4/27/2020

An Exploration of Homelessness and Bias Motivation Impacts on Hate Crime Incident

ALEXTYR Group Members:

Tram Ngoc Le, Yanyan Deng, Lujie Li, Rishi Purwar

BG-7975-002| Business Analytics for manager

PROFESSOR R.P.

1. **Matrix**

Research question: What were the relationship between homelessness, bias motivation and hate-crim incidents?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hypothesis | Question | Analytic  type | Variable(s) | Chart |
| 1.The higher number of homelessness, the higher number of hate crime incidents | How were the number of homeless and hate crime incidents distributed per state over the country? | Descriptive | Overall Homeless,  States,  Number of Incidents (LOD-Total number of incidents by Year} | US Map -Chart 1 |
|  | How were the changes in homelessness from 2007 to 2018 across U.S. regions | Descriptive,  Time series | Overall Homeless (LOD-Overall Homeless by Year), Region, Year | Stacked Columns, Part of A Whole – Chart 2 |
|  | How was the trend of homelessness and incidents from 2007 to 2018? | Descriptive,  Time series, | Number of Incidents (LOD – Total number of incidents by Year),  Overall homeless,  States (LOD-Overall Homeless by Year), Year | Combination of line and bar charts – Chart 3 |
|  | What was the correlation between the number of hate-crime individual victims and individual homelessness based on Region? | Statistical,  Regression Model | Total individual victims,  Overall individual homeless,  State,  Region | Scatterplot and Trend line (Logarithm Transformation) – Chart 4 |
|  | Is there a correlation between individual homelessness and individual victims in the United State? | Statistical,  Regression Model | Total individual victims, Overall individual homeless | Regression Model (Logarithm) – Chart C. |
| 2.Anti-race, anti-religion, and anti-transgender victims were the major target of hate-crime | How was bias motivation distributed among hate-crime victims across region? | Descriptive Analytics | Total victims (LOD – Victims by Year),  Biased motivation,  Region | Tree map – Chart 5 |
|  | What were the top 10 hate crimes with the highest number of offender and victims in the United States? | Descriptive Analytics (drill down by Bias Motivation) | Total victims,  Total offenders,  Bias Motivation | Butterfly (bar) chart – Chart 6 |
|  | How was the offender race related to the bias motivation crime? | Descriptive Analytics (Drilldown by Bias Motivation and Offender races) | Total Victims,  Offender Race,  Bias Motivation | Sankey Chart – Chart 7 |
| 3. Hate Crime would increase in the next 2 years | How were individual hate-crime victims distributed among the states | Descriptive Analytics,  Drill Down | Region, State, Total individual victims | Lollipop Chart – Chart 8 |
|  | What was the trend of hate crime incident in the 2019 and 2020? | Predictive analysis, Time series | Incident ID, Month | Line Chart – Chart 9 |
| 4.Deserted or private locations tended to have the highest number of hate-crime cases | How was the trend of hate-crime cases distributed based on location? | Descriptive Analytics | Incident Locations,  Total victims,  Region | Heatmap – Chart 10 |
|  | What was the relationship between incident locations and bias motivation? Where was the most popular location of incidents for each crime? | Descriptive Analytics | Incident Locations,  Number of Incidents, Bias Motivation | Highlighted Table – Chart 11 |
|  | What was monthly the trend of hate crimes incidents and offenders by location across region? | Descriptive Analytics | Incident Location, Incident Date (Month), Number of Incident, Number of Offenders,  Region | Combination area and line chart – Chart 12 |
|  | What was the most popular of offense types of hate crime in the U.S.? | Descriptive Analytics | Offense Types, Number of Victims | Donut Chart – Chart 13 |
|  | What was the most popular location for white offender to committed crime | Predictive Analytics,  Machine Learning Model (Decision Tree) | Total offenders,  Incident Locations,  Offenders’ race | Bubble Chart – Chart 14 |
| 5. The urban areas tended to have more hate crime than the suburban area | How hate crime victims were distributed across cities, counties level? | Statistical Analytics,  Drill Down Statistical Analysis,  Median,  Max,  Min | Population Group Description,  Total victims, month | Boxplot – Chart 15 |
|  | Which population range do the area have the highest number of victims? | Descriptive Analytics,  Drill down,  Trend analysis | Population Group Description,  Total Victims, Region, State | Bar Code Chart – Chart 16 |
|  | How were the top 5 offense types distributed among cities and all counties level? | Descriptive, Analytics,  Drilldown on area level | Population group description (Group by Cities, MSA counties, Non-MSA Counties),  Offense Type,  Total Victims (LOD Level),  Sum of Accidents (LOD Level) | Bar-in-Bar Chart – Chart 17 |
| 6. Individual victims tended to suffer from multiple offense | Do individual victims of hate crime suffer from multiple offense more than single offense? | Descriptive Analytics | Multiple Offense,  Total Individual Victims | Pie Chart – Chart 18 |
|  | Which states has the highest trend of individual victims from 2007 to 2018? | Descriptive, Analytics,  Time series,  Trend Analysis | Total Individual Victims,  States,  Incident Date | Cumulative Curves – Chart 19 |
|  | How were the number of victims distributed among population group? | Descriptive Analytics | Total Individual Victims, Total number of description group | Treemap – Chart 20 |
|  | How to classify the cases as multiple bias or single bias by incident locations? | Predictive  Analytics,  Machine Learning, Multi-classification | Multiple bias, Incident Location, Offender Races | Random Forest  SPSS Results – Chart D |

1. **Charts**

A close up of a map

Description automatically generatedChart 1 – U.S. Map of 12-Year Average Homelessness and Hate Crime Incidents (2007-2018)

* The chart showed the 12-year average number of hate crime incidents and homelessness on the U.S. map for each State.
* The chart insight was that the higher the number of homeless, the highest the number of hate crime incidents and hate crime victims. In the West and the Northeast were the hotspot of homelessness and hate crime incidents as well.
* The chart implied there was approximately positive correlation between number of homelessness and number of hate crime.

A screenshot of a cell phone

Description automatically generatedChar 2 – Stacked Columns Chart of 12-year Changes of Homelessness by Region (2007-2018)

* The chart showed the percentage of homelessness by Midwest, Northeast, South, and West for each year from 2007 to 2018.
* The chart insight was the change in distribution of homelessness across region in the United States from 2007 to 2018.
* The chart implied that the West has the highest number of homelessness and increases in 2017 and 2018 while the Northeast has reduced the number of homelessness among that period. The homelessness in the Midwest and the South were moving at a stable rate as these two regions has the lowest of homelessness in the country.

A screenshot of a cell phone

Description automatically generatedChart 3 – Combination Chart of Incidents and Homelessness Changes Over Time

* The chart showed the number of incidents as trend line, overall homelessness as columns and the year from 2007 to 2018.
* The chart insight was to understand if hate crime incidents and homelessness would have the same changes in volumes from 2007 to 2018.
* The chart implied that hate crime incidents and homelessness overall approximately move in the same direction and there was possible positive correlation between homelessness and hate crime.

A close up of a map

Description automatically generatedChart 4 – Logarithm Scatterplot of Individual Victims and Overall Homeless Individual

* The chart showed the logarithm of overall homeless individuals as explanatory variable, total individual victim as response variables, region and the logistic regression model in logarithm form (trend line by region). The trend line of the West fit perfectly to the points. There were small variances between the logarithm trend line in the Northeast, South, and Midwest.
* The chart insight was to explore the relationship between individual victims and overall homeless individual between regions. The higher the overall homeless individuals, the higher the total individual victims in the West.
* The char implied that there was a positive correlation between individual victims and overall homeless individual in West. P-value was smaller than 0 in the model description, which meant the model was significant.

A screenshot of a cell phone

Description automatically generatedChart 5 – Tree Map of Bias Motivation

* The chart showed the accumulative number of victims from 2008 to 2012 of difference hate-crime type based on bias motivation across regions.
* The chart insight was to understand the relationship between hate crime victimization and bias motivation across each type of crime.
* The chart implied that race-based crime was the most popular crime in the United States across Regions, especially in the West. In the Northeast, Religion seemed to be the second bias motivation with the highest number of victims comparing to other regions, which sexual orientation was the second bias motivation with the highest number of victims.

A screenshot of a cell phone

Description automatically generatedChart 6. Butterfly Chart of Top 10 Types of Hate Crime in the U.S.

* The chart showed the number of victim and offender in comparison regarding the top 10 hate-crime types in the U.S.
* The chart aimed to specify the type of hate crime with the highest number of victims and the relationship between victim and offender.
* The chart implied that Anti-Black or African American was the top problem of hate crime with the highest number of victims and offenders. Mostly, number of offenders were equal to number of victims. However, Anti-Jewish hate crime offenders was approximately one third of number of victims.

A picture containing umbrella

Description automatically generated

Chart 7. Sankey Chart of The Victim Flow From Bias Motivation to Offender Profile

* The chart showed the percentage of victims by bias motivation and offender races and the flow of victims from bias motivation to offender races.
* The chart insight was the connection of bias motivation and offender races in order to identify the kinds of offenders that committed certain crime.
* The chart implied that the number of racism victims was the highest in the U.S. and white offenders, taking up to 41% of other offender races, played a major part in hurting these people. Also, the racism victims were mostly attacked by white offenders or unknown offenders.

A screenshot of a cell phone

Description automatically generatedChart 8. Lollipop Chart of Individual Victims and Overall Individual Homeless based on State

* The chart showed the total number of victims in the circle size, the total number of individual homeless as the length of the bar chart by each state.
* The chart insight was to understand the relationship between the number of individual hate crimes and the individual homeless across all states. California, New York, Florida, and Texas had a high number of individual homelessness however the number of hate crime individual victims was not necessarily correlated with the homelessness number. There were a lot of states with low number of homelessness yet a high number of hate crime.
* The chart implied that there was not necessarily a correlation between Individual victims and overall individual homeless for every state. If there is, probability the number was biased based on some outlier such as California.

A screenshot of a cell phone

Description automatically generatedChart 9. Line Chart of Hate Crime Incident Prediction

* The chart showed the number of hate crime incidents from 2007 to 2018 and the prediction of those incidents in 2019 and 2020 on a monthly basis.
* The chart insight was the trend of hate crime victimization in the United States through the time. Overall, hate crime incidents has been decreasing during these 12 years. However, there were peaks and plummets during those months.
* The chart implied that the number of hate crime incidents will decrease in 2019 and 2020, around 400 to roughly 600 cases.

A screenshot of a social media post

Description automatically generated

Chart 10 – Heat Map of Incident Location Across Region

* The chart showed the number of hate crime victims across Midwest, Northeast, South, and West for each of the incident location.
* The chart insight was a descriptive analysis of incident location across regions, which answer the question of which location has the highest potential of incidents to prevent what will happen in the future.
* The chart implied that the West has the highest number of victims in Residence/Home and on Highway as well. The South has the highest number of victims in Residence. In general, Residence, Unknown, Highway places were the top incident locations that have the highest number of victims.

A screenshot of a cell phone

Description automatically generated

Chart 11 – Highlighted Table of Incident Location and Hate Crime Type

* The chart showed the number of incidents happened for each incident locations across with bias motivation
* The chart insight was to understand which type of crime happen at what kind of place the most.
* The chart implied anti-race crime has the highest number of incidents at residence/home, highway/street/field, unknown place, school, bus and gas station. Religion related crime were likely to happen at residence home and unknown place. Sexual orientation also had the highest incident at residence and highway, which were deserted places with less people involved.

A close up of an animal

Description automatically generatedChart 12 - Area Chart of Monthly Incidents Patterns by Location (2008-2012)

* The chart showed the monthly accumulated number of hate crime incidents and offenders by incident location from 2008 to 2012, drilled down by region.
* The chart insight was to demonstrate the monthly trend of hate crime incidents in the top 5 incident locations with the highest number of crimes committed.
* The chart implied that deserted area such as highway, street field and lake as well as resident home has the highest number of hate crime, especially during Summer. Northeast had the highest “unknown” incidents location. At the same time, the number of offenders in residence home was lower than the number of incidents, which indicated that the offenders committed the crime again. Hate crime tended to peak at school when the semester started and plummeted during summer.

A screenshot of text

Description automatically generatedChart 13 – Donut Chart of Offense Type

* The chart showed the accumulated hate crime victims of 76,696 and the percentage of offense type as part of a whole.
* The chart aimed to identify the distribution of top 10 offense types with the highest number of victims in the U.S.
* The chart implied that destruction/vandalism of property, intimidation, simple assault and aggravated assault were causing the most concern in the U.S., especially Destruction of Property. There were 30.35% of victims were attack by destruction of property.

A screenshot of a cell phone

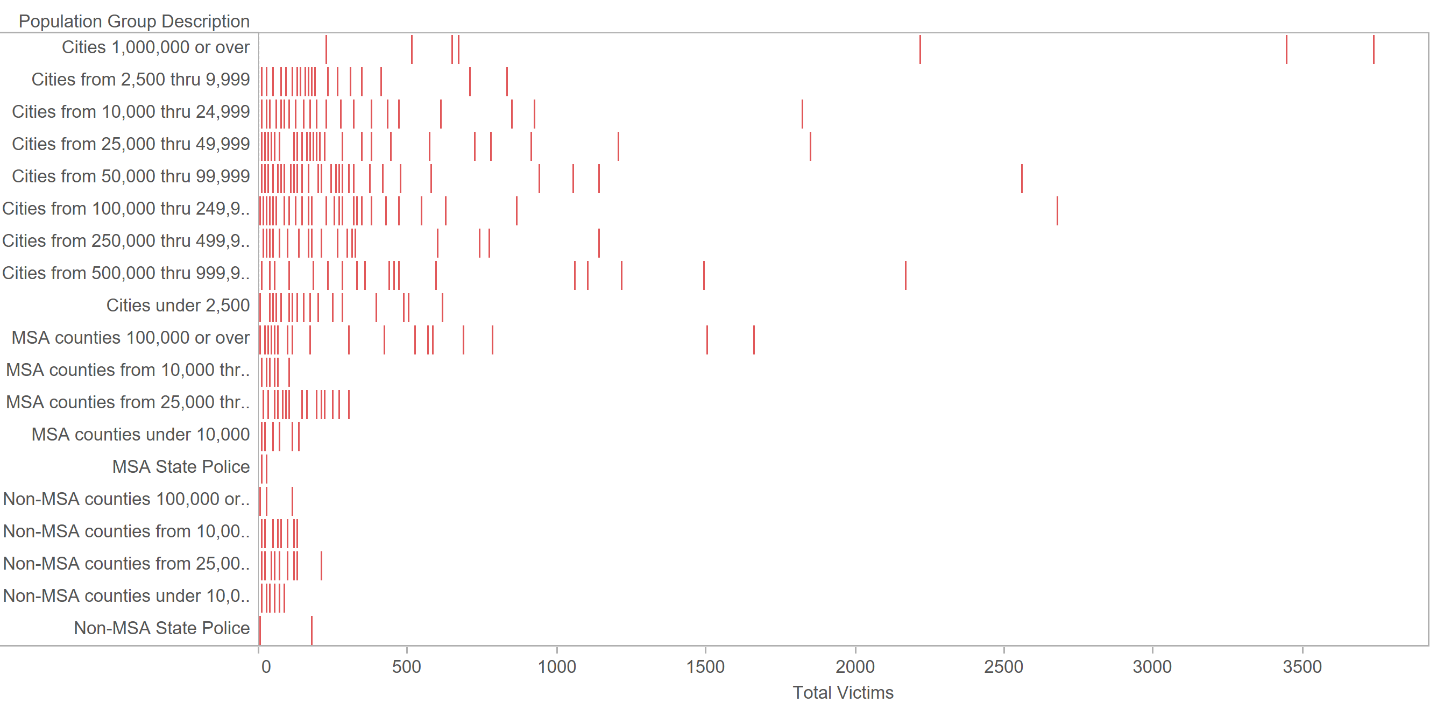
Description automatically generatedChart 14 – Bubble Chart of Offender Races and Incidents Location

* The chart showed the accumulative number of incidents over the last 12 years based of each offender types and incident locations.
* The chart insight was to explore where the certain offender would like to commit crime.
* The chart implied that White offender tended to commit crime the most at residence/home while back African or American offender tended to commit crime the most on highway or street. Offenders with unknown race tended to commit crime at home since they might focus on intimidation offense types.

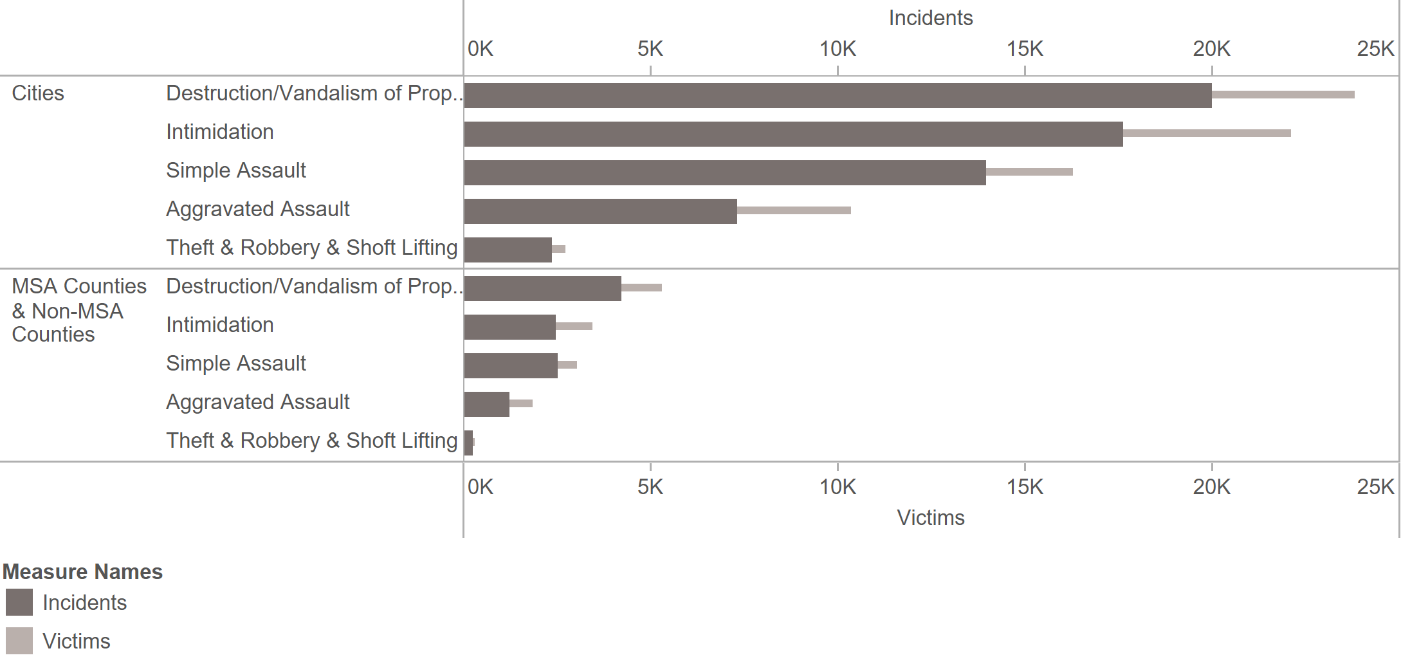
A screenshot of a video game

Description automatically generatedChart 15 – Boxplot of Monthly Victims by Cities, MSA Counties and Non-MSA Counties

* The chart showed the 12-year accumulative number of victims distributed by each metropolitan type, including cities, MSA counties and Non-MSA Counties, from 2007 to 2018 per month.
* The chart insight was the comparison of number victims on monthly basis between cities, MSA (Metropolitan statistics area) counties and Non-MSA (Non-Metropolitan statistic area) counties. On a monthly basis, the smallest number of victims of cities was 345, the largest number of victims of cities was 804, and the median number of victims of cities was roughly 562. Those numbers were much higher than the number of victims in MSA counties and Non-MSA counties. In the MSA counties area, on a monthly basis, the number of victims reached a minimum of 35, a maximum of 151, and a median of 83. At the same time, Non-MSA counties area reached the minimum victims of 5, maximum victims of 42, and median victims of 22.
* The chart implied that cities, or urban area, had much more victims of hate crime than counties areas do.

Chart 16 – Bar Code Chart of Total Victims Distribution by Population Level

* The chart showed the total number of hate crime victims accumulated from 2007 to 2008 of each state distributed by different population of urban area, cities, suburban area, counties and State police.
* The chart insight was to compare the number of victims between different population range, as a drill down on urban and suburban areas.
* The chart implied that cities with more population tended to have more hate crime victim, the rationale went the same with counties level. In general, the higher the population of each area segments, cities or counties, the higher the hate crime victims. However, this rationale did not apply with Non-MSA counties and State Police.

Chart 17 – Bar in Bar Chart of Offense Types on Cities and Counties

* The chart showed the number of incidents and the number of victims drilled down on the offenses types that happens in these areas.
* The chart insight was to identify the most popular type of offense in each area.
* The chart implied that in cities and counties, the most serious offense type was destruction/ vandalism of property. However, in the city intimidation was more serious than simple assault while in the counties the number of cases of these two offense types were roughly the same. Therefore, we in the city we should care more about intimidation than simple assault while in the country we would have to take these two offense types into account at the same level of importance.

A close up of a card

Description automatically generatedChart 18 – Pie Chart of Multiple and Single Offense Types on Individual Victims

* The chart showed the number of individual victims as percentage on two types of offense, multiple or single.
* The chart insight was to understand the situation of individual victims in the United States. Multiple offense on individual victims only takes 8.30% of all the individuals yet single offense on individual victims took up to 91.7% of all the individual victims in the country.
* The chart implied that individual victims tended to suffer from single offense in the United States.

A close up of a map

Description automatically generatedA close up of a map

Description automatically generatedChart 19 – Cumulative Curve of Individual Victims from 2007 to 2018

* The chart showed the cumulative curve of individual victims for each state from 2007 to 2018. Each curve represented the running sum of total individual victims based on the date of the incidents.
* The chart insight was to understand the trend of the hate crime individual victimization situation for each state for a 12-year period.
* The chart implied that California has an extremely strong trend in hate crime victimization. A lot of incidents happened in California from 2007 to 2018. On the east side of the country, Miami was the state that showed a strong trend in the increase of individual victims. California and Miami were the top two state that the government should pay attention too.

A screenshot of a cell phone

Description automatically generated

Chart 20 – Tree Map of Population Group Description

* The chart showed the accumulative number of victims from 2008 to 2012 of difference hate-crime type based on bias motivation across regions.
* The chart insight was to understand the relationship between hate crime victimization and bias motivation across each type of crime.
* The chart implied that race-based crime was the most popular crime in the United States across Regions, especially in the West. In the Northeast, Religion seemed to be the second bias motivation with the highest number of victims comparing to other regions, which sexual orientation was the second bias motivation with the highest number of victims.

**C. Statistical Model (Hypothesis 1 from the matrix)**

* **A screenshot of a computer

  Description automatically generatedA screenshot of a cell phone

  Description automatically generatedHypothesis:** There is a correlation between homeless individuals and hate crime individual victims

Chart C. Scatterplot of Statistical Model and Regression Model (Logarithm)

* The chart showed the trend line between number of homeless individuals and the number of total individual victims in logarithm form.
* The chart insight was to test the hypothesis if there was a correlation between homeless and hate crime incidents. P-value < 0.0001, which means that the slopes were significant.
* The chart implied that there was a positive correlation between individual victims and homeless individuals.

**D. Machine Learning – Random Forest**

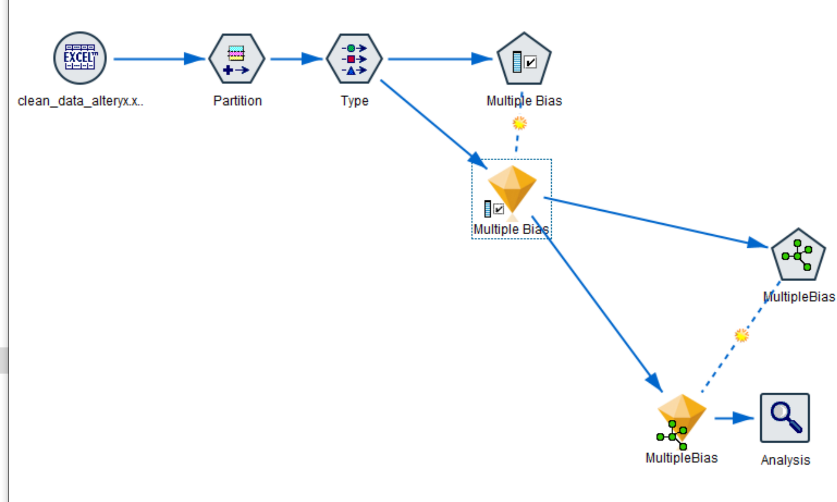
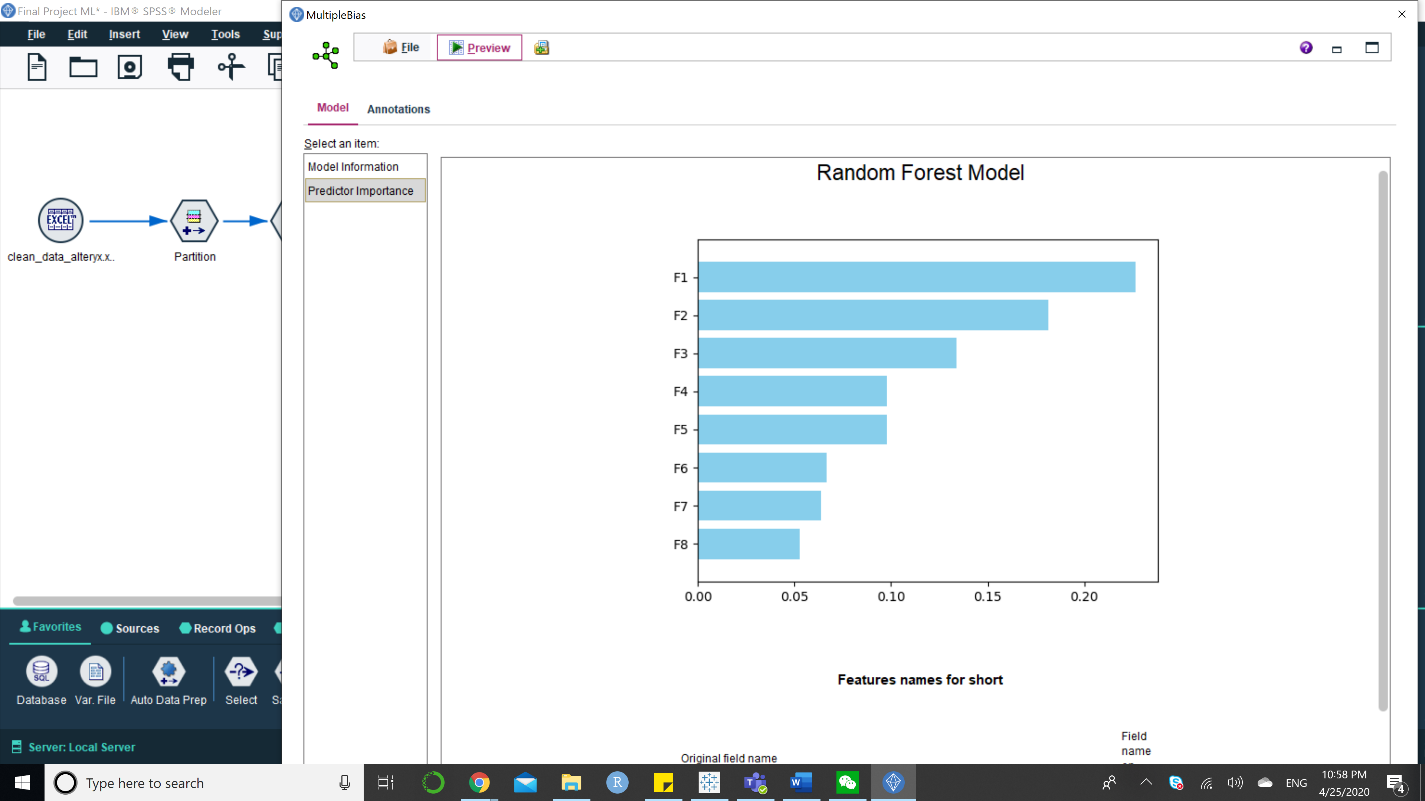
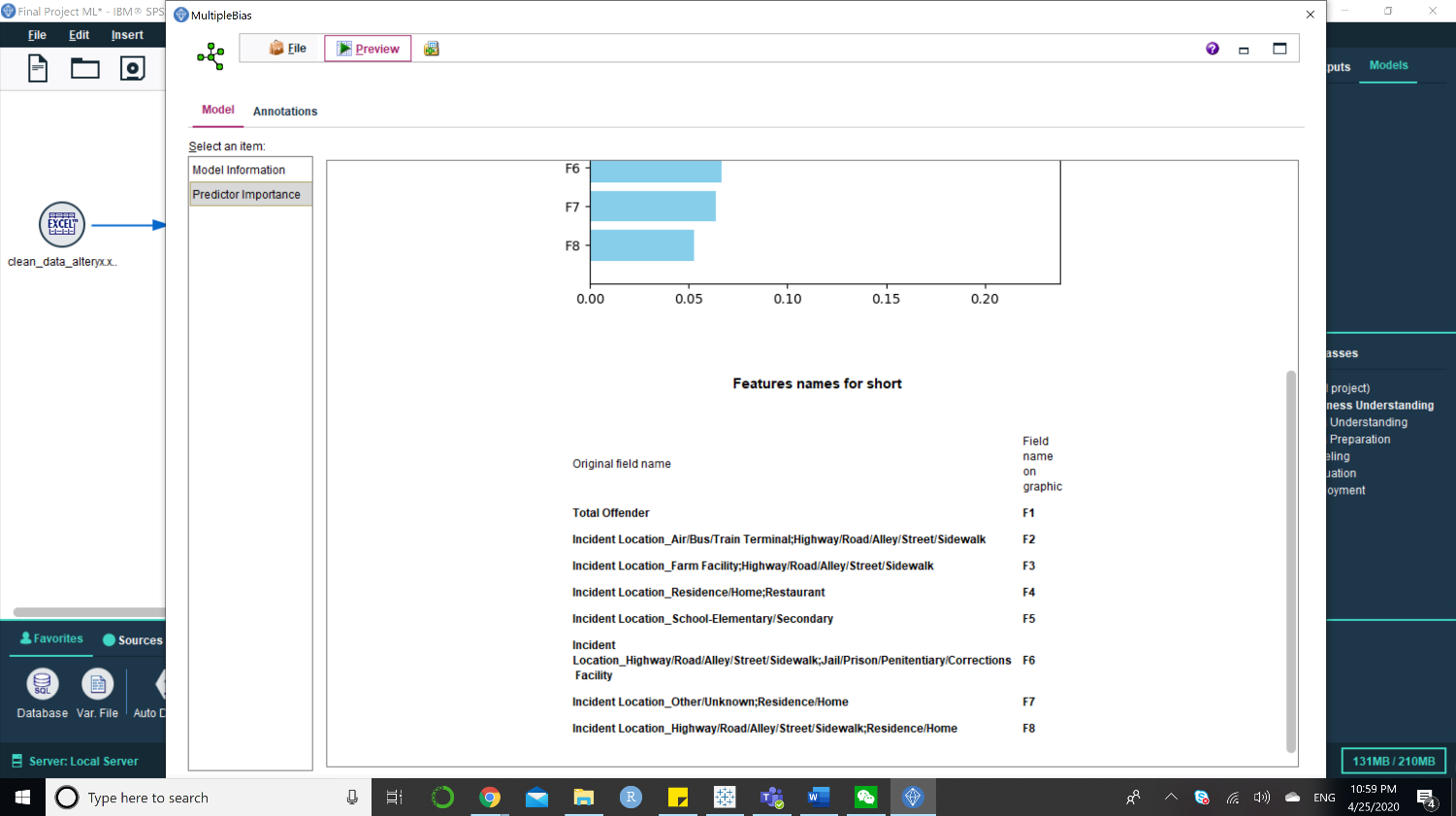
* ******Hypothesis**: Individual victims tend to suffer from multiple offense in deserted place such as highway, road, alley

Chart D of Machine Learning Model - Random Forest Summary

* The chart showed the summary of importance predictor about if an individual would have to suffer from a single bias or multiple bias type of hate-crime.
* The chart insight was to understand which factor can predict if the hate crime incident was single or multiple bias. The most important factor was Total offender and second was incident location.
* The chart implied that the top incident location was at high-way or bus terminal, which was very deserted places with less people. On the other hand, school at elementary and secondary can be another place for hate crime incidents to happen.