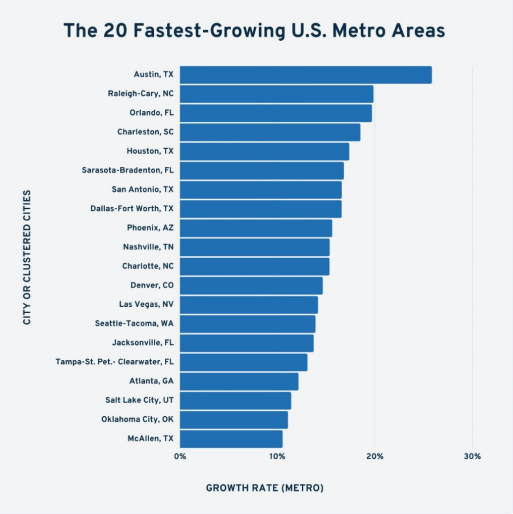
Which cities in North Carolina have the least crime?

Why are we looking into this topic?

- North Carolina is a growing state with Charlotte and Raleigh leading the way
- Industries such as science and finance are becoming a hub in the state
- Majority of those in class live in NC and may move around the state to pursue new opportunities.

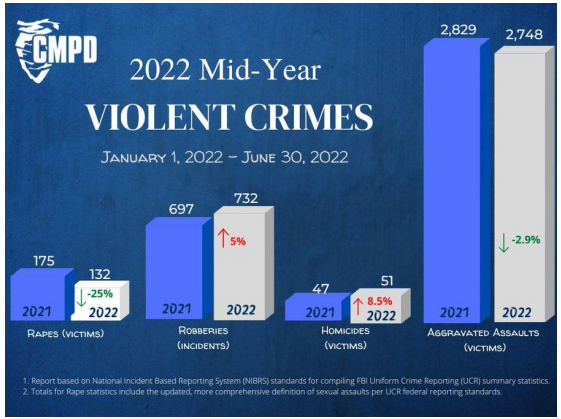


Every city offers something different

- Cost of living differs from city to city
- One may want a more affordable house/cheaper rent but work in a higher cost of living area
- One may want a suburb feel or a city feel
- A common example we have seen is up north. Many way work in NYC however live in New Jersey where the cost is cheaper. (New Jersey is on average about 30% cheaper to live in than NYC)

Why does this deal with city safety?

- For those wanting to move anywhere in the state or to the state, it is important to have an understanding of where you're going.
- Safety should be a top concern as the environment is new
- Seeing trends in city safety overs years can give direction to the future



What we are looking to explore?

1. Which cities have the least crime?
2. What cities are safe in North Carolina?
3. What is the difference in crime from 2016 and 2019?

Data Source

- The data sets we used to explore were from the FBI and we choose to use the years 2019 and 2016.
- These data sets contained violent and non-violent crimes.

Data Exploration

What we were looking for in our data

- Recent crime reports
- Credible website we can download information from (in our case it was an excel file)

Table 6

NORTH CAROLINA
Offenses Known to Law Enforcement
by City, 2016

Data Declaration Download Excel Table 6 State Listing

City	Population	Violent crime	Murder and nonnegligent manslaughter	Rape (revised definition[1])	Rape (legacy definition[2])	Robbery	Aggravated assault	Property crime	Burglary	Larceny-theft	Motor vehicle theft	Arson
Aberdeen	7,549	25	0		3	10	12	289	52	226	11	1
Ahoskie	4,883	43	2		3	8	30	291	81	201	9	0
Albemarle	16,024	107	1		6	24	76	771	201	538	32	4
Angier	5,097	14	0		2	2	10	129	62	59	8	3
Apex	47,324	40	1		2	13	24	601	83	501	17	1
Asheboro	26,230	104	1		11	39	53	1,501	300	1,161	40	10
Asheville	89,546	537	10		48	144	335	4,357	752	3,345	260	7
Atlantic Beach	1,503	14	0		4	1	9	130	42	84	4	0

Data Analysis Phase

- The first cleaned our data to drop columns/rows that were titles and double counted values
- We determined a threshold that the model would learn to detect a safe and unsafe city in NC using excel by calculating a crime score.
- Total crimes = Property Crimes + Violent crimes

Total Number of crimes/Population of the city*100= Crime Index Score

- We needed to create definition of “safe” which we determined by using the third quartile.

Data Analysis Phase

- In SQL, we created three tables and a database that our machine learning model would pull from.
- We used a Logistic Regression Model because it predicts binary outcomes (ie. safe or unsafe)
- This model analyzes the available data, and when presented a new sample, mathematically determines its probability of belonging to a class. If the probability is above a certain cutoff point, the sample is assigned to that class. If the probability is less than the cutoff point, the sample is assigned to the other class.

Machine Learning Model

Logistic Regression Model

- **Limitations:**
  - It is difficult to capture complex relationships using logistic regression. More powerful and complex algorithms such as Neural Networks can easily outperform this algorithm.
- **Benefits:**
  - Good accuracy for many simple data sets and it performs well when the dataset is linearly separable.
  - It is simple to implement and doesn’t require high computation power.