CMPT 353 Final Project

Natalie Woods
Alvin Tsang
Github Repo

Problem Statement and Project Scope

In the age of digital money, many people only carry a credit card with them instead of the inconvenience of loose change. With this in mind, we investigate the following question:

> *Do parking meters with a credit card payment option reduce the number of parking tickets caused by non-payment in Vancouver? If so, can we predict the parking meters that are at a "high-risk" of receiving a parking ticket from non-payment?*

Here, we define "non-payment" as a parking meter not being paid or an expired parking meter.

Using geospatial parking ticket data and meter attributes, we quantify the effect of credit card convenience on compliance, identify high-risk locations for intervention, and provide actionable recommendations for urban policy.

Data Collection

1. **Parking Tickets Dataset**
    a. Parking tickets issued in the Metro Vancouver area between 2020 and 2024
    b. Parking tickets — City of Vancouver Open Data Portal
2. **Parking Meters Dataset**
    a. Locations of city-owned parking meters in the Metro Vancouver Area
    b. Parking meters — City of Vancouver Open Data Portal
3. **Local Area Boundary Dataset**:
    a. Geospatial information of the boundaries between Neighbourhoods of Metro Vancouver
    b. Local area boundary — City of Vancouver Open Data Portal
4. **Latitude and Longitiude Given Block and Street**
    a. The parking tickets dataset and parking meters dataset did not have a common key to combine the two datasets cleanly. Our workaround was to use Nominatim and OpenStreetMap to identify the latitude and longitude of each parking ticket issued and use it to determine the closest parking meter.
    b. Latitude and Longitude of Block + Street
    c. Collection script

Data Cleaning

1. **Parking Ticket Dataset**
    a. Isolated parking tickets caused by a violation of Bylaw 2952, Section 5(4)(A)(ii) or Section 5(4)(B) — payment expiration for pay stations or meter head flashing zeros
    b. Filter for tickets that were issued only

c. Inserted the latitude and longitude values from each parking ticket's issued block and street
d. Identify the neighbourhood in which the parking ticket was issued with the Local Area Boundary dataset

2. **Parking Meter Dataset**
    a. Removed irrelevant columns (i.e. payment for different times, max parking time for different times)
    b. Extract the latitude and longitude coordinates from the geometry data

---

Visualizations and Initial Observations

Before diving into modeling, we created spatial visualizations to identify patterns in ticket distribution across Vancouver:

1. **Geo Local Area Boundaries Map**: To orient ourselves geographically, we visualized the official neighborhood boundaries of Vancouver (Figure 1). We also annotated each area with its name, helping us connect later results back to recognizable regions.

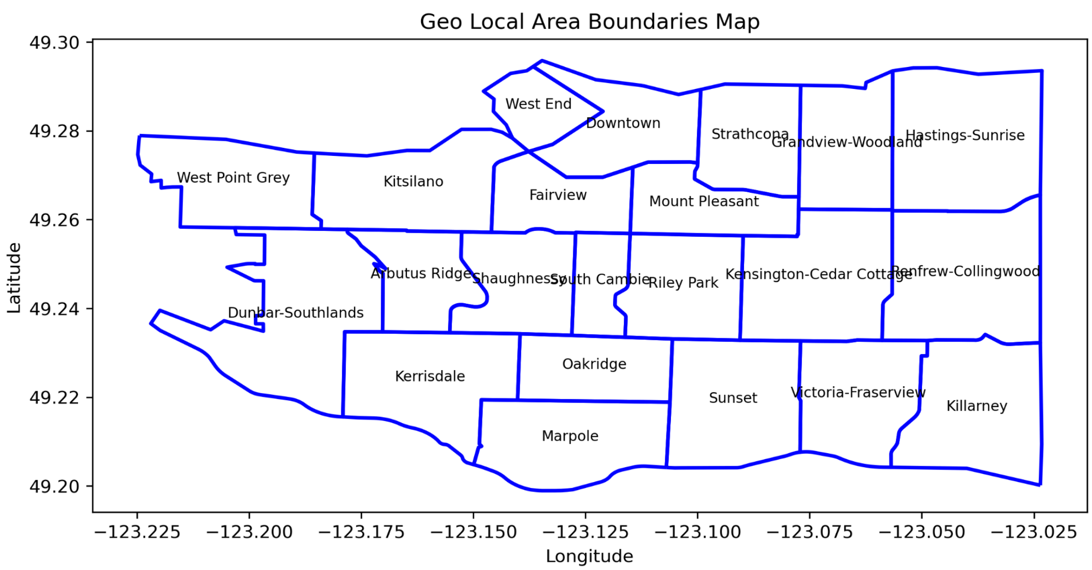

Figure 1

2. **Parking Meters by Area**: We mapped the parking meters onto the boundary map layer, with red representing parking meters with credit card options available and green representing parking meters without credit card payment available (Figure 2).

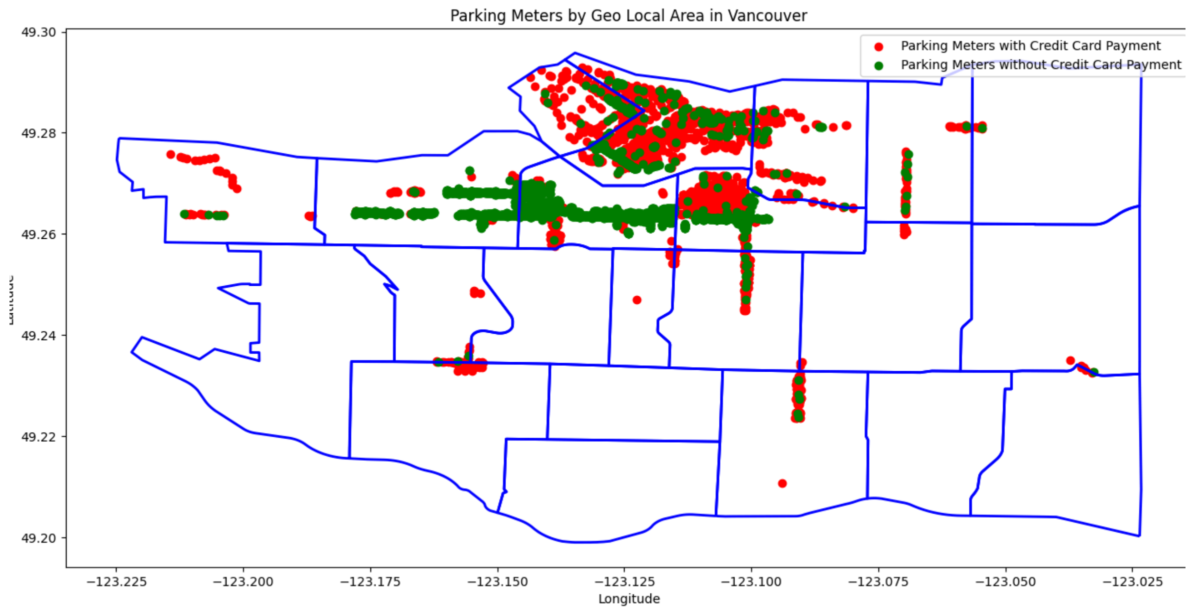

Figure 2

3. **Parking Tickets by Neighborhood**: We grouped ticket counts by neighborhood and plotted a choropleth map to visualize ticket density (Figure 3).

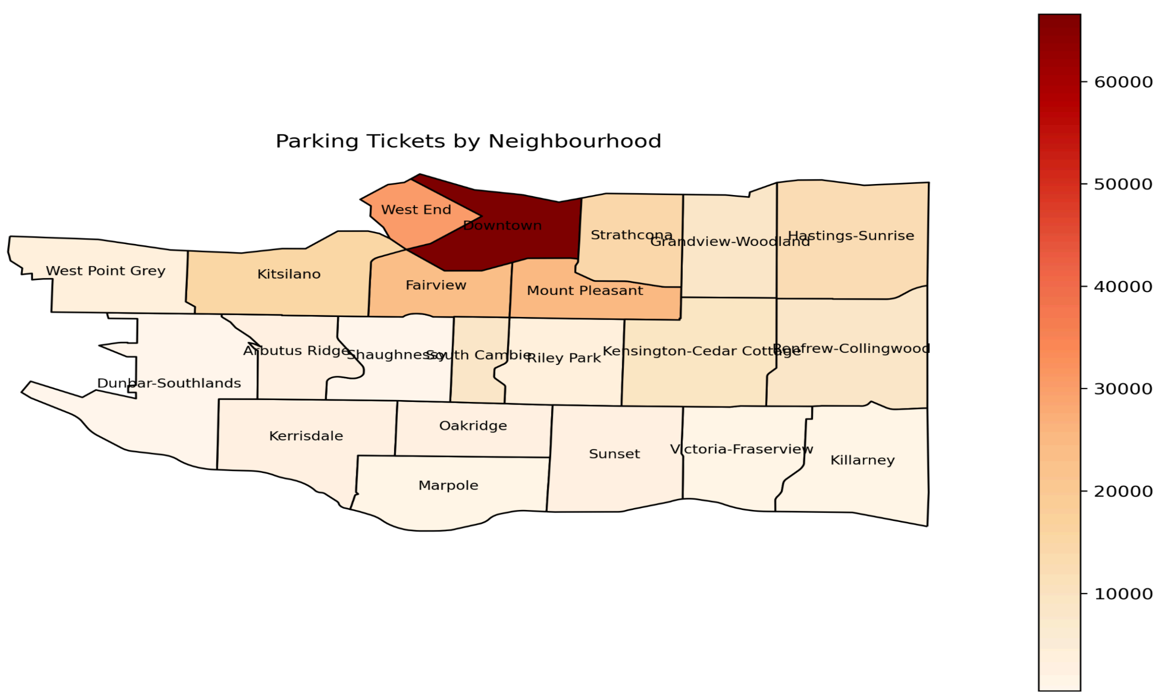

Figure 3

The next few visualizations are neighbourhood-specific to provide a sense of what the dataset looks like. Specifically, the Downtown neighbourhood was used here.

4. **Tickets by Day of the Week**: We grouped tickets by day of the week to examine weekly trends. Figure 4 revealed which days had the highest ticket volumes, helping us anticipate when enforcement is most active.

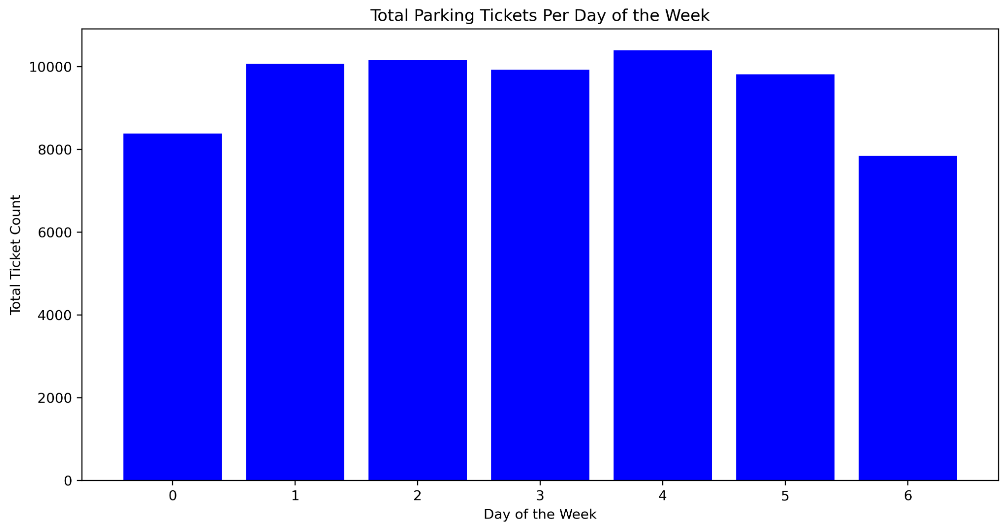

Figure 4

6. **Heatmap by Day of Week and Street**: Figure 5 shows how ticketing volume varies by both street and day of week. It highlights consistent enforcement patterns across specific streets.

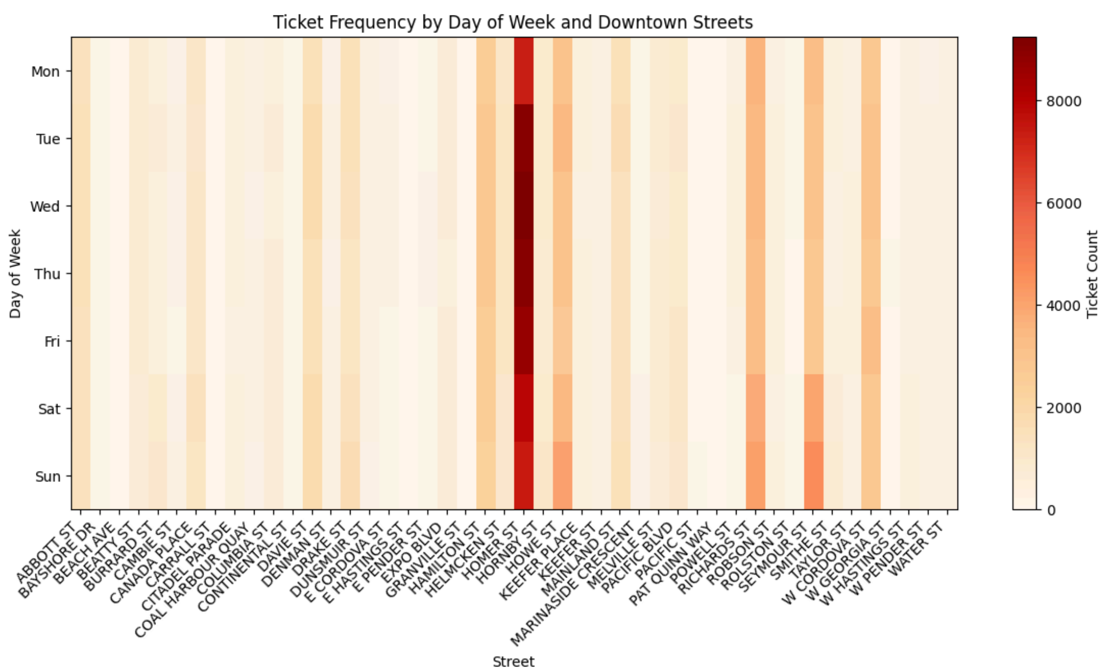

Figure 5

Analysis and Modeling Approach

The source code for the modeling and analysis can be found here.

1. **Feature Selection**
   a. *CREDITCARD*:
      i. This is our main variable of interest. This feature is a binary value (0 for False, 1 for True) indicating whether or not a credit card payment option is available at the associated parking meter
   b. *Street, Neighourhood*
      i. The location of each entry
   c. METERHEAD
      i. The type of meter captured (i.e. pay station, twin)
   d. *dayofweek*:
      i. The temporal patterns throughout a week

2. **Statistical Testing**
   a. Before blindly building a model, we want to verify that there is evidence suggesting our hypothesis using the features above. To do this, we first look at the average number of parking tickets received between the two types of parking meters
      i. *Mean tickets for credit card meters*: 3.8
      ii. *Mean tickets for non-credit card meters*: 5.6
   b. Indeed, there is some (but incredibly weak) evidence suggesting that parking meters with credit card payment available receive fewer non-payment parking tickets. We also performed an independent t-test and a Mann-Whitney U test. We will say that there is a statistical significance for $p < 0.05$. Below are the results
      i. *t-statistic*: -8.33187700606638
      ii. *t-test p-value*: 9.244768566984192e-1
      iii. *Mann-Whitney U test p-value*: 1.3492288493980663e-29
      iv. Both p-values are less than 0.05, suggesting that there is a statistical significance between parking tickets and parking meters with credit card payment options.

3. **Predictive Model**
   a. With some support suggesting our research question is correct, we move onto creating a predictive model to identify "high risk" areas for parking tickets and see if a credit card option has any effect.
   b. To quantify the predictive power of a credit card payment option versus other features, we used a gradient-boosted tree. After training and testing the model, we gathered the following results:
      i. *Overall Accuracy*: 89%
         1. *Not High-Risk Prediction Precision: 90%*
         2. *High-Risk Prediction Precision: 87%*
      ii. *Overall Recall*: 89%
         1. *Not High-Risk Recall: 87%*
         2. *High-Risk Recall: 90%*

4. **Feature Importance**
    a. Analyzing the importance of each feature, we note that *CREDITCARD* ranks 29/233 features with an importance value of 0.010087. This shows that *CREDITCARD* is a semi-important feature, but features like neighbourhood are even more important to this classification.

Results

1. From statistical testing, we see that there is a significant statistical difference between parking tickets due to non-payment at credit card-enabled parking meters and parking tickets due to non-payment at parking meters without a credit card option.
2. From our predictive model and feature analysis, we realize that information about a parking meter's credit card payment ability can help predict whether or not a non-payment will lead ot a parking ticket. However, features like the parking meter neighbourhood and street have more influence.

Limitations and Future Work

1. From our definition of "non-payment", we considered a parking meter not being paid and an expired parking meter to be the same. However, our research question is looking at the more strict definition of "non-payment" to only be the former option.
2. Our mapping between a parking ticket and a parking meter is a rough estimate. Our estimation looks for the closest parking meter within 0.1m of the parking ticket location. From fiddling with our model, we noticed that the credit card importance would change depending on the value set during the mapping.
3. The current model only uses a handful of features. It would be worthwhile to investigate other features and how they affect our model and analysis.

Project Experience Summary

1. **Natalie Woods**
    a. Summary
        i. Co-designed project scope and selected datasets from the City of Vancouver.
        ii. Implemented key data-cleaning scripts in Python, including spatial transformations and feature engineering.
        iii. Led the visualization efforts using Matplotlib.
        iv. Participated in model validation and report writing.
    b. Accomplishment Statement
        i. Co-led project scoping and dataset selection. Developed core data-cleaning pipelines in Python, including spatial joins and temporal feature engineering. Trained RandomForest models and created data visualizations using Seaborn and Matplotlib. Contributed to model validation and collaborative report writing.. Implemented key data-cleaning scripts in Python, including spatial transformations and time-based feature engineering. Led the visualization efforts using

Seaborn and Matplotlib. Participated in model validation and report writing to support urban planning and parking enforcement.

2. **Alvin Tsang**
    a. Co-led a geospatial analysis project from conception to implementation, utilizing Python (Pandas, Scikit-learn & GeoPandas) and machine learning to evaluate how payment methods affect parking violations ($p < 0.0001$)
    b. Engineered a unified data pipeline integrating 3 city datasets (parking tickets, parking meter local area boundaries), performing spatial joins and feature engineering to enable street and neighborhood-level insights
    c. Co-designed the statistical approach, implementing both parametric (t-test) and non-parametric (Mann-Whitney U) tests to validate findings with 99.99% confidence
    d. Developed a Gradient Boosting Classifier that identified high-risk parking meter locations with 89% accuracy