ADVANCED MATHEMATICAL STATISTICS

MTH 522

Final project report – I

By

Venkata Karthik Varma Sagi

Student Id: 02078584

Objective

The objective of this project is to examine a dataset on a specific topic and extract valuable insights from it. The project involves utilizing statistical and data analysis techniques to investigate the dataset, detect patterns and trends, and draw informed conclusions based on the results. The ultimate objective is to present the findings of the analysis in a clear and succinct manner, utilizing relevant visual aids and statistical measures to substantiate the conclusions. The main goal of the project is to produce a detailed and informative report that answers the research question and enriches the current knowledge on the topic.

Dataset:

The Global Terrorism Database (GTD) is a comprehensive dataset that contains information on more than 180,000 terrorist attacks that occurred worldwide between 1970 and 2017. This dataset, which is an open-source database, contains events of both domestic and foreign terrorism. Researchers from the National Consortium for the Study of Terrorism and Responses to Terrorism (START), which is based at the University of Maryland, have meticulously gathered, and are maintaining the data in the database. The dataset offers extensive details on each attack, including the time and place of the attack, the weaponry employed, the number of casualties, and other information. The GTD, one of the most complete datasets on the subject, is frequently utilized in terrorist research. For a number of uses, including academic research, data analysis, and machine learning, this dataset is available on Kaggle.

Data Analysis

In this data set there were 27% null values, the total number of null values can be estimated by multiplying the total number of values in the dataset by 0.27. This data is valuable for evaluating the dataset's quality and identifying any concerns with missing data that could impair the accuracy of analysis or modeling.

We divided the data into major, minor, and minor attacks depending on the number of victims, and we used the following conditions:

Minor attacks: Those with less than 3 casualties.

Small attacks: Those with 3 to 10 casualties.

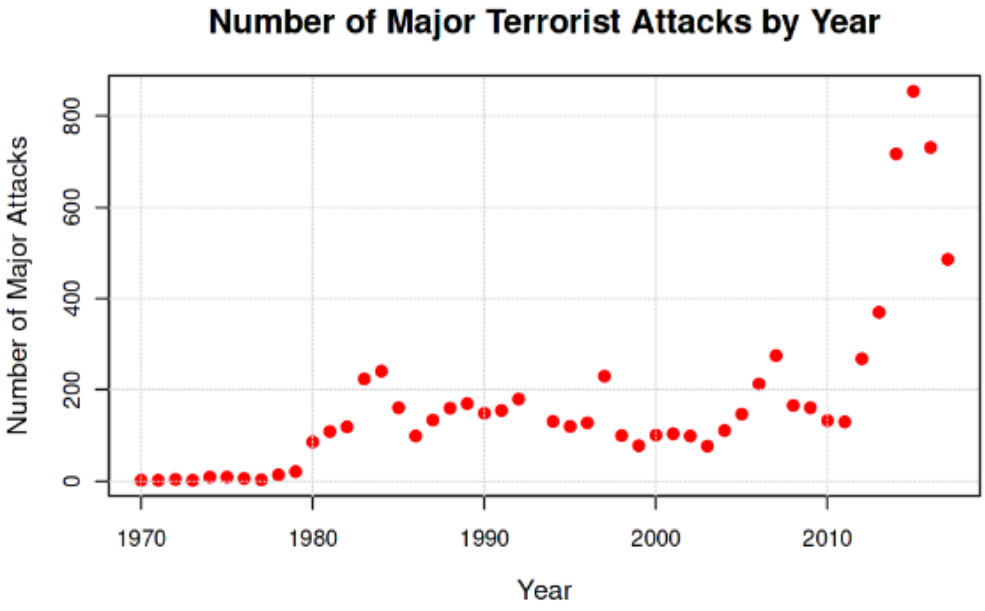
To calculate the number of attacks we have grouped the data by year and calculated the number of observations as each observation represents an attack.

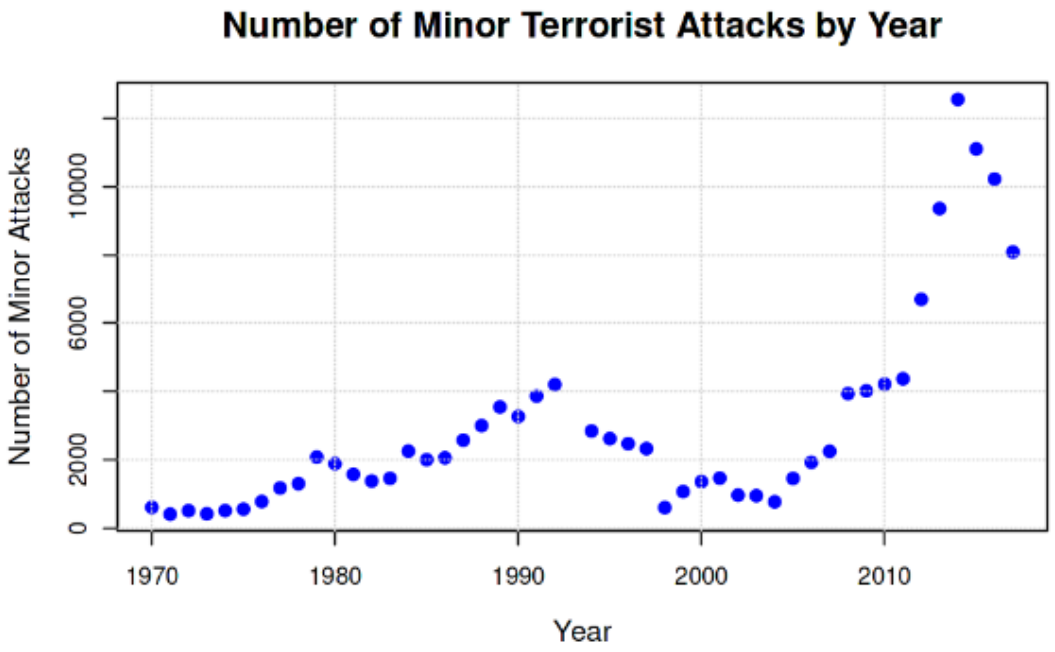
We have done the same for both major and minor attacks.

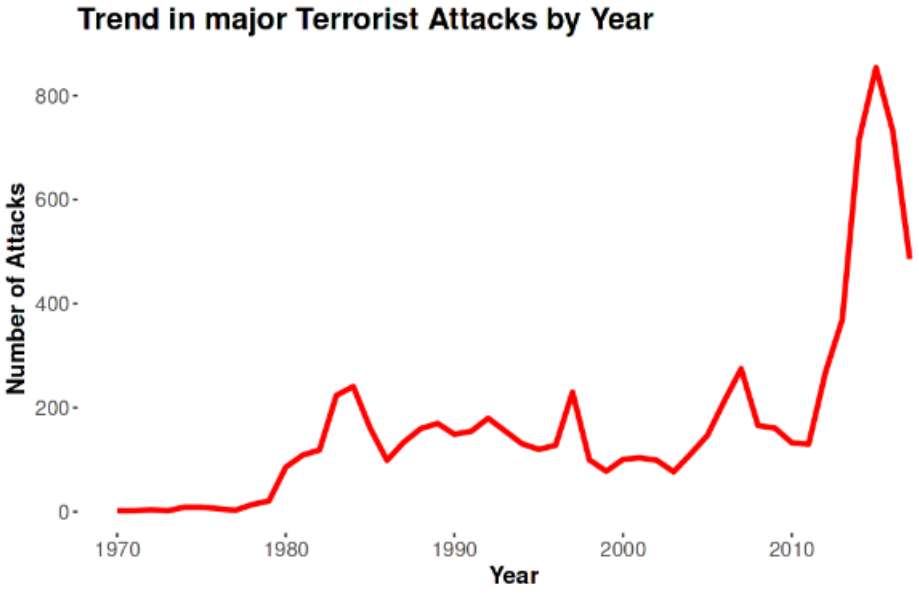
```
data_major <- train_major %>%
  group_by(iyear) %>%
  summarise(count = n())
data_minor <- train_minor %>%
    group_by(iyear) %>%
   summarise(count = n())
```

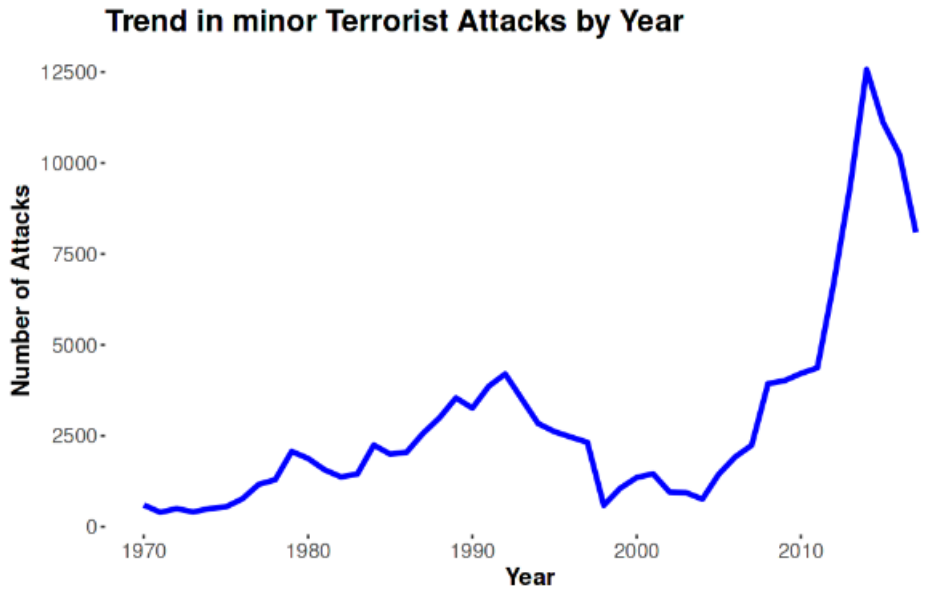
The datasets data_major and data_minor contain information on the number of major and minor attacks, respectively, that occurred in each year. Both datasets include the year as well as the corresponding number of major or minor attacks in that year. These datasets can be useful for analyzing trends and patterns in terrorist activity and identifying potential areas of concern for counterterrorism efforts.

By producing scatter plots for both major and minor attacks the results are as follows




Number of Minor Terrorist Attacks by Year
Number of Minor Attacks
0
2000
6000
10000
1970
1980
1990
2000
2010
Year


Trend in major Terrorist Attacks by Year
Number of Attacks
0
200
400
600
800
1970
1980
1990
2000
2010
Year



Based on the scatterplots, it appears that there has been an overall increasing trend in both the number of major and minor terrorist attacks over the years. In addition, there is a large spike in the number of attacks in the most recent years, which is a cause for concern.

For major attacks, the plot shows that there was a gradual increase in the number of attacks from the 1970s until the early 2000s, followed by a relatively stable period until around 2010. However, in the past few years, there has been a sharp increase in the number of major attacks, with a significant spike in 2014 and 2015.

Similarly, for minor attacks, the plot shows a steady increase in the number of attacks from the 1970s until the early 2000s, followed by a relatively stable period until around 2010. However, in recent years, there has been a sudden surge in the number of minor attacks, with a significant spike in the most recent years. This pattern is similar to that observed for major attacks, indicating a concerning trend in overall terrorist activity.

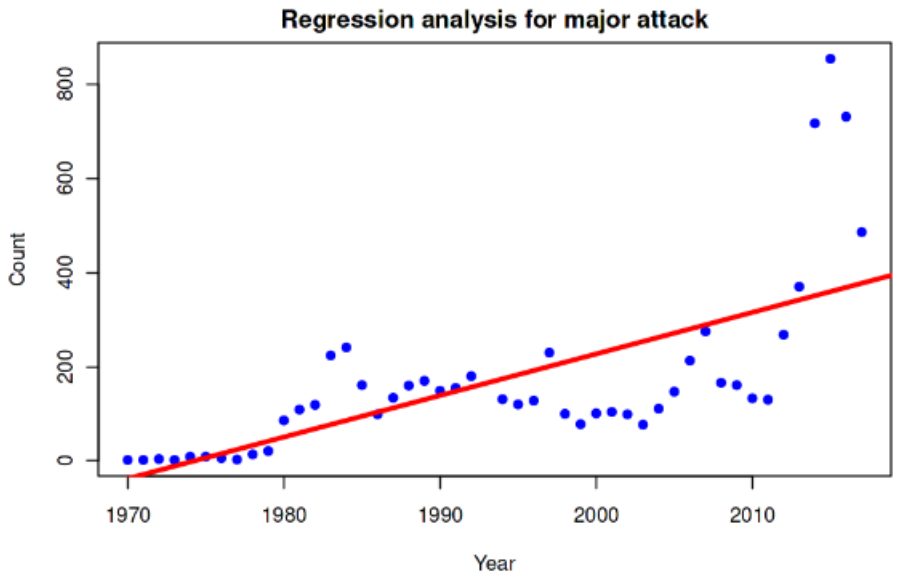
Regression analysis on Major Terrorist Attacks and year:

```
Call:
lm(formula = count ~ iyear, data = data_major)

Residuals:
    Min      1Q  Median      3Q     Max
-194.34  -88.40    2.39   36.96  494.36

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -17424.454   2893.416  -6.022 2.90e-07 ***
iyear            8.826      1.451   6.081 2.37e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 139.3 on 45 degrees of freedom
Multiple R-squared:  0.4511,    Adjusted R-squared:  0.4389
F-statistic: 36.98 on 1 and 45 DF,  p-value: 2.367e-07
```



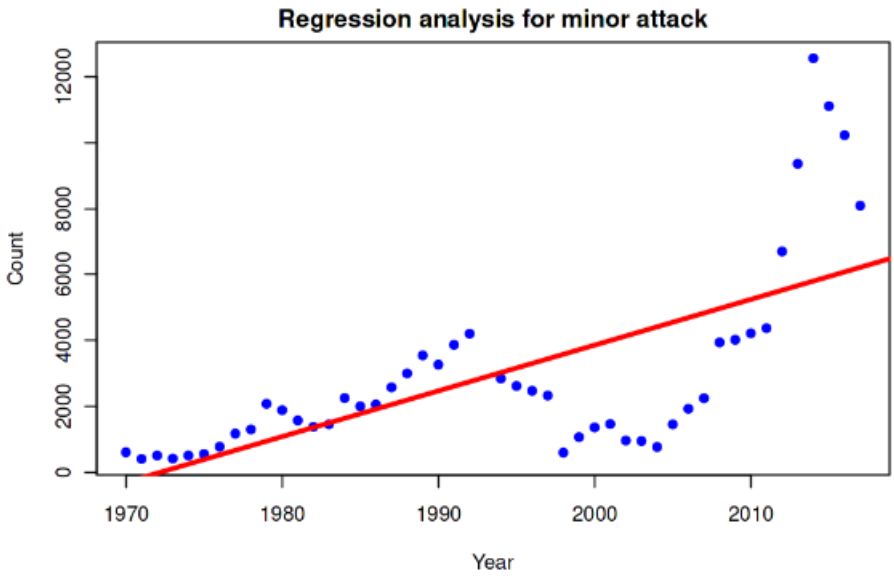
Regression analysis for minor terrorist attack and year:

```
Call:
lm(formula = count ~ iyear, data = data_minor)

Residuals:
    Min      1Q  Median      3Q     Max
-3650.0 -1060.2   248.1   805.1  6765.6

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -274009.3    45249.2  -6.056 2.58e-07 ***
iyear           138.9       22.7   6.121 2.06e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2178 on 45 degrees of freedom
Multiple R-squared:  0.4543,    Adjusted R-squared:  0.4422
F-statistic: 37.47 on 1 and 45 DF,  p-value: 2.064e-07
```



Code:

```
train <- read.csv('/kaggle/input/gtd/globalterrorismdb_0718dist.csv')

100 * sum(is.na(train)) / (nrow(train) * ncol(train))

str(train)

train_major <- subset(train,nkill>10)
```

```
train_minor <- subset(train,nkill<3)
train_small <- subset(train,nkill >= 3 & nkill <= 10)
data_major <- train_major %>%
  group_by(iyear) %>%
  summarise(count = n())
data_minor <- train_minor %>%
    group_by(iyear) %>%
   summarise(count = n())

par(mar = c(5, 5, 4, 2) + 0.1, cex.lab = 1.2, cex.main = 1.5)
options(repr.plot.width=8, repr.plot.height=5)

# Create the scatterplot with labels and titles
plot(data_major$iyear, data_major$count,
     main = "Number of Major Terrorist Attacks by Year",
     xlab = "Year", ylab = "Number of Major Attacks",
     col = "red", pch = 16, cex = 1.2)

# Add gridlines to the plot
grid()
par(mar = c(5, 5, 4, 2) + 0.1, cex.lab = 1.2, cex.main = 1.5)
options(repr.plot.width=8, repr.plot.height=5)

# Create the scatterplot with labels and titles
plot(data_minor$iyear, data_minor$count,
     main = "Number of Minor Terrorist Attacks by Year",
     xlab = "Year", ylab = "Number of Minor Attacks",
     col = "blue", pch = 16, cex = 1.2)

# Add gridlines to the plot
grid()
ggplot(data_major, aes(x = iyear, y = count)) +
  geom_line(color = "red", size = 1.2) +
  labs(title = "Trend in major Terrorist Attacks by Year", x = "Year", y = "Number of
Attacks") +
  theme_bw() +
```

```
  theme(plot.title = element_text(face = "bold", size = 18),
        axis.title = element_text(face = "bold", size = 14),
        axis.text = element_text(size = 12),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())


ggplot(data_minor, aes(x = iyear, y = count)) +
  geom_line(color = "blue", size = 1.2) +
  labs(title = "Trend in minor Terrorist Attacks by Year", x = "Year", y = "Number of
Attacks") +
  theme_bw() +
  theme(plot.title = element_text(face = "bold", size = 18),
        axis.title = element_text(face = "bold", size = 14),
        axis.text = element_text(size = 12),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank())


model_major <- lm(count ~ iyear, data = data_major)
summary(model_major)
par(mar = c(5, 5, 2, 2)) # Adjust margins
plot(data_major$iyear, data_major$count,
     xlab = "Year", ylab = "Count",
     main = "Regression analysis for major attack", pch = 16, col = "blue")


# Add regression line
abline(lm(data_major$count ~ data_major$iyear), lwd = 3, col = "red")


model_minor <- lm(count ~ iyear , data = data_minor)
summary(model_minor)
par(mar = c(5, 5, 2, 2)) # Adjust margins
plot(data_minor$iyear, data_minor$count,
     xlab = "Year", ylab = "Count",
```

```
     main = "Regression analysis for minor attack", pch = 16, col = "blue")

# Add regression line
abline(lm(data_minor$count ~ data_minor$iyear), lwd = 3, col = "red")
```