DS 4002

03/24/2025

# Anomaly Detection of Parking Ticket Issuance in Charlottesville

Group 14: Kyle Tran (Leader), Balkees Rekik, Evan Stewart

# Project Details

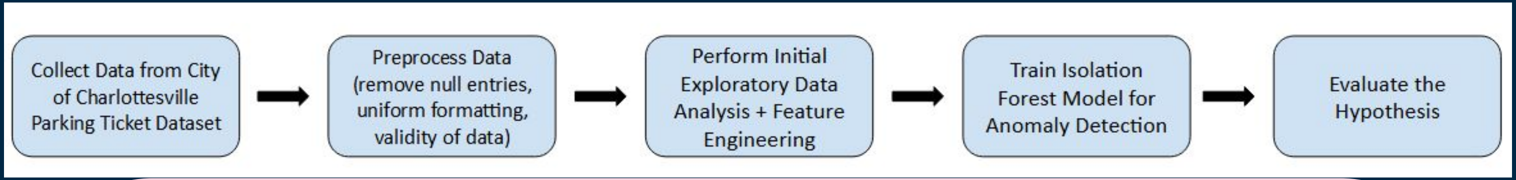
- **Motivation -** Identify the key factors that influence the likelihood of receiving a parking ticket, helping drivers avoid unnecessary expenses
- **Research Question -** Does the time of day, day of week, and/or location influence the frequency of a parking ticket being issued, and can we identify high-risk periods and areas for parking violations?
- **Hypothesis**
  - Null hypothesis (H0): There is no significant difference in ticket issuance frequency between anomalous and normal periods. Any deviations are due to random variation.
  - Hypothesis (H1): There is a significant difference in ticket issuance frequency between anomalous and normal periods, suggesting the presence of real anomalies (e.g., day of the week, time of day).
- **Modeling approach -** Anomaly Detection with Isolation Forest
- **Goal -** Perform an anomaly detection to identify factors that influence the frequency of parking tickets

# The Data

- Dataset: Parking Tickets Issued in Charlottesville
  - OpenData From City of Charlottesville
  - Time-Series Data (1 data file)
  - Data was cleaned by uniform formatting, conversion to datetime object, regex filtering, etc.
  - Cleaned Data contained 375,476 data entries

| Column | Description | Responses |
|---|---|---|
| RecordID | Number given to each record used as a unique identifier | Integer, Ex: 0, 1, 4 |
| TicketNumber | Number given to each ticket used as a unique identifier | Integer, Ex: 69692, 0880773, 0881858 |
| DateIssued | Date the ticket was issued, in YYYY/MM/DD | Datetime Value, Ex. 2015/10/30, 2022/01/27, 2021/06/29 |
| StreetName | Street the ticket was issued | String, Ex: W WATER ST, 14TH ST NW, JEFFERSON PARK AVE |
| TimeIssued | Time the ticket was issued | String, Ex: 9:58, 12:21, 11:00 |
| StreetNumber | Street number the ticket was issued | Integer, Ex: 100, 22, 1700 |
| LicenseState | State of the violating vehicle's license plate | String, Ex: VA, PA, NJ |
| ViolationDescription | Short description of violation's nature | String, Ex: Void, Curb Painted Yellow, No Parking any time |
| Location | Composite location where the ticket was issued | String, Ex: 100 W WATER ST, 22 14TH ST NW, 1700 JEFFERSON PARK AVE |
| LicensePlateAnon | Anonymized number related to the violating vehicle's license plate | Integer, Ex: 23644, 11385, 142588 |

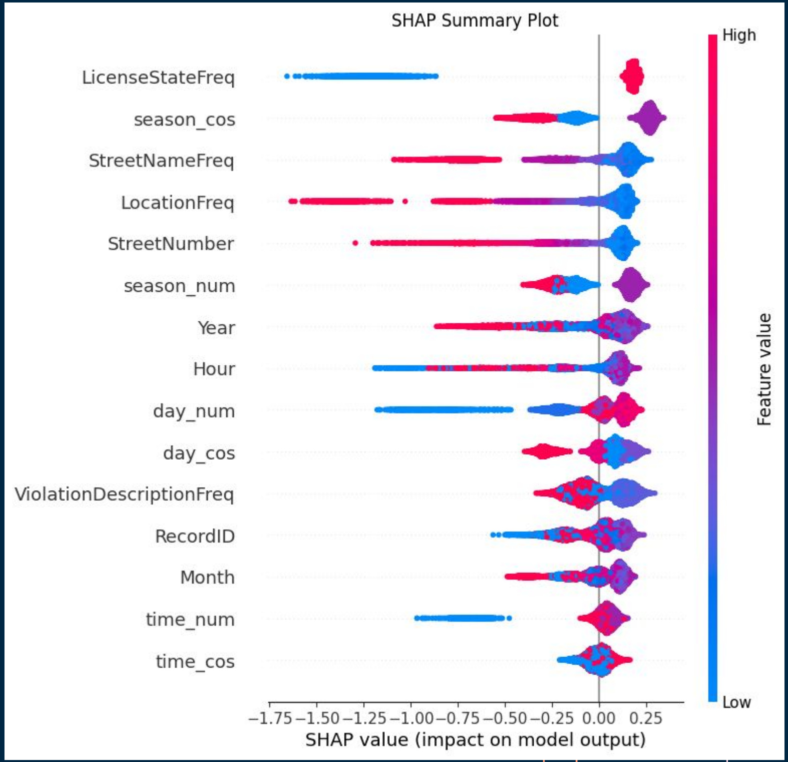
# Analysis Plan




- Utilized Isolation Forest for anomaly detection
  - Often used for Fraud detection, Network Traffic Monitoring, Patient Vital Monitoring
  - Identify data points that are anomalies (outliers) which deviate from expected behavior
  - Unsupervised anomaly detection algorithm, focusing on isolating anomalies
    - Uses binary trees and randomness to isolate data points
  - Produces an anomaly label for each data entry
    - -1 = Anomaly, 1 = Normal
  - Hyperparameters:
    - Contamination, Number of Trees

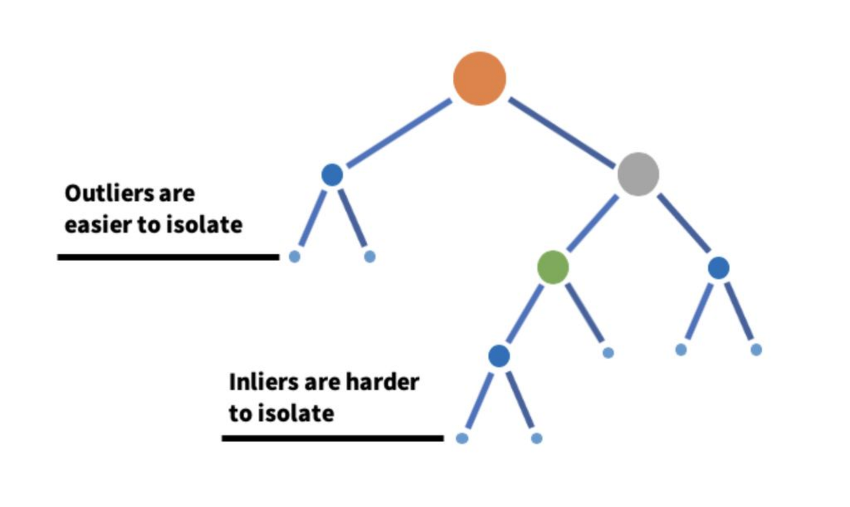
# Analysis Plan Cont.

- SHAP Values were used to identify feature importance
  - SHapley Additive exPlanations
  - Based on game theory
  - Used to see how much a feature contributes to the model output
- We performed a statistical analysis in the ticket issuance rate per year between the Anomalous and Normal tickets
  - Used a t-test



# Tricky Analysis Decision

- We had initially planned to use LSTM and Random Forest models to perform predictive model analysis
- After initial EDA, we realized we didn't have any labelled data to perform prediction
  - We pivoted to Anomaly Detection with Isolation Forest
  - The initial goal was to identify times/areas parking tickets are likely to happen
  - Instead, we looked at outliers to see if we could gain insight on parking ticket issuance

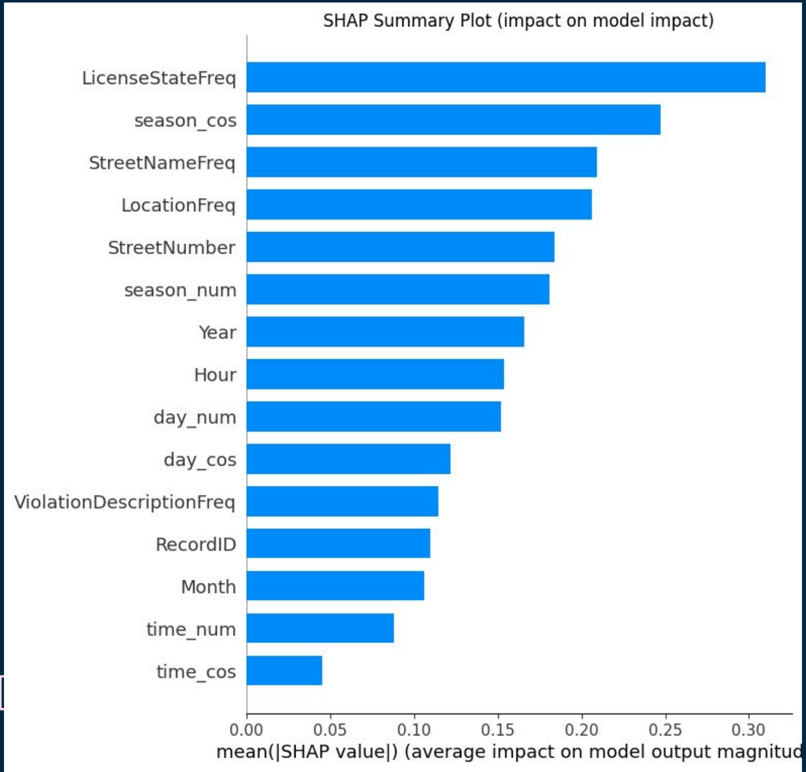


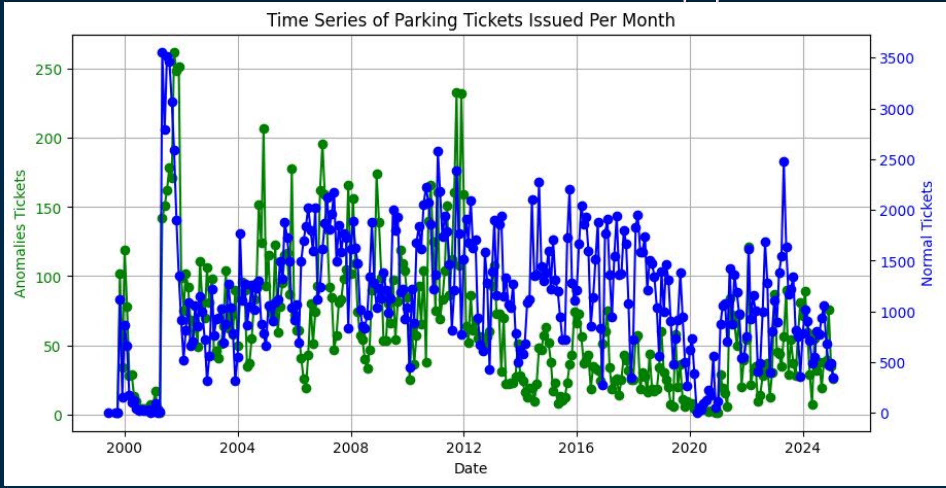
# Bias and Uncertainty Validation

- Biases:
  - The source where data was collected from
    - potential typos, data entered wrong
  - The accuracy of the data (timestamps, location)
- To combat some of the bias, data that had invalid timestamps (invalid formatting) were removed
- Additionally, data cleaning was used to make sure locations had uniform formatting
- Entries that had the same ticket number were removed as ticket numbers were used for unique identification

| | 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 |

# Results and Conclusions



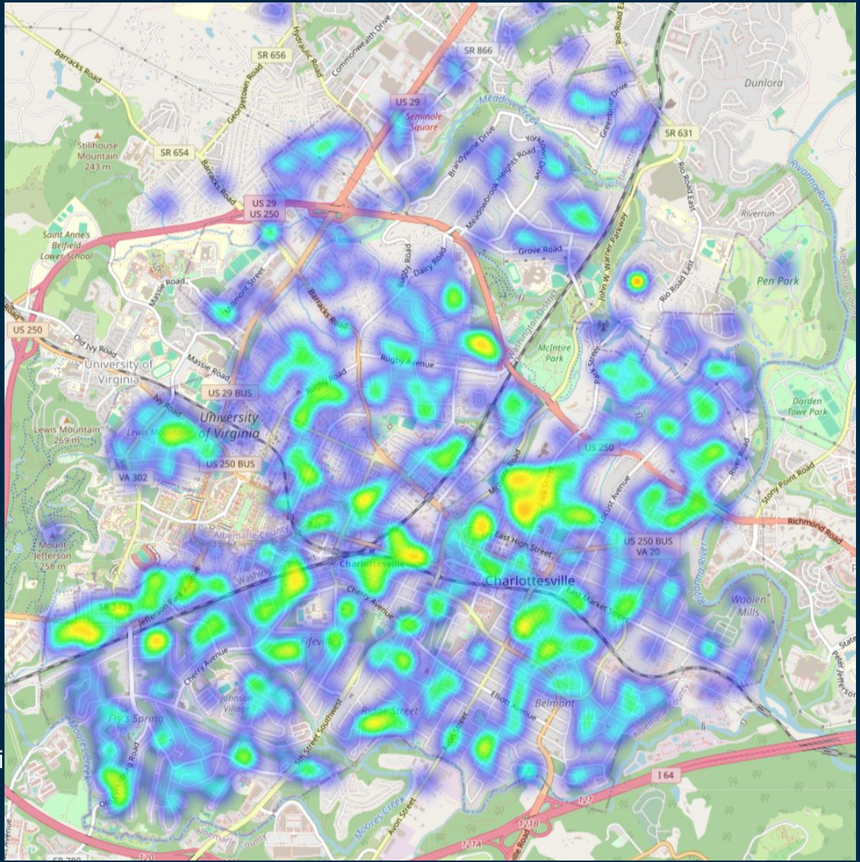

- There is a significant difference in ticket issuance rate per year between anomalous and normal periods
- License Plate State, Street Name, and Season were the most important features that influenced whether data points were considered an anomaly or normal
- T-statistic: 11.12
- P-value: 2.28 x 10^-15

# Next Steps

- New lines of exploration
  - K-means clustering/DBSCAN clustering to explore more trends among tickets being issued
  - Comparing results between different Anomaly Detection models [Autoencoders, Local Outlier Factor (LOF), etc]
- Improvements
  - More in-depth statistical analysis besides t-test
  - The use of Isolation Forest models from different libraries
- New Questions
  - Does adjusting hyperparameters (e.g., contamination rate, number of trees) significantly change the results?
  - Are anomalies clustered in specific time periods (e.g., holidays, weekends, rush hours)?
  - Is it possible to predict the likelihood of a parking ticket being issued based on previous data?

Do you have any questions?

# Thank You!



# References


[1] L. Mucciolo, F. Walker, A. Boddupalli, “The Cost of Parking: A Preliminary Analysis of Parking Tickets Data in Austin, Minneapolis, and Portland,” January, 2023. https://www.urban.org/sites/default/files/2023-01/The%20Cost%20of%20Parking.pdf. [Accessed February 28, 2025].

[2] A. Javaheri et al., "Evaluating the Impact of Smart Parking Systems on Parking Violations," in IEEE Access, vol. 12, pp. 175585-175596, 2024, doi: 10.1109/ACCESS.2024.3503513. https://ieeexplore.ieee.org/document/10758649. [Accessed February 28, 2025].

[3] “Anomaly detection using Isolation Forest,” April, 2024.https://www.geeksforgeeks.org/anomaly-detection-using-isolation-forest/#. [Accessed March 17, 2025].

[4] J. Verbus, “Detecting and preventing abuse on LinkedIn using isolation forests,” August, 2019. https://www.linkedin.com/blog/engineering/data-management/isolation-forest. [Accessed March 22, 2025].

[5] V. Kumar, “Model Explainability For Isolation Forest Using SHAP and LIME,” July, 2023. https://medium.com/@vikaasnreva/model-explainability-using-shap-and-lime-d818e37fd310. [Accessed March 22, 2025]