(Read readme.md file first)
N Yogeesh
MBA(BA)
IIFT Delhi

Report

Analytical Dashboard on Total cybercrimes in metropolitan cities from 2019 to 2019

Introduction:

This R Shiny dashboard provides an interactive platform to explore and analyze cybercrime data in metropolitan cities across India from 2019 to 2021. The dashboard offers a variety of features, including:

Data exploration: View the data structure, summary statistics, and raw data.

Visualization: Create histograms, identify top cities with high crime rates, and analyse trends over time.

Hypotheses testing: Conduct ANOVA tests to assess if there are statistically significant differences in cybercrime cases and arrests across the years.

This dashboard can be used by researchers, policymakers, and other stakeholders to gain insights into cybercrime patterns and trends, and to inform data-driven strategies for prevention and mitigation.

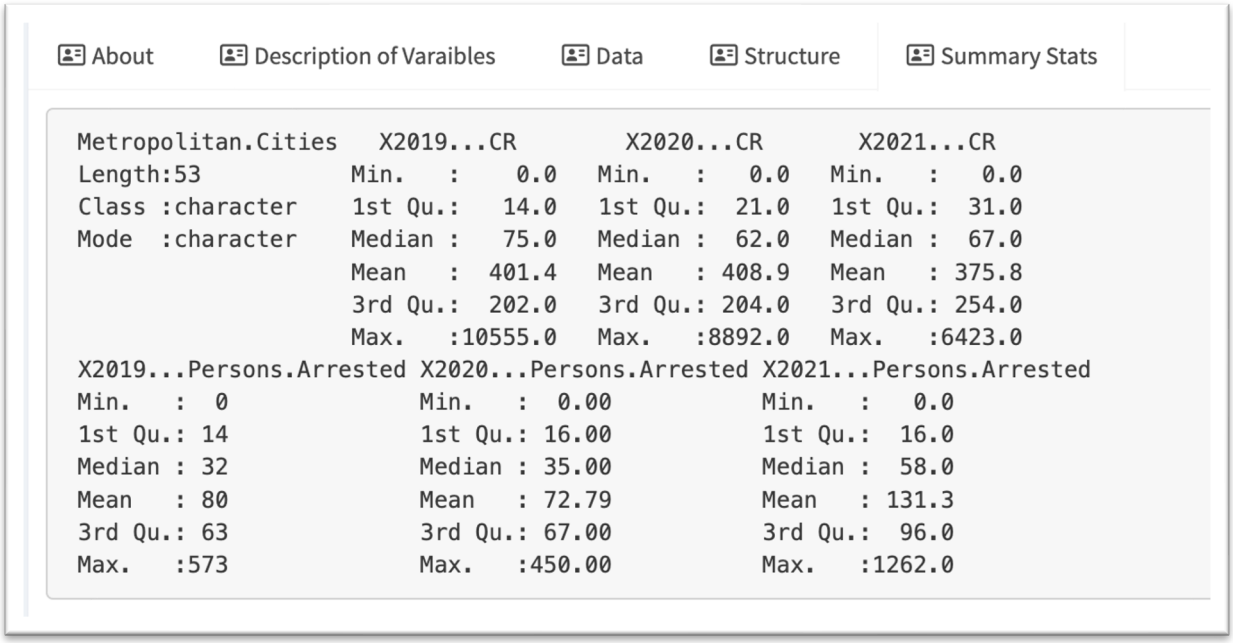
Summary of Dashboard: First tab i.e. Data tab covers summary of data. Second Tab i.e. Visualization tab covers interactive graphs and third tab i.e. Analysis tab covers ANOVA tests to check if number of crime cases registered and number of accused persons arrested over the years increased or decreased.

Detailed Analysis

Data Tab:

- **About:** This tab provides information about the dataset, including its source, content, and goals of the analysis. It also includes an image for visual reference.
- **Description of Variables:** This tab explains the meaning and purpose of each variable in the dataset.
- **Data:** This tab displays the entire dataset in a data table format.

- **Structure:** This tab shows the data structure, including data types and dimensions.
- **Summary Stats**: This tab displays summary statistics for each variable, such as mean, median, standard deviation, etc.
    - ***Analysis:*** Here summary function of R is used on all variables of data frame. Result:
        - Number of cybercrime cases registered in 2019 have mean 401 in 2019, 408 in 2020 and 375 in 2021. Also, there are states with 0 cases and maximum case for a state is 10,555 in 2019, 8892 in 2020 and 6423 in 2021.
        - Number of accused persons arrested in cybercrime cases in 2019 have mean 80 in 2019, 73 in 2020 and 131 in 2021. Also, there are states with 0 accused persons arrested and maximum Number of accused persons arrested in cybercrime cases is 573 in 2019, 450 in 2020 and 1262 in 2021.



    - ***Implications:***
        - Varied Cybercrime Rates: The variation in the number of cybercrime cases across states indicates a diverse landscape of cyber threats. Understanding the factors contributing to this variation can help tailor cybersecurity measures based on regional needs.
        - Yearly Fluctuations: The slight increase in the mean number of cases

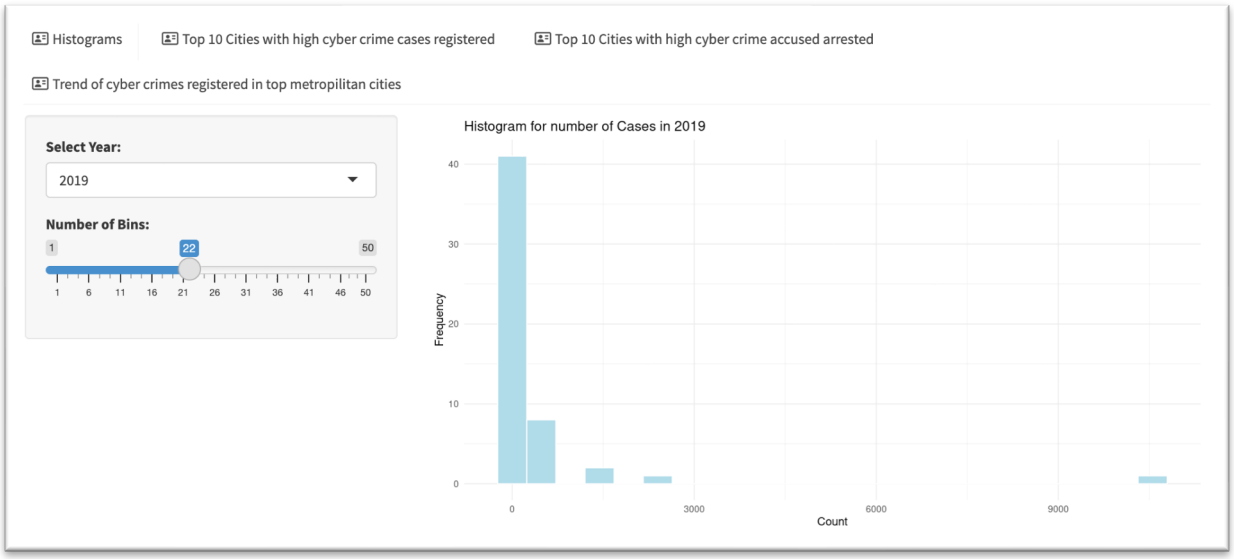
from 2019 to 2020 followed by a decrease in 2021 suggests a potential trend. Further investigation into the causes of these fluctuations could inform cybersecurity strategies and resource allocation.

- Resource Allocation for Investigations: States with consistently high numbers of cases and arrests may require additional resources for cybercrime investigations and law enforcement activities. Identifying these states is crucial for allocating resources effectively.
- Targeted Enforcement Strategies: States with a low number of reported cases or arrests may benefit from targeted awareness campaigns and enhanced cybersecurity training to prevent cybercrimes. Understanding the characteristics of states with 0 cases is essential for developing preventive strategies.
- Capacity Building: The significant variability in the maximum number of cases and arrests highlights the need for capacity building in states facing higher cybercrime rates. Sharing best practices and providing support can enhance the overall cybersecurity posture.
- Strategic Planning for Law Enforcement: Law enforcement agencies should consider the fluctuating trends in cybercrime rates when developing long-term strategic plans. Adapting to evolving threats requires a proactive approach in terms of technology, training, and collaboration.
- Public Awareness and Education: Enhancing public awareness about cyber threats and implementing educational programs can contribute to reducing cybercrime rates. States with lower awareness may need targeted initiatives to educate the public about online risks.

In conclusion, the insights from the summary statistics analysis provide a foundation for strategic planning and decision-making in the realm of cybersecurity. Addressing the implications outlined above can contribute to a more effective and targeted approach in combating cybercrimes at both regional and national levels.

Visualization Tab:

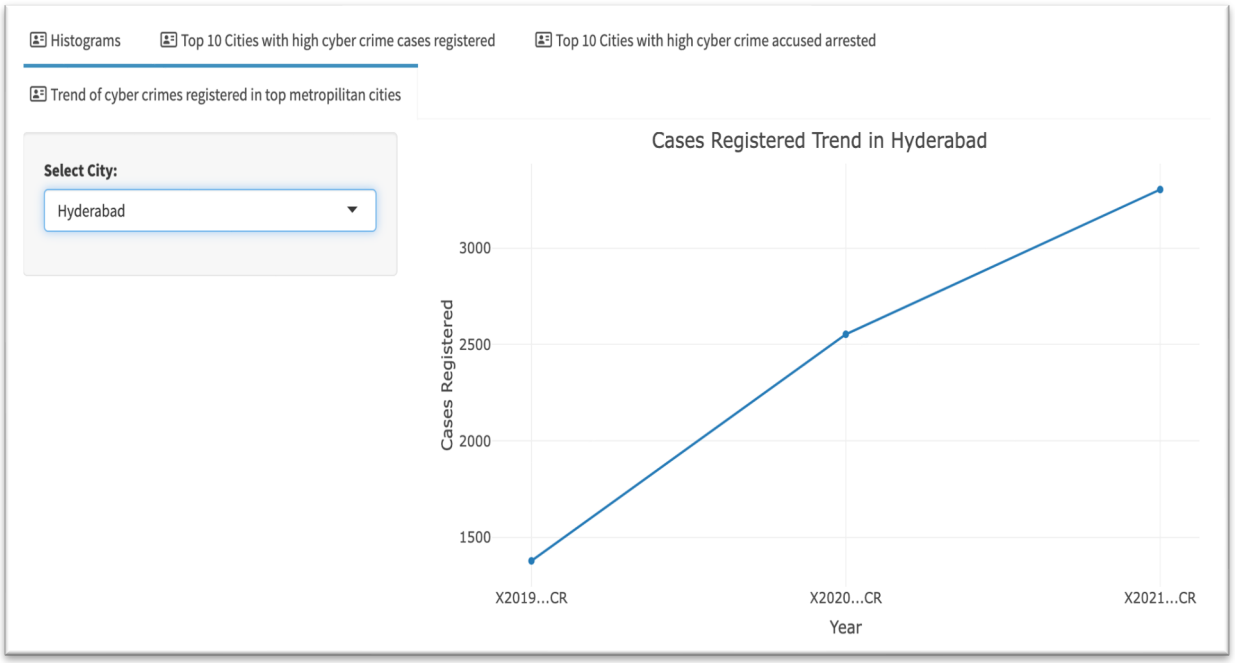
- **Histograms**: This tab allows you to create histograms for the number of cases registered in different years. You can select the year and adjust the number of bins to customize the visualization.



  - *Implications:*
    - The choice of bin size in histograms influences the granularity of the data presentation. A smaller bin size provides a more detailed view, potentially revealing nuances in the distribution. Conversely, a larger bin size may smooth out variations, highlighting broader trends. Careful consideration of bin size is essential for accurate interpretation. If there are spikes or gaps in the histograms, it suggests specific periods or ranges with notable variations in the number of cases. Adjusting the bin size can help uncover patterns and identify whether certain periods experience a disproportionate number of cybercrimes.
    - If there are recurring peaks in certain periods, law enforcement and cybersecurity agencies should consider allocating additional resources, increasing surveillance, and implementing targeted awareness campaigns during those times. This proactive approach can enhance the response to potential spikes in cybercrimes.
- **Top 10 Cities with high cybercrime cases registered**: This tab identifies the top 10 cities with the highest number of cybercrime cases registered in a selected year.

- **Top 10 Cities with high cybercrime accused arrested**: Similar to the previous tab, this one identifies the top 10 cities with the highest number of accused arrested in a selected year.
    - *Implications:*
        - The visualization of the top 10 cities with the highest number of cybercrime cases registered and accused arrested provides valuable insights into the geographic distribution of cybercrimes. Identifying cities that consistently appear in the top 10 implies a higher risk of cybercrimes in those regions. Law enforcement and cybersecurity agencies can use this information to prioritize resource allocation, implement targeted interventions, and enhance collaborative efforts with local authorities in these high-risk areas.
        - The top 10 cities graphs highlight the need for tailoring cybersecurity strategies to specific regions. Cities with a consistently high number of cybercrime cases may require specialized awareness campaigns, community engagement initiatives, and local law enforcement partnerships. Recognizing the unique challenges faced by each city allows for the development of targeted strategies that address the specific dynamics contributing to cybercrime prevalence.
- **Trend of cybercrimes registered in top metropolitan cities:** This tab allows you to track the trend of cybercrime cases registered over time for a selected city. You can choose any of the listed metropolitan cities to analyse its trend.
    - *Implications :*
        - The trend graph for cybercrime cases in top metropolitan cities allows for a city-specific threat analysis over time. Identifying trends, spikes, or patterns in cybercrime occurrences for a selected city provides crucial insights into the city's evolving threat landscape. Understanding whether cybercrimes are increasing, decreasing, or following specific patterns enables law enforcement and cybersecurity agencies to tailor interventions to the unique challenges faced by each metropolitan area.
        - Tracking the trend of cybercrimes in top metropolitan cities helps in

strategic resource allocation and preparedness. If there is a consistent upward trend in cybercrimes, allocating additional resources, enhancing cybersecurity infrastructure, and increasing law enforcement capabilities become imperative. On the other hand, a decreasing trend may require evaluating the effectiveness of existing preventive measures and adjusting resources accordingly. The graph serves as a valuable tool for ensuring that resources are deployed where they are most needed.



Hypotheses Testing Tab:

- **ANOVA test on Cases Registered from 2019 to 2021**: This tab performs an ANOVA test to determine if there is a statistically significant difference in the number of cybercrime cases registered across the years 2019, 2020, and 2021. It displays the test results in text format.
    - **Objective of the Test:**
        - The objective of the ANOVA test on cases registered from 2019 to 2021 is to determine whether there is a statistically significant difference in the mean number of cybercrime cases across the three years. The test assesses whether any observed differences in the means are likely due to actual differences in the population means or if they could be the result of random variation.
    - **Null Hypothesis (H_o):**
        - The null hypothesis for the ANOVA test is that there is no significant difference in the mean number of cybercrime cases registered across the years 2019, 2020, and 2021. Mathematically, this can be expressed as: H_o: mean_{2019} = mean_{2020} = mean_{2021}
        - Where mean_{2019} = mean_{2020} = mean_{2021}) are the population means for the number of cybercrime cases in 2019, 2020, and 2021, respectively.
    - **Alternate Hypothesis (H_1):**
        - The alternative hypothesis for the ANOVA test is that there is a significant difference in the mean number of cybercrime cases across at least two of the three years.
    - **Analysis:**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Year	2	31898	15949	0.01	0.99
Residuals	156	257122522	1648221		

- **Interpretation**
  - F-statistic: The F-statistic is very small (0.01), suggesting that the variability between the group means (across the years) is negligible compared to the variability within the groups (residuals).
  - p-value: The p-value associated with the F-statistic is 0.99, which is much greater than the common significance level of 0.05. Therefore, we fail to reject the null hypothesis.
- **Managerial Implications:**
  - The analysis leads to the conclusion that there is no significant difference in the mean number of cybercrime cases registered across the years 2019, 2020, and 2021. The p-value is substantially higher than the significance level, indicating that any observed differences in the means are likely due to random variation rather than actual differences in population means.
  - **Consistency in Cybercrime Rates:** The lack of statistical significance suggests a consistency in cybercrime rates over the three years. Law enforcement and cybersecurity strategies may not need significant adjustments based on the annual variation in the number of cases.
  - **Resource Allocation:** Resource allocation for cybersecurity measures can be distributed evenly across the three years, with an emphasis on continuous improvement rather than responding to significant year-to-year variations.
  - **Long-Term Planning:** Policymakers may focus on long-term planning rather than reacting to short-term fluctuations. This could involve developing sustainable and adaptive cybersecurity policies that consider the overall stability in cybercrime rates.
  - In summary, the ANOVA test results indicate that the observed variations in the mean number of cybercrime cases across the years are likely due to random fluctuations, and there is no evidence to suggest a statistically significant difference in cybercrime rates between 2019, 2020, and 2021.

- **ANOVA test on number of accused arrested from 2019 to 2021**: Similar to the previous tab, this one performs an ANOVA test to assess the statistical significance of differences in the number of accused arrested across the years.
    - **Objective of the Test:**
        - The objective of the ANOVA test on the number of accused arrested from 2019 to 2021 is to determine whether there is a statistically significant difference in the mean number of accused arrested across the three years. The test assesses whether any observed differences in the means are likely due to actual differences in the population means or if they could be the result of random variation.
    - **Null Hypothesis (H_o):**
        - The null hypothesis for the ANOVA test is that there is no significant difference in the mean number of accused arrested across the years 2019, 2020, and 2021. Mathematically, this can be expressed as:
          H_o: mean_{2019} = mean_{2020} = mean_{2021}
        - Where mean_{2019} = mean_{2020} = mean_{2021} are the population means for the number of accused arrested in 2019, 2020, and 2021, respectively.
    - **Alternate Hypothesis (H_1):**
        - The alternative hypothesis for the ANOVA test is that there is a significant difference in the mean number of accused arrested across at least two of the three years.
    - **Analysis :**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Year	2	107967	53983	2.06	0.131
Residuals	156	4087380	26201		

    - **Interpretation**
        - **F-statistic:** The F-statistic is 2.06, indicating that there is some variability between the group means (across the years) relative to the variability within the groups (residuals).

- **p-value:** The p-value associated with the F-statistic is 0.131, which is greater than the common significance level of 0.05. Therefore, we fail to reject the null hypothesis.

- **Managerial Implications:**
  - The analysis suggests that there is no statistically significant difference in the mean number of accused arrested across the years 2019, 2020, and 2021. The p-value is higher than the significance level, indicating that any observed differences in means are likely due to random variation.
  - **Consistency in Arrest Rates**: Similar to the cybercrime cases analysis, the lack of statistical significance implies a consistency in the number of accused arrested over the three years.
  - **Resource Allocation**: Law enforcement and cybersecurity strategies may not need major adjustments based on the annual variation in the number of accused arrested.
  - **Long-Term Planning**: Policymakers may focus on long-term planning, considering sustainable and adaptive strategies for law enforcement in response to cybercrimes.
  - **Monitoring Trends**: While not statistically significant, the moderate F-statistic suggests a slight variability in the number of accused arrested. Ongoing monitoring of trends may be valuable to detect any emerging patterns or shifts over time.