

# Traffic Collisions in North Hollywood

By: Andrea Rosenberger, Stephen Yang & Zoey Charif



# Project Overview

Our project seeks to uncover patterns among traffic accidents in Los Angeles (North Hollywood), utilizing traffic collision data provided by the LAPD.

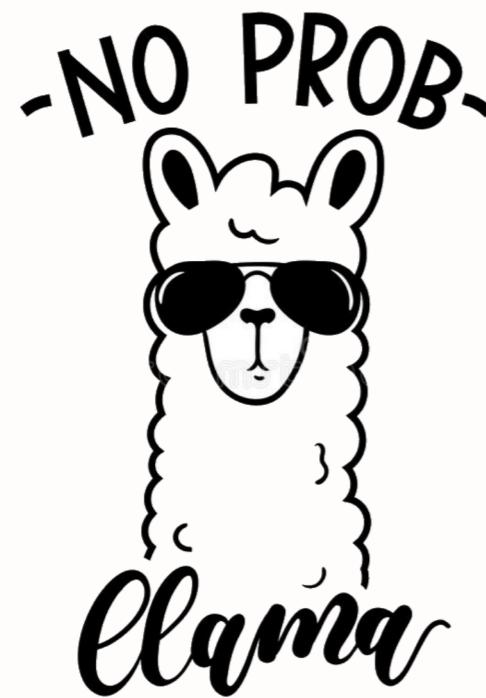
Our mission is to provide:

- 1) Collision frequency per location,
- 3) Collision analytics for holidays
- 4) Collision type analytics

We will present statistics, charts, maps to illustrate our results.

Our study may be useful to city officials, planners, and engineers to design and implement infrastructural changes to reduce or eliminate future accidents.

# Our Process



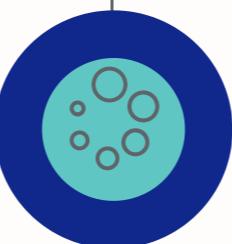
Step 1



## Where we found our data

We found a study of LA Traffic Collisions on Kaggle.  
We traced it to their source - the Los Angeles Data Catalogue: <https://data.lacity.org/>

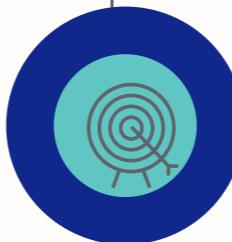
Step 2



## How we explored our data

We exported it to CSV and reviewed in Excel using pivot tables.

Step 3



## Problems we aim to solve

We seek to provide data analytics to help identify locations which require review of infrastructure or police presence to reduce or eliminate future accidents.

# CSV clean-up

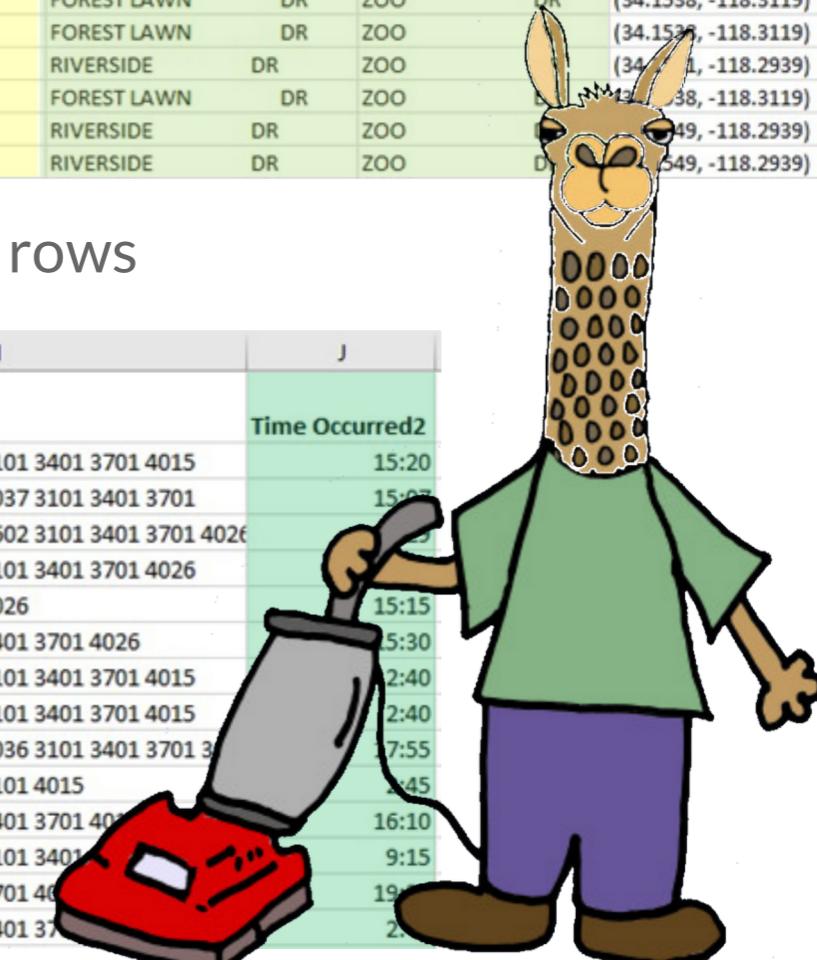
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	DR Number	Date Reported	Date Occurred	Time Occurred	Area ID	Area Name	Reporting District	Crime Code	Crime Description	MO Codes	Victim Age	Victim Sex	Victim Descent	Premise Code	Premise Description	Address	Cross Street	Location
2	110514304	7/21/2011	7/21/2011	215	5	Harbor	551	997 TRAFFIC COLLISION			21	F	O	101	STREET	1ST ST	ZORANA RD	PL (33.7441, -118.3127)
3	110612341	4/28/2011	4/28/2011	1745	6	Hollywood	622	997 TRAFFIC COLLISION			M	O		101	STREET	WILLOW GLEN RD	ZORADA DR	(34.1185, -118.3655)
4	180616035	7/9/2018	7/9/2018	1710	6	Hollywood	622	997 TRAFFIC COLLISION	3011 3026 3035 3037 3102 3401 370		17			101	STREET	NICHOLS CANYON RD	ZORADA CT	(34.1146, -118.3607)
5	201107508	3/19/2020	3/19/2020	400	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 0605 3034 3011		M	H		101	STREET	WESTERN HERI WY	ZOO DR	(34.1374, -118.2941)
6	201107183	3/11/2020	3/11/2020	900	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 0605 3004 3034		64	M	W	101	STREET	RIVERSIDE DR	ZOO DR	(34.1551, -118.2939)
7	191119563	11/16/2019	11/9/2019	700	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 3004 3026 3036 060		47	M	H	101	STREET	RIVERSIDE	ZOO DR	(34.1551, -118.2939)
8	191518805	10/13/2019	10/11/2019	2145	15	N Hollywood	1567	997 TRAFFIC COLLISION	3004 3037 3029 3026 3101 3401 370		22	M	H	101	STREET	FOREST LAWN DR	ZOO DR	(34.1499, -118.3635)
9	191518134	9/28/2019	9/28/2019	1710	15	N Hollywood	1589	997 TRAFFIC COLLISION	4026 3004 3030 3028 3033 3037 310		28	M	A	101	STREET	FOREST LAWN DR	ZOO DR	(34.1524, -118.3179)
10	191115752	8/28/2019	8/28/2019	1110	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 3004 3002 3034		46	F	B	101	STREET	FOREST LAWN DR	ZOO DR	(34.1538, -118.3119)
11	191115674	8/21/2019	8/21/2019	2000	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 3004 3030 0605		58	F	W	101	STREET	FOREST LAWN DR	ZOO DR	(34.1538, -118.3119)
12	191114437	7/29/2019	7/29/2019	1430	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 3004 3026		23	M	H	101	STREET	RIVERSIDE DR	ZOO DR	(34.1538, -118.2939)
13	191113659	7/16/2019	7/16/2019	1600	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 3004 3026 3036		45	M	F	101	STREET	FOREST LAWN DR	ZOO DR	(34.1538, -118.3119)
14	191112671	6/26/2019	6/26/2019	1405	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 3004 3026 3036		23	F	W	101	STREET	RIVERSIDE DR	ZOO DR	(34.1538, -118.2939)
15	191109182	4/18/2019	4/18/2019	1715	11	Northeast	1101	997 TRAFFIC COLLISION	3101 3401 3701 3008 3025		29	F	W	101	STREET	RIVERSIDE DR	ZOO DR	(34.1549, -118.2939)

18 cols x 21906 rows



9 columns x 13070 rows

	A	B	C	D	E	F	G	H	I	J
1	DR Number	Date Reported	Date Occurred	Time Occurred	Reporting District	Address	Cross Street	Location	MO Codes	Time Occurred2
2	191519215	10/17/2019	10/17/2019	1520		1526 LANKERSHIM BL	ERWIN ST	(34.1831, -118.3856)	3003 3026 3029 3037 3101 3401 3701 4015	15:20
3	191518670	10/5/2019	10/4/2019	1507		1526 ERWIN ST	LANKERSHIM BL	(34.1831, -118.3878)	4015 3008 3025 3029 3037 3101 3401 3701	15:07
4	191515782	8/17/2019	8/17/2019	1529		1505 TROOST AV	SHERMAN WY	(34.2012, -118.3866)	3034 3006 3037 3028 3602 3101 3401 3701 4026	
5	191515783	8/17/2019	8/17/2019	1530		1505 TROOST AV	SHERMAN WY	(34.2012, -118.3866)	3034 3006 3037 3026 3101 3401 3701 4026	
6	191520926	11/21/2019	11/21/2019	1515		1546 RIVERSIDE DR	BECK AV	(34.1577, -118.3834)	3003 3025 3036 3101 4026	15:15
7	191517096	9/9/2019	9/9/2019	1530		1525 RADFORD AV	GILMORE ST	(34.1875, -118.3921)	3003 3037 3026 3101 3401 3701 4026	15:30
8	191519466	10/24/2019	10/21/2019	740		1506 SATICOY ST	ELMER AV	(34.2085, -118.3769)	3003 3025 3029 3036 3101 3401 3701 4015	2:40
9	191521801	12/11/2019	10/21/2019	740		1506 SATICOY ST	ELMER AV	(34.2085, -118.3769)	3003 3036 3029 3025 3101 3401 3701 4015	2:40
10	191520053	11/4/2019	11/4/2019	655		1504 LANKERSHIM BL	RUNNYMEDE ST	(34.2058, -118.3878)	4026 3008 3025 3035 3036 3101 3401 3701 3	7:55
11	191521912	12/13/2019	12/13/2019	745		1549 CAHUENGA BL	CHANDLER BL	(34.1687, -118.3615)	3003 3025 3029 3036 3101 4015	1:45
12	191508576	3/29/2019	3/28/2019	1610		1524 LAUREL CANYON BL	KITTRIDGE ST	(34.1903, -118.3965)	3036 3003 3025 3101 3401 3701 4015	16:10
13	191520559	11/13/2019	11/13/2019	815		1506 VALERIO ST	VINELAND AV	(34.2048, -118.3703)	3003 3036 3029 3026 3101 3401 3701 4015	9:15
14	191512455	6/13/2019	6/12/2019	1925		1545 RIVERSIDE DR	COLFAX AV	(34.1577, -118.3877)	3004 3036 3026 3101 3701 4015	19:45
15	191518277	10/1/2019	9/26/2019	740		1522 VICTORY BL	COLDWATER CANYON AV	(34.1867, -118.4137)	3008 3025 3036 3101 3401 3701 4015	2:30



# Andrea's Mission

To investigate the total number of collisions over periods of time

- Do some days of the week have higher collisions than others?
- Do some months have higher collision rates than other months?
- Which holidays have higher collision rates than others?

# Process

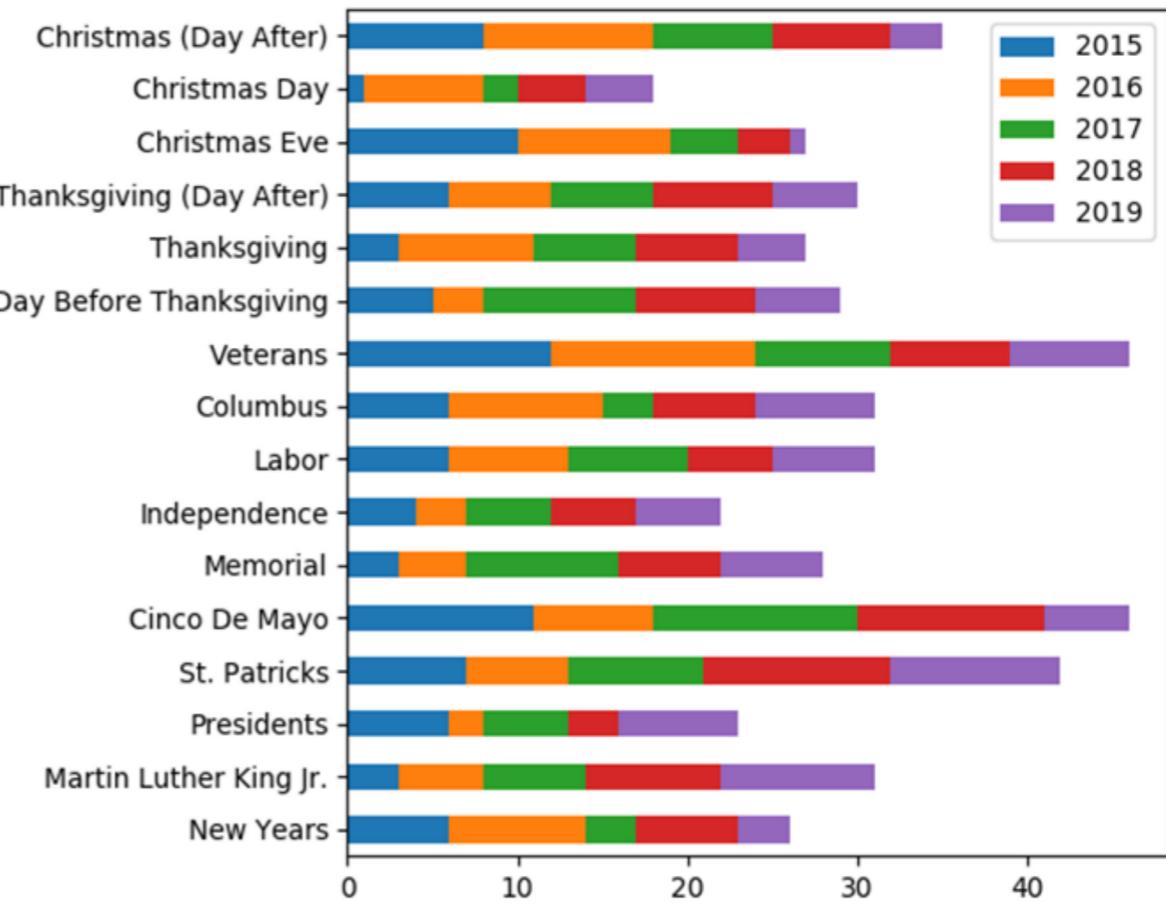
```

1 ##NEW YEAR'S DAY
2 nyd15 = traffic_df2[(traffic_df2["Date Occurred"] == '2015-01-01')]['DR Number'].count()
3 nyd16 = traffic_df2[(traffic_df2["Date Occurred"] == '2016-01-01')]['DR Number'].count()
4 nyd17 = traffic_df2[(traffic_df2["Date Occurred"] == '2017-01-01')]['DR Number'].count()
5 nyd18 = traffic_df2[(traffic_df2["Date Occurred"] == '2018-01-02')]['DR Number'].count()
6 nyd19 = traffic_df2[(traffic_df2["Date Occurred"] == '2019-01-01')]['DR Number'].count()
7 nyd = [nyd15,nyd16,nyd17,nyd18,nyd19]

```



	2015	2016	2017	2018	2019
New Years	6	8	3	6	3
Martin Luther King Jr.	3	5	6	8	9
Presidents	6	2	5	3	7
St. Patricks	7	6	8	11	10
Cinco De Mayo	11	7	12	11	5
Memorial	3	4	9	6	6
Independence	4	3	5	5	5
Labor	6	7	7	5	6
Columbus	6	9	3	6	7
Veterans	12	12	8	7	7
Day Before Thanksgiving	5	3	9	7	5
Thanksgiving	3	8	6	6	4
Thanksgiving (Day After)	6	6	6	7	5
Christmas Eve	10	9	4	3	1
Christmas Day	1	7	2	4	4
Christmas (Day After)	8	10	7	7	3



# Key Findings

According to the data analysis, Veteran's Day and Cinco de Mayo have higher rates of accidents than other holidays in the year.

# Obstacles

Trust, but verify

Reporting Distr	Address	Cross Street	Location
1596 VENTURA	EUREKA	(34.1406, -118.376)	
1596 VENTURA	EUREKA	(34.1406, -118.376)	
1500 VENTURA BL	EUREKA AV	(34.1406, -118.376)	
1585 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1585 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1585 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1585 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1585 VENTURA BL	EUREKA AV	(34.1406, -118.376)	
1585 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	E EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA AV	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA ST	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA AV	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	
1596 VENTURA BL	EUREKA DR	(34.1406, -118.376)	

	A	B	C	D	E	F	G	H	I	J
1	DR Number	Date Reported	Date Occurred	Time Occurred	Reporting District	Address	Cross Street	Location	MO Codes	Time Occurred2
2	191519215	10/17/2019	10/17/2019	1520	1526	LANKERSHIM BL	ERWIN ST	(34.1831, -118.3856)	3003 3026 3029 3037 3101 3401 3701 4015	15:20
3	191518670	10/5/2019	10/4/2019	1507	1526	ERWIN ST	LANKERSHIM BL	(34.1831, -118.3878)	4015 3008 3025 3029 3037 3101 3401 3701	15:07
4	191515782	8/17/2019	8/17/2019	1529	1505	TROOST AV	SHERMAN WY	(34.2012, -118.3866)	3034 3006 3037 3028 3602 3101 3401 3701 4026	15:29
5	191515783	8/17/2019	8/17/2019	1530	1505	TROOST AV	SHERMAN WY	(34.2012, -118.3866)	3034 3006 3037 3026 3101 3401 3701 4026	15:30
6	191520926	11/21/2019	11/21/2019	1515	1546	RIVERSIDE DR	BECK AV	(34.1577, -118.3834)	3003 3025 3036 3101 4026	15:15
7	191517096	9/9/2019	9/9/2019	1530	1525	RADFORD AV	GILMORE ST	(34.1875, -118.3921)	3003 3037 3026 3101 3401 3701 4026	15:30
8	191519466	10/24/2019	10/21/2019	740	1506	SATICOY ST	ELMER AV	(34.2085, -118.3769)	3003 3025 3029 3036 3101 3401 3701 4015	2:40
9	191521801	12/11/2019	10/21/2019	740	1506	SATICOY ST	ELMER AV	(34.2085, -118.3769)	3003 3036 3029 3025 3101 3401 3701 4015	2:40
10	191520053	11/4/2019	11/4/2019	655	1504	LANKERSHIM BL	RUNNymeDE ST	(34.2058, -118.3878)	4026 3008 3025 3035 3036 3101 3401 3701 3033	17:55
11	191521912	12/13/2019	12/13/2019	745	1549	CAHUENGA BL	CHANDLER BL	(34.1687, -118.3615)	3003 3025 3029 3036 3101 4015	2:45
12	191508576	3/29/2019	3/28/2019	1610	1524	LAUREL CANYON BL	KITTRIDGE ST	(34.1903, -118.3965)	3036 3003 3025 3101 3401 3701 4015	16:10



# Stephen's Mission

Mapping incident locations and  
generating a heat map.

# Process

After initial clean-up, mapping of incident locations was to be made.

Conversion of the Location field within the data frame was made to break out to separate Latitude and Longitude coordinates.

The screenshot shows a Jupyter Notebook interface with the title "jupyter Mapping Last Checkpoint: 5 hours ago (autosaved)". The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. Below the menu is a toolbar with various icons. The code editor contains the following Python code:

```
kaw_Lat, kaw_Lon = incidents_at.Location[0].split(',')
print(Raw_Lat)
print(Raw_Lon)
```

Output:

```
(34.1831, -118.3856)
(34.1831
 -118.3856)
```

In [10]: # okay, we've split it in two. Now lets get rid of the parens

```
Lat = Raw_Lat[1:]
print(Lat)
Lon = Raw_Lon[:-1]
print(Lon)
```

Output:

```
34.1831
-118.3856
```

In [11]: float(Lon)

Out[11]: -118.3856

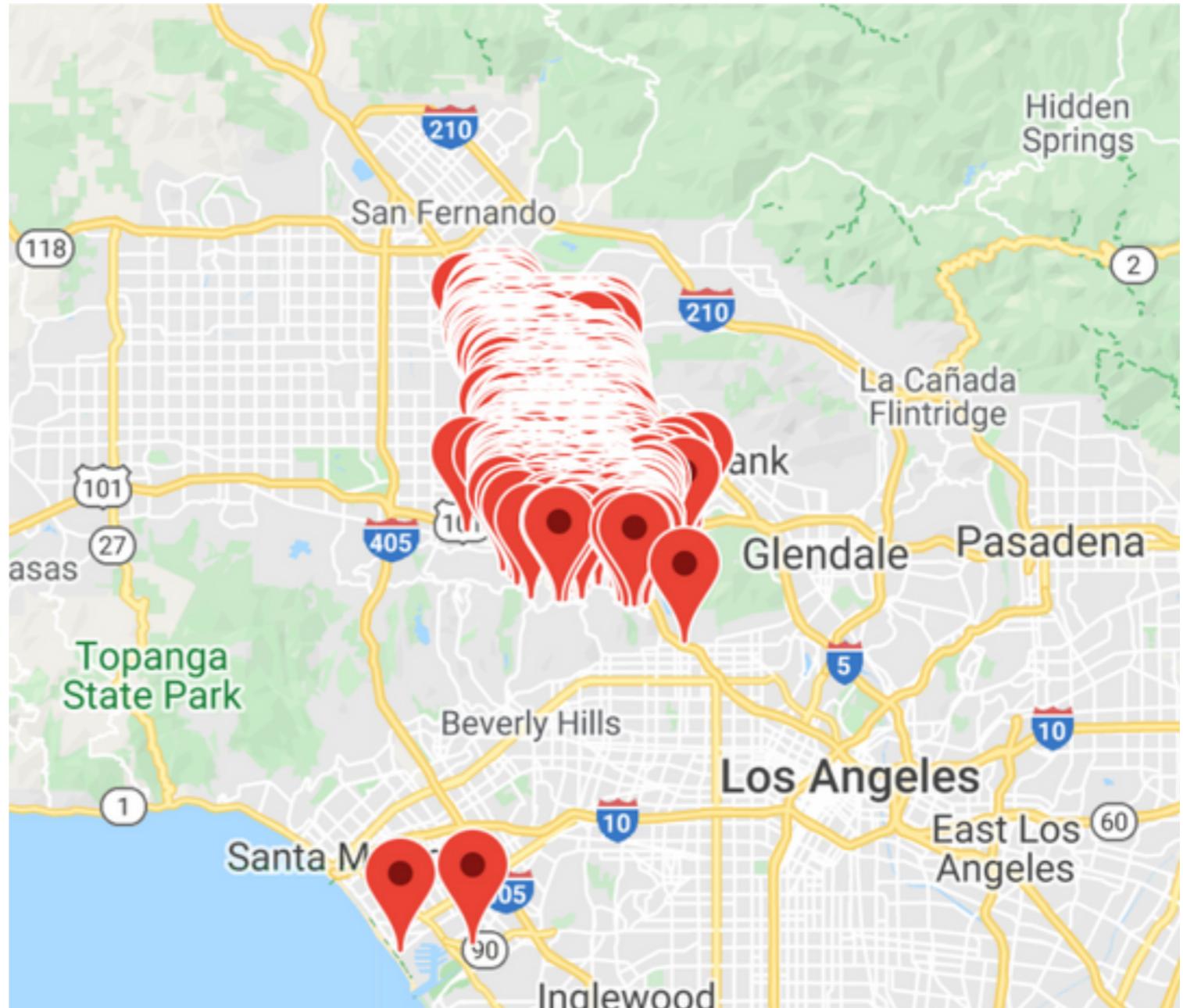
In [12]: # Let's add 2 new columns of map coordinates

```
Incidents_df['Lat'] = 1.1
Incidents_df['Lon'] = 1.1

# And fill them in

for x in range(len(Incidents_df)):
    Raw_Lat, Raw_Lon = Incidents_df.Location[x].split(',')
    Lat = Raw_Lat[1:]
    Lon = Raw_Lon[:-1]
    Incidents_df['Lat'][x] = float(Lat)
    Incidents_df['Lon'][x] = float(Lon)
```

The initial map showed some locations were found to be well outside the boundaries of North Hollywood (in the LAX area), so those locations were removed by filtering them out by latitude.



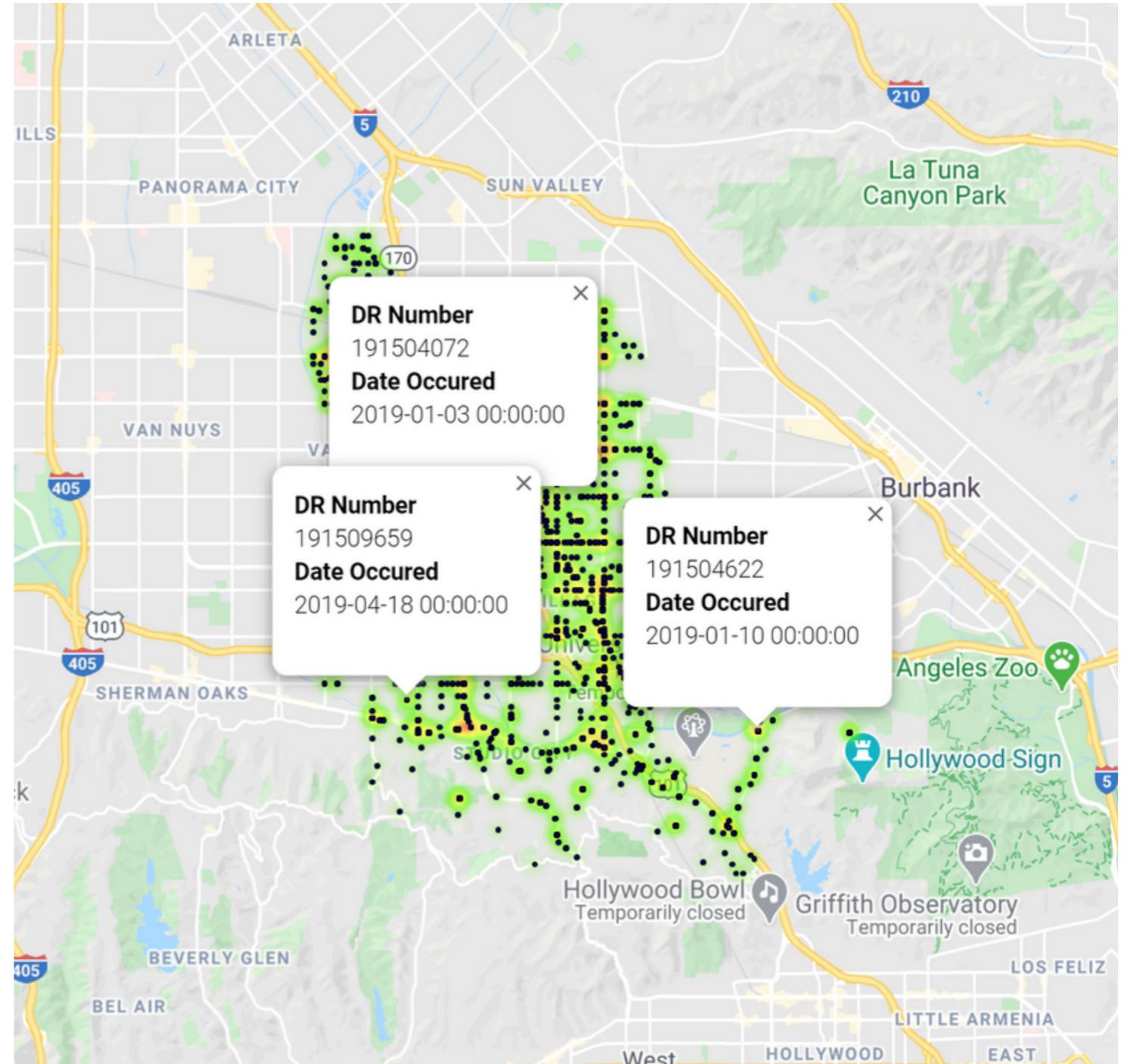
Preliminary maps were made using Google Maps for a single year's worth of data, to make sure the coding was working.

Created new data frame of unique locations for the heat map, and made a count for all the incidents per location.

```
# now to create a column for the count  
HeatBins_df[ 'Count' ]=0  
  
# and now iterate through the data  
  
for x in range(len(Cleaned_5Year_df)):  
    for y in range(len(HeatBins_df)):  
        try:  
            if Cleaned_5Year_df.Location[x]==HeatBins_df.Location[y]:  
                HeatBins_df.Count[y]+=1  
                break  
        except:  
            print(Cleaned_5Year_df[ 'DR Number' ][x])
```

# Final Map

Final maps using all 5 years worth of data were then created, and saved.



# Obstacles



Live navigation, while possible, performs very poorly due to the amount of data, making any system it's run on very very slow.

# Zoey's Mission

To find the most prevalent types of collisions in North Hollywood's top 10 North Hollywood hot spots



# Process

Find top 10 collision hotspots using Stephen's Heatbins file

Out[5]:

	Location	Lat	Lon	Count
0	(34.2012, -118.4052)	34.2012	-118.4052	165
1	(34.2012, -118.4138)	34.2012	-118.4138	137
2	(34.1867, -118.3703)	34.1867	-118.3703	123
3	(34.2012, -118.4094)	34.2012	-118.4094	107
4	(34.1867, -118.4052)	34.1867	-118.4052	100
5	(34.1867, -118.4137)	34.1867	-118.4137	99
6	(34.1867, -118.3965)	34.1867	-118.3965	99
7	(34.2012, -118.3878)	34.2012	-118.3878	99
8	(34.2012, -118.4225)	34.2012	-118.4225	88
9	(34.1649, -118.3704)	34.1649	-118.3704	88

```
top10_series_df  
=hotspot_data_df.sort_value  
s('Count',ascending =  
False).head(10)
```

```
top10_series_df.index=r  
ange(len(top10_series_df  
))
```

```
hotspot1_df =  
collision_data_df.loc[collision_da  
ta_df['Location']=='(34.2012,  
-118.4052)']
```



	DR Number	Date Reported	Date Occurred	Time Occurred	Reporting District	Address	Cross Street	Location	MO Codes	Lat	Lon
0	191521665	12/6/2019	2019-12-06	2215	1513	SHERMAN WY	WHITSETT AV	(34.2012, -118.4052)	3006 3025 3037 3101 4015	34.2012	-118.4052
1	161513566	6/22/2016	2016-06-22	1130	1513	SHERMAN WY	WHITSETT AV	(34.2012, -118.4052)	3004 3037 3030 3028 3101 3401 3701 4026	34.2012	-118.4052
2	161513567	6/22/2016	2016-06-22	1135	1513	SHERMAN WY	WHITSETT AV	(34.2012, -118.4052)	3004 3037 3030 3028 3101 3401 3701 4026	34.2012	-118.4052
3	151514680	7/10/2015	2015-07-09	1850	1513	WHITSETT AV	SHERMAN WY	(34.2012, -118.4052)	3004 3037 3025 3101 3401 3701 4026 3602	34.2012	-118.4052
4	181515320	7/30/2018	2018-07-30	35	1513	SHERMAN WY	WHITSETT AV	(34.2012, -118.4052)	3004 3037 3030 3028 3101 3401 3701 4026	34.2012	-118.4052

# Process

Separate MO Codes and count them using following process:

```
hotspot1_df['MO Codes'] =  
hotspot1_df['MO  
Codes'].values.tolist()
```



```
df1=hotspot1_df['MO  
Codes'].str.get_dummi  
es(sep=' ')
```



```
hotspot1_df=hotspot1  
_df.join(df1)
```

	DR Number	Date Reported	Date Occurred	Time Occurred	Reporting District	Address	Cross Street	Location	MO Codes	Lat	...	3039	3101	3102	3104	3401	3602	3701
0	191521665	12/6/2019	2019-12-06	2215	1513	SHERMAN WY	WHITSETT AV	(34.2012, -118.4052)	3006 3025 3037 3101 4015	34.2012	...	0	1	0	0	0	0	0
1	161513566	6/22/2016	2016-06-22	1130	1513	SHERMAN WY	WHITSETT AV	(34.2012, -118.4052)	3004 3037 3030 3028 3101 3401 3701 4026	34.2012	...	0	1	0	0	1	0	1
2	161513567	6/22/2016	2016-06-22	1135	1513	SHERMAN WY	WHITSETT AV	(34.2012, -118.4052)	3004 3037 3030 3028 3101 3401 3701 4026	34.2012	...	0	1	0	0	1	0	1

# Process

Create new df with sum of each MO count for every hot spot.

MO 3003	MO 3004	MO 3006	MO 3008	MO 3009	MO 3011
3	127	13	5	1	7

hotspot1\_df.sum()

```
hotspot1_MO3003 =  
hotspot1_df['3003'].sum()
```

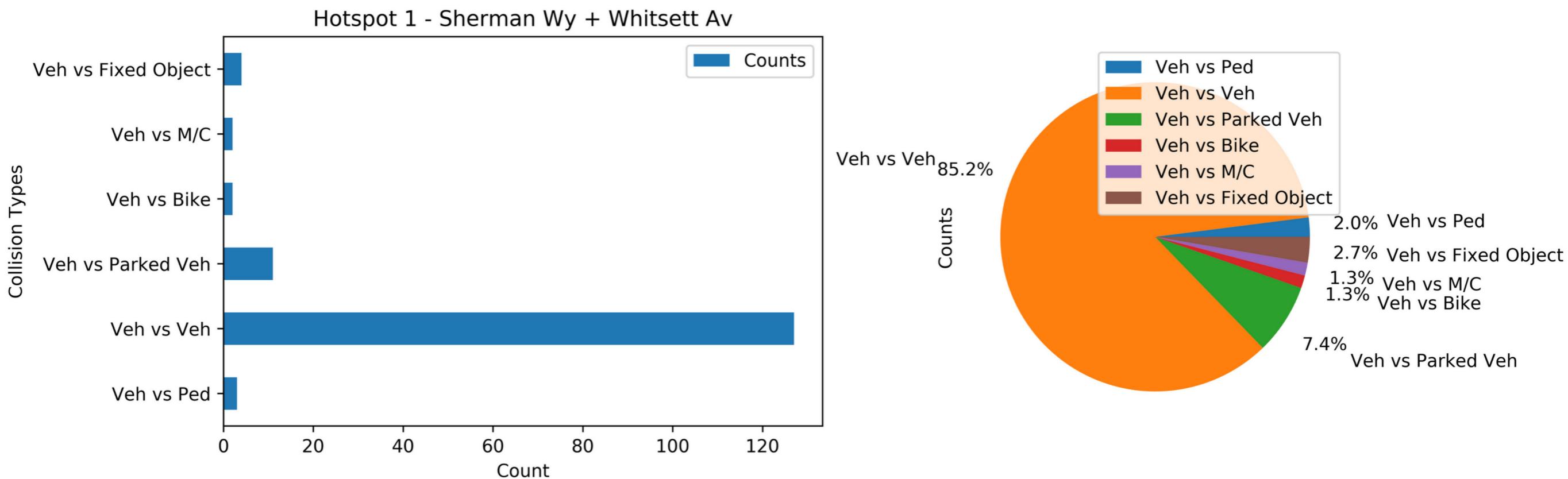
```
summary1_df = pd.DataFrame({"MO 3003": hotspot1_MO3003},  
index=[""])
```



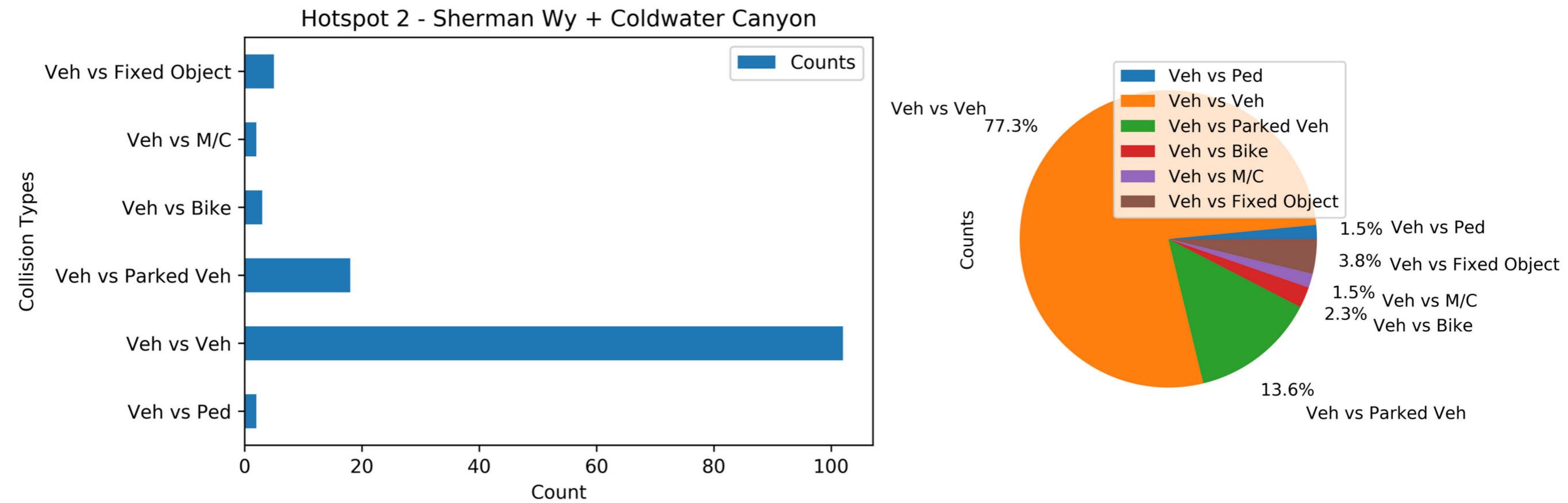
	Counts
Veh vs Ped	3
Veh vs Veh	127
Veh vs Parked Veh	11
Veh vs Bike	2
Veh vs M/C	2
Veh vs Fixed Object	4

```
summary1_df = pd.DataFrame({"Veh vs Ped":  
hotspot1_MO3003, "Veh vs Veh":  
hotspot1_MO3004,  
"Veh vs Parked Veh":hotspot8_MO3006, "Veh vs  
Bike": hotspot8_MO3008,  
"Veh vs M/C": hotspot8_MO3009, "Veh vs Fixed  
Object": hotspot8_MO3011},  
index=[""]).T  
updated1_df = summary1_df.rename(columns =  
{"" : "Counts"})
```

# HotSpot 1

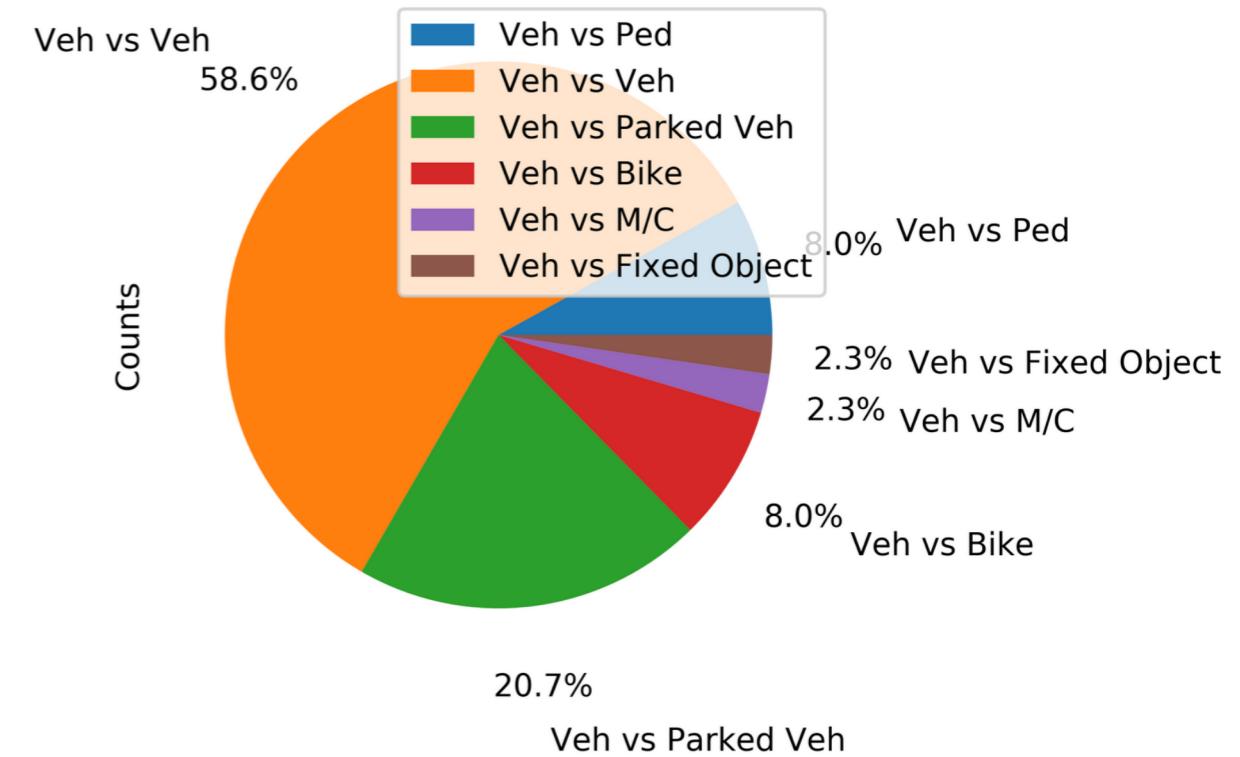
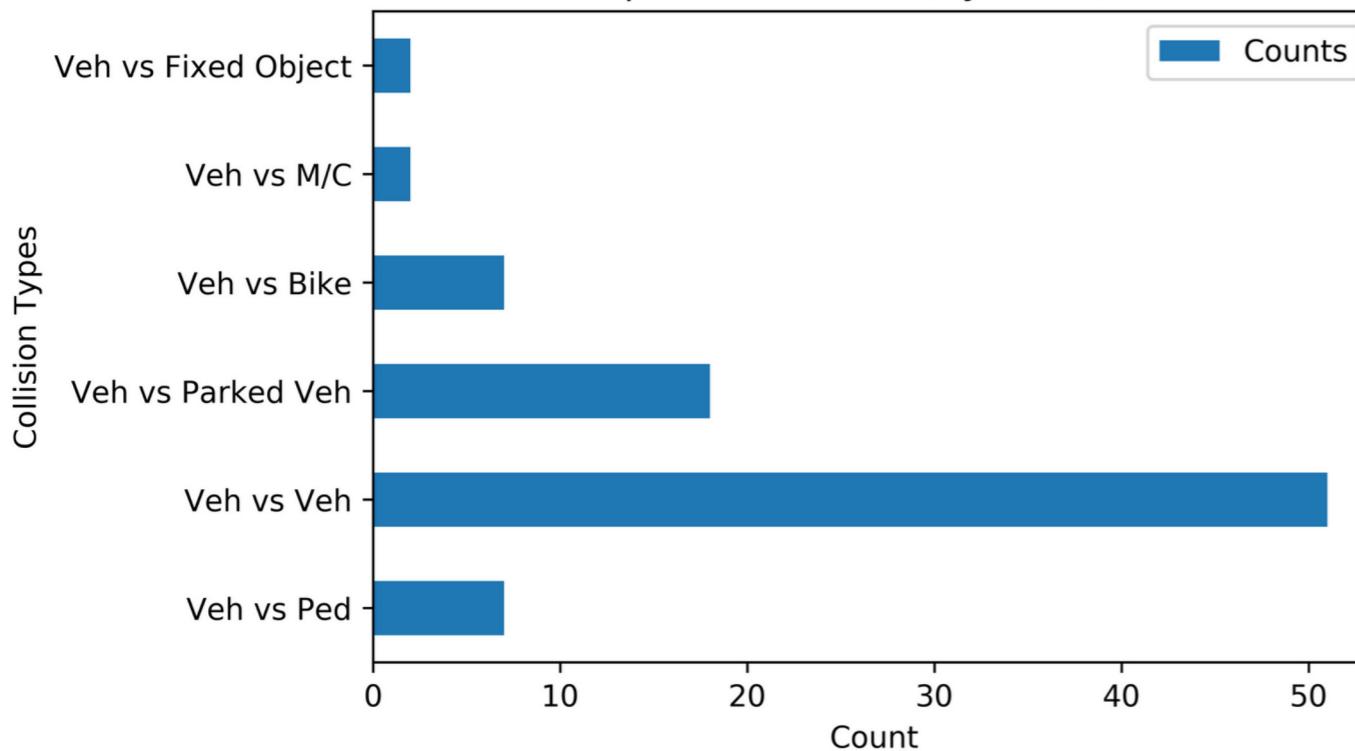


# Hotspot 2



# Hotspot 9

Hotspot 9 - Sherman Wy + Fulton Av



# Findings

Based on our analysis, we determined that Veh vs. Veh collision types are far more common than others. The top 10 hotspots all indicate very high percentages, ranging from 58% to 85%.

We also determined the second most prevalent collision is Veh. vs. Parked Vehicle. In some cases, the percentage was as high as 20%.

The reason for these accidents are unknown based on the provided data, and further investigation will be required.

# Obstacles



Initially planned to determine the most prevalent MO Codes; however soon after realized that the top 10 MO Codes were very generic and non-descriptive Codes.

Pivoted to new set of MO Codes (3003-3012) which would provide specific collision related information.



Initially planned to discover which MOs were most prevalent at specific times of the day (divided into hourly time slots). However, soon realized there was not enough data to draw any conclusions.

For example, Hotspot 9 which has 50 car vs car collisions was in a span of 5 years. This would not have been enough data to spread over 24 hour spans over 5 years.

# Questions?



Q: But why llamas?

A:

