AZAVEA SUMMER of MAPS

Clean Air Council

An Analysis of Bike Parking Demand in Center City Philadelphia

Final Report

CLEAN AIR COUNCIL

Prepared by

Jianting Zhao

azavea

990 Spring Garden Street, 5th Floor

Philadelphia, PA 19123

(215) 925-2600

http://www.azavea.com

Table of Contents

Executive Summary

Clean Air Council is a member-supported environmental organization serving the Mid-Atlantic Region. The Council is dedicated to protecting and defending everyone's right to breathe clean air. In this project, the Council initiated a crowdsourcing effort to collect workday bike parking data in Center City Philadelphia. With the data, the Council will better understand the landscape of bike parking across Center City.

This project is composed of two parts. The first part is a visualization of the bike parking data through different aspects. In the second part, Azavea constructed linear regression models to predict for bike parking demand. The study shows that land use, employment, distance to bike lanes and average interpolated Indego bike share trips contribute the most to the predict for bike parking demand. The study concludes with recommendations for areas to add new bike racks and a cross validation for each of the statistical methods.

Project Overview

In the first section of this report, there will be a series of maps showing the bike parking data through different perspectives. The second section is the construction of the prediction model. Two finalized prediction models will be addressed. Following this section is the limitation section, where I will discuss the limitations of this study. In the Preliminary Conclusion section, I will highlight areas that need additional parking installation. Lastly, the possible next steps will be discussed.

Data Collection

Previous studies have collected bicycle parking data, but this was not available at the time of this project. Therefore, the Clean Air Council initiated a crowdsourcing effort to collect workday bike parking data, aiming to recreate the data that the City of Philadelphia once had. In just 3 days, 12 volunteers collected nearly 4,000 data points on Fulcrum, a data collection web and mobile application (Figure 1). When volunteers collected the data, they added attribute information based on: the type of the parking, and the location of the parking and rack. Figure 2 shows examples of formal and informal parking that we used to guide the data entry.

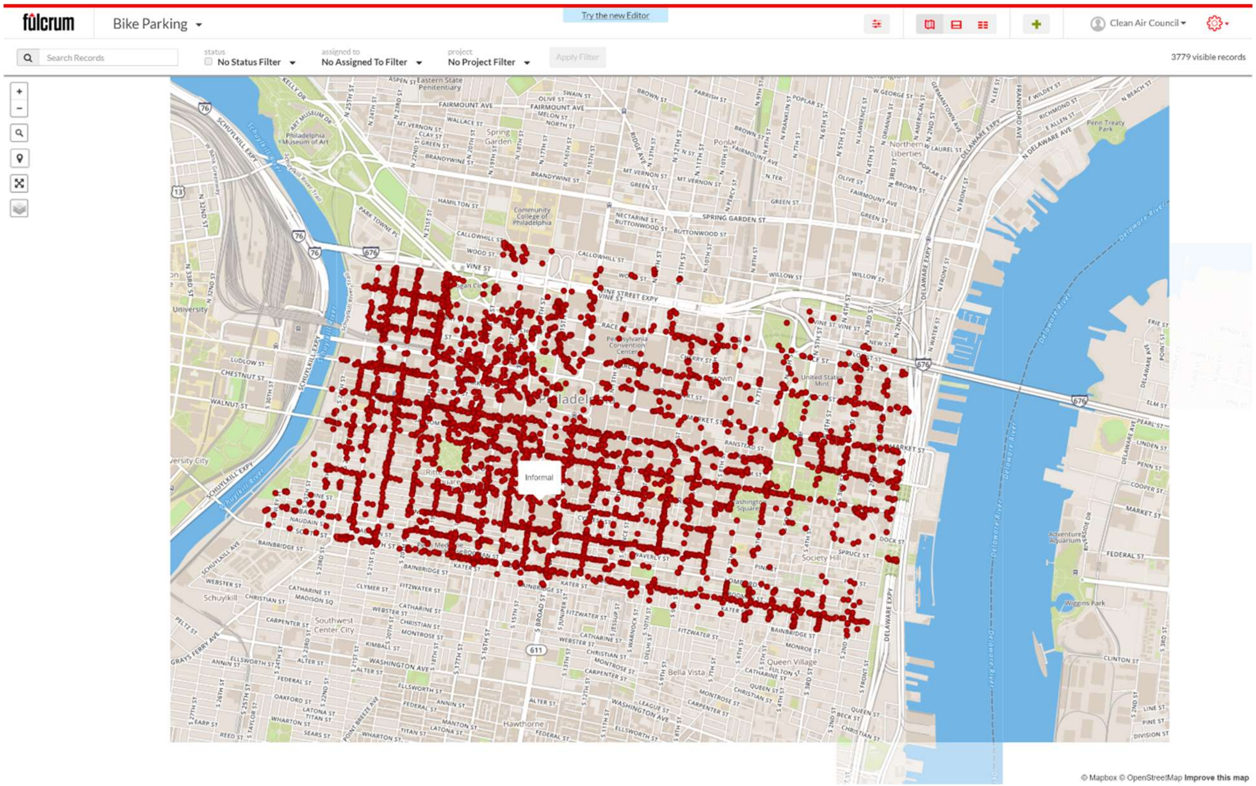


Figure 1. Data Collection on Fulcrum

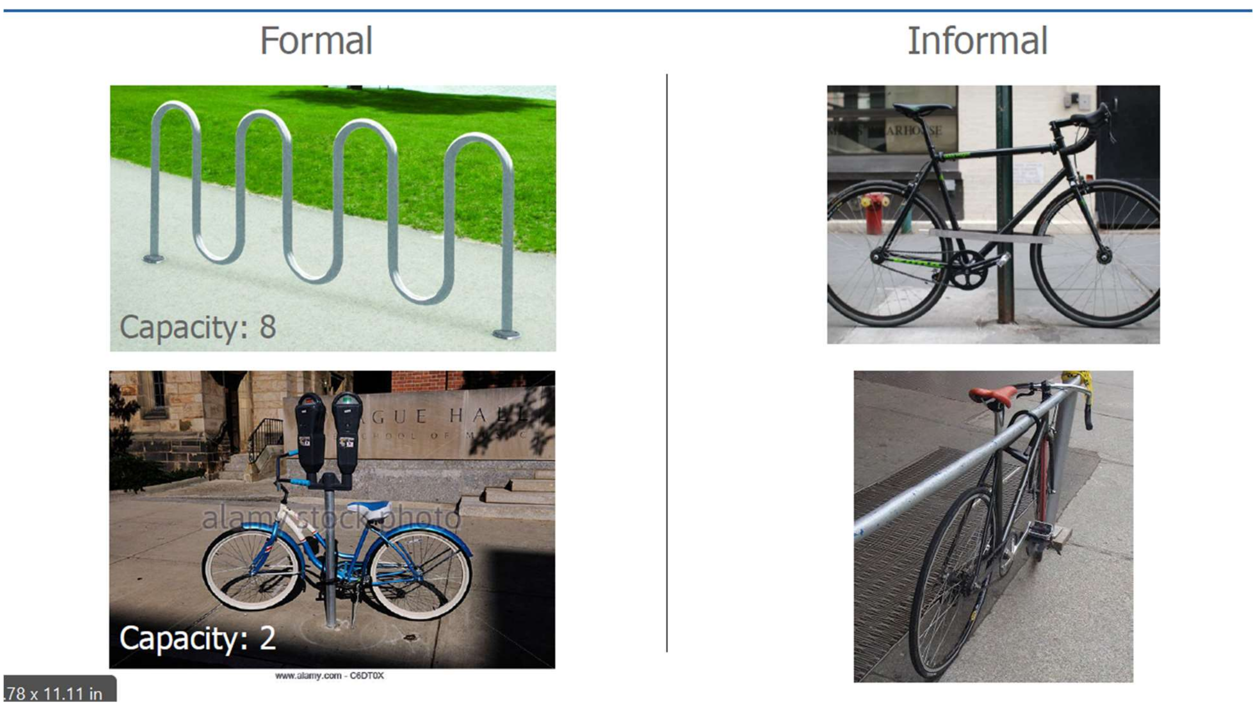


Figure 2. Formal and Informal Parking Examples

With the data on bike parking facilities, the Council wants to answer three questions:

1. How is bike parking demand relevant to factors like employment, land use and transportation?
2. Where do we need to increase the number of bike racks?
3. Is there a formula to predict work day bike parking demand?

This analysis will incorporate two parts to address the questions from the Council.

1. Visualize the current bike parking demand, predicted demand, and different factors.
2. Create a formula to guide recommending bike parking installation.

Visualizing Bike Parking Data

Figure 3 provides an overview of the crowdsourced bike parking data. There are clusters of bike parking in the blocks between Logan Square and City Hall. In contrast, there is an absence of bike parking and bike racks in the Old City area between Second and 5th street.

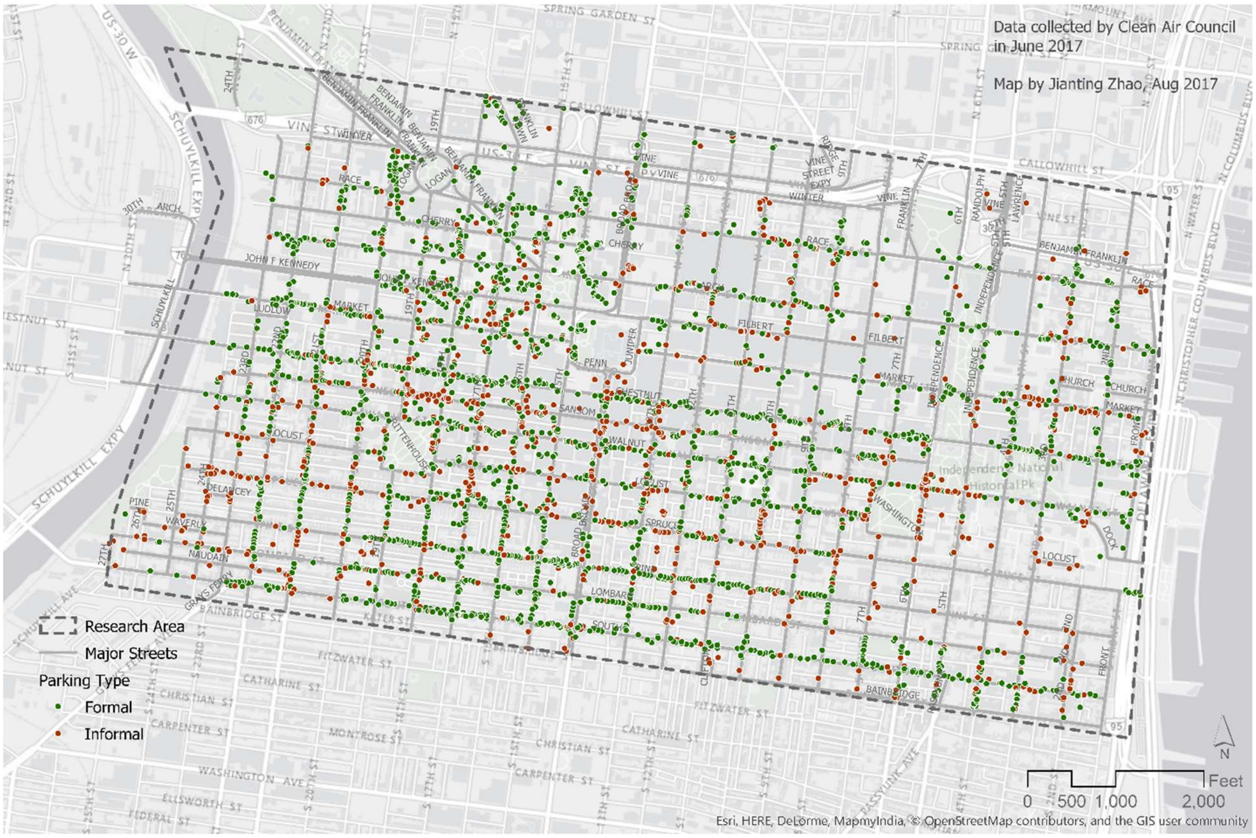


Figure 3. Original Bike Parking Points

Due to the imprecision of data collection via users' mobile phones, the raw coordinate point data had too much noise. Therefore, I needed to aggregate them in some way such that I can bind each point with a unit of analysis. I chose to aggregate to blocks. They are precise enough for our purpose of study but also easy to reconcile with other datasets, such as census data.

Figures 4 to 7 present different aspects of the parking data. Figure 4 shows the total use aggregated to blocks. The total use is the sum of informal and formal bike parking. The blocks where the Municipal Services Building and Lubert Plaza are located have the most amount of total use.

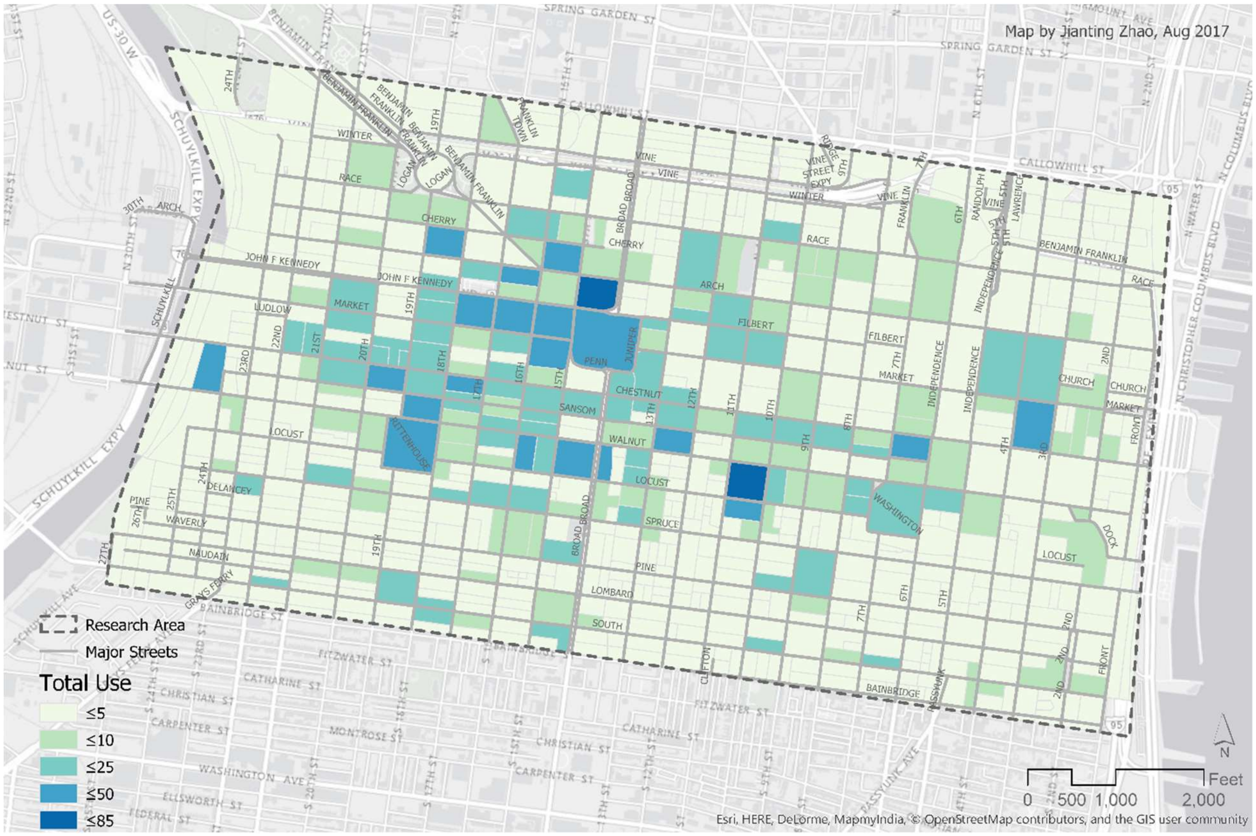


Figure 4. Total Use Aggregated to Blocks

In the histogram, it is apparent that there are lots of blocks with zero bike parking (chart 1).

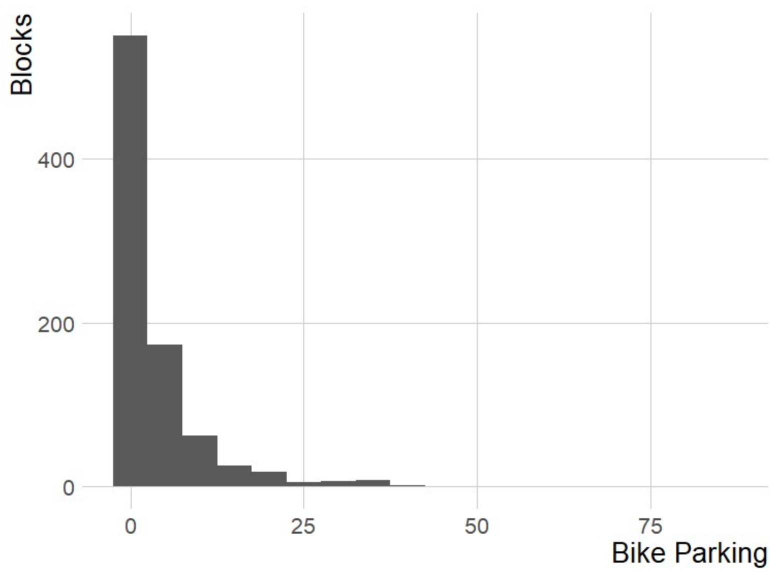


Chart 1. Total Use Histogram

Figure 5 shows the number of formal parking in each block.

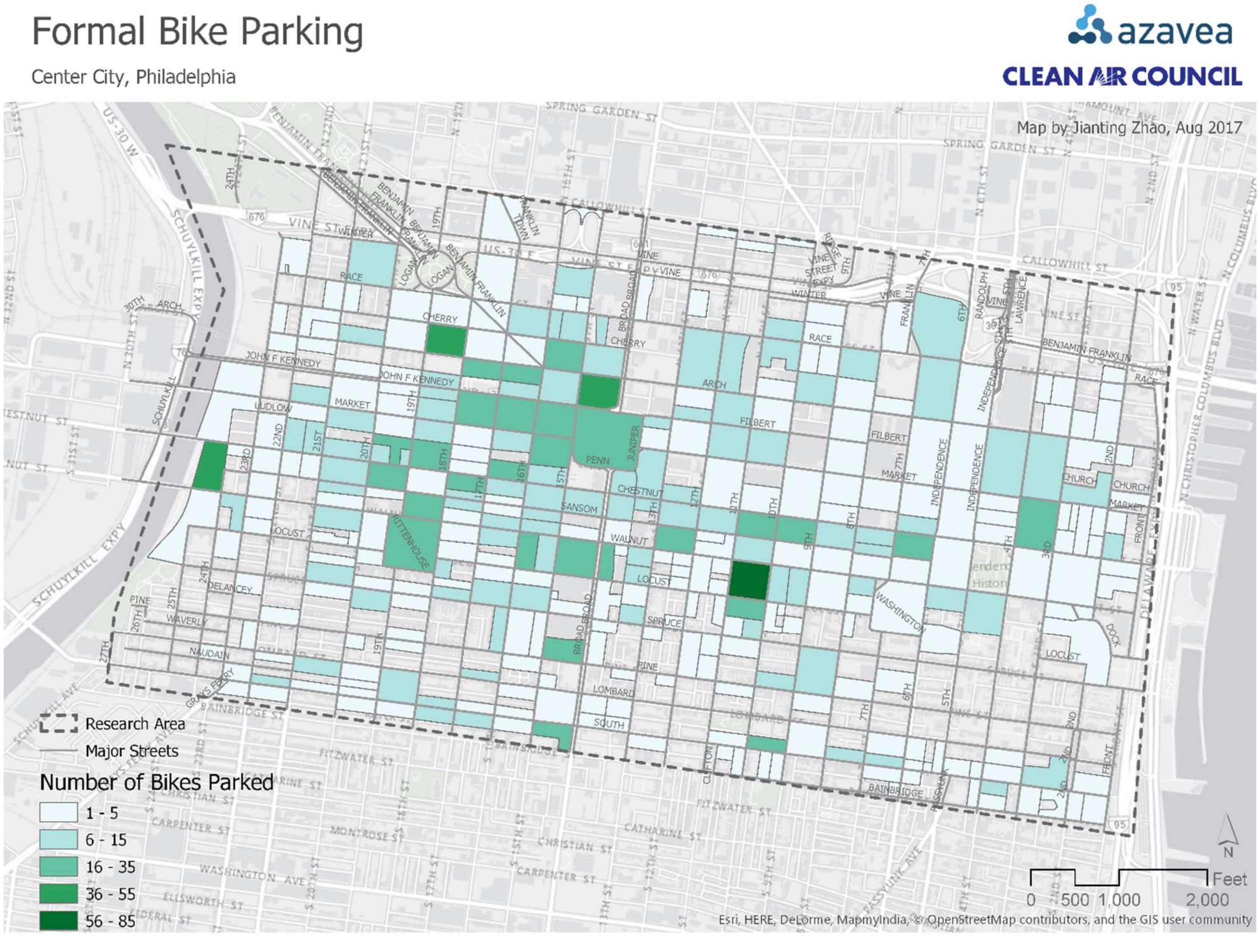


Figure 5. Formal Parking per Block

Chart 2 is the histogram of formal parking. Similar to chart 1, there are lots of zero values.

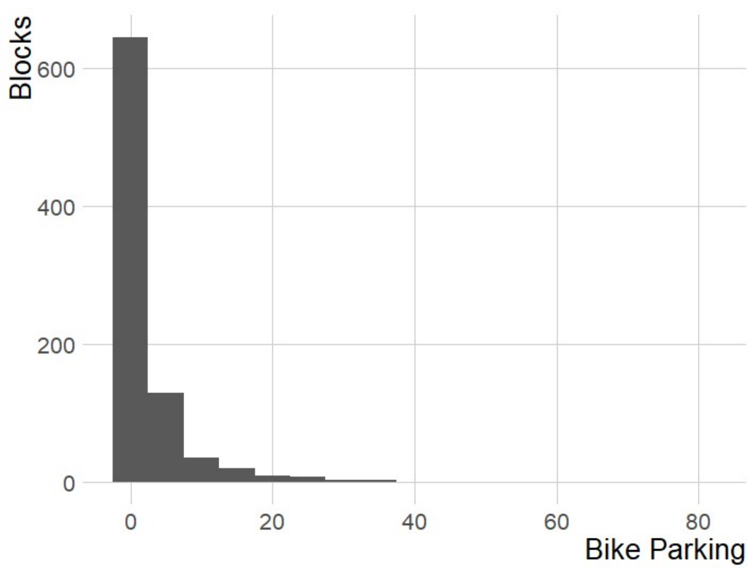


Chart 2. Formal Parking Histogram

Figure 6 shows the number of informal parking in each block.

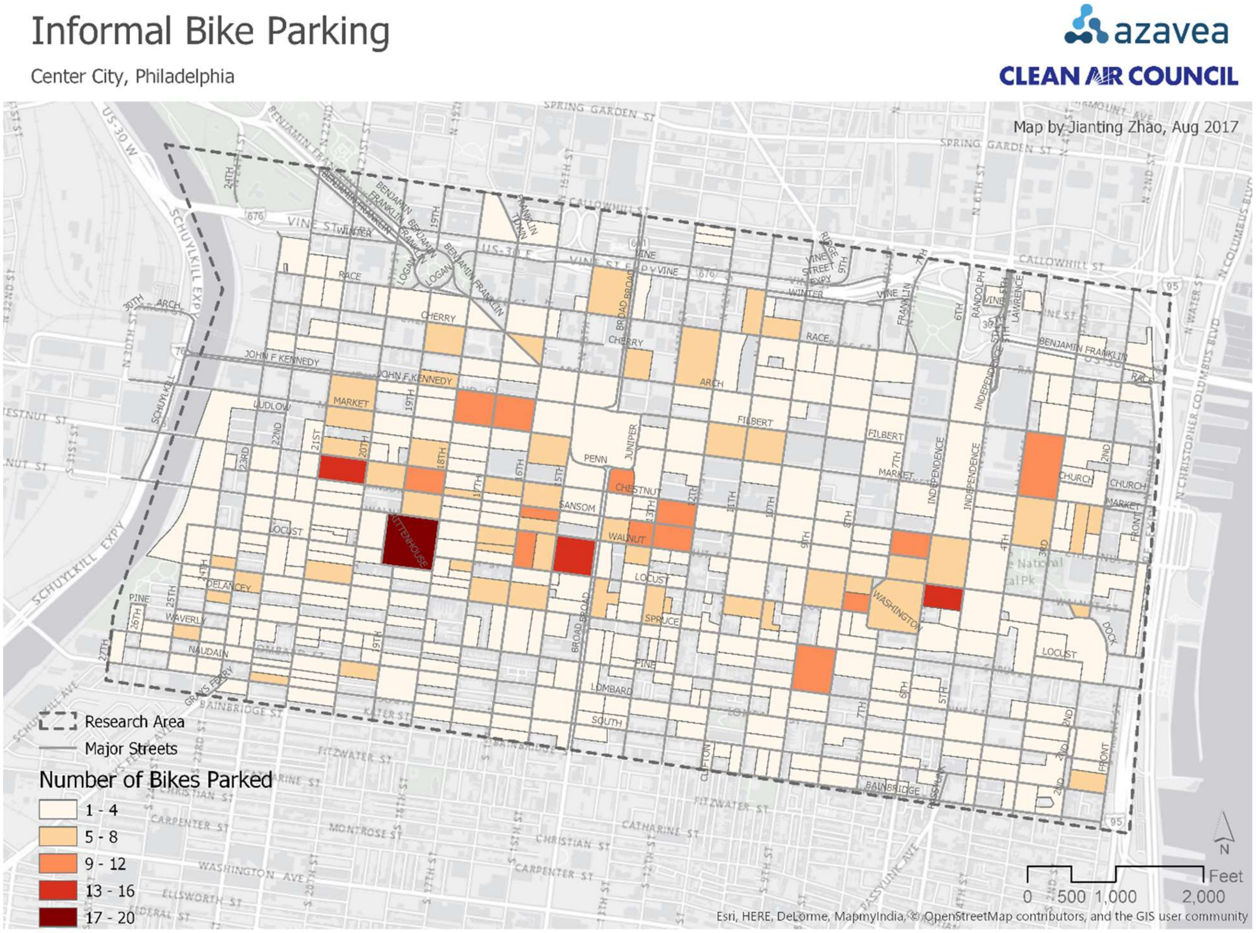


Figure 6. Informal Parking per Block

Chart 3 shows the histogram of informal parking.

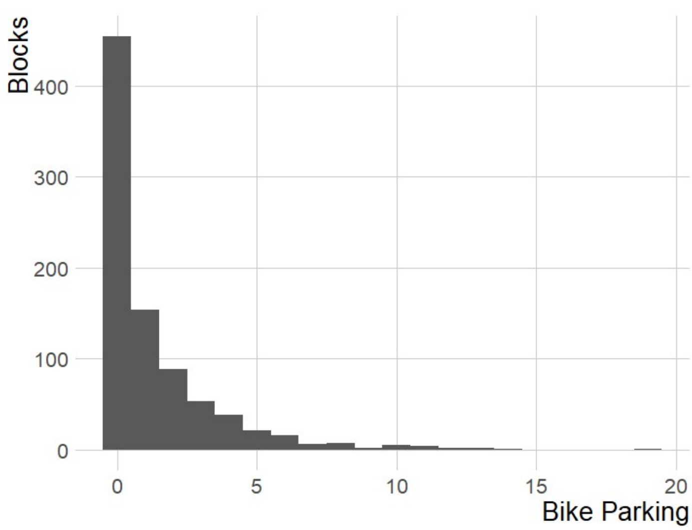


Chart 3. Informal Parking Histogram

Figure 7 shows the total capacity in each block. Capacity of a block is the maximum number of bikes that can be parked on formal racks in a block. This is determined by the total number of racks. Usually, one bike rack has a capacity of 2 and a corral has 8.

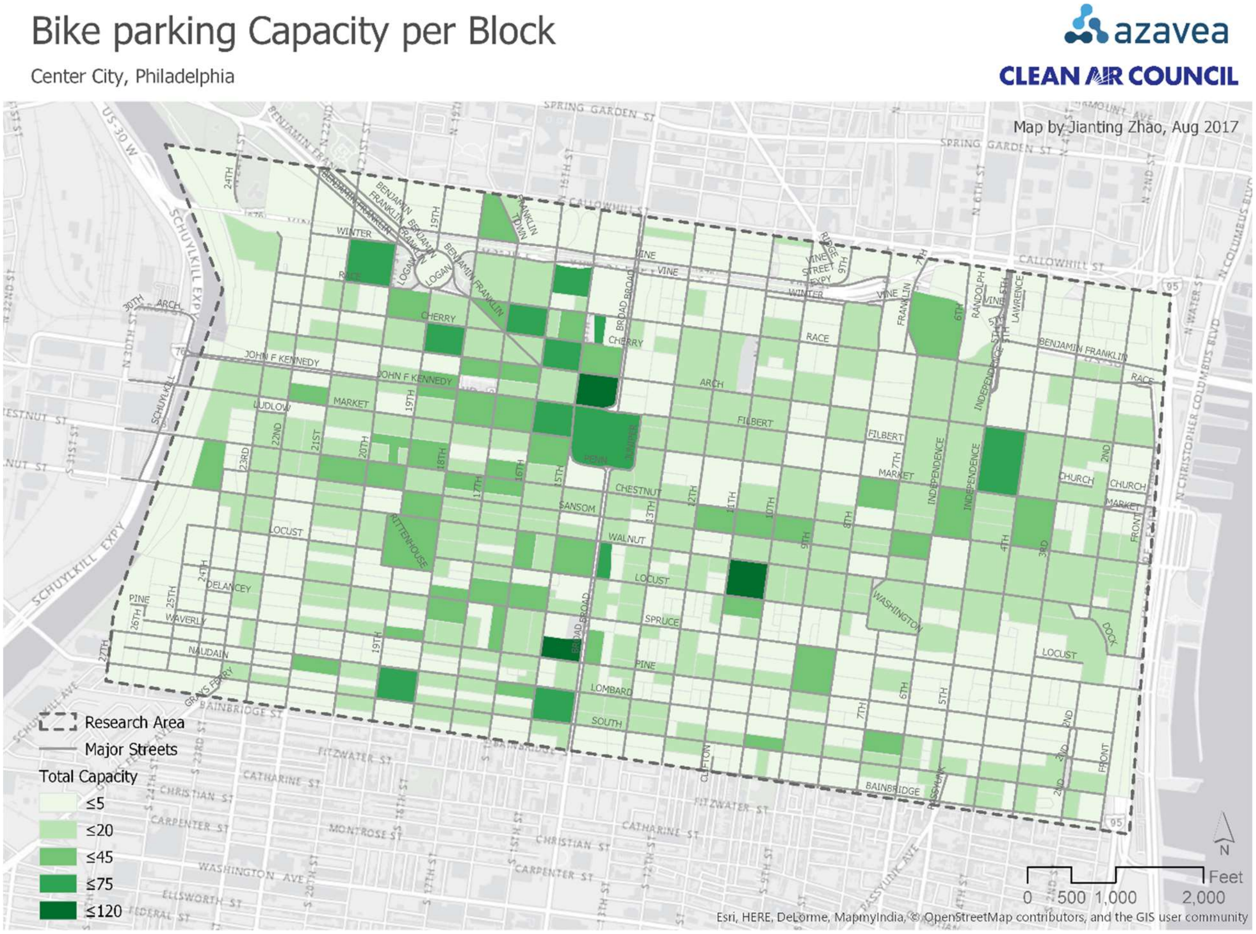


Figure 7. Capacity per Blocks

Chart 4 shows the histogram of total capacity.

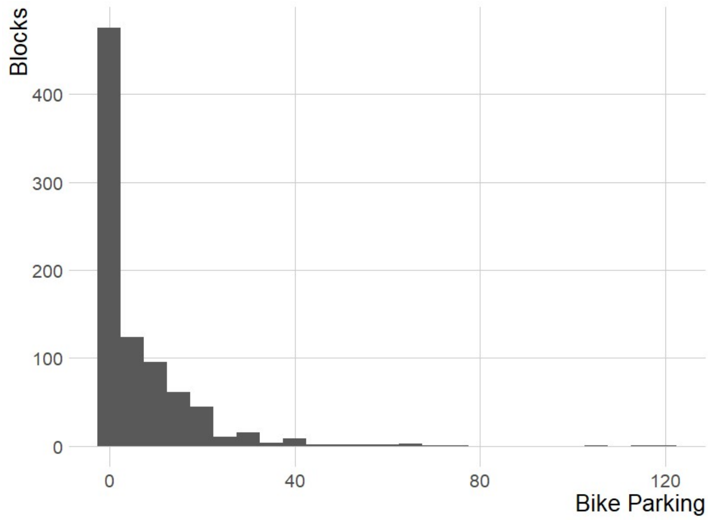


Chart 4. Total Capacity Histogram

After looking at each aspect, let us combine the informal and formal parking uses. Figure 8 shows two types of blocks. The ones in blue are the blocks with the most informal and formal uses. These blocks satisfy the following criteria:

Blocks in Blue = top quintile in formal parking usage AND top quintile in informal parking usage

The other ones in red are the blocks with the high usage in informal parking but low usage in formal parking. These blocks satisfy the following criteria:

Blocks in Red = bottom quintile in formal parking usage AND top quintile in informal parking usage

These blocks may be worthwhile to focus on in terms of informal parking regulation.

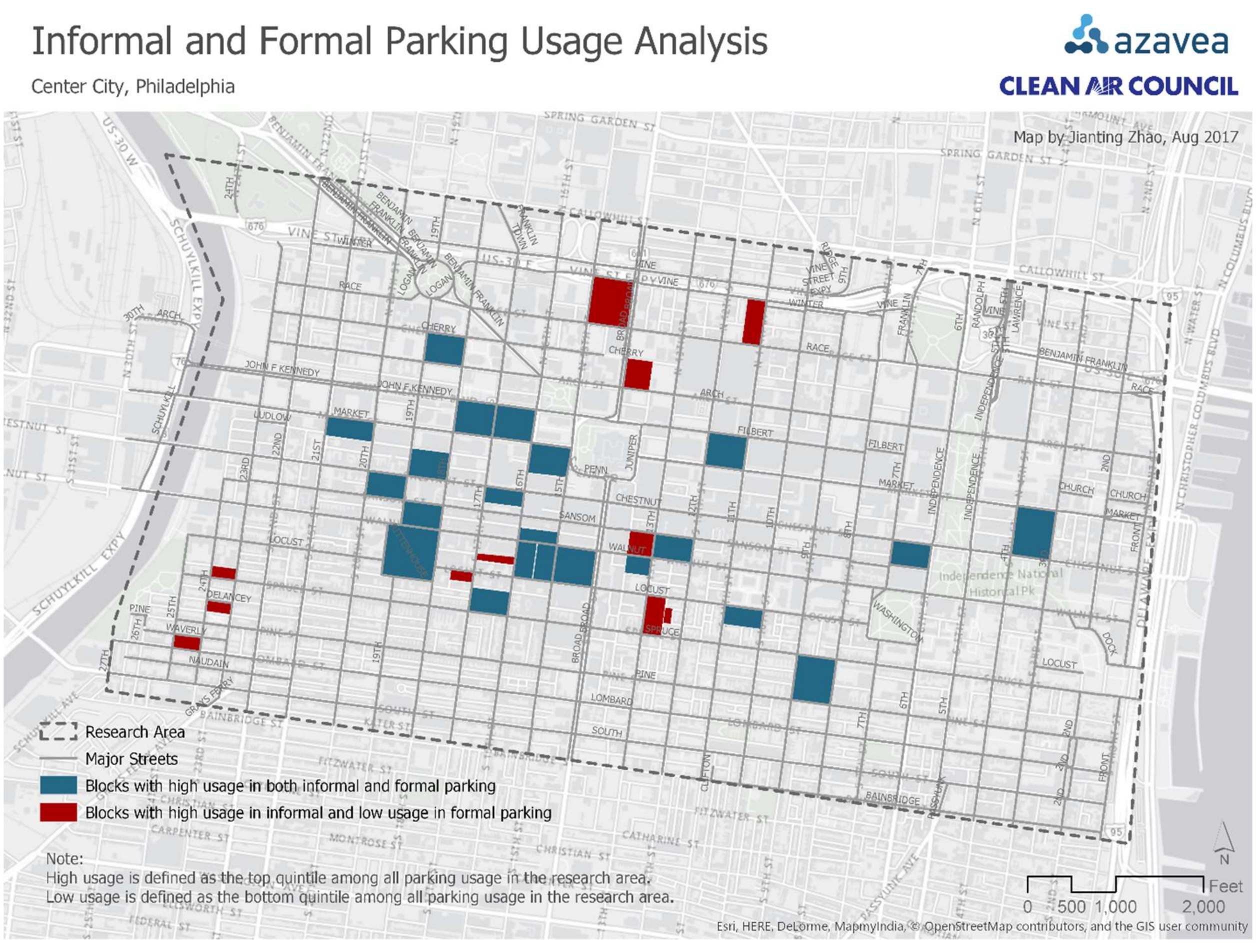


Figure 8. Parking Usage Analysis

Predicting Bike Parking Demand

With a better understanding of the bike parking data, we can study the prediction for the bike parking demand. As the Clean Air Council pointed out, they would like to incorporate land use, employment, and transportation factors in this study. In addition, in order to produce a more predictive model, several additional datasets were gathered. The next section displays several major factors that are used in the model.

Data Wrangling

Factor 1: Total Jobs per Block (Figure 9)

One on the factors that may affect bike parking demand during workdays is employment. Therefore, the employment data from the Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) is used.[12]

As shown in figure 9, jobs are concentrated along the Market Street and Chestnut Street corridors.

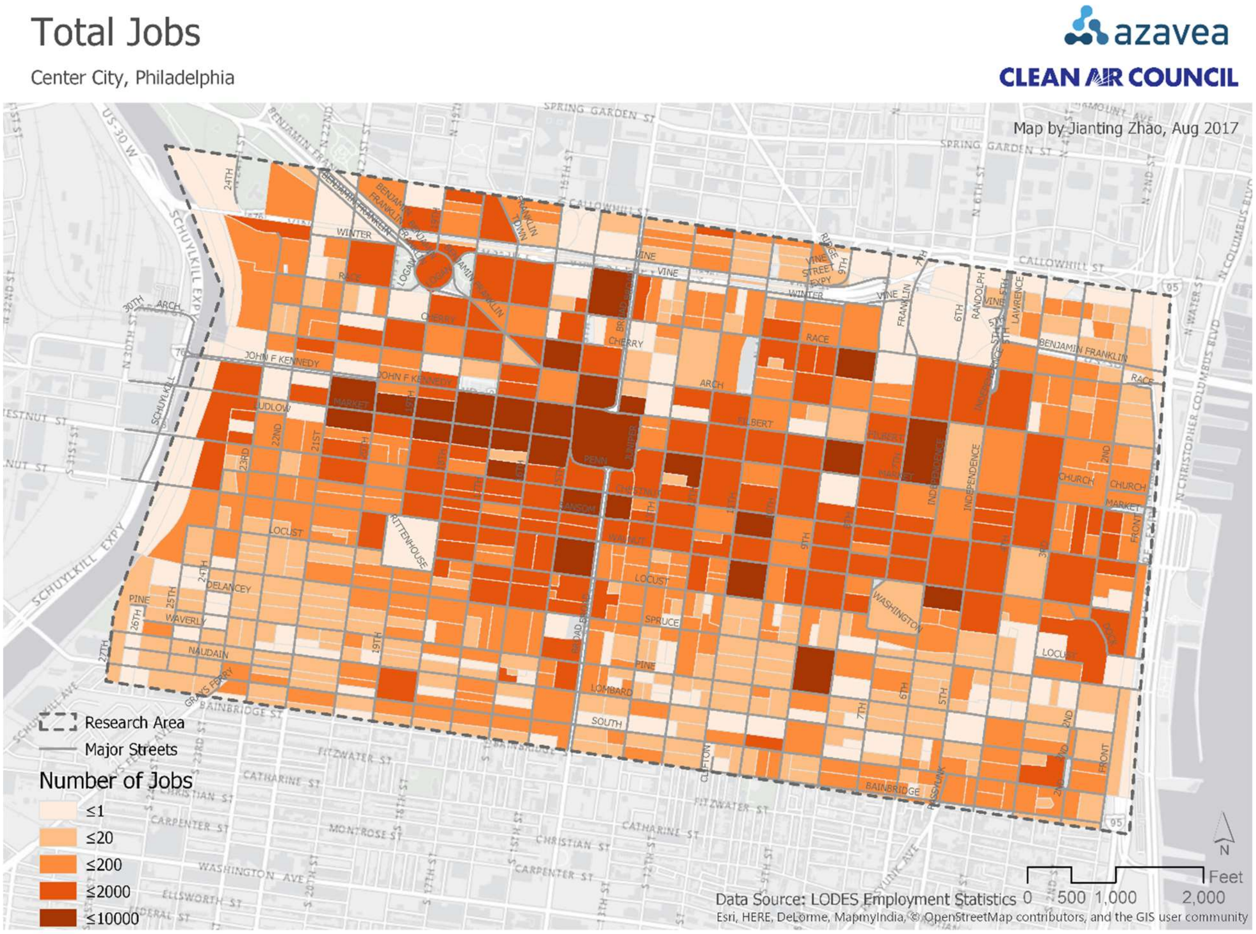


Figure 9. Total Jobs per Block

Factor 2: Number of Low Income Jobs per Block (Jobs with Monthly Income below $1,250) (Figure 10)

Only taking total jobs into account may not be enough. Therefore, I also incorporated other job categories. The major one is the low-income jobs category.[3] This factor has a similar pattern as the total jobs.

---

[1] There are a series of tables, and the "pa_wac_S000_JT00_2014_42101.csv" table is used for this study.

[2] The column C000 in this table is the total number of jobs per block. Please refer to the LODES manual for their metadata.

[3] This factor comes from the column CE01 in the same table as the total jobs.

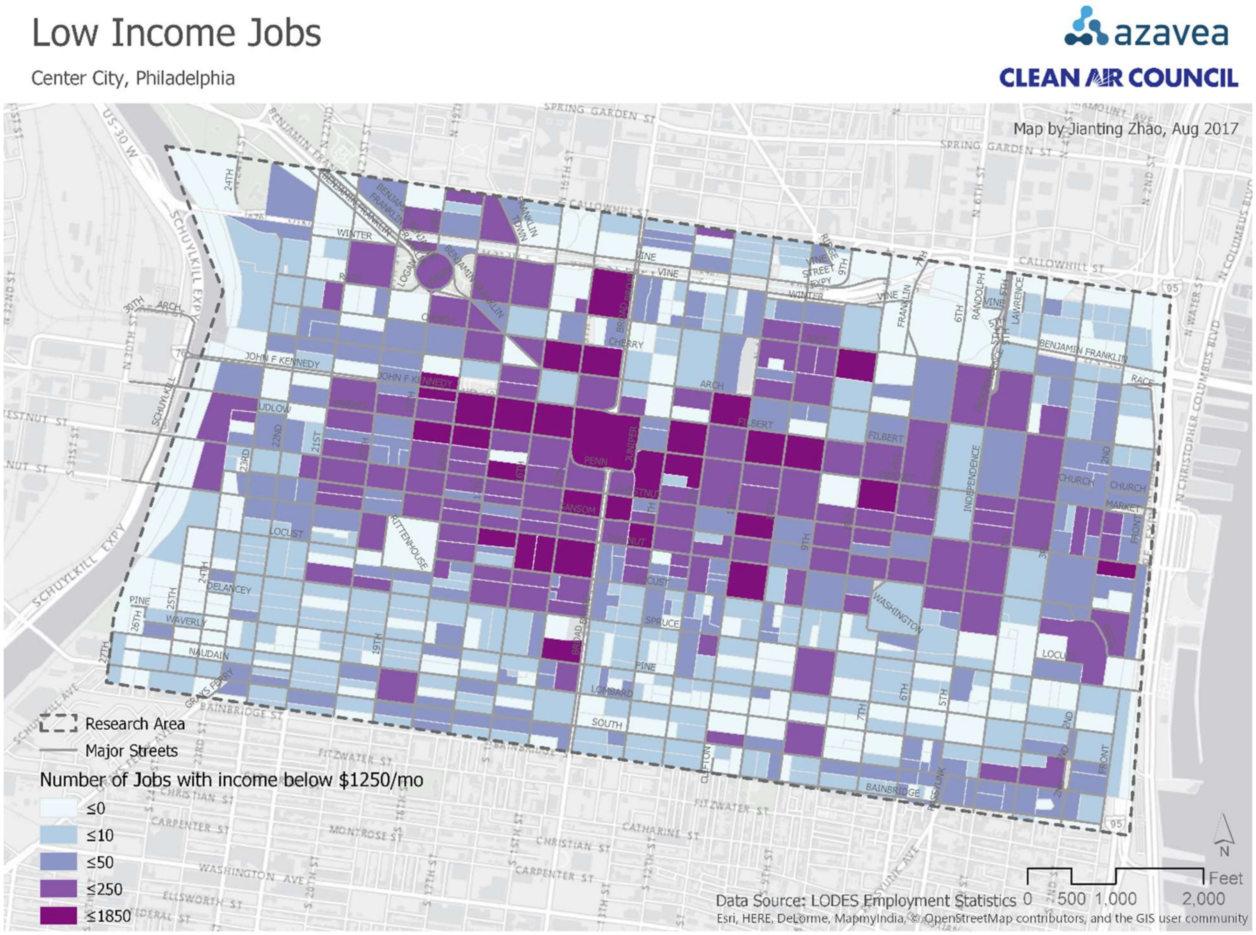


Figure 10. Number of Jobs with Monthly Income below $1,250 per Block

Besides the low-income jobs, the following job-related factors were also considered in the analysis:

1. number of medium-income jobs (monthly income between $1250 and $3333)
2. number of high-income jobs (monthly income greater than $3333)
3. number of jobs for worker under age 29
4. number of jobs for workers 30-54
5. number of jobs for workers over 55.

However, they were not shown in this report because they were not effective in terms of predicting for bike parking.

Factor 3: Average Nearest Distance to Bike Lanes

Center City Philadelphia has several bike lanes.[4] It is intuitive to think that there is a close relationship between the distance to bike lanes and the bike parking demand. Therefore, this variable is included in the analysis.

[4] The latest bike lane data was obtained from OpenDataPhilly, last updated on November 28, 2016.

Figure 11 shows the original bike lane data.

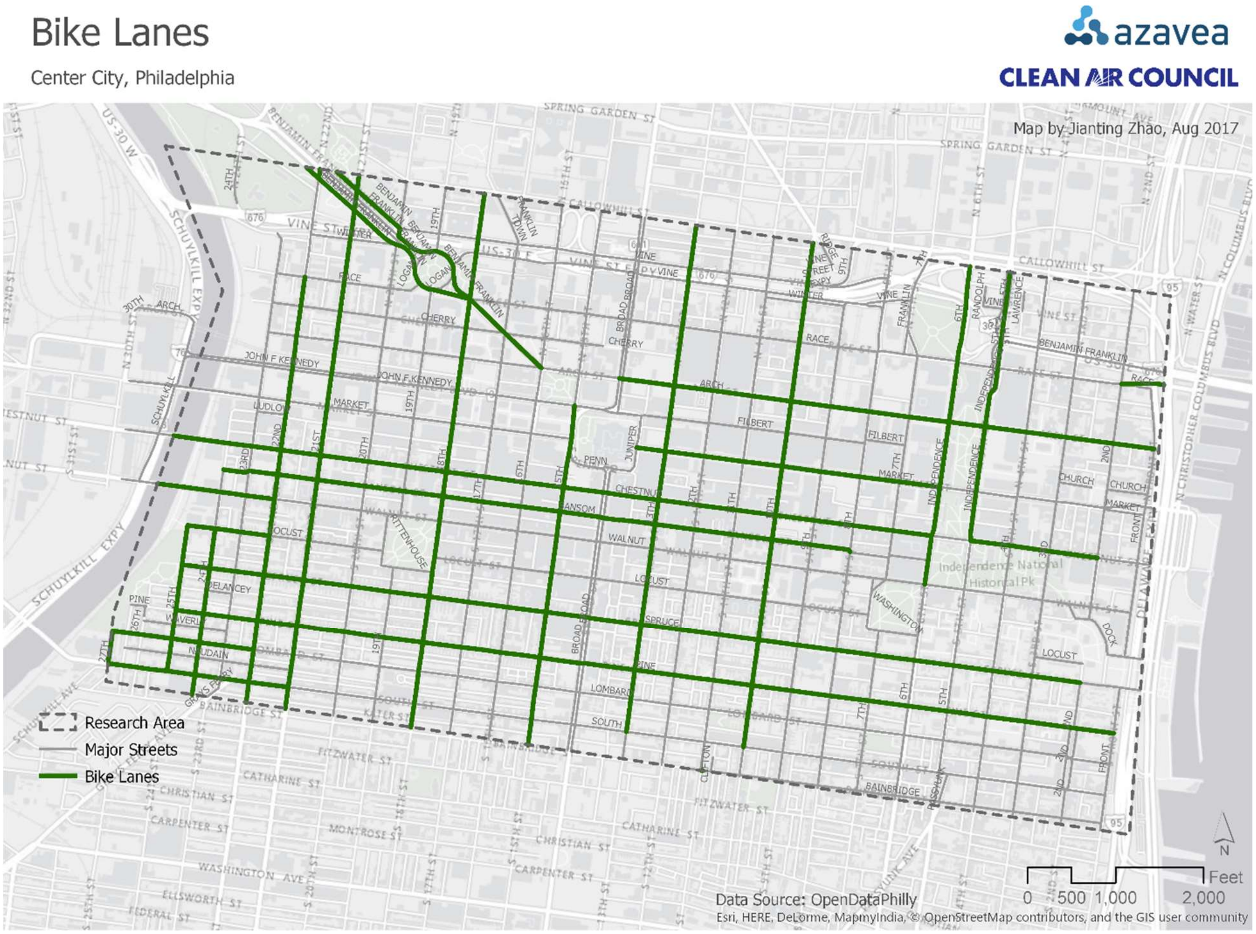


Figure 11. Bike Lanes in Center City

In order to bind the bike lanes to blocks, I calculated the average nearest distance from each block to its three closest bike lanes. As figure 12 shows, for each block, measure the average distance of three nearest bike lanes, and bind that number to that block.

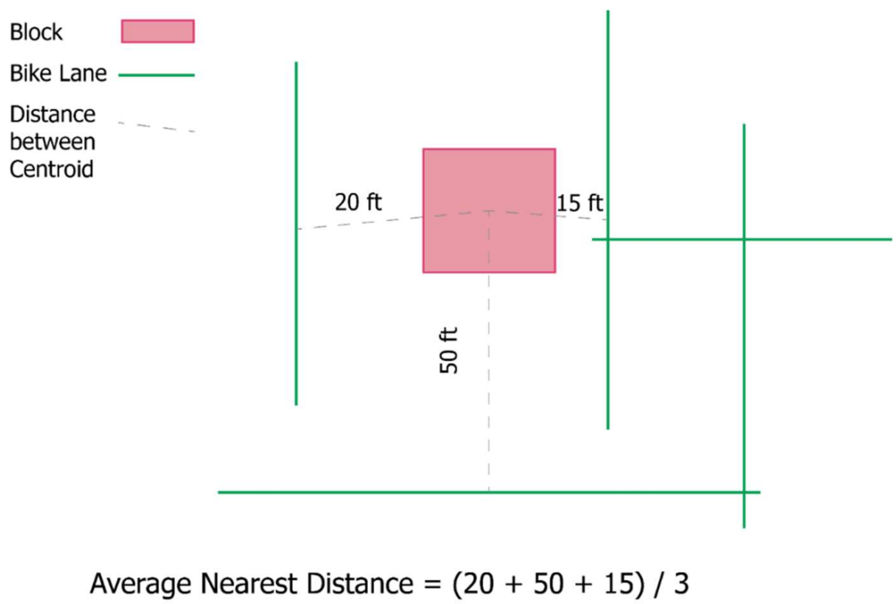


Figure 12. Average Nearest Distance Calculation Illustration

After calculating average distance for all blocks, I produced the Average Distance to Bike Lane variable as shown in figure 13.

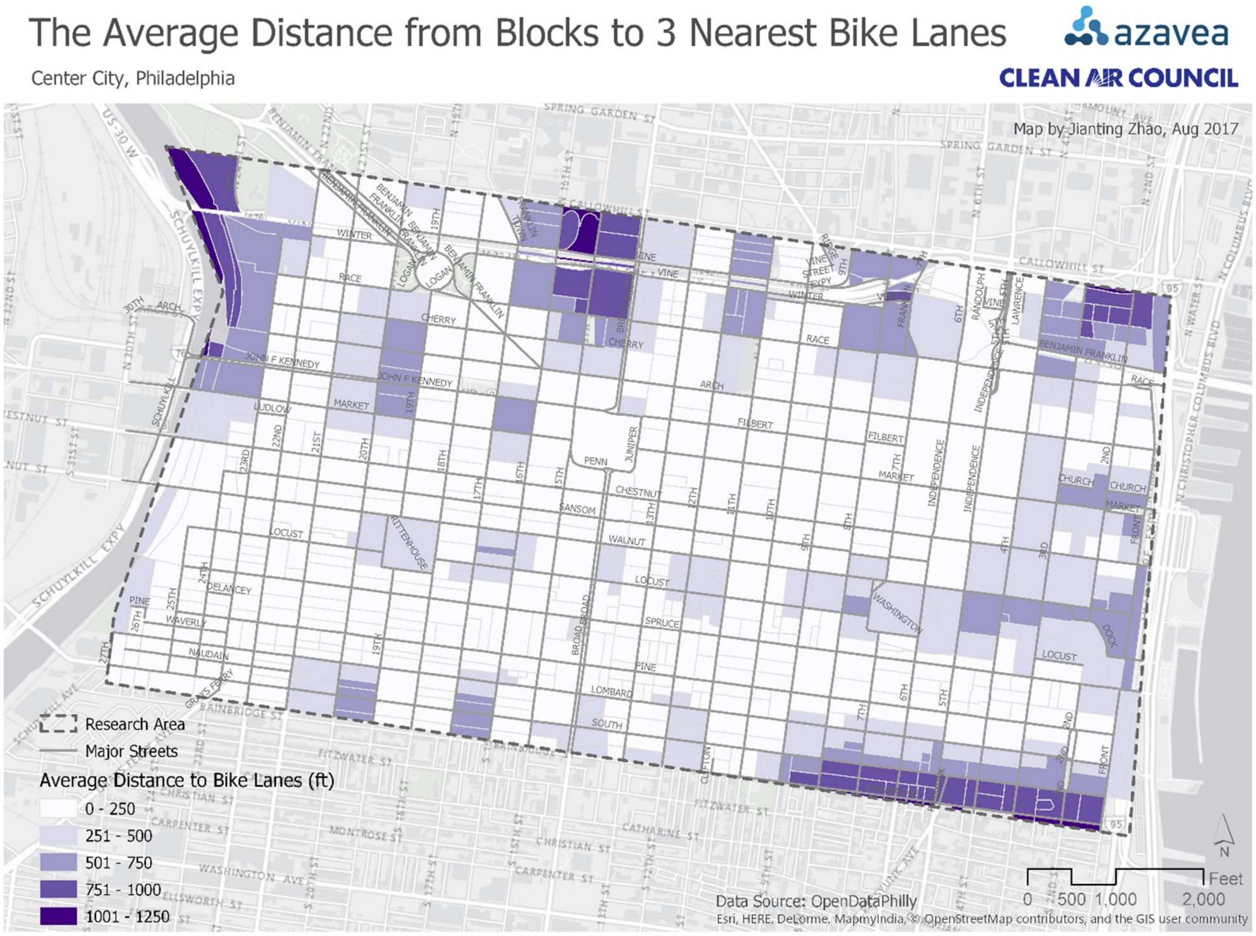


Figure 13. Average Nearest Distance to Bike Lanes Aggregated to Blocks

Factor 4: Average Distance to Trolley, Subway and Railway stations

One other factor that may influence bike parking demand is transportation, especially the trolley, subway and PATCO stations.[5] The reason why bus stations are not included is that they almost cover all street intersections, making it difficult to contrast the parking difference with or without them.

However, since the Suburban Station and Jefferson Station have multiple entrances that spread out several blocks, it is necessary to note their entrance locations because bikes can be parked at any of them. I manually added several points to represent their locations. With a similar approach to the bike lane dataset, I created the average distance to stations and aggregated them to blocks (figure 14).

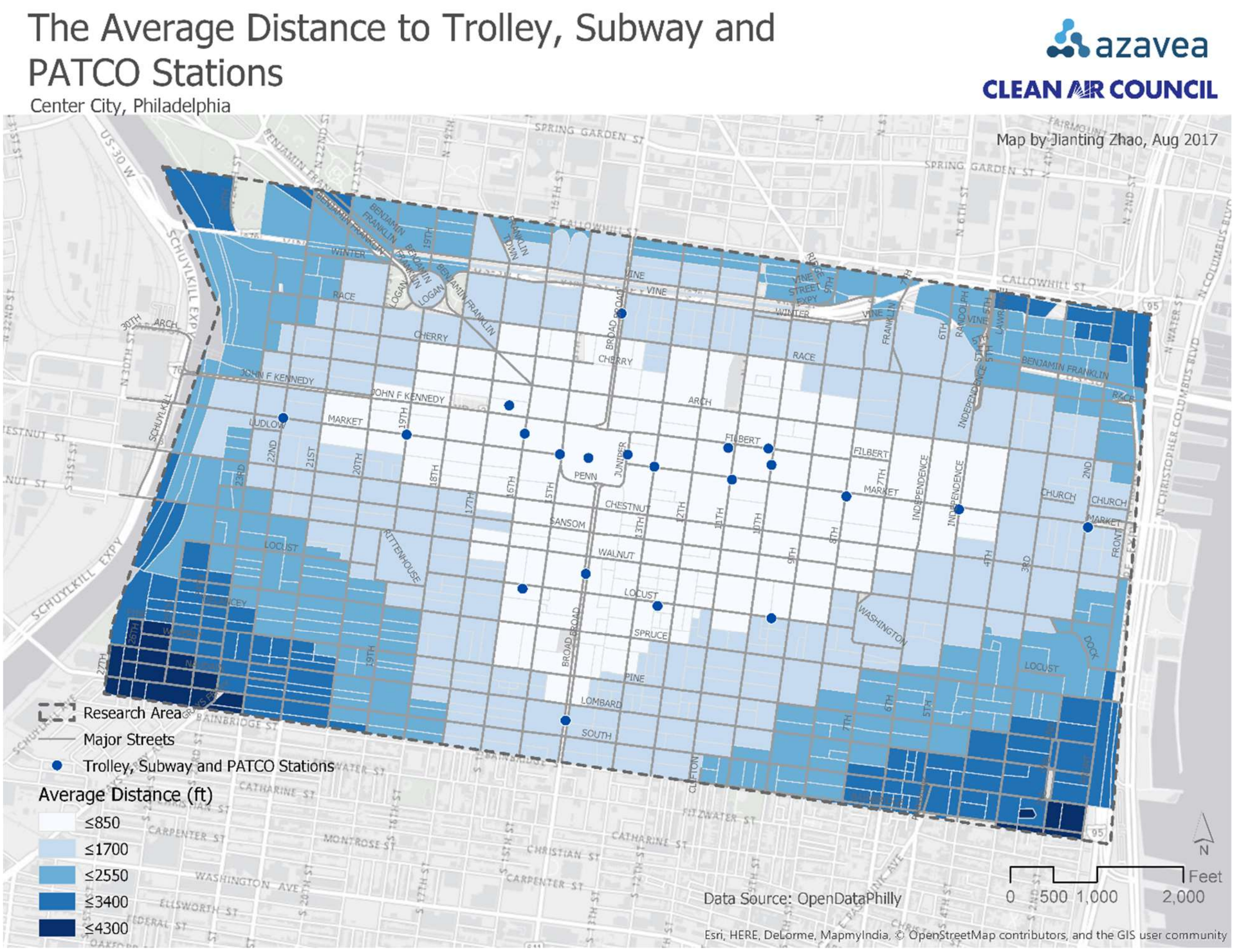


Figure 14. Average Distance to Stations Aggregated to Blocks

Factor 5: Land Use

Different land use may lead to differences in bike parking demand. For example, a block with a lot of business buildings may need more bike parking than residential blocks during working hours. There are

[5] The trolley, subway and PATCO stations are obtained from Opendataphilly portal. Data last updated on August 15, 2016.

16 land types. [6] [7] However, the types "Culture/Amusement", "Active Recreation", "Park/Open Space", "Cemetery", "Water", and "Vacant land" have a very low count, which would make regression analysis difficult. In addition, they are all open space land types. Therefore, I merged all of them can gave a new category called "100 – Open Space".

It is also worth noting that all streets and parking lots are categorized as "transportation" in the original land use map. However, the streets polygons make aggregation to blocks harder and inaccurate; therefore, I eliminated all the streets and left the parking spaces.

Figure 15 is the finalized land use map. The lower one third of the research area is mostly residential and the upper two thirds of the research area are commercial.

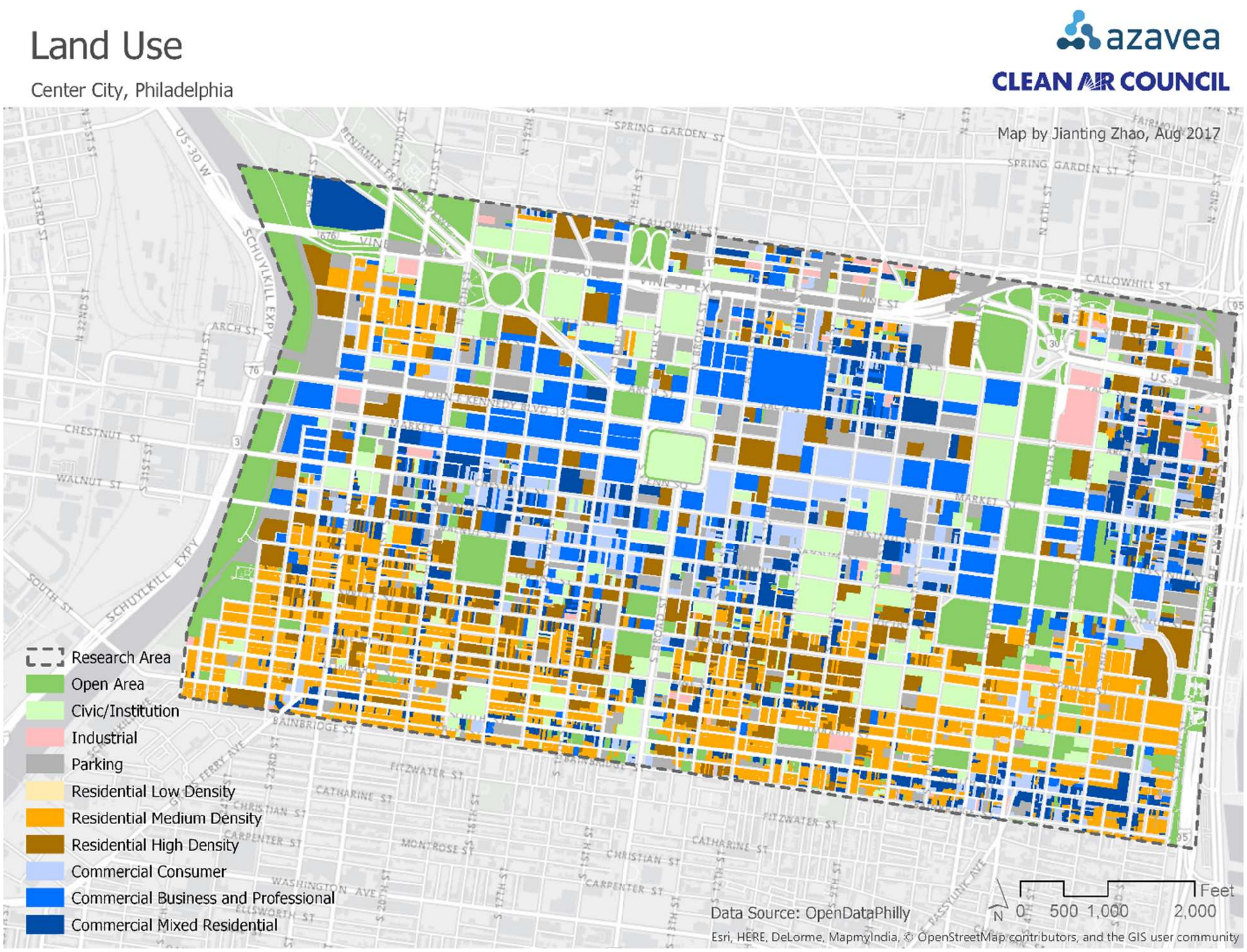


Figure 15. Land Use Map

[6] The land use data are also obtained from the Opendataphilly portal. Data last updated on April 24, 2017
[7] Refer to the land use metadata (http://metadata.phila.gov/#home/datasetdetails/5543864420583086178c4e74/?view_219_sort=field_15|asc) for details.

The land use parcels are much smaller than blocks, but there are a dozen of different types. I used each of the land types as a variable for the prediction model. For each land type variable, it is the area it occupies in each block. Figure 16 illustrates how the calculation works.

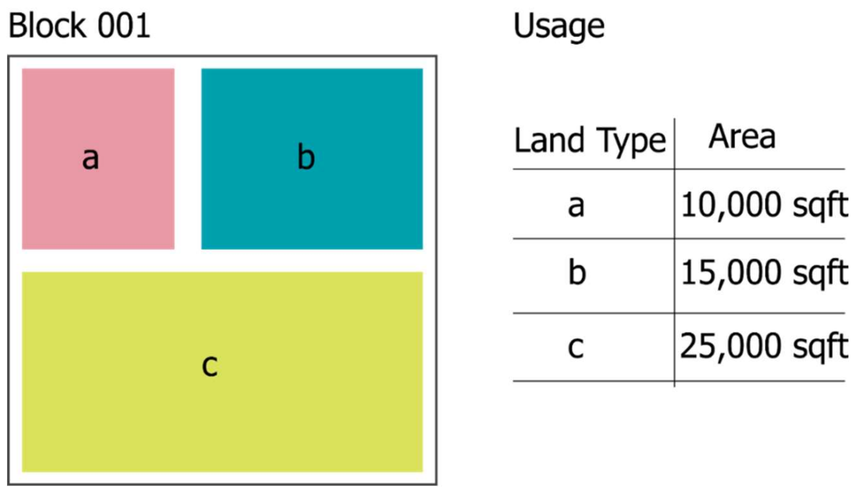


Figure 16. Land Use Area Bind to Blocks

Factor 6: Indego Bike Trips

Indigo bike trips can be a proxy for bike parking demand, because there are detailed records about the bike parking usage at each station.[8] I used the 2016 Q2 and Q3 Indego bike trips data. In addition, since we are focused on the bike parking pattern in Center City during workdays, I set three criteria to best filter out the data. They are:

1. Workday trips (Mon – Fri)
2. Trips between 7a and 11a
3. One-way trips going inbound to the research area

[8] The information can be found on https://www.rideindego.com/about/data/.

Figure 17 shows the trips with the filter applied.

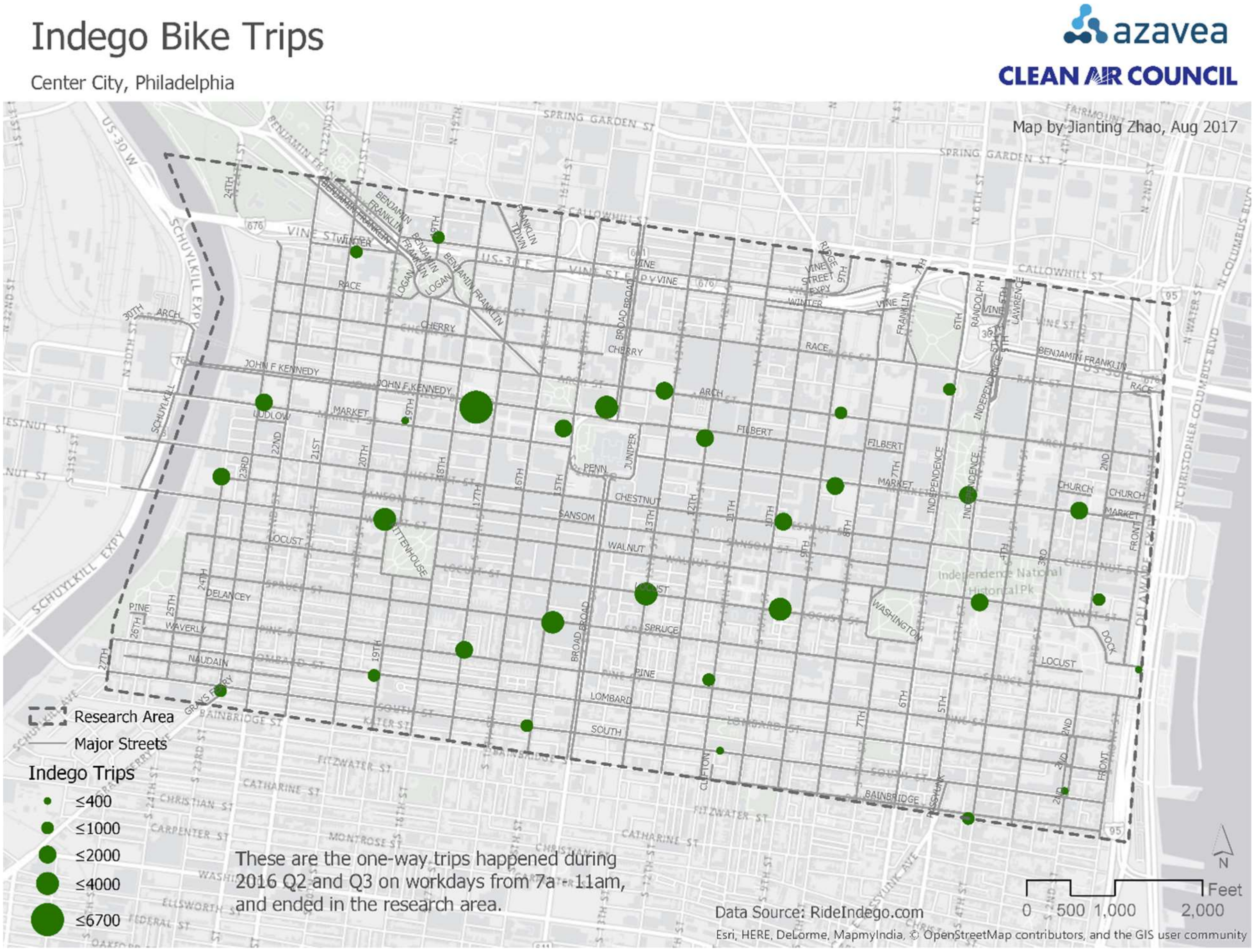


Figure 17. Indego Bike Trips

To bind the Indego trips point data to the blocks, I used a tool called "inverse distance weighted (IDW) interpolation"[9]. As figure 18 shows, the IDW interpolation is a way to create a trip estimation surface based on the theory that *nearer things are more alike than farther things*.

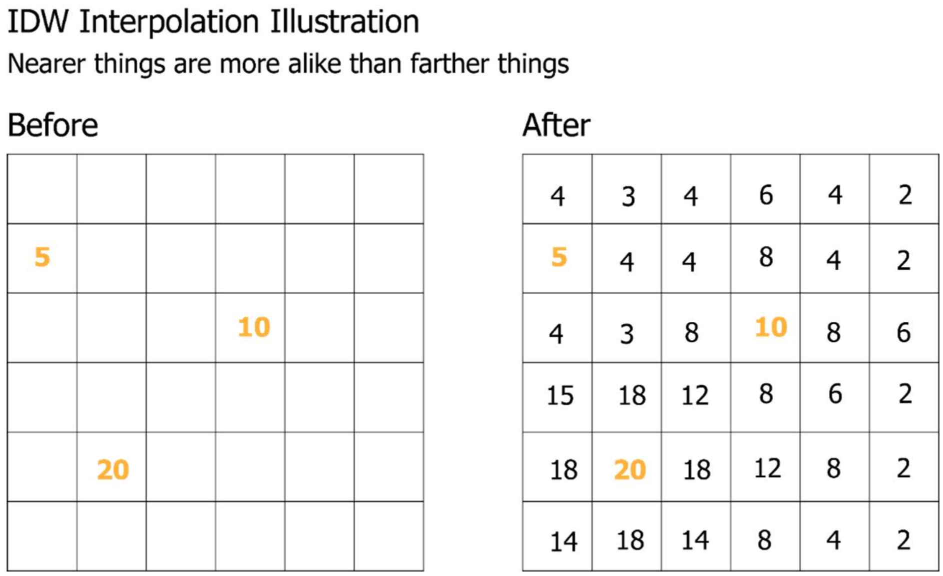


Figure 18. Illustration of IDW Interpolation

[9] Documentation can be found on http://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm

Figure 19 shows the interpolated bike trip layer. Figure 20 shows the average interpolated trips aggregated to blocks.

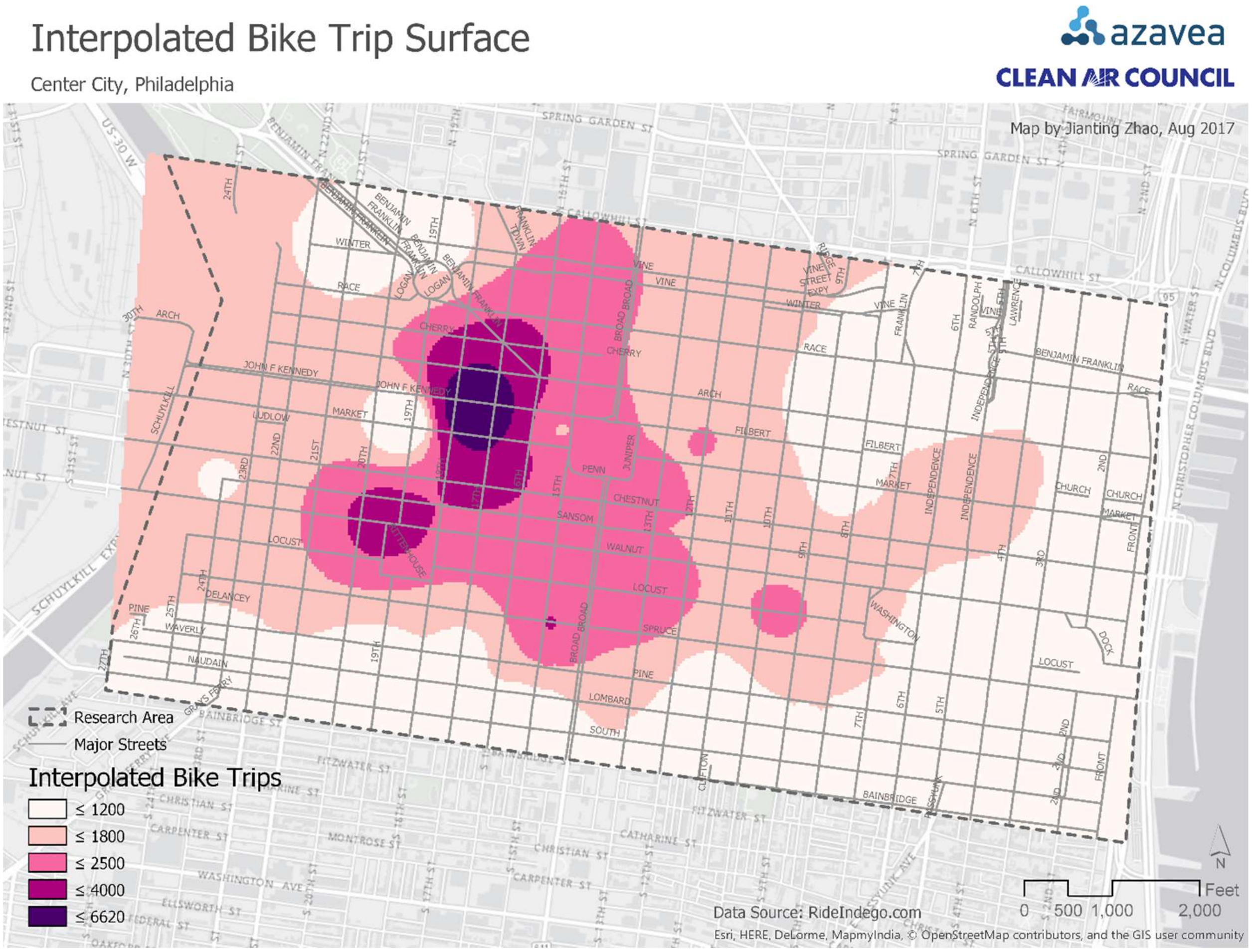


Figure 19. Interpolated Bike Trip Surface

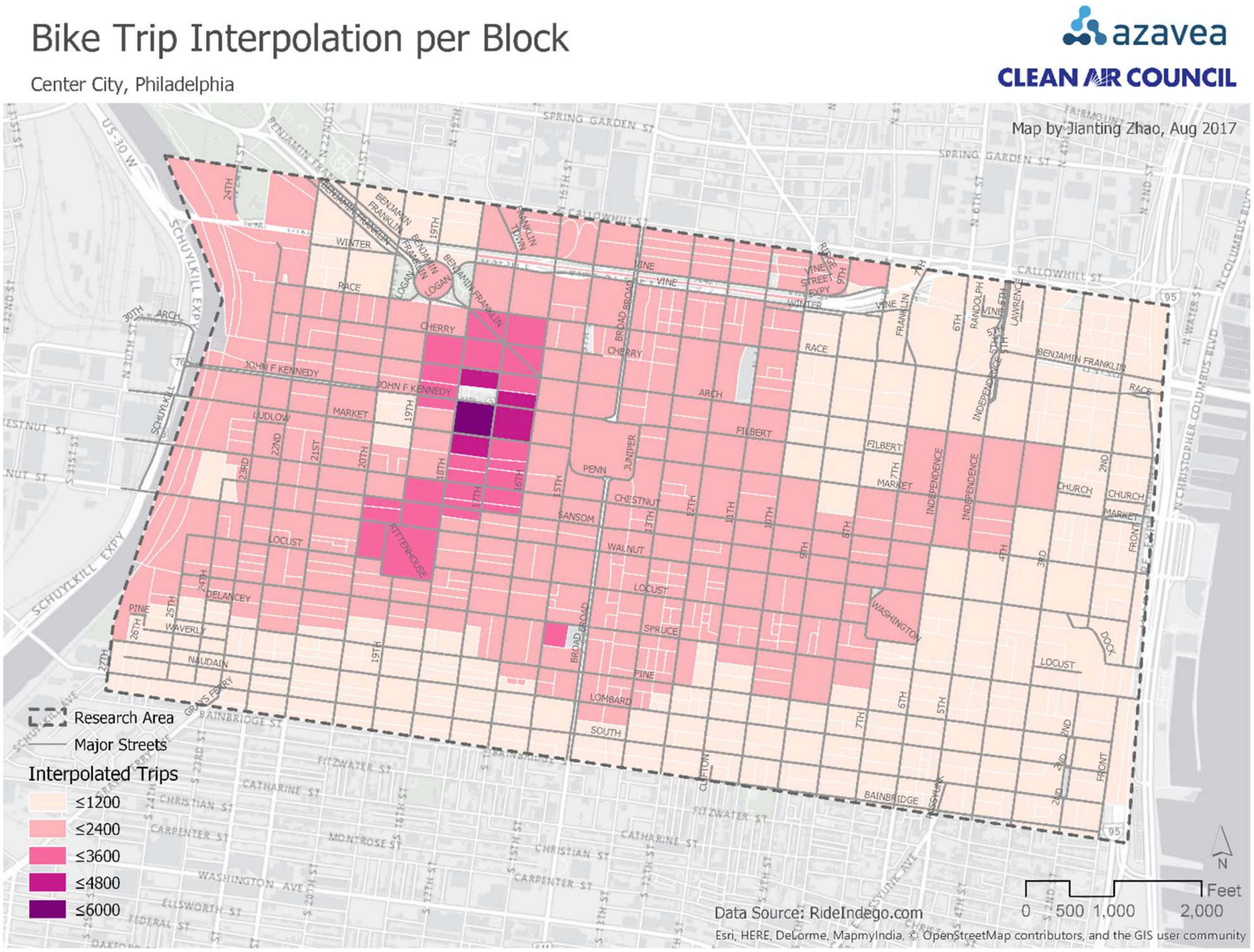


Figure 20. Interpolated Bike Trips Aggregated to Blocks

These six factors sum up the major data sources that I have used for this bike parking study. Other data sets tested include the percent of people using bike as a means of commuting in each census tract, average distance to end of bike lanes, and bus shelter locations. Unfortunately, they were not significant enough to be displayed in the report. With all the factors, we are ready to move on to the next step – utilize the factors to build the prediction model.

Analysis Approach

In the prediction model, I plan to use all six factors and the total capacity to predict for total use (figure 21). The reason to include total capacity as one of the predictor variables is that I assumed the existing bike racks reflect the general parking pattern, and therefore it can become a predictor for bike use.

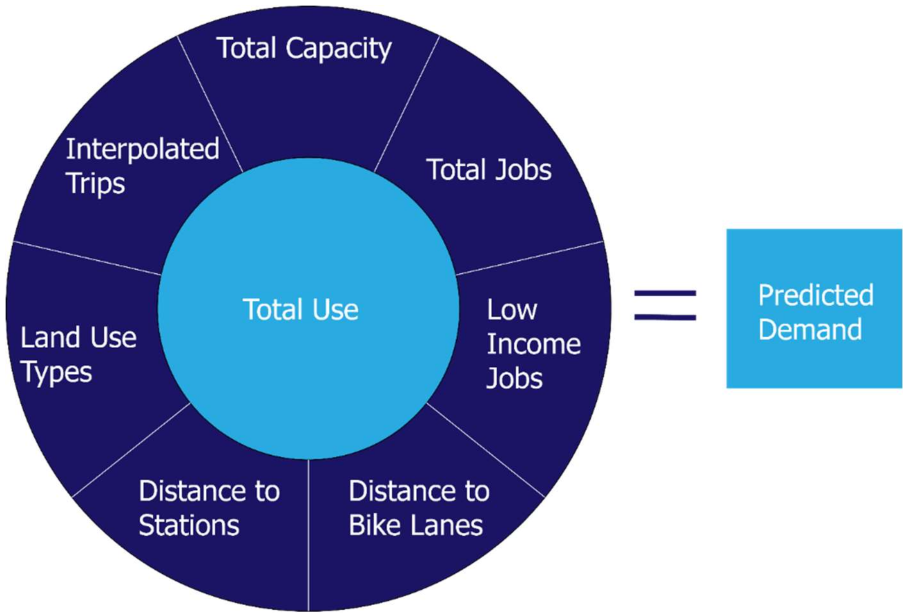
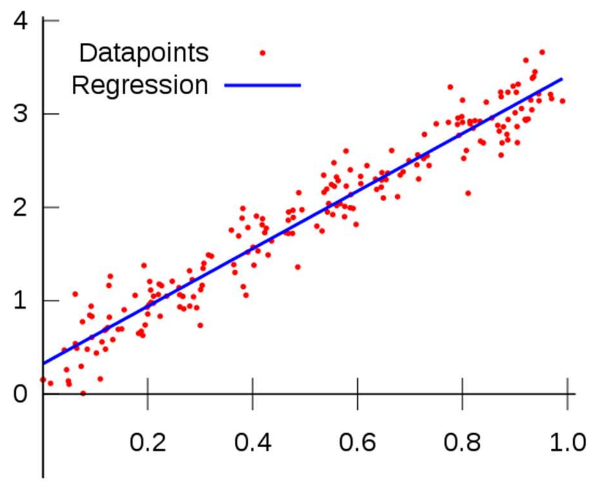


Figure 21. Variables Relationship

To find the best prediction model, I tried a list of regression models:

1. Linear Regression
2. Spatial Error Regression
3. Spatial Lag Regression
4. Geographically Weighted Regression
5. Poisson Regression
6. Negative Binomial Regression
7. Zero Inflated Regression

Even though the Poisson Regression, Negative Binomial Regression and Zero Inflated Regression are designed to work with count data with lots of zeros, they were not very predictive. The most effected model is the linear regression. Furthermore, linear regression can be written as a formula and is relatively easy to interpret.



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Figure 22. Linear Regression Illustration

Figure 23. Best-fit-line Equation

Linear regression is a statistical model that is frequently used to study influential factors and predict values. As shown in figure 22, its goal is to find the best-fit-line for all the red dots, i.e. the data points. The equation to represent this best-fit-line is shown in figure 23. Y is the response variable, in this case it is the Total Use. X_1, X_2, and X_3 are predictor variables, in this case are total jobs, land use, average distance to bike lanes and others. The β_1, β_2, and β_3, are the coefficients for the predictor variables. Lastly, the Ɛ is the prediction error, meaning the vertical distance between data points and the red line.

Finalized Formulae

After running the linear regression with different variables combinations, I achieved two best model candidates.

Model 1:

```
Parking_demand_in_each_block =
                    -1.80 + 0.399 * formal_capacity + 0.00114 * total_job +
                    0.00166 * interpolated_indego_bike_trips +
                    0.0000513 * commercial_consumer_landuse (sqft) +
                    0.0000217 * commercial_business_and_professional (sqft) +
                    0.0000202 * commercial_mixed_residential (sqft)
```

Interpretation:

The table 1 provides a guide for interpretation.

Amount	Factor	Trend	Amount	Factor	Trend
10	Total capacity	Increase	3.99	Parking demand	Increase
1000	Total job	Increase	1.14	Parking demand	Increase

1000	Interpolated trips	Increase	1.66	Parking demand	Increase
100,000	Commercial consumer (sqft)	Increase	5.13	Parking demand	Increase
100,000	Commercial business and professional (sqft)	Increase	2.17	Parking demand	Increase
100,000	Commercial mixed residential (sqft)	Increase	2.02	Parking demand	Increase

Table 1. Model 1 Interpretation

The table explains how much the predictor variable would bring how much change in the response variable. A sample interpretation would be: holding all other variables constant, if there are 10 more additional parking spaces, there will be demand for 4 more bike parking spaces.

Predicted Demand Overflow

In the Demand Overflow map, each increment means 1 capacity, which is half of a regular bike rack. The red blocks in the demand overflow map indicates need of installing several additional bike racks (figure 24). This map will be addressed further in the conclusion section.

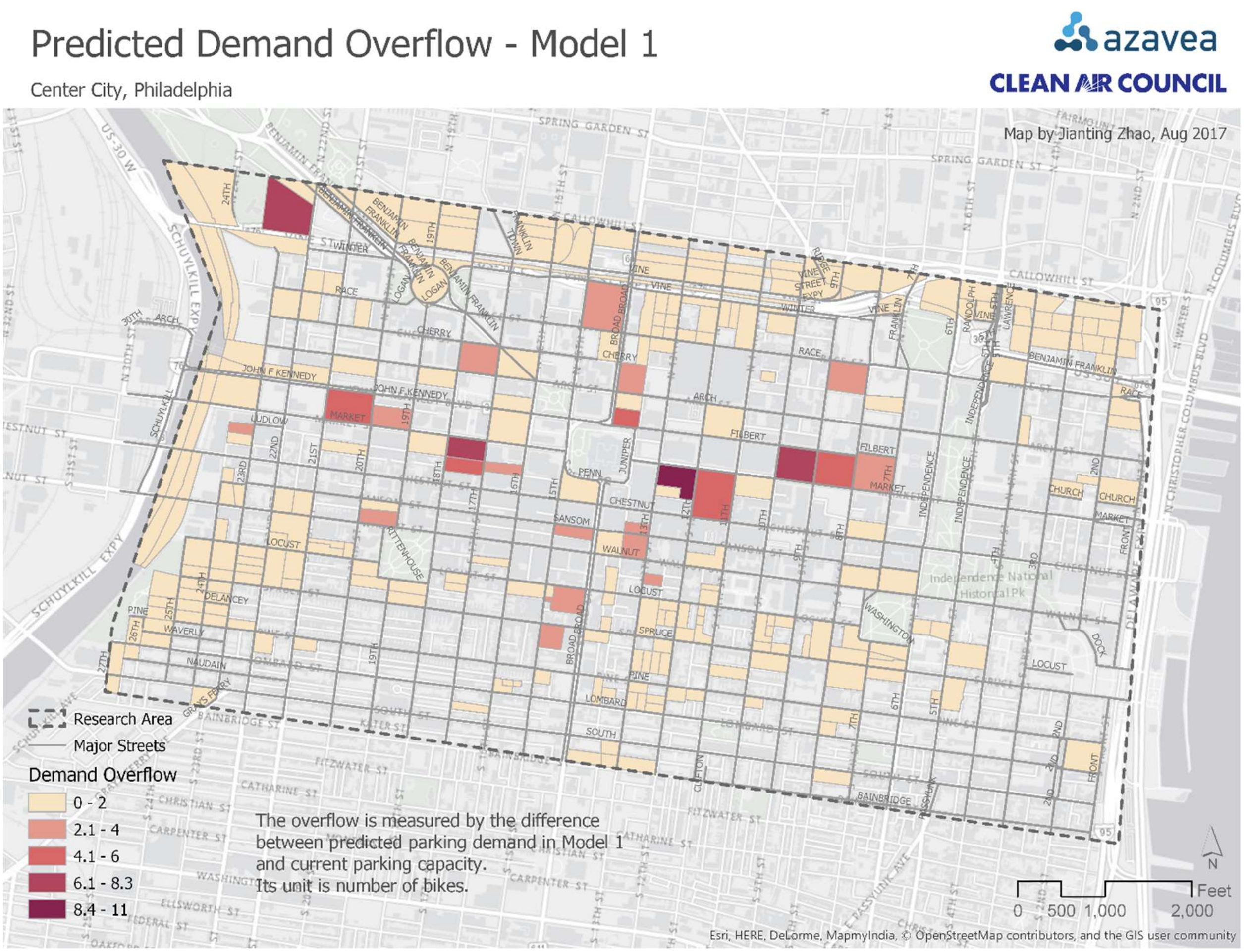


Figure 24. Model 1 Predicted Demand Overflow

Residual Map

Looking at the residual map in Figure 25, there are some areas that are worth noting.

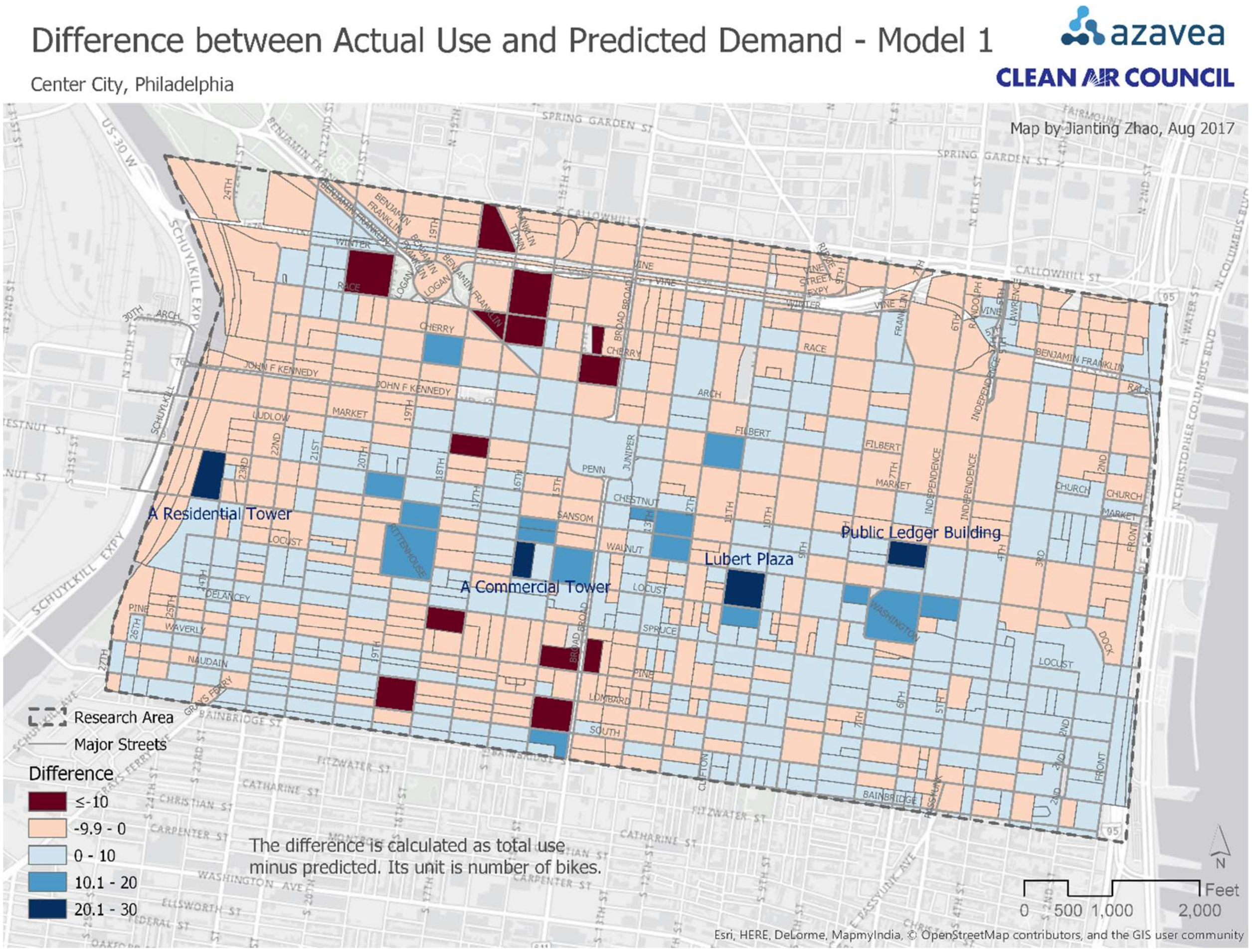


Figure 25. Model 1 Residual Map

The blocks represented in dark blue are under-predicted by 20 to 30 bikes and the dark red blocks are over-predicted by more than 10 bikes. As labeled on Figure 25, blocks like the ones for the Public Ledger Building, Lubert Plaza, a commercial tower, and a residential tower are significantly under-predicted.

Model 2:

```
Parking_demand_in_each_block =
                    0.684 + 0.405 * formal_capacity + 0.000929 * total_job +
                    0.0000439 * commercial_consumer_landuse (sqft) +
                    0.0000276 * commercial_business_and_professional (sqft) +
                    0.0000198 * commercial_mixed_residential (sqft) +
                    0.00503 * #jobs_under$1250/mo_or_less +
                    -0.00157 * average_distance_from_bike_lane
```

Interpretation:

The table 2 provides a guide for interpreting the regression model.

Amount	Factor	Trend	Amount	Factor	Trend
10	Total capacity	Increase	4.05	Parking demand	Increase
1000	Total job	Increase	0.93	Parking demand	Increase
1000	Low income jobs	Increase	5.03	Parking demand	Increase
100,000	Commercial consumer (sqft)	Increase	4.39	Parking demand	Increase
100,000	Commercial business and professional (sqft)	Increase	2.76	Parking demand	Increase
100,000	Commercial mixed residential (sqft)	Increase	1.98	Parking demand	Increase
1,000	Feet away from bike lane	increase	1.57	Parking demand	Increase

Table 2. Model 2 Interpretation

The interpretation for model 2 is similar to that of model 1. However, the last variable is slightly different. It should be interpreted as “holding other variables constant, if a block is 1000 feet farther away from bike lane, there is a 1.57 decrease in parking demand.

Predicted Demand Overflow

The predicted demand overflow for model 2 is similar to model 1, even if the variables are different. Both maps will be addressed in conclusion section.

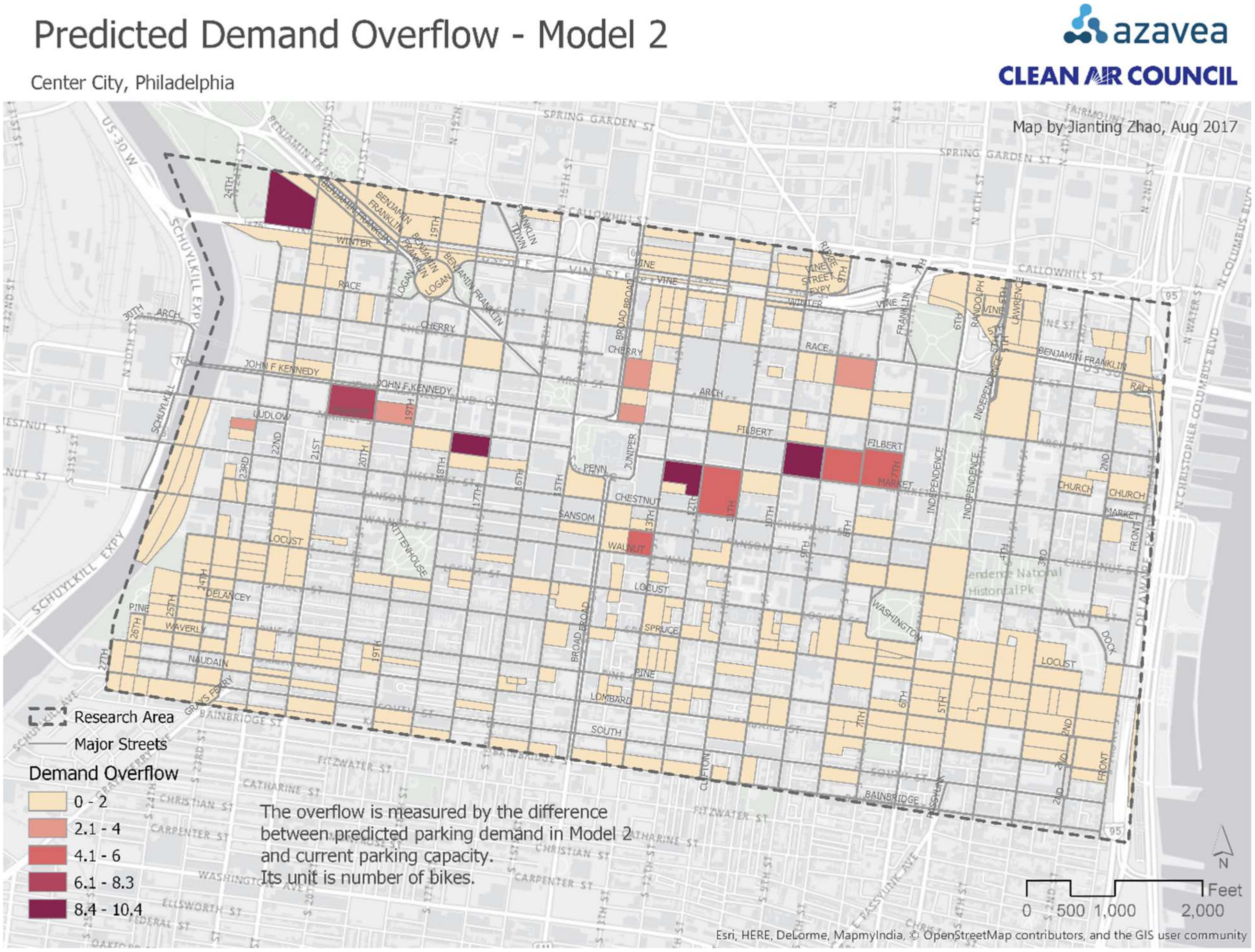


Figure 26. Model 2 Predicted Demand Overflow

Residual Map

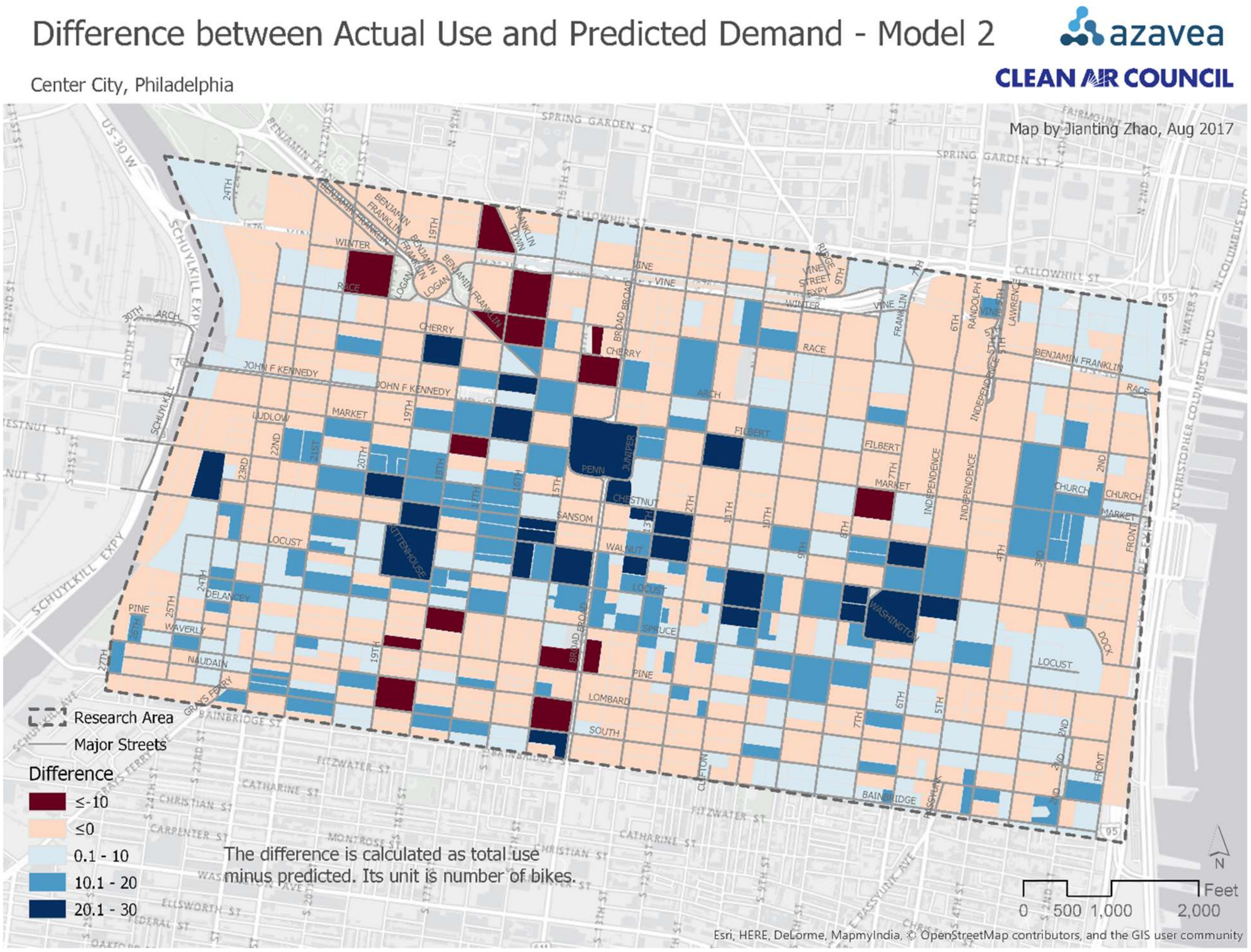


Figure 27. Model 2 Residual Map

The blocks represented in dark blue are under-predicted by 20 to 30 bikes and the dark red blocks are over-predicted by more than 10 bikes. Overall, the model 2 seems to be less accurate than model 1, since there are a lot more overpredicted and underpredicted blocks. Many blocks along Walnut Street are significantly under-predicted. Whereas several blocks near northern and southern border of the research area are significantly over-predicted.

Limitation

This study has several limitations in terms of the research design and the methodology. First, the bike parking data is only a snapshot of the parking situation. Each place has only been recorded once. The pattern would be significantly different if it was recorded during a different time of day or a different time of year. Therefore, it may cause some bias in our result.

Second, there is not enough detailed information about bike commuting. Even though the American Community Survey asks about bicycle commuting to work, it only shows where people bike from, not where they bike to. In addition, that data is not available at the block level, therefore too broad for this study.

Third, aggregating the original point data to blocks can cause some minor inaccuracy. As shown in Figure 28, the data points that align perfectly with a street would be difficult to split to two blocks. The parking in each block may not reflect the actual number.

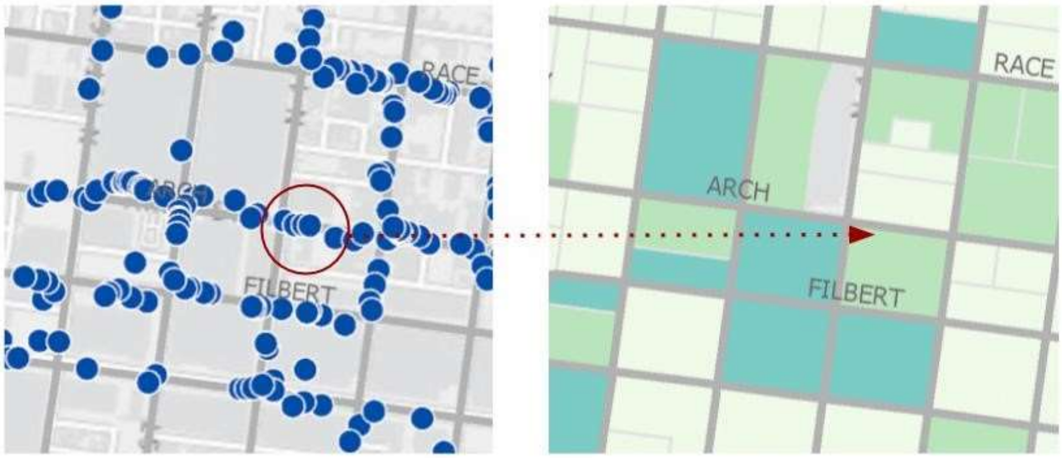


Figure 28. Point to Block Aggregation Inaccuracy

Lastly, even though linear regression provides the best prediction result, it may not be the best model due to the characteristics of the response variable. The linear regression works best with continuous response variables. However, the variable we have, i.e. the total use, is a discrete and positive variable. This type of variable should work with Poisson regression the best, because Poisson regression predicts for positive and discrete values.

Preliminary Conclusion

The Predicted Demand Overflow maps for both models provide similar results. Both maps suggest four areas that need 8 to 10 additional parking spaces. They are highlighted in Figures 29 to 32. The original bike parking data is brought in as a validation tool.

1. An apartment in front of the Art Museum (Figure 29)

   The map on the left highlights the recommended block (dark red) from two prediction models. The map on the right shows the actual parking use of that block. Since the crowdsourced data did not cover this block, there was no parking recorded on the map on the right. But even in the blocks underneath the highlighted block, we do not see a lot of bike parking usage.

Figure 29. Parking installation suggestion contrast with actual parking situation

2. A block next to the Liberty Place (Figure 30)

   The map on the left highlights the recommended block (dark red) from two prediction models. Since this block is next to Liberty Place, it is possible that it is a popular place to park bikes. Yet, the map on the right shows that there is not a lot of parking activity in this block.

Figure 30. Parking installation suggestion contrast with actual parking situation

3. A block near Market-Frankford Line 13th Station (Figure 31)

   The map on the right in Figure 31 shows that are 3 informal parking spaces. Even though it aligns with the suggestion by the models, 8 – 10 additional parking spaces seems too many for this block.

Figure 31. Parking installation suggestion contrast with actual parking situation

4. Jefferson Station block (figure 32)

   The dark red block is where the Jefferson Station and Market-Frankford Line 11th St. Station are located. The adjacent orange red blocks are department stores. It is likely that there is a lot of bike traffic and parking demand. However, the parking data on the right does not reflect the high usage.

Figure 32. Parking installation suggestion contrast with actual parking situation

To sum up, the parking recommendation does not align with the original parking data's current parking situation. Therefore, more inspection and observation are needed before applying the model recommendations.

Future Considerations

There are several things that could be done in the near future. The quickest would be to aggregate the original point data to fishnet polygons (Figure 33) and try a Poisson regression. This may provide a more precise result. Second, it would be beneficial to collect more data about the indoor bike parking capacity. Third, as the Council indicates there is an increasing need for covered bike parking, it would be beneficial to find the optimal spots for large covered parking installation. Fourth, study further the parking pattern at the popular urban parks such as Rittenhouse Square, Washington Square, Lubert Plaza, and others. The current model does not seem to capture the high parking usage at those popular areas. Last, experiment the parking pattern by installing temporary bike racks.

Figure 33. Fishnet Polygon Aggregation