Crimes Against Indian Women Project (2022)

Sase, Kingsley Kator

2025-03-10

library(tidyverse) library(readxl) library(ggplot2) library(rvest) library(tidytext)

##Introduction This report analyzes crimes committed against women and girl children in India in 2022, using data from the National Crime Records Bureau and scraped data fro statisticstimes.com. The goal was to identify patterns, high-risk regions, and key issues that need urgent attention. This dataset contains the number of cases registered across all Indian states/ UTs against the crimes committed against women (includes adults and minors) that are recognizable within the Indian Penal Code.

Importing the dataset

data <- read_excel("C:/Users/DJ DABIZ/Desktop/#Analyst League Community/Indian Women/CRIMES_AGAINST_W

View column names and first few rows

colnames(data) head(data)

Checking the structure of the data

str(data)

Questions to Answer

#Question 1 #How many states are in the dataset? num_states <- n_distinct(data$`State/UT`) num_states

#Question 2 #Most Common Crimes Against Women crime_summary <- data %>% select(-`State/UT`, -`Total Crime against Women (IPC &SLL)`) %>% summarise(across(everything(), sum, na.rm = TRUE)) %>% pivot_longer(cols = everything(), names_to = "Crime", values_to = "Total") %>% arrange(desc(Total))

#Bar Chart ggplot(crime_summary, aes(x = reorder(Crime, Total), y = Total)) + geom_bar(stat = "identity", fill = "brown") + geom_text(aes(label = Total), hjust = -0.1, size = 3) + coord_flip() + labs(title = "Most Common Crimes Against Women", x = "Crime Type", y = "Total Cases") + theme_minimal()

#Question 3 #Total Crime Against Women total_crime <- sum(data$'Total Crime against Women (IPC &SLL)', na.rm= TRUE) total_crime

#Question 4 #What individual crime has the highest total cases? # The most common single crime committed top_crime <- crime_summary %>% slice(1) top_crime

#Question 5 #State with the highest total crime unsafe_state <- data %>% select(`State/UT`, `Total Crime against Women (IPC &SLL)`) %>% arrange(desc(`Total Crime against Women (IPC &SLL)`)) %>% slice(1)

#Question 6 #Web scraping the total number of women in India from statisticstimes.com

library(rvest)

#Reading the web page url <- "https://statisticstimes.com/demographics/country/india-sex-ratio.php" page <- read_html(url)

#Extracting all the tables tables <- page %>% html_table(fill = TRUE) length(tables) # how many tables were scraped

#Viewing the tables View(tables[[1]]) View(tables[[2]]) view(tables[[3]]) view(tables[[4]]) view(tables[[5]]) view(tables[[6]])

#Getting the female population for 2025 # Inspecting the structure of the table pop_table <- tables[[5]]

female_row_2022 <- pop_table %>% filter(Year == 2022)

female_row_2022

#Extracting and cleaning th female column female_population <- female_row_2022$Female %>% gsub(",", " ", .) %>% as.numeric()

female_population

#Question 7 #Calculate % and number of women/girl child being abused #Using the scraped population abuse_count <- total_crime abuse_percent <- (abuse_count / female_population) * 100

abuse_count abuse_percent

#Question 8 #Does indecent dressing cause rape? # Checking correlation between 'Rape' and 'Indecent Representation of Women' correlation <- cor(data$Rape, data$`Indecent Representation of Women`, use = "complete.obs") correlation

#Question 9 #Analysis and visualization of key crimes (rape, kidnapping, cruelty, domestic violence) key_crimes <- c("Rape", "Kidnapping/Abduction", "Cruelty by Husband/relatives", "Domestic violence", "Trafficking")

Filtering and reshaping the data for plotting

key_crimes_data <- data %>% select(`State/UT`, all_of(key_crimes)) %>% pivot_longer(cols = -`State/UT`, names_to = "Crime", values_to = "Count")

Graph

ggplot(key_crimes_data, aes(x = reorder(`State/UT`, -Count), y = Count, fill = Crime)) + geom_bar(stat = "identity", fill = "brown") + facet_wrap(~ Crime, scales = "free") + geom_text(aes(label = Count), hjust = -0.1, size = 3) + coord_flip() + labs(title = "Key Crimes Against Women by State", x = "State/UT", y = "Reported Cases")

install.packages("tidytext") library(tidytext)

#Showing just top 10 of the states per crime top10_key_crimes <- key_crimes_data %>% group_by(Crime) %>% slice_max(order_by = Count, n = 10) %>% ungroup()

#Graph Viz. ggplot(top10_key_crimes, aes(x = reorder_within(`State/UT`, Count, Crime), y = Count)) + geom_bar(stat = "identity", fill = "brown") + geom_text(aes(label = Count), hjust = -0.1, size = 3) +

coord_flip() + facet_wrap(~ Crime, scales = “free”) + scale_x_reordered() + # Handles grouped ordering labs(title = “Top 10 States by Key Crimes Against Women”, x = “State/UT”, y = “Reported Cases”) + theme_minimal()

#Question 10 #Who is more sexually abused, the girl child or women? data %>% summarise(Women_Total = sum(Rape, na.rm = TRUE), GirlChild_Total = sum(`Sexual Violence towards girl child`, na.rm = TRUE))

#Question 11 #Which states are easily duped? duping_columns <- c(“Trafficking”, “Selling of Minor Girls”, “Buying of Minor Girls”, “Kidnapping/Abduction”)

duped_states <- data %>% select(`State/UT`, all_of(duping_columns)) %>% mutate(Duped_Total = rowSums(across(all_of(duping_columns)), na.rm = TRUE)) %>% arrange(desc(Duped_Total))

head(duped_states, 10) # Top 10 duped states

#Visualization of the top ten states ggplot(duped_states %>% slice(1:10), aes(x = reorder(`State/UT`, Duped_Total), y = Duped_Total)) + geom_bar(stat = “identity”, fill = “brown”) + geom_text(aes(label = Duped_Total), hjust = -0.1, size = 3) + coord_flip() + labs(title = “Top 10 Duped States”, x = “State”, y = “Total Cases”) + theme_minimal()

#Question 12 #What is the percentage of deaths of abused girl child and women? death_cols <- c(“Dowry Deaths”, “Murder with Rape/Gang Rape”)

total_deaths <- data %>% summarise(across(all_of(death_cols), (x) sum(x, na.rm = TRUE))) %>% rowSums()

total_deaths

#Percentage of deaths death_percent <- (total_deaths / total_crime) * 100 death_percent

#Question 13 #From this data, what caused the abuse to women and girl child? top_causes <- c(“Cruelty by Husband/relatives”, “Dowry Deaths”, “Assault due to Dowry”, “Sexual Violence towards girl child”, “Cyber Crimes committed against women”)

data %>% summarise(across(all_of(top_causes), (x) sum(x, na.rm = TRUE)))

#Question 14 #States with High Domestic Violence but Low Dowry Deaths (Top 10 by Ratio) high_dv <- data %>% select(`State/UT`, `Domestic violence`, `Dowry Deaths`) %>% mutate(Ratio = `Domestic violence` / (`Dowry Deaths` + 1)) %>% arrange(desc(Ratio)) %>% slice(1:5)

ggplot(high_dv, aes(x = reorder(`State/UT`, Ratio), y = Ratio)) + geom_bar(stat = “identity”, fill = “brown”) + geom_text(aes(label = round(Ratio, 1)), hjust = -0.1, size = 3) + coord_flip() + labs(title = “Top 5 States: High Domestic Violence to Dowry Death Ratio”, x = “State”, y = “High Domestic Violence”) + theme_minimal()

#Question 15 #States with Highest Attempted Crimes (Acid, Rape) # Prepare the data attempts_top5 <- data %>% mutate(TotalAttempts = `Attempt to Acid Attack` + `Attempt to Commit Rape`) %>% arrange(desc(TotalAttempts)) %>% select(`State/UT`, TotalAttempts) %>% slice(1:5)

Plot

ggplot(attempts_top5, aes(x = reorder(`State/UT`, TotalAttempts), y = TotalAttempts)) + geom_bar(stat = “identity”, fill = “brown”) + geom_text(aes(label = TotalAttempts), hjust = -0.1, size = 3) + coord_flip() + labs(title = “Top 5 States with Highest Attempted Crimes Against Women”, x = “State”, y = “Total Attempted Crimes”) + theme_minimal()

#Question 16 # States with More Crimes Against Girls than Adult Women (Sexual) girl_vs_women <- data %>% mutate(Diff = `Sexual Violence towards girl child` - Rape) %>% filter(Diff > 0) %>% arrange(desc(Diff)) %>% slice(1:10)

ggplot(girl_vs_women, aes(x = reorder(`State/UT`, Diff), y = Diff)) + geom_bar(stat = “identity”, fill = “brown”) + geom_text(aes(label = Diff), hjust = -0.1, size = 3) + coord_flip() + labs(title = “States Where Girl Child Sexual Abuse > Adult Rape”, x = “State”, y = “Difference (Girl Child - Adult)”) + theme_minimal()

##Conclusion This analysis reveals patterns in crime distribution, key danger zones, and emerging forms of abuse such as Rape and child exploitation. These insights can inform prevention policies and help focus awareness campaigns.