Georgian

Amith Heiden, Karthikeyan Suresh Kumar, Srilekha Sampath Kumar

Hamilton

Analysis on Hamilton Public Parking Data

Problem Statement:

With rapid urbanization, public places such as parks, parking lots, etc., are being overshadowed by big buildings. This reduces their visibility (even signs don't help) in such a way that the public may have a hard time finding them. Underutilization causes loss of revenue to the city. We believe the analysis of available transaction data with more details about the parking lots can benefit both the public and the city.

Tools Used:

1) Microsoft Excel (with Data Analysis ToolPak add-in) for data manipulation and analysis
2) Real Statistics using Excel (XRealStats add-in from Charles Zaiontz, https://realstatictics.com, using Solver add-in) for further data analysis
3) Microsoft Power BI for data visualization
4) Microsoft PowerPoint for the presentation of data and analysis

Data Collection and Pre-processing

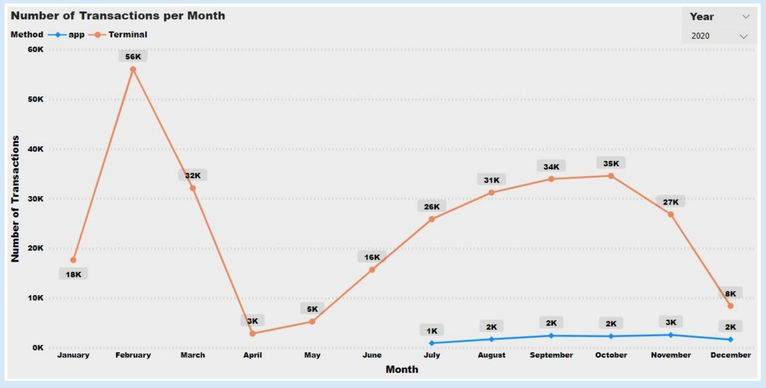
1) Parking lot data like community, Location, Minimum rate per hour rate, Daily rate, Monthly Rate were extracted from the Hamilton Parking lot map provided on the competition website.
2) Educational places, Religious places, Hospitals and Public places closer to the parking lots were identified using Google maps and the latitude and longitude were captured.
3) The population density around the parking lots was captured from a third-party map.
4) The number of terminals for the available parking lots was added along with the parking lots numbers from the terminals-dataset that were not present in the initial dataset.
5) Space count attribute has been added into the Parking lot spreadsheet which was separate initially.

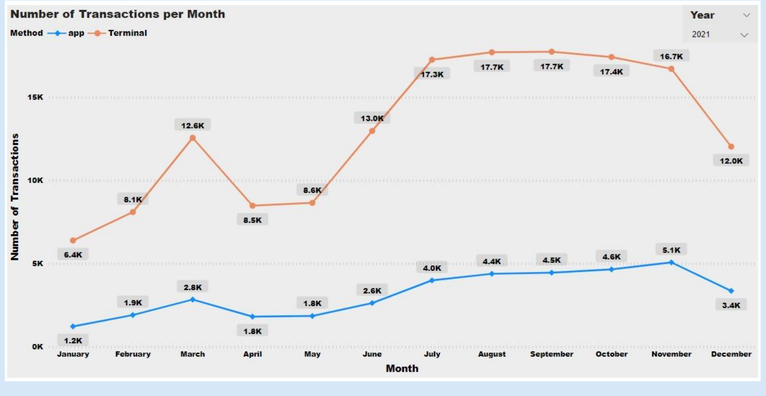
Frequency Analysis

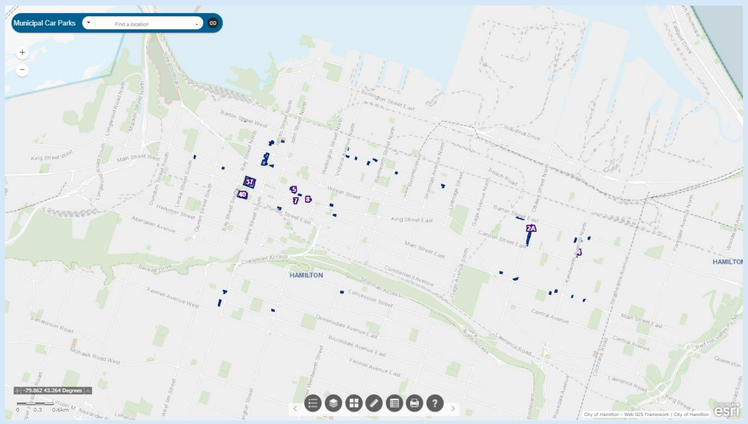
1) All six transaction datasets plus Terminal transaction data are combined into a single Excel spreadsheet.
2) Rate Name was split into several variables: Rate Name, Hourly rate, Working days, Operating hours, Flexible timing, and Full day charge.
3) Time Stamp was split into several variables: Entry Time, Exit time, Entry Day, Exit Day.
4) For cluster analysis, Entry Time, Exit time, Entry Day, Exit Day were converted to numerical values.
5) The duration column in minutes was converted into duration in hours.
6) A column named 'Total Charge' has been created by multiplying Duration * Hourly Rate
7) LotNo, Lot, Space Count, Street Name, and Population Density were added to transaction data from the parking lot list.

Does the city need to install more terminals in all parking lots?

Although four different methods (app, web, voice, and terminal) are available, terminals are by far the most widely used. Below, we can see that app-based transactions were introduced in Jul/20, during the pandemic. Initially, the use was low, but it has increased to about 20% of transactions. **Rather than installing terminals, this trend shows that the city should invest in improving and promoting the parking app.** This not only saves money but also makes it easy for the customers to make transactions from their phones.



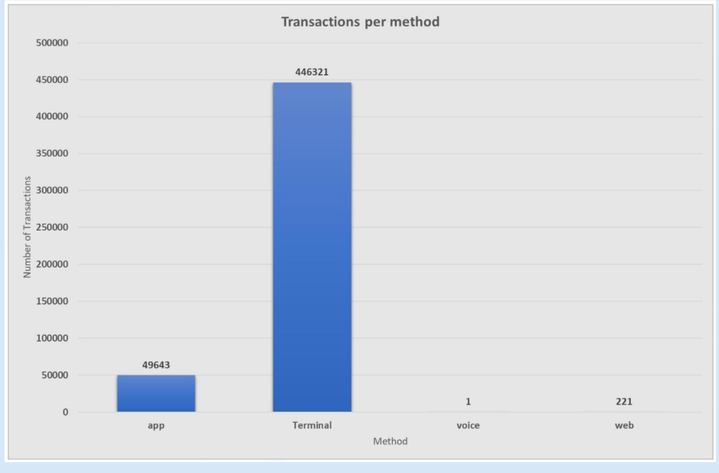




Breaking News!!!! No one is using Voice or Web payment!

City can save money by discontinuing these modes of payment.

The money could be used to help pay for other improvements.



References:

Data, problem and questions from https://datacompetition.mohawkcollege.ca/2022-data-links/

Population density map from CensusMapper, https://censusmapper.ca/#14/43.2438/-79.8560

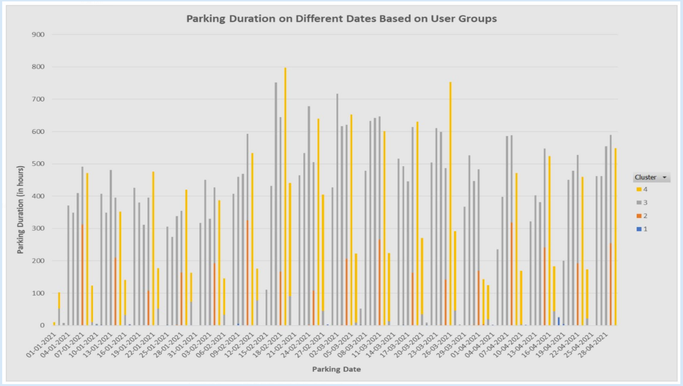
Cluster Analysis

1) Transaction data was split in 6 groups of 4 months each: Jan-Apr, May-Aug, and Sep-Dec in each year.
2) K-Means Nearest Neighbour Cluster Analysis was performed on New Entry Time, New Entry Date, New Exit Time, and New Exit Date; we tried to identify four clusters of parking lot users by day of the week and time of use.
3) Due to processing limitations, only the Jan/21-Apr/21 group was chosen for cluster analysis, and we were unable to try other numbers of groups or validate cluster separation using MANOVA.

Clusters are separated largely by day of entry. Cluster 1 included Sun, Mon, and Tues arrivals during business hours. Cluster 2 arrivals were mostly Thurs daytime, with some on Wed afternoon. Cluster 3 arrivals were Mon-Thurs daytime and evening with some on Saturday. Cluster 4 arrivals were on Friday and Saturday daytime and evening.

This grouping of customers could be useful in identifying patterns in parking lot usage.

The city could increase revenue by creating loyalty schemes to attract and retain parking lot customers.



The total parking hours of the different clusters show us how efficient the parking lot would be with certain groups of customers. If the pattern shows a particular user group often using the parking lot, then it might be helpful if the parking lot can be expanded, or several new parking lots can be constructed with the behavior of those groups that can be found across Hamilton. **Big data analytics using more powerful tools and processes with machine learning techniques would allow the city to tighten their focus on customer groups using cluster analysis.**