**Федеральное государственное бюджетное образовательное учреждение профессионального образования**

**Нижегородский государственный**

**технический университет имени Р.Е. Алексеева**

**Лабораторная работа**

“ **Предварительная обработка данных**”

Работу выполнил

Воробьев М.И.

Работу принял

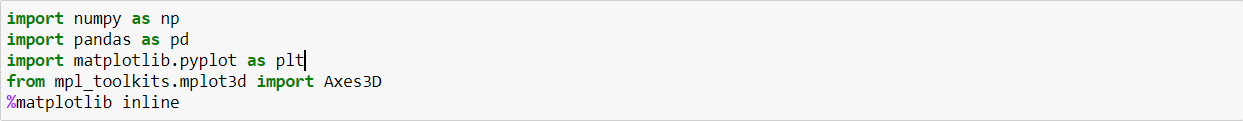
Гай В. Е.

г. Нижний Новгород

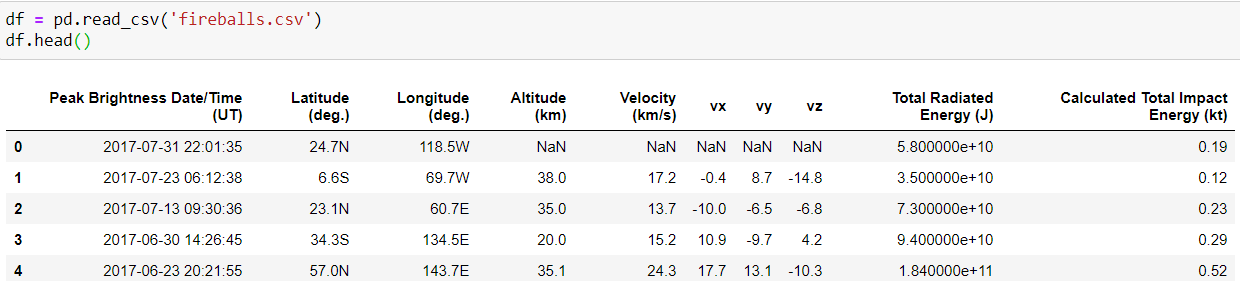
2020 г.

**Цель лабораторной работы** – получение навыков работы с библиотеками анализа данных, предварительной обработки данных, средствами визуализации.

**Вариант 5. Метеоритная защита земли**

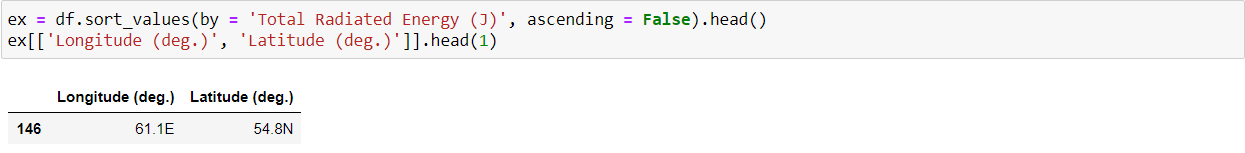
****

Для начало импортируем все необходимые библиотеки



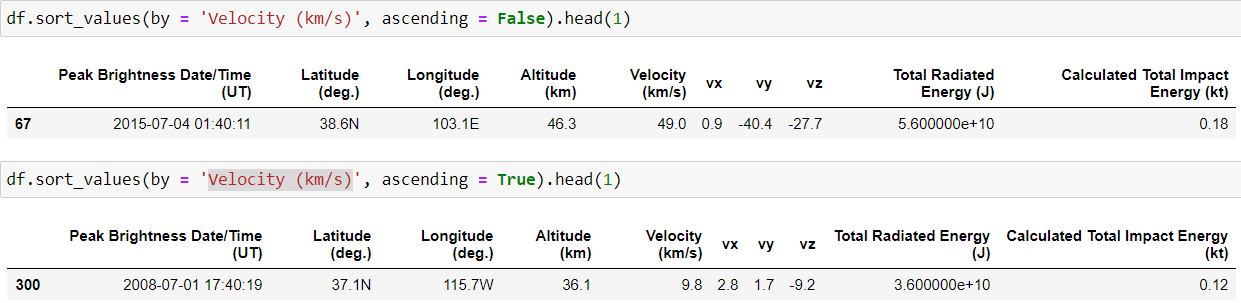
Создаем переменную df в которую поместим значения данных с которыми будем работать. С помощью метода head() выводим заданное количество строк на экран.

1. Определите местоположение появления метеорита с наибольшим количеством высвободившейся энергии.



В переменную ех помещаем отсортированную методом sort\_values, по столбцу Total Radiated Energy (J) таблицу, выводим 1 строку у столбцов координат, т. к. сортировка по убыванию.

2. Определите метеорит с максимальной и минимальной скоростью.



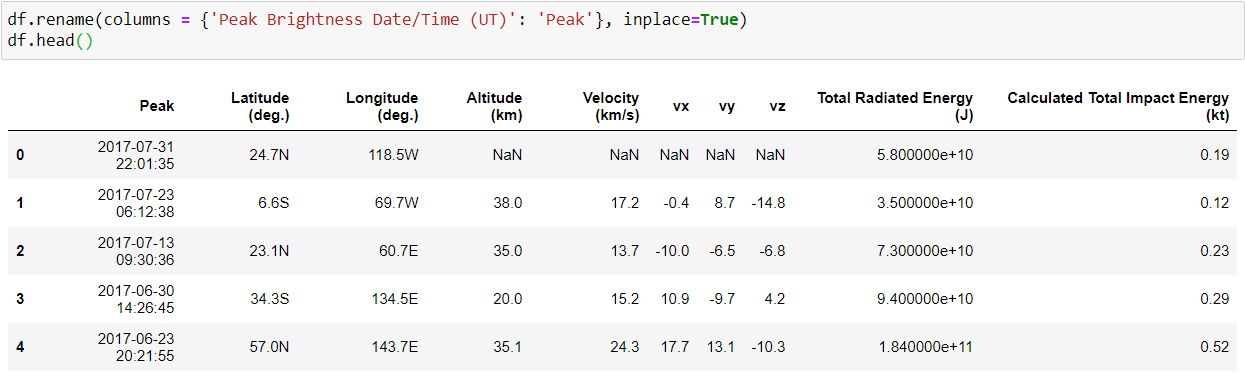
Сортируем таблицу по столбцу Velocity (km/s) с начало по убыванию потом по возрастанию и выводим 1 строку.

3. Определите место максимальной концентрации метеоритов.

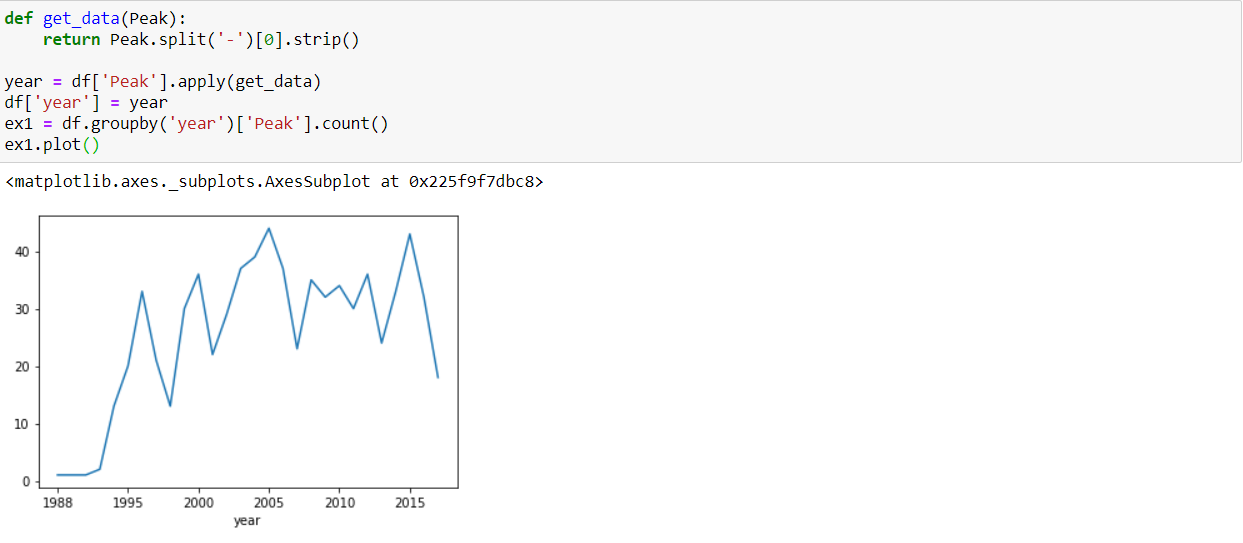


Для начало удаляем все пустые значения из столбцов координат. С помощью метода shape делим сетку координат на 8 частей и подсчитываем количество точек в каждой части. Максимальная концентрация метеоритов в зоне координат x>0, y<0, z<0.

4. Определите время года, в которое вероятность появления метеоритов максимальна. Постройте график.



Для удобства переименуем столбец Peak Brightness Date/Time (UT) в Peak.



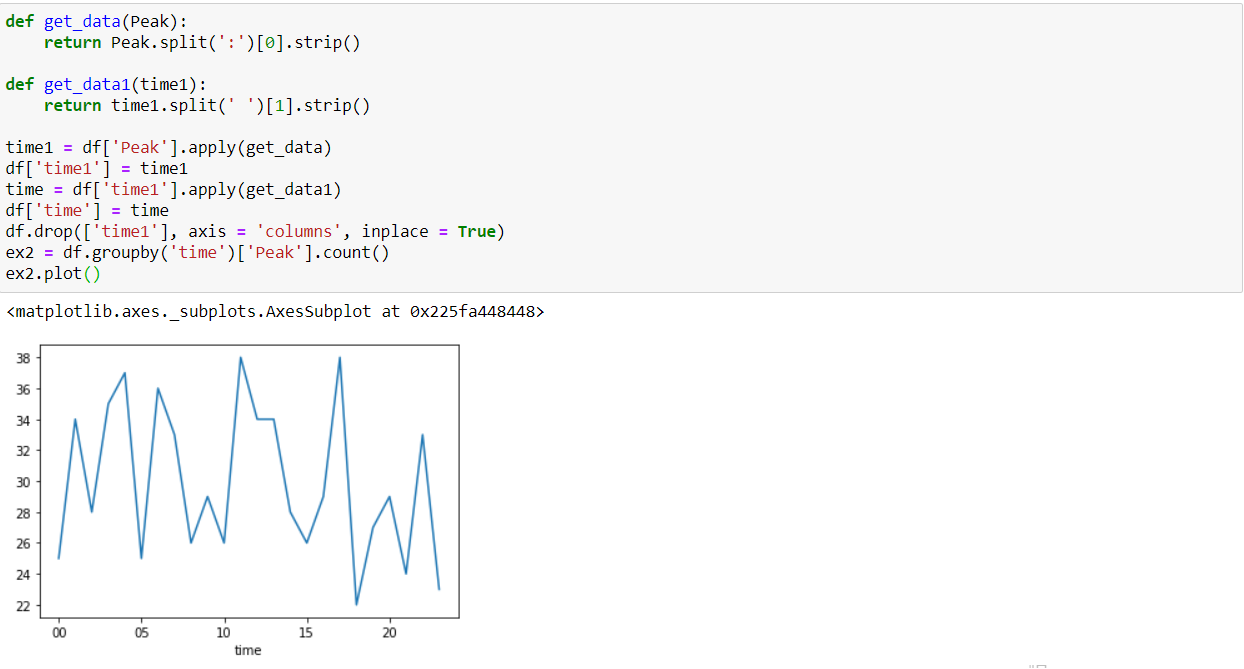
Создаем функцию get\_data которая из столбца Peak выделит год и поместим полученные данные в столбец year. Методом groupby сгруппируем в переменной ex1 данные столбца Peak по столбцу year. С помощью метода plot построим график.

5. Удалите строки данных, в которых для метеоритов не указана скорость. Выведите размер набора данных до и после удаления.



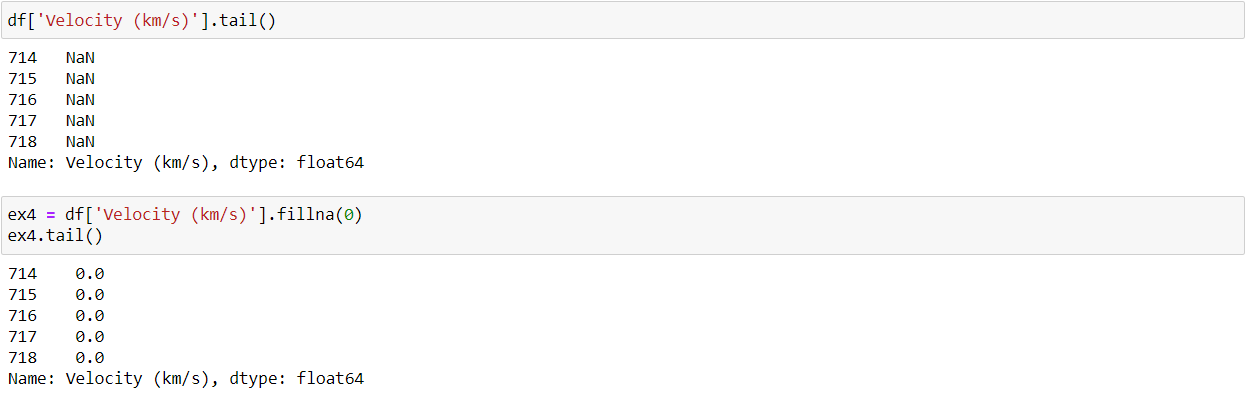
С помощью метода shape выведем размер таблицы df. Методом dropna удалим из столбца Velocity (km/s) все пустые значения. Выводим размер ex на экран.

6. Построить график, на котором по оси ОХ отложено время суток, по OY – частота появления метеоритов.



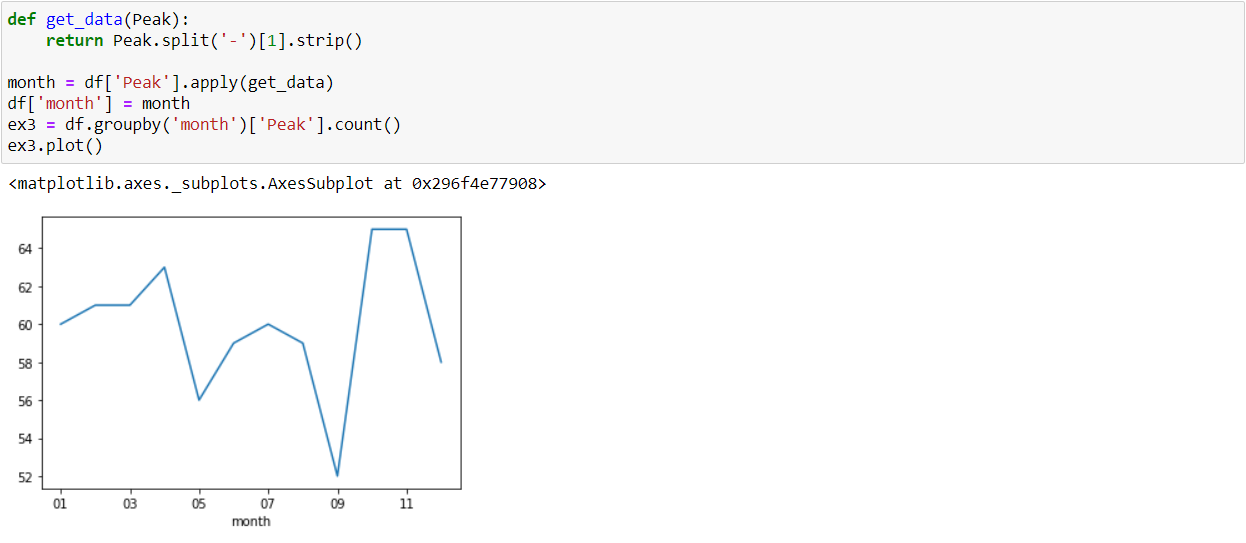
С помощью функций get\_data и get\_data1 выделяем время из столбца Peak и добавляем новый столбец time. Группируем данные столбца Peak по time. Выводим график.

7. Заполните строки в которых для метеоритов не указана скорость.



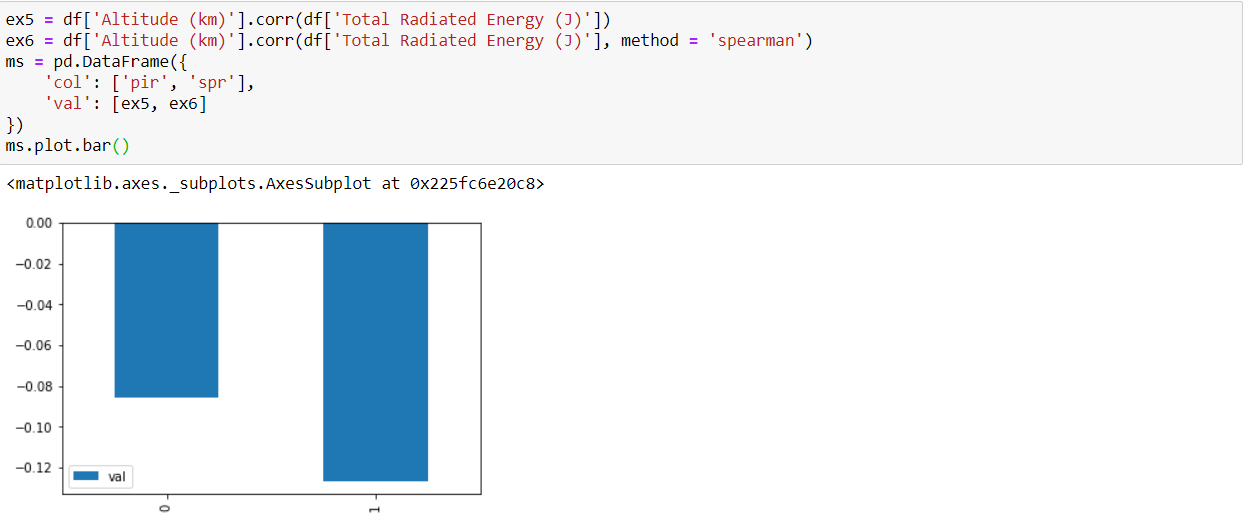
С помощью метода fillna заполняем нулями пустые строки в столбце Velocity (km/s).

8. Определите месяц года, для которого появление метеоритов наиболее вероятно (постройте необходимые графики).



C помощью функции get\_data выделяем из столбца Peak месяц. Создаем новый столбец month в котором будут хранится выделенные данные. Группируем данные Peak по месяцу и выводим график.

9. Вычислите корреляцию Пирсона и Спирмена между признаками Altitude и Total Radiated Energy. Сравните полученные величины корреляции. Постройте гистограммы для указанных признаков. Оцените значение p-value.



С помощью метода corr вычисляем значение корреляции Пирсона, для Спирмена метод нужно указывать отдельно method = spearman. Сохраняем полученный результат в таблицу. Строим гистограмму с помощью plot.bar. Результаты говорят о слабой обратной связи между признаками.