E-scooter Share Pilot Program in Chicago

Background & Motivation:

Shared e-scooter services, such as Lime and Bird, first appeared in the US in 2017 in California. Since then, it has been introduced in many cities around the world to reduce traffic congestion and air pollution and enhance community relations as well as the resilience of urban transport, which are important assets for sustainable urban transport. In addition, the use of shared electric scooters as an alternative to public transport for short trips is increasing during the COVID-19 pandemic. Different from bike sharing, the scooter is not limited by the dock and benefits for its small size for short trips. Users simply download an app and find a nearby scooter to start riding. Once at destination, users can leave the scooter anywhere. Scooter sharing can be a great way to help communities connect to transit stops. For residents who are less likely to live within walking distance of a bus or train, shared e-scooters and e-bikes could fill the missing link, giving people a way to reach transit stops previously only accessible by expensive cars. In 2019 and 2020, the city of Chicago conducted a four-month scooter sharing pilot to evaluate the feasibility of Chicago scooter sharing, test technology and rules, and better understand how to form a more durable program. I only focus on the 2020 pilot program here. The time is from August to December. There are three companies involved in the program. Compared to the 2019 program, there is a priority area added to 2020 pilot program.

Research Question: Who currently has access to E-shared scooters?

Focus on the equity implications of the systems: who currently has access to scooters, and who will have access if we keep following the business-as-usual approach. As these vehicles rapidly integrate into cities, the city clearly needs to forecast the potential demand for e-scooter travel. If users don't have a walkable scooter when they want to use it, they're likely to give up looking. To improve service levels, we should think about how to forecast demand so that we can effectively redistribute scooters to ensure that everyone can use them. Whether there is an imbalance in the demand for locomotives, resulting in unfair rights of residents to use locomotives. In this study, I use Chicago 2020 scooter trip data to construct a model for predicting ridership for this year.

Literature Review:

Kim's article studied a demand prediction model for the use and development of shared electric scooters in Seoul, South Korea, in October 2020. The space unit of analysis is set as a grid of 200 square meters, and the hourly demand of each grid is summarized according to the electric scooter travel data. The community structure approach was used to cluster the spatial areas into five communities before predicting the needs. They built a demand prediction model based on long and short-term memory (LSTM) and compared the prediction results based on activation function.

Ursaki et al. compared the social and economic characteristics of U.S. Census Bureau block groups within and outside bike-sharing service areas in seven cities based on the American Community Survey. By creating a 500-meter buffer zone around each site in ArcGIS, the location of the bike-sharing site was used to define the bike-sharing service area. Using the student T-test to compare socioeconomic characteristics within and outside bike-sharing service areas, significant differences in visits based on race and income variables were found in Boston, Chicago, Denver, Seattle, and New York City.

Data Source and Feature Engineering :

To estimate the e-scooter trip forecast model, the build environment, social-economic, and demographic data are used. Data source, the e-scooter trip data is from the Chicago open data portal. And data for build environment features is from open street map. And also used 2017 American Community Survey (ACS) to obtain demographic and economic statistics. These variables would reflect both the underlying demand for scooters in a census tract and the likelihood that a provider would make more vehicles available in a tract. I chose census areas as the units of spatial analysis because they represent the highest level of geographic aggregation in the scooter passenger data set. Compared to 2019, the e-scooter trips decline from 710,000 to 630,000. It is mainly due to the covid-19 pandemic.

For build environments, mainly chose the following 9 features (Retail Stores, Restaurants, Leisure, Tourism, Public Transportation, Offices, College and Universities, Bike path) [Exhibit 1]. During the feature engineering process, I tried four different variations of the built environment variables. Density, Count, KNN: The distance from the tract centroid to the nearest k location (here I used 5 nearest), Ratio. I chose KNN for the final panel because KNN displayed the greater correlation with user pickups in each census tract.

For demographic variables, I used total population, median age, percentage of white population and percentage of female population.

For socio-economic variables, I chose these 11 features (Household Income, Home Values, Median Rental Prices, Commute Mode share, Commute Distance, Housing Units, Occupancy Rates, Vehicle Ownership, The number of jobs located in this tract, the number of workers who live in this tract). The final panel including total 23 features and I defined a function to count the origins or the destinations of trips in a census tract. I made some correlation plots and a correlation matrix with the final features, but for the most part, the explanatory variables show little collinearity [Exhibit 2, 3].

Whether there is an imbalance for Scooter and whether it will give residents an unfair right to use the scooter. vehicle flow imbalance happens when riders take more vehicles from one place than other riders bring in. but we could see the inflow and outflow maps that the trips mainly happened in some same limited community [Exhibit 4]. A big problem in the dataset is that most census tract do not have shared trips happened. There are nearly 800 census tracts in

Chicago, but there are less than 80 tracts with shared trips origins, which leads to a large number of zeros in the final panel.

Model Building---Modeling Strategy

We used the final set of features to construct several models that predict raw trip counts in each census tract.

Linear Model: an OLS linear regression

Random Forest: an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

XGBoost: another tree-based ensemble method. Unlike random forest, which creates many trees at once and aggregates their results at the end, this method builds trees iteratively, employing boosting in the bagging process to address large prediction errors in previous trees.

Poisson model: is log-linear model. particularly useful for predicting counting data. Since our data is counting data with lots of zeros and the predicted residuals are not normally distributed, Poisson might be a good choice.

And then split the data into a 75%/25% training set and test set, within the 75% training set, use a grid search to select the optimal hyperparameters. To improve the robustness of the grid search, used random k-fold (k=20) cross-validation within the training set. Finally, test these models on the 25% test set, compare mean absolute error (MAE) & root mean square error (RMSE) to select final model framework.

Model Performance:

The plots show the average RMSE produced by each modeling framework [Exhibit 5]. We see here that Poisson, the model I finally chose because it had the lowest MAE on the test set, had the higher RMSE. One interpretation of this is that the model was predicting well for many census tracts compared to the other frameworks, which yielded a lower MAE, but it may also produce a few predictions with very large errors that drove up the RMSE, which is more sensitive to outliers than MAE. In general, the errors were quite high. This reflects the difficulty of using a small sample set with many 0 values to make predictions of raw ridership counts.

Prediction:

The mean of the prediction on test set is much higher than the actual mean trip counts. And I did a second step to include data only with the non-zero data. And did the prediction on test set again, the prediction is much closer to the actual situation. We can also see from the error map that the overall absolute error is pretty high and especially higher in some census tract and trend to overestimate the distribution of scooter riders in the city [Exhibit 6]. Used the model I chose, the prediction map on this year is not very good result [Exhibit 7]. The forecast of scooter trips is still very uneven. I think it has a lot to do with my use of the 2020 data. In the original data, many census tract did not have records of e-scooter data, which seriously affected our final prediction results. Most predicted trips are still concentrated in the city downtown area. There is also a certain connection between the company's initial placement in the urban area. Because the scooter tends to travel on short distances, if the vehicles are not evenly distributed in every place at the beginning, the vehicles will always be concentrated in certain areas. For future works, I may try another city or different year to see if the model will have a better result.

References:

Kim S, Choo S, Lee G, Kim S. Predicting Demand for Shared E-Scooter Using Community Structure and Deep Learning Method. *Sustainability*. 2022; 14(5):2564. https://doi.org/10.3390/su14052564

Ursaki, Julia, Aultman-Hall Lisa. Quantifying the Equity of Bikeshare Access in U.S. Cities. TRC Report 15-011. 2015. https://rosap.ntl.bts.gov/view/dot/36739

Exhibits:

Exhibit 1 Visualization of built environment

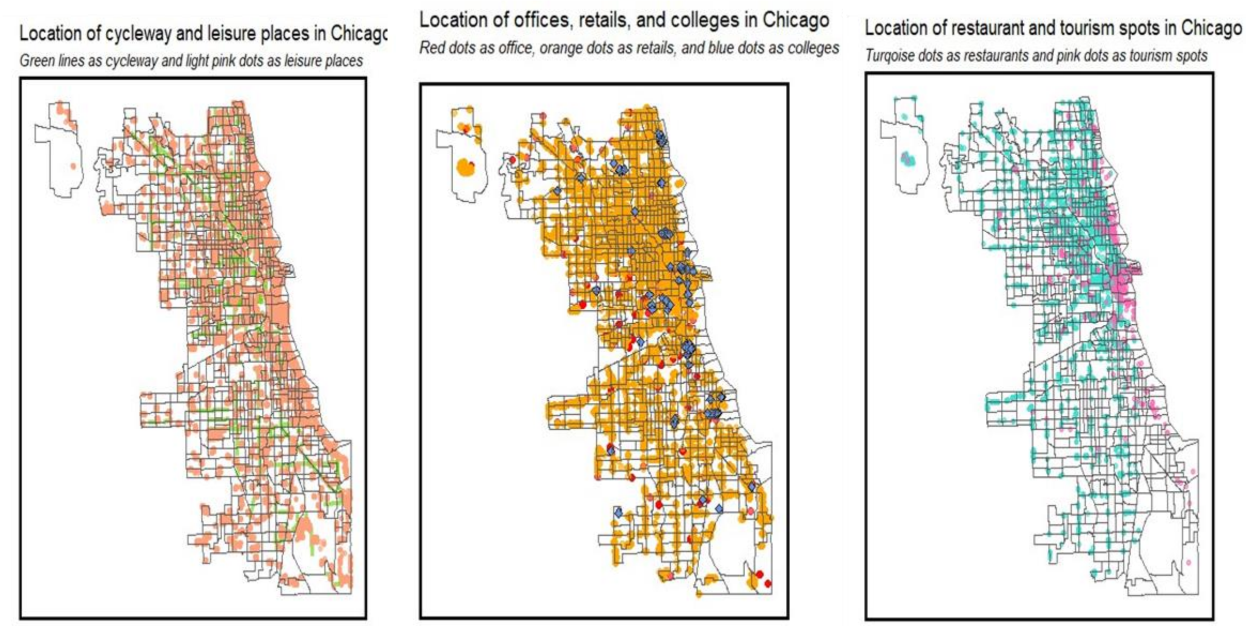


Exhibit 2.1 Correlation Plot--- Socio-economic factors

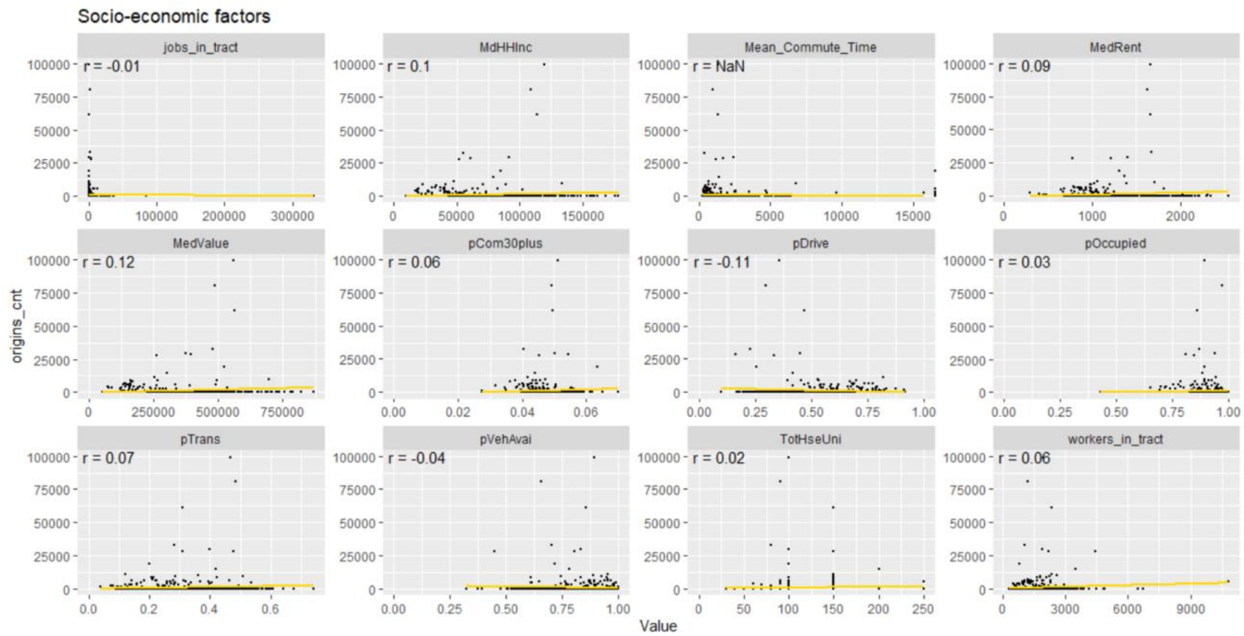


Exhibit 2.2 Correlation Plot--- Demographic factors

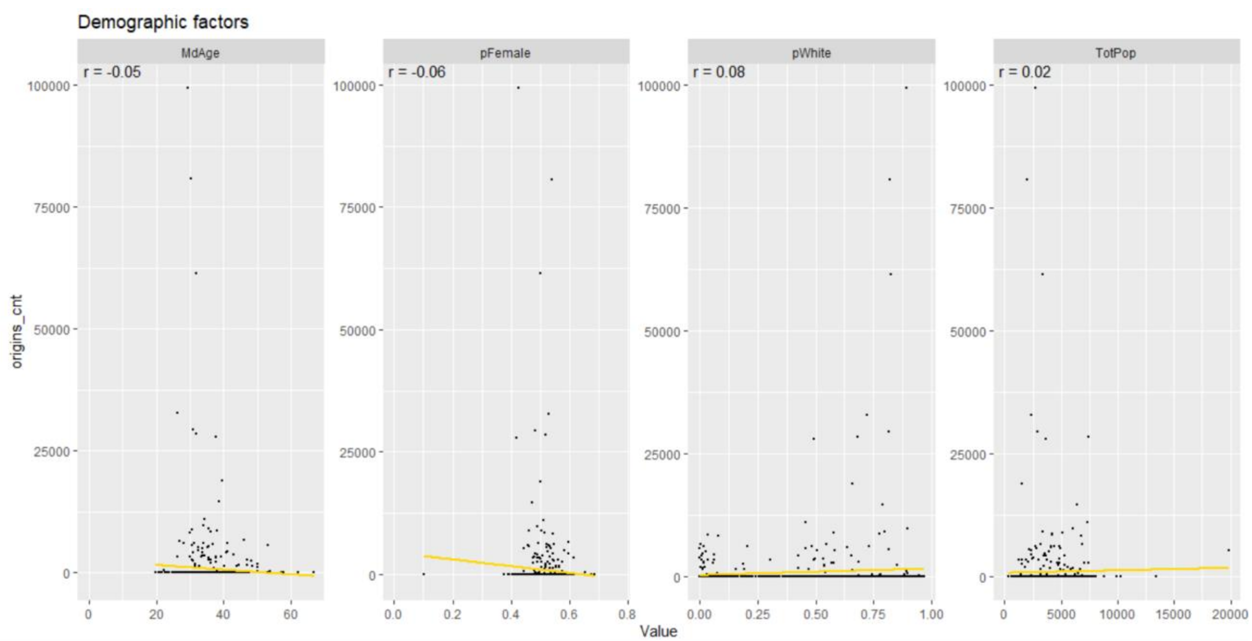


Exhibit 2.3 Correlation Plot--- Build Environment factors

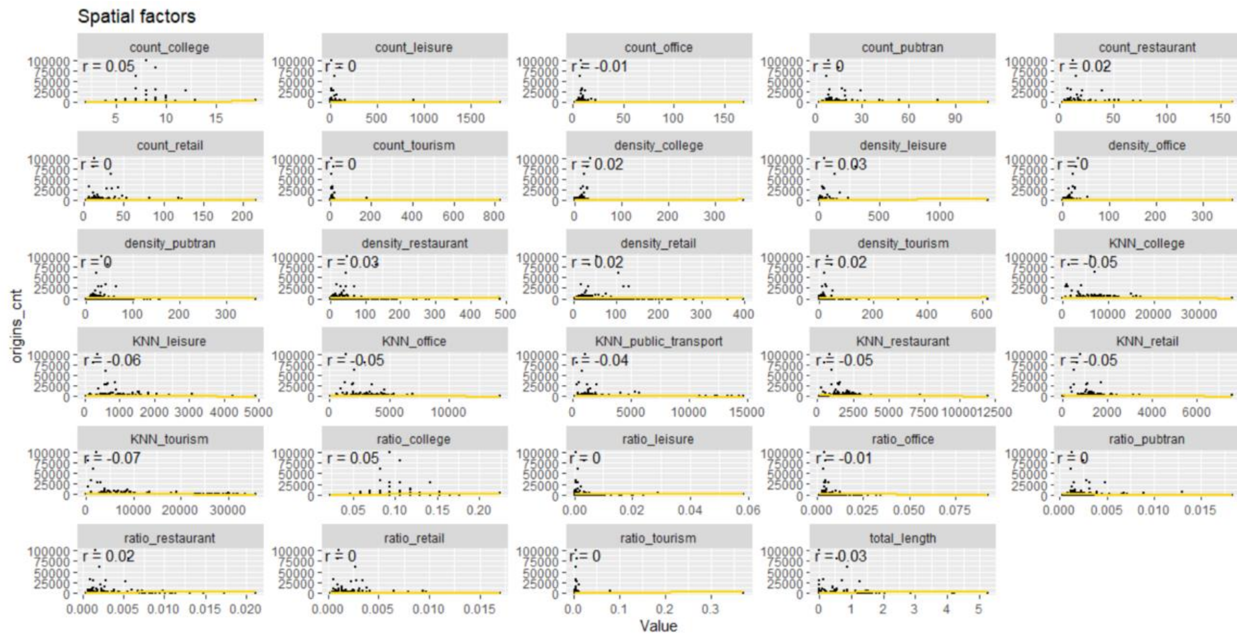


Exhibit 3 Correlation Matrix

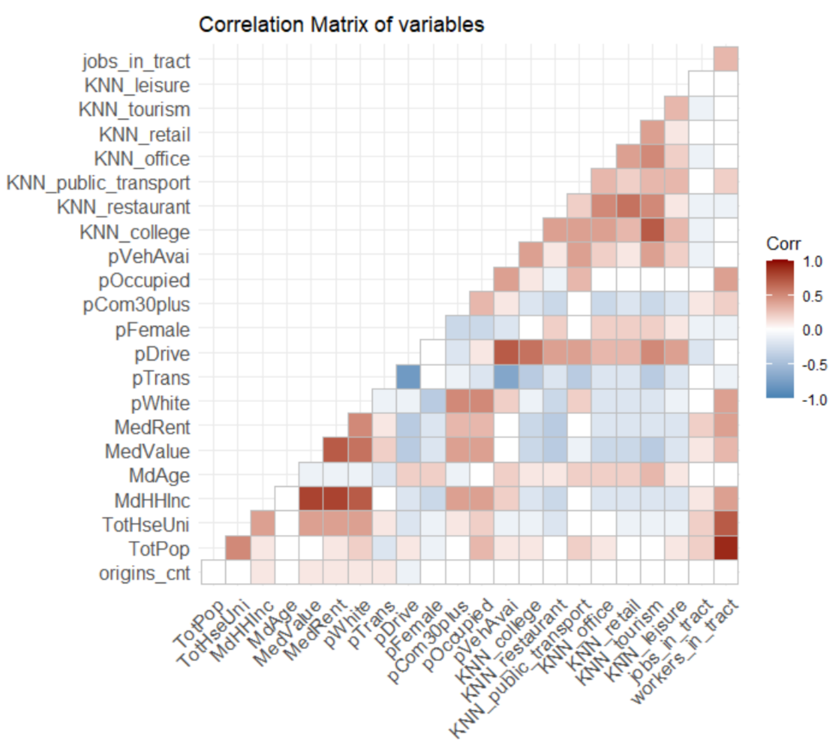
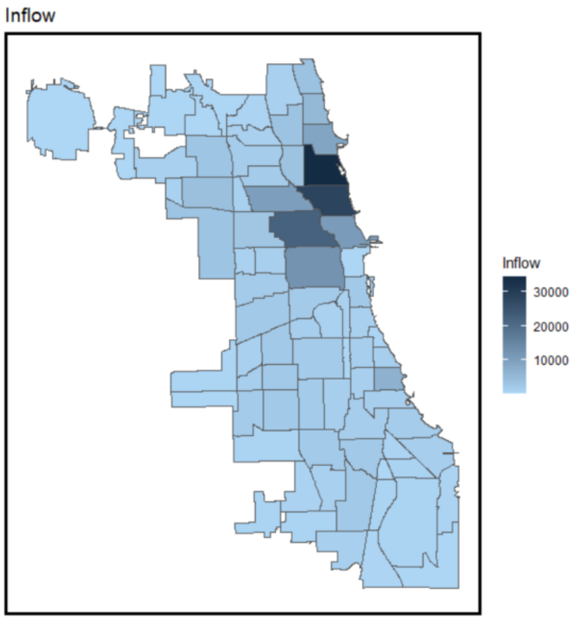


Exhibit 4 Inflow and Outflow of Scooter Trips



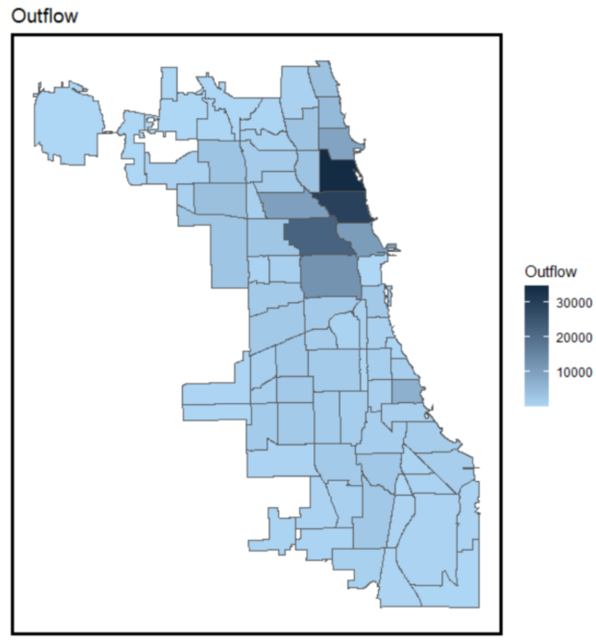


Exhibit 5.1 MAE by Model type

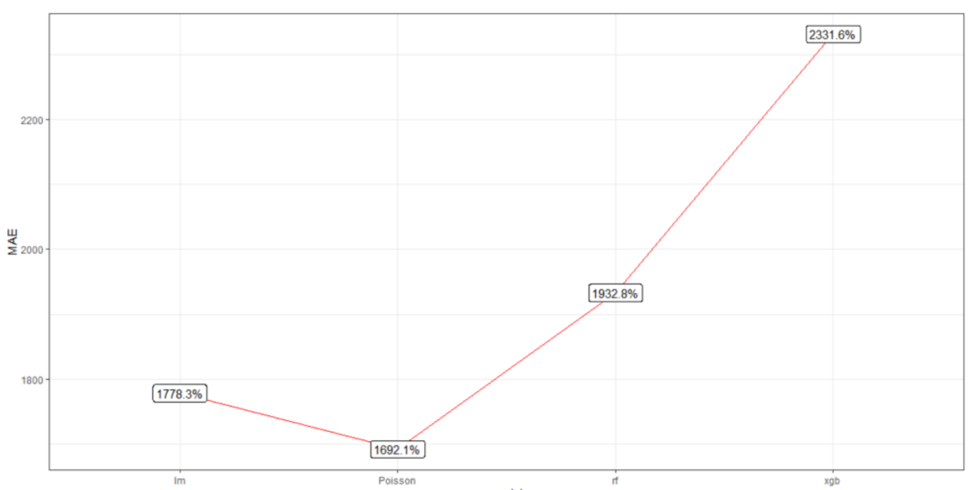


Exhibit 5.2 RMSE by Model Type

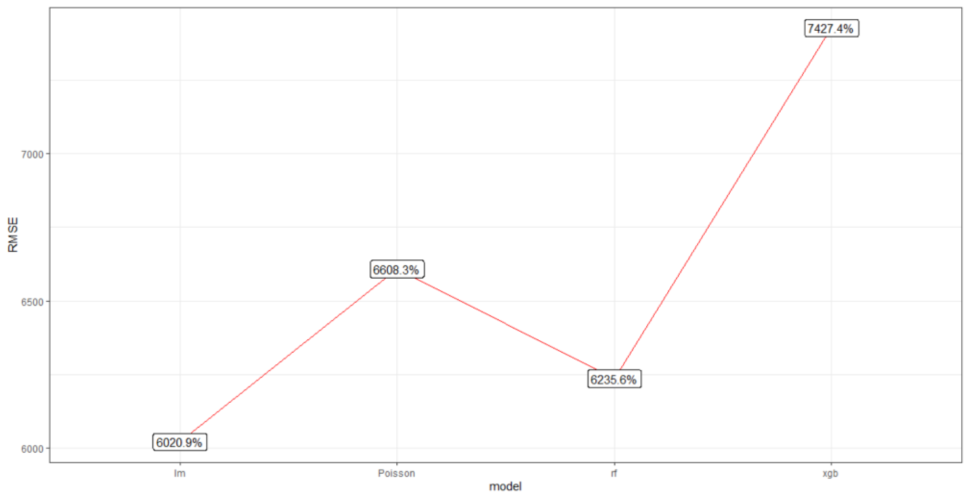


Exhibit 5.3 Actual vs Predicted Prediction

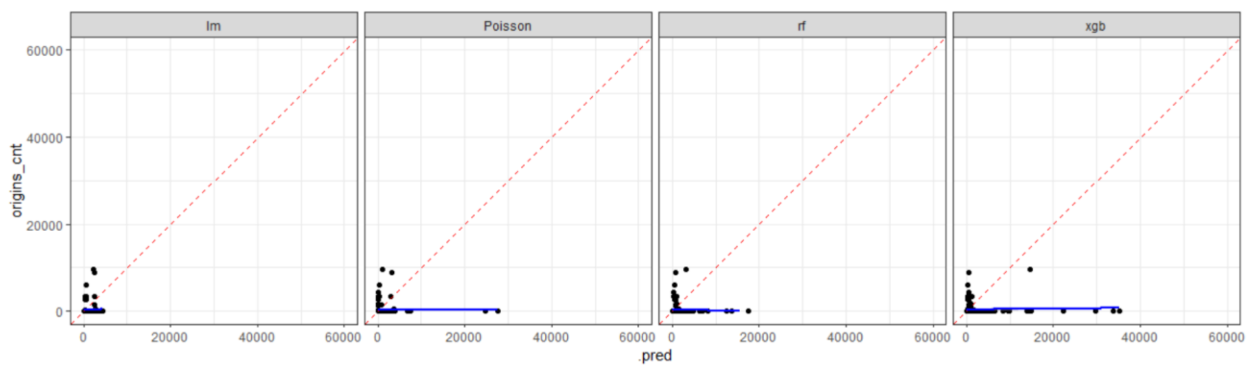


Exhibit 6 Error Map

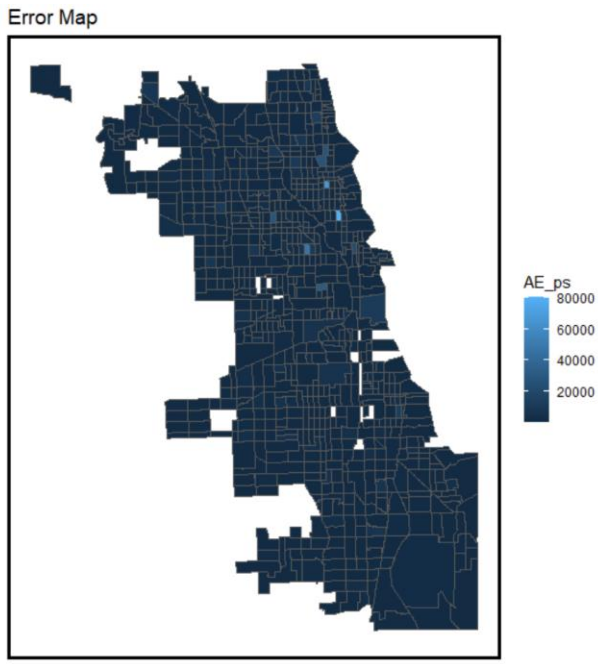


Exhibit 7 Final Prediction