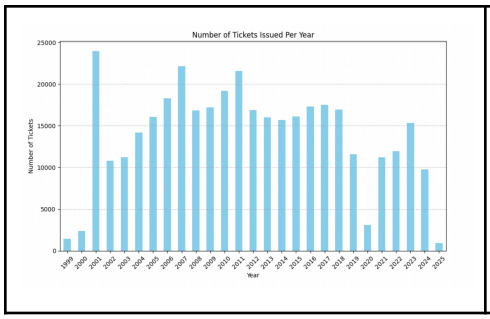
Data Appendix

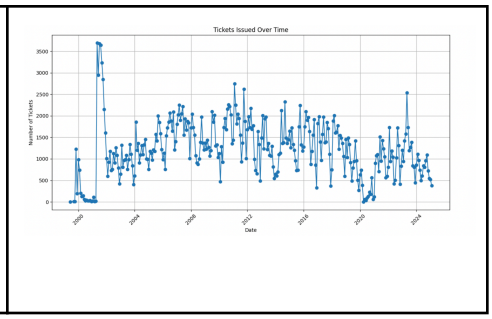
For this project, data was obtained from the City of Charlottesville, which details parking tickets issued from 2009 to 2025. This data is available as Parking_Tickets_Original.zip. The data were cleaned, resulting in the cleaned_parking_tickets.zip file.

cleaned_parking_tickets.zip (cleaned_parking_tickets.csv)

- This dataset consists of 10 columns: [“RecordID”, “Ticket Number”, “DateIssued”, “StreetName”, “TimeIssued”, “StreetNumber”, “LicenseState”, “ViolationDescription”, “Location”, “LicensePlateAnon”]. Each row of the data represents a unique traffic ticket issued by the City. The dataset contains 375,477 entries

- Details of Each Variable:
    1. RecordID:
        a. Definition: Number given to each record used as unique identifier
        b. Data Type: Integer
        c. Processing Steps: N/A Values Removed
        d. Responses: 0, 1, 4, etc
    2. Ticket Number
        a. Definition: Number given to each ticket used as unique identifier
        b. Data Type: Integer
        c. Processing Steps:
            i. N/A Values Removed
            ii. Duplicate Values Removed
        d. Responses: 69692, 0880773, etc
    3. DateIssued
        a. Definition: Date the ticket was issued, in YYYY/MM/DD
        b. Data Type: Datetime Value
        c. Processing Steps:
            i. N/A Values Removed
            ii. Converted to python datetime value
            iii. Restricted to 2000-2024
        d. Responses: 2015/10/30, 2022/01/27, etc
        e. Number of tickets issued over time





        f. Statistics

	DateIssued
count	453013
mean	2010-04-07 10:58:33.307713024+00:00
min	1999-02-10 05:00:00+00:00
25%	2003-06-13 04:00:00+00:00
50%	2009-10-20 04:00:00+00:00
75%	2016-03-11 05:00:00+00:00
max	2208-11-08 05:00:00+00:00

4. StreetName
    a. Definition: Street the ticket was issued
    b. Data Type: String
    c. Processing Steps:
        i. N/A Values Removed
        ii. Street Names Standardized with Regular Expression Patterns
            1. A python function was created to standardize street names so that variations in the name (ex. ST vs Street, etc) were accounted for

```
# Standardize common abbreviations
replacements = {
    r'\b(\d+)(st|nd|rd|th)\b': r'\1',  # Remove ordinal suffixes
    r'\bstreet\b': 'ST',
    r'\bavenue\b': 'AVE',
    r'\bav\b': 'AVE',
    r'\bboulevard\b': 'BLVD',
    r'\broad\b': 'RD',
    r'\bdrive\b': 'DR',
    r'\bcourt\b': 'CT',
    r'\blane\b': 'LN',
    r'\bplace\b': 'PL',
    r'\bparkway\b': 'PKWY',
    r'\bterrace\b': 'TER',
    r'\bhighway\b': 'HWY',
    r'\bsquare\b': 'SQ',
    r'\bway\b': 'WAY',
    r'\bnortheast\b': 'NE',
    r'\bnorthwest\b': 'NW',
    r'\bsoutheast\b': 'SE',
    r'\bsouthwest\b': 'SW',
    r'\bnorth\b': 'N',
    r'\bsouth\b': 'S',
    r'\beast\b': 'E',
    r'\bwest\b': 'W'
}
```

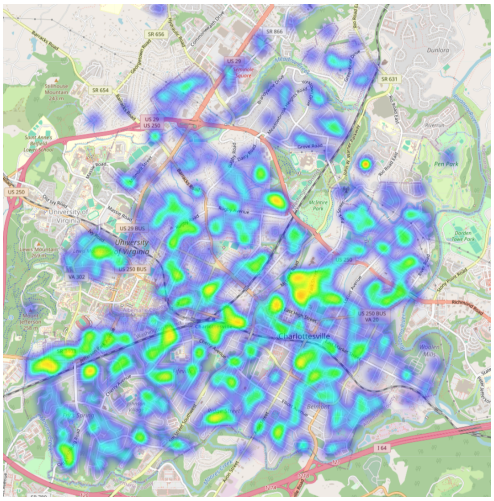
    d. Responses: W WATER ST, 14TH ST NW, etc
5. TimeIssued
    a. Definition: Time the ticket was issued
    b. Data Type: String
    c. Processing Steps:
        i. N/A Values Removed
        ii. Uniform formatting enforced, coercing to HOUR:MINUTE

    d. Responses: 9:58, 12:21, etc
6. StreetNumber
    a. Definition: Street number the ticket was issued at
    b. Data Type: Integer
    c. Processing Steps: N/A Values Removed
    d. Responses: 100, 22, 1700, etc
7. LicenseState
    a. Definition: State of the violating vehicle's license plate
    b. Data Type: String
    c. Processing Steps:
        i. N/A Values Removed
        ii. Invalid Values removed (not a common state abbreviation)
    d. Responses: VA, PA, NJ, etc
    e. Count of 5 States with the Most Violations

LicenseState	count
VA	336681
MD	5652
NC	4373
FL	3680
PA	3589

8. ViolationDescription
    a. Definition: Short description of violation's nature
    b. Data Type: String
    c. Processing Steps: N/A Values Removed
    d. Responses: Void, Curb Painted Yellow, No Parking any time
9. Location
    a. Definition: Composite location where the ticket was issued
    b. Data Type: String
    c. Processing Steps: N/A Values Removed
    d. Responses: 100 W WATER ST, 22 14TH ST NW, etc

e. Heatmap of tickets issued by location



10. LicensePlateAnon
    a. Definition: Anonymized number related to the violating vehicle's license plate
    b. Data Type: Integer
    c. Processing Steps: N/A Values Removed
    d. Responses: 23644, 11385, 142588, etc