Camera Traffic Counts

By: John Trelford, Lucas Chiang, Travis Welsh, Brian Pham, and Ritesh Penumatsa
(Group 11)

Dataset Overview

Camera Traffic Counts:

- 15-minute interval traffic by intersection in Austin
- Collected by GRIDSMART optical traffic detectors

Subsetted Data

How?

- Filter to only rows where year = 2019

Why?

- Prevents interference from the COVID-19 pandemic period
- The full Camera Traffic Counts dataset was 82.1 million rows
    - Overwhelms RAM
    - Computationally Intensive

Features

Shape - (Rows; Columns): (14,717,624; 19)

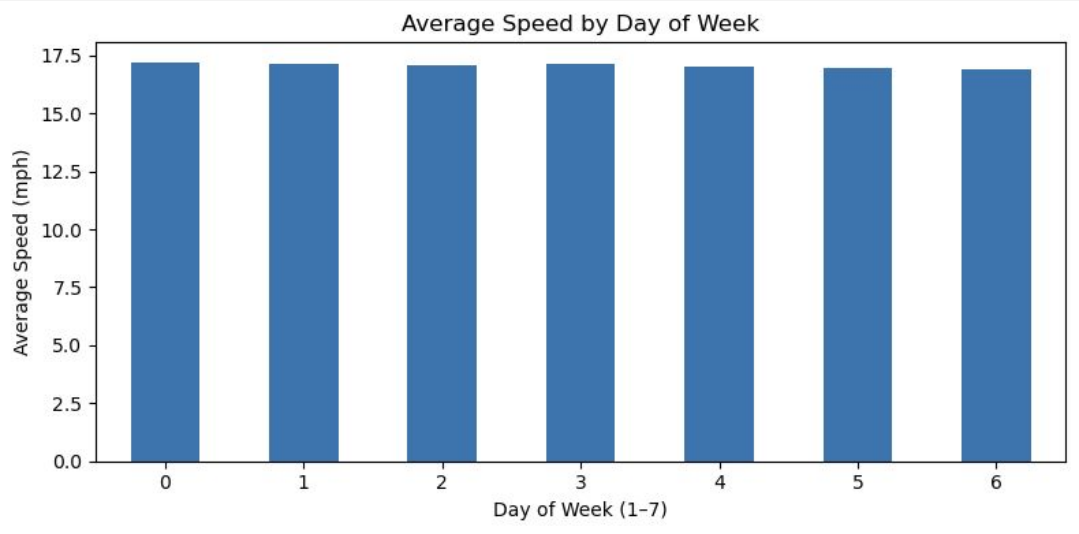
Non-Exhaustive List of Features:

- **Month**, **Day**, **Year**, **Hour**, **Minute**, **Day of Week**
- **Speed Average** - average vehicle speed entering the intersection during the interval
- **Seconds in Zone Average** - average time vehicles spent in the intersection
- **Heavy Vehicle** - Vehicles equal to or longer than 17 feet
- **Volume** - The number of vehicles traveling in the respective direction and making the respective turning movement during the 15-minute interval
- **Direction** (approaching the intersection)

Hypothesis #1:

People tend to drive slower toward the end of the week

Results

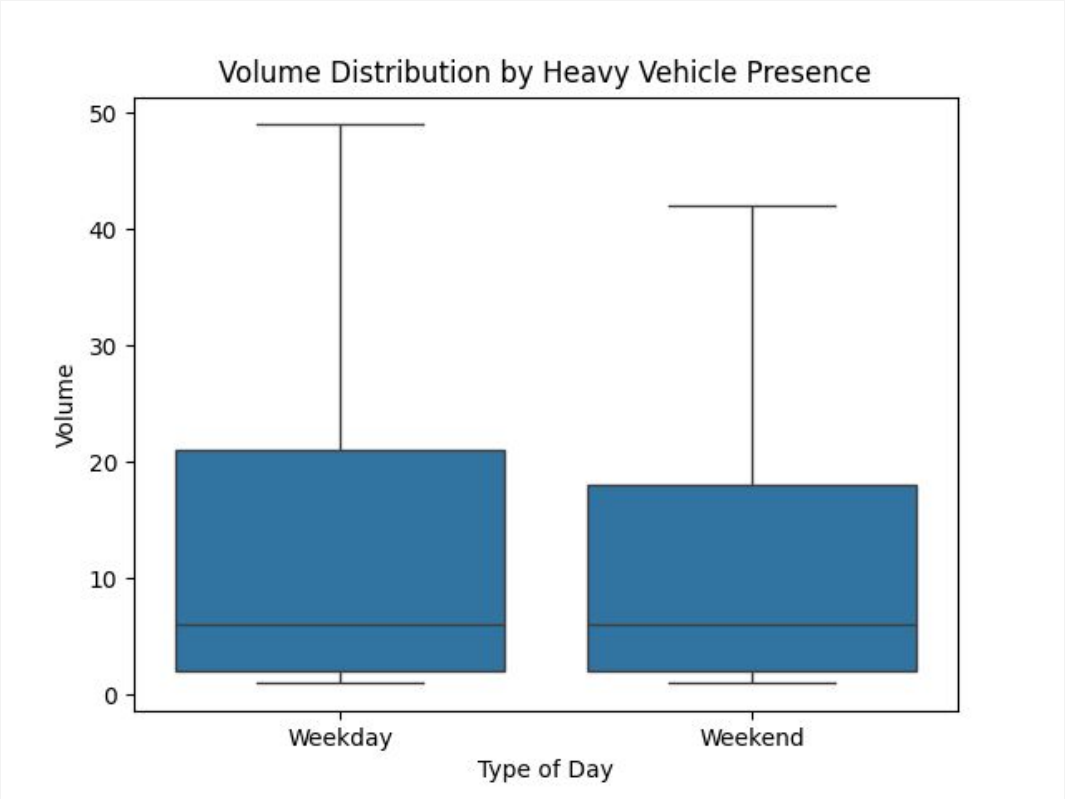


Day of Week	
0	17.200082
1	17.165757
2	17.075055
3	17.107241
4	17.021098
5	16.946072
6	16.888311

Hypothesis #2

People tend to drive less on weekends

Box Plot



- They appear to be visually similar at the lower half of the data
- Difference becomes evident as you view the higher percentiles
- Likely still significant due to the large size of the dataset (14 million rows)

Numerical Analysis

Day Type	Weekday	Weekend
Lower Whisker	1.000000	1.000000
Q1	2.000000	2.000000
Median	6.000000	6.000000
Q3	21.000000	18.000000
Upper Whisker	49.000000	42.000000
Mean	26.866897	22.173453

- Median and Lower Quartile are the same, likely due to late hours and off roads receiving minimal traffic regardless of day
- Difference becomes evident when looking at the mean and upper quartile

T-Test

```
--- T-Test Results ---
T-statistic: 22.754
P-value: 1.7811741950211357e-114
```

- Because the p-value is less than .05, there is significant evidence to support that the means for volume on weekends and weekdays are significantly different

Hypothesis 3

Vehicles drive more slowly on average around 9AM and 5PM.

Hypothesis Test

(Null Hypothesis): The average vehicle speed at around 9AM and 5PM is greater than or equal to the average speed during non-rush hour.

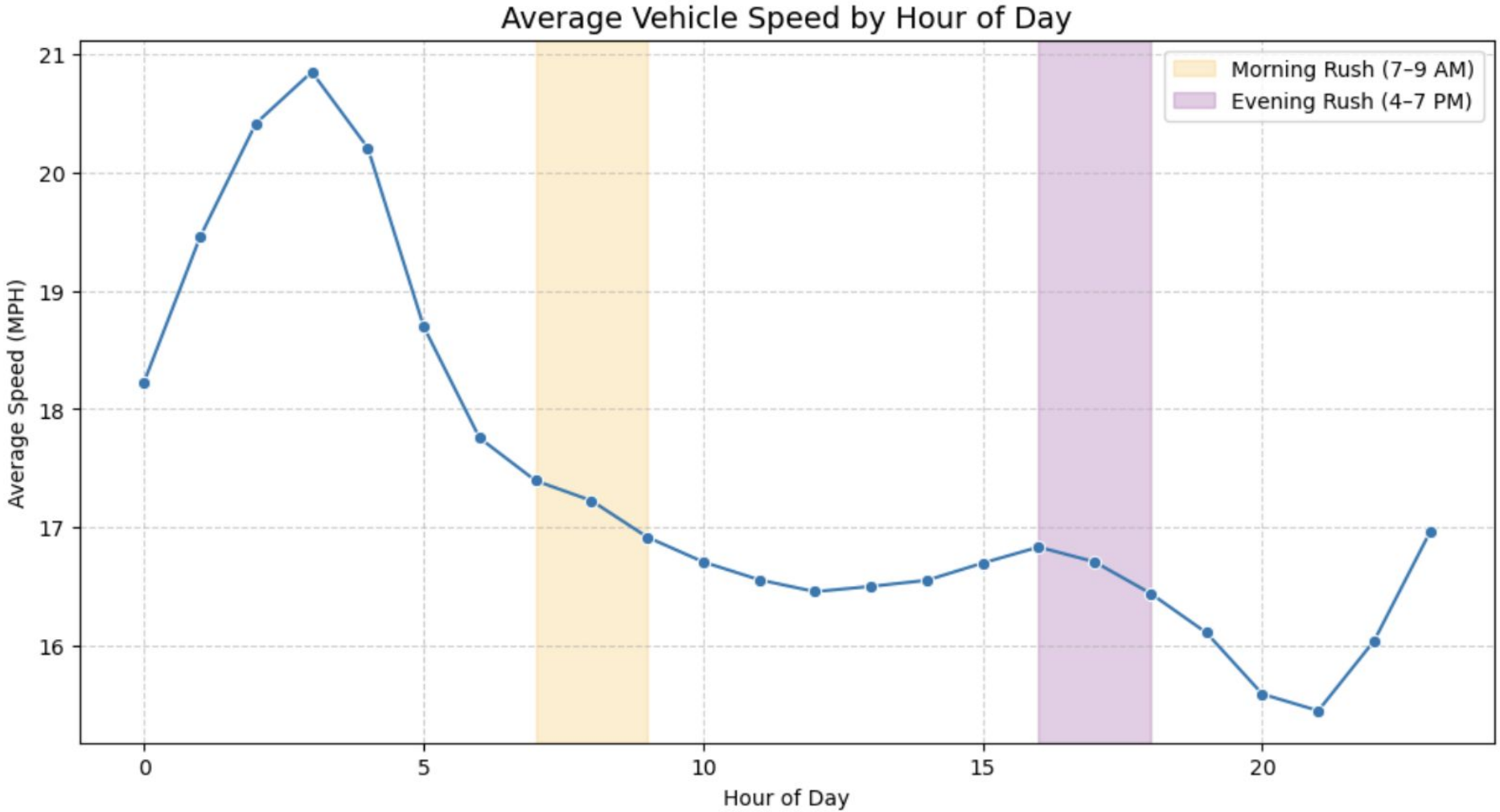
(Alternative Hypothesis): The average vehicle speed at around 9AM and 5PM is significantly lower than during non-rush hour.

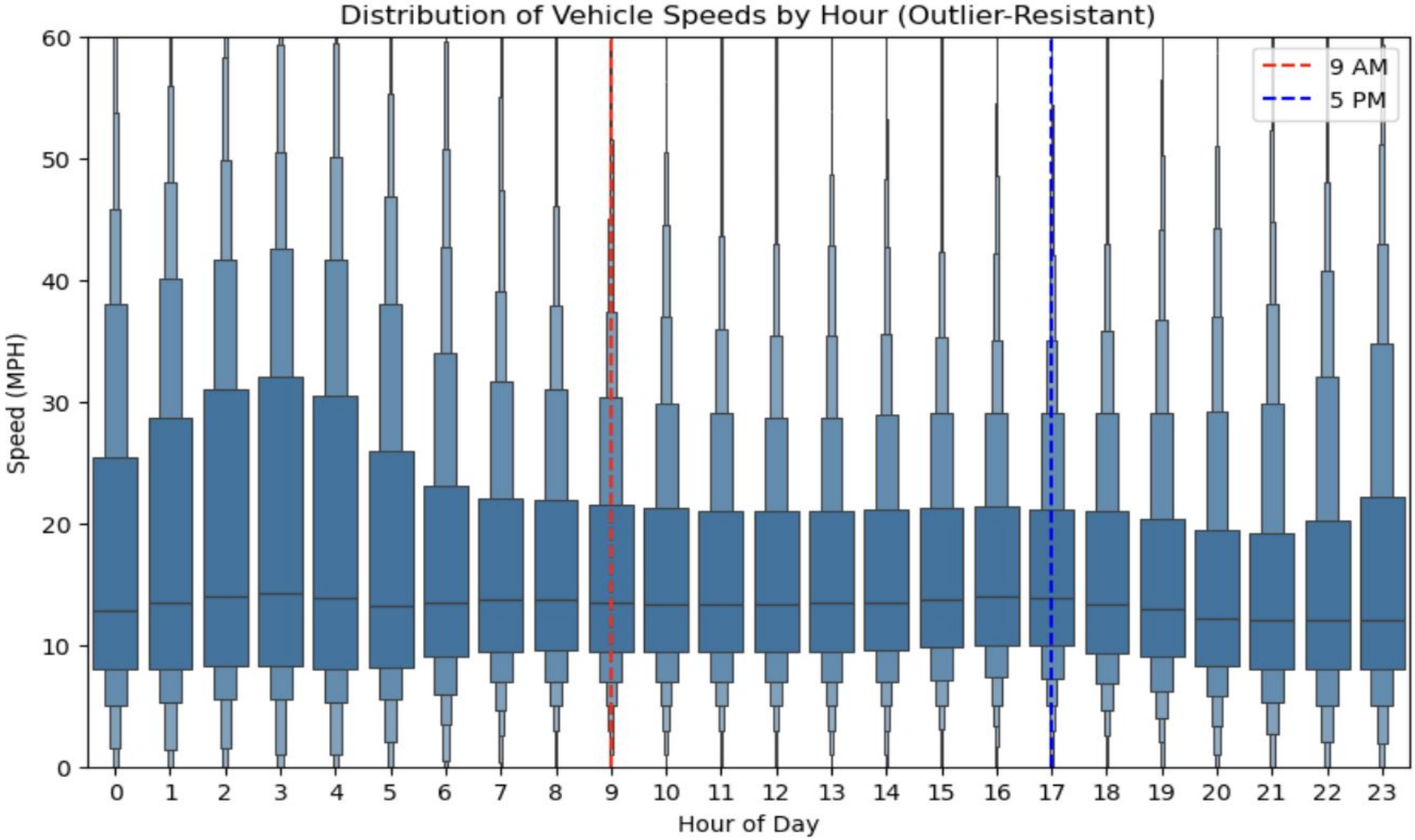
Significance Level (α): 0.05

Methodology

Two-Sample T-Test (Statistical Analysis)

This test was conducted to determine if the average speeds of two distinct groups—**morning** traffic and **evening** traffic—are significantly different compared to other times.


Average Vehicle Speed by Hour of Day
Morning Rush (7–9 AM)
Evening Rush (4–7 PM)
Average Speed (MPH)
Hour of Day
21
20
19
18
17
16
0
5
10
15
20


Distribution of Vehicle Speeds by Hour (Outlier-Resistant)
9 AM
5 PM
Speed (MPH)
Hour of Day
0
10
20
30
40
50
60
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Methodology cont.

Results: The test produced a **t-statistic** and a **one-tailed p-value** for our alternative hypothesis.

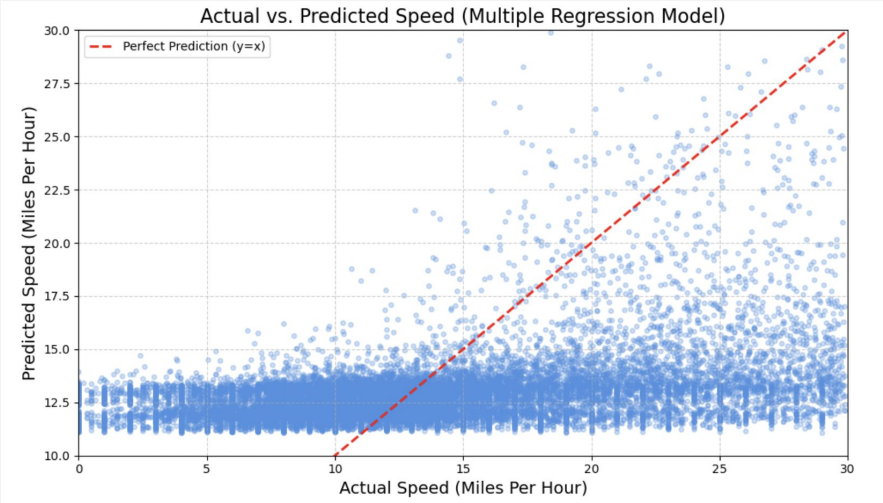
- **t-statistic:** -29.564346384176176
- **p-value:** 2.1945207631138 96e-192

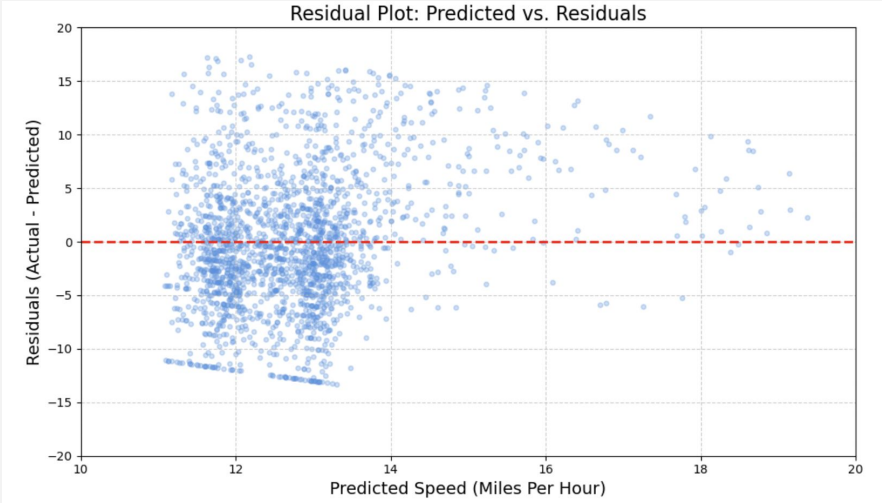
Conclusion: With a p-value of 2.2e-192 (less than $\alpha = 0.05$), we **reject the null hypothesis**. The results provide overwhelming statistical evidence that vehicles drive significantly slower during the designated rush hour periods.

Linear Regression Model

Data Preparation: We selected the following input features and the target variable for our model.

- **Input (X)**: Hour, Day of Week, Volume, Heavy Vehicle, Direction.
- **Target (y)**: Speed Average (Miles Per Hour).





Model Evaluation (Account for Confounders)

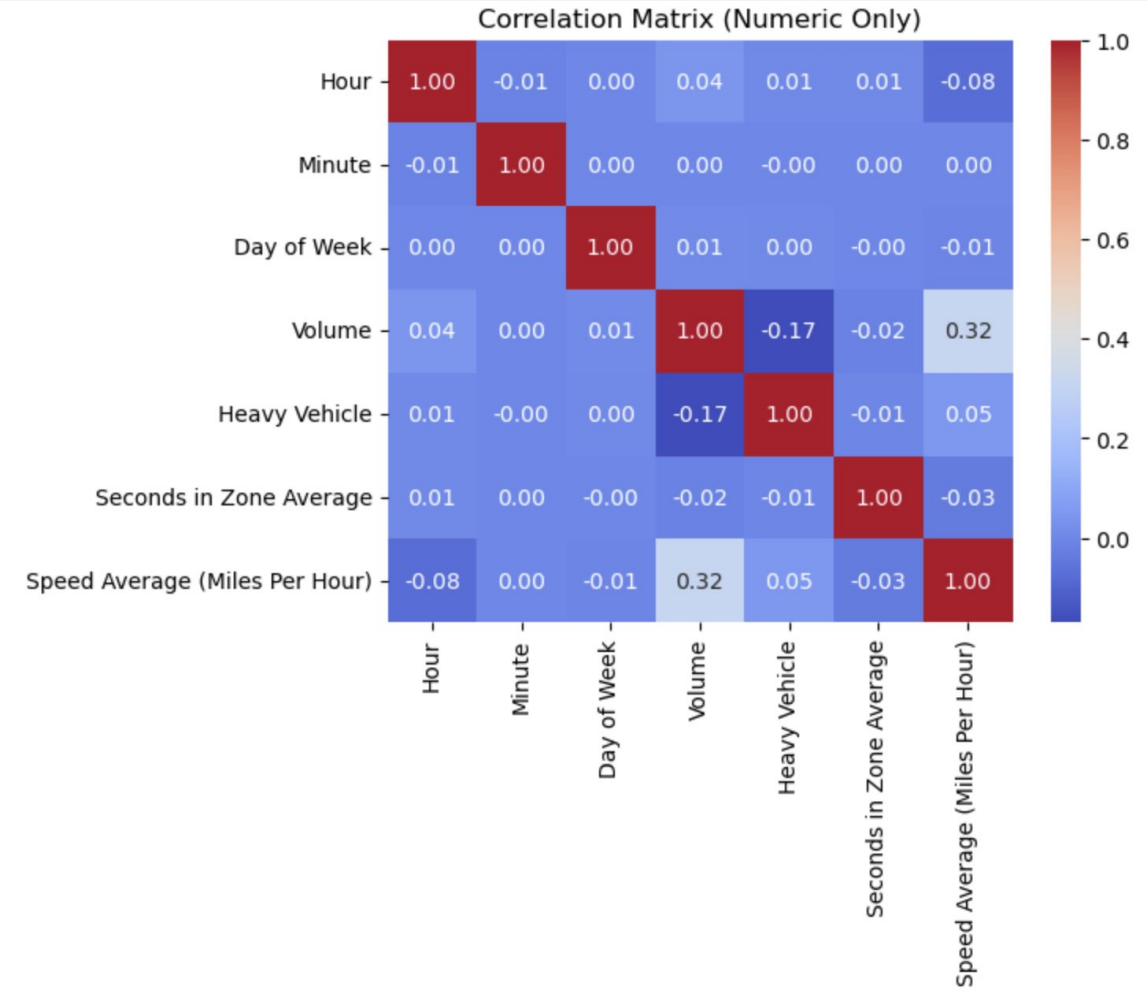
-- Regression Performance Report --
R-squared (R^2): 0.1465

Mean Absolute Error (MAE): 6.3918 MPH

Mean Squared Error (MSE): 8.4158 MPH

--- Model Coefficients (Feature Importance) ---
Direction_SOUTHBOUND -1.087741
Direction_NORTHBOUND -0.475226
Direction_WESTBOUND -0.380772
Hour -0.186419
Day of Week -0.072805
Volume 0.067215
Direction_UNASSIGNED 2.548512
Heavy Vehicle 2.752015

The 'Unassigned Direction' and 'Heavy Vehicle' coefficients are typically the most important. A negative coefficient (e.g., for Hour) means as that factor increases, speed decreases. The MAE indicates that the model's predictions are, on average, off by that many MPH.


Correlation Matrix (Numeric Only)
Hour
Minute
Day of Week
Volume
Heavy Vehicle
Seconds in Zone Average
Speed Average (Miles Per Hour)
1.00 -0.01 0.00 0.04 0.01 0.01 -0.08
-0.01 1.00 0.00 0.00 -0.00 0.00 0.00
0.00 0.00 1.00 0.01 0.00 -0.00 -0.01
0.04 0.00 0.01 1.00 -0.17 -0.02 0.32
0.01 -0.00 0.00 -0.17 1.00 -0.01 0.05
0.01 0.00 -0.00 -0.02 -0.01 1.00 -0.03
-0.08 0.00 -0.01 0.32 0.05 -0.03 1.00
1.0
0.8
0.6
0.4
0.2
0.0

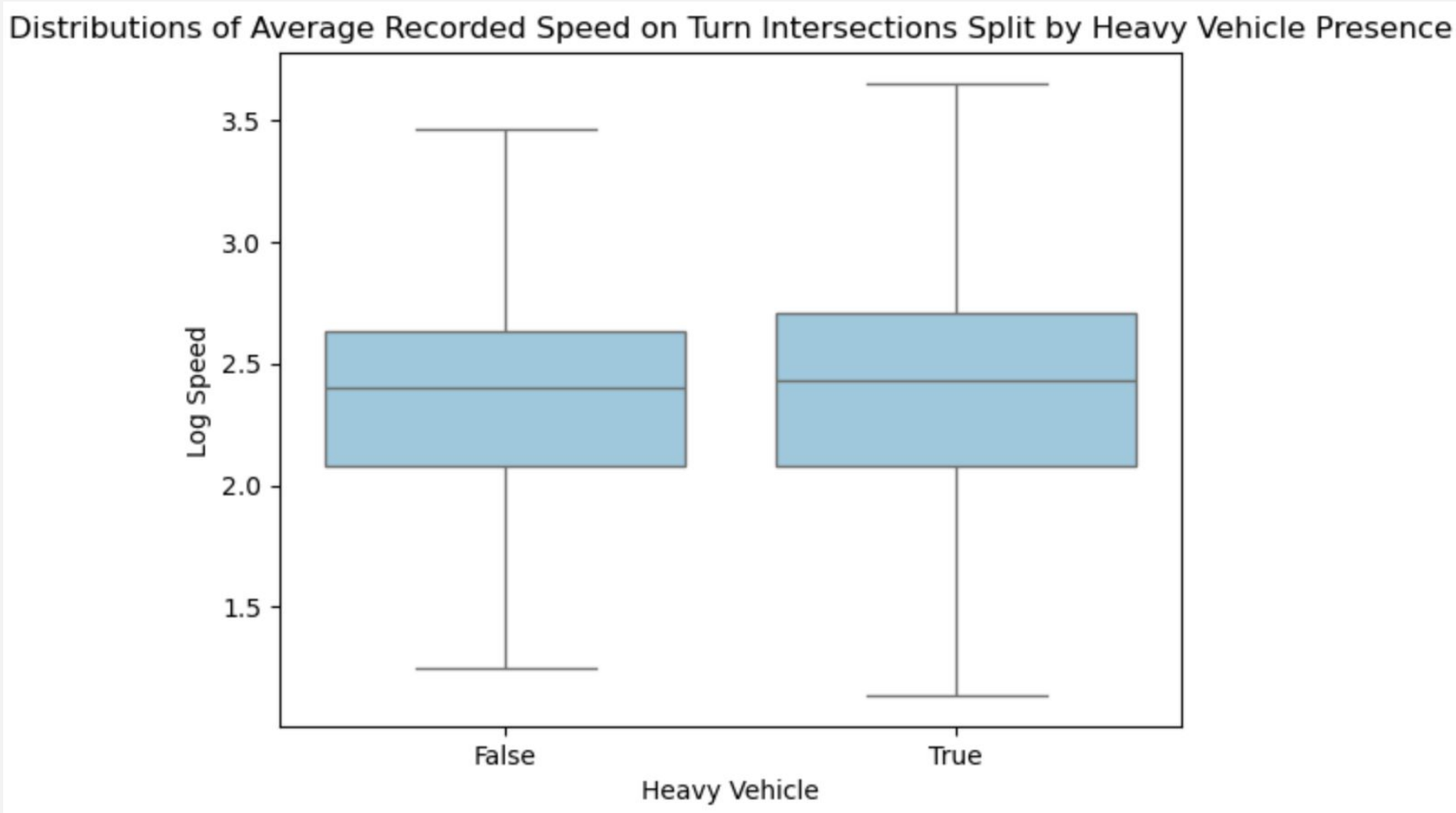
Hypothesis 4

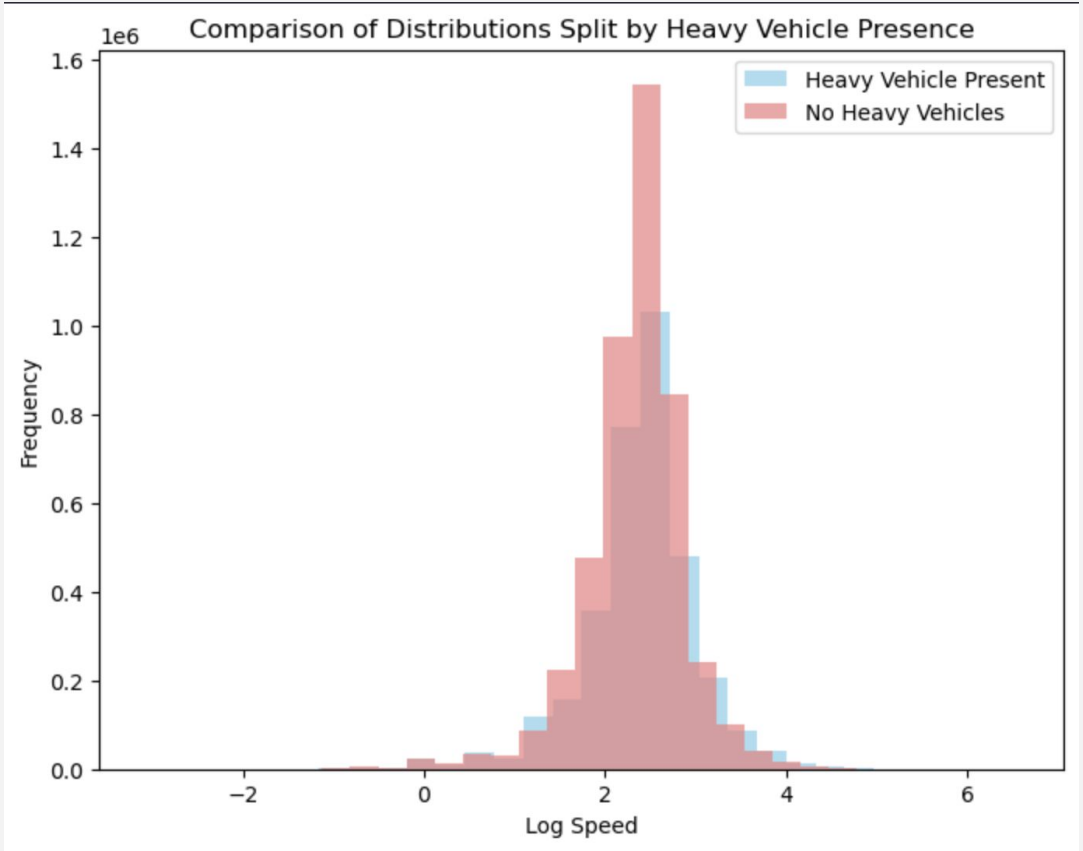
Cars tend to move faster at turns when there is a heavy vehicle present

Methodology

- Filtered the records based on the movement column (the vehicle movement the detector is positioned to capture) and extracted the records that indicated a turn (left turn or right turn)
- Created charts comparing the average speeds recorded when there are heavy vehicles present and when there isn't and conducted one-sided Welch's T-Test

ATD Device ID	Read Date	Intersection Name	Direction	Movement	Heavy Vehicle	Volume	Speed Average (Miles Per Hour)	Speed StdDev	Seconds in Zone Average	Seconds in Zone StdDev	Month	D
6549	11/07/2019 12:00:00 AM	WEST GATE BLVD / JONES RD	EASTBOUND	LEFT TURN	True	1	11.000	0.000	3.800	0.000	11	
6549	11/07/2019 12:00:00 AM	WEST GATE BLVD / JONES RD	NORTHBOUND	LEFT TURN	False	1	8.000	0.000	0.100	0.000	11	
6549	11/07/2019 12:00:00 AM	WEST GATE BLVD / JONES RD	NORTHBOUND	THRU	False	9	27.000	5.766	1.811	0.697	11	
6549	11/07/2019 12:00:00 AM	WEST GATE BLVD / JONES RD	NORTHBOUND	THRU	True	2	32.000	4.243	1.850	0.071	11	
6549	11/07/2019 12:00:00 AM	WEST GATE BLVD / JONES RD	SOUTHBOUND	LEFT TURN	False	1	12.000	0.000	2.700	0.000	11	


Distributions of Average Recorded Speed on Turn Intersections Split by Heavy Vehicle Presence
3.5
3.0
2.5
2.0
1.5
Log Speed
False
True
Heavy Vehicle



Heavy Vehicle Presence = True:

Mean Log Average Recorded Speed = 2.379

Heavy Vehicle Presence = False:

Mean Log Average Recorded Speed = 2.334

Welch's T-Test

(Null Hypothesis): The Mean Average Recorded Speed at turn intersections is **equal** when Heavy Vehicle Presence is True and when Heavy Vehicle Presence is False.

(Alternative Hypothesis): Mean Average Recorded Speed is **greater** when Heavy Vehicle Presence is True

Significance Level (α): 0.05

:

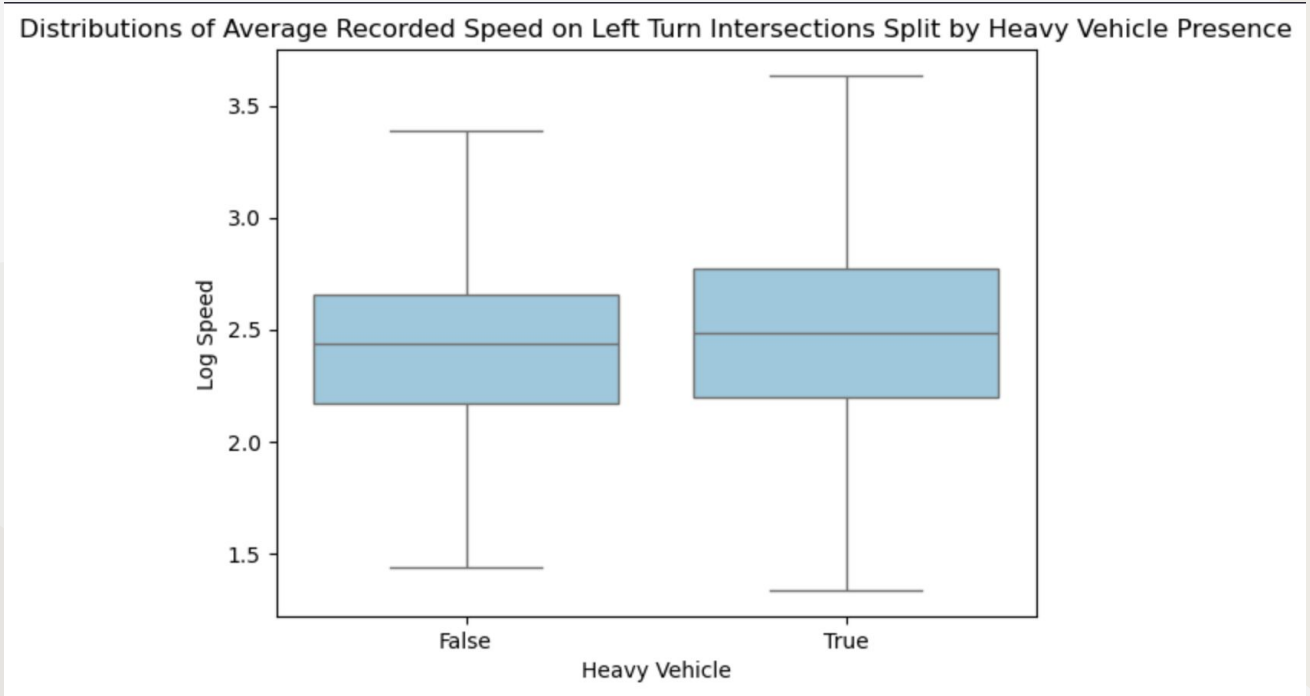
Welch's T-Test Results

```
# T-Test for Turns

from scipy.stats import ttest_ind

a = turns[turns["Heavy Vehicle"] == True]["Speed Average (Miles Per Hour)"]
b = turns[turns["Heavy Vehicle"] == False]["Speed Average (Miles Per Hour)"]

ttest_ind(a, b, alternative = "greater")
```

✓ 9.7s

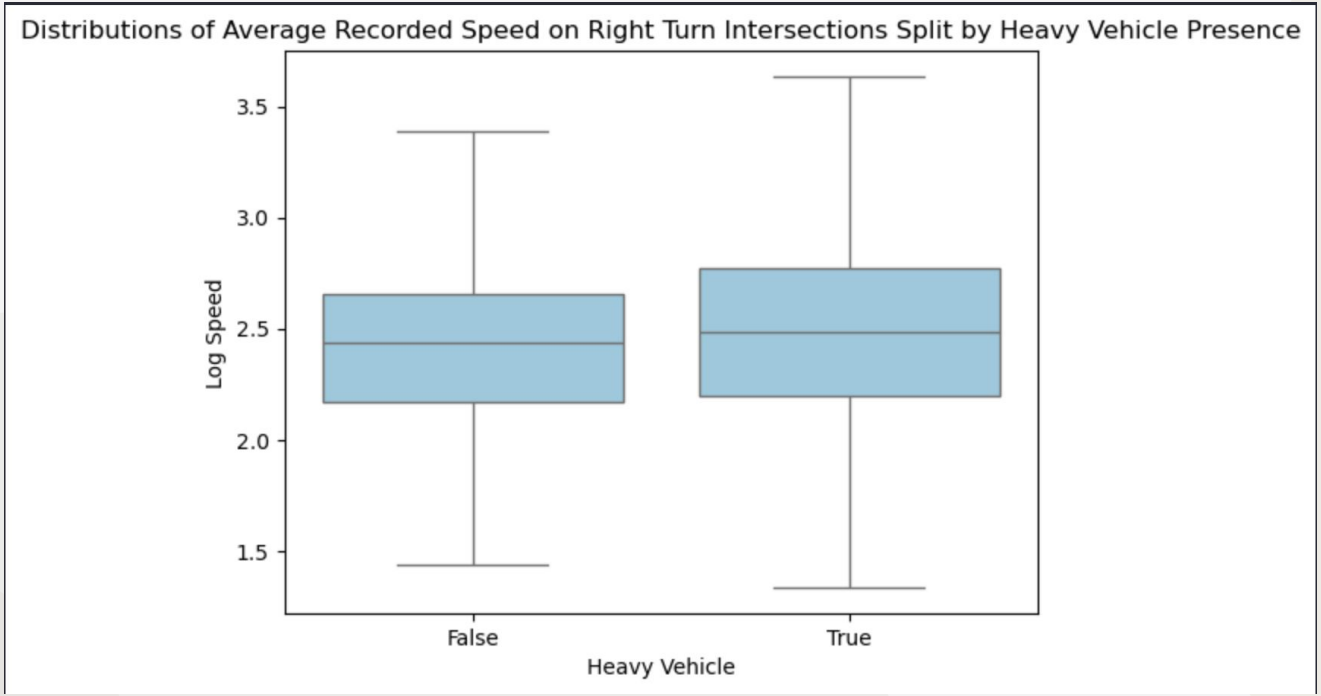
```
TtestResult(statistic=187.88018567721417, pvalue=0.0, df=8057031.0)
```

Results for Left Turn




```
TtestResult(statistic=164.77153923825853, pvalue=0.0, df=4319555.0)
```

Results for Right Turn




```
TtestResult(statistic=100.40133407916369, pvalue=0.0, df=3737474.0)
```

Takeaways

- There is statistical evidence to suggest that drivers are driving faster at turns when there is a heavy vehicle.

- Possible Explanations: Maybe the drivers in heavy vehicles are driving faster or people may want to pull ahead of heavy vehicles at intersections

Summary

Hypothesis #1

- People drive slower later in the week

Hypothesis #2

- People drive less on weekends

Hypothesis #3

- People drive slower during rush hour

Hypothesis #4

- People drive at higher speeds during turns when a heavy vehicle is present

Thank you!
Questions?