Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»



**Отчет**

**Лабораторная работа № 1**

**По курсу «Технологии машинного обучения»**

## «Разведочный анализ данных. Исследование и визуализация данных»

**ИСПОЛНИТЕЛЬ:**

Матиенко Андрей

Группа ИУ5-61Б

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

**ПРЕПОДАВАТЕЛЬ:**

Гапанюк Ю.Е.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

Москва 2020

**Цель лабораторной работы:**

Изучение различных методов визуализация данных.

**Краткое описание.**

Построение основных графиков, входящих в этап разведочного анализа данных.

1. **Текстовое описание набора данных**

**Dataset (www.kaggle.com):**

2019 Coronavirus dataset (January - February 2020)

Tracking the spread of 2019-nCoV

**Context:**

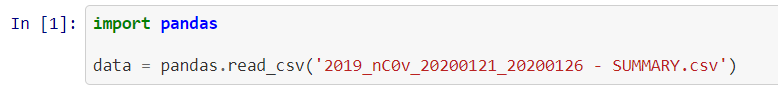
The 2019-nCoV is a contagious coronavirus that hailed from Wuhan, China. This new strain of virus has striked fear in many countries as cities are quarantined and hospitals are overcrowded. This dataset will help us understand how 2019-nCoV is spread aroud the world.

Датасет содержит следующие колонки:

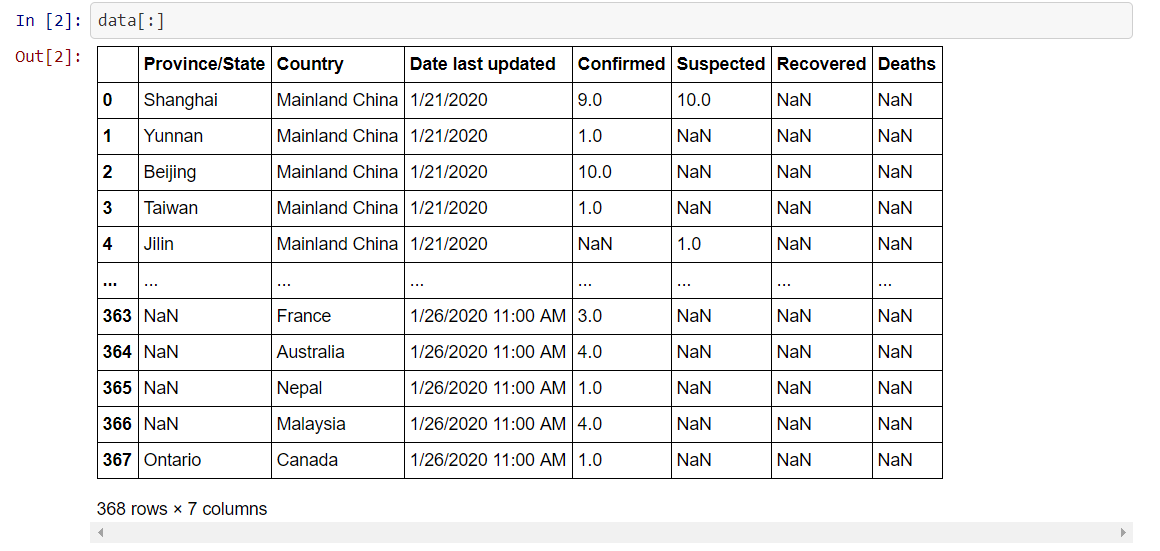
* Province/State – провинция, город
* Country – страна
* Date last updated – последнее обновление данных
* Confirmed – подтвержден вирус, сколько заболевших
* Suspected – ожидают лечение
* Recovered – выздоровело людей
* Deaths – смертей от вируса

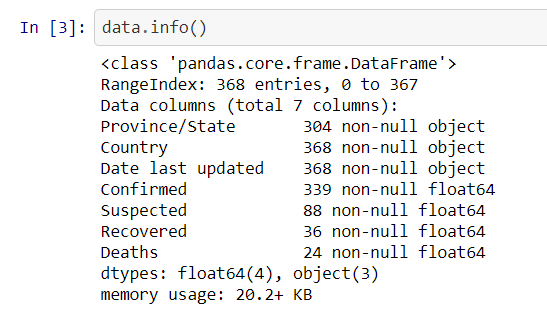
1. **Загрузка данных**

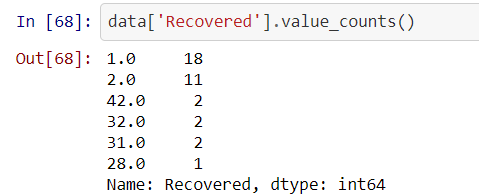
Загрузим файлы датасета в помощью библиотеки Pandas.

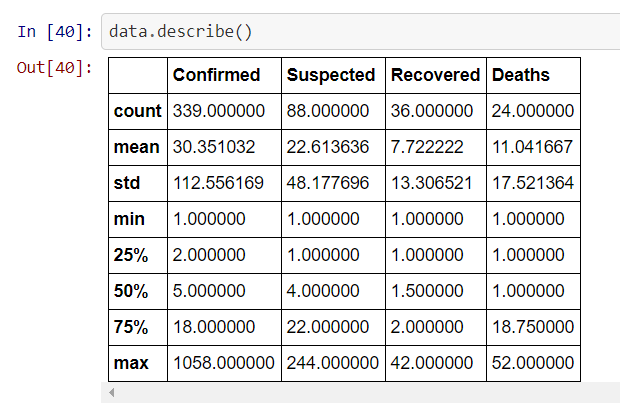


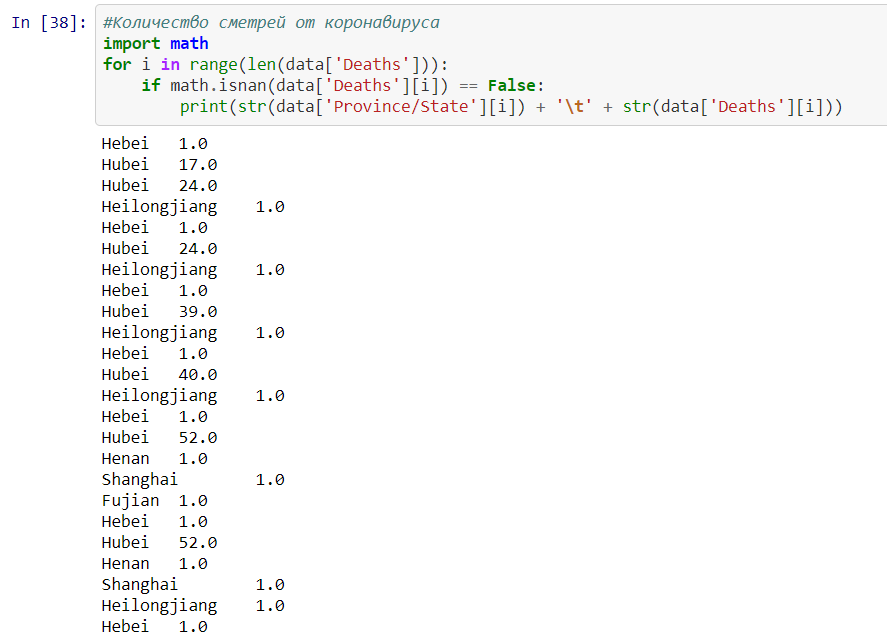
1. **Основные характеристики датасета**

****

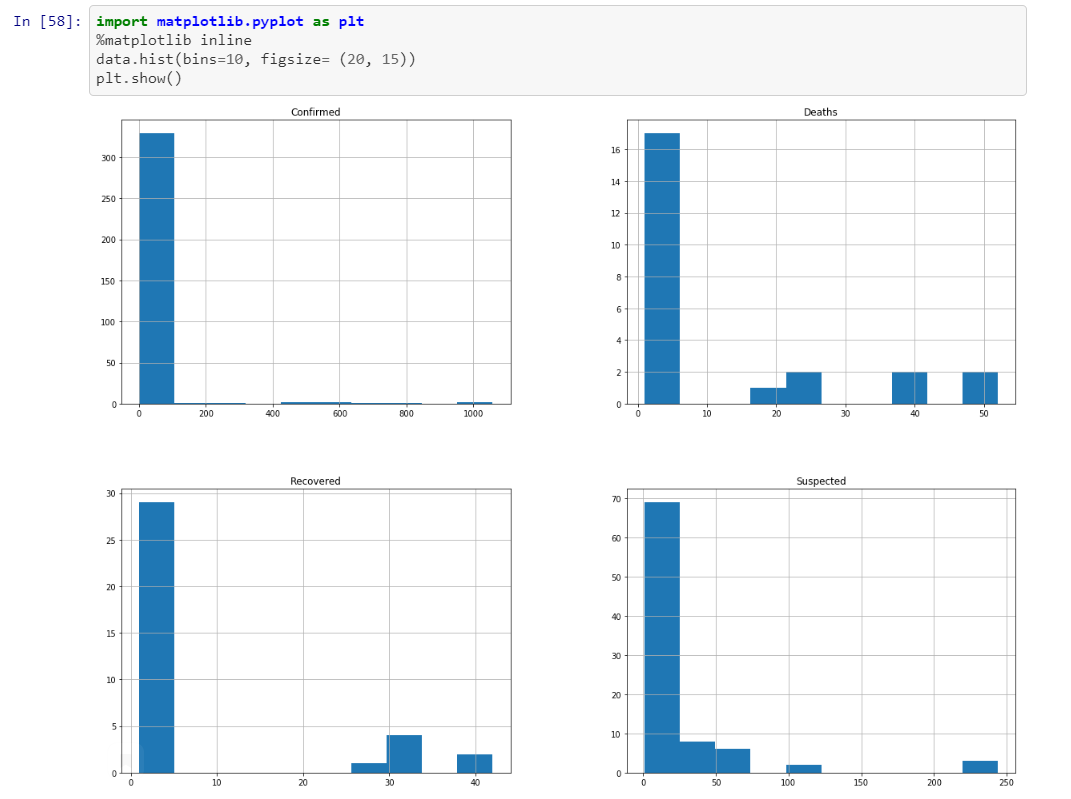
****

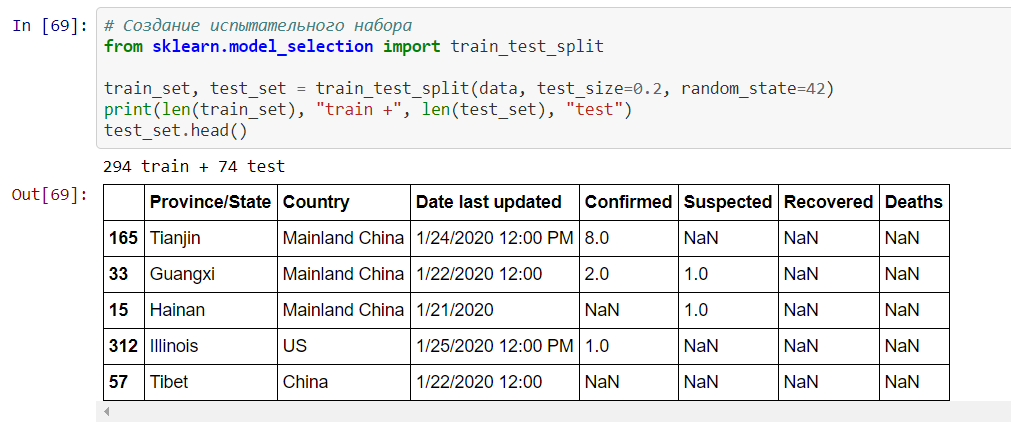
****

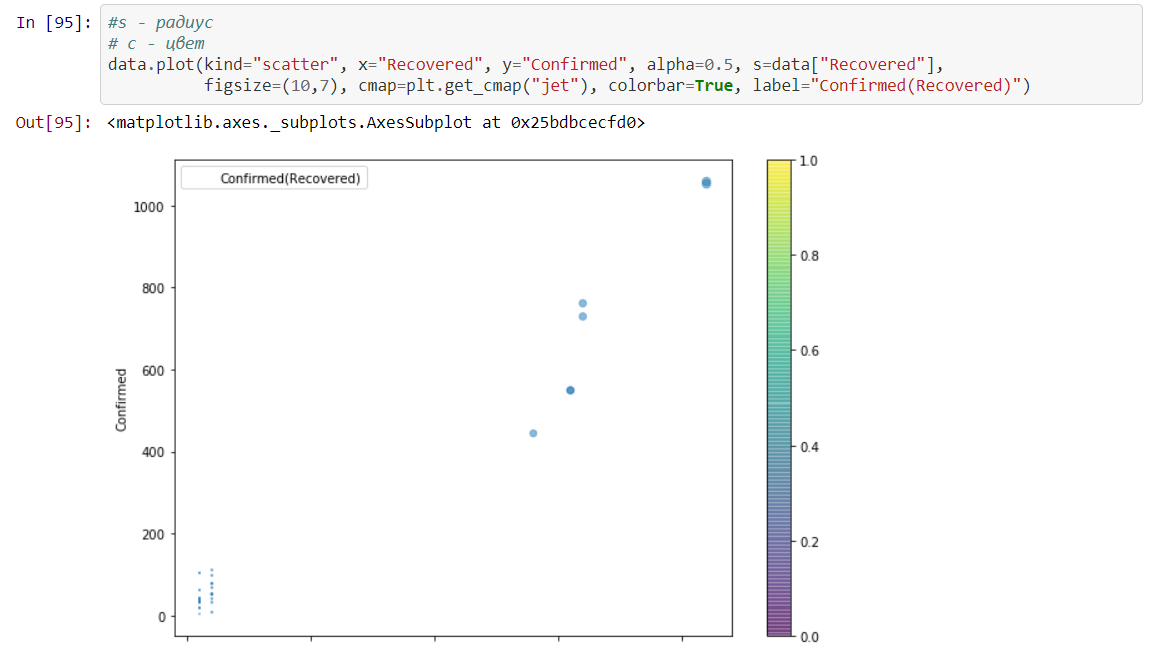
****

****

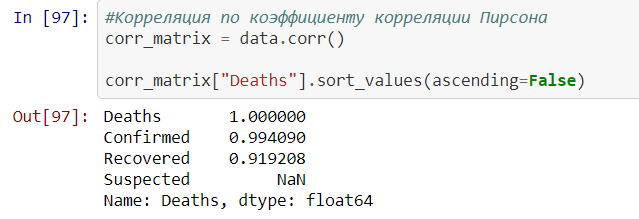
1. **Визуальное исследование датасета**

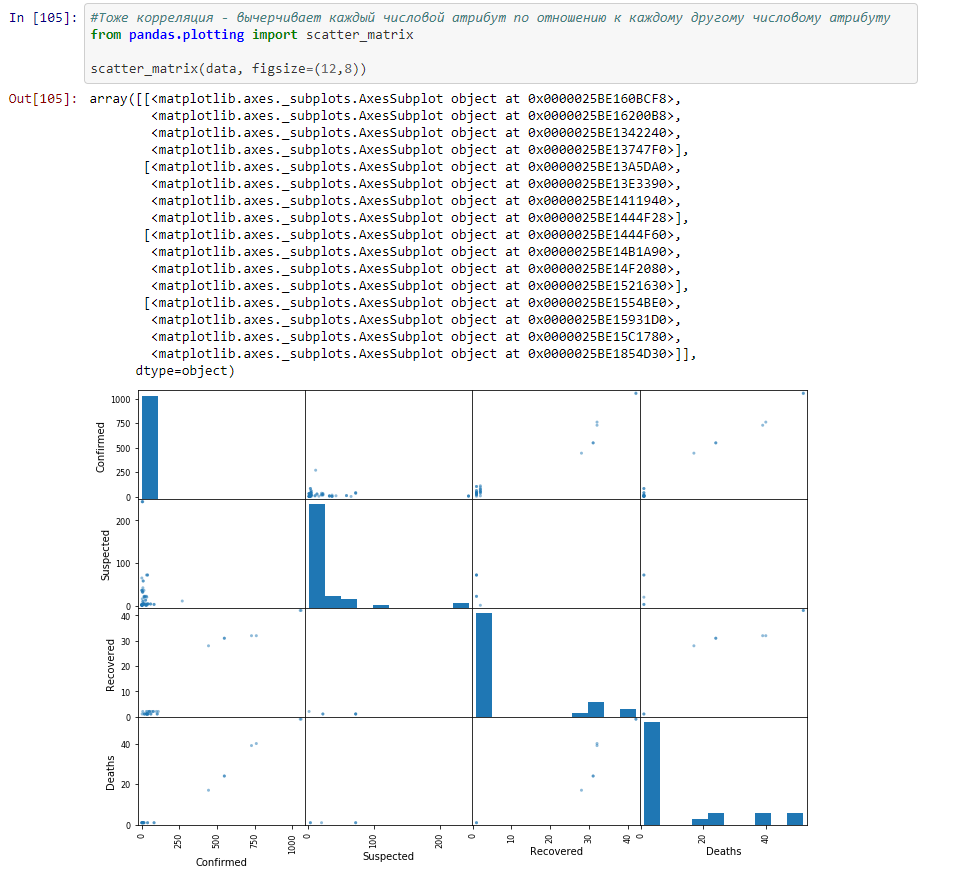
****

****

****

1. **Корреляция**

****

****

**Вывод:**

Научился работе в Jupyter notebook. Познакомился с ML.