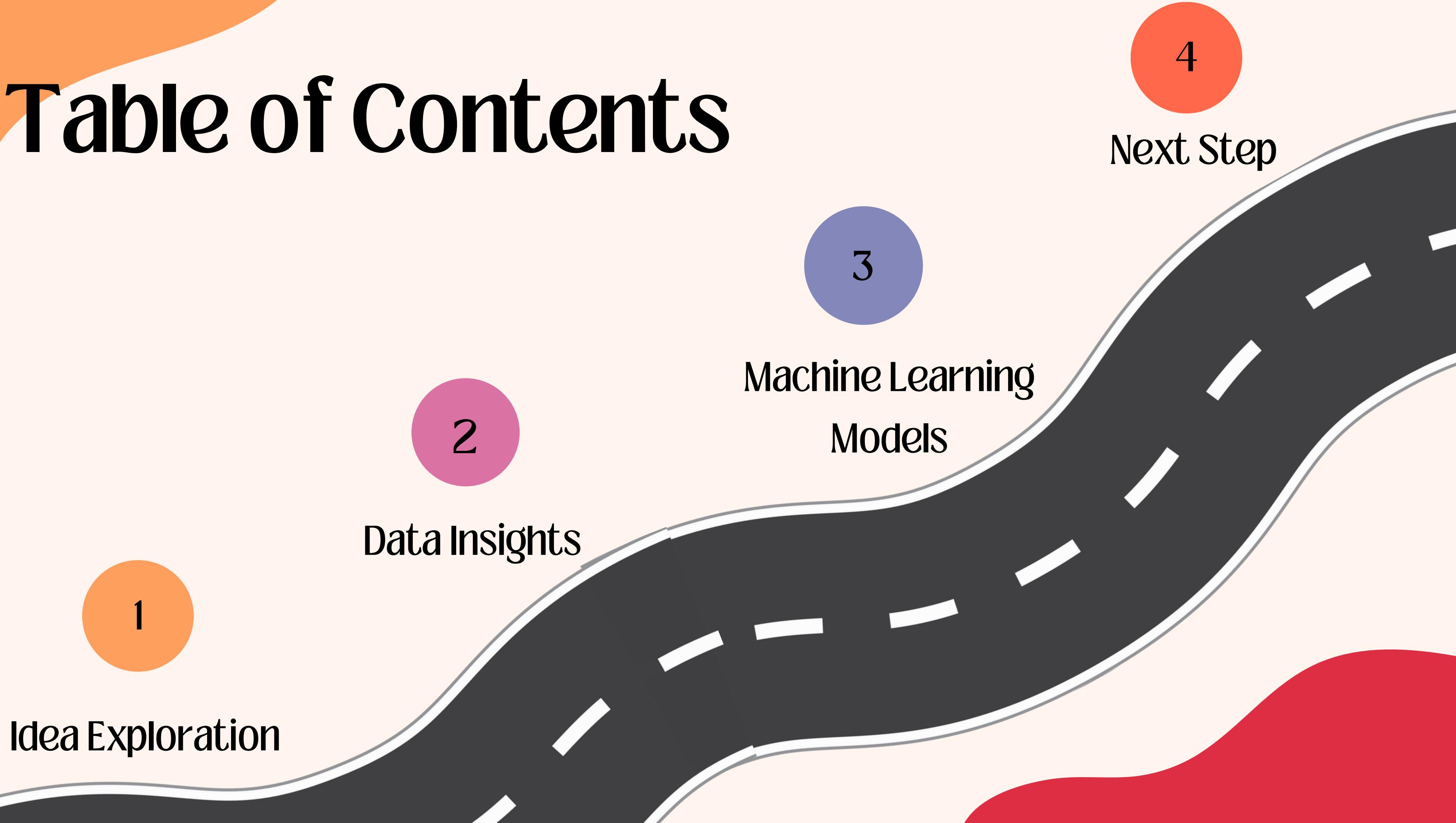


# Predicting TTC Streetcar Delays

Jessica Seo  
November 2023

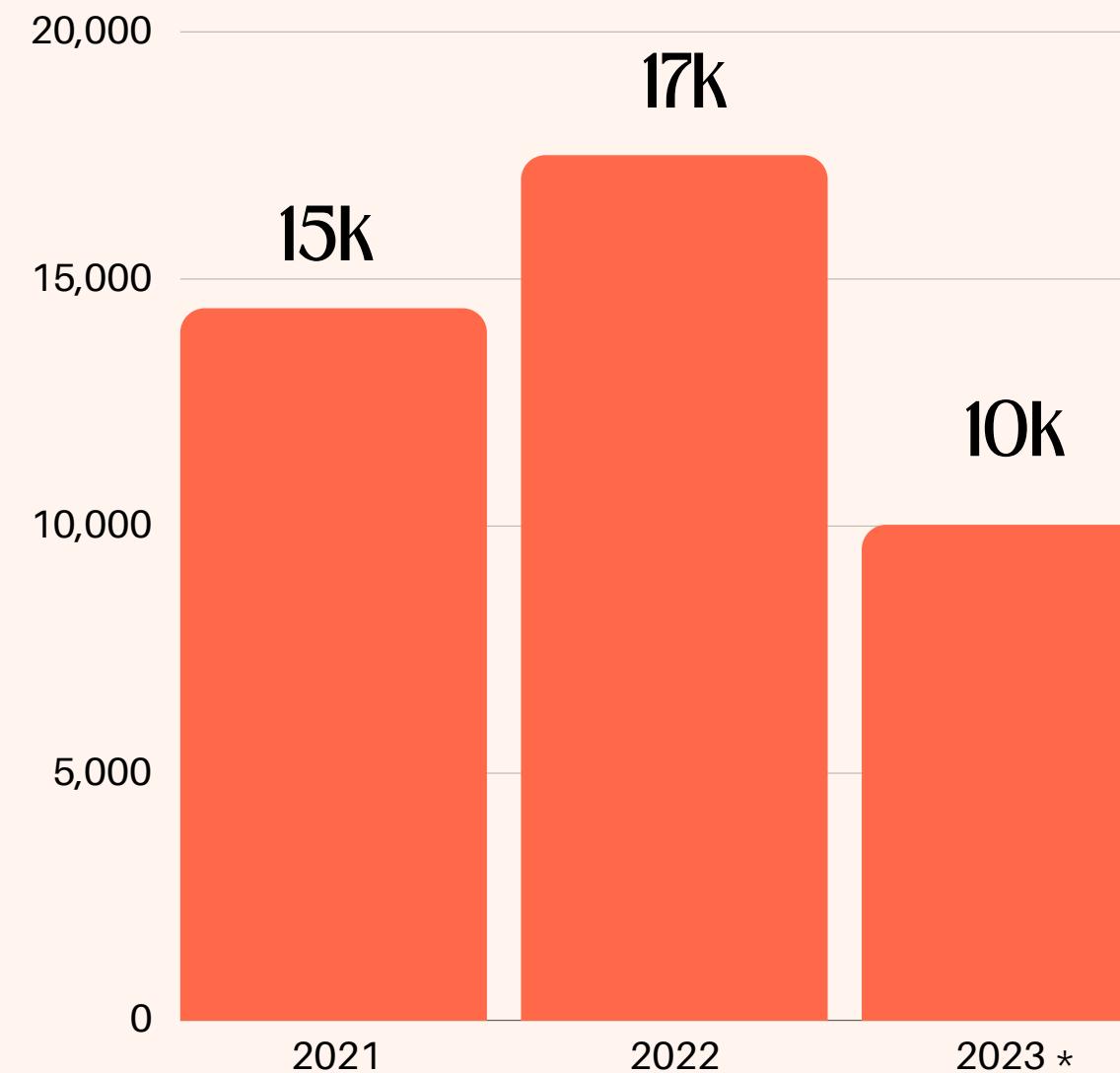


# Table of Contents

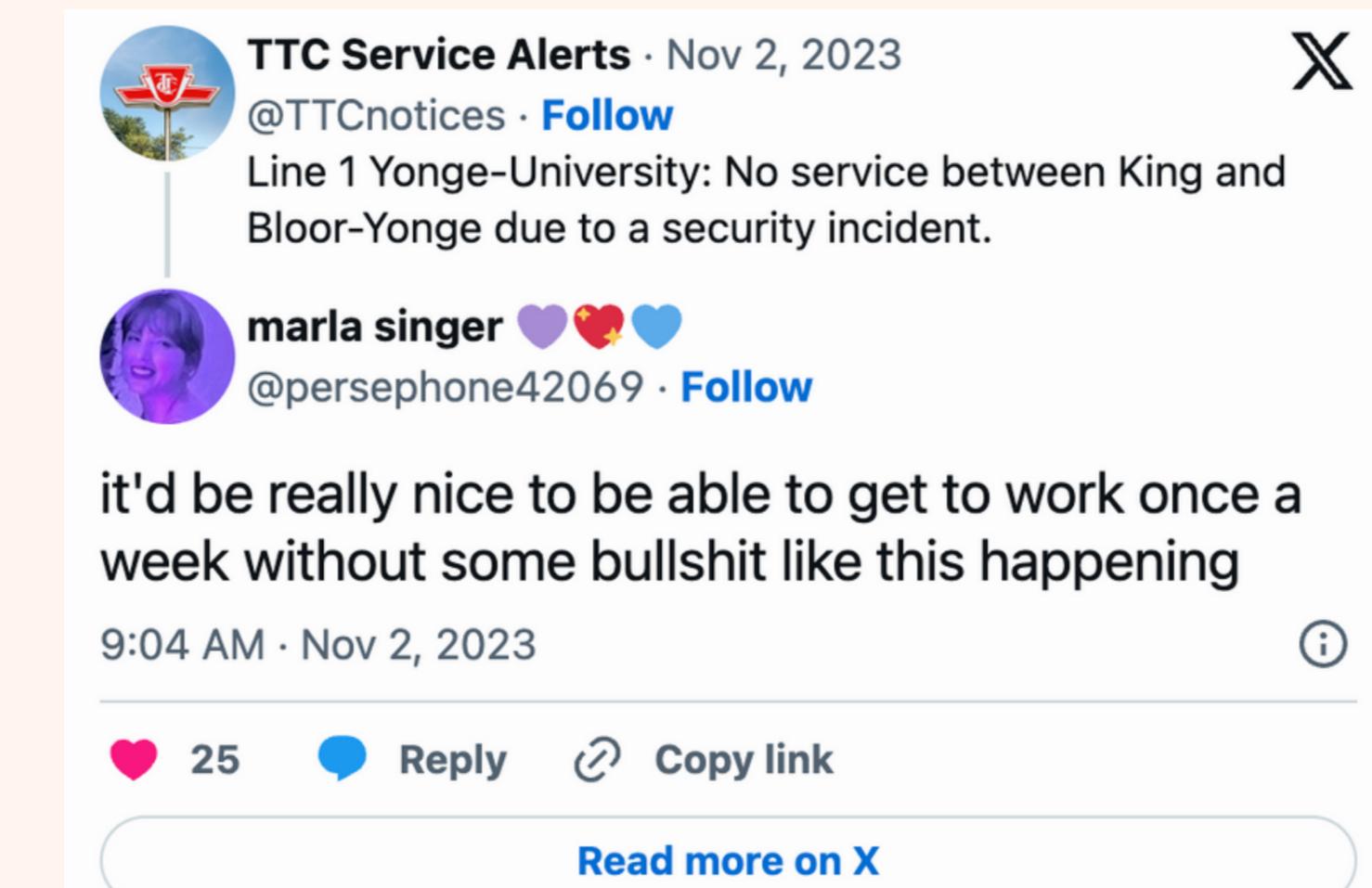


# The Problem Area

## Annual Streetcar delays



\*2023: Jan-Sept30



(1)

# Areas of Impact

General Public:

- Provide transparent information regarding delays and decrease commuters' frustration and answer their core question:  
**'So how long is this going to take?!**

City of Toronto:

- Regain the reputation of being the best public transit agency in North America, and increase ridership leading to higher fare revenue. (2)

# Meet the Data



Streetcar Data:  
Line, Delay Reason, Time



Environment  
Canada

Weather Data:  
Temperature, Precipitation,  
Weather Conditions

# Things to note

- 41,000 rows and 40 columns
- 3 years of data from 2021 January - 2023 October
- Multi classification Target
- Handling 80% of null values
- Feature Engineering
  - Encoding categorical variables
  - Extracting time feature

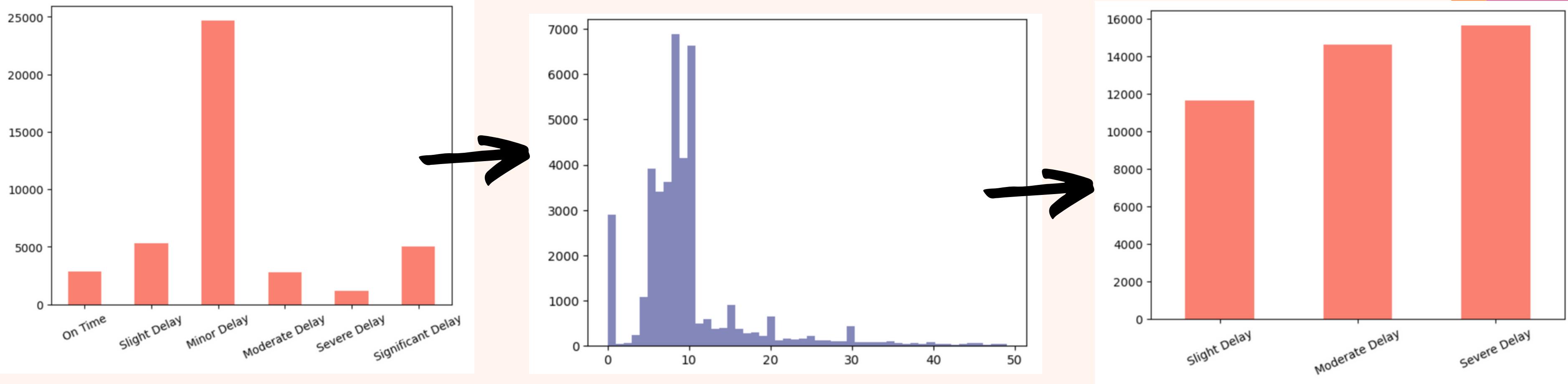
# Imputation Method



Datetime	Temperature	Precipitation	Summer_PT	Winter_PT	Wind_Speed	Visibility	Weather
2021-04-08 07:00:00	10.5	0.0	NaN	NaN	32.0	16.1	LegendNANA
2021-04-08 13:00:00	13.9	0.0	NaN	NaN	39.0	16.1	LegendNANA
2021-04-08 14:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN

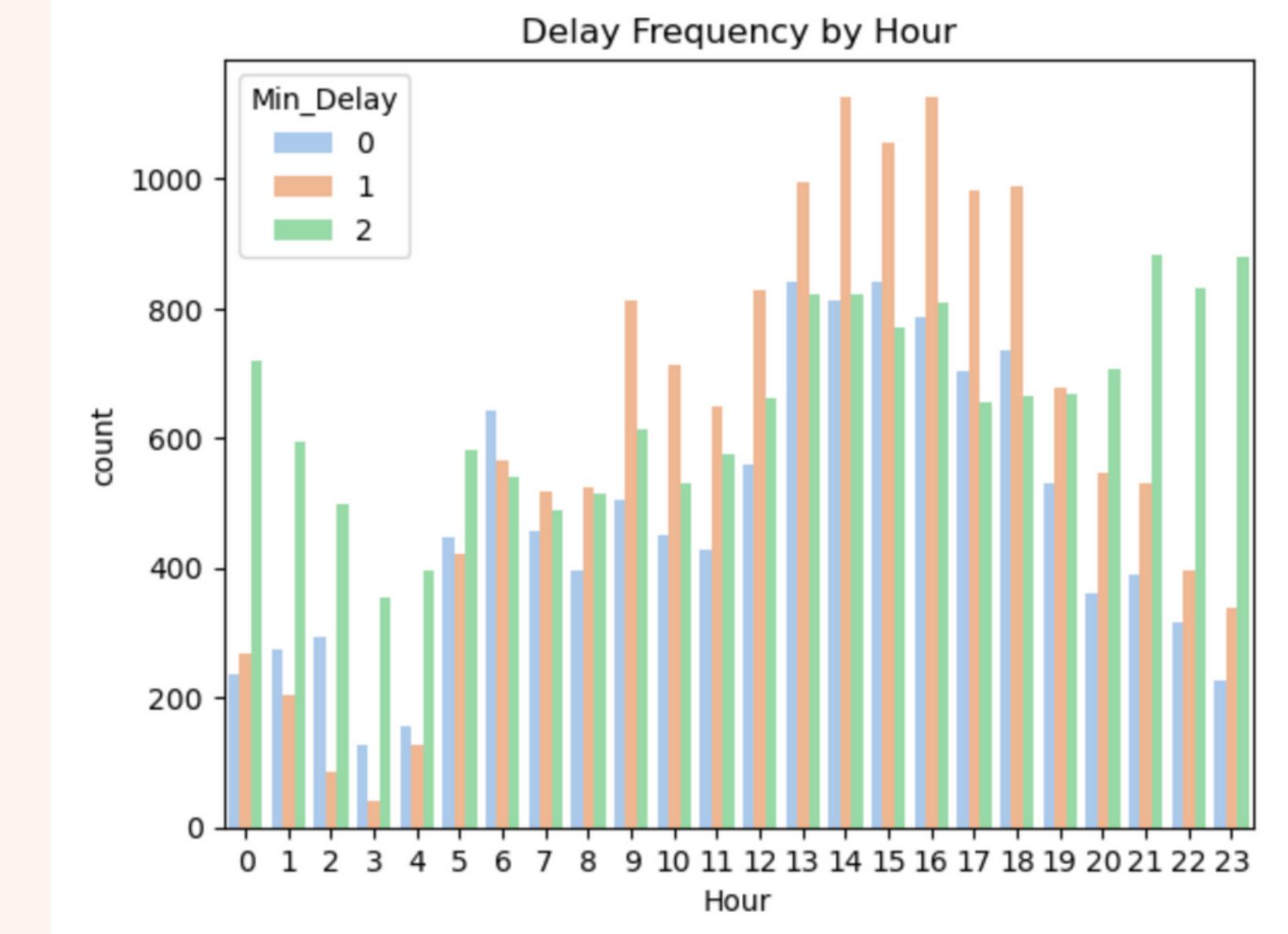
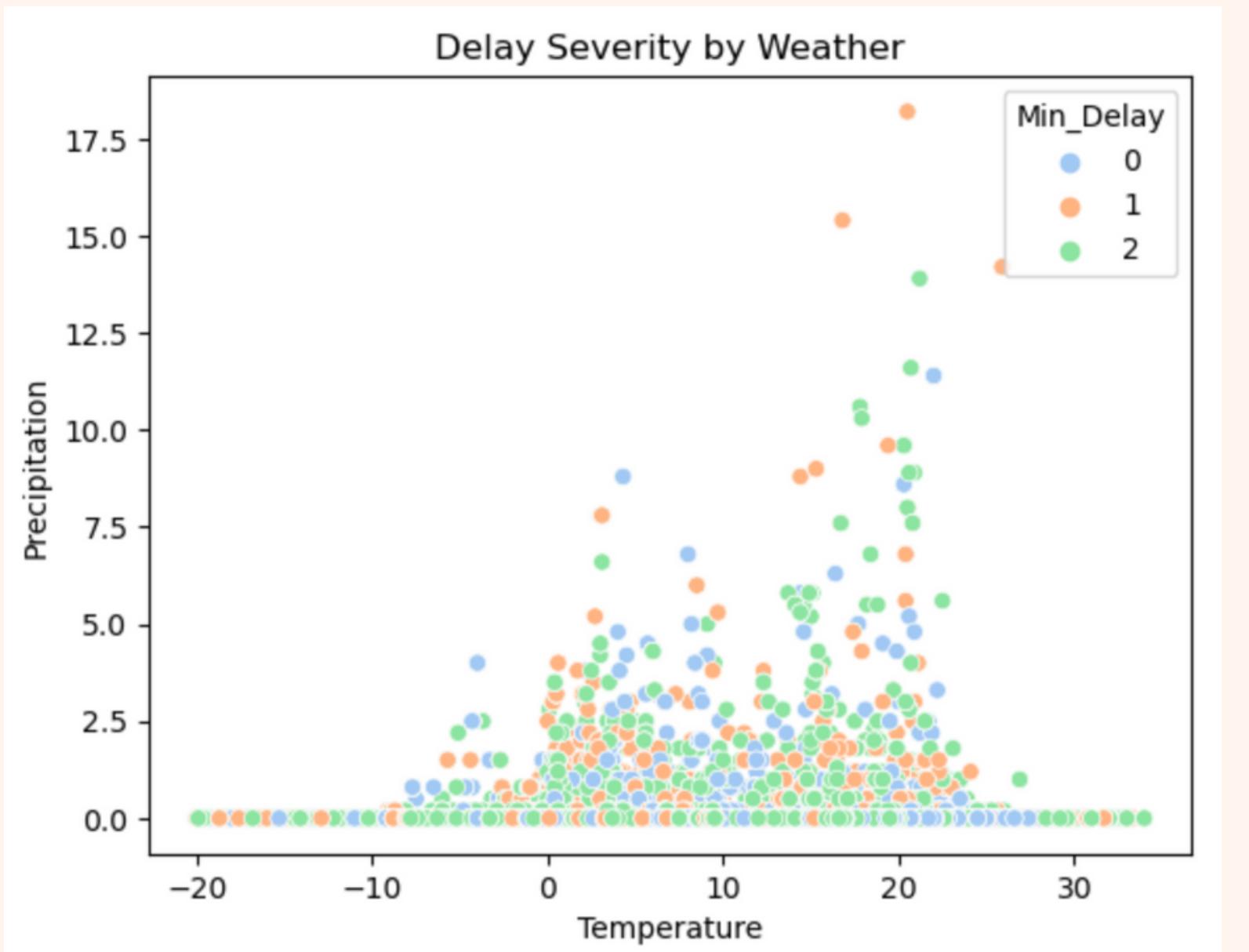
- Weather column had 85% null values:
  - Tactic 1: Explore the data
  - Tactic 2: Investigate feature relationships
    - e.g. Imputing weather data using precipitation, wind speed, and visibility

# Labelling Target Variable



- Highly concentrated in 'Minor Delay' category.
- 8 min delay - 7k approx.
- 10 min delay - 6.5k approx.
- Slight Delay - 0-6 mins
- Moderate Delay - 7-9 mins
- Severe Delay - 10-981 mins

# Interesting Insights



- Rainy days during warmer weather appear to influence the severity of streetcar delays.

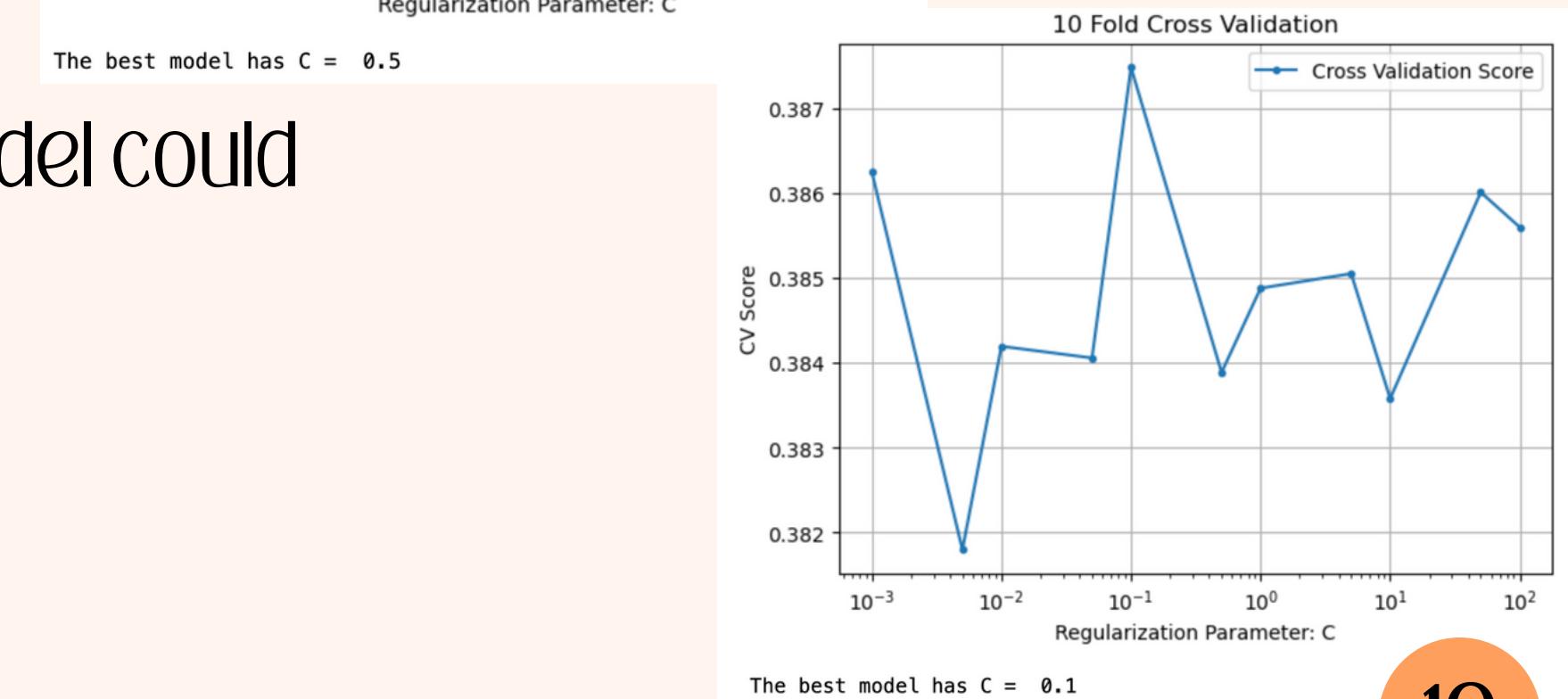
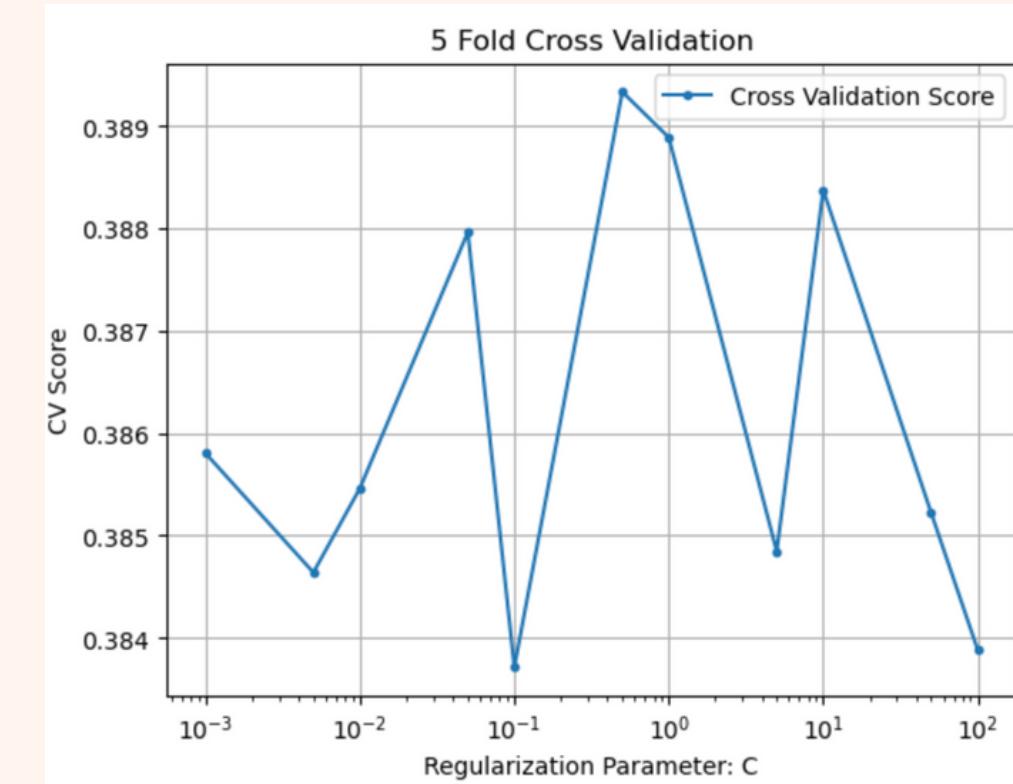
- Streetcars tend to arrive late by 7-9 minutes during rush hour.

Using machine learning, how might **we predict TTC streetcar delays** such that TTC can provide reliable service for all passengers?

# Baseline Model

## Logistic Regression:

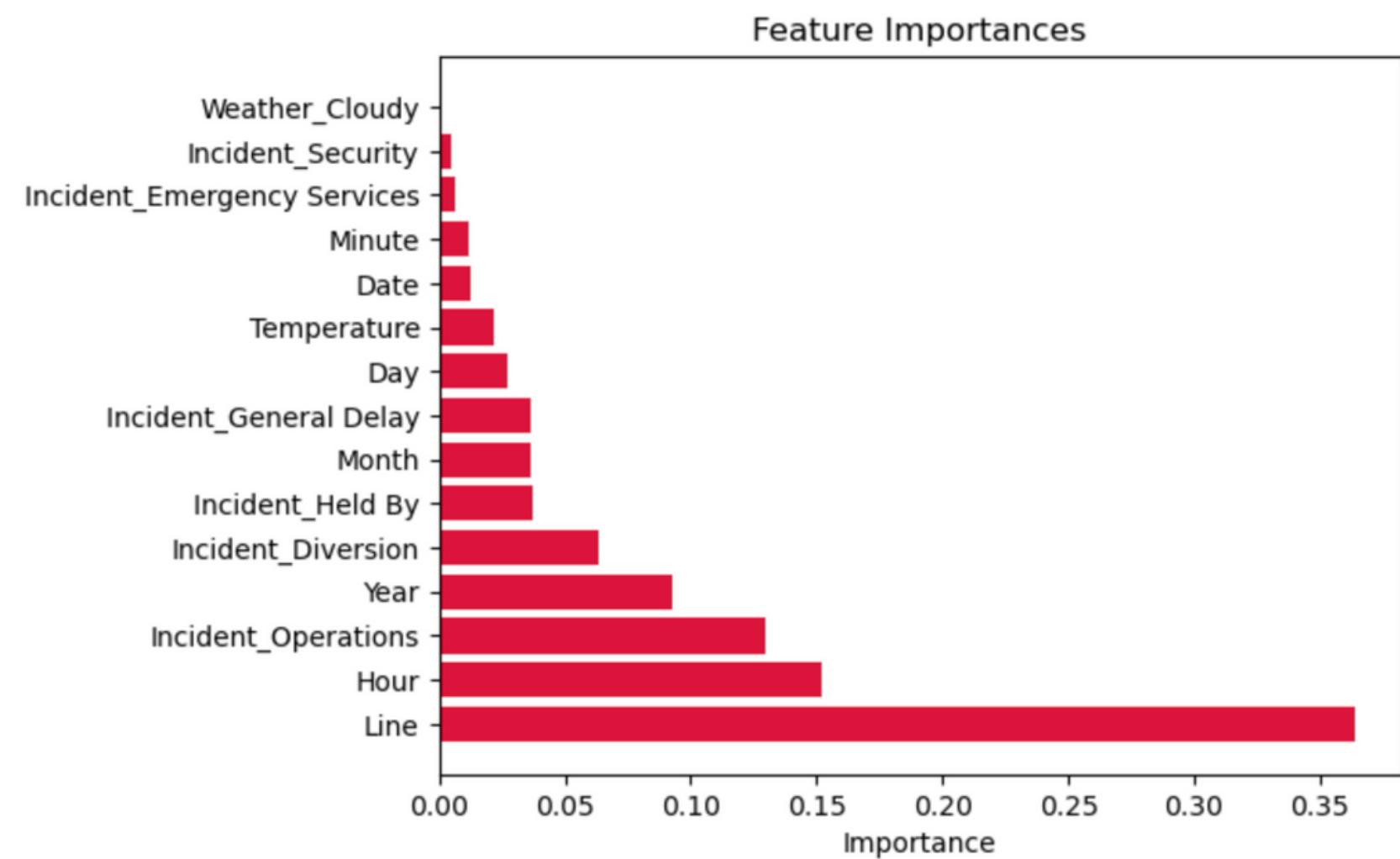
- Multi-class classification data
- Inconsistent C value ( 0.01, 0.5, 1.0)
- Default C value (1.0)
  - Train set Accuracy: 45%
  - Test set Accuracy: 45%
- The accuracy score tell us that the model could be reliable, but...



# Advance Model

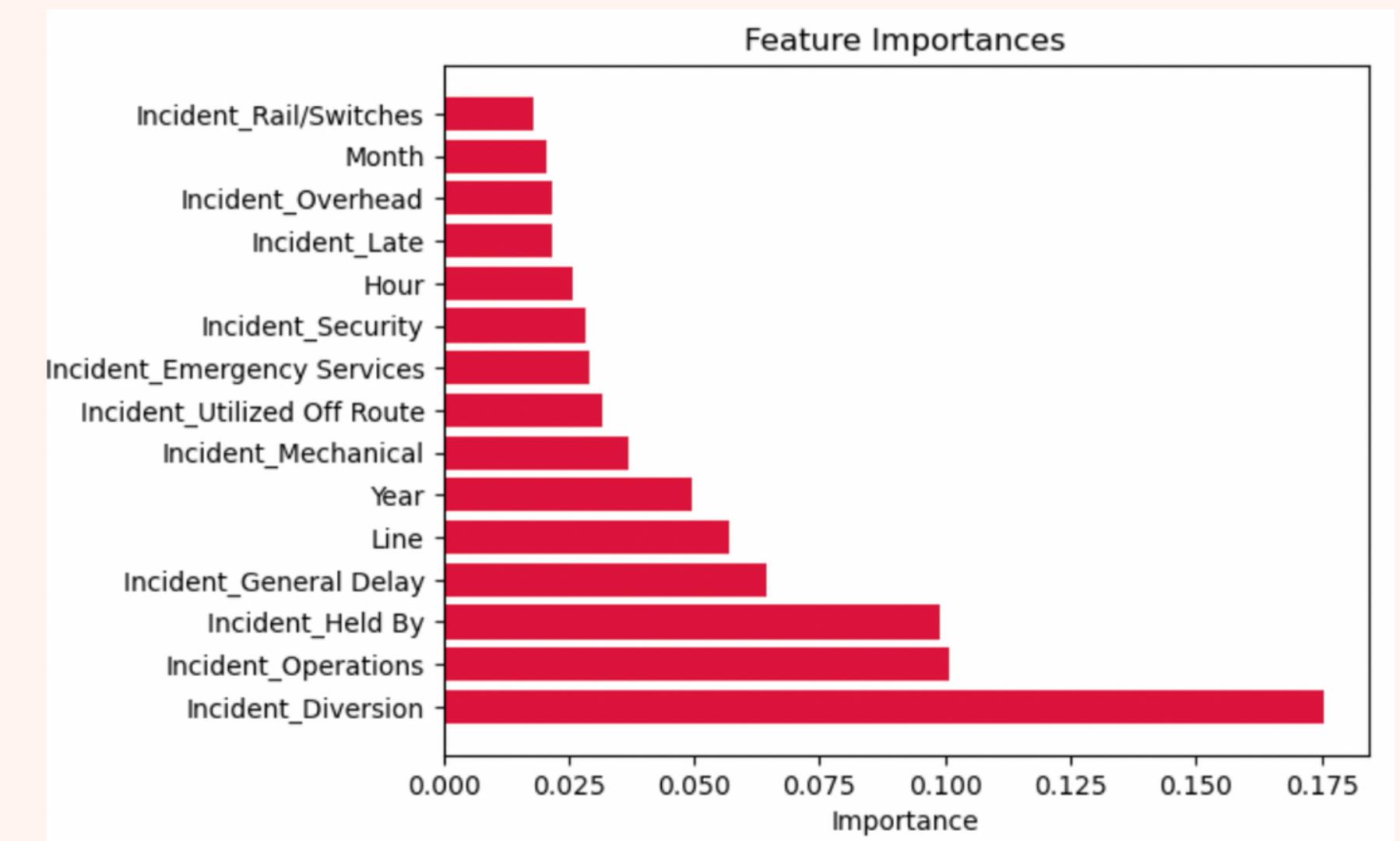
- **Decision Tree Model**

- Default parameters:
- Train set Accuracy: 99%
- Test set Accuracy: 55%
- After finding the best parameters...
- Train set Accuracy: 61%
- Test set Accuracy: 59%
- Line, Time, and Incident has the most influence.



# Advance Model

- **XGBoost**
  - Regularize overfitting data
  - Distribute weights of data points
  - Learn from mistakes
  - After finding the best parameters...
  - Train Set Accuracy: 71%
  - Test Set Accuracy: 63%



# Next Step:

- Streamlit:
  - Be able to select features such as Time, Weather condition, and Incidents to predict the severity of streetcar delays.



Happy Commuting!

# References:

- (1):
  - <https://www.blogto.com/city/2023/11/ttc-overwhelmed-complaints/>
- (2):
  - <https://www.ttc.ca/news/2017/June/TTC-named-North-Americas-best-transit-agency-for-2017#:~:text=The%20TTC%20has%20been%20named,of%20the%20people%20of%20Toronto.>
  -