

## C740 Assessment

### Part 1:

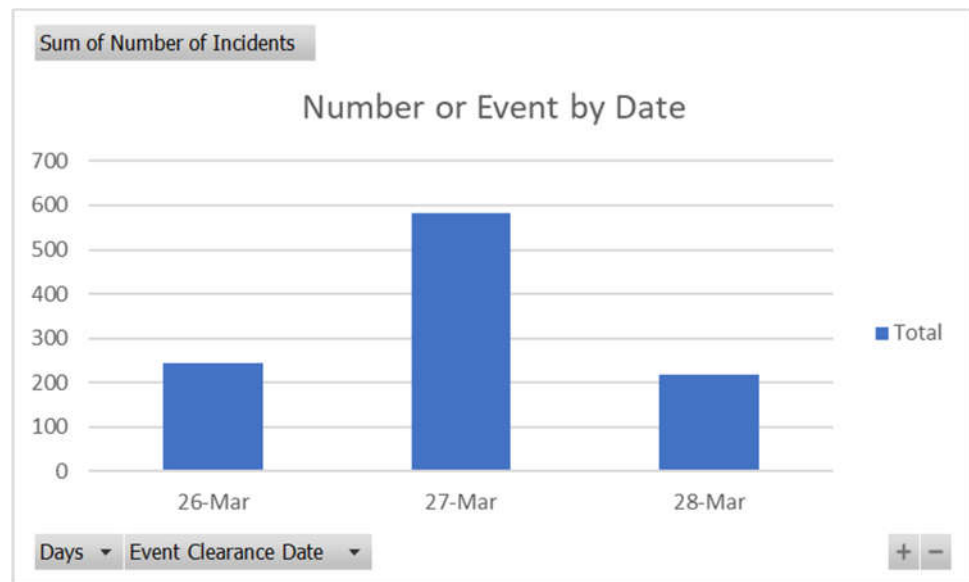
#### A. & B. Preparing a dataset from the data provided. The Datasheet is attached as an Excel file.

I am going ahead and remove the following columns from the original data due to duplicate /missing records:

- Column M (Longitude) and Column N (Latitude) contain partial information. We can extract them from Column O (Incident Location). Having Columns M and N is redundant.
- Column F (Event Clearance SubGroup) and G (Event Clearance Group) contain the same information as listed in column E (Event Clearance Description).
- Similarly, Column P, Column Q, and Column R have duplicate information with Column E. We can remove them and keep Column E (Event Clearance Description).
- On the other hand, I will also remove Column S (At Scene Time) and Row 226 (CAD CDW ID: 1702543) due to missing values.

#### C. Create data sheets using the cleaned data.

Date	Number of Events
03/26/16	243
03/27/16	583
03/28/16	219
<b>Total</b>	<b>1045</b>



By using Pivot Table, I created a new table named *“Number of Incidents by the Type of Event”*

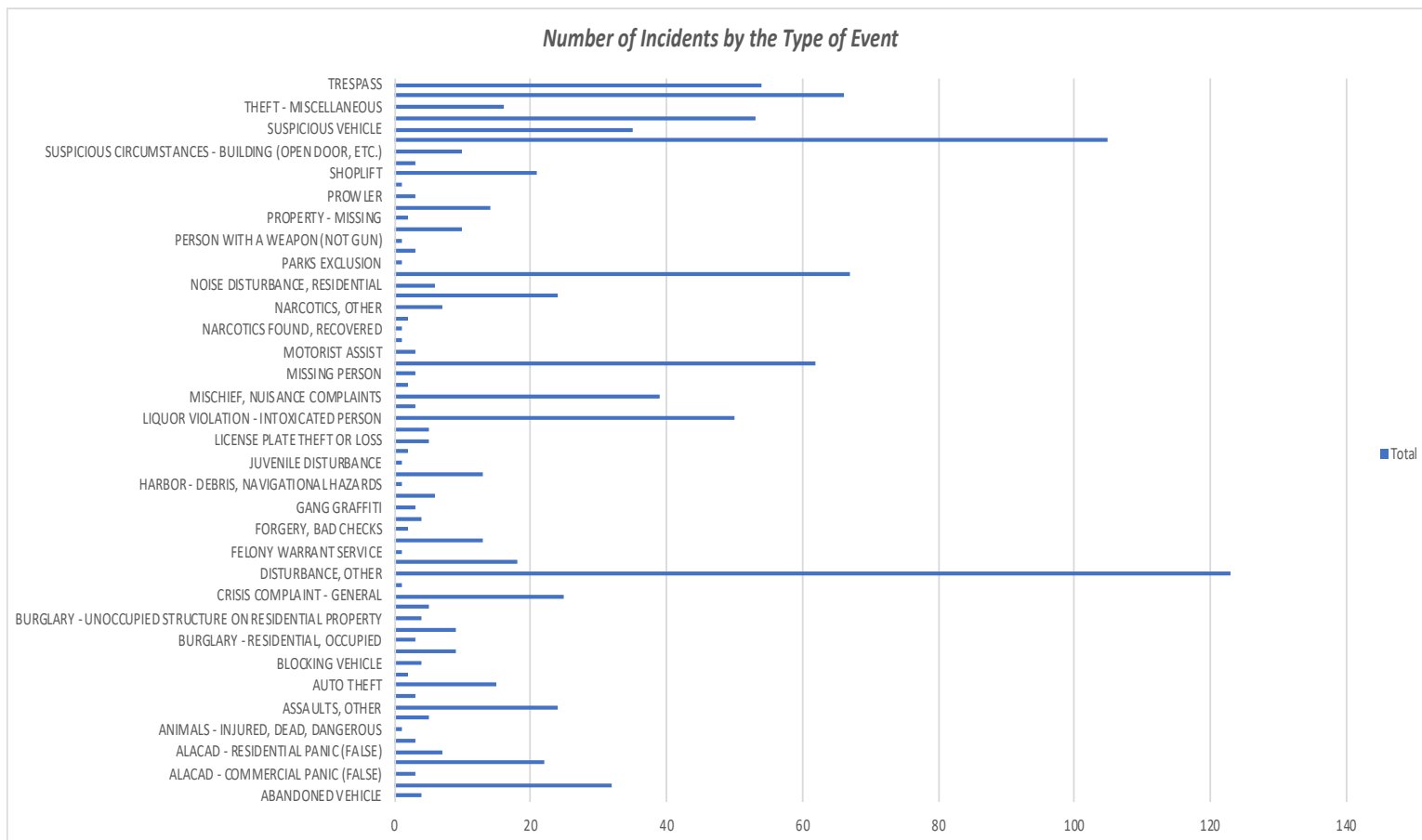
Event Type	Number of Incidents
ABANDONED VEHICLE	4
ALACAD - COMMERCIAL BURGLARY (FALSE)	32
ALACAD - COMMERCIAL PANIC (FALSE)	3
ALACAD - RESIDENTIAL BURGLARY (FALSE)	22
ALACAD - RESIDENTIAL PANIC (FALSE)	7
ANIMAL NOISE, STRAYS, BITES	3
ANIMALS - INJURED, DEAD, DANGEROUS	1
ARMED ROBBERY	5

ASSAULTS, OTHER	24
AUTO RECOVERY	3
AUTO THEFT	15
BICYCLE THEFT	2
BLOCKING VEHICLE	4
BURGLARY - COMMERCIAL	9
BURGLARY - RESIDENTIAL, OCCUPIED	3
BURGLARY - RESIDENTIAL, UNOCCUPIED	9
BURGLARY - UNOCCUPIED STRUCTURE ON RESIDENTIAL PROPERTY	4
CASUALTY ( <u>NON-CRIMINAL</u> /TRAFFIC) - MAN DOWN, SICK PERSONS, INJURED, DOA)	5
CRISIS COMPLAINT - GENERAL	25
CRISIS COMPLAINT - PICK-UP OR TRANSPORT	1
DISTURBANCE, OTHER	123
DRIVING WHILE UNDER INFLUENCE (DUI)	18
FELONY WARRANT SERVICE	1
FIGHT DISTURBANCE	13
FORGERY, BAD CHECKS	2
FRAUD (INCLUDING IDENTITY THEFT)	4
GANG GRAFFITI	3
HARASSMENT, THREATS	6
HARBOR - DEBRIS, NAVIGATIONAL HAZARDS	1
HAZARDS	13
JUVENILE DISTURBANCE	1
LEWD CONDUCT	2
LICENSE PLATE THEFT OR LOSS	5
LIQUOR VIOLATION - ADULT	5
LIQUOR VIOLATION - INTOXICATED PERSON	50
MARIJUANA PUBLIC USE (NOT DISPENSARY)	3
MISCHIEF, NUISANCE COMPLAINTS	39
MISDEMEANOR WARRANT SERVICE	2
MISSING PERSON	3
MOTOR VEHICLE COLLISION	62
MOTORIST ASSIST	3
NARCOTICS ACTIVITY REPORT	1
NARCOTICS FOUND, RECOVERED	1
NARCOTICS, DRUG TRAFFIC LOITERING	2
NARCOTICS, OTHER	7
NOISE DISTURBANCE	24
NOISE DISTURBANCE, RESIDENTIAL	6
PARKING VIOLATION (EXCEPT ABANDONED VEHICLES)	67
PARKS EXCLUSION	1

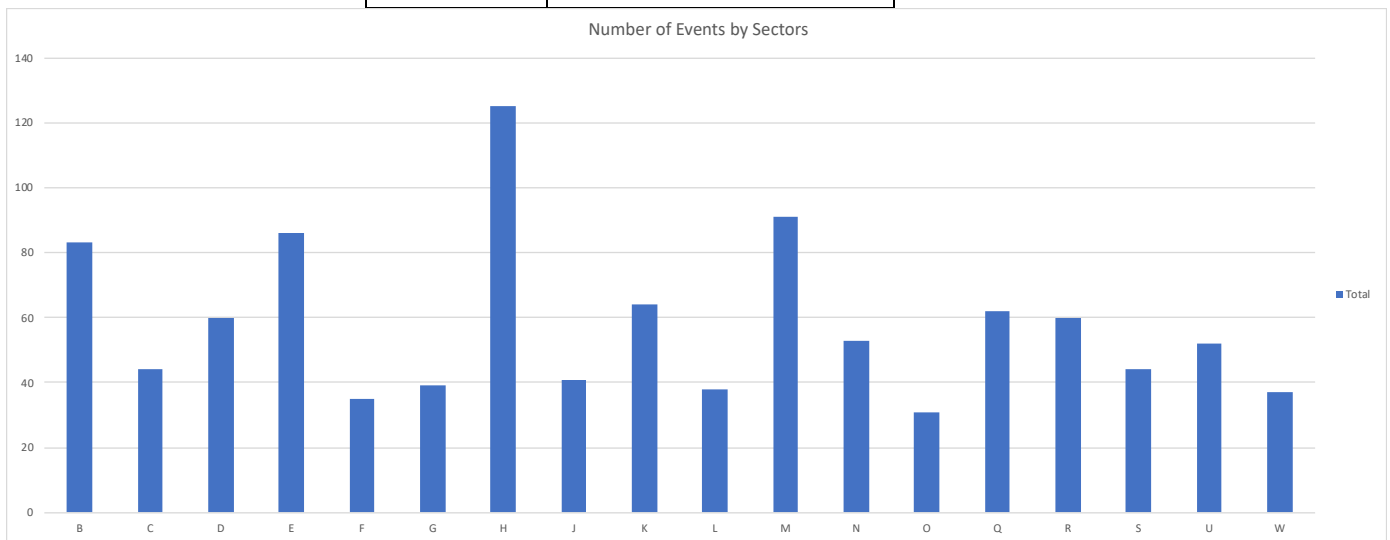
PEDESTRIAN VIOLATION	3
PERSON WITH A WEAPON (NOT GUN)	1
PROPERTY - FOUND (FOLLOW UP TO SPD CASE)	10
PROPERTY - MISSING	2
PROPERTY DESTRUCTION	14
PROWLER	3
RECKLESS ENDANGERMENT, LITTERING, PARKS CODE VIOLATIONS	1
SHOPLIFT	21
STRONG ARM ROBBERY	3
SUSPICIOUS CIRCUMSTANCES - BUILDING (OPEN DOOR, ETC.)	10
SUSPICIOUS PERSON	105
SUSPICIOUS VEHICLE	35
THEFT - CAR PROWL	53
THEFT - MISCELLANEOUS	16
TRAFFIC (MOVING) VIOLATION	66
TRESPASS	54

GRAND TOTAL

1045



Sector	Sum of Number of Events
B	83
C	44
D	60
E	86
F	35
G	39
H	125
J	41
K	64
L	38
M	91
N	53
O	31
Q	62
R	60
S	44
U	52
W	37
<b>Grand Total</b>	<b>1045</b>



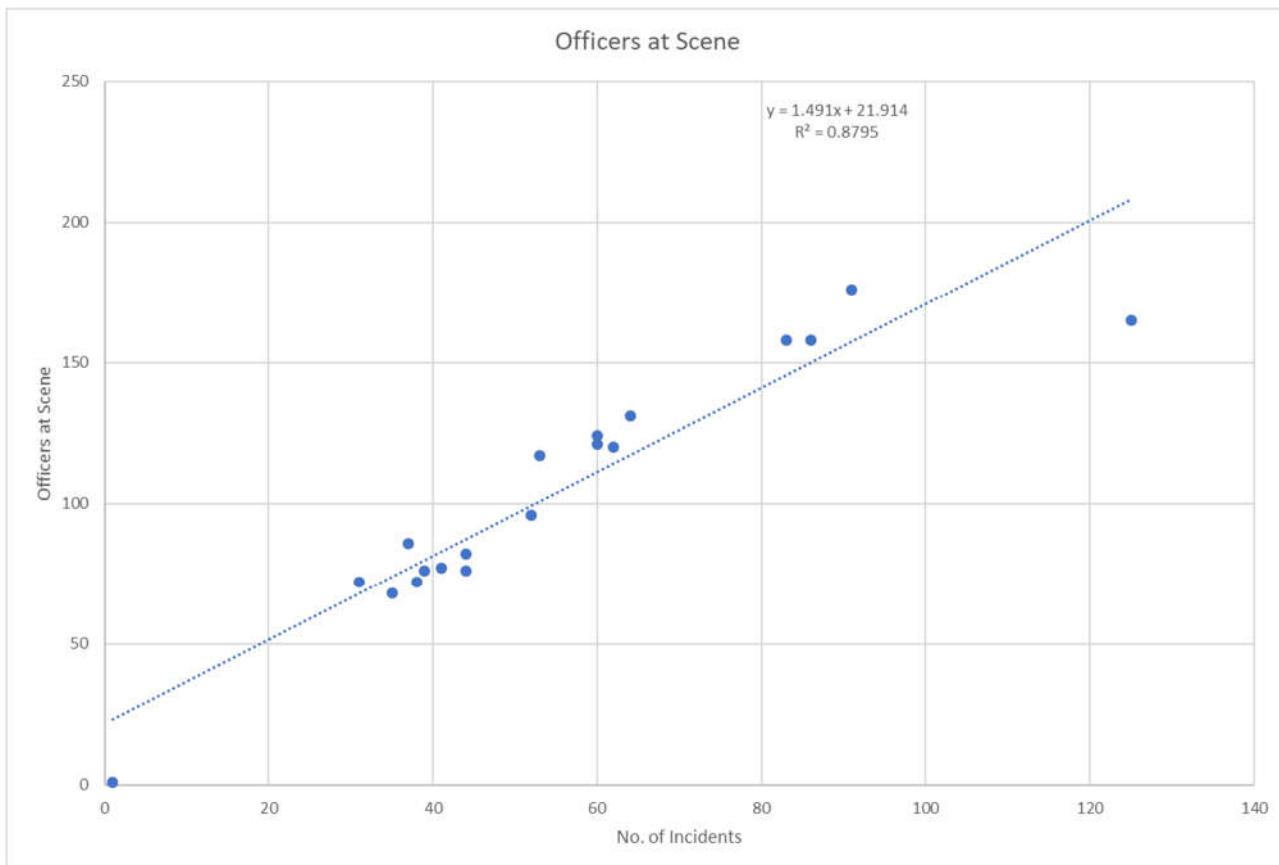
**D. Observation Summary:** The datasheet provided is from 03/26/2016 to 03/28/2016. It is a weekend period along with the Easter Sunday holiday (03/27/2016). A high number of recorded incidents is unusual as Seattle's crime rate is often lower than national average in the United States (US News, 2019). So, the holiday may be the reason why they recorded a high level of incidents because alcohol consumption often raises during holidays. Otherwise, there are two outliers here which may be a result of a false 911 call or a typo.

On the other hand, the highest number of incidents by type belongs to Disturbance and Suspicious Person with an amount of 123 and 105 incidents respectively.

Last but not least, the H sector presents the highest number of incidents by sector with an amount of 125 incidents.

**Part2:**

***E. Describing the fit of the linear regression line to the dataset provided.***



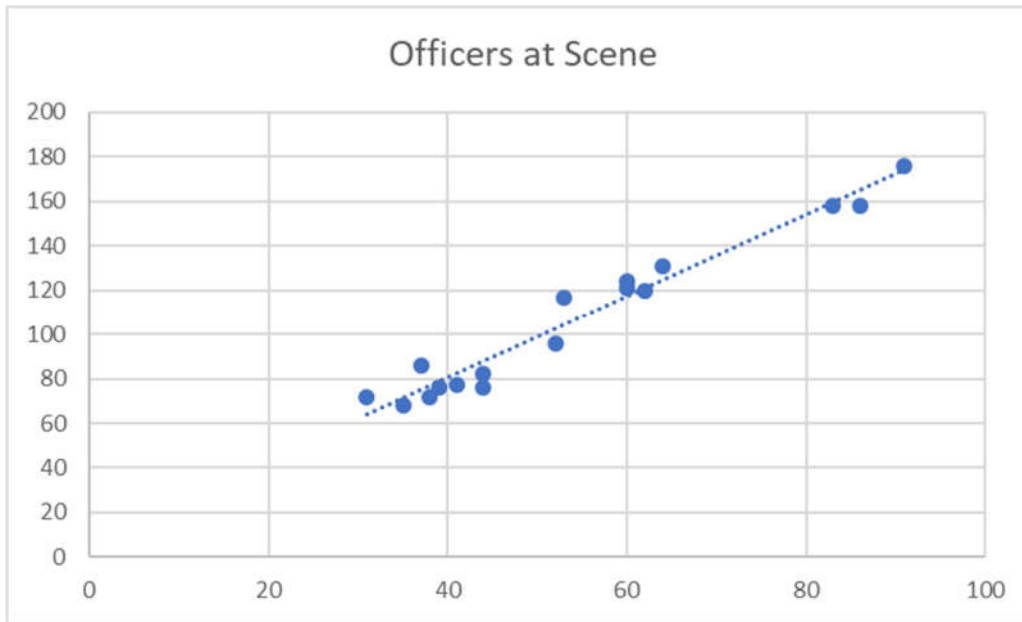
This graph is trending up demonstrates that these 2 variables have a positive relationship. In other words, the higher the Number of Incidents, the more Police Officers will be presented.

With  $x = 0$ ,  $y$  is equal to 21.914, indicating the slope of the graph.  $R^2 = 87.9\%$  also suggests that it is a strong relationship between two variables.

However, there are some outliers and they should be removed from the dataset. **Here are calculated values based on the formula**

x	y
0	21.914
1	23.405
2	24.896
3	26.387
...	...
124	206.798
125	208.289

***F. Describing the effect of the outliers on the dataset.***



x	y
0	7.3058
1	9.1382
2	10.9706
3	12.803
...	...
90	172.2218

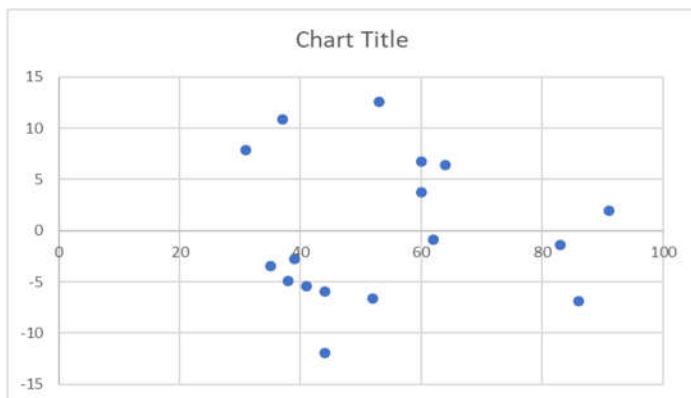
$$Y = 1.8324 * x + 7.3058$$

$$R^2 = 0.9591$$

After removing two outliers from both sides, the slope reduced by almost 3 times to 7.306.

On the other hand, the R<sup>2</sup> rise to 95.9%, suggesting a much stronger relationship between the two variables.

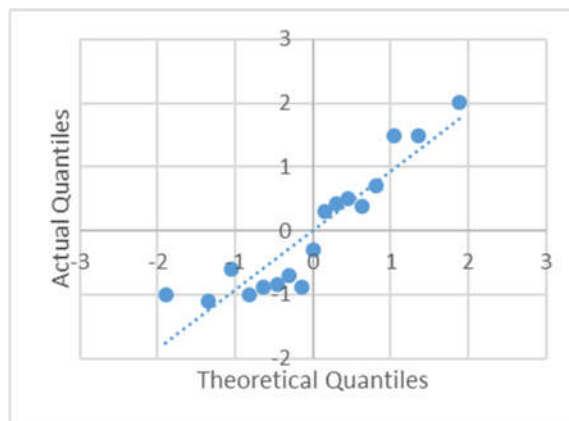
***G. Providing a residual plot and explaining how to improve the linear-regression model.***



The Residual Plot

The residual plot on the left shows the mean to be close to 0, but residuals don't look that they are created randomly or having constant variance. There aren't any residuals on the left side. The majority of the residuals are gathered in the middle of the plot with only a few on the right side.

The residual normal probability plots stand on the right side to assess the normality assumption. As illustrate, the residuals are not even close to the diagonal line. In conclusion, the normality assumption doesn't hold for this data.



***H. Current Qualification: Determine if the police department currently qualifies for the funding.***

Average Police Officers/Incident	1.9453588
Standard Deviation	0.2250085
Incidents > 2.5 Officers	0
Probability > 2.5	0

*Calculating the probability that the department will or won't qualify for the funding in the future.*

Average Police Officers/Incident	2.1280964
Standard Deviation	5.3980703
Incidents > 2.5 Officers	289
Probability > 2.5	28.90%

The table above is the results of the MonteCarlo simulation base on the Numbers of Officers at the scene per incident. The average number of police officers present at the scene is rounded to 2.13. It shows to have about 289 incidents that have more than 2.5 Police Officers at the scene. The probability is equal to 28.9%. This indicates that more than 2.5 Police Officers presented at the scene are only 28.9% of the total cases. In other words, ***the department doesn't qualify for funding*** because of the probability of 28.9% is too low for consideration.

In my opinion, the department should keep the same number of Police Officers as they are doing now.

**I. Describe the precautions that should be taken into consideration while working with and communicating about the sensitive data in this situation.**

The sensitive information in this project is the location and address data. However, based on the longitude & latitude data provided, the reader can easily find a specific location. Anything like that indicating specific people or places should not be included.

**J. References source:**

US News. (2019). *Seattle Crime Rate*. Retrieved from  
<https://realestate.usnews.com/places/washington/seattle/crime>