Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»



**Отчет**

**По лабораторной работе №1**

**По курсу «Технологии машинного обучения»**

**ИСПОЛНИТЕЛЬ:**

Группа ИУ5-65Б

Гурова М.Д.

Москва 2023

**СОДЕРЖАНИЕ**

ВВЕДЕНИЕ 3

1. Определение данных для анализа 4
2. Описание данных 5
3. Изучение общей информации 6
4. Изучение общей информации 7
5. Агрегирование данных 8
6. Исследовательский анализ данных 12

6.1 Доля звездных ресторанов в гиде Мишлен 12

6.2 Регионы с наибольшим количеством звезд Мишлен 12

6.3 Топ городов с наибольшим количеством звездных ресторанов 14

6.4 Определение самой популярной кухни 15

6.5 Расположение ресторанов по миру 16

6.6 Самые популярные тэги в названиях ресторанов 17

6.7 Кухня 17

6.8 Расположение ресторанов самой популярной кухни 19

1. Общий вывод 20

# ВВЕДЕНИЕ

В наше работе нам нужно исследовать выбранный набор данных.

В исследование данных входит работа с данными в Python и выявление статистических закономерностей, возникающих в них.

Цели:

1. определение данных;
2. формулирование гипотез;
3. загрузка данных в Python;
4. проверка данных;
5. очистка данных;
6. преобразование данных;
7. выбор данных для анализа;
8. агрегирование данных;
9. визуализация данных;
10. подтверждение или опровержение поставленных гипотез;
11. формулировка ограничений и выводов.

# Определение данных для анализа

В качестве данных для анализа был выбран dataset «Рестораны гида Мишлен 2018-2019 годов», который является актуальным и интересным для людей, работающих в ресторанной и рекламной сфере, так как попадание в этот гид для ресторана равно увеличению потока гостей, инвесторов, приглашению к участию в различных премиях, то есть в ресторанной сфере – это самый большой успех.

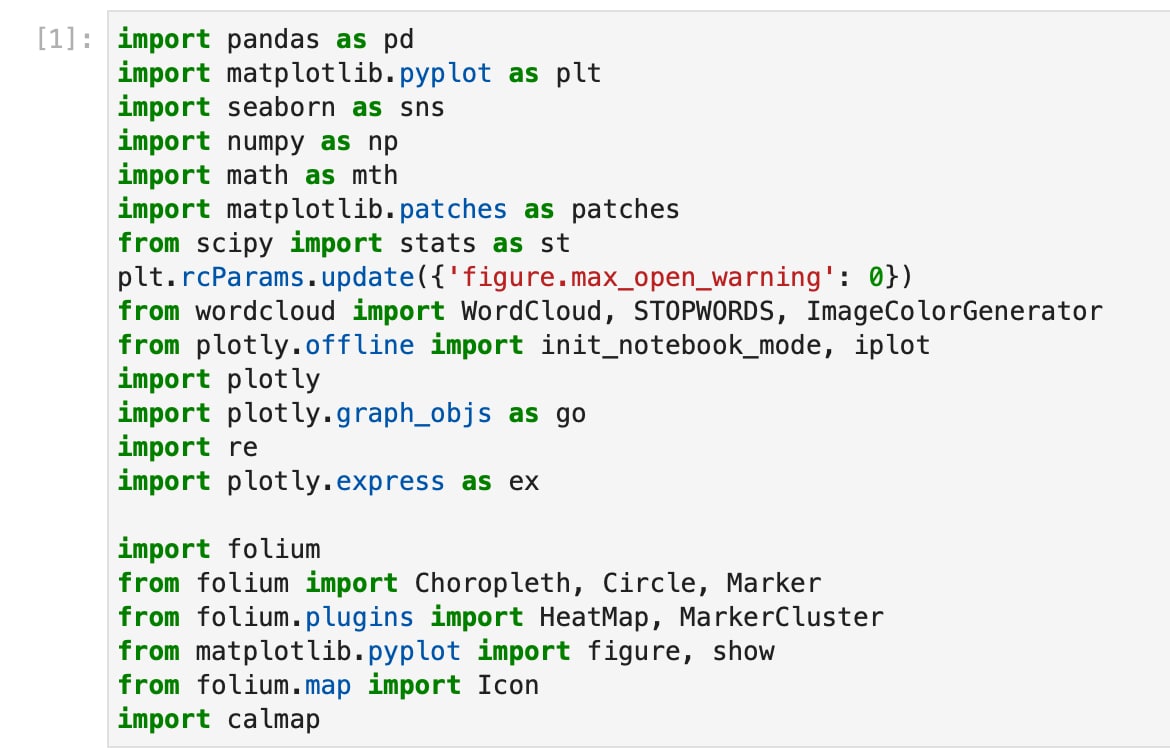
В наборе данных, который был выбран для анализа представлены списки ресторанов, попавших в гид Мишлен в 2018 и 2019 годах, то есть получили большое признание в ресторанной сфере, с информацией о каждом ресторане – названием, типом кухни и основными их характеристиками.

# Описание данных

Для анализа были собраны данные обо всех ресторанах, включенных в гид. В наборе данных содержатся:

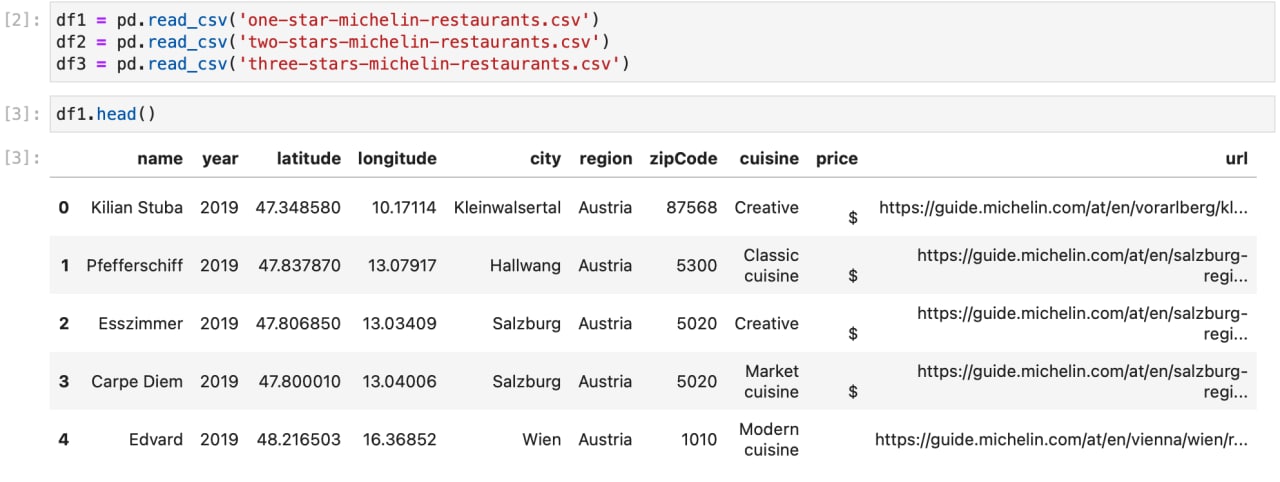
1. name – название ресторана;
2. year – год попадания в гид;
3. latidute – широта;
4. longitude – долгота;
5. city – город;
6. region – регион;
7. zipCode – индекс;
8. cuisine – кухня;
9. price – ценовой диапазон;
10. url – ссылка.
11. Изучение общей информации

В начале работы нам нужно подключить необходимые нам для работы библиотеки.

****

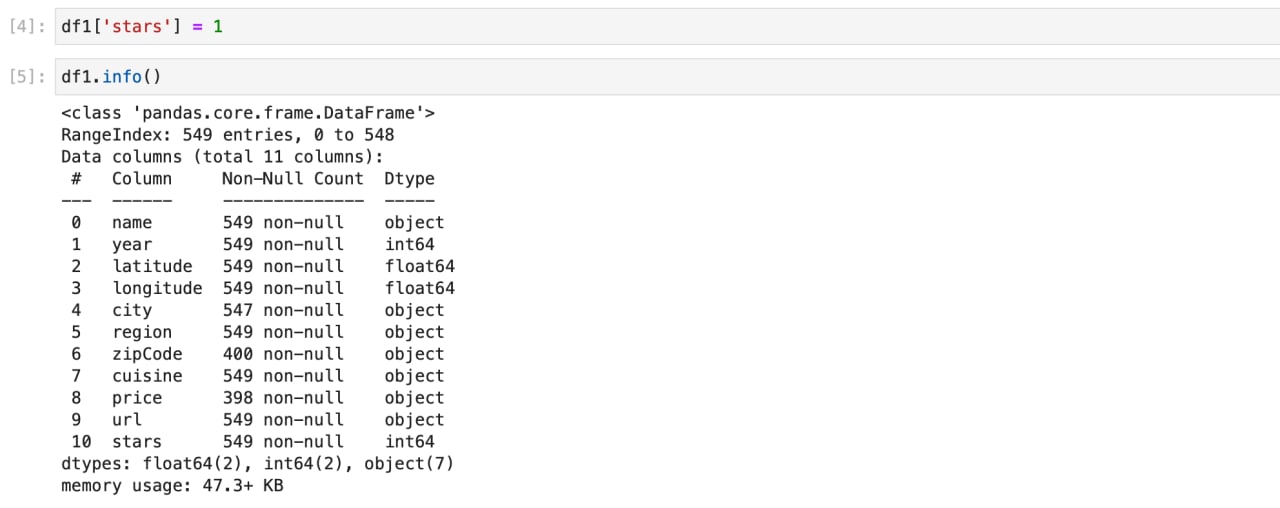
1. Изучение общей информации

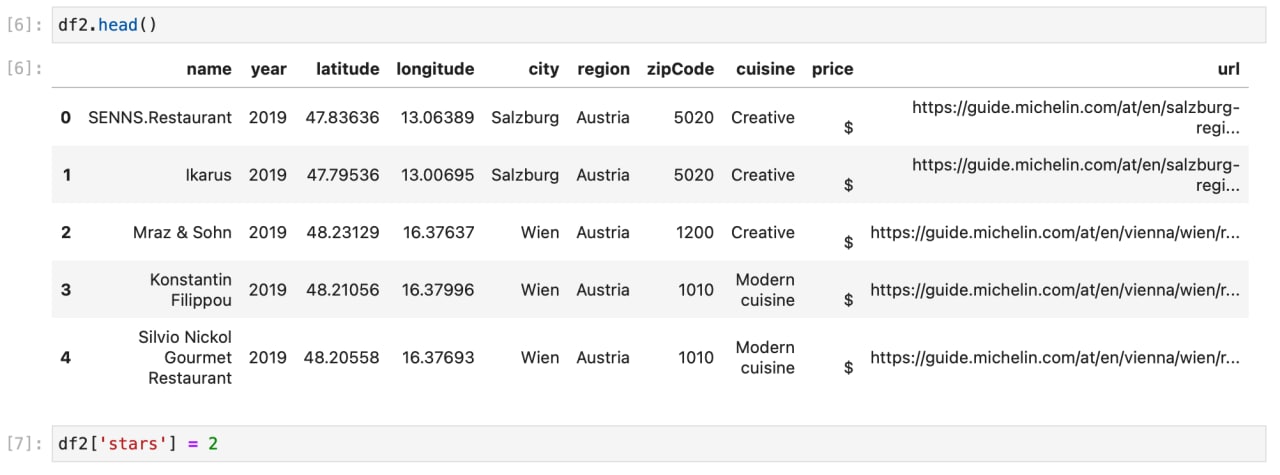
Нам необходимо загрузить наш датасет, а именно три таблицы с ресторанами, в которых содержится вся наша информация:

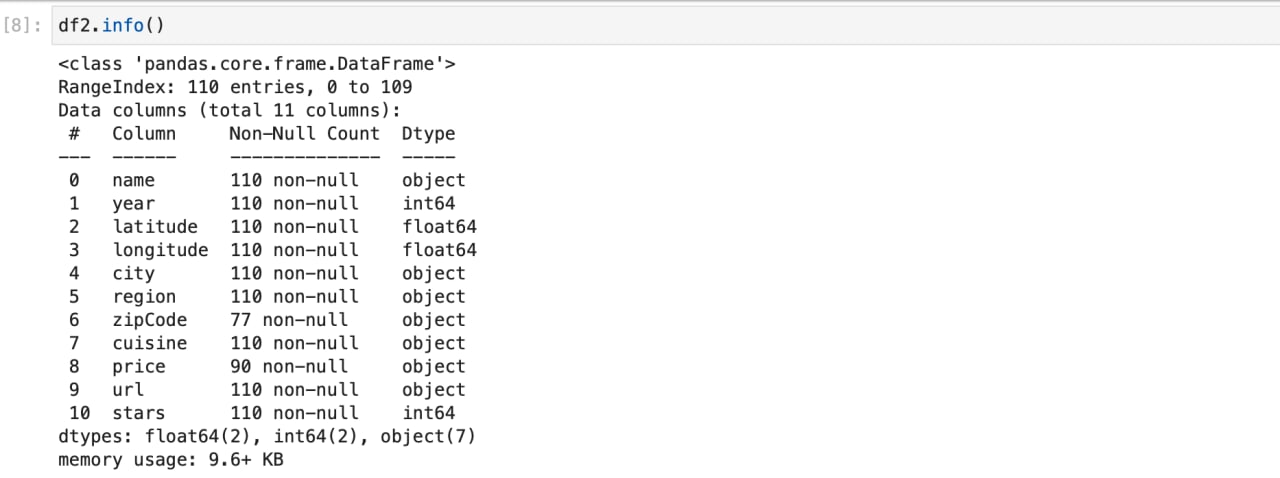
****

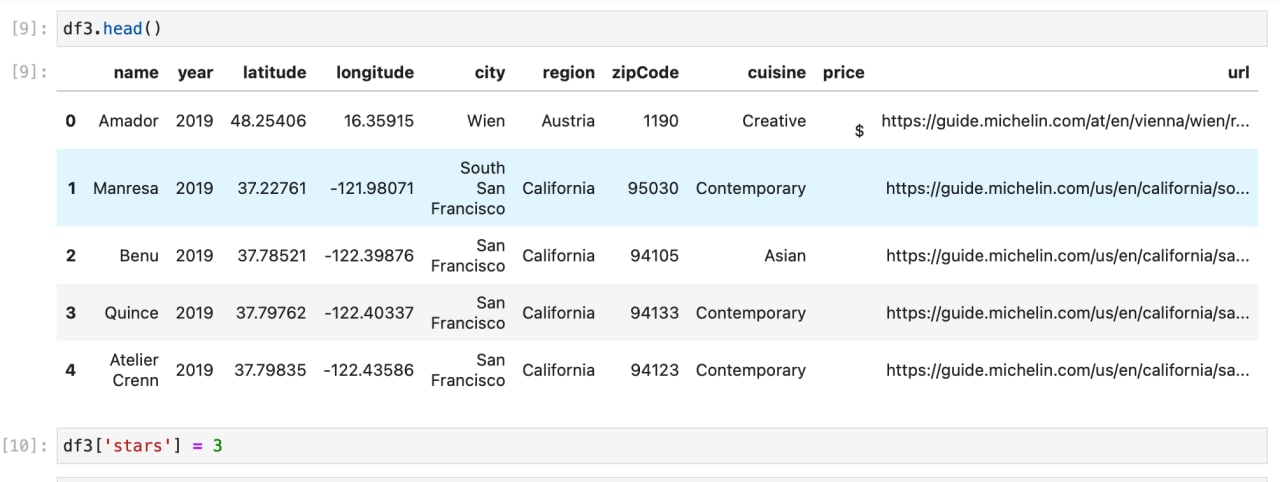
1. Агрегирование данных

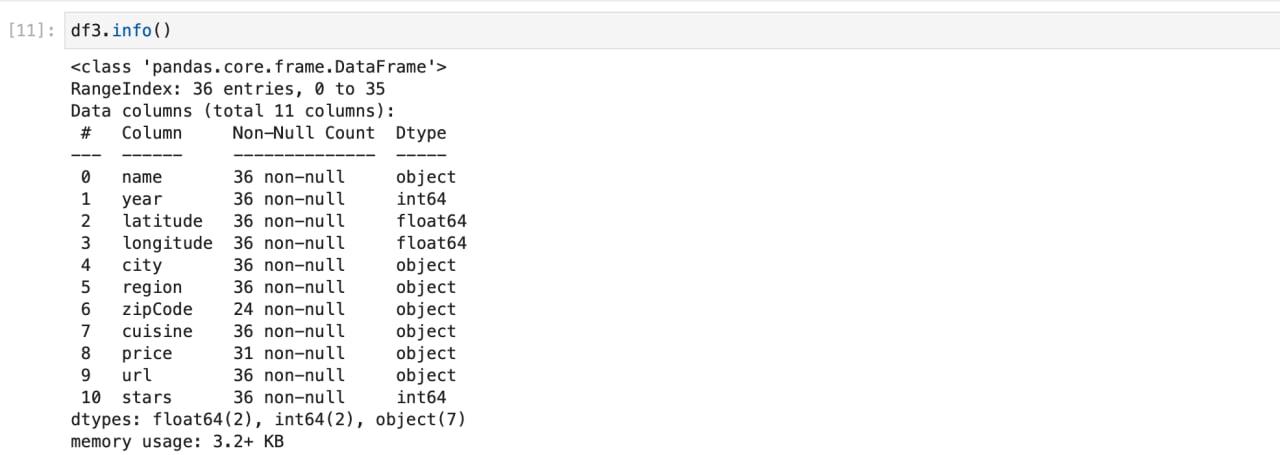
Особенностью датасета является то, что мы можем соединить три таблицы в одну, добавив столбец stars, так как во всех таблицах столбцы одинаковые.

****

****

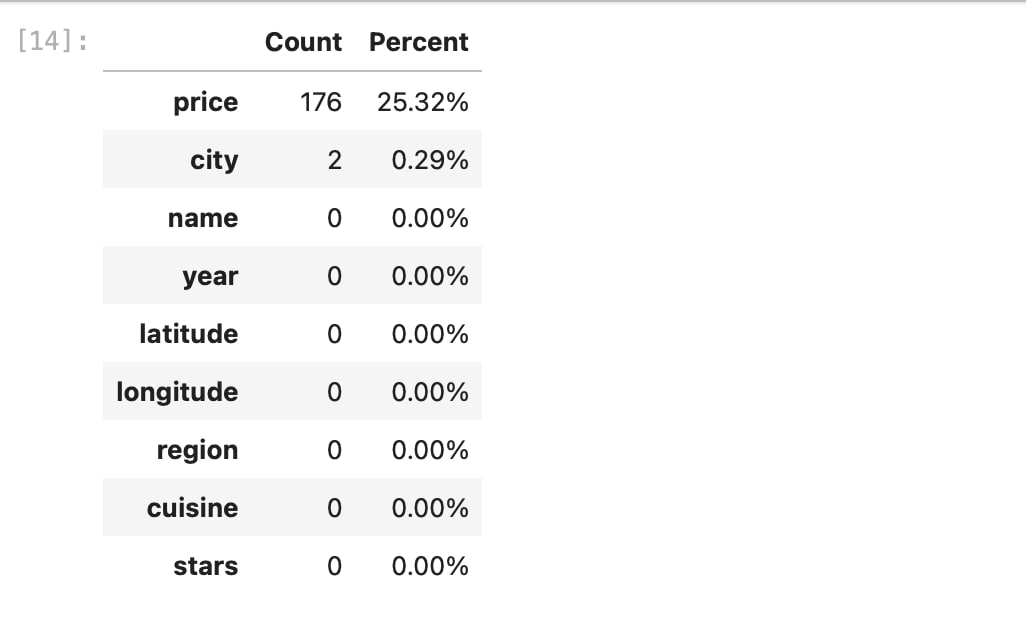
****



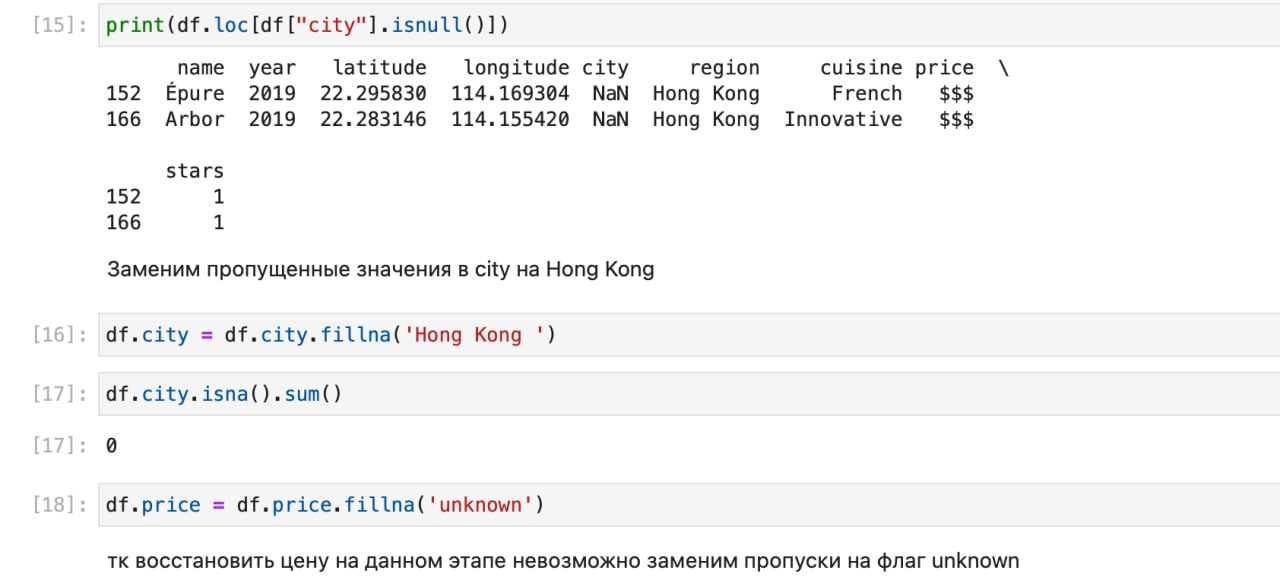


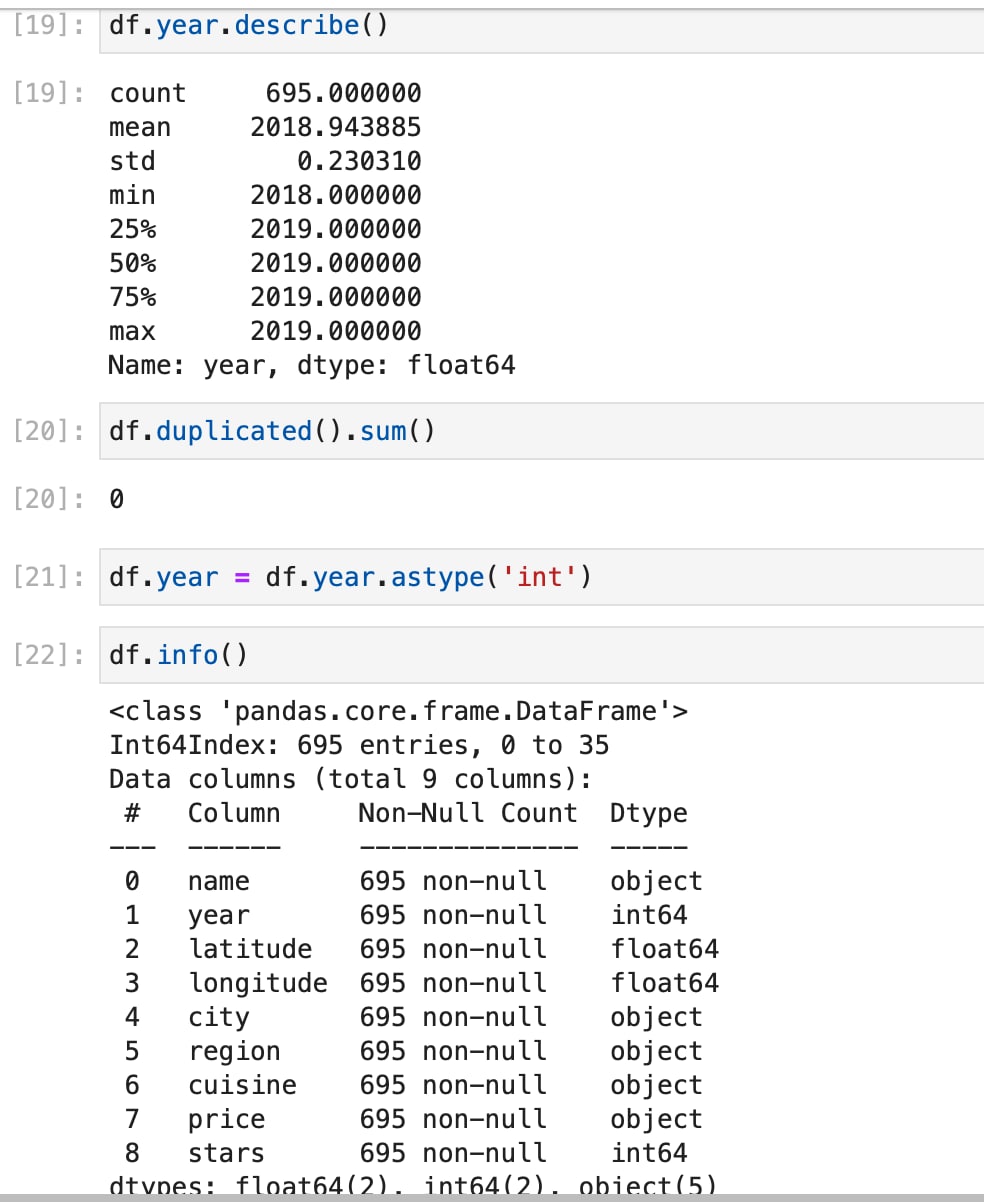
Так же заметим, что столбцы url и zipCode не интересны нам для анализа, поэтому не будем их использовать в нашем датасете.****

Подсчитаем количество пустых строк.



Видим, что основной процент пропусков приходится на цену.





Видим, что данные у нас за период 2018-2019 года. Посмотрим, как изменялось количество ресторанов за этот период.

****

**Вывод:**

1. нулевых данных не имеется;
2. все данные уникальные.
3. Исследовательский анализ данных

### 6.1 Доля звездных ресторанов в гиде Мишлен

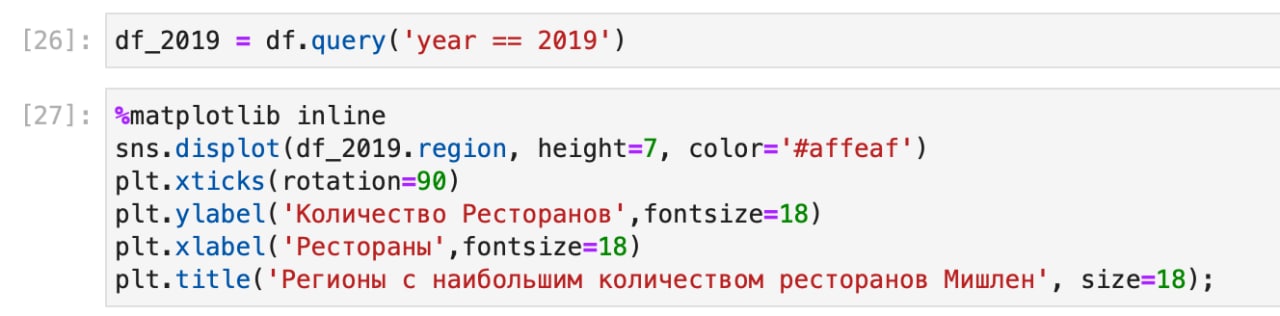


Рисунок 6.1 – Доля ресторанов

**Вывод:**

По диаграмме на рисунке 6.1 можно заметить, что 79% в нашем наборе данных представляют рестораны Мишлен с 1 звездой, и очень небольшое количество представляет рестораны Мишлен с 3 звездами, конечно, из-за требований получения звезд.

### 6.2 Регионы с наибольшим количеством звезд Мишлен



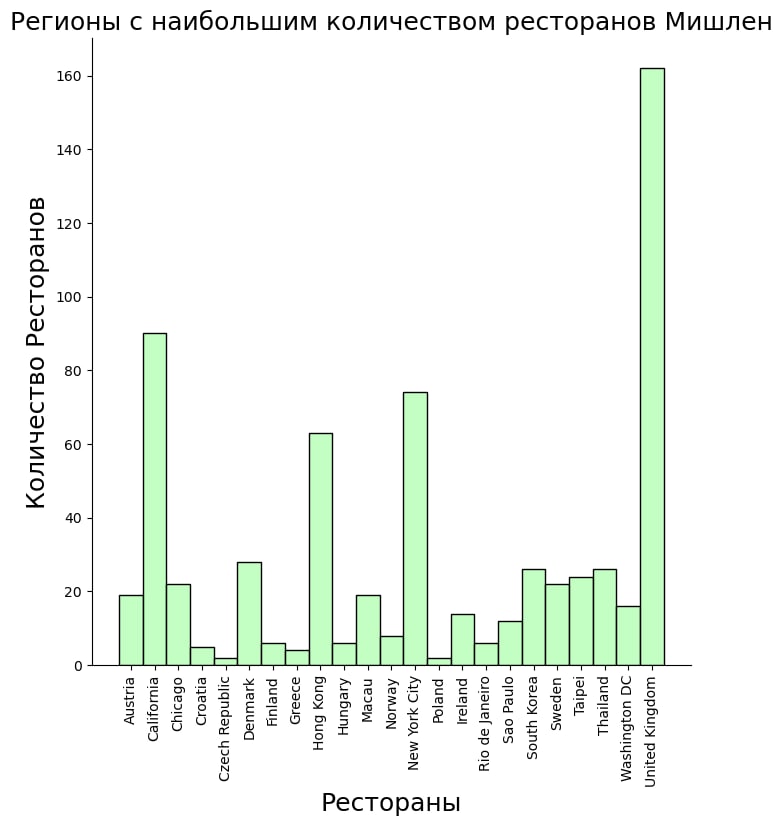


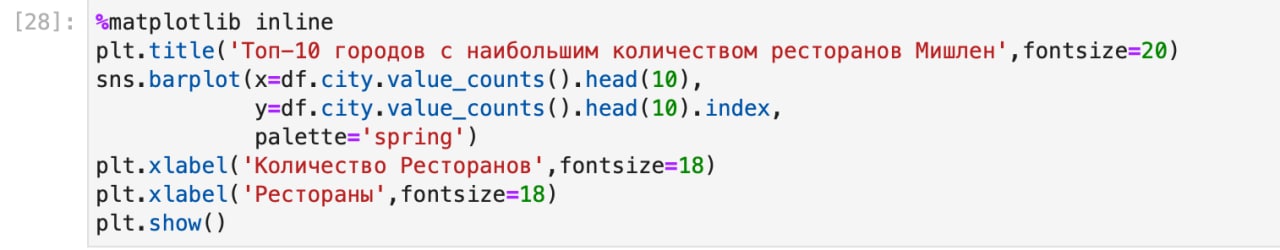
Рисунок 6.2 – Распределение ресторанов по регионам

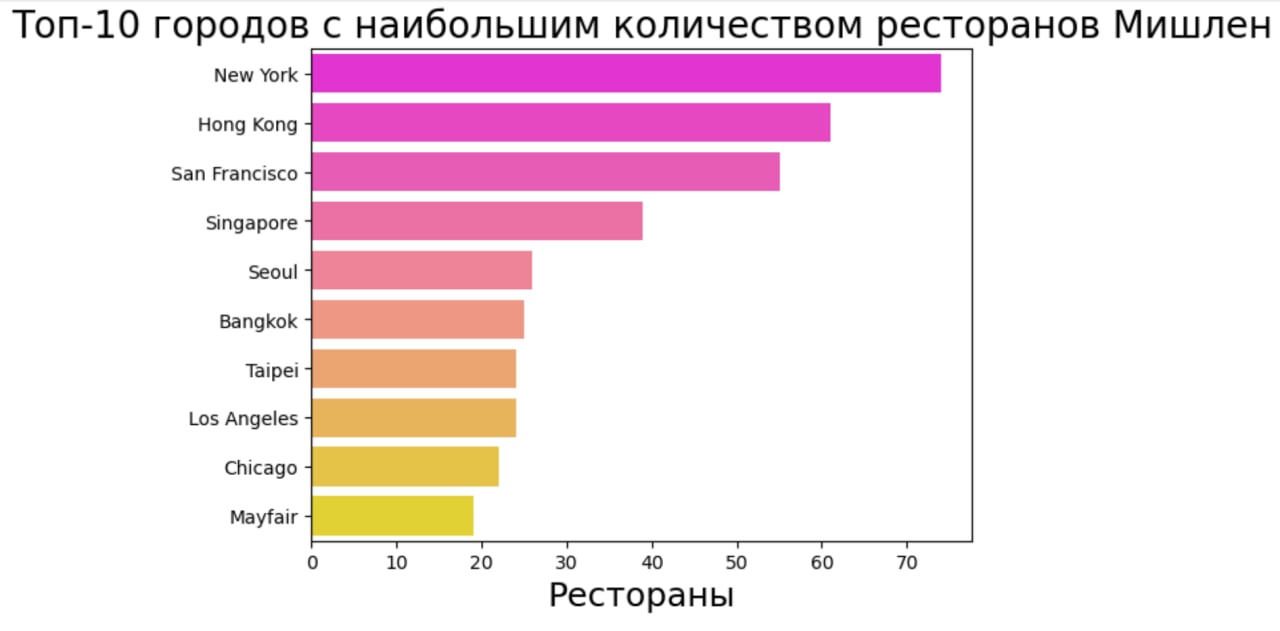
**Вывод:**

Из графика на рисунке 6.2 видно, больше всего ресторанов находится в развитых обеспеченных странах – Англии, США, Гонконге, так как там высокая покупательная способность населения и много туристов.

### 6.3 Топ городов с наибольшим количеством звездных ресторанов

Рассмотрим, в каких городах находится наибольшее количество ресторанов, попавших в рейтинг гида Мишлен.





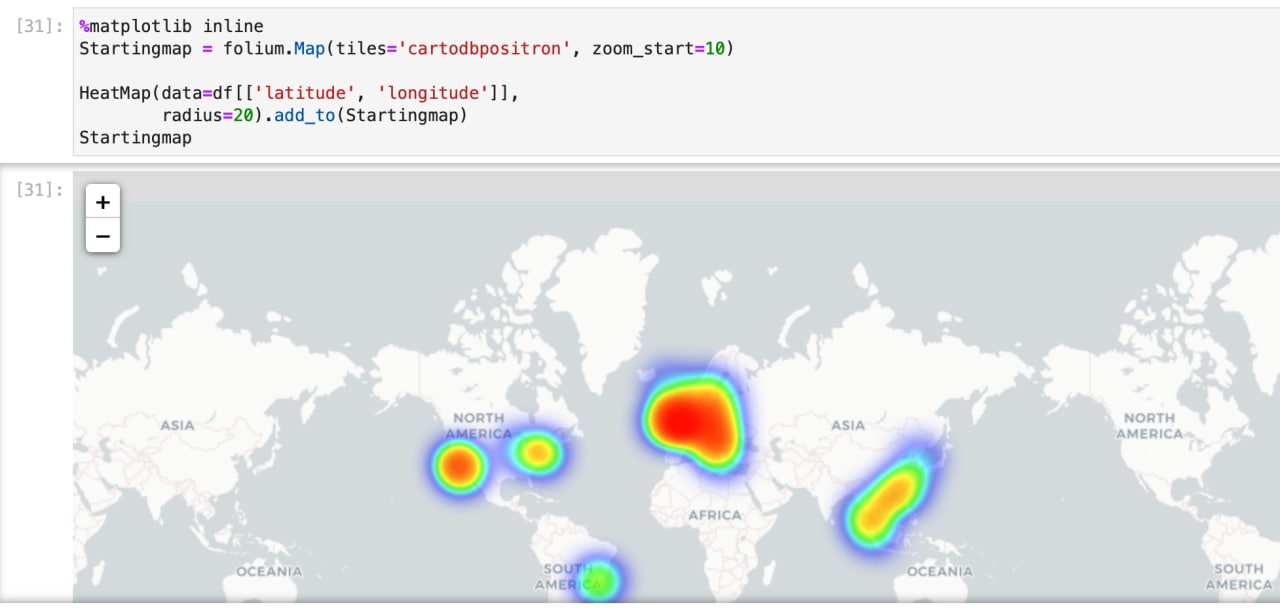


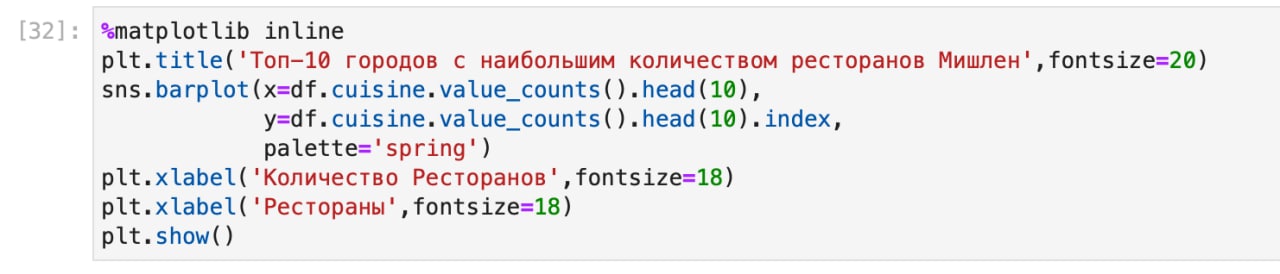
Рисунок 6.3 – Топ городов

**Вывод:**

Согласно диаграмме на рисунке 6.3 больше всего ресторанов в Нью-Йорке и Гонконге.

### 6.4 Определение самой популярной кухни

Определим, какая кухня самая популярная, среди ресторанов, имеющих звезды Мишлен.



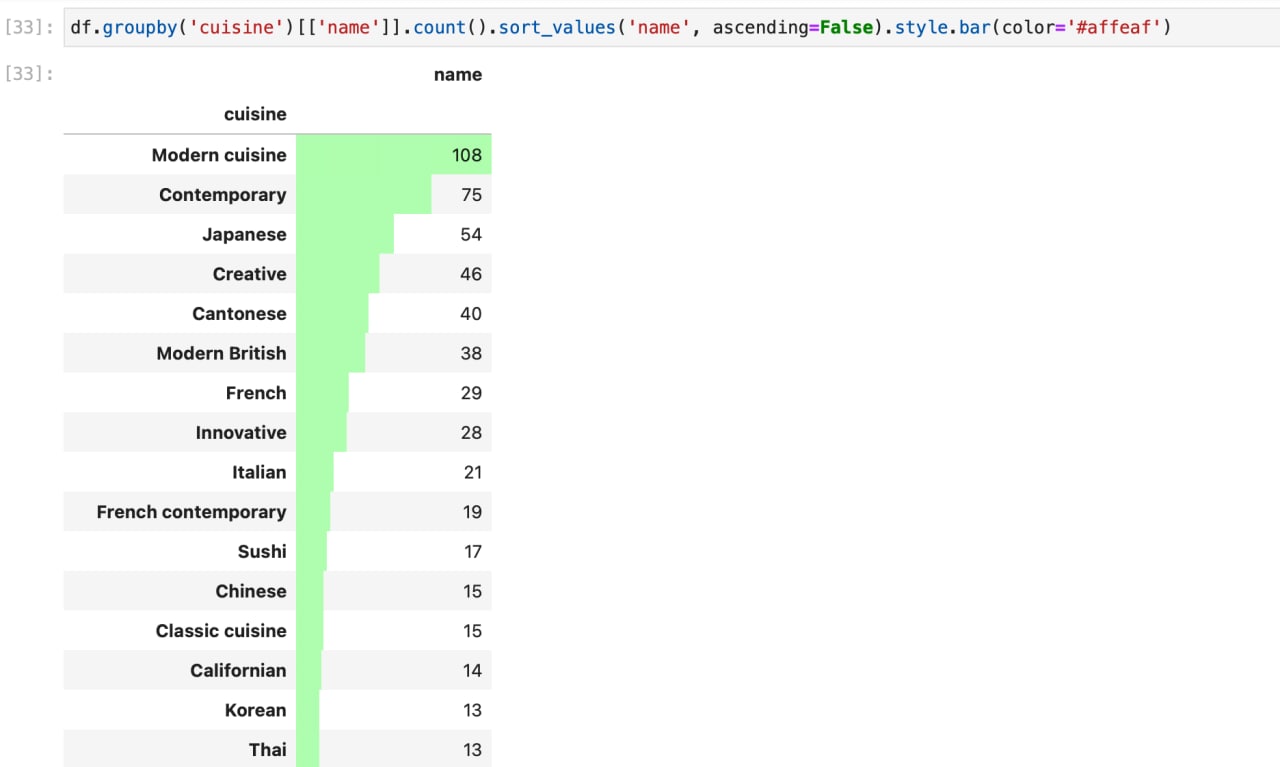


Рисунок 6.4 – Топ кухонь

**Вывод:**

Глядя на приведенный выше график, мы можем заметить, что кухни, которые являются наиболее популярными в ресторанах Мишлен. Данные удивляют, популярные по всему миру кухни - французские, итальянские, британские находятся в самом низу мишленовского рейтинга. Топ-3 составляют современная кухня и японская. Как мы видим наше предположение оказалось частично верно: Итальянская кухня оказалась предпоследней, Японская на 3 месте, а Китайская кухня вовсе отсутствует в нашем топе.

### 6.5 Расположение ресторанов по миру

Рассмотрим, как расположены рестораны из гида Мишлен территориально.

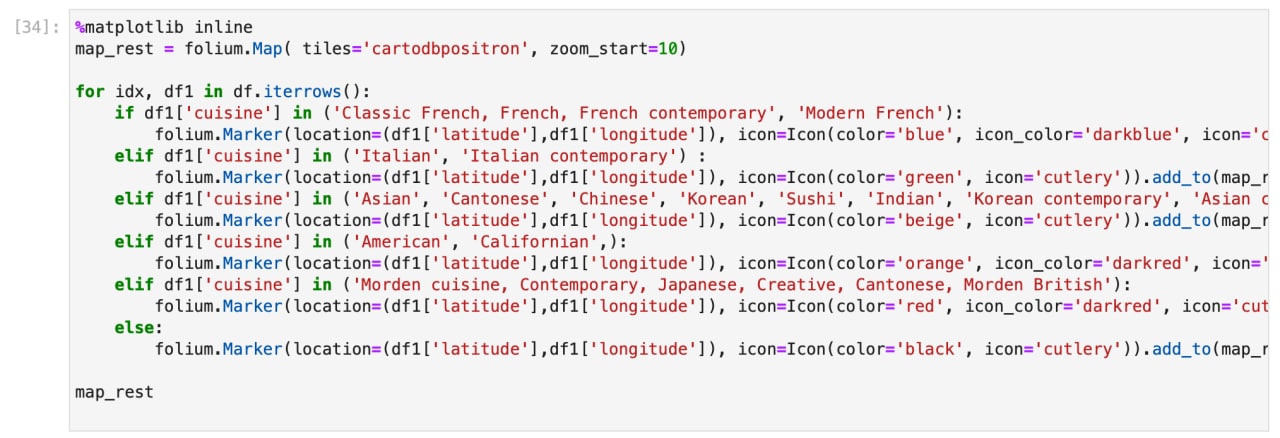




Рисунок 6.5 – Расположение ресторанов популярных видов кухни по миру

### 6.6 Самые популярные тэги в названиях ресторанов



Рисунок 6.6 – Самые популярные тэги

**Вывод:**

Хоть и считается, что ресторан с звездой Мишлен должен иметь уникальное название, как и кухню, но все равно можно выявить самые популярные теги, которые встречаются в имени ресторана: sushi (что не удивительно, ведь японская кухня входит в топ-3 популярных кухонь).

### 6.7 Кухня



Выделим из данного списка ряд категорий кухонь "Вегетарианский", "Интернациональный", "Стейкхаус", "Гастропаб", "Средиземноморский", "Морепродукты", "Суши", "современный", "креативный", "Уличная еда", "Мясо и гриль", "Европейская", "Региональная кухня". Они явно выделяются из общего списка.

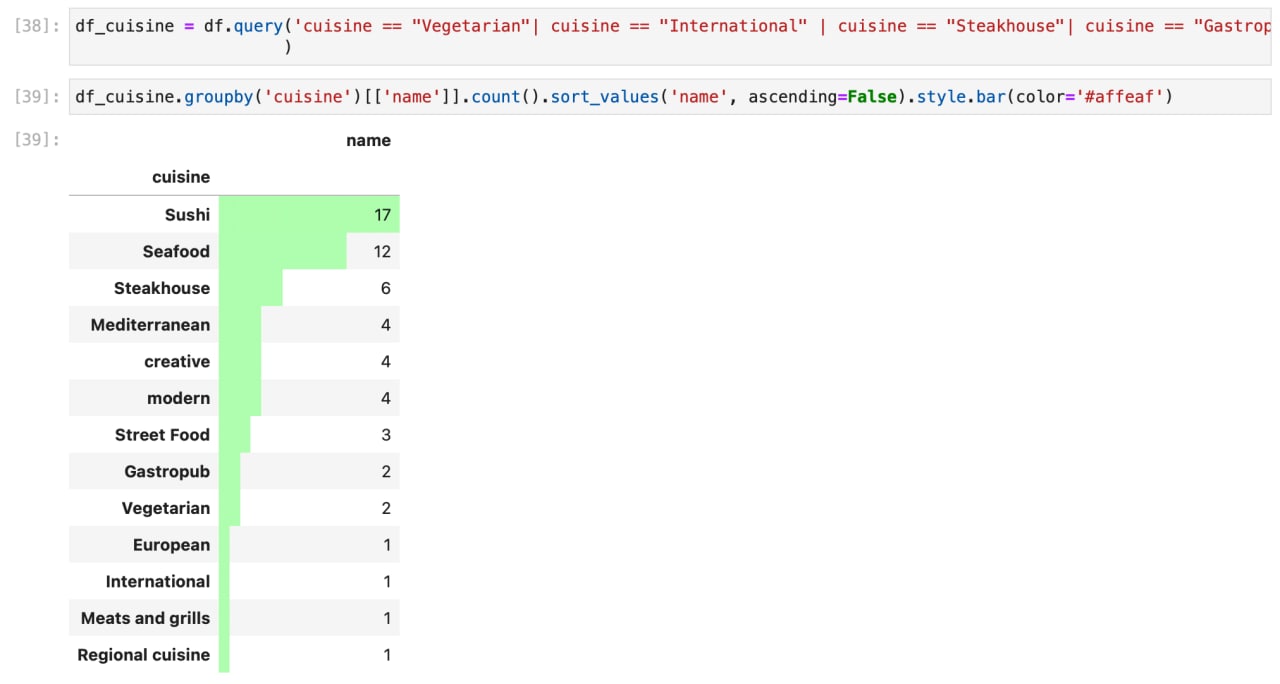


Рисунок 6.7 – Кухня





Рисунок 6.8 – Расположение видов кухни по миру

### 6.8 Расположение ресторанов самой популярной кухни

Сделаем вывод, как расположены рестораны Суши.

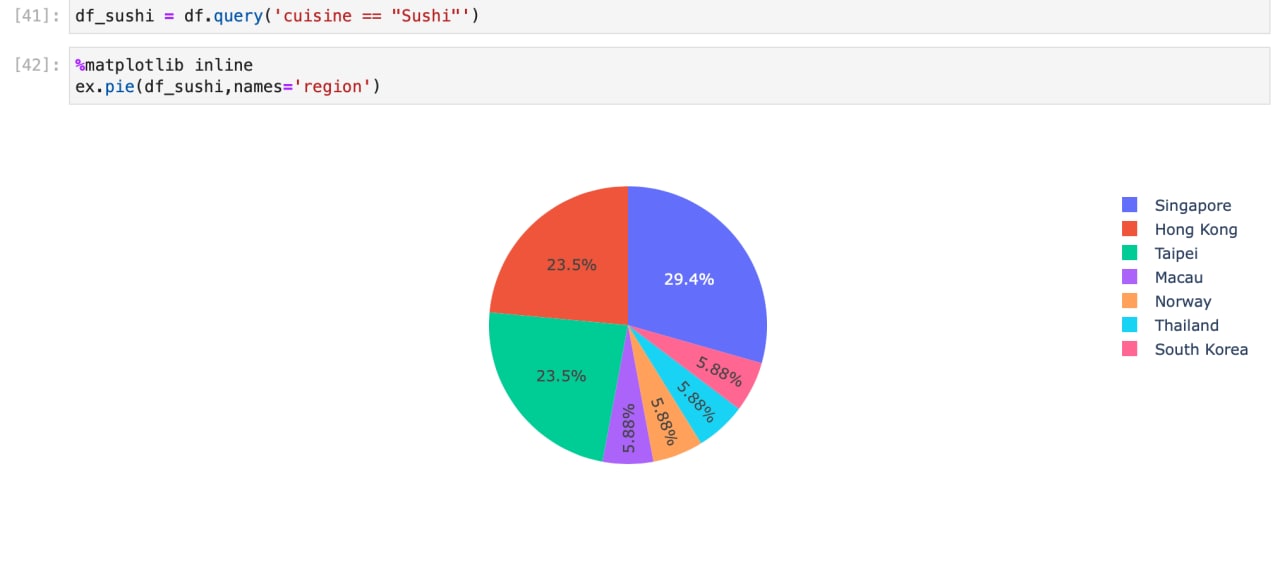


Рисунок 6.9 – Расположение суши-ресторанов

**Вывод:**

Как мы и предполагали большинство Мишлен - ресторанов с кухней суши располагаются в Азии. Это доказывает и карта и круговая диаграмма (94,12%).

1. Общий вывод

Перед анализом мы оптимизировали наш датасет с помощью соединения трех таблиц в одну, а также избавления его от ненужных столбцов. Также мы избавились от нулевых значений и выяснили, что все данные уникальны, что позволило нам провести необходимый анализ и получить знания по предметной области «Рестораны гида Michelin».

В ходе анализа мы выяснили:

1. больше всего имеется однозведных ресторанов, а меньше всего трехвездочных;
2. в топ-3 кухни входят современная и японская кухни;
3. в топ-3 городов входят Нью-Йорк с 74 ресторанами, Гонконг с 61 рестораном и Сан-Франциско с 55 ресторанами.