Fatal Crashes in North Carolina for the Year 2021

Presentation by Aline Vo, Brian Stumm, and Annie Josephrajan

Introduction

1,659 persons were killed in car crashes in NC in 2021

When are the high-risk times that these fatal accidents occur?

- Hour of the day
- Day of the week
- Month of the year

Where are the accidents occurring?

- By County

Who would benefit?

- General public
- Traffic enforcement
- Civil planning
- Legislature
- Hospitals
- Insurance companies

The Data

Sources and Interpretation

Data Sources

- NHTSA – The National Highway Transportation Safety Administration
- FARS – Fatality Analysis Reporting System
  - FARS is a nationwide census providing NHTSA, Congress and the American public yearly data regarding fatal injuries suffered in motor vehicle traffic crashes.
- Data goes as far back as 1975 and files are organized by year

Data Cleaning

- - Retrieved FARS data from nhtsa.gov
- - Removed extra columns from file
- - Narrowed search to just North Carolina
- - Dropped rows that contained an hour range as “Unknown Hours”
- - Create CSV file of fatalities for North Carolina only

```
filtered_df = ncstate_df[ncstate_df['Hour_Range'] == "Unknown Hours"]

# Display the first 15 rows of the filtered DataFrame
filtered_df.head(5)
```

County	Monthid	Month	Day	Day_of_Month	Day_of_Week	Weekday	Year	Hour	Hour_Range	Minute	Minute_Digit	Latitude	L
AY (43)	2	February	26	26	6	Friday	2021	99	Unknown Hours	99	Unknown Minutes	35.035628	
_FORD (81)	4	April	6	6	3	Tuesday	2021	99	Unknown Hours	99	Unknown Minutes	35.948942	
/ILKES (193)	6	June	5	5	7	Saturday	2021	99	Unknown Hours	99	Unknown Minutes	36.149217	

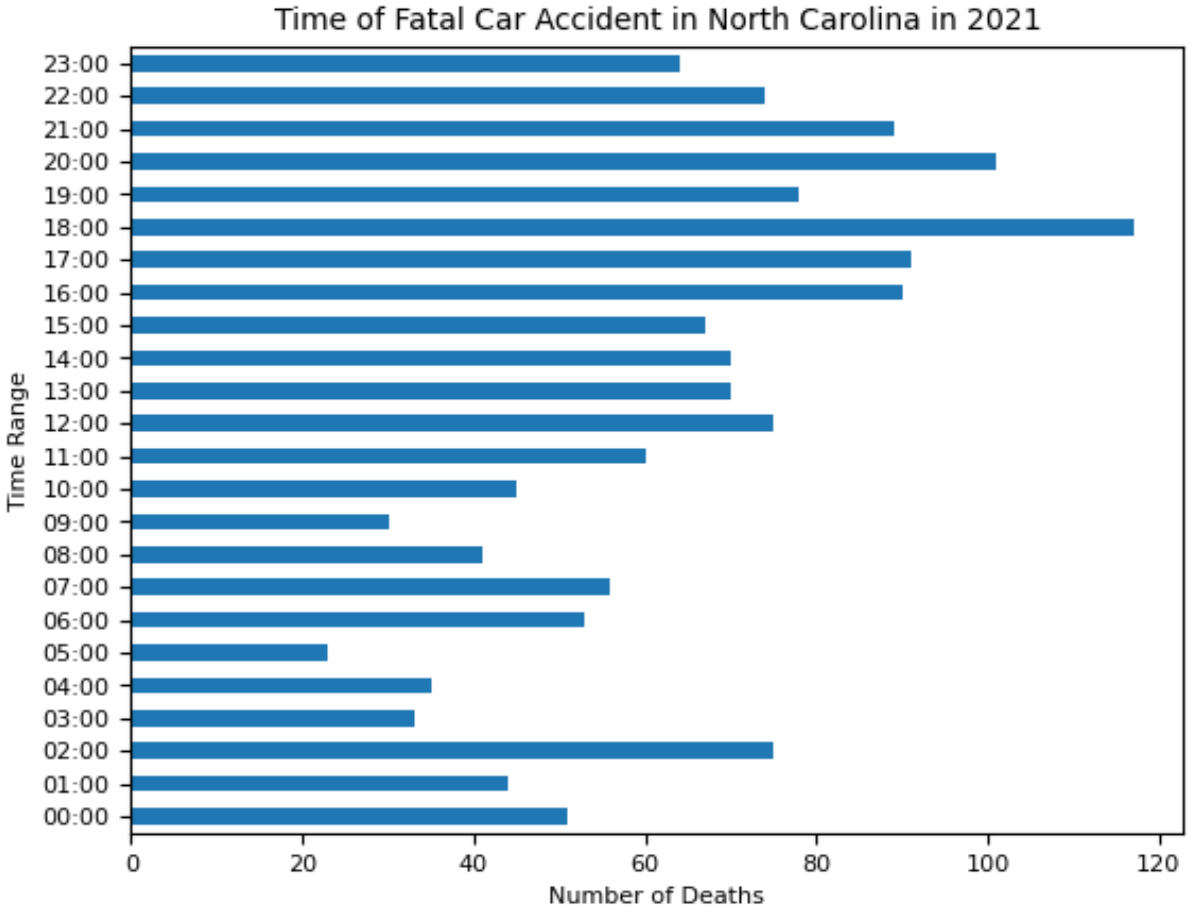
SQL Database

- Created table and imported the data into SQL
- Wrote SQL scripts used to generate graphs for the dashboard (Slemma)
- Built interactive reports using the SQL scripts generated

Data Analysis

Visualizations

Time



Method:

- Imported Matplotlib.pyplot
- Split ‘Hour Range’ Column to show the first hour, then formatted to military time
- Created new data frame with groupby and count

Results:

- The most fatal accidents occur at 6:00PM (117) with 8:00PM (101) coming in second
- The least amount of fatal accidents is at 5:00AM (23)

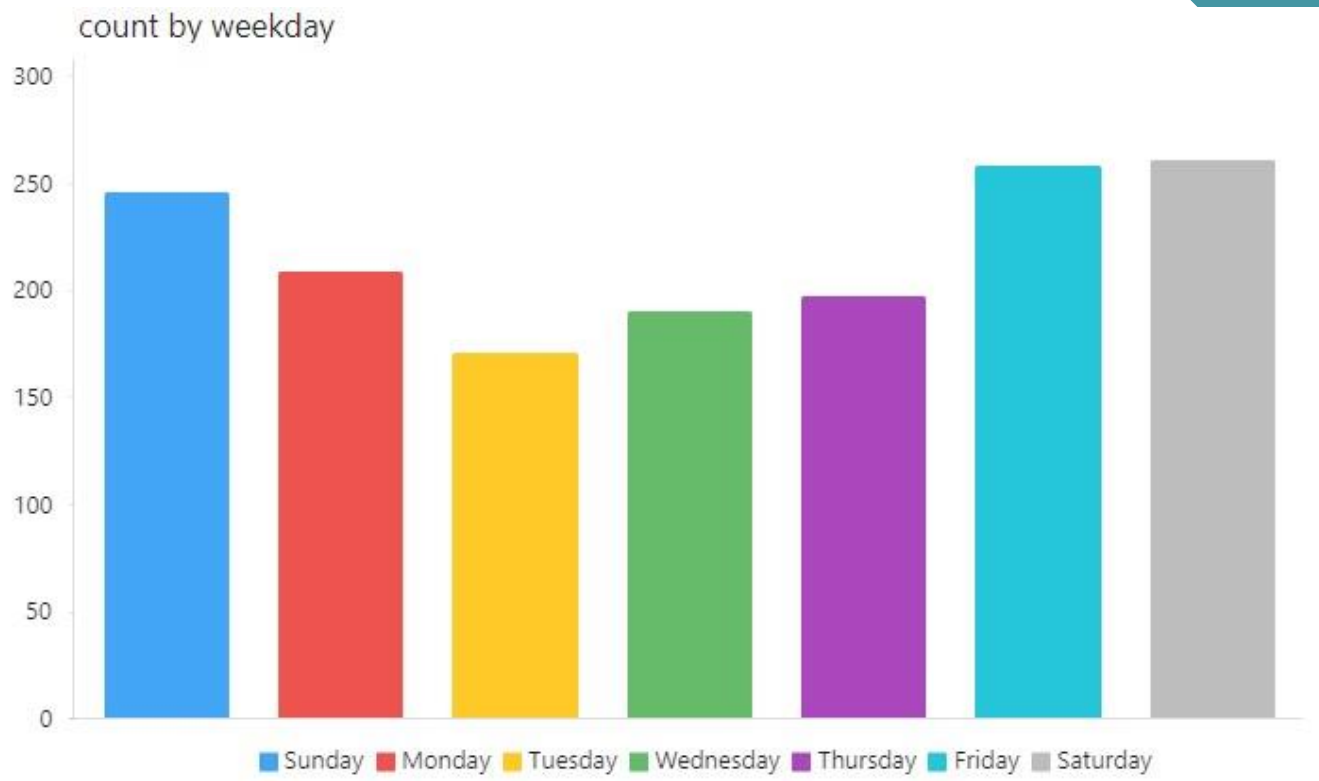
Days of the Week

Method:

-Imported csv file from sql script
-Grouped and ordered by weekday
-Imported data:
    -Matplotlib.pyplot
    -Visualization tool Slemmacount

Results:

- -The most fatal accidents occur on the weekends



Comparison of Years for Weekdays

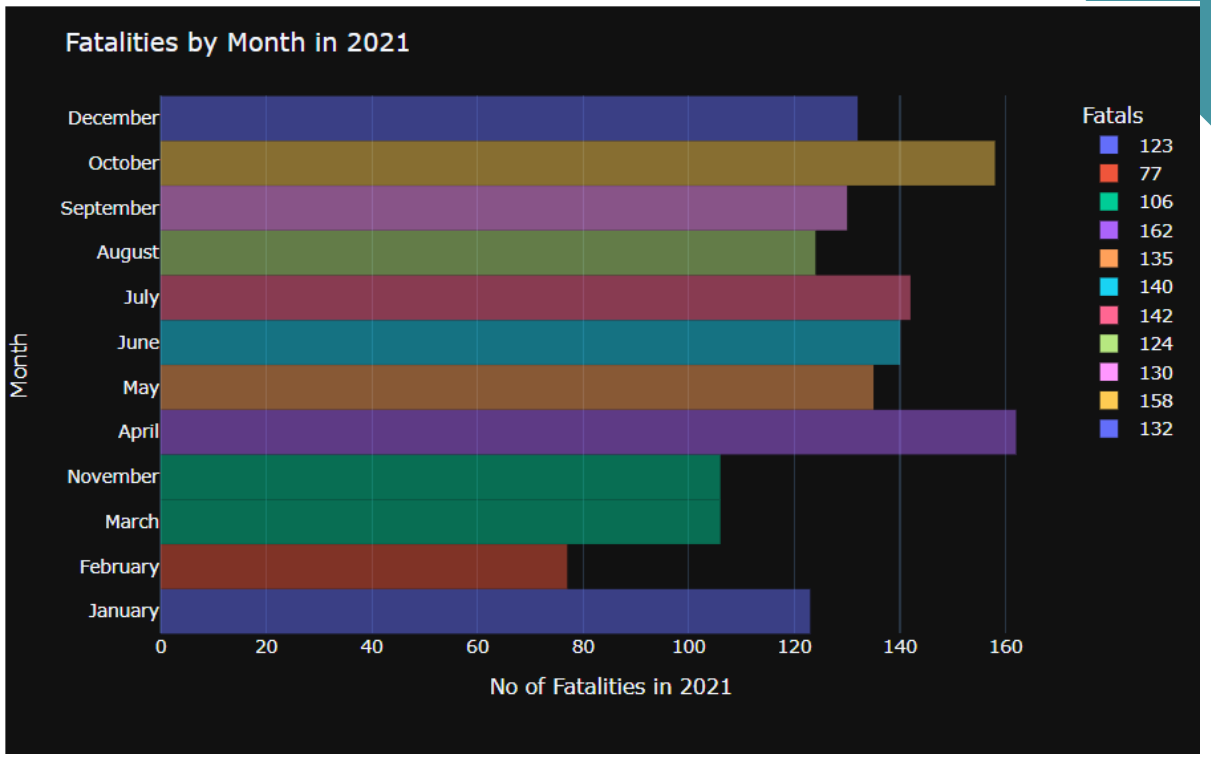
Months of the Year

Tool Used: plotly.express module

- High level data visualization package that allows you to create interactive plots with very little code
- provides functions to visualize a variety of types of data.
- Easy to use

Data Clean Up:

- Import the data
- Created new data frame with desired columns
- Changing the data types, rearranging the columns, extract the month names from Date
- Using groupby(), count() method to get the desired DataFrames
- Using the Data Frames visualize the data in different charts

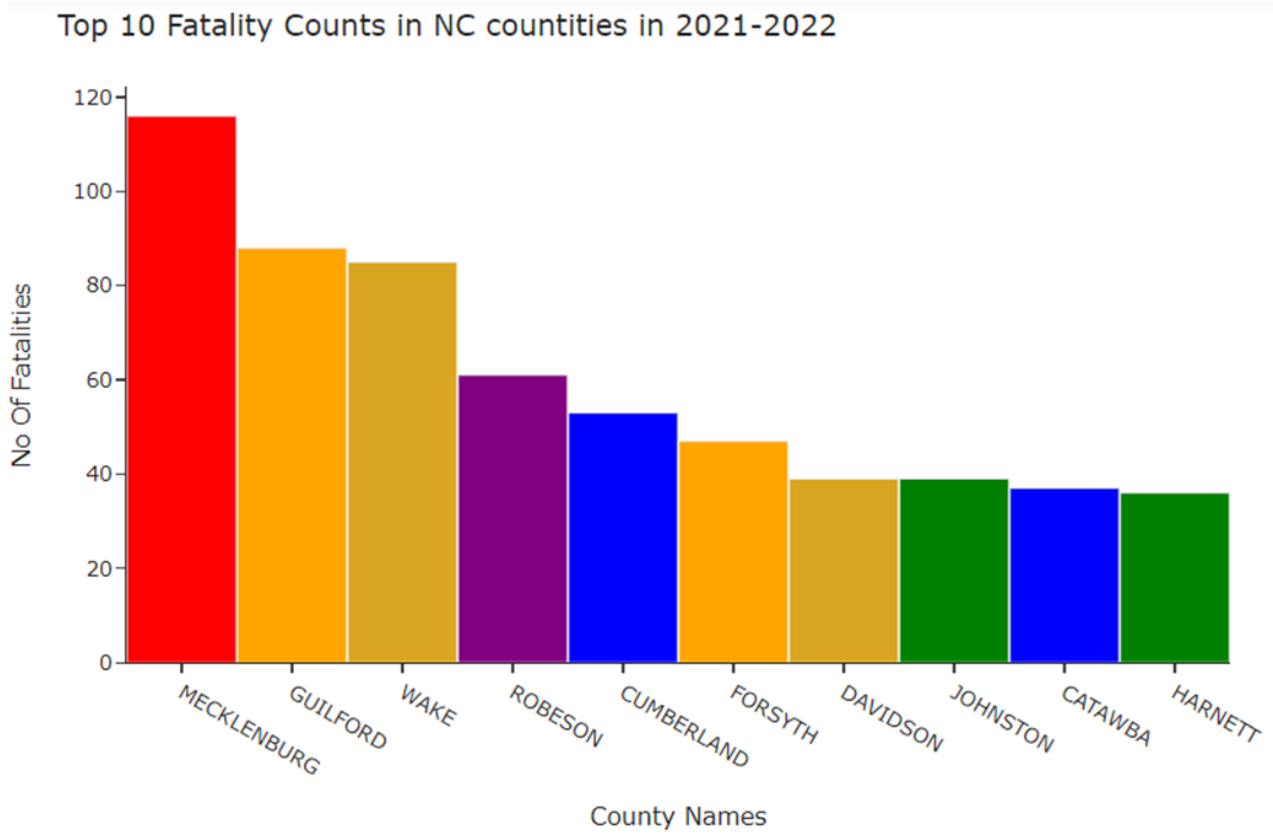


The highest being April - 162 and the lowest being February 77

Top 10 Counties

Results:

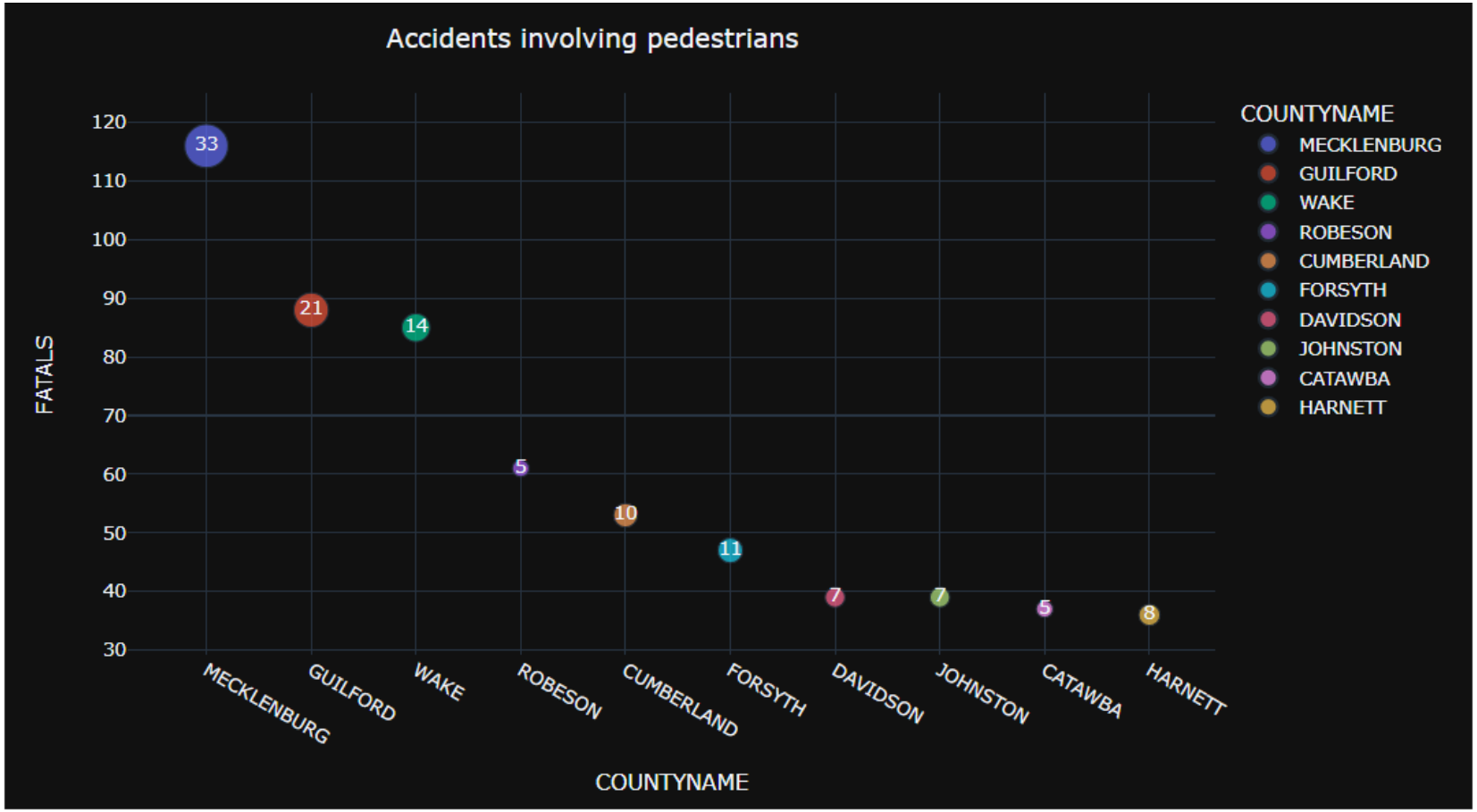
- Bar Chart using the counties and number of fatalities
- Highest - Mecklenburg – 116
- Lowest - Harnett – 36 (of the top 10)



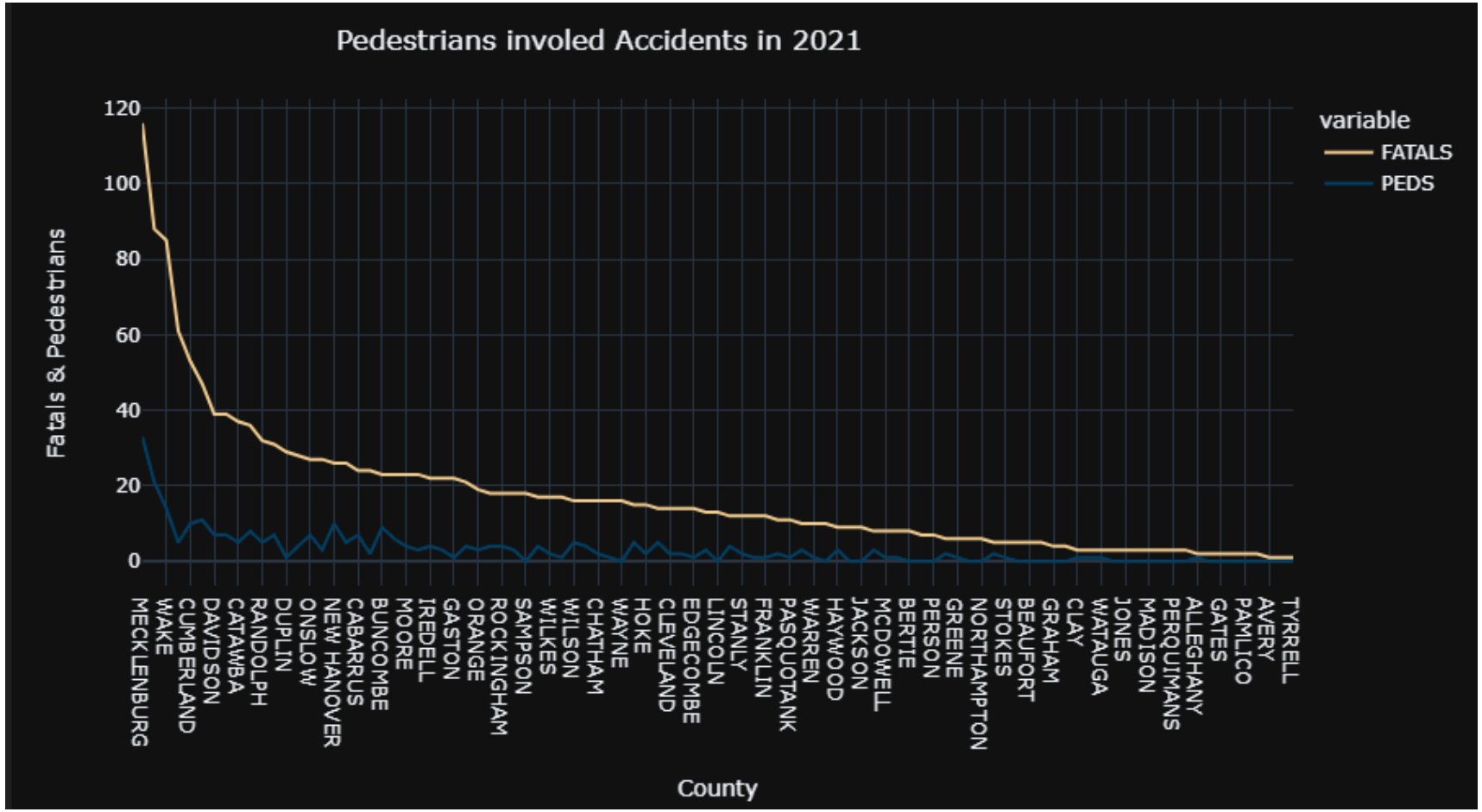
Pedestrians

Results:

- Bubble chart fatal involving pedestrians
- Bubble size represents the percentage of pedestrians involved
- Largest - Mecklenburg – 28%
- Smallest - Robeson – 8% (of the top 10)



Multiple Line Chart



Ethical Considerations

- Gathered by law enforcement
- Did not include identifying data of vehicles or persons
- Possible to find out who was in the accident by searching the date and location

Maps

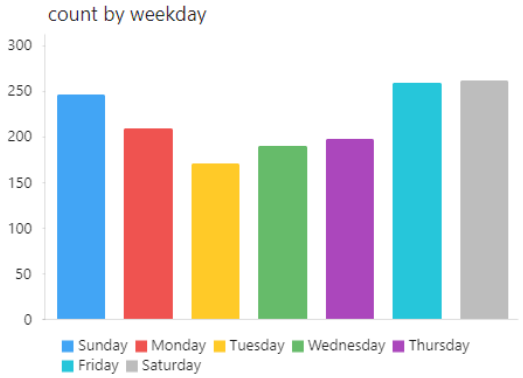
Folium

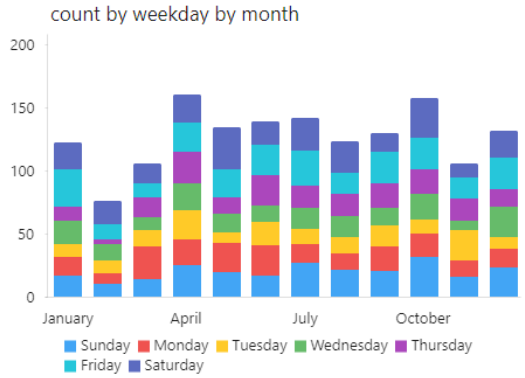
- A Python Library that helps create Leaflet maps
- Data manipulation in Python
- Visualize data in Leaflet Map

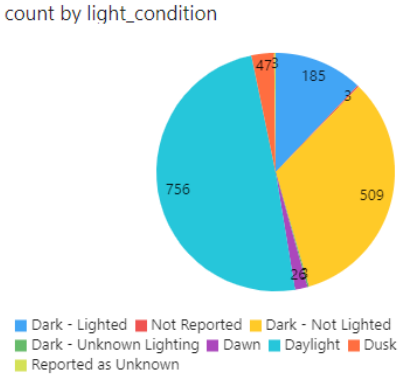
Conclusion

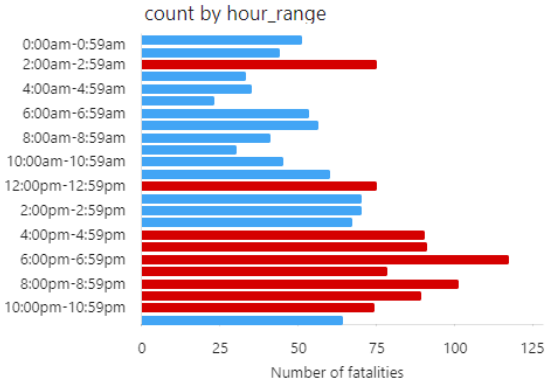
Findings:

- The most fatalities occurred in the evening hours, not in the morning hour commutes
- The most dangerous days of the week for motor vehicle fatalities was the weekend (Friday – Sunday)
- April and October highest fatality months, but February had the lowest fatalities


count by weekday
300
250
200
150
100
50
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday


count by weekday by month
200
150
100
50
0
January
April
July
October
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday


count by light_condition
473
185
3
509
26
756
Dark - Lighted
Not Reported
Dark - Not Lighted
Dark - Unknown Lighting
Dawn
Daylight
Dusk
Reported as Unknown


count by hour_range
0:00am-0:59am
2:00am-2:59am
4:00am-4:59am
6:00am-6:59am
8:00am-8:59am
10:00am-10:59am
12:00pm-12:59pm
2:00pm-2:59pm
4:00pm-4:59pm
6:00pm-6:59pm
8:00pm-8:59pm
10:00pm-10:59pm
0
25
50
75
100
125
Number of fatalities

Slemma Demo

Q&A

Thank you