Storm and Climate Data Record (SCDR)

Implementation Plan

Date of record:
Jan 1, 2017 to Dec 1, 2019

Principal Investigator: Daniel Palhano

I. Introduction

1.1 Overview and purpose

The dataset we are given is related to the Miami PD's crime statistics regarding crime activity during storms within the Miami Dade County area. This Storm and Crime Data Report (SCDR) is a set of data that gives us insight into all these events recorded within the Jan 1, 2017 and Dec 1, 2019 period. The main goal that we are looking for is looking at the insights of what are the most common crimes within this period the dataset is recorded within and to help prepare Miami PD to counteract the potential of future occurrences during storms from happening. This analysis will go over the main points of concern and problem areas the department may want to look into to prevent a crime spree from occurring during storm and harsh weather events.

1.2 Define why you need data analysis

The reason for having this analysis done on the SCDR is due to the fact that the department simply has had a problem with crime activity during harsh weather and storms. This analysis will give clarity to any problem areas and the type of weather along with the kinds of crimes that are more popular during this time period and how it may shift during the type of incoming weather and storms. Analysis not only will give us insights on the crime statistics in general, but also include filtering out specific data points not necessary for our analysis and involve specifying the parameters needed to get the desired result.

II. Data Preparations

2.1 Name data sources

Data sources for this kind of SCDR would range from not only crime reports from officers reporting to the scenes of crimes, but also the weather report for the day. The combination of the Miami PD's reports allows us to gain the information of crime, the date of the crime, the zone and city of the crime. This type of data allows us to see the severity of the type of crime and compare it alongside the storm event and activity and see problem areas easily throughout the dataset. The weather report for the necessary days from either sources such as the NOAA, NWS, or even local reports will give us the storm and weather data to identify the severity of the storm at hand.

Another source of raw data to aid in the investigation is other crime reports within metropolitan Florida areas such as Fort Lauderdale, Tampa, and Orlando to also give us a picture of the kinds of crimes to compare our current report with. Incorporating other SCDR or reports similar to the one being done by us for the Miami PD in other regions similar to Miami Dade will help us compare to what types of weather events are most common for crime and concern to the department.

2.2 Filter through unnecessary data

Filtering through the SCDR allows us to take away unnecessary data that is not needed or necessary for our analysis. We see a possible couple of columns that would not be used such as Zone, CrimeEventID, and StormEventID. The reason for not using Zone is fairly self explanatory since it simply just reiterates the county in which the crimes are taking place, since the SCDR is all crimes reported within Miami Dade only, this column is not needed. CrimeEventID is also not needed since all the IDs are the exact same and does not have an individualized number to identify the specific crime that was reported, which then means it will not be used or necessary for analysis. Lastly, StormEventID is also unused for our analysis since it does not point to

anything. Looking at reports from the same date with the same weather event does not have a matching ID, which means they are not IDs identifying the storm or weather type. If it is to be pointing to the event recorded along with the storm type, it should take the place of the CrimeEventID and combine its purpose of identifying the storm type and crime type in 1 singular ID column. Despite this, the ID does not identify anything for our analysis that would be necessary and thus get filtered out.

2.3 Define your parameters

Our parameters for analysis are guided by the goals of what the analysis is to give insight to Miami PD to make changes on how they approach crime based on weather type. For analysis, we can now assume our parameters will involve the dates of our incident, crime activity and what occurred on the date, storm activity, and lastly the city that said crime occurred within. Our dates will give us a window into what periods are most active for crime within the county, running this alongside the storm activity during the time periods will help give the PD storm patterns or trends to potentially look into. The city will allow us to see problem areas which could be more active to pay more attention to when forecasts trend towards a bigger storm. Finally, our crime activity will tell us the most common crimes and what to look out for during these periods of storm activity during potential times of the year which could see upticks in crime.

2.4 Identify measurement priorities

The priority for measurement for this analysis is finding strong correlations between crime types, storm severity, and city the crimes are occurring within. While correlation does not always mean causation, it will ultimately give the Miami PD what they are looking for when it comes to figuring out whether or not there is a surge in crime during specific storm activity. Another measurement to look out for is contextually and looking into how the type of storm and city affects the severity of crime. Looking into the storm type along with cities will tell us the at risk areas which will potentially see an uptick in crime due to poor weather conditions or storm forecasts. Looking at this in a contextual way will allow for us to differentiate the shifts in crime depending on storm types and looking at it objectively instead.

2.5 Ensure collected data fits the need

The data given within the SCDR fits the basic necessities for analysis and will allow us to further analyze to give a response to the Miami PD in what kind of changes could be made to circumvent the issue of crime stopping depending on storm type. The collected data stays on topic with the goals of what we are trying to achieve and is not bloated with filler data that would cause poor quality data. We have quality columns and bits of data accurately recorded for further analysis.

III. Data Analysis

3.1 Identify scripts used

For our analysis, we took the crime data we were given identifying crime data identifying victim losses between the periods to figure out how losses compared from when there are storms to when there is no storm. The scripts we used involved having to use the second method, the work around, compared to the original method due to packages not installing correctly from the first method.

install.packages("ggplot2");

```
install.packages("dplyr");
library(ggplot2)
library(dplyr)


crimestormdataQ <- read.csv("crimeStormQ.csv")
crimenostormdataQ <- read.csv("crimenostormQ.csv")

crimestormdataQ$Date    <-    seq(as.Date("2017-01-01"),    by    =    "month",    length.out    =
nrow(crimestormdataQ))
crimenostormdataQ$Date    <-    seq(as.Date("2017-01-01"),    by    =    "month",    length.out    =
nrow(crimenostormdataQ))

crimestormdataQ$StormStatus <- "Storm"
crimenostormdataQ$StormStatus <- "No Storm"
```

> Looking at our first couple of scripts after installing the necessary packages and loading the libraries, we imported the data from the necessary CSVs which will allow us to compare the storm vs no storm data. This will allow us to see the correlation of how much more the losses of the victims are when there is a storm to no storm. After importing the data, we assigned the data in a monthly sequence due to the data's nature. Since we are doing this analysis across a 2 year period, assigning it monthly will give us a better visual compared to if we were to do so with any other period breakdown. The last bit of this set of scripts labels our storm and non storm data for our plot.

```
combined_data <- bind_rows(
  crimestormdataQ[, c("Date", "Loss", "StormStatus")],
  crimenostormdataQ[, c("Date", "Loss", "StormStatus")])

combined_data <- combined_data %>%
  group_by(StormStatus) %>%
  arrange(Date) %>%
  mutate(CumLoss = cumsum(Loss) / 1000)
```

> This set of scripts consolidates our data for easier visualization and our second by doing our calculation to further display it on our visual. The calculation is what provides us with the lines to see how much more victims lost when there was a storm compared to no storm.

```
ggplot(combined_data, aes(x = Date, y = CumLoss, color = StormStatus)) +
  geom_line(size = 1.2) +
  labs(
    title = "Victim Loss From Crimes for Jan 2017 - Dec 2019",
    subtitle = "Cumulative Loss in Thousands of Dollars",
    x = "By Month by Year",
    y = "Victim Loss (K$)",
    color = "Crime Condition"
  ) +
  theme_minimal()
```

The last lines of our script is our visual which provides us with the insight on how much more victims of our crimes were lost when there was a storm compared to no storm or weather conditions.

3.2 Run the scripts to analyze the data and validate the output

Victim Loss From Crimes for Jan 2017 - Dec 2019
Cumulative Loss in Thousands of Dollars
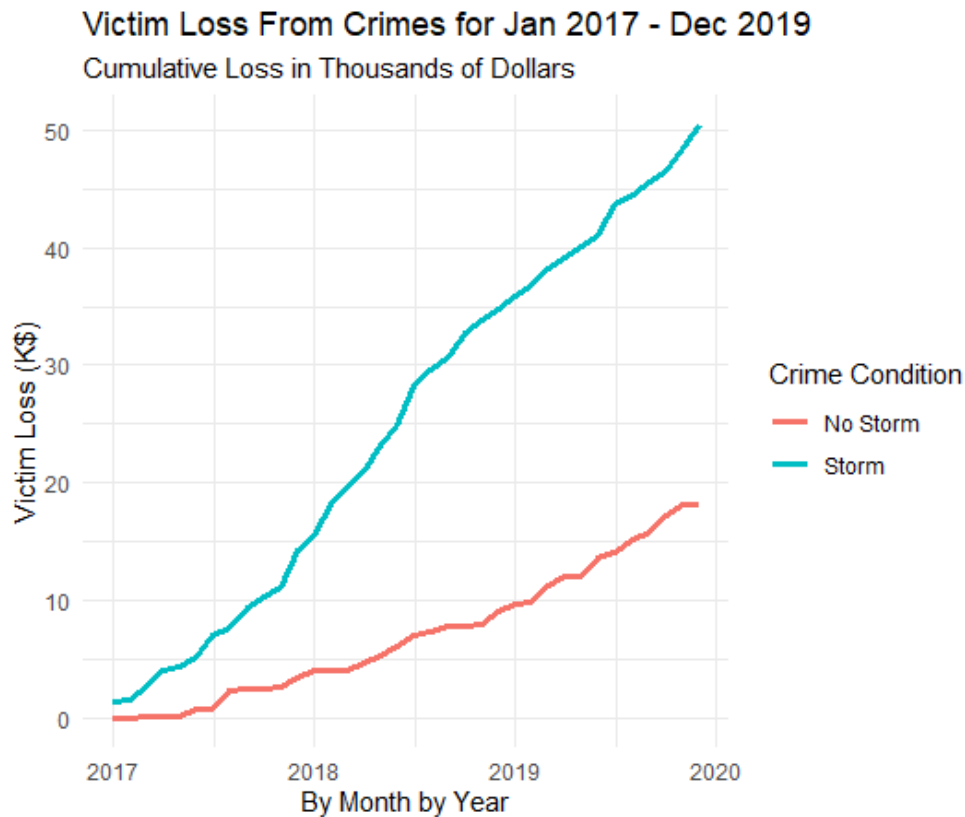
IV. Drawing Conclusions

4.1 Present the results of the analysis to stakeholders

From our analysis we see that roughly victims from crimes when there were no storms lost considerably less than those who were victims when a storm was active. We see victims from crimes during storms lose a little over $50,000 compared to the non-storm victims which lose a little less than $20,000. We could see a large gap widen especially between the 2018 year period indicating a potential uptick in harsh weather the region experienced and how much this could have affected the region. We see that during that period in specific, losses for storm victims climbed roughly a whole $20,000 which indicated an increase in losses and potential increase in individual incidents as well.

4.2 Determine whether the problem was addressed, including any challenges and limitations

The problem overall points to crime increasing overall in the region of Miami Dade when storm activity was present. The amount of the losses of victims increased drastically during stormy weather and overall the data shows the suspicions of the Miami PD as being proven true. Challenges and limitations do exist with the visual though, however, the data set used for the

visual only gives us insight into the losses of victims and does not show how many individual events occurred relative to population. If we were able to compare the change in population to see how an increase in population could have led to more incidents, then we could have compared this number to previous years or even other counties with similar makeup of Miami such as Fort Lauderdale and Tampa. We also run into the issue of potentially having outliers in our visualization as well which could cause the losses as looking much worse for specific time periods within our dataset. Not filtering out the outliers fully would mean we have poor quality analysis and distorted view of the issue at hand.

4.3 Report potential new findings

Overall, our analysis covers the base problem at hand which Miami PD wants to figure out, but we do have insights they could use from the analysis to either predict future losses or prevent future crimes. Looking at the losses, we can use this info to look into the cities or areas within the county which could be most at risk for high volume or cost of losses. If the county were to look at cities or regions most at risk for high volume of losses, they could plan more accordingly to prevent these crimes from occurring and create a system to protect citizens as well. Finding what kind of storms could potentially create the scenarios for crime to occur will allow them to also create action plans to prevent crime from occurring in the first place as well. Using the data will allow them to establish a baseline for the types of crime to expect and further anticipate to prevent from occurring along with tracking potential patterns of criminals as well.