1.Замените пробел в названиях на знак нижнего подчеркивания.

2. Вас попросили узнать, сколько записей (строк) в датафрейме относятся к району Манхэттен (Manhattan). Получить ответ на данный вопрос можно было бы, например, с помощью следующей команды:

taxi.query("borough == 'Manhattan'").shape[0]

А если мы хотим посмотреть, сколько раз встречается каждый из районов? Неужели придется использовать подобную конструкцию для каждого уровня переменной? Здесь на помощь приходит метод value\_counts()

Сколько раз в данных встречается район Бруклин (Brooklyn)?

3.  выяснить, из какого района было совершено наибольшее количество поездок за весь период. Для начала, посчитайте общее количество поездок (pickups), без группировки.

4. А теперь сгруппируйте данные по району (borough) и укажите, из какого пункта было совершено больше всего поездок.

5. Еще несколько полезных методов в pandas! Возможно, в предыдущем степе вы просто посмотрели на данные или отсортировали значения. Ускорить данный процесс в дальнейшем помогут методы idxmin() и idxmax(), которые возвращают индекс минимального или максимального значения.

[idxmin](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.idxmin.html?highlight=idxmin#pandas.Series.idxmin) – индекс минимального значения

[idxmax](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.idxmax.html) – индекс максимального значения

Сохраните название района с наименьшим числом поездок в переменную min\_pickups, применив подходящий метод.

6. Продолжим изучение данных и посмотрим на число поездок в выходные дни. Сгруппируйте данные по двум признакам: району города и является ли день выходным (колонки borough и hday). Сравните среднее число поездок, и выберите районы, из которых по праздникам в среднем поступает больше заказов, чем в обычные дни.

7. Для каждого района посчитайте число поездок по месяцам. Отсортируйте полученные значения по убыванию и сохраните результирующий датафрейм в pickups\_by\_mon\_bor.

Обратите внимание, что итоговый датасет должен состоять из 3-х колонок - pickup\_month, borough, pickups.

Данные сохранены в переменную taxi.

8. Теперь задача посложнее! Попрактикуемся в написании собственных функций и их применении к датафрейму. Поскольку данные о поездках в Нью-Йорке, температура представлена в градусах Фаренгейта.

Напишите функцию temp\_to\_celcius, которая получает на вход колонку с температурой в °F, и возвращает значения, переведенные в градусы Цельсия.

Формула:

Пример работы функции:

> taxi['temp\_C'] = temp\_to\_celcius(taxi['temp'])

> taxi['temp\_C'][:5]

0 -1.111111

1 -1.111111

2 -1.111111

3 -1.111111

4 -1.111111