## CAPSTONE ABSTRACT

## TTC Bus Delay Data

## <https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#bb967f18-8d90-defc-2946-db3543648bd6>

No doubt that buses present very important mean of transportation in metropolitan cities such as city of Toronto, thousands of Torontonians ride them daily to get to their work, schools and many different places. The repeated delays of the buses are equal to enormous waste of time and money and lead to a lot of frustration of the buses riders who facing these delays.

In my capstone I plan to use TTC Bus Delay Data provided by TTC to form a pattern of these delays over past 56 months from Jan 2014 to Aug 2018, study the main incidents causing these delays, correlate the delays with different seasons of the year, different times of the day, road construction works etc…

I will use R-programming language in discovering the data, building the decision tree as a classification model, then verifying it by using Weka, also, applying time series principals to study the changes of the delay contributors and show their trends over the past few years.

**The Theme:**

Predictive Analytics (Pattern mining, Time-series, Causality, etc.)

**The dataset:**

TTC Bus Delay Data

Owner Toronto Transit Commission

Currency August 2018

Format xls

Refresh Rate Monthly

Contact Open Data Team [opendata@toronto.ca](mailto:opendata@toronto.ca)

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/TTC_Bus_Delay_Data_Metadata.xlsx>

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/Bus%202014.xlsx>

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/Bus%202015.xlsx>

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/Bus%202016.xlsx>

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/Bus%202017.xlsx>

<https://www.toronto.ca/ext/open_data/catalog/data_set_files/Bus_2018.xlsx>

**Research questions:**

* 1. What are the main incidents causing the delay of TTC buses?
  2. What are the old vehicles having frequent delays that need to be retired and replaced with new ones?
  3. What are the routes suffering either longer or repeated delays and what are the alternative solutions?
  4. How is the pattern of the delay varied over the past 56 months 2014-2018, how it looks like in different seasons?

<https://developers.google.com/transit/gtfs/reference/?csw=1#stopstxt>