

# Hypotheses

MBTA Transportation

Python for Data Science and AI

**GROUP 4**

- Suzuki Kaori 21224038
- Shrestha Roji 22224045



# Table Of Content

◆ 01

Introduction

◆ 02

First  
Hypothesis

◆ 03

Second  
Hypothesis

◆ 04

Third  
Hypothesis

◆ 05

Conclusion

# What is MBTA?

- The Massachusetts Bay Transportation Authority (MBTA) provides public transit services including subway, bus, commuter rail, and ferry routes in the Greater Boston area.
- The CR-Providence line is part of the commuter rail network connecting Boston with southern Massachusetts and Rhode Island.



```

#First we imported the libraries we needed for analysis.
import requests
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

#This is the API which we worked on. We got it from MBTA.
test_url = 'https://api-v3.mbta.com/route_patterns?filter[route]=CR-Providence&include=representative_trip&fields[trip]=headsign'
response = requests.get(test_url)
data = response.json()

# Extract relevant information into a DataFrame
route_patterns = [
    {
        'route_pattern_id': pattern['id'],
        'direction_id': pattern['attributes']['direction_id'],
        'name': pattern['attributes']['name'],
        'typicality': pattern['attributes']['typicality'],
    }
    for pattern in data['data']
]
df = pd.DataFrame(route_patterns)

# Display the DataFrame to verify the structure
df.head()

```

	route_pattern_id	direction_id	name	typicality
0	CR-Providence-9cf54fb3-0	0	South Station - Wickford Junction via Back Bay	1
1	CR-Providence-9515a09b-0	0	South Station - Stoughton via Back Bay	1
2	CR-Providence-743ff41d-0	0	South Station - Providence via Back Bay	2
3	CR-Providence-cf299c9f-0	0	South Station - Stoughton via Back Bay	3
4	CR-Providence-e7e93ca2-0	0	South Station - Wickford Junction via Back Bay	3

**Research question 1:** Is there a correlation between direction (inbound or outbound) and the typicality of routes in the CR-Providence route patterns?

**Hypothesis 1:** Routes with higher typicality scores are more likely to be inbound ( $\text{direction\_id} = 0$ ).

**Null-hypothesis 1:** There is no correlation between direction (inbound or outbound) and the typicality of routes in the CR-Providence route patterns.



```
# New Research Question and Hypothesis
# Research Question: Is there a correlation between direction and typicality in the CR-Providence route patterns?
# Hypothesis: Routes with higher typicality scores are more likely to be inbound (direction_id = 0).

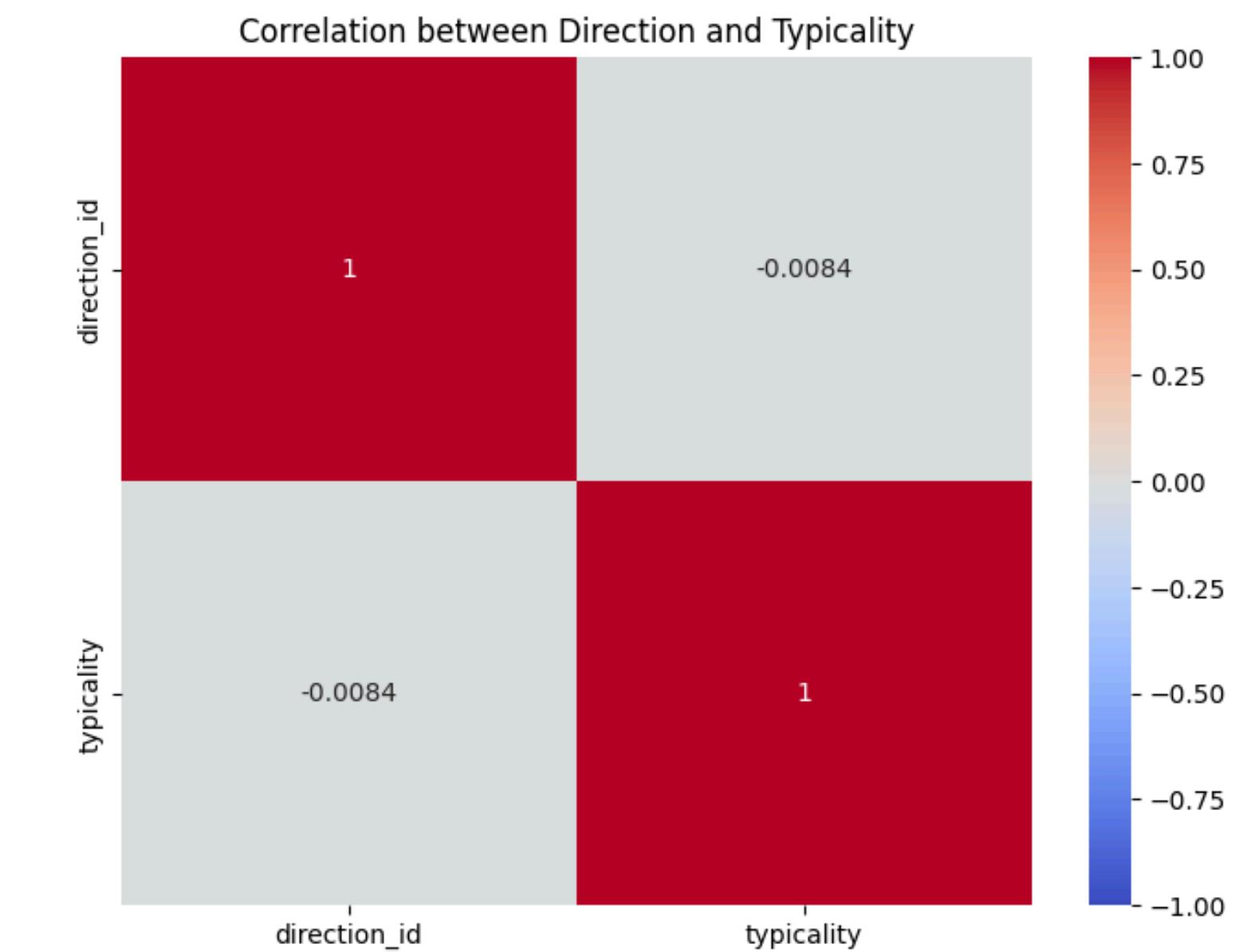
# Exploring correlation using basic statistics
correlation = df[['direction_id', 'typicality']].corr()

# Visualizing the correlation between direction and typicality
plt.figure(figsize=(8, 6))
sns.heatmap(correlation, annot=True, cmap="coolwarm", vmin=-1, vmax=1)
plt.title('Correlation between Direction and Typicality')
plt.show()

correlation
```

# Hypothesis 1

- The correlation coefficient between **direction** and **typicality** is **-0.008**, which indicates almost **no linear correlation** between these two variables.
- The direction (whether the route is inbound or outbound) does not appear to be related to the typicality score. Thus, the hypothesis is not supported by the data .



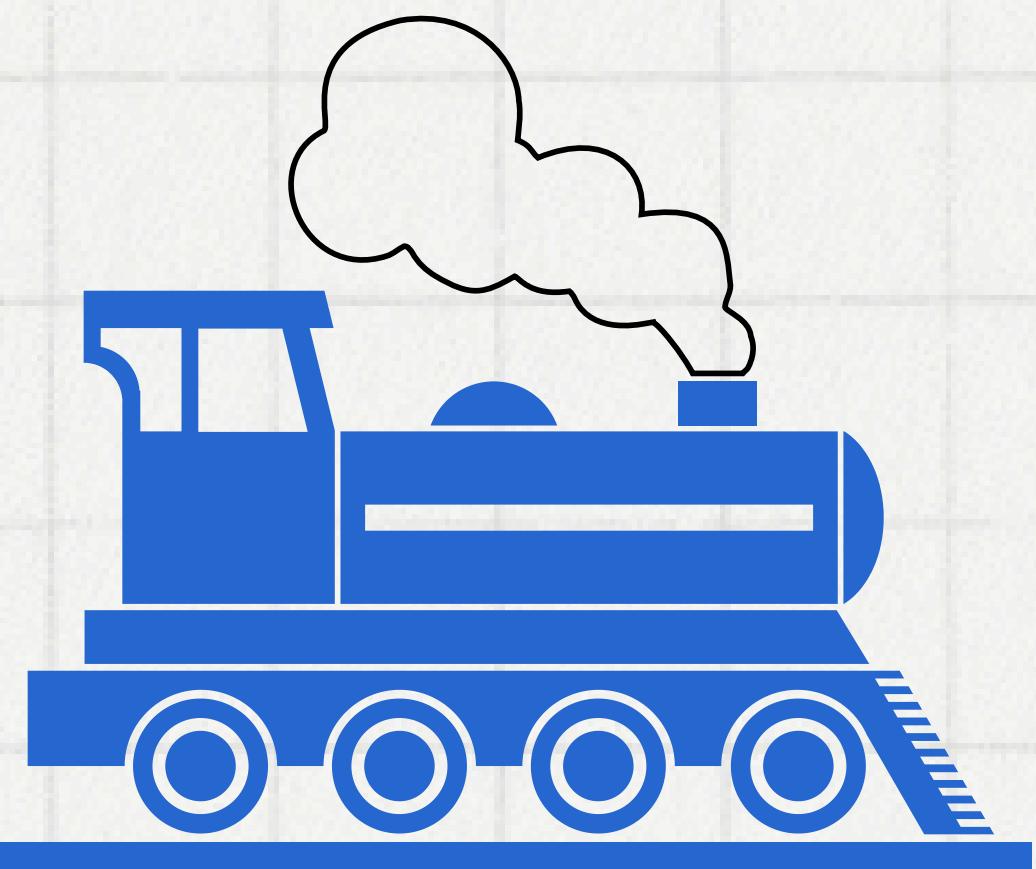
**Research question 2:** Are certain route patterns more frequent in the data?

**Hypothesis 2:** Specific patterns such as "South Station - Wickford Junction" appear more frequently.

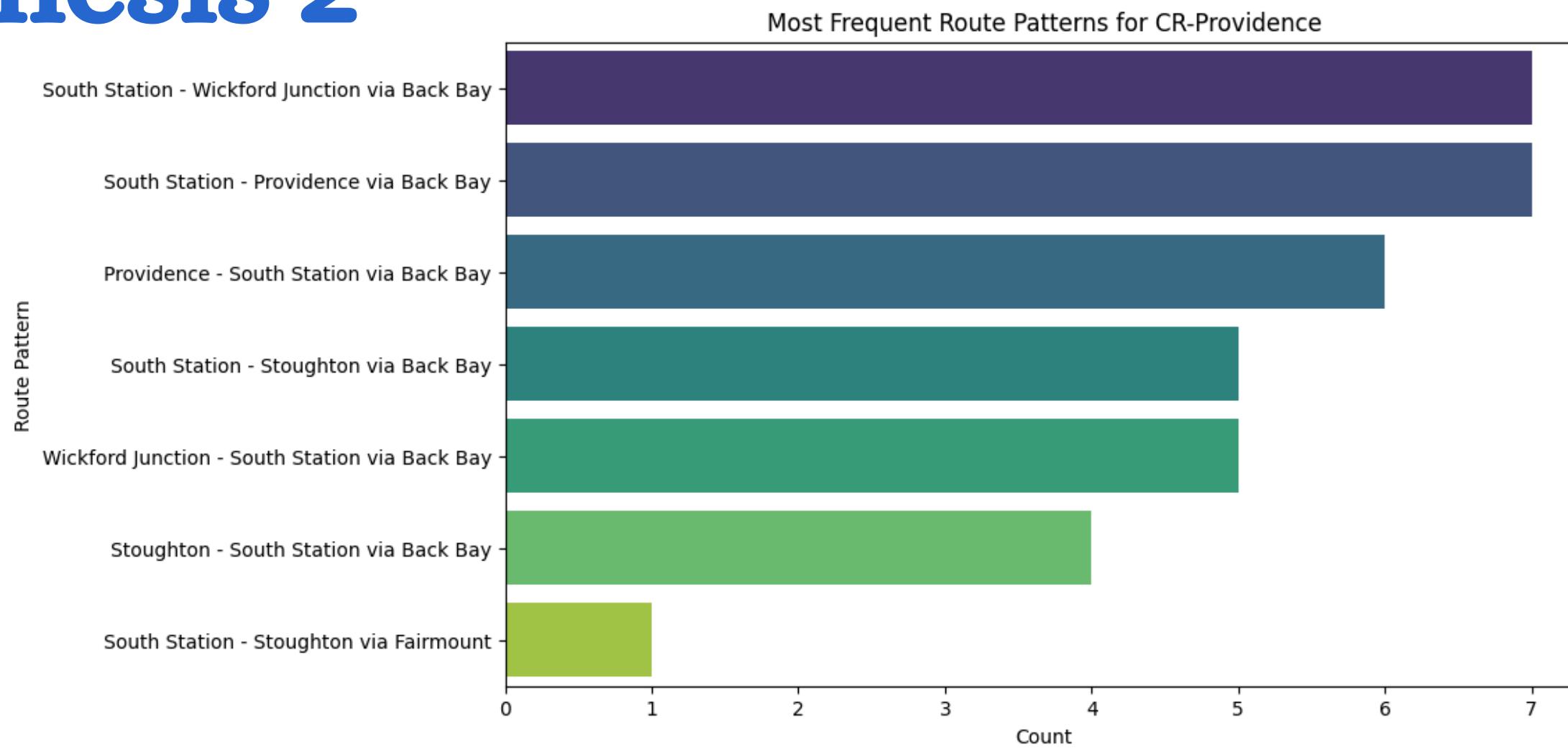
**Null Hypothesis 2:** All route patterns in the CR-Providence data are equally frequent, with no specific patterns appearing more frequently than others.

```
# Preparing data for the bar chart
frequent_patterns = df['name'].value_counts().reset_index()
frequent_patterns.columns = ['Route Pattern', 'Count']

plt.figure(figsize=(10, 6))
sns.barplot(data=frequent_patterns, x='Count', y='Route Pattern', hue='Route Pattern', dodge=False, palette="viridis", legend=False)
plt.title('Most Frequent Route Patterns for CR-Providence')
plt.xlabel('Count')
plt.ylabel('Route Pattern')
plt.show()
```



# Hypothesis 2



The bar plot of route patterns shows that "South Station - Wickford Junction via Back Bay" and similar patterns have higher number. This confirms that certain route patterns are indeed more frequent than others. Our hypothesis is supported.

**Research question 3:** Does the CR-Providence route have more inbound or outbound routes?

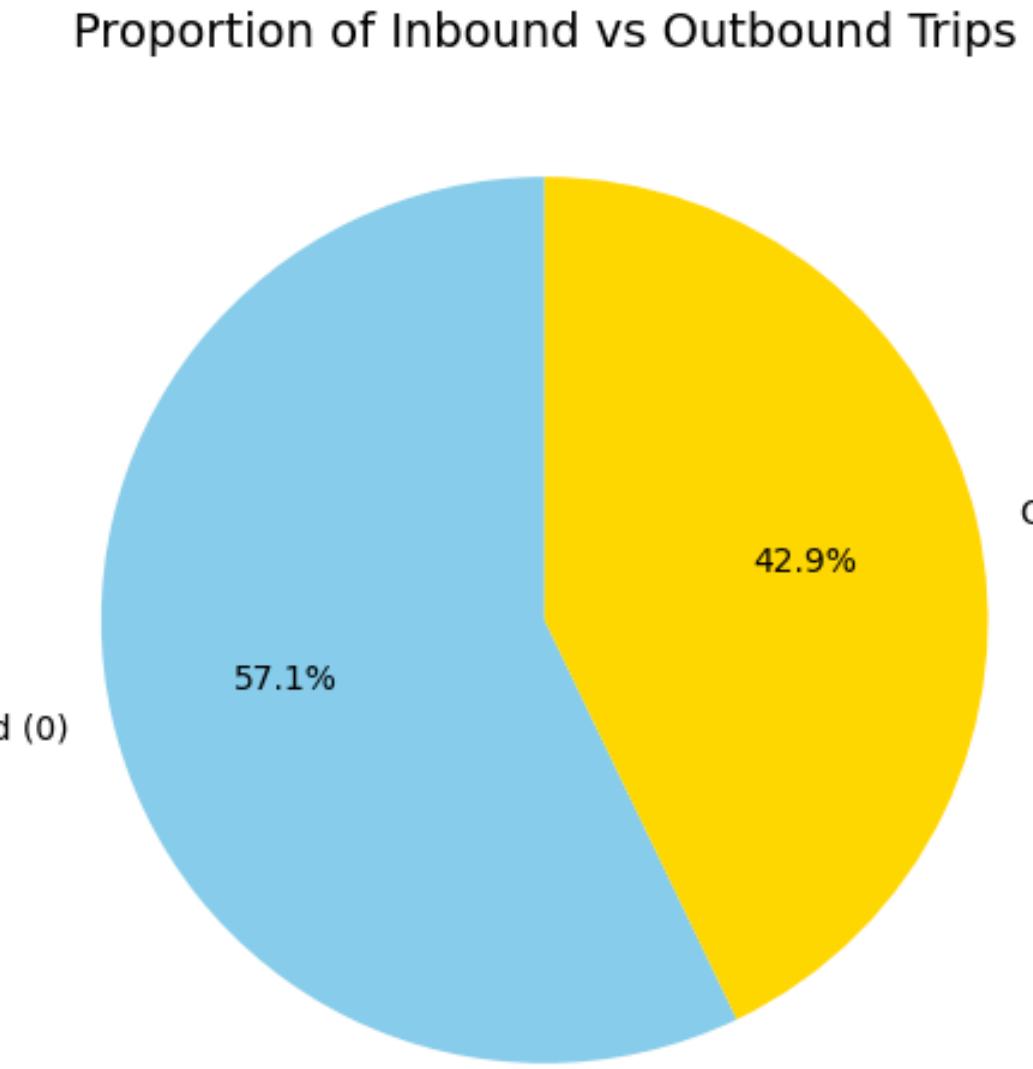
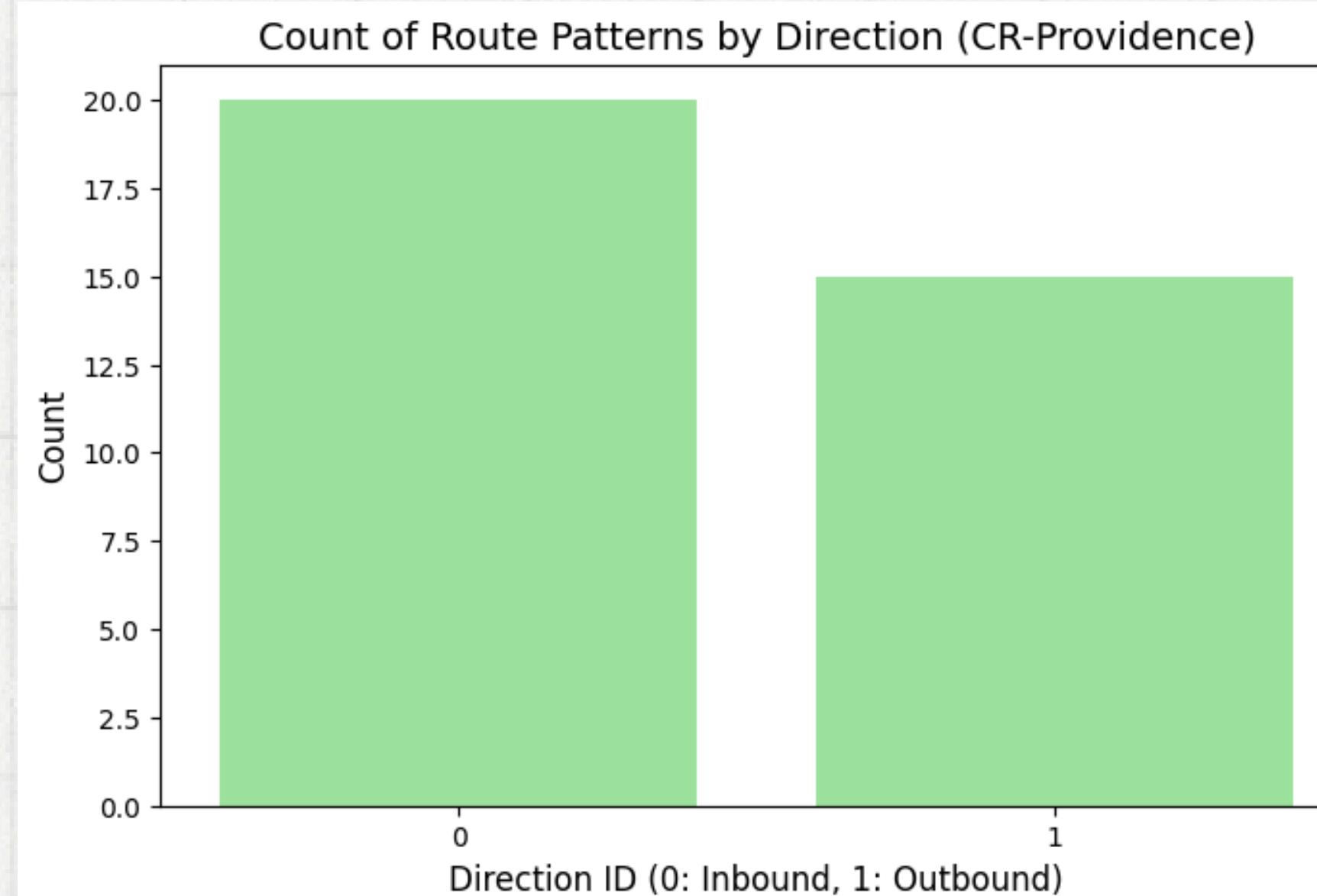
**Hypothesis 3:** There are more inbound routes.

**Null-hypothesis 3:** There is no difference in the number of inbound and outbound routes on the CR-Providence line.

To begin, we imported the necessary libraries, including requests to access the API, pandas for data handling, and matplotlib and seaborn for data visualization.

```
# Visualize the count of route patterns by direction
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='direction_id', color='lightgreen')
plt.title('Count of Route Patterns by Direction (CR-Providence)', fontsize=14)
plt.xlabel('Direction ID (0: Inbound, 1: Outbound)', fontsize=12)
plt.ylabel('Count', fontsize=12)
plt.show()
```

```
# Pie chart for proportion of direction usage
direction_counts = df['direction_id'].value_counts()
labels = ['Inbound (0)', 'Outbound (1)']
plt.figure(figsize=(6, 6))
plt.pie(direction_counts, labels=labels, autopct='%.1f%%', startangle=90, colors=['#87CEEB', '#FFD700'])
plt.title('Proportion of Inbound vs Outbound Trips', fontsize=14)
plt.show()
```



- The graph for count of route patterns by direction shows that the inbound is higher than outbound.
- The piechart confirms that **57% of the routes are inbound**, while **43% are outbound**, supporting our hypothesis.



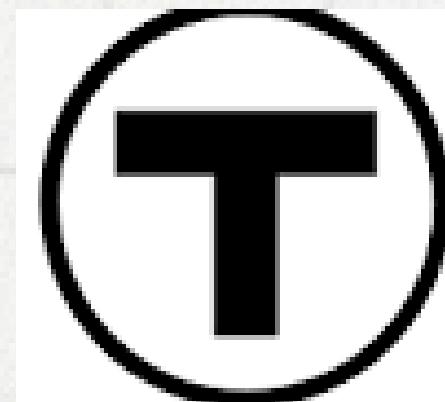
# Conclusion

From the visualizations, we found out that:

01. There is almost **no linear correlation** between direction and typicality score. (null-hypothesis)
02. Specific patterns such as "South Station - Wickford Junction" appear **more frequently**. (hypothesis)
03. The CR-Providence route have **more inbound routes** than outbound routes. (hypothesis)

# Reference (APA)

Massachusetts Bay Transportation  
Authority. (n.d.). MBTA API v3. [https://api-  
v3.mbta.com/](https://api-v3.mbta.com/) Retrieved October 13, 2024.



**Massachusetts Bay  
Transportation Authority**

**Thank you  
very much!**