Всероссийский конкурс научно-технологических проектов "Большие вызовы"

**Обработка и предсказание и Визуализация данных по covid-19 в различных странах мира.**

Ульяновск 2022

**Содержание:**

1. **Введение**
2. **Сбор и анализ данных**
3. **Преобразования данных(описываете преобразования в pandas)**
4. **Визуализация данных. (со скриншотами и все таким) Отдельно рассказываем про dash но где неизвестно => в 4**
5. **Нормирование и все его варианты**
6. **Корреляции и всё такое**
7. **Обучение линейной регрессии по самым похожим. Предположение и построение прогнозов**
8. **Перспективы.**

**Введение:**

В настоящее время весь мир терроризирует коронавирус, из-за которого на данный момент уже заболело 328 миллионов человек, из которых 5.54 миллиона - это летальные исходы. Какое-то время назад многие страны мира боролись с ним на пределе возможностей: постоянно забитые больницы, измученные врачи, работающие по две смены и, конечно же, постоянная разработка вакцины против covid-19.

Кроме всего этого, на данный момент ситуация примерно такова: активно пропагандируется вакцинация, 1.2 миллиона заболевших за день в США, QR-коды и обязательное ношение маски.

Исходя из всего этого, **гипотеза** такова: «Количество вакцинировавшихся напрямую коррелирует с количеством заболевших, а страны с населением, либо правительством с более мягким отношением к covid-19 имеют более высокие показатели процентальной заболеваемости, относительно населения».

**Актуальность:**

Как было уже сказано выше, все наши данные и изыскания взяты напрямую с недавних дат. Прямо сейчас всем нам было бы важно узнать, насколько важна вакцинация, а также же, её влияние непосредственно на заболеваемость.

**Проблематика:**

Здоровье - это самый важный человеческий ресурс, который очень дорого и трудозатратно восстанавливать. Никто бы не хотел потратить его впустую, для этого все мы должны понимать уровень опасности болезни, которая нас постигла, а также все мы должны уметь ей противостоять.

**Цели и задачи:**

Целью нашего исследования является создание графиков смертности и заболеваемости в разных странах, а также поиск и сравнение стран с похожими ситуациями. На основе этих данных, прогноз количества заболевших на несколько дней вперёд и выяснение, действительно ли прививка влияет на количество заболевших.

**Сбор и анализ данных:**

Начали свой поиск мы с небольшого **CSV** файла, в котором находились данные о заболеваемости и смертности на 2020-ый год. Для начала этого было достаточно, но лишь для начала. Со временем, мы поняли, что актуальность этих данных очень низкая, так что, перешли к файлу побольше, котором уже были данные по сей день. Это очень помогла в работе с данными, так как мы стали лучше представлять то, о чём идёт речь и данные стали более осязаемыми, даже новости пошли в бой.

Но это был один из самых простых наших шагов. Дальше же пошла настоящая работа.

**Обработка данных:**

Для преобразования всех наших данных мы использовали Python и конкретно его библеотеку - Pandas. Так как изначально у нас был просто CSV файл, с которым было абсолютно неудобно работать, мы стали его преобразовывать в две разные таблицы - это таблица смертности и таблица заболеваний. Кроме того, эти две таблицы у нас были отсортированы по дням, что нам очень помогло в будущем. Также, так вышло, что нам пришлось удалить несколько стран, которые не имели веса в дальнейшем будущем, так как просто нагружали наш код лишними операциями, но при этом были без кого-либо заболевшего или умершего. К тому же, были страны, которые не имели данных в начале, но имели все остальные данные, из-за них нам пришлось обрезать всю нашу таблицу в начале. Зато, все эти действия нам позволили более продуктивно работать, так что, определённо, это стоило потраченного времени.

**Визуализация данных:**

Для визуализации данных изначально мы использовали библиотеку Numpy, но она не позволила бы нам вывести все наши диаграммы на отдельную страничку, так что, в итоге мы перешли к набору библиотек Dash и Plotly, что уже позволило задуманное нами воплотить в реальность. И тут, мы начали сталкиваться с одной проблемой за другой проблемой.

Прежде всего, мы совсем не знали, как работать с этими библиотеками, что было решить проще всего, но дальше были проблемы, которые решить было в разу трудней….