****Московский государственный технический университет им. Н.Э. Баумана  
Кафедра «Системы обработки информации и управления»

Лабораторная работа №1

по дисциплине «Методы машинного обучения»

на тему«Создание "истории о данных"»

Выполнил:  
студент группы： ИУ5-23М  
Аимань Мухэяти

Москва — 2020 г.

1. **Цель лабораторной работы**изучение различных методов визуализация данных и создание истории на основе данных.
2. **Задание**

* Выбрать набор данных (датасет).
* Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:

1. История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.
2. На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.
3. Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.
4. Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.
5. История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.

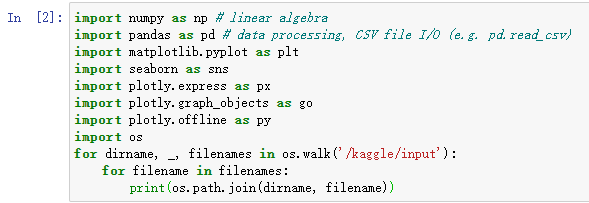
* Сформировать отчет и разместить его в своем репозитории на github.

1. **Ход выполнения работы**Побочные реакции вакцины (VAERS)

VAERS, система отчетности о нежелательных явлениях вакцины, введенная в действие законом 1986 года. В 2016 году VAERS получила 59 117 сообщений следующим образом: 432 смерти, 1091 постоянная инвалидность, 4132 госпитализации, 10 284 посещения отделений неотложной помощи, 59 117 общих нежелательных явлений, зарегистрированных в 2016 году.

Поскольку я также была вакцинирована против COVID-19, были и побочные реакции, но я знала, что это нормально.Поэтому я хотела бы сделать доклад о побочных реакциях на вакцину COVID-19.

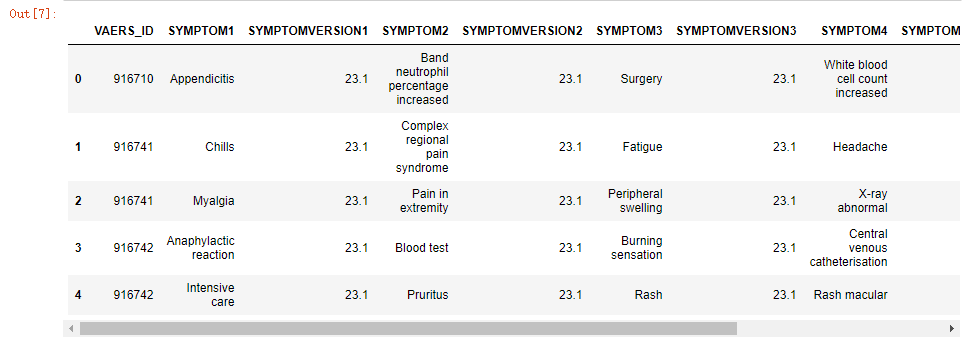
Подключим все необходимые библиотеки



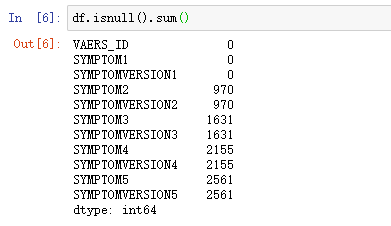
Загрузим непосредственно данные(2021vaerssymptoms.csv):



Посмотрим на данные в данном наборс данных:



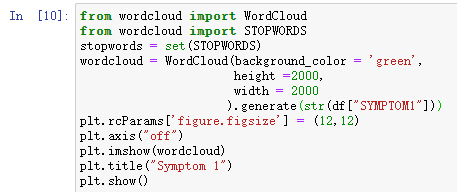
Проверяем количество пропущенных данных в каждом столбце



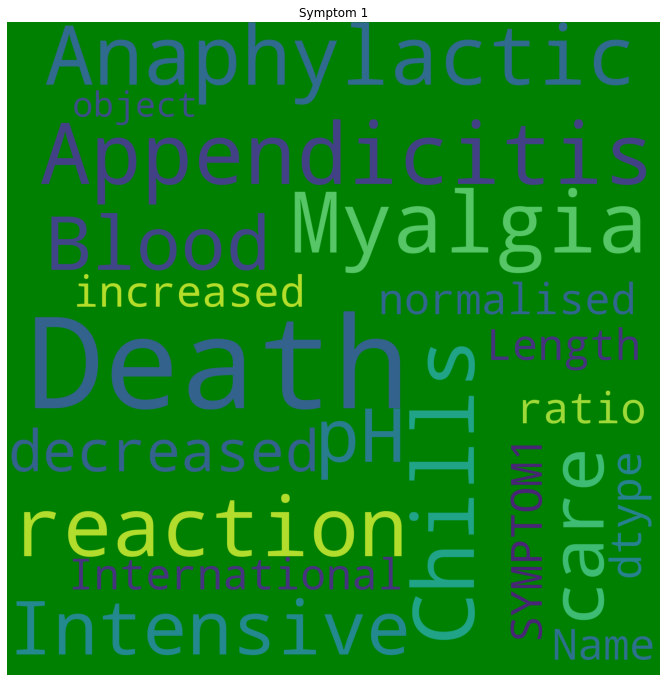
Заполняем недостающие данные с помощью режима



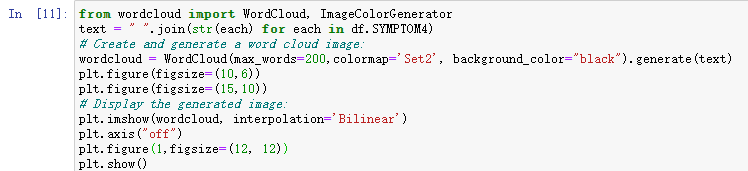
С помощью библиотеки "облако слов" мы составляем диаграмму симптома 1. Интуитивно покажите нам, как часто появляется каждый симптом.

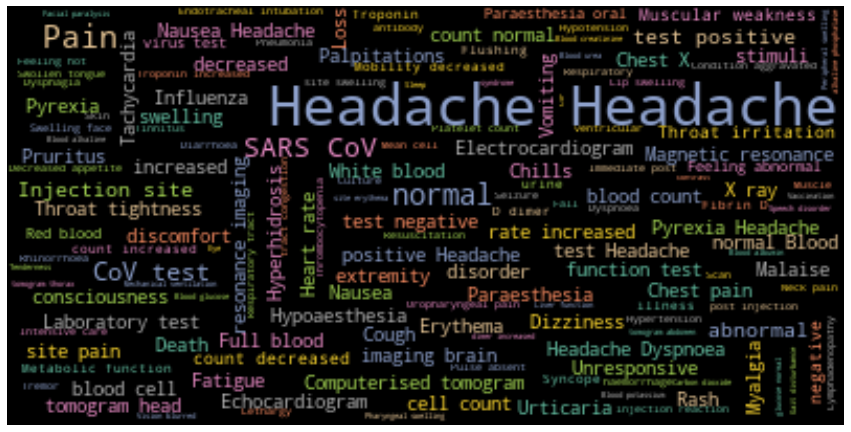


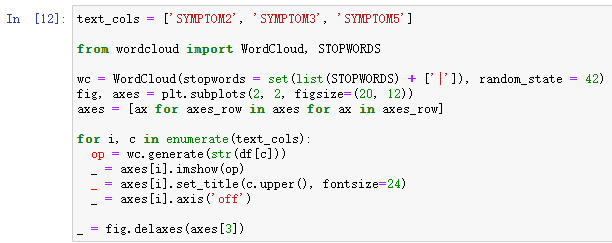
Мы видим самую высокую долю смертей, за которыми следуют миалгия и аппендицит



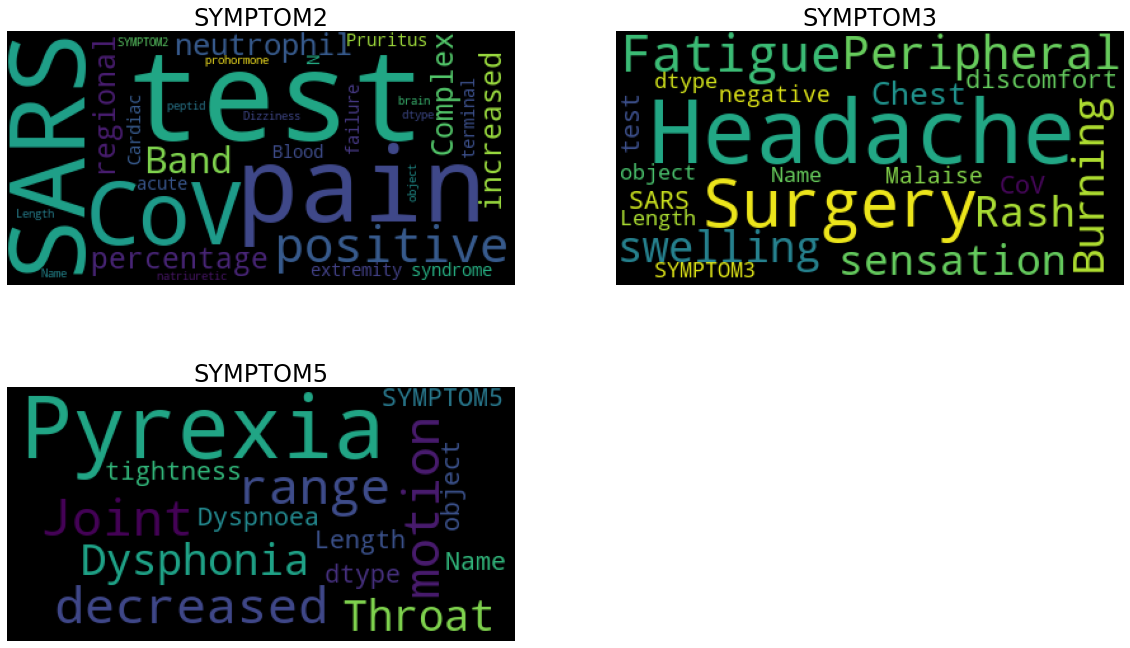
Затем мы провели аналогичную операцию над симптомами 4,2,3,5 соответственно, для создания и генерации образов облака слов



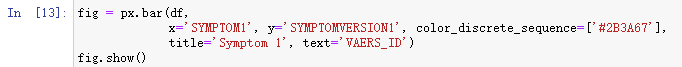


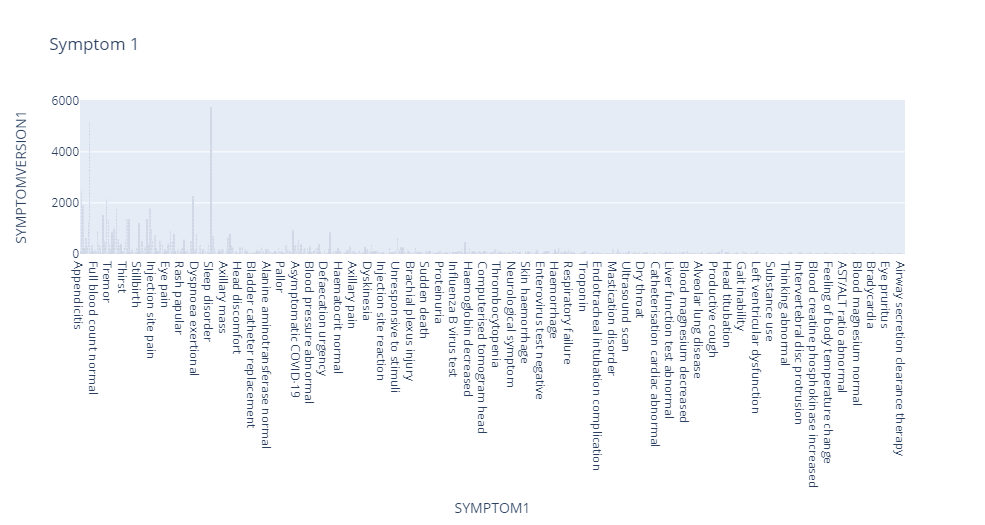


Видно,что головные боли,лихорадка также являются наиболее распространенными побочными реакциями



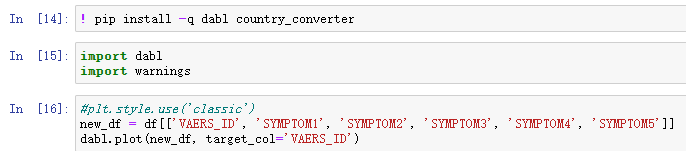
создаем и генерируем гистограмму симптома 1

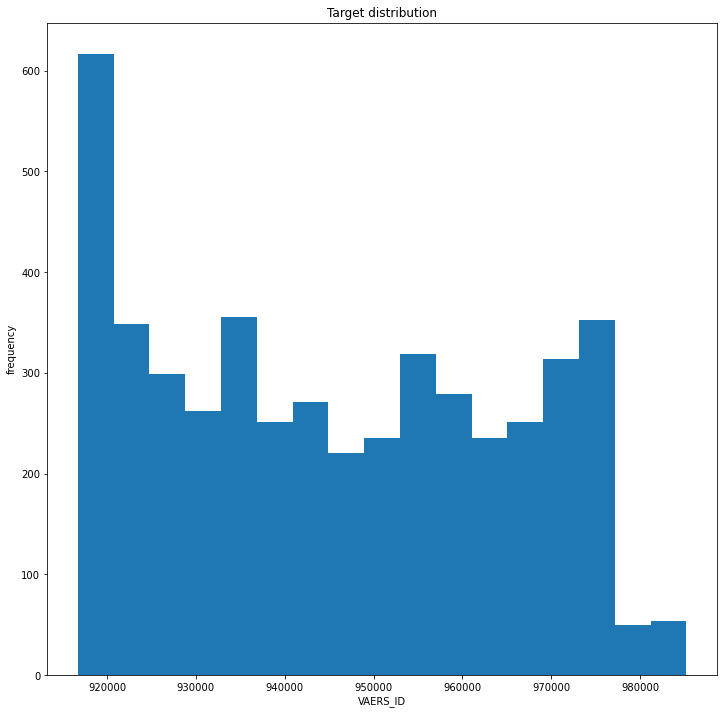




Для обработки данных используйте базовую библиотеку анализа данных.(data analysis baseline library) dabl

Исследовательский данные анализ дает мне быстрое понимание данных

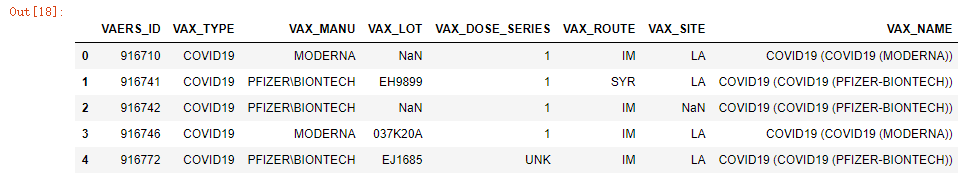




Загрузим непосредственно данные(2021vaersvax.csv):

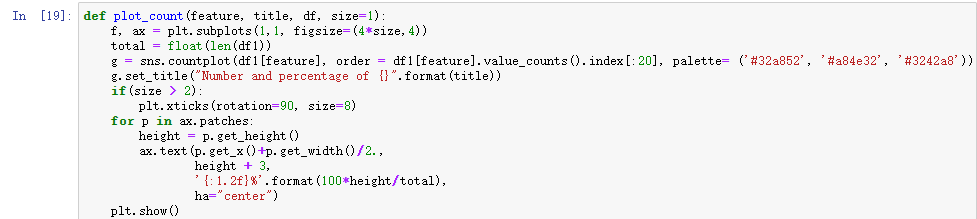


Посмотрим на данные в данном наборс данных:

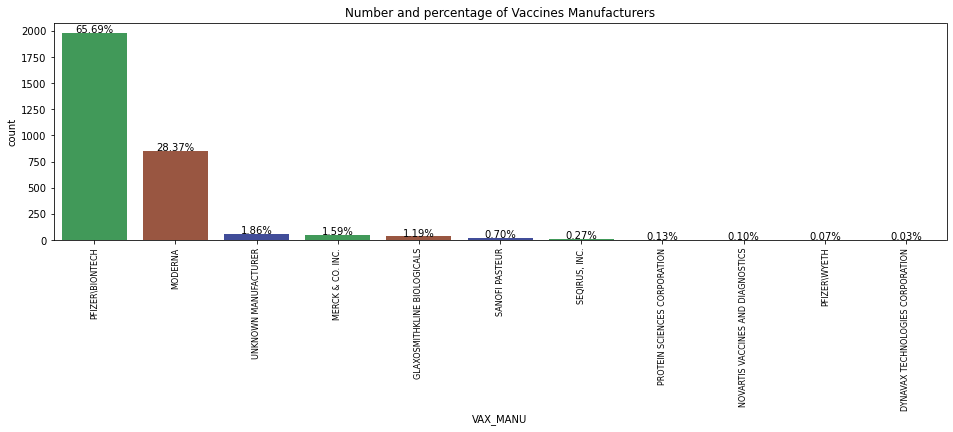


Определение функции

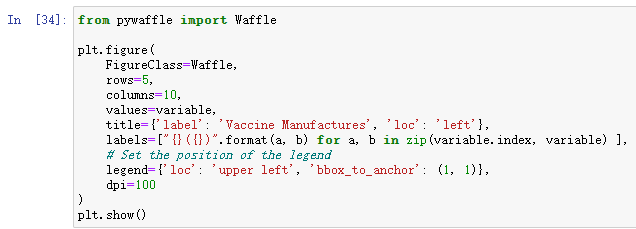
Количество отдельных производителей вакцин и их процентное соотношение,и в-изуализируем эту информацию

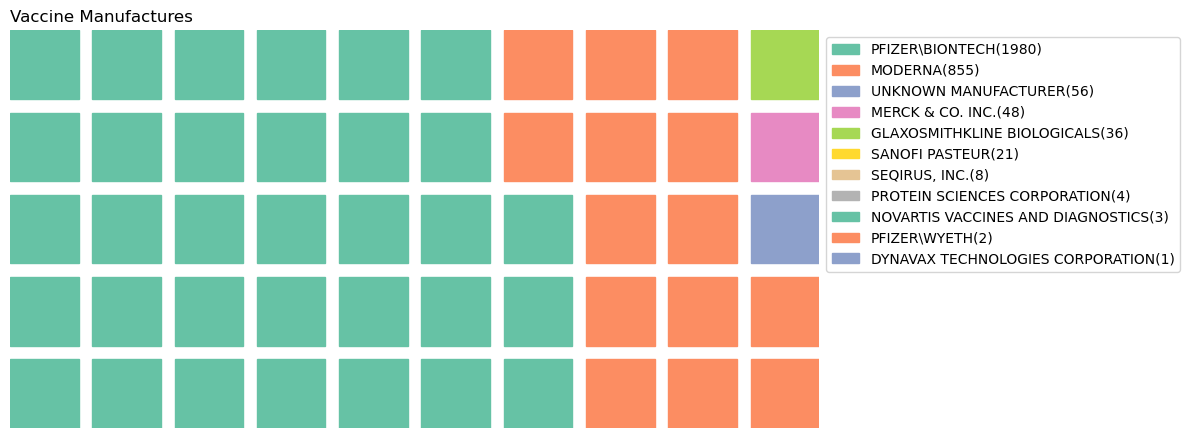


Как видно,Производитель вакцин PFIZER/ BIONTECH имеет самые неблагоприятные реакции на вакцину.За ним следует производитель вакцин MODERNA.Остальные производители вакцин составляли небольшую долю

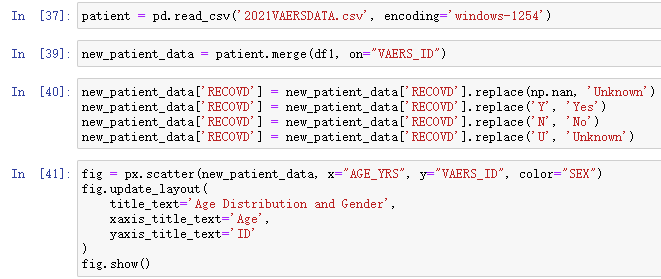


Другой способ показать, теперь мы создаем и генерируем "Вафельные" изображения



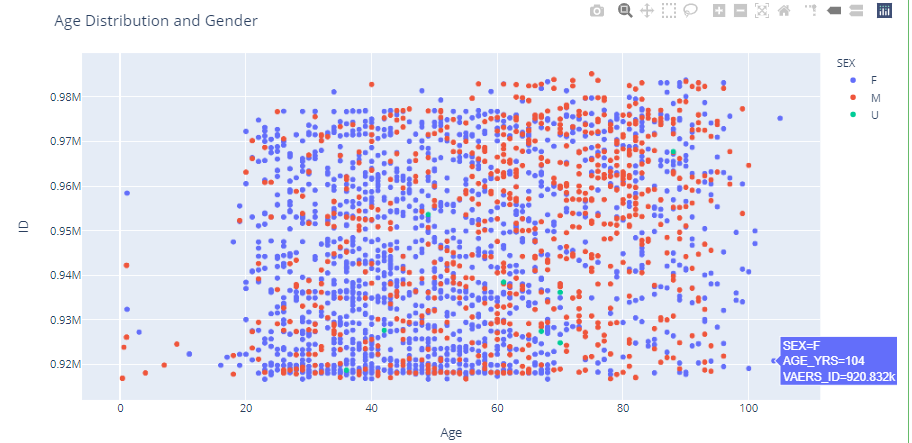


Загрузим непосредственно данные(2021vaersdata.csv):

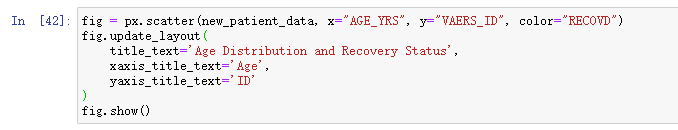


В этой таблице мы видим распределение по возрасту и полу.Синий для женщин и красный для мужчин.Зеленый для неизвестного

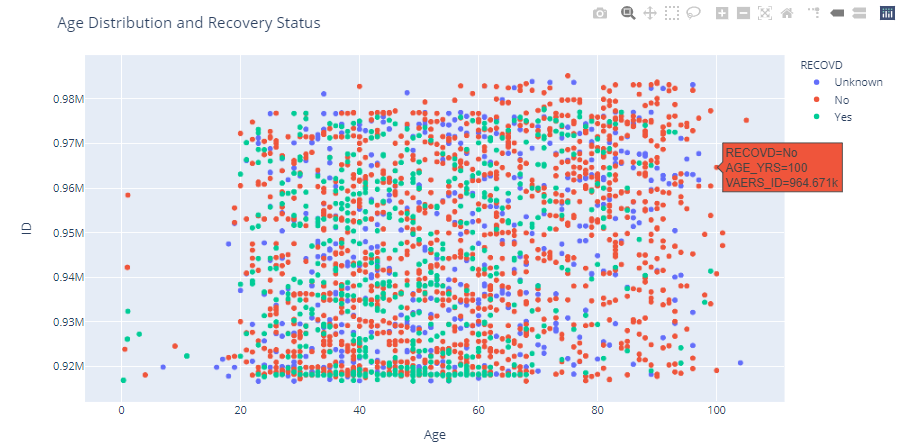
Есть больше женщин с побочными реакциями и больше мужчин старше 60 лет.

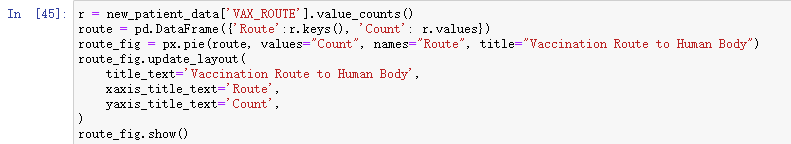


Этот график представляет собой распределение распределения по возрасту и состоянию восстановления.



Красный означает отсутствие восстановления,зеленый означает восстановление.Синий цвет символизирует неизвестность.Видно, что общее частичное выздоровление не больше, а население старше 60 лет выздоравливает намного меньше.

Хотим знать распределение иммунных путей в организме человека для каждой вакцины.

Видно, что IM иммунный путь Замби больше всего, за ним следует ОТ.

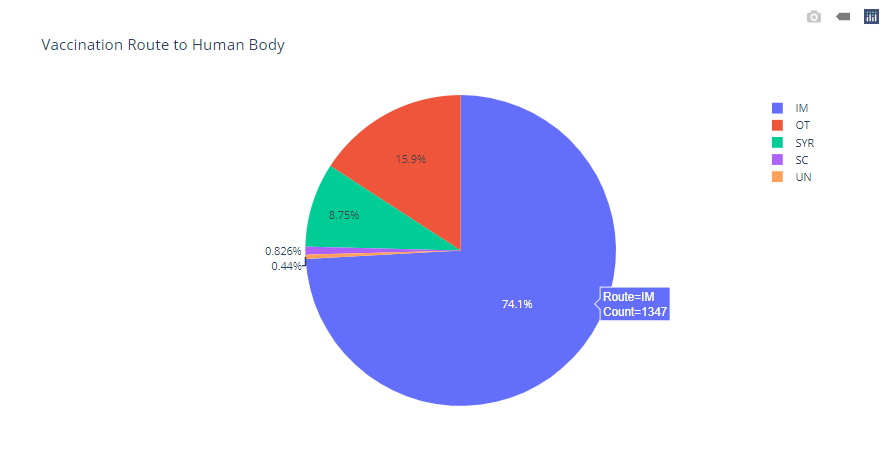
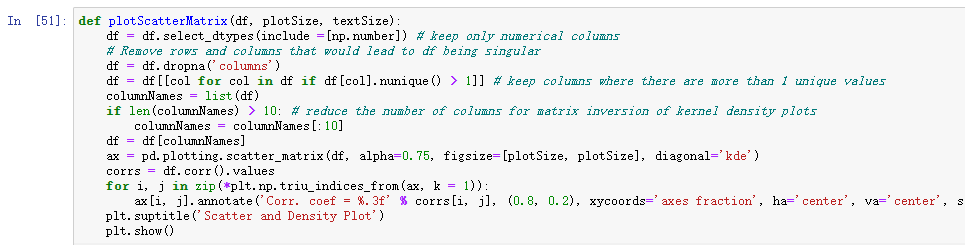
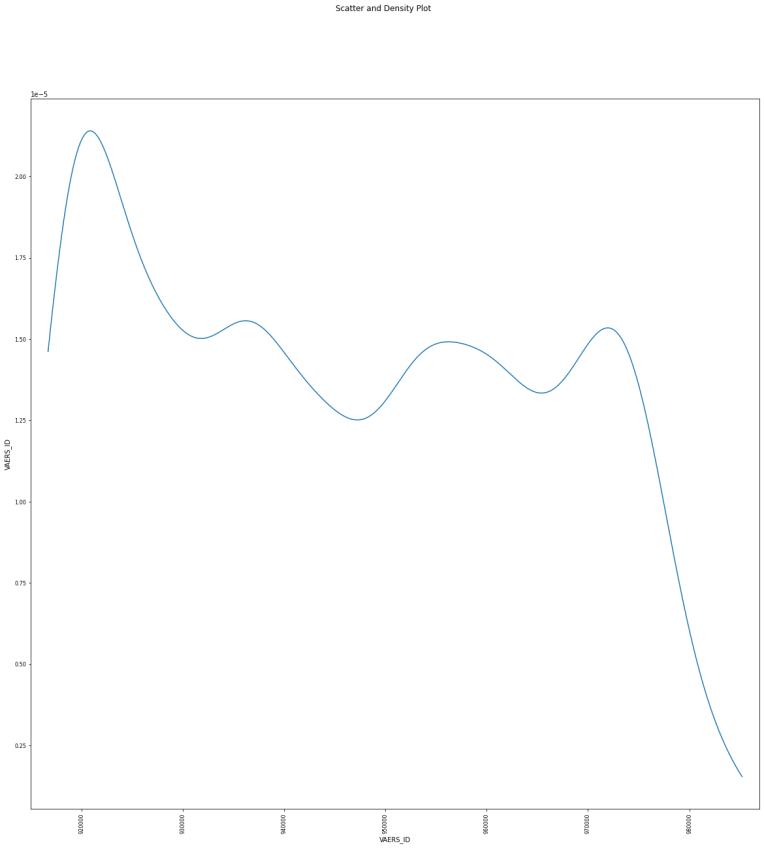
Эта круго-вая диаграмма показывает долю и общее количество иммунных путей

Диаграмма рассеяния и карты плотности для каждой вакцины





Вакцины могут вызывать побочные реакции, но побочные реакции делятся на различные степени и различные уровни, как правило, легкие, умеренные и тяжелые.В целом вакцины COVID-19 сейчас на рынке очень много.Среди производителей вакцин доминирует компания Rycon.Побочные реакции различаются по возрасту и полу.Женщины и пожилые люди более склонны к побочным реакциям, а пожилые вакцинаторы не так легко восстанавливаются после побочных реакций.

**Список литературы**

## Гапанюк Ю. Е. Лабораторная работа «истории о данных» [Электронный ресурс] // GitHub. –– 2021. –– Режим доступа: <https://github.com/ugapanyuk/ml_course_2021/wiki/COURSE_MMO>