

Описание проекта

Вы решили открыть небольшое кафе в Москве. Оно оригинальное — гостей должны обслуживать роботы. Проект многообещающий, но дорогой. Вместе с партнёрами вы решились обратиться к инвесторам. Их интересует текущее положение дел на рынке — сможете ли вы снискать популярность на долгое время, когда все зеваки насмотрятся на роботов-официантов?

Вы — гуру аналитики, и партнёры просят вас подготовить исследование рынка. У вас есть открытые данные о заведениях общественного питания в Москве.

Составим небольшой план по работе:

- Загрузим данные и ознакомимся с ними
 - Выполним предобработку, если это понадобится
 - Проведем анализ данных
 - Дадим рекомендации исходя из проведенного анализа
 - Составим презентацию
-

Знакомство с данными

Таблица `rest_data`:

- `id` — идентификатор объекта;
- `object_name` — название объекта общественного питания;
- `chain` — сетевой ресторан;
- `object_type` — тип объекта общественного питания;
- `address` — адрес;
- `number` — количество посадочных мест.

```
In [1]: # Импортируем необходимые библиотеки
import pandas as pd
import seaborn as sns
import matplotlib as plt
import plotly.express as px
from plotly import graph_objects as go
import re
```

```
In [2]: pd.set_option('display.max_colwidth', None)
```

```
In [3]: # Чтения файла
try:
    df = pd.read_csv('rest_data.csv')
except:
    df = pd.read_csv('https://code.s3.yandex.net/datasets/rest_data.csv')
```

```
In [4]: df.head()
```

Out [4]:

| | id | object_name | chain | object_type | address | number |
|---|--------|------------------------|-------|-------------|---|--------|
| 0 | 151635 | СМЕТАНА | нет | кафе | город Москва, улица Егора Абакумова, дом 9 | 48 |
| 1 | 77874 | Родник | нет | кафе | город Москва, улица Талалихина, дом 2/1, корпус 1 | 35 |
| 2 | 24309 | Кафе «Академия» | нет | кафе | город Москва, Абельмановская улица, дом 6 | 95 |
| 3 | 21894 | ПИЦЦЕТОРИЯ | да | кафе | город Москва, Абрамцевская улица, дом 1 | 40 |
| 4 | 119365 | Кафе «Вишневая метель» | нет | кафе | город Москва, Абрамцевская улица, дом 9, корпус 1 | 50 |

In [5]:

```
# Используем функцию для ознакомления с данными
def df_info(data):
    print(data.info(), '\n\n')
    print('Кол-во пропусков\n', data.isna().sum(), '\n\n')
    print('Кол-во дубликатов\n', data.duplicated().sum(), '\n\n')
    print('Стат\n', data.describe())
df_info(df)

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15366 entries, 0 to 15365
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id               15366 non-null  int64
1   object_name      15366 non-null  object
2   chain            15366 non-null  object
3   object_type      15366 non-null  object
4   address          15366 non-null  object
5   number           15366 non-null  int64
dtypes: int64(2), object(4)
memory usage: 720.4+ KB
None

Кол-во пропусков
id           0
object_name  0
chain        0
object_type  0
address      0
number       0
dtype: int64

Кол-во дубликатов
0

Стат
              id          number
count  15366.000000  15366.000000
mean   119720.066901    59.547182
std     73036.130732    74.736833
min      838.000000     0.000000
25%    28524.000000    12.000000
50%   144974.500000    40.000000
75%   184262.250000    80.000000
max   223439.000000   1700.000000
```

Дубликатов и пропусков нет. Типы данных привели к нужным. Всего датафрейм содержит 15366 записей и 6 колонок. Среднее кол-во посадочных мест почти 60, медиана - 40.

Предобработка

Заменим тип данных в колонке `chain` на логический

```
In [6]: df['chain'] = df['chain'].replace('нет', 0)
df['chain'] = df['chain'].replace('да', 1)
```

```
In [7]: df['chain'] = df['chain'].astype('bool')
```

```
In [8]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15366 entries, 0 to 15365
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id               15366 non-null  int64
1   object_name      15366 non-null  object
2   chain            15366 non-null  bool
3   object_type      15366 non-null  object
4   address          15366 non-null  object
5   number           15366 non-null  int64
dtypes: bool(1), int64(2), object(3)
memory usage: 615.4+ KB
```

Также я бы посмотрел на неявные дубликаты, ведь названия сетей явно могут быть записаны по-разному

```
In [9]: sorted(df[df['chain'] == 1]['object_name'].unique())
```

```
Out[9]: ['Beverly Hills Diner',
        'Bierloga',
        'Black & White',
        'Bocconcino',
        'BooBo',
        'Bubbleology',
        'Burger Club',
        'COFIX',
        'CORREAS',
        'CoffeeShop',
        'Coffeeshop Company',
        'Cofix',
        'Cookhouse',
        'Correas',
        'Costa Coffee',
        'Deli by Prime Прайм-кафе',
        'Dunkin Donuts',
        'FRIDAYS',
        'Florentini',
        'Fridays KFC',
        'GRAND УПЮК',
        'GlowSubs Sandwiches',
        'Goodman Гудман',
        'Grand Cru',
        'HEALTHY Food',
        'Healthy Food',
        'Healthy food',
        'IL Forno Иль Форно',
        'Jeffreys Coffee',
        'Jeffreys coffee',
        'KFC',
        'KFC Волгоградский',
        'Krispy Creme',
        'Krispy Krem',
        'LAVKALAVKA',
        'Luciano',
        'M Cafe Хинкальная',
        'Maki Maki',
        'Marmalato',
        'Marrakesh Хинкальная',
        'MosKalyan',
        'Movenpick',
        'My box',
        'Osteria Mario',
        'PAUL Поль',
        'PRIME STAR Прайм стар',
        'PRIME Прайм стар Прайм-кафе',
        'Panda Express',
        'Pizengof99 СПб',
        'Prime Прайм стар',
        'STARBUCKS COFFEE',
        'SUBWAY',
        'Starbucks',
        'Starbucks Старбакс',
        'Starbucks кофе',
        'Starbucks, СТАРБАКС КОФЕ',
        'Starlite Diner',
        'Subway',
        'Sushilka',
        'TAJJ MAHAL',
        'TGI FRIDAYS',
        'TGI Fridays',
        'TUTTI FRUTTI FROZEN YOGURT',
        'The Terrace',
        'Tokyo bay',
```

'Torro Grill Toppo Гриль',
'Travelers Coffe',
'Travelers Coffee',
'Tutti frutti',
'UPSIDE DOWN CAKE',
'Upside Down',
'Upside Down Cake',
'Van Wok Ванвок',
'Vanwok Ванвок',
'VietCafe Вьеткафе',
'White Rabbite (Белый кролик)',
'Wok & Box',
'Wokker',
'Yogurt Frenzy',
'«Кафе «Кулинарная лавка братьев Караваевых»',
'«Кафе «Кулинарная лавка братьев Караваевых»»',
'Ёрш',
'АВ-Дейли Азбука Вкуса',
'АКАДЕМИЯ',
'АЛЛО ПИЦЦА',
'АНДЕРСОН',
'АРИАНА',
'Азбука Вкуса',
'Азбука вкуса',
'Академия',
'Алло Пицца',
'Алло Пицца 2',
'Американ Сити Пицца ЦЕХ ПО ИЗГОТОВЛЕНИЮ ПИЦЦЫ',
'АндерСон',
'Андерсон',
'Анти-кафе кальянная «MosKalyan»',
'БАКИНСКИЙ БУЛЬВАР',
'БАР Азбука Вкуса',
'ББ энд БУРГЕРС В&В Бургер',
'БИСТРО ПРОНТО',
'БОТИК ПЕТРА',
'БРАВА Коста кофе',
'БРУСНИКА',
'БУРГЕР КИНГ',
'Бабай Клаб',
'Бакинский бульвар',
'Бар «Суши Wok»',
'Бар Боулинг-Космик',
'Бар Мята Lounge',
'Бар- буфет «Николай»',
'Бар-буфет Николай',
'Баракат',
'Барашка',
'Барбарис',
'Барбекю',
'Баскин Роббинс',
'Бенто WOK',
'Билла',
'Бир Хаус Паб',
'Бир хаус',
'Блинная',
'Бодрый день Кофе с собой',
'Ботик Петра',
'Братья Караваевы',
'Братья Караваевых',
'Брудер',
'Брусника',
'Бургер Кинг',
'Бургер Кинг Burger King',
'Бургер Клаб',

'Бургер кинг',
'Бургер клуб',
'Бутчер Бизон',
'B&B Бургер',
'Black Burger Чайхона №1',
'ВГости. Урюк',
'ВЕНЕЦИЯ',
'ВОККЕР',
'Ваби-Саби',
'Ваби-саби',
'Ванвок',
'Вареничная №1',
'Васаби',
'Венеция',
'Виват-Пицца',
'Виктория',
'Вкусняшка',
'Воккер',
'Волконский',
'Восточный базар',
'ВьетКафе',
'Вьеткафе',
'ГЕНАЦВАЛИ',
'ГРАБЛИ',
'Газпромнефтьцентр',
'Гамбринус',
'Ганс и Марта',
'Гино-но-таки',
'Городские автокофейни',
'Готика',
'Готовые блюда Милти',
'Грабли',
'Граци рагацци Зю Кафе',
'Грузинская кухня ЭЗО Хинкальная',
'Грузинские каникулы Барбарис',
'Гудман',
'Гудман Гудвин',
'Гурман',
'Гурмания',
'ДАНКИН ДОНАТС',
'ДВЕ ПАЛОЧКИ',
'ДЕТСКИЙ КЛУБ «АНДЕРСОН»',
'ДЕТСКОЕ КАФЕ «АНДЕРСОН»',
'ДЖОН ДЖОЛИ',
'ДОДО ПИЦЦА',
'ДОДО Пицца',
'ДОДО пицца',
'ДОМИНО'С ПИЦЦА",
'Да Пино',
'Да Пино (Da Pino)',
'Дабл Би',
'Дабл Би блэк',
'Дабл би',
'Данкин Донатс',
'Две палочки',
'Де Марко',
'Детское кафе «Космик»',
'Джаганнат',
'Джон Джоли',
'ДоДо Пицца',
'ДоДо пицца',
'Добрынинский и партнёры',
'Додо Пицца',
'Додо пицца',
'Домашнее Кафе сеть городских кафе',

"Домино'с Пицца",
"Домино'с Пицца Domino,s Pizza",
"Домино'с пицца",
'Дорогая я перезвоню',
'Дюшес',
'Елки-Палки',
'Жан Жак',
'ЗОЛОТАЯ ВОБЛА',
'Закусочная KFC',
'Закусочная «KFC»',
'Закусочная «Американ Сити Пицца»',
'Закусочная «Баскин Роббинс & Стардогс»',
'Закусочная «Бургер Кинг»',
'Закусочная «Крошка Картошка»',
'Закусочная «Пицца Хат»',
'Закусочная «СтардогS»',
'Закусочная «СушиШоп»',
'Закусочная «Теремок»',
'Зодиак',
'Золотая вобла',
'Зю Кафе',
'ИЛЬ ПАТИО',
'ИЛЬ ФОРНО',
'Изба',
'Икура Паб',
'Иль Патио',
'Иль Патио Планета Суши',
'Иль Форно',
'Иль-патио',
'Илья Муромец',
'Ирландский Паб',
'Ичибан Боши',
'КАКТУС',
'КАРО',
'КАФЕ KFC',
'КАФЕ Starbucks',
'КАФЕ «Андерсон»',
'КАФЕ «Восточный базар»',
'КАФЕ «МАКДОНАЛДС»',
'КАФЕ «МУ-МУ»',
'КАФЕ «ТЕРЕМОК»',
'КАФЕ «Теремок»',
'КАФЕ «Хлеб насущный»',
'КАФЕ «ШТОЛЛЕ»',
'КАФЕ «Шоколадница»',
'КАФЕ «ЯКИТОРИЯ»',
'КАФЕ Кофе-Хаус',
'КАФЕ МАКДОНАЛДС',
'КАФЕ-ПЕКАРЕЯ «ВОЛКОНСКИЙ»',
'КОЛБАСОФФ',
'КОРЧМА «ТАРАС БУЛЬБА»',
'КОРЧМА ТАРАС БУЛЬБА',
'КОФЕ С СОБОЙ',
'КОФЕ ХАУС',
'КОФЕ ХАУС. ЭКСПРЕССО И КАПУЧИНО БАР',
'КОФЕЙНЯ «Costa Coffee»',
'КОФЕЙНЯ «ДЕ МАРКО»',
'КОФЕЙНЯ «ШОКОЛАДНИЦА»',
'КОФЕМАНИЯ',
'КОФЕПОРТ',
'КРОШКА КАРТОШКА',
'КРУЖКА',
'КУЛИНАРНАЯ ЛАВКА БРАТЬЕВ КАРАВАЕВЫХ',
'КУЛИНАРНОЕ БЮРО',
'Кальян- бар «Мята Lounge»',

'Кальян-бар MosKalyan',
'Кальян-бар «Мята Lounge»',
'Кальянная F-lounge',
'Кальянная «Мята Lounge»',
'Кальянная Мята Lounge',
'Кафе PRIME Прайм стар',
'Кафе Prime Прайм стар Прайм-кафе',
'Кафе «Costa coffee»',
'Кафе «KFC»',
'Кафе «PRIME» Прайм стар',
'Кафе «Prime» (Прайм стар)',
'Кафе «Prime» Прайм стар',
'Кафе «SUBWAY»',
'Кафе «Starbucks»',
'Кафе «Subway»',
'Кафе «АндерСон»',
'Кафе «Андерсон»',
'Кафе «БУРГЕР КИНГ»',
'Кафе «Баскин Роббинс»',
'Кафе «Бургер Кинг»',
'Кафе «Ваби-Саби»',
'Кафе «Данкин Донатс»',
"Кафе «Домино 'с Пицца»",
'Кафе «Иль Патио»',
'Кафе «Кофе Хаус»',
'Кафе «Кофейня Старбакс»',
'Кафе «Крошка Картошка»',
'Кафе «Кружка»',
'Кафе «Кулинарная лавка братьев Караваевых»',
'Кафе «Кулинарное бюро»',
'Кафе «МАКДОНАЛДС»',
'Кафе «МЕЛЕНКА»',
'Кафе «Макдоналдс»',
'Кафе «Марукамэ»',
'Кафе «Менза»',
'Кафе «Му-Му»',
'Кафе «Му-му»',
'Кафе «Нияма»',
'Кафе «Папа Джонс»',
'Кафе «Пицца Фабрика»',
'Кафе «Пицца Хат»',
'Кафе «Поль Бейкери»',
'Кафе «Прайм Стар»',
'Кафе «Сабвей» Старбакс',
'Кафе «Сим-Сим»',
'Кафе «Старбакс»',
'Кафе «Суши Wok»',
'Кафе «Суши Вок»',
'Кафе «Суши Сет»',
'Кафе «ТО ДА СЁ»',
'Кафе «Тануки»',
'Кафе «Тапчан»',
'Кафе «Теремок»',
'Кафе «То Да Сё»',
'Кафе «Торнадо»',
'Кафе «Урюк» Хивинская чайхона',
'Кафе «Хинкальная»',
'Кафе «Чайхана Тапчан»',
'Кафе «Шоколадница»',
'Кафе «Штолле»',
'Кафе «Якитория»',
'Кафе БУРГЕР КИНГ',
'Кафе Бургер Кинг',
'Кафе Волконский пекарня',
'Кафе Иль Патио «Il Patio»',

'Кафе КОФЕ ТУН',
'Кафе Космик',
'Кафе Кофемания',
'Кафе МУ-МУ',
'Кафе ПРОНТО',
'Кафе Пекарня Хачапури',
'Кафе Песто и Митлес',
'Кафе Пронто',
'Кафе Рецепттор',
'Кафе Суши Wok',
'Кафе Чайхона №1',
'Кафе Штолле',
'Кафе Якитория',
'Кафе быстрого питания Му-му',
'Кафе мороженого «Tutti Frutti»',
'Кафе при АЗС',
'Кафе при АЗС «ВР»',
'Кафе при АЗС Газпромнефть',
'Кафе при АЗС Лукойл',
'Кафе теремок',
'Кафе шоколад',
'Кафе-закусочная «Бургер Кинг»',
'Каффе «Wokker»',
'Квартира 44',
'Кебаб Хаус',
'Кебаб хаус',
'Колбасофф',
'Кондитерия Тирольские пироги',
'Кондитерская Волконский',
'Кондитерская-пекарня Волконский',
'Корчма Тарас Бульба',
'Космик',
'Коста Кофе',
'Кофе',
'Кофе Хаус',
'Кофе Хаус,Экспресс',
'Кофе с Собой',
'Кофе с собой',
'Кофе с собой Gate 42',
'Кофе с собой «Кофейня»',
'Кофе тун',
'Кофе-Бин',
'Кофе-Хаус',
'Кофейня «Дабл Би»',
'Кофейня «Кофе Хаус»',
'Кофейня «Кофе хаус»',
'Кофейня «СТАРБАКС КОФЕ»',
'Кофейня «Старбакс»',
'Кофейня «Шоколадница»',
'Кофейня Кофемания, пиццерия Бармалини',
'Кофейня ШОКОЛАДНИЦА',
'Кофейня Шоколадница',
'Кофемания',
'Кофемания Starbucks',
'Кофепорт',
'Кофетун',
'Кофешоп Coffeeshop Company',
'Крепери де Пари',
'Крепери де пари',
'Крошка Картошка',
'Крошка картошка',
'Кружка',
'Кулинарная Лавка Братьев Караваевых Братья Караваевы',
'Кулинарная лавка братьев Караваевых',
'Кулинарная лавка братьев Караваевых Братья Караваевы',

'Кулинарное бюро',
'Кулинарное бюро Kitchen',
'Кулинарное бюро Китчен',
'Лаундж-бар «Мята Lounge»',
'Лепешка',
'Лето',
'Ливан-Хаус',
'Ливан-хаус',
'Лукойл',
'Лукойл-Центрнефтьпродукт',
'МАГНОЛИЯ',
'МАКДОНАЛДС',
'МАКИ-МАКИ',
'МЕНЗА',
'МИ ПЬЯЧЕ',
'МИЛТИ',
'МИМИНО',
'МОСКВА – МАКДОНАЛДС',
'МСК Московская сеть кальянных',
'МСК Московская сеть кальянных на Шаболовке',
'МУ-МУ',
'МЯСОR00B',
'МагБургер',
'Магазин готовой еды «Милти»',
'Магазин-кулинария «Суши Сет»',
'Магбургер',
'Магбургер АЗС №16',
'Магбургер АЗС №17',
'Макдоналдс',
'Маки-Маки',
'Маки-маки',
'Макс Бреннер',
'Марчеллис',
'Меленка',
'Менза',
'Метро K&K',
'Ми Пьяче',
'Милано пицца',
'Милти',
'Мимино',
'Мимино Хинкальная',
'Мираторг',
'Мистер Картошка',
'Молли гвинз',
'Му-Му',
'Му-му',
'Мюнгер',
'МясоR00B',
'Мята',
'Мята Lounge',
'Мята Lounge Автозаводская',
'Мята Lounge Октябрьская',
'Мята Lounge Шаболовка',
'Мята lounge',
'НИЯМА',
'Николай',
'Нияма',
'Нияма Пицца Пи',
'Нияма. Пицца Пи',
'Ньокки',
'Обжорный ряд',
'Оникс',
'ПАПА ДЖОНС',
'ПАПА ДЖОНС ПИЦЦА',
'ПБО «KFC Покрышкина»',

'ПБО «KFC»',
'ПБО «Теремок»',
'ПИВКО',
'ПИЛЗНЕР',
'ПИРОГИ ШТОЛЛЕ',
'ПИЦЦЕРИЯ «Папа Джонс»',
'ПИЦЦЕТОРИЯ',
'ПОМИДОР',
'ПРАВДА Кофе',
'ПРАЙМ',
'ПРАЙМ СТАР',
'ПРАЙМкафе Прайм стар',
'ПРИМАВЕРА',
'ПРОНТО',
'Паб Кружка',
'Павлин Мавлин Чайхона №1',
'Павлин-Мавлин',
'Панчо Пицца',
'Папа Джонс',
'Перекресток',
'ПивКо',
'Пикколо',
'Пиппони',
'Пироговая Штолле',
'Пиццерия «Папа Джонс»',
'Пицца Pomodoro',
'Пицца Pomodoro и Суши Дзен',
'Пицца «Паоло»',
'Пицца ПИПОНИ',
'Пицца Паоло',
'Пицца Паоло и Бенто WOK',
'Пицца Пипони',
'Пицца Фабрика',
'Пицца Хат',
'Пицца Экспресс',
'Пицца хат',
'Пицца экспресс',
'Пиццерия «АКАДЕМИЯ»',
'Пиццерия «Додо Пицца»',
'Пиццерия «Домино'с Пицца»',
'Пиццерия «ПАПА ДЖОНС»',
'Пиццерия «Папа Джонс»',
'Пиццерия «Пицца Паоло»',
'Пиццерия «Пицца Хат»',
'Пиццерия Донателло',
'Пиццерия Папа Джонс',
'Пиццерия Пиу дель Чибо',
'Пиццетория',
'Планета Суши',
'Планета суши',
'Поль Бейкери',
'Пончиковое кафе ИКЕА Ikea',
'Порто Мальтезе',
'Пражечка',
'Прайм',
'Прайм Стар',
'Прайм Стар кафе',
'Прайм стар',
'Прайм-кафе',
'Прайм-кафе Прайм стар',
'Праймстар',
'Предприятие быстрого обслуживания «KFC»',
'Предприятие быстрого обслуживания «Бенто WOK»',
'Предприятие быстрого обслуживания «Бургер Кинг»',
'Предприятие быстрого обслуживания «ДОДО Пицца»',

"Предприятие быстрого обслуживания «Домино'с Пицца»",
'Предприятие быстрого обслуживания «Крошка Картошка»',
'Предприятие быстрого обслуживания «Магбургер»',
'Предприятие быстрого обслуживания «Макдоналдс»',
'Предприятие быстрого обслуживания «Милти»',
'Предприятие быстрого обслуживания «СушиШоп»',
'Предприятие быстрого обслуживания «Теремок»',
'Предприятие быстрого питания «Крошка Картошка»',
'Примавера',
'Пронто',
'Простые Вещи',
'Простые вещи',
'РЕСТОРАН «ДЖОН ДЖОЛИ»',
'РЕСТОРАН «Космик»',
'РЕСТОРАН «ТАНУКИ»',
'РЕСТОРАН «Тануки»',
'РЕСТОРАН «УРЮК»',
'РЕСТОРАН «ЧАЙХОНА №1»',
'РЕСТОРАН «Чайхона №1»',
'РЕСТОРАН«МАКДОНАЛДС»',
'РОСТИКС KFC',
'Ресторан KFC',
'Ресторан «Florentini»',
'Ресторан «KFC»',
'Ресторан «Ёрш»',
'Ресторан «БРУДЕР»',
'Ресторан «БУРГЕР КИНГ»',
'Ресторан «Бакинский бульвар»',
'Ресторан «Брудер»',
'Ресторан «Бургер Кинг»',
'Ресторан «Вареничная №1»',
'Ресторан «Вьеткафе»',
'Ресторан «Джардино Да Пино»',
'Ресторан «Джон Джоли»',
"Ресторан «Домино'с Пицца»",
'Ресторан «Золотая Вобла»',
'Ресторан «ИЛЬ ПАТИО»',
'Ресторан «Иль Патио»',
'Ресторан «Илья Муромец»',
'Ресторан «Колбасофф»',
'Ресторан «Корчма Тарас Бульба»',
'Ресторан «Космик»',
'Ресторан «Макдоналдс»',
'Ресторан «Ньюкки»',
'Ресторан «Пицца Хат»',
'Ресторан «Планета Суши»',
'Ресторан «Пронто»',
'Ресторан «ТАНУКИ»',
'Ресторан «Тануки»',
'Ресторан «Темпл Бар»',
'Ресторан «Теремок»',
'Ресторан «Территория»',
'Ресторан «Урюк»',
'Ресторан «Хинкальная»',
'Ресторан «Чайхана Урюк»',
'Ресторан «Чайхона №1»',
'Ресторан «ЯКИТОРИЯ»',
'Ресторан «Якитория»',
'Ресторан «Ян Примус»',
'Ресторан Джон Джоли',
'Ресторан ИКЕА Ikea',
'Ресторан Маки-Маки',
'Ресторан ТАНУКИ',
'Ресторан Тануки',
'Ресторан Торро Гриль',

'Ресторан Хинкальная',
'Ресторан быстрого питания KFC',
'Ресторан быстрого питания «Крошка Картошка»',
'Ресторан японской кухни «Маки-Маки»',
'Ресторан«ВьетКафе»',
'Рецептор',
'Роллофф',
'Роснефть',
'Ростикс KFC',
'Руккола',
'САБВЕЙ',
'СПБ',
'СТАРИНА МЮЛЛЕР Старина Миллер',
'СТЕЙК-ХАУС ГУДМАН',
'СУШИ WOK',
'СУШИ ВОК',
'Сабвей',
'Сабвей Subway',
'Сабвей СушиВок',
'Сбарро',
'Сварня',
'Семейное кафе Андерсон',
'Семейное кафе и кондитерская «Андерсон»',
'Семейный ресторан «Космик»',
'Сеть магазинов «Суши Wok»',
'Сити',
'Сити Пицца',
'Сказка',
'Советские времена',
'Советские времена Чебуречная СССР',
'Спорт-бар «Космик»',
'Старбакс',
'Старбакс Starbucks',
'Старбакс Кофе',
'СтардогS',
'СтардогS и шаурма',
'Стардогс',
'Стардогс СтардогS',
'Старина Миллер',
'Стейк Хаус ГУДМАН',
'Сувлаки',
'Суп Кафе',
'Суши WOK',
'Суши WoK',
'Суши Wok',
'Суши «Маки-Маки»',
'Суши Вок',
'Суши Сет',
'Суши бар «Маки-маки»',
'Суши вок',
'Суши сет',
'Суши тун',
'Суши тун, Хруст pizza, Кофе тун',
'СушиМан',
'СушиШоп',
'Сытая Утка',
'ТАНУКИ',
'ТАШИР ПИЦЦА',
'ТЕРЕМОК',
'ТЕРЕМОК ИНВЕСТ',
'ТЕРРИТОРИЯ ЯСЕНЕВО',
'Тайм Авеню',
'Тамаси Суши',
'Тануки',
'Тапчан',

'Ташир пицца',
'Темпл Бар',
'Темпл бар',
'Теремок',
'Теремок–Инвест',
'Территория',
'Территория TIMBIGFAMILY',
'Тирольские пироги',
'То Да Сё',
'Торро Гриль',
'Точка продажи готовой еды Милти',
'Траттория Semplice',
'Траттория Примавера',
'УРЮК',
'Урожай',
'Урюк',
'Урюк Вавилова',
'ФАНТОЦЦИ РУС',
'ФИТНЕС–БАР «WORLD CLASS»',
'Французская выпечка',
'ХАЧАПУРИ',
'ХИНКАЛЬНАЯ',
'ХЛЕБ НАСУЩНЫЙ',
'Хачапури',
'Хачапури, Одесса –мама',
'Хижина',
'Хинкальная',
'Хинкальная City',
'Хинкальная КИНТО',
'Хинкальная №1',
'Хлеб Насущный',
'Хлеб насущный',
'Хлеб&Со',
'Хлебница пекарня',
'Цинандали Хинкальная',
'ЧАЙХОНА №1',
'Чайхона №1',
'Чайхона №1 Бутово',
'Чебуречная СССР',
'Чешская пивная «Козловица»',
'Чешская пивная «Пилзнер»',
'Чешская пивная ПИЛЗНЕР',
'Чин Чин',
'ШАНТИМЕЛЬ',
'ШАШЛЫК–МАШЛЫК',
'ШВАРЦВАЛЬД',
'ШОКОЛАДНИЦА',
'ШТОЛЛЕ',
'Шантимель',
'Шантимель (кондитерские)',
'Шикари Иль Патио',
'Шоколад',
'Шоколадница',
'Шоколадница Кофе Хаус',
'Шоколадница Кофемания',
'Штирбирлиц',
'Штолле',
'Югос',
'ЯКИТОРИЯ',
'ЯКИТОРИЯ И СПОРТ БАР',
'ЯММИ МИКС',
'ЯПОША',
'Якитория',
'Якитория и Ян Примус',
'Ян Примус',

```
'Японский ресторан «Ваби-Сабид»,
'Япоша',
'кафе «Му-Му»',
'кафе «Нияма»',
'кафе «Папа Джонс»',
'кафе «Старбакс»',
'кафе «Шоколадница»',
'ресторан «Брудер»',
'ресторан «Якитория»']
```

```
In [10]: # Сводник с кол-вом заведений
(
    df.query('chain == 1')
    .pivot_table(index='object_name', values='id', aggfunc='count')
    .reset_index()
)
```

```
Out[10]:
```

| | object_name | id |
|-----|---------------------|-----|
| 0 | Beverly Hills Diner | 1 |
| 1 | Bierloga | 1 |
| 2 | Black & White | 1 |
| 3 | Bocconcino | 3 |
| 4 | BooBo | 1 |
| ... | ... | ... |
| 719 | кафе «Папа Джонс» | 1 |
| 720 | кафе «Старбакс» | 2 |
| 721 | кафе «Шоколадница» | 1 |
| 722 | ресторан «Брудер» | 1 |
| 723 | ресторан «Якитория» | 1 |

724 rows x 2 columns

Неявные дубликаты явно есть, но пока оставим их без изменений, тк в данный момент неясно помешают ли они нам в данной работе. Плюс в данный момент мы можем потерять много времени на обработку такого огромного массива данных вручную. Обидно, если окажется, что мы сделали это впустую

```
In [11]: df[df.duplicated(['object_name', 'chain', 'object_type', 'address'])].sort_values('
```

Out [11]:

| | id | object_name | chain | object_type | address | number |
|-------|--------|-------------|-------|-----------------------------------|---|--------|
| 5190 | 107480 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 5 |
| 5189 | 107471 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 4 |
| 5188 | 107466 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 5 |
| 5187 | 107462 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 5 |
| 5191 | 107485 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 6 |
| ... | ... | ... | ... | ... | ... | ... |
| 14223 | 205506 | Кафе | False | кафе | город Москва, улица Трофимова, владение 36 | 40 |
| 6422 | 19758 | БАР | False | бар | город Москва, улица Улофа Пальме, дом 5, строение 1 | 16 |
| 6421 | 19757 | БАР | False | бар | город Москва, улица Улофа Пальме, дом 5, строение 1 | 12 |
| 6423 | 19759 | БАР | False | бар | город Москва, улица Улофа Пальме, дом 5, строение 1 | 10 |
| 10510 | 173865 | Шаурма | False | предприятие быстрого обслуживания | город Москва, улица Черняховского, дом 4А | 0 |

170 rows x 6 columns

Можно заметить, что есть казалось бы одинаковые заведения по одному адресу. Но кол-во посадочных мест разное. Скорее всего это бизнес-центры или ТРЦ.

In [12]:

```
df[df.duplicated(['object_name', 'chain', 'object_type', 'address', 'number'])].sort
```


Out [12]:

| | id | object_name | chain | object_type | address | number |
|-------|--------|--------------------------------------|-------|-----------------------------------|---|--------|
| 5186 | 107574 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 10 |
| 5190 | 107480 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 5 |
| 5188 | 107466 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 5 |
| 5185 | 107442 | Кафе | False | кафе | город Москва, 1-я Останкинская улица, дом 55 | 20 |
| 2381 | 28611 | КОМБИНАТ ПИТАНИЯ МГТУ ИМ.Н.Э.БАУМАНА | False | столовая | город Москва, 2-я Бауманская улица, дом 5, строение 1 | 100 |
| ... | ... | ... | ... | ... | ... | ... |
| 12744 | 183901 | Гацоева Анжела Камалдиновна | False | предприятие быстрого обслуживания | город Москва, улица Суцёвский Вал, дом 5, строение 5 | 2 |
| 11066 | 183855 | Каменев Василий Владимирович | False | кафе | город Москва, улица Суцёвский Вал, дом 5, строение 6 | 1 |
| 14223 | 205506 | Кафе | False | кафе | город Москва, улица Трофимова, владение 36 | 40 |
| 6423 | 19759 | БАР | False | бар | город Москва, улица Улофа Пальме, дом 5, строение 1 | 10 |
| 10510 | 173865 | Шаурма | False | предприятие быстрого обслуживания | город Москва, улица Черняховского, дом 4А | 0 |

82 rows × 6 columns

Это уже больше похоже на дубликаты. Предлагаю сразу избавиться от них

In [13]:

df = df.drop_duplicates(subset=['object_name', 'chain', 'object_type', 'address', ''])

Анализ данных

Исследуйте соотношение видов объектов общественного питания по количеству. Постройте график.

In [14]:

df.head()

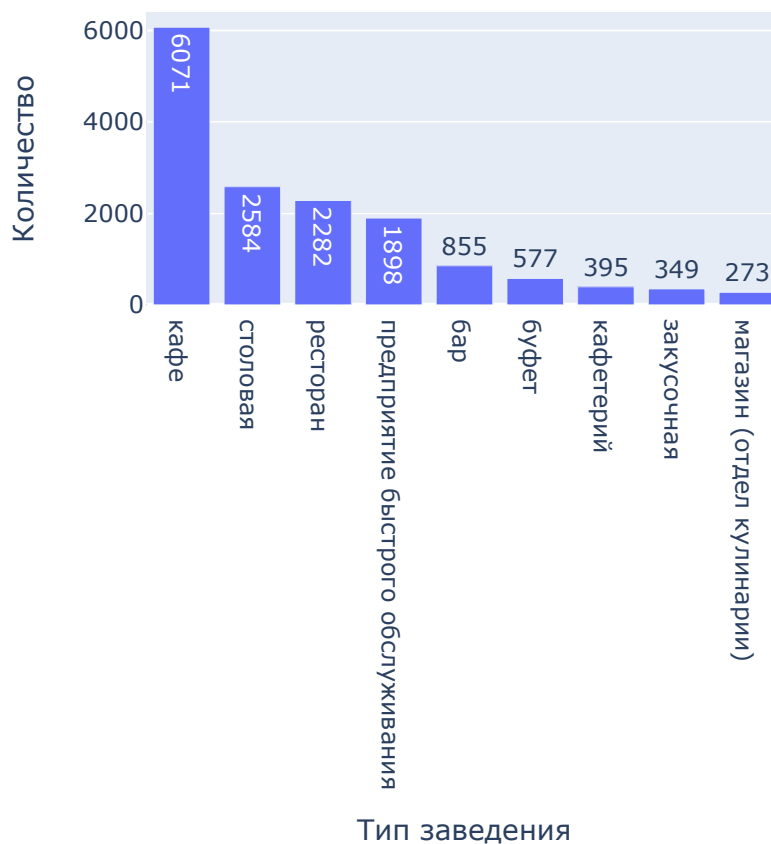
| | id | object_name | chain | object_type | address | number |
|---|--------|------------------------|-------|-------------|---|--------|
| 0 | 151635 | СМЕТАНА | False | кафе | город Москва, улица Егора Абакумова, дом 9 | 48 |
| 1 | 77874 | Родник | False | кафе | город Москва, улица Талалихина, дом 2/1, корпус 1 | 35 |
| 2 | 24309 | Кафе «Академия» | False | кафе | город Москва, Абельмановская улица, дом 6 | 95 |
| 3 | 21894 | ПИЦЦЕТОРИЯ | True | кафе | город Москва, Абрамцевская улица, дом 1 | 40 |
| 4 | 119365 | Кафе «Вишневая метель» | False | кафе | город Москва, Абрамцевская улица, дом 9, корпус 1 | 50 |

```

In [15]: fig = px.bar(df.groupby('object_type')['id'].count().reset_index().sort_values('id',
        x='object_type',
        y='id',
        text='id',
        title='Количество объектов по типу заведения',
        labels={'id': 'Количество', 'object_type': 'Тип заведения'})
fig.show()

```

Количество объектов по типу заведения



Подавляющее большинство заведений в Москве - кафе. Примерно на одном уровне рестораны, столовые и предприятия быстрого обслуживания. Меньше всего кафетериев, закусочных и отделов кулинарии в магазинах

Исследуйте соотношение сетевых и несетевых заведений по количеству. Постройте график.

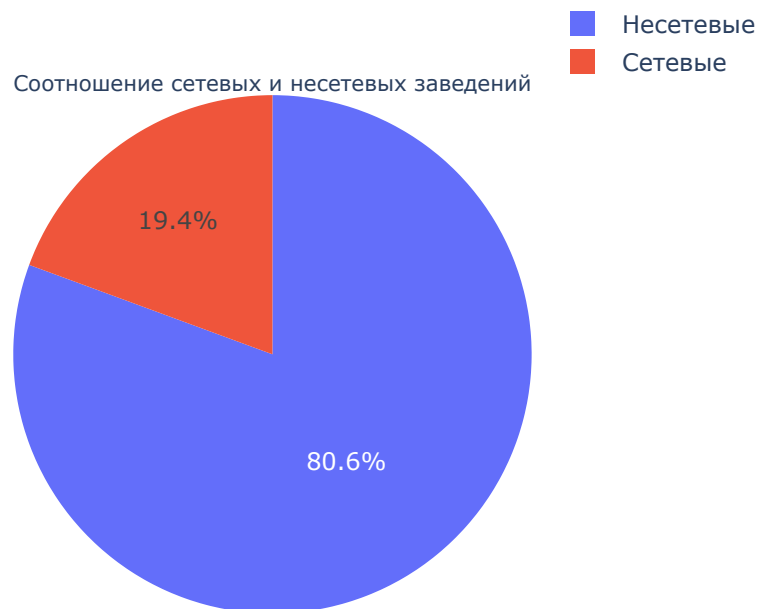
```
In [16]: # Сводная таблица по кол-ву сетевых/несетевых заведений
chain_count = df.groupby('chain')['id'].count().reset_index()
chain_count
```

```
Out[16]:
```

| | chain | id |
|---|-------|-------|
| 0 | False | 12320 |
| 1 | True | 2964 |

```
In [17]: # Круговая диаграмма лучше отражает соотношения
fig_1 = go.Figure(data=[go.Pie(labels=chain_count['chain'].replace([False, True],
values=chain_count['id'],
title='Соотношение сетевых и несетевых заведений')

fig_1.show()
```

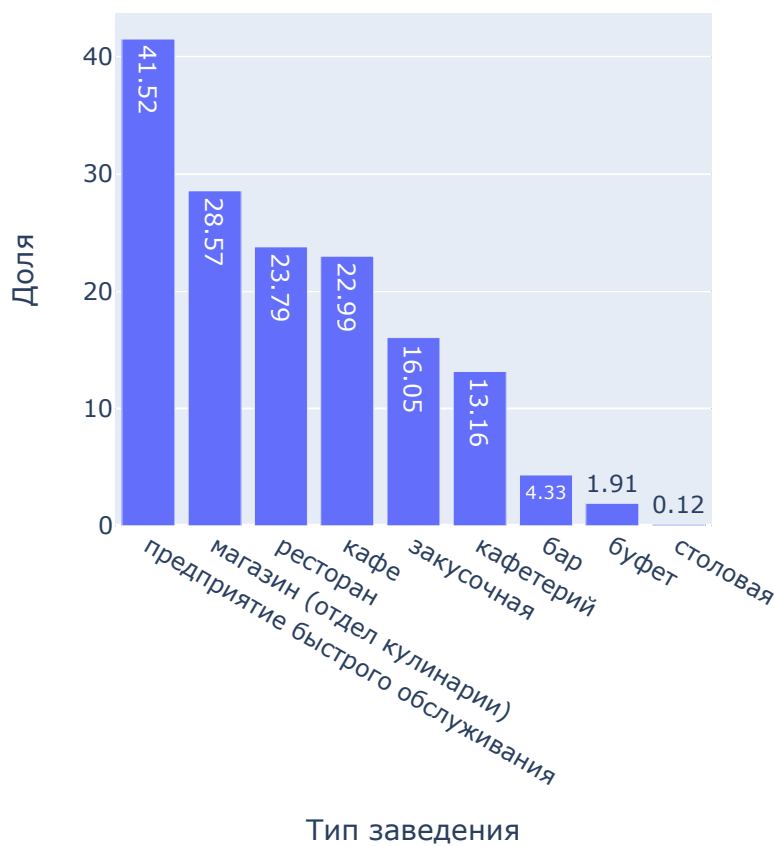


В Москве в 4 раза больше несетевых заведений. 80.7 против 19.3 процентов

Для какого вида объекта общественного питания характерно сетевое распространение?

```
In [18]: fig_2 = px.bar(
    (df.groupby('object_type', as_index=False)
     .apply(lambda x: round(x['chain'].mean()*100,2))
     .rename(columns={None : 'ratio'}))
     .sort_values('ratio', ascending=False)
    ),
    x='object_type',
    y='ratio',
    text='ratio',
    title='Доля сетевых заведений для каждого типа заведений',
    labels={'ratio': 'Доля', 'object_type': 'Тип заведения'}
)
fig_2.show()
```

Доля сетевых заведений для каждого типа завед



Что и следовало ожидать: Почти половина предприятий быстрого обслуживания - сетевые. Столовые и буфеты почти всегда несетевые.

Что характерно для сетевых заведений: много заведений с небольшим числом посадочных мест в каждом или мало заведений с большим количеством посадочных мест?

```
In [19]: chain_stat = (
    df.query('chain == 1')
    .groupby('object_name')['number']
    .agg(['count', 'mean'])
    .reset_index()
    .sort_values('count', ascending=False)
```

```
)  
chain_stat
```

Out [19]:

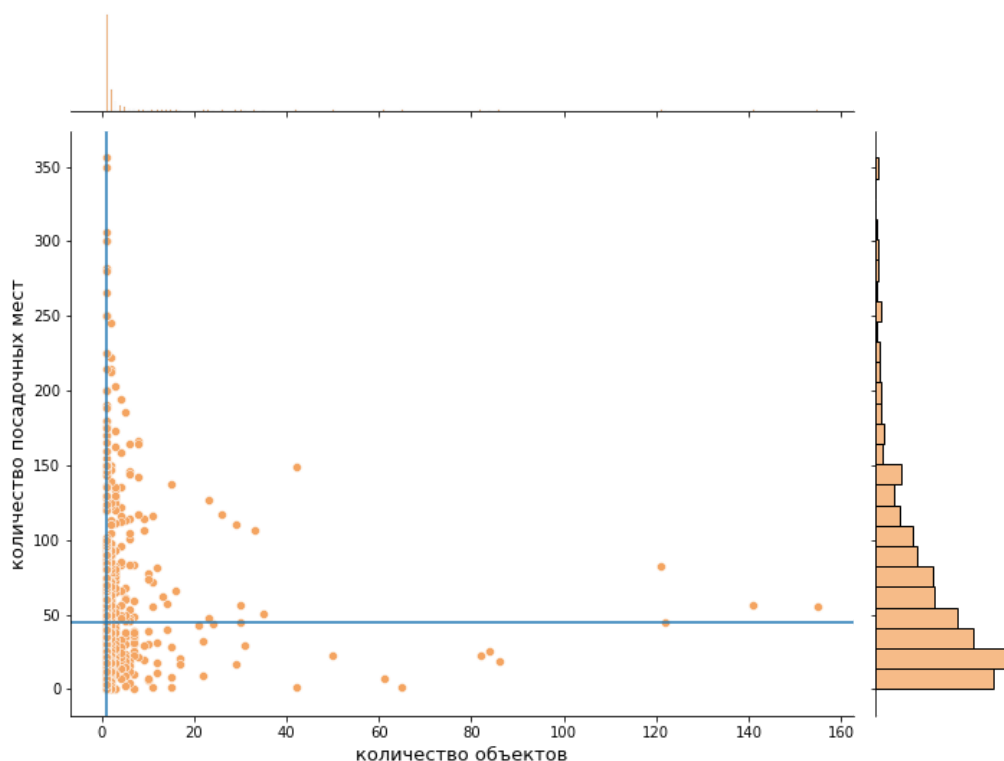
| | object_name | count | mean |
|-----|------------------------|-------|------------|
| 30 | KFC | 155 | 55.341935 |
| 702 | Шоколадница | 141 | 56.170213 |
| 127 | Бургер Кинг | 122 | 45.016393 |
| 418 | Макдоналдс | 121 | 82.173554 |
| 195 | Домино'с Пицца | 86 | 18.209302 |
| ... | ... | ... | ... |
| 293 | Кафе «Марукамэ» | 1 | 40.000000 |
| 291 | Кафе «МЕЛЕНКА» | 1 | 60.000000 |
| 290 | Кафе «МАКДОНАЛДС» | 1 | 40.000000 |
| 289 | Кафе «Кулинарное бюро» | 1 | 18.000000 |
| 723 | ресторан «Якитория» | 1 | 124.000000 |

724 rows x 3 columns

In [20]:

```
#построим два распределения на одном графике  
g=sns.jointplot(x=('count'), y=('mean'), data=chain_stat, color='#f4a460',kind='scatter')  
g.fig.set_figwidth(10)  
g.fig.set_figheight(8)  
g.fig.suptitle('Распределение сетевых заведений по количеству объектов и посадочных мест')  
g.set_axis_labels('количество объектов', 'количество посадочных мест', fontsize=12)  
g.fig.tight_layout()  
  
g.ax_joint.axvline(x=chain_stat['count'].median(), label="Медианное число объектов")  
g.ax_joint.axhline(y=chain_stat['mean'].median(), label="Медианное количество посадочных мест")
```

Распределение сетевых заведений по количеству объектов и посадочных мест



Судя по результатам, в выборке преобладают заведения с небольшим числом объектов (до 10) и небольшим количеством посадочных мест (20-50).

Заведения с большим количеством объектов обладают небольшой вместимостью (40-70 мест)

Для каждого вида объекта общественного питания опишите среднее количество посадочных мест. Какой вид предоставляет в среднем самое большое количество посадочных мест? Постройте графики.

```
In [21]: df.groupby('object_type', as_index=False)['number'].mean()
```

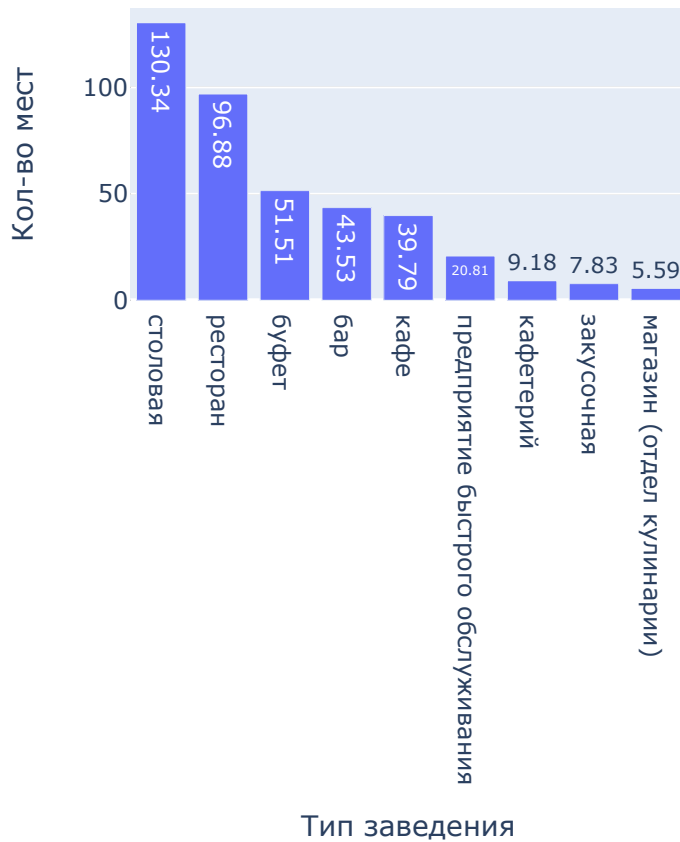
```
Out[21]:
```

| | object_type | number |
|---|-----------------------------------|------------|
| 0 | бар | 43.533333 |
| 1 | буфет | 51.509532 |
| 2 | закусочная | 7.825215 |
| 3 | кафе | 39.785703 |
| 4 | кафетерий | 9.184810 |
| 5 | магазин (отдел кулинарии) | 5.589744 |
| 6 | предприятие быстрого обслуживания | 20.808219 |
| 7 | ресторан | 96.884750 |
| 8 | столовая | 130.337848 |

```
In [22]: fig_3 = px.bar(
    (df.groupby('object_type', as_index=False)
     .apply(lambda x: round(x['number'].mean(),2))
     .rename(columns={None: 'number'})
     .sort_values('number', ascending=False)
    ),
    x='object_type',
    y='number',
    text='number',
    title='Среднее кол-во посадочных мест',
    labels={'number': 'Кол-во мест', 'object_type': 'Тип заведения'})

fig_3.show()
```

Среднее кол-во посадочных мест



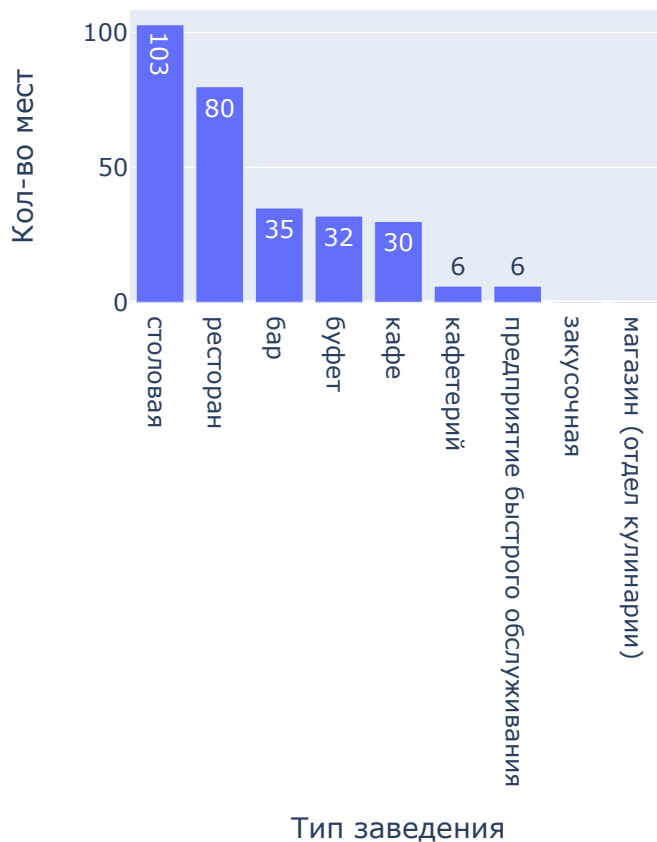
Самые большие по количеству посадочных мест - столовые (что в целом неудивительно) Второе место занимают рестораны. В закусочных и в отделах кулинарии при магазинах меньше всего посадочных мест.

Также посмотрим на медианное значение посадочных мест по типу заведений

```
In [23]: fig_3 = px.bar(
    (df.groupby('object_type', as_index=False)
     .apply(lambda x: round(x['number'].median(),2))
     .rename(columns={None : 'number'}))
     .sort_values('number', ascending=False)
    ),
    x='object_type',
    y='number',
    text='number',
    title='Медианное кол-во посадочных мест',
    labels={'number': 'Кол-во мест', 'object_type': 'Тип заведения'
    })

fig_3.show()
```

Медианное кол-во посадочных мест



В целом результаты не отличаются от средних. Хотя, например, медианное кол-во посадочных мест в барах больше чем в буфетах, что не скажешь о среднем количестве

Выделите в отдельный столбец информацию об улице из столбца address .

```
In [24]: # Разделим строку с адресом по запятой с пробелом
# df['street'] = df['address'].str.split(r', ', n=-1, expand=True)[1]
```

Для выделения улицы из адреса воспользуемся регулярным выражением

```
In [25]: words = ['проезд', 'пр.', 'прзд.', 'шоссе', 'шос.', 'ш.', 'улица', 'ул.', 'переулок',
                 'мкад', 'Московской Кольцевой Автодороги', 'деревня', 'проспект', 'пр.',
                 'набережная', 'наб.', 'тупик', 'линия', 'посёлок', 'территория', 'кварта']
str_pat = r"*.*,\s*\b(?:^,)*?(?:{words})\b(?:^,)*[,,$]+".format("|".join(words))
df["street"] = df["address"].str.extract(str_pat, flags=re.I)
df.sample(15)
```


Out [25] :

| | id | object_name | chain | object_type | address | number | street |
|-------|--------|-------------------------------|-------|-----------------------------------|---|--------|----------------------------|
| 11217 | 178260 | Starhit cafe | False | предприятие быстрого обслуживания | город Москва, Семёновская площадь, дом 1 | 1 | Семёновская площадь |
| 11898 | 182144 | САБВЕЙ | True | кафе | город Москва, улица Лобачевского, дом 112А, строение 2 | 10 | улица Лобачевского |
| 7631 | 85339 | Макдоналдс | True | предприятие быстрого обслуживания | город Москва, проспект Андропова, дом 8 | 20 | проспект Андропова |
| 10716 | 29480 | ГБОУ г.Москвы школа № 609 | False | столовая | город Москва, город Зеленоград, корпус 314 | 180 | NaN |
| 6657 | 139668 | JW Джерри Вингс (Jerry Wings) | False | предприятие быстрого обслуживания | город Москва, площадь Киевского Вокзала, дом 2 | 20 | площадь Киевского Вокзала |
| 13268 | 202506 | Алиса | False | кафе | город Москва, поселение Сосенское, посёлок Коммунарка, улица Александры Монаховой, дом 96, корпус 2 | 15 | улица Александры Монаховой |
| 7513 | 147435 | Выпечка | False | предприятие быстрого обслуживания | город Москва, Шипиловский проезд, владение 39, корпус 3А | 0 | Шипиловский проезд |
| 7344 | 29334 | ШКОЛА 510 | False | столовая | город Москва, Ореховый бульвар, дом 3 | 75 | Ореховый бульвар |
| 11015 | 185013 | Amore Grande | False | кафе | город Москва, Дмитровское шоссе, дом 73, строение 1 | 12 | Дмитровское шоссе |
| 5253 | 83700 | Крошка Картошка | True | кафе | город Москва, Алтуфьевское шоссе, дом 86, корпус 1 | 12 | Алтуфьевское шоссе |
| 5797 | 106261 | Кофе | True | кафе | город Москва, Рязанский проспект, дом 2, корпус 2 | 2 | Рязанский проспект |
| 6090 | 24867 | Кофе Хаус | True | кафе | город Москва, улица Миклухо-Маклая, дом 32А | 20 | улица Миклухо-Маклая |
| 1717 | 150811 | Тапки | False | ресторан | город Москва, улица Вавилова, дом 81, корпус 1 | 45 | улица Вавилова |

| | id | object_name | chain | object_type | address | number | street |
|------|-------|----------------------------------|-------|-------------|---|--------|-----------------------|
| 5622 | 23213 | Шоколадница | True | кафе | город Москва, Свободный проспект, дом 33 | 80 | Свободный проспект |
| 6831 | 23707 | Столовая ГБОУ Школа № 1298 | False | столовая | город Москва, Юровская улица, дом 97 | 192 | Юровская улица |

В сете остались адреса без указания города в адресе. Изменим регулярное выражение для данных адресов и попробуем вытащить улицу

```
In [26]: str_pat = r"^([\,]*?(?:{ })\b([\,]*)[,\$]+".format("|".join(words))
df.loc[df['street'].isna(), 'street'] = df.loc[df['street'].isna(), 'address']
df[df['street'].isna()]
```

| | id | object_name | chain | object_type | address | number | street |
|-------|--------|---------------------------------|-------|-------------|---|--------|--------|
| 859 | 23601 | Буфет - ГБОУ СОШ № 64 | False | буфет | Партизанская улица, дом 30 | 100 | NaN |
| 860 | 23604 | Буфет - ГБОУ 60 | False | буфет | Партизанская улица, дом 30 | 100 | NaN |
| 1405 | 20199 | ЭТО | False | кафе | город Москва, город Зеленоград, корпус 401 | 35 | NaN |
| 1406 | 68113 | Френдс | False | кафе | город Москва, город Зеленоград, корпус 435 | 34 | NaN |
| 1407 | 20105 | Кафе «Граф Монте- Кристо» | False | кафе | город Москва, город Зеленоград, корпус 436 | 40 | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 14468 | 205933 | Кафе «Империя бургеров» | False | кафе | город Москва, город Зеленоград, корпус 1106E | 24 | NaN |
| 14470 | 220595 | Пекарня № 1 | False | кафетерий | город Москва, город Зеленоград, корпус 1627 | 6 | NaN |
| 14473 | 220978 | Бар суши TAKAI | False | кафе | город Москва, город Зеленоград, корпус 1602A | 28 | NaN |
| 14478 | 211676 | Никольский | False | ресторан | город Москва, город Зеленоград, корпус 532 | 120 | NaN |
| 15184 | 220587 | Пицца Паоло | True | кафетерий | город Москва, город Зеленоград, корпус 1627 | 0 | NaN |

192 rows x 7 columns

Замена почему-то не происходит. Создадим таблицу с названием улиц с

пропусками и из нее вытащим название улиц в наш датафрейм

```
In [27]: nan_street = df.loc[df['street'].isna(), 'address'].str.extract(str_pat, flags=re.IGNORECASE)
nan_street.columns = ['street']
nan_street
```

```
Out[27]:
```

| | street |
|-------|--------------------|
| 859 | Партизанская улица |
| 860 | Партизанская улица |
| 1405 | NaN |
| 1406 | NaN |
| 1407 | NaN |
| ... | ... |
| 14468 | NaN |
| 14470 | NaN |
| 14473 | NaN |
| 14478 | NaN |
| 15184 | NaN |

192 rows × 1 columns

```
In [28]: # Замена
df.loc[df['street'].isna(), 'street'] = nan_street['street']
```

```
In [29]: # группируем данные по улицам
rest_address = df.groupby('street').agg({'id': 'count'}).sort_values(by='id')
rest_address
```

```
Out[29]:
```

| | street | id |
|------|------------------------------------|-----|
| 0 | проспект Мира | 204 |
| 1 | Профсоюзная улица | 182 |
| 2 | Ленинградский проспект | 172 |
| 3 | Пресненская набережная | 167 |
| 4 | Варшавское шоссе | 165 |
| ... | ... | ... |
| 1981 | улица Академика Семёнова | 1 |
| 1982 | Боровая улица | 1 |
| 1983 | Малая Тульская улица | 1 |
| 1984 | Большой Трёхсвятительский переулок | 1 |
| 1985 | Выползов переулок | 1 |

1986 rows × 2 columns

```
In [30]: df.head()
```

Out [30]:

| | id | object_name | chain | object_type | address | number | street |
|---|--------|------------------------------|-------|-------------|---|--------|--------------------------|
| 0 | 151635 | СМЕТАНА | False | кафе | город Москва, улица Егора Абакумова, дом 9 | 48 | улица Егора Абакумова |
| 1 | 77874 | Родник | False | кафе | город Москва, улица Талалихина, дом 2/1, корпус 1 | 35 | улица Талалихина |
| 2 | 24309 | Кафе «Академия» | False | кафе | город Москва, Абельмановская улица, дом 6 | 95 | Абельмановская улица |
| 3 | 21894 | ПИЦЦЕТОРИЯ | True | кафе | город Москва, Абрамцевская улица, дом 1 | 40 | Абрамцевская улица |
| 4 | 119365 | Кафе «Вишневая метель» | False | кафе | город Москва, Абрамцевская улица, дом 9, корпус 1 | 50 | Абрамцевская улица |

In [31]:

```
# Проверим пропуски
df.isnull().sum()
```

Out [31]:

```
id          0
object_name 0
chain       0
object_type 0
address     0
number      0
street     129
dtype: int64
```

In [32]:

```
df[df['street'].isnull()]
```

Out [32]:

| | id | object_name | chain | object_type | address | number | street | |
|--|-------|-------------|--------------------------|-------------|-----------|--|--------|-----|
| | 1405 | 20199 | ЭТО | False | кафе | город Москва, город Зеленоград, корпус 401 | 35 | NaN |
| | 1406 | 68113 | Френдс | False | кафе | город Москва, город Зеленоград, корпус 435 | 34 | NaN |
| | 1407 | 20105 | Кафе «Граф Монте-Кристо» | False | кафе | город Москва, город Зеленоград, корпус 436 | 40 | NaN |
| | 1408 | 22982 | Альфорно | False | кафе | город Москва, город Зеленоград, корпус 438 | 49 | NaN |
| | 1409 | 151267 | Альфорно | False | кафе | город Москва, город Зеленоград, корпус 438 | 45 | NaN |
| | ... | ... | ... | ... | ... | ... | ... | ... |
| | 14468 | 205933 | Кафе «Империя бургеров» | False | кафе | город Москва, город Зеленоград, корпус 1106E | 24 | NaN |
| | 14470 | 220595 | Пекарня № 1 | False | кафетерий | город Москва, город Зеленоград, корпус 1627 | 6 | NaN |
| | 14473 | 220978 | Бар суши TAKAI | False | кафе | город Москва, город Зеленоград, корпус 1602A | 28 | NaN |
| | 14478 | 211676 | Никольский | False | ресторан | город Москва, город Зеленоград, корпус 532 | 120 | NaN |
| | 15184 | 220587 | Пицца Паоло | True | кафетерий | город Москва, город Зеленоград, корпус 1627 | 0 | NaN |

129 rows × 7 columns

В основном остался только Зеленоград. В Зеленограде нет улиц, поэтому предлагаю так и оставить

Постройте график топ-10 улиц по количеству объектов общественного питания. Воспользуйтесь внешней информацией и ответьте на вопрос — в каких районах Москвы находятся эти улицы?

```
In [33]: top_streets = df['street'].value_counts().reset_index().head(10)
top_streets
```

Out [33]:

| | index | street |
|---|-------|------------------------|
| 0 | | проспект Мира |
| 1 | | Профсоюзная улица |
| 2 | | Ленинградский проспект |
| 3 | | Пресненская набережная |
| 4 | | Варшавское шоссе |
| 5 | | Ленинский проспект |
| 6 | | проспект Вернадского |
| 7 | | Кутузовский проспект |
| 8 | | Каширское шоссе |
| 9 | | Кировоградская улица |

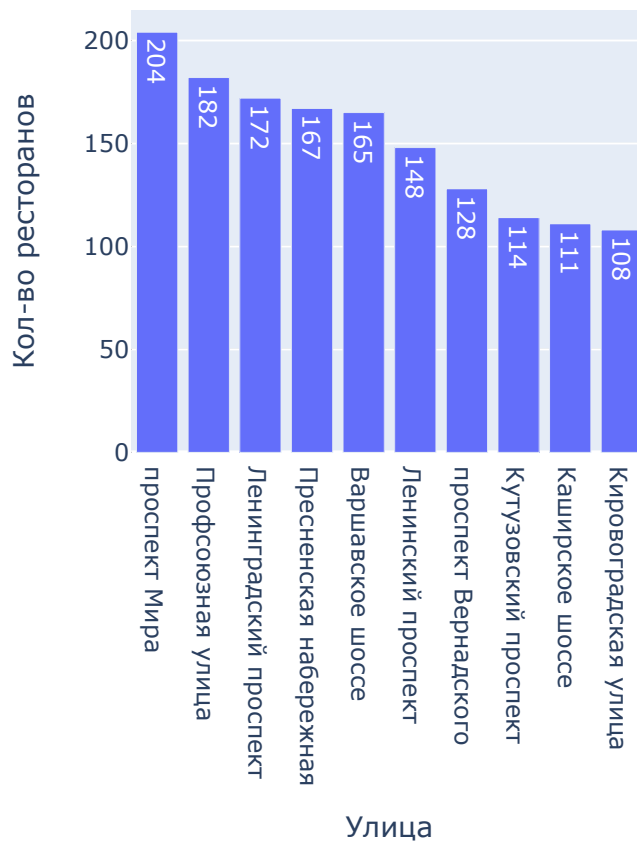
```
In [34]: top_streets = (  
    df['street'].value_counts()  
    .reset_index().head(10)  
    .rename(columns={'index': 'street', 'street': 'count'}))  
top_streets
```

Out [34]:

| | street | count |
|---|------------------------|-------|
| 0 | проспект Мира | 204 |
| 1 | Профсоюзная улица | 182 |
| 2 | Ленинградский проспект | 172 |
| 3 | Пресненская набережная | 167 |
| 4 | Варшавское шоссе | 165 |
| 5 | Ленинский проспект | 148 |
| 6 | проспект Вернадского | 128 |
| 7 | Кутузовский проспект | 114 |
| 8 | Каширское шоссе | 111 |
| 9 | Кировоградская улица | 108 |

```
In [35]: fig_4 = px.bar(top_streets,  
    x='street',  
    y='count',  
    text='count',  
    title='Топ-10 улиц с ресторанами',  
    labels={'count': 'Кол-во ресторанов', 'street': 'Улица'})  
  
fig_4.show()
```

Топ-10 улиц с ресторанами



Больше всего ресторанов находится на Проспекте мира - 204. Также лидерами являются Профсоюзная улица, Ленинградский проспект, Пресненская набережная. В целом почти все улицы из топ-10 - длинные Московские артерии. За счет чего и такое кол-во ресторанов

Подгрузим сторонний датафрейм с информацией о районах Москвы

```
In [36]: try:
          streets = pd.read_csv('mosgaz-streets.csv')
        except:
          streets = pd.read_csv('https://frs.noosphere.ru/xmlui/bitstream/handle')
```

```
In [37]: streets.head()
```

```
Out[37]:
```

| | streetname | areaid | okrug | area |
|---|--------------------------------|--------|-------|-------------------|
| 0 | Выставочный переулок | 17 | ЦАО | Пресненский район |
| 1 | улица Гашека | 17 | ЦАО | Пресненский район |
| 2 | Большая Никитская улица | 17 | ЦАО | Пресненский район |
| 3 | Глубокий переулок | 17 | ЦАО | Пресненский район |
| 4 | Большой Гнездиновский переулок | 17 | ЦАО | Пресненский район |

```
In [38]: top_streets_area = (
          top_streets.merge(streets, left_on='street', right_on='streetname', how='left')
          .drop(columns = ['streetname', 'areaid'])
          )
```

```
In [39]: top_streets_area
```


Out [39] :

| | street | count | okrug | area |
|----|------------------------|-------|-------|-----------------------------|
| 0 | проспект Мира | 204 | СВАО | Алексеевский район |
| 1 | проспект Мира | 204 | СВАО | Ярославский Район |
| 2 | проспект Мира | 204 | СВАО | Район Марьино |
| 3 | проспект Мира | 204 | СВАО | Останкинский район |
| 4 | проспект Мира | 204 | СВАО | Район Ростокино |
| 5 | проспект Мира | 204 | СВАО | Район Свиблово |
| 6 | проспект Мира | 204 | ЦАО | Мещанский район |
| 7 | Профсоюзная улица | 182 | ЮЗАО | Академический район |
| 8 | Профсоюзная улица | 182 | ЮЗАО | Район Черемушки |
| 9 | Профсоюзная улица | 182 | ЮЗАО | Район Ясенево |
| 10 | Профсоюзная улица | 182 | ЮЗАО | Район Коньково |
| 11 | Профсоюзная улица | 182 | ЮЗАО | Обручевский район |
| 12 | Профсоюзная улица | 182 | ЮЗАО | Район Теплый Стан |
| 13 | Ленинградский проспект | 172 | САО | Район Аэропорт |
| 14 | Ленинградский проспект | 172 | САО | Район Беговой |
| 15 | Ленинградский проспект | 172 | САО | Хорошевский район |
| 16 | Ленинградский проспект | 172 | САО | Район Сокол |
| 17 | Пресненская набережная | 167 | ЦАО | Пресненский район |
| 18 | Варшавское шоссе | 165 | ЮАО | Район Чертаново Центральное |
| 19 | Варшавское шоссе | 165 | ЮАО | Район Чертаново Северное |
| 20 | Варшавское шоссе | 165 | ЮАО | Район Чертаново Южное |
| 21 | Варшавское шоссе | 165 | ЮАО | Донской район |
| 22 | Варшавское шоссе | 165 | ЮАО | Район Нагатино-Садовники |
| 23 | Варшавское шоссе | 165 | ЮАО | Нагорный район |
| 24 | Варшавское шоссе | 165 | ЮЗАО | Район Северное Бутово |
| 25 | Варшавское шоссе | 165 | ЮЗАО | Район Южное Бутово |
| 26 | Ленинский проспект | 148 | ЗАО | Район Проспект Вернадского |
| 27 | Ленинский проспект | 148 | ЗАО | Район Тропарево-Никулино |
| 28 | Ленинский проспект | 148 | ЮАО | Донской район |
| 29 | Ленинский проспект | 148 | ЮЗАО | Район Гагаринский |
| 30 | Ленинский проспект | 148 | ЮЗАО | Ломоносовский район |
| 31 | Ленинский проспект | 148 | ЮЗАО | Обручевский район |
| 32 | Ленинский проспект | 148 | ЮЗАО | Район Теплый Стан |
| 33 | Ленинский проспект | 148 | ЦАО | Район Якиманка |
| 34 | проспект Вернадского | 128 | ЗАО | Район Проспект Вернадского |
| 35 | проспект Вернадского | 128 | ЗАО | Район Раменки |
| 36 | проспект Вернадского | 128 | ЗАО | Район Тропарево-Никулино |
| 37 | проспект Вернадского | 128 | ЮЗАО | Район Гагаринский |
| 38 | проспект Вернадского | 128 | ЮЗАО | Ломоносовский район |
| 39 | Кутузовский проспект | 114 | ЗАО | Район Дорогомилово |

| | street | count | okrug | area |
|----|----------------------|-------|-------|---------------------------------|
| 40 | Кутузовский проспект | 114 | ЗАО | Район Фили-Давыдково |
| 41 | Каширское шоссе | 111 | ЮАО | Район Москворечье-Сабурово |
| 42 | Каширское шоссе | 111 | ЮАО | Район Нагатино-Садовники |
| 43 | Каширское шоссе | 111 | ЮАО | Район Орехово-Борисово Южное |
| 44 | Каширское шоссе | 111 | ЮАО | Район Орехово-Борисово Северное |
| 45 | Кировоградская улица | 108 | ЮАО | Район Чертаново Центральное |
| 46 | Кировоградская улица | 108 | ЮАО | Район Чертаново Северное |
| 47 | Кировоградская улица | 108 | ЮАО | Район Чертаново Южное |

Как мы и отмечали ранее, топ улиц составляют длинные артерии Москвы. Поэтому и районов, к которым принадлежит улица - много

Найдите число улиц с одним объектом общественного питания. Воспользуйтесь внешней информацией и ответьте на вопрос — в каких районах Москвы находятся эти улицы?

```
In [40]: low_streets = (
            df['street'].value_counts()
            .reset_index()
            .rename(columns={'index': 'street', 'street': 'count'}))
low_streets = low_streets.query('count == 1')
low_streets
```

```
Out[40]:
```

| | street | count |
|------|---------------------------------|-------|
| 1392 | посёлок Газопровод | 1 |
| 1393 | Пионерская улица | 1 |
| 1394 | Народный проспект | 1 |
| 1395 | улица Кухмистерова | 1 |
| 1396 | Мишина улица | 1 |
| ... | ... | ... |
| 1981 | Ростокинская улица | 1 |
| 1982 | 1-й Сельскохозяйственный проезд | 1 |
| 1983 | Юрловский проезд | 1 |
| 1984 | 1-й Ботанический проезд | 1 |
| 1985 | улица Евгения Родионова | 1 |

594 rows x 2 columns

```
In [41]: low_streets_area = (
            low_streets.merge(streets, left_on='street', right_on='streetname',
                               .drop(columns = ['streetname' , 'areaid'])
            )
low_streets_area
```

Out [41]:

| | street | count | okrug | area |
|-----|---------------------------------|-------|-------|-------------------|
| 0 | посёлок Газопровод | 1 | NaN | NaN |
| 1 | Пионерская улица | 1 | ЗАО | Район Внуково |
| 2 | Народный проспект | 1 | ВАО | Район Измайлово |
| 3 | улица Кухмистерова | 1 | ЮВАО | Район Печатники |
| 4 | Мишина улица | 1 | САО | Савеловский район |
| ... | ... | ... | ... | ... |
| 635 | Ростокинская улица | 1 | СВАО | Район Ростокино |
| 636 | 1-й Сельскохозяйственный проезд | 1 | СВАО | Район Ростокино |
| 637 | Юрловский проезд | 1 | СВАО | Район Отрадное |
| 638 | 1-й Ботанический проезд | 1 | СВАО | Район Свиблово |
| 639 | улица Евгения Родионова | 1 | NaN | NaN |

640 rows x 4 columns

В "городах" Москвы типа Зеленограда, Троицка и тд своеобразная система записи адресов. Из-за этого, в таблицы образовались пропуски

In [42]:

```
(
low_streets_area
.groupby('area', as_index=False)['count'].count()
.sort_values('count', ascending=False).head(10)
)
```

Out [42]:

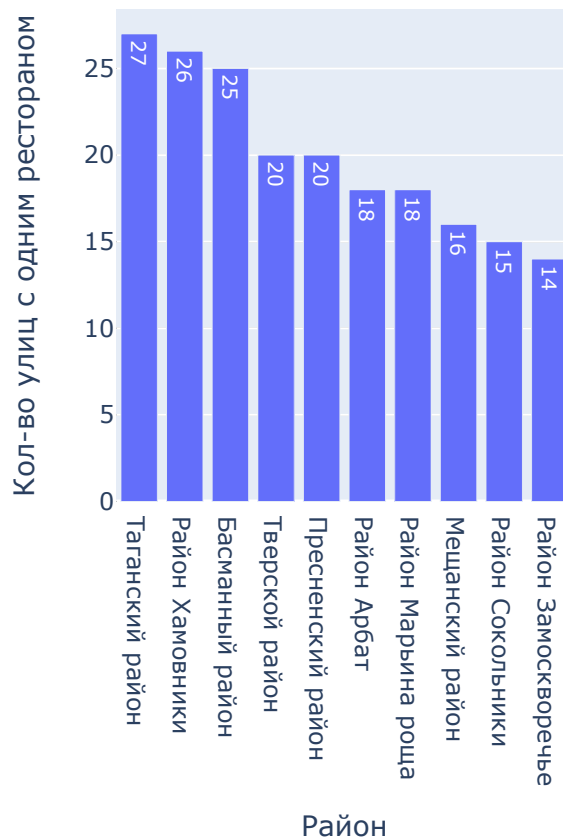
| | area | count |
|----|---------------------|-------|
| 95 | Таганский район | 27 |
| 81 | Район Хамовники | 26 |
| 3 | Басманный район | 25 |
| 96 | Тверской район | 20 |
| 21 | Пресненский район | 20 |
| 22 | Район Арбат | 18 |
| 52 | Район Марьяна роща | 18 |
| 13 | Мещанский район | 16 |
| 75 | Район Сокольники | 15 |
| 34 | Район Замоскворечье | 14 |

In [43]:

```
fig_5 = px.bar((
    low_streets_area
    .groupby('area', as_index=False)['count'].count()
    .sort_values('count', ascending=False).head(10)
),
x='area',
y='count',
text='count',
title='Районы с минимальным кол-вом ресторанов на одной
labels={'count': 'Кол-во улиц с одним рестораном', 'area':
)
```

```
fig_5.show()
```

Районы с минимальным кол-вом рестора-



При всем при этом, достаточно много улиц с одним заведением в центре Москвы. Почти все улицы из нашего "топ" находятся в центре города.

Выводы и рекомендации

Мы провели обширную работу по анализу рынка общественного питания Москвы:

- Ознакомились с данными
- Выполнили предобработку
- Выполнили анализ данных
- Добавили информацию с привлечением внешних источников

И теперь мы готовы дать рекомендации по открытию собственного заведения

В Москве сильно преобладают кафе над остальными видами заведений. Этот вид заведений явно пользуется большой популярностью у горожан

Также в городе в 4 раза больше несетевых заведений.

В основном сетевые заведения характерны для предприятий быстрого обслуживания

Среднее число посадочных мест в Москве – 60. Больше всего посадочных мест в столовых, меньше – в закусочных и кулинарии при магазинах

Были выявлены улицы с самым большим количеством заведений. В основном это длинные Московские улицы.

Улицы с наименьшим кол-вом заведений, наоборот, сосредоточены в центре города

Я бы рекомендовал открывать в Москве именно кафе, тк этот тип заведений явно популярен в городе. Что касается кол-во посадочных мест, то я бы советовал не выделяться от общей массы заведений данного типа и остановиться на отметки около 40 посадочных мест – думаю, это оптимально для обслуживания роботами. Относительно локации – я бы советовал центр города. Высокая стоимость аренды помещения с лихвой окупится из-за большой проходимости в данных районах и относительно небольшом кол-ве заведений на одной улице. Плюсом мы можем стать туристической фишкой города: где еще гостей обслуживают роботы? Также наше заведение сможет сэкономить на персонале, тк обслуживать гостей будут роботы. По поводу развития сети – Москва, на мой взгляд, наиболее хорошо подходит для стартовой точки. В городе нет так много сетевых кафе – всего 22%. Да и в целом развитие заведения с роботами в сеть – хорошая инвестиция, тк мир стоит на пороге роботизации

Также предлагаю вашему вниманию презентацию по проделанной работе <https://drive.google.com/file/d/1u5yuJhmvLDILl2noaWOdSheIH3G9l2tU/view?usp=sharing>