USA Traffic Accidents Data Analysis

-Komal Patel

11/23/2021

Introduction

- NHTSA's June'21 report shows that an estimated 38,680 people died in motor vehicle traffic crashes—the largest projected number of fatalities since 2007. Source : NHTSA Web page
- Reducing traffic accidents is an important public safety challenge.
- Accidents results in death, injuries and economic loss.
- Accidents cause a negative impacts on involved parties such as long or short term mental and physical health issues, productivity loss etc.
- Accidents also brings significant financial impact like medical costs, many resources involvement such as Emergency Medical Services, First Responders etc., Insurance administration cost, property damage etc.

Purpose of the Project

- Study of USA Traffic Accident Data received from Kaggle website.
  - Data from Feb'2016 to Dec.'2020

- **Goal :** From the data if we can predict any major pattern for causes of accident, it might be useful to identify any safety improvements and how better to operate and maintain the transportation facilities and resources.

- From the received data, Project focused on four key areas:

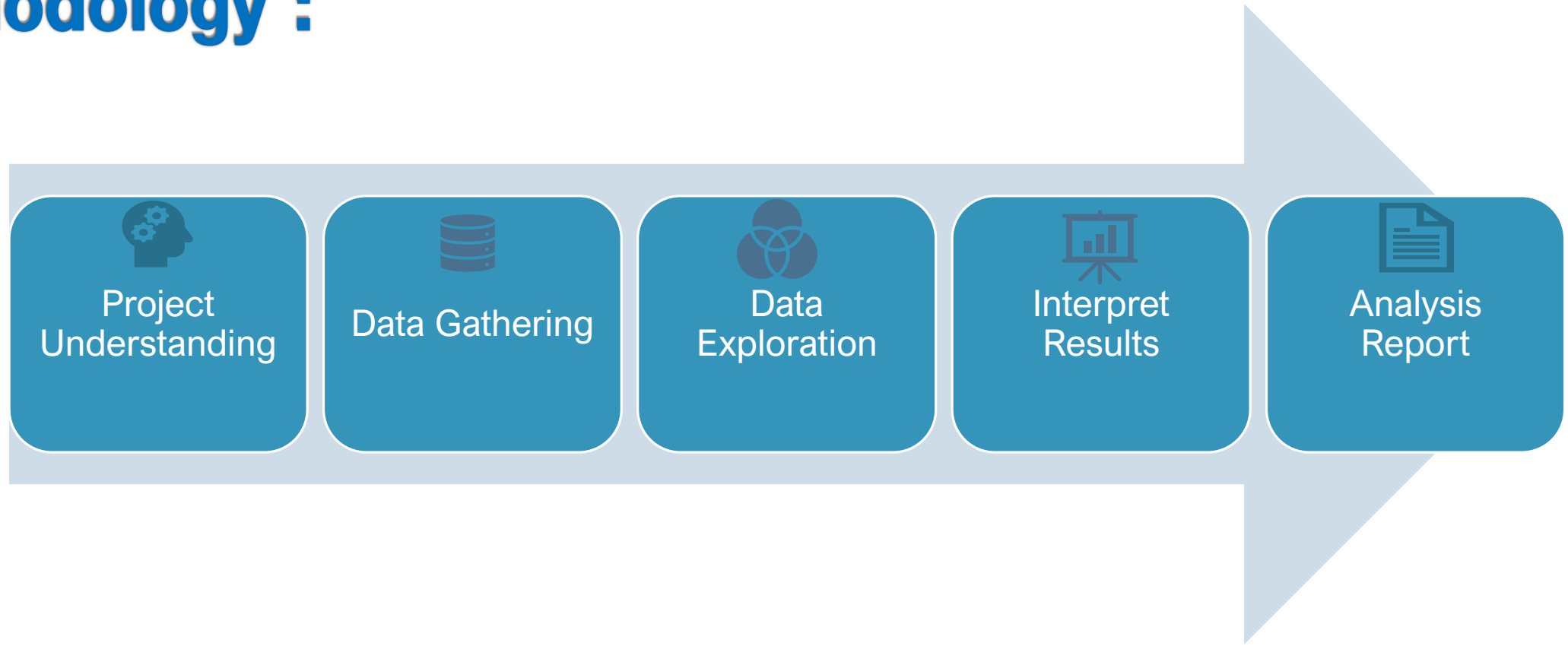
Locations
Identify commons areas with higher probability of accidents

Time
Finding a trend of accident and any relation with a time of the day or year

Severity
Evaluate to impact of Accidents on Traffic

Weather Condition
Explore if there is any impact of weather conditions on accidnets

Methodology :



Data Source :

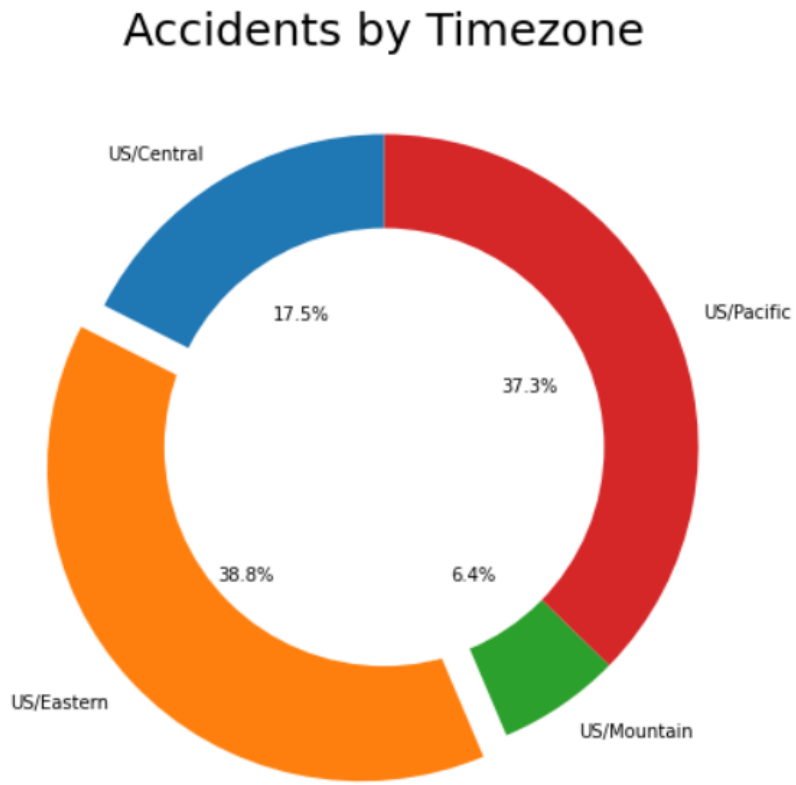
- Kaggle

Tools/Libraries :

- Colab
- Seaborn
- Matplotlib
- Pandas
- Numpy
- Scipy

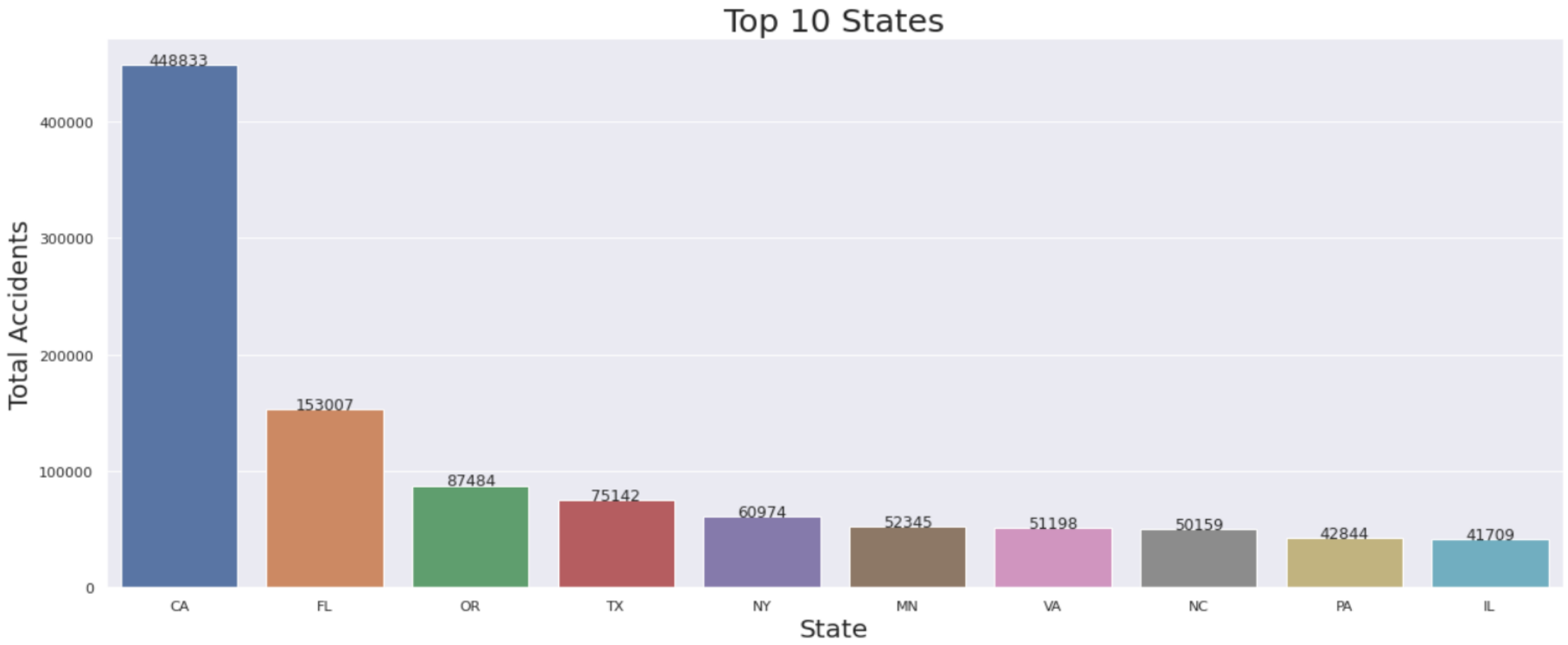
Location Analysis

Highest Number of accidents in Eastern and Pacific Time zones



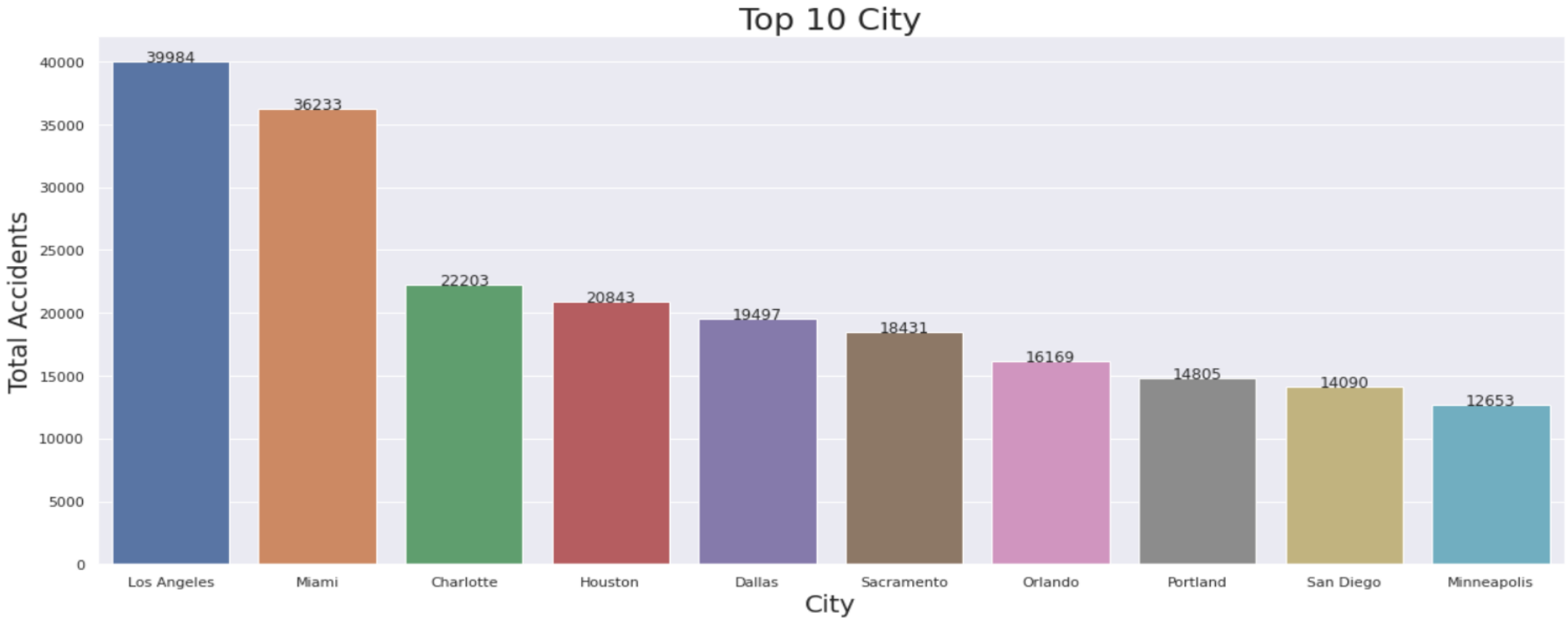
- Eastern time zone had the highest accidents no. 587961 (38.8%)
- Pacific time zone had the 2nd highest accidents no. 564358 (37.3%)
- Mountain time zone had the least accidents no. 96239 (6.4%)

~ 70% Accidents recorded only from the top 10 out of 49 states



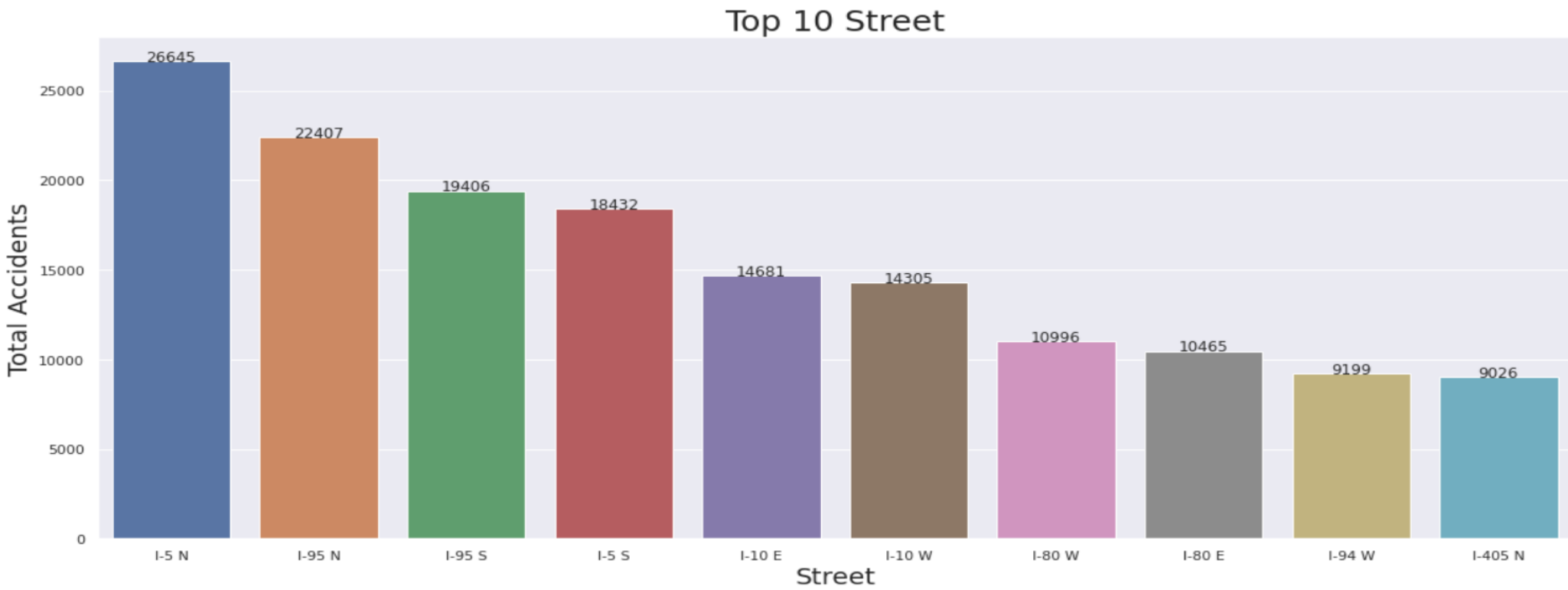
- **California had the highest** 448833 (29.60%) no. of accidents; averaging 7738 accidents monthly.
- **Florida had the 2nd highest** 153007 (10.09%) no. of accidents; averaging 2638 accidents monthly.

~14% accident recorded only from the top 10 out of 10,658 cities



- **Los Angeles** city **had the highest** 39984 (2.63%)no. of accidents; averaging 689 accidents monthly.
- **Miami** city **had the 2nd highest** 36233 (2.39%) no. of accidents; averaging 624 accidents monthly.

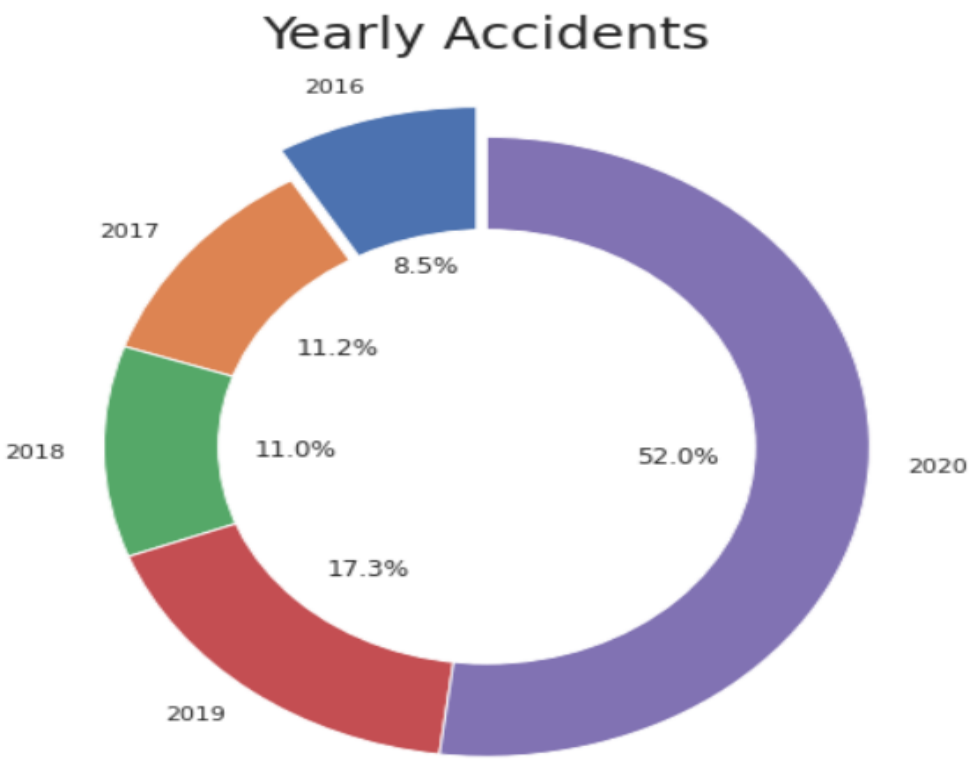
~10% accidents from the top 10 out of 93K+ highways



- **I-5 N** had **the highest** 26645 (1.75%) no. of accidents; averaging 459 accidents monthly.
- **I-95 N** had **the 2nd highest** 22407 (1.47%) no. of accidents; averaging 386 accidents monthly.

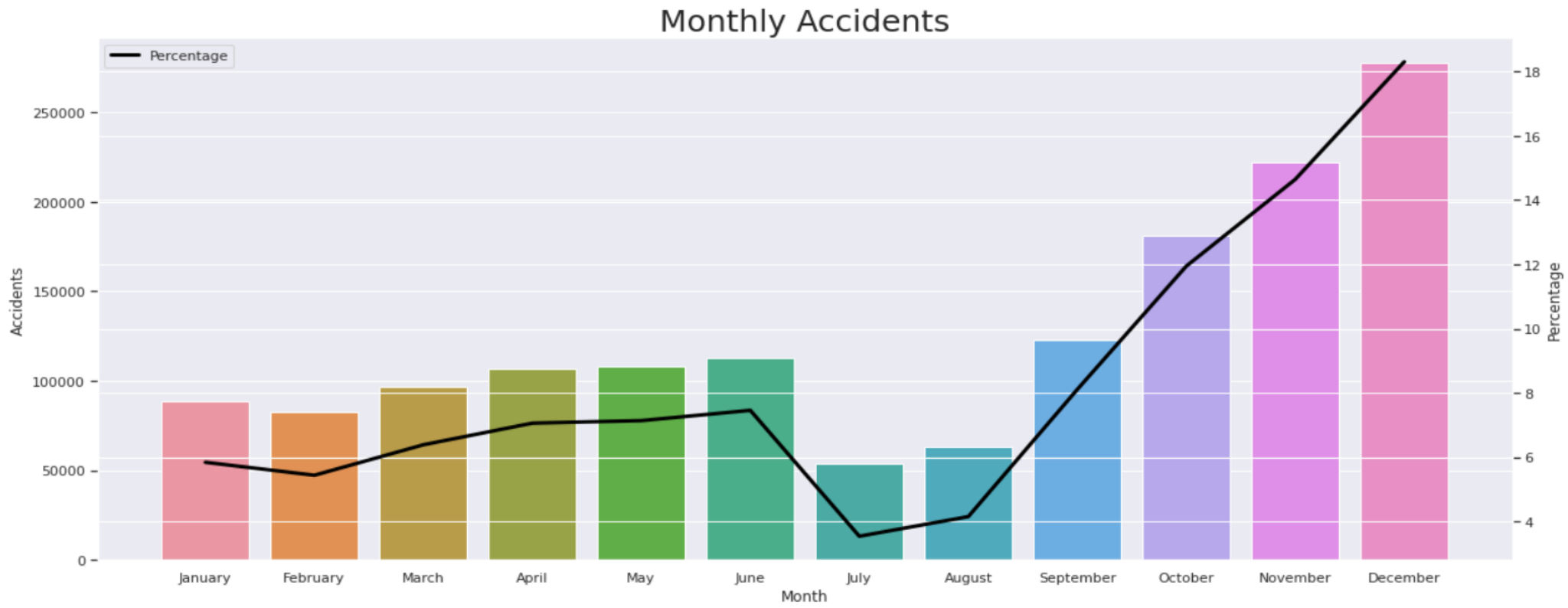
Time Analysis

Number of Accidents incremented from 2016 to 2020*



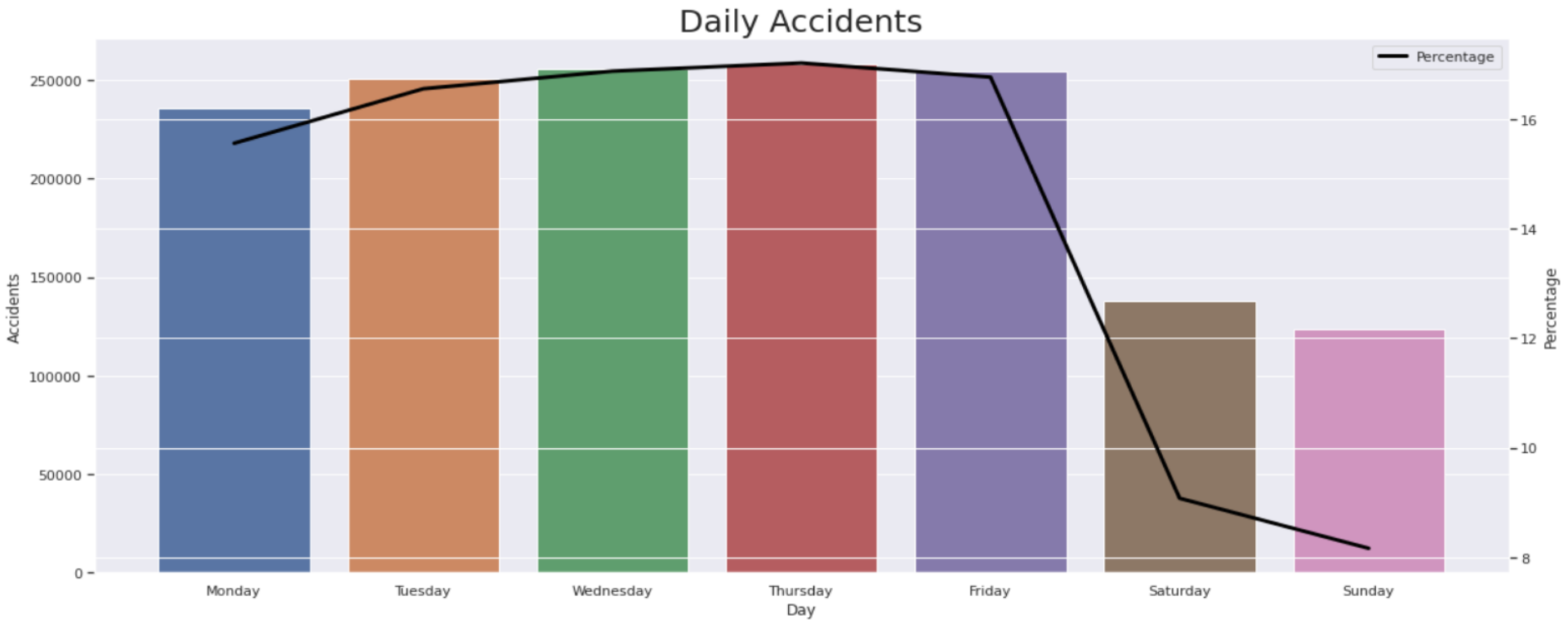
- Year **2020** had **the highest** 787932 (52.0%) and ~**half of the no. of accidents.**
- Year **2016** had **the least** 129325 (8.5%) no. of accidents.
- No. of accidents had been increasing from year 2016 to 2020 annually*.
  - *Year 2018 had a 0.2% drop in accident from year 2017.

Avg. ~52% Accidents observed between Sept. to Dec.



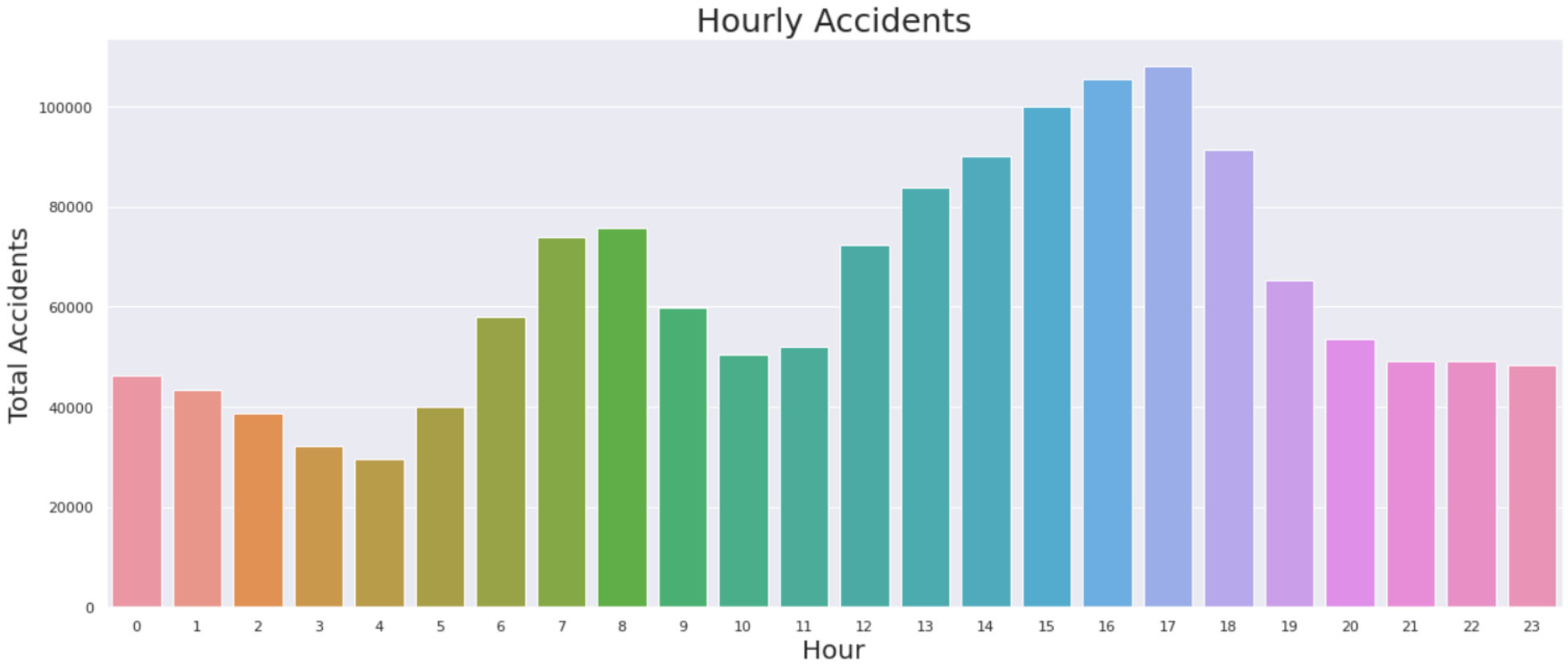
- **December** had **the highest** 277489 (5.84%) no. of accidents.
- No. of accidents had been increasing **from Sept. to Dec.** accounting for **half of the accidents avg. 52%.**
- **July** had **the least** 53650 (3.53%) no. of accidents.

Weekdays show the highest and most no. of accidents ~ avg. 82%



- **Thursday** had **the highest** 258036 (17.02%) no. of accidents.
- **Saturday** and **Sunday** had **the least** 261395 (17.26%) no. of accidents.

Analysis shows majority accidents happened during rush hours

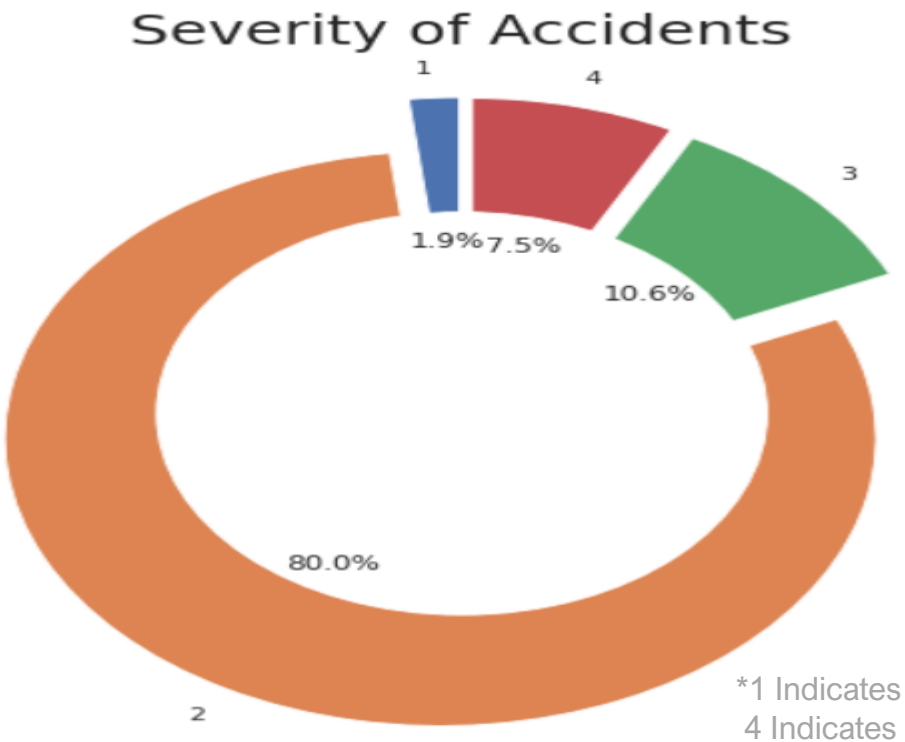


- **Nearly 45.0%** of the accidents happened **in peak traffic hours.**
  - During morning (between 6 to 9 am; 18.0%) and evening (between 3 to 6 pm; 27.0%) peak traffic hours.

- The highest 108011 (7.12%) no. of accidents happened at 5 pm.

Accident Severity Analysis

Majority Accidents showed an average impact on traffic

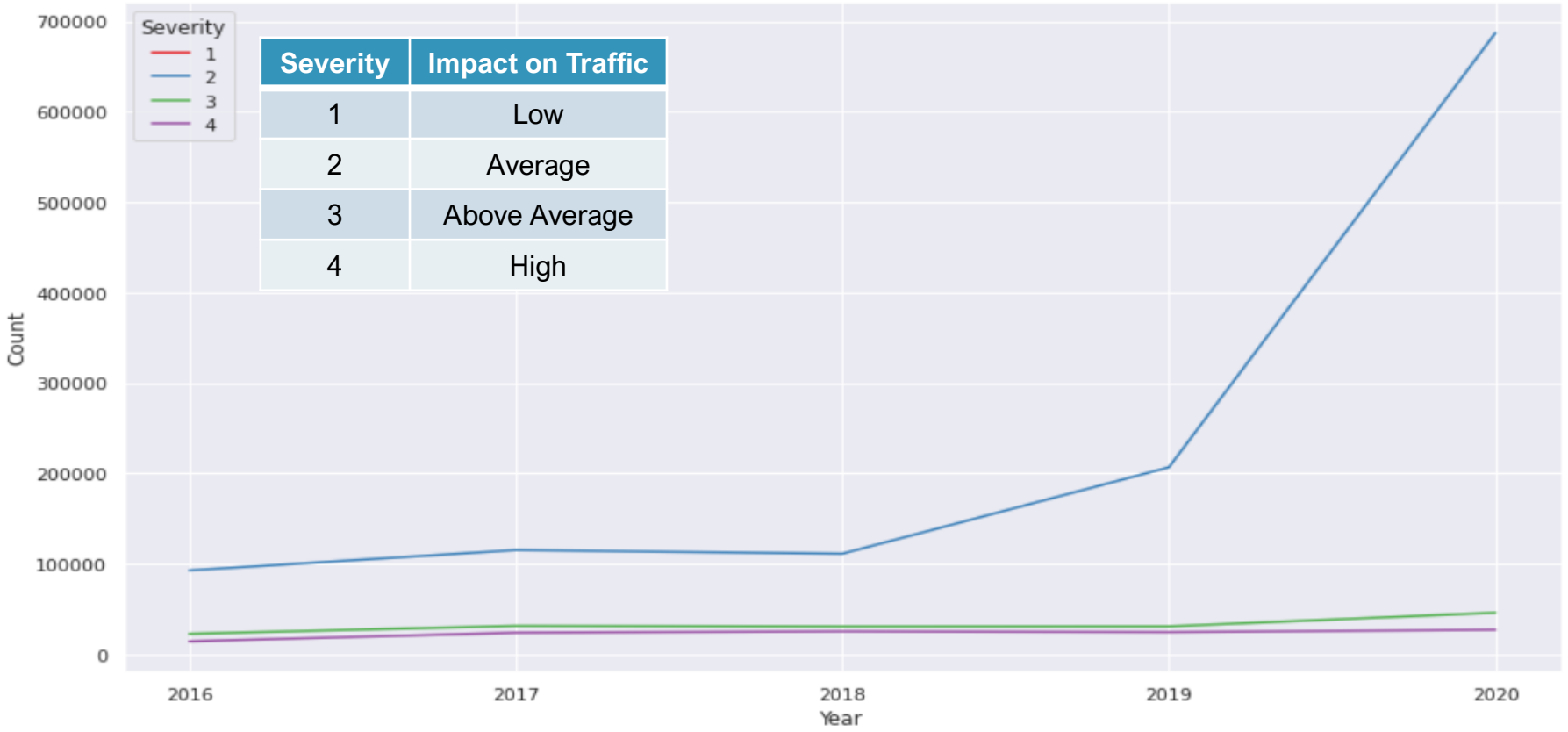


Severity	Impact on Traffic
1	Low
2	Average
3	Above Average
4	High

*1 Indicates least impact on traffic (i.e. Shorter delay due to an accident) &
4 Indicates significant impact on traffic (i.e. longer delay)

- Nearly 80.0% of the accidents shown a moderate impact on the traffic.
- 7.5% of the accidents with a severe impact on the traffic.

Significant Increment in Severity 2 pattern Accidents

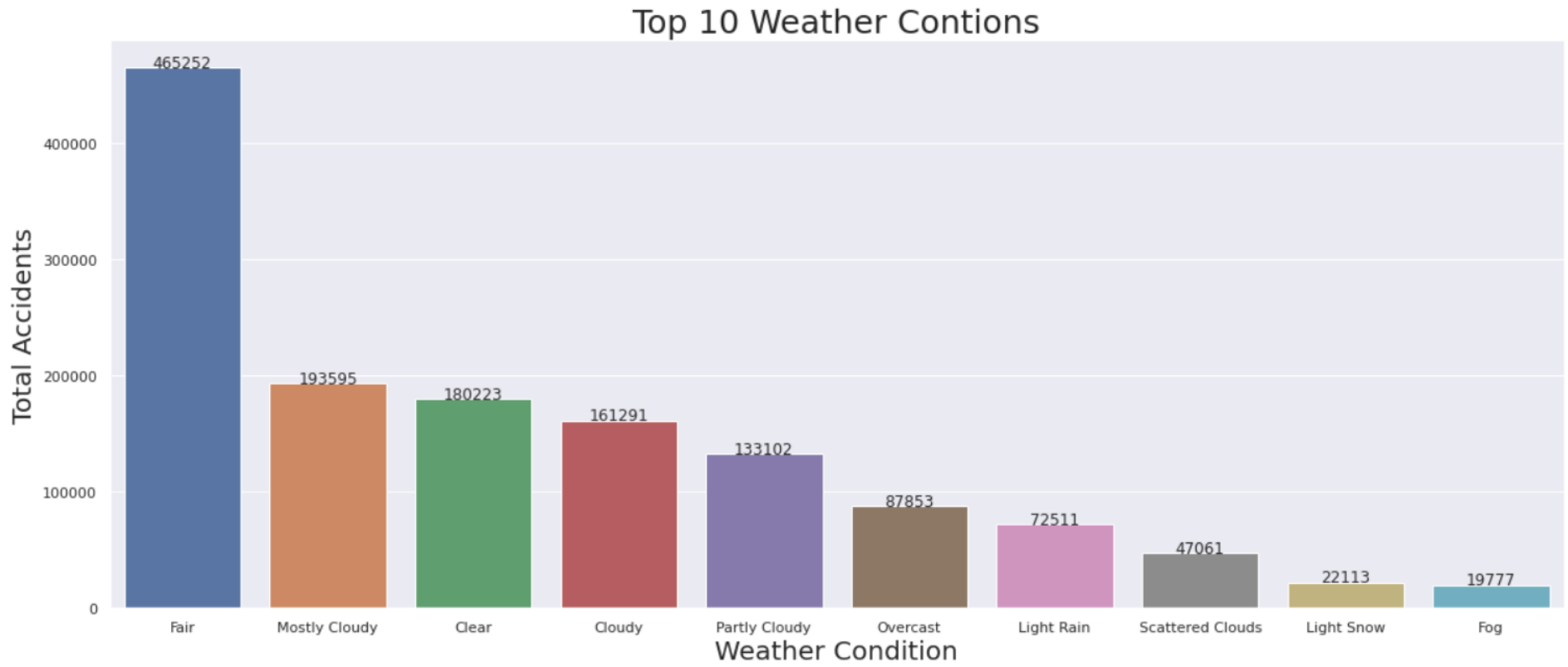


Severity	Impact on Traffic
1	Low
2	Average
3	Above Average
4	High

- For each severity, number of accidents shows increment for each year
- Severity 2 accidents are most occurred accidents every year and increased drastically in 2020
- Severity 3 accidents also increased a bit from year 2019 to 2020.
- Severity 1 and 4 showing relatively low numbers of accidents compared to Severity 2

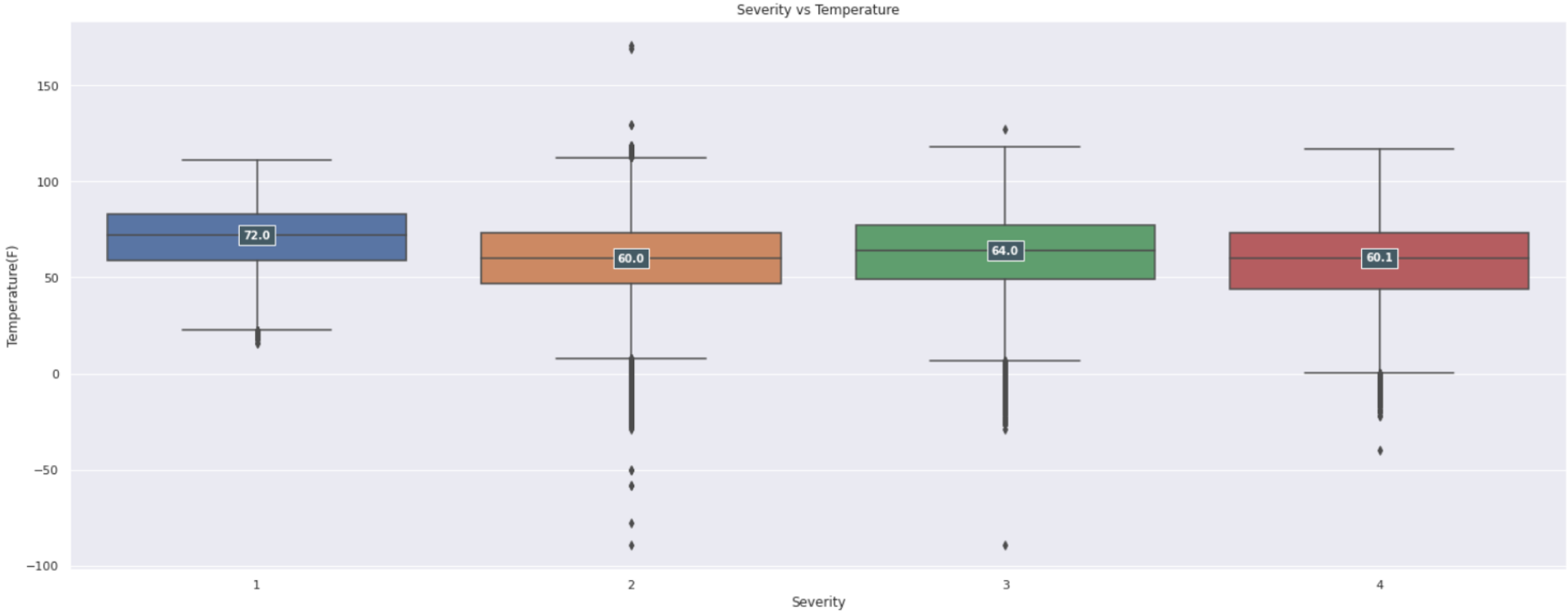
Weather Condition Analysis

Weather conditions displayed minor role in accidents



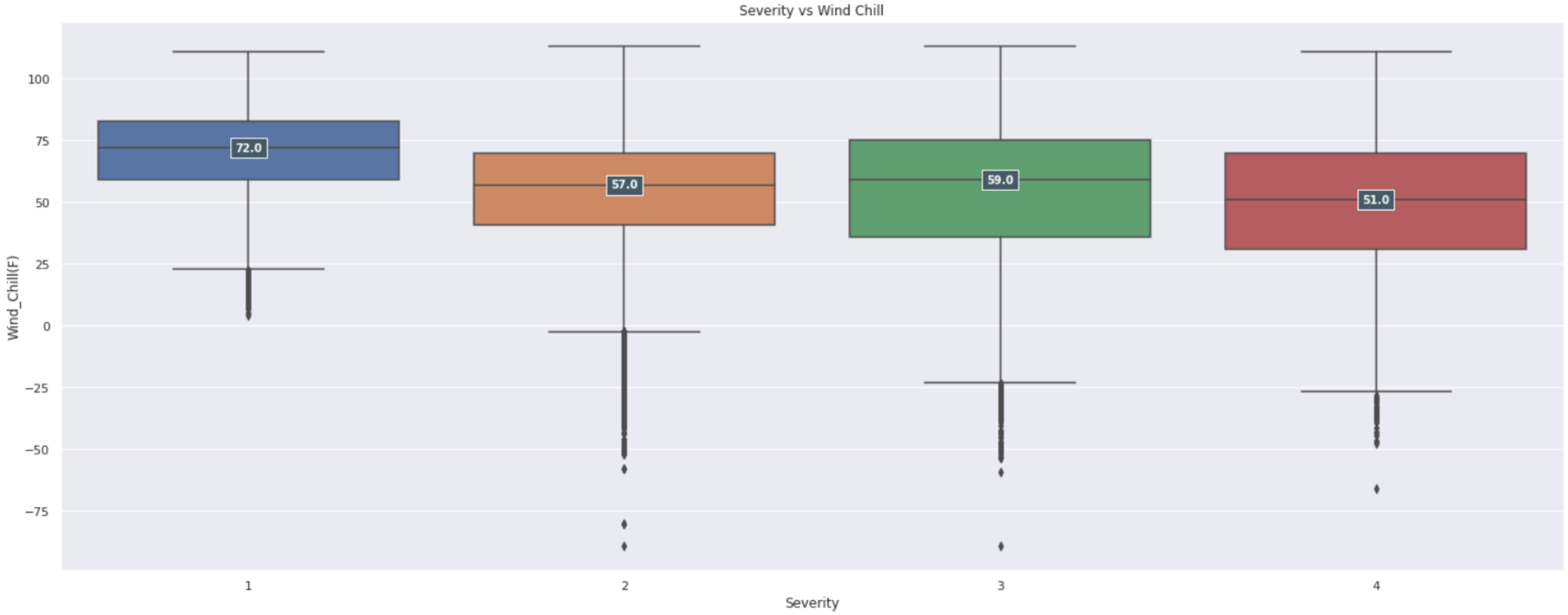
- **Most accidents** happens **in a fair-weather condition** 465252 (31.60%).
- Mostly cloudy weather condition rates 2nd with the 193595 (13.15%).

Temperature played a minor role on Severity of Accidents



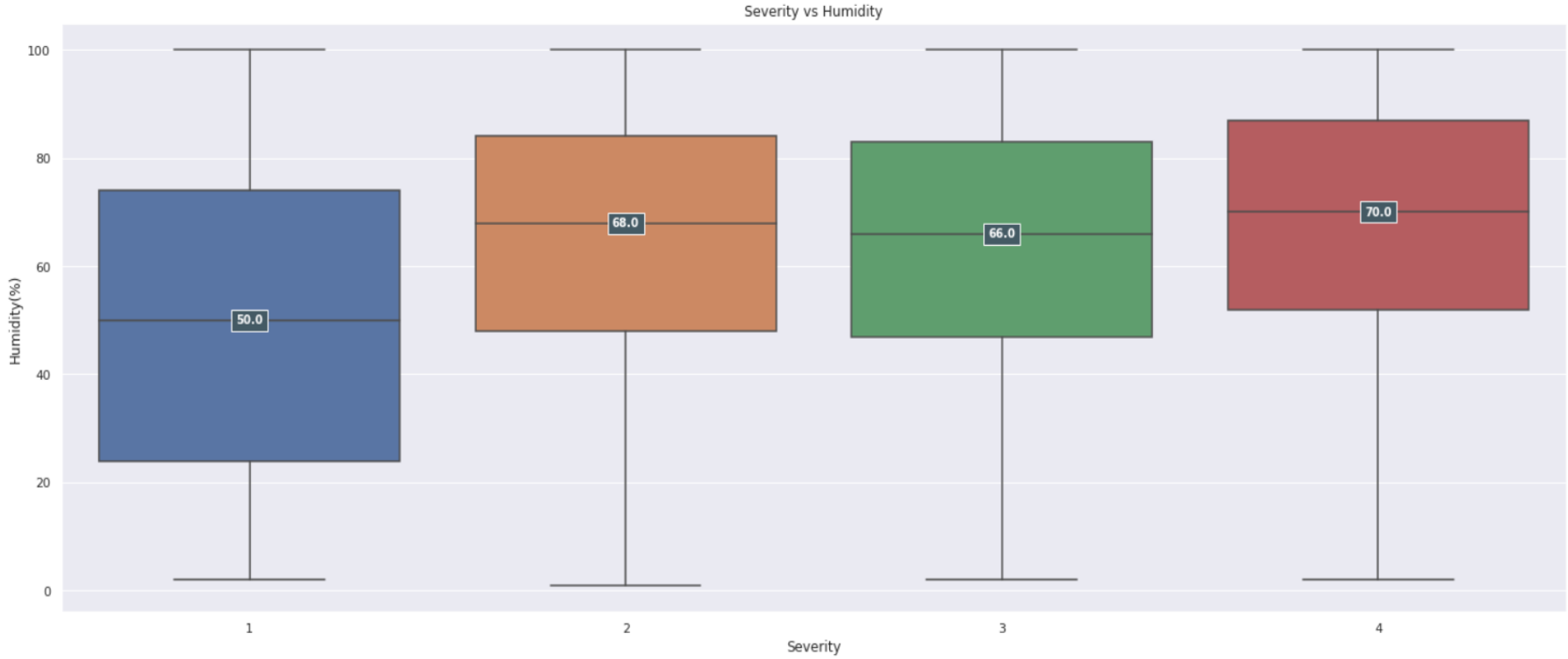
- Severity 2, 3 and 4 have small difference in median temperature.
- severity 4 indicates that lower temperature might result to more severe accidents.

Minor correlation between Severity and Wind Chill



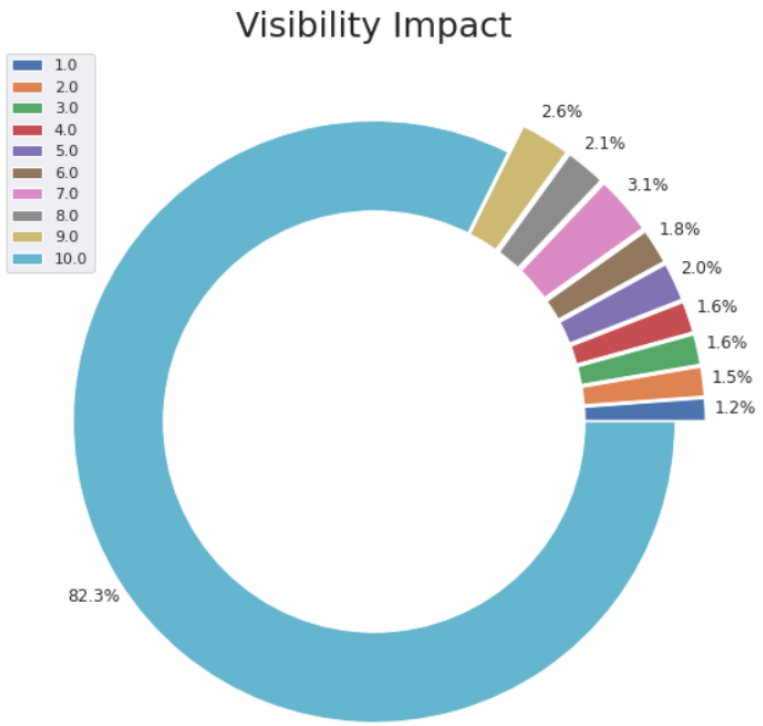
- severity 4 indicates that lower wind chill might result in more severe accidents.

Accident severity increased with higher humidity



- Severity 3 accidents lowers a little bit as humidity decreases.

Visibility is not significant concern for accidents.



- Nearly 1178150 (82%) of accidents occurs when the visibility is the best 10mi.
- Only 16643 (1.2%) of accidents occurs when the visibility is the worst 1mi.

Key Take Away

- Majority of accidents happen in east and pacific time zones on highway I-5 and 95.
- California (LA city) and Florida (Miami) states has highest no of accidents
- Nearly 52% of accidents occurred in year 2020.
- Majority of accidents happened during weekdays and rush-hours.
- Significant increment in severity 2 accidents over the years and showed an impact on traffic.
- Weather conditions and visibility played minor role in accidents.

Potential Area of Further Investigation/Next Steps

- Analysis with large-scale & most recent data
- Analysis based on -
  - Road factors (e.g. Speed limit, Pavement condition, Conflict access points etc.)
  - Human Factors (e.g. Age, Gender, Impaired Driving, Distractions, avoiding safety restraints etc.)
  - Vehicle Factors (e.g. Vehicle type, year, maneuverability (Size), safety features etc.)

Thank you !