

|  |
| --- |
| МИНОБРНАУКИ РОССИИ  Федеральное государственное бюджетное образовательное учреждение высшего образования  **«МИРЭА – Российский технологический университет»**  **РТУ МИРЭА** |
| Институт информационных технологий (ИТ) |
| Кафедра прикладной математики (ПМ) |

**ОТЧЁТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ 4**

**по дисциплине «Технологии и инструментарий анализа больших данных»**

Выполнил студент группы ИКБО-02-20 Злобин А.О.

Принял

*Ассистент*

Горячев А.А.

Практические работы выполнены « » 2023 г.

(подпись студента)

«Зачтено» « » 2023 г.

(подпись руководителя)

Москва 2023

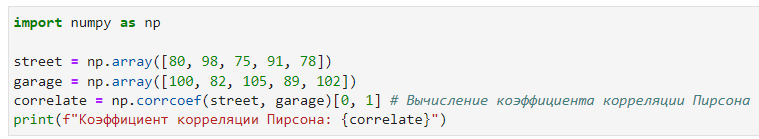
# Определить два вектора, представляющие собой число автомобилей, припаркованных в течении 5

**рабочих дней у бизнес-центра на уличной стоянке и в подземном гараже.**

* 1. **Найти и интерпретировать корреляцию между переменными**

**«Улица» и «Гараж» (подсчитать корреляцию по Пирсону).**

**Код**

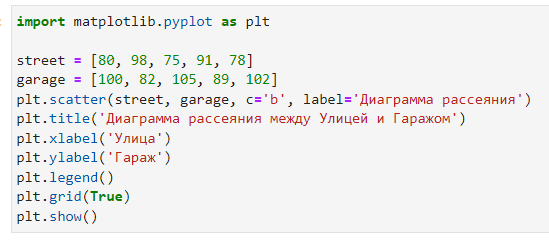
****

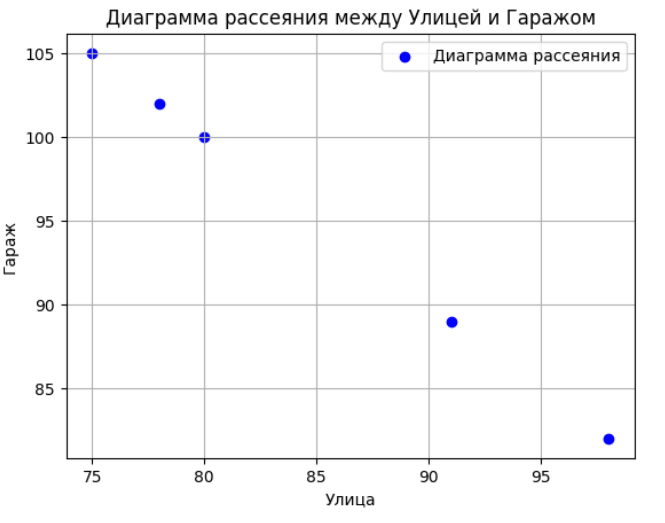
**Результат**



* 1. **Построить диаграмму рассеяния для вышеупомянутых переменных.**

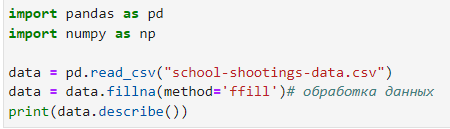
**Код**

****

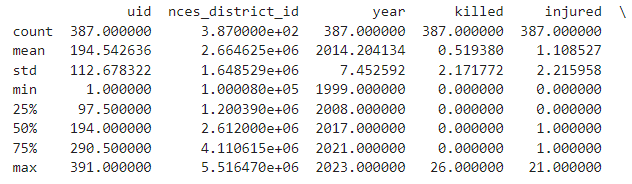
**Результат**

# Найти и выгрузить данные. Вывести, провести предобработку и описать признаки.

**Код**

****

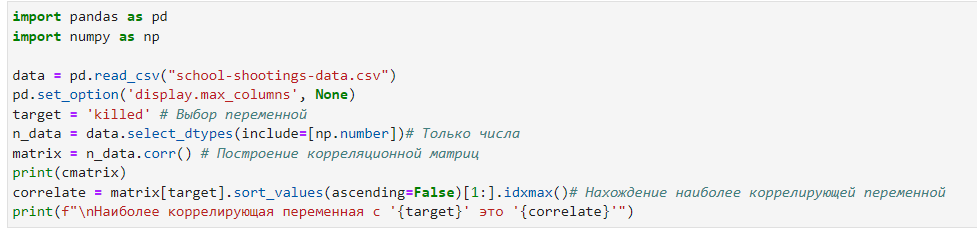
**Результат**



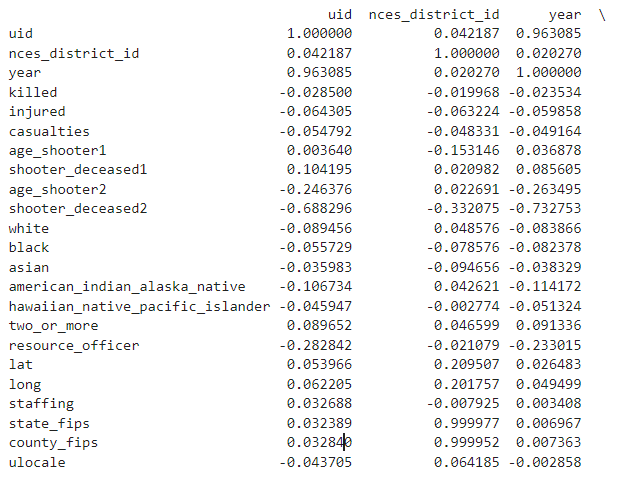
* 1. **Построить корреляционную матрицу по одной целевой**

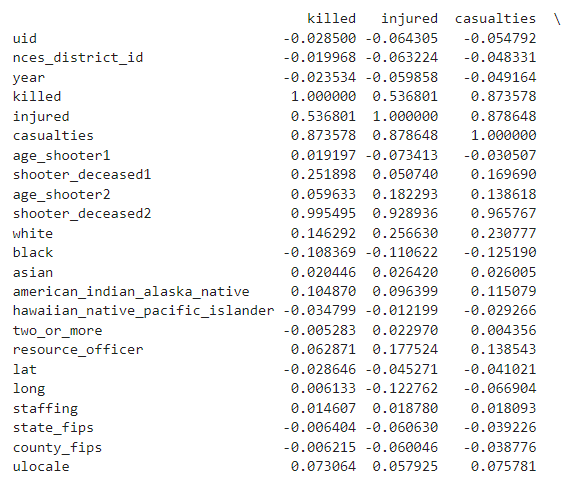
**переменной. Определить наиболее коррелирующую переменную, продолжить с ней работу в следующем пункте.**

**Код**

****

**Результат**



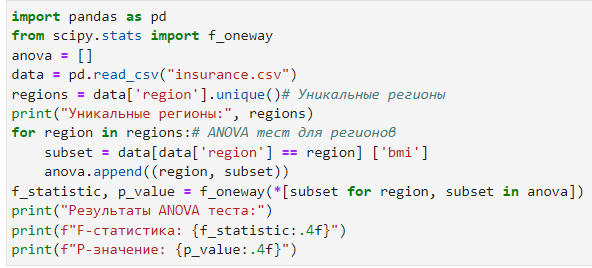


Наиболее коррелирующая переменная с killed это shooter\_deceased2

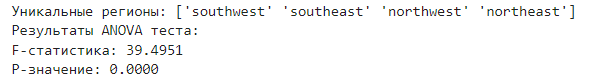
# Загрузить данные: 'insurance.csv'. Вывести и провести предобработку. Вывести список уникальных регионов.

* 1. **Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя первый способ, через библиотеку Scipy.**

**Код**



**Результат**

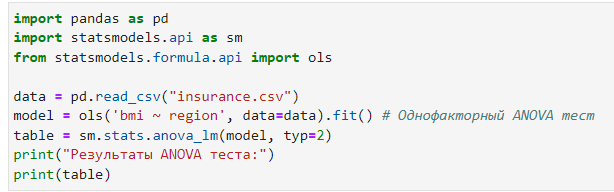
****

Так как p-значение меньше альфа значения, то гипотеза BMI отвергается.

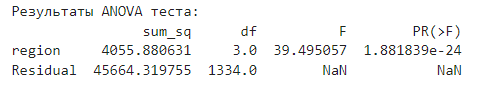
* 1. **Выполнить однофакторный ANOVA тест, чтобы проверить влияние региона на индекс массы тела (BMI), используя второй**

**способ, с помощью функции anova\_lm() из библиотеки statsmodels.**

**Код**

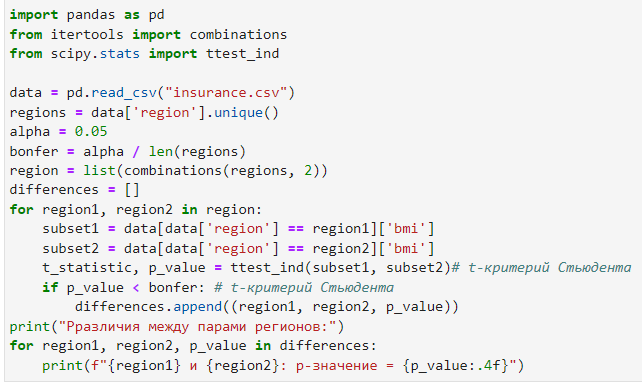
****

**Результат**

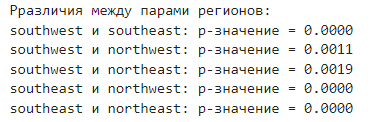
****

* 1. **С помощью t критерия Стьюдента перебрать все пары. Определить поправку Бонферрони. Сделать выводы.**

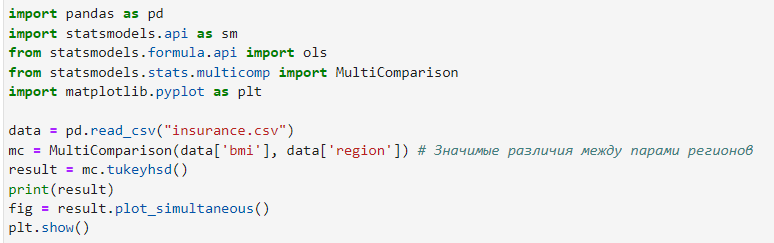
**Код**

****

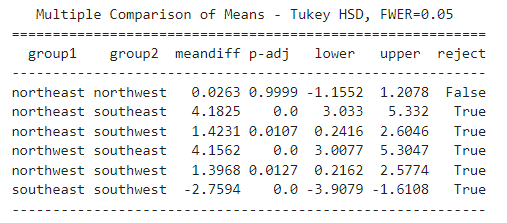
**Результат**

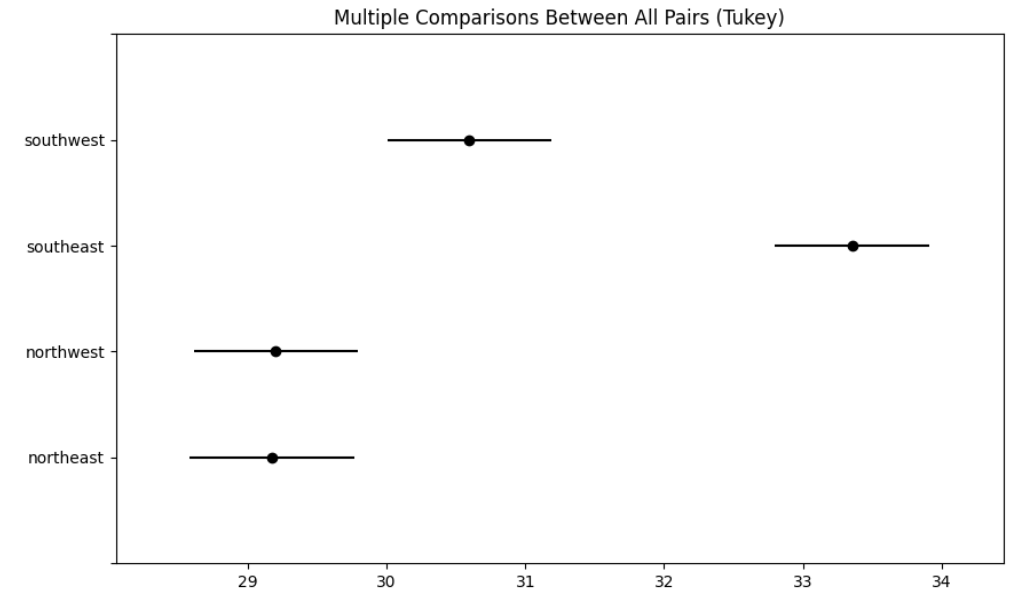
****

* 1. **Выполнить пост-хок тесты Тьюки и построить график. Код**



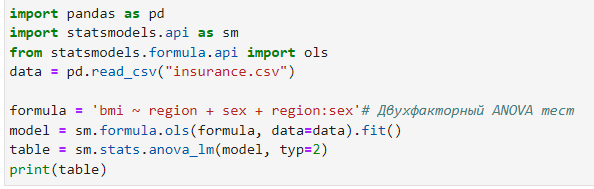
**Результат**

****

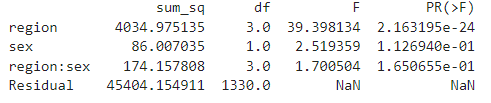
****

* 1. **Выполнить двухфакторный ANOVA тест, чтобы проверить влияние региона и пола на индекс массы тела (BMI), используя функцию anova\_lm() из библиотеки statsmodels.**

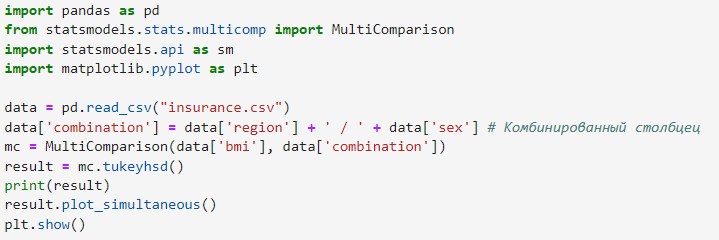
**Код**



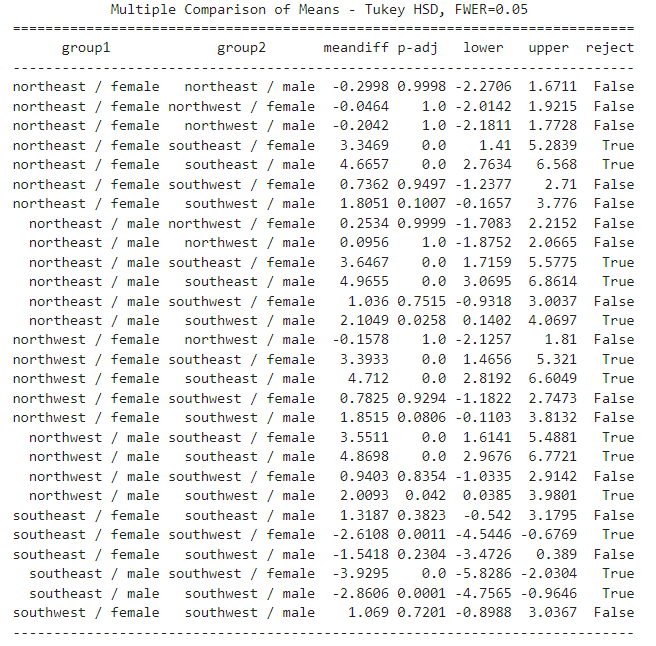
**Результат**

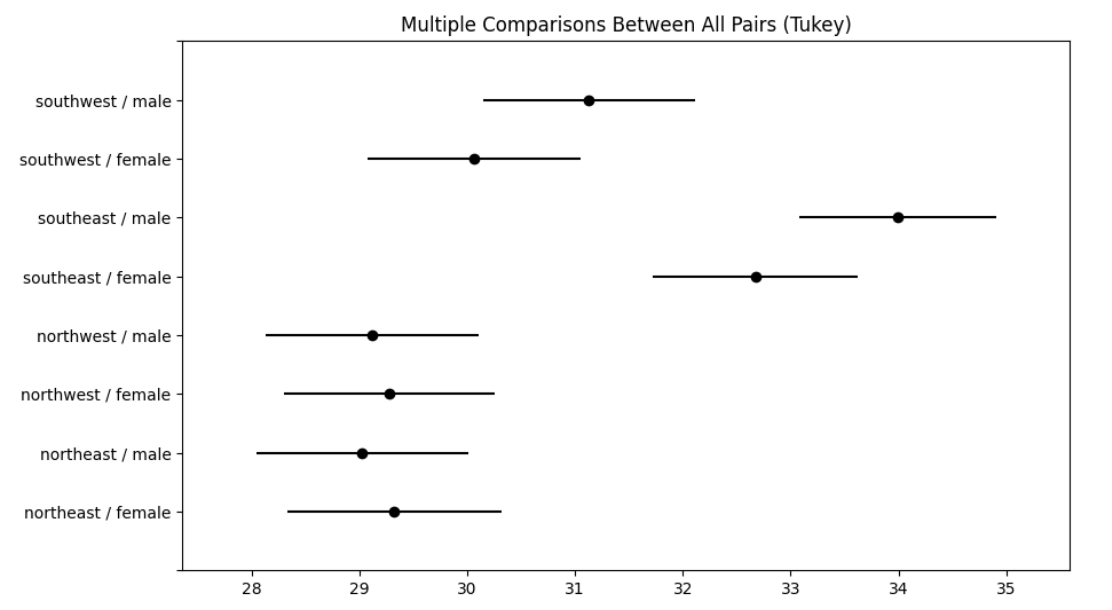
****

* 1. **Выполнить пост-хок тесты Тьюки и построить график. Код**



**Результат**

****

****