Vision over Transit Incidents and Claims

Data Driven Approaches to Reducing Insurance Costs to TransLink

Merve Sahin, Brayden Tang, Simardeep Kaur, Xugang Zhong
UBC MDS Capstone

Overview

- Business Questions
- Research Questions of Interest to address 1.
- Data Overview
- Data Product
- Proposed Methodology
- Rough Timeline

Business Question - high costs

- Insurance premium is one of the largest spendings in TransLink's budget
- In the past five years, claim costs have increased by about **122.5%**
- Therefore, we have been asked to find:
  - potential strong predictors of claim severity/frequency that TransLink can leverage to help reduce costs

Research Questions

- What are the main predictors of the frequency and severity of bus accidents?
  - Driver characteristics (probation period, experience)
  - Claim types and other accident descriptions
  - Bus model/model year
  - Bus routes
  - Acceleration/decceleration
  - Weather
  - Time
  - Geographic location
  - and more

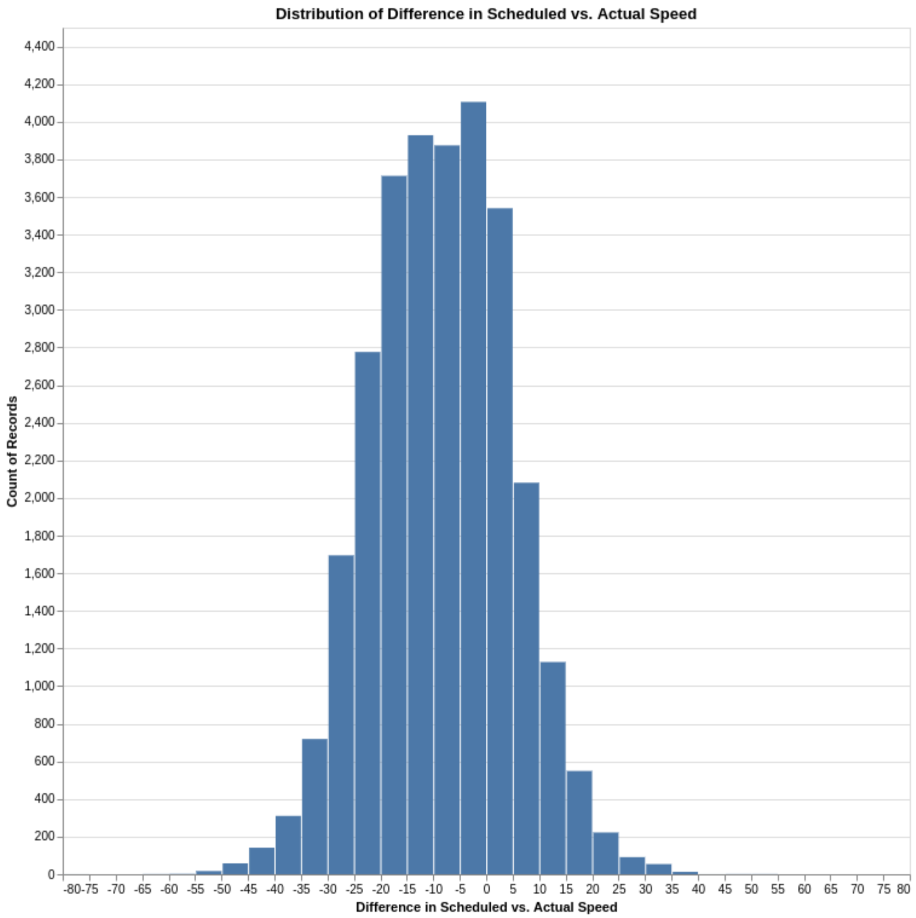
Research Questions (cont)

- Within specific categorical features (such as claim type codes), are there specific clusters or groupings that are particularly noteworthy for having worse or better claims/accident experience?

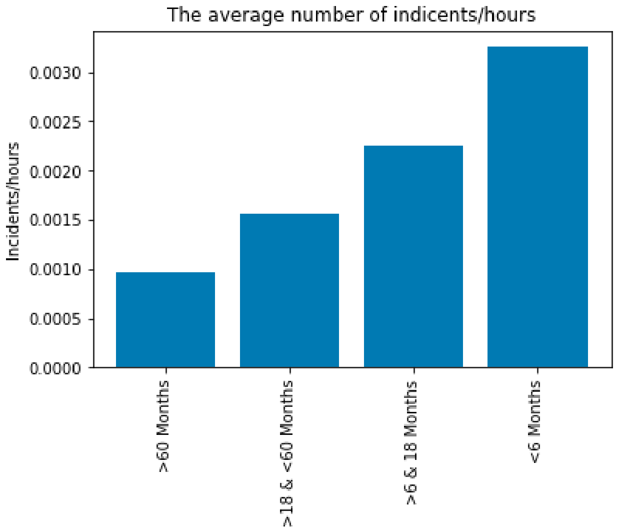
High Level Data Descriptions

- Bus Speeds for All Routes, Route Information
- Actual Incidient Reports
- Collisions (Preventable and Non-Preventable)
- Claims

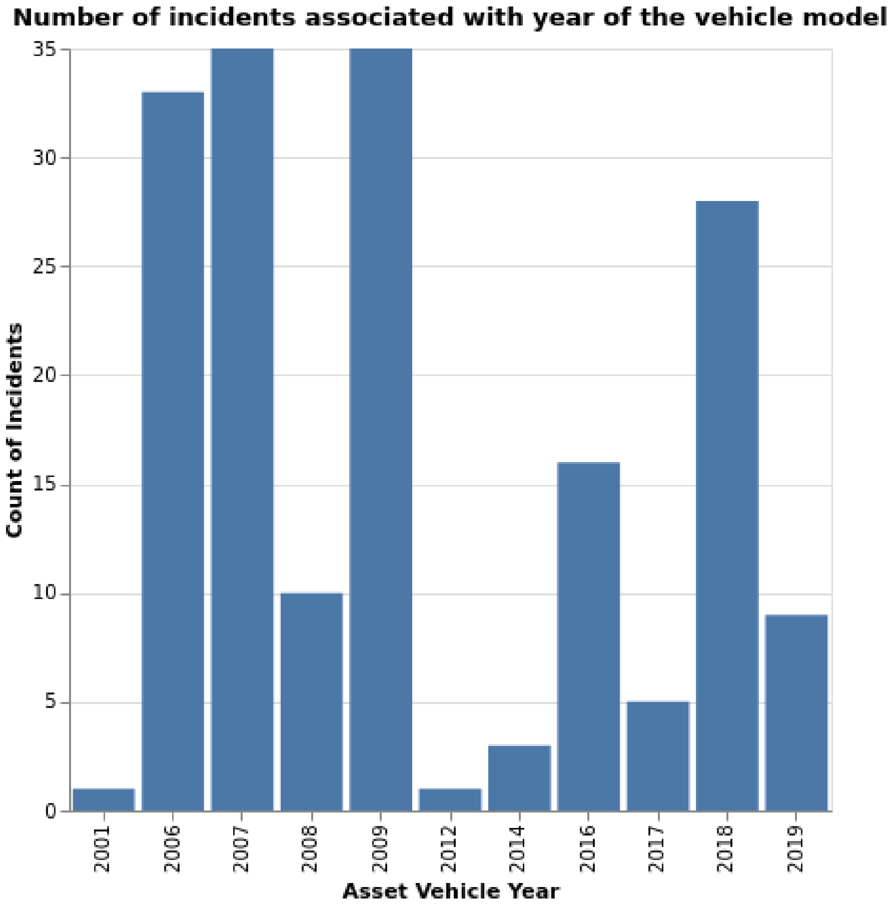
Bus Speeds for All Routes, Route Information



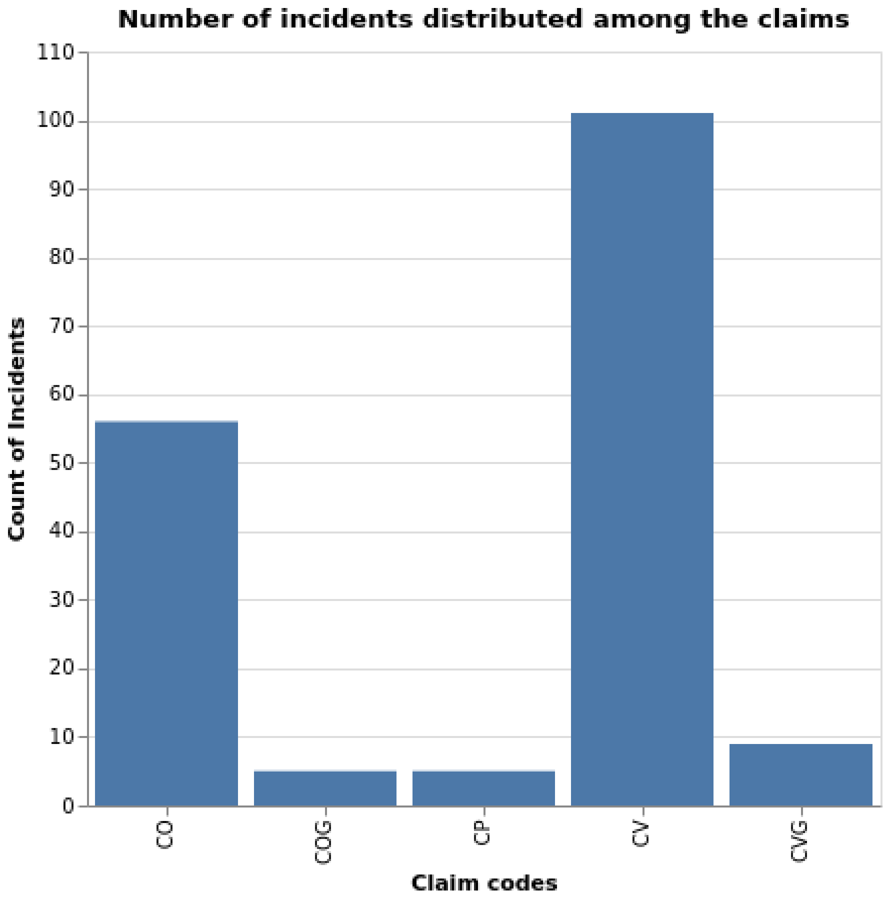
Actual Incidient Reports



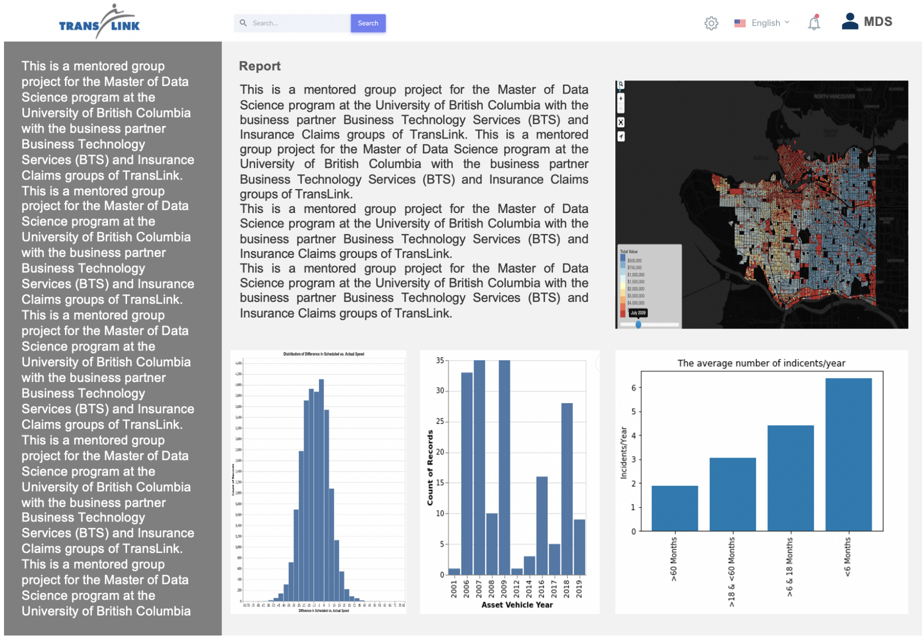
Collisions (Preventable and Non-Preventable)



Claims


Number of incidents distributed among the claims
Count of Incidents
0
10
20
30
40
50
60
70
80
90
100
110
CO
COG
CP
CV
CVG
Claim codes

Data Product

- A reproducible, report that allows the reader to:
  - visualize relationships between claim frequency/severity and specific variables interactively
- 

This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink. This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink. This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink. This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink. This is a mentored group project for the Master of Data Science program at the University of British Columbia

Report

This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink. This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink.

This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink.

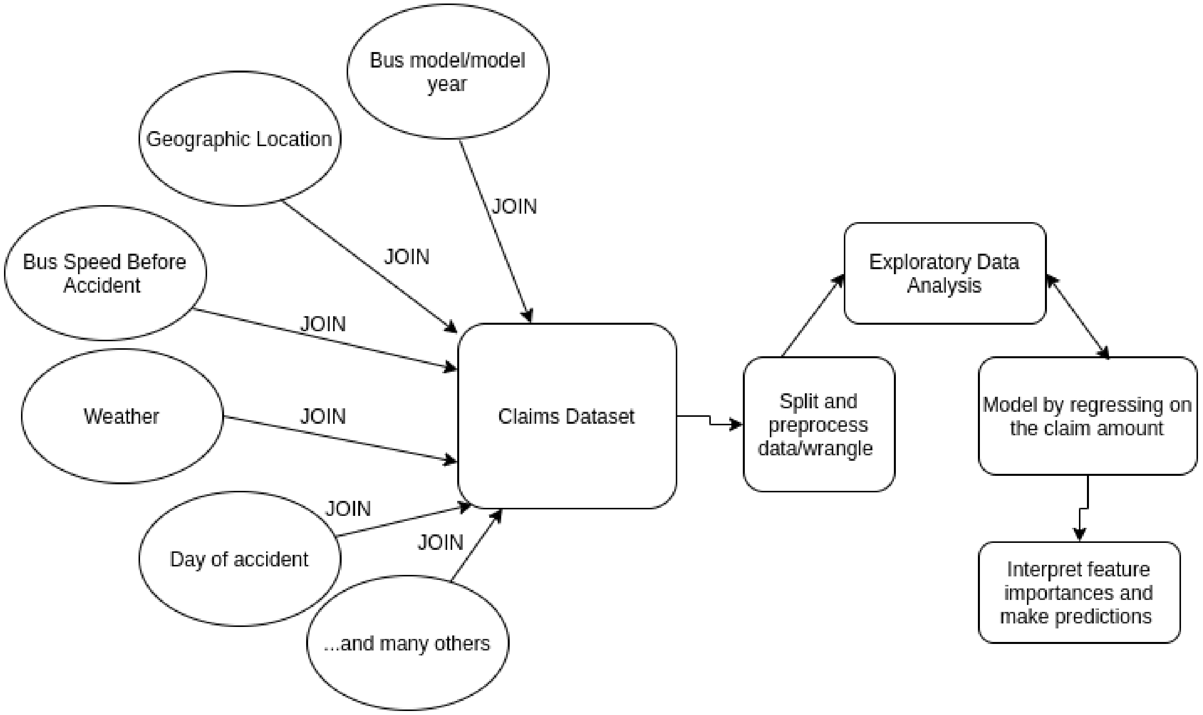
This is a mentored group project for the Master of Data Science program at the University of British Columbia with the business partner Business Technology Services (BTS) and Insurance Claims groups of TransLink.

Data Product (cont)

- A fully reproducible data pipeline
  - user-friendly way to run the entire data analysis front to back using simple Make commands
  - stored on a Docker container
  - detailed documentation describing how to run the analysis and the code

Methodology

- Join all datasets together, split, exploratory data analysis, model, interpret
- Machine learning model that takes in multiple inputs at once
- Emphasis on models that can be interpreted (feature importance scores)

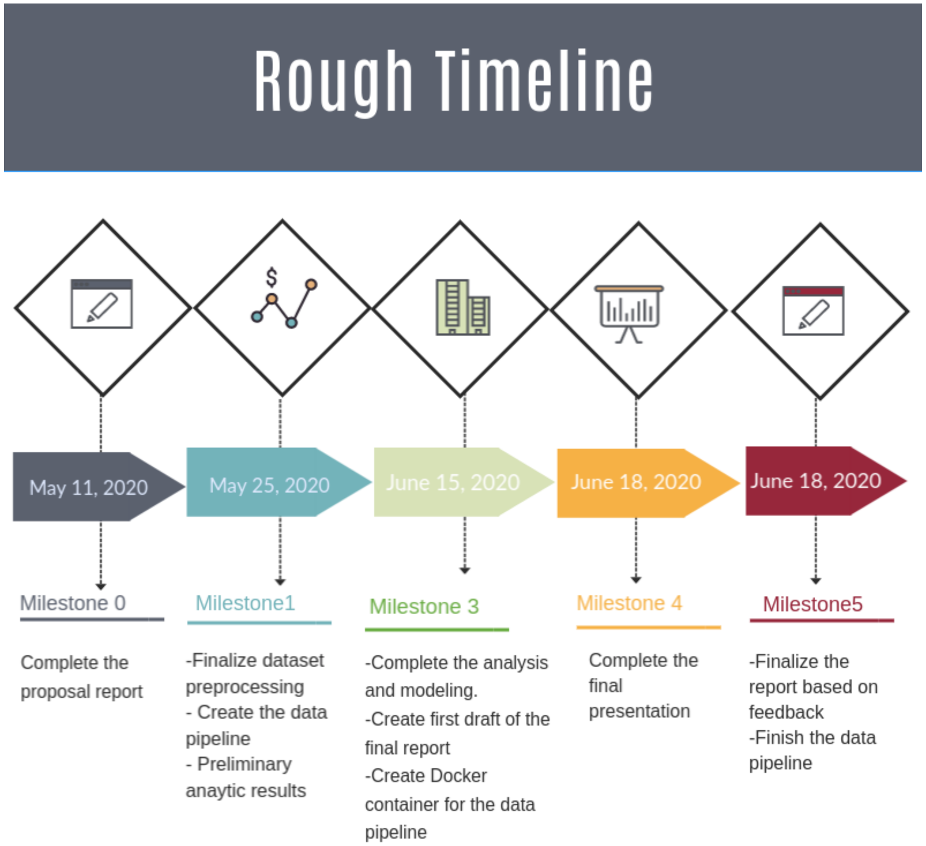

Bus model/model year
Geographic Location
Bus Speed Before Accident
Weather
Day of accident
...and many others
JOIN
JOIN
JOIN
JOIN
JOIN
JOIN
Claims Dataset
Split and preprocess data/wrangle
Exploratory Data Analysis
Model by regressing on the claim amount
Interpret feature importances and make predictions

Methodology (cont)

- Cluster Analysis for analysis of specific categorical features like claim type code, claim description (LDA, DBSCAN)

Rough Timeline

.



Rough Timeline

May 11, 2020

Milestone 0

Complete the proposal report

May 25, 2020

Milestone1

-Finalize dataset preprocessing
- Create the data pipeline
- Preliminary anaytic results

June 15, 2020

Milestone 3

-Complete the analysis and modeling.
-Create first draft of the final report
-Create Docker container for the data pipeline

June 18, 2020

Milestone 4

Complete the final presentation

June 18, 2020

Milestone5

-Finalize the report based on feedback
-Finish the data pipeline

/