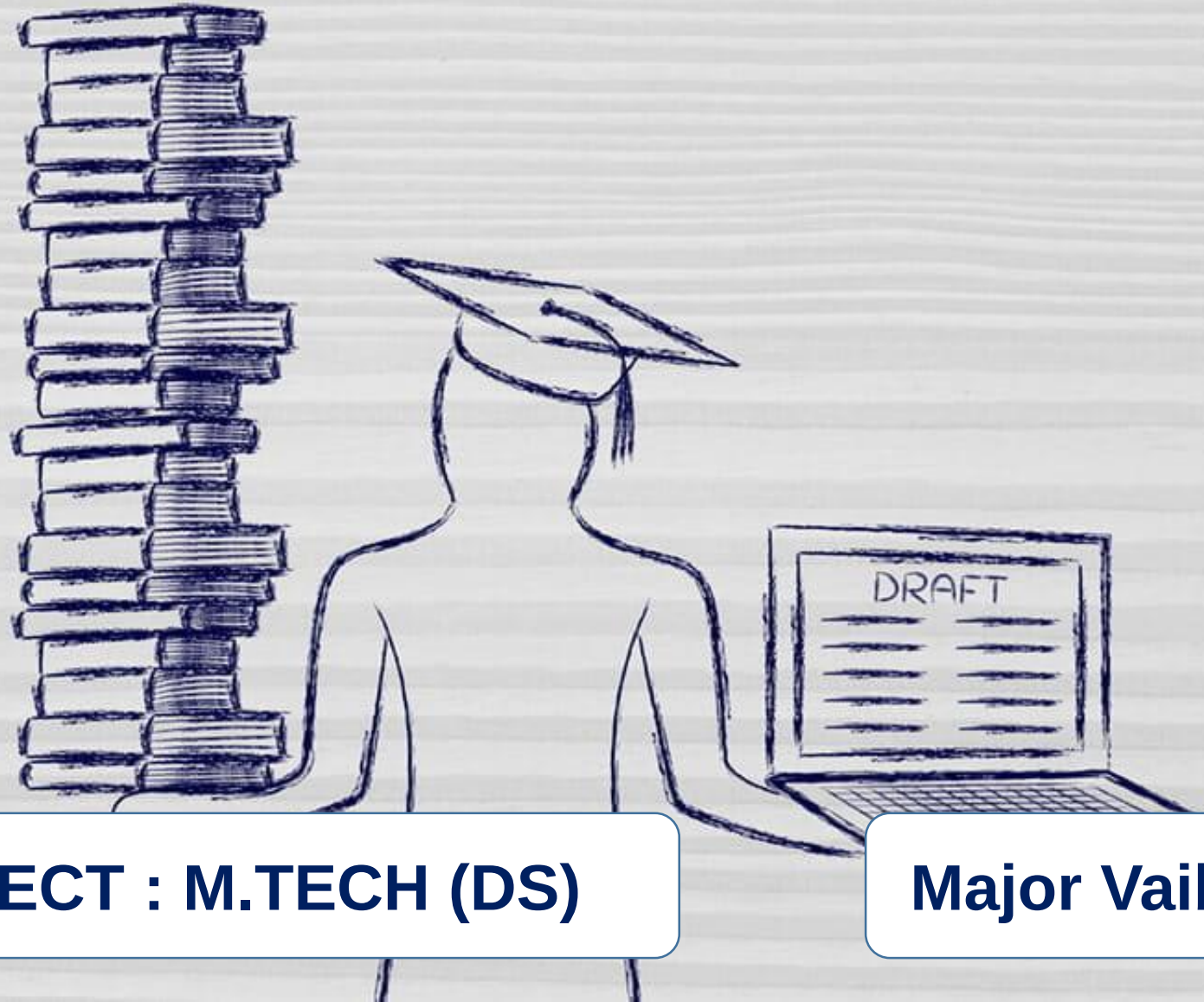
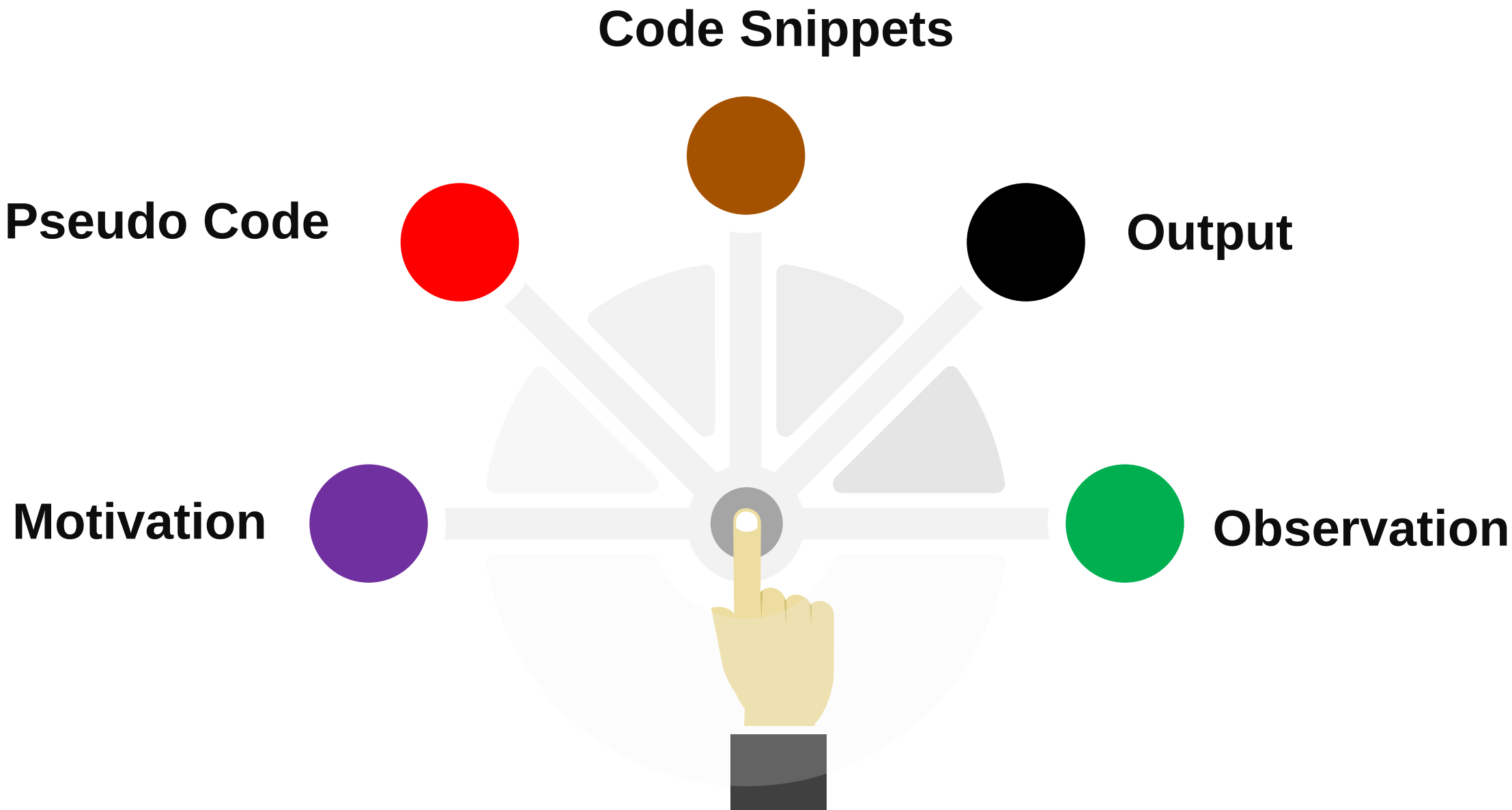
Decoding Nation's Health: Patterns and Trends with Python



PYTHON PROJECT : M.TECH (DS)

Major Vaibhav Mishra

SCOPE



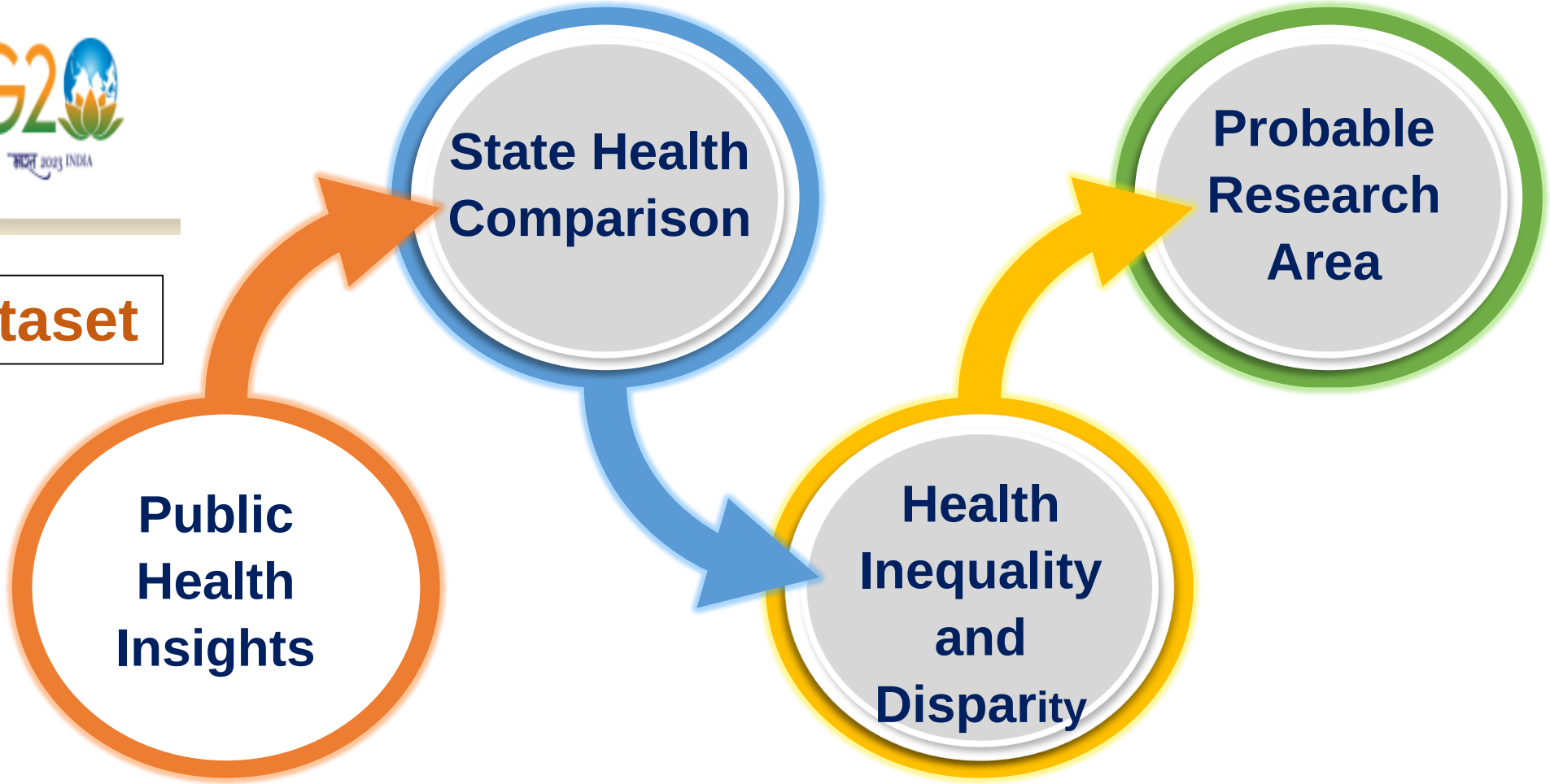
MOTIVATION

A Digital India Initiative

data.gov.in Open Government Data (OGD) Platform India

G20 भारत 2023 INDIA

Authentic Dataset



PROJECT WORKFLOW

Identifying the Data Source

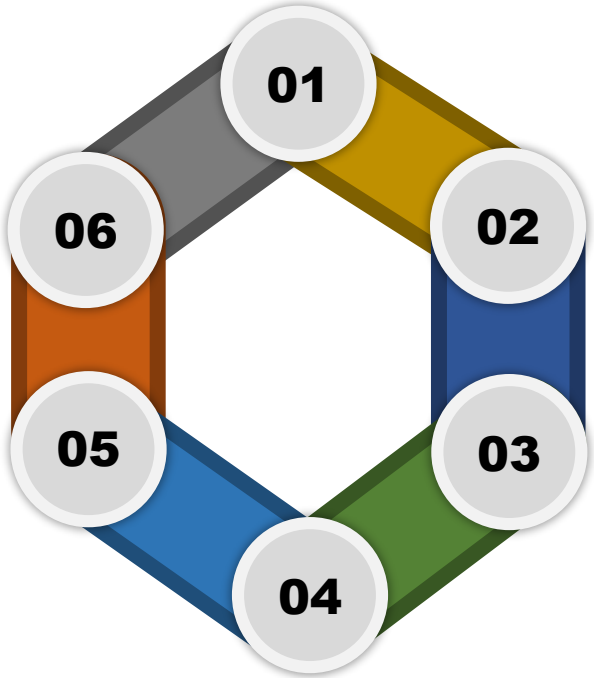
Build a Problem Statement

Data Handling

Building Data Visualization

Code Re-organisation

Code Testing
And Improvement



Pseudo Code

```
# Data Analysis Class
DataAnalysis:
    analyze_sex_ratios():
        extract_relevant_columns()
        calculate_mean_sex_ratios()
        combine_mean_values()
        plot_horizontal_bar_chart()
        identify_states_with_higher_birth_ratio()

    analyze_alcohol_consumption():
        extract_alcohol_columns()
        median_alcohol_consumption()
        alcohol_consumption_chart()

    analyze_blood_pressure():
        calculate_mean_blood_pressure()
        blood_pressure_comparison_table()
        plot_blood_pressure_comparison_chart()
```

```
# Choropleth Map Class
ChoroplethMap:
        load_and_clean_data()
        replace_state_names(replace_dict)
        randomly_columns_for_visualization()
        group_and_merge_data()

    plot_choropleth_map():


# Scatter Plot Generator Class
ScatterPlotGenerator:
    load_and_clean_data()
    plot_scatter_plot(col1, col2):

    execute_scatter_plot():
        display_available_columns()
        get_user_input_for_columns()
        plot_scatter_plot_based()
```

Code Snippets

Data Cleaning

```
self.df.iloc[:, 2:] = self.df.iloc[:, 2:].apply(pd.to_numeric, errors='coerce')
self.df = self.df.replace('*', pd.NA)  # Replacing '*' with NaN
self.df.iloc[:, 2:] = self.df.iloc[:, 2:].abs()  #
self.df.fillna(self.df.mean(), inplace=True)
```

Data Grouping

```
grouped_data = self.df.groupby('State/UT')[self.selected_feature].mean().reset_index()
self.gdf = pd.merge(self.gdf, grouped_data, left_on='ST_NAME', right_on='State/UT', how='left')
self.gdf = self.gdf.dropna(subset=[self.selected_feature])
```

Code Snippets

Choropleth plot

```
colors = [(0, 0, 1), (1, 1, 1), (1, 0, 0)] # Defining the color gradient
cmap = LinearSegmentedColormap.from_list('custom', colors, N=256)
fig, ax = plt.subplots(1, 1, figsize=(15, 10))
self.gdf.plot(column=self.selected_feature, cmap=cmap, linewidth=0.8, ax=ax, edgecolor='0.8', legend=True)
```

Handling Column Inputs

```
for i, col in enumerate(self.df.columns[5:16]):
    print(f"{i + 1}. {col}")

selected_column_indices = [int(input(f"Enter the index of the first column: ")) - 1,
                           int(input(f"Enter the index of the second column: ")) - 1]

col1, col2 = self.df.columns[5:16][selected_column_indices]
self.plot_scatter_plot(col1, col2)
```

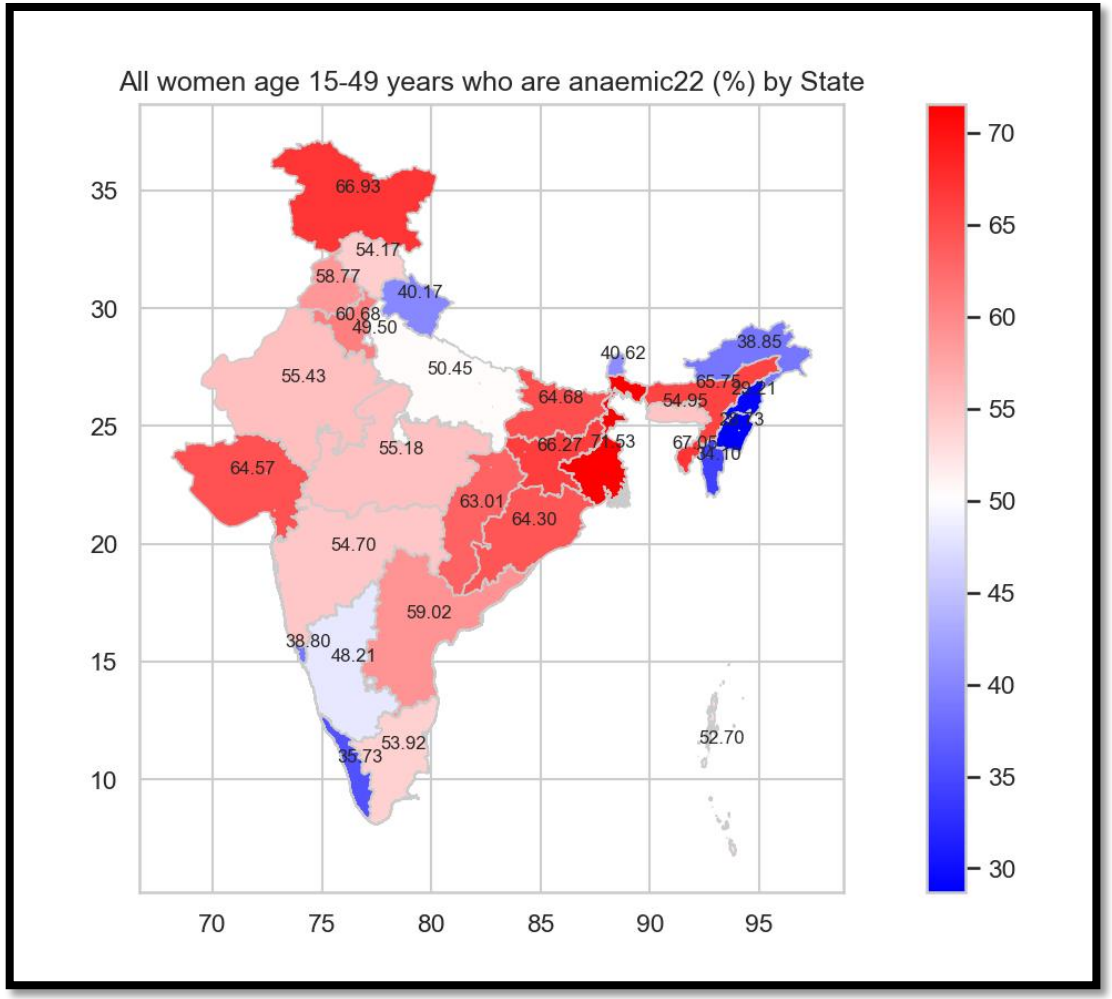
OUTPUT

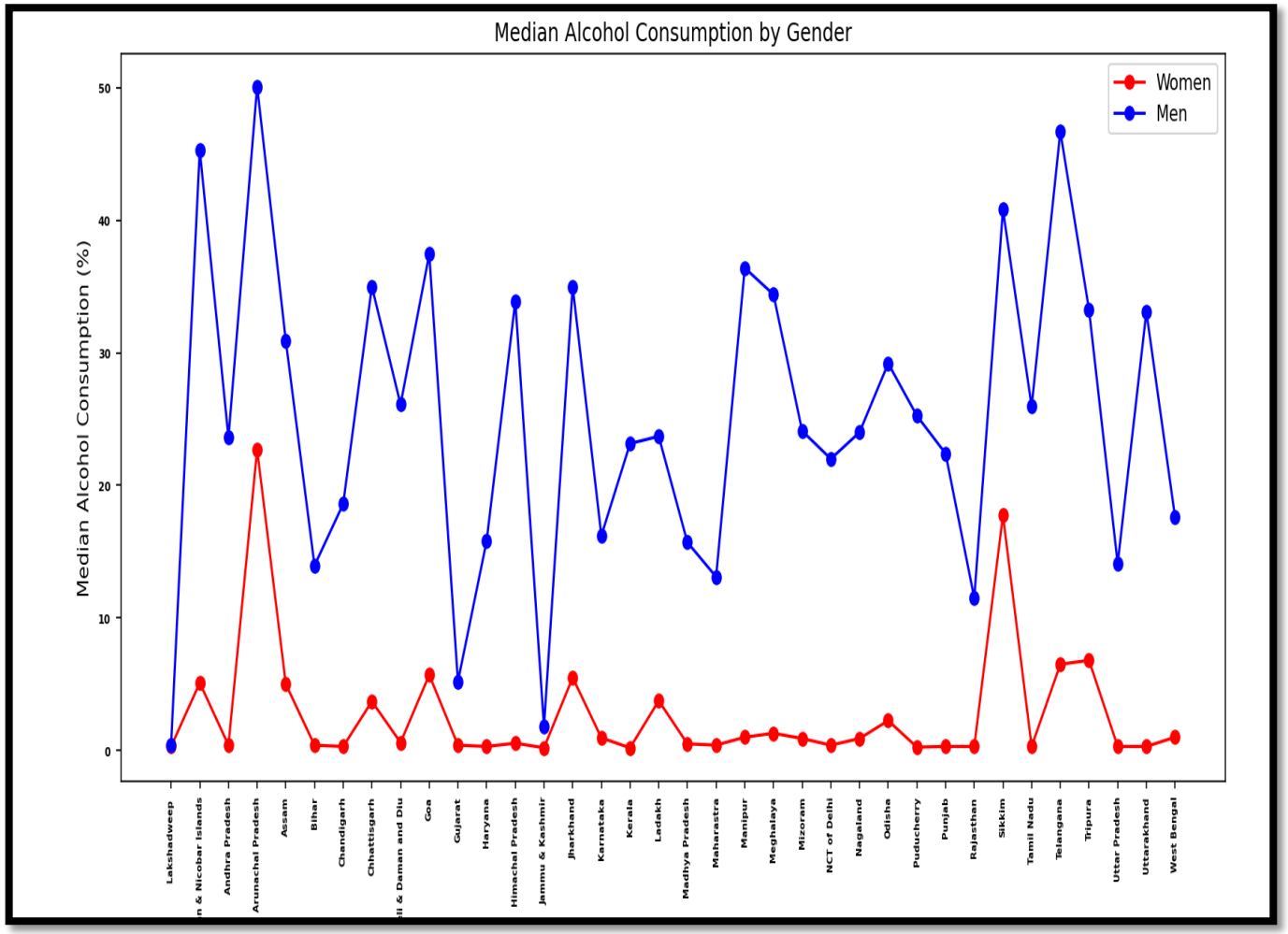
```
**********************************************************************************************************
*                                        National Health Data Analysis Project
*                                        Made By: Major Vaibhav Mishra
**********************************************************************************************************
Enter Respective Digits to Analyze


1 Sex Ratios Analysis

2 Alcohol Consumption Analysis

3 Blood Pressure Analysis

4 Data on Choropleth Map With 15 Options

5 Scatter Plot : Select your own coloumn

Enter 'Exit' to quit:
```

DASHBOARD VIEW

```
Available columns for scatter plot:
1. Female population age 6 years and above who ever attended school (%)
2. Population below age 15 years (%)
3. Sex ratio of the total population (females per 1,000 males)
4. Sex ratio at birth for children born in the last five years (females per 1,000 males)
5. Children under age 5 years whose birth was registered with the civil authority (%)
6. Deaths in the last 3 years registered with the civil authority (%)
7. Population living in households with electricity (%)
8. Population living in households with an improved drinking-water source1 (%)
9. Population living in households that use an improved sanitation facility2 (%)
10. Households using clean fuel for cooking3 (%)
11. Households using iodized salt (%)
```

OUTPUT

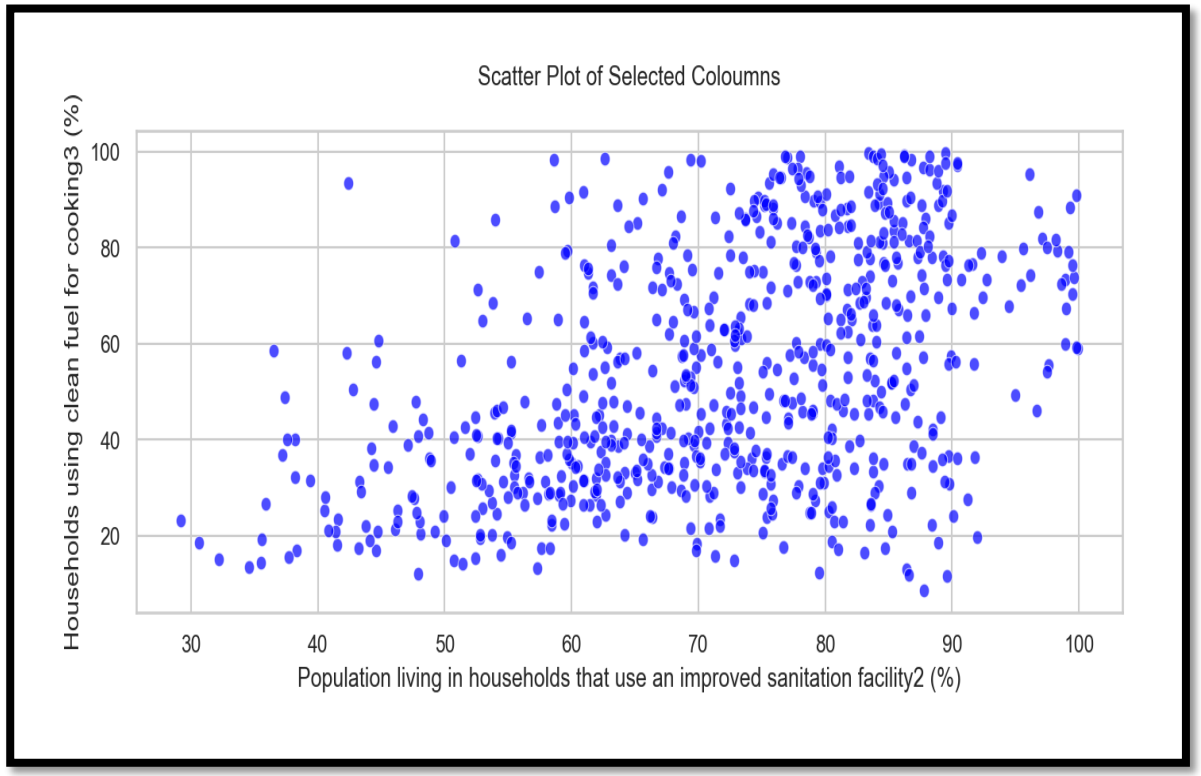


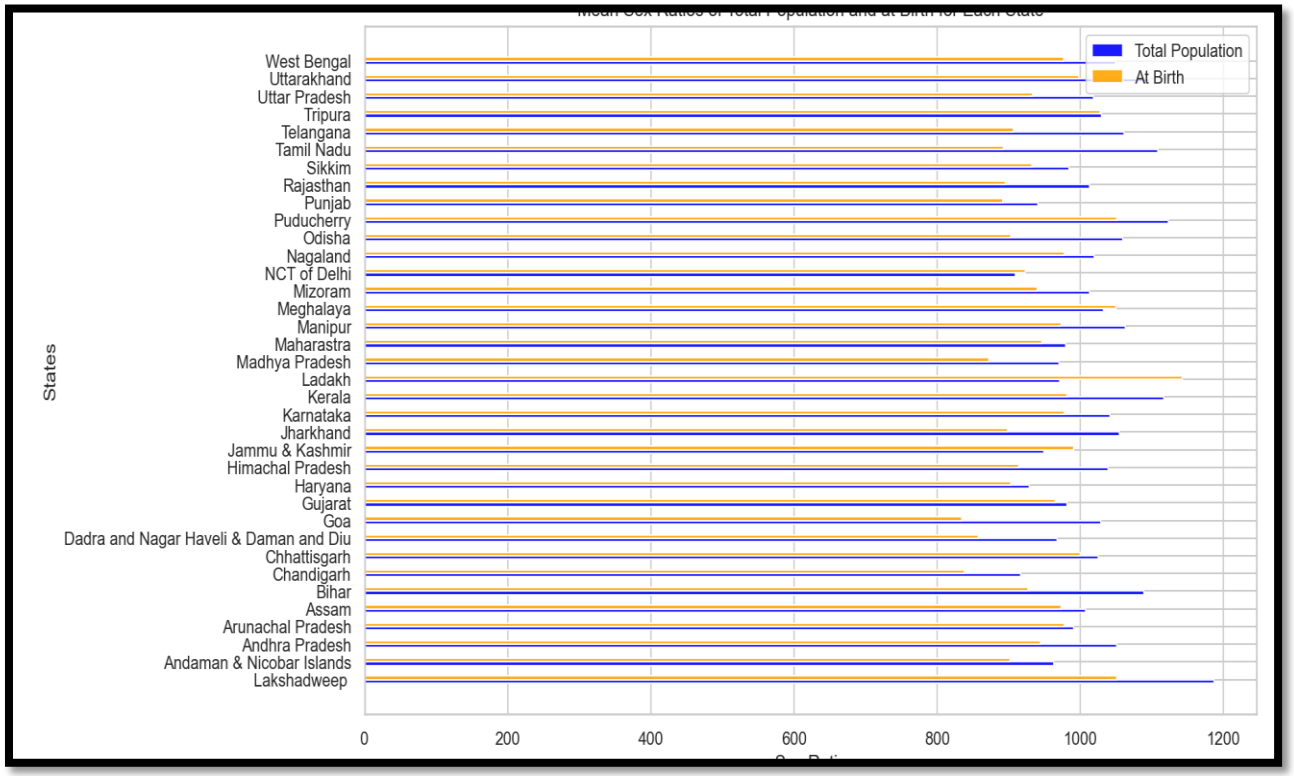


CHOROPLETH

LINE PLOT

OUTPUT





SCATTER PLOT

BAR PLOT

OBSERVATION

Understanding Nature of Data is Important

Object Oriented Approach

Use of Geopandas , Pandas , Matplotlib

User Interaction

Importance of Debugging

https://github.com/vaibhav2404/NFHS-5.git

THANK YOU

QUESTIONS