The Worst Streets in New York

Timur Boskailo, Tyler Cranmer, Hitomi Imai

Description:

Our team aims to examine the dataset of NYC parking tickets and address the following questions to consider whether there are any noteworthy patterns between violations and types of cars, seasonalities and locations:

- Are there any patterns between types of violations (violation codes) and features of cars (such as vehicle types, makers and colors of cars)
- Are there any patterns in which agencies issue tickets?
- Are there any seasonalities in violation? Which months show more violations?
- Are there any trends in the locations (registration states and locations of the violation)?

The Datasets

NYC Parking Tickets

- Data is collected by the NYC Department of Finance and disclosed on Kaggle's platform
- 42M Rows of Parking ticket Data
- Dataset will be stored on Google Colab or Jupyter Notebook on local systems for team members.
- https://www.kaggle.com/new-york-city/nyc-parking-tickets

ATTRIBUTES

```
df_2017.head()
```

	Plate ID	Registration State	Plate Type	Issue Date	Violation Code	Vehicle Body Type	Vehicle Make	Issuing Agency	Vehicle Expiration Date	Violation Precinct	Issuer Precinct	Issuer Code	Issuer Squad	Violation Time	Violation County	Street Name	Vehicle Color	Vehicle Year	Meter Number	Violation Description
0	GZH7067	NY	PAS	07/10/2016	7	SUBN	TOYOT	V	0	0	0	0	NaN	0143A	BX	ALLERTON AVE (W/B) @	GY	2001	NaN	FAILURE TO STOP AT RED LIGHT
1	GZH7067	NY	PAS	07/08/2016	7	SUBN	TOYOT	V	0	0	0	0	NaN	0400P	BX	ALLERTON AVE (W/B) @	GY	2001	NaN	FAILURE TO STOP AT RED LIGHT
2	FZX9232	NY	PAS	08/23/2016	5	SUBN	FORD	V	0	0	0	0	NaN	0233P	BX	SB WEBSTER AVE @ E 1	BK	2004	NaN	BUS LANE VIOLATION
3	66623ME	NY	COM	06/14/2017	47	REFG	MITSU	T	20180630	14	14	359594	J	1120A	NY	7th Ave	WH	2007	NaN	47-Double PKG-Midtown
4	37033JV	NY	COM	11/21/2016	69	DELV	INTER	T	20170228	13	13	364832	M	0555P	NY	6th Ave	WHITE	2007	NaN	69-Failure to Disp Muni Recpt

List of tool(s)

We intend to use:

- Python 
- Pandas 
- NumPy 
- Matplotlib 

Data Preparation

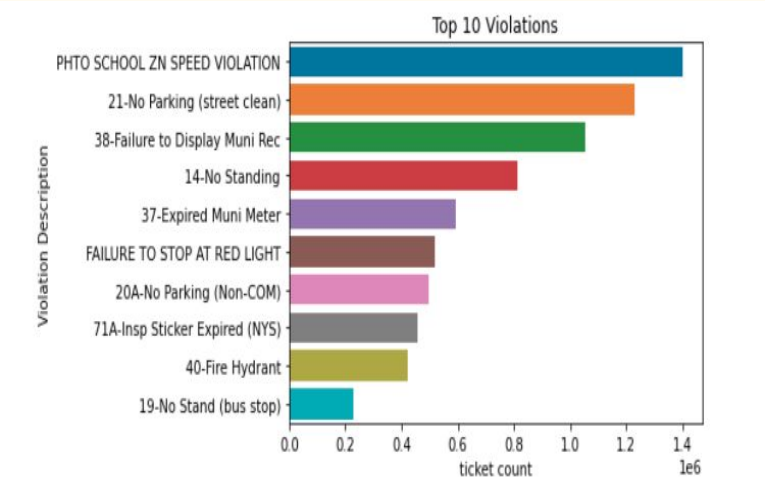
We examined the 44 attributes and removed the ones which were incomplete, duplicative or irrelevant for our analysis. The following is a list of removed attributes.

```
df_2017.drop('Summons Number', axis=1, inplace=True)
df_2017.drop('Street Code1', axis=1, inplace=True)
df_2017.drop('Street Code2', axis=1, inplace=True)
df_2017.drop('Street Code3', axis=1, inplace=True)
df_2017.drop('Violation Location', axis=1, inplace=True)
df_2017.drop('Issuer Command', axis=1, inplace=True)
df_2017.drop('Time First Observed', axis=1, inplace=True)
df_2017.drop('Violation In Front Of Or Opposite', axis=1, inplace=True)
df_2017.drop('House Number', axis=1, inplace=True)
df_2017.drop('Intersecting Street', axis=1, inplace=True)
df_2017.drop('Date First Observed', axis=1, inplace=True)
df_2017.drop('Law Section', axis=1, inplace=True)
df_2017.drop('Sub Division', axis=1, inplace=True)
df_2017.drop('Violation Legal Code', axis=1, inplace=True)
df_2017.drop('Days Parking In Effect    ', axis=1, inplace=True)
df_2017.drop('From Hours In Effect', axis=1, inplace=True)
df_2017.drop('To Hours In Effect', axis=1, inplace=True)
df_2017.drop('Unregistered Vehicle?', axis=1, inplace=True)
df_2017.drop('No Standing or Stopping Violation', axis=1, inplace=True)
df_2017.drop('Hydrant Violation', axis=1, inplace=True)
df_2017.drop('Double Parking Violation', axis=1, inplace=True)
```

	Nan Entry	% Filled Entry	Unique Entry	% Unique Entries
Summons Number	False	100.000000	10803028	100.000000
Plate ID	True	99.993261	3059045	28.316552
Registration State	False	100.000000	67	0.000620
Plate Type	False	100.000000	86	0.000796
Issue Date	False	100.000000	2063	0.019096
Violation Code	False	100.000000	100	0.000926
Vehicle Body Type	True	99.604639	1847	0.017097
Vehicle Make	True	99.323801	5702	0.052781
Issuing Agency	False	100.000000	17	0.000157
Street Code1	False	100.000000	6561	0.060733
Street Code2	False	100.000000	6855	0.063454
Street Code3	False	100.000000	6730	0.062297
Vehicle Expiration Date	False	100.000000	4074	0.037712
Violation Location	True	80.816490	212	0.001962
Violation Precinct	False	100.000000	213	0.001972
Issuer Precinct	False	100.000000	655	0.006063
Issuer Code	False	100.000000	36929	0.341839
Issuer Command	True	80.906788	3250	0.030084
Issuer Squad	True	80.898494	49	0.000454
Violation Time	True	99.999417	1743	0.016134
Time First Observed	True	7.782512	2131	0.019726
...	...	...	...	...
House Number	True	78.815032	51575	0.477412
Street Name	True	99.962890	84215	0.779550
Intersecting Street	True	31.172325	226774	2.099171
Date First Observed	False	100.000000	537	0.004971
Law Section	False	100.000000	8	0.000074
Sub Division	True	99.992845	143	0.001324
Violation Legal Code	True	19.093786	4	0.000037
Days Parking In Effect	True	74.892077	180	0.001666
From Hours In Effect	True	49.542425	484	0.004480
To Hours In Effect	True	49.542452	556	0.005147
Vehicle Color	True	98.589821	2367	0.021911
Unregistered Vehicle?	True	10.437777	1	0.000009
Vehicle Year	False	100.000000	100	0.000926
Meter Number	True	16.527524	29377	0.271933
Feet From Curb	False	100.000000	17	0.000157
Violation Post Code	True	70.469511	1157	0.010710
Violation Description	True	89.563389	107	0.000990
No Standing or Stopping Violation	True	0.000000	0	0.000000
Hydrant Violation	True	0.000000	0	0.000000
Double Parking Violation	True	0.000000	0	0.000000
ticket count	False	100.000000	10803028	100.000000

Knowledge is Power

Our team found that from the 107 different violations that were given out, over 66% of tickets were from 10 different violations. Where speeding in a school zone was the top contender with 13% of total tickets and street cleaning consisted of 11% tickets



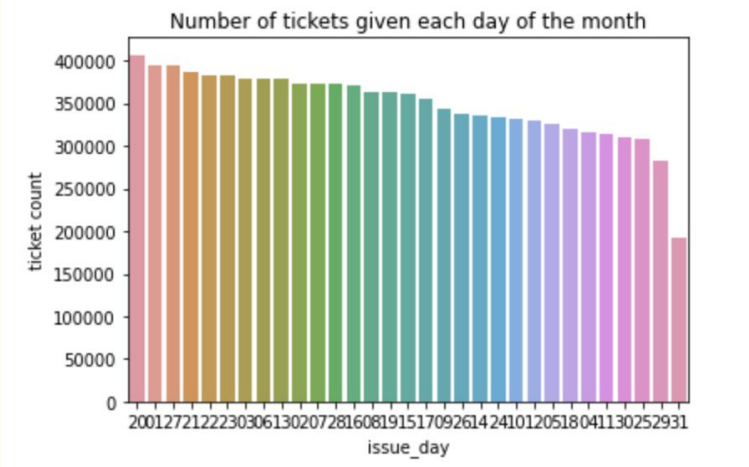
Registration States of Violating Cars

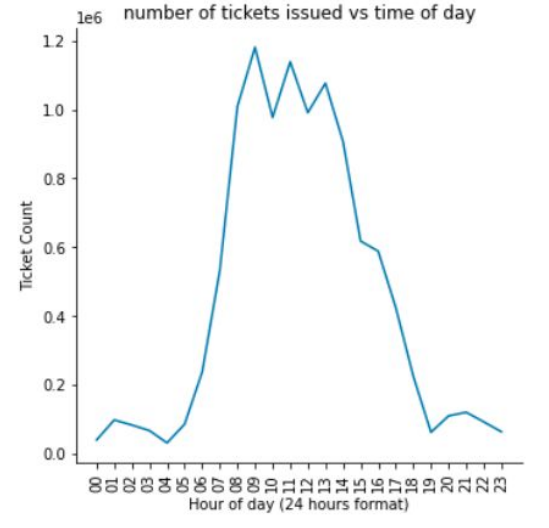
We examined if there is any tendency between violating codes and registration states. For the sake of analysis, we broke down by home areas, specifically ‘cars registered in NY States’ or ‘cars registered outside NY States’. Our team focused on top 3 violations codes (i.e, 21 - Street Cleaning - No parking where parking is not allowed by sign, street marking or traffic control device; 36 - Exceeding the posted speed limit in or near a designated school zone; 38 - Failing to show a receipt or tag in the windshield) and found that while approximately 78% of the violating cars are registered in NY, there are certain tendency that some codes are more likely violated by local cars registered in NY States. For example, Code 36 is relatively more violated by cars registered in NY States, while Code 21 is relatively less likely to be violated by cars registered in NY States. The result is shown in the table below:

	Total Violations	Violation Code 21	36	38
NY	0.784	0.748	0.862	0.793
Not NY	0.216	0.252	0.138	0.207

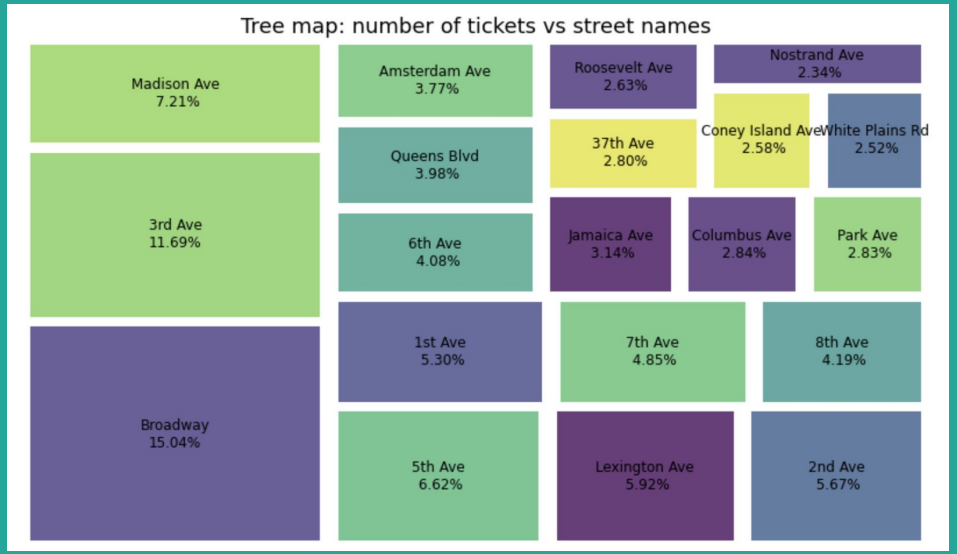
Ticket Numbers During the Month

Our team found that police agency T had by far the most tickets issued throughout the course of the year. What's interesting to see is that the ticket issuances on a monthly basis slowly declined through time and had a swift drop off at the end of each month. One can think that the infamous theory of police ticket quotas might exist.



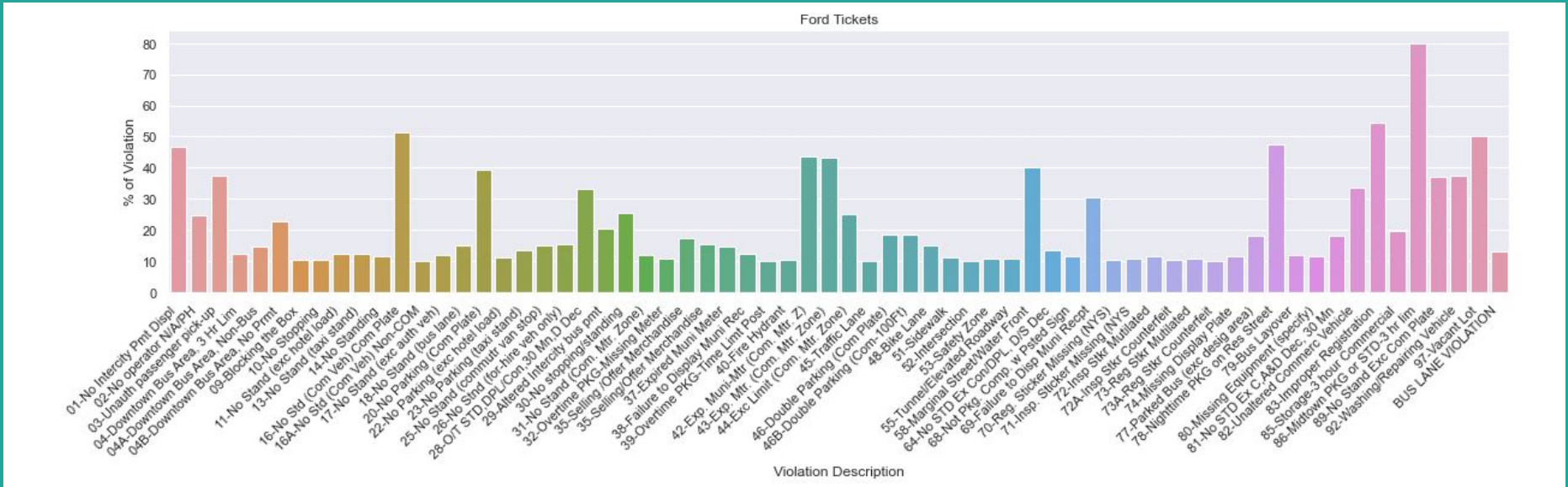


Dispersion of Tickets Between City Streets

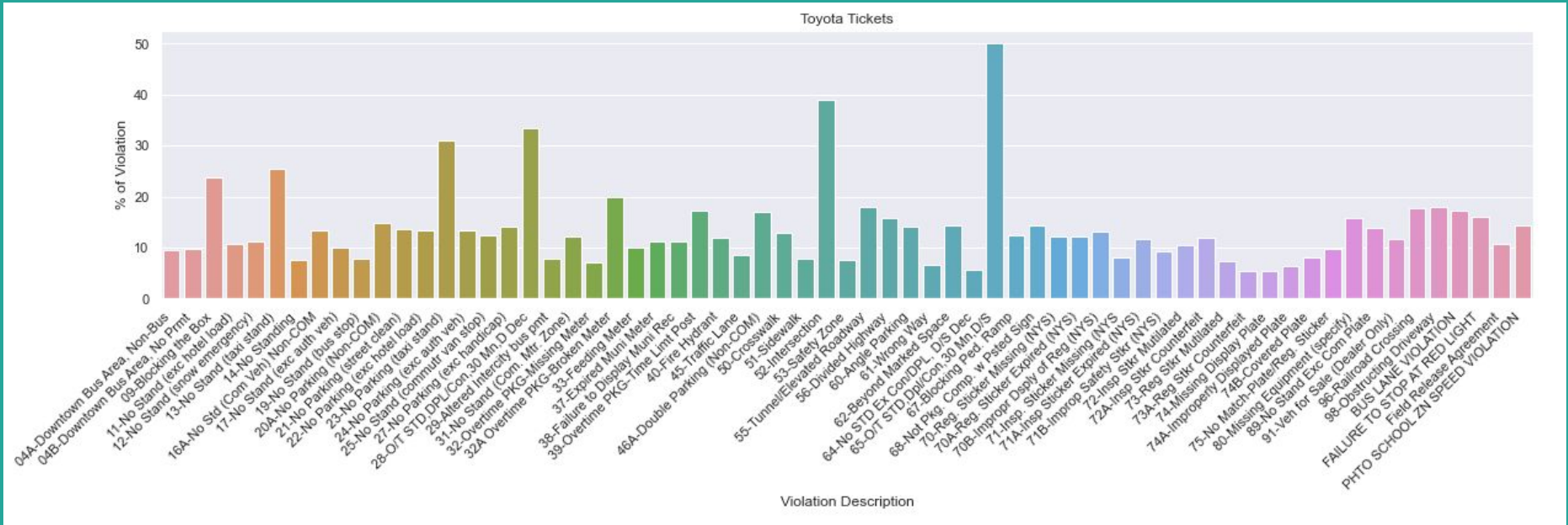


As clearly visible in the graphic above, Broadway, 3rd Avenue, and Madison avenue, had the highest percentage of tickets, with 15.04%, 11.69%, and 7.21% of all tickets handed out respectively. Combined, these 3 streets make up 33.94% of all tickets handed out.

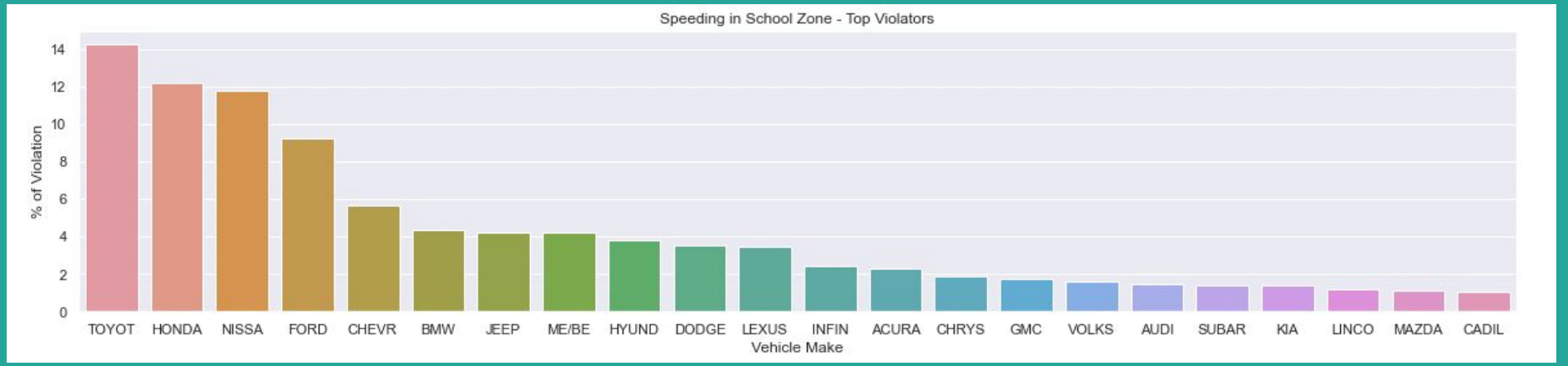
BREAKDOWN OF FORD TICKETS

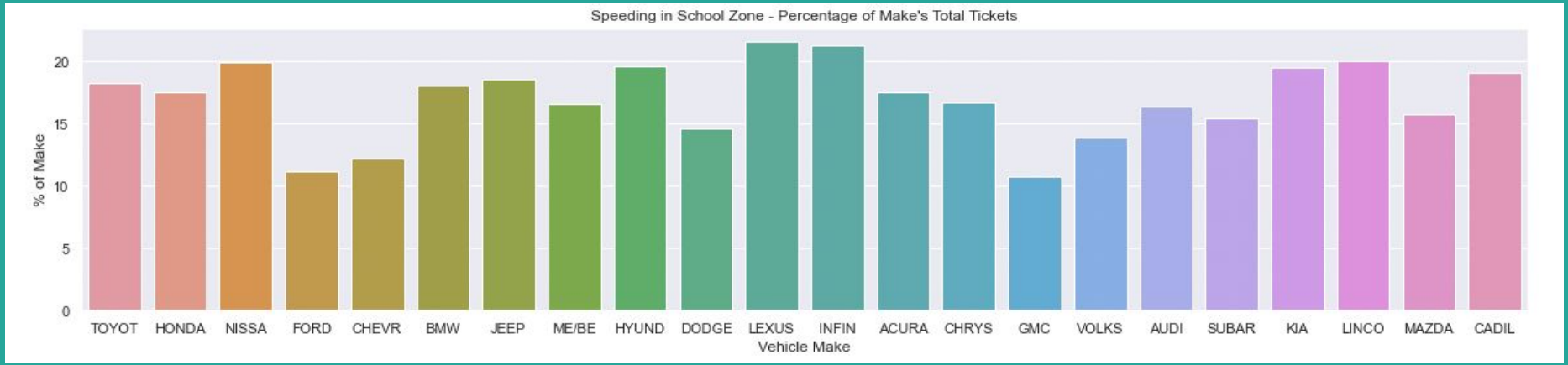


BREAKDOWN OF TOYOTA TICKETS



SPEEDING IN A SCHOOL ZONE





RESULTS, CONCLUSION, AND APPLICATIONS

- General Takeaways
- Future Work
- Applications
    - Two Perspectives
    - City
    - People