Toronto Cyclist Accidents – An Analytics Study on KSI Accidents

GMMA 860 – Acquisition and Management of Data

New York Analytics Group

Smith SCHOOL OF BUSINESS | Queen's University

Table of Contents

Recent Context

Cycling is one of the fastest growing modes of transportation in Toronto

According to the 2016 Census, the number of bicycle commuters in all of Toronto rose 58% - from 1.7% in 2006 to 2.7% in 2016.

Over 10 period (2008-2018) there were a total of **548 Killed and Seriously Injured (KSI) cyclist** in the city of Toronto.

The City of Toronto also allocated $54M in 2016 for its **Vision Zero Road Safety Plan**, the goal of which is to reduce traffic-related fatalities and serious injuries on city streets via several “emphasis areas”, one of which is cyclists.

Safety being a prime concern of both existing and aspirational bike commuters, Toronto City Council approved capital funding in 2016 for bicycle infrastructure of $16M annually for 10 years

City of Toronto. [2018] 2018 Cycling in Toronto. Retrieved from https://www.toronto.ca/city-government/council/2018-council-issue-notes/cycling-in-toronto/

Objectives

Team New York will investigate past bicycle accident data (from 2008-2018 inclusive) to determine additional opportunities for the Vision Zero program to make cycling safer in Toronto.

VISIONZERO

"**Vision Zero** is a comprehensive five year (2017-2021) action plan focused on reducing traffic-related fatalities and serious injuries on Toronto's streets. With over 50 safety measures across our six emphasis areas, the Plan prioritizes the safety of our most vulnerable road users, through a range of initiatives."

Vehicle Action Plan

- Aggressive Driving and Distraction
- Slowing-Down Campaign
- Slowing-Down Measures
- Improved Traffic Control

Bicycle Action Plan

- Enhanced Safety Measures
- Cycling Lanes
- Safety Education
- Safety Campaign

Are there additional initiatives or opportunities to improve the VisionZero program?

*Killed and Seriously Injured

City of Toronto. [2018] 2018 Cycling in Toronto. Retrieved https://www.toronto.ca/services-payments/streets-parking-transportation/road-safety/vision-zero/vision-zero-plan-overview/

Recommendation Overview

Reduce KSIs in Toronto

Vision Zero Emphasis in the Suburbs

Review Toronto's investment in suburbs to improve cycling safety.

Targeted Safety Education Campaign

Promote road sharing, reduce aggressive and distracted driving/cycling.

Traffic Control, Intersection and Light Sequencing

Control and manage bicyclist and driver interactions during rush hour.

Toronto should revisit cycling safety in the suburbs to address a stagnant progression in lowering KSI accidents

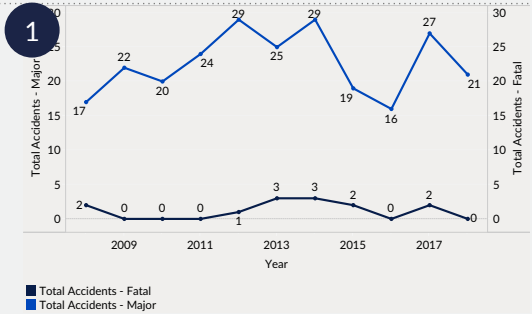
VisionZero Emphasis on the Suburbs

The Vision Zero plan has been focusing on security measures and awareness campaigns that targets zones within the city core of Toronto. However, our analysis shows that while downtown accidents are trending downward suburban accidents are stagnant, especially in the Scarborough and Don Valley areas.

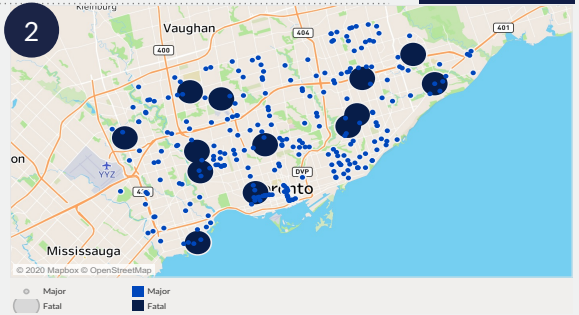
Moreover, the highest proportion of accident involves a vehicle performing a turn maneuver and colliding with a cyclist. We suspect there is a lack of cycling infrastructure.

The current Vision Zero infrastructure plan should be reviewed with suburbs in mind. Therefore, understanding what measures were most effective will allow to tailor an action plan for the suburbs.

1



2



1 While city center (Ward 4 & 9-14) had a decrease in KSI since 2014, KSI accidents in suburbs has remain stagnant over the same period.

2 From the historical dataset, we have observed an increasing trend of accidents involving cyclists aged 50+, especially in the suburban Wards.

3 Data shows that most common conditions of accidents in the suburbs are involving a cyclist going forward and a vehicle turning.

4 Our predictive model indicates that being in the suburbs increases your probability of a KSI by just over 50%.

We recommend a review of the cycling infrastructure project plan in the suburb area and determine if further investments are required.

Toronto should deploy a transportation safety awareness campaign to reduce preventable accidents and eliminate fatalities

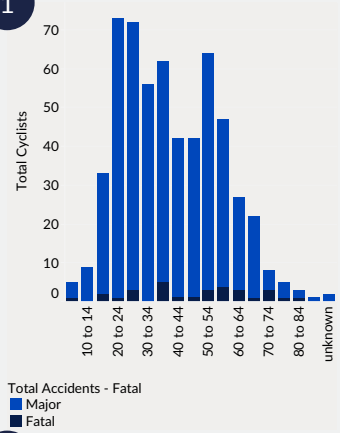
Targeted Safety Education Campaign for 20-30 age group

We recommend deploying **transportation safety awareness campaigns** focused on pedestrian, cyclist, and driver safety. The campaigns should be delivered through **traditional** and **digital** means with the primary objective of **reducing accidents and eliminating fatalities**.

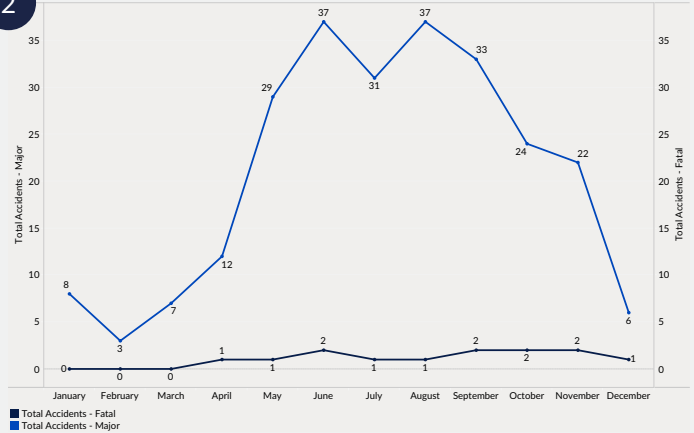
A sample video storyline could be between 15-30 seconds which allows for quick consumption (for younger demographics) via social media or traditional TV. This example can be adapted for each district in Toronto.

- A young professional leaves work at 5pm for their Thursday summer cocktail hour. A driver is commuting home while distracted on a call and misses the traffic signal turned red. A cyclist rushing home equally ignores the red light. All 3 parties collide fatally with the pedestrian seriously injured and cyclist killed.

1



2



1 The awareness campaign should target Torontonians across the **GTA** between the **ages of 20-30**. The bar chart shows over 25% of cyclists injured are within this age group.

2 The campaigns should take place prior to the summer, between **April-June**, and **August-October**, in preparation for the back to school rush. Chart 2 shows that highest number of accidents occur during this time period.

3 Bicycle aggression was also correlated with driver actions, aggression, traffic violations indicating that aggressive bicycling has a link to vehicle state of mind.

The goal of the campaign should be to promote road sharing and reducing situations with multiple parties in aggressive or distracted states of mind. Further education on cycling in adverse weather conditions.

Toronto needs to manage and control bicycle & driver interactions at intersections during peak hours

Traffic Control, Intersection and Light Sequencing

To tackle the issues of dangerous and potentially fatal actions of both cyclists and drivers, Toronto needs to re-evaluate the concept of "Right of Way". This is particularly heightened during rush hour.

Some possible solutions include the following:

- Scramble Intersections
- Provide right of way to one group of commuters at a time
- Traffic Control

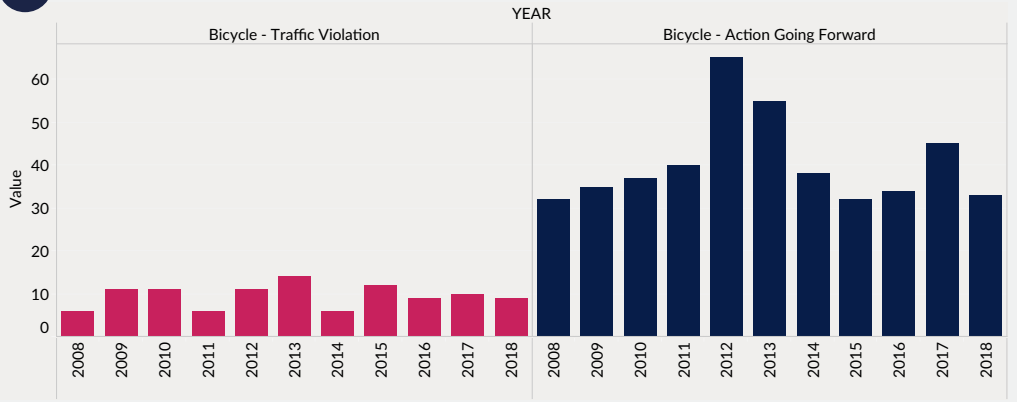
1

Over the last 10 years, we see **steady** volumes of **accidents** involving a **cyclist** going **forward** and committing **traffic violations**.

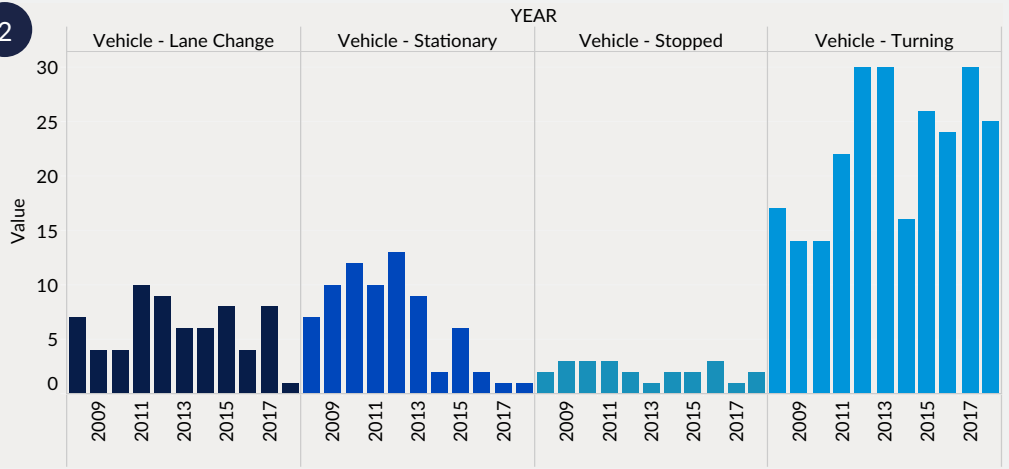
2

Over the last 10 years, we also see an **increase** in accident volumes involving **vehicles turning**.

1



2



Toronto needs to manage and control bicycle & driver interactions at intersections during peak hours (cont'd)

Traffic Control, Intersection and Light Sequencing

To tackle the issues of dangerous and potentially fatal actions by both cyclists and drivers, Toronto needs to re-evaluate the concept of "Right of Way". This is particularly heightened during rush hour.

Some possible solutions include the following:

- Scramble Intersections
- Provide right of way to one group of commuters at a time
- Traffic Control

3

According to our analysis, **38%** of major and fatal accidents **occur during rush hour** times (between 8am-10am & 4pm-6pm)

3

Hour	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Grand Total
0	2		1	1	2	2	3	11
1		1	1	1	1	2	2	8
2				1	1	2	1	5
3					2		2	4
4							1	1
5			2		2	1	1	6
6	2	1	3	3	3		1	13
7	4	4	6	2	3			19
8	3	9	7	6	5	1		31
9	1	7	6	7	4	1	1	27
10	12	4	4	3	6	1	1	31
11	3	4	4	1	3	3	1	19
12	2	7		4	1	6	6	26
13	6	3	3	9	4	2	2	29
14	6	5	4	9	3	5	6	38
15	3	11	3	8	6	4	3	38
16	3	5	3	8	8	3	3	33
17	6	10	10	6	7	4	3	46
18	3	6	11	5	7	3	10	45
19	7	4	6	8	2	2	4	33
20	5	1	5	4	1	4	2	22
21	2	6	2	10	5	3	3	31
22	1		4	3	4	2	1	15
23	2	2	2	2	1	4	4	17
Grand..	73	90	87	101	81	55	61	548

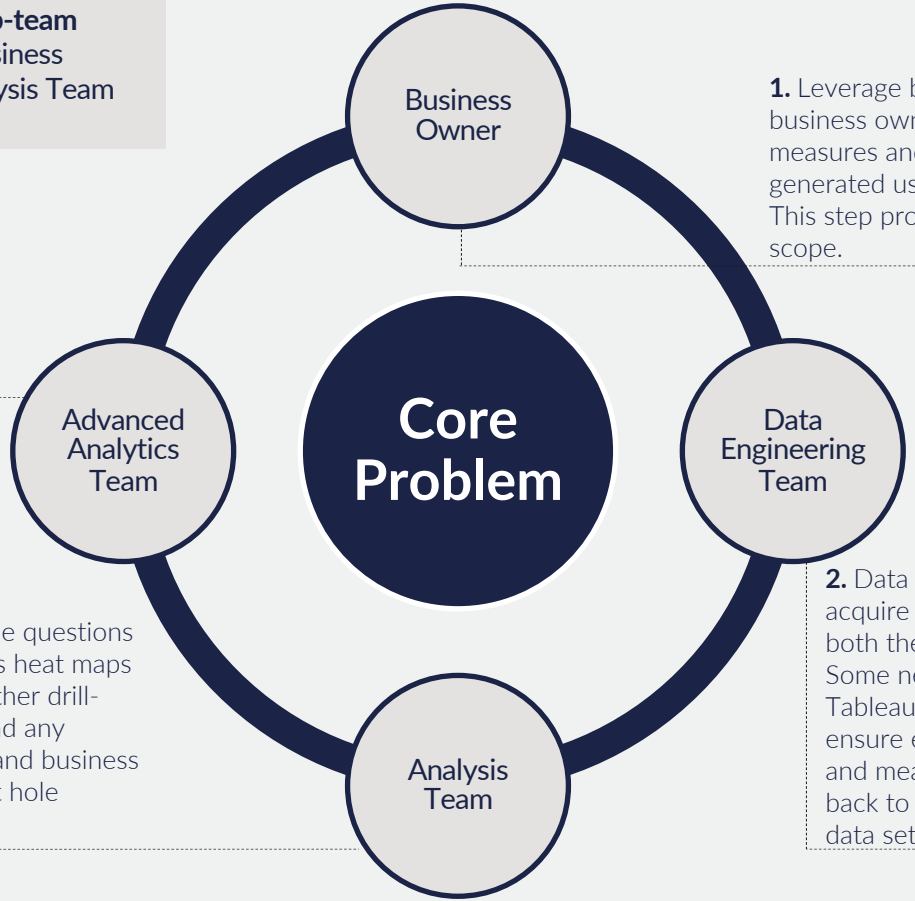
More protections and emphasis around controlling bicycle and driver interactions during peak commute times need to be put in place, as accidents that occur during the morning rush hour are about 25% more likely to be fatal/major and accidents that occur during the evening rush hour are about 80% more likely to be fatal/major.

TECHNICAL SUMMARY

Table of Contents

Team Organization

The team worked was built into 4 sub-team based on the following functions: Business Owner, Data Engineering Team, Analysis Team and Advanced Analytics Team.



1. Leverage business questions posed by the business owner to explore variables and measures and see what insights could be generated using various visualization methods. This step provided the team with guidance and scope.

2. Data Engineering team was responsible to acquire data and clean datasets to be used by both the business team and analytics team. Some new measures were created ad-hoc in Tableau as the data exploration was ongoing to ensure efficiency. Ultimately, the final variables and measures that were created were reported back to the DE team to be included in the final data set.

3. Fluid analysis process to answer top-line questions using the appropriate visualization such as heat maps for time of day/day of week analysis. Further drill-down was performed if team members had any questions, keeping in mind the objective and business questions so as not to go down the rabbit hole (analysis-paralysis).

4. Complete feedback loop with the Advanced Analytics team to validate predictive insights team generated from to ensure accuracy and relevancy.

Methodology

Define

The project scope was defined based on articles and blogs addressing the issues of killed cyclist in the city of Toronto, and the Vision Zero. This led to the development of the business problem, and the structure of the analytics research. *

Discover

Dataset research was conducted to gather data relevant to the business questions. The Data Engineering team conducted the joining, cleaning and structuring of the accident dataset.

Explore

The Analytics team and the Data Engineering team focused on descriptive analytics to better understand the current situation related to cycling accidents. This led to the development of the initial recommendations and supported the development of future regression model.

Analyze

The Analytics Team built a logit regression model to predict the probability of injuries, and determined significant conditions leading to such injury.

Communicate

This stage is the consolidation of all the information into a comprehensive SCR presentation that is concise, but informative. Our approach was to combine both statistics and visual aids to support the recommendation.

Operationalize

The objectives is not to operationalize the regression model, but to provide strong recommendation on how the city can adapt their Vision Zero objectives. Those recommendation are targeted high-risk conditions to prevent accidents.

*Business Questions are included in appendix H

Technique 1 to 3 – Data Acquisition and Cleaning

Data Summary

The cyclist dataset is data sourced from the Toronto Police open data portal. It is a subset of the Killed and Seriously Injured (KSI) dataset containing all serious or fatal cyclist accidents over a 10-year period between 2008 to 2018. The dataset includes 1341 observations, 60 variables and is distinct at the accident participant level. The dataset mainly contains categorical variables. We joined an hourly Toronto weather dataset derived from data available through Environment and Climate Change Canada to add additional weather variables to the accident dataset. The tables were joined at the date and hour level.

Data Cleaning & Joining

Conversion of missing data to binary key

- Initial assessment of data indicated many missing data points around variables describing participant actions/conditions, secondary location variables, and incident flags. These missing values related to other variables within the dataset meaning that the absence of a value had meaning. These were transformed into a binary key to maintain information value.

Redundant variables dropped from data frame

- Many redundant variables such as multiple coordinate values and variables with more than 99.5% of observations missing were removed from the dataset. Variables not utilized for modelling or visualization were also dropped.

Feature transformations

- Date variables in timestamp format were converted into short date format while time was rounded to the nearest hour. This transformation was mainly done to simply visualizations but also create two anchor variables to join against the hourly weather dataset.

Joining hourly weather data

- The hourly weather date and time variables were converted into short date and rounded hour to match the cleaned cyclist dataset. A left join was performed on the weather data to the cyclist data on the date and hour variables.

Technique 1 to 3 – Data Acquisition and Cleaning

Accident Level Dataset: Created to facilitate our regression modelling and to remove the high number of duplicate data values attributed to the participant level data.

Isolated accident level data

- Accident specific variables (location, type, time and severity variables) were isolated into a new data frame to compile unique observations at the accident level. The resultant dataset was 569 unique observations.

Aggregation of high-level participant data

- Participant data around vehicle types, participant general categories and police incident flags (traffic violation, DUI, etc.) were aggregated into the accident level by taking the maximum binary value for each flag. This consolidated these binary flags at the accident level with no loss of information.

Feature Engineering for Modelling Variables

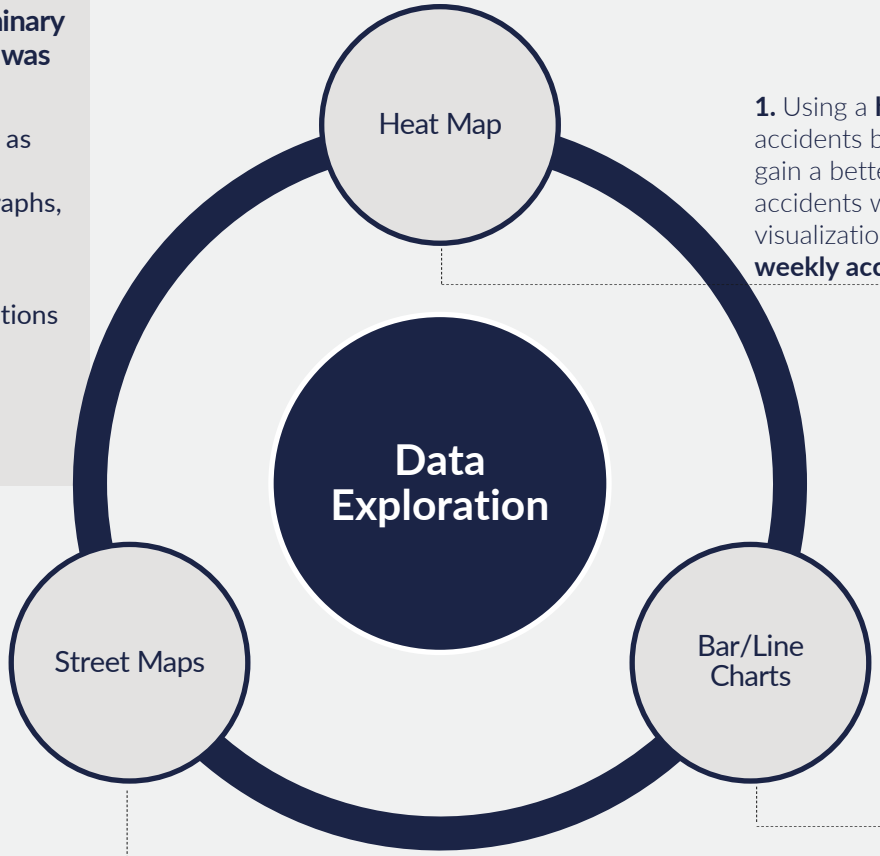
- New features were created from aggregating observations from the participant level actions and conditions for driver, pedestrian and cyclist. These new variables allowed us to model what groups of actions the participant were performing or what conditions the participants were under at the time of the accident. As our analysis focused on cyclist, we created additional action and condition variables to focus our modelling.
- Ward variables were created by delineating the string of numbers from the ward fields at the accident level. These variables were turned into binary flags per ward and weighted against the other wards when an accident occurred in more than one ward. This was done to try to remove collinearity between ward variables.
- Participants were separated into groups (driver, pedestrian, and cyclist) to create participant counts at the accident level
- The participants age categories were converted into ranked numeric values and then summarized into the accident level data, separated by participant type. As our analysis was specifically focused on cyclist, only a cyclist age feature was created.

Technique 4 – Data Exploration

The Analysis Team provided the preliminary insights to all Teams. Data Exploration was guided using the business questions.

To understand the landscape and answer as many business question, we developed different charts such as bar charts, line graphs, and heat maps were created.

Insight was provided across the teams to ensure team understanding. Certain questions were drilled-down further based on team feedback.



1. Using a **heat map** we were able to map total accidents by **day of week** and **time of day** to gain a better understanding of specifically when accidents were occurring. This type of visualization allowed us to gain insight into **weekly accident patterns**.

2. Bar charts and **line charts** were used to plot trends, create categorical breakouts, and build histograms. **Accident trend graphs**, **accident type breakouts**, and **age histograms** were created to provide a **fundamental understanding** of the situation before diving deeper into Advanced Analytics.

3. Street maps allowed us to visualize where **major and fatal accidents** are occurring. This allowed us to make the inference that **most fatal accidents are happening at or near major intersections**. The street map also allows us to determine **where to deploy preventative campaigns**.

Technique 5 to 8 – Regression Modeling Overview

Modeling Process – 4 phases

1. Data Sampling – creating a training and validation dataset
2. Regression Modelling – Methodology and model selection
3. Model Testing – Validation Results
4. Coefficient Interpretation – Interpretation of the results to generate insights

Technique 5 to 8 cont'd – Data Sampling

Data Sampling

Selected variables that we wanted to test in the regression model

- Used a combination of existing knowledge and referencing factors listed in Vision Zero report

Sample 70% of the dataset using the Major/Fatal incidents as a marker

- Although we had panel data we sampled across all years because Major/Fatal incidents were relatively constant YoY
- The remaining 30% of the dataset was saved as a test sample to validate our final regression model

Class Imbalance Issue – Resolution: Up sampling

- Our dataset and sample were highly imbalanced almost 95% of observations were 1's because most observations were Major/Fatal
- We used an up sampling technique (part of the 'caret' package in r) to create a more balanced dataset to train our model (Do to the
- limited number of 0's minor accidents we chose to up sample rather than down sampling)
- This arbitrarily will increase our models reported "fit" something we considered as we fitted our regression model

Technique 5 to 8 cont'd – Regression Model

Building Regression Model

Selected variables from existing dataset that we wanted to test in the model

- Used a combination of existing knowledge and referencing factors listed in Vision Zero report variables are a subset of variables included in Police Cyclist Accident Report and Weather Report

Logistic Regression Modelling

- Our dependent variable was binary (Accident was classified as Major or Fatal)
- We chose to use a binomial logistic regression to build our model (as linear regression would not have given us the best fit)
- Model is therefore predicting the probability of an accident in Toronto from 2008-2018 being Major or Fatal to a bicyclist

Test Test Test Methodology*

- Selected relevant variables and ran logistic binomial regression
- Reviewed variable t-tests (p-values) and eliminated variables that did not pass hypothesis tests
- Conducted joint F-tests to remove variables with minimal significance and high correlations
- Used combination of AIC score, McFadden Pseudo R-squared and confusion matrix to evaluate model fit

Technique 5 to 8 cont'd – Regression Model

Building Regression Model

Final Equation

BIC_INJURY_MAJOR_FATAL =

Variable	Coefficient
Intercept	+23.57
RUSHHOUR_MOR	– 1.05
RUSHHOUR_NI	+ 1.55
UNDER44	–1.83
DRIV_AGGRO	+2.43
DRIV_TRAFFIC_VIOLATION	+ 1.55
DRIV_DISTRACT	1.53
DRIV_ALCOHOL	+ 21.45
PED_CROSSING	– 0.68
BIC_TRAFFIC_VIOLATION	– 0.80
BIC_AGGRO	+ 0.72
BIC_SPEED	– 7.05
BIC_LOSS_CTRL	+ 1.09
BIC_ALCOHOL	– 2.01
BIC_DISTRACT	+ 0.65
VEH_TURNING	+ 0.28
VEH_STATIONARY	+ 1.91

Variable	Coefficient
VEH_LANE_CHG	– 2.81
BIC_ACTN_FORWARD	– 1.45
BIC_ACTN_TURNING	– 1.93
wind_speed	– 0.11
temperature	– 0.01
ROAD_WET	+ 2.81
VISIBILITY_VALUE	+ 0.0001
PPL_CNT	– 1.19
WARD_1	– 24.17
WARD_2	+14.98
WARD_3	– 17.63
WARD_4	+ 0.99
WARD_5	+ 2.92
WARD_6	– 0.19
WARD_7	– 22.05
WARD_8	+ 15.64
WARD_9	– 16.6

Variable	Coefficient
WARD_10	– 20.6
WARD_11	– 19.21
WARD_12	– 21.16
WARD_13	– 14.65
WARD_14	– 19.85
WARD_15	+ 12.39
WARD_16	– 17.83
WARD_17	– 16.78
WARD_18	– 0.23
WARD_19	+ 15.89
WARD_20	+ 2.87
WARD_21	– 17.13
WARD_22	– 20.31
WARD_23	– 0.18
WARD_24	+ 2.19

Technique 5 to 8 cont'd – Regression Variable Discussion

Selecting Coefficients/Variables

Wards 1-25

- Although none of the wards came out significant on their own when Joint Hypothesis tests were conducted their inclusion jointly contributed to the overall fit of the model and were kept
- There were 25 wards total, dropped Ward 25 to avoid perfect collinearity

Weather Variables

- This includes temperature, ROAD_WET, visibility_value
- Originally we had included relative humidity but we found the correlation between the two variables to be high and removing the humidity variable did not impact the model significantly

PED_, DRIV_/VEH_ & BIC_ Variables

- After realizing that many of the turning and action variables were not significant on their own we tried to combine them more (ie. Right, left and U turns are all classified as a turn)
- This improved the model slightly (lower # of variables) but there is another underlying factor that makes these variables significant to the model

```
Coefficients:

(Intercept)
RUSHHOUR_MOR
RUSHHOUR_NI
UNDER44
DRIV_TRAFFIC_VIOLATION
DRIV_AGGRO
DRIV_DISTRACT
DRIV_ALCOHOL
PED_CROSSING
BIC_TRAFFIC_VIOLATION
BIC_AGGRO
BIC_SPEED
BIC_LOSS_CTRL
```

```
BIC_ALCOHOL
BIC_DISTRACT
VEH_TURNING
VEH_STATIONARY
VEH_LANE_CHG
BIC_ACTN_FORWARD
BIC_ACTN_TURNING
wind_speed
temperature
ROAD_WET
visibility_value
PPL_CNT
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1059.13  on 763  degrees of freedom
Residual deviance:  454.47  on 715  degrees of freedom
AIC: 552.47

Number of Fisher Scoring iterations: 18

> PseudoR2(final_regr)
 McFadden
0.5709006
```

Technique 5 to 8 cont'd – Model Testing

Model Testing

Use validation data to test predictions

- Use regression equation
- Input values for independent variables using the validation dataset
- Predict dependent variable or parameter

```
> pred_mod <- predict(final_regr, newdata = test_da
ta, type = "response")
>
> y_pred_num <- ifelse(pred_mod > 0.5, 1, 0)
> y_pred <- factor(y_pred_num, levels=c(0, 1))
> y_act <- test_data$BIC_INJURY_MAJOR_FATAL
>
> mean(y_pred_num)
[1] 0.8529412
> mean(y_act)
[1] 0.9764706
> mean(y_pred_num == y_act) #Resulting accuracy
[1] 0.8647059
```

Compare results to actuals (from validation dataset)

- Model equation spits out probabilities so have to convert probs into 0,1's to compare predicted to actuals
- Used p(Major_Fatal)=0.5 as threshold, where p>0.5 is a 1, p<=0.5 is a 0.
- Compared results from model prediction to actuals (from validation dataset)
- Calculated accuracy using confusion matrix as well.

Confusion Matrix

```
> pred_mod <- predict(final_regr, newdata = test_data, type = "response")
> table(test_data$BIC_INJURY_MAJOR_FATAL, pred_mod>0.5)

   FALSE TRUE
  0     3    1
  1    22  144
> (3+144)/(3+22+1+144)
[1] 0.8647059
```

Technique 5 to 8 cont'd – Coefficient Interpretation

Interpreting Coefficients/Variables

1. Referencing the Correlation Matrix

- After looking at descriptive analytics insights was able to refer to correlation matrix to identify correlations between variables of interest and other relevant variables
- Variables with large correlations

2. Calculating the Odds Ratio

- Using the formula exp(coef(regression)) we were able to convert the coefficients of our model into an odds ratio

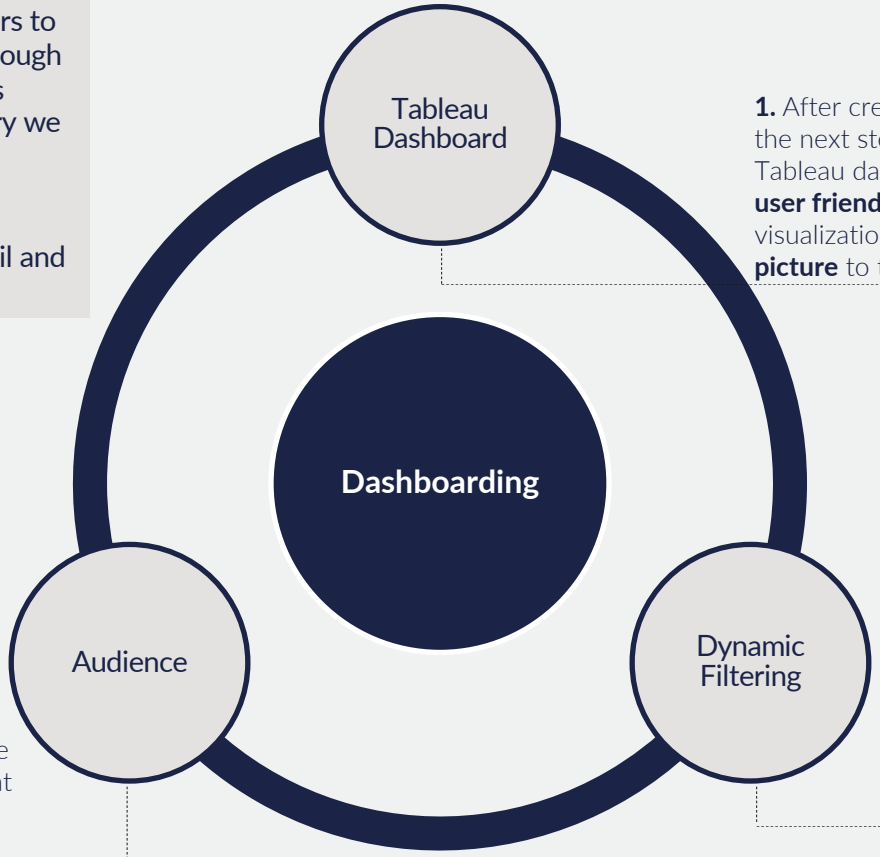
3. Calculating the Probability

- Using the formula odds/(1+odds) we were able to convert the coefficients of our model into probabilities and extrapolate

Technique 9 – Dashboarding

The **Tableau dashboard** allows end-users to curate their own controlled insights through dynamic filtering. Specific visualizations were selected to curate the overall story we wanted to tell.

This dashboard was intended to be consumed by both Toronto City Council and the average Torontonian Citizen.



1. After creating the different data visualizations, the next step was to **incorporate** them into a Tableau dashboard. The dashboard needs to be **user friendly**, **engaging**, and **clear**. The right visualizations need to be included to **paint a picture** to tell a linear story.

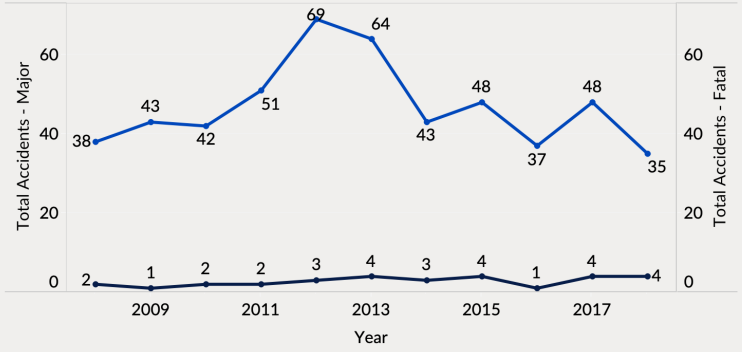
2. In order create a **dynamic** dashboard, we enabled each visualization to act as a filter, in addition to adding traditional filters. This allows the end-user to **point and click** on any element to get a **specific view** of the element they want to explore.

3. Audience is an important part to consider when creating a dashboard. We wanted to provide **top-level insights** that both the average **Citizen** and **City Council** could understand if provided without context.

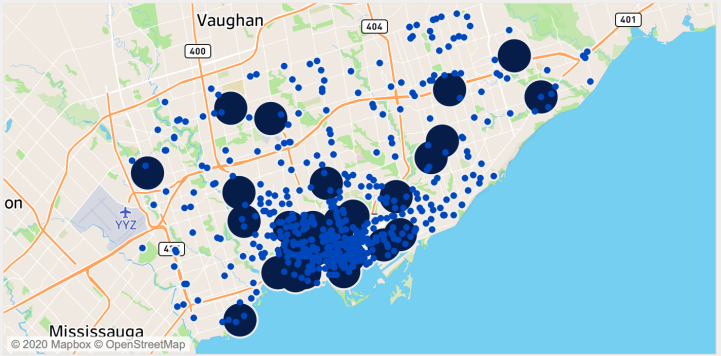
Technique 9 – Dashboarding

Toronto Killed or Seriously Injured Bicyclists Dashboard

10-Year Trend of Total Accidents



Accidents in the Greater Toronto Area



District
All

Neighbourhood
All

Year
All

Month
All

Hour
All

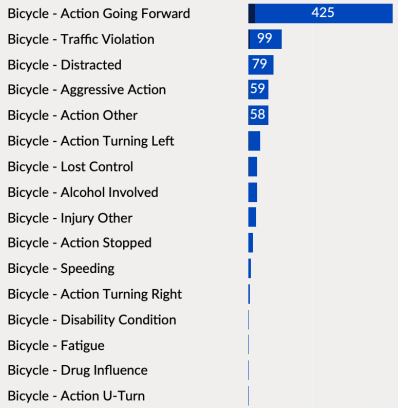
Ward Number
All

Total Accidents by Day/Hour

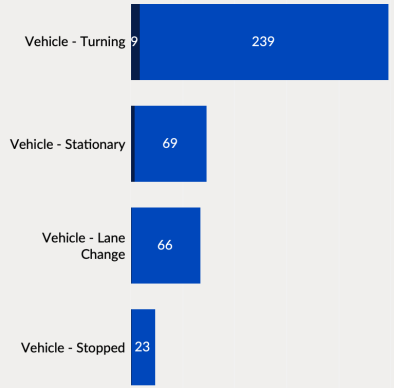
Hour	Mond..	Tuesd..	Wed..	Thurs..	Friday	Satur..	Sunda..	Gran..
0	2		1	1	2	2	3	11
1		1	1	1	1	2	2	8
2				1	1	2	1	5
3					2		2	4
4							1	1
5			2		2	1	1	6
6	2	1	3	3	3		1	13
7	4	4	6	2	3			19
8	3	9	7	6	5	1		31
9	1	7	6	7	4	1	1	27
10	12	4	4	3	6	1	1	31
11	3	4	4	1	3	3	1	19
12	2	7		4	1	6	6	26
13	6	3	3	9	4	2	2	29
14	6	5	4	9	3	5	6	38
15	3	11	3	8	6	4	3	38
16	3	5	3	8	8	3	3	33
17	6	10	10	6	7	4	3	46
18	3	6	11	5	7	3	10	45
19	7	4	6	8	2	2	4	33
20	5	1	5	4	1	4	2	22
21	2	6	2	10	5	3	3	31
22	1		4	3	4	2	1	15
23	2	2	2	2	1	4	4	17
Grand..	73	90	87	101	81	55	61	548

Major Fatal Major Fatal

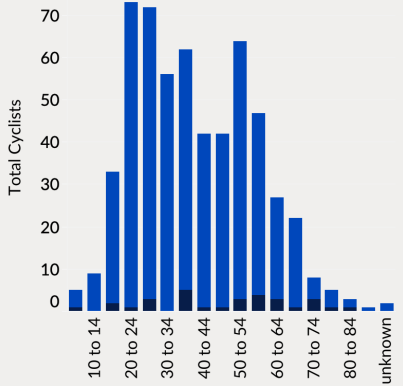
Number of Accidents Involving a Bicyclist



Number of Accidents Involving a Vehicle



Age Distribution of Accidents



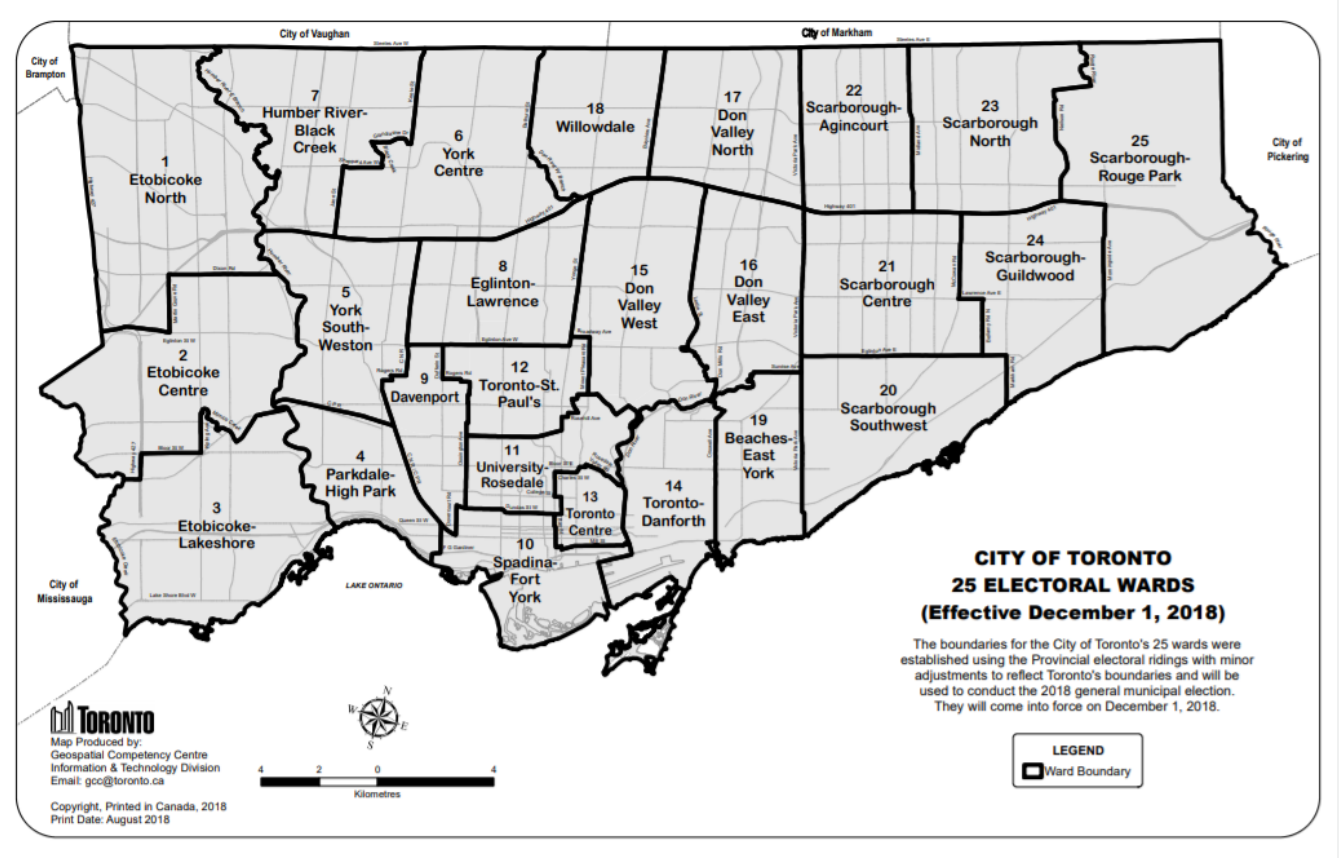
Limitations and Assumptions

- The dataset is specific to only reported cyclist accidents in Toronto for the ten-year period. We are limited to the data provided and assume the same trend going forward.
- The data was reported by Police Officers on the scene however there were incomplete observations (as not all details were filled out) these observations or variables were dropped (e.g. Cyclist_Type).
- Our dataset was truncated, the majority of observations were major and fatal accidents. We assume this is because people are less likely to report minor accidents so our insights could only focus on more severe accidents.
- The predictive power of our model was limited by our small sample size.

Appendix

A. **Accidents Interactive Dashboard (Tableau)** – NewYork_Data Viz_Final.twbx.
B. **Toronto City Ward Map** (Slide 28)
C. **R Script Data Cleaning and Regression** – GMMA860_R_Script_Team_NewYork.txt
D. **Accident Dataset** – Accident_level_summary_v3.csv
E. **Data Dictionary** – NewYork_Cyclist_Data_Dictionnary.xls
F. **Regression Model** (Slide 30 to 38) – Supporting Documents
G. **Business Questions** (Slide 39)

Appendix B – Toronto Ward Map



City of Toronto. [2020] 2020 Detailed Ward Maps. Retrieved from https://www.toronto.ca/city-government/elections/city-elections/ward-maps/

Appendix F1 - Regression Model Output A

Model A With Dummy Variable for Each Individual Ward

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.357e+01	3.249e+03	0.007	0.994212	
RUSHHOUR_MOR	-1.050e+00	5.393e-01	-1.946	0.051630	.
RUSHHOUR_NI	1.549e+00	5.563e-01	2.785	0.005351	**
UNDER44	-1.832e+00	4.456e-01	-4.113	3.91e-05	***
DRIV_TRAFFIC_VIOLATION	2.519e+00	6.105e-01	4.126	3.69e-05	***
DRIV_AGGRO	2.431e+00	6.767e-01	3.593	0.000327	***
DRIV_DISTRACT	1.530e+00	5.553e-01	2.755	0.005862	**
DRIV_ALCOHOL	2.145e+01	3.474e+03	0.006	0.995073	
PED_CROSSING	-6.819e-01	1.452e+00	-0.469	0.638736	
BIC_TRAFFIC_VIOLATION	-8.033e-01	7.325e-01	-1.097	0.272759	
BIC_AGGRO	7.228e-01	7.679e-01	0.941	0.346605	
BIC_SPEED	-7.049e+00	1.586e+00	-4.445	8.80e-06	***
BIC_LOSS_CTRL	1.091e+00	8.450e-01	1.291	0.196653	
BIC_ALCOHOL	-2.007e+00	8.294e-01	-2.420	0.015505	*
BIC_DISTRACT	6.500e-01	5.798e-01	1.121	0.262253	
VEH_TURNING	2.784e-01	5.597e-01	0.497	0.618907	
VEH_STATIONARY	1.908e+00	7.111e-01	2.683	0.007287	**
VEH_LANE_CHG	-2.809e+00	7.080e-01	-3.968	7.26e-05	***
BIC_ACTN_FORWARD	-1.455e+00	6.283e-01	-2.315	0.020602	*
BIC_ACTN_TURNING	-1.931e+00	1.022e+00	-1.890	0.058726	.
wind_speed	-1.129e-01	2.552e-02	-4.425	9.65e-06	***
temperature	-6.498e-03	1.992e-02	-0.326	0.744213	
ROAD_WET	2.813e+00	9.649e-01	2.916	0.003547	**
visibility_value	9.438e-05	4.352e-05	2.169	0.030096	*
PPL_CNT	-1.186e+00	3.236e-01	-3.665	0.000248	***

WARD_1	-2.417e+01	3.249e+03	-0.007	0.994066
WARD_2	1.498e+01	5.182e+03	0.003	0.997693
WARD_3	-1.763e+01	3.249e+03	-0.005	0.995672
WARD_4	9.932e-01	3.792e+03	0.000	0.999791
WARD_5	2.915e+00	4.334e+03	0.001	0.999463
WARD_6	-1.863e-01	4.323e+03	0.000	0.999966
WARD_7	-2.205e+01	3.249e+03	-0.007	0.994585
WARD_8	1.564e+01	4.925e+03	0.003	0.997466
WARD_9	-1.660e+01	3.249e+03	-0.005	0.995924
WARD_10	-2.060e+01	3.249e+03	-0.006	0.994943
WARD_11	-1.921e+01	3.249e+03	-0.006	0.995282
WARD_12	-2.116e+01	3.249e+03	-0.007	0.994804
WARD_13	-1.465e+01	3.249e+03	-0.005	0.996404
WARD_14	-1.986e+01	3.249e+03	-0.006	0.995124
WARD_15	1.239e+01	5.376e+03	0.002	0.998162
WARD_16	-1.783e+01	3.249e+03	-0.005	0.995621
WARD_17	-1.678e+01	3.249e+03	-0.005	0.995879
WARD_18	-2.306e-01	5.162e+03	0.000	0.999964
WARD_19	1.589e+01	4.085e+03	0.004	0.996896
WARD_20	2.869e+00	4.061e+03	0.001	0.999436
WARD_21	-1.713e+01	3.249e+03	-0.005	0.995793
WARD_22	-2.031e+01	3.249e+03	-0.006	0.995013
WARD_23	-1.777e-01	4.514e+03	0.000	0.999969
WARD_24	2.196e+00	5.474e+03	0.000	0.999680

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Appendix F1 - Regression Model Output B

Model B With Dummy Variable for Suburbs

```
Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              3.956e+00  1.055e+00   3.751 0.000176 ***
RUSHHOUR_MOR            -1.219e+00  3.381e-01  -3.606 0.000311 ***
RUSHHOUR_NI              7.053e-01  2.891e-01   2.440 0.014689 *
UNDER44                 -1.814e+00  2.750e-01  -6.599 4.15e-11 ***
DRIV_TRAFFIC_VIOLATION   1.328e+00  3.403e-01   3.901 9.59e-05 ***
DRIV_AGGRO               2.132e+00  4.328e-01   4.925 8.42e-07 ***
DRIV_DISTRACT            5.160e-01  3.310e-01   1.559 0.119078
DRIV_ALCOHOL             1.577e+01  5.552e+02   0.028 0.977335
PED_CROSSING            -2.445e+00  1.217e+00  -2.009 0.044575 *
BIC_TRAFFIC_VIOLATION    9.338e-02  3.734e-01   0.250 0.802522
BIC_AGGRO                7.704e-01  4.996e-01   1.542 0.123039
BIC_SPEED               -4.000e+00  9.344e-01  -4.281 1.86e-05 ***
BIC_LOSS_CTRL            1.092e-01  4.616e-01   0.237 0.812964
BIC_ALCOHOL             -1.591e-01  4.769e-01  -0.334 0.738596
BIC_DISTRACT             3.524e-01  3.534e-01   0.997 0.318769
VEH_TURNING              8.464e-01  3.181e-01   2.660 0.007806 **
VEH_STATIONARY           2.219e+00  4.687e-01   4.735 2.19e-06 ***
VEH_LANE_CHG            -1.449e+00  4.146e-01  -3.495 0.000474 ***
BIC_ACTN_FORWARD        -1.394e+00  4.064e-01  -3.431 0.000601 ***
BIC_ACTN_TURNING        -1.267e+00  6.057e-01  -2.091 0.036515 *
wind_speed              -5.499e-02  1.358e-02  -4.050 5.13e-05 ***
temperature             -9.527e-03  1.143e-02  -0.833 0.404668
ROAD_WET                 2.425e+00  6.391e-01   3.794 0.000148 ***
visibility_value         5.404e-05  2.822e-05   1.915 0.055536 .
PPL_CNT                 -1.210e+00  2.124e-01  -5.697 1.22e-08 ***
BURBS                    7.629e-02  2.198e-01   0.347 0.728498
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix F2 – Regression Model Variable Rationale

WARDS 1-25

As mentioned previously, there were 25 wards. Ward 25 was dropped so that there wasn't perfect collinearity between the variables. The wards individually were not significant but jointly they were. We tested changing the location variable to include only the downtown wards (wards 4 & 9-14) and only the suburb wards.

– 24.17WARD_1 +14.98WARD_2
– 17.63WARD_3 + 0.99WARD_4
+ 2.92WARD_5 – 0.19WARD_6
– 22.05WARD_7 + 15.64WARD_8
– 16.6WARD_9 – 20.6WARD_10
– 19.21WARD_11 – 21.16WARD_12
– 14.65WARD_13 – 19.85WARD_14
+ 12.39WARD_15 – 17.83WARD_16
– 16.78WARD_17 – 0.23WARD_18
+ 15.89WARD_19 + 2.87WARD_20
– 17.13WARD_21 – 20.31WARD_22
– 0.18WARD_23 + 2.19WARD_24

Pedst., Cyclist, Driver State of Mind

State of mind was evaluated through variables such as aggression, traffic violation, distraction and alcohol. Overall driver state of mind was more significant than bicyclist or pedestrian though all were jointly significant.

+ 2.43DRIV_AGGRO + 1.53DRIV_DISTRACT + 21.45DRIV_ALCOHOL – 0.68PED_CROSSING + 0.72BIC_AGGRO – 2.01BIC_ALCOHOL + 0.65BIC_DISTRACT

Weather Variables

Wind Speed, road wet and visibility value all were significant in our model. Visibility value and road wet were highly correlated but jointly significant so they were retained in the model. Temperature was included as a proxy for the number of bicyclist on the road (only a few months of the year in Toronto where cyclists can ride).

– 0.11wind_speed – 0.01temperature + 2.81ROAD_WET + 0.0001VISIBILITY_VALUE

Bicycle and Vehicle Actions

Variables for turning (right, left, u), stationary, lane change and forward actions were initially included. Turning variable was combined as individual types of turns did not result in significance. The remaining actions are significant in predicting KSIs.

+ 2.52DRIV_TRAFFIC_VIOLATION
+ 0.28VEH_TURNING +1.91VEH_STATIONARY – 2.81VEH_LANE_CHG –1.45BIC_ACTN_FORWARD – 1.93BIC_ACTN_TURNING – 7.05**BIC_SPEED**
– 0.80BIC_TRAFFIC_VIOLATION+ 1.09**BIC_LOSS_CTRL**

Other Variables and Intercept

Time of day, age and number of parties involved in accident were also included in the model. All reported relatively high significance in our model indicating that their inclusion benefited the overall fit.

23.57 – 1.05RUSHHOUR_MOR + 1.55RUSHHOUR_NI – 1.83UNDER44– 1.19PPL_CNT

Appendix F3 – Regression Coefficient Odds Ratio

Model A With Dummy Variable for Each Individual Ward

```
> format(odd, scientific = F) #Print out Odds Ratios
                (Intercept)                  RUSHHOUR_MOR                    RUSHHOUR_NI                       UNDER44          DRIV_TRAFFIC_VIOLATION
"17243803970.35390090942382812" "          0.35006189984948749" "          4.70882198936571150" "          0.16002171901057421" "          12.41566954582550686"
                 DRIV_AGGRO                DRIV_DISTRACT                   DRIV_ALCOHOL                    PED_CROSSING          BIC_TRAFFIC_VIOLATION
"         11.37588013435072476" "          4.61853639239157321" " 2071981391.63888239860534668" "          0.50567597528890273" "          0.44782786638182076"
                 BIC_AGGRO                    BIC_SPEED                   BIC_LOSS_CTRL                   BIC_ALCOHOL                   BIC_DISTRACT
"          2.06010120306840827" "          0.00086838826003484" "          2.97740237773989636" "          0.13433649704984096" "          1.91561558054005499"
               VEH_TURNING                VEH_STATIONARY                   VEH_LANE_CHG               BIC_ACTN_FORWARD              BIC_ACTN_TURNING
"          1.32104049368336751" "          6.74153314258031511" "          0.06025032998136451" "          0.23348941409319024" "          0.14494726563944296"
               wind_speed                   temperature                     ROAD_WET                visibility_value                       PPL_CNT
"          0.89324002627579246" "          0.99352306162982840" "         16.66661478066583868" "          1.00009438425664743" "          0.30545097182334302"
                   WARD_1                       WARD_2                       WARD_3                       WARD_4                       WARD_5
"          0.00000000003198686" "    3209470.80125023890286684" "          0.00000002213970923" "          2.69984574892281293" "         18.45699068701010859"
                   WARD_6                       WARD_7                       WARD_8                       WARD_9                      WARD_10
"          0.83005468527336834" "          0.00000000026457864" "    6193015.30741257406771183" "          0.00000006181223319" "          0.00000000113608644"
                  WARD_11                      WARD_12                      WARD_13                      WARD_14                      WARD_15
"          0.00000000453212141" "          0.00000000064698080" "          0.00000043624785706" "          0.00000000237943315" "     239756.58786945202155039"
                  WARD_16                      WARD_17                      WARD_18                      WARD_19                      WARD_20
"          0.00000001800093135" "          0.00000005147431245" "          0.79403595446584885" "    7968263.05019460618495941" "         17.62344983776996798"
                  WARD_21                      WARD_22                      WARD_23                      WARD_24
"          0.00000003631209610" "          0.00000000151085293" "          0.83722618211958466" "          8.98489621531819438"
>
```

Model B With Dummy Variable for if part of Ward

```
> odd<-exp(coef(test_wards))
> format(odd, scientific = F)
      (Intercept)          RUSHHOUR_MOR            RUSHHOUR_NI               UNDER44 DRIV_TRAFFIC_VIOLATION           DRIV_AGGRO         DRIV_DISTRACT
"     52.24133665" "      0.29542496" "      2.02439363" "      0.16292204" "      3.77161607" "      8.42754414" "      1.67528790"
     DRIV_ALCOHOL          PED_CROSSING BIC_TRAFFIC_VIOLATION           BIC_AGGRO           BIC_SPEED       BIC_LOSS_CTRL         BIC_ALCOHOL
"7088667.48363981" "      0.08676803" "      1.09788008" "      2.16061864" "      0.01831145" "      1.11540478" "      0.85288415"
     BIC_DISTRACT          VEH_TURNING       VEH_STATIONARY          VEH_LANE_CHG     BIC_ACTN_FORWARD    BIC_ACTN_TURNING          wind_speed
"      1.42240761" "      2.33116355" "      9.19910636" "      0.23479018" "      0.24796856" "      0.28181405" "      0.94649262"
      temperature             ROAD_WET      visibility_value              PPL_CNT                BURBS
"      0.99051801" "     11.29900564" "      1.00005404" "      0.29812896" "      1.07927875"
```

Odds ratio calculation

Odds<- exp(coef(ModelA))

Appendix F3 – Regression Coefficient Probabilities

Model A With Dummy Variable for Each Individual Ward

```
prob <- (odd / (1 + odd))
format(prob,scientific = F) #Print out Probabilities
```

(Intercept)	RUSHHOUR_MOR	RUSHHOUR_NI	UNDER44	DRIV_TRAFFIC_VIOLATION	DRIV_AGGRO	DRIV_DISTRACT
"0.99999999994200817"	"0.25929322195412990"	"0.82483251328158746"	"0.13794717494346806"	"0.92546029875108504"	"0.91919766601290998"	"0.82201770529524998"
DRIV_ALCOHOL	PED_CROSSING	BIC_TRAFFIC_VIOLATION	BIC_AGGRO	BIC_SPEED	BIC_LOSS_CTRL	BIC_ALCOHOL
"0.99999999951737018"	"0.33584647931429989"	"0.30931015818956459"	"0.67321342215829816"	"0.00086763481614650"	"0.74857962433053205"	"0.11842737794227776"
BIC_DISTRACT	VEH_TURNING	VEH_STATIONARY	VEH_LANE_CHG	BIC_ACTN_FORWARD	BIC_ACTN_TURNING	wind_speed
"0.65701925635382574"	"0.56915874465720617"	"0.87082662031119429"	"0.05682651377474553"	"0.18929178590871157"	"0.12659732896823958"	"0.47180495546192946"
temperature	ROAD_WET	visibility_value	PPL_CNT	WARD_1	WARD_2	WARD_3
"0.49837550452893270"	"0.94339606017252442"	"0.50002359495066584"	"0.23398118996128597"	"0.00000000003198686"	"0.99999968842225073"	"0.00000002213970874"
WARD_4	WARD_5	WARD_6	WARD_7	WARD_8	WARD_9	WARD_10
"0.72971846183286326"	"0.94860459070540537"	"0.45356824140442403"	"0.00000000026457864"	"0.99999983852779484"	"0.00000006181222937"	"0.00000000113608644"
WARD_11	WARD_12	WARD_13	WARD_14	WARD_15	WARD_16	WARD_17
"0.00000000453212139"	"0.00000000064698080"	"0.00000043624766674"	"0.00000000237943315"	"0.99999582912053431"	"0.00000001800093102"	"0.00000005147430980"
WARD_18	WARD_19	WARD_20	WARD_21	WARD_22	WARD_23	WARD_24
"0.44259757029354679"	"0.99999987450215078"	"0.94630425572538590"	"0.00000003631209478"	"0.00000000151085293"	"0.45570120340528103"	"0.89984873368379503"

Model B With Dummy Variable for Suburbs

```
> prob <- (odd / (1 + odd))
> format(prob,scientific = F) #Probability Table
```

(Intercept)	RUSHHOUR_MOR	RUSHHOUR_NI	UNDER44	DRIV_TRAFFIC_VIOLATION	DRIV_AGGRO	DRIV_DISTRACT
"0.98121760"	"0.22805255"	"0.66935521"	"0.14009713"	"0.79042740"	"0.89392784"	"0.62620845"
DRIV_ALCOHOL	PED_CROSSING	BIC_TRAFFIC_VIOLATION	BIC_AGGRO	BIC_SPEED	BIC_LOSS_CTRL	BIC_ALCOHOL
"0.99999986"	"0.07984043"	"0.52332833"	"0.68360624"	"0.01798217"	"0.52727723"	"0.46030085"
BIC_DISTRACT	VEH_TURNING	VEH_STATIONARY	VEH_LANE_CHG	BIC_ACTN_FORWARD	BIC_ACTN_TURNING	wind_speed
"0.58718756"	"0.69980459"	"0.90195219"	"0.19014581"	"0.19869776"	"0.21985564"	"0.48625544"
temperature	ROAD_WET	visibility_value	PPL_CNT	BURBS		
"0.49761821"	"0.91869261"	"0.50001351"	"0.22966051"	"0.51906400"		

Probability Odds Ratio Calculation
prob<- odd / (1+odd)

Appendix F4 – Regression Coefficient Probabilities for Age

Model B With Dummy Variable for if part of Ward

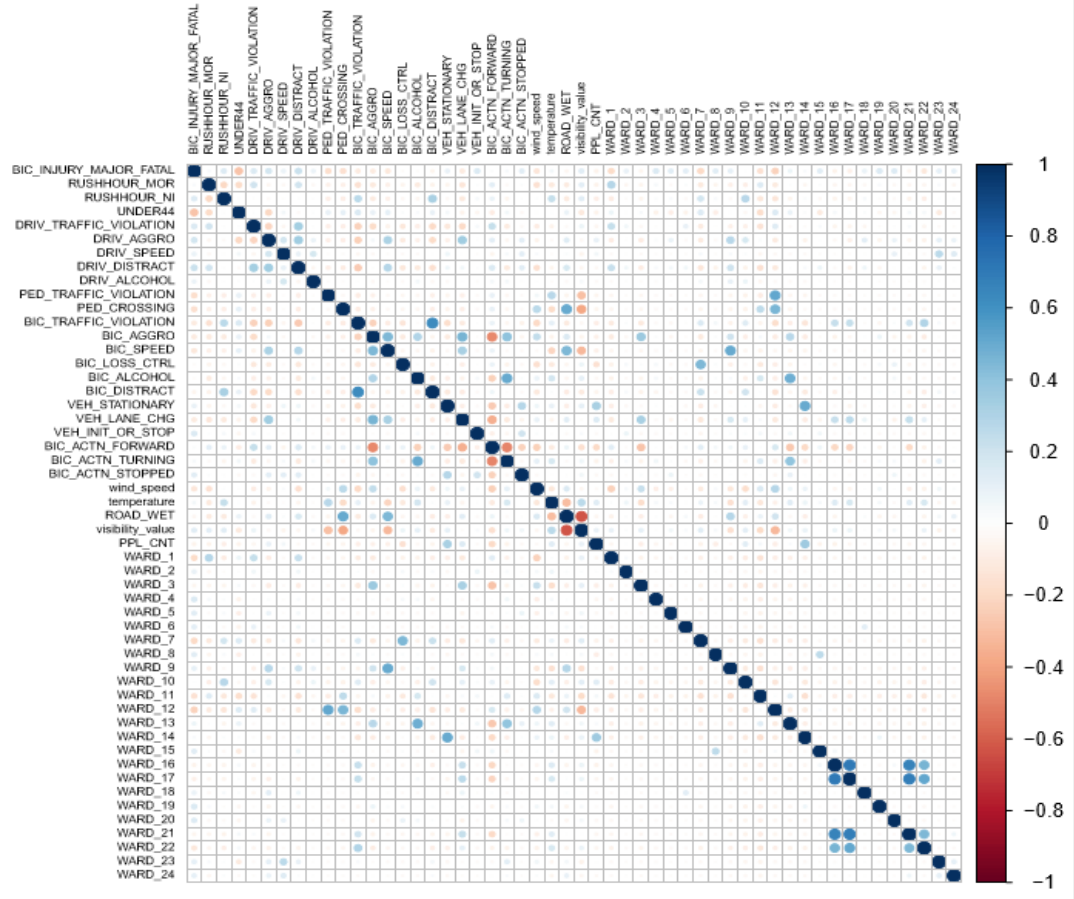
(Intercept)	RUSHHOUR_MOR	RUSHHOUR_NI	AGE20S	DRIV_TRAFFIC_VIOLATION	DRIV_AGGRO
1.0000	0.3178	0.7828	0.4822	0.8609	0.8948
DRIV_DISTRACT	DRIV_ALCOHOL	PED_TRAFFIC_VIOLATION	PED_CROSSING	BIC_TRAFFIC_VIOLATION	BIC_AGGRO
0.7713	1.0000	0.0000	0.1398	0.1606	0.7036
BIC_SPEED	BIC_LOSS_CTRL	BIC_ALCOHOL	BIC_DISTRACT	BIC_DRUGS	BIC_DISABILITY
0.0001	0.4476	0.0508	0.6734	0.0600	1.0000
BIC_FATIGUE	VEH_TURNING	VEH_STATIONARY	VEH_LANE_CHG	VEH_INIT_OR_STOP	BIC_ACTN_FORWARD
1.0000	0.6661	0.8714	0.1238	1.0000	0.4892
BIC_ACTN_TURNING	BIC_ACTN_STOPPED	wind_speed	temperature	ROAD_WET	visibility_value
0.2551	1.0000	0.4676	0.4997	0.9704	0.5000
PPL_CNT	WARD_1	WARD_2	WARD_3	WARD_4	WARD_5
0.2368	0.0000	1.0000	0.0000	0.9553	0.9996
WARD_6	WARD_7	WARD_8	WARD_9	WARD_10	WARD_11
0.4899	0.0000	1.0000	0.0000	0.0000	0.0000
WARD_12	WARD_13	WARD_14	WARD_15	WARD_16	WARD_17
0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
WARD_18	WARD_19	WARD_20	WARD_21	WARD_22	WARD_23
0.2600	1.0000	0.9975	0.0000	0.0000	0.6813
WARD_24					
0.9958					

Probability Odds Ratio Calculation
prob<- odd / (1+odd)

Appendix F5: Correlation Matrix A

This is the correlation matrix for every variable that was considered for our model.

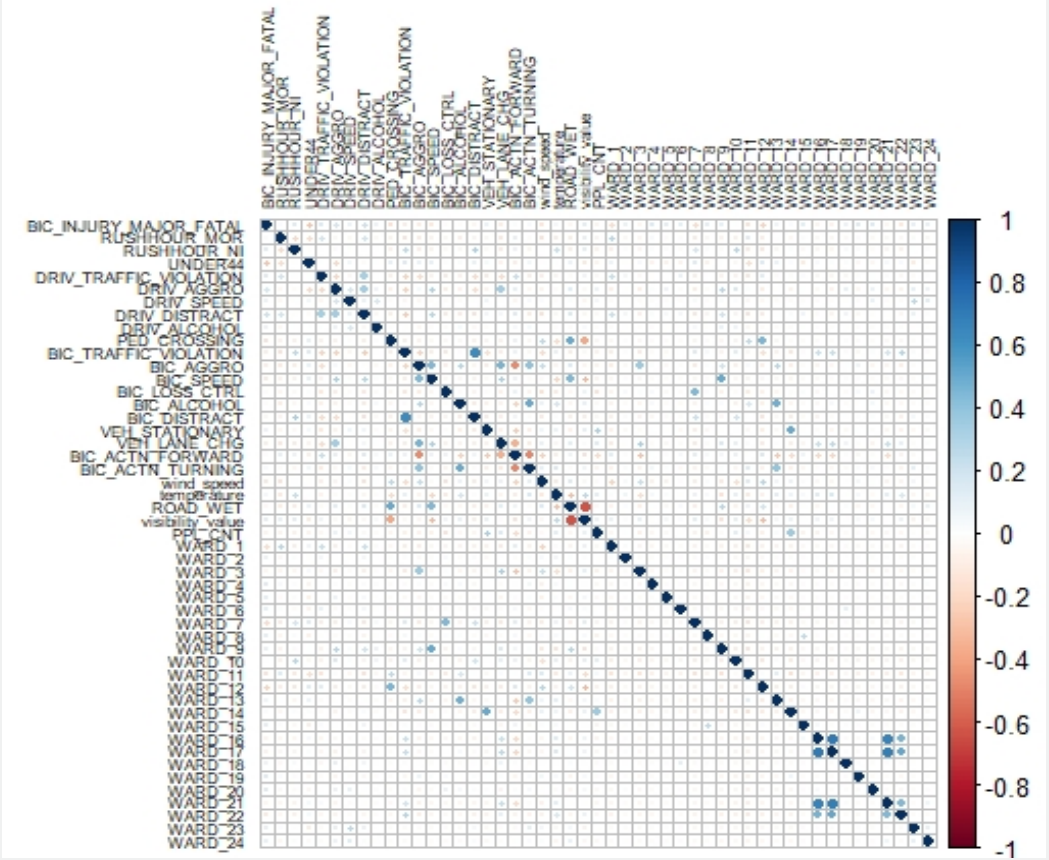
Larger Darker Circles (either red or Blue) indicate correlations close to 1 or -1



Appendix F5: Correlation Matrix B

This is the correlation matrix for our final model.

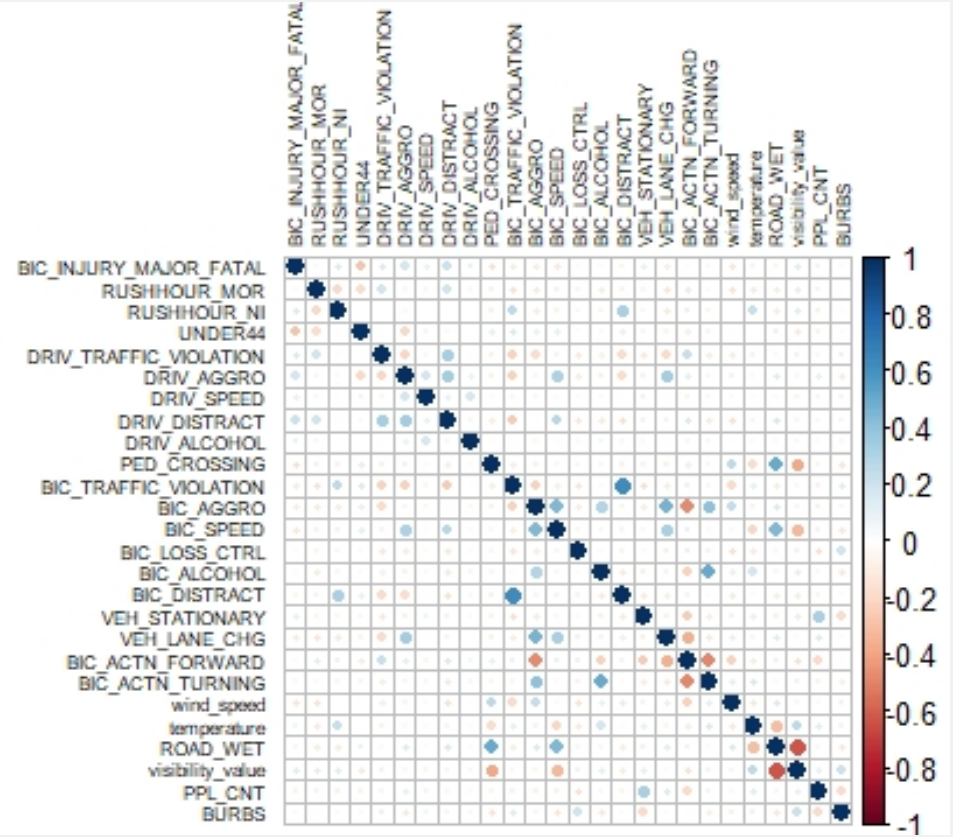
Larger Darker Circles (either red or Blue) indicate correlations close to 1 or -1



Appendix F5: Correlation Matrix C

This is the correlation matrix for the model with wards that are a part of the suburbs of Toronto.

Larger Darker Circles (either red or Blue) indicate correlations close to 1 or -1



Appendix H – Business Questions

Overarching Question	What series of recommendations will help further eliminate fatalities and serious injuries in Toronto's transportation system?
Detailed Questions	Where are these fatal incidents are taking place (Downtown, Arteries)?
	What variables are likely to contribute to major/fatal accidents?
	Are there recurring accident patterns like start of the week at 9am, or Thursday at 5pm? What about the time of the year? (i.e. Summer peak, back to school) Which day of the week has the most accidents?
	What demographics were involved in reported cycling accidents?
	What were they doing? What are the main accident categories? (i.e. distracted cycling, DUI) What maneuvers did they perform (turning, crossing street)?
	What conditions lead to accidents?