Methods of Advanced Data Engineering (MADE)

23076230, Ismail Halil Kuzu

July 4, 2024

1 Introduction

This report explores the relationship between weather conditions and traffic accidents in New York City. Understanding how weather influences accident rates can inform safety measures and improve traffic management strategies. The key question addressed is: How do different weather conditions affect the frequency and severity of traffic accidents in NYC?

2 Used Data

Datasource1: NYC Traffic Accidents

- **Metadata URL:** NYC Traffic Accidents - Metadata
- **Data URL:** NYC Traffic Accidents - Data
- **Data Type:** CSV

This dataset includes information about traffic accidents recorded in NYC from January 2020 to August 2020. It provides details such as the date of the accident and the number of accidents on that date. Key value types: **date** and **accident count** for combined dataset.

Datasource2: Weather Data for NYC

- **Metadata URL:** NYC Weather - Metadata
- **Data URL:** NYC Weather - Data
- **Data Type:** CSV

This dataset covers weather parameters recorded in NYC during the same period, including average temperature, total precipitation, total rain, and average cloud cover. Key value types: **date**, **avg temperature**, **total precipitation**, **avg cloudcover**, and **avg windspeed** for combined dataset.

Both datasets were obtained from reliable sources and processed to ensure accuracy and consistency in the analysis. The NYC Traffic Accidents Data helps to understand the frequency of accidents over time, while the NYC Weather Data provides insights into weather conditions prevailing during these periods.

Data sources licenses and obligations: Accident and weather datasets have been collected from public sources and are available for free use.

#	date	accident_count	avg_temperature	total_precipitation	total_rain	avg_cloudcover	avg_cloudcover_low	avg_cloudcover_mid	avg_cloudcover_high	avg_windspeed	avg_winddirection
1	2020-01-01	336	3.63	0	0	43.79	30.21	30.83	9.33	15.36	260.17
2	2020-01-02	409	2.23	0	0	10.04	0	0	33.63	10.24	232.46
3	2020-01-03	452	6.47	3.5	3.5	86.42	54.92	80	91.04	7.31	234.46
4	2020-01-04	347	7.84	4.9	4.9	99.88	93.75	83.75	86.67	5.34	186.21
5	2020-01-05	286	4.5	1.1	1.1	51.75	29.92	37.58	9.21	21.03	310.67
6	2020-01-06	455	2.52	0.6	0	49.38	19.46	51.38	17	11.16	247.71
7	2020-01-07	402	3.1	0.1	0.1	40.38	0	37.75	58.83	9.54	209.92
8	2020-01-08	451	1.59	0.8	0.5	32.33	2.5	43.96	12.5	15.91	251.96
9	2020-01-09	517	-1.74	0	0	19.67	11.13	9.67	17.5	13.59	289.08
10	2020-01-10	520	2.4	0	0	55.79	24.38	30.5	73.83	11.42	206.92
11	2020-01-11	438	11	0	0	63.33	30.5	39.17	68.42	15.71	210.96
12	2020-01-12	387	15.78	5.8	5.8	63	53.29	28.5	71.46	23.25	230.29
13	2020-01-13	473	5.25	0	0	58.17	41.88	4	90.63	8.32	98.88
14	2020-01-14	497	4.17	4.3	4.3	94.08	83.08	29.88	97.92	5.84	178.75
15	2020-01-15	505	6.81	0	0	53.46	54.29	6.54	12.38	7.84	234.92
16	2020-01-16	547	6.4	3.4	3.4	47.38	26.96	43.71	0.63	17.65	237.79
17	2020-01-17	565	-0.95	0	0	2.83	2.96	0.29	0	23.78	324.25
18	2020-01-18	773	-4.7	4	0	64.25	7.08	58.46	91.38	10.26	256.42
19	2020-01-19	360	2.75	6	1.4	77.67	72.88	46.96	15.46	16.75	261.88
20	2020-01-20	389	-1.95	0	0	27.04	30.04	0	0	18.18	322.67
21	2020-01-21	554	-3.12	0	0	7.75	0	8.29	9.13	10.3	329.33
22	2020-01-22	488	-1.49	0	0	3.5	0	0.83	9.92	4.83	288.42

Figure 1: Output of the combined and filtered datasets.

3 Analysis

Data Integration and Preparation: The traffic accidents and weather datasets were filtered and then merged based on the common date field to facilitate analysis.

Exploratory Data Analysis (EDA)

Monthly Traffic Accidents: The distribution of traffic accidents over months was visualized to identify any seasonal patterns. (See Figure 2)
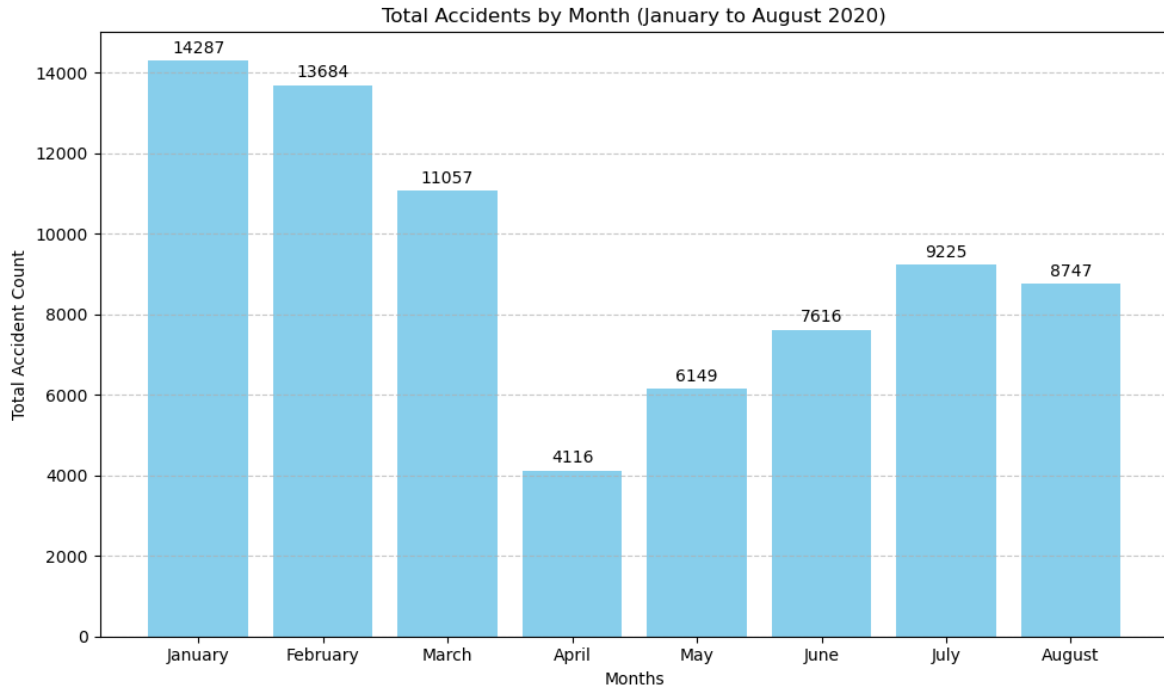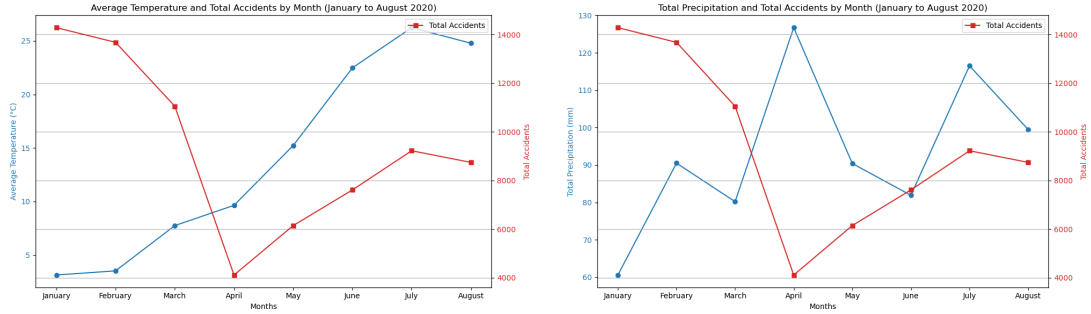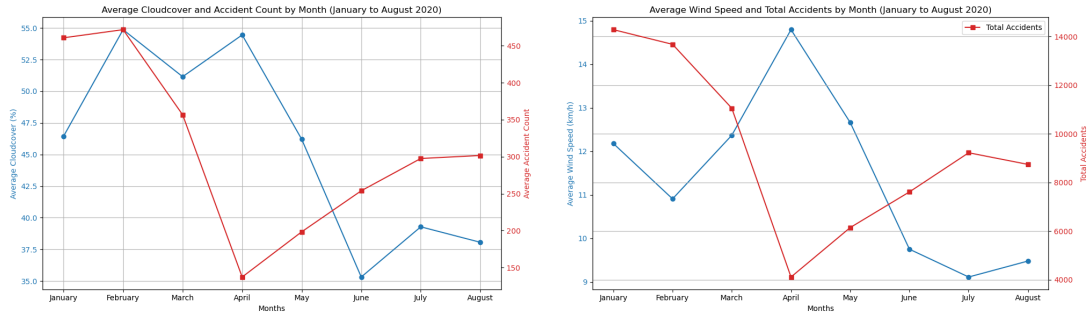

Figure 2: Total Accidents by Month.

Weather Variables: The variation of weather variables (temperature, precipitation, cloud cover, wind speed) across different months was analyzed. (See Figures 3 and 4)

We conducted correlation analyses between weather variables such as temperature, precipitation, cloud cover, and wind speed and the frequency of traffic accidents. These analyses provide insights into the influence of various weather conditions on accident rates.

(a) Average Temperature and Total Accidents by Month (b) Total Precipitation and Total Accidents by Month

Figure 3: Analysis of Temperature and Precipitation.

(a) Average Cloudcover and Accident Count by Month (b) Average Wind Speed and Total Accidents by Month

Figure 4: Analysis of Cloud Cover and Wind Speed.

Statistical techniques were employed to quantify the effects of weather variables on traffic accident rates. The results are presented using descriptive statistics, figures and tables to elucidate significant associations.
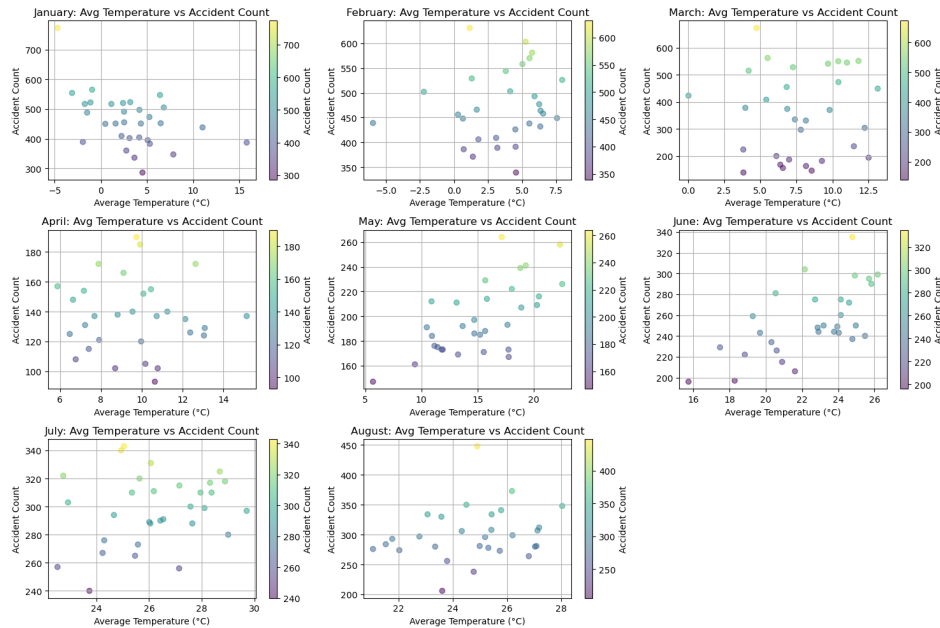

Figure 5: Average Tempature vs Total Accidents for Each Month

4 Conclusions

Based on the analysis:

- **Temperature:** Higher temperatures correlate with increased traffic accidents, potentially due to increased pedestrian and vehicle activity.

- **Precipitation:** Heavy precipitation and rain contribute to higher accident rates, highlighting the impact of wet road conditions on driving safety.

- **Cloud Cover:** Moderate levels of cloud cover showed mixed effects, indicating a need for further investigation into its specific impact.

- **Wind Speed:** Higher wind speeds showed a slight increase in accident rates, possibly due to reduced visibility and control issues for drivers.

Limitations and Uncertainties

- The analysis focused on correlational relationships and did not establish causation between weather conditions and accident rates.

- The data is limited to NYC and a specific time period (January 2020 to August 2020), which may not generalize to other regions or different years.

In conclusion, while weather conditions indeed influence traffic accident rates in NYC, further research with larger datasets would provide deeper insights into these relationships. This report serves as a foundational analysis to guide future studies and policy decisions aimed at improving road safety under varying weather conditions.