Executive Report

‘TPS_Incident_Data_2010-2018’ Analysis

Zi Wang
#217095118

Introduction

The dataset "TPS_Incident_Data_2010-2018" was created by Toronto Police Service (TPS), focusing on the quantities, types, priority, and regions of incidents handled by TPS from 2010 to 2018. Through this data set, City of Toronto can have a clear view about the current volume of TPS's service activities, the future incident trend and the improvements needed for TPS to keep Toronto the best and safest place to be. Furthermore, they can have a better insight on the future budget on TPS.

Dataset Exploration & Preparation

The dataset contains 9 worksheets, which are classified by year. There are around 200,000 to 290,000 observations (rows) in each worksheet. Each observation represents an incident. In each worksheet, there are 6 variables (columns), which are 'ID', 'Dispatch_Time', 'Incident_Type', 'Priority_Number', 'Units_Arrived_At_Scene' and 'Forward_Sortation_Area'.

- 'Incident_Type' includes 'Airport Standy', 'Emergency Transfer', 'Fire', 'Medical' and 'Motor Vehicle Accident'.
- 'Priority_Number' covers 1, 3, 4, 5, 9, 11, 12, and 13.

The column name of 'Forward_Sortation_Area' is different in the 9 worksheets: some of them are 'Forward_Sortation_Area', and the others are 'FSA'. To avoid confusion, this column name is converted to 'Forward_Sortation_Area' in each worksheet.

There are several observations, which lack of 'Forward_Sortation_Area' information. They will be filtered during the incident-region analysis.

Moreover, there are several incidents happened at the midnight on December 31, 2018; therefore, the dispatch time is on January 1, 2019. They will be filtered for the incident-time analysis.

Since this dataset only covers the 'Forward_Sortation_Area', the postal code, an Ontario region dataset will be left joined to this dataset for a deeper analysis on relationship between incident types and region.

Dataset Analysis & Dashboard Content

The TPS dataset analysis will concentrate on two factors: the incident types and the regions the incident happened.

To have a basic concept on the dataset, the first dashboard will be an overview of the whole dataset, including:

- total number of incidents over 9 years;
- number of incidents by year and by month;
- number and percentage of incidents by type;
- number and percentage of incidents by city and by region;
- top cities / regions with high incident records;

Then, the second dashboard will be related to incident type analysis, which will cover:

- trend of each incident type by year from 2010 to 2018 and forecast on 2019;
- relationship between month and incident type;
- relationship between time and incident type;
- relationship between priority number and incident type;
- relationship between units arrived at the scene and incident type;

The second dashboard can provide different combinations of data through the filter of incident type.

Lastly, the last dashboard will show the incident rates of four regions: GTA, southcentral, southwest, and southeast, separately, and also the relationship between incident types and these four regions. Also, there will be a map showing the quantity of incident of each region. This map can be filtered by incident types, by years incident happened, and by months incident happened.

Recommendations

This dataset can be merged with the dataset of TPS locations and employees to see whether the number of employees in each location is reasonable and how to assign incidents to each TPS location to decrease the resource wastes.