Exploratory Data Analysis of TTC Bus Delay Data to Reduce Waiting Time at the Bus Stop

Sarthak Oke, Akef Ali

***Abstract*— In large cities, transportation systems have a significant impact on residents' quality of life. Despite numerous policies and strategies to increase service quality and dependability, commuters continue to criticize public transportation. This is primarily caused by the buses arriving at incorrect times, either early or late. These anomalies can result in commuters missing crucial appointments, having to wait a long time at bus stops in bad weather, or even being late for work. Since time is a valuable resource, bus passengers would like to have accurate arrival time information. To accurately analyze bus arrival times, a number of research techniques have been developed but few do so with results that are extremely precise and based on open data. In this research, we look at data that is readily accessible on transportation, including data on bus delays. For urban analytics, In order to analyze information about public transportation, such as historical bus arrival times, the system specifically accesses open data, performs exploratory data analysis and visualization to look for recurrent patterns. According to the patterns it has found, it analyzes whether the bus will arrive on time or not and how much delay you should expect. The outcomes of this research are positive for the objective of creating smart cities.**

1. INTRODUCTION

It has been understood how crucial it is to model and forecast bus arrival times for public transportation [1]. In the last decade, much effort has gone into developing delivery systems that are more reliable and faster. [2]. However, commuters continued to criticize public transportation authorities for inconsistencies between the time of expected and actual vehicle arrival. Naturally, these anomalies have a detrimental effect on commuters' daily lives. Commuters may get late for work, miss important appointments, and spend a lot of time waiting at bus stops in inclement weather. Due to the availability of extensive and pervasive data, we can analyze the time of actual arrival of public transportation and analyze and provide them with solutions to deal with their differences from scheduled times. In order to model the irregularities in public transportation bus arrival times, this study uses time of historical bus arrival, location of bus stops, schedules of bus, and weather information. One of two ways that irregularities can manifest itself is through leads like the bus being early or delays. In the City of Toronto, where Toronto Transit Commission issued notes to the commuters who turned up late to work as a result of incorrect scheduling times, emphasis is made on analyzing the anomalies of the buses. [3].

2. METHODOLOGY

2.1. Data Collection

For the data collection we use the open data provided by government of Canada. Where they took initiative to provide the bus delay data and other information about delay in bus between stops. The open data initiative in Canada is a way of opening the city data to public that many major cities are undertaking. They took this initiative as a way to support research which things can get improvised and solve these issues. Since Durham Region Open Data was not able to provide us with bus delay data we procure our data from one of the biggest Public Transportation of Canada Toronto Transit Commission (TTC) from Toronto Open Data (https://open.toronto.ca/dataset/) where we were able to obtain the Bus Delay Information with delay time and also reason for delays. The Dataset that we gathered consisted of Minimum Delay and Incident Due to which the delay was occurred.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	Route	Time	Day	Location	Incident	Min Delay	Min Gap	Direction	Vehicle		
2	1-Jan-22	320	02:00	Saturday	YONGE AND DUNDAS	General Delay	0	0		8531		
3	1-Jan-22	325	02:00	Saturday	OVERLEA AND THORCLIFFE	Diversion	131	161	W	8658		
4	1-Jan-22	320	02:00	Saturday	YONGE AND STEELES	Operations - Operator	17	20	S	0		
5	1-Jan-22	320	02:07	Saturday	YONGE AND STEELES	Operations - Operator	4	11	S	0		
6	1-Jan-22	320	02:13	Saturday	YONGE AND STEELES	Operations - Operator	4	8	S	0		
7	1-Jan-22	363	02:16	Saturday	KING AND SHAW	Operations - Operator	30	60		0		
8	1-Jan-22	96	02:18	Saturday	HUMBERLINE LOOP	Security	0	0	N	3536		
9	1-Jan-22	320	02:38	Saturday	STEELES AND YONGE	Operations - Operator	4	8		0		
10	1-Jan-22	320	02:55	Saturday	YONGE AND STEELES	Operations - Operator	4	8		0		
11	1-Jan-22	300	03:18	Saturday	KENNEDY STATION	Emergency Services	0	0	E	8094		
12	1-Jan-22	300	03:32	Saturday	BLOOR AND INDIAN	Security	17	34	E	8452		
13	1-Jan-22	47	03:34	Saturday	LANSDOWNE AND ST CLAIR	Operations - Operator	15	30	S	0		
14	1-Jan-22	45	03:52	Saturday	DANFORTH AND DANFORTH	Operations - Operator	30	60	W	1325		
15	1-Jan-22	32	04:21	Saturday	EGLINTON STATION	Operations - Operator	16	33		1130		
16	1-Jan-22	32	04:39	Saturday	YONGE AND BERWICK	Emergency Services	0	0	S	1267		
17	1-Jan-22	39	04:42	Saturday	FINCHDENE AND FINCH	Operations - Operator	30	60	W	0		
18	1-Jan-22	32	04:53	Saturday	RENFORTH STATION	Operations - Operator	30	0	E	1073		
19	1-Jan-22	53	04:58	Saturday	STEELES AND BAYVIEW	Security	30	60	E	3299		
20	1-Jan-22	29	05:01	Saturday	DUFFERIN AND LAWRENCE	Operations - Operator	10	20		9149		

Figure 1

2.2. Data Cleaning

First we had calculated the total number of null values in the columns of the dataset using *df.isnull().sum(),* where we were able to see more than 8000 null values in the Direction column and this column cannot be filled, and we do not require the Direction in relation to the route column. So we dropped the column for the directions. Also we were able to depict that there were 164 null values in the Route or the Route Number in which the bus was moving were also missing. After analyzing the Route data we saw that most of them had 0 Delay so we dropped these column using *df = df[df['Route'].notnull()]*.Thus we were able to clean our data.

Also the date format was not proper so we converted it into proper date-month-year numeric form using *df['Date'] =df['Date'].apply(pd.to_datetime).*

3. RESULTS

For visualizing the results, we first find the most common cause of delay. For doing this we plot a bar chart.

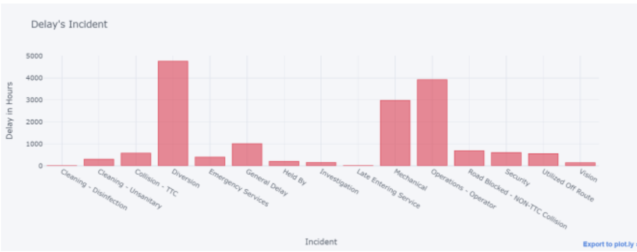

Figure 2

- Diversion is found to be the most common cause of delay with 4764 hours in delay
- Followed by operations delay which has 3922 hours of total delay in the 10 months of 2022
- Delays due to mechanical reasons come at the third place with 2979 hours of delay
- We can comprehend from this that there is a lot of construction going on currently in Toronto and that is the main cause of delay

Next we find out the most problematic routes i.e. the routes with the most delays in Toronto.

Route_20

✓ 0.3s

	Route	Min Delay
135	Lawrence West	28055
100	Finch West	22456
80	Eglinton West	22057
223	Yong	19051
220	Wilson	15988
69	Dufferin	15878
59	2	15852
137	Lawrence East	15797
23	121	15504
107	39	14708

Figure 3

From figure 3 it can be seen that Lawrence west is the most problematic route with a delay of 28055 minutes followed by finch west with 22456 minutes and Eglinton west with 22057 minutes of total delay in the first 10 months of 2022.

To better understand these routes we replace the route numbers with the names.

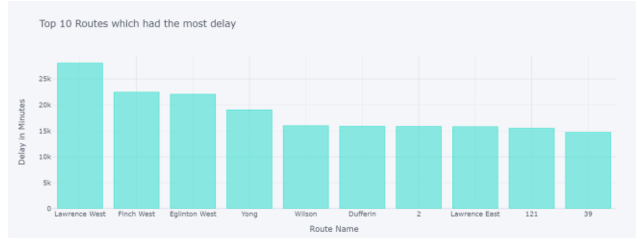

Figure 4

Now we look at the delays based on the day of the week.

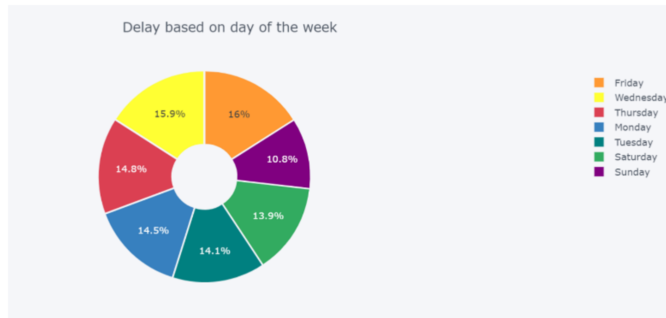

Figure 5

Even though it is apparent from figure 5 that delays are almost the same on every day of the week, Saturday and Sunday has the least delay and Wednesday and Friday has the most delay during the week.

Now lets take a look at the month wise delays from january to october which gives nifty insights.

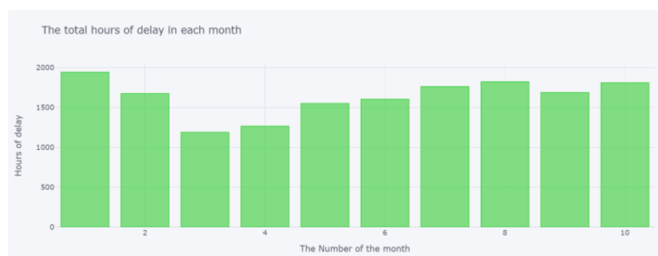

Figure 6

- January has the most delays and the delay time gradually decreases from February to April and then starts to increase again from May till august. The delays suddenly go down in September and then up in October.

Since January has the most delay in all the 10 months, lets investigate deeper into this month.

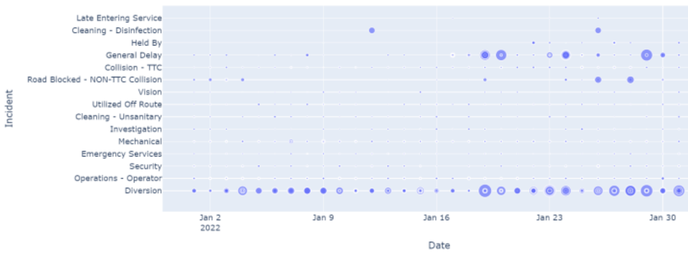

Figure 7

When January was analysed the most delay was in the second half of the month which was due to the unprecedented snow storm. The news extract about this snow storm can be seen in figure 8.

TORONTO

REPORT FOR ACTION

January 16-17, 2022 Major Snow Event Post-Operational Report

Date: March 17, 2022
To: Infrastructure and Environment Committee
From: General Manager, Transportation Services
Wards: All

SUMMARY

On January 16-17, 2022, the City of Toronto experienced an extraordinary winter storm that involved extreme cold temperatures, very rapid snowfall, and an ultimate snow accumulation of 55 centimetres in just 15 hours. The below freezing temperatures that followed the storm and lasted for more than two weeks created a unique set of challenges for storm clean up. Responses to this winter weather event required additional efforts above and beyond typical salting and snow plowing activities, and included significant involvement from Transportation Services, Strategic Public and Employee Communications, 311, and Fleet Services.

Figure 8

4. CONCLUSION

Canada has a lot of students who rely on public transportation. It can get difficult to wait at the bus stops during harsh winters. By use of this analysis we can reduce waiting time at bus stops by providing the results of this analysis to the Transportation authority of Toronto to provide more number of vehicles in between the busiest Bus Stops which can help commuters to face less difficulties in harsh weather. Also the most profound reason for bus delays is due to diversions i.e. construction work. The government needs to take firm steps on bolstering the transit system in areas with dense student population. The identified problematic routes must be looked into and resolved.

5. REFERENCES

[1] Kumar V, Kumar BA, Vanajakshi L, Subramanian SC (2014) Comparison of model based and machine learning approaches for 1 bus arrival time prediction. Transportation Research Board 93rd Annual Meeting.

[2] Hua X, Wang W, Wang Y, Ren M (2018) Bus arrival time prediction using mixed multi-route arrival time data at previous stop. Trans-port 33(2):543–554.

[3] Star TT (2020) Ttc gives notes for affected customers arriving late for work. https ://www.thestar.com/news/gta/2017/12/01/late-for-work-the-ttc-can-give-you-a-note-for-that.html Accessed 2020.

[4] Altinkaya, M. and Zontul, M. (2013). Urban bus arrival time prediction: A review of computational models. International Journal of Recent Technology and Engineering (IJRTE), 2(4):164–169.

[5] M. Kormksson, L. Barbosa, M. R. Vieira, and B. Zadrozny, “Bus travel time predictions using additive models,” in 2014 IEEE International Conference on Data Mining, Dec 2014.

[6] Ang, R.P., & D.H. Goh (2013) “Predicting juvenile offending: a comparison of data mining methods,” Int. J. Offender Therapy and Comparative Criminology 57(2).

[7] Ashraf, N., R.R. Haque, M.A. Islam, C.F. Ahmed, C.K. Leung, J.J. Mai, & B.H. Wodi (2019) “WeFreS: weighted frequent subgraph mining in a single large graph,” in ICDM 2019.

[8] Chowdhury, N.K., & C.K. Leung (2011) “Improved travel time prediction algorithms for intelligent transportation systems,” in KES 2011, Part II.

[9] Chowdhury, N.K., R.P.D. Nath, H. Lee, & J. Chang (2009) “Development of an effective travel time prediction method using modified moving average approach,” in KES 2009, Part I.

[10] Lin, Y., X. Yang, N. Zou, & L. Jia (2013) “Real-time bus arrival time prediction: case study for Jinan,” China J. Transport. Eng. 139(11).