Driving Behaviour Analytics for Monitoring Changes in Behaviour

DATA5000: Introduction to Data Science
Final Report: April 16, 2017

K. Jacobs
Department of Geography and Environmental Studies
Carleton University
Ottawa, Canada
Kent.Jacobs@carleton.ca

N. Fung
Department of Systems and Computer Engineering
Carleton University
Ottawa, Canada
Nathanael.Fung@carleton.ca

ABSTRACT

The existing methods of screening medically unfit older drivers in Canada (i.e., cognitive and physically impaired) needs to be improved by providing more evidence-based, objective measures of driving ability. As older drivers often rely on driving for independence (e.g., visit friends, grandchildren or social clubs), revoking their driver's license can negatively impact their quality of life. By analyzing the driving history from sensor data on real-world driving, two methods were developed to monitor the changes in driving behaviour to detect a decline in driving ability. Our methods were developed from a dataset that included thousands of trips from 28 older drivers over a few recent years from the Candrive research study. The first method includes the development of total driving area and primary "mean" locations for the further proposal of a real-time GIS monitoring application. For the second method, a detection algorithm was created to track the proportion of unusual trips per month and provides an alert when a consistently high proportion of unusual trips were observed, which may indicate that the older driver is losing the sense of direction and have decline in memory. The methods in this study can improve the screening process for medically unfit older drivers and contribute to reducing healthcare costs and improve physician-patient relationship for older adults.

Keywords

Driving Behaviour, Driver Analytics, Elderly, Sensor Data, Geographic Information System (GIS), Unsupervised Learning, Anomaly Detection, Cognitive and Physical Impairment

1. INTRODUCTION

The demographic of older adults, aged 65 years old and older, is increasing in Canada. This population is expected to become 20% of the total Canadian population by 2025 [1] and is largely contributed by the 'baby boomers' generation. Since many of these people have a valid driver's license, the increase in older drivers would place a larger demand for governmental services related to older drivers.

As older adults experience cognitive and physical decline, their driving ability can be negatively affected and this decline in driving ability can result in increased risks of motor vehicle collisions [2]. Preventing at-risk and unsafe older adults from driving can not only protect themselves, who are often more vulnerable, but also the general public (i.e., other drivers and pedestrians). However, driving is an essential part of their independence and revoking their driver's license can have a negative impact on their quality of life, including increased social isolation which can lead to depression [3]. With the growing population of older drivers, it will be an even greater challenge to improve traffic safety while maintaining a high quality of life for older drivers.

The screening process of older drivers is not standardized across Canada. Each province has its own legislation for assessing the driving ability of older adults. For example, older drivers aged 80 years and older are required to renew their driver's license every 2 years in Ontario [4], while medical report from primary-care physicians are required every 2 years in British Columbia [5]. Regardless of the province, the primary-care physicians are obligated to report their patients who they believe to be medically unfit to drive at any time.

Problems can arise during the screening process because it is often difficult to determine when exactly to report the older drivers. Most physicians use in-office cognitive and physical assessments to determine medical fitness to drive, which often consist of relatively simple and timed tasks. For example, one task is to draw a picture of a clock within a time limit, and another is to recall a pattern of numbers or walk for a specific distance. Although these tests were designed to assess cognitive and physical ability, they may not always be an accurate reflection of their overall driving ability (e.g., patient may have an off day). As a result, situations may arise where the physician deems the patient medically unfit to drive, but the patient is in disagreement. For example, the medical assessment may show a negative score, but the patient's driving history may indicate otherwise. In another scenario, the patient may have spent some time practicing for the in-office medical assessment and obtains a passing score on the tests, but in reality has a low driving ability. On-road driving tests can be more accurate in assessing driving ability, but the stress from being observed by an examiner combined with the possibility of imminently losing their driver's license can result in an unusually poor performance. Thus, there is a need for more objective and unbiased, evidence-based screening tools for identifying medically unfit older drivers.

A possible solution to this problem is to track the everyday driving of older drivers using in-car sensors and build a record of their driving habits over time. This method creates a history and unbiased representation of their driving behaviours in natural conditions. At the time of the in-office medical assessments for fitness to drive, the records of the past and recent driving history could provide supporting evidence for the physician to make better-informed decisions and also detect low or declining driving ability earlier than the current assessments.

To make use of the in-car sensor data, computer algorithms must be created to model the driving behaviours and detect when the older driver may be experiencing a change in driving behaviour, which may indicate a decline in driving ability associated with cognitive or physical impairments.

2. OBJECTIVES

With sensor data from older drivers' trips, we aim to develop two methods to monitor changes in driving behaviour to detect possible cases of declining driving ability.

Since the individual sensor readings at one point in time are not informative, we need to develop a processing algorithm to extract useful information from the sensor data and obtain information about their driving behaviour. Once the useful information has been extracted, we develop two methods to monitor changes in driving behaviour:

- A combination of sensor analytics and spatio-temporal analysis to visually observe for any large or obvious deviations in values.
- An anomaly detection algorithm to identify unusual trips, where a consistently high frequency of unusual trips may indicate a declining memory and sense of direction.

3. DATASET

The dataset was provided by the Candrive research study, formally known as the Canadian Driving Research Initiative for Vehicular Safety in the Elderly [3]. The study used a data logging device to the record in-car sensor measurements from everyday driving of older adults, who were at least 70 years old. Each of the participant's personal automobile was outfitted with the data logging device that recorded GPS, On-Board Diagnostic II (OBDII) engine, and GIS data. The signals that were selected for this study are listed in Table 1.

A total of 28 drivers were provided with varying years of data ranging from 2009 to 2015. These drivers resided in the same city and were considered regular drivers who did not share their vehicles. This dataset represents a realistic scenario where a group of elderly patients is under the care of a physician. All other information about the participants was withheld.

Below is a list of properties of the sensor data

- 1 Hz sampled data (data recorded every second)
- 4 month intervals of CSV files (structured data)
- Average file size is 1 MB CSV file
- Average number of samples (rows) per file is 292,800
- Total number of files is 374 files (~370 MB)

4. METHOD

4.1 Data Preprocessing

The processing steps in Figure 1 were used to process the raw selected sensor signals in Table 1.

Table 1. Selected sensor signals from the raw data

Signals	Measurement	Units/Format
GPS	Date & Time	Date String
	Longitude & Latitude (Anonymized)	Decimal Degrees
	Speed	km/hr
GIS	Posted Speed Limit	km/hr
OBDII	Speed	km/hr
Device	Trip Number	Integer

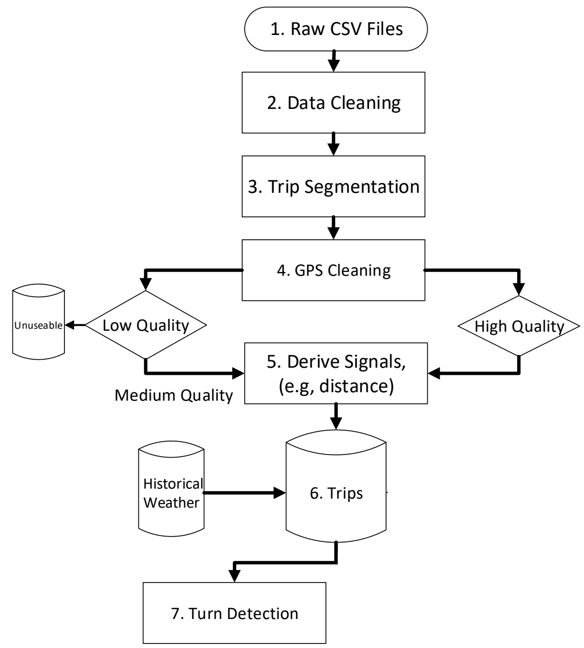

Figure 1. Block diagram showing the processing steps for the raw sensor data

Each step are described as follows:

1) A unique identification number was assigned to each driver driverID = {1,2,..,28} and the CSV files were organized by these driverID. All files were verified to be CSV files.
2) Each measurement was verified to have realistic values (e.g., < 150 km/hr vehicle speed). If not, measurements were replaced with the nearest realistic value.
3) The sensor data was divided into *trips*, which were the periods between the automatic starting of the data logging when the car turned on until the automatic stopping of the logging when the car turned off. Trips that were < 100 m or < 2 min were removed, such as periods of idling or readjusting parking.
4) The GPS trace drifted when the car temporarily stopped (e.g., during a red traffic light) which caused a phantom speed reported by the GPS. The secondary measurement of speed from the vehicle's engine sensor did not drift, so it was used to detect when the car was actually at rest. The GPS coordinates were set to the location before the drifting, and the speed was correct to 0 km/hr. For a few trips, there were dropped samples that occurred at the beginning of the trip (i.e. when the data logging device was initializing) and areas of low GPS reception (e.g., in a tunnel). The trips were sorted into three groups:
   - High Quality: absence of dropped samples (~98% of the data)
   - Medium Quality: A few dropped samples occurred and were filled-in by linearly interpolation.
   - Low Quality: Too many dropped samples; the trip was considered unreliable and was discarded (~1% of the data)
5) From the raw sensor signals, Table 2 lists the derived variables that were used to summarize each trip.
6) The historical daily and monthly weather CSV files for each respective year was downloaded from the Government of Canada website [6], and missing values were filled-in using the nearest neighbours. Since only the city was known, the weather was approximated by

using the weather recorded from the city's main airport. The corresponding weather information (e.g., temperature, outlook, rain level) was integrated by adding the weather to each trip.

7) For each trip, the number of turns in the route was counted to describe the complexity of the trips.

4.2 Features Extraction

Table 2 lists all the extracted information from the sensor data on each trip, which is effectively converting the stream of sensor data into trip summaries.

Table 2. Features extracted from each trip

Features	Description/Format
Day of the Week	Relative Frequency Distribution with 7 bins: [Sun,Mon,Tue,Wed,Thu,Fri,Sat]
Hour of the Day	Relative Frequency Distribution with 24 bins: [0,1,...,23]
Time of Day (Rush Hour)	Relative Frequency Distribution with 4 bins: [7am-9am,9am-4pm,4pm-6pm,6pm-7am]
Stop Time (min)	Count the number of samples <= 5 km/hr measured speed
Moving time (min)	Count the number of samples with > 5 km/hr measured speed
Total distance travelled in the trip (km)	Haversine Formula
Average Moving Speed (m/s)	$\frac{Distance\ Travelled\ (m)}{Total\ Moving\ Time\ (s)}$
Destination Bearing (°)	Azimuth angle from start to end destination of the trip
Displacement (km)	The distance between the start and end destination of the trip
Bounding Box (km)	Apply a map projection with universal transmercator (UTM) to convert decimal degrees into meters and form a tight box around the route of the trip. Length = (max(x) – min(x)) Wdith = (max(y) – min(y))
Route (km/hr)	Relative frequency distribution of the roadways, which is the proportion of each speed limit occurring in the route with 5 bins: [40,50,60,80,100] km/hr. These values are restricted to between 0 and 1, and each bin is standardized with mean of 0 and standard deviation of 1.
Number Of Turns	Created vectors along each adjacent coordinate in the route and calculate the point-to-point using change in heading using the following formula: $\theta = cos^{-1}\left(\frac{\vec{A}\cdot\vec{B}}{\lvert\vec{A}\rvert\lvert\vec{B}\rvert}\right)$ The threshold $\theta > 60^{o}$ was used for detecting prominent turns in the route.

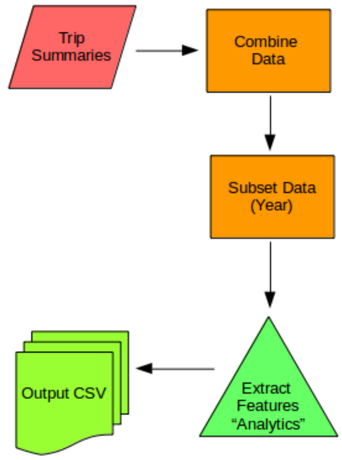

Figure 2. Designed workflow chart for the development of driver analytics.

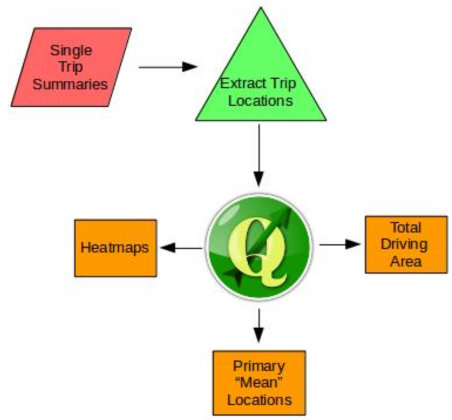

Figure 3. Designed workflow chart for the development of total driving area and primary "mean" locations (Spatio-Temporal Analysis/Geovisualization).

4.3 Driver Analytics

By using the derived variables (i.e., moving speed, distance) and extracted features (i.e., average moving speed, total trip distance) for each driver and their respective trips, advanced driver analytics were developed to monitor changes in driving behaviour. This process was done entirely with the open-source statistical software package R. Figure 2 illustrates the entire pre-processing and cleaning stages for the development of driver analytics.

Using the extracted trip summaries (i.e., trip duration, distance, weather data) advanced driver analytics were developed to compare each of the drivers and their behaviour. An example of generated driver analytics includes number of short, medium and long distance trips, number of short, medium, long duration trips and number of rush hour trips. Table A.1 (Appendix A) includes a comprehensive list and descriptions of the generated driver analytics used to compare drivers.

4.4 Spatio-Temporal Analysis

Single trip summaries were used to develop methods of analyzing driver behaviour and changes through space and time scales. Even with the anonymized latitude and longitude coordinates, it is still possible to visualize changes in driving area and distance using GIS software. Figure 3 outlines the workflow for processing and cleaning the single trip summaries (with anonymized coordinates) to generate output maps used for spatio-temporal analysis. R was also the statistical software package used for the process along with the open-source GIS software package, QGIS.

Using methods developed by Babulal et al. [7], the total driving area and primary "mean" driving locations were extracted for two drivers. Total driver area is defined as the smallest polygon that includes all the anonymized coordinates for a driver over a given year. Total driving area polygons were generated using the convex hull algorithm in QGIS. Primary "mean" locations is defined as the geographical mean centre for all of the anonymized coordinates. For example; mean x-coordinate = (X1 + X2 + X3...+ Xn) / N, mean y-coordinate = (Y1 + Y2 + Y3...+ Yn) / N. Primary "mean" locations were generated using the mean coordinates algorithm in QGIS. Wagner et al. [8] examine methods to analyze speed along single trips. Using QGIS and coloured symbol styling, it is possible to examine how fast a vehicle is moving along one single trip.

The development of "heatmaps" were used to visualize and analyze arrival and departure locations for each of the drivers. Heatmaps are used to visualize high and low concentrations of data using colour. For example, heatmaps are typically used to analyze high and low crime regions within urban areas. Using R, the first and last anonymized coordinates were extracted from two driver's trips and imported into QGIS. Using the Heatmap plug-in, maps were generated to visualize departure and arrival locations over three years for five drivers. Five drivers with the highest quality of data were analyzed, but additional analysis could have been conducted for subsequent drivers in the future.

4.5 Detecting Unusual Trips

To develop a detection algorithm for the anomalous trips, the analysis was limited to 3 specific years (i.e., 2011, 2012, 2013) because these were the most common years among the drivers since each driver began and concluded their enrollment in the Candrive study in different years; 22 of the 28 drivers fit this criteria. We treat the first year, 2011, as the baseline year to model their normal behaviour.

4.5.1 Definition of an Unusual Trips

Unusual trips are anomalous trips that contain extreme values that are not commonly observed in the baseline year, 2011, and these represent examples of driving behaviour that deviate from their normal behaviour. If there is a consistent increase in the number of unusual trips, this provides evidence of a change in driving behaviour and could provide evidence for a decline in driving ability.

4.5.2 Driving Behaviour Model

Since each driver has a unique driving history, the modeling of the normal behaviour must be appropriate for each driver. Table 3 lists the features that were used to model the driving behaviour and the corresponding probability density functions (PDF) that were fitted to the distributions. The PDF was used to generalize the attributes of the driver's trips and describe the probabilities of observing a particular trip.

4.5.2.1 Distance

The total distance travelled had a characteristic unimodal distribution with right skew. This may indicate that the older drivers preferred to walk for short distances rather than drive; most of the trips were made within a specific radius of their homes (i.e., trips to the local mall or grocery store); and long distance trips are less common (i.e. travelling to the city, special events, or out of town trips).

Table 3. Probability density functions for statistical modeling of normal driving behaviour (year 2011)

Features	Probability Density Function
Distance (km)	Lognormal (continuous)
Destination Bearing (°)	Kernel Density Estimation with Guassian Kernel (continous)
Displacement (km)	Lognormal (continuous)
Length of the Trip (km)	Lognormal (continuous)
Width of the Trip (km)	Lognormal (continuous)
Area (km^2)	Exponential (continuous)
Route (km/hr)	Gaussian Mixture Model using DBSCAN for initial cluster seeds (continuous)
Number Of Turns	Relative frequency distribution (discrete)

4.5.2.2 Destination Bearing

These distributions exhibit a multimodal distribution that is limited between 0^0 and 360^0. This may indicate that the driver frequently visits specific regions (e.g., a large shopping center contains a grocery store, clothing stores, gym, and social clubs) and the back-and-forth trips (e.g., trips between home and work). Kernel density estimation with Gaussian kernel was used to fit a PDF to the multimodal distributions as shown in Figure 6.

4.5.2.3 Displacement

The displacements represents the distance of the shortest path (i.e. "As the crow flies") of the trip and had similar distribution to distance in Figure 4.

4.5.2.4 Length of Trip

The length of the bounding box per trip represents the distance travelled along north-south direction and had similar distribution to distance in Figure 4.

4.5.2.5 Width of Trip

The width of the bounding box per trip represent the distance travelled along the east-west direction and had similar distribution to distance in Figure 4.

4.5.2.6 Area of Trip

The distribution of the area of the bounding box per trip represents the size of the trip and Figure 7 shows that most of the trips have short distances.

4.5.2.7 Number of Turns

The number of turns represents the complexity of the trip. Figure 10 shows that the distribution of the number of turns is unimodal with right skew.

4.5.2.8 Route

Figure 8 shows that there are natural clusters of routes, which indicate that the older drivers have preferred routes. To identify the prominent clusters, density-based spatial clustering of applications with noise (DBSCAN) algorithm was used to separate noise from the prominent clusters. DBSCAN requires two parameters: the radius to check for neighbours and the number of neighbours that must be within the radius. The radius was set to a tenth of the average squared Euclidean distance between every datapoint and the number of neighbours was set to 10. The datapoints that were labeled as noise were used as test cases for determining a threshold probability density value (Table 4).

After the prominent clusters were determined, the mean datapoint of each cluster was used as the initial seeds for the expectation-maximization algorithm (EM) to create a Gaussian Mixture Model (GMM). GMM is a combination of Gaussian distributions to model the probability density of the routes.

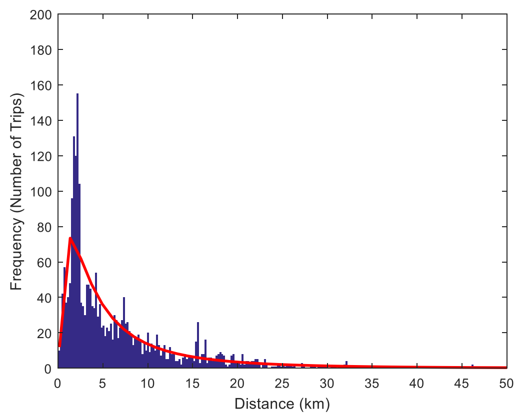

Figure 4. The distribution of the total distance travelled per trip and its log normal PDF for an example driver. Similar shape of the distribution was observed for the displacement, length, and width distributions.

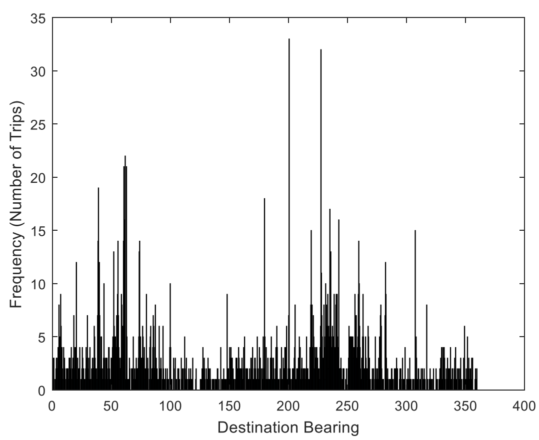

Figure 5. The distribution of the destination bearing per trip for an example driver.

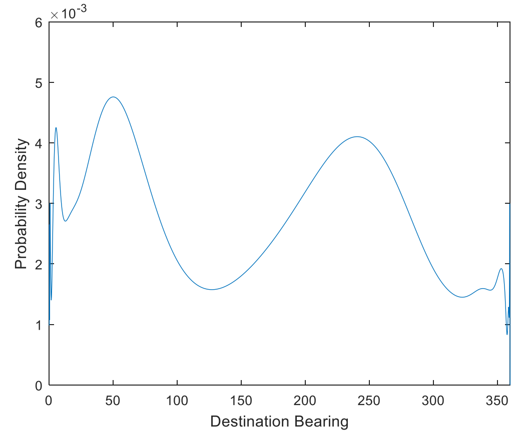

Figure 6. The KDE PDF of the destination bearing distribution for an example driver. The endpoints have error due to numerical approximations.

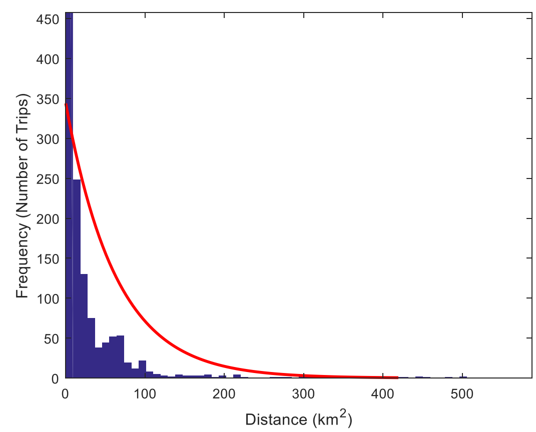

Figure 7. The distribution of the bounding box area per trip and its exponential PDF for an example driver.

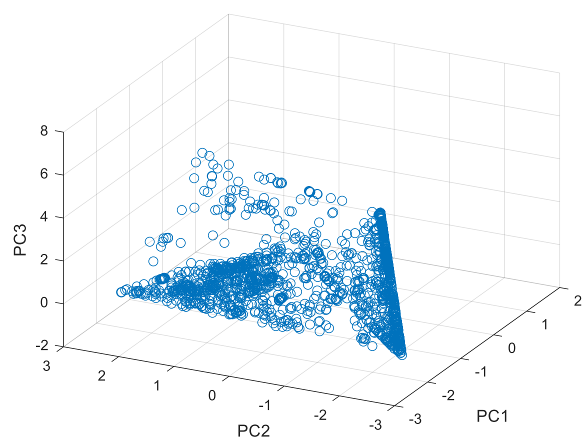

Figure 8. The distribution of the routes for an example driver. The 5 dimensional data vector is projected onto 3 dimensional space with Principal Component Analysis (PCA) for visualization. The triangular shape is due to many values being zero in the bins for the route feature.

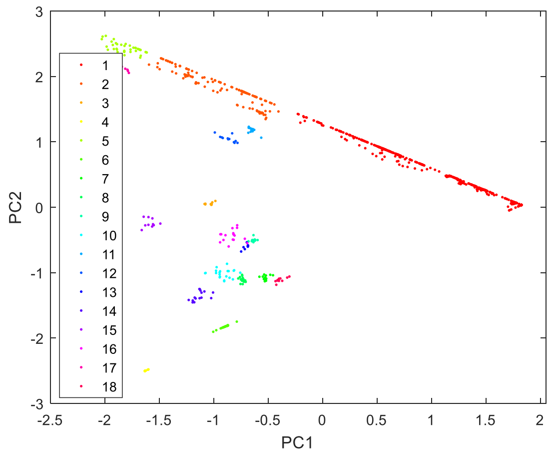

Figure 9. The clusters of routes for an example driver. The 5D data vector was projected onto 2D space with PCA for visualization. A GMM was created with these clusters.

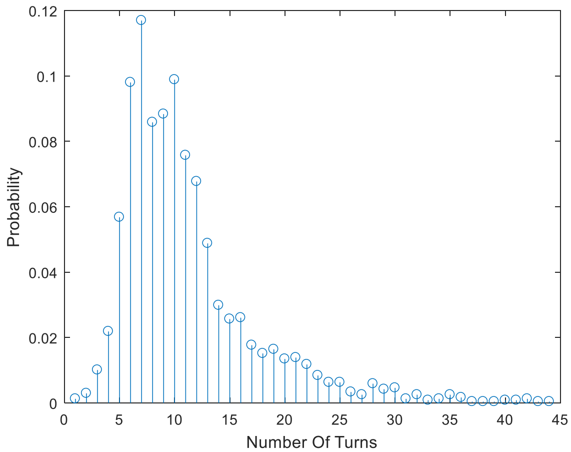

Figure 10. The probability distribution of the number of turns per trip for an example driver.

Table 4. Definitions of extreme values for each feature

Features	Detection Rules
Distance (d)	P(d < 0.01) or P(d > 0.99)
Destination Bearing (b)	PDF(b) < ε The threshold was obtained from calculating the probability density at which P(d < 0.01)
Displacement (dp)	P(dp < 0.01) or P(dp > 0.99)
Length of the Trip (l)	P(l < 0.01) or P(l > 0.99)
Width of the Trips (w)	P(w < 0.01) or P(w > 0.99)
Area (a)	P(a < 0.01) or P(a > 0.99)
Route (r)	PDF(r) < ε The threshold was obtained by using the noise datapoints as test cases for unusual trips, and finding the density value that provided equal false negative and false positives rates in detecting the noise.
Number Of Turns (t)	P(t < 0.01) or P(t > 0.99)

4.5.2.9 Rules for Detecting Unusual Trips
Any extreme values observed in any of the features indicate an unusual trip, but to detect only the highly unusual trips, two of the features must exhibit extreme values to be considered an unusual trip. The rules for detecting an unusual trip for each feature are outlined in Table 4.

4.5.2.10 Rule for Detecting A Change in Behaviour
It is expected that drivers would take unusual trips from time to time (i.e. vacation or visiting a friend), thus a better indication of a change in behaviour is to observe the number of unusual trips with respect to the total number of trips per month. A sustained increase in the proportion of unusual trips for 6 consecutive months would indicate a change in behaviour and raise an alert.

5. RESULTS AND DISCUSSION

5.1 Driver Analytics
Using the designed driver analytics in Table A.1, resulting descriptive statistics were generated (Table A.2, A.3, A.4). Table A.2-A.4 do not use all of the developed driver analytics outlined in Table A.1, rather driver analytics that were thought to be most indicative of one's driving behaviour. For example, while a driver average moving speed may describe how fast a driver is moving throughout their trips, it does not indicate if a driver is actually speeding. If a driver has a high average moving speed, that driver may just be driving more often on highways (with high speed limits). Rather than analyzing average moving speed, number of times found going 30 and 50 km/hr over the speed limit is more descriptive of how fast and how often a particular driver is speeding. Number of long distance and duration drivers was deemed an indicative descriptive statistic based on the idea that it would be easy to determine drivers based on how high or low distance or duration are. For example, if one particular driver has a low number of long distance trips and a high number of short distance trips, it is possible to conclude that this driver does more urban driving (longer stop times).

The incorporation of weather data was also included in the driver analytics portion of this project. Speeding during hazardous weather is often seen as dangerous driving behaviour. Average moving speed during rain and snow weather events were also believed to indicate how drivers react to potentially dangerous driving conditions. Higher average moving speed in rain and snow events can indicate drivers who are driving faster in these weather conditions as well as the type of roads they are driving on (higher average moving speed = major highways). If a driver decides to drive on a major highway during a rain or snow event, it may indicate they are more confident in their driving abilities during these events.

5.2 Spatio-Temporal Analysis
By processing total driving area using a convex hull algorithm, it is possible to gain an understanding on the entire spatial extent of how far a driver travels. A primary "mean" location can be calculated to determine an average area (within their driving area) that they tend to travel around. Table 2 outlines the findings from Figures A.6-A.10. The total driving area calculations are represented in kilometres squared (km^2) and perimeter in kilometres (km). It should be noted that driver 26 did not participate in this study during the year 2009.

Table 1. Total driving area (km^2) and perimeter (km) for five drivers between the years 2009 to 2011. Driver 26 did not participate in the study in 2009.

DriverID	Year	Total Driving Area (km^2)	Perimeter (km)
3	2009	8,996	1,083
	2010	13,476	1,624
	2011	135	117
5	2009	183	92
	2010	1,579	399
	2011	456	234
20	2009	216	160
	2010	5,382	664
	2011	10,862	678
21	2009	4,723	622
	2010	20,312	1,312
	2011	13,579	1,149
26	2010	52,407	2,086
	2011	21,218	1,521

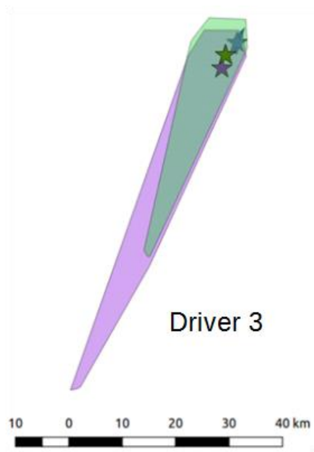

Figure 11. Driver 3 total driving area (polygon) and primary "mean" coordinate locations (star) through years 2009 (green), 2010 (teal) and 2011 (blue). Attached to Appendix A.

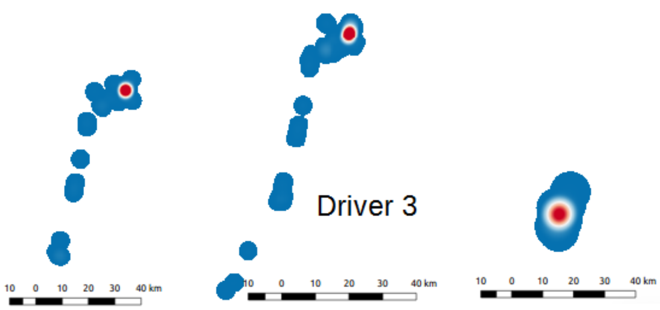

Figure 12. Driver 3 heatmaps for arrival and departure locations through years 2009, 2010 and 2011 (left to right). Attached to Appendix A.

Figure 11 illustrates an example for what the total driving area and primary "mean" locations can look like through a map or GIS. As previously mentioned, total driving area was calculated using the convex hull algorithm within QGIS. From here, it is possible to calculate the area and perimeter using QGIS.

Figure 12 shows the generate heatmaps from the derived departure and arrival locations for each of the five drivers analyzed over space and time.

5.3 Detecting Unusual Trips

With the detection algorithm, the proportion of unusual trips per month was calculated for 3 years (36 months) for 22 drivers.

5.3.1 Detecting Normal Driving Behaviour

Figure 13 shows the proportion of unusual trips per month for driverID = 22 over 36 months, where no change in behaviour was detected. This driver seems to have a normal driving behaviour.

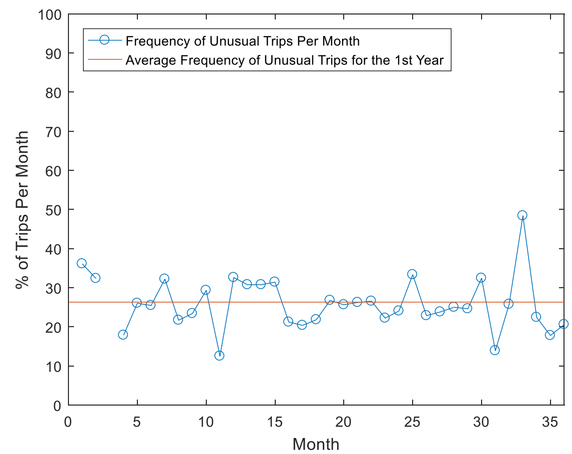

Figure 13. The proportion of unusual trips per month for driverID = 22. The missing values indicate that the driver did not have any trips for that month.

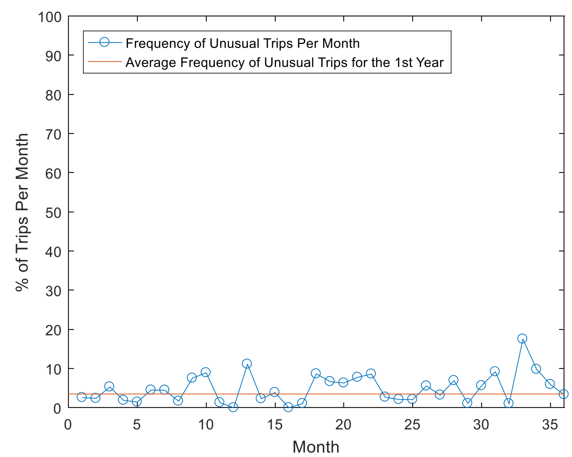

Figure 14. The proportion of unusual trips per month for driverID = 1. This driver has an overall low amount of unusual trips, indicating more consistent driving routine and behaviour.

Figure 14 shows the proportion of unusual trips per month for driverID = 22 over 36 months, where no change in behaviour was detected. This driver seems to have lower proportion of unusal trips which is an example of a more consistent driving behaviour.

5.3.2 Detecting Change in Driving Behaviour

Figure 15 shows the proportion of unusual trips per month for driverID = 11 over 36 months, where there was a change in behaviour. This driver has a gradual increase in the proportion of unusal trips, which may be an example of a steady decline in driving ability.

Figure 16 shows the proportion of unusual trips per month for driverID = 18 over 36 months. This driver is an example of large change in behaviour, which may be an example of a drastic decline in driving ability.

Figure 17 shows the proportion of unusual trips per month for driverID = 19 over 36 months, where there is subtle change in behaviour, which may be an example of early onset of impairment.

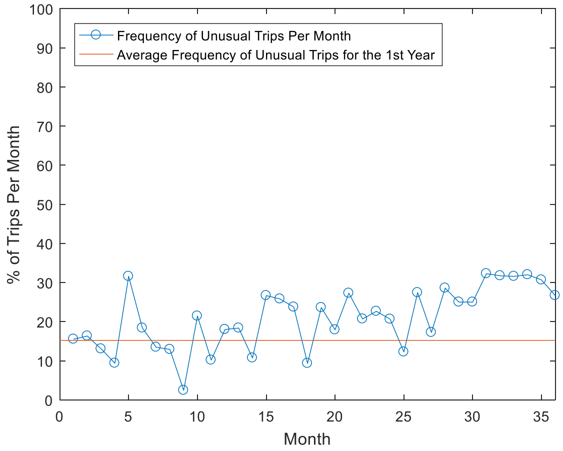

Figure 15. The proportion of unusual trips per month for driverID = 11. This driver has a gradual increase in the proportion unusual trips in the 3rd year, indicating a change in behaviour and possibly a decline in driving ability.

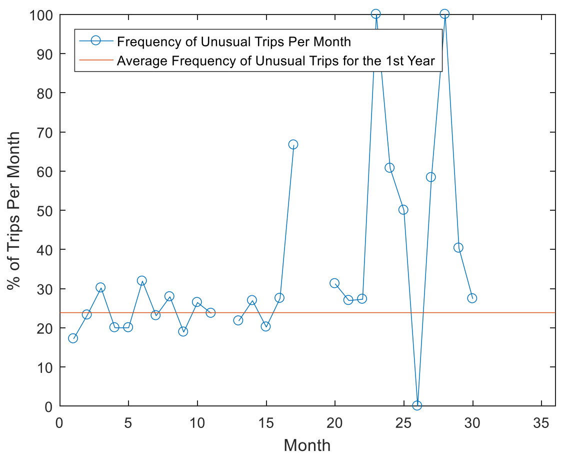

Figure 16. The proportion of unusual trips per month for driverID = 18. This driver has an increased proportion of unusual trips in the 3rd year, indicating a large change in behaviour and possibly a drastic decline in driving ability.

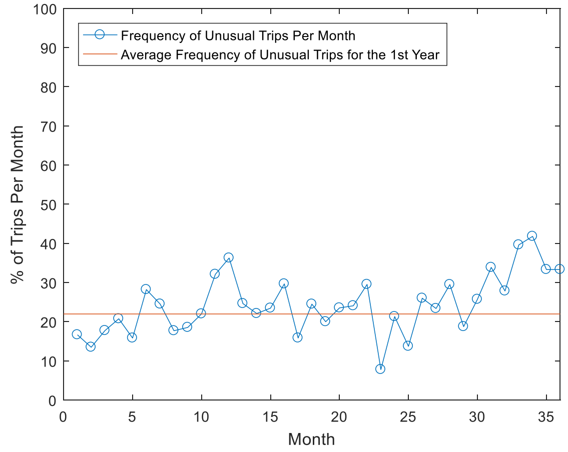

Figure 17. The proportion of unusual trips per month for driverID = 19. This driver has a subtle increase in unusual trips in the 3rd year, possibly an early onset of impairment.

5.3.2.1 Driving Behaviour Report

An example of a report for physicians in monitoring changes in behaviour is Table 5 which can provide alerts and the time of the detection. This report can easily mark which drivers to follow-up for further assessments when changes in behaviour were detected.

Table 5. An example of a report generated from screening older drivers for changes in driving behaviour

DriverID	Alert	Time Period
1	No	-
2	No	-
3	No	-
4	No	-
5	No	-
6	No	-
7	No	-
8	No	-
9	No	-
10	No	-
11	Yes	Feb 2012 to Dec 2013
12	No	-
13	Yes	May 2011 to Oct 2011
14	No	-
15	No	-
16	No	-
17	No	-
18	Yes	Aug 2012 to Jan 2013
19	Yes	July 2013 to Dec 2013
20	No	-
21	No	-
22	Yes	March 2013 to Nov 2013

6. IMPLICATIONS

The ability to process sensor signals and extract useful information to monitor older drivers can be beneficial for detecting when an older driver is experiencing a change in driving behavior related to a decline in driving ability. Understanding more about the driving behaviour of older adults can contribute to developing personalized driving cessation programs based on their driving history and behaviour.

The methods proposed in this study can be incorporated into a larger scheme and design of a more complex screening tool for cognitive and physical impairment detection, and thereby improving the safety of older drivers and the public while maintaining a high quality of life. The sensor data on driving history can become one of several modes of information (i.e., medical tests, police driving records) to comprehensively assess medical fitness to drive. Utilization of sensor data can be a key component of a driving safety program in the future; this can revolutionize screening process of older drivers. Older adults would enroll into this program at a specific age to have their driving monitored. They could be rewarded for good driving behaviour with prolonged driver's licenses or lower insurance premiums and early prevention of at-risk drivers can reduce motor vehicle collisions.

The potential positive impact includes lower costs of healthcare services related to older drivers and improved physician-patient relationship. The physician can choose to allocate resources to specific patients to speed up the screening process, rather than fully test each patient equally which is time-consuming. This ultimately will contribute to alleviating the healthcare costs and the tax burdens on the younger generations in the society, as there are a growing number of retired adults.

The existing screening process can create distrust between the physician and patient because of discrepancies in medical assessment scores and the on-road driving history. By having supporting evidence from real-world driving, both parties can have a better understanding and come into a better agreement over these objective measures of driving ability. This improved physician-patient relationship can contribute to reducing the number of unnecessary follow-up tests caused by patients' disagreements.

6.1 Driver Analytics

The development of driver analytics allows for the monitoring of changes in driving behaviour. Tables A.2-A.4 outline a set of driving analytics that were most descriptive of how drivers behave on roadways. While Table A.1 outlines all of the driving analytics developed, this report outlines ten metrics used for driver monitoring.

One of the top ten metrics analyzed is average stop time (minutes) during each trip. By examining Tables A.2-A.4, it is clearly shown that drivers 4 and 5 exhibit the largest average stop time per trip (minutes) through 2009 to 2011. It can be further concluded that these two particular elderly drivers mainly drive in urban locations. This conclusion does merit value because within urban locations there are a large number of intersections where vehicles eventually have to stop at. There are a number of reasons why these elderly drivers exhibit higher average stop times per trip than other drivers. One reason may be due to the fact that they are still employed or run a personal business (self-employed). Driving across the urban regions without merging onto highway ramps will exaggerate the average stop time per trip for drivers 4 and 5.

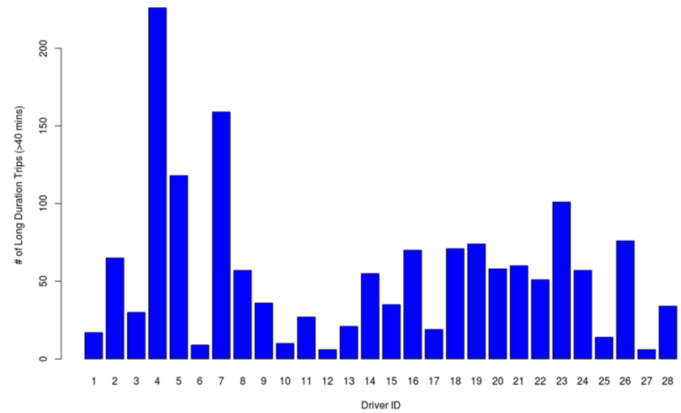

Figure 18. Total number of long duration trips for all 28 drivers (over 40 minutes).

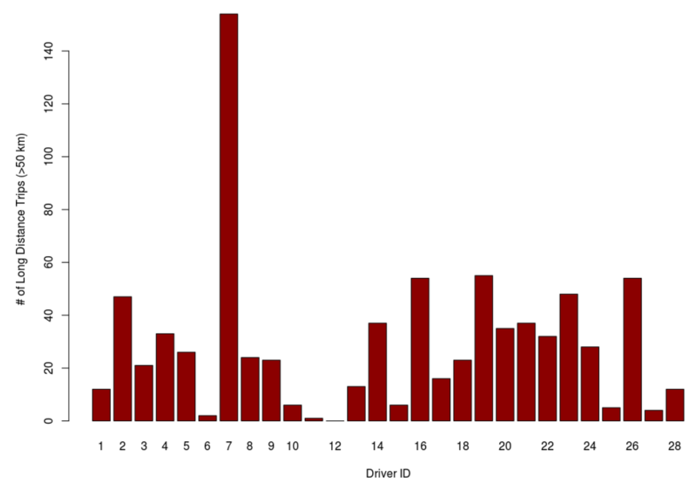

Figure 19. Total number of long distance trips for all 28 drivers (over 40 km).

By examining the number of long duration (Figure 19) and long distance trips (Figure 20), it is evident that drivers 4 and 5 display larger long duration trips (over 40 minutes) and few number of long distance trips (over 40 km). Figures 19 and 20 also show some drivers who driver a large number of long duration trips and long distance trips. One example of this instance is driver 7. Driver 7 travelled over 150 long duration trips and 140 long distance trips during the duration of their time in this research study. As a result, driver 7 may exhibit a greater confidence in their long duration trip driving abilities and feel much more comfortable driving greater distances from their homes or primary "mean" locations.

7. CONCLUSION

With the expected growth of older adults in Canada, the number of older drivers will increase and existing methods of screening medically unfit older drivers (i.e., cognitive and physically impaired) need to improve to provide more evidence-based, objective measures of driving ability. By using sensor data collected from real-world driving of older adults, our goals were to process the sensor data and develop methods for monitoring changes in driving behaviour. In this study, we developed a method to automate the processing of real-world sensor data and proposed two methods for monitoring changes in behaviour for detecting declining driving ability. Our methods were developed from a data set that included thousands of trips from 28 older drivers over a few years from the Candrive research study.

With the development of health GIS applications for monitoring disease movement, we propose a further GIS application that can be used for monitoring elderly drivers. By gathering real-time sensor data, it is possible to monitor the movements and trajectories of elderly drivers in real-time. By monitoring total driving area and commonly located areas, it is possible to gain an understanding how the lifestyles of these people are changing by analyze their total driving spatial extent. This method analyzed the trips of 5 drivers over 3 years. The results demonstrate that each of the drivers analyzed display unique characteristics in terms of total driving area. By examining total driving area, it is possible to conclude that drivers are more comfortable driving long distances from home or primary "mean" locations.

An algorithm was created to detect the anomalous trips and count the proportion of unusual trips per month. An alert is raised when the older driver had a consistently high proportion of unusual trips for 6 consecutive months, which may indicate a decline in driving ability. Features were extracted from the drivers' trips and statistical models were used to generalize the normal driving behaviours from their baseline year. Of the 22 older drivers that were screened, 5 drivers were flagged to have an increased proportion of unusual trips. This demonstrates that each older driver has unique characteristics in their driving behaviour and that this proposed algorithm can detect older driver with changes in these behaviours. In practice, the physician could use this tool to generate a report and identify which patients to follow-up for further assessments. Medical assessments are often time-consuming and costly, thus reducing the number of tests is beneficial for lowering healthcare costs. The weather condition was expected to influence driving behaviour, but it was not considered for this algorithm.

As older drivers often rely on driving for independence (e.g., visit friends, grandchildren or social clubs), revoking their driver's license can negatively impact their quality of life. By analyzing the driving behaviour from sensor data on real-world driving, the screening process for medically unfit older drivers can become more objective measure of driving ability, which can contribute to reducing healthcare costs and improve physician-patient relationship.

Since the dataset did not include any medical assessment scores, further investigations is required for identifying associations between the observed changes in behaviour from the proposed screening methods and the actual health status of the individuals.

8. ACKNOWLEDGMENTS
We thank the Candrive research study (www.candrive.ca) and Lynn Maclean for providing the data on the older drivers. Without their support, this study would not have been possible.

9. REFERENCES
[1] Statistics Canada, "Canada's Population Estimates: Age And Sex, July 1, 2015". Sept 29, 2015. [Online]. Available: http://www.statcan.gc.ca/daily-quotidien/150929/dq150929b-eng.htm

[2] S. C. Marshall, "The Role of Reduced Fitness to Drive Due to Medical Impairments in Explaining Crashes Involving Older Drivers," *Traffic Inj. Prev.*, vol. 9, no. 4, pp. 291–298, 2008

[3] S. C. Marshall, *et al.*, "Protocol for Candrive II / Ozcandrive, a multicentre prospective older driver cohort study," *Accid. Anal. Prev.*, vol. 61, pp. 245–252, 2013.

[4] Ontario Ministry of Transportation, "Senior driver's license renewal program," March 23, 2017. [Online]. Available: http://www.mto.gov.on.ca/english/driver/senior-driver-licence-renewal-program.shtml

[5] Government of British Columbia, "Senior Drivers," 2017. [Online]. Available: http://www2.gov.bc.ca/gov/content/transportation/driving-and-cycling/driver-medical/driver-medical-fitness/senior-drivers

[6] Government of Canada, "Historical Weather Data," 2017. [Online]. Available: http://climate.weather.gc.ca/historical_data/search_historic_data_e.html

[7] G.M. Babulal, C.M. Traub, M. Webb *et al.*, "Creating a driving profile for older adults using GPS devices and naturalistic driving methodology," *F1000Research*, 2016, **5**:2376 (doi: 10.12688/f1000research.9608.1)

[8] M. Jensen, J. Wagner, and K. Alexander, "Analysis of in-vehicle driver behaviour data for improved safety," International Journal of Vehicle Safety, vol 5, pp. 197-212, 2011.

Appendix A

Table A.1. Designed driver analytics to monitor changes between drivers

Summary Statistic	Description	Time
Distance (km)	Number of short (<20), medium (>= 20 & <= 50) and long (>50) distance trips	Total Per month, week and day
Duration (minutes)	Number of short (>20), medium (>= 20 & <= 40) and long (>40) duration trips	Total Per month, week and day
Morning Rush Hour	Number of trips during morning rush hour (7-9am)	Total Per month, week and day
Day Rush Hour	Number of trips during day rush hour (9am-4pm)	Total Per month, week and day
Evening Rush Hour	Number of trips during evening rush hour (4-6pm)	Total Per month, week and day
Night Rush Hour	Number of trips during evening rush hour (6pm-7am)	Total Per month, week and day
Sunday Trips	Number of trips on Sunday	Total Per month
Saturday Trips	Number of trips on Saturday	Total Per month
Weekday Trips	Number of trips during weekdays (Monday-Friday)	Total Per month and week
30 km/hr Speed Limit (SpeedAbsDiff4)	Number of times found going 30 km/hr over the posted limit	Total Per month
30 km/hr Speed Limit (AvgMvSpAbsDiff4)	Average moving speed when found going over 30 km/hr speed limit	
50 km/hr Speed Limit (SpeedAbsDiff5)	Number of times found going 50 km/hr over the posted limit	Total Per month
50 km/hr Speed Limit (AvgMvSpAbsDiff5)	Average moving speed when found going over 50 km/hr speed limit	
Rain Moving Speed	Average moving speed when it is raining	
Snow Moving Speed	Average moving speed when it is snowing	
Stop Time (minutes)	Average stop time for all trips	

Table A.2. Driver analytics results for 2010

	DriverID	TotalTrips	AvgSpT	LongDist	LongDur	TSAdiff4	TSAdiff4Avg	TSAdiff5	TSAdiff5Avg	RainAS	SnowAS
1	[redacted]	1073	2.04	16	30	10	64	1	38.15	25.32	25.95
2	[redacted]	1541	2.79	44	74	15	71.07	20	83.91	30.02	29.08
3	[redacted]	1243	2.62	45	52	31	70.39	17	79.09	36.95	33.82
4	[redacted]	2049	5.81	40	190	134	53.89	54	79.5	35.93	29.96
5	[redacted]	952	5.57	12	58	117	69.35	70	74.37	56.91	40.41
6	[redacted]	1425	2.13	7	22	123	65.13	6	65.66	47.62	45.54
7	[redacted]	1468	1.59	175	176	106	82.41	71	85.19	59.76	55.24
8	[redacted]	659	2.06	44	78	50	70.01	39	83.94	54.63	48.2
9	[redacted]	2219	1.72	43	62	58	63.19	31	78.39	42.04	34.89
10	[redacted]	738	2.84	19	23	67	54.91	3	67.44	38.83	31.94
11	[redacted]	745	1.68	8	36	16	63.89	6	74.97	25.88	24.83
12	[redacted]	1311	2.41	0	5	3	39.99	0	NaN	34.21	33.64
13	[redacted]	1175	2.22	0	11	12	54.41	7	62.91	36.33	32.58
14	[redacted]	1380	2.58	57	89	239	78.46	183	83.21	44.45	37.24
15	[redacted]	939	2.73	6	47	125	60.08	3	73.84	38.12	42.07
16	[redacted]	1119	1.98	39	49	30	66.89	0	NaN	47.55	37.9
17	[redacted]	566	1.52	15	33	227	69.48	6	72.09	52.1	50.19
18	[redacted]	1586	3.08	17	48	76	50.58	11	78.73	32.34	29.09
19	[redacted]	1427	2.02	54	70	110	63.84	74	72.62	42.42	39.13
20	[redacted]	967	3.02	29	51	191	73.01	163	77.59	42.73	40.33
21	[redacted]	1011	1.82	57	69	58	67.45	37	89.02	35.95	43.07
22	[redacted]	1205	1.55	31	48	100	65.96	12	73.98	33.66	36.07
23	[redacted]	2745	2.12	161	207	57	71.05	51	82.36	38.61	35.57
24	[redacted]	1321	2.78	20	55	49	66.07	11	80.92	36.09	36.85
25	[redacted]	451	3.1	0	3	18	45.41	4	64.6	40.02	33.29
26	[redacted]	1043	2.38	50	64	57	66.35	48	69.83	45.81	36.42
27	[redacted]	593	1.14	5	11	72	57.24	14	76.82	37.84	54.3
28	[redacted]	828	2.2	14	33	58	69.69	118	79.7	42.17	25.79

Table A.3. Driver analytics results for 2011

	DriverID	TotalTrips	AvgSpT	LongDist	LongDur	TSAdiff4	TSAdiff4Avg	TSAdiff5	TSAdiff5Avg	RainAS	SnowAS
1	[redacted]	1282	1.63	12	17	8	67.76	0	NaN	24.28	24.97
2	[redacted]	1361	2.71	47	65	18	73.69	16	81.52	26.76	31.39
3	[redacted]	1371	3.67	21	30	25	62.12	13	86.13	31.55	32.13
4	[redacted]	1849	7.22	33	226	116	46.11	50	79.15	35.09	33.42
5	[redacted]	1723	5.44	26	118	186	72.82	128	77.17	38.66	32.88
6	[redacted]	1416	1.97	2	9	87	63.21	8	67.48	48.99	43.19
7	[redacted]	1514	1.63	154	159	79	84.33	54	90.36	54.14	56
8	[redacted]	559	1.9	24	57	44	73.54	21	74.95	56.05	54.72
9	[redacted]	1797	1.82	23	36	41	62.92	19	88.3	36.58	42.97
10	[redacted]	490	3.11	6	10	42	52.84	3	76.41	34.42	30.16
11	[redacted]	1071	1.62	1	27	21	65.22	2	42.58	27.84	23.77
12	[redacted]	970	2.35	0	6	8	41.2	0	NaN	31.23	33.72
13	[redacted]	1272	2.56	13	21	20	70.05	17	75.12	35.13	38.96
14	[redacted]	1142	2.16	37	55	226	83.03	186	85.8	40.77	43.8
15	[redacted]	995	2.85	6	35	190	59.49	4	66.03	44.99	42.97
16	[redacted]	1148	2.11	54	70	44	65.85	2	101.93	39.23	41.53
17	[redacted]	426	1.2	16	19	169	69.65	2	77.04	48.82	47.88
18	[redacted]	1351	3.27	23	71	96	63.94	47	74.02	30.21	27.82
19	[redacted]	1252	2.13	55	74	89	62.6	31	71.3	46	38.63
20	[redacted]	1133	3.15	35	58	257	75.86	236	79.89	39.2	45.11
21	[redacted]	1314	2.19	37	60	44	62.16	35	81.1	37.11	39.13
22	[redacted]	839	1.66	32	51	53	64.68	9	77.01	36.86	33.15
23	[redacted]	2377	2.16	48	101	34	67.93	27	83.2	35.27	31.86
24	[redacted]	1429	2.52	28	57	57	68.64	5	85.85	36.51	36.06
25	[redacted]	387	2.9	5	14	58	49.34	6	79.89	33.36	34.78
26	[redacted]	1196	2.63	54	76	61	65.93	61	72.9	37.21	41.99
27	[redacted]	643	1.32	4	6	111	58.79	23	75.72	36.23	39.95
28	[redacted]	1371	2.13	12	34	78	67.55	136	79.34	27.82	30.6

Table A.4. Driver analytics results for 2012

	DriverID	TotalTrips	AvgSpT	LongDist	LongDur	TSAdiff4	TSAdiff4Avg	TSAdiff5	TSAdiff5Avg	RainAS	SnowAS
1	[redacted]	1137	1.82	15	26	11	73.95	0	NaN	28.71	25.81
2	[redacted]	1406	2.47	24	37	22	72.22	12	82.15	29.88	27.52
3	[redacted]	1714	2.79	25	33	20	65.35	5	76.44	37.09	31.67
4	[redacted]	1684	8.16	25	216	99	48.25	36	77.88	31.84	31.51
5	[redacted]	2121	5.47	40	141	243	76.91	208	79.43	48.4	37.15
6	[redacted]	1374	1.94	6	23	70	61.37	7	64.2	45.61	44.62
7	[redacted]	1677	1.58	135	139	65	81.33	43	87.61	59.16	47.27
8	[redacted]	631	1.99	36	65	61	71.94	40	77.35	57.75	54.54
9	[redacted]	1532	2.03	47	63	40	64.32	27	91.73	34.85	46.04
10	[redacted]	589	3.14	13	20	43	55.29	4	61.14	32.13	28.74
11	[redacted]	1028	1.7	4	18	21	63.65	7	50.52	27.47	26.01
12	[redacted]	165	2.29	0	3	1	41.17	0	NaN	30.3	29.17
13	[redacted]	1336	2.69	8	21	19	62.4	14	66.11	34.21	34.09
14	[redacted]	1103	2.2	42	67	236	84.15	196	86.81	47.73	46.43
15	[redacted]	884	3.05	11	53	173	59.61	5	59.24	42.41	44.76
16	[redacted]	1089	2.12	57	80	36	64.63	2	84.79	43.55	36.54
17	[redacted]	43	3.45	0	0	0	NaN	0	NaN	NaN	25.39
18	[redacted]	897	2.1	44	55	66	63.97	28	78.61	43.33	36.77
19	[redacted]	1434	3.11	29	64	300	79.1	287	81.48	45.54	45.63
20	[redacted]	1022	1.73	15	20	39	61	11	73.9	41.58	35.82
21	[redacted]	502	1.36	20	29	29	67.81	8	81.6	36.15	16.84
22	[redacted]	2135	2.18	80	120	61	67.89	55	79.36	35.69	40.62
23	[redacted]	1347	2.45	17	38	41	62.9	2	81.42	35.35	29.94
24	[redacted]	105	3.18	0	1	13	46.88	0	NaN	21.95	37.38
25	[redacted]	973	2.53	27	35	57	57.72	44	62.72	38.97	34.77
26	[redacted]	750	1.49	13	24	124	57.82	13	72.38	35.68	37.2
27	[redacted]	1110	2.12	5	24	55	67.63	91	80.49	40.17	26.73

Figure A.1. Driver 3 heatmaps for arrival and departure locations through years 2009, 2010 and 2011 (left to right).

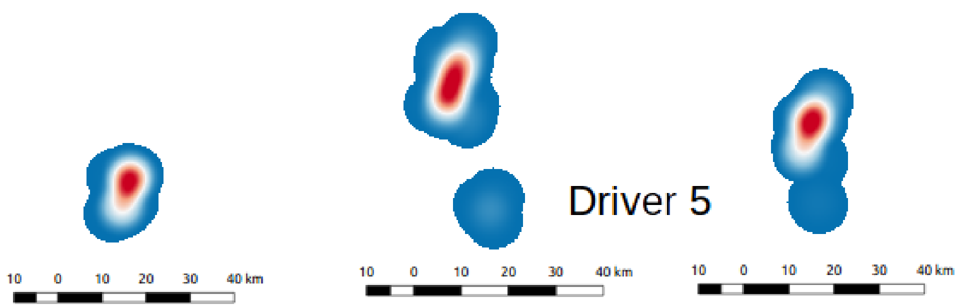

Figure A.2. Driver 5 heatmaps for arrival and departure locations through years 2009, 2010 and 2011 (left to right).

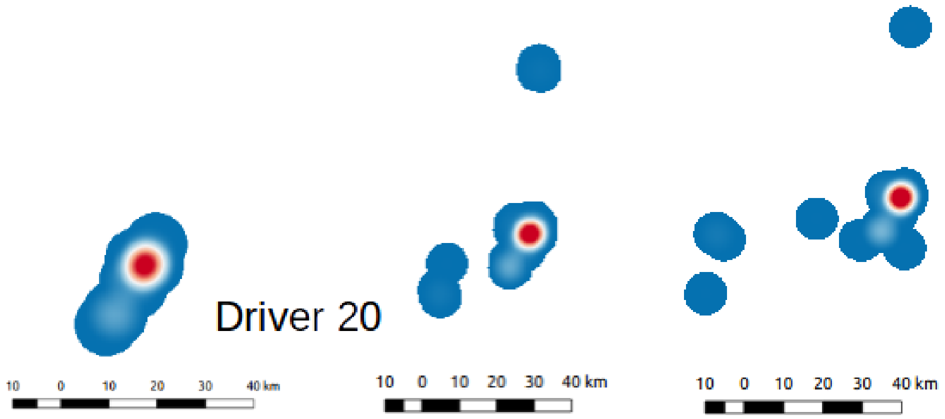

Figure A.3. Driver 20 heatmaps for arrival and departure locations through years 2009, 2010 and 2011 (left to right).

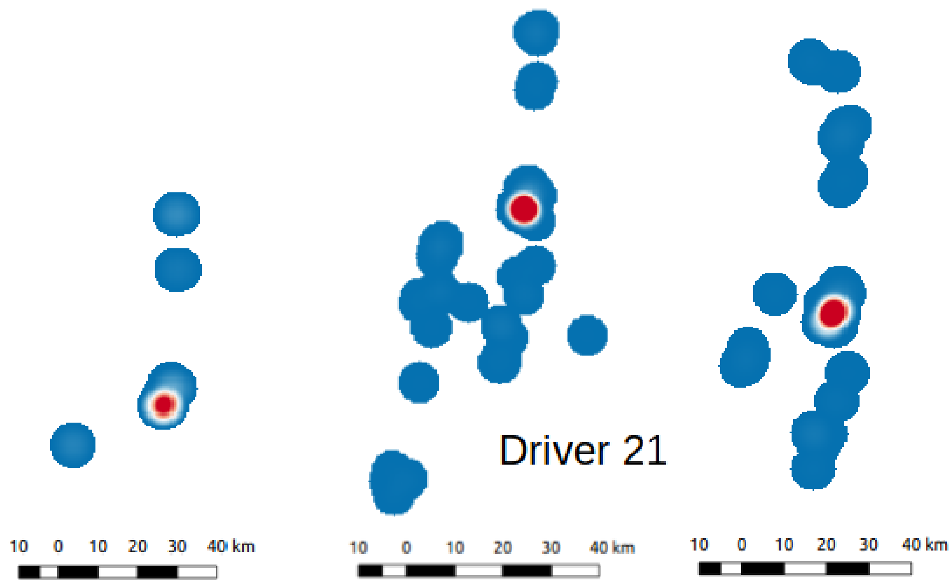

Figure A.4. Driver 21 heatmaps for arrival and departure locations through years 2009, 2010 and 2011 (left to right).

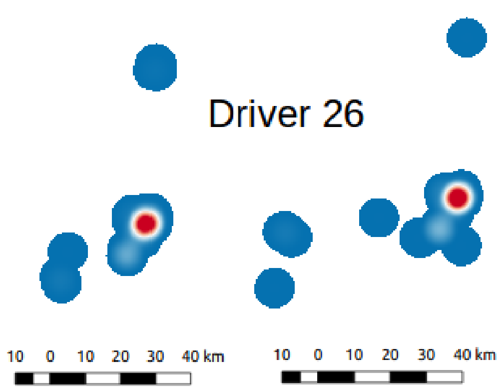

Figure A.5. Driver 26 heatmaps for arrival and departure locations through years 2009, 2010 and 2011 (left to right).

Figure A.6. Driver 3 total driving area (polygon) and primary "mean" coordinate locations (star) through years 2009 (green), 2010 (purple) and 2011 (blue).

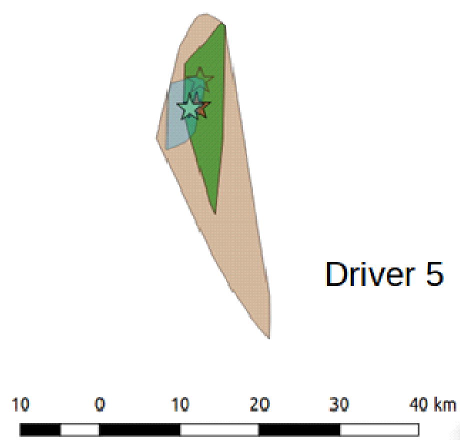

Figure A.7. Driver 5 total driving area (polygon) and primary "mean" coordinate locations (star) through years 2009 (teal), 2010 (brown) and 2011 (green).

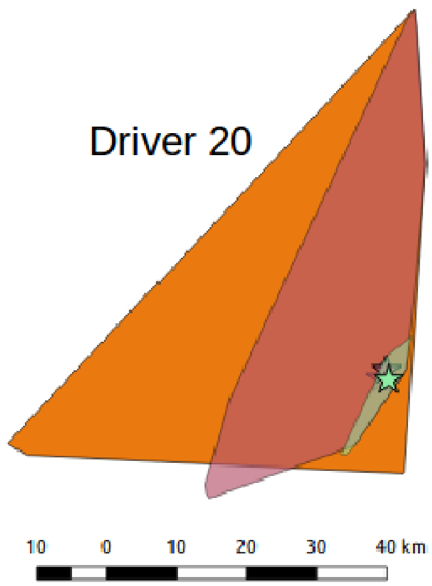

Figure A.8. Driver 20 total driving area (polygon) and primary "mean" coordinate locations (star) through years 2009 (teal), 2010 (purple) and 2011 (orange).

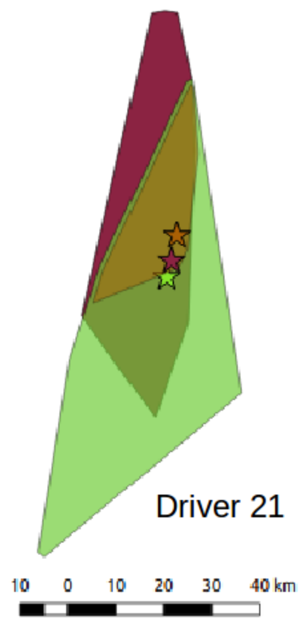

Figure A.9. Driver 21 total driving area (polygon) and primary "mean" coordinate locations (star) through years 2009 (brown), 2010 (green) and 2011 (maroon).

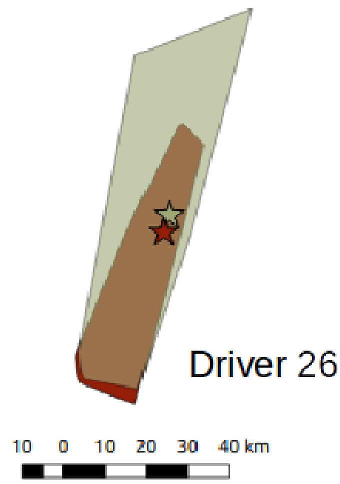

Figure A.10. Driver 26 total driving area (polygon) and primary "mean" coordinate locations (star) through years 2010 (grey) and 2011 (brown).