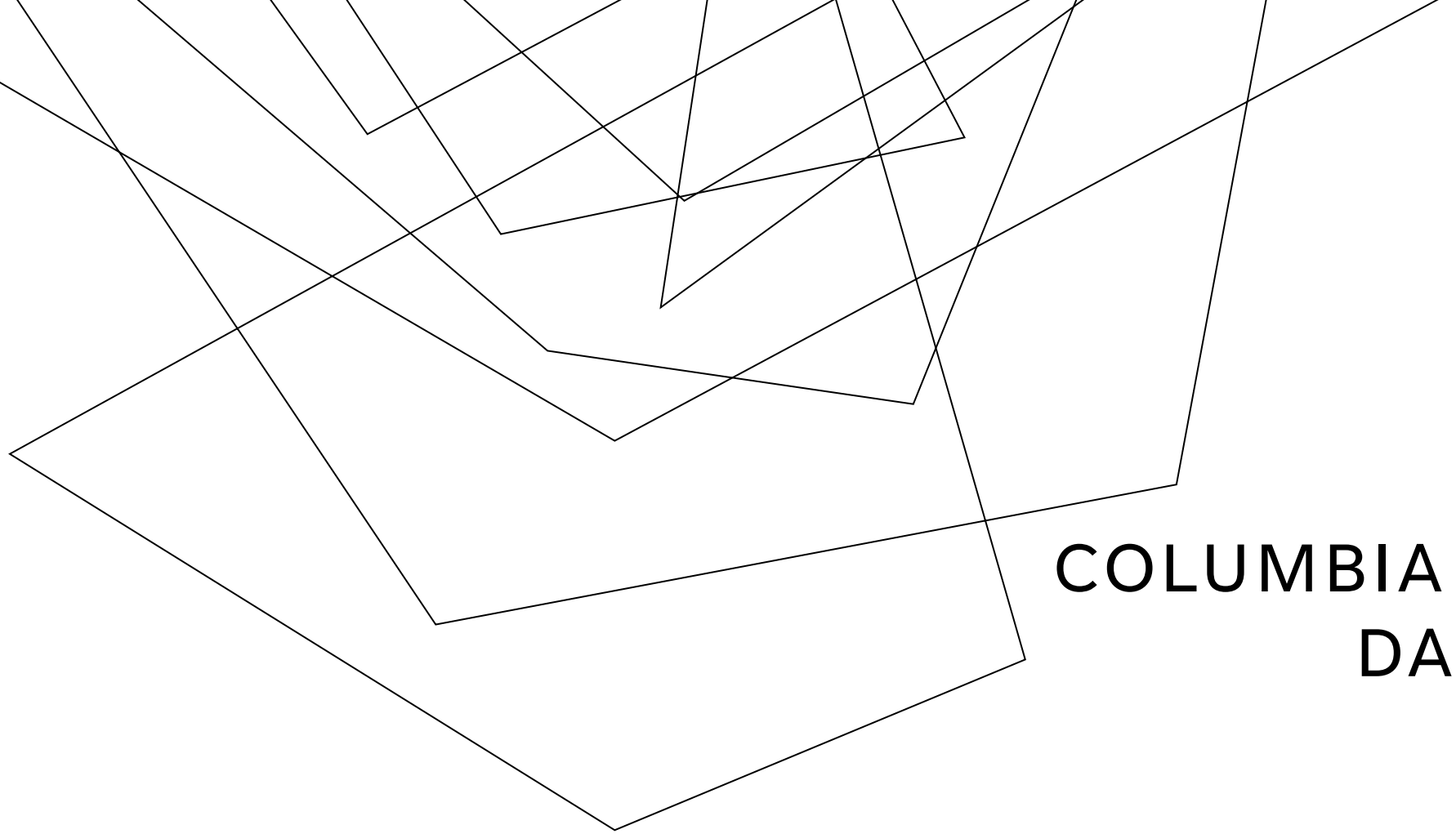


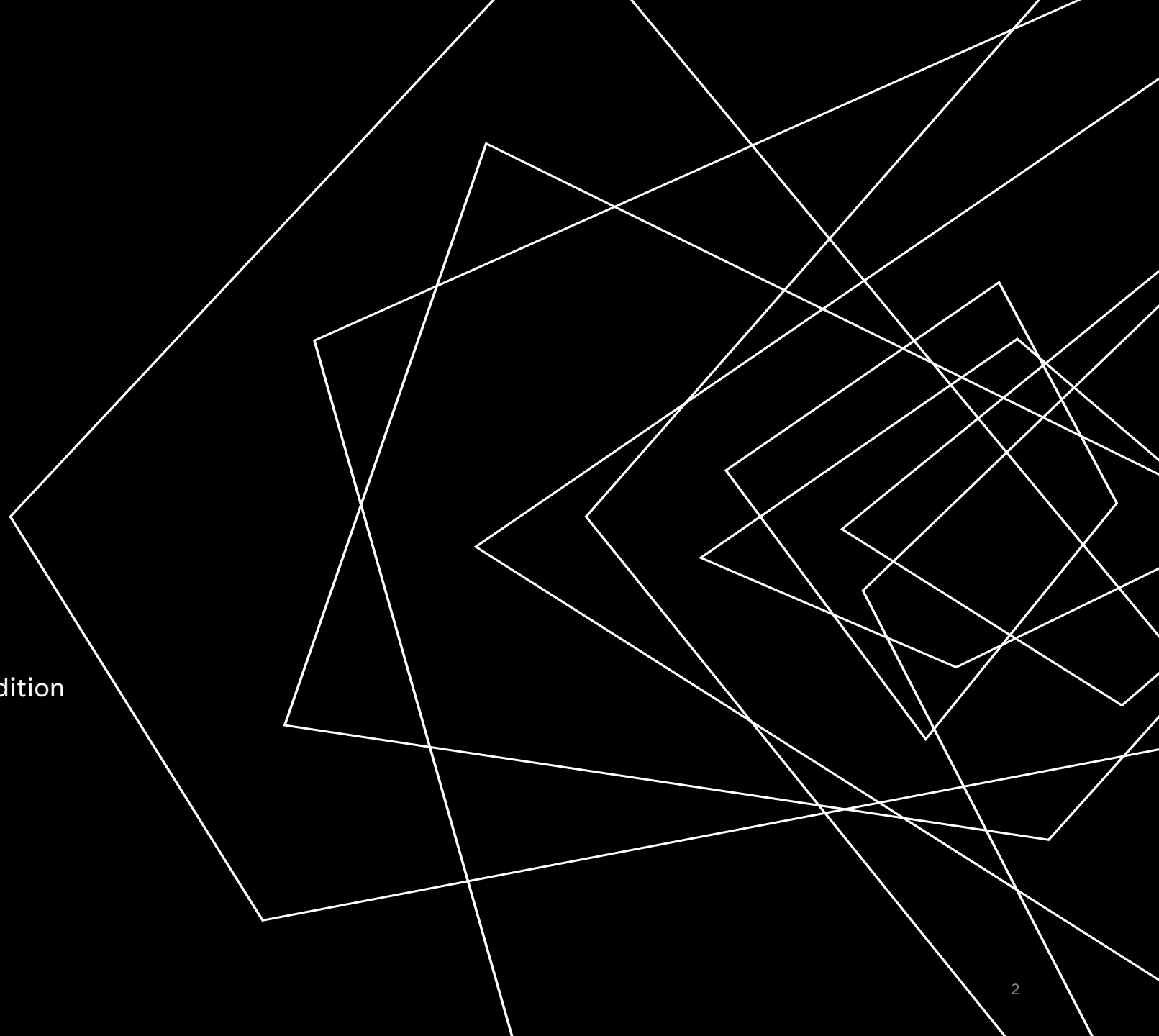
COLUMBIA ENGINEERING DATA ANALYTICS BOOT CAMP PROJECT I

Date: 6/17/2024

Karen Lin, Kevin Zhang, Sunjae Youm

AGENDA

- Introduction
- Data Processing
- Analysis
    - Data exploring
    - ROCKVILLE dataset
        - Speed Limit
        - Traffic control
        - Weather & surface condition
        - Hourly intervals
        - Timeseries
- Conclusion



INTRODUCTION

Background

Our data team assumed that we received a data request from the Montgomery County local government in Maryland. The purpose of the request was to determine how to use the county budget to reduce car accidents. Additionally, Montgomery County wanted to choose a specific municipality and asked us, as data analysts, to identify which roads need immediate improvements. To achieve this, we investigated traffic accident records from January 2015 to March 2024, analyzing the main car crash accident factors and variables from the dataset.

Data Source:
https://catalog.data.gov/dataset/crash-reporting-drivers-data/resource/9851a37f-4f32-464e-8ba6-c23023653a7f

QUESTIONS

Q1. How can we identify the area with the highest frequency of car crashes?

Q2. Which factors in the dataset should we prioritize to effectively reduce car accidents?

Q3. What insightful patterns can we uncover from the dataset?

DATA PROCESSING

DATA EXPLORING – RAW DATA

```
car_crash_df.columns
```

```
Out[3]: Index(['Report Number', 'Local Case Number', 'Agency Name', 'ACRS Report Type',
       'Crash Date/Time', 'Route Type', 'Road Name', 'Cross-Street Type',
       'Cross-Street Name', 'Off-Road Description', 'Municipality',
       'Related Non-Motorist', 'Collision Type', 'Weather',
       'Surface Condition', 'Light', 'Traffic Control',
       'Driver Substance Abuse', 'Non-Motorist Substance Abuse', 'Person ID',
       'Driver At Fault', 'Injury Severity', 'Circumstance',
       'Driver Distracted By', 'Drivers License State', 'Vehicle ID',
       'Vehicle Damage Extent', 'Vehicle First Impact Location',
       'Vehicle Second Impact Location', 'Vehicle Body Type',
       'Vehicle Movement', 'Vehicle Continuing Dir', 'Vehicle Going Dir',
       'Speed Limit', 'Driverless Vehicle', 'Parked Vehicle', 'Vehicle Year',
       'Vehicle Make', 'Vehicle Model', 'Equipment Problems', 'Latitude',
       'Longitude', 'Location'],
      dtype='object')
```

```
In [21]: car_crash_df.dtypes
```

```
Out[21]: Report Number                    object
         Local Case Number                object
         Agency Name                      object
         ACRS Report Type                 object
         Crash Date/Time                  object
         Route Type                       object
         Road Name                        object
         Cross-Street Type                object
         Cross-Street Name                object
         Off-Road Description             object
         Municipality                     object
         Related Non-Motorist             object
         Collision Type                   object
         Weather                          object
         Surface Condition                object
         Light                            object
         Traffic Control                  object
         Driver Substance Abuse           object
         Non-Motorist Substance Abuse     object
         Person ID                        object
         Driver At Fault                  object
         Injury Severity                  object
         Circumstance                     object
         Driver Distracted By             object
         Drivers License State            object
         Vehicle ID                       object
         Vehicle Damage Extent            object
         Vehicle First Impact Location    object
         Vehicle Second Impact Location   object
         Vehicle Body Type                object
         Vehicle Movement                 object
         Vehicle Continuing Dir           object
         Vehicle Going Dir                object
         Speed Limit                       int64
         Driverless Vehicle               object
         Parked Vehicle                   object
         Vehicle Year                      int64
         Vehicle Make                     object
         Vehicle Model                    object
         Equipment Problems               object
         Latitude                        float64
         Longitude                       float64
         Location                         object
         dtype: object
```

	A	B	C	D	E	F	G
1	Report Number	Local Case Number	Agency Name	ACRS Report Type	Crash Date/Time	Route Type	Road Name
2	MCP3040003N	190026050	Montgomery County Police	Property Damage Crash	5/31/2019 15:00		
3	EJ78850038	230034791	Gaithersburg Police Depar	Property Damage Crash	7/21/2023 17:59	Maryland (State)	FREDERICK RD
4	MCP2009002G	230034583	Montgomery County Police	Property Damage Crash	7/20/2023 15:10	Maryland (State)	GEORGIA AVE
5	MCP3201004C	230035036	Montgomery County Police	Property Damage Crash	7/23/2023 12:10	County	CRYSTAL ROCK DR
6	MCP23290028	230035152	Montgomery County Police	Property Damage Crash	7/24/2023 6:10	County	MONTGOMERY VILLAGE AVE
7	MCP295200DV	230032956	Montgomery County Police	Property Damage Crash	7/11/2023 7:40	County	WAYNE AVE
8	MCP33510013	230033282	Montgomery County Police	Property Damage Crash	7/12/2023 20:28	Maryland (State)	COLESVILLE RD
9	EJ7869003F	230032124	Gaithersburg Police Depar	Injury Crash	7/5/2023 23:25	Maryland (State)	CLOPPER RD
10	MCP3244002K	230034697	Montgomery County Police	Property Damage Crash	7/21/2023 7:14	US (State)	GEORGIA AVE
11	MCP2863002V	230034445	Montgomery County Police	Property Damage Crash	7/19/2023 19:00	Maryland (State)	WOODFIELD RD
12	MCP2456007L	230034690	Montgomery County Police	Property Damage Crash	7/20/2023 17:00	County	OLD COLUMBIA PIKE
13	MCP2009002G	230034583	Montgomery County Police	Property Damage Crash	7/20/2023 15:10	Maryland (State)	GEORGIA AVE
14	MCP9365001V	230030221	Montgomery County Police	Injury Crash	6/24/2023 12:39	Maryland (State)	SANDY SPRING RD
15	EJ78860034	230034298	Gaithersburg Police Depar	Property Damage Crash	6/19/2023 22:19		

Raw Data Summary

- No quantitative data
- 43 columns
- Car crash date/time is in the same cell

DATA EXPLORING – A MULTIPLE PIVOT TABLES

Row Labels	Count of Report Number
2015	20286
2016	21778
2017	21541
2018	21042
2019	20941
2020	13809
2021	16208
2022	17582
2023	18916
2024	2
Grand Total	172105

Collision Type	Count of Report Number
SAME DIR REAR END	55755
STRAIGHT MOVEMENT ANGLE	30340
OTHER	19030
SAME DIRECTION SIDESWIPE	16226
SINGLE VEHICLE	15869
HEAD ON LEFT TURN	12926
SAME DIRECTION RIGHT TUR	3832
HEAD ON	3786
SAME DIRECTION LEFT TURN	3715
OPPOSITE DIRECTION SIDES\	2883
ANGLE MEETS LEFT TURN	2033
ANGLE MEETS RIGHT TURN	1204
SAME DIR REND LEFT TURN	743
SAME DIR REND RIGHT TURN	720
SAME DIR BOTH LEFT TURN	719
UNKNOWN	717
ANGLE MEETS LEFT HEAD ON	700
N/A	585
OPPOSITE DIR BOTH LEFT TUF	322
Grand Total	172105

Municipality	Count of Report Number
N/A	137042
(blank)	15937
ROCKVILLE	9320
GAITHERSBURG	6437
TAKOMA PARK	1685
KENSINGTON	413
CHEVY CHASE #4	325
CHEVY CHASE #3	156
FRIENDSHIP HEIGHTS	133
POOLESVILLE	108
CHEVY CHASE VIEW	93
CHEVY CHASE VILLAGE	87
CHEVY CHASE #5	74
SOMERSET	53
GARRETT PARK	53
WASHINGTON GROVE	41
NORTH CHEVY CHASE	34
GLEN ECHO	33
MATINS ADDITION	28
LAYTONSVILLE	25
BROOKEVILLE	16
DRUMMOND	12
Grand Total	172105

Light	Count of Report Number
DAYLIGHT	116140
DARK LIGHTS ON	39549
DARK NO LIGHTS	4967
DUSK	3935
DAWN	3472
DARK -- UNKNOWN LIGHTI	1579
N/A	1445
UNKNOWN	676
OTHER	342
Grand Total	172105

Weather	Count of Report Number
CLEAR	116774
RAINING	20608
CLOUDY	17294
N/A	13354
SNOW	1444
UNKNOWN	692
FOGGY	676
WINTRY MIX	391
OTHER	370
SLEET	218
SEVERE WINDS	154
BLOWING SNOW	115
BLOWING SAND, SOIL, DIRT	15
Grand Total	172105

Row Labels	Count of Report Number
Oct	16596
Dec	15745
Nov	15561
Sep	15304
May	14905
Jan	13971
Jun	13916
Jul	13731
Aug	13693
Mar	13320
Apr	12846
Feb	12517
Grand Total	172105

Surface Condition	Count of Report Number
DRY	120569
WET	28426
(blank)	15935
N/A	4183
ICE	1058
SNOW	938
UNKNOWN	509
SLUSH	204
OTHER	164
MUD, DIRT, GRAVEL	46
WATER(STANDING/M	41
OIL	28
SAND	4
Grand Total	172105

Traffic Control	Count of Report Number
NO CONTROLS	68626
TRAFFIC SIGNAL	58892
N/A	25469
STOP SIGN	12488
FLASHING TRAFFIC SIGNAL	2132
OTHER	1991
YIELD SIGN	1727
UNKNOWN	284
PERSON	282
WARNING SIGN	154
RAILWAY CROSSING DEVICE	44
SCHOOL ZONE SIGN DEVICE	16
Grand Total	172105

Driver Substance Abuse	Count of Report Number
NONE DETECTED	122532
N/A	31324
UNKNOWN	11993
ALCOHOL PRESENT	4084
ALCOHOL CONTRIBUTED	1435
ILLEGAL DRUG PRESENT	258
MEDICATION PRESENT	117
ILLEGAL DRUG CONTRIBUTED	102
COMBINED SUBSTANCE PRESI	92
MEDICATION CONTRIBUTED	64
OTHER	58
COMBINATION CONTRIBUTED	46
Grand Total	172105

Route Type	Count of Report Number
Maryland (State)	77074
County	55568
(blank)	16973
Municipality	9379
US (State)	7567
Interstate (State)	3149
Other Public Roadway	1129
Government	627
Ramp	579
Service Road	40
Unknown	20
Grand Total	172105

Related Non-Motorist	Count of Report Number
(blank)	166642
PEDESTRIAN	3880
BICYCLIST	1169
OTHER	246
OTHER CONVEYANCE	84
MACHINE OPERATOR/RIDEI	39
OTHER PEDALCYCLIST	26
OTHER, PEDESTRIAN	9
BICYCLIST, OTHER	4
BICYCLIST, PEDESTRIAN	3
IN ANIMAL-DRAWN VEH	1
OTHER CONVEYANCE, PEDI	1
OTHER, OTHER CONVEYANC	1
Grand Total	172105

Surface Condition	Count of Report Number
PASSENGER CAR	119102
(SPORT) UTILITY VEHI	15991
PICKUP TRUCK	6786
VAN	4957
TRANSIT BUS	3642
SCHOOL BUS	2975
(blank)	2470
POLICE VEHICLE/NOI	2116
OTHER LIGHT TRUCK	1909
CARGO VAN/LIGHT TF	1858
OTHER	1539
POLICE VEHICLE/EME	1498
MEDIUM/HEAVY TRUC	1469

Traffic Control	Count of Report Number
TOYOTA	23171
HONDA	18870
FORD	17138
TOYT	8840
NISSAN	8525
HOND	5765
DODGE	4488
HYUNDAI	3935
CHEVROLET	3796
JEEP	3769
UNKNOWN	3705
CHEV	3670
BMW	3310

Driver Substance Abuse	Count of Report Number
35	50479
40	33589
25	24045
30	23377
45	12454
15	6125
0	4749
50	4669
5	4213
55	3978
10	3094
20	1192
60	76

Route Type	Count of Report Number
NO MISUSE	123464
N/A	34141
UNKNOWN	14148
AIR BAG FAILED	174
OTHER	112
BELT(S) MISUSED	34
BELTS/ANCHORS BROKE	12
STRAP/TETHER LOOSE	8
FACING WRONG WAY	6
NOT STREPPED RIGHT	4
SIZE/TYPE IMPROPER	2
Grand Total	172105

Related Non-Motorist	Count of Report Number
2015	11861
2014	11632
2016	11207
2013	10937
2012	9337
2017	9093
2011	8368
2007	7926
2008	7906
2010	7430
2006	7413
2009	6783
2018	6738

Data Cleaning – Removed Not available rows

In [3]:
```
# show what columns the dataframe has
car_crash_df.columns
```

Out[3]:
```
Index(['Report Number', 'Local Case Number', 'Agency Name', 'ACRS Report Type',
       'Crash Date/Time', 'Route Type', 'Road Name', 'Cross-Street Type',
       'Cross-Street Name', 'Off-Road Description', 'Municipality',
       'Related Non-Motorist', 'Collision Type', 'Weather',
       'Surface Condition', 'Light', 'Traffic Control',
       'Driver Substance Abuse', 'Non-Motorist Substance Abuse', 'Person ID',
       'Driver At Fault', 'Injury Severity', 'Circumstance',
       'Driver Distracted By', 'Drivers License State', 'Vehicle ID',
       'Vehicle Damage Extent', 'Vehicle First Impact Location',
       'Vehicle Second Impact Location', 'Vehicle Body Type',
       'Vehicle Movement', 'Vehicle Continuing Dir', 'Vehicle Going Dir',
       'Speed Limit', 'Driverless Vehicle', 'Parked Vehicle', 'Vehicle Year',
       'Vehicle Make', 'Vehicle Model', 'Equipment Problems', 'Latitude',
       'Longitude', 'Location'],
      dtype='object')
```

In [5]:
```
# Check the total number of Municipality records
car_crash_df["Report Number"].count()
```

Out[5]:
```
172105
```

In [6]:
```
# Remove rows where Municipality is "N/A"
car_crash_df = car_crash_df[car_crash_df["Municipality"] != "N/A"]

# Remove rows where Municipality is NaN
car_crash_df = car_crash_df.dropna(subset=["Municipality"])

# Check data cleaning process has been properly applied
car_crash_df["Report Number"].count()
```

Out[6]:
```
19126
```

DATA TRANSFORMATION

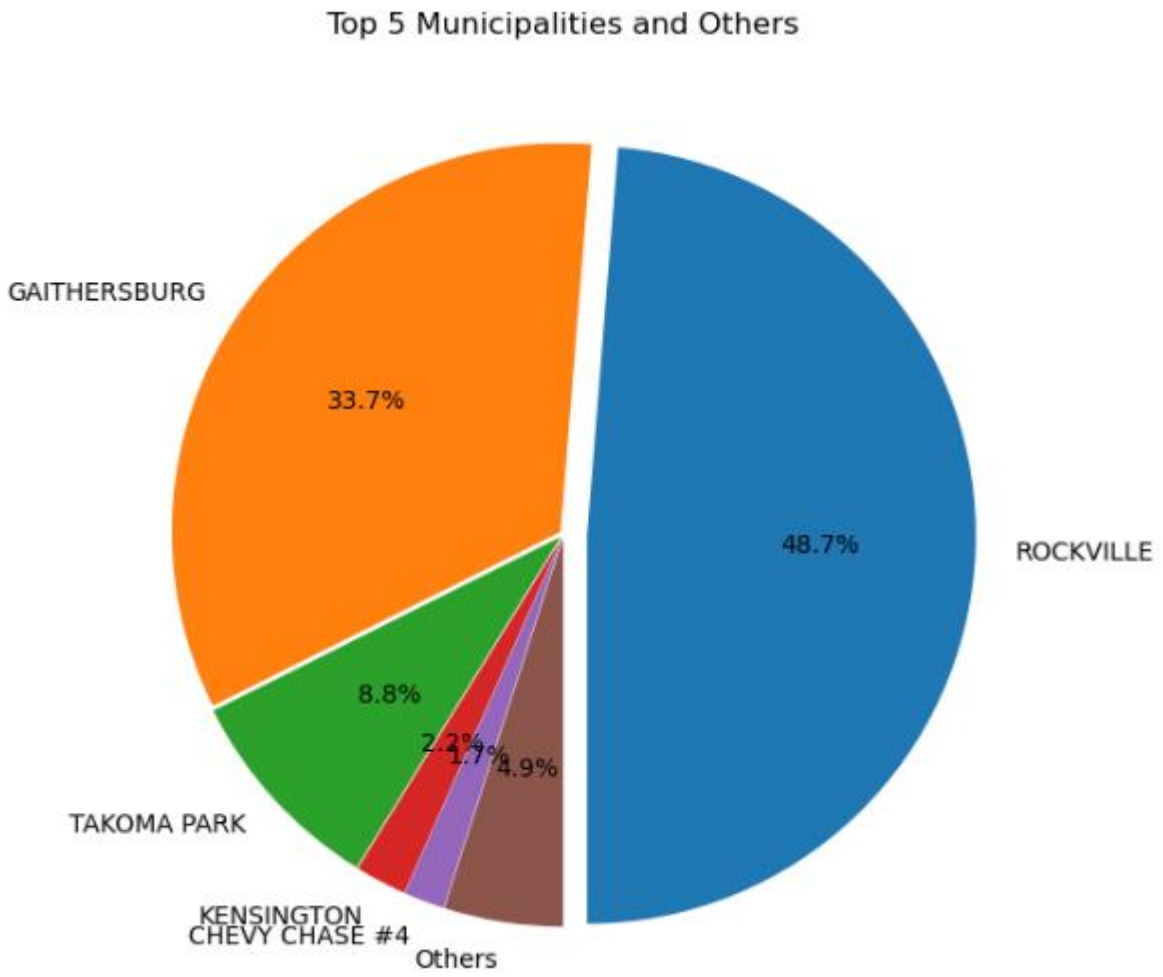
	A	B	C	D	E	F	G
1	Report Number	Local Case Number	Agency Name	ACRS Report Type	Crash Date/Time	Route Type	Road Name
2	MCP3040003N	190026050	Montgomery County Police	Property Damage Crash	5/31/2019 15:00		
3	EJ78850038	230034791	Gaithersburg Police Depar	Property Damage Crash	7/21/2023 17:59	Maryland (State)	FREDERICK RD
4	MCP2009002G	230034583	Montgomery County Police	Property Damage Crash	7/20/2023 15:10	Maryland (State)	GEORGIA AVE
5	MCP3201004C	230035036	Montgomery County Police	Property Damage Crash	7/23/2023 12:10	County	CRYSTAL ROCK DR
6	MCP23290028	230035152	Montgomery County Police	Property Damage Crash	7/24/2023 6:10	County	MONTGOMERY VILLAGE AVE
7	MCP295200DV	230032956	Montgomery County Police	Property Damage Crash	7/11/2023 7:40	County	WAYNE AVE
8	MCP33510013	230033282	Montgomery County Police	Property Damage Crash	7/12/2023 20:28	Maryland (State)	COLESVILLE RD
9	EJ7869003F	230032124	Gaithersburg Police Depar	Injury Crash	7/5/2023 23:25	Maryland (State)	CLOPPER RD
10	MCP3244002K	230034697	Montgomery County Police	Property Damage Crash	7/21/2023 7:14	US (State)	GEORGIA AVE
11	MCP2863002V	230034445	Montgomery County Police	Property Damage Crash	7/19/2023 19:00	Maryland (State)	WOODFIELD RD
12	MCP2456007L	230034690	Montgomery County Police	Property Damage Crash	7/20/2023 17:00	County	OLD COLUMBIA PIKE
13	MCP2009002G	230034583	Montgomery County Police	Property Damage Crash	7/20/2023 15:10	Maryland (State)	GEORGIA AVE
14	MCP9365001V	230030221	Montgomery County Police	Injury Crash	6/24/2023 12:39	Maryland (State)	SANDY SPRING RD
15	EJ78860034	230034298	Gaithersburg Police Depar	Property Damage Crash	6/19/2023 22:19		

```
rame
 = car_crash_df[selected_columns]
[['Date', 'Time']] = reduced_car_crash_df['Crash Date/Time'].str.split(' ', expand=True, n=1)

o military format
['Time'] = pd.to_datetime(reduced_car_crash_df['Time'], format='%I:%M:%S %p').dt.strftime('%H:%M:%S')

"Crash Date/Time" and "Time" columns if needed
.drop(columns=['Crash Date/Time'], inplace=True)

['Date'] = pd.to_datetime(reduced_car_crash_df['Date'], format='%m/%d/%Y')
['Year'] = reduced_car_crash_df['Date'].dt.year
['Month'] = reduced_car_crash_df['Date'].dt.month
['Day'] = reduced_car_crash_df['Date'].dt.day
.head()
```

treet ame	Collision Type	Municipality	Weather	Surface Condition	Traffic Control	Speed Limit	Latitude	Longitude	Date	Time	Year	Month	Day
INCE ARD BLVD	STRAIGHT MOVEMENT ANGLE	GAITHERSBURG	CLEAR	DRY	STOP SIGN	30	39.141990	-77.224371	2023-07-20	14:48:00	2023	7	20
T AVE	SAME DIR REAR END	CHEVY CHASE #4	CLEAR	DRY	NO CONTROLS	25	38.977186	-77.088324	2023-07-27	16:14:00	2023	7	27
N LA	SAME DIRECTION RIGHT TURN	TAKOMA PARK	CLEAR	DRY	NaN	35	38.985947	-76.988159	2023-07-22	17:10:00	2023	7	22
R AVE	SAME DIR REAR END	GAITHERSBURG	CLEAR	DRY	NO CONTROLS	25	39.133463	-77.203386	2023-07-15	10:47:00	2023	7	15
W STON DR	STRAIGHT MOVEMENT ANGLE	ROCKVILLE	NaN	DRY	TRAFFIC SIGNAL	40	39.074431	-77.135645	2023-07-22	04:07:00	2023	7	22

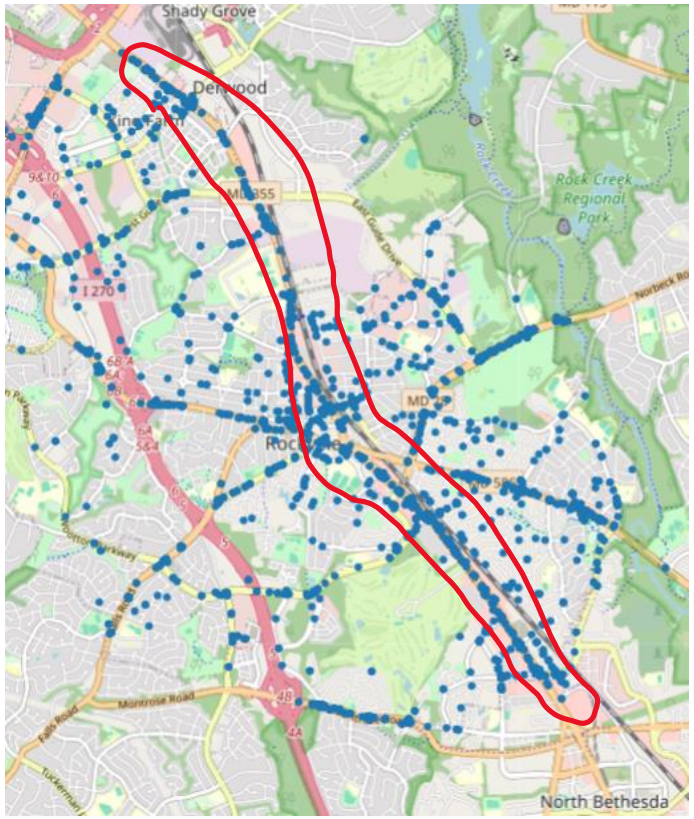
ANALYSIS

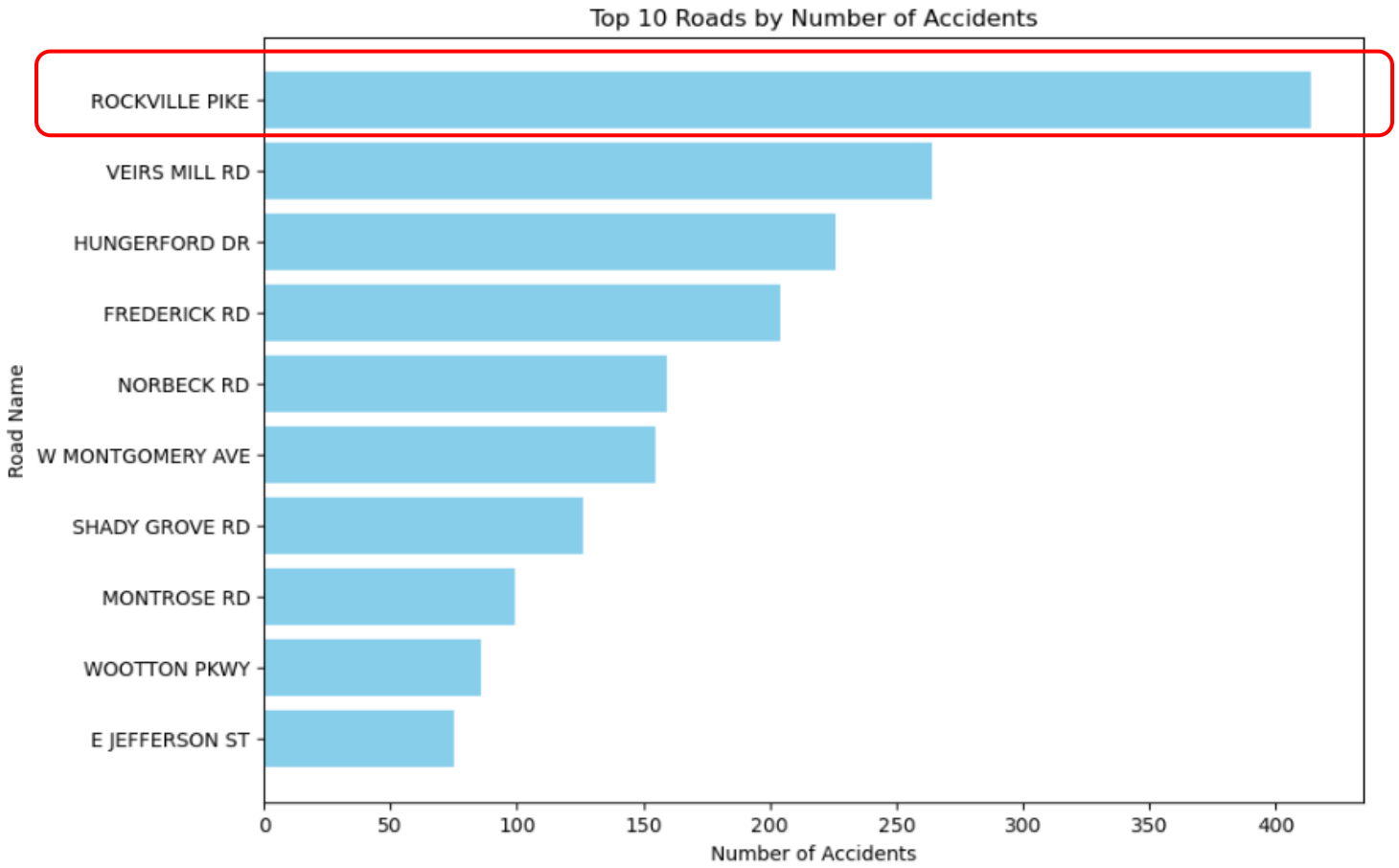
EXPLORING DATA



- ROCKVILLE Municipality has the highest number of car accidents in Montgomery county

EXPLORING DATA

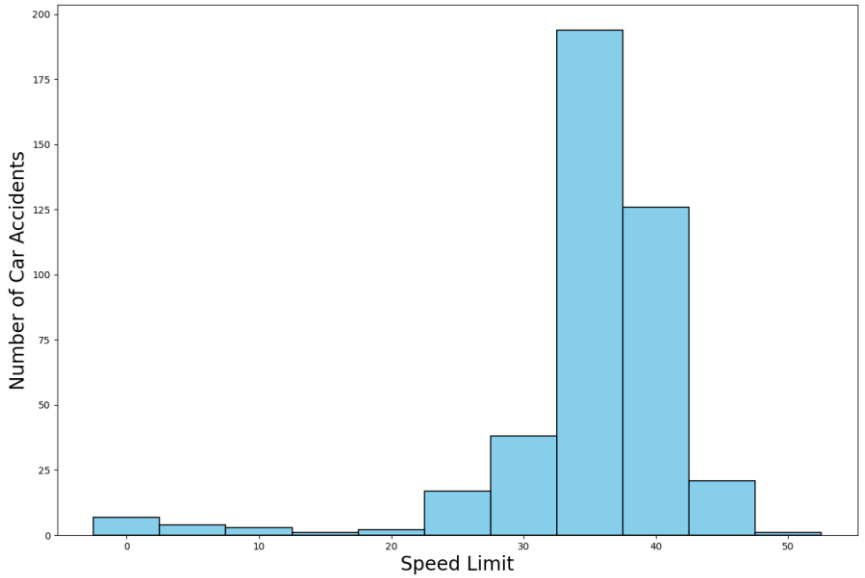




- Used Latitude and Longitude in the dataset to draw the map plot to see where the car accidents most happen
- Aggregated the number of car crash accidents by road name in Rockville Municipality

ROCKVILLE DATA SET – BY SPEED LIMIT

	Speed Limit	num_of_accident
0	35	194
1	40	126
2	30	38
3	45	21
4	25	17
5	0	7
6	5	4
7	10	3
8	20	2
9	15	1
10	50	1

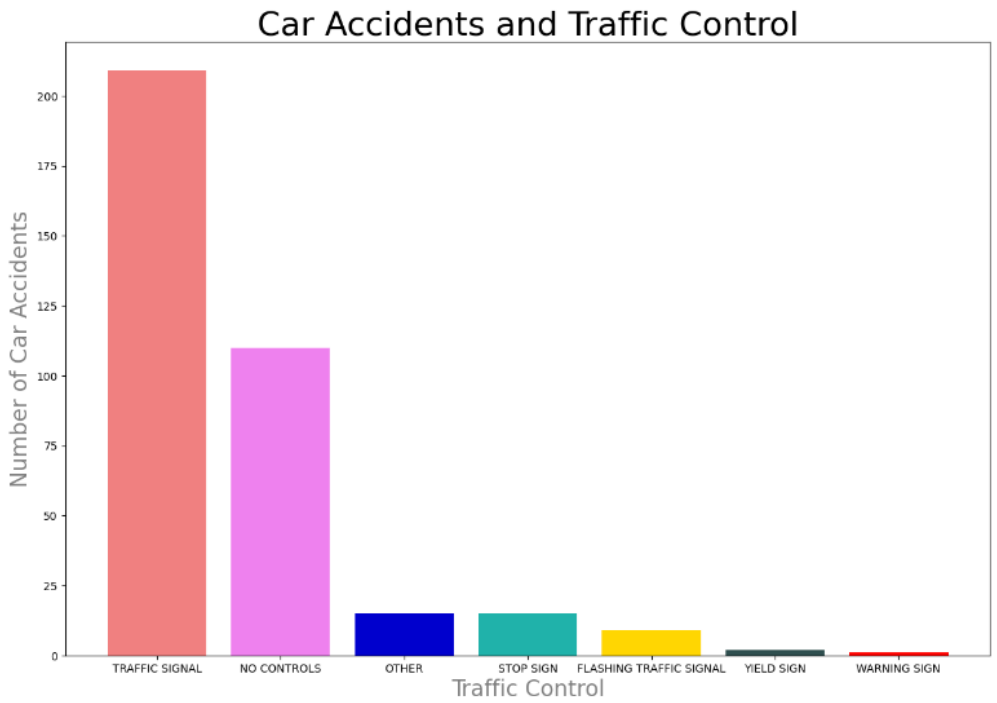


Observation:
- 80% of the number of accidents were occurred between 35 and 40 limit.

Recommendation:
- For car accidents that occur on roads with speed limit 35 and 40 mph, One suggestion is to reduce the speed limit, but further study needs to be done to see which roads the accidents occurred, to tailor the response and take into account congestion times.

ROCKVILLE DATA SET – BY TRAFFIC CONTROL

	Traffic Control	num_of_accident
0	TRAFFIC SIGNAL	209
1	NO CONTROLS	110
2	OTHER	15
3	STOP SIGN	15
4	FLASHING TRAFFIC SIGNAL	9
5	YIELD SIGN	2
6	WARNING SIGN	1



Observation:

- Car accident near Traffic Signal was the number one reason among other traffic controls

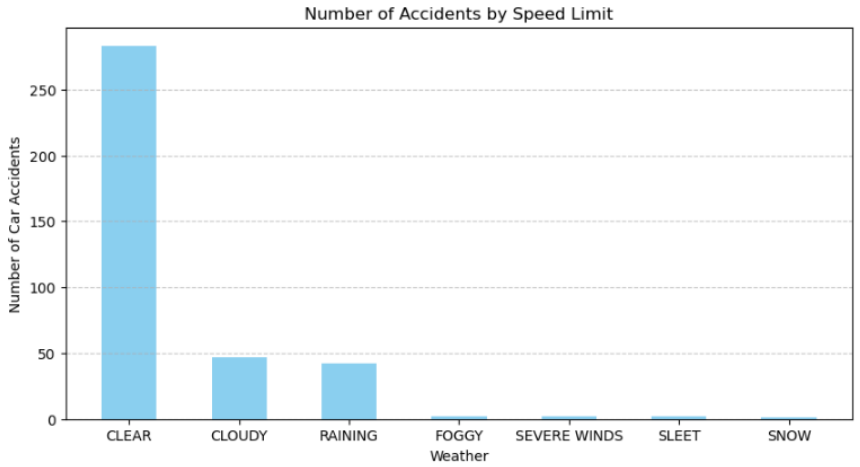
Recommendation:

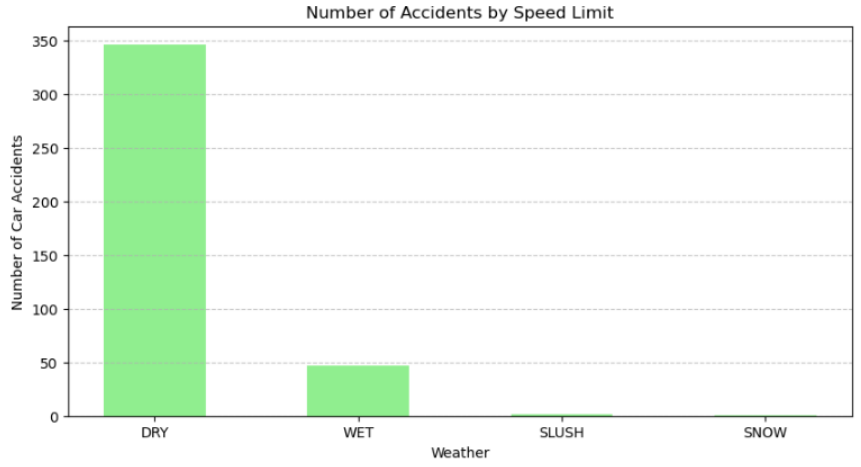
- The road has higher speed limit and normally drivers get into the car accident when they try to pass the yellow light. With that said, shorting the yellow light time would help improve the number of car accident

ROCKVILLE DATA SET – BY WEATHER & SURFACE CONDITION

	Weather	num_of_accident
0	CLEAR	283
1	CLOUDY	47
2	RAINING	42
3	FOGGY	2
4	SEVERE WINDS	2
5	SLEET	2
6	SNOW	1

	Surface Condition	num_of_accident
0	DRY	346
1	WET	47
2	SLUSH	2
3	SNOW	1



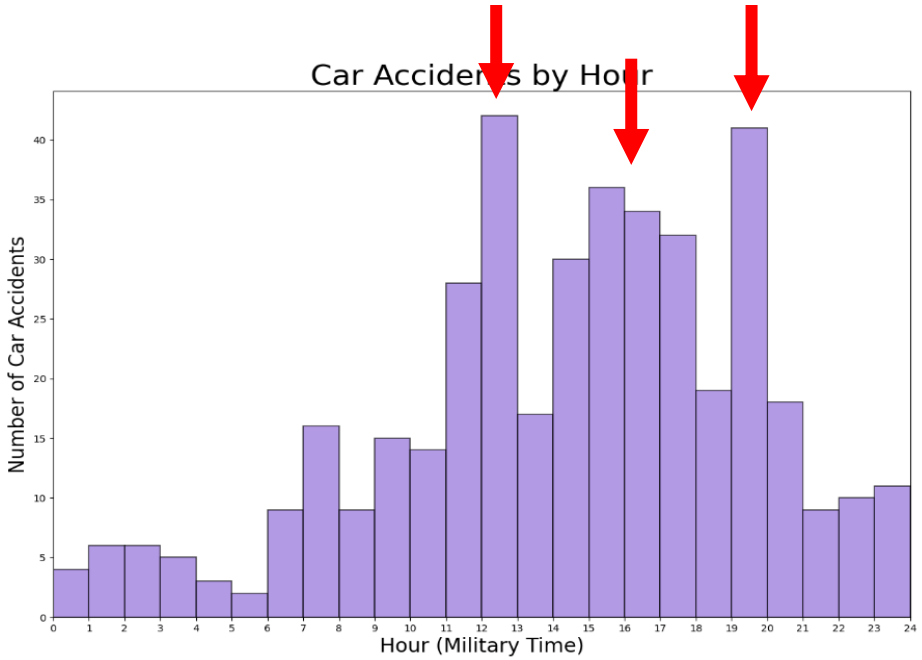


Observation:

- **Can't** find any correlation with the number of car accidents

ROCKVILLE DATA SET – BY HOURLY INTERVALS

Hour Interval	num_of_accident
00:00:00 - 00:59:59	2
01:00:00 - 01:59:59	6
02:00:00 - 02:59:59	6
03:00:00 - 03:59:59	5
04:00:00 - 04:59:59	3
05:00:00 - 05:59:59	2
06:00:00 - 06:59:59	9
07:00:00 - 07:59:59	16
08:00:00 - 08:59:59	9
09:00:00 - 09:59:59	15
10:00:00 - 10:59:59	14
11:00:00 - 11:59:59	28
12:00:00 - 12:59:59	42
13:00:00 - 13:59:59	17
14:00:00 - 14:59:59	30
15:00:00 - 15:59:59	36
16:00:00 - 16:59:59	34
17:00:00 - 17:59:59	32
18:00:00 - 18:59:59	19
19:00:00 - 19:59:59	41
20:00:00 - 20:59:59	18
21:00:00 - 21:59:59	9
22:00:00 - 22:59:59	10
23:00:00 - 23:59:59	11



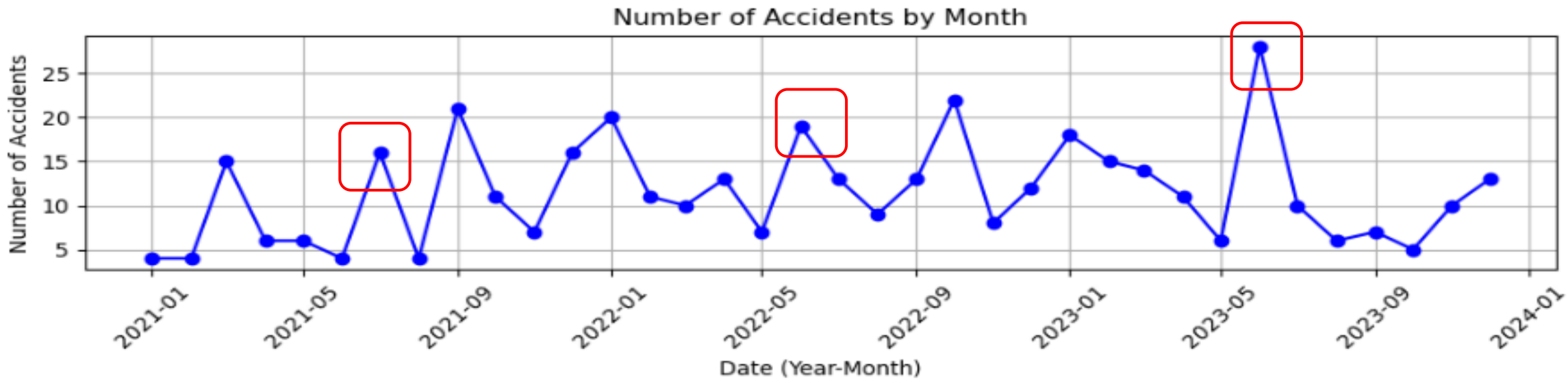
Observation:

- The number of car accidents were higher than another hour interval at between "12:00 PM and 1:00PM", "02:00PM and "06:00PM" and "07:00 PM and 08:00PM".

Recommendation:

- Place more traffic polices on the ROCKVILLE PIKE on the time intervals have the most car accidents

ROCKVILLE DATA SET – TIME SERIES



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
YearMonth	2021-01-01	2021-02-01	2021-03-01	2021-04-01	2021-05-01	2021-06-01	2021-07-01	2021-08-01	2021-09-01	2021-10-01	2021-11-01	2021-12-01	2022-01-01	2022-02-01	2022-03-01	2022-04-01	2022-05-01	2022-06-01	2022-07-01	2022-08-01	2022-09-01	2022-10-01	2022-11-01	2022-12-01	2023-01-01	2023-02-01	2023-03-01	2023-04-01	2023-05-01	2023-06-01	2023-07-01	2023-08-01	2023-09-01	2023-10-01	2023-11-01	2023-12-01
num_of_accident	4	4	15	6	6	4	16	4	21	11	7	16	20	11	10	13	7	19	13	9	13	22	8	12	18	15	14	11	6	28	10	6	7	5	10	13

Observation:

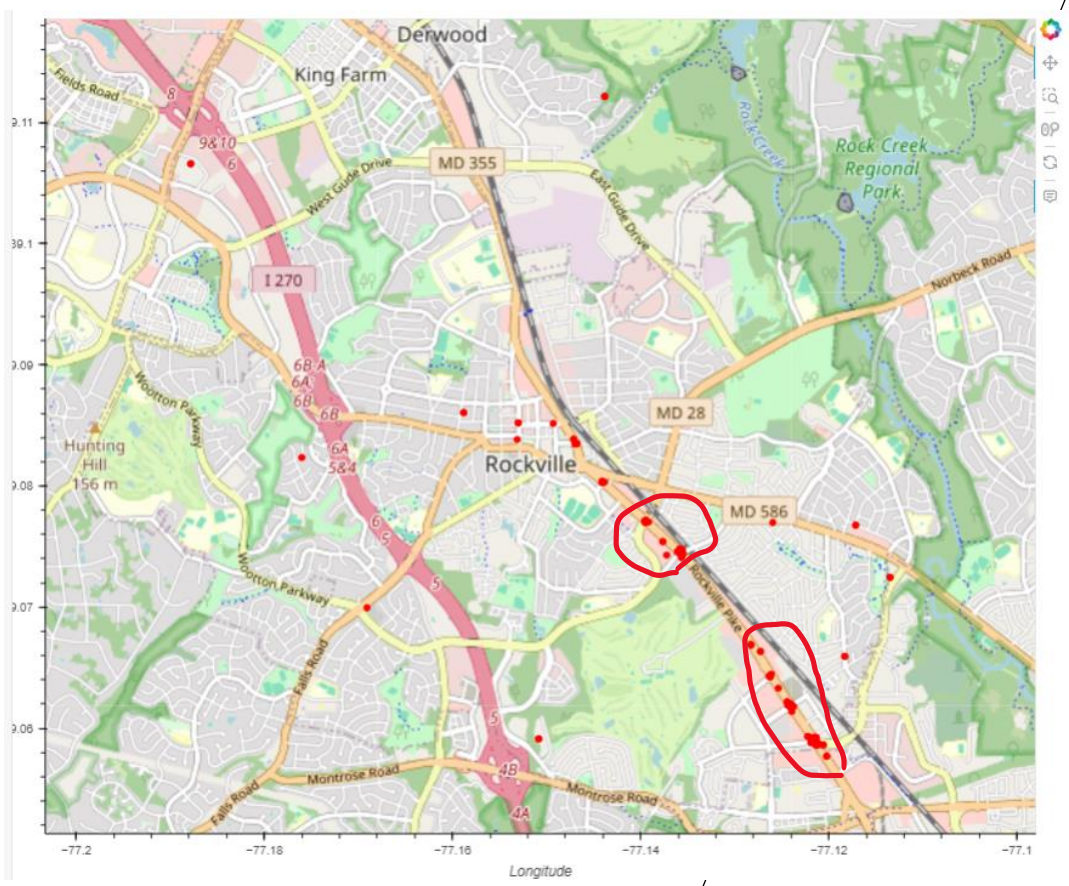
- Every year June had high increase of car accidents
- Research shows that the traffic congestion is getting higher due to the increase of population (https://empowermontgomery.com/transit-power/)

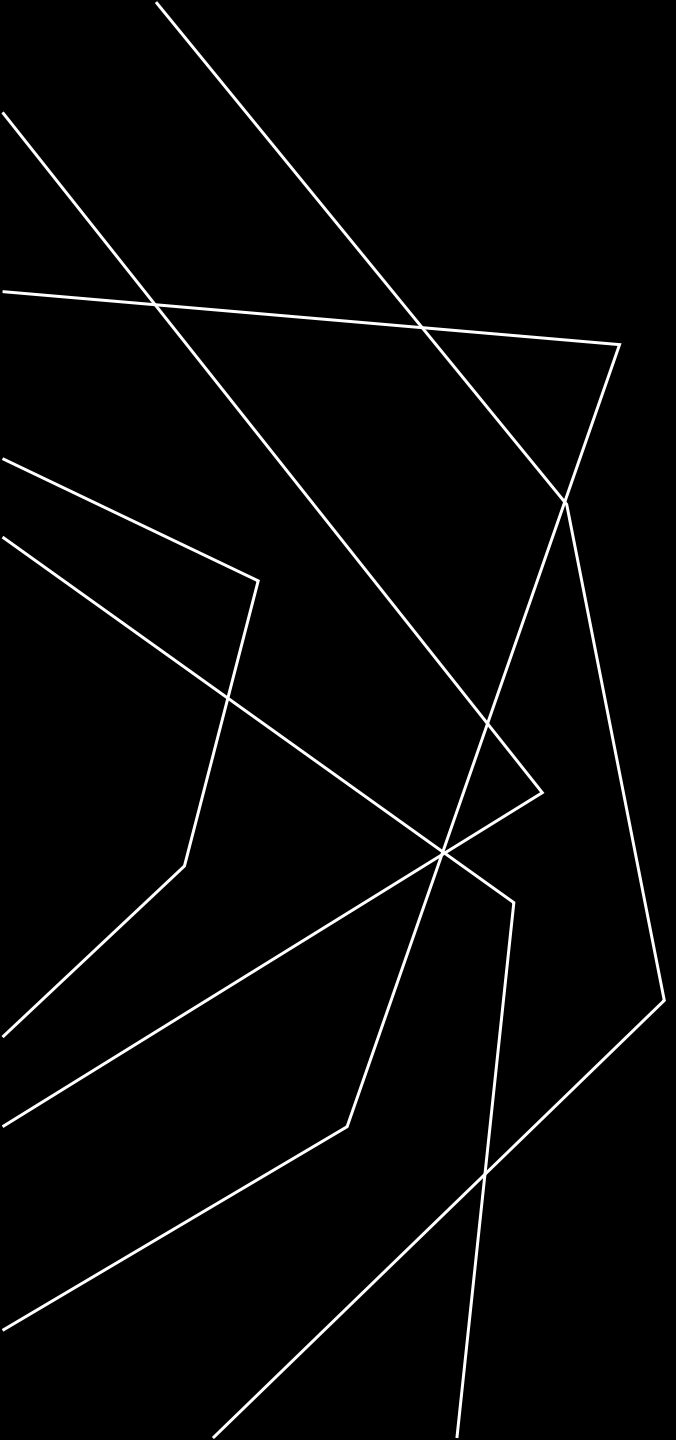
Recommendation:

- Expand the highway right next to Rockville Pike or put more traffic police in June to control the traffic

CONCLUSION

- **We noticed that the two main perspectives could narrow down the location requires improvement**
- **With that filter, 155 car accidents records found.**
- **Final Recommendation to local government is lower the speed limit on the spotted area in the map or control traffic signal light**





THANK YOU