876	4.4	средние	Средний счёт:350–500 Р	425.0	NaN	0	96.0	ленинградское шоссе	True
19	пекарня	булочная	москва, ижорский проезд, 5	Северный административный округ	ежедневно, круглосуточно	55.887969	37.515688	4.4	NaN	NaN	NaN	NaN	1	NaN	ижорский проезд	True
24	drive café	кафе	москва, улица дыбенко, 9ас1	Северный административный округ	ежедневно, круглосуточно	55.879992	37.481571	4.0	NaN	NaN	NaN	NaN	1	NaN	улица дыбенко	True
49	2и-ту-ю	пиццерия	москва, ижорская улица, 8а	Северный административный округ	ежедневно, круглосуточно	55.886160	37.508784	2.7	NaN	Средний счёт:900 Р	900.0	NaN	0	NaN	ижорская улица	True
...
8394	намангале	кафе	москва, ферганская улица, вл17-21	Юго-Восточный административный округ	ежедневно, круглосуточно	55.705332	37.819244	4.3	NaN	NaN	NaN	NaN	0	NaN	ферганская улица	True
8399	пекарня, кафе-гриль	булочная	москва, болотниковская улица, 52, корп. 2	Юго-Западный административный округ	ежедневно, круглосуточно	55.662866	37.582572	4.2	NaN	Средний счёт:50–250 Р	150.0	NaN	0	50.0	болотниковская улица	True
8403	самовар	кафе	москва, люблинская улица, 112а, стр. 1	Юго-Восточный административный округ	ежедневно, круглосуточно	55.648859	37.743219	3.9	NaN	Средний счёт:от 150 Р	150.0	NaN	0	150.0	люблинская улица	True
8404	чайхана sabr	кафе	москва, люблинская улица, 112а, стр. 1	Юго-Восточный административный округ	ежедневно, круглосуточно	55.648849	37.743222	4.2	NaN	NaN	NaN	NaN	1	150.0	люблинская улица	True
8405	kebab time	кафе	москва, россoshанский проезд, 6	Южный административный округ	ежедневно, круглосуточно	55.598229	37.604702	3.9	NaN	NaN	NaN	NaN	0	12.0	россошанский проезд	True

772 rows × 16 columns

Вывод: В ходе предобработки данных были обнаружены и обработаны дубликаты, по итогу осталось 8403 заведения в датасете, также было принято решение оставить пропуски по причине невозможности корректно заполнить их, созданы столбец `street` с названиями улиц и столбец `is_24/7` с обозначением, что заведение работает ежедневно и круглосуточно.

Шаг. Анализ данных

Количество объектов общественного питания по категориям

In [16]:

```
# Посчитаем количество заведений по категориям
cat_count = data.groupby('category').agg({'address': 'count'}).rename(columns={'address': 'count'}).sort_values('count', ascending=False).reset_index()
cat_count
```

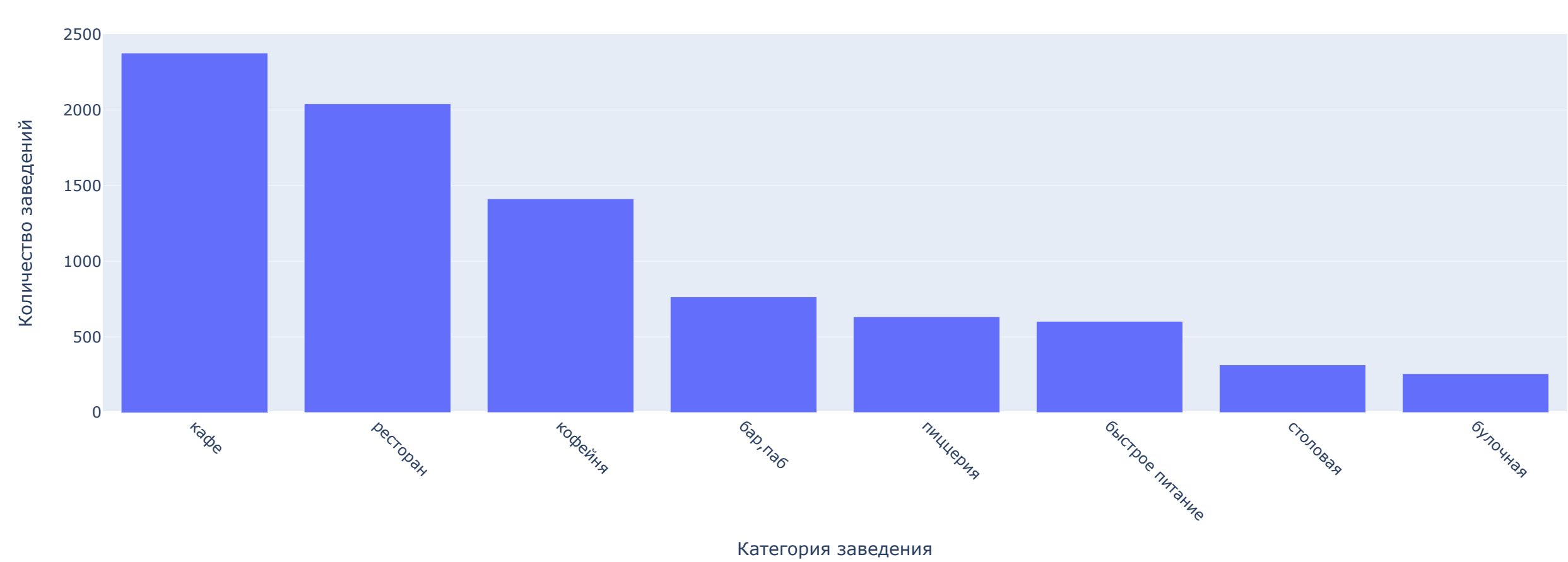
Out[16]:

	category	count
0	кафе	2377
1	ресторан	2041
2	кофейня	1413
3	бар,паб	765
4	пиццерия	633
5	быстрое питание	603
6	столовая	315
7	булочная	256

In [17]:

```
# Визуализируем количество заведений по категориям
fig = px.bar(
    cat_count,
    x='category',
    y='count',
    title='Количество заведений по категориям',
    labels={'category': 'Категория заведения', 'count': 'Количество заведений'})
fig.update_xaxes(tickangle=45)
fig.show()
```


Количество заведений по категориям



В столице больше всего кафе(2377), ресторанов(2041) и кофеен(1413), менее распространены бары/пабы(765), пиццерии(633) и заведения быстрого питания(603), самыми нераспространёнными заведениями являются столовые(315) и булочные(256).

Количество посадочных мест в заведениях по категориям

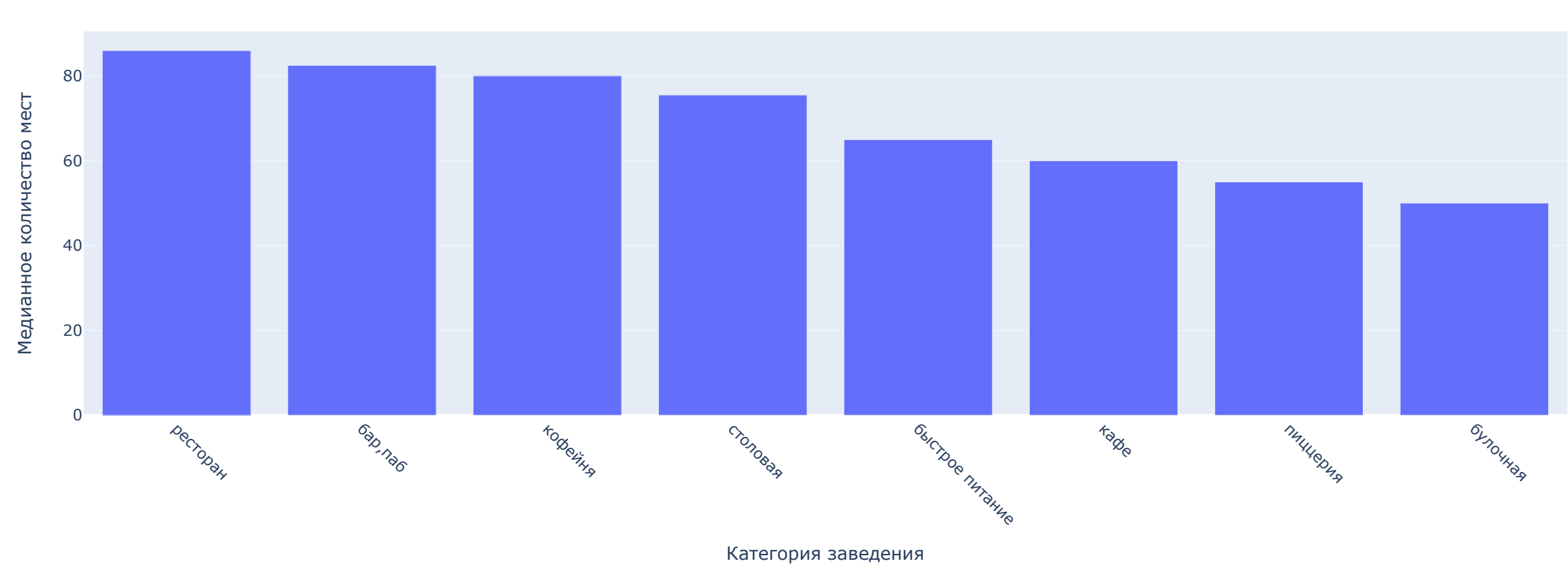
```
In [18]: # Посмотрим медианное количество мест по категориям заведений
median_seats = data.groupby('category').agg({'seats': 'median'}).sort_values('seats', ascending=False).reset_index()
median_seats
```

Out[18]:

	category	seats
0	ресторан	86.0
1	бар, паб	82.5
2	кофейня	80.0
3	столовая	75.5
4	быстрое питание	65.0
5	кафе	60.0
6	пиццерия	55.0
7	булочная	50.0

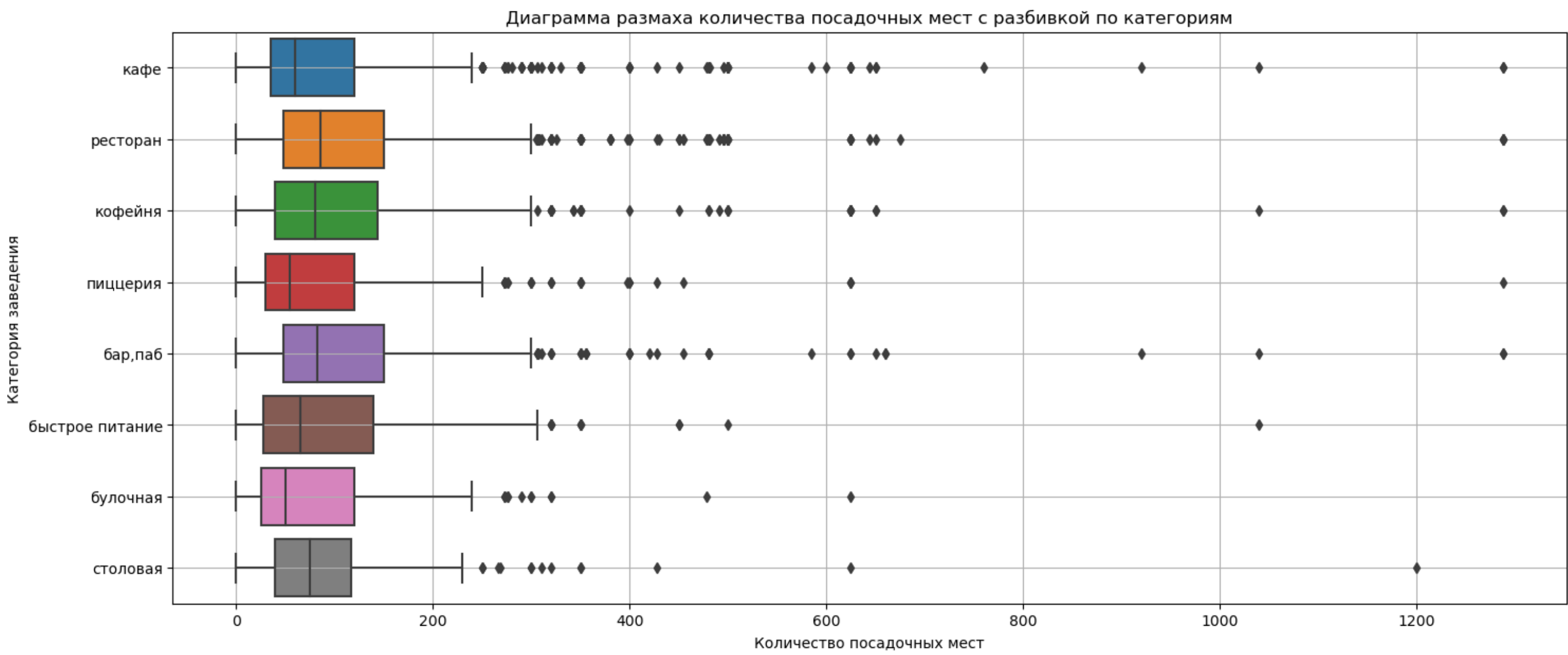
```
In [19]: # Визуализируем медианное количество мест по категориям заведений
fig = px.bar(
    median_seats,
    x='category',
    y='seats',
    title='Медианное количество посадочных мест в заведениях по категориям',
    labels={'category': 'Категория заведения', 'seats': 'Медианное количество мест'})
fig.update_xaxes(tickangle=45)
fig.show()
```

Медианное количество посадочных мест в заведениях по категориям



Больше всего посадочных мест в ресторанах(86), барах/пабах(82) и кофейнях(80), поменьше посадочных мест в столовых(75) и заведениях быстрого питания(65). Меньше всего посадочных мест в кафе(60), в пиццериях(55) и в булочных(50).

```
In [20]: # Также посмотрим на диаграмме размаха количество мест с разбивкой по категориям
plt.figure(figsize=(17, 7))
ax = sns.boxplot(x=data['seats'], y=data['category'])
plt.title('Диаграмма размаха количества посадочных мест с разбивкой по категориям')
plt.xlabel('Количество посадочных мест')
plt.ylabel('Категория заведения')
plt.grid()
plt.show()
```



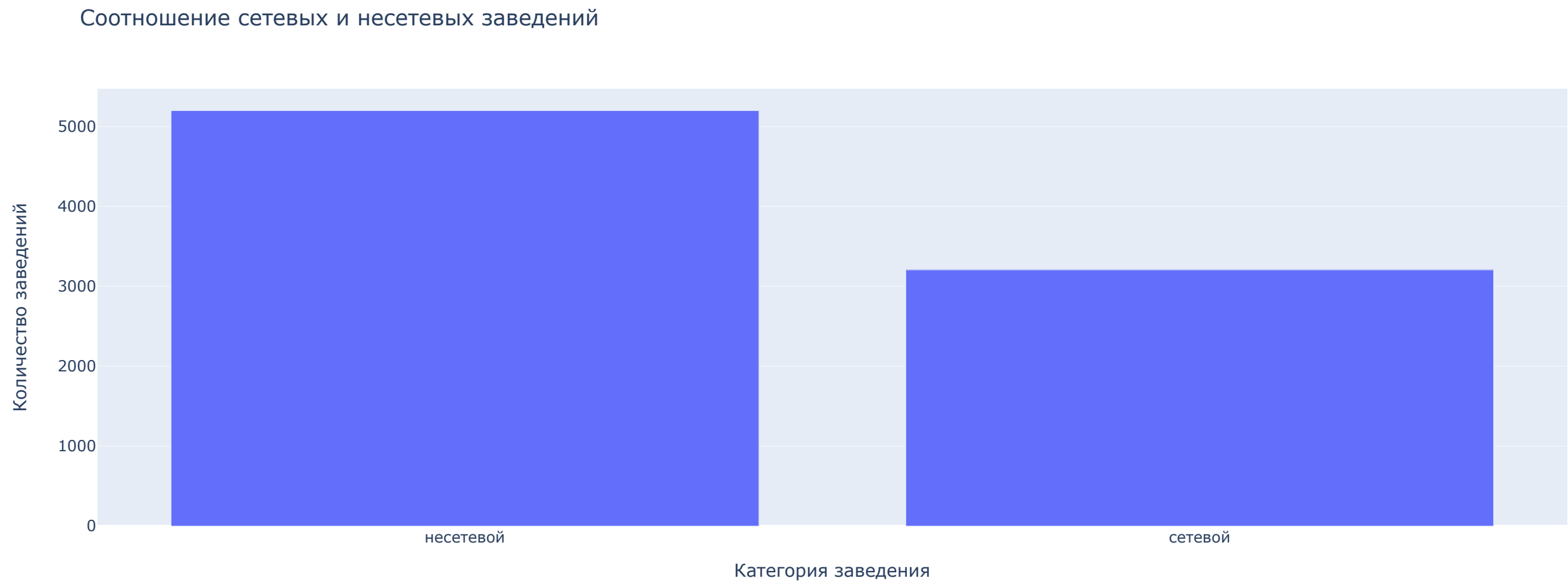
По графикам видно, что по категории заведения варьируется медианное количество посадочных мест. Верхняя граница у всех типов заведений в районе 200-300 мест, остальное выбросы очень больших заведений, медианное же количество посадочных мест варьируется в районе от 50 до 80 мест.

Соотношение сетевых и несетевых заведений в датасете

```
In [21]: # Переименуем "0" и "1" на "несетевой" и "сетевой"
data['chain'] = data['chain'].replace(0, 'несетевой')
data['chain'] = data['chain'].replace(1, 'сетевой')
# Посчитаем количество сетевых и несетевых заведений
chain = data['chain'].value_counts()
chain

Out[21]: несетевой    5198
сетевой      3205
Name: chain, dtype: int64

In [22]: # Визуализируем соотношение сетевых и несетевых заведений
fig = px.bar(
    data, x=chain.index,
    y=chain,
    title='Соотношение сетевых и несетевых заведений',
    labels={'x': 'Категория заведения', 'y': 'Количество заведений'})
fig.show()
```



По полученному графику видно, что в москве преобладают несетевые заведения, 5198 против 3205.

Какие категории заведений чаще являются сетевыми

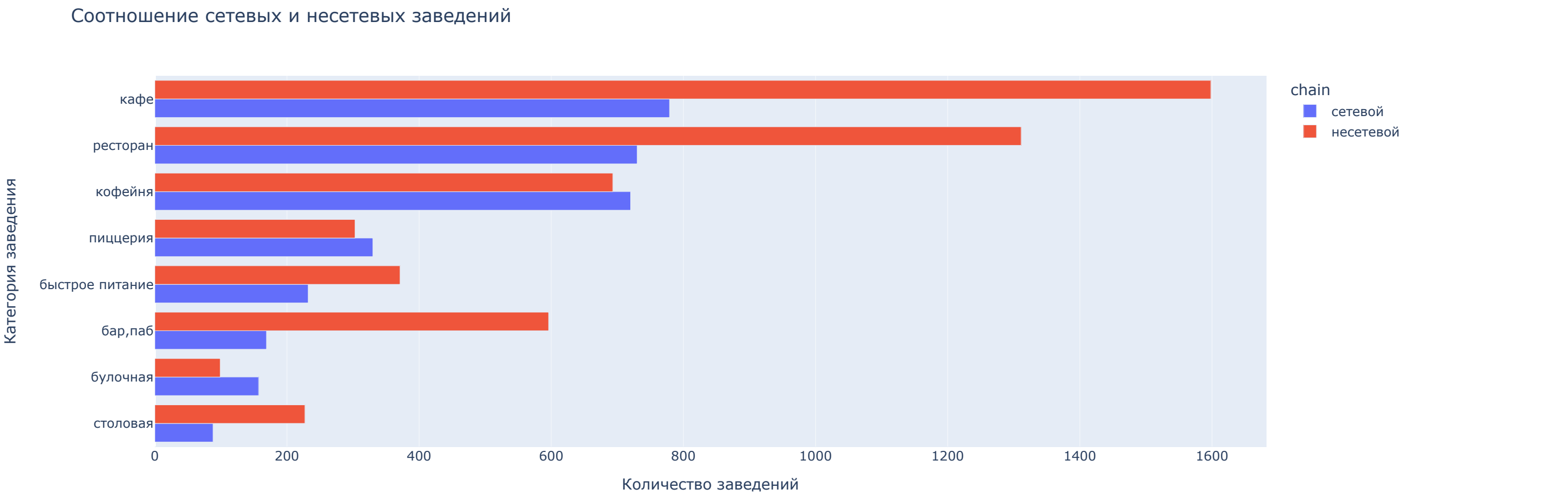
```
In [23]: # Посчитаем количество сетевых и несетевых заведений по категориям
chain_cat = data.pivot_table(
    index=['category', 'chain'],
    values='address',
    aggfunc='count').reset_index().rename(columns={'address': 'count'}).sort_values(by='count')
chain_cat
```

Out[23]:

	category	chain	count
15	столовая	сетевой	88
2	булочная	несетевой	99
3	булочная	сетевой	157
1	бар,паб	сетевой	169
14	столовая	несетевой	227
5	быстрое питание	сетевой	232
10	пиццерия	несетевой	303
11	пиццерия	сетевой	330
4	быстрое питание	несетевой	371
0	бар,паб	несетевой	596
8	кофейня	несетевой	693
9	кофейня	сетевой	720
13	ресторан	сетевой	730
7	кафе	сетевой	779
12	ресторан	несетевой	1311
6	кафе	несетевой	1598

In [24]:

```
# Визуализируем количество сетевых и несетевых заведений по категориям
fig = px.bar(
    chain_cat,
    x='count',
    y='category',
    color='chain',
    title='Соотношение сетевых и несетевых заведений',
    labels={'category': 'Категория заведений', 'count': 'Количество заведений'},
    barmode='group')
fig.show()
```



По графику видно, что в категориях кофейни(720 с:693 нс), пиццерии(330 с:303 нс) и булочные(157 с:99 нс) количество сетевых заведений преобладает над несетевыми, в остальных случаях ситуация обратная. У баров(169 с:596 нс) доля несетевых заведений больше, чем в других категориях.

Топ-15 популярных сетей в Москве

In [25]:

```
# Посчитаем количество самых популярных сетевых заведений
top_chain = data[data['chain']=='сетевой'].pivot_table(index=['name', 'category'],
    values=['rating', 'address'],
    aggfunc={'rating': 'median', 'address': 'count'}).reset_index().rename(columns={'address': 'count'}).sort_values(by='count', ascending=False).head(15)

top_chain
```

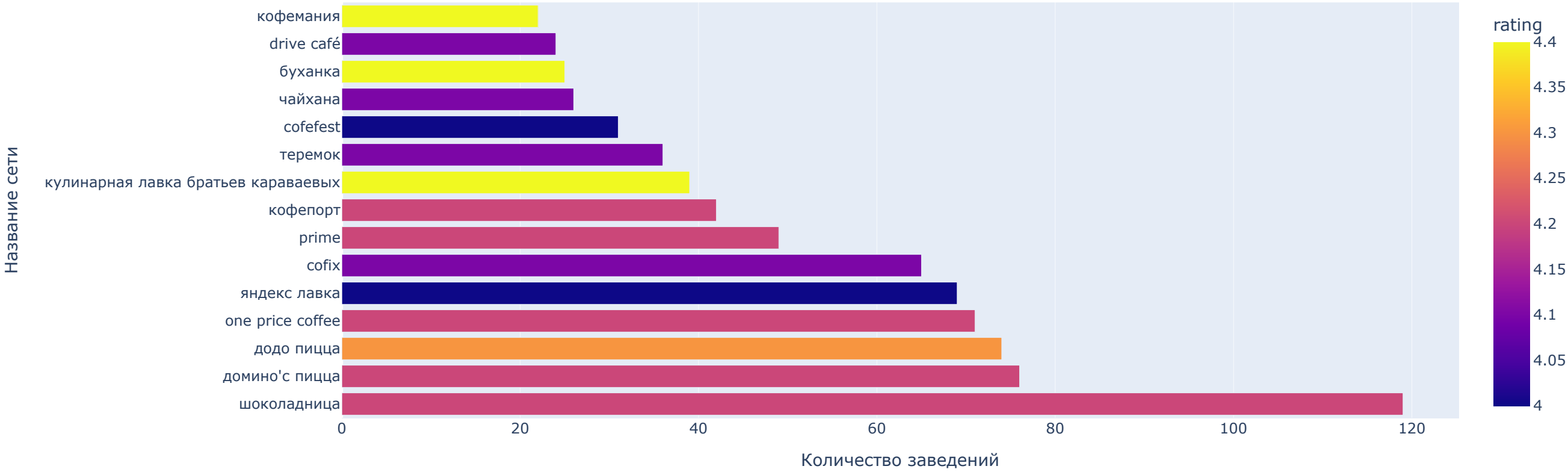
Out[25]:

	name	category	count	rating
1125	шоколадница	кофейня	119	4.2
496	домино'с пицца	пиццерия	76	4.2
489	додо пицца	пиццерия	74	4.3
205	one price coffee	кофейня	71	4.2
1141	яндекс лавка	ресторан	69	4.0
72	sofix	кофейня	65	4.1
241	prime	ресторан	49	4.2
617	кофепорт	кофейня	42	4.2
632	кулинарная лавка братьев караваевых	кафе	39	4.4
962	теремок	ресторан	36	4.1
50	cofefest	кофейня	31	4.0
1049	чайхана	кафе	26	4.1
385	буханка	булочная	25	4.4
89	drive café	кафе	24	4.1
615	кофемания	кофейня	22	4.4

In [26]:

```
# Визуализируем количество самых популярных сетевых заведений с рейтингом
fig = px.bar(
    top_chain,
    x='count',
    y='name',
    color='rating',
    title='Топ-15 популярных сетей в Москве',
    labels={'name': 'Название сети', 'count': 'Количество заведений'})
fig.show()
```

Топ-15 популярных сетей в Москве



Самой распространённой сетью в Москве является кофейня Шоколадница(119), второе и третье место делят пиццерии Домино'с пицца(76) и Додо пицца(74), также можно отметить, что среди 15 самых популярных сетей преобладают кофейни(шоколадница, one price coffee, cofix, кофепорт, cofefest, кофемания).

Какие административные районы Москвы присутствуют в датасете?

```
In [27]: # Посчитаем количество заведений по районам
districts = data.groupby('district').agg({'address': 'count'}).rename(columns={'address': 'total'}).sort_values('total', ascending=False).reset_index()
districts
```

Out[27]:

	district	total
0	Центральный административный округ	2242
1	Северный административный округ	899
2	Южный административный округ	892
3	Северо-Восточный административный округ	890
4	Западный административный округ	850
5	Восточный административный округ	798
6	Юго-Восточный административный округ	714
7	Юго-Западный административный округ	709
8	Северо-Западный административный округ	409

```
In [28]: # Также посчитаем количество заведений представленных категорий по районам
districts_cat = data.groupby(['district', 'category']).agg({'address': 'count'}).rename(columns={'address': 'count'}).sort_values('count', ascending=False).reset_index()
districts_cat = districts_cat.merge(districts, on='district', how='left').sort_values('total', ascending=False).reset_index()
districts_cat
```

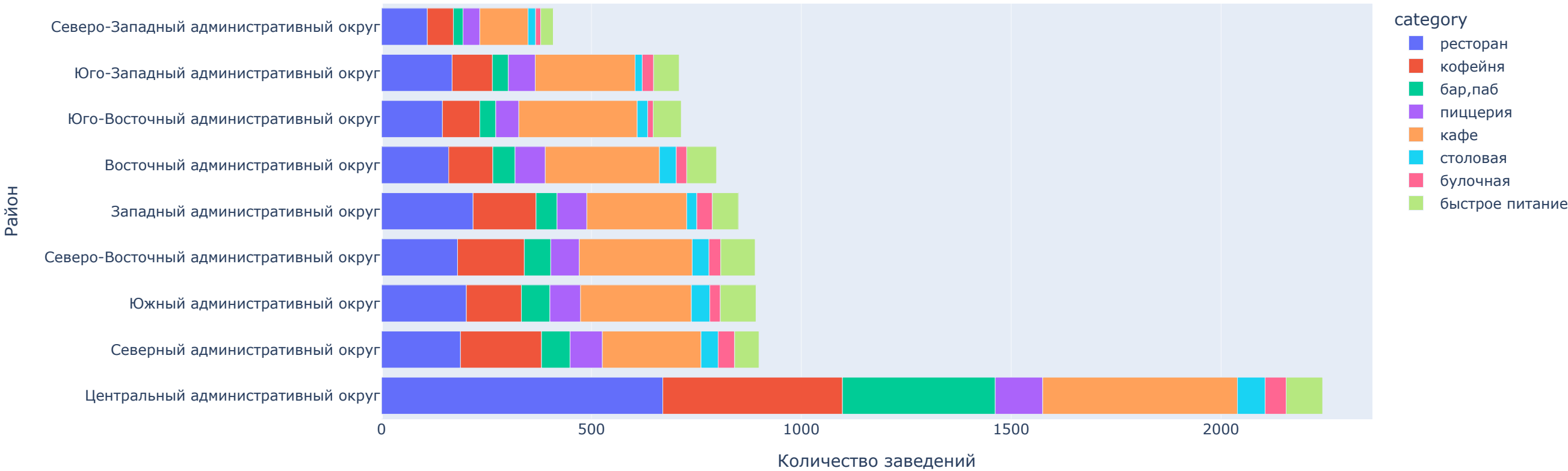
Out[28]:

	index	district	category	count	total
0	0	Центральный административный округ	ресторан	670	2242
1	2	Центральный административный округ	кофейня	428	2242
2	3	Центральный административный округ	бар,паб	364	2242
3	23	Центральный административный округ	пиццерия	113	2242
4	1	Центральный административный округ	кафе	464	2242
...
67	22	Северо-Западный административный округ	кафе	115	409
68	67	Северо-Западный административный округ	бар,паб	23	409
69	68	Северо-Западный административный округ	столовая	18	409
70	24	Северо-Западный административный округ	ресторан	109	409
71	71	Северо-Западный административный округ	булочная	12	409

72 rows × 5 columns

```
In [29]: # Визуализируем количество заведений представленных категорий по районам
fig = px.bar(
    districts_cat,
    x='count',
    y='district',
    color='category',
    title='Распределение заведений по районам Москвы',
    labels={'district': 'Район', 'count': 'Количество заведений'})
fig.show()
```

Распределение заведений по районам Москвы



Ожидаемо, что в центре сосредоточено больше всего заведений всех категорий, где предположительно наибольшая проходимость, меньше всего заведений на северо-западе. В большинстве районов преобладают кафе и рестораны, также можно отметить значительное количество кофеен и заведений быстрого питания во всех районах, бары наибольшей популярностью пользуются в центре, столовых и булочных мало во всех районах.

Средние рейтинги заведений по категориям

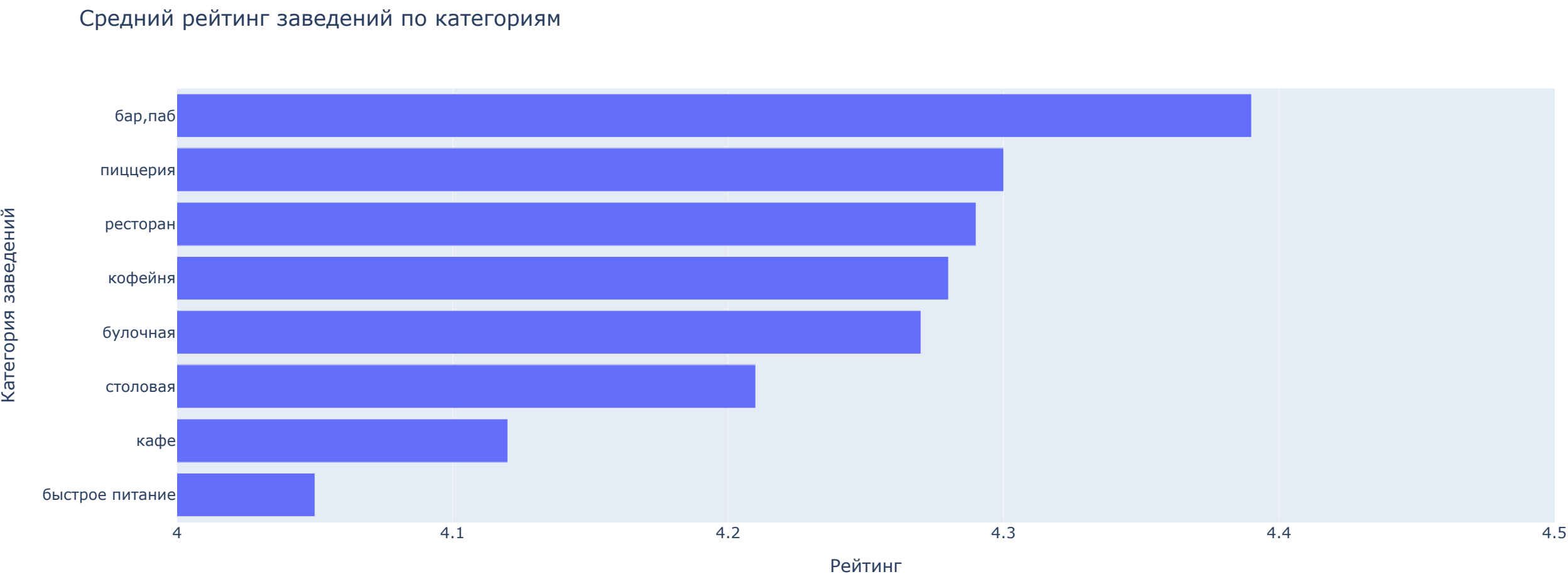
```
In [30]: # Посчитаем средний рейтинг заведений по категориям
rating = data[data['rating']>0].pivot_table(index='category',
                                             values='rating',
                                             aggfunc='mean').round(2).sort_values('rating').reset_index()

rating
```

Out[30]:

	category	rating
0	быстрое питание	4.05
1	кафе	4.12
2	столовая	4.21
3	булочная	4.27
4	кофейня	4.28
5	ресторан	4.29
6	пиццерия	4.30
7	бар,паб	4.39

```
In [31]: # Визуализируем средний рейтинг заведений по категориям
fig = px.bar(
    rating,
    x='rating',
    y='category',
    title='Средний рейтинг заведений по категориям',
    labels={'rating': 'Рейтинг', 'category': 'Категория заведений'})
fig.update_xaxes(range=[4, 4.5])
fig.show()
```



Можно сказать, что откровенно плохих категорий в Москве нет, все средние рейтинги от 4,05 до 4,39. Самый высокий средний рейтинг у баров(4,39), а самый низкий у быстрого питания(4,05), это можно объяснить наличием в этой категории шаурмечных без стандартов. Можно предположить, что такие высокие средние рейтинги обусловлены тем, что в Москве высокая конкуренция, которая вынуждает заведения придерживаться высоких стандартов, чтобы бороться за внимание клиентов.

Фоновая картограмма (хороплет) со средним рейтингом заведений каждого района

```
In [32]: # Посчитаем средний рейтинг заведений по районам
rating_by_district = data[data['rating']>0].pivot_table(index='district',
                                                         values='rating',
                                                         aggfunc='mean').round(2).sort_values('rating').reset_index()

rating_by_district
```

Out[32]:

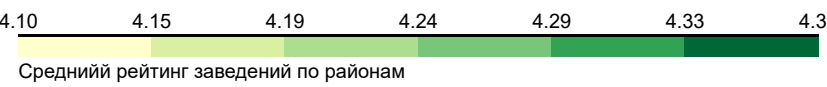
	district	rating
0	Юго-Восточный административный округ	4.10
1	Северо-Восточный административный округ	4.15
2	Восточный административный округ	4.17
3	Юго-Западный административный округ	4.17
4	Западный административный округ	4.18
5	Южный административный округ	4.18
6	Северо-Западный административный округ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


```
In [33]: # загружаем JSON-файл с границами округов Москвы
state_geo = 'C:/Users/dimch/OneDrive/Рабочий стол/Яндекс Практикум/Datasets/admin_level_geomap.geojson'
# 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, tiles='Cartodb Positron')

# создаём хороплет с помощью конструктора Choropleth и добавляем его на карту
Choropleth(
    geo_data=state_geo,
    data=rating_by_district,
    columns=['district', 'rating'],
    key_on='feature.name',
    fill_color='YlGn',
    fill_opacity=0.8,
    legend_name='Средний рейтинг заведений по районам',
).add_to(m)

# выводим карту со средним рейтингом заведений по районам
m
```



 Leaflet (<https://leafletjs.com>) | © OpenStreetMap (<https://www.openstreetmap.org/copyright>) contributors © CARTO (<https://carto.com/attributions>)

Самый высокий рейтинг в заведениях в Центральном административном округе(4,38), а самый низкий в Юго-Восточном административном округе(4,1).

Все заведения датасета на карте

```
In [34]: # создаём карту Москвы
moscow_all = Map(location=[moscow_lat, moscow_lng], zoom_start=10, tiles='Cartodb Positron')
# создаём пустой кластер, добавляем его на карту
marker_cluster = MarkerCluster().add_to(moscow_all)

def create_clusters(row):
    # сохраняем URL-адрес изображения со значком торгового центра с icons8,
    # это путь к файлу на сервере icons8
    icon_url = 'https://img.icons8.com/neon/96/restaurant-table.png'
    # создаём объект с собственной иконкой размером 30x30
    icon = CustomIcon(icon_url, icon_size=(30, 30))

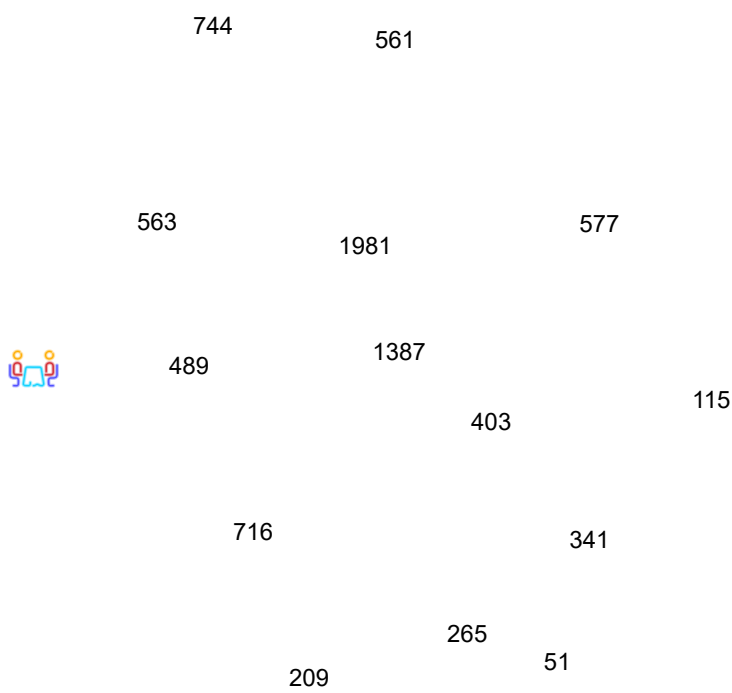
    # создаём маркер с иконкой icon и добавляем его в кластер
    Marker(
        [row['lat'], row['lng']],
        popup=f"{row['name']} {row['rating']}",
        icon=icon,
    ).add_to(marker_cluster)

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

# выводим карту
moscow_all
```

+

−



Leaflet (https://leafletjs.com) | © OpenStreetMap (https://www.openstreetmap.org/copyright) contributors © CARTO (https://carto.com/attributions)

Топ-15 улиц по количеству заведений

```
In [35]: # Посчитаем количество заведений на улицах с наибольшим их количеством
top_st = data.groupby('street').agg({'address': 'count'}).rename(columns={'address': 'total'}).sort_values('total', ascending=False).head(15).reset_index()
top_st
```

Out[35]:

	street	total
0	проспект мира	183
1	профсоюзная улица	122
2	проспект вернадского	108
3	ленинский проспект	107
4	ленинградский проспект	95
5	дмитровское шоссе	88
6	каширское шоссе	77
7	варшавское шоссе	76
8	ленинградское шоссе	70
9	мкад	65
10	люблинская улица	60
11	улица вавилова	55
12	кутузовский проспект	54
13	улица миклухо-маклая	49
14	пятницкая улица	48

```
In [36]: # Посчитаем также с разбивкой по категориям
top_st_cat = data.groupby(['street', 'category']).agg({'address': 'count'}).rename(columns={'address': 'count'}).sort_values('count', ascending=False).reset_index()
top_st_cat = top_st_cat.merge(top_st, on='street', how='left').sort_values('total', ascending=False).dropna().reset_index()
top_st_cat
```

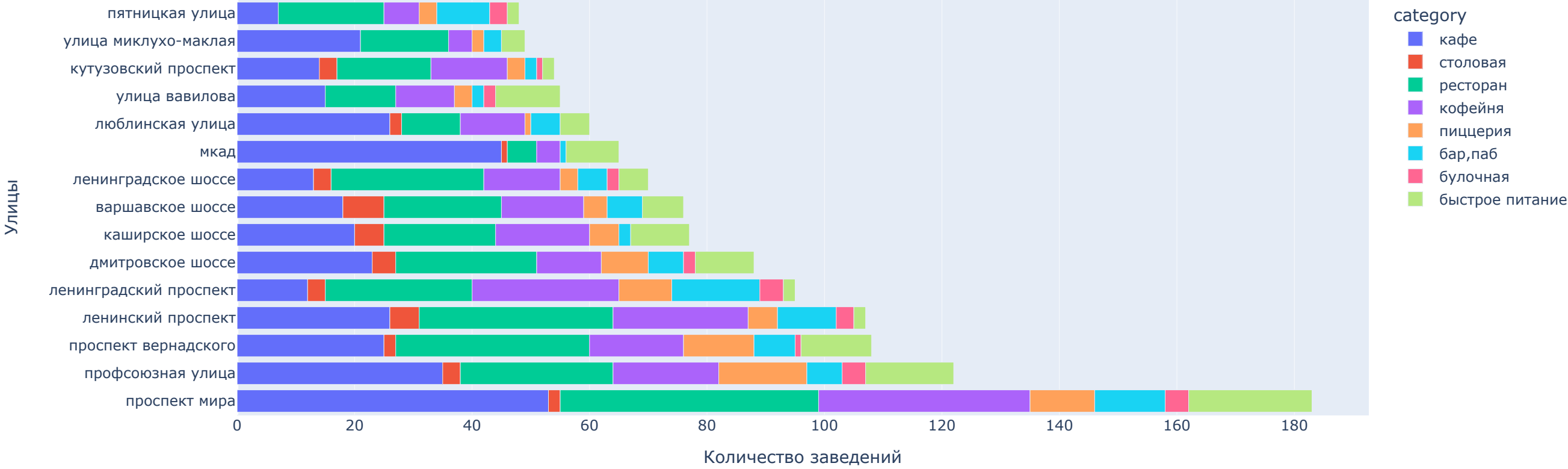
Out[36]:

	index	street	category	count	total
0	0	проспект мира	кафе	53	183.0
1	1202	проспект мира	столовая	2	183.0
2	2	проспект мира	ресторан	44	183.0
3	3	проспект мира	кофейня	36	183.0
4	65	проспект мира	пиццерия	11	183.0
...
106	109	пятницкая улица	бар,паб	9	48.0
107	221	пятницкая улица	кофейня	6	48.0
108	28	пятницкая улица	ресторан	18	48.0
109	148	пятницкая улица	кафе	7	48.0
110	750	пятницкая улица	булочная	3	48.0

111 rows × 5 columns

```
In [37]: fig = px.bar(
    top_st_cat,
    x='count',
    y='street',
    color='category',
    title='Топ-15 улиц Москвы по количеству заведений',
    labels={'street': 'Улицы', 'count': 'Количество заведений'})
fig.show()
```

Топ-15 улиц Москвы по количеству заведений



Проспект Мира лидирует по количеству заведений, он сильно выделяется на фоне других улиц. На этой улице преобладают кафе, рестораны и кофейни. Чаше всего на представленных улицах встречаются кафе и рестораны.

Улицы, на которых находится только один объект общепита

```
In [38]: # Отберём улицы с одним заведением
one_cafe_st = data.groupby('street').agg({'address': 'count'}).rename(columns={'address': 'total'}).sort_values('total').reset_index()
one_cafe_st = one_cafe_st.query('total == 1')
one_cafe_st = one_cafe_st['street'].tolist()
one_cafe_st
```



```
Out[38]: ['1-й автозаводский проезд',
'новосущевская улица',
'новошуйкинская улица',
'новоясенеvский тупик',
'одинцовская улица',
'октябрьский переулок',
'ордынский тупик',
'оренбургская улица',
'новороссийская улица',
'орловский переулок',
'островная улица',
'отрадный проезд',
'павелецкая набережная',
'панкратьевский переулок',
'парк алтуфьево',
'парк борисовские пруды',
'парк братеевская набережная',
'оршанская улица',
'новопоселковая улица',
'новомосковская улица',
'новгородская улица',
'малый кисловский переулок',
'малый николапесковский переулок',
'малый патриарший переулок',
'малый толмачевский переулок',
'мансуровский переулок',
'медовый переулок',
'мерзляковский переулок',
'миусская площадь',
'михайловский проезд',
'молдавская улица',
'москворецкая набережная',
'московская улица',
'московский проспект',
'мурановская улица',
'мясницкий проезд',
'нащокинский переулок',
'новая улица',
'парк дружбы',
'парк зюзино',
'парк имени артема боровика',
'парк искусств музеон',
'померанцев переулок',
'поперечный просек',
'посланников переулок',
'пр-т комсомольский',
'прибрежный проезд',
'природно-исторический парк измайлово',
'проезд донелайтиса',
'проезд одоеvского',
'проезд ольминского',
'проезд якушкина',
'проектируемый проезд № 5265',
'промышленный проезд',
'пронская улица',
'проспект академика сахарова',
'проспект лихачева',
'просторная улица',
'проточный переулок',
'полоцкая улица',
'малый казенный переулок',
'пожарский переулок',
'площадь журавлева',
'парк красная пресня',
'парк левобережный',
'парк сад будущего',
'парк технических видов спорта',
'парк тюфелева роща',
'парк этнографическая деревня бибирево',
'парусный проезд',
'певческий переулок',
'переведеноvский переулок',
'перекопская улица',
'переулок капранова',
'пермская улица',
'перовский парк культуры и отдыха',
'пестовский переулок',
'петроверигский переулок',
'петровско-разумовская аллея',
'платовская улица',
'площадь савеловского вокзала',
'пулковская улица',
'малый ивановский переулок',
'малый дровяной переулок',
'дорогобужская улица',
'дурасовский переулок',
'ереванская улица',
'живарев переулок',
'жуков проезд',
'загорьевский проезд',
'задонский проезд',
'дмитровский переулок',
'замоскворецкая линия',
'зарайская улица',
'звенигородская улица',
'золотая улица',
'зоологическая улица',
'зубарев переулок',
'ивановская улица',
'иваньковское шоссе',
'заозерная улица',
'дербеневская улица',
'денежный переулок',
'дендропарковая улица',
'выползов переулок',
'гаврикова улица',
'гагаринский переулок',
'гагаринский тоннель',
'газгольдерная улица',
'гжельский переулок',
'главный ботанический сад имени н.в. циицина российской академии наук',
'глубокий переулок',
'городецкая улица',
'городская улица',
'гороховский переулок',
'графский переулок',
'грузинский сквер',
'гурьевский проезд',
'даев переулок',
'девятикин переулок',
'делегатская улица',
'игарский проезд',
'измайловская площадь',
'инициативная улица',
'казарменный переулок',
'курсовой переулок',
'кутузовский проезд',
'лазаревский переулок',
'ландшафтный заказник лианозовский',
'ленинский проспект (дублер)',
'леонтьевский переулок',
'лечебная улица',
'лианозовский парк культуры и отдыха',
'лианозовский проезд',
'липецкая улица (дублер)',
'мажоров переулок',
'малая андроньевская улица',
'малая красносельская улица',
```

' малая почтовая улица',
' малая тульская улица',
' маленковская улица',
' малый гнездиковский переулок',
' кузьминская улица',
' малый златоустинский переулок',
' кубанская улица',
' красноярская улица',
' каланчевская улица',
' каргопольская улица',
' клинская улица',
' коломенская набережная',
' кольмажный переулок',
' кольцевая линия',
' коммунистический переулок',
' композиторская улица',
' конюшковская улица',
' кооперативная улица',
' коптевский бульвар д 18 а стр 1',
' костомаровский переулок',
' костромская улица',
' котляковская улица',
' красноармейская улица',
' красностуденческий проезд',
' краснохолмская набережная',
' крылатский мост',
' ворошиловский парк',
' путевой проезд',
' раушская набережная',
' улица маршала соколовского',
' улица мжд киевское 5-й км',
' улица мусоргского',
' улица неверовского',
' улица николая химушина',
' улица олеко дундича',
' улица острякова',
' улица маршала баграмяна',
' улица паустовского',
' улица пивченкова',
' улица поликарпова',
' улица полины осипенко',
' улица раевского',
' улица расковой',
' улица расплетина',
' улица ремизова',
' улица петра романова',
' улица малая якиманка',
' улица малая полянка',
' улица максимова',
' улица ильинка',
' улица капотня',
' улица кашенкин луг',
' улица кирпичные выемки',
' улица климашкина',
' улица клочкова',
' улица комдива орлова',
' улица кондратюка',
' улица коненкова',
' улица константина федина',
' улица костякова',
' улица кутузова',
' улица ленивка',
' улица лефортовский вал',
' улица литвина-седого',
' улица лобачика',
' улица луиджи лонго',
' улица речников',
' улица рогова',
' улица розанова',
' улица рокотова',
' федеративный проспект',
' хитровский переулок',
' хлыновский тупик',
' хоромный тупик',
' хохловский переулок',
' центральный парк культуры и отдыха имени м. горького',
' чапаевский переулок',
' челобитьевское шоссе',
' челюскинская улица',
' черемушкинский проезд',
' чермянский проезд',
' чонгарский бульвар',
' шоссейный проезд',
' электрический переулок',
' юго-западный административный округ',
' якиманский переулок',
' яковоапостольский переулок',
' уссурийская улица',
' улица ивана франко',
' улица юннатов',
' улица шулева 4',
' улица россолимо',
' улица садовники',
' улица саяма адия',
' улица самеда вургунa',
' улица саморы машела',
' улица седова',
' улица советской армии',
' улица станиславского',
' улица старые кузьминки',
' улица стасовой',
' улица усиевича',
' улица уткина',
' улица фадеева',
' улица фотиевой',
' улица чечулина',
' улица чистова',
' улица шулева',
' улица шухова',
' пушкарев переулок',
' улица зорге',
' улица демьяна бедного',
' старый петровско-разумовский проезд',
' старый толмачевский переулок',
' стахановская улица',
' сторожевая улица',
' стрелецкая улица',
' стромынский переулок',
' ступинский проезд',
' старомонетный переулок',
' сумская',
' сытинский переулок',
' сытинский тупик',
' таганский парк культуры и отдыха',
' таежная улица',
' талдомская улица',
' таможенный проезд',
' тарутинская улица',
' сумской проезд',
' староконюшенный переулок',
' старокаширское шоссе',
' ставропольский проезд',
' савеловская линия',
' сад эрмитаж',
' самарская улица',
' самокатная улица',
' светлогорский проезд',
' северный бульвар',

' сеченовский переулок',
' симферопольский проезд',
' скатертный переулок',
' сквер имени м.и. калинина',
' скотопрогонная улица',
' смоленская-сенная площадь',
' советская улица',
' солянский тупик',
' софийская набережная',
' средний тишинский переулок',
' ставропольская улица',
' терлецкий лесопарк',
' товарищеский переулок',
' токмаков переулок',
' третье транспортное кольцо',
' улица ватутина',
' улица вешних вод',
' улица вильгельма пика',
' улица водников',
' улица всеволода вишневого',
' улица вучетича',
' улица высоцкого',
' улица габричевского',
' улица гастелло',
' улица гашека',
' улица генерала глаголева',
' улица генерала дорохова',
' улица генерала ермолова',
' улица говорова',
' улица годовикова',
' улица грекова',
' улица губкина',
' улица васильцовский стан',
' улица дудинка',
' улица бусиновская горка',
' улица большая молчановка',
' трехпрудный переулок',
' триумфальная площадь',
' троилинский переулок',
' турчанинов переулок',
' тучковская улица',
' тюменский проезд',
' угличская улица',
' ул. профсоюзная',
' ул. ярославская',
' улица 8 марта',
' улица 800-летия москвы',
' улица айвазовского',
' улица академика ильюшина',
' улица академика комарова',
' улица антонова-овсеенко',
' улица артюхиной',
' улица богородский вал',
' улица буженинова',
' воробьевская набережная',
' № 7',
' будайский проезд',
' 2-й лучевой просек',
' 2-й новоподмосковный переулок',
' 2-й обыденский переулок',
' алтайская улица',
' 2-й переулок петра алексеева',
' аллея молодоженов',
' большой волоколамский проезд',
' 2-й полевой переулок',
' братиславский парк',
' 2-й рошинский проезд',
' 2-й силикатный проезд',
' 2-й тверской-ямской переулок',
' 2-й хорошевский проезд',
' 2-й щемилловский переулок',
' 2-я владимирская улица',
' брянская улица',
' андреевский пешеходный мост',
' 2-й крестовский переулок',
' 2-й красносельский переулок',
' 17-й проезд марьиной рощи',
' 2-й балтийский переулок',
' 2-й боткинский проезд',
' архангельский переулок',
' 2-й верхний михайловский проезд',
' 2-й вышеславцев переулок',
' арбатский переулок',
' 2-я звенигородская улица',
' булатниковский проезд',
' анненский проезд',
' 4-й рошинский проезд',
' 2-й зачатьевский переулок',
' 2-й квартал капотни',
' 2-й квартал капотня',
' 2-й кожуховский проезд',
' андроньевская площадь',
' большая тихоновская улица',
' 16-я парковая улица',
' 2-я карачаровская улица',
' 2-я пугачевская улица',
' 3-й михалковский переулок',
' 3-й новомихалковский проезд',
' 3-й проезд перова поля',
' 3-й проезд подбельского',
' большой николипесковский переулок',
' 5-й проезд подбельского',
' большой сухаревский переулок',
' большой строченовский переулок',
' 3-й хорошевский проезд',
' 3-я радиаторская улица',
' 3-я тверская-ямская улица',
' 3-я фрунзенская улица',
' 4-я улица 8 марта',
' 4-й воробьевский проезд',
' 4-й лесной переулок',
' большой трехгорный переулок',
' большой козловский переулок',
' 3-й лихачевский переулок',
' 5-я магистральная улица',
' братеевский парк',
' 2-я рошинская улица',
' 2-я рыбинская улица',
' боярский переулок',
' большой кисловский переулок',
' ботанический сад московского государственного университета',
' ботанический переулок',
' 2-я магистральная улица',
' 9-я улица соколиной горы',
' 2-я улица марьиной рощи',
' 6-я парковая улица',
' 6-й лучевой просек',
' борисовская улица',
' 2-я фрезерная улица',
' 5-я улица ямского поля',
' 3-й кадашевский переулок',
' боровское шоссе',
' бутиковский переулок',
' 4-я кабельная улица',
' 1-й колобовский переулок',
' 1-й земельный переулок',
' 1-й капотнинский проезд',
' 1-я парковая улица',

```
' 1-й кирпичный переулок',
' барвихинская улица',
' басманный тупик',
' 1-я линия хорошевского серебряного бора',
' 1-я курьяновская улица',
' веселая улица',
' 1-й котляковский переулок',
' 1-й самотечный переулок',
' 1-я боевская улица',
' вагоноремонтная улица',
' 1-й рижский переулок',
' 1-й курьяновский проезд',
' бобров переулок',
' 1-й спасоналивковский переулок',
' беговая аллея',
' богословский переулок',
' 1-й сетуньский проезд',
' верхнелихоборская улица',
' верхний сусальный переулок',
' 1-й щипковский переулок',
' 1-я стекольная улица',
' 1-й николащеповский переулок',
' 1-й грайвороновский проезд',
' валдайский проезд',
' большая калитниковская улица',
' верхоянская улица',
' 1-й балтийский переулок',
' 14-я парковая улица',
' 1-й варшавский проезд',
' 12-я парковая улица',
' бесединское шоссе',
' 1-й дербенеvский переулок',
' 11-я парковая улица',
' 1-я ямская улица',
' 1-й вешняковский проезд',
' 1-я фрунзенская улица',
' верхняя улица',
' васьильевская улица',
' 1-й голутвинский переулок',
' 1-й новокузнецкий переулок',
' балакиревский переулок',
' бибиревская улица',
' большая пионерская улица']
```

```
In [39]: # Создадим датасет только с единственными на улице заведениями
data_ocs = data.query('street in @one_cafe_st').reset_index().drop('index', axis=1)
data_ocs
```

Out[39]:

	name	category	address	district	hours	lat	lng	rating	price	avg_bill	middle_avg_bill	middle_coffee_cup	chain	seats	street	is_24/7
0	дом обеда	столовая	москва, улица бусиновская горка, 2	Северный административный округ	пн-пт 08:30–18:30; сб 10:00–20:00	55.885890	37.493264	4.1	средние	Средний счёт:300–500 р	400.0	NaN	несетевой	180.0	улица бусиновская горка	False
1	7/12	кафе	москва, прибрежный проезд, 7	Северный административный округ	ежедневно, 10:00–22:00	55.876805	37.464934	4.5	NaN	NaN	NaN	NaN	несетевой	NaN	прибрежный проезд	False
2	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	сетевой	NaN	парк левобережный	False
3	coffeekaldi's	кофейня	москва, угличская улица, 13, стр. 8	Северо-Восточный административный округ	ежедневно, 09:00–22:00	55.900316	37.570558	4.1	средние	Средний счёт:500–800 р	650.0	NaN	сетевой	NaN	угличская улица	False
4	чебуречная история	кофейня	москва, ландшафтный заказник лианозовский	Северо-Восточный административный округ	ежедневно, 10:00–22:00	55.899845	37.570488	4.9	NaN	NaN	NaN	NaN	сетевой	NaN	ландшафтный заказник лианозовский	False
...
452	мираж	ресторан	москва, улица шкулева, 2а	Юго-Восточный административный округ	пн-пт 11:00–23:00; сб,вс 11:00–00:00	55.693340	37.746231	4.7	NaN	NaN	NaN	NaN	несетевой	90.0	улица шкулева	False
453	wtfcoffee	кофейня	москва, улица артюхиной, 14/8с1	Юго-Восточный административный округ	ежедневно, 10:00–22:00	55.699044	37.737790	4.4	средние	NaN	NaN	NaN	несетевой	NaN	улица артюхиной	False
454	беседка	кафе	москва, кузьминская улица, 10	Юго-Восточный административный округ	ежедневно, 11:00–23:00	55.690691	37.787735	4.1	NaN	NaN	NaN	NaN	сетевой	NaN	кузьминская улица	False
455	лагман хаус	кафе	москва, новороссийская улица, 14, стр. 2	Юго-Восточный административный округ	ежедневно, 10:00–00:00	55.678772	37.759422	4.3	NaN	NaN	NaN	NaN	сетевой	40.0	новороссийская улица	False
456	практика кофе	кофейня	москва, чонгарский бульвар, 26а, корп. 1	Юго-Западный административный округ	пн-пт 08:00–22:00; сб,вс 09:00–22:00	55.654289	37.600722	4.4	NaN	NaN	NaN	NaN	несетевой	55.0	чонгарский бульвар	False

457 rows × 16 columns

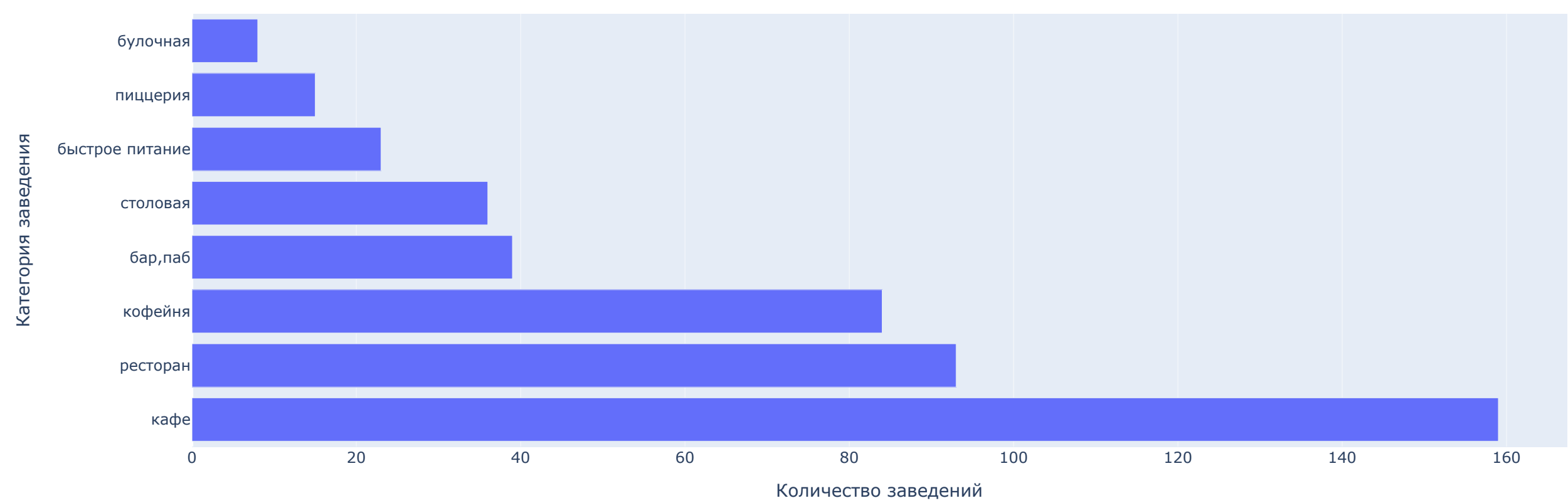
```
In [40]: # Посчитаем среди этих заведений количество по категориям
category_ocs = data_ocs.groupby('category').agg({'address': 'count'}).rename(columns={'address': 'count'}).sort_values('count', ascending=False).reset_index()
category_ocs
```

Out[40]:

	category	count
0	кафе	159
1	ресторан	93
2	кофейня	84
3	бар,паб	39
4	столовая	36
5	быстрое питание	23
6	пиццерия	15
7	булочная	8

```
In [41]: # Визуализируем соотношение единственных на улице сетевых и несетевых заведений по категориям
fig = px.bar(
    category_ocs,
    x='count',
    y='category',
    title='Соотношение единственных на улице заведений по категориям',
    labels={'category': 'Категория заведения', 'count': 'Количество заведений'},)
fig.show()
```


Соотношение единственных на улице заведений по категориям



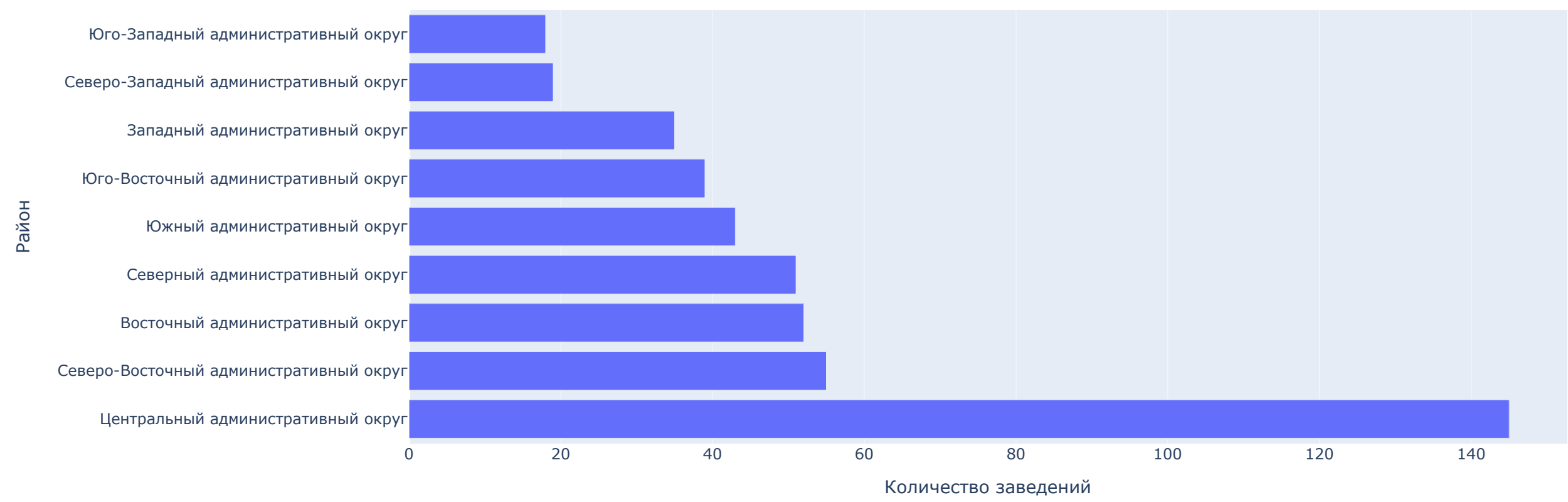
Распределение похоже на распределение всех заведений. Лидируют"одинокие" кафе(159), их почти вдвое больше, чем ресторанов(93) или кофеев(84), меньше всего булочных(8).

```
In [42]: # Посчитаем оличество по районам
districts_ocs = data_ocs.groupby('district').agg({'address': 'count'}).rename(columns={'address': 'count'}).sort_values('count', ascending=False).reset_index()
districts_ocs
```

	district	count
0	Центральный административный округ	145
1	Северо-Восточный административный округ	55
2	Восточный административный округ	52
3	Северный административный округ	51
4	Южный административный округ	43
5	Юго-Восточный административный округ	39
6	Западный административный округ	35
7	Северо-Западный административный округ	19
8	Юго-Западный административный округ	18

```
In [43]: # Визуализируем соотношение единственных на улице заведений по районам
fig = px.bar(
    districts_ocs,
    x='count',
    y='district',
    title='Соотношение единственных на улице заведений по районам',
    labels={'district': 'Район', 'count': 'Количество заведений'})
fig.show()
```

Соотношение единственных на улице заведений по районам



Центральный административный округ с большим отрывом лидирует по количеству единственных на улице заведений.

```
In [44]: # Посчитаем средний рейтинг по категориям
rating_ocs = data_ocs.groupby('category').agg({'rating': 'mean'}).rename(columns={'rating': 'mean_rating'}).round(2).sort_values('mean_rating', ascending=False).reset_index()
rating_ocs
```

	category	mean_rating
0	бар,паб	4.48
1	пиццерия	4.37
2	булочная	4.34
3	кофейня	4.31
4	столовая	4.26
5	ресторан	4.25
6	кафе	4.14
7	быстрое питание	4.08

```
In [45]: # Визуализируем соотношение рейтингов единственных на улице заведений по категориям
fig = px.bar(
    rating_ocs,
    x='mean_rating',
    y='category',
    title='Соотношение рейтингов единственных на улице заведений по категориям',
    labels={'category': 'Категория заведения', 'mean_rating': 'Средний рейтинг заведений'})
fig.update_xaxes(range=[4, 4.5])
fig.show()
```

Соотношение рейтингов единственных на улице заведений по категориям



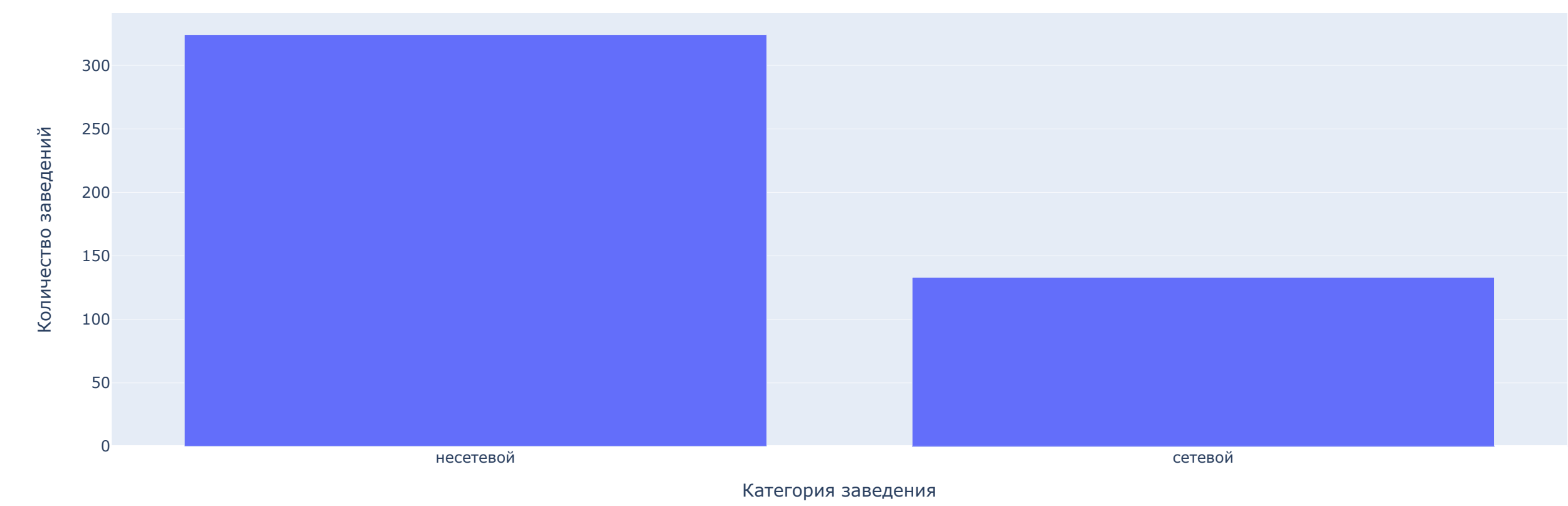
Средние рейтинги единственных на улице заведений по категориям примерно такие же как при расчёте по вскму датасету.

```
In [46]: # Посчитаем количество сетевых и несетвых заведений
chain_ocs = data_ocs['chain'].value_counts()
chain_ocs

Out[46]: несетевой    324
сетевой    133
Name: chain, dtype: int64

In [47]: # Визуализируем соотношение единственных на улице сетевых и несетевых заведений
fig = px.bar(
    chain_ocs, x=chain.index,
    y=chain_ocs,
    title='Соотношение единственных на улице сетевых и несетевых заведений',
    labels={'x': 'Категория заведения', 'y': 'Количество заведений'})
fig.show()
```

Соотношение единственных на улице сетевых и несетевых заведений



Количество сетевых заведений(324) над несетевыми(133) также преобладает среди единственных на улице.

По полученной информации можно сказать, что единственное на своей улице заведение в Москве скорей всего окажется несетевым кафе в центре с рейтингом 4,1.

Исследование средних чеков

```
In [48]: # Посчитаем медианный средний чек по районам
median_bill_dist = data.query('middle_avg_bill != 0').groupby('district').agg({'middle_avg_bill': 'median'}).sort_values('middle_avg_bill', ascending=False).reset_index()
median_bill_dist

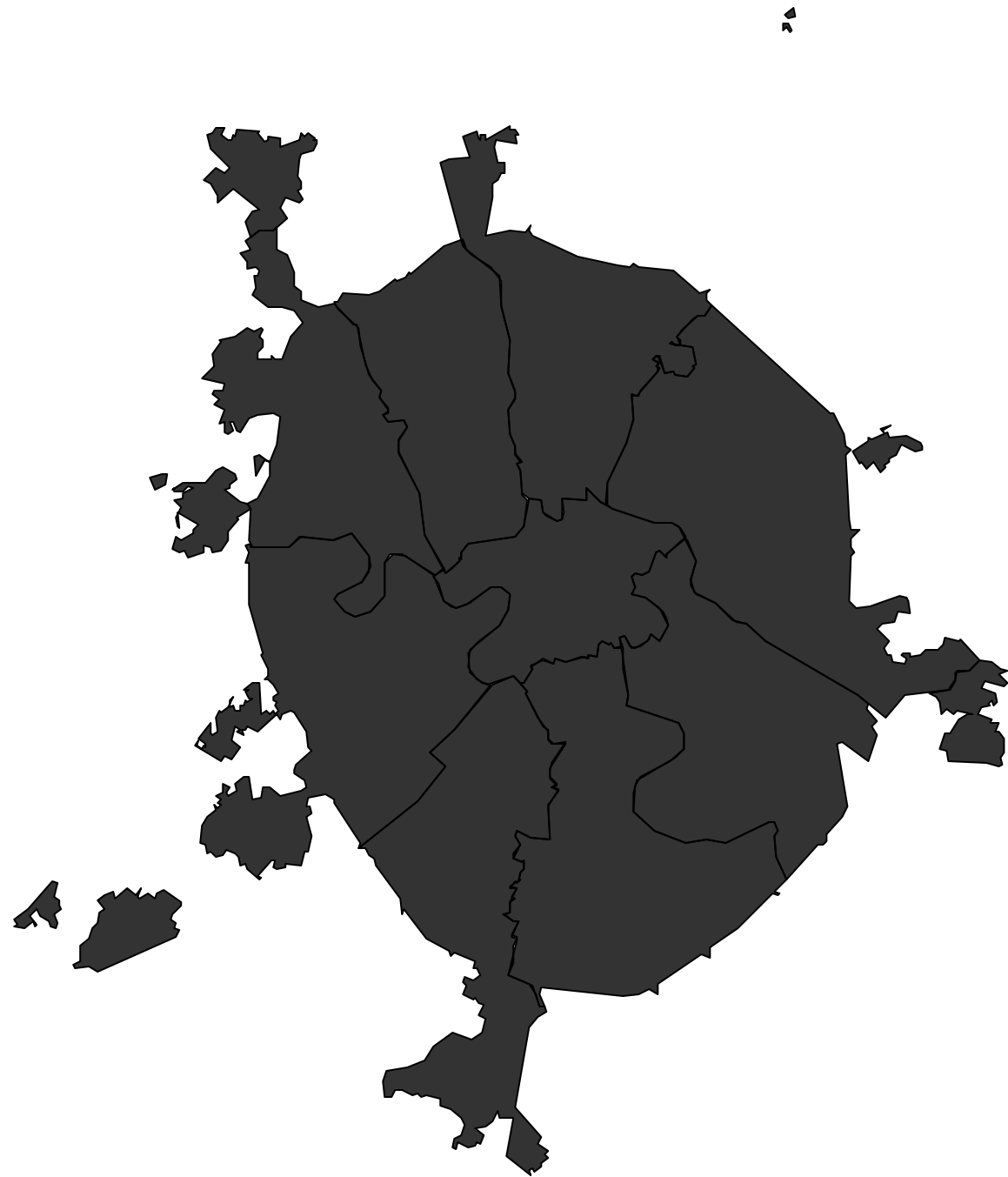
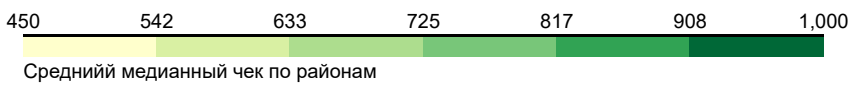
Out[48]:
```

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

```
In [49]: # создаём карту Москвы
median_bill_map = Map(location=[moscow_lat, moscow_lng], zoom_start=10, tiles='Cartodb Positron')

# создаём хороплет с помощью конструктора Choropleth и добавляем его на карту
Choropleth(
    geo_data=state_geo,
    data=median_bill_dist,
    columns=['district', 'middle_avg_bill'],
    key_on='feature.name',
    fill_color='YlGn',
    fill_opacity=0.8,
    legend_name='Средний медианный чек по районам',
).add_to(median_bill_map)

# выводим карту
median_bill_map
```



Leaflet (<https://leafletjs.com>) | © OpenStreetMap (<https://www.openstreetmap.org/copyright>) contributors © CARTO (<https://carto.com/attributions>)

По мере удаления от центра медианный средний чек падает, но только не в Западном административном округе(1000), там чек как в центре. Заметней всего падение медианного среднего чека в Северо-Восточном административном округе(500), Южном административном округе(500) и Юго-Восточном административном округе(450).

Вывод:

- В столице больше всего кафе(2377), ресторанов(2041) и кофейен(1413), менее распространены бары/пабы(765), пиццерии(633) и заведения быстрого питания(603), самыми нераспространёнными заведениями являются столовые(315) и булочные(256);
- Больше всего посадочных мест в ресторанах(86), барах/пабах(82) и кофейнях(80), поменьше посадочных мест в столовых(75) и заведениях быстрого питания(65). Меньше всего посадочных мест в кафе(60), в пиццериях(55) и в булочных(50);
- В Москве преобладают несетевые заведения, 5198 против 3205;
- В категориях кофейни(720 с:693 нс), пиццерии(330 с:303 нс) и булочные(157 с:99 нс) количество сетевых заведений преобладает над несетевыми, в остальных случаях ситуация обратная. У баров(169 с:596 нс) доля несетевых заведений больше, чем в других категориях;
- Топ-15 популярных сетей в Москве попало шесть кофейен, лидируют по количеству заведений Шоколадница(119), Домино'с пицца(76) и Додо пицца(74);
- В центре сосредоточено больше всего заведений, меньше всего заведений на северо-западе. В большинстве районов преобладают кафе и рестораны;
- Все средние рейтинги по категориям варьируются от 4,05 до 4,39, можно предположить, что это вызвано высокой конкуренцией, которая вынуждает заведения придерживаться высоких стандартов;
- Самый высокий рейтинг в заведениях в Центральном административном округе(4,38), а самый низкий в Юго-Восточном административном округе(4,1);
- В топ-15 улиц по количеству заведений попали крупные транспортные артерии города, это может свидетельствовать о важности проходимости для расположения заведений общественного питания;
- Средний чек меняется в зависимости от района, ближе к центру высокий, по мере удаления он падает, самый низкий на юго-востоке, это может быть связано с позиционированием заведений в данных районах, а также со стоимостью аренды.

Шаг. Детализируем исследование: открытие кофейни

Кофейни в москве

```

In [50]: # Оставим в датасете только кофейни и посчитаем их
data_coffee = data.query('category == "кофейня"')
print('Всего кофеен в москве:', data_coffee['name'].count())
print('Сетевых кофеен в москве:', data_coffee.query('chain == "сетевой")['name'].count())
print('Несетевых кофеен в москве:', data_coffee.query('chain != "сетевой")['name'].count())

Всего кофеен в москве: 1413
Сетевых кофеен в москве: 720
Несетевых кофеен в москве: 693

```

```
In [51]: # Посчитаем количество кофеен по районам
coffee_district = data_coffee.groupby('district').agg({'address': 'count'}).rename(columns={'address': 'count'}).sort_values('count', ascending=False).reset_index()
coffee_district
```

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

```

in [52]: # Создаем карту Москвы с начальным масштабом
coffee_map = Map(location=[moscow_lat, moscow_lng], zoom_start=10)

# Создаем пустой кластер и добавляем его на карту
marker_cluster = MarkerCluster().add_to(coffee_map)

def create_clusters(row):
    # сохраняем URL-адрес изображения со значком торгового центра с icons8,
    # это путь к файлу на сервере icons8
    icon_url = 'https://img.icons8.com/doodle/48/coffee-to-go.png'
    # создаём объект с собственной иконкой размером 30x30
    icon = CustomIcon(icon_url, icon_size=(30, 30))

    # создаём маркер с иконкой icon и добавляем его в кластер
    Marker(
        [row['lat'], row['lng']],
        popup=f"{row['name']} {row['rating']}",
        icon=icon,

```

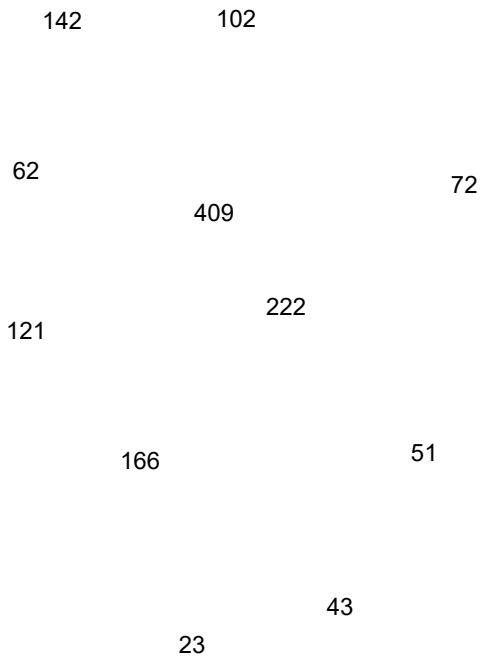
```

    ).add_to(marker_cluster)

# применяем функцию для создания кластеров к каждой строке датафрейма
data_coffee.apply(create_clusters, axis=1)

# Возвращаем объект карты для отображения
coffee_map
```

Out[52]: Make this Notebook Trusted to load map: File -> Trust Notebook



Логично, что в центре расположено больше всего кофеен, так как там большая проходимость, много офисов и мест для прогулок. В спальных районах число кофеен ниже, большинству людей не нужно бодриться кофе по дороге домой.

Круглосуточные кофейни

```
In [53]: data_coffee_24 = data_coffee[data_coffee['is_24/7']==True]
print('Круглосуточных кофеен в москве:', data_coffee_24['address'].count())
```

Круглосуточных кофеен в москве: 76

```
In [54]: # Посчитаем круглосуточные кофейни по районам
coffee_24 = data_coffee_24.groupby('district').agg({'address': 'count'}).rename(columns={'address': 'count'}).sort_values('count', ascending=False).reset_index()
coffee_24
```

Out[54]:

	district	count
0	Центральный административный округ	38
1	Западный административный округ	11
2	Юго-Западный административный округ	7
3	Северный административный округ	6
4	Восточный административный округ	5
5	Северо-Восточный административный округ	4
6	Северо-Западный административный округ	3
7	Юго-Восточный административный округ	1
8	Южный административный округ	1

```
In [55]: # Добавим на карту круглосуточные кофейни
# Создаем карту Москвы с начальным масштабом
coffee_24_map = Map(location=[moscow_lat, moscow_lng], zoom_start=10)

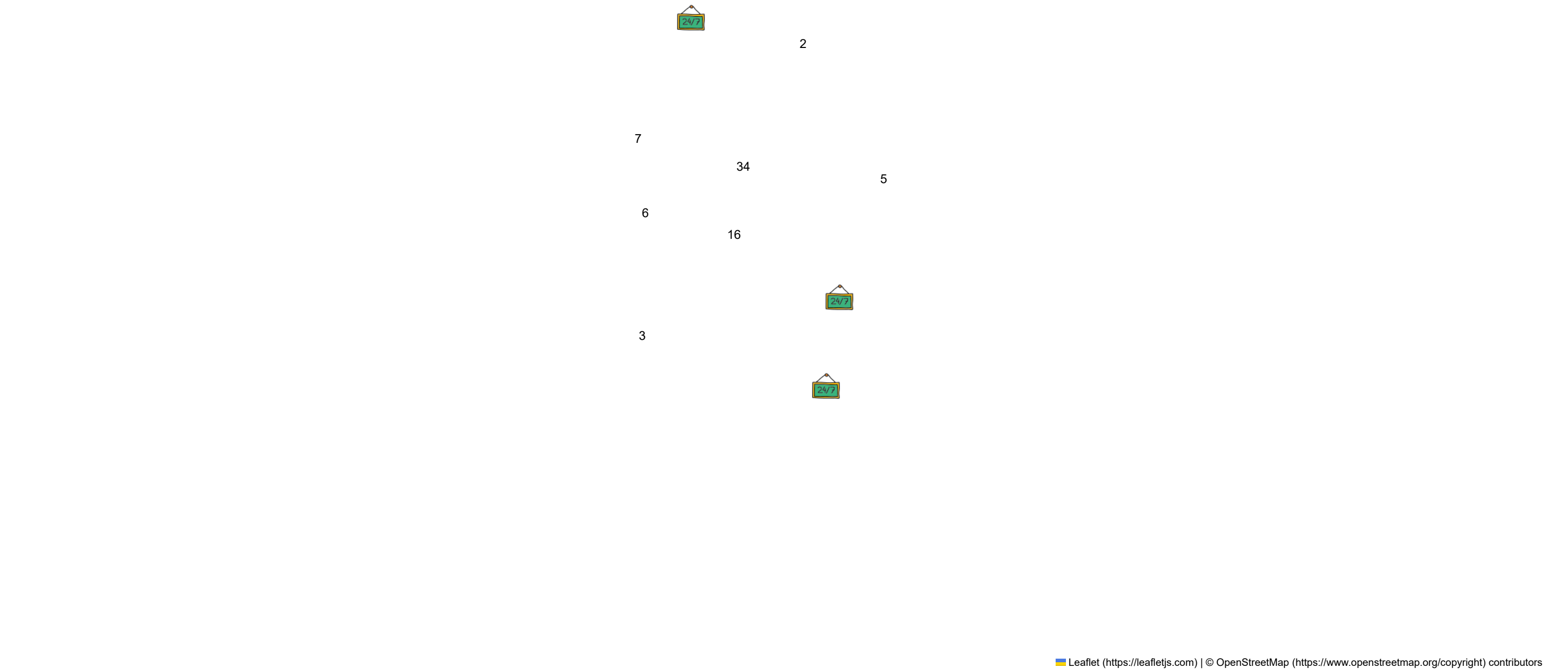
# Создаем пустой кластер и добавляем его на карту
marker_cluster = MarkerCluster().add_to(coffee_24_map)

def create_clusters(row):
    # сохраняем URL-адрес изображения со значком торгового центра с icons8,
    # это путь к файлу на сервере icons8
    icon_url = 'https://img.icons8.com/doodle/48/24-7-open-sign-.png'
    # создаём объект с собственной иконкой размером 30x30
    icon = CustomIcon(icon_url, icon_size=(30, 30))

    # создаём маркер с иконкой icon и добавляем его в кластер
    Marker(
        [row['lat'], row['lng']],
        popup=f"{row['name']} {row['rating']}",
        icon=icon,
    ).add_to(marker_cluster)

# применяем функцию для создания кластеров к каждой строке датафрейма
data_coffee_24.apply(create_clusters, axis=1)

# Возвращаем объект карты для отображения
coffee_24_map
```

76 круглосуточных кофеен, в центре как всегда больше.

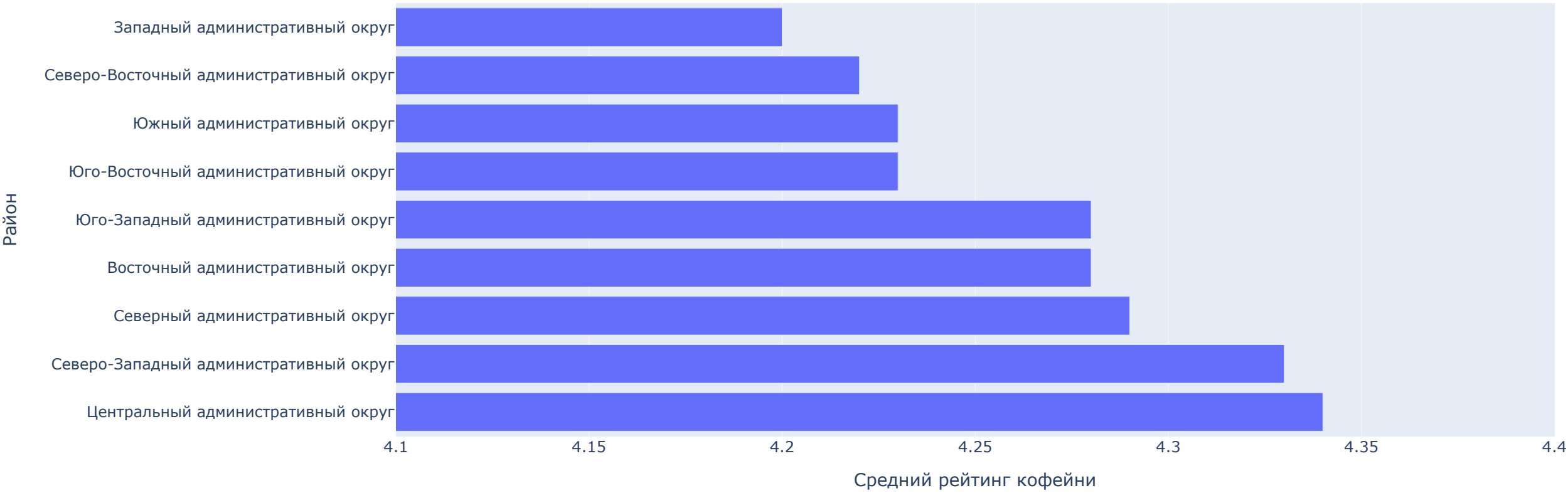
Рейтинги кофеен Москвы

```
In [56]: # Посчитаем средний рейтинг кофеен по районам
coffee_rating = data_coffee.groupby('district').agg({'rating':'mean'}).round(2).rename(columns={'rating':'mean_rating'}).sort_values('mean_rating', ascending=False).reset_index()
coffee_rating
```

	district	mean_rating
0	Центральный административный округ	4.34
1	Северо-Западный административный округ	4.33
2	Северный административный округ	4.29
3	Восточный административный округ	4.28
4	Юго-Западный административный округ	4.28
5	Юго-Восточный административный округ	4.23
6	Южный административный округ	4.23
7	Северо-Восточный административный округ	4.22
8	Западный административный округ	4.20

```
In [57]: # Построим диаграмму распределения среднего рейтинга кофеен по районам
fig = px.bar(
    coffee_rating,
    x='mean_rating',
    y='district',
    title='Соотношение рейтингов кофеен по районам',
    labels={'district':'Район', 'mean_rating':'Средний рейтинг кофейни'})
fig.update_xaxes(range=[4.1, 4.4])
fig.show()
```

Соотношение рейтингов кофеен по районам



Самые высокие рейтинги у кофеен в центре и на северо-западе(4,3), ниже всего на западе(4,2).

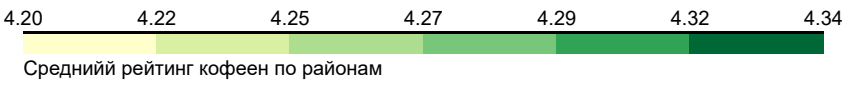
```
In [58]: # создаём карту Москвы
rating_coffee_map = Map(location=[moscow_lat, moscow_lng], zoom_start=10, tiles='Cartodb Positron')

# создаём хороплет с помощью конструктора Choropleth и добавляем его на карту
Choropleth(
    geo_data=state_geo,
    data=coffee_rating,
    columns=['district', 'mean_rating'],
    key_on='feature.name',
```

```
fill_color='YlGn',
fill_opacity=0.8,
legend_name='Средний рейтинг кофеен по районам',
).add_to(rating_coffee_map)

# выводим карту
rating_coffee_map
```

Out[58]: Make this Notebook Trusted to load map: File -> Trust Notebook



 Leaflet (<https://leafletjs.com>) | © OpenStreetMap (<https://www.openstreetmap.org/copyright>) contributors © CARTO (<https://carto.com/attributions>)

В целом по всем районам у кофеен высокий рейтинг от 4,2 до 4,3.

Стоимость чашки капучино в Москве

```
In [59]: # Посчитаем среднюю стоимость чашки капучино
cup_cost = data.groupby('district')['middle_coffee_cup'].agg(['min', 'mean', 'max']).round(2).rename(columns={'min': 'Минимальная стоимость', 'mean': 'Средняя стоимость', 'max': 'Максимальная стоимость'}).s
print('Средняя стоимость чашки капучино в Москве:', data['middle_coffee_cup'].mean().round(2))
cup_cost
```

Средняя стоимость чашки капучино в Москве: 174.72

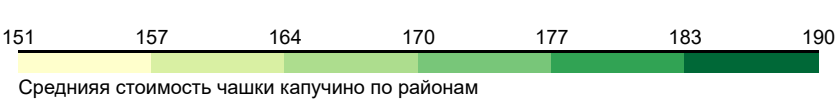
Out[59]:


	district	Минимальная стоимость	Средняя стоимость	Максимальная стоимость
0	Западный административный округ	60.0	189.64	300.0
1	Центральный административный округ	60.0	188.21	328.0
2	Юго-Западный административный округ	95.0	183.49	291.0
3	Восточный административный округ	60.0	174.02	1568.0
4	Северо-Восточный административный округ	60.0	165.33	297.0
5	Северный административный округ	60.0	165.09	325.0
6	Северо-Западный административный округ	60.0	160.46	270.0
7	Южный административный округ	60.0	157.83	275.0
8	Юго-Восточный административный округ	60.0	150.77	375.0

```
In [60]: # создаём карту Москвы
cup_cost_map = Map(location=[moscow_lat, moscow_lng], zoom_start=10, tiles='Cartodb Positron')

# создаём хороплет с помощью конструктора Choropleth и добавляем его на карту
Choropleth(
    geo_data=state_geo,
    data=cup_cost,
    columns=['district', 'Средняя стоимость'],
    key_on='feature.name',
    fill_color='YlGn',
    fill_opacity=0.8,
    legend_name='Средняя стоимость чашки капучино по районам',
).add_to(cup_cost_map)

# выводим карту
cup_cost_map
```



 Leaflet (<https://leafletjs.com>) | © OpenStreetMap (<https://www.openstreetmap.org/copyright>) contributors © CARTO (<https://carto.com/attributions>)

Дороже всего за среднюю чашку капучино придётся заплатить на западе(189,64) и в центре(188,21) Москвы, самая низкая стоимость средней чашки на юго-востоке(150,77).

Рекомендации по открытию кофейни: Целевой аудиторией для кофеен являются в основном студенты, туристы и офисные работники, поэтому могу рекомендовать открыть кофейню в местах большой проходимости данных категорий людей, к ним относятся: прилегающие к метро территории, офисные здания, университеты, парки и объекты культурного досуга. Привлекательным для открытия является Западный административный округ, так как конкуренция там ниже, чем в Центральном административном округе, а средняя стоимость чашки капучино высокая, средний рейтинг кофеен в этом районе ниже, чем в других, можно выделиться чуть меньшими ценами и сервисом лучше. По поводу формата работы необходимы дополнительные исследования, так как у нас недостаточно информации необходимости круглосуточных кофеен, возможно людям их не хватает, а конкуренция в этом сегменте намного ниже, а может их мало, потому что они не нужны.

Презентация: <https://drive.google.com/file/d/1VLbD6adKFPCyChE3s2guN9QcQbktppP0/view?usp=sharing>