

The factors to get your car lost in Toronto and Vancouver

XiaoSheng (Mouder) Chen
20191125

1. Introduction

Thieves are every where. They may steal your wallet, your phone or your bag. No matter what kind of belonging get stolen, it will cause you trouble, big or small. Think about what makes you most trouble if you lose it. A key? Not really, you should always have a backup key. Money? No, it is unlikely that you will bring all of your money with you and get stolen. Your ID card? Yes, it could make you trouble, but you can still report loss and get a new card back. The answer could be your vehicle. If your vehicle gets lost, it could make you a lot of trouble even though the police can find your car back eventually or insurance company can compensate your lost.

1.1 Background

Why vehicles are stolen:

- **To sell the vehicle or parts** - Often thieves will take your car out of the Province or Country. Identification numbers may be altered or removed, ownership certificates forged, parts sold or the vehicle repainted.
- **For transportation** - Thieves will use your vehicle to get where they want to go and later abandon the vehicle. They may just drive it around to gain 'status' with their friends.
- **To commit other crimes** - Vehicles are stolen and used in other crimes such as robberies, break and enters, drug dealing or drive-by shootings.
- **For vehicle cloning** - Vehicles are stolen, given a fraudulent identification number and then sold to unsuspecting persons who are defrauded of their money. Fraudulent vehicles will be located and seized by the Police.

We can see there is a dark industry behind auto theft. As a normal citizen, what we can do is to stay alert in order to minimize the likelihood of losing your vehicle.

1.2 Problem

How can we avoid auto theft? If we are visiting a place that we know there could be higher chance for this kind of crime. We can be more careful or add extra protection when parking our car there. The problem is how we know if that area is safe or not? This assignment is to tell you what is the top ten venues to get your vehicle lost.

1.3 Target audience

The target audience is almost for everyone: e.g. citizen, traveller, owner of the venue, policemen or even government.

2. Data acquisition and cleaning

2.1 data source

Two big cities(Vancouver and Toronto) in Canada are chosen for this study. Fortunately, the auto theft data can be downloaded from the Vancouver police department website and Toronto Police service respectively. Our analyze will be base on the data from 2017-2018.

Vancouver police department : <https://vancouver.ca/police/>

Toronto police service : <http://data.torontopolice.on.ca/pages/open-data>

For the venues information, they can be retrieved from vendor source: Foursquare.

As a bonus section, this assignment will also try to consider if weather condition will impact the chance to lose your car.

2.2 data cleaning

Regarding the auto theft data of Toronto, it is in CSV(Comma Separated Value) and including the latitude and longitude for each incidents. As the data set contains more data than needed, extra records need to be filtered out and kept those data from year 2017-2018

Attributes

Chart • Map Visualization

<div><div><div><div></div><div></div><div></div></div><div></div></div><div>Division</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>event_unique_id</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>Hood_ID</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>Index</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>Lat</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>Long</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>MCI</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>Neighbourhood</div><div>Text</div></div>
<div><div><div><div></div><div></div><div></div></div><div></div></div><div>ObjectId</div><div>Unique ID</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>occurrence_date</div><div>Date or Time</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>occurrence_day</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>occurrence_day_of_week</div><div>Text</div></div>		<div><div><div><div></div><div></div><div></div></div><div></div></div><div>occurrence_day_of_year</div><div>Number</div></div>		
<div><div><div><div></div><div></div><div></div></div><div></div></div><div>occurrence_hour</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>occurrence_month</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>occurrence_year</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>offence</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>premise_type</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>reported_date</div><div>Date or Time</div></div>		
<div><div><div><div></div><div></div><div></div></div><div></div></div><div>reported_day</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>reported_day_of_week</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>reported_day_of_year</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>reported_hour</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>reported_month</div><div>Text</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>reported_year</div><div>Number</div></div>		
<div><div><div><div></div><div></div><div></div></div><div></div></div><div>ucr_code</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>ucr_ext</div><div>Number</div></div>	<div><div><div><div></div><div></div><div></div></div><div></div></div><div>SHOW FEWER</div><div>Attributes</div></div>					

For the auto theft data of Vancouver, it is also in csv format. However, it only provides the X , Y coordinate values which are projected in UTM Zone 10. That requires a conversion from UTM format to latitude and longitude format.

- **X:** Coordinate values are projected in UTM Zone 10. All data must be considered offset and users should not interpret any locations as related to a specific person or specific property.
- **Y:** Coordinate values are projected in UTM Zone 10. All data must be considered offset and users should not interpret any locations as related to a specific person or specific property.

For the venue information, it can be retrieved from Foursquare. However, there is 5000 requests limit per hour. So it need to split into 3 batches at least in order to extract all of the venue data.