SAN FRANCISCO POLICE DEPARTMENT INCIDENT ANALYSIS

Final Project

MARCH 23, 2024

IST 687

Peter Henry, Noel Ford, Kevin Hansen

Contents

Business Introduction 2

Code Descriptions: 2

Data Cleansing and Trend Analysis 2

Analyzing Overall Crime and Violent Crime per District. 4

Analysis of Overall Days and Times in Districts 5

Conclusion 7

Appendix 8

Business Introduction

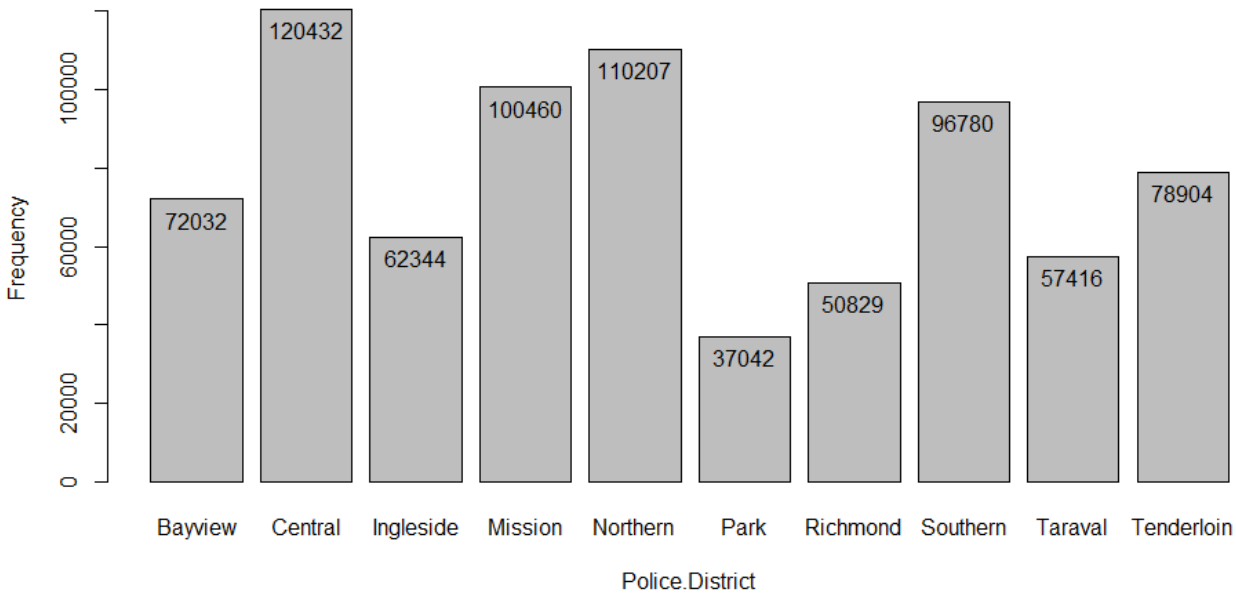
The San Francisco police department has more funds that they are using to hire five new officers to join their police force. They asked us to analyze their incident report data from the entire San Francisco area from 2018 to present day to decide which police district would most benefit from more officers.

Code Descriptions:

Data Cleansing and Trend Analysis

First, we cleaned the dataset and kept only columns that give relevant information about the incidents such as the date, time, location, incident category, and police district. We also removed anything that did not take place in the San Francisco Police Departments jurisdiction and that happened in 2024, since this is not a full year of data.

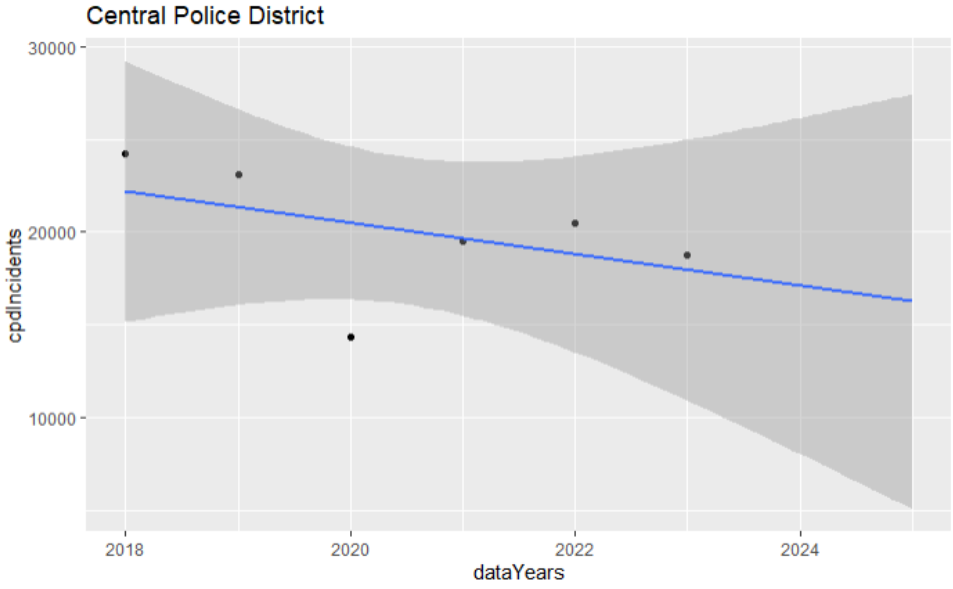
Using this new data frame, we created a bar graph showing the number of total incidences per police district. This gave us a better idea of what police districts to focus on since the districts with the higher incident rate will need more officers. The districts with the most incidents were Central, Northern, Mission, and Southern in that order. These districts had more total incidents than the other districts by almost 20,000.

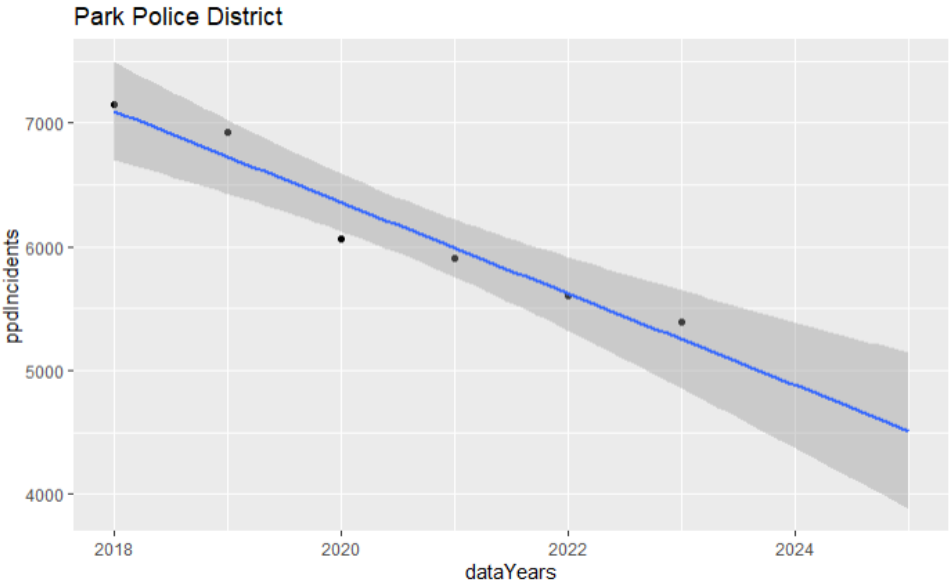


Next, we wanted to be able to figure out if each police district's crime was going to increase or decrease going forward into 2024 and 2025. This was done by breaking down the data into separate data

frames based on year and police districts. Then we created a data frame of vectors from these broken-down data frames.

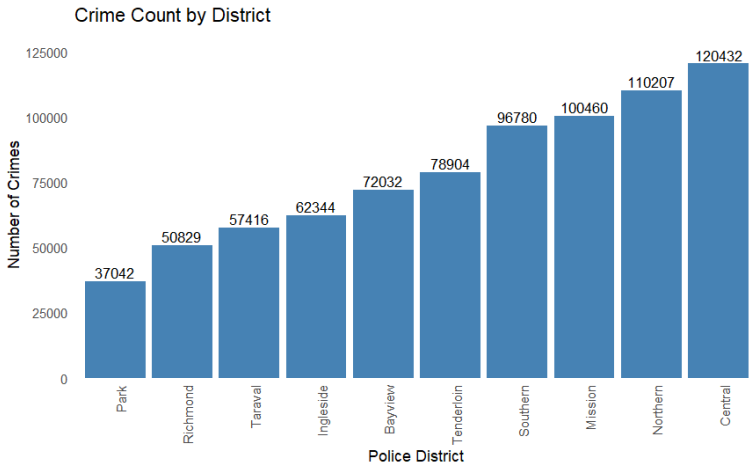
These were used with ggplot to create a linear model and extrapolate the data to form a trendline for each police district. This was helpful and did give us a general idea of the trend going into the next couple of years, but since there are so few data points, since we only have data for 6 years, the minimum and maximum of the extrapolation for most of the plots are not close. For example, Central according to this method could potentially have more than 26,000 incidents in 2025 or as low as 5,000. To make a more correct prediction we would need more historical data so that the data points looked more like Park District, which gives us a much better idea of what is going to occur in the future in that district. Finally, COVID data is playing a big role in trend analysis. Almost all the districts had less overall crime in the year of 2020. Central District is a great example, 2020 drops below 15,000 incidents, but then jumps right back up in 2021 and 2022. We need more historical data in order for COVID to not create such heavily weighted data points.



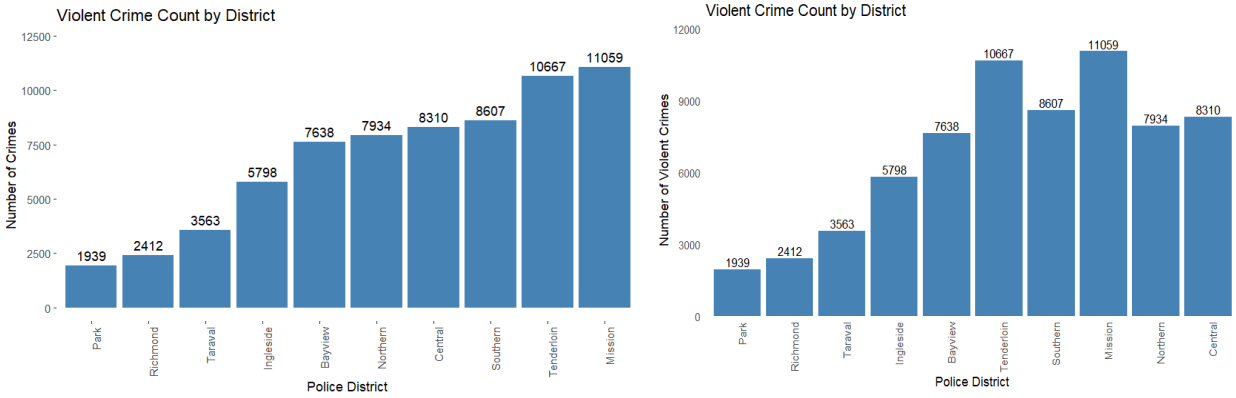


Analyzing Overall Crime and Violent Crime per District.

We decided that looking further into the district crime rate would be an ideal way to see where the 5 additional police officers could be located. Using the new data frame, we remade the bar graph depicting overall crime per district. This time we reordered the graph in ascending order to allow ease of viewing. Again, 2024 data has been removed as well as any data related to outside San Franciso. The bar graph was made using ggplot. Below is a view of the bar graph.



Violent crime is an area that may need more attention. To elaborate, stopping violent crimes could be considered more important than stopping petty crimes such as purse snatching. We decided to break up the crime data into associated violent crimes. The data frame was filtered on the following violent Crimes: Assault; Homicide; Human Trafficking (A) Commercial Sex Acts; Human Trafficking (B) Involuntary Servitude; Human Trafficking; Commercial Sex Acts; Rape; Robbery; and Suicide. Below are the referenced bar graphs.

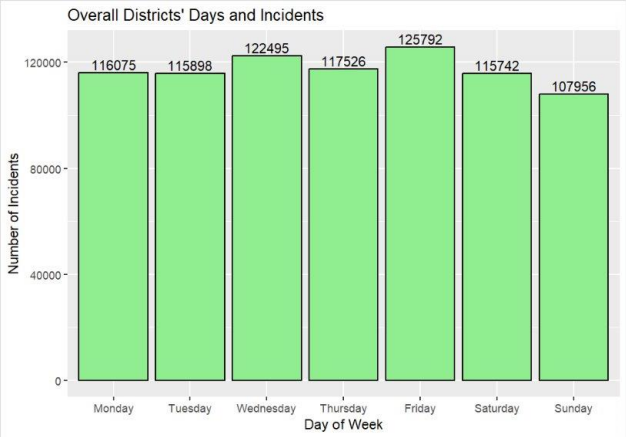


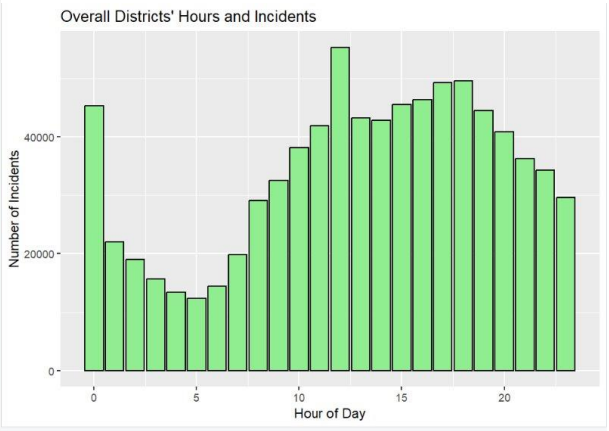
First, the left bar graph represents violent crime per district in ascending order. Second, the right bar graph keeps the rank for overall crime per district but depicts violent crime. Both bar graphs were made using ggplot. As seen in the right graph, Central District is ranked #1 for overall crime but #4 in

violent crime in the left graph. 5 police officers could be allocated to topmost violent districts to try and combat violent crime. These districts are as follows: Mission; Tenderloin; Southern; Central; and Northern.

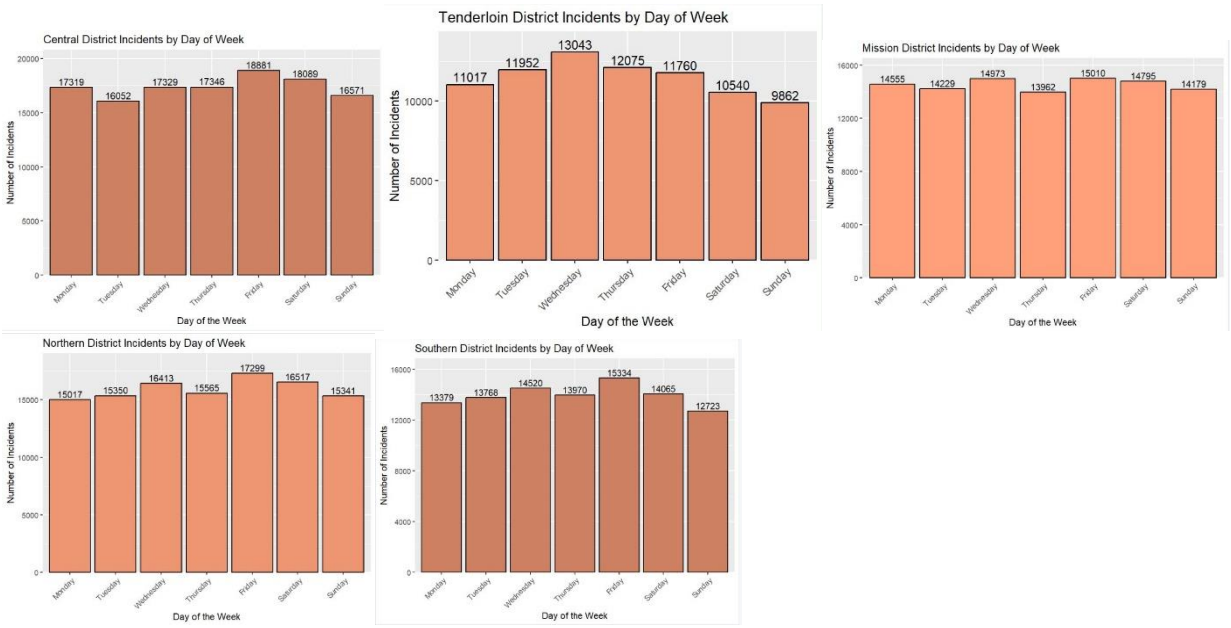
Analysis of Overall Days and Times in Districts

After having an idea of where we wanted to send these officers, we evaluated when we wanted these officers on duty. First, an analysis was made about all the districts, evaluating the sum of days and hours of when most incidents occurred. We noticed some interesting results in the overall data such as the top days were Friday and Wednesday, which were slightly different than what we had originally hypothesized, which was that Saturday and Sunday would be most active. We were even more surprised at the overall hours because 12:00pm was unmistakably higher in incidents than any other time.

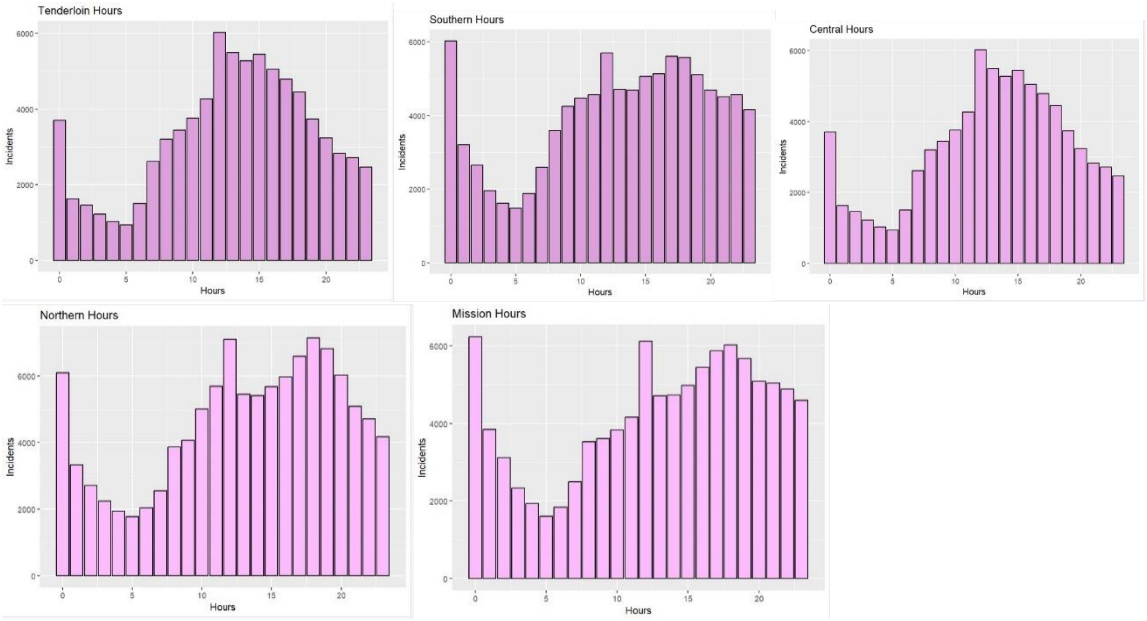




Since the analysis of crimes, violent crime, and crime rate happened within the same districts as noted in the Analysis of Crime section, we looked at the top 5 districts to see if this trend is related to the overall districts and to compare them to one another. According to the data, the days mostly correlated with the overall districts analysis with Friday and Wednesday being most active, and in some cases, Saturday was also a more active day.



Hours, too, reflected similar results to the overall incident time where 12pm (noon) was one of the most active times across all 5 districts with similar trends of activities and patterns in the afternoon and evening hours. Surprisingly, most incidents began waning around 8pm with a huge spike at 12am (midnight), especially in the Southern and Mission districts, but again dropping off until morning.



Further analysis would need to be done on specific district days and hours of incident activity to create a specific schedule for those officers. It should also be noted that knowing the number of officers already posted in the various districts would also be useful, so districts in need would be getting the necessary help on the necessary days and hours. With that said, after analyzing the trends we see in the data, we

would hypothesize and recommend that Friday, Wednesday, and Saturday between 12pm and 12am would be the most useful for officers to work as these are the most active hours and days of the week.

Conclusion

Following our analyses, our recommendation based off the San Francisco Crime Data is for the police department to staff 5 officers on Wednesday, Friday, and Saturday in 12-hour shifts from 12pm to 12am in the following districts: Central; Tenderloin; Mission; Northern; and Southern. This will enable maximum utilization of the officers in the most crime ridden districts following the spike in crime rate at 12pm. In addition, it would address the overall increase in crime rate in the evening hours. Violent Crime took priority in our data and is why we created a separate analysis for the said crimes. Through our analysis, we found that the most violent districts are typically in the most crime ridden districts. With this result, this allowed us to focus on our time related data for when officers should be staffed.

Given that there are only 5 years of crime data, we are not able to predict a trend in crime rate with accuracy. To further support, we recommend the San Francisco Police department release at minimum 15 years of crime data; however, 30 years of data would be more sufficient. With this additional information, we would be able to spot seasonal and yearly trends. In addition, COVID data would not create such a heavily weighted factor in our trend analyses. Finally, the data would provide further insight if there is an uptick or downtick in crime rate based on the current staffing with officers.

In conclusion, our data shows that 5 officers should be staffed in the above selected districts during from 12pm to 12am on Wednesday, Friday, and Saturday. However, additional years of crime data would increase our confidence in the data accuracy. Following our recommendation, overall crime rate is likely to decrease in the selected districts as well as facilitate crime prevention to create a safer community.

Appendix

Final Project

#Kevin Hansen, Peter Henry, Noel Ford

```{r}
# Reading in the CSV file as a data frame
library(tidyverse)
library(readr)
df <-
read_csv("C:\\Users\\kchan\\OneDrive\\Desktop\\Police_Department_Incident_Reports__2018_to_Present_20240208.csv")
```

```{r}
#Creating a subset of the original data frame with only relevant data
incidentFrame <- df[c("Incident Date", "Incident Time", "Incident Year", "Incident Day of Week",
                      "Incident ID", "Incident Number", "Report Type Code", "Report Type Description", "Incident Code",
                      "Incident Category", "Incident Subcategory", "Incident Description", "Police District")]
```

```{r}
#Table of the police districts
table(incidentFrame$`Police District`)
#Remove Incidences that were not in SF
incidentFrame2 <- subset(incidentFrame, !(incidentFrame$`Police District` %in% c("Out of SF")))
#Confirm Removal of Out of SF
table(incidentFrame2$`Police District`)
```

```{r}
# Remove Incident code, report type/description, ID, subcategory, and description.
incidentFrame2 <- incidentFrame2[,-5:-9]
incidentFrame2 <- incidentFrame2[,-6:-7]
glimpse(incidentFrame2)
# Remove Data from 2024 because incomplete data for the year
incidentFrame2 <- incidentFrame2[incidentFrame2$`Incident Year` < 2024,]
# Table showing the Number of incidences per incident category
table(incidentFrame2$`Incident Category`)
```
```{r}
# Break up data set based on police district
bayviewDF <- incidentFrame2[incidentFrame2$"Police District" == "Bayview",]
centralDF <- incidentFrame2[incidentFrame2$"Police District" == "Central",]
inglesideDF <- incidentFrame2[incidentFrame2$"Police District" == "Ingleside",]
missionDF <- incidentFrame2[incidentFrame2$"Police District" == "Mission",]
northernDF <- incidentFrame2[incidentFrame2$"Police District" == "Northern",]
parkDF <- incidentFrame2[incidentFrame2$"Police District" == "Park",]
RichmondDF <- incidentFrame2[incidentFrame2$"Police District" == "Richmond",]
southernDF <- incidentFrame2[incidentFrame2$"Police District" == "Southern",]
TaravalDF <- incidentFrame2[incidentFrame2$"Police District" == "Taraval",]
tenderloinDF <- incidentFrame2[incidentFrame2$"Police District" == "Tenderloin",]

```

```{r}
#Break up data set based on year
year_2018 <- incidentFrame2[incidentFrame2$"Incident Year" == 2018,]
year_2019 <- incidentFrame2[incidentFrame2$"Incident Year" == 2019,]
year_2020 <- incidentFrame2[incidentFrame2$"Incident Year" == 2020,]
year_2021 <- incidentFrame2[incidentFrame2$"Incident Year" == 2021,]
year_2022 <- incidentFrame2[incidentFrame2$"Incident Year" == 2022,]
year_2023 <- incidentFrame2[incidentFrame2$"Incident Year" == 2023,]
```

```{r}
library(Rcmdr)
#Bar Graph of Number of Incidents per Police District
with(incidentFrame2, Barplot(incidentFrame2$`Police District`, xlab="Police District", ylab="Frequency",
label.bars=TRUE))

# Bar Graphs of Incident cases per Police District per category
with(incidentFrame2, Barplot(incidentFrame2$`Incident Category`, by=incidentFrame2$`Police District`,
  style="divided", legend.pos="above", xlab='Incident Category',
  ylab="Frequency", label.bars=FALSE))

# Bar Graphs of Incident cases per Police District per category
with(incidentFrame2, Barplot(incidentFrame2$`Police District`, by=incidentFrame2$`Incident Category`,
  style="divided", legend.pos="above", xlab='Police District',
  ylab="Frequency", label.bars=FALSE))

```

```{r}
# Central PD Forecast for number of crimes going forward

# Years
dataYears <- c(2018, 2019, 2020, 2021, 2022, 2023)

# Number of Incidents per Year
cpdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Central",]),
                  nrow(year_2019[year_2019$`Police District` == "Central",]),
                  nrow(year_2020[year_2020$`Police District` == "Central",]),
                  nrow(year_2021[year_2021$`Police District` == "Central",]),
                  nrow(year_2022[year_2022$`Police District` == "Central",]),
                  nrow(year_2023[year_2023$`Police District` == "Central",]))
# Create data frame of vectors
cpdDF <- data.frame(dataYears, cpdIncidents)
# Create linear model and extrapolate
ggplot(cpdDF, aes(x=dataYears, y=cpdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T) +
  xlim(2018, 2025) + ggtitle("Central Police District")

```
```{r}
# Mission PD Forecast for number of crimes going forward

# Number of Incidents per Year
mpdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Mission",]),
                  nrow(year_2019[year_2019$`Police District` == "Mission",]),
                  nrow(year_2020[year_2020$`Police District` == "Mission",]),
                  nrow(year_2021[year_2021$`Police District` == "Mission",]),
                  nrow(year_2022[year_2022$`Police District` == "Mission",]),
```

```
            nrow(year_2023[year_2023$`Police District` == "Mission",]))
# Create data frame of vectors
mpdDF <- data.frame(dataYears, mpdIncidents)
# Create linear model and extrapolate
ggplot(mpdDF, aes(x=dataYears, y=mpdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T) +
  xlim(2018, 2025) + ggtitle("Mission Police District")
```
```{r}
# Northern PD Forecast for number of crimes going forward

# Number of Incidents per Year
npdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Northern",]),
            nrow(year_2019[year_2019$`Police District` == "Northern",]),
            nrow(year_2020[year_2020$`Police District` == "Northern",]),
            nrow(year_2021[year_2021$`Police District` == "Northern",]),
            nrow(year_2022[year_2022$`Police District` == "Northern",]),
            nrow(year_2023[year_2023$`Police District` == "Northern",]))
# Create data frame of vectors
npdDF <- data.frame(dataYears, npdIncidents)
# Create linear model and extrapolate
ggplot(npdDF, aes(x=dataYears, y=npdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T) +
  xlim(2018, 2025) + ggtitle("Northern Police District")
```
```{r}
# Southern PD Forecast for number of crimes going forward

# Number of Incidents per Year
spdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Southern",]),
            nrow(year_2019[year_2019$`Police District` == "Southern",]),
            nrow(year_2020[year_2020$`Police District` == "Southern",]),
            nrow(year_2021[year_2021$`Police District` == "Southern",]),
            nrow(year_2022[year_2022$`Police District` == "Southern",]),
            nrow(year_2023[year_2023$`Police District` == "Southern",]))
# Create data frame of vectors
spdDF <- data.frame(dataYears, spdIncidents)
# Create linear model and extrapolate
ggplot(spdDF, aes(x=dataYears, y=spdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T)+
  xlim(2018, 2025) + ggtitle("Southern Police District")
```
```{r}
# Bayview PD Forecast for number of crimes going forward

# Number of Incidents per Year
bpdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Bayview",]),
            nrow(year_2019[year_2019$`Police District` == "Bayview",]),
            nrow(year_2020[year_2020$`Police District` == "Bayview",]),
            nrow(year_2021[year_2021$`Police District` == "Bayview",]),
            nrow(year_2022[year_2022$`Police District` == "Bayview",]),
            nrow(year_2023[year_2023$`Police District` == "Bayview",]))
# Create data frame of vectors
bpdDF <- data.frame(dataYears, bpdIncidents)
# Create linear model and extrapolate
ggplot(bpdDF, aes(x=dataYears, y=bpdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T) +
xlim(2018, 2025) + ggtitle("Bayview Police District")
```
```{r}
# Ingleside PD Forecast for number of crimes going forward
```

```
# Number of Incidents per Year
ipdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Ingleside",]),
                  nrow(year_2019[year_2019$`Police District` == "Ingleside",]),
                  nrow(year_2020[year_2020$`Police District` == "Ingleside",]),
                  nrow(year_2021[year_2021$`Police District` == "Ingleside",]),
                  nrow(year_2022[year_2022$`Police District` == "Ingleside",]),
                  nrow(year_2023[year_2023$`Police District` == "Ingleside",]))
# Create data frame of vectors
ipdDF <- data.frame(dataYears, ipdIncidents)
# Create linear model and extrapolate
ggplot(ipdDF, aes(x=dataYears, y=ipdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T) +
  xlim(2018, 2025) + ggtitle("Ingleside Police District")
```
```{r}
# Park PD Forecast for number of crimes going forward

# Number of Incidents per Year
ppdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Park",]),
                  nrow(year_2019[year_2019$`Police District` == "Park",]),
                  nrow(year_2020[year_2020$`Police District` == "Park",]),
                  nrow(year_2021[year_2021$`Police District` == "Park",]),
                  nrow(year_2022[year_2022$`Police District` == "Park",]),
                  nrow(year_2023[year_2023$`Police District` == "Park",]))
# Create data frame of vectors
ppdDF <- data.frame(dataYears, ppdIncidents)
# Create linear model and extrapolate
ggplot(ppdDF, aes(x=dataYears, y=ppdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T) +
  xlim(2018, 2025) + ggtitle("Park Police District")
```
```{r}
# Richmond PD Forecast for number of crimes going forward

# Number of Incidents per Year
rpdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Richmond",]),
                  nrow(year_2019[year_2019$`Police District` == "Richmond",]),
                  nrow(year_2020[year_2020$`Police District` == "Richmond",]),
                  nrow(year_2021[year_2021$`Police District` == "Richmond",]),
                  nrow(year_2022[year_2022$`Police District` == "Richmond",]),
                  nrow(year_2023[year_2023$`Police District` == "Richmond",]))
# Create data frame of vectors
rpdDF <- data.frame(dataYears, rpdIncidents)
# Create linear model and extrapolate
ggplot(rpdDF, aes(x=dataYears, y=rpdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T) +
  xlim(2018, 2025) + ggtitle("Richmond Police District")
```
```{r}
# Taraval PD Forecast for number of crimes going forward

# Number of Incidents per Year
tarapdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Taraval",]),
                  nrow(year_2019[year_2019$`Police District` == "Taraval",]),
                  nrow(year_2020[year_2020$`Police District` == "Taraval",]),
                  nrow(year_2021[year_2021$`Police District` == "Taraval",]),
                  nrow(year_2022[year_2022$`Police District` == "Taraval",]),
                  nrow(year_2023[year_2023$`Police District` == "Taraval",]))
# Create data frame of vectors
tarapdDF <- data.frame(dataYears, tarapdIncidents)
# Create linear model and extrapolate
```

```
ggplot(tarapdDF, aes(x=dataYears, y=tarapdIncidents)) + geom_point() + geom_smooth(method="lm", fullrange=T)
+
  xlim(2018, 2025) + ggtitle("Taraval Police District")
```
```{r}
# Tenderloin PD Forecast for number of crimes going forward

# Number of Incidents per Year
tendpdIncidents <- c(nrow(year_2018[year_2018$`Police District` == "Tenderloin",]),
                     nrow(year_2019[year_2019$`Police District` == "Tenderloin",]),
                     nrow(year_2020[year_2020$`Police District` == "Tenderloin",]),
                     nrow(year_2021[year_2021$`Police District` == "Tenderloin",]),
                     nrow(year_2022[year_2022$`Police District` == "Tenderloin",]),
                     nrow(year_2023[year_2023$`Police District` == "Tenderloin",]))
# Create data frame of vectors
tendpdDF <- data.frame(dataYears, tendpdIncidents)
# Create linear model and extrapolate
ggplot(tendpdDF, aes(x=dataYears, y=tendpdIncidents)) + geom_point() + geom_smooth(method="lm",
fullrange=T) +
  xlim(2018, 2025) + ggtitle("Tenderloin Police District")
```
```{r}
# Looking at the unique Police districts in SF.
police_districs <- unique(incidentFrame$`Police District`)
police_districs
```
```{r}
# Summing up the crime count for each district
crime_count <- incidentFrame %>% group_by(`Police District`) %>%
  summarise(count = n())
crime_count <- crime_count %>% arrange(count)
crime_count

crime_count <- crime_count %>%
  mutate(`Police District` = fct_reorder(`Police District`, count))
crime_count
```

```{r}
# Plotting the crime count to get a better picture
ggplot(crime_count, aes(x=`Police District`, y=count)) +
  geom_bar(stat="identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Police District", y = "Number of Crimes", title = "Crime Count by District")

# This plot presents the overall crime per Police district in San Francisco.
```
```{r}
# Looking at violent crimes and see what districts they take place in
violent_crime <- unique(incidentFrame$`Incident Category`)
violent_crime
view(violent_crime)
```
```{r}
# Creating a DF with only Violent crimes.
violent_DF <- incidentFrame %>% filter(`Incident Category` %in% c("Assault", "Homicide", "Human Trafficking
(A), Commercial Sex Acts", "Human Trafficking (B), Involuntary Servitude", "Human Trafficking, Commercial Sex
Acts", "Rape", "Robbery","Suicide"))
```

```
violent_incident_counts <- violent_DF %>%
  group_by(`Police District`) %>%
  summarise(Count = n())

```
```{r}
#Creating a bar chart off of violent crimes
violent_incident_counts <- violent_incident_counts %>%
  mutate(`Police District` = fct_reorder(`Police District`, Count))

ggplot(violent_incident_counts, aes(x=`Police District`, y=Count)) +
  geom_bar(stat="identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Police District", y = "Number of Crimes", title = "Crime Count by District")
```

```{r}
# Plot for all crime data and plot for violent crime data

#All Crime data
ggplot(crime_count, aes(x=`Police District`, y=count)) +
  geom_bar(stat="identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Police District", y = "Number of Crimes", title = "Crime Count by District")


#Violent Crime Data
ggplot(violent_incident_counts, aes(x=`Police District`, y=Count)) +
  geom_bar(stat="identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Police District", y = "Number of Crimes", title = "Violent Crime Count by District")
```

```{r}
# I want the bar charts to match on the x label. Below is the code for that.
district_order <- crime_count$`Police District`

violent_incident_counts <- violent_incident_counts %>%
  mutate(`Police District` = factor(`Police District`, levels = district_order))

#All Crime data
ggplot(crime_count, aes(x=`Police District`, y=count)) +
  geom_bar(stat="identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Police District", y = "Number of Crimes", title = "Crime Count by District")

#Violent Crime Data
ggplot(violent_incident_counts, aes(x=`Police District`, y=Count)) +
  geom_bar(stat="identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Police District", y = "Number of Violent Crimes", title = "Violent Crime Count by District")

# The bar charts now follow the Police Districts with the most overall crime. You can see that Violent crimes do not follow the same order once you get to the higher crime rate areas. 'Tenderloin' and 'Mission' Districts stand out the most.
```
```{r}
```

```
library(arules)
library(tidyverse)
library(readr)
library(lubridate)
library(data.table)

#Create database with only days and times
timeDF <- df[ , 2:5]

```
```{r}
#Overall Hours / Days for All Districts Combined
# What day or days do incidents happen the most
days <- timeDF %>% count(`Incident Day of Week`, sort = TRUE)

#Graph days
days$`Incident Day of Week` <- factor(days$`Incident Day of Week`,
      levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))


daysPlot <- ggplot(data=days, aes(x=days$`Incident Day of Week`, y=days$n)) +
  geom_bar(stat="identity", color = "black", fill = "lightgreen") + labs(y= "Number of Incidents", x = "Day of
Week") +
  ggtitle("Overall Districts' Days and Incidents") +
  geom_text(aes(label = n), vjust = -0.3, color = "black")

daysPlot
```
```{r}
#What hour or hours do the incidents happen the most

timeDF$hour <- hour(timeDF$`Incident Time`)

hours <- timeDF %>% count(hour, sort = TRUE)

#Graph hours

hourPlot <- ggplot(data=hours, aes(x=hours$hour, y=hours$n)) +
  geom_bar(stat="identity", color = "black", fill = "lightgreen") +
  labs(y= "Number of Incidents", x = "Hour of Day") +
  ggtitle("Overall Districts' Hours and Incidents")

hourPlot
```


```{r}
#District Overall Days and Hours

#District Hours
districthdf <- df[ , c('Incident Time', 'Police District')]

districthdf$hour <- hour(districthdf$`Incident Time`)

dhours <- districthdf %>% count(hour, sort = TRUE)

#Days / Hours based on Top Five Most Incident Districts
districtdf <- df[, c('Incident Day of Week', 'Police District')]
```
```
```

```{r}
#Mission District Days
missionDistrict <- districtdf[districtdf$`Police District` == "Mission",]

missionDistrict$`Incident Day of Week` <- factor(missionDistrict$`Incident Day of Week`,
                                    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
"Sunday"))
missionCount <- missionDistrict %>% count(`Incident Day of Week`)

missionPlot <- ggplot(data = missionCount, aes(x = `Incident Day of Week`, y = n)) +
  geom_bar(stat = "identity", color = "black", fill = "lightsalmon") +
  geom_text(aes(label = n), vjust = -0.3, color = "black") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  ggtitle("Mission District Incidents by Day of Week") +
  labs(y = "Number of Incidents", x = "Day of the Week") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Mission District Hours

missionTwo <- districthdf[districthdf$`Police District` == "Mission",]

missionHours <- missionTwo %>% count(hour, sort=TRUE)

missionHPlot <- ggplot(data=missionHours, aes(x=missionHours$hour, y=n)) +
  geom_bar(stat="identity", color="black", fill="plum1") +
  ggtitle ("Mission Hours") + labs(y= "Incidents", x = "Hours")

missionHPlot
```
```{r}
#Tenderloin District Days
tenderloinDistrict <- districtdf[districtdf$`Police District` == "Tenderloin",]

tenderloinCount <- tenderloinDistrict %>% count(tenderloinDistrict$`Incident Day of Week`)

tenderloinDistrict$`Incident Day of Week` <- factor(tenderloinDistrict$`Incident Day of Week`,
                                    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
"Sunday"))

tenderloinPlot <- ggplot(data = tenderloinCount, aes(x = `tenderloinDistrict$\`Incident Day of Week\``, y = n)) +
  geom_bar(stat = "identity", color = "black", fill = "lightsalmon2") +
  geom_text(aes(label = n), vjust = -0.3, color = "black") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  ggtitle("Tenderloin District Incidents by Day of Week") +
  labs(y = "Number of Incidents", x = "Day of the Week") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Tenderloin District Hours
tenderloinTwo <- districthdf[districthdf$`Police District` == "Tenderloin",]

tenderloinHours <- tenderloinTwo %>% count(hour, sort=TRUE)

tenderloinHPlot <- ggplot(data=tenderloinHours, aes(x=tenderloinHours$hour, y=n)) +
  geom_bar(stat="identity", color="black", fill="plum") +
  ggtitle ("Tenderloin Hours") + labs(y= "Incidents", x = "Hours")
```
```{r}
#Central District Days
```

```
centralDistrict <- districtdf[districtdf$`Police District` == "Central",]

centralDistrict$`Incident Day of Week` <- factor(centralDistrict$`Incident Day of Week`,
                                        levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
"Sunday"))

centralCount <- centralDistrict %>% count(centralDistrict$`Incident Day of Week`)

centralPlot <- ggplot(data=centralCount, aes(x=`centralDistrict$\`Incident Day of Week\``, y=n)) +
  geom_bar(stat="identity", color="black", fill="lightsalmon3") +
  ggtitle ("Central Days") +
  geom_text(aes(label = n), vjust = -0.3, color = "black") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  ggtitle("Central District Incidents by Day of Week") +
  labs(y = "Number of Incidents", x = "Day of the Week") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Central District Hours

centralTwo <- districthdf[districthdf$`Police District` == "Central",]

centralHours <- tenderloinTwo %>% count(hour, sort=TRUE)

centralHPlot <- ggplot(data=centralHours, aes(x=centralHours$hour, y=n)) +
  geom_bar(stat="identity", color="black", fill="plum2") +
  ggtitle ("Central Hours") + labs(y= "Incidents", x = "Hours")
```
```{r}
#Northern District Days

northernDistrict <- districtdf[districtdf$`Police District` == "Northern",]

northernDistrict$`Incident Day of Week` <- factor(northernDistrict$`Incident Day of Week`,
                                        levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
"Sunday"))

northernCount <- northernDistrict %>% count(northernDistrict$`Incident Day of Week`)

northernPlot <- ggplot(data=northernCount, aes(x=`northernDistrict$\`Incident Day of Week\``, y=n)) +
  geom_bar(stat="identity", color="black", fill="lightsalmon2") +
  ggtitle ("Northern District") +
  geom_text(aes(label = n), vjust = -0.3, color = "black") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  ggtitle("Northern District Incidents by Day of Week") +
  labs(y = "Number of Incidents", x = "Day of the Week") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Northern District Hours

northernTwo <- districthdf[districthdf$`Police District` == "Northern",]

northernHours <- northernTwo %>% count(hour, sort=TRUE)

northernHPlot <- ggplot(data=northernHours, aes(x=northernHours$hour, y=n)) +
  geom_bar(stat="identity", color="black", fill="plum1") +
  ggtitle ("Northern Hours") + labs(y= "Incidents", x = "Hours")
```

```{r}
#Southern District Days

southernDistrict <- districtdf[districtdf$`Police District` == "Southern",]

southernDistrict$`Incident Day of Week` <- factor(southernDistrict$`Incident Day of Week`,
                                        levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday",
"Sunday"))

southernCount <- southernDistrict %>% count(southernDistrict$`Incident Day of Week`)

southernPlot <- ggplot(data=southernCount, aes(x=`southernDistrict$\`Incident Day of Week\``, y=n)) +
  geom_bar(stat="identity", color="black", fill="lightsalmon3") +
  ggtitle ("Southern District") +
  geom_text(aes(label = n), vjust = -0.3, color = "black") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  ggtitle("Southern District Incidents by Day of Week") +
  labs(y = "Number of Incidents", x = "Day of the Week") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Southern District Hours

southernTwo <- districthdf[districthdf$`Police District` == "Southern",]

southernHours <- southernTwo %>% count(hour, sort=TRUE)

southernHPlot <- ggplot(data=southernHours, aes(x=southernHours$hour, y=n)) +
  geom_bar(stat="identity", color="black", fill="plum") +
  ggtitle ("Southern Hours") + labs(y= "Incidents", x = "Hours")
```