

ВСТУПЛЕНИЕ

Добрый день, уважаемые коллеги! Рада вас приветствовать на презентации хода исследования и его результатов.

Описание проекта:

Анализ рынка общественного питания Москвы для инвесторов из фонда «Shut Up and Take My Money», планирующих открытие нового заведения. Исследование проводится на основе датасета заведений общественного питания Москвы, составленного на основе данных сервисов Яндекс Карты и Яндекс Бизнес за лето 2022 года.

Цель проекта:

Предоставить инвесторам информацию о рынке общественного питания Москвы для принятия обоснованного решения об открытии нового заведения. Анализ должен выявить перспективные направления, категории заведений, локации и ценовые сегменты, а также особенности конкурентной среды.

Ожидаемые результаты проекта:

- Презентация с результатами анализа, содержащая:
 - Обзор категорий заведений и их распределения по городу.
 - Анализ количества посадочных мест и их распределения по категориям.
 - Соотношение сетевых и несетевых заведений.
 - Топ популярных сетей и их характеристик.
 - Распределение заведений по административным районам.
 - Анализ средних рейтингов и чеков по категориям и районам.
 - Картограммы с отображением рейтингов и цен по районам.
 - Топ улиц с наибольшей концентрацией заведений.
 - Рекомендации по выбору категории, локации и ценового сегмента для нового заведения.
 - Выводы о ключевых особенностях рынка общественного питания Москвы.
 - Детализация исследования

ПЛАН ИССЛЕДОВАНИЯ:

1. Загрузка и подготовка данных

- 1.1 Предварительное знакомство с данными. Переименование столбцов
- 1.2 Обработка пропусков и дубликатов
- 1.3 Изменение таблицы, добавление необходимых полей
- 1.4 Изучение и проверка данных

2. Анализ данных

- 2.1 Категоризация заведений
- 2.2 Изучение сетевых и несетевых заведений. Выявление топ-15
- 2.3 Подсчёт распределения заведений по административным районам
- 2.4 Распределение средних рейтингов по районам и категориям
- 2.5 Геоаналитика заведений

3. Детализация исследования

- 3.1 Ответы на вопросы заказчика
- 3.2 Выводы и рекомендации заказчику

4. Презентация

Загрузка и подготовка данных

```
In [1]: import warnings
warnings.simplefilter('ignore')
```

```
In [2]: # !pip install folium
```

```
In [3]: # Ипорты необходимых библиотек
import pandas as pd
pd.options.display.max_colwidth = 100
from matplotlib import pyplot as plt
from datetime import timedelta, date
import json
from folium import Map, Choropleth, Marker
import seaborn as sns
import numpy as np
import plotly.express as px
from folium.plugins import MarkerCluster
import warnings
warnings.simplefilter('ignore')
```

```
In [4]: # Функция для извлечения наименования улицы
def extract_street(address):
    if isinstance(address, str):
        parts = address.split(',')
        if len(parts) > 1:
            return parts[1].strip()
    return None
```

```
In [5]: # Функцию, которая создает булев режим работы
def operating_mode(hours):
    if isinstance(hours, str):
        if 'ежедневно' and 'круглосуточно' in hours:
            return True
    else:
```

```
        return False
    return None
```

```
In [6]: # создание маркера
def create_clusters(row):
    Marker(
        [row['lat'], row['lng']],
        popup=f"{row['name']} {row['rating']}",
    ).add_to(marker_cluster)
```

```
In [7]: # Спроектируем графику
def param_coffee_of_district(param, data, title):
    plt.figure(figsize=(11, 8))
    ax = sns.barplot(x=param[0], y=param[1], data=data.sort_values(by=param[1],

    for p in ax.patches:
        height = p.get_height()
        ax.annotate(f'{height:.2f}',
                    (p.get_x() + p.get_width() / 2., height),
                    ha='center', va='bottom')

    plt.title(title)
    plt.xlabel('Округ')
    plt.ylabel('Количество')
    plt.xticks(rotation=45, ha='right')
    plt.tight_layout()
    plt.show()
```

```
In [8]: places_data = pd.read_csv('moscow_places.csv')
```

Предварительное знакомство с данными.

```
In [9]: places_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 [10]: places_data.head(5)
```

Out[10]:

	name	category	address	district	hours	lat	
0	WoWфли	кафе	Москва, улица Дыбенко, 7/1	Северный административный округ	ежедневно, 10:00–22:00	55.878494	37
1	Четыре комнаты	ресторан	Москва, улица Дыбенко, 36, корп. 1	Северный административный округ	ежедневно, 10:00–22:00	55.875801	37
2	Хазри	кафе	Москва, Клязьминская улица, 15	Северный административный округ	пн-чт 11:00– 02:00; пт,сб 11:00– 05:00; вс 11:00–02:00	55.889146	37
3	Dormouse Coffee Shop	кофейня	Москва, улица Маршала Федоренко, 12	Северный административный округ	ежедневно, 09:00–22:00	55.881608	37
4	Иль Марко	пиццерия	Москва, Правобережная улица, 15	Северный административный округ	ежедневно, 10:00–22:00	55.881166	37

In [11]: `places_data['name'].nunique()`

Out[11]: 5614

В данных имеем 8406 строк в них представлено 5614 заведений в г. Москва. Видим корректное наименование столбцов и соответствие типов для всех полей кроме "seats", так как количество посадочных мест не может быть вещественным, данное поле необходимо перевести в целочисленный тип. Так же предварительно можно увидеть множество пропусков в данных.

In [12]: `places_data['seats'] = places_data['seats'].astype('Int64')`
`places_data.dtypes`

Out[12]:

name	object
category	object
address	object
district	object
hours	object
lat	float64
lng	float64
rating	float64
price	object
avg_bill	object
middle_avg_bill	float64
middle_coffee_cup	float64
chain	int64
seats	Int64
dtype:	object

In [13]: `places_data.head(5)`

Out[13]:

	name	category	address	district	hours	lat	
0	WoWfli	кафе	Москва, улица Дыбенко, 7/1	Северный административный округ	ежедневно, 10:00–22:00	55.878494	37
1	Четыре комнаты	ресторан	Москва, улица Дыбенко, 36, корп. 1	Северный административный округ	ежедневно, 10:00–22:00	55.875801	37
2	Хазри	кафе	Москва, Клязьминская улица, 15	Северный административный округ	пн-чт 11:00– 02:00; пт,сб 11:00– 05:00; вс 11:00–02:00	55.889146	37
3	Dormouse Coffee Shop	кофейня	Москва, улица Маршала Федоренко, 12	Северный административный округ	ежедневно, 09:00–22:00	55.881608	37
4	Иль Марко	пиццерия	Москва, Правобережная улица, 15	Северный административный округ	ежедневно, 10:00–22:00	55.881166	37

Обработка пропусков и дубликатов.

In [14]: `places_data.duplicated().sum()`

Out[14]: `np.int64(0)`

In [15]: `name_add = pd.concat([places_data['name'].str.lower(), places_data['address'].str.lower()], axis=1)`
`name_add['name_add'] = name_add['name'] + ' ' + name_add['address']`

In [16]: `len(places_data) - len(name_add['name_add'].unique())`

Out[16]: `4`

Полученной значение 4, это говорит о том что в датасете имеется 4 заведения расположенных по одному адресу с одним названием.

In [17]: `duplicates = name_add['name_add'].value_counts().sort_values().tail(4)`
`duplicates.index`

Out[17]: `Index(['more poke москва, волоколамское шоссе, 11, стр. 2',
'хлеб да выпечка москва, ярцевская улица, 19',
'раковарня клешни и хвосты москва, проспект мира, 118',
'кафе москва, парк ангарские пруды'],
dtype='object', name='name_add')`

In [18]: `name_add[name_add['name_add'].isin(duplicates.index)][['name', 'address']].index`

Out[18]: `Index([215, 1511, 2420, 3109], dtype='int64')`

In [19]: `places_data.drop([215, 1511, 2420, 3109], inplace=True)`

```
plases_data.reset_index(drop=True)
```

Out[19]:

	name	category	address	district	hours	
0	WoWfli	кафе	Москва, улица Дыбенко, 7/1	Северный административный округ	ежедневно, 10:00–22:00	55.878
1	Четыре комнаты	ресторан	Москва, улица Дыбенко, 36, корп. 1	Северный административный округ	ежедневно, 10:00–22:00	55.875
2	Хазри	кафе	Москва, Клязьминская улица, 15	Северный административный округ	пн-чт 11:00– 02:00; пт,сб 11:00–05:00; вс 11:00–02:00	55.889
3	Dormouse Coffee Shop	кофейня	Москва, улица Маршала Федоренко, 12	Северный административный округ	ежедневно, 09:00–22:00	55.881
4	Иль Марко	пиццерия	Москва, Правобережная улица, 1Б	Северный административный округ	ежедневно, 10:00–22:00	55.881
...
8397	Суши Мания	кафе	Москва, Профсоюзная улица, 56	Юго-Западный административный округ	ежедневно, 09:00–02:00	55.670
8398	Миславнес	кафе	Москва, Пролетарский проспект, 19, корп. 1	Южный административный округ	ежедневно, 08:00–22:00	55.640
8399	Самовар	кафе	Москва, Люблинская улица, 112А, стр. 1	Юго-Восточный административный округ	ежедневно, круглосуточно	55.648
8400	Чайхана Sabr	кафе	Москва, Люблинская улица, 112А, стр. 1	Юго-Восточный административный округ	ежедневно, круглосуточно	55.648
8401	Kebab Time	кафе	Москва, Россошанский проезд, 6	Южный административный округ	ежедневно, круглосуточно	55.598

8402 rows × 14 columns



In [20]: `name_add[name_add['name_add'].isin(duplicates.index)][['name', 'address']]`

Out[20]:

	name	address
189	кафе	москва, парк ангарские пруды
215	кафе	москва, парк ангарские пруды
1430	more poke	москва, волоколамское шоссе, 11, стр. 2
1511	more poke	москва, волоколамское шоссе, 11, стр. 2
2211	раковарня клешни и хвосты	москва, проспект мира, 118
2420	раковарня клешни и хвосты	москва, проспект мира, 118
3091	хлеб да выпечка	москва, ярцевская улица, 19
3109	хлеб да выпечка	москва, ярцевская улица, 19

In [21]: `places_data.isna().sum()`

Out[21]:

name	0
category	0
address	0
district	0
hours	535
lat	0
lng	0
rating	0
price	5087
avg_bill	4586
middle_avg_bill	5253
middle_coffee_cup	7867
chain	0
seats	3610

dtype: int64

In [22]: `print(f'В поле со средней стоимостью заказа "avg_bill" {8406 - 4590} записей')`

В поле со средней стоимостью заказа "avg_bill" 3816 записей

In [23]: `print(f'Сумме на 2 поля "middle_avg_bill" и "middle_coffee_cup" {(8406 - 5257) +`

Сумме на 2 поля "middle_avg_bill" и "middle_coffee_cup" 3684 записей

In [24]: `print(f'Разница числа записей в данных полях {3816 - 3684}')`

Разница числа записей в данных полях 132

In [25]: `places_data[['avg_bill', 'middle_avg_bill', 'middle_coffee_cup']][places_data['a
places_data['middle_avg_bill'].isna() &
places_data['middle_coffee_cup'].isna()]`

Out[25]:

	avg_bill	middle_avg_bill	middle_coffee_cup
10	Цена бокала пива:250–350 ₽	NaN	NaN
67	Цена бокала пива:120–350 ₽	NaN	NaN
97	Цена бокала пива:90–230 ₽	NaN	NaN
241	Цена бокала пива:160–499 ₽	NaN	NaN
417	Цена бокала пива:199–300 ₽	NaN	NaN
...
8129	Цена бокала пива:280–410 ₽	NaN	NaN
8141	Цена бокала пива:190–390 ₽	NaN	NaN
8144	Цена бокала пива:220–250 ₽	NaN	NaN
8316	Цена бокала пива:160–390 ₽	NaN	NaN
8387	Цена бокала пива:от 140 ₽	NaN	NaN

132 rows × 3 columns

В поле "avg_bill" нет значений, которые могли бы прописать в столбцы "middle_avg_bill" и "middle_coffee_cup" так как разница в 132 записи нивелируется наличием записей с ценой бокала пива, для которой нет отдельного поля и при учете данных строк количество пропусков для "avg_bill" и "middle_avg_bill" + "middle_coffee_cup" равняется. Явные дубликаты в данном датасете выявить не удалось.

Изменение таблицы, добавление необходимых полей.

```
In [26]: # Проверим, присутствуют ли ошибки в указании, является ли заведение частью сети
single_names = places_data['name'].value_counts()[places_data['name'].value_coun

# Фильтруем DataFrame: имена должны встречаться только один раз И chain должно б
filtered_df = places_data[(places_data['name'].isin(single_names)) & (places_dat
filtered_df['name']]
```

```
Out[26]: 133      Крепери
206      Vintage
382      Halal food
613      Wayрма
745      В своей тарелке
...
7818     Кофейник
7874     SUШи
7899     Роллофф
7905     Easy Pizza
7955     Донер хаус
Name: name, Length: 66, dtype: object
```

```
In [27]: # Используем .loc для изменения значений в столбце 'chain' по указанным индексам
places_data.loc[filtered_df.index, 'chain'] = 0
```



```
In [28]: # Перепроверяем
single_names = places_data['name'].value_counts()[places_data['name'].value_counts() > 1]

# Фильтруем DataFrame: имена должны встречаться только один раз И chain должно быть
filtered_df = places_data[(places_data['name'].isin(single_names)) & (places_data['chain'] == 0)]
filtered_df['name']
```

```
Out[28]: Series([], Name: name, dtype: object)
```

После проведенной замены видим, что в данных больше нет ошибки и все заведения, которые в датасете встречаются лишь 1 раз в записи "chain" получили значение "0".

```
In [29]: # Использую функцию extract_street для извлечения наименования улицы (функция об
places_data['street'] = places_data['address'].apply(extract_street)
places_data['street']
```

```
Out[29]: 0          улица Дыбенко
1          улица Дыбенко
2      Клязьминская улица
3      улица Маршала Федоренко
4      Правобережная улица
...
8401      Профсоюзная улица
8402      Пролетарский проспект
8403      Люблинская улица
8404      Люблинская улица
8405      Россошанский проезд
Name: street, Length: 8402, dtype: object
```

```
In [30]: # Использую функцию operating_mode для для создание булев (функция объявлена в н
places_data['is_24_7'] = places_data['hours'].apply(operating_mode)
places_data[['hours', 'is_24_7']].head(11)
```

```
Out[30]:
```

	hours	is_24_7
0	ежедневно, 10:00–22:00	False
1	ежедневно, 10:00–22:00	False
2	пн-чт 11:00–02:00; пт,сб 11:00–05:00; вс 11:00–02:00	False
3	ежедневно, 09:00–22:00	False
4	ежедневно, 10:00–22:00	False
5	ежедневно, 10:00–23:00	False
6	пн 15:00–04:00; вт-вс 15:00–05:00	False
7	пн-чт 10:00–22:00; пт,сб 10:00–23:00; вс 10:00–22:00	False
8	ежедневно, 10:00–22:00	False
9	ежедневно, 12:00–00:00	False
10	ежедневно, круглосуточно	True

Заменяли значения является ли сетью для заведений, которые сетью не являются. Для удобства дальнейшей работы в датафрейме создали два новых столбца: 'street', содержащий название улицы, и 'is_24_7', принимающий булевы значения.

Все необходимые действия для предобработки данных совершены можем далее приступить к анализу данных.

Анализ данных

```
In [31]: # Задаю единый стиль оформления графиков
sns.set_style('whitegrid')
```

Категоризация заведений

- Подсчет количества заведений по каждой категории.
- Визуализация
- Группировка данных по категориям и подсчет суммы посадочных мест
- Визуализация

```
In [32]: # Считаю количества заведений по каждой категории
category_counts = places_data['category'].value_counts()
category_counts
```

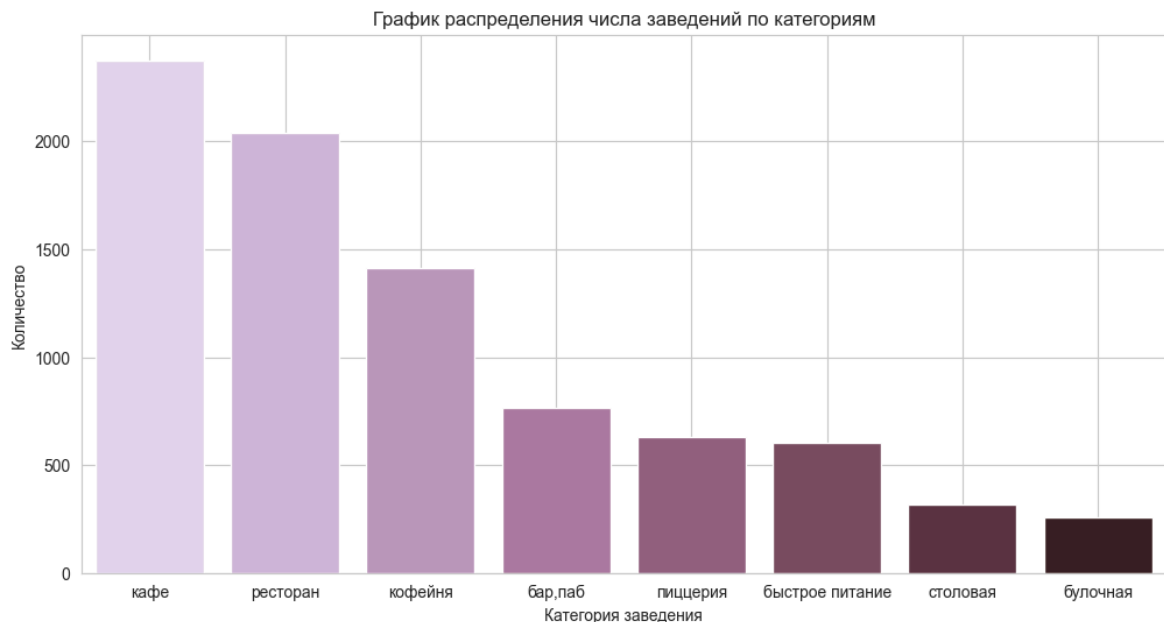
```
Out[32]: category
кафе                2376
ресторан            2042
кофейня             1413
бар, паб             764
пиццерия             633
быстрое питание     603
столовая             315
булочная            256
Name: count, dtype: int64
```

```
In [33]: places_data['category'].count()
```

```
Out[33]: np.int64(8402)
```

В данных представлено 8402 заведений, разделенных на 8 категорий заведений общепита.

```
In [34]: plt.figure(figsize=(12, 6))
sns.barplot(x=category_counts.index, y=category_counts.values, palette="ch:start")
plt.xlabel('Категория заведения')
plt.ylabel('Количество')
plt.title('График распределения числа заведений по категориям')
plt.grid(True)
plt.show()
```

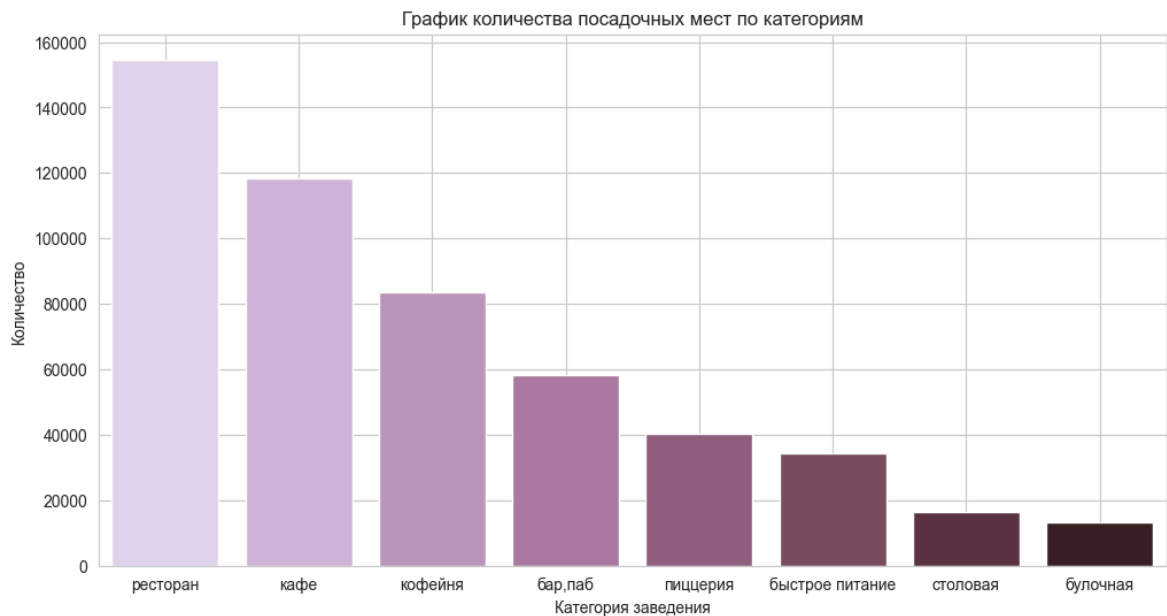


Видим что в г Москва очень популярны кафе и рестораны. Быстрые перекусы менее популярны столовые и булочные встречаются редко.

```
In [35]: # Общее кол-во посадочных мест на все заведения Москвы
seat_count = places_data.groupby('category')['seats'].sum().sort_values(ascending=True)
```

```
Out[35]: category
ресторан      154681
кафе          118494
кофейня       83511
бар, паб      58131
пиццерия      40350
быстрое питание 34513
столовая      16359
булочная     13229
Name: seats, dtype: Int64
```

```
In [36]: plt.figure(figsize=(12, 6))
sns.barplot(x=seat_count.index, y=seat_count.values, palette="ch:start=1,rot=-.3")
plt.xlabel('Категория заведения')
plt.ylabel('Количество')
plt.title('График количества посадочных мест по категориям')
plt.grid(True)
plt.show()
```



```
In [37]: places_data.groupby('category')['seats'].median().sort_values(ascending=False)
```

```
Out[37]: category
ресторан      86.0
бар, паб      82.0
кофейня       80.0
столовая      75.5
быстрое питание 65.0
кафе          60.0
пиццерия      55.0
булочная     50.0
Name: seats, dtype: Float64
```

- Форматы с длительным пребыванием: Рестораны, бары/пабы и ориентированы на длительное пребывание посетителей, что отражается в более высоком медианном количестве посадочных мест.
- Форматы с быстрым обслуживанием: Кафе, пиццерии, быстрое питание, столовые и булочные в основном ориентированы на быстрое обслуживание и заказы на вынос, поэтому имеют меньшее количество посадочных мест.

Медианное количество мест в барах/пабах (82.5) выше, чем у кафе (60), несмотря на то, что кафе имеют гораздо большее общее количество мест. Это может говорить о том, что бары/пабы ориентированы на более плотную посадку и большее количество посетителей на квадратный метр.

Прослеживается закономерность: чем выше медианное количество мест, тем выше общее количество мест (за исключением кофеен и баров). Это означает, что размеры заведений определенной категории в целом соответствуют их общему количеству.

Изучение сетевых и несетевых заведений. Выведение топ-15

- Подсчет количества сетевых и несетевых заведений

- Визуализация
- Группировка данных по категориям и подсчет количества сетевых заведений в каждой
- Визуализация
- Группировка по названию заведения и подсчет количества заведений
- Визуализация

```
In [38]: print(f'Несетевых заведений: {places_data["chain"].value_counts().values[0]}\nСетевых заведений: {places_data["chain"].value_counts().values[1]}')
```

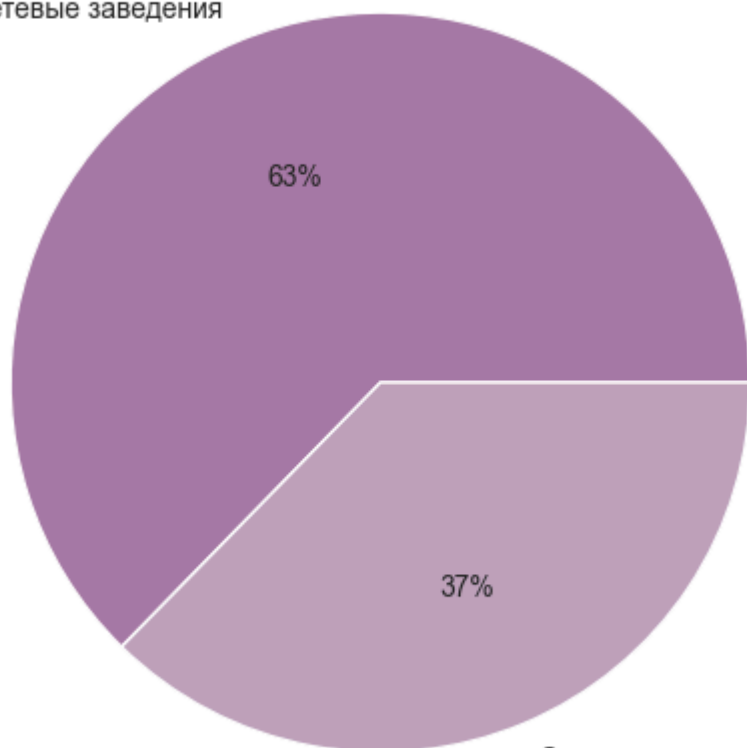
Несетевых заведений: 5265

Сетевых заведений: 3137

```
In [39]: plt.figure(figsize=(12, 6))
plt.pie(places_data["chain"].value_counts(), labels=['Несетевые заведения', 'Сетевые заведения'], autopct='%1.1f%%')
plt.title('График соотношения сетевых и несетевых заведений')
plt.grid(True)
plt.show()
```

График соотношения сетевых и несетевых заведений

Несетевые заведения



Сетевые заведения

Несетевые заведения встречаются чаще

```
In [40]: # Создаю табл с разделением по "сеть, не сеть" и категориям заведений с количеством
chain_of_category = places_data.groupby(['chain', 'category'], as_index=False)['count'].sum()
chain_of_category['total_name'] = chain_of_category['category'].map(chain_of_category['count'])
chain_of_category['fraction'] = chain_of_category['name'] / chain_of_category['total_name']
```

```
In [41]: # Заменяем 0 и 1 на соответствующие 0: 'несетевые', 1: 'сетевые'
chain_of_category['chain'] = chain_of_category['chain'].replace({0: 'несетевые', 1: 'сетевые'})
```

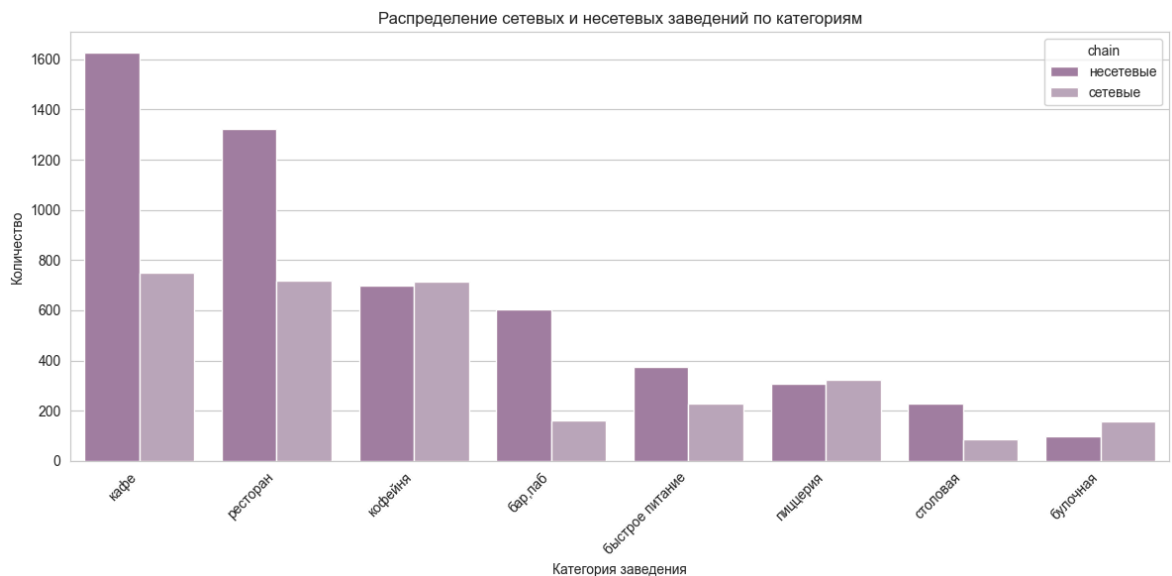
chain_of_category

Out[41]:

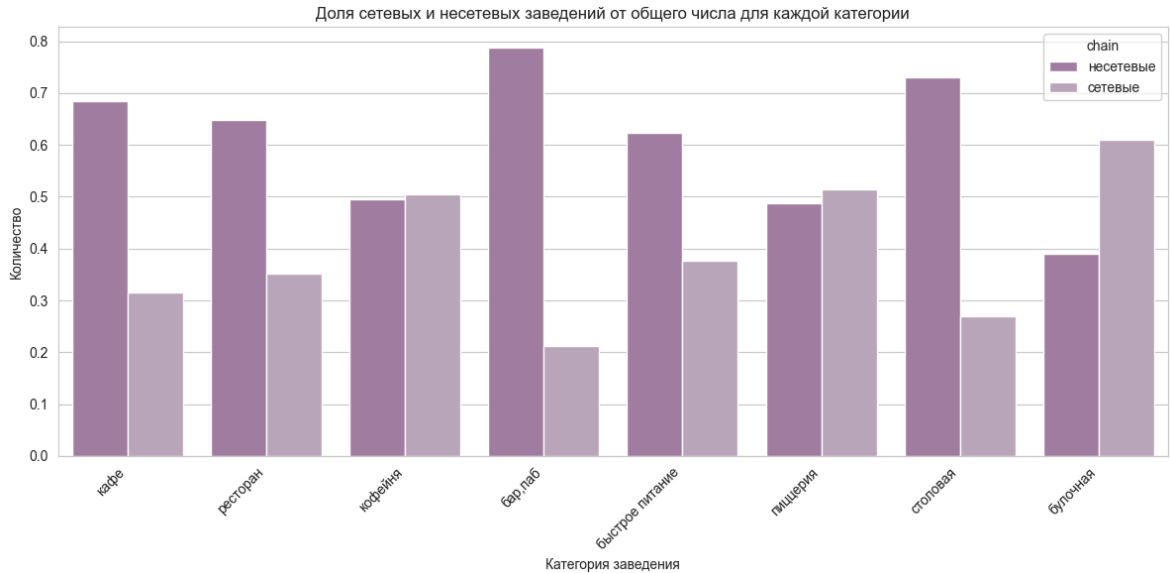
	chain	category	name	total_name	fraction
3	несетевые	кафе	1627	2376	0.684764
6	несетевые	ресторан	1323	2042	0.647894
11	сетевые	кафе	749	2376	0.315236
14	сетевые	ресторан	719	2042	0.352106
12	сетевые	кофейня	714	1413	0.505308
4	несетевые	кофейня	699	1413	0.494692
0	несетевые	бар,паб	602	764	0.787958
2	несетевые	быстрое питание	376	603	0.623549
13	сетевые	пиццерия	325	633	0.513428
5	несетевые	пиццерия	308	633	0.486572
7	несетевые	столовая	230	315	0.730159
10	сетевые	быстрое питание	227	603	0.376451
8	сетевые	бар,паб	162	764	0.212042
9	сетевые	булочная	156	256	0.609375
1	несетевые	булочная	100	256	0.390625
15	сетевые	столовая	85	315	0.269841

In [42]:

```
plt.figure(figsize=(12, 6)) # Задаем размер графика
sns.barplot(x='category', y='name', hue='chain', data=chain_of_category, palette
plt.title('Распределение сетевых и несетевых заведений по категориям')
plt.xlabel('Категория заведения')
plt.ylabel('Количество')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



```
In [43]: plt.figure(figsize=(12, 6)) # Задаем размер графика
sns.barplot(x='category', y='fraction', hue='chain', data=chain_of_category, pal
plt.title('Доля сетевых и несетевых заведений от общего числа для каждой категор
plt.xlabel('Категория заведения')
plt.ylabel('Количество')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



В большинстве случаев лидируют несетевые заведения, исключением становятся кофейни пиццерии и булочные. По графику долей несетевых заведений видим, что для булочных подавляющее большинство составляют сетевые заведения.

```
In [44]: # Выводи топ-15 по количеству заведений
top_places = places_data.groupby(['category', 'name', 'chain'], as_index=False).
```

```
In [45]: top_places
```

Out[45]:

	category	name	chain	size
2083	кафе	Кафе	0	158
3861	кофейня	Шоколадница	1	119
4063	пиццерия	Домино'с Пицца	1	76
4058	пиццерия	Додо Пицца	1	74
3343	кофейня	One Price Coffee	1	71
5783	ресторан	Яндекс Лавка	1	69
3181	кофейня	Cofix	1	65
4572	ресторан	Prime	1	49
3560	кофейня	КОФЕПОРТ	1	42
2249	кафе	Кулинарная лавка братьев Караваевых	1	39
5541	ресторан	Теремок	1	36
5332	ресторан	Ресторан	0	33
1236	быстрое питание	Шаурма	0	32
3123	кофейня	CofeFest	1	31
2879	кафе	Чайхана	1	26

In [46]: color_map = {'кафе': '#dedede', 'кофейня': '#d4c1d4', 'пиццерия': '#a883a8', 'ре

```
fig = px.bar(top_places,
              x='name',
              y='size',
              text='size',
              color='category',
              color_discrete_map=color_map,
              hover_data=['category'],
              title='Топ заведения по количеству')

fig.update_traces(hovertemplate='<b>{x}</b><br>Количество: {y}<br>Категория: %')

fig.update_layout(xaxis_title="Название заведения",
                  yaxis_title="Количество",
                  xaxis_categoryorder='total descending',
                  plot_bgcolor='white',
                  yaxis=dict(showgrid=True, gridcolor='lightgrey', gridwidth=1))

fig.show()
```



```
In [47]: top_places['category'].value_counts()
```

```
Out[47]: category
кофейня      5
ресторан     4
кафе         3
пиццерия     2
быстрое питание 1
Name: count, dtype: int64
```

```
In [48]: top_places['chain'].value_counts()
```

```
Out[48]: chain
1      12
0       3
Name: count, dtype: int64
```

Самые многочисленные заведения это кофейни 30 % после них рестораны 27 % от общего числа заведений. Большинство из представленных заведений являются сетевыми 80 % в топе от общего числа топа.

Подсчёт распределения заведений по административным районам.

- Подсчет общего количества заведений по районам
- Подсчет количества заведений каждой категории по районам

- Визуализация

```
In [49]: # Подсчёт заведений по округам
places_data.groupby(['district'])['name'].count().sort_values(ascending=False)
```

```
Out[49]: district
Центральный административный округ      2242
Северный административный округ          898
Южный административный округ             892
Северо-Восточный административный округ  890
Западный административный округ          850
Восточный административный округ         798
Юго-Восточный административный округ     714
Юго-Западный административный округ      709
Северо-Западный административный округ   409
Name: name, dtype: int64
```

```
In [50]: # Подсчёт заведений по округа и категориям
district_places = places_data.groupby(['district', 'category'], as_index=False).head(5)
```

```
Out[50]:
```

	district	category	count
46	Центральный административный округ	ресторан	670
43	Центральный административный округ	кафе	464
44	Центральный административный округ	кофейня	428
40	Центральный административный округ	бар,паб	364
51	Юго-Восточный административный округ	кафе	282

В центральном округе расположено наибольшее число заведений, самые распространенные заведения по категории это рестораны примерно 30 от всего числа заведений в Центральном округе. В Северо-Западном административном округе представлено наименьшее число заведений.

```
In [51]: # for i in district_places['district'].unique():
#         print(f"{i}\n{round(district_places[district_places['district'] == i]['cou
```

```
In [52]: district_places
```

Out[52]:

	district	category	count
46	Центральный административный округ	ресторан	670
43	Центральный административный округ	кафе	464
44	Центральный административный округ	кофейня	428
40	Центральный административный округ	бар,паб	364
51	Юго-Восточный административный округ	кафе	282
...
32	Северо-Западный административный округ	бар,паб	23
39	Северо-Западный административный округ	столовая	18
63	Юго-Западный административный округ	столовая	17
49	Юго-Восточный административный округ	булочная	13
33	Северо-Западный административный округ	булочная	12

72 rows × 3 columns

In [53]:

```
district_totals = district_places.groupby('district')['count'].sum()
district_places['district_total'] = district_places['district'].map(district_tot
district_places['percentage'] = round((district_places['count'] / district_place
district_places.sort_values(['district', 'count'], ascending=False).iloc[:8]
```

Out[53]:

	district	category	count	district_total	percentage
67	Южный административный округ	кафе	264	892	30.0
59	Юго-Западный административный округ	кафе	238	709	34.0
51	Юго-Восточный административный округ	кафе	282	714	39.0
46	Центральный административный округ	ресторан	670	2242	30.0
35	Северо-Западный административный округ	кафе	115	409	28.0
27	Северо-Восточный административный округ	кафе	269	890	30.0
19	Северный административный округ	кафе	234	898	26.0
11	Западный административный округ	кафе	238	850	28.0
3	Восточный административный округ	кафе	272	798	34.0

In [54]:

```
district_places.sort_values(['district', 'count'], ascending=False).iloc[7::8]
```

Out[54]:


	district	category	count	district_total	percentage
65	Южный административный округ	булочная	25	892	3.0
63	Юго-Западный административный округ	столовая	17	709	2.0
49	Юго-Восточный административный округ	булочная	13	714	2.0
41	Центральный административный округ	булочная	50	2242	2.0
33	Северо-Западный административный округ	булочная	12	409	3.0
25	Северо-Восточный административный округ	булочная	28	890	3.0
17	Северный административный округ	булочная	39	898	4.0
15	Западный административный округ	столовая	24	850	3.0
1	Восточный административный округ	булочная	25	798	3.0

In [55]:

```
# Создаем график:
fig = px.bar(district_places,
             x='district',
             y='count',
             text='percentage',
             color='category',
             title='Распределение категорий заведений по районам',
             labels={'district': 'Район', 'count': 'Количество заведений', 'category': 'Категория'},
             hover_data=['percentage'])

fig.update_layout(
    yaxis_title='Количество заведений',
    xaxis_title='Район',
    title_x=0.5,
    legend_title='Категория',
    xaxis_tickangle=-45,
    plot_bgcolor='white', yaxis=dict(
        showgrid=True, gridcolor='lightgrey', gridwidth=1),
    width=1000,
    height=900
)

fig.show()
```



```
In [56]: fig = px.bar(district_plases,  
                    x='district',  
                    y='count',  
                    text='count',  
                    color='category',
```

```
        hover_data=['count'],
        labels={'count': 'Количество заведений', 'district': 'Район', 'cate
        title='Количество заведений по районам и категориям')
fig.update_layout(plot_bgcolor='white',
                  xaxis_tickangle=-45,
                  yaxis=dict(showgrid=True, gridcolor='lightgrey', gridwidth=1),
                  width=1000,
                  height=900)
fig.show()
```



Наиболее развитый в плане пищевой инфраструктуры является Центральный административный округ. Северо-Западный наименее развитый. В остальных округах примерно одинаковое количество заведений и примерно на одном уровне представлена каждая категория. Так же заметно, что каждой категории заведений

представлена примерно одинаковым количеством доли от всего числа заведений в каждом отдельном округе. Может показаться что рестораны резко преобладают в Центральном округе, но это не так он в пределах 20 % -30 % процентного соотношения в других районах, хотя и является лидером. Зато отличается число баров и пабов в Центральном округе, по процентному соотношению их в 2 раза больше, чем в остальных округах, это может быть связано с тем что Центральный округ является популярным местом для отдыха и туризма, и в нём наиболее активна ночная жизнь.

Распределение средних рейтингов по районам и категориям

- Группировка данных по категориям и вычисление среднего рейтинга
- Визуализация
- Вычисление среднего рейтинга по районам
- Фоновая картограмма со средним рейтингом по районам

```
In [57]: # Смотрю основные статистические показатели рейтинга всех заведений
print('Медиана рейтингов:', places_data['rating'].mean())
places_data['rating'].describe()
```

Медиана рейтингов: 4.230016662699356

```
Out[57]: count      8402.000000
mean         4.230017
std          0.470320
min          1.000000
25%          4.100000
50%          4.300000
75%          4.400000
max          5.000000
Name: rating, dtype: float64
```

```
In [58]: # Перцентили, в данный интервал попадает 90 % оценок
np.quantile(places_data["rating"], [0.05, 0.95])
```

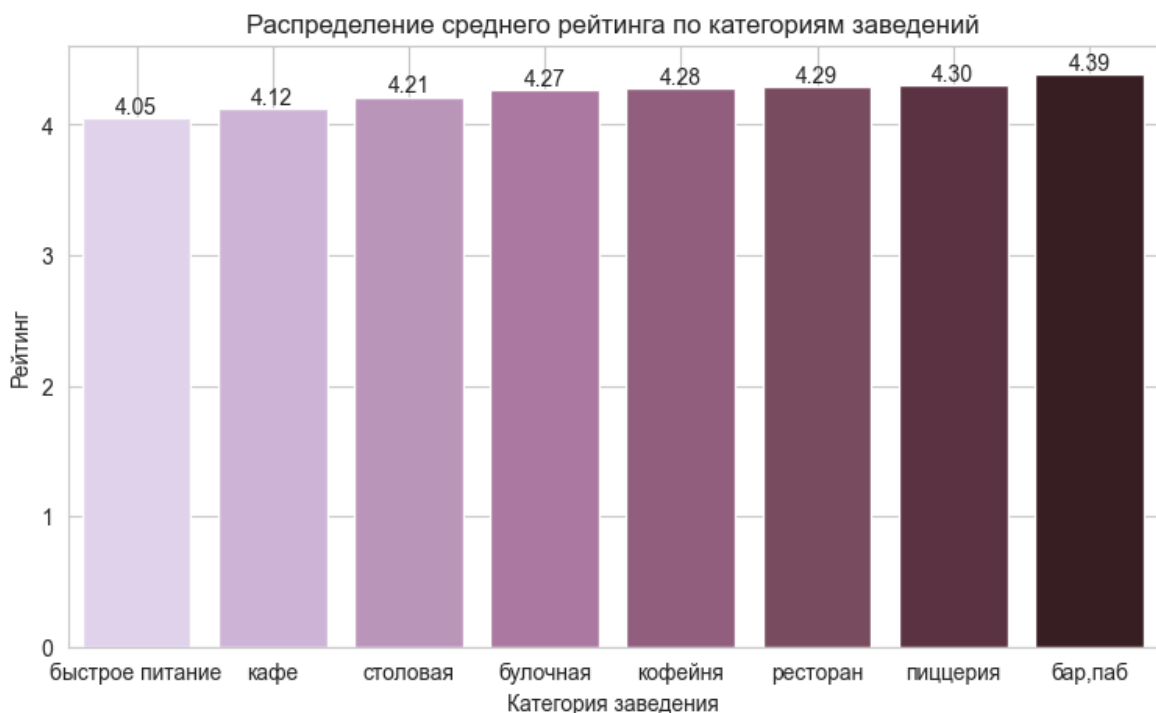
```
Out[58]: array([3.4, 4.8])
```

1. В целом, отзывы положительные: Большинство заведений получают хорошие оценки (в среднем выше 4,23 баллов).
2. Небольшой разброс: Оценки достаточно стабильны, без сильных колебаний.
3. Есть заведения с очень низкими оценками: Несмотря на общий положительный тренд, есть заведения, которые получают крайне низкие оценки. Это требует отдельного анализа, чтобы понять причины.
4. Тот факт, что 75% заведений имеют рейтинг 4.4 и ниже, может говорить о том, что сложно получить оценки выше этого уровня, или потребители редко ставят максимальные баллы.
5. Стандартное отклонение (std) равно 0.47, что говорит о том, что значения рейтинга достаточно кучно сгруппированы вокруг среднего.
6. 90 % оценок держатся в интервале от 3.4 - 4.8.


```
In [59]: # Создаю таблицу для среднего расчёта рейтинга по категориям
category_rating = places_data.groupby('category')['rating'].mean().sort_values()
```

```
In [60]: plt.figure(figsize=(9, 5))
ax = sns.barplot(x=category_rating.index, y=category_rating.values, palette="ch:

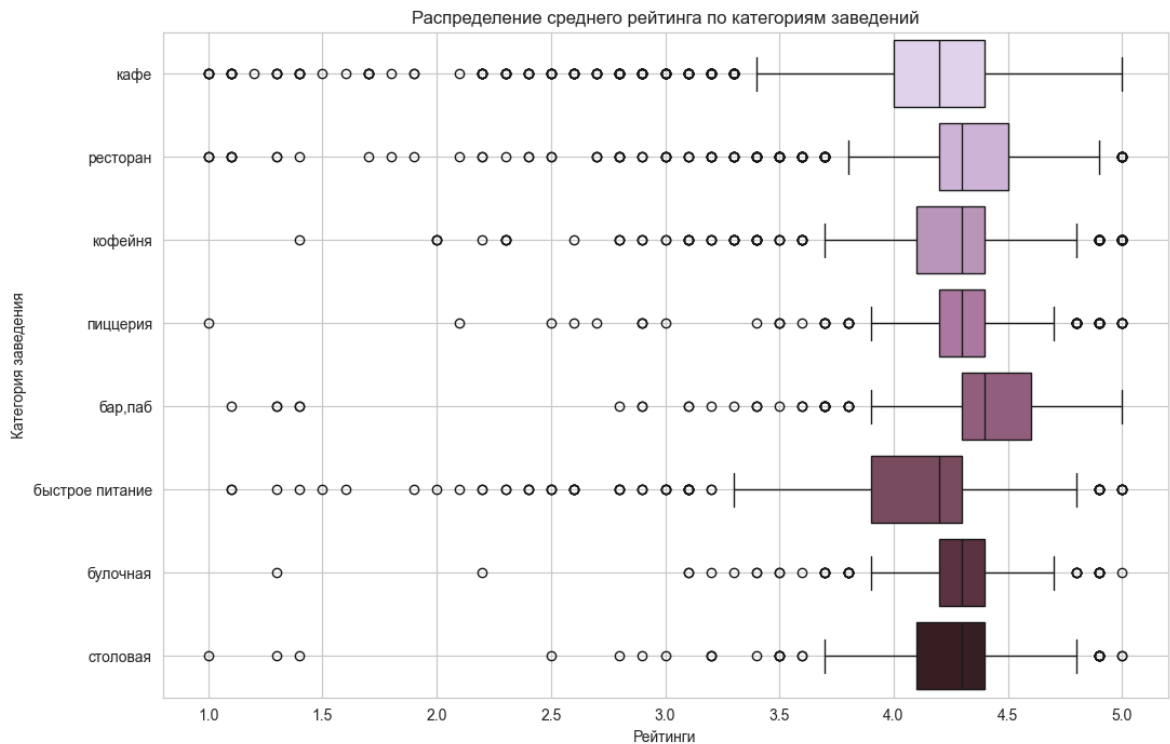
# Добавляем текст для каждого столбца
for p in ax.patches:
    height = p.get_height()
    ax.annotate(f'{height:.2f}', # Текст
                (p.get_x() + p.get_width() / 2., height),
                ha='center', va='bottom')
plt.xlabel('Категория заведения')
plt.ylabel('Рейтинг')
plt.title('Распределение среднего рейтинга по категориям заведений')
plt.grid(True)
plt.show()
```



Сильных отличий в распределении среднего рейтинга по категориям заведений найти не удалось, размах в 0.34 балла не так высок. Средняя оценка для всех категорий ниже 4 баллов не снижается.

```
In [61]: plt.figure(figsize=(12, 8))

sns.boxplot(x='rating', y='category', data=places_data, palette="ch:start=1,rot=
plt.xlabel('Рейтинги')
plt.ylabel('Категория заведения')
plt.title('Распределение среднего рейтинга по категориям заведений')
plt.grid(True)
plt.show()
```



In []:

Ящик с усами отображает распределение рейтинга для каждой категории заведений. Видно, что:

- Средние значения рейтинга у всех категорий находятся в диапазоне 4.0-4.5.
- Заведения фаст фуда имеют самые низкие средние значения рейтинга. И так же самый низкий нижний квартиль и достаточное большое число низких выбросов, даже при учёте не очень многочисленное представление данной категории в данных меньше 7 % от всех, что говорит и крайне низком сервисе, качестве обслуживания и предлагаемых блюд.
- Так же важно не забывать о количестве представленных заведений в каждой категории и делать категоричные выводы об оценках не стоит, так как в данных представленно, например в 10 раз меньше булочных, чем ресторанов, такие отличия в количественном выражение сказываются на результатах исследования средних рейтингов и количестве и размерах выбросов.
- Видим не равномерное распределение выбросов у пиццерий, баров и столовых, разрыв превышает 1 балл, это показывает, что в рамках этих категорий существуют заведения, которые либо сильно выбиваются из общей картины, либо не соответствуют ожиданиям основной массы клиентов.

In [62]: `category_rating = round(places_data.groupby('district', as_index=False)['rating'`

Приступаю к геоаналитике.

```

In [63]: with open('admin_level_geomap.geojson', 'r') as f:
          geo_json = json.load(f)
          # print(json.dumps(geo_json, indent=2, ensure_ascii=False, sort_keys=True))

In [64]: state_geo = 'admin_level_geomap.geojson'
          # moscow_lat - широта центра Москвы, moscow_lng - долгота центра Москвы
          moscow_lat, moscow_lng = 55.751244, 37.618423

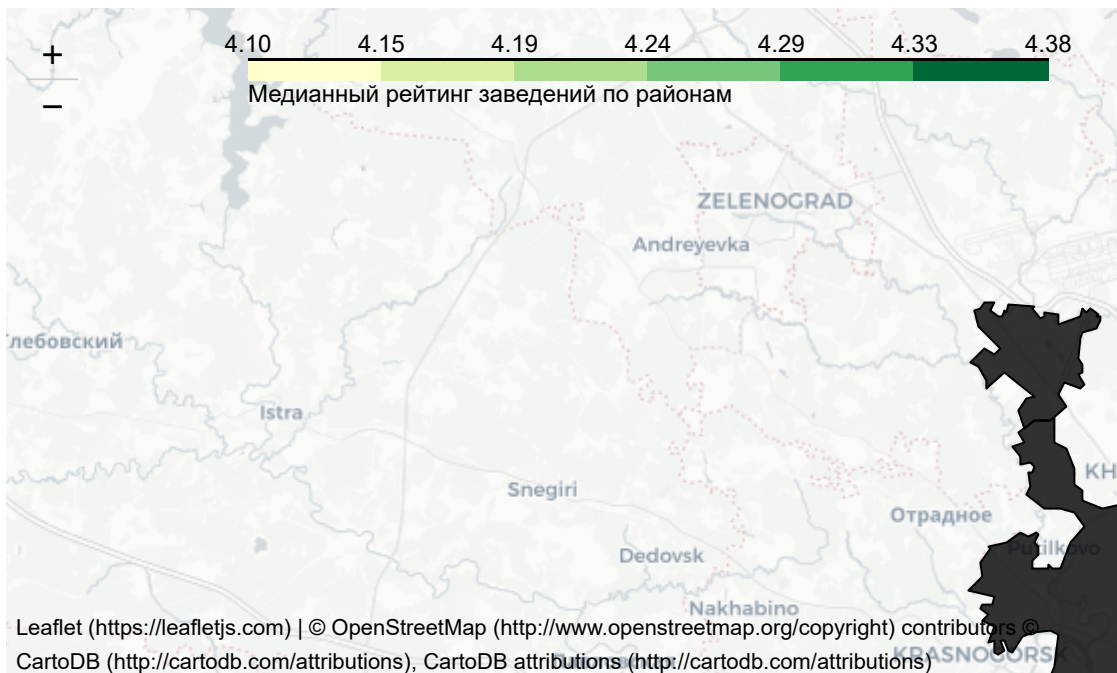
          # создаём карту Москвы
          m = Map(location=[moscow_lat, moscow_lng], zoom_start=10, tiles='Cartodb Positron')

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

          # выводим карту
          m

```

Out[64]:



Центральный административный округ имеет самый высокий средний рейтинг по всем заведениям. Юго-Восточный административный округ самые низкие оценки. Видим закономерное снижение рейтинга от Северного округа к противоположному к Юго-восточному

Геоаналитика заведений

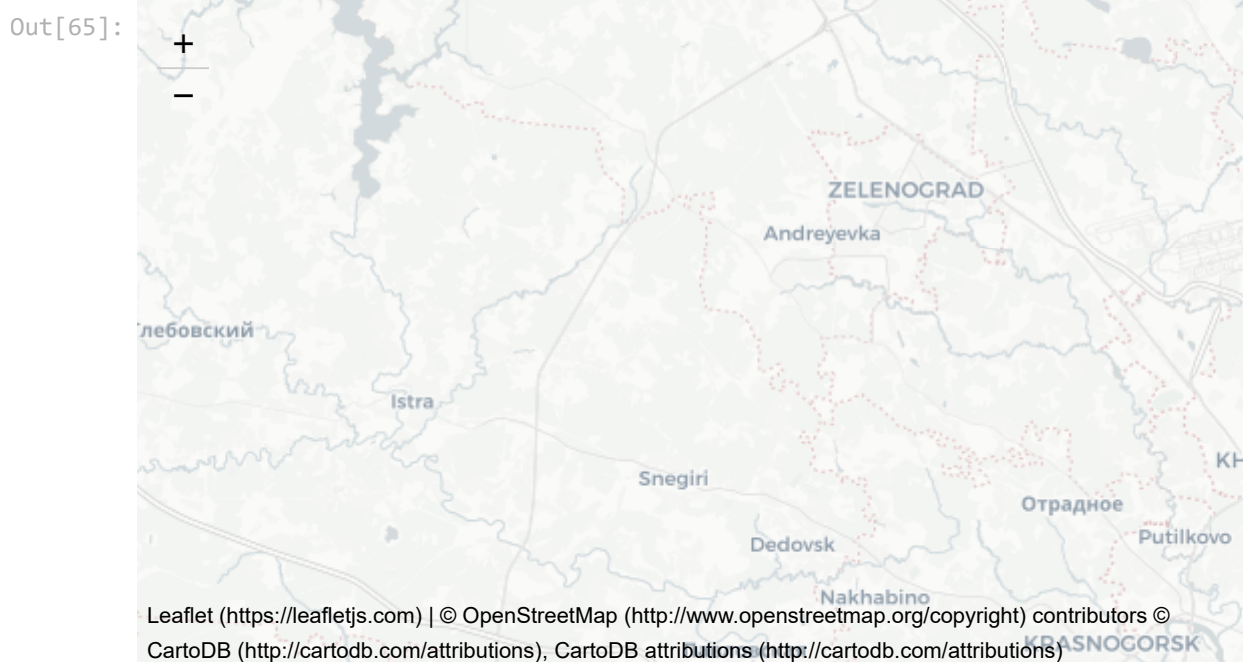
- Отображение заведений на карте с кластерами
- Топ-15 улиц по количеству заведений
- Визуализация

- Улицы с одним заведением
- Медиана среднего чека по районам
- Визуализация

```
In [65]: # создаём карту Москвы
m = Map(location=[moscow_lat, moscow_lng], zoom_start=10, tiles="Cartodb Positron")
# создаём пустой кластер, добавляем его на карту
marker_cluster = MarkerCluster().add_to(m)

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

# выводим карту
m
```



```
In [66]: top_15_streets = places_data.groupby('street', as_index=False)['name'].count().sort_values(ascending=False)
top_15_streets
```

Out[66]:

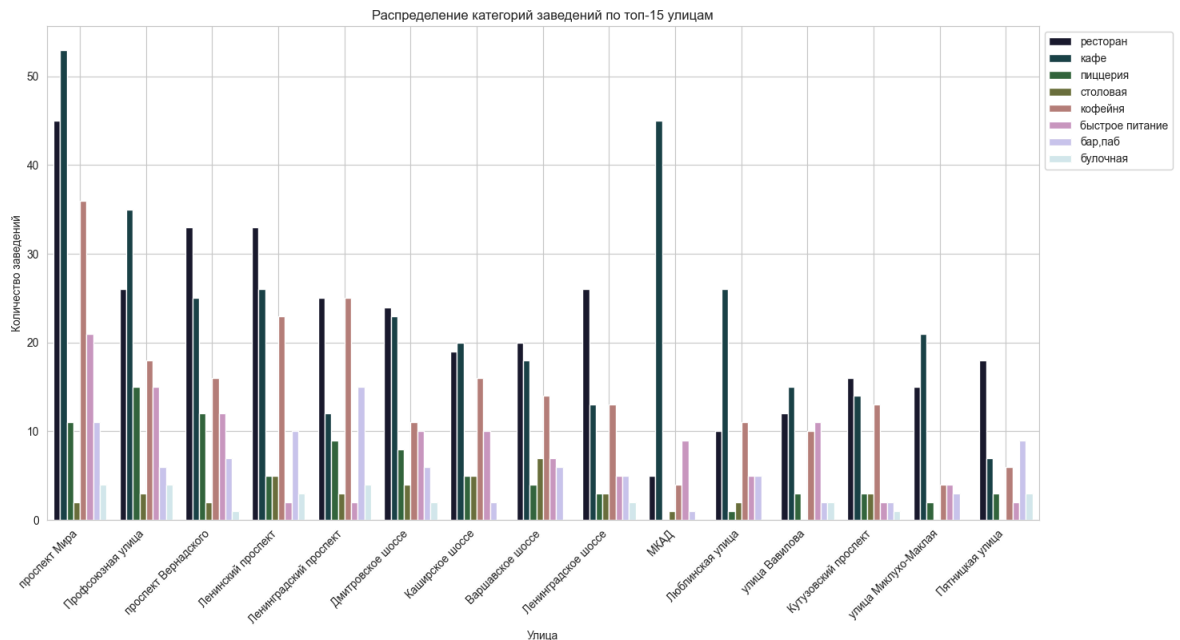
	street	name
1090	проспект Мира	183
773	Профсоюзная улица	122
1087	проспект Вернадского	108
525	Ленинский проспект	107
523	Ленинградский проспект	95
373	Дмитровское шоссе	88
455	Каширское шоссе	77
298	Варшавское шоссе	76
524	Ленинградское шоссе	70
550	МКАД	65
547	Люблинская улица	60
1154	улица Вавилова	55
517	Кутузовский проспект	54
1309	улица Миклухо-Маклая	49
781	Пятницкая улица	48

```
In [67]: df_top15 = places_data[places_data['street'].isin(top_15_streets['street'])]
```

```
In [68]: plt.figure(figsize=(16, 8))

sns.countplot(x='street',
              hue='category',
              data=df_top15,
              order=df_top15['street'].value_counts().index, palette="cubehelix")

plt.title('Распределение категорий заведений по топ-15 улицам')
plt.xticks(rotation=45, ha='right')
plt.ylabel('Количество заведений')
plt.xlabel('Улица')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.grid(True)
plt.show()
```



Самая большая улица по числу заведений это проспект Мира и там мы видим, что большинство заведений представлено в категориях кафе и рестораны, так же данные категории является лидерами почти во всех улицах топа.

In [69]: `# Найдите улицы, на которых находится только один объект общепита. Что можно ска`

In [70]: `street_one_object = places_data.groupby('street', as_index=False)['name'].count()
street_one_object = street_one_object[street_one_object['name'] == 1]
street_one_object = places_data[places_data['street'].isin(street_one_object['st
street_one_object.head(5)`

Out[70]:

	name	category	address	district	hours	lat
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
21	7/12	кафе	Москва, Прибрежный проезд, 7	Северный административный округ	ежедневно, 10:00–22:00	55.876805
25	В парке вкуснее	кофейня	Москва, парк Левобережный	Северный административный округ	ежедневно, 10:00–21:00	55.878453
58	Coffeekaldi's	кофейня	Москва, Угличская улица, 13, стр. 8	Северо-Восточный административный округ	ежедневно, 09:00–22:00	55.900316
60	Чебуречная история	кофейня	Москва, ландшафтный заказник Лианозовский	Северо-Восточный административный округ	ежедневно, 10:00–22:00	55.899845

In [71]: `street_one_object.groupby('district')['name'].size().sort_values()`

```
Out[71]: district
        Юго-Западный административный округ      18
        Северо-Западный административный округ   19
        Западный административный округ          35
        Юго-Восточный административный округ     39
        Южный административный округ             43
        Восточный административный округ         52
        Северный административный округ          53
        Северо-Восточный административный округ  55
        Центральный административный округ      145
        Name: name, dtype: int64
```

```
In [72]: print(street_one_object['category'].value_counts())
        print(street_one_object['rating'].describe())
```

```
category
кафе          160
ресторан       94
кофейня       84
бар, паб       39
столовая       36
быстрое питание 23
пиццерия       15
булочная       8
Name: count, dtype: int64
count      459.000000
mean        4.236383
std         0.461492
min         1.000000
25%         4.100000
50%         4.300000
75%         4.500000
max         5.000000
Name: rating, dtype: float64
```

- Преобладание популярных форматов: Кафе (160), рестораны (93) и кофейни (84) являются наиболее распространенными категориями заведений на таких улицах. За ними следуют бары/пабы и столовые (39 и 36 соответственно).
- Меньшая представленность фастфуда: Заведения быстрого питания (23), пиццерии (15) и булочные (8) встречаются значительно реже, что может указывать на другую целевую аудиторию и потребности.
- В целом высокие рейтинги: Средний рейтинг заведений составляет 4.24, что свидетельствует о положительном восприятии клиентами.
- Разброс оценок: Рейтинги варьируются от 1.0 до 5.0, но основная масса заведений (50%) имеет оценки в диапазоне от 4.1 до 4.5.

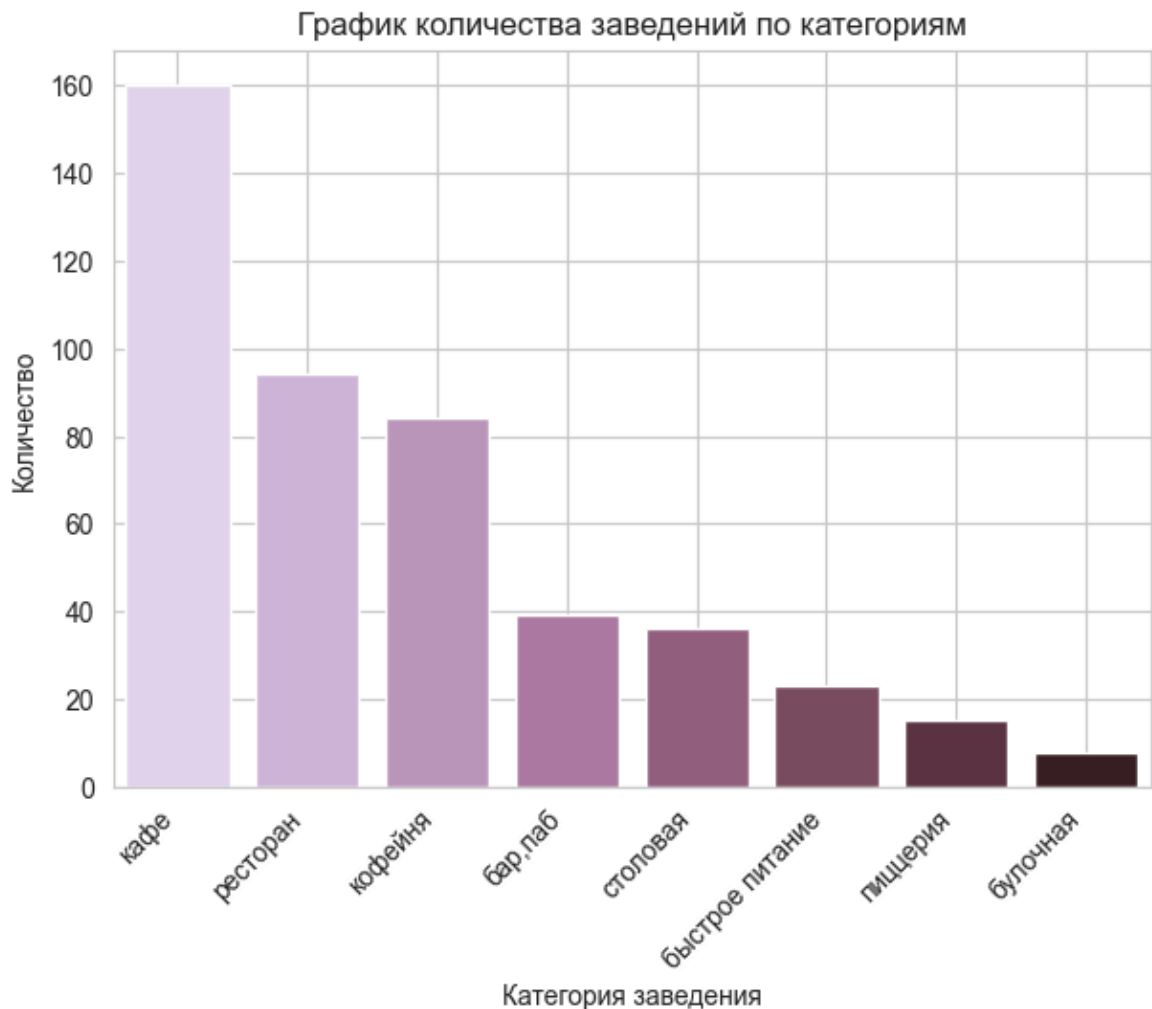
```
In [73]: print(f'Всего {round(len(street_one_object) / len(places_data) * 100, 1)} % заве
```

Всего 5.5 % заведений расположены в одиночестве на улице

```
In [74]: print(f'Их средний рейтинг равняется: {round(street_one_object["rating"].mean(),
```

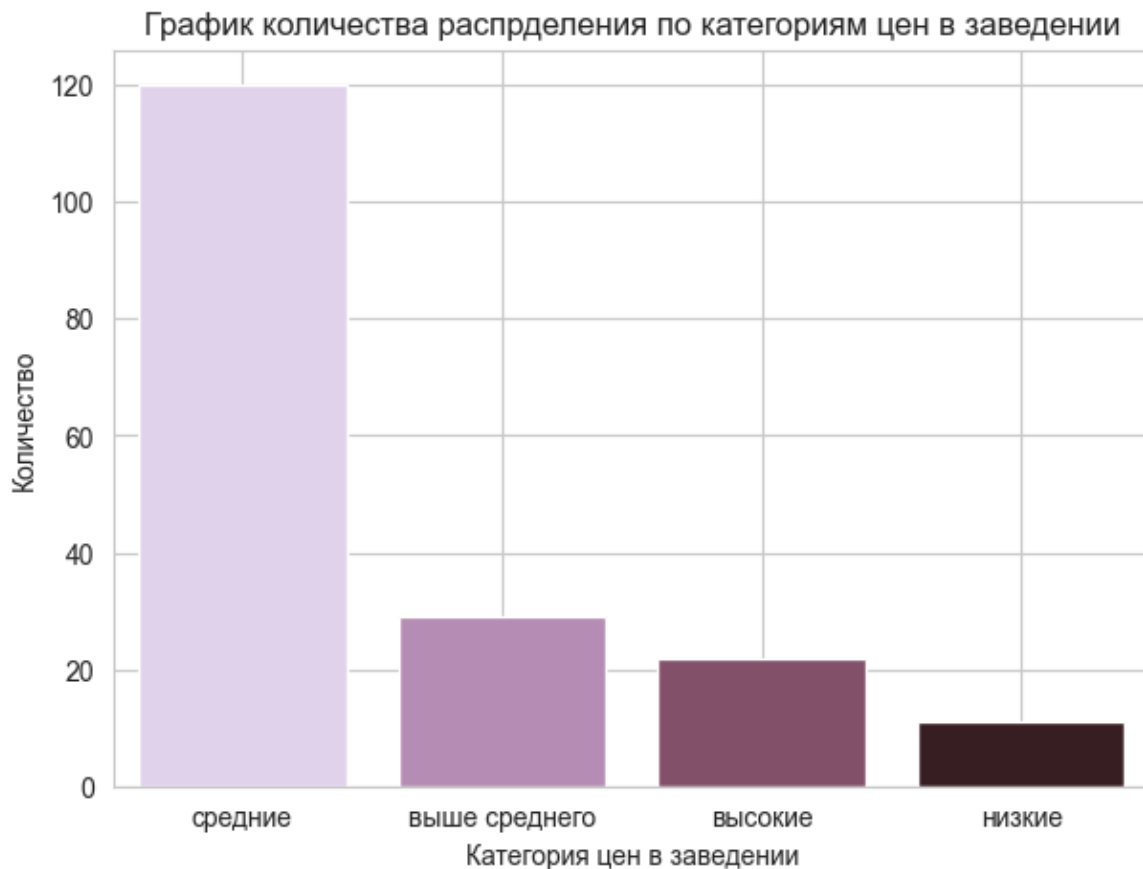
Их средний рейтинг равняется: 4.24

```
In [75]: plt.figure(figsize=(7, 5))
sns.barplot(x=street_one_object['category'].value_counts().index, y=street_one_o
plt.xlabel('Категория заведения')
plt.ylabel('Количество')
plt.title('График количества заведений по категориям')
plt.xticks(rotation=45, ha='right')
plt.grid(True)
plt.show()
```



В основном представлены кафе в минимальном количестве представлены булочные.

```
In [76]: plt.figure(figsize=(7, 5))
sns.barplot(x=street_one_object['price'].value_counts().index, y=street_one_obje
plt.xlabel('Категория цен в заведении')
plt.ylabel('Количество')
plt.title('График количества распрделения по категориям цен в заведении')
plt.grid(True)
plt.show()
```

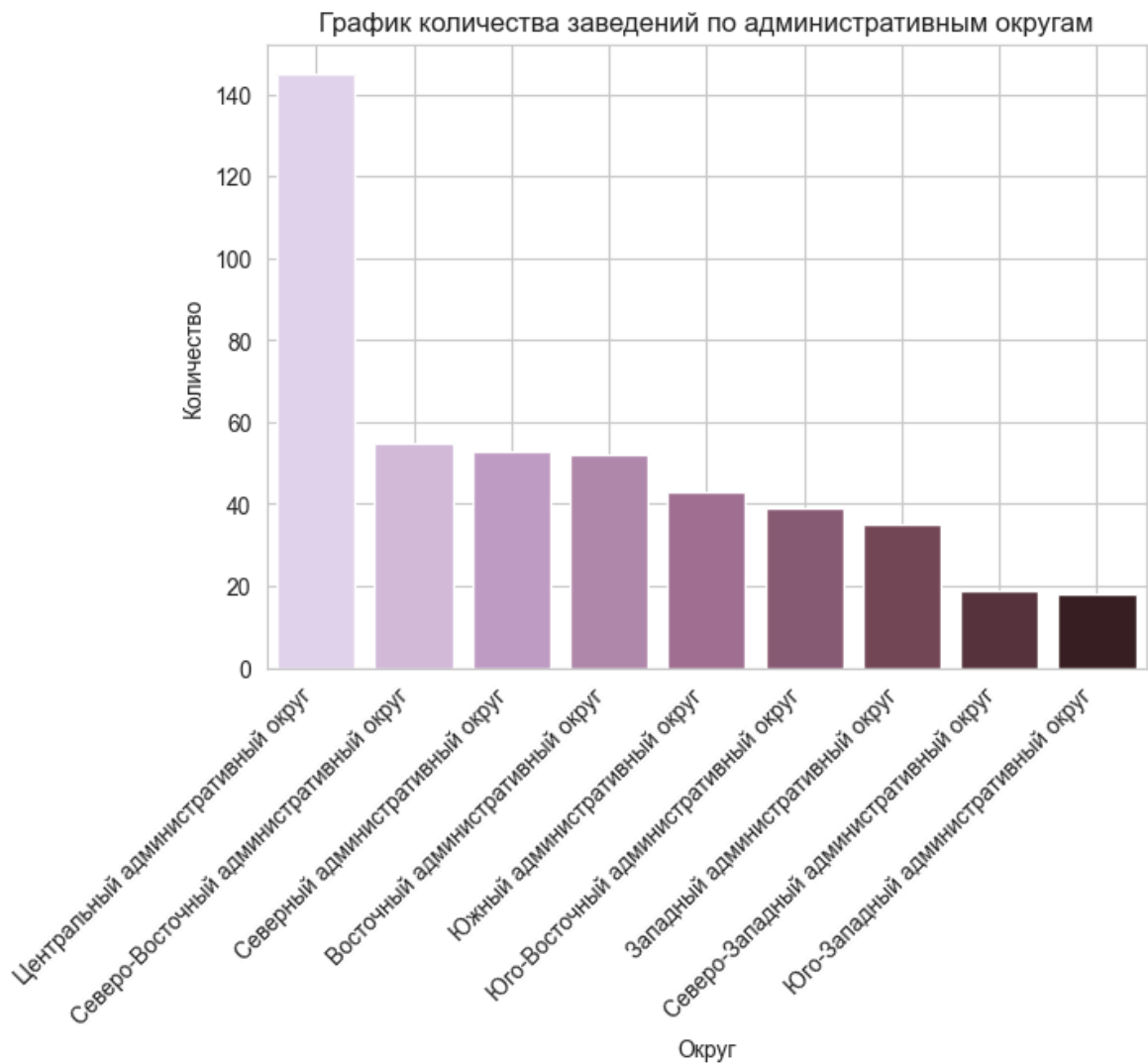



Явно преобладают заведения со средней категорией цен и можем увидеть, что заведений с низкими ценами меньшинство.

```
In [77]: street_one_object['chain'].value_counts()
```

```
Out[77]: chain
0      328
1      131
Name: count, dtype: int64
```

```
In [78]: plt.figure(figsize=(7, 5))
sns.barplot(x=street_one_object['district'].value_counts().index, y=street_one_o
plt.xlabel('Округ')
plt.ylabel('Количество')
plt.title('График количества заведений по административным округам')
plt.xticks(rotation=45, ha='right')
plt.grid(True)
plt.show()
```



Центральный округ явным лидер в данной категории заведений, Северо-западный и Юго-западный в меньшинстве.

```
In [79]: # Количество заведений работающих круглосуточно
round(street_one_object['is_24_7'].sum() / len(street_one_object) * 100, 2)
```

Out[79]: 6.75

```
In [80]: # Количество заведений работающих круглосуточно от всех данных
round(places_data['is_24_7'].sum() / len(places_data) * 100, 2)
```

Out[80]: 9.19

Несомненный лидер по количеству одиночных заведений общепита — Центральный административный округ. На его территории встречается 145 заведений, расположенных единственными на улице.

В выборке заведений на улицах с одним объектом общепита наиболее часто встречаются кафе (160) и рестораны (93). Это может говорить о том, что на таких улицах более востребованы места для проведения времени, нежели просто быстрого перекуса. Возможно, эти улицы менее ориентированы на транзитный трафик и больше — на локальных жителей. Что подтверждается малым числом заведений быстрого питания (23), пиццерий (15) и булочных (8). Это может

указывать на то, что на улицах с одним объектом общепита меньше спрос на быстрый перекус "на ходу", или что другие факторы (например, высокая арендная плата) делают открытие таких заведений менее привлекательным.

Сильных различий в данной выборке в процентном соотношении для заведений работающих круглосуточно выявить не удалось, на 2,5 % меньше заведений работает круглосуточно от всего числа заведений.

```
In [81]: district_middle_avg_bill = places_data.groupby('district', as_index=False)['midd
```

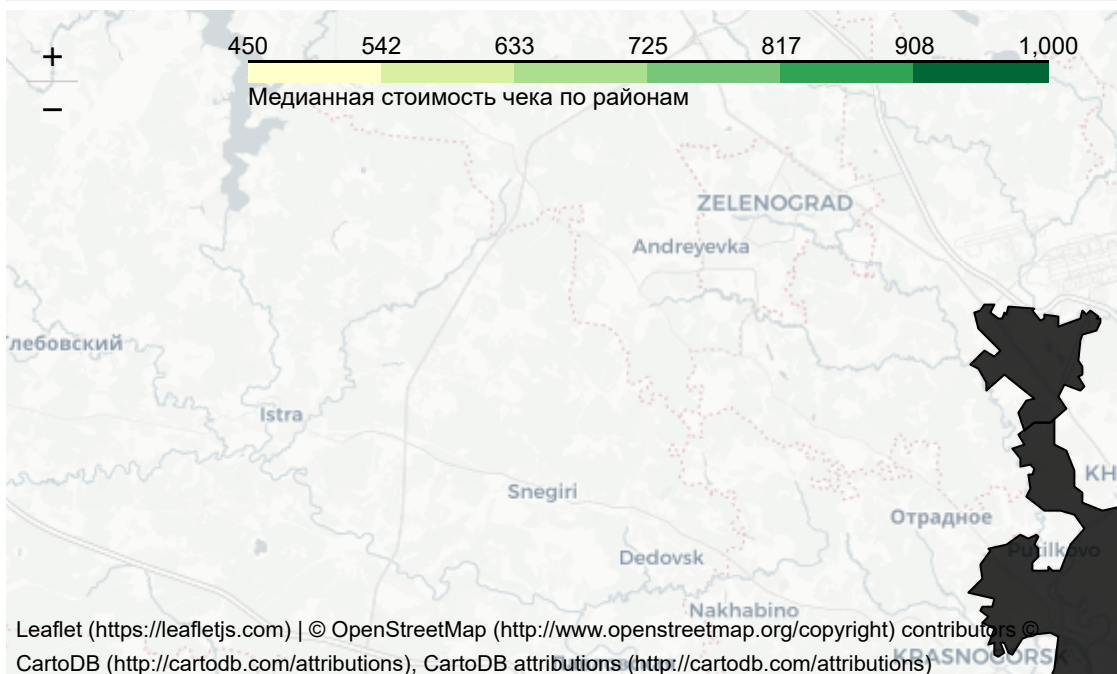
```
In [82]: # moscow_lat - широта центра Москвы, moscow_lng - долгота центра Москвы
moscow_lat, moscow_lng = 55.751244, 37.618423

# создаём карту Москвы
m = Map(location=[moscow_lat, moscow_lng], zoom_start=10, tiles='Cartodb Positro

# создаём хороплет с помощью конструктора Choropleth и добавляем его на карту
Choropleth(
    geo_data=state_geo,
    data=district_middle_avg_bill,
    columns=['district', 'middle_avg_bill'],
    key_on='feature.name',
    fill_color='YlGn',
    fill_opacity=0.8,
    legend_name='Медианная стоимость чека по районам',
).add_to(m)

# выводим карту
m
```

Out[82]:



```
In [83]: district_middle_avg_bill
```

Out[83]:

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

Можем видеть самый высокий средний чек в Западном и Центральном административный округа, Юго-Восточный административный округ имеет самый низкий чек. Я бы отметила, что высота среднего чека уменьшается по мере отдаления не от центра, а от запада города, исключение Восточный округ, но и он от ближайшего южного отличается на 75 р.

Высокий чек в западном округе можно объяснить несколькими факторами: - расположены одни из самых престижных и дорогих районов Москвы - находятся крупные бизнес-центры и офисные комплексы - расположен МГУ им. М.В. Ломоносова, а также ряд других престижных вузов

Дополнительный анализ

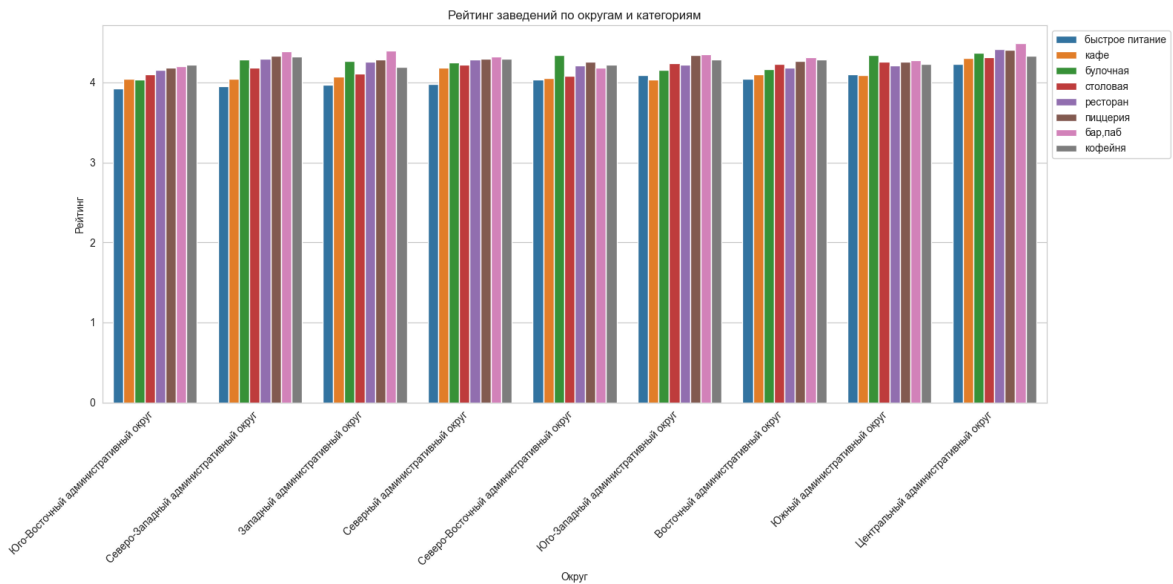
Проиллюстрируем другие взаимосвязи, которые мы нашли в данных. Исследуем часы работы заведений и их зависимость от расположения и категории заведения. Также можно исследовать особенности заведений с плохими рейтингами, средние чеки в таких местах и распределение по категориям заведений.

```
In [84]: with_category = places_data.groupby(['district', 'category']).agg(count=('name',
                                                                              round_cloak=('i
                                                                              rating=('ratin
                                                                              middle_avg_bil
                                                                              seats=('seats'
                                                                              cup_coffee=('m
                                                                              ).reset_index().s
with_category['seats'] = with_category['seats'].astype('Int64')
```

```
In [85]: # Создаем график
plt.figure(figsize=(16, 8))
sns.barplot(x='district',
            y='rating',
            hue='category',
            data=with_category)

plt.title('Рейтинг заведений по округам и категориям')
```

```
plt.xlabel('Округ')
plt.ylabel('Рейтинг')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Категория заведения')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.tight_layout()
plt.show()
```



```
In [86]: lowest = places_data[(places_data['rating'] < 3.4)]
highest = places_data[(places_data['rating'] > 4.8)]
```

В целом, рейтинги заведений довольно высокие (больше 3.5) и не сильно отличаются между округами и категориями, хотя есть небольшие колебания. Это может говорить о равномерном распределении качественных заведений по городу. Категория "Быстрое питание" имеет самый низкий рейтинг среди представленных. В Западном административном округе почти все категории показывают стабильно высокие результаты.

Рассмотрим заведения с крайними значениями рейтингов, самые низкие меньше 3.4 и самые высокие больше 4.8

```
In [87]: lowest.groupby('district')['name'].count().sort_values()
```

```
Out[87]: district
Центральный административный округ    17
Северо-Западный административный округ  20
Юго-Западный административный округ    32
Северный административный округ        33
Южный административный округ          36
Западный административный округ        47
Восточный административный округ       54
Северо-Восточный административный округ  59
Юго-Восточный административный округ    67
Name: name, dtype: int64
```

```
In [88]: highest.groupby('district')['name'].size().sort_values()
```

```
Out[88]: district
Северо-Западный административный округ    16
Юго-Западный административный округ        20
Западный административный округ            29
Северо-Восточный административный округ    30
Восточный административный округ           35
Южный административный округ               38
Юго-Восточный административный округ       46
Северный административный округ             48
Центральный административный округ         86
Name: name, dtype: int64
```

Юго-Восточный административный округ лидер по самым низким рейтингам у заведений

Центральный административный округ лидер по числу заведений с наивысшим рейтингом

```
In [89]: lowest.groupby('district')['middle_avg_bill'].mean().sort_values()
```

```
Out[89]: district
Юго-Западный административный округ    425.000000
Южный административный округ           485.000000
Центральный административный округ     490.000000
Юго-Восточный административный округ   527.500000
Восточный административный округ       613.333333
Северный административный округ        700.000000
Северо-Восточный административный округ 901.000000
Западный административный округ        925.000000
Северо-Западный административный округ 1440.000000
Name: middle_avg_bill, dtype: float64
```

```
In [90]: highest.groupby('district')['middle_avg_bill'].mean().sort_values()
```

```
Out[90]: district
Юго-Восточный административный округ    212.500000
Северо-Западный административный округ   341.666667
Северо-Восточный административный округ  462.500000
Северный административный округ          606.250000
Южный административный округ             643.750000
Восточный административный округ        1050.000000
Западный административный округ          1050.000000
Юго-Западный административный округ     1200.000000
Центральный административный округ      1819.545455
Name: middle_avg_bill, dtype: float64
```

Северо-Западный административный округ в данном анализе видим очень интересный результат для заведений с самыми низкими рейтингами в Северо-Западном административном округе выходит очень высокий средний чек 1440, для рейтинга заведений с наивысшим рейтингом эта сумма была на втором месте после лидера Центрального административного округа с размером среднего чека 1819.5

```
In [91]: lowest.groupby('district')['is_24_7'].sum().sort_values()
```

```
Out[91]: district
Центральный административный округ      1
Северный административный округ          4
Северо-Западный административный округ    4
Юго-Западный административный округ       4
Западный административный округ           6
Северо-Восточный административный округ    7
Южный административный округ              8
Восточный административный округ          11
Юго-Восточный административный округ      12
Name: is_24_7, dtype: object
```

```
In [92]: highest.groupby('district')['is_24_7'].sum().sort_values()
```

```
Out[92]: district
Северо-Западный административный округ    0
Юго-Западный административный округ       0
Западный административный округ           1
Северный административный округ           1
Северо-Восточный административный округ    1
Юго-Восточный административный округ       3
Южный административный округ              3
Восточный административный округ           4
Центральный административный округ         4
Name: is_24_7, dtype: object
```

По количеству заведений работающих круглосуточно в рейтинге с самыми низкими оенками побеждает Юго-Восточный административный округ Центральный административный округ становится лидером и в данном рейтинге.

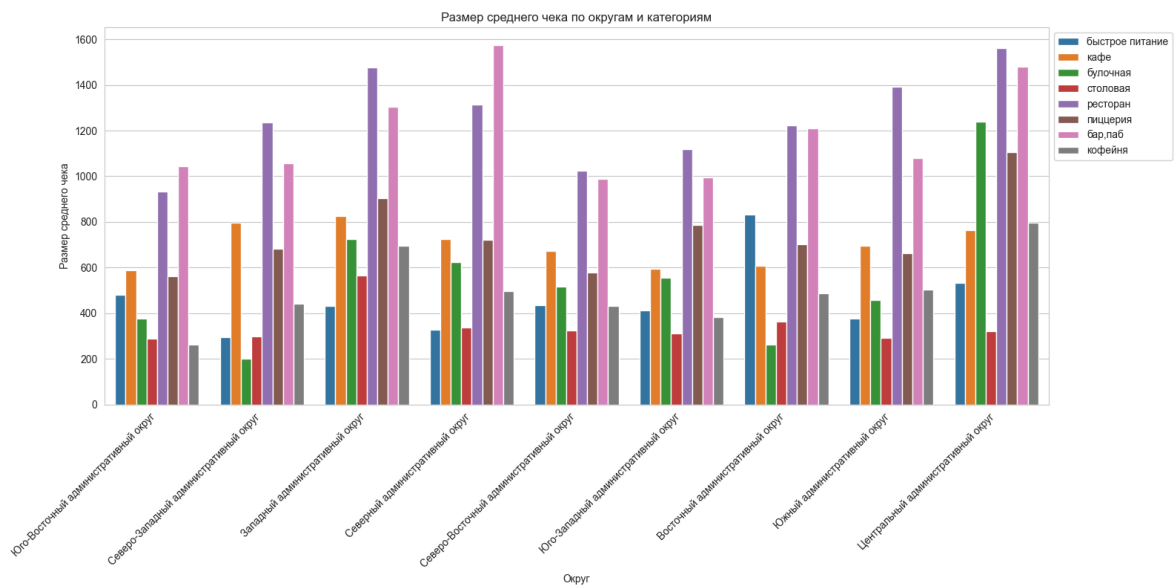
```
In [93]: places_data['rating'].corr(places_data['middle_avg_bill'])
```

```
Out[93]: np.float64(0.1832379714762454)
```

Между показателями рейтинга и среднего чека выявили очень слабую корреляцию. Невозможно сказать, что рейтинг зависит от размера среднего чека.

```
In [94]: # Создаем график
plt.figure(figsize=(16, 8))
sns.barplot(x='district',
            y='middle_avg_bill',
            hue='category',
            data=with_category)

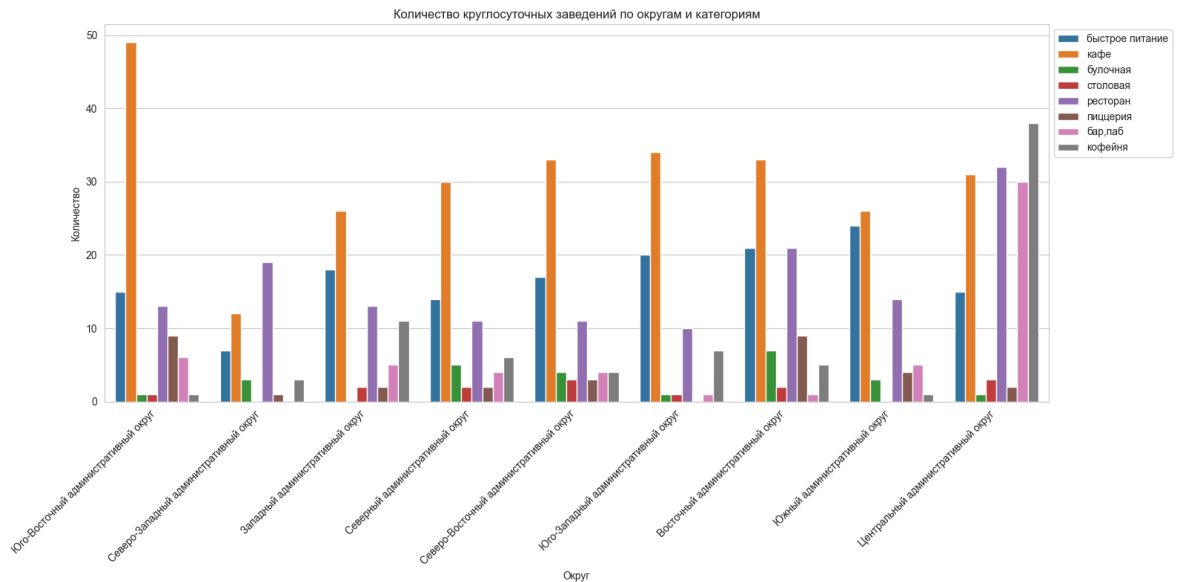
plt.title('Размер среднего чека по округам и категориям')
plt.xlabel('Округ')
plt.ylabel('Размер среднего чека')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Категория заведения')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.tight_layout()
plt.show()
```



В большинстве округов, средний чек в ресторанах и барах/пабах значительно выше, чем в других категориях. Эти заведения, как правило, предлагают более широкий ассортимент блюд и напитков, а также более высокий уровень сервиса, что отражается на стоимости посещения. Столовые демонстрируют самый низкий средний чек в каждом из округов. Это также соответствует ожиданиям, поскольку столовые ориентированы на предоставление доступного питания. В Западном и Центральном административных округах средний чек в ресторанах заметно выше, чем в других округах. Это может быть связано с особенностями аудитории в этих районах или с более высоким уровнем заведений, представленных там, это так же подтверждается на графике с рейтингами заведений.

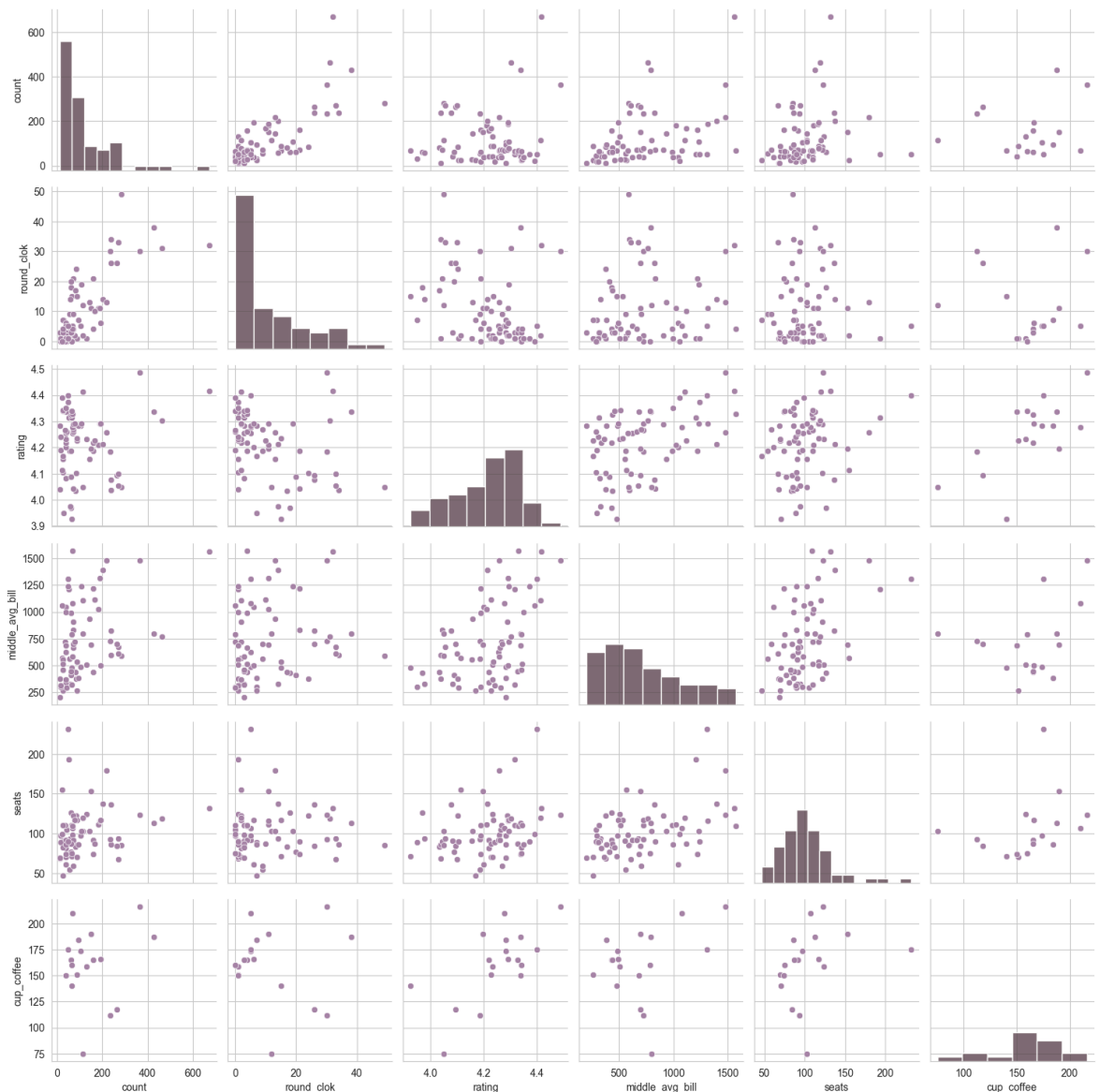
```
In [95]: # Создаем график
plt.figure(figsize=(16, 8))
sns.barplot(x='district',
            y='round_clok',
            hue='category',
            data=with_category)

plt.title('Количество круглосуточных заведений по округам и категориям')
plt.xlabel('Округ')
plt.ylabel('Количество')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Категория заведения')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.tight_layout()
plt.show()
```

Кафе безусловно лидирует по количеству круглосуточных заведений в большинстве округов. Особенно выделяется Юго-Восточный округ, где количество круглосуточных кафе значительно превосходит другие категории и округа. За ним следует "быстрое питание", хотя в некоторых округах количество круглосуточных кафе и заведений быстрого питания примерно одинаково. В Центральном административном округе можно заметить, что в лидеры выходят заведения категории "кофейни". Это может быть связано с новыми трендами, популярностью и доступностью в ценовой категории у молодёжи, которая предпочитает проводить время в центральной части города, а также с возможностью работать из этих заведений.

```
In [96]: plot = {'color': '#a883a8'}
scat = {'color': '#523c47'}
sns.pairplot(with_category, plot_kws=plot, diag_kind="hist", diag_kws=scat)
plt.tight_layout()
plt.show()
```



- Большинство значений количества заведений сосредоточено в диапазоне низких чисел, есть небольшое количество заведений с очень большим количеством заведений.
- Большинство заведений работают не круглосуточно.
- Рейтинги преимущественно высокие, около 4.2.
- Средний чек в основном находится в диапазоне 250-750.
- Заметна некоторая положительная связь между рейтингом и средним чеком, а также между количеством заведений и режимом работы, но это просто объяснить тем, что чем больше заведений всего, тем больше всего из них круглосуточно.

Общий вывод по всем вопросам

В данных представлены такие заведения категорий:

- кафе
- ресторан

- кофейня
- бар, паб
- пиццерия
- быстрое
- столовая
- булочная

Наиболее популярными (встречаются чаще остальных) в данных это кафе 28,2 % заведений и ресторан 23,4 % заведений.

Среди представленных данных кафе среди кафе больше всего представлено сетевых заведений 31,5 %, но по отношению внутри категории лидером становится заведение булочные примерно 61 % заведений.

Среди топа заведений самые частовстречаемые это кофейни на них приходится 30 % от всего топа. В топ так же попали такие категории как кафе, ресторан, пиццерия и заведения быстрого питания.

Самое популярное заведение это Кафе из категории кафе не являющегося сетевым, всего в топ попали 3 заведения не являющимися сетевыми остальные 80 % это сетевые заведения.

В датасете представлено 9 административных районов

- Центральный административный округ
- Северный административный округ
- Южный административный округ
- Северо-Восточный административный округ
- Западный административный округ
- Восточный административный округ
- Юго-Восточный административный округ
- Юго-Западный административный округ
- Северо-Западный административный округ

Сильных различий в средних рейтингах заведений по категориям типа общепита выявить не удалось. Размах не превышает 0,34 балла.

Центральный административный округ лидирует по количеству (145) одиночных заведений общепита. Среди них преобладают кафе (160) и рестораны (93), что говорит о востребованности мест для времяпрепровождения, а не быстрого перекуса. Небольшое число заведений быстрого питания (23), пиццерий (15) и булочных (8) подтверждает меньший спрос на "еду на ходу" или влияние других факторов, таких как аренда.

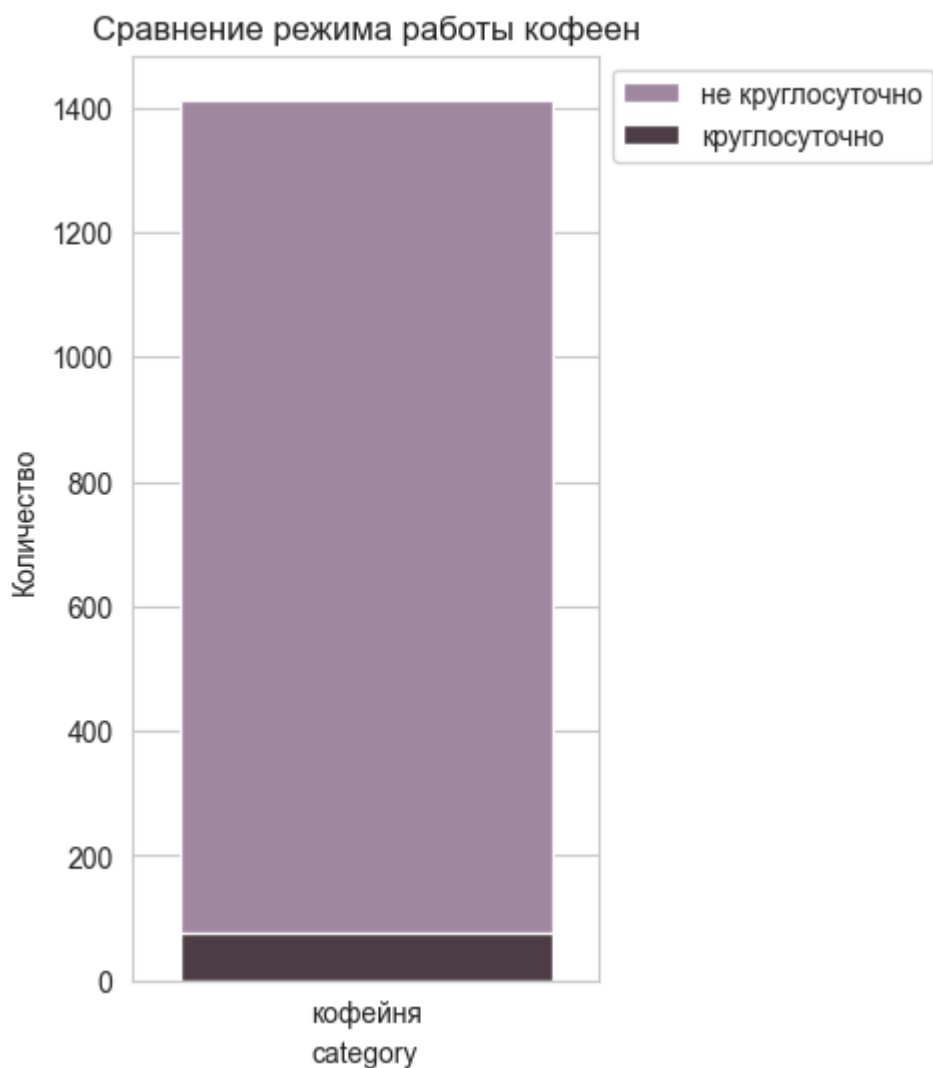
Самый высокий средний чек - в Западном и Центральном округах. Самый низкий - в Юго-Восточном округе. Чек снижается по мере отдаления от запада, с исключением в Восточном округе (но незначительно). Высокий чек в Западном округе объясняется престижными районами, бизнес-центрами и престижными вузами.

Центральный район в своем составе имеет самые высококлассные заведения с наивысшими оценками и самыми высокими средними чеками по городу. В данном округе высокая конкуренция и высокие доходы поэтому необходимо обладать высоким сервисом, высоким уровнем обслуживания.

Детализация исследования

```
In [97]: coffee_shop_select = places_data[places_data['category'] == 'кофейня'].groupby('
        round_clok=('is_24_7', 'sum'),
        rating=('rating', 'mean'),
        middle_avg_bill=('middle_avg_bill', 'mean')
        ).reset_index().sort_values(by='rating')
```

```
In [98]: plt.figure(figsize=(3, 6))
sns.barplot(x='category', y='count', label='не круглосуточно', data=coffee_shop_)
sns.barplot(x='category', y='round_clok', data=coffee_shop_select, label='кругло
plt.legend(title='Категория заведения')
plt.legend(loc='upper left', bbox_to_anchor=(1, 1))
plt.ylabel('Количество')
plt.title('Сравнение режима работы кофейен')
plt.show()
```



```
In [99]: category_coffee = with_category[with_category['category'] == 'кофейня']
```

```
In [100... print(f'В данных представлено всего: {coffee_shop_select["count"].values[0]} коф
```

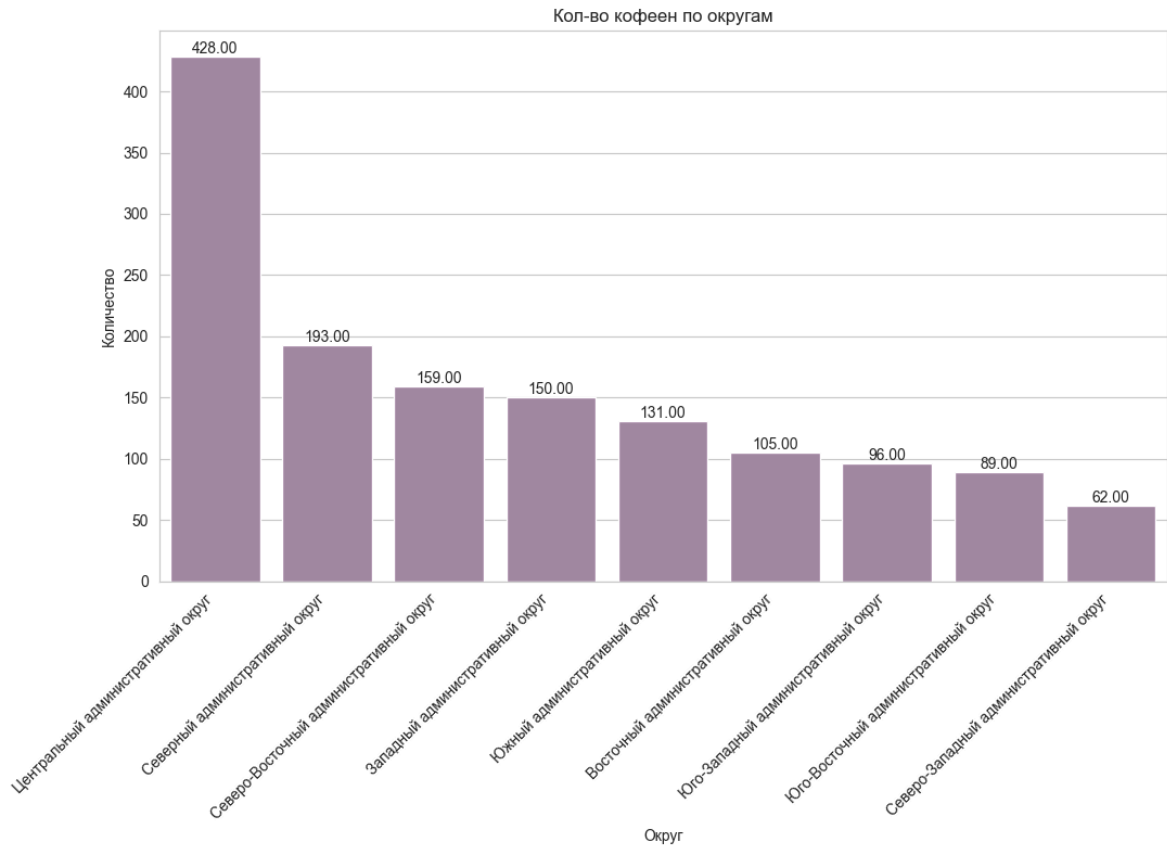
В данных представлено всего: 1413 кофеен

Из них работает круглосуточно: 76

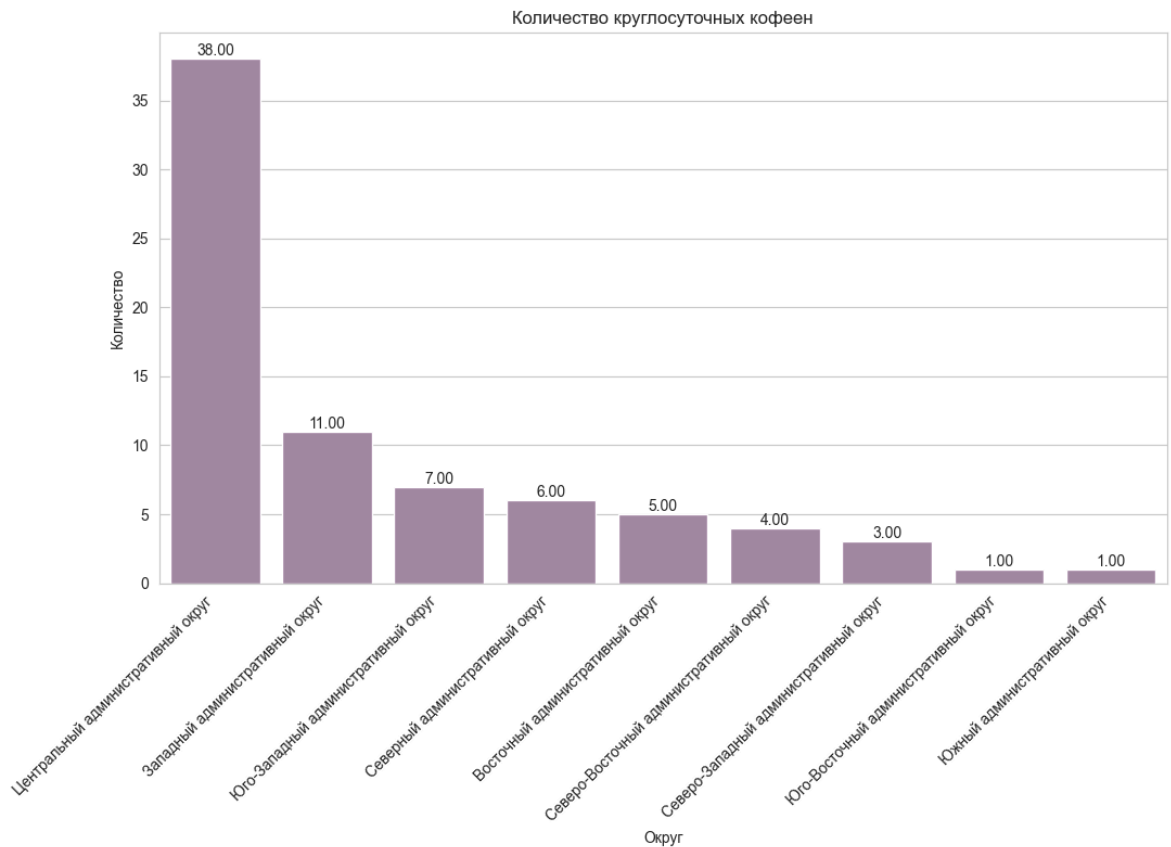
В среднем рейтинг у кофеен по всему городу держится на уровне 4.28

90 % оценок находятся в диапазоне [3.7 4.9]

```
In [101... param_coffee_of_district(['district', 'count'], category_coffee, 'Кол-во кофеен
```

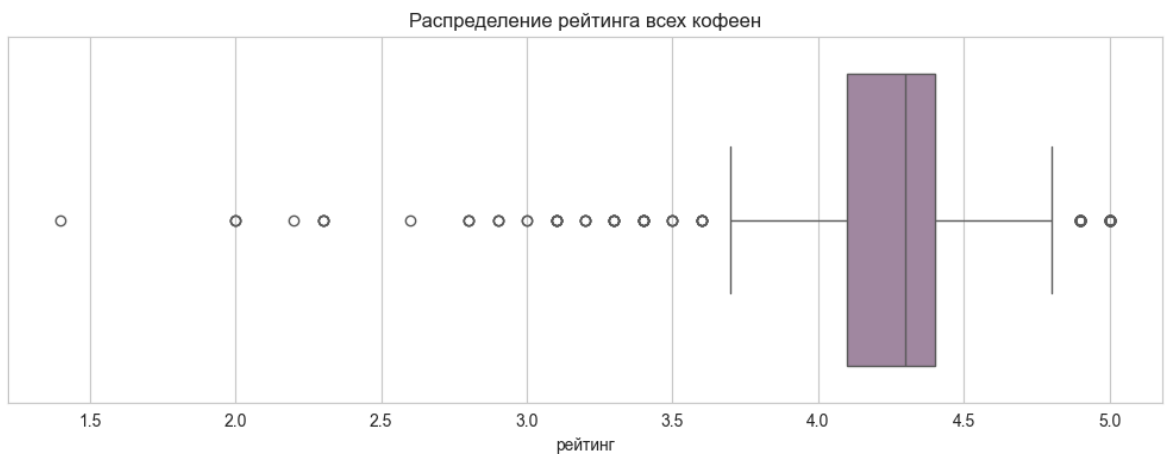


```
In [102... param_coffee_of_district(['district', 'round_clok'], category_coffee, 'Количеств
```



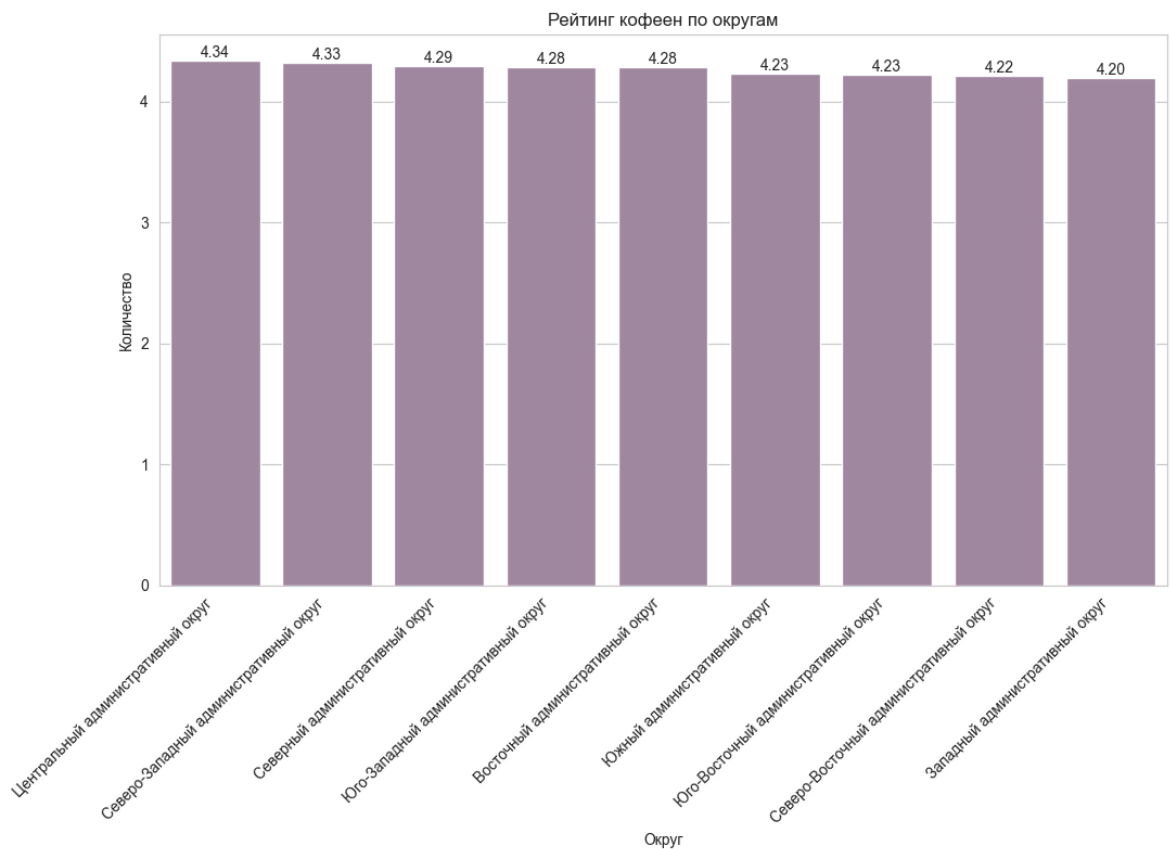
In [103...

```
plt.figure(figsize=(10, 4))
sns.boxplot(x=places_data[places_data['category'] == 'кофейня']['rating'], color=
plt.title('Распределение рейтинга всех кофеен')
plt.xlabel('рейтинг')
plt.tight_layout()
plt.show()
```

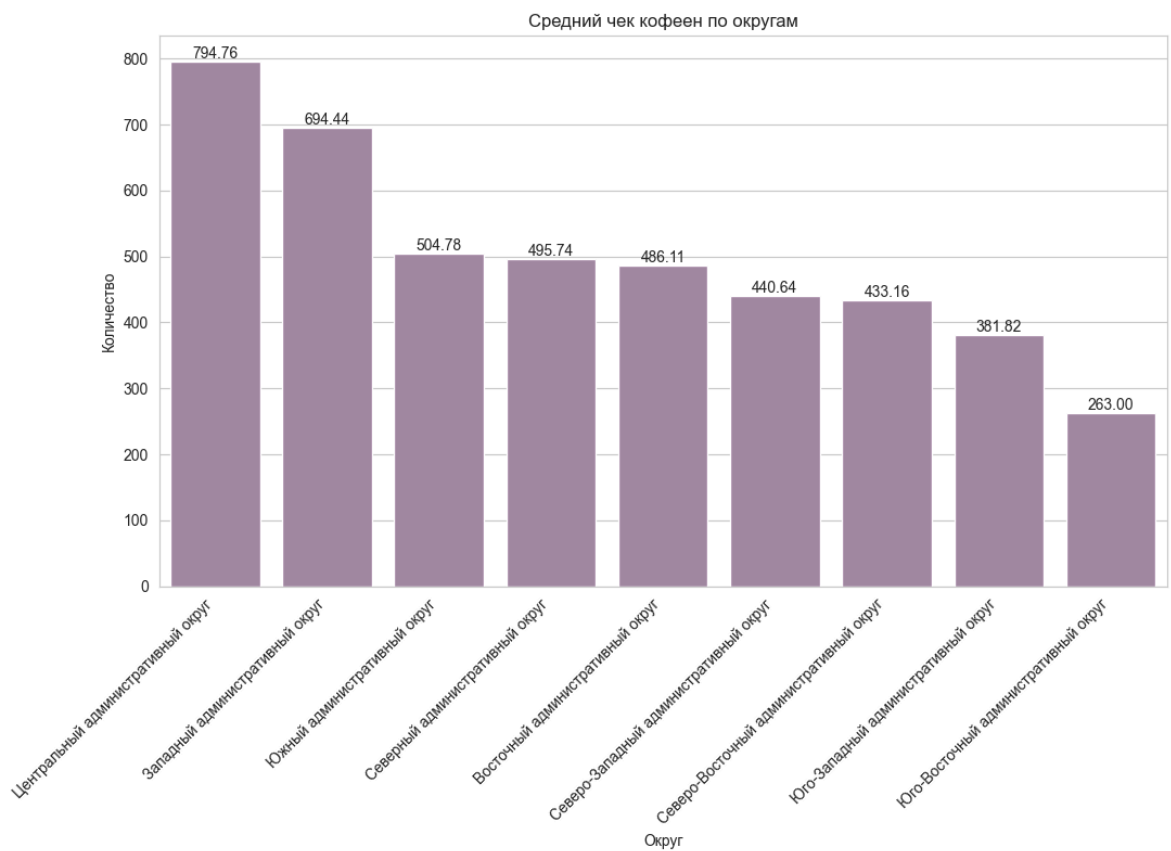


In [104...

```
param_coffee_of_district(['district', 'rating'], category_coffee, 'Рейтинг кофе
```



In [105... `param_coffee_of_district(['district', 'middle_avg_bill'], category_coffee, 'Сред`



In [106... `lowest_rating = places_data[(places_data['category'] == 'кофейня') & (places_dat`

In [107... `lowest_rating`

Out[107...

	district	name
4	Северо-Западный административный округ	1
7	Юго-Западный административный округ	4
5	Центральный административный округ	5
0	Восточный административный округ	6
1	Западный административный округ	6
2	Северный административный округ	8
8	Южный административный округ	8
3	Северо-Восточный административный округ	9
6	Юго-Восточный административный округ	9

In [108...

```
highest_rating = places_data[(places_data['category'] == 'кофейня') & (places_da
```

In [109...

```
highest_rating
```

Out[109...

	district	name
3	Северо-Восточный административный округ	1
7	Юго-Западный административный округ	1
8	Южный административный округ	2
1	Западный административный округ	3
6	Юго-Восточный административный округ	4
4	Северо-Западный административный округ	4
0	Восточный административный округ	4
5	Центральный административный округ	5
2	Северный административный округ	13

В Северном административном округе находится самое большое количество кофеен с наивысшим рейтингом, но там так же достаточно кофеен с очень низким рейтингом. Северо-Восточном и Юго-Восточном административном округе наибольшее число с самым низким рейтингом

In [110...

```
category_coffee[['district', 'seats']].sort_values('seats')
```


Out[110...

	district	seats
52	Юго-Восточный административный округ	70
60	Юго-Западный административный округ	86
36	Северо-Западный административный округ	87
28	Северо-Восточный административный округ	92
4	Восточный административный округ	97
44	Центральный административный округ	113
20	Северный административный округ	117
68	Южный административный округ	124
12	Западный административный округ	153

В целом, прослеживается тенденция к увеличению количества мест в кофейнях, если двигаться от Юго-Восточного округа к Западному.

In [111...

```
category_coffee[['district', 'cup_coffee']].sort_values('cup_coffee')
```

Out[111...

	district	cup_coffee
52	Юго-Восточный административный округ	151.088235
68	Южный административный округ	158.488372
28	Северо-Восточный административный округ	165.333333
36	Северо-Западный административный округ	165.523810
20	Северный административный округ	165.789474
4	Восточный административный округ	174.023810
60	Юго-Западный административный округ	184.176471
44	Центральный административный округ	187.518519
12	Западный административный округ	189.938776

Подобно предыдущему анализу, наблюдается общая тенденция к увеличению стоимости кофе по мере продвижения от Юго-Восточного округа к Западному.

С целью привлечения и удержания целевой аудитории, готовой платить за качественный кофейный напиток и соответствующий сервис в Северо-Западный административный округ, необходимо провести анализ цен на кофе в заведениях аналогичного класса. Заниженная стоимость может привлечь клиентов, не являющихся нашей целевой аудиторией и не способных оценить премиальность предлагаемого продукта.

In [112...

```
middle_coffee_cup = places_data[(places_data['category']=='кофейня') & (places_d
```

In [113...

```
print('Медианная стоимость средней чашки кофе', middle_coffee_cup.median())
```

Медианная стоимость средней чашки кофе 182.5

```
In [114... print('Средняя стоимость средней чашки кофе', middle_coffee_cup.mean())
```

Средняя стоимость средней чашки кофе 187.5

Ответы на вопросы заказчика

Сколько всего кофеен в датасете? В каких районах их больше всего, каковы особенности их расположения?

В датасете представлено 1413 кофеен. Наибольшая концентрация кофеен наблюдается в Центральном административном округе (ЦАО), где сосредоточено около 30% всех кофеен города.

Есть ли круглосуточные кофейни?

Да, в датасете присутствуют круглосуточные кофейни, однако они составляют лишь небольшую долю - около 5.5% от общего числа кофеен в Москве. Примечательно, что половина (50%) этих круглосуточных заведений расположена в Центральном административном округе.

Какие у кофеен рейтинги? Как они распределяются по районам?

Большинство кофеен имеют рейтинг в диапазоне от 3.7 до 4.9. В целом, кофейни с самыми высокими рейтингами преимущественно расположены в Центральном административном округе, а кофейни с самыми низкими рейтингами - в Западном административном округе. Если рассматривать концентрацию заведений с крайними значениями рейтингов, то в Северном административном округе наблюдается наибольшая концентрация кофеен с максимальными оценками, в то время как в Северо-Восточном и Юго-Восточном административных округах отмечается наибольшее количество кофеен с рейтингом ниже 3.7.

На какую стоимость чашки капучино стоит ориентироваться при открытии и почему?

При определении стоимости чашки капучино рекомендуется ориентироваться на среднюю стоимость в районе, где планируется открытие кофейни. Согласно данным, цены на капучино в кофейнях Москвы варьируются в диапазоне от 150 до 190 рублей, поэтому целесообразно установить цену, конкурентоспособную в выбранной локации. Дальнейшая корректировка цены может зависеть от рейтинга заведения, качества обслуживания и уникальности предлагаемого продукта.

Выводы и рекомендации заказчику

Рекомендация: Открытие кофейни в Северо-Западном административном округе.

Обоснование:

- Ненасыщенный рынок кофеен: В Северо-Западном округе наблюдается относительно низкая концентрация кофеен, что представляет собой возможность для захвата доли рынка.
- Высокая платёжеспособность: Анализ других категорий заведений в округе показывает высокий средний чек и хорошие рейтинги. Это свидетельствует о платёжеспособной аудитории, готовой тратить деньги на качественный сервис и продукты. Открытие кофейни в этом округе может быть успешным, если предложить качественный продукт по конкурентоспособной цене.
- Привлечение аудитории из соседних округов: Близость Центрального и Западного округов, известных своей развитой кофейной культурой, позволит привлечь новую аудиторию. Любители кофе из этих районов могут заинтересоваться новым заведением с уникальной концепцией в Северо-Западном округе.
- Рекомендованная ценовая политика: На начальном этапе рекомендуется установить цену на чашку кофе в пределах среднего значения по району для заведений с наивысшим рейтингом (около 180-190 рублей). По мере роста рейтинга и завоевания лояльности клиентов, стоимость можно будет постепенно увеличивать.

Ключевые преимущества:

- Возможность занять нишу на рынке с низким уровнем конкуренции.
- Привлечение платёжеспособной аудитории.
- Потенциал для расширения клиентской базы за счет жителей соседних районов.
- Гибкая ценовая политика, позволяющая адаптироваться к рыночным условиям.

Презентация

[Презентация. ЯНДЕКС-ДИСК.](#)