|  |  |
| --- | --- |
| https://lh6.googleusercontent.com/ENp5iVIZzIHv2b4n4mkS4vHt4UkEBlWYmz738CY-LLW4hqPglKHQ6jUh7UR0d4Ymmrr709Maa1nSxWlKGCsAkajArO1QhFOMQzxwHhEhk0zqlMRi4_H7oj2RIxwzkiz_qKiY_Giy | **Work Integrated Learning Programmes Division**  **M.Tech (Data Science and Engineering)** |

**Data Visualization & Interpretation   
(DSECL ZG555)**

**Assignment 1 – [COVID-19 DATA ANALYSIS] - [Weightage 12%]**

1. **Problem Statement**

Coronavirus disease 2019 (COVID‑19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic. As of 17 September 2020, more than 29.8 million cases have been reported across 188 countries and territories with more than 940,000 deaths; more than 20.3 million people have recovered. As the number of cases were increasing day by day, Indian Government has taken multiple measures to handle the situation which includes the lockdown for the entire nation. However, the primary purpose of the lockdown was to slow down the spread of the disease and give time to governments, both Centre and states, to be better prepared on all fronts to deal with the fallout of rising infections and a drastically slowing economy. Assuming the trajectory of new cases starts showing a flattening curve, the costs of extending the lockdown will far exceed the benefits. Now there are discussions about partial reopening of Colleges /Schools as well.

With the Government planning to reopen colleges, the principal of the Engineering Collegewhere you studied wanted to give an awareness webinar to **the students** before they start attending classes. Knowing that you are a Data Science Consultant, he thought that you would be the best candidate for the purpose.

With the given context, you need to create a dashboard using TABLEAU. (Use the concepts learned in the class).

[Datasets](https://api.covid19india.org/documentation/csv/)



**Data preparation has to be done**. If required, feel free to make proper assumptions about the data and the analysis process to be followed. Don’t forget to mention the assumptions at the right places.

We have created a new column name as **Active Adjustment** using Custom calculated field. We used below calculation to get the Active cases of COVID-19 (state wise).

IF [Status] = "Confirmed" THEN [Count]

ELSEIF [Status] = "Recovered" THEN -[Count]

ELSEIF [Status] = "Deceased" THEN -[Count]

ELSE 0

END.

Consider adding the below disclaimer to your dashboard.

**Disclaimer:**

**I am a**data visualization student**,** not an infectious disease expert or epidemiologist**. This dashboard was created**only to fulfil the academic requirements of the course and should not be used for medical decision making. *.*I bear no responsibility towards the correctness of the data**.**

The disclaimer can be added as a footnote or other text on the page, as a tooltip, or even as a collapsible floating container that almost requires that a reader review and close the box to view the visualizations.

**The objectives include**

**Demonstrate the VISUALISATION CONTEXT**

1. **KNOW YOUR AUDIENCE (First question is answered for you)**
2. List the primary groups or individuals to whom you’ll be communicating.

**The students of Engineering College**

1. If you had to narrow that to a single person, who would that be?

**A final-year student who is anxious about returning to college and wants to understand the safety and risk factors.**

1. What does your audience care about?
   * **Trends in COVID-19 cases in their state**
   * **The safety of students attending offline classes.**
   * **Understanding how COVID-19 has evolved and how it impacts educational institutions.**
   * **Making informed decisions about college reopening based on data.**
   * **Preventive measures and risk zones**
2. What action does your audience need to take?
   * **Understand the gravity of the pandemic through data.**
   * **Make informed decisions about attending offline classes and following guidelines.**
   * **Follow health and safety protocols**
   * **Stay informed about COVID trends**
3. What is at stake? What is the benefit if the audience acts in the way you want them to? What are the risks if they don’t?
   * **Benefit: Informed and alert students are less likely to contribute to virus spread; safer campus.**
   * **Risk: Unawareness can lead to outbreaks in colleges, risking lives and disrupting academic sessions.**

## WHAT?

* What are you trying to communicate? What questions are you trying to answer/display in your visualizations? Write these as specific questions. You need to come up with 3 questions at least, each of which will be answered using one Viz.

**Answer:** We aim to communicate the impact and trends of COVID-19 in India through clear and insightful data visualizations. The goal is to help engineering students develop a factual understanding of how the pandemic unfolded—both over time and across different states.

The primary objective is to raise awareness among students about the seriousness of the pandemic using data-driven insights, helping them grasp the scale of the crisis and the effectiveness of the measures taken to control it. By visualizing patterns in the spread of infections, state-wise comparisons, and healthcare outcomes such as recovery, we aim to promote informed decision-making and encourage adherence to safety protocols.

These insights are especially crucial as educational institutions move toward reopening. A well-informed student body will be better equipped to take personal and collective precautions, reducing the risk of infection and supporting a safer return to campus life.

**Question 1 (Chart 1): How has the number of COVID-19 active cases changed over time in different regions?**

**Question 2 (Chart 2): Which states in India have been most affected by COVID-19?**

**Question 3 (Chart 3): How have COVID-19 recovery trends varied over time in India?**

* **Data preparation needed to answer the specific queries must be done.**

1. **Present the BIG IDEA**.

* It should: (1) articulate your point of view, (2) convey what’s at stake, and (3) be a complete (and single!) sentence.

**Understanding India's COVID-19 trends through interactive visualizations empowers students to make informed decisions as they return to campus post-lockdown.**

## HOW?

* 1. Chart 1: What type of viz did you create? Why did you select the viz that you did?

### **Chart 1 Type:**

### Multi-Line Time Series Chart

### **Why this visualization?**

* Allows comparison of case trends across multiple states on a common time scale.
* Highlights the differences in peak periods, wave patterns, and how regions responded over time.

**Key Observations:**

1. A smaller rise can be observed around August–September 2020, showing the impact of the first wave.
2. A sharp peak appears between March and May 2021, representing the second wave of COVID-19 in India. This spike is the most prominent in the entire graph.
3. Maharashtra shows a very sharp increase during both waves.
4. After May 2021, there is a sharp decline in active cases across all states, showing effective control measures post-second wave.

**Note:** From the given state wise daily dataset, we found three entries which are not considered as a state in India such as TT (Total count respective to date), UN (Unknown) and DD (Daman & Diu, which is no longer a separate code). Hence, in the first sheet, we excluded those entries and showing data for only 36 out of the 39 entries—representing the actual states and UTs of India.

* 1. Chart 2: What type of viz did you create? Why did you select the viz that you did?

### **Chart 2 Type:**

### **Symbol Maps (Geo Map)**

### **Why this visualization?**

* Helps to show a clear, sortable comparison of states based on COVID-19 case counts.
* Exact **comparative ranking** is the goal.

**Key Observations:**

1. Colour gradient is used to show variation from lowest (light blue) to highest (dark blue) case counts.
2. Maharashtra shows the darkest shade, indicating it has the highest number of confirmed COVID-19 cases in India.
3. North-eastern states, Jammu & Kashmir, and smaller union territories like Ladakh, Chandigarh, and Andaman & Nicobar Islands are shaded the lightest, reflecting lower case counts.
   1. Chart 3: What type of viz did you create? Why did you select the viz that you did?

### **Chart 3 Type:**

### Line Chart (Time Series) of Recovered Cases Over Time

### **Why this visualization?**

* A line chart is the most effective way to show how values change over a continuous variable.
* It allows the viewer to easily identify rises, peaks, and plateaus in recovery numbers.

E.g. A sharp upward trend in recovered cases during May 2021 (post-second wave), A slower rise during early 2020 due to limited testing and isolation.

**Key Observations:**

1. The recovery count started very slowly, reflecting limited infections and recoveries in the early stages (April 2020 – August 2020).
2. The curve continues to rise in a relatively linear fashion, indicating consistent recoveries (September 2020 – March 2021).
3. A very steep increase in recoveries is observed during (April 2021 – June 2021) period. This is due to the aggressive recovery efforts like vaccinations, better treatment protocols.
4. A flattening curve near the end is a positive sign, showing that the new cases were decreasing, most patients from earlier waves had recovered and the recovery trend was stabilizing.
   1. For each of the Visualisation, identify at least 3 Gestalt principles employed.

**1. COVID-19 Active Cases Trend Over Time by State (Line Chart)**

**Gestalt Principles Used:**

1. **Similarity**:
   * Each state is represented by a **different coloured line**, helping viewers differentiate and mentally group the data by state.
2. **Continuity**:
   * The lines follow a **smooth temporal path** (from left to right), allowing the viewer’s eye to follow trends naturally over time.
3. **Proximity**:
   * The lines are drawn **close together**, which encourages comparison and shows that all states belong to a common category (COVID active cases over time).

**2. State-Wise Spread of COVID-19 in India (Map View)**

**Gestalt Principles Used:**

1. **Proximity**

* States geographically close are visually grouped naturally on the map, helping users relate regional spread patterns.

1. **Similarity**

* States with similar COVID-19 case counts are often coloured using a gradient or categorical colour scheme to represent severity. Similar colours imply similar data values, helping viewers quickly categorize regions by case levels.

1. **Enclosure**

* The map boundary itself acts as an enclosure grouping all states within India. Sometimes, borders or state outlines are emphasized to distinguish between states clearly.

1. **Figure/Ground**

* The map (figure) stands out against a neutral background (ground), focusing attention on the data rather than on irrelevant surroundings.

**3. India’s COVID-19 Recovery Journey Over Time (Line Chart)**

**Gestalt Principles Used:**

1. **Continuity**

* The line chart connects data points smoothly over time, helping viewers naturally follow the trend of recoveries day by day or month by month.

1. **Figure/Ground**

* The line (figure) stands out against a clear, uncluttered background (ground), ensuring the trend is the focus.

1. **Enclosure**

* Axes, gridlines, or chart borders enclose the data area, visually grouping the information and aiding interpretation.
  1. For each of the Visualisation, mention how you strategically used pre-attentive attributes to draw the audience's attention.

**1. COVID-19 Active Cases Trend Over Time by State (Multi-Line Chart)**

**Pre-attentive Attributes Used:**

1. **Colour (Hue & Saturation)**

* Different states’ lines are usually coloured distinctly (different hues) to separate them visually. Saturation or brightness might be used to emphasize certain states over others.

1. **Position**

* Lines are plotted along the x-axis (time) and y-axis (active cases), allowing viewers to quickly understand trends by their position.

1. **Orientation**

* The slope or direction of each line indicates increase or decrease in active cases, helping viewers immediately grasp trends.

**2. State-Wise Spread of COVID-19 in India (Map View)**

**Pre-attentive Attributes Used:**

1. **Colour Gradient (Intensity)**:

* States with higher case counts are shown in **darker shades**, immediately drawing the viewer’s attention to the **most affected regions**.

1. **Position**

* Geographic location of states on the map helps viewers instantly relate data to real-world regions.

1. **Shape**

* Distinct state boundaries and shapes help differentiate regions.

**3. India’s COVID-19 Recovery Journey Over Time (Single Line Chart)**

**Pre-attentive Attributes Used:**

1. **Position**

* Data points are positioned along the x-axis (time) and y-axis (recovery counts), allowing quick perception of the trend.

1. **Length/Size**

* The **length and steepness** of the curve in early 2021 highlights the rapid increase in recoveries without needing labels or explanation.

1. **Orientation**

* The upward or downward slope of the line conveys increases or decreases in recoveries instantly.

1. **Create your dashboard**

## Deliverables

Zipped file containing

1. The word doc with answers to question 1 through 4 above.
2. The tableau workbook(.twbx) with 3 Sheets(Each sheet should have 1 visual) and 1 Dashboard comprising all three visuals.
3. The source file after preprocessing(if any).

The file name should be the respective group name.

## Deadline

* The strict deadline for submission of the assignment is <**June 22, 2025**> **EoD.**
* Late submissions won’t be evaluated.

## How to submit

* This is a group assignment.
* All members of the group will work on the same problem statement.
* Each group should zip the deliverables and upload in CANVAS in respective locations.
* Assignment submitted via means other than through CANVAS will not be graded.

## Evaluation

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Criteria** | **Description** |
| 1 | Know your audience.  (10%) | * Getting to know our audience and understanding their needs and what drives them is an important early part of the process for successfully communicating with data. |
| 2. | WHAT?  Effective Formulation of Contextual Questions  (25%) | * Identify what needs to be communicated very clearly and frame the questions accordingly justifying the context. |
| 3 | BIG IDEA (10%) | * The Big Idea can help us get clear and succinct on the main message we want to get across to our audience |
| 4 | Choice of appropriate visuals  (25%) | * Identify the appropriate visuals for communicating the message |
| 5 | No clutter in the visuals  (10%) | * The visuals presented should not have any unwanted elements that reduces the understanding of data |
| 6 | Audience attention  (10%) | * The visuals presented should have the right kind of visual cues that helps the audience to focus the attention wherever required. |
| 7 | Dashboard in Tableau  (10%) | * Use Principles of Effective Dashboard Design to come up with an interesting Dashboard |

**ALL GROUP MEMBERS WILL BE CREDITED THE SAME MARKS. ITS INDIVIDUAL’S RESPONSIBILITY TO ENSURE HIS/HER PARTICIPATION AS WELL AS TEAM’S RESPONSIBILITY TO ENSURE EVERYONE’S PARTICIPATION.**

|  |  |  |
| --- | --- | --- |
| **BITS ID** | **Name** | **Participation** |
| 2024DA04077 | Sumanth Reddy | 100% |
| 2024DA04078 | Phadke Swarali Sachin Smita | 100% |
| 2024DA04079 | Doddi Dileep Kumar | 100% |
| 2024DA04076 | Vanjari Mahesh | 100% |