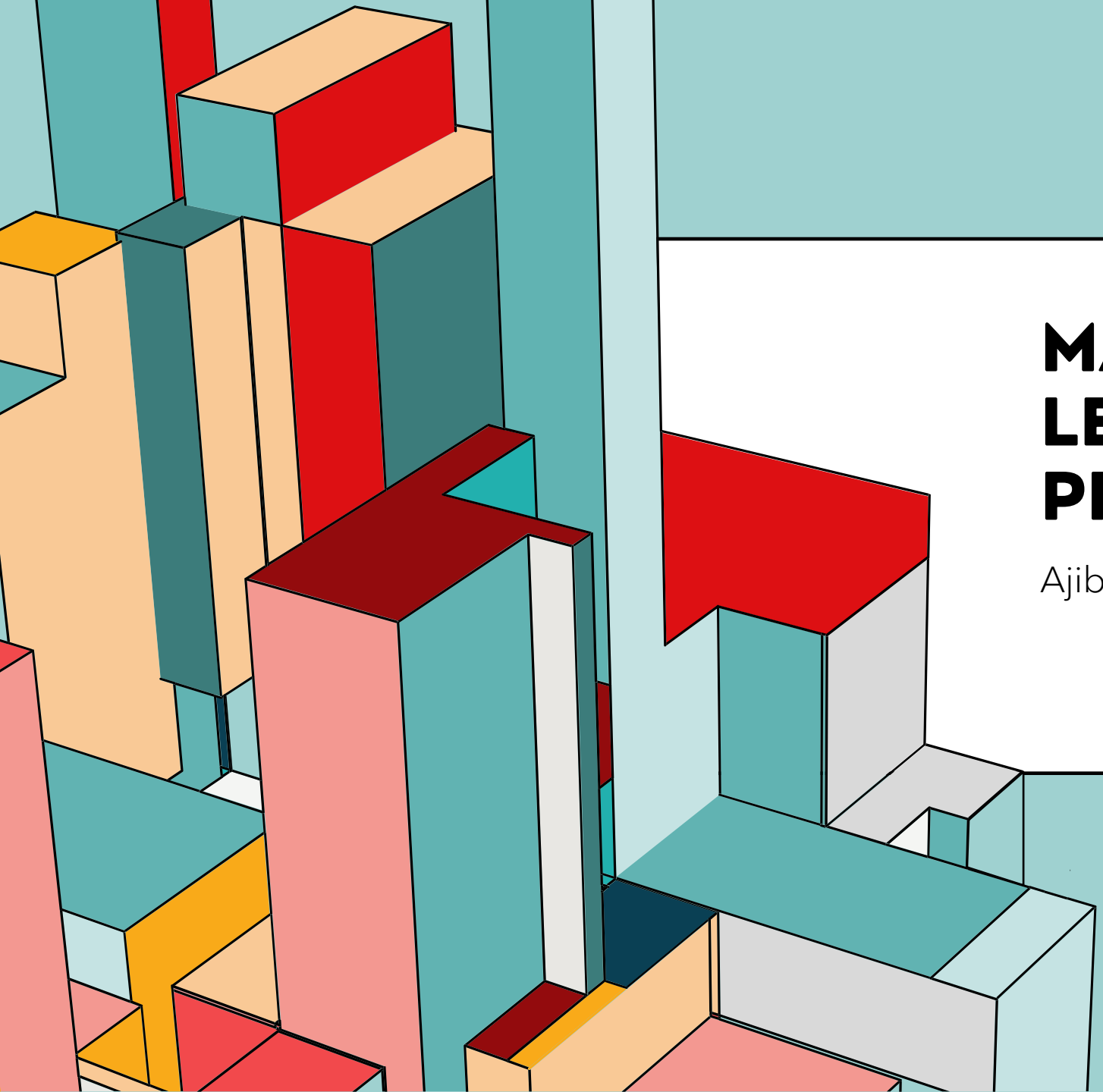
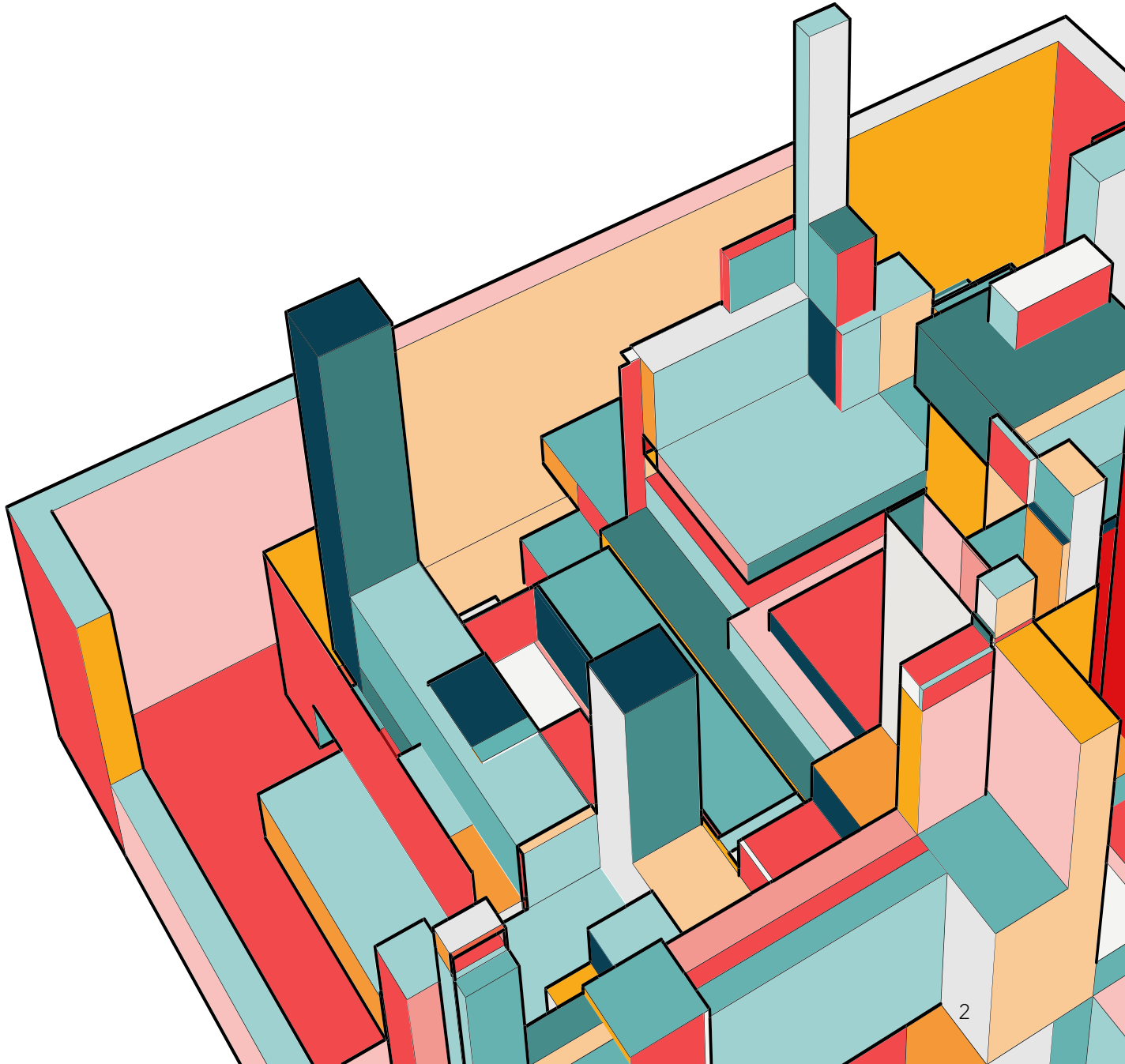


MACHINE LEARNING PROJECT

Ajibola Ganiyu & Madison Davis

ABOUT OUR DATA

We chose a unique dataset from the ones provided called “Motor_Vehicle_Collisions_-_Crashes.csv”. It was a very complex but interesting dataset with records of vehicle accidents in the five boroughs of NYC mostly from 2016 to 2022. Our aim was to understand the huge data very well so we would be able explain it well to the class. By filtering the data thoroughly, we were able to work with about 154,000 records out of the original dataset that contained over a million records.



MAKING SENSE OF THE DATA

How we filtered the data

STEP 1

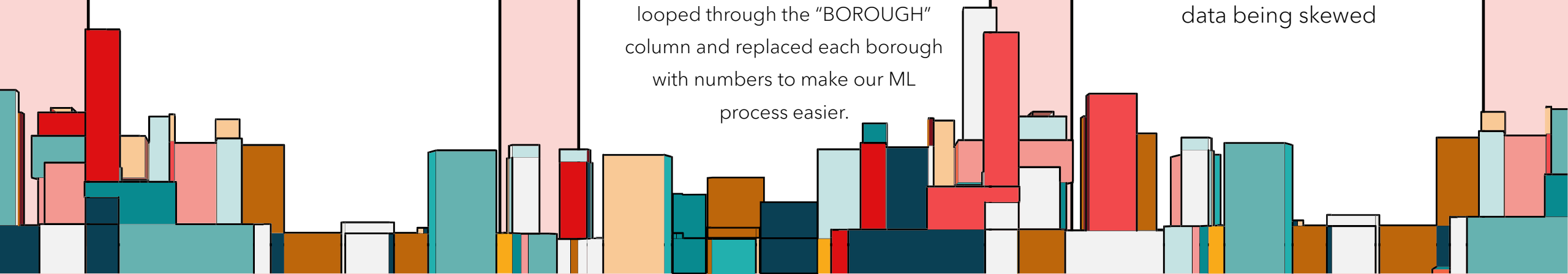
To make the data less ambiguous, we first dropped some columns we didn't need and then dropped other rows that the boroughs weren't specified in the data.

STEP 2

At this point, we had cut off about half the data. Then, we looped through the dates, that were in a MM/DD/YYYY format, to get only the value of the years. Then we replaced it as the "CRASH DATE". We also looped through the "BOROUGH" column and replaced each borough with numbers to make our ML process easier.

STEP 3

For the final step, we added the "INJURED" and "KILLED" values to a single column and streamlined the dataset to be from 2016-2022 to avoid the data being skewed



OUR MODEL

SPLITTING

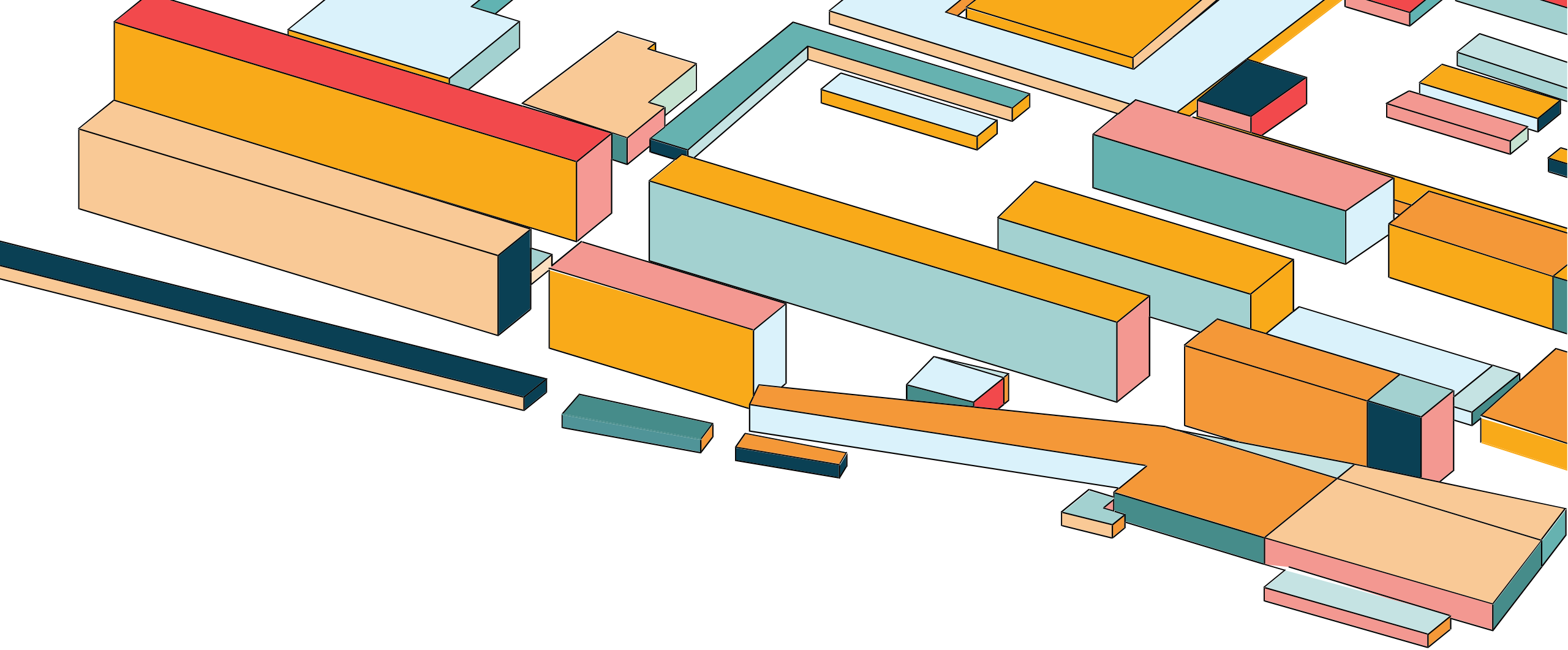
For the ML process, we decided to use the values from “INJURED OR KILLED” to predict the “CRASH DATE”. We believe our model could be used to determine a fair estimation of the years of other missing values in the dataset.

TRAINING

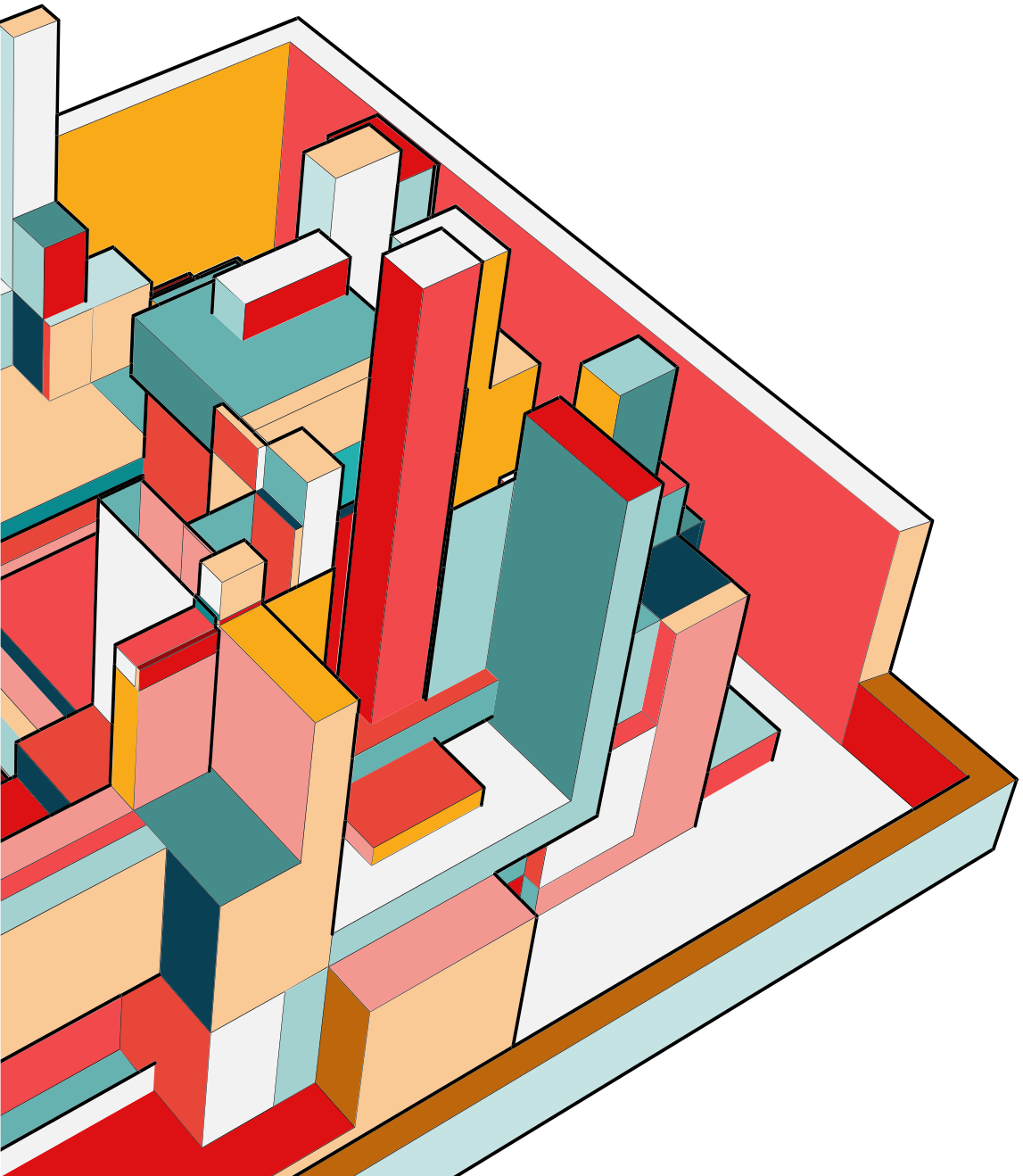
We used a regression model for this process. We trained 80% and left 20% for testing.

PREDICTING

Given that the values we had were very close, our model predicted years very close to the mean with almost no error. Mission accomplished!



WHAT WE LEARNED FROM THE DATA

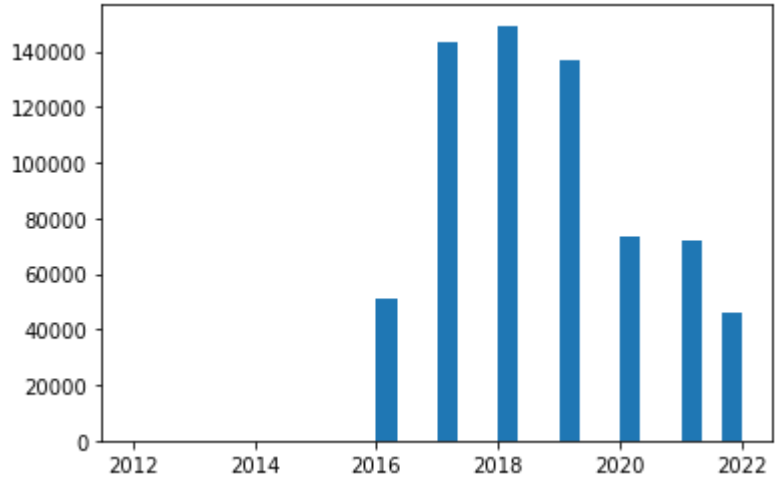


HOW WE VIEWED THE DATA

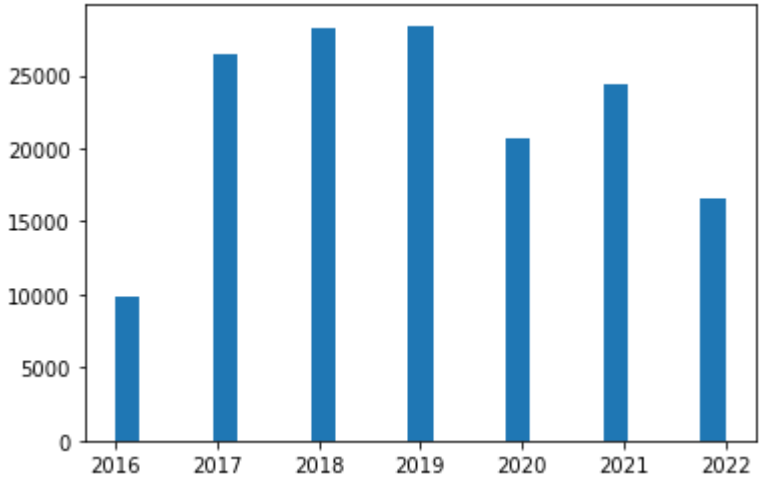
Since this was a huge dataset, we looked at it from two perspectives and studied the trends. First, we studied the general cases of vehicle accidents in the five boroughs of NYC, and then we studied the FATAL cases of those that had at least one person injured or killed.

OVER THE YEARS.....

TREND IN GENERAL ACCIDENTS



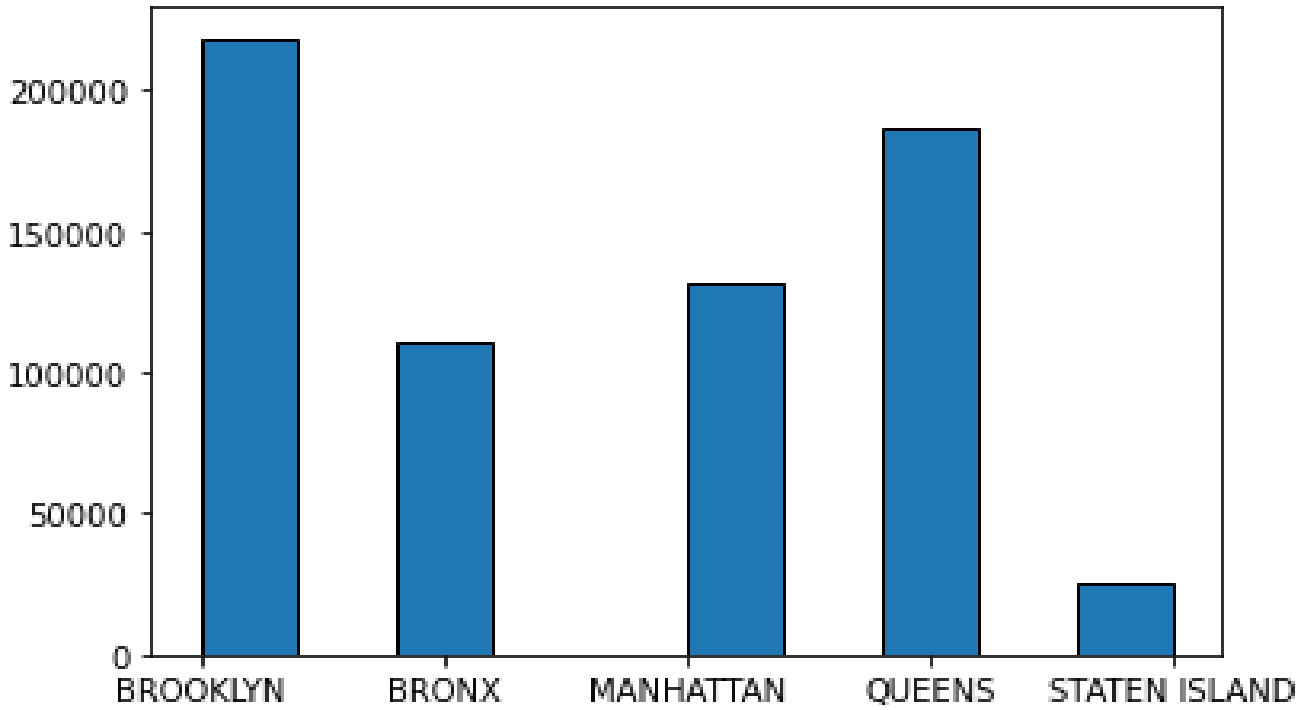
TREND IN FATAL ACCIDENTS



We observed 2018 to be the year with the highest vehicle accidents in the city.

BASED ON THE BOROUGHS

TREND IN GENERAL ACCIDENTS

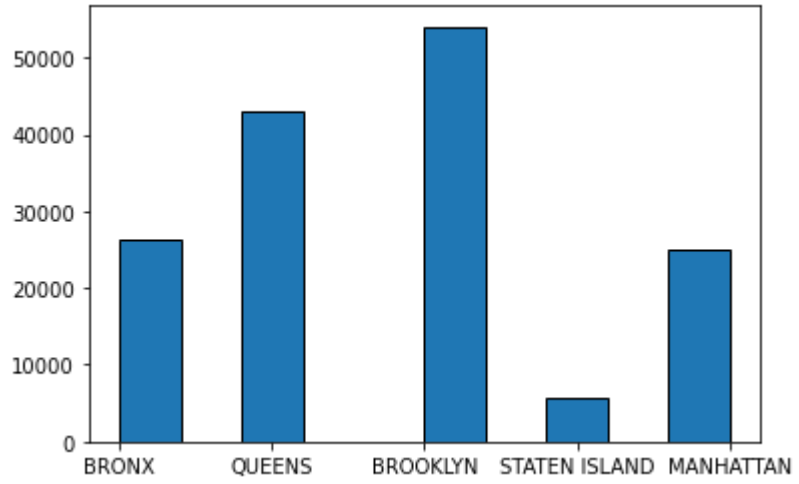


Be careful in Brooklyn!

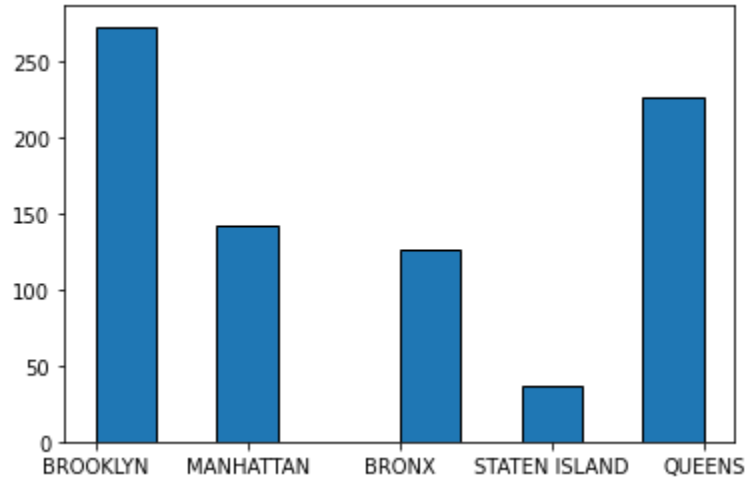
BASED ON THE BOROUGHS

TREND IN FATAL ACCIDENTS

AT LEAST ONE INJURED



AT LEAST ONE KILLED



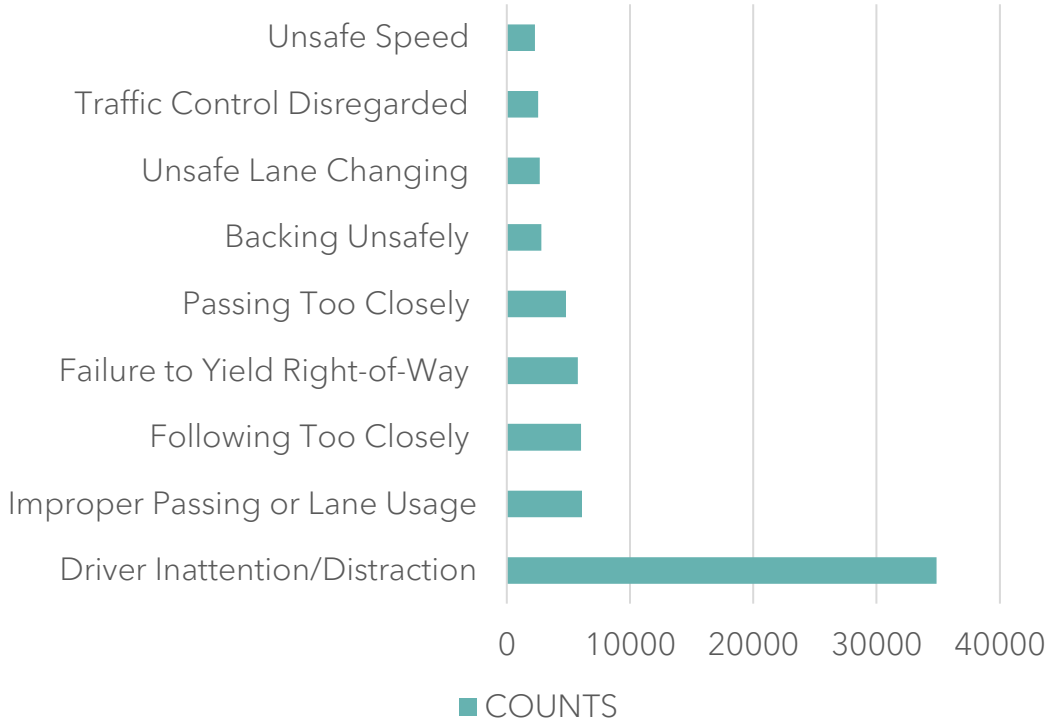
According to our data, 23% of all vehicle accidents from 2016-2022 in NYC were fatal.

MAJOR CAUSES OF THE FATAL VEHICLE ACCIDENTS IN OUR DATA

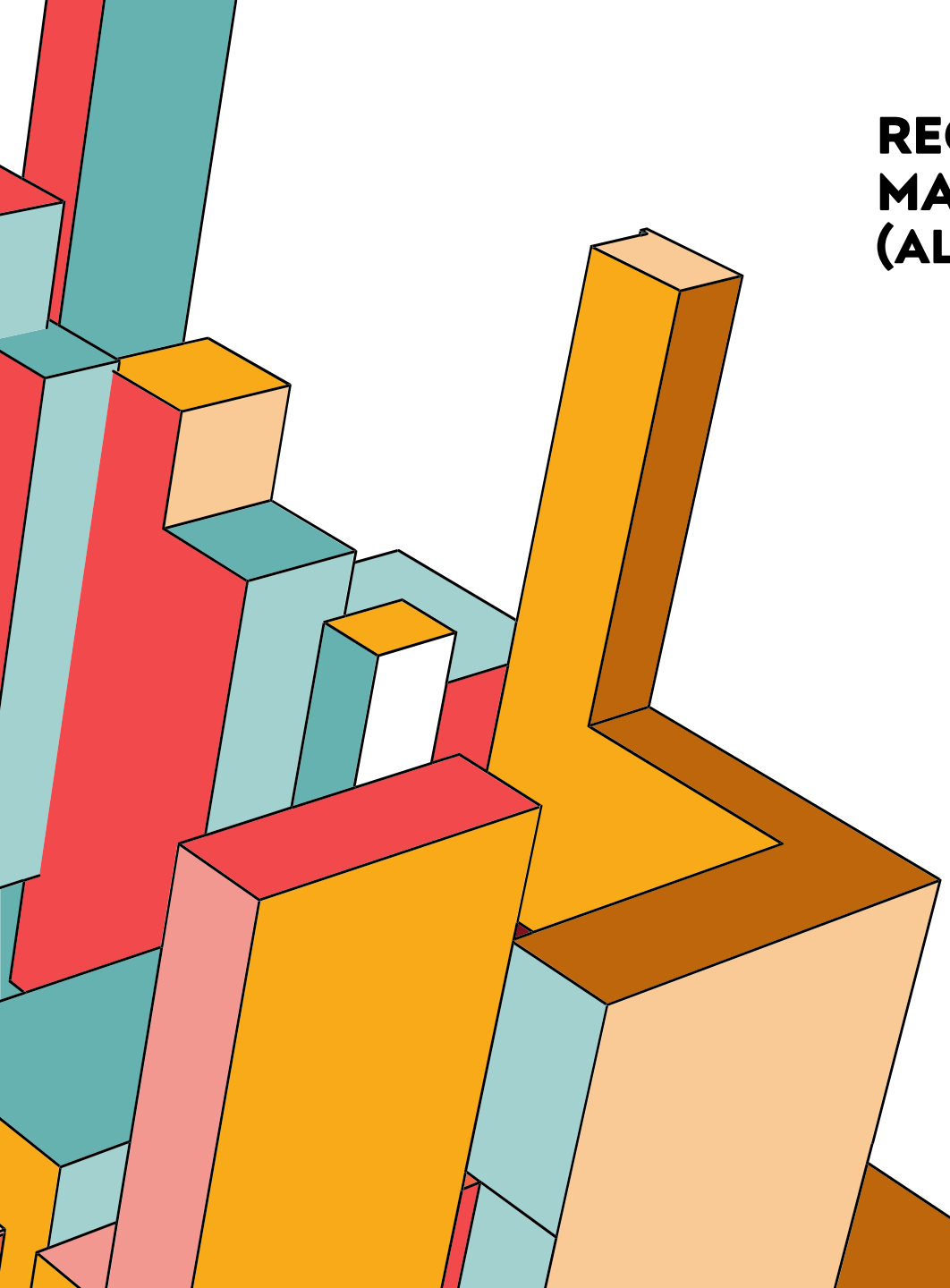
(AT LEAST ONE PERSON WAS INJURED OR KILLED)

CAUSES	AMOUNT OF INCIDENTS
Driver Inattention/Distraction	34,879
Improper Passing or Lane Usage	6,119
Following Too Closely	6,030
Failure to Yield Right-of-Way	5,766
Passing Too Closely	4,812
Backing Unsafely	2,804
Unsafe Lane Changing	2,682
Traffic Control Disregarded	2,539
Unsafe Speed	2,302

VISUAL AID

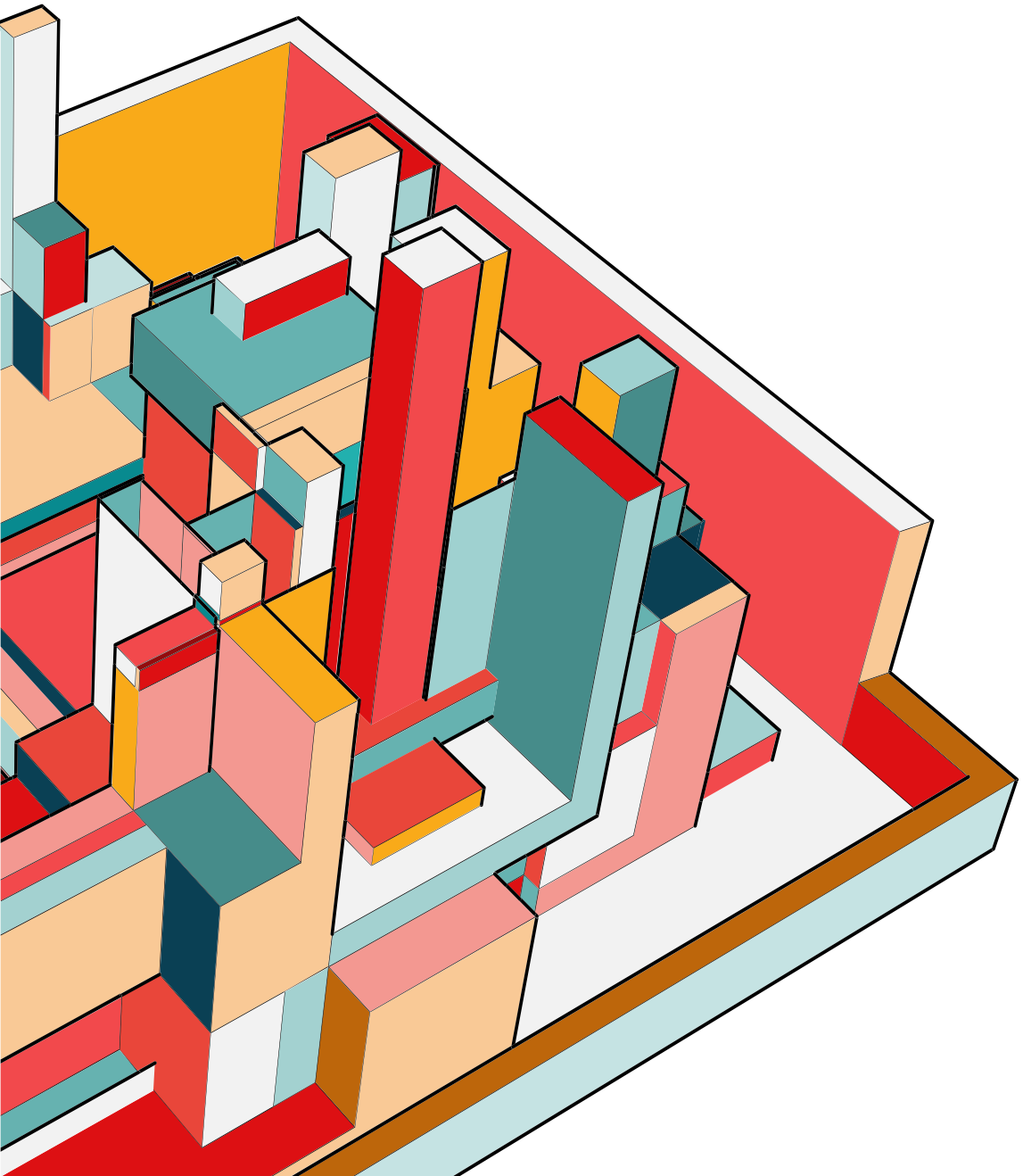


PS: There were about 460,000 cases in the dataset that the causes could not be determined



REGULAR MISTAKES THE AVERAGE ROAD USER MAKES THAT REFLECTED IN OUR DATA (ALSO, AT LEAST ONE PERSON INJURED OR KILLED)

CAUSES	AMOUNT OF INCIDENTS
Texting or use of cell phone	43
Eating or Drinking	4
Use of other Electronic Device	10
Fatigued/Drowsy	45
Fell Asleep	32



BUT WHAT KIND OF VEHICLES ARE USUALLY INVOLVED IN ACCIDENTS IN NYC?

FROM OUR DATA, THESE WERE THE TOP 95TH PERCENTILE OF NYC ACCIDENTS BY VEHICLE CLASS. THEREFORE, FOR EVERY VEHICLE ACCIDENT IN NYC, HERE ARE THE CHANCES OF THESE VEHICLE TYPES TO BE INVOLVED

Sedan

47.1%

SUV

34.8%

Taxi

4.3%

Pick-up Truck

2.7%

Box Truck

1.98%

Bus

1.7%

Bike

1.15%

Van

0.72%

HOWEVER, THESE WERE THE SAFEST

Limousine

0.02% in fatal accidents

Garbage Truck

0.13% in fatal accidents

School Bus

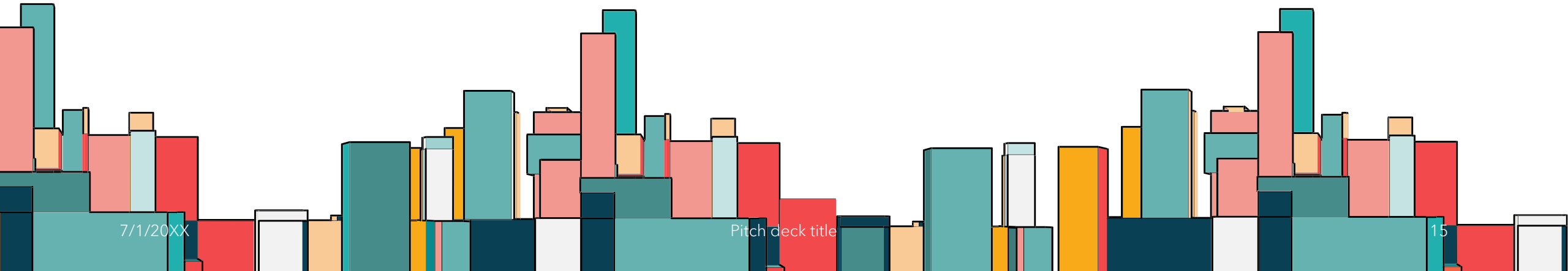
0.02% in fatal accidents

Ambulance Truck

0.17% in fatal accidents

SUMMARY

- 2018 was the year with the highest accidents in the dataset having a maximum of 22 people injured in a single accident.
- However, an accident in 2017 recorded 8 people killed, which was the highest in the whole dataset.
- 2020 had a drastic decrease in accidents compared to the years prior. This could have been caused by many variables in real life. Most likely the pandemic!
- About 23% of all accidents recorded from 2016-2022 involved at least one person injured or killed.
- As much as the law tries to regulate speed limits on roads, those that obey those rules but don't pay proper attention to driving are 15 times more likely to severely injure or kill other road users than those that violate speed limits (and pay attention).
- Fatal accidents in the city involving sedans and SUVs are 8 times more likely to occur than those involving taxis, buses and bikes combined!



THANK YOU

Any questions?