Исследование рынка заведений общественного питания Москвы

Содержание

- 1 Импорт библиотек и определение функций проекта
 - 1.1 Импорт библиотек
 - 1.2 Функции загрузки и обзора данных
 - 1.3 Функции предварительной обработки
- 2 Обзор данных
 - 2.1 Вывод
- 3 Предварительная обработка данных
 - 3.1 Вывод
- 4 Анализ данных
 - 4.1 Сравнение количества заведений общественного питания разных форматов
 - 4.2 Сравнение количества сетевых и несетевых заведений общественного питания
 - 4.3 Определение видов объектов общественного питания, для которых характерно сетевое распространение
 - 4.4 Изучение соотношения количества сетевых заведений и количества мест в них
 - 4.5 Оценка минимального, максимального и среднего количества посадочных мест для различных видов объектов общественного питания
 - 4.6 Анализ распределения заведений общественного питания по улицам Москвы
 - 4.6.1 Выделение улицы из адреса заведения
 - 4.6.2 Определение топ-10 улиц по количеству объектов общественного питания и районов Москвы, в которых они расположены
 - 4.6.3 Определение районов Москвы, в которых расположены улицы с единственным объектом общественного питания
 - 4.7 Вывод
- 5 Общий вывод исследования и рекомендации
- 6 Презентация для партнёров

Прежде чем запускать дорогостоящий проект в Москве - кафе с роботизированным обслуживанием - требуется провести исследование рынка.

Исследование проводится на основе открытых данных о заведениях общественного питания в Москве.

Требуется:

- 1. Определить текущее положение дел на рынке.
- 2. Спронозировать популярность заведения на длительный срок.

Данные хранятся в файле rest_data.csv

В ходе работы предполагается:

- загрузить и подготовить данные;
- исследовать соотношение видов объектов общественного питания по количеству;
- исследовать соотношение сетевых и несетевых заведений по количеству;
- определить, для какого вида объекта общественного питания характерно сетевое распространение;
- определить, что характерно для сетевых заведений: много заведений с небольшим числом посадочных мест в каждом или мало заведений с большим количеством посадочных мест;
- описать среднее количество посадочных мест для каждого вида объекта общественного питания;
- обогатить данные из открытых источников названиями районов Москвы;
- проанализировать районы с большой и малой плотностью объектов питания;
- сделать выводы и сформулировать рекомендации по стратегии создания кафе с роботизированным обслуживанием.

Результаты исследования необходимо оформить в виде презентации для партнёров в формате PDF.

Импорт библиотек и определение функций проекта

Импорт библиотек

```
In [1]: import os
import pandas as pd
import numpy as np
from plotly import graph_objects as go
import matplotlib.pyplot as plt
import seaborn as sns
import requests
from urllib.parse import urlencode
```

Функции загрузки и обзора данных

```
elif os.path.exists(pth2):
               return pd.read_csv(pth2, sep=sep)
            else:
               print("ERROR: Neither Yandex nor local path is reacable...")
            return pd.DataFrame() # в случае ошибки чтения вернём пустой DataFrame
In [3]: # определение функции обзора данных
                                         ._____
         .
# на вход подаётся датафрейм df
        # на выходе:
             - 10 случайных строк df
             - информация df.info()
- количество явных дубликатов в строках df
             - процент пропусков данных в столбцах df
        {\tt def} \ {\tt data\_observe(df)} \colon
                          # количество отображаемых строк таблицы
            row_num = 5
            print('Произвольные строки таблицы:')
            display(df.sample(row_num))
            else:
               display(df)
            print('\nИнформация о таблице:')
            print('=====')
df.info()
            print('\nКоличество явных дубликатов в таблице:')
            ._____')
            print('\nПроцент пропусков в столбцах:')
            .background_gradient('coolwarm')
In [4]: # определение функции загрузки данных с Яндекс.Диска
         .
# на вход подаётся:
              file_url - публичный URL файла на Я.Диске параметры для read\_csv
        # на выходе - датафрейм с загруженными данными
        def open_file_from_yadisk(file_url, sep=',', encoding='utf-8'):
            # используем арі
            base_url = 'https://cloud-api.yandex.net/v1/disk/public/resources/download?'
            public_key = file_url
            # получаем url
            final_url = base_url + urlencode(dict(public_key=public_key))
response = requests.get(final_url)
download_url = response.json()['href']
            # загружаем файл в df
            download_response = requests.get(download_url)
            df = pd.read_csv(download_url, sep=sep, encoding=encoding)
```

Функции предварительной обработки

```
In [5]: def str_to_bool(value):
                corr_dict={'ma':True, 'met':False}
return corr_dict[value]
```

Обзор данных

Откроем и изучим файл rest_data.csv :

```
In [6]: rest_data = open_file('rest_data.csv', sep=',')
        if not rest_data.empty:
           data_observe(rest_data)
```

Произвольные строки таблицы:

	id	object_name	chain	object_type	address	number
6050	76495	Чайхона Ходжент	нет	кафе	город Москва, улица Миклухо-Маклая, дом 8	40
8867	27866	СТОЛОВАЯ ГЕНШТАБА	нет	столовая	улица Знаменка, дом 14	90
7178	20637	Буфет в «БОК на Дубнинской»	нет	буфет	город Москва, Дубнинская улица, дом 6	6
4569	29260	Столовая 894	нет	столовая	город Москва, Ферганский проезд, дом 10, корпус 4	200
3802	21001	БУФЕТ	нет	буфет	город Москва, Погодинская улица, дом 1, строен	16

```
Информация о таблице:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15366 entries, 0 to 15365
Data columns (total 6 columns):
# Column Non-Null Count Dtype
                          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
5 number 15366 non-
dtypes: int64(2), object(4)
memory usage: 720.4+ KB
Количество явных дубликатов в таблице:
Процент пропусков в столбцах:
                 NaNs, %
            id
 object_name
        chain
  object_type
```

Таблица rest_data состоит из 15366 строк, 6 столбцов. Типы данных в столбцах - object, int64.

Согласно описанию данных, столбцы хранят следующую информацию:

- id идентификатор объекта;
- object_name название объекта общественного питания;
- chain сетевой ресторан;
- object_type тип объекта общественного питания;
- address адрес;

address number

• number — количество посадочных мест.

Столбцы поименованы в хорошем стиле snake_case.

В таблице, предварительно, отсутствуют явные дубликаты и пропуски данных.

Вывод

Данные, на первый взгляд, достаточно качественные.

На этапе предварительной обработки целесообразно:

- изменить тип столбца chain на булевский:
- изучить столбец object_name на наличие неявных дубликатов в названиях точек питания;
- изучить состав значений столбца object_type;
- изучить столбцы number и id на предмет возможности сокращения размерности чисел до int32 в целях экономии памяти.

Предварительная обработка данных

Изменим тип столбца chain на булевский. Для этого сперва убедимся, что он может принимать только два значения:

```
In [7]: rest_data.chain.sort_values().unique()
 Out[7]: array(['да', 'нет'], dtype=object)
         Действительно, поле принимает только значения "да" и "нет". Преобразуем к двоичному типу:
 In [8]: rest_data.chain = rest_data.chain.apply(str_to_bool)
         rest_data.chain.dtype
Out[8]: dtype('bool')
         Итак, мы имеем 2968 сетевых точек питания:
 In [9]: rest_data.chain.sum()
Out[9]: 2968
         Перейдём к изучению значений в столбце object_name :
In [10]: print(list(rest_data.object_name.sort_values().unique()))
```

Moscow foco service market research.

(1-14) Moscow foco service m

Moscow food service market research

Tr, Sciolit LOVT, 'Solit MollD', 'Solit Gyer', 'Solit DILL', 'Solit Mill', 'S ENDET IMI, 197, 199ET IMI, 1971, 199ET IMI, 1971, 199ET IMI, 1971, 199ET IMION 127, 199ET IMION 127, 199ET IMION 127, 199ET INDIA 1872, 199ET

Moscow food service market research

райс', 'Бургер MIX', 'Бургер Кинг', 'Бургер Кинг Burger King', 'Бургер Клаб', 'Бургер Пицца', 'Бургер Рус', 'Бургер Сити Клаб', 'Бургер шеф', 'Бургер герой пивной бар', 'Бургер клаб', 'Бургерная "Кургерная «Нідһ Ground Киберспорт Лаунж» Интернет-кафе, Кальян-бар', 'Бургерная «Хищник. Стейк & Б ургерна», 'Бургерная МЕАТ лав', 'Бургерная ФАРШ', 'Бургерная Хищник', 'Бургерная Электроугли', 'Бургеры', 'Бургеры и Стейки', 'Бургерная «Хищник', 'Бургерная "Кургерная Кургерная "Кургерная Куштик', 'Бургерная "Кургерная Куштик', 'Бургерная "Кургерная Куштик', 'Бургерная Суштик', 'Бургерная Суштик, Суштик', 'Бургет Кому Суштик Суштик, Суштик, Суштик Суштик, Суштик, Суштик Суштик, Суштик, Суштик, Суштик Суштик, Суштик, Суштик, Суштик, Суштик, Суштик ОРОДА МОСКВЫ ШКОЛА № 853', 'ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБЩЕОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ГОРОДА МОСКВЫ ШКОЛА № 1528', 'ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБЩЕОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ГОРОДА МОСКВЫ ШКОЛА № 1528', 'ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБЩЕОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ГОРОДА МОСКВЫ ШКОЛА № 1528', 'ГОТОВ', 'ГОУ ВПО РОССИЙСКИ И ГОСУДАРСТВЕННОЕ БИДЖЕТНОЕ ОБЩЕОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ГОРОДА МОСКВЫ ШКОЛА № 1528', 'ГООВ', 'ГОУ ВПО РОССИЙСКИ И ГОСУДАРСТВЕННОЕ БИДЖЕТНОЕ ОБЩЕОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ГОРОДА МОСКВЫ ШКОЛА № 1528', 'ГООВ', 'ГОУ ВПО РОССИЙСКИ И ГОСУДАРСТВЕННЫЕ ИЗБОЕ, 'ГОУ СПО КОЛЛЕДЖ', 'ГОУ СПО К

POPM

Moscow food service market research

pop. Ca2, ' Topoacea Crysto Britains', Topoacea nonskinnins 229', 'Topoacea crysto arrivet and the pop. 'Topoacea service market research

pop. Ca2, ' Topoacea Crysto Britains', Topoacea nonskinnins 229', 'Topoacea service market', 'Topoacea service was news, sper-poins', 'Topoacea service was news, sper-poins', 'Topoacea service was news, 'Topoacea service, 'Topoa | 1.235. "SATOPSE BX. 942." "340TOPE BX. 948." "340TOPE BX. 187." "340XAGAGASE", "340TAM SADO." "340XAGAGASE", "340TAM SADO." SADOTAM SADO." SOUTH Moscow food service market research

yn', 'Kananu kryd', 'Kankackas kywie', 'Kankacakan nneuwaja', 'Kanani', 'Kanani Kanan', ' фе «Вейк ал»", "Кафе «Венахи»", "Кафе «Венаки»", "Кафе «Веранда»", "Кафе «Верона»", "Кафе «Вокомания»", "Кафе «Вокосточныя уктафи «Вокомания»", "Кафе «Вокосточныя уктафи «Вокосточныя уктафи «Вокомания»", "Кафе «Покоточныя уктам»", "Кафе «Деревяма»", "Кафе «Каревер от", "Кафе «Малима", "Кафе «Марк МИА»", "Кафе «Марко дворик»", "Кафе «Маркода»", "Кафе «Мартили»", "Кафе «Маркода»", "Кафе «Меду делом»", "Кафе «Маркод»", "Кафе «Меду делом»", "Кафе «Маркод»", "Кафе «Каркод»", "Кафе «Сар

ОРОВ»', 'Кафе «Царица»', 'Кафе «Цинандали»', 'Кафе «Циники»', 'Кафе «Цитадель»', 'Кафе «Циккини»', 'Кафе «Чайхана Айва»', 'Кафе «Чайхана Тапчан»', 'Кафе «Ча айхана»', 'Кафе «Чеминон»', 'Кафе «Чистая линия»', 'Кафе «Шашлычный дворик»', 'Кафе «Шарар»', 'Кафе «Шарар»', 'Кафе «Шокупадница»', 'Кафе «Шокупадница»', 'Кафе «Штолле»', 'Кафе «Варц кайзер»', 'Кафе «Вокорадница»', 'Кафе «Вокорадна»', 'Кафе Арарат (при гостинице «Арарат Парк Хаятт»)', 'Кафе Аршонн', 'Кафе Аршоны А.А.', 'Кафе Бургсу Кафе Бургсу Кинг', 'Кафе Бургарат (при гостинице «Арарат Парк Хаятт»)', 'Кафе Аршоны А.А.', 'Кафе Бурбосноь', 'Кафе Бургся Кинг', 'Кафе Бордко,', 'Кафе Бордко,', 'Кафе Бордко,', 'Кафе Бордко,', 'Кафе Бордко,', 'Кафе Борсонь', 'Кафе Боксор', 'Кафе Борсонь', 'Кафе Боксор', 'Кафе Борсонь', 'Ка adan 244 ", Yade - 20x06yrc", Yade - 30xces", Yade - 30xces", Yade - 40xcesses Treed втосалоне Вольво на Тульской', 'Кафетерий на заправке', 'Кафетерий при АЗК 317', 'Кафетерий при АЗС', 'Кафетерий, 'Кафетерий, 'Кафетерий, 'Кафетерий, 'Кафетерий, 'Кафетерий, 'Кафишка', 'Кафк «Качевник»', 'Кафк ПИАЛА', 'Каффа', 'Каффетерий, 'Кафетерий, 'Качаторе', 'Кашаторе', 'Кашаторе', 'Кашаторе', 'Кашаторе', 'Кашамалаша', 'Каши от Аркаши', 'Кашатор', 'Квартира 44', 'Квартира № 5', 'Квартира № 7', 'Кворум', 'Кебаб 24', 'Кебаб 24', 'Кебаб Каус', 'Кинаа', ' -век', 'Колос пекарня Осетинские пироги', 'Колхи', 'Колхида', 'Комбинат дошкольного питания', 'Комбинат питаная ВАГШ', 'Комбинат питания мГТУ им. Н.Э. Баумана', 'Комбинат питания МПТХТ «ИНФРАСТРУКТУРА Холдинг»', 'Комбинат питания марии', 'Комбинат хлебопродуктов', 'Ком еди кафе', 'Комета', 'Комета бар', 'Комменти ресторан', 'Компасс групп рус', 'Комплексное предприятие', 'Компот', 'Кому жить хорошо', 'Конак', 'Кондитерская "Кондитерская", 'Кондитерская "Кондитерская "Кондитерская "Кондитерская "Кондитерская "Кондитерская "Кондитерская "Кондитерская «Констанци я»', 'Кондитерская «Кузина»', 'Кондитерская «Пряник.Рогалик»', 'Кондитерская Варовая Индустрия', 'Кондитерская Марта', 'Кондитерская фермак "Кондитерская кондитерская обртутти', 'Кондитерская кафе Пушкины', 'Кондитерская лавка Александра»', 'Кондитерская пекарня', 'Кондитерская кафе "Прижины', 'Кондитерская лавка Александра»', 'Кондитерская пекарня, 'Кондитерская пекарня Волконский', 'Кондитерская лавка Александра»', 'Консерватория им. П.И. Чайковского', 'Коперник', 'Коптильня', 'Корнетская кухня', 'Корейский ресторан', 'Корейское бистро «И нсам»', 'Коржовь', 'Корифей', 'Корима', 'Корнфиза Диканька', 'Корнфиза Диканька', 'Корнма Тарас Бульба', 'Корн доги', 'Корнетарас Бульба', 'Корнма Диканька', 'Корема Тарас Бульба', 'Корнма "Корема Карэль', 'Кор чолета, 'Кофе', 'Кот-Ч', 'Коте Азур', 'Котет ет', 'Кофе Баристо', 'Кофе Бин', 'Кофе Бин', 'Кофе Бир о', 'Кофе Еда', 'Кофе Еда', 'Кофе Еда', 'Кофе Баристо', 'Кофе Вин', 'Кофе Бир о', 'Кофе Еда', 'Кофе Варисто', 'Кофе Моментс', 'Кофе Каус', 'Кофе Паб', 'Кофе Премиум', 'Кофе Вынтос', 'Кофе Стайл', 'Кофе в Тамла, 'Кофе и пирожок', 'Кофе Ваннос', 'Кофе Памлава', 'Кофе ос собой «Соаби «Со offee Point»', 'Кофе с собой «Кофейня»', 'Кофе с собой «Сладкая карамель»', 'Кофе с собой «Тоа Кофе»', 'Кофе с собой Пирогово', 'Кофе с собой и кулинария', 'Кофе тун', 'Кофе-Бар', 'Кофе-Бар', 'Кофе-бар', 'Кофе-бар', 'Кофе-бар', 'Кофе тун', 'Кофе кики "Кофектор", 'Кофеварня', 'Кофейня (Кофейна Кофеин), 'Кофейна (Кофейна, 'Кофейна Кофеварня', 'Кофейна Кофейна Кофей Кофей Семеания Кофейна фейня ЦДМ', 'Кофемаус', 'Кофемоверс', 'Кофемания Жоколадница', 'Кофемания', 'Кофемания', 'Кофемания', 'Кофемания', 'Кофемания', 'Кофемания', 'Кофемания', 'Кофемания', 'Кофемания', 'Кофетерий «Кондитерский дом Александра Селезнева»', 'Кофетерий «Кофетерий', 'Кофетерий «Кондитерский дом Александра Селезнева»', 'Кофетерий «Кофетерий', 'Кофетерий "Кофецевт', 'Кофешоп Соffeeshop Сотралу', 'Кофетерий "Кофетерий "Кофетерий", 'Крабта", 'Крабта", 'Крабта", 'Крабта", 'Краснодар', 'Крафт & Бургер', 'Крафт Бар', 'Крафт Стейшн', 'Крафт-Паб «69 Pints»', 'Крафт-бар Колбаса', 'Крафтмен', 'Крафтмен', 'Крафтово пиво', 'Крафтовый бар «Кампания»', 'Крафтовый кофе-бар', 'Креветки', 'Крейзи Дейзи', 'Кренделькофф', 'Крепери де Пари', 'Крепери де пари', 'Кровая', 'Кр рошка Картошка', 'Крошка Картошка', 'Крошка Картошка', 'Крошка Картошка', 'КрумасанКафе', 'Крулый кот', 'Кружка', 'Круиз', 'Крылы шко-ножка', 'Крумьа', 'Куулый кот', 'Кружка', 'Кууза', 'Крылы шко-ножка', 'Кулинария", 'Кулинария", 'Кулинария", 'Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария "Кулинария "Кулинария "Кулинария Три сестры', 'Кулинария Фасоль', 'Кулинария пан Пан', 'Кулинария Три сестры', 'Кулинария", 'Кулинария", 'Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария "Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария "Кулинария "Кулинария "Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария", 'Кулинария "Кулинария "Кулинария", 'Кулинария "Кулинария", 'Кули

Mose and ophieses. "Symmyton research for Pair." "Symmetres areas through "Symmyton research Symmetry (Symmyton research Symmyton Symmyton

Moscow food service market research

balant imagene story coulders in freezpoin disystemation in Therapies demonstrated audits in financial exploration in the process and process of the PECTOPAM «Tanysum», "PECTOPAM (Tanysum», "PECTOPAM (Tanysum»), "PECTOPAM (Tanysum», "Panysum», "Pan и»', 'Ресторан «Белый город»', 'Ресторан «Белый дом»', 'Ресторан «Биджо»', 'Ресторан «Бирхоф»', 'Ресторан «Бистро»', 'Ресторан «Борисовский пассаж»', 'Ресторан «Ботиран «Ботиран»', 'Ресторан «Ботиран»', 'Ресторан «Ботиран»', 'Ресторан «Брудер»', 'Ресторан «Буйабез»', 'Ресторан «Бульвар»', 'Ресторан «Бургер Кинг»', 'Ресторан «Балап Паста»', 'Ресторан «В старом городе»', 'Ресторан «ВЕТАР»', 'Ресторан «ВЕТЕРНИЙ БАКУ»', 'Ресторан «Винотема «Ветиран»', 'Ресторан «Бильчиран»', 'Ресторан «Бильчиран»', 'Ресторан «Бильчиран «Ботиран»', 'Ресторан «Гариран»', 'Ресторан «Гариран»', 'Ресторан «Гариран»', 'Ресторан «Гриль и вино»', 'Ресторан «Грузин»', 'Ресторан «Грузинская кухн я»', 'Ресторан «ДЖОРДЖИЯ»', 'Ресторан «ДНК»', 'Ресторан «ДОНН»', 'Ресторан «ДОННА КЛАРА»', 'Ресторан «Де Пари»', 'Ресторан «Де Пихто»', 'Ресторан «Джорджино»', 'Ресторан «Деторан «Дихто»', 'Ресторан «Дамардино "Донамардино", 'Ресторан «Деторан «Деторан»', 'Ресторан «Деторан «Деторан»', 'Ресторан «Дихто»', 'Ресторан «Дихто»',

Моском food service market research

Да Пино»', 'Ресторан «Джон Джоли»', 'Ресторан «Джон Сильвер»', 'Ресторан «Дичь»', 'Ресторан «Добрыня»', 'Ресторан «Дам на Патриарших»', 'Ресторан «Домжу р»', "Ресторан «Домино': Пища»", 'Ресторан «Донна Маргарита»', 'Ресторан «Дружба»', 'Ресторан «Двасма Хо»', 'Ресторан «Срам»', 'Ресторан «Донна Маргарита»', 'Ресторан «Дружба»', 'Ресторан «Дружба»', 'Ресторан «Дамасма Хо»', 'Ресторан «Дружба»', 'Ресторан «Дру рия»', 'Ресторан «Тики Бар»', 'Ресторан «Токио»', 'Ресторан «Трио»', 'Ресторан «Тропарево»', 'Ресторан «Тосторан «Уроко»', 'Ресторан «Усадьба»', 'Ресторан «Фокачерия»', 'Ресторан «Фокачерия»', 'Ресторан «Фокачерия»', 'Ресторан «Хинкальи. Вино. Мимино»', 'Ресторан «Хинкальи. Вино. Мимино»', 'Ресторан «Хинкальи. Вино. Мимино»', 'Ресторан «Хинкальи. Вино. Мимино»', 'Ресторан «Царъград»', 'Ресторан «Даръград»', 'Ресторан Каръград»', | «Ποτορια «Πακονα» (Σε', "Pecropa» «ΠΕΣΠΑΙΔΙΑΤ» (ΤΟΘΕΑ» ("Pecropa» «ΠΕΣΠΑΙΑ"), "Pecropa» «ΠΕΣΠΑΙΑ", "Pecropa» «ΠΕ OBAR OBK «ЭКОПОЛИС» № 1288', 'СТОЛОВАЯ ПОИ МЕТОДОВАЯ ПОИ МЕТОДОВАЯ ПОВК «ЭКОПОЛИС» № 1583', 'СТОЛОВАЯ ПОВ «ВТОДОРОЖНОМ КОЛЛЕДЖЕ", 'СТОЛОВАЯ ПРИ АВТОСАЛОНЕ «АВТОМИР»', 'СТОЛОВАЯ ПРИ АО ВНИИЗТО', 'СТОЛОВАЯ ПРИ АТОМЭНЕРГОПРОЕКТЕ", 'СТОЛОВАЯ ПРИ ВОЕНТОРГБАЗ ФКП «5113 ЦТЗБ»', 'СТОЛОВАЯ ПРИ ГБОУ ШКОЛА № 1503', 'СТОЛОВАЯ ПРИ ГИЗНАЗИИ МПС РФ', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1664', 'СТОЛОВАЯ ПРИ ГОУ ГОИ № 1452', 'СТОЛОВАЯ ПРИ ГОУ ГОШ № 1455', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1464', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1464', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1464', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1465', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1455', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1465', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1455', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1465', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1455', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1504', 'СТОЛОВАЯ ПРИ ГОУ СОШ № 1505', 'СТОЛОВАЯ ПРИ МЕМЬТЬТЬЕ МЕМЬТЬЕ М А, СТОЛОВАЯ ПРИ МОСКОВСКОМ ЭЛЕКТРОМЕХАНИЧЕСКОМ КОЛЛЕДЖЕ МЭБ:, "СТОЛОВАЯ ПРИ МОСТЬАНКЕ", "СТОЛОВАЯ ПРИ МОСКОВСКОМ ЭНДОКРИННОМ ЗАВОДЕ", "СТОЛОВАЯ ПРИ НИИАА ИМ. СЕМЕНИХИНА", "СТОЛОВАЯ ПРИ НУ АКУШЕРСТВА И ГИНЕКОЛОГИИ", "СТОЛОВАЯ ПРИ ОАО «ЗАВОДЬНЫ", "СТОЛОВАЯ ПРИ ОАО «КАЛИБРЬ", "СТОЛОВАЯ ПРИ ОАО «КАЛИБРЬ", "СТОЛОВАЯ ПРИ ОАО «КАЛИБРЬ", "СТОЛОВАЯ ПРИ ОАО «КАЛИБРЬ", "СТОЛОВАЯ ПРИ ОАО МОСГИПРОТРАНЕ", "СТОЛОВАЯ ПРИ ОАО МОСГИПРОТРАНЕ", "СТОЛОВАЯ ПРИ ОВОМЕТЬ В ДЕТЕРАНОВ ТРУДА N 31", "СТОЛОВАЯ ПРИ ОВОМЕТЬ В ДЕТЕРАНОВ ТРУДА N 31", "СТОЛОВАЯ ПРИ ОВОМЕТЬ В ДЕТЕРАНОВ ТРУДА N 6", "СТОЛОВАЯ ПРИ ПОЛИГРАФИЧЕСКОМ КОЛЛЕДЖЕ", "СТОЛОВАЯ ПРИ ОВОМЕТЬ В ДЕМЕТЬ В ДЕТЕРАНОВ ТРУДА N 6", "СТОЛОВАЯ ПРИ ОВОМЕТЬ "СТОЛОВАЯ ПРИ ОВОМЕТЬ "СТОЛОВАЯ ПРИ ОВОМЕТЬ "СТОЛОВАЯ ПРИ СТОЛОВАЯ ПРИ СПОТИВНЕННОМ КОЛЛЕДЖЕ", "СТОЛОВАЯ ПРИ СТ. ТОВАРНАЯ-ЯРОСЛАВСКАЯ", "СТОЛОВАЯ ПРИ СТРОИТЕЛЬНОМ КОЛЛЕДЖЕ

19 PM

NOSCOW food Service market research

N12*, 'CTONDBAR IPIU TSU-8*, 'CTONDBAR IPIU YBGL CAND', 'CTONDBAR IPIU YNCH WOCCOPTERIO', 'CTONDBAR IPIU WIN-11 BOC', 'CTONDBAR IPIU WIN-12*, 'CTONDBAR IWI WIN-12*, 'CTO N12', 'СТОЛОВАЯ ПРИ ТЭЦ-8', 'СТОЛОВАЯ ПРИ УВД СЗАО', 'СТОЛОВАЯ ПРИ УПКИ МОСГОРТЕПЛО', 'СТОЛОВАЯ ПРИ УПП-11 ВОС', 'СТОЛОВАЯ ПРИ УПРАВЛЕНИИ СУДЕБНОГО ДЕПАРТ с. "СТОЙОВАЯ-ДОГОТОВОЧНАЯ ПРИ ШКОЛЕ W 237", 'СТОЛОВАЯ-ДОГОТОВОЧНАЯ ПРИ ШК. 289", 'СТОЛОВАЯ-РАЗДАТОЧНАЯ ПРИ ШК. 289", 'СТОЛОВАЯ-РАЗДАТОЧНАЯ ПРИ ШК. 289", 'СТОЛОВАЯ-РАЗДАТОЧНАЯ ПРИ ШК. 289", 'СТОЛОВАЯ-РАЗДАТОЧНАЯ ПРИ ШК. 389", 'СТОИТЕЛЬНЫЙ КОЛЛЕДЖ', 'СТУШИ МОК", 'СУШИ МОК", 'СУШИ МОК", 'СУМИОТ", 'СУМОСТ", 'СУМИСТО", 'СУМИ СТОР", 'СУШИ СОР", 'СУШИ КОР", 'СУШИ КОР", 'СУШИ КОР", 'СУШИ МОК", 'СУШИ КАР", 'СУШИ МОК", 'СУШИ КАР", 'СУШИ КАР", 'СУШИ КОР", 'СОВОРА ", ' рая усадьба', 'Старбагс', 'Старбакс', 'Старбакс Starbucks', 'Старбакс Кофе', 'Стардаст кофе-кофе с собой', 'Стардог5', 'Стард a»', 'Столовая «Саntina City»', 'Столовая «Јust Kitchen»', 'Столовая «АНОКПУМГКЭМТ»', 'Столовая «АПЕЛЬСИН»', 'Столовая «АРТ-столовая»', 'Столовая «ВНИИХ Т»', 'Столовая «Ватрушка»', 'Столовая «Вытушка»', 'Столовая «Гьоу сош 1158»', 'Столовая «Гьоу сош 1179»', 'Столовая «Гьоу сош 1623»', 'Столовая «Гьоу сош 851»', 'Столовая «Гьоу сош 856»', 'Столовая «Гьоу сош 1623»', 'Столовая «Гьоу сош 851»', 'Столовая «Гьоу сош 856»', 'Столовая «Гьоу сош 1179»', 'Столовая «Гьоу сош 1623»', 'Столовая «Гьоу сош 856»', 'Столовая «Гьоу школа №1692»', 'Столовая «Гьоу школа №1692»', 'Столовая «Гьоу школа №1692»', 'Столовая «Гьоу школа №1528»', 'Столовая «Гьоу школа №1528»', 'Столовая «Гьоу школа №153»', 'Столовая «Гьоу сош №1749»', 'Столовая «Гьоу сош №1749»', 'Столовая «Гоу сош №1749»', 'Столовая «Гоу сош №1749»', 'Столовая «Гоу сош №1748»', 'Столовая «Гоу сош №1749»', 'Столов ам «Гостиница МЗП»", "Столовая «ГринЛондон»", "Столовая «Дежене»", "Столовая «Дежене»", "Столовая «Дежене»", "Столовая «Кастрюлька»", "Столовая «Компоная «Кастрюлька»", "Столовая «Компоная «Кастрюлька»", "Столовая «Компоная «Компона»", "Столовая «Компоная «Компона»", "Столовая «Московский исихолого-педагогический колледж»", "Столовая «Профессиональный Заз», "Столовая «Правда»", "Столовая «Правда»", "Столовая «Профессиональный заз», "Столовая «Правда»", "Столовая «Столовая «Столова»," "Столовая «Столовая «Обра»,", "Столовая «Обра»," 50 / COU № 1538', 'CTOЛОВАЯ ГБОУ COU № 1544', 'CTОЛОВАЯ ГБОУ COU № 1564', 'CTОЛОВАЯ ГБОУ COU № 1571', 'CТОЛОВАЯ CБОУ COU № 1571', 'CТОЛОВАЯ

6 1.1', 'Столовая ГБОУ Школа № 1101 здание 1.3', 'Столовая ГБОУ Школа № 1101 здание 2.1', 'Столовая ГБОУ Школа № 1298', 'Столовая ГБОУ Школа № 1387', 'Столовая ГБОУ Школа № 1507', 'Столовая ПБОУ Школа № 1507', 'Сто укрово 1828 кдлп - 1, "столовая Що № 1798", "столовая Що 1298", "столовая Цо 1298", "столовая Боли 1299", "столовая Боли МСХА им. К.А. Тимирязева', 'Столовая №15', 'Столовая №21', 'Столовая №8, при ГКНПЦ ИМ.ХРУНИЧЕВА М.В', 'Столовая, кулинария', 'Столовая-Буфет', 'Столовая-Пе кария', 'Столовая-Пельменная', 'Столовая-Бар Хутор', 'Столовая-Кария', 'Столовая-Кария', 'Столовая-Пельменная', 'Столовая-Бар Хутор', 'Столовая-Кария', 'Столовая-Карии', 'Страна Которой нет', 'Стрелка', 'Стрит Флеш', 'Стрит Еда.', 'Строительный колледж № 26 СП 6', 'Строительный колледж № 26, СП 2', 'Строительный техникум', 'Студенте (закрытое)', 'Суши Мак е', 'Суши Карсс', 'Суши Карсс', 'Суши «Вооте', 'Суши Карсс', 'Суши Карсо', 'Суши Карсс', 'Суш 'ΤΑΚ', 'ΤΑΚ', 'ΤΑΚ' ΘΟ', 'ΤΑΜΩΑΛ, 'ΤΑΜΟΧΗЯ ДΑΕΤ ДΟБРО', 'ΤΑΗ', 'ΤΑΗΙΜΡ', 'ΤΑΤΙΤΕΡ ΚΠΑΕ', 'ΤΑΙΜΡ ΠΝΙΙΙΑ', 'ΤΕΜΠΚ (ΜΑΘΕ «ΦΕ-СТИВАЛЬНОΘ» БЛИНЬ"), 'Τ

Д ГУМ Кафе 1 Гастроном 1', 'ΤД ГУМ Кафе 2 Гастроном 1', 'ΤД ГУМ Кафе фестивальное БАР КИНОЗА", 'ΤД ГУМ Кафе 1 Гастроном 1', 'ΤД ГУМ Кафе 2 Гастроном 1', 'ТД ГУМ Кафе фестивальное БАР КИНОЗА", 'ТЕМЕ, 'ТЕРЕМОК', 'ТЕКАС-ИКЕН', 'ТЕКНОЛОГИИ РЕСТОРАННОГО БИЗНЕСА', 'ТЕКНОЛОГИИ ВАСТОЛОВЫХ УПРАВЛЕНИЯ МИНОБОРОНЫ РФ', 'ТРИУМО', 'ТОКИО ВА ВУ', 'ТОМ ТЭЙЛОР РУС', 'ТОМ ЯНГ', 'ТОТОРО', 'ТОЧКА ОТРЫВА', 'ТРАКТИР ВАГАНТ', 'ТРАКТИР У ТИМОВЕЕЧА', 'ТРАТОРИЯ ІТАLIAN «ПаставЛицца»', 'ТРАТТОРИЯ', 'ТЕДЕТ ВО ВОВЫХ УПРАВЛЕНИЯ МИНОБОРОНЫ РФ', 'ТРИУМО', 'ТОКИО ВА ВУ', 'ТОМ ТЭЙЛОР РУС', 'ТОМ ЯНГ', 'ТОТОРО', 'ТОЧКА ТОТОРИМОТЬ ВАТОРИМОТЬ ВА

Moscow food service market research

ex', 'Утёс', 'Ухтомская усадьба', 'Ухты-Пухты', 'Уют', 'Уют Востока', 'Ф-КА ГОСЗНАК', 'Ф-Л 17АВТ.ПАРК. ГУП «Мосгортранс»', 'ФА-СОЛЬ-КА', 'ФАКУЛЬТЕТ ПЕДАГОГ ИКИ МГПУ', 'ФАНТОЦЦИ РУС', 'ФАРЕНГЕЙТ', 'ФАРЕН', 'ФАРЕН-МЕТОВО ЗАВЕСТИЯ УПД РФ', 'ФТОР МЕРИД-МОРДЫВЫ ПОЧТАЯТНТ', 'ФТОР УАТ МИНРО ПО АТОМНОЙ ЭНЕРГИИ', 'ФЕР РГАНА', 'ФИЛИАЛ ШБС ЛЮБЛИНО ШК. 1819', 'ФОРМА БИСКОСТИТ', 'ФИРМА АВРЕЛЬ', 'ФИРМА БИСКОСТИТ', 'ФИРМА ВИСКОСТИТ', 'ФИРМА ВИСКОСТИТ', 'ФОРМА ВЕС ЛЮБЛИНО ШК. 713', 'ФИЛИАЛ ШБС ЛЮБЛИНО ШК. 1819', 'ФИЛИАЛ ШБС ЛЮБЛИНО ШК. 1819', 'ФИРМА ВИСКОСТИТ', 'ФОРМА ВЕС ЛОБЛИНО ШК. 1819', 'ФОРМА ВЕС ЛОБЛИНО ШК. 1819', 'ФОРМА ВЕС ЛОБЛИНО ШК. 1819', 'ФИРМА ВЕС ЛОБЛИНО ШК. 1819', 'ФОРМАСТЕР', 'ФОРМА ВЕС ЛОБЛИНО ШК. 1819', 'ФОРМА ВЕС ЛОБЛИНО ШК. | Total Conference | Total Conf

'Ян Примус', 'Яндек 'Чоъ', 'Ясмин', "Ля 'Ян Примус', 'Яндекс Еда', 'Янис', 'Японситальяна', 'Японская кулинария', 'Японский пес', 'Японский ресторан «Ваби-Саби»', 'Япончик', 'Япоша', 'Япошкин Ко т', 'Ярь', 'Ясмин', 'Яуза', 'бар «Центральный»', 'бар гостиницы «Татьяна»', 'буфет при школе № 443', 'буфет при школе № 1269', 'кафе «Кафин', 'кафе «Ия Фермы', 'кафе «Михайло»', 'кафе «Мухайло»', 'кафе «Нияма»', 'кафе «Папа Джонс»', 'кафе «Сквознячок»', 'кафе «Старбакс»', 'кафе «Шокайло»', 'кафе «Михайло»', 'кафе «Михайло»', 'кафе «Папа Джонс»', 'кафе «Сквознячок»', 'кафе «Старбакс»', 'кафе «Шокайло»', 'кафе Аэробар при Павелецком вокзале', 'кафе Баловень', 'кафе при кинотеатре', 'лищеблок в хоровой академии', 'ресторан «Баровень', 'ресторан «Попаз»', 'ресторан «Якитория»', 'ресторан ТЕРРАСА', 'ресторан ТЕРРАСА', 'ресторан ТЕРРАСА', 'столовая при МТГ', 'столовая при гостинице', 'столовая при школе № 1317', 'столовая при школе № 467', 'столовая при школе № 467 тем № 1329', 'столовая при школе № 406', 'столовая при школе № 422', 'столовая при школе № 423', 'столовая при школе № 672', 'школа 1143', 'школа 1186', 'школа 1877', 'школа 1986', 'школа 2010', 'школа 2011', 'школа 2012', 'школа 335', 'школа 338', 'школа 393', 'школа 460', 'школа 476', 'школа 772', 'школа 775', 'школа 913', 'школа 919']

На первый взгляд, в столбце имеются неявные дубликаты. Тем не менее, рассматривать название точки (которое даже при регистрации юридического лица может немного отличаться от названия, например, сети или франшизы) в отрыве от её уникального идентификатора не следует. Проверим на дубликаты все встречающиеся в таблице пары (id, object_name):

```
In [11]: rest_data[['id', 'object_name']].duplicated().sum()
Out[11]: 0
```

Итак, дубликатов точек, которые имели бы различия в остальных столбцах, не найдено. При этом, так же нормальной ситуацией является расположение нескольких точек с одним названием (object_name) по одному адресу. Например:

```
In [12]: (
              rest data
              .query('address == "город Москва, улица Земляной Вал, дом 33"')
               sort_values(by='object_name')
              .tail()
```

```
id object_name chain object_type
                                                                         address number
15254 209240 Чайхона №1 True
                                       кафе город Москва, улица Земляной Вал, дом 33
                                                                                      15
14716 209239
               Чача room False
                                      кафе город Москва, улица Земляной Вал, дом 33
                                                                                     20
15363 209186 Шоколадница True
                                                                                     20
                                      кафе город Москва, улица Земляной Вал, дом 33
15362 209264 Шоколадница True кафе город Москва, улица Земляной Вал, дом 33
                                                                                     10
 7961 26632 Этаж прожект False
                                      кафе город Москва, улица Земляной Вал, дом 33
```

В больших торговых комплексах различные заведения одной сети могут располагаться в нескольких местах.

Это нормально, если заведения различаются параметрами (например, числом посадочных мест). Однако наличие по одному адресу двух полностью одинаковых по параметрам заведений маловероятно: достаточно сложно на разных этажах арендовать одинаковые по характеристикам площади

Вместе с тем, при сборе информации из различных источников возможны ошибки, и одно и то же заведение может быть зарегистрировано под разными id.

Проверим эту гипотезу:

```
In [13]: (
              rest data
              .set_index(['id'])
              .duplicated()
              ·sum()
Out[13]: 82
```

Действительно, обнаружено 82 заведения, дублирующих уже имеющиеся, но с другими іd. Удалим дубликаты и восстановим іd в виде столбца:

```
In [14]: rest_data =
                rest data
                .set_index(['id'])
                .drop_duplicates()
.reset_index()
           .
# проверим результат
                rest data
                .set_index(['id'])
                .duplicated()
Out[14]:
```

Неявные полные дубликаты заведений удалены.

Как уже упоминалось выше, в столбце 'object_name' возможны неявные дубликаты названий. Наиболе ярко это должно быть заметно на примере сетей:

```
In [15]: list(
              rest data
               .query('chain')
               'object_name'
               .sort values()
               .unique()
```

```
Out[15]: ['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 YPHOK',
'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'
                       'Marmalato',
'Marrakesh Хинкальная',
'MosKalyan',
'Movenpick',
'My box',
'Osteria Mario',
'PAUL Поль',
'PRIME STAR Прайм стар',
'PRIME Прайм стар Прайм-кафе',
                       'Panda Express',
'Pizengof99 CΠБ',
                       'Prime Прайм стар',
'STARBUCKS COFFEE',
                        'SUBWAY',
'Starbucks'
                        'Starbucks Старбакс',
                        'Starbucks κοφe',
'Starbucks, CTAPEAKC ΚΟΦΕ',
'Starlite Diner',
                       'Subway',
                        'TAJJ MAHAL
                       'TGI FRIDAYS',
                        'TUTTI FRUTTI FROZEN YOGURT',
                        'The Terrace',
                       'Tokyo bay',
'Torro Grill Торро Гриль',
                       'Travelers Coffe',
'Travelers Coffee',
                       'Tutti frutti',
'UPSIDE DOWN CAKE',
                       'Upside Down',
'Upside Down Cake',
                       'Van Wok Ванвок',
'Vanwok Ванвок',
                       Vainuok bahabuk,
'VietCafe Вьеткафе',
'White Rabbite (Белый кролик)',
'Wok & Box',
'Wokker',
                       'Vogurt Frenzy',
'«Кафе «Кулинарная лавка братьев Караваевых»',
'«Кафе «Кулинарная лавка братьев Караваевых»»',
                       '«Кафе «Кулинарная лавка
'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',
 Баский Робойис, , 'Бенто МОК', 'Билла', 'Бир Хаус Паб', 'Бир хаус', 'Блинная', 'Бодый день Кофе с собой',
 'Ботик Петра',
'Братья Караваевы',
'Братья Караваевы',
'Братья Караваевых',
'Брудер',
'Брусника',
'Бургер Кинг',
'Бургер Кинг Burger King',
'Бургер Клаб',
'Бургер клаб',
'Бургер клаб',
'Бургер клаб',
'Бутер Бизон',
'888 Бургер',
'Black Burger Чайхона №1',
'BEHELUЯЯ',
'BOKKEP',
  'BOKKEP'
  '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 ЭZO Хинкальная',
 'Грузинские каникулы Барбарис',
 'Гудман',
'Гудман Гудвин',
 'Гурман',
'Гурмания',
'ДАНКИН ДОНАТС',
 ДВЕ ПАЛОЧКИ',
'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',
 дапкин донатс,
'Две палочки',
'Де Марко',
'Детское кафе «Космик»',
'Джаганнат',
  'Джон Джоли',
  'ДоДо Пицца'
 'ДоДо пицца',
'Добрынинский и партнёры',
  'Додо Пицца',
  'Доло 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',
'Елки-Палки',

"Жан Жак',

"ЗОЛОТАЯ ВОБЛА',

"Закусочная КFС',

"Закусочная «КРС',

"Закусочная «Баскин Роббинс & Стардогс»',

"Закусочная «Бругер Кипг»',

"Закусочная «Крошка Картошка»',

"Закусочная «Пицца Хат»',

"Закусочная «СишПоль",
 'Закусочная «СушиШоп»',
 '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',
'KAPO',
'KAΦE KFC',
'КАФЕ Starbucks',
'КАФЕ «Андерсон»'
'КАФЕ «Восточный базар»',
'КАФЕ «МАКДОНАЛДС»',
'ΚΑΦΕ «MY-MY»',
'ΚΑΦΕ «ΤΕΡΕΜΟΚ»'
'КАФЕ «Теремок»',
'КАФЕ «Хлеб насущный»',
'КАФЕ «ШТОЛЛЕ»',
 'КАФЕ «Шоколадница»
'КАФЕ «ЯКИТОРИЯ»',
'КАФЕ МАКДОНАЛДС',
'КАФЕ-ПЕКАРЕЯ «ВОЛКОНСКИЙ»',
'КОЛБАСОФФ',
'КОРЧМА «ТАРАС БУЛЬБА»',
'КОРЧМА ТАРАС БУЛЬБА',
КОФЕЙНЯ «ДЕ МАРКО»',

'КОФЕЙНЯ «ДЕ МАРКО»',
 'КОФЕЙНЯ «ШОКОЛАДНИЦА́»',
 'КОФЕМАНИЯ',
 'КОФЕПОРТ',
'КРОШКА КАРТОШКА',
'КРУЖКА',
'КУЛИНАРНАЯ ЛАВКА БРАТЬЕВ КАРАВАЕВЫХ',
 'КУЛИНАРНОЕ БЮРО',
'КУЛИНАРНОЕ БИРО',
'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 Prime Прайм стар Прайм

'Кафе «Costa coffee»',

'Кафе «KFC»',

'Кафе «PRIME» Прайм стар',

'Кафе «Prime» (Прайм стар',

'Кафе «SuBWAY»',

'Кафе «Starbucks»',

'Кафе «Starbucks»',
 'Kaфe «Subwav»'
'Кафе «АндерСон»'
 'Кафе «Андерсон»'
'Кафе «БУРГЕР КИНГ»'
'Кафе «Баскин Роббинс»'.
'Кафе «Бургер Кинг»',
 'Кафе «Ваби-Саби»',
'Кафе «Данкин Донатс»'
кафе «динкин Донатс»',
"Кафе «Домино'с Пицца»",
'Кафе «Иль Патио»',
'Кафе «Кофе Хаус»',
'Кафе «Кофейня Старбакс»',
'Кафе «Комия»
'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 «Урик» 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 Кофемани
'Кафе МУ-МУ',
'Кафе ПРОНТО',
```

```
'Кафе Пекарня Хачапури',
'Кафе Песто и Митлес',
'Кафе Пронто',
'Кафе Рецептор'
'Кафе Суши Wok',
'Кафе Чайхона №1',
'Кафе Штолле',
'Кафе Якитория'
'Кафе Якитория',
'Кафе быстрого питания Му-му',
'Кафе мороженного «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',
'Кофе С Собой',
'Кофе с Собой',
'Кофе с собой Gate 42',
'Кофе с собой «Кофейня»',
Кофе с собой «Кофейня»',
'Кофе тун',
'Кофе-Кан',
'Кофе-Кан',
'Кофейня «Дабл Би»',
'Кофейня «Кофе Хаус»',
'Кофейня «СТАРБАКС КОФЕ»',
'Кофейня «СТАРБАКС КОФЕ»',
'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',
'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'.
 'MEH3A',
'МИ ПЬЯЧЕ',
'МИЛТИ',
'мимино'
 'МОСКВА - МАКДОНАЛДС',
'МСК Московская сеть кальянных',
'МСК Московская сеть кальянных на Шаболовке',
'MУ-MУ',
'MЯCOROOB'
 'МагБургер',
'Магазин готовой еды «Милти»',
'Магазин-кулинария «Суши Сет»',
 'Магбургер',
'Магбургер',
'Магбургер АЗС №16',
'Магбургер АЗС №17',
'Макдоналдс',
'Маки-Маки',
 'Маки-маки',
'Макс Бренне́р',
'Марчеллис',
 'Меленка',
'Менза',
'Метро К&К',
'Ми Пьяче',
 'Милано пицца'.
 'Милти'
 'Мимино'
 'Мимино Хинкальная',
'Мираторг',
'Мистер Картошка',
'Молли гвинз',
'Му-Му',
 'Му-му'
 'Мюнгер'
 'МясоROOB',
'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 «КFC Покрышкина»',
'ПБО «КFC»',
"ПБО «Теремок»',
"ПИВКО',
"ПИЛЗНЕР',
"ПИРОГИ ШТОЛЛЕ',
'ПИЦЦЕРИЯ «Папа Джонс»',
'ПОМИДОР',
'ПРАВDA Кофе',
'ПРАЙМ',
'ПРАЙМ СТАР',
'ПРАЙМкафе Прайм стар',
'ПРИМАВЕРА',
'ПРОНТО',
ПРОПТО ,
'Паб Кружка',
'Павлин Мавлин Чайхона №1',
'Павлин-Мавлин',
'Панчо Пицца',
'Папа Джонс'
'Перекресток'
'ПивКо',
'Пикколо',
'Пиппони',
'Пироговая Штолле',
'Пицерия «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 «КFC»', 
'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 «Макдоналдс»', 
"Предприятие быстрого обслуживания «СушиШоп»', 
"Предприятие быстрого обслуживания «СушиШоп»',
'Предприятие быстрого обслуживания «Теремок»
предприятие быстрого обслуживания «Теремок»; 
'Предприятие быстрого питания «Крошка Картошка»', 
'Примавера',
'Пронто',
'Простые Вещи',
Простые вещи',
'РЕСТОРАН «ДЖОН ДЖОЛИ»',
'РЕСТОРАН «КОСМИК»',
'РЕСТОРАН «ТАНУКИ»',
'PECTOPAH «ТаНУКИ»',
'PECTOPAH «ТаНУКИ»',
'PECTOPAH «УРЮК»',
'PECTOPAH «ЧАЙХОНА №1»',
'PECTOPAH «Чайхона №1»',
'PECTOPAH«МАКДОНАЛДС»',
'PECTOPAH«MAKДOHAЛДC»',
'POCTUKC KFC',
'PectopaH KFC',
'PectopaH «Florentini»',
'PectopaH «KFC»',
'PectopaH «Ëpu»',
'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 «Ньокки»',
'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 ИКЕА ІКЕА',
'Ресторан Маки-Маки',
'Ресторан ТАНУКИ',
'Ресторан Тануки',
'Ресторан Торро Гриль',
'Ресторан Хинкальная', 
'Ресторан быстрого питания КFC',
'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',
'Сабвей',
'Сабвей Subway',
'Сабвей СушиВок',
'C6appo',
'Сварно',
'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',
Старракс Кофе',
'СтардогS',
'СтардогS и шаурма',
'СтардогS',
'СтардогС СтардогS',
'Старина Миллер',
'Старина Стардог 
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 сет',
'Суши тун',
'Суши тун, Хруст ріzza, Кофе тун',
'СушиМан',
'СушиШоп',
'СушиШоп',
'Сытая Утка',
'ТАНУКИ',
'ТАНУКИ',
'ТЕРЕМОК',
'ТЕРЕМОК ИНВЕСТ',
'ТЕРЕМОК ИНВЕСТ',
 'Тайм Авеню'
'Тамаси Суши',
'Тануки',
'Тапчан',
'Ташир пицца',
'Темпл Бар',
'Темпл бар',
'Теремок',
'Теремок-Инвест',
'Территория',
'Территория ТІМВІGFAMILY',
'Тирольские пироги',
'То Да Сё',
'Торро Гриль',
```

```
'Точка продажи готовой еды Милти'.
'Тратория 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',
'ШВДРИВДЛЬЛ'
'ШОКОЛАДНИЦА',
'ШТОЛЛЕ',
'Шантимель
'Шантимель (кондитерские)',
'Шикари Иль Патио',
'Шоколад',
'Шоколадница'
'Шоколадница Кофе Хаус',
'Шоколадница Кофемания',
'Штирбирлиц',
'Штолле',
'Югос',
'ЯКИТОРИЯ',
'ЯКИТОРИЯ Й СПОРТ БАР',
'ямми микс',
'ЯПОША',
'Якитория'
'Якитория и Ян Примус',
'Ян Примус'
'Японский ресторан «Ваби-Саби»',
'Япоша',
'кафе «Му-Му»',
'кафе «Нияма»',
'кафе «Папа Джонс»',
'кафе «Старбакс»',
'кафе «Шоколадница»',
'ресторан «Брудер»'
'ресторан «Якитория»']
```

Действительно, мы видим разные варианты написания одних и тех же названий сетей. Однако в контексте настоящего исследования не предполагается изучать отдельные сети, поэтому оставим пока названия отдельных точек, как они есть, не приводя к единому имени сети. Тем более, что вопрос выбора общего наименования часто дискуссионный:

- ДЕТСКИЙ КЛУБ «АНДЕРСОН», или
- ДЕТСКОЕ КАФЕ «АНДЕРСОН»?

Изучим состав значений столбца object_type :

```
In [16]: rest_data.object_type.sort_values().unique()
           array(['бар', 'буфет', 'закусочная', 'кафе', 'кафетерий', 'магазин (отдел кулинарии)', 'предприятие быстрого обслуживания',
                    'ресторан', 'столовая'], dtype=object)
           Категориальный столбец object_type может принимать 9 значений. Неявных дубликатов не обнаружено.
```

Изучим столбцы number и id:

```
In [17]: rest_data[['number', 'id']].describe()
```

14 000000 28494 500000 40.000000 144805.500000

80.000000 184123.000000 max 1700.000000 223439.000000

25%

75%

```
Out[17]:
                     number
          count 15284.000000 15284.000000
                   59.784350 119506.877126
                   74.838657 73042.951997
                    0.000000
                              838.000000
           min
```

```
Диапазоны значений, которые принимают number и id , позволяют привести тип данных столбцов к int32 для экономии памяти:
```

```
In [18]: rest_data.id = rest_data.id.astype('int32')
    rest_data.number = rest_data.number.astype('int32')
```

```
print(rest_data.id.dtype)
print(rest_data.number.dtype)
int32
int32
```

Вывод

Мы изучили данные и провели их предварительную обработку:

- преобразовали столбец chain к двоичному типу;
- удалили полные дубликаты точек питания;
- столбец object_type может принимать 9 категориальных значений, дубликатов нет;
- привели number и id к типу int32 для экономии памяти.

Анализ данных

Проанализируем имеющиеся в наличии данные и определим текущее положение дел на рынке услуг общественного питания.

Сравнение количества заведений общественного питания разных форматов

Формат заведения указан в столбце object_type . Сгруппируем данные и посчитаем количество заведений каждого типа:

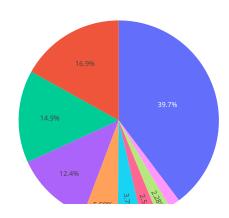
```
In [19]: object_formats = (
                     ret_tormats =
rest_data
.groupby(by='object_type')
.agg({'id' : 'nunique'})
.sort_values(by='id', ascending=False)
               object_formats.columns = ['object_count']
               object_formats
```

Out[19]:		object_count
	object_type	
	кафе	6071
	столовая	2584
	ресторан	2282
	предприятие быстрого обслуживания	1898
	бар	855
	буфет	577
	кафетерий	395
	закусочная	349
	магазин (отдел кулинарии)	273

Невооружённым глазом видно, что наиболее популярно открывать кафе. Визуализируем полученные данные, определив долю каждого формата:

```
layout=go.Layout(
    title='Доля заведений каждого формата в Москве'
      fig.show()
```

Доля заведений каждого формата в Москве



На полученной диаграмме легко выделить Тор-5 форматов заведений питания в Москве:

1. кафе (39.7%);

```
2. столовая (16.8%);
3. ресторан (14.9%):
4. предприятие быстрого обслуживания (12.5%);
5. бар (5.57%).
```

Сравнение количества сетевых и несетевых заведений общественного питания

Признак принадлежности заведения к сети указан в столбце chain . Сгруппируем данные и посчитаем количество заведений каждого типа:

```
In [21]: chain_objects = (
                rest data
                 .groupby(by='chain')
                 .agg({'id' : 'nunique'})
.sort_values(by='id', ascending=False)
            chain_objects.columns = ['object_count']
           chain_objects
```

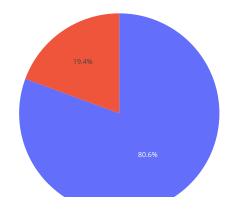
object_count

chain	
False	12320
True	2964

Как видим, в москве количественно преобладают несетевые заведения общественного питания. Визуализируем полученные данные:

```
values=chain_objects.object_count)],
         layout=go.Layout(
            title='Доля сетевых и несетевых заведений в Москве'
      fig.show()
```

Доля сетевых и несетевых заведений в Москве



Вывод: Диаграмма соотношения количества сетевых и несетевых заведений питания свидетельствует о более чем 4-кратном преобладании индивидуальных точек питания над сетевыми.

Вероятно, это связано с тем, что малый и средний бизнес неохотно инвестируют в сетевые франшизы, т.к. это не тлько накладывает ограничения на формат и ассортимент заведения, но и может предусматривать регулярные выплаты (роялти) за пользование франшизой, что приводит к снижению прибыльности дела.

Определение видов объектов общественного питания, для которых характерно сетевое распространение

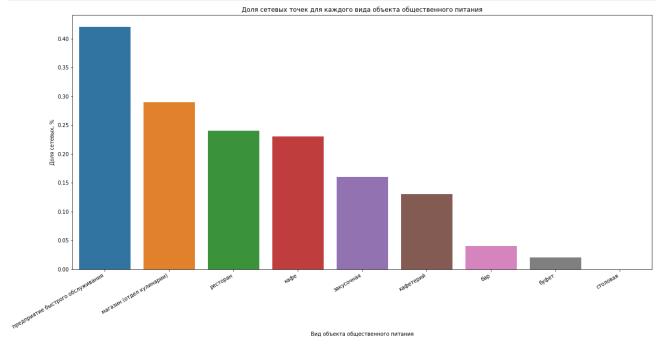
Для определения видов объектов общественного питания, для которых характерно сетевое распространение, посчитаем долю сетевых точек по тношению к общему количеству для каждого вида объекта:

```
In [23]: chain_objects = (
                    rest_data
.pivot_table(
                         values='id',
index='object_type',
                         columns='chain',
aggfunc='nunique',
fill_value=0
                     reset_index()
             chain_objects['total'] = chain_objects[True] + chain_objects[False]
chain_objects['percent'] = round(chain_objects[True] / chain_objects['total'], 2)
              chain_objects = chain_objects.sort_values(by='percent', ascending=False)
              chain_objects
```

chain	object_type	False	True	total	percent
6	предприятие быстрого обслуживания	1110	788	1898	0.42
5	магазин (отдел кулинарии)	195	78	273	0.29
7	ресторан	1739	543	2282	0.24
3	кафе	4675	1396	6071	0.23
2	закусочная	293	56	349	0.16
4	кафетерий	343	52	395	0.13
0	бар	818	37	855	0.04
1	буфет	566	11	577	0.02
8	столовая	2581	3	2584	0.00

Визуализируем полученный результат:

```
sns.barplot(
                  x='object_type',
y='percent',
                  data=chain_objects
            plt.title('Доля сетевых точек для каждого вида объекта общественного питания')
plt.xlabel('Вид объекта общественного питания')
            fig.autofmt_xdate(rotation=30) plt.ylabel('Доля сетевых, %')
            fig.savefig('Доля сетевых точек.png')
plt.show()
```



Итак, диаграмма ниглядно демонстрирует, что в наибольшей степени сетевое распространение характерно для предприятий быстрого обслуживания (фаст-

Тем не менее, в значительной степени сетевое распространение характерно также и для отделов кулинарии, ресторанов и кафе.

Изучение соотношения количества сетевых заведений и количества мест в них

Исследуем, что характерно для сетевых заведений: большое количество заведений с небольшим числом посадочных мест в каждом, небольшое количество заведений с большим количеством посадочных мест, или иное. Для этого построим диаграмму рассеяния количества посадочных мест для различных сетевых точек питания:

```
In [25]: fig = plt.figure(
                    figsize=(15, 8)
              sns.scatterplot(
                   x='id',
y='number',
                    hue='chain'
                   data=rest_data.query('chain'),
              plt.title('Количество посадочных мест для сетевых объектов общественного питания')
             plt.xlabel('Идентификаторы сетевых объектов общественного питания') fig.autofmt_xdate(rotation=45) plt.ylabel('Количество посадочных мест') fig.savefig('Количество посадочных мест.png')
              plt.show()
```

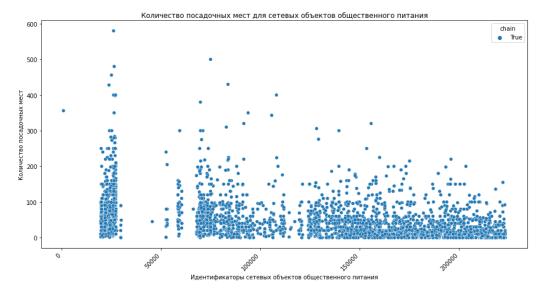


Диаграмма рассеяния наглядно демонстрирует, что для сетевых объектов в большей степени характерно большое количество объектов с относительно малым количеством посадочных мест.

Посмотрим на этот факт с другой стороны - сгруппируем точки по имени сети, посчитаем количество точек и среднее количество посадочных мест:

```
In [26]: quantity_vs_seats = (
    rest_data.query('chain')
                   .groupby(by='object_name')
.agg({'id':'count', 'number':'median'})
                   rename(columns={'id':'num_places', 'number':'median_seats'})
             quantity_vs_seats.head()
```

Out[26]: num_places median_seats

object_name		
Beverly Hills Diner	1	88.0
Bierloga	1	75.0
Black & White	1	40.0
Bocconcino	3	68.0
ВооВо	1	46.0

Найдём границы понятий "много" и "мало":

```
In [27]: num_places_tresch = round(quantity_vs_seats.num_places.mean())
            median_seats_tresch = round(quantity_vs_seats.median_seats.median())
print('Пороговое значение количества заведений:', num_places_tresch)
             \verb"print('Пороговое значение количества мест:', median_seats\_tresch)"
```

Пороговое значение количества заведений: 4 Пороговое значение количества мест: 45

Определим функцию классификации:

```
In [28]: def get_chain_class(row, places_tresch=4, seats_tresch=45):
    if row['num_places'] <= places_tresch: # мало забедений
    if row['median_seats'] <= seats_tresch: # мало мест
    return 'мало заведений / мало мест'
                                else: # много мест
                         return 'мало заведений / МНОГО мест'
else: # много заведений
                                if row['median_seats'] <= seats_tresch: # мало мест
                                return 'МНОГО заведений / мало мест' else: # много мест
                         return 'МНОГО заведений / МНОГО мест'
return 'undefined' # сюда не должны попасть никогда
```

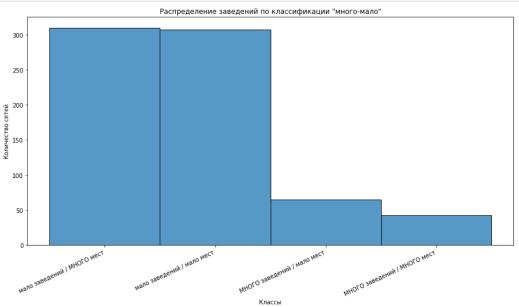
Классифицируем сети по введённым признакам:

```
In [29]: quantity_vs_seats['class'] = quantity_vs_seats.apply(
               get_chain_class,
args=(num_places_tresch,
                      median_seats_tresch),
          quantity vs seats.head()
```

Out[29]:		num_places	median_seats	class
	object_name			
	Beverly Hills Diner	1	88.0	мало заведений / МНОГО мест
	Bierloga	1	75.0	мало заведений / МНОГО мест
	Black & White	1	40.0	мало заведений / мало мест
	Bocconcino	3	68.0	мало заведений / МНОГО мест
	ВооВо	1	46.0	мало заведений / МНОГО мест

Визуализируем распределение заведений по введённой классификации "много-мало":

```
In [30]: fig = plt.figure(
    figsize=(15, 8)
               sns.histplot(
                     x='class'
                    data=quantity_vs_seats,
              plt.title('Распределение заведений по классификации "много-мало"')
plt.xlabel('Классы')
fig.autofmt_xdate(rotation=25)
              plt.ylabel('Количество сетей')
fig.savefig('классификация много-мало.png')
              plt.show()
```



Попробуем визуализировать данное распределение диаграммой рассеяния:

```
In [31]: # восстановим индексы
          quantity_vs_seats = quantity_vs_seats.reset_index()
         quantity_vs_seats
```

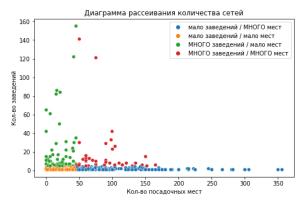
)ut[31]:		object_name	num_places	median_seats	class
	0	Beverly Hills Diner	1	88.0	мало заведений / МНОГО мест
	1	Bierloga	1	75.0	мало заведений / МНОГО мест
	2	Black & White	1	40.0	мало заведений / мало мест
	3	Bocconcino	3	68.0	мало заведений / МНОГО мест
	4	ВооВо	1	46.0	мало заведений / МНОГО мест
	719	кафе «Папа Джонс»	1	24.0	мало заведений / мало мест
	720	кафе «Старбакс»	2	63.5	мало заведений / МНОГО мест
	721	кафе «Шоколадница»	1	50.0	мало заведений / МНОГО мест
	722	ресторан «Брудер»	1	86.0	мало заведений / МНОГО мест
	723	ресторан «Якитория»	1	124.0	мало заведений / МНОГО мест

724 rows × 4 columns

```
In [32]: # nocmpowm scatterpLot
plt.figure(figsize=(8, 5))
sns.scatterplot(data = quantity_vs_seats, x='median_seats', y='num_places', hue='class')
plt.title('Диаграмма рассеивания количества сетей')
plt.ylabel('Кол-во заведений')
plt.xlabel('Кол-во посадочных мест')
               plt.legend()
                # группируем данные по категориям с показателями
               quantity_vs_seats.groupby('class').agg({'class':'count', 'num_places':'median', 'median_seats':'median'})
```

Out[32]: class num_places median_seats

class			
МНОГО заведений / МНОГО мест	42	8	81.5
МНОГО заведений / мало мест	65	9	20.0
мало заведений / МНОГО мест	310	1	85.0
мало заведений / мало мест	307	1	20.0



Из полученного распределения можно сделать следующие выводы:

- для сетей более характерно небольшое количество заведений:
- для небольшого количества заведений более характерно большое количество посадочных мест.

Оценка минимального, максимального и среднего количества посадочных мест для различных видов объектов общественного питания

Как мы видели выше, на примере сетевых заведений, для параметра количество посадочных мест характерно наличие выбросов. Поэтому для оценки среднего количества посадочных мест для различных видов точек питания будем использовать медиану, как более устойчивую к выбросам.

Сгруппируем данные по виду заведения питания и посчитаем минимальное, максимальное и медианное значения количества посадочных мест:

```
groupby(by='object_type')['number']
            .agg(
               min_seats='min',
               max_seats='max'
               median_seats='median'
            reset_index()
        obj_median_seats.sort_values(by='median_seats', ascending=False)
```

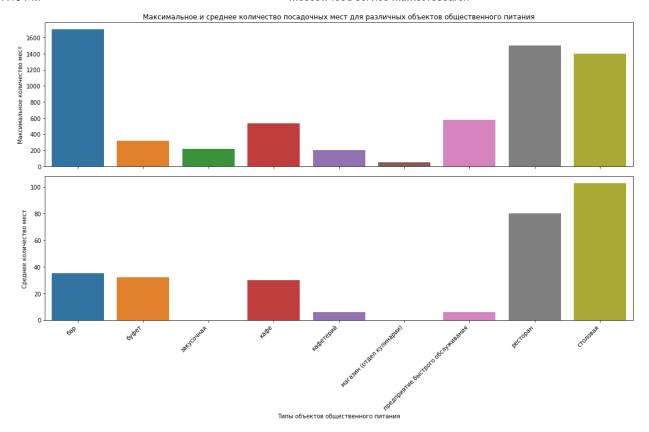
:	object_type	min_seats	max_seats	median_seats
8	столовая	0	1400	103
7	ресторан	0	1500	80
0	бар	0	1700	35
1	буфет	0	320	32
3	кафе	0	533	30
4	кафетерий	0	200	6
6	предприятие быстрого обслуживания	0	580	6
2	закусочная	0	216	0
5	магазин (отдел кулинарии)	0	50	0

Из полученных данных следует, что в среднем самое большое количество посадочных мест (103) предоставляют столовые, за ними следуют рестораны и, с большим отрывом, бары, буфеты и кафе.

Визуализируем данные с использованием столбчатых диаграмм для максимального и медианного количества мест, расположенных на общей оси х:

```
In [34]: # задаём размер сетки для графиков
           fig = plt.figure(
    figsize=(15, 10)
           # 2. во второй ячейке строим диаграмму для MEDIAN
# горизонтальная ось — от графика из первой ячейки
            ax2 = plt.subplot(2, 1, 2)
            sns.barplot(
                x='object_type'
y='median_seats
                 data=obj_median_seats
           plt.xlabel('Типы объектов общественного питания') ax2.set_ylabel('Среднее количество мест')
            # 1. в первой ячейке строим диаграмму для МАХ
            ax1 = plt.subplot(2, 1, 1, sharex=ax2)
            sns.barplot(
                x='object_type',
y='max_seats',
                 data=obj_median_seats
            ax1.set_ylabel('Максимальное количество мест')
            fig.autofmt_xdate(rotation=45)
            plt.title('Максимальное и среднее количество посадочных мест для различных объектов общественного питания')
            plt.tight_layout()
            fig.savefig('Среднее количество посадочных мест.png')
           plt.show()
```

Out[35]



Сделанные выше выводы подтвердились:

- несмотря на то, что максимальное количество мест обеспечивают бары, в большинстве случаев это выбросы;
- наибольшее в среднем количество посадочных мест обеспечивают столовые

Анализ распределения заведений общественного питания по улицам Москвы

Выделение улицы из адреса заведения

Для начала, выделим улицу из адреса. Для этого разобъём строку адреса по разделителю ', ' (запятая и следующий за ней пробел), а затем выделим второй элемент получившегося списка - это и будет название улицы:

```
rest_data.sample(5)
```

street	number	address	object_type	chain	object_name	id	:
улица Хачатуряна	2	город Москва, улица Хачатуряна, дом 20А	закусочная	False	Пицца Гирос, Шаурма	204101	12709
Старокачаловская улица	28	город Москва, Старокачаловская улица, дом 5	кафе	True	ПОМИДОР	89115	6203
Локомотивный проезд	28	город Москва, Локомотивный проезд, дом 4	предприятие быстрого обслуживания	True	Макдоналдс	115493	7112
поселение Сосенское	15	город Москва, поселение Сосенское, посёлок Ком	кафе	False	Кафе Тайм аут	194008	13251
Ленинский проспект	40	город Москва, Ленинский проспект, дом 109	кафе	True	Крошка Картошка	71252	6161

Итак, название улицы успешно выделено в отдельный столбец.

Определение топ-10 улиц по количеству объектов общественного питания и районов Москвы, в которых они расположены

Сгруппируем данные по улице и посчитаем количество точек питания на каждой из них. Выберем топ-10:

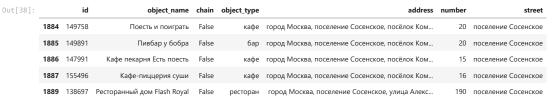
```
In [36]: (
                                res_uatd
groupby(by='street')
agg({'id' : 'count'})
.rename(columns={'id':'obj_count'})
.sort_values(by='obj_count', ascending=False)
                                 head(10)
```

obj count город Зеленоград 232 проспект Мира 204 181 Профсоюзная улица Ленинградский проспект 170 Пресненская набережная 167 Варшавское шоссе 161 Ленинский проспект 147 поселение Сосенское 137 проспект Вернадского 128 Кутузовский проспект 114

В результаты попали город Зеленоград и поселение Сосенское. Вероятно потому, что в этих населённых пунктах, входящих в состав Москвы, не указаны улицы, либо иная структура адреса. Проверим эту гипотезу:

In [37]: rest_data.query('street == "город Зеленоград"').head() Out[37]: object_name chain object_type address number **1386** 22331 СТОЛОВАЯ МИЭТ False столовая город Москва, город Зеленоград, улица Юности, ... 56 город Зеленоград кафе город Москва, город Зеленоград, корпус 401 35 город Зеленоград **1405** 20199 ЭТО False **1406** 68113 Френдс False город Москва, город Зеленоград, корпус 435 34 город Зеленоград кафе кафе город Москва, город Зеленоград, корпус 436 40 город Зеленоград **1407** 20105 Кафе «Граф Монте-Кристо» False **1408** 22982 Альфорно False город Москва, город Зеленоград, корпус 438 49 город Зеленоград

In [38]: rest_data.query('street == "поселение Сосенское"').head()

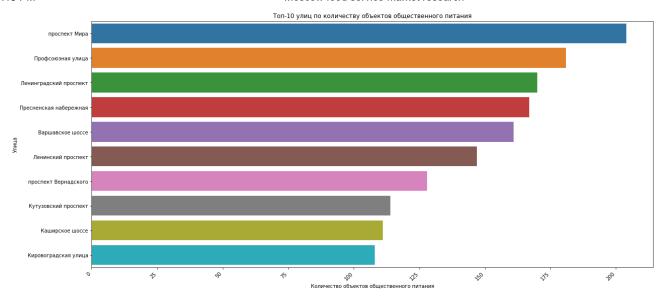


Наша гипотеза подтвердилась отчасти:

- в поле 'address' для некоторых точек питания в Зеленограде указана улица (идёт 3-м номером в списке элементов строки);
- для других же точек в Зеленограде улица не указана;
- для точек питания в поселении Сосенское название улицы указано 3-м и далее номером в списке элементов строки.

Поскольку партнёров в первую очередь интересуют локации непосредственно в городе Москва, исключим точки, расположенные в Зеленограде и Сосенском, и заново построим топ-10 улиц по количеству заведений питания:

```
In [39]: top_10_streets = (
                rest_data
.query('street != "город Зеленоград" & street != "поселение Сосенское"')
                .groupby(by='street')
.agg({'id' : 'count'})
                reset_index()
                .rename(columns={'id':'obj_count'})
                 sort_values(by='obj_count', ascending=False)
                .head(10)
           fig = plt.figure(
   figsize=(20, 10)
           sns.barplot(
               y='street',
x='obj_count'
                data=top_10_streets,
               orient='h
           plt.title('Ton-10 улиц по количеству объектов общественного питания')
           plt.xlabel('Количество объектов общественного питания')
           fig.autofmt_xdate(rotation=45)
plt.ylabel('Улица')
           fig.savefig('топ-10 улиц.png')
           plt.show()
```



Итак, топ-10 улиц по количеству точек общественного питания в Москве выглядит следующим образом:

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

Для определения районов, в которых располагаются эти улицы, используем внешние данные. Например открытые данные Правительства Москвы. Для работы нам потребуются:

- 1. Общемосковский классификатор улиц Москвы ОМК УМ (ОМК 001-2013)
- 2. Общемосковский классификатор территориальных единиц Москвы ОМК ТЕ (ОМК 002-2013)

Для удобства работы данные датасеты были скачаны и выложены в публичный доступ на Яндекс.Диск.

Загрузим информацию об улицах и выделим столбцы с названиями улиц и кодами районов:

```
Іп [40]: # загрузим датафрейм
          streets_class = open_file_from_yadisk(
              'https://disk.yandex.ru/d/BHTNopnmOmybGw',
              sep=
              encoding='cp1251'
          # удалим дублирующий заголовок на русском языке
          streets_class.drop(index=[0], axis=0, inplace=True)
           # приведём названия столбцов к snake_case
          streets_class.columns = streets_class.columns.str.lower()
          # выделим нужные столбиы
          streets_class = streets_class[['um_namef', 'um_te']]
         # удалим явные дубликаты streets_class = streets_class.drop_duplicates()
          streets_class.tail()
```

	um_namef	um_te
5324	улица Беллы Ахмадулиной	1101
5325	улица Лётчика Осканова	0203; 0209
5326	улица Летовского Ополчения	1109
5327	улица Быстролётова	1109
5328	площадь Донецкой Народной Республики	0106

Out[40]

Некоторым улицам соответствуют несколько районов. В идеальной ситуации для точного определения района следовало бы воспользоваться запросами к платным картографическим сервисам, однако для целей настоящего исследования не будет большим искажением отнести улицу ко всем районам, по которым она пролегает.

Разделим в разные строки перечисления в столбце 'um_te' и приведём значения кодов к целому типу:

```
In [41]: # разбиваем строки для значений, разделённых '; '
          streets_class = (
              pd.concat(
    [pd.Series(row['um_namef'], row['um_te'].split('; '))
                               for _, row in streets_class.iterrows()]
               reset_index()
          .
# переименовываем столбиы
          streets_class.columns = ['um_te', 'um_namef']
```

```
# разбиваем строки для значений, разделённых ';' (без пробела)
          streets_class = (
              pd.concat(
                 [pd.Series(row['um_namef'], row['um_te'].split(';'))
                              for _, row in streets_class.iterrows()]
              .reset_index()
          .
# пепеименовываем столбиы
         streets_class.columns = ['um_te', 'um_namef']
          # приводим 'um_te' к целому типу
         # приоодим аште к целлму marry
streets_class['um_te'] = streets_class['um_te'].astype('int32')
# удалим явные дубликаты
         streets_class = streets_class.drop_duplicates()
streets_class.tail()
               um te
                                               um namef
                                 улица Лётчика Осканова
         5798 209 улица Лётчика Осканова
         5799
               1109
                               улица Летовского Ополчения
                       улица Быстролётова
         5800 1109
               106 площадь Донецкой Народной Республики
         5801
In [42]: # проверим, есть ли пропуски в кодах
         streets_class['um_te'].isna().sum()
Out[42]: 0
         Загрузим информацию о районах и выделим столбцы с названиями районов и их кодами:
In [43]: # загрузим датафрейм
         regions_class = open_file_from_yadisk(
    'https://disk.yandex.ru/d/XkjaS7Ah8RsXQw',
              sep='
              encoding='cp1251'
          # приведём названия столбцов к snake case
         regions_class.columns = regions_class.columns.str.lower()
          # выделим нужные столбцы
          regions_class = regions_class[['kod', 'name']]
         # удалим явные дубликаты
regions_class = regions_class.drop_duplicates()
         regions class.sample(10)
              kod
                           район Лефортово
          75 515
                      Южнопортовый район
          22 212
                      Молжаниновский район
          79 608
                     район Братеево
          50
              406
                            район Гольяново
          64
              504
                     район Выхино-Жулебино
          136 1101
                       поселение Внуковское
           8 108
                         Тверской район
          96 703
                            район Зюзино
         126 906 район Хорошёво-Мнёвники
         Объединим top 10 streets и streets class:
In [44]: top_10_streets =
              top 10 streets
              .merge(
streets_class,
                 left_on='street',
right_on='um_namef',
                  how='left'
         top_10_streets.head(10)
Out[44]:
                      street obj_count um_te
                                                      um_namef
         0
                проспект Мира 204 105
                                                  проспект Мира
         1
                проспект Мира 204 301 проспект Мира
         2
                проспект Мира
                                  204 309
                                                  проспект Мира
                проспект Мира 204 310 проспект Мира
         3
                проспект Мира 204 312
         4
                                                  проспект Мира
         5
                проспект Мира 204 313 проспект Мира
                проспект Мира
                                  204 317
                                                  проспект Мира
                                  181 701 Профсоюзная улица
         7 Профсоюзная улица
                                  181
                                         704 Профсоюзная улица
          8 Профсоюзная улица
                                181 707 Профсоюзная улица
         9 Профсоюзная улица
```

Очевидно, большое количество объектов питания находится на длинных улицах, которые могут проходить через несколько районов.

Добавим названия районов в таблицу top_10_streets:

```
regions_class,
                  left_on='um_te',
right_on='kod',
                  how='left'
          ,
# удаляем лишнее
top_10_streets = top_10_streets.drop(columns=['um_te', 'um_namef'], axis=1)
          top_10_streets = top_10_streets.rename(columns={'name':'region_name'})
          top 10 streets.head(10)
Out[45]:
```

	street	obj_count	kod	region_name
0	проспект Мира	204	105	Мещанский район
1	проспект Мира	204	301	Алексеевский район
2	проспект Мира	204	309	район Марьина Роща
3	проспект Мира	204	310	Останкинский район
4	проспект Мира	204	312	район Ростокино
5	проспект Мира	204	313	район Свиблово
6	проспект Мира	204	317	Ярославский район
7	Профсоюзная улица	181	701	Академический район
8	Профсоюзная улица	181	704	район Коньково
9	Профсоюзная улица	181	707	Обручевский район

Теперь мы можем перечислить районы, в которых расположены топ-10 улиц по количеству объектов общественного питания:

```
In [46]: top_10_streets.region_name.sort_values().unique()
Out[46]: array(['Академический район', 'Алексеевский район', 'Гагаринский район',
                                            академическии район , Алексевскии район , Тагаринскии райо , 'Донской район', 'Ломоносовский район', 'Мещанский район', 'Нагорный район', 'Обручевский район', 'Останкинский район', 'Пресненский район', 'Хорошёвский район', 'Ярославский район', госеление Шербинка', 'район Аэропорт', 'район Беговой', 'район Дорогомилово', 'район Коньково', 'район Марьина Роща', 'район Москворечье-Сабурово', 'район Нагатино-Садовники',
                                              'район Орехово-Борисово Северное', 'район Орехово-Борисово Южное',
                                             'район Проспект Вернадского', 'район Раменки', 'район Ростокино', 'район Свиблово', 'район Северное Бутово', 'район Сокол', 'район Тропарёво-Никулино', 'район Тёплый Стан',
                                            раион пропарево-никулино, раион геплыи стан, 
'район Фили-Давыдково', 'район Чертаново Северное', 
'район Чертаново Южное', 'район Чертаново Южное', 
'район Черёмушки', 'район Южное Бутово', 'район Якиманка', 
'район Ясенево'], dtype=object)
```

Отберём топ-5 районов, через которые проходят топ-10 улиц:

```
In [47]: (
                 top_10_streets
                  .groupby(by='region_name')
.agg({'street':'count'})
                  sort_values(by='street', ascending=False)
```

Out[47]:

```
region name
район Нагатино-Садовники
      Обручевский район
район Тропарёво-Никулино
                             2
       район Тёплый Стан
район Чертаново Северное
```

Определение районов Москвы, в которых расположены улицы с единственным объектом общественного питания

Определим, на скольких улицах Москвы находится всего по 1 заведению общественного питания:

```
In [48]: len(
                       rest data
                       les_data
.query('street != "город Зеленоград" & street != "поселение Сосенское"')
.groupby(by='street')
.agg({'id' : 'count'})
.rename(columns={'id':'obj_count'})
                        .query('obj_count == 1')
Out[48]: 579
```

Итак, по одной точке питания расположено предположительно на 579 улицах Москвы, не считая Зеленограда и поселения Сосенское, которые мы ранее исключили из рассмотрения. Тем не менее, другие подобные искажения данных возможны, и нам следует исключить их из рассмотрения:

```
In [49]: # сохраним список улиц с 1 точкой общепита
           street_list = list(
    rest_data
                 .query('street != "город Зеленоград" & street != "поселение Сосенское"')
                 .agg({'id' : 'count'})
.rename(columns={'id':'obj_count'})
                 .query('obj_count == 1')
```

```
index
   # зададим корректные типы улиц
# зистров | Toppes | Str_types | Fig. | Str_types | St
   good_street_list = []
   # отберём из street_list только корректные названия for street in street_list:
                                     for str_type in str_types:
    if street.find(str_type) != -1:
                                                                                                  good_street_list.append(street)
break
 print(len(good_street_list))
```

Итак, 540 улиц в Москве содержат по 1 заведению общественного питания. Отберём для дальнейшего рассмотрения только точки на этих улицах:

```
In [50]: rest_regions = (
              rest data
              .query('street in @good_street_list')
         rest_regions.head()
```

Out[50]:	: id		object_name	chain	object_type	address	number	street
	19	153644	Американская Лаборатория Десертов	False	кафе	город Москва, Филипповский переулок, дом 15/5	20	Филипповский переулок
	61	21214	Столовая	False	столовая	город Москва, Бабаевская улица, дом 3	90	Бабаевская улица
	69	22139	ХуанХэ	False	ресторан	город Москва, 1-й Балтийский переулок, дом 3/25	60	1-й Балтийский переулок
	81	25833	АЗЕРБАЙДЖАН	False	ресторан	город Москва, улица Демьяна Бедного, дом 4	180	улица Демьяна Бедного
	131	150881	вгик	False	столовая	город Москва, Будайская улица, дом 3	45	Будайская улица

Объединим rest_regions и streets_class:

```
In [51]: rest_regions = (
                rest_regions
                 .merge(
streets_class,
                     left_on='street',
right_on='um_namef',
                     how='left'
           rest regions.head()
```

f	um_name	um_te	street	number	address	object_type	chain	object_name	id	L]:
<	Филипповский переуло	101.0	Филипповский переулок	20	город Москва, Филипповский переулок, дом 15/5	кафе	False	Американская Лаборатория Десертов	153644	(
а	Бабаевская улиц	418.0	Бабаевская улица	90	город Москва, Бабаевская улица, дом 3	столовая	False	Столовая	21214	1
	1-й Балтийски переуло	201.0	1-й Балтийский переулок	60	город Москва, 1-й Балтийский переулок, дом 3/25	ресторан	False	ХуанХэ	22139	2
)	улица Демьяна Бедного	906.0	улица Демьяна Бедного	180	город Москва, улица Демьяна Бедного, дом 4	ресторан	False	АЗЕРБАЙДЖАН	25833	3
а	Будайская улиц	312.0	Будайская улица	45	город Москва, Будайская улица, дом 3	столовая	False	вгик	150881	4

Изучим уникальные коды территориальных образований для улиц с одним предприятием питания:

```
In [52]: rest_regions['um_te'].sort_values().unique()
Out[52]: array([ 101., 102., 103., 104., 105., 106., 107., 108., 109.,
                 110., 201.,
                              202., 203.,
                 211., 212.,
                              213., 214.,
                                            215., 217.,
                                                          301.,
                                                                 302..
                                                                        303..
                 305., 307., 308., 309., 310., 311., 312.,
                 315., 316., 317., 403., 405., 406., 407., 408.,
                                                                        409.,
                 410., 411., 413., 414., 415., 416., 417., 418.,
                 505., 506.,
                              507., 508., 509., 510., 511., 512.,
                                                                        513. .
                 514., 515.,
                              606., 607., 609., 610., 612., 613., 615.,
                                                                        708.,
                 618., 619., 621., 702., 703., 704., 705., 707., 708., 709., 710., 711., 712., 805., 807., 809., 811., 813.,
```

905.,

815., 816., 818., 819., 820., 901., 902., 903., 905. 906., 907., 908., 1010., 1102., 1104., 1105., 1106., 1108. 1111., 1201., 1202., 1203., 1204., 1206., 1207., 1209., nan Как видим, в итоговой таблице есть пропуски в кодах районов. Изучим их причины:

```
In [53]: rest regions[rest regions['um te'l.isna()]
```

III [33].		c6_05	cgrangl.coc_regrangl_am_cc_1/12/md//1							
Out[53]:		id	object_name	chain	object_type	address	number	street	um_te	um_namef
	208	23514	ОСИПОВ А.С. (прод.база «Очаково»)	False	столовая	город Москва, проектируемый проезд № 1980, дом 4	60	проектируемый проезд № 1980	NaN	NaN
	494	162646	Кафе «Мастер Ланч»	False	кафетерий	город Троицк, 42-й километр Калужского шоссе,	23	42-й километр Калужского шоссе	NaN	NaN
	567	192097	Ветерок	False	кафе	город Москва, улица Старые Кузьминки, дом 10	40	улица Старые Кузьминки	NaN	NaN

Пропуски наблюдаются всего в 3 строках итогового датафрейма rest_regions :

- 1. Столовая ОСИПОВ А.С. (прод.база «Очаково») расположена на проектируемом проезде № 1980, который по понятным причинам не находится в Общемосковском классификаторе улиц.
- 2. Кафетерий «Мастер Ланч» расположен в г. Троицк и по понятным 42-й километр Калужского шоссе также отсутствует в Общемосковском классификаторе улиц.
- 3. Кафе Ветерок расположено на короткой улице Старые Кузьминки, которая находится в центре парка Кузьминки и, вероятно, по этой причине не попала в Общемосковский классификатор улиц.

Обработаем данные исключительные ситуации следующим образом:

- исключим из данных кафетерий в Троицке;
- данные по остальным двум точкам заполним вручную на основании поиска в Яндекс.Картах:
 - кафе Ветерок расположено в районе Кузьминки (код 506)
 - столовая ОСИПОВ А.С. в районе Очаково-Матвеевское (код 820)

```
In [54]: # удаляем строку с индексом 494 rest_regions.drop(index=[494], axis=0, inplace=True)
             # заполняем кафе Ветерок
            rest_regions.loc[567, 'um_te'] = 506
            # заполняем столовая ОСИПОВ А.С. rest_regions.loc[208, 'um_te'] = 820
            rest_regions.query('id in [23514, 162646, 192097]')
```

Out[54]: id object_name chain object_type address number street um te um namef **208** 23514 ОСИПОВ А.С. (прод.база «Очаково») False столовая город Москва, проектируемый проезд № 1980, дом 4 60 проектируемый проезд № 1980 820.0 NaN Ветерок False **567** 192097 кафе город Москва, улица Старые Кузьминки, дом 10 40 улица Старые Кузьминки 506.0 NaN

```
In [55]: rest_regions['um_te'].isna().sum()
Out[55]: 0
```

Пропусков в кодах районов больше нет. Добавим названия районов в таблицу rest_regions :

```
In [56]: rest_regions = (
              rest_regions
               .merge(
    regions_class,
                   left_on='um_te',
right_on='kod',
                   how='left'
              )
          rest_regions = rest_regions.drop(columns=['um_te', 'um_namef'], axis=1)
                    гновываем сто
          rest_regions = rest_regions.rename(columns={'name':'region_name'})
          rest_regions.head()
```

[56]:		id	object_name	chain	object_type	address	number	street	kod	region_name
	0 1	53644	Американская Лаборатория Десертов	False	кафе	город Москва, Филипповский переулок, дом 15/5	20	Филипповский переулок	101	район Арбат
	1	21214	Столовая	False	столовая	город Москва, Бабаевская улица, дом 3	90	Бабаевская улица	418	район Сокольники
	2	22139	ХуанХэ	False	ресторан	город Москва, 1-й Балтийский переулок, дом 3/25	60	1-й Балтийский переулок	201	район Аэропорт
	3	25833	АЗЕРБАЙДЖАН	False	ресторан	город Москва, улица Демьяна Бедного, дом 4	180	улица Демьяна Бедного	906	район Хорошёво- Мнёвники
	4 1	50881	ВГИК	False	столовая	город Москва, Будайская улица, дом 3	45	Будайская улица	312	район Ростокино

Проверим пропуски районов в итоговой таблице:

```
In [57]: rest_regions['kod'].isna().sum()
Out[57]: 0
```

Пропусков не наблюдается. Теперь мы можем перечислить районы, в которых расположены улицы с единственным объектом общественного питания:

```
In [58]: rest_regions.region_name.sort_values().unique()
```

```
Out[58]: array(['Алексеевский район', 'Алтуфьевский район', 'Бабушкинский район', 'Басманный район', 'Бескудниковский район', 'Бутырский район', 'Войковский район', 'Гагаринский район', 'Головинский район', 'Даниловский район', 'Дмитровский район', 'Донской район', 'Красносельский район', 'Лосиноостровский район', 'Мещанский район', 'Можайский район', 'Молжаниновский район', 'Нагорный район', 'Можайский район', 'Обручевский район', 'Останкинский район', 'Пресненский район', 'Рязанский район', 'Савёловский район', 'Таганский район', 'Тимирязевский район', 'Таганский район', 'Можнопортовый район', 'Ярославский район', 'Тоселение Вороновское', 'поселение Воскресенское', 'поселение Киевский', 'поселение Карсионахорское', 'поселение Карушкинское', 'поселение Марушкинское', 'поселение Московский', 'поселение Новофедоровское', 'поселение Московский', 'поселение Новофедоровское', 'поселение Московский', 'поселение Новофедоровское',
                                                                                                                                                                                     'поселение Московский', 'поселение Новофедоровское' поселение Первомайское', 'поселение Рязановское',
                                                                                                                                                                                   поселение первоманское, поселение Рязановское, поселение Троицк', поселение Шербинка', 'район Арбат', 'район Аэропорт', 'район Беговой', 'район Бирюлёво Восточное', 'район Бирюлёво Восточное', 'район Босточное Измайлово', 'район Восточный',
                                                                                                                                                                                       район Восточное измалиово, район восточный,
'район Выхино-Жулебино', 'район Кольяново', 'район Дорогомилово',
'район Замоскворечье', 'район Западное Дегунино', 'район Зюзино',
                                                                                                                                                                                   район Замоскворечье', 'район Западное Дегунино', 'район Зюзино', 'район Ивановское', 'район Измайлово', 'район Капотня', 'район Коньково', 'район Коптево', 'район Косино-Ухтомский', 'район Котловка', 'район Кузьминки', 'район Кунцево', 'район Куркино', 'район Лефортово', 'район Люблино', 'район Лефортово', 'район Люблино', 'район Марфино', 'район Марьина Роща', 'район Марьино', 'район Матино', 'район Маркино', 'район Матино', 'район Мекворечье-Сабурово', 'район Нагатино-Садовники', 'район Некрасовка', 'район Ново-Переделкино', 'район Новогиреево', 'район Отрадное', 'район Очаково-Матвеевское', 'район Перово', 'район Перображенское', 'район Раменки', 'район Ростокино', 'район Савёлки', 'район Северное Бутово', 'район Северное Измайлово', 'район Северное Музыйлово', 'район Северное Музыйно Коколиная Гора',
                                                                                                                                                                                   район Северное Измайлово', 'район Северное Медведково', 'район Северное Медведково', 'район Соколиная Гора', 'район Сокольники', 'район Солицево', 'район Строгино', 'район Текстильщики', 'район Теплый Стан', 'район Фили-Давыдково', 'район Филевский Парк', 'район Хамовники', 'район Филевский Парк', 'район Хамовники', 'район Фиртаново Южное', 'район Чертаново Северное', 'район Чертаново Южное', 'район Черёмушки', 'район Щукино', 'район Южное Бутово', 'район Южное Можное Можн
                                                                                                                                                                                       'район Ясенево'], dtype=object)
```

Оценим количество районов, в которых находятся улицы с единственной точкой питания:

```
In [59]: len(list(rest_regions.region_name.sort_values().unique()))
Out[59]: 116
         Общее количество районов:
In [60]: len(regions class)
```

Вывод: улицы с недостатком объектов питания расположены практически по всей территории Москвы.

Отберём топ-5 районов, через которые проходят улицы с единственным объёктом питания:

```
In [61]: (
                 rest regions
                  .groupby(by='region_name')
                  .agg({'street':'count'})
.sort_values(by='street', ascending=False)
                  .head()
```

street

Out[60]: 158

region_name						
Таганский район	28					
район Хамовники	26					
Басманный район	25					
Тверской район	20					
Пресненский район	19					

Вывод

Проведённый анализ данных об объектах общественного питания в Москве показал, что:

1. Наибольшее количество точек питания работают в форматах:

```
А. кафе (39.7%);
В. столовая (16.8%);
С. ресторан (14.9%):
D. предприятие быстрого обслуживания (12.5%);
Е. бар (5.57%).
```

- 1. Диаграмма соотношения количества сетевых и несетевых заведений питания свидетельствует о более чем 4-кратном преобладании индивидуальных точек питания над сетевыми.
- 1. В наибольшей степени сетевое распространение характерно для предприятий быстрого обслуживания (фаст-фуд). Кроме того, в значительной степени сетевое распространение характерно также и для отделов кулинарии, ресторанов и кафе.
- 1. Для сетей более характерно небольшое количество заведений. Для сетей с небольшим количеством заведений более характерно большое количество

- 1. В среднем самое большое количество посадочных мест (103) предоставляют столовые, за ними следуют рестораны и, с большим отрывом, бары, буфеты и кафе. Несмотря на то, что максимальное количество мест обеспечивают бары, в большинстве случаев это выбросы.
- 1. Топ-10 улиц по количеству точек общественного питания в Москве выглядит следующим образом:
 - А. проспект Мира
 - В. Профсоюзная улица
 - С. Ленинградский проспект
 - D. Пресненская набережная
 - Е. Варшавское шоссе
 - F. Ленинский проспект
 - G. проспект Вернадского
 - Н. Кутузовский проспект
 - I. Каширское шоссе
 - Кировоградская улица
- 1. Топ-10 улиц по количеству точек общественного питания расположены в следующих районах:

```
'Академический район', 'Алексеевский район', 'Гагаринский район',
'Донской район', 'Ломоносовский район', 'Мещанский район',
'Нагорный район', 'Обручевский район', 'Останкинский район'
'Пресненский район', 'Хорошёвский район', 'Ярославский район',
'поселение Щербинка', 'район Аэропорт', 'район Беговой',
'район Дорогомилово', 'район Коньково', 'район Марьина Роща',
'район Москворечье-Сабурово', 'район Нагатино-Садовники',
'район Орехово-Борисово Северное', 'район Орехово-Борисово Южное',
'район Проспект Вернадского', 'район Раменки', 'район Ростокино',
'район Свиблово', 'район Северное Бутово', 'район Сокол',
'район Тропарёво-Никулино', 'район Тёплый Стан',
район Фили-Давыдково', 'район Чертаново Северное',
'район Чертаново Центральное', 'район Чертаново Южное'
'район Черёмушки', 'район Южное Бутово', 'район Якиманка',
'район Ясенево'
```

2. Районы, в которых расположены улицы с единственным объектом общественного питания:

'Алексеевский район', 'Алтуфьевский район', 'Бабушкинский район', 'Басманный район', 'Бескудниковский район', 'Бутырский район', 'Войковский район', 'Гагаринский район', 'Головинский район', 'Даниловский район', 'Дмитровский район', 'Донской район', 'Красносельский район', 'Лосиноостровский район', 'Мещанский район', 'Можайский район', 'Молжаниновский район', 'Нагорный район', 'Нижегородский район', 'Обручевский район', 'Останкинский район', 'Пресненский район', 'Рязанский район', 'Савёловский район', 'Таганский район', 'Тверской район', 'Тимирязевский район', 'Хорошёвский район', 'Южнопортовый район', 'Ярославский район', 'поселение Вороновское', 'поселение Вокресенское', 'поселение Киевский', 'поселение Кленовское', 'поселение Кокошкино', 'поселение Краснопахорское', 'поселение Марушкинское', 'поселение Московский', 'поселение Новофедоровское', 'поселение Первомайское', 'поселение Рязановское', 'поселение Троицк', 'поселение Щербинка', 'район Арбат', 'район Аэропорт', 'район Беговой', 'район Бирюлёво Восточное', 'район Бирюлёво Западное', 'район Богородское', 'район Внуково', 'район Восточное Измайлово', 'район Восточный', 'район Выхино-Жулебино', 'район Гольяново', 'район Дорогомилово', 'район Замоскворечье', 'район Западное Дегунино', 'район Зюзино', 'район Ивановское', 'район Измайлово', 'район Капотня', 'район Коньково', 'район Коптево', 'район Косино-Ухтомский', 'район Котловка', 'район Кузьминки', 'район Кунцево', 'район Куркино', 'район Левобережный', 'район Лефортово', 'район Люблино', 'район Марфино', 'район Марьина Роща', 'район Марьино', 'район Метрогородок', 'район Митино', 'район Москворечье-Сабурово', 'район Нагатино-Садовники', 'район Некрасовка', 'район Ново-Переделкино', 'район Новогиреево', 'район Отрадное', 'район Очаково-Матвеевское', 'район Перово', 'район Печатники', 'район Покровское-Стрешнево', 'район Преображенское', 'район Раменки', 'район Ростокино', 'район Савёлки', 'район Свиблово', 'район Северное Бутово', 'район Северное Измайлово', 'район Северное Медведково', 'район Северный', 'район Сокол', 'район Соколиная Гора', 'район Сокольники', 'район Солнцево', 'район Строгино', 'район Текстильщики', 'район Тёплый Стан', 'район Фили-Давыдково', 'район Филёвский Парк', 'район Хамовники', 'район Хорошёво-Мнёвники', 'район Царицыно', 'район Чертаново Северное', 'район Чертаново Южное', 'район Черёмушки', 'район Щукино', 'район Южное Бутово', 'район Южное Медведково', 'район Южное Тушино', 'район Якиманка', 'район Ясенево'

1. Улицы с недостатком объектов питания расположены практически по всей территории Москвы.

Общий вывод исследования и рекомендации

Проведённое исследование посвящено изучению рынка общественного питания Москвы с целью найти место для инновационного проекта - кафе с роботизированным обслуживанием - и включает в себя:

- обзор данных;
- предварительную обработку данных;
- обогащение данных с использованием внешних открытых источников;
- анализ особенностей рынка рынка общественного питания.

Основные выводы по каждому этапу приведены в соответствующих разделах

Основываясь на них, можно сделать следующие выводы и дать следующие рекомендации:

- 1. Наиболее популярной формой заведения общественного питания в Москве является кафе. Популярность её обусловлена, по-видимому, наиболее оптимальным балансом характеристик бизнеса: размер помещения, размер инвестиций, численность команды, соотношение прибыли и затрат и т.д. Поэтому представляется рациональным начать бизнес с открытия кафе.
- 1. На первых порах обособленное кафе выглядит предпочтительнее создание с нуля сети заведений требует огромных вложений, которые могут не окупиться, а существующие сети вряд-ли с энтузиазмом воспримут инновационную идею роботизированного обслуживания.
- 1. Если в перспективе всё же рассмотреть развитие в виде сети роботизированных кафе, то рациональнее всего выбрать формат фаст-фуда:
 - во-первых, ввиду ограниченного ассортимента, приготавливаемого из полуфабрикатов, данный формат проще реализовать и контролировать в виде сети;
 - во-вторых, по тем же причинам, упрощается программное обеспечение обслуживающих посетителей роботов, что приведёт к снижению затрат на их разработку и производство.
- 1. Общее количество мест в пилотном заведении целесообразно выбрать равным среднему для формата кафе 30 мест. По ходу его эксплуатации будут уточнены способности обслуживающих роботов по работе с различным количеством клиентов, а также различным ассортиментом. Кроме того, количество посадочных мест напрямую влияет на площадь арендуемого помещения (затраты на этапе пилотного проекта имеют особый вес). При необходимости, для второго и последующих заведений сети количество мест и площадь помещения могут быть откорректированы в любую сторону - от среднестатистического фастфуда (в среднем, 5 мест) до ресторана (в среднем, 80 мест).

- 1. На улицах, входящих в топ-10 по количеству заведений питания, расположенных на них, наиболее высока конкуренция, но и наболее велика проходимость (поток потенциальных клиентов). В случае инновационного высокотехнологичного бизнеса повышены риски не выдержать конкурентную борьбу в долгосрочной перспективе, когда "вау"-эффект от роботизированного обслуживания птускнеет. Поэтому значительно рациональнее выбрать для размещения пилотного кафе одну из улиц с единственным заведением общественного питания, близлежащих к улицам, входящим в топ-10. Таким образом, за счёт некоторого снижения проходимости может существенно снизиться конкуренция, что будет поддерживать интерес клиентов в период становления и развития бизнеса.
- 1. Районы поиска улицы размещения пилотного заведения определяются пересечением множеств районов, в которых расположены топ-10 улиц, и районов, в которых расположены улицы с единственной точкой питания. Для точного выбора улицы требуются дополнительные исследования с применением картографических сервисов.

Презентация для партнёров

Презентация для визуального сопровождения озвучивания выводов и рекомендаций размещена в файле формата PDF по следующей ссылке.