Link_Of_Problem_Statement - https://www.google.com/url?q=https://www.notion.so/aman-vats-masai-school/COVID19-India-b1204116434d40a7aec1044c3f18482b&sa=D&source=editors&ust=1664387981014619&usg=AOvVaw1AhlgL8RRddN8ZI3c0qFCC

- Use `python` to get the data of covid situation in India (Use python `requests` module to do that - https://requests.readthedocs.io/en/latest/) - from below given source
    - https://data.covid19india.org/v4/min/data.min.json
    - https://data.covid19india.org/v4/min/timeseries.min.json
- Use python to parse this json data, below given references can be helpful:
    - https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Objects/JSON
    - https://docs.python.org/3/library/json.html
- In order to better understand the data, use the below given resources:
    - https://data.covid19india.org/documentation/v4_data.html
    - https://data.covid19india.org/documentation/timeseries.min.html

---

- Data Preparation
    - Get comfortable with the data
    - Clean the data :
        - Remove the irrelevant data
        - Keep things in consistent structure and remove inconsistencies (if any)
        - Fix missing data (missing value imputation)
        - Remove outliers (if any)
- Use `SQL` to generate aggregation and then present the aggregated result in an Excel dashboard (please note minimal aggregations has to be done in Excel, most of the aggregations have to be done in SQL and just the report being presented in excel)
    - Your aggregated reports and that shown in Excel should be able to bring following insights:
        - Weekly evolution of number of confirmed cases, recovered cases, deaths, tests. For instance, your dashboard should be able to compare Week 3 of May with Week 2 of August
        - Let's call `testing ratio(tr) = (number of tests done) / (population)`, now categorise every district in one of the following categories:
            - Category A: $0.05 \leq tr \leq 0.1$
            - Category B: $0.1 < tr \leq 0.3$
            - Category C: $0.3 < tr \leq 0.5$
            - Category D: $0.5 < tr \leq 0.75$
            - Category E: $0.75 < tr \leq 1.0$

            Now perform an analysis of number of deaths across all category. Example, what was the number / % of deaths in Category A district as compared for Category E districts

    - Generate 2 - 3 insights that is very difficult to observe

- Compare delta7 confirmed cases with respect to vaccination
- Make at least 2 such KPI that presents the severity of case in different states (example: Any numerical measure to comment on how severe were the cases in Bihar as compared to that of Kerala)
- Categorise total number of confirmed cases in a state by Months and come up with that one month which was worst for India in terms of number of cases

---