COLUMBIA MATHEMATICAL MODELING CONTEST

DIVIDE AND CONQUER: MULTI-MODEL APPROACH TO BIKING INFRASTRUCTURE

February 6, 2023

Phillip Yan, Maksym Bondarenko, Caden Lin

For office use only	Team Control Number	For office use only
T1 ________________		F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________		F4 ________________

2023

MCM/ICM

Summary Sheet

Divide and Conquer: Multi-model Approach to Biking Infrastructure

In the past decade, New York City has rapidly expanded its bicycle lane network through a series of initiatives and funding packages. Most recently, NYC has received a $7.25 million dollar grant focused on expanding both the current greenway network and equitable access. Our team has been tasked to prepare a plan that maximizes the expansion of bike lanes and improvements to current infrastructure. To address each of the problem statements, our team developed a series of three linked models which support our recommendations in our plan.

First, we developed a model to geographically optimize our spending by concentrating our investments into a few select zip codes. Zip codes were chosen based a **development value** which was determined by a **trifecta of indices** reflecting current access to infrastructure, overall socioeconomic status, and health. The index values for each zip code were calculated based on government data available for each zip code and normalized against the corresponding average value across NYC to obtain a standardized value tha was comparable to other zip codes. After conducting parameter testing with a variety of weights for each of the indices, we selected five zip codes which appeared the most often.

Second, we considered the tradeoffs and decided our mix between protected and unprotected bike infrastructure by using a two-step model. Starting with determining necessary protected bike lanes, we constructed a **decision tree** which on determined on a road-to-road basis roads that were too dangerous (e.g. had multiple previous bike-related accidents) without a protected lane. We then determined the allocation of the leftover budget based on the ratio of existing bike lane density and average NYC bike lane density. Prioritizing conventional bike lane construction in neighborhoods with limited existing infrastructure and vice versa, we utilized a **sliding scale** model to determine an optimal mix of protected and conventional bike lanes.

Finally, our third model considered the impact of our plan on traffic, as measured by car usage, as well as future bike ridership. This analysis was conducted through the compartmental of a **compartmental population model** that divided the given population of commuters into car users and bike riders. Then, existing literature on bike riders in NYC was analyzed to compute appropriate **parameter value**, which were then **tuned** to investigate the impact of policy changes on car usage across the city.

The confluence of the three models provide a basis to recommend a comprehensive plan including a series of specific zip codes to focus our investments, a specific mix of protected and conventional bike infrastructure, and an analysis of its impact on both traffic and future ridership.

Key Words: Optimization, data analysis, multivariable analysis, decision tree, compartmental modeling, parameterization, parameter tuning

Contents

1 Introduction

1.1 Restatement of problems

Cycling throughout New York City has been one of the most rapidly growing modes of transport, resulting in a correspondingly sharp increase in demand for bike lanes. Although in an ideal world we would be able to address all shortfalls in bike infrastructure across the city, for this challenge we have only a limited grant of $7.25 million to develop a recommended course of action. With this budget, we aim to address several focal points of the plan.

- Identify one region in the city where the expansion or improvement of current infrastructure would have the greatest impact. Prioritize between high-density cycling infrastructure and lower-income communities without access to transportation
- Propose the best mix between protected and conventional bike lanes by considering the benefits associated with protected lanes and their increased cost when compared to conventional lanes
- Assess the overall impact of the cycling plan on city traffic. Weigh the benefits and drawbacks of allocating more/less road space to cyclists versus pedestrians and engined vehicles.

  We choose to address each one of those with a separate model. We evaluate the implications of this decision in discussions.

1.2 Order of Analysis

We address each of the problem statements in order, each with a linked, but unique model.

1. First, we develop a model which selects an optimal area for bike infrastructure investment based on potential ridership, social impact, and safety improvements. The model is then applied to select a single zip code for investment. The description of this model can be found in Section 2.
2. Second, we develop a safety-oriented generalized model for NYC to determine an appropriate mix of protected and unprotected bike lanes. Using data collected for the chosen area from part 1, we apply this cost-benefit derived model to determine an appropriate mix in protected and unprotected lanes
3. Third, we develop a model to determine the potential change of drivers to bikers as a proxy to determine the impact of our plan on both overall traffic in NYC and the number of new bikers.
4. After developing the three models, we discuss the strengths and weaknesses of each model alongside the ramifications of dividing our analysis among three distinct models.

2 Determining priority investment area

2.1 Assumptions and methodology

Map of NYC Zip Codes Constructed Using GeoPandas

Our goal in this section to determine areas of New York investing in which will yield the highest socioeconomic impact. We believe the potential socioeconomic benefits of cycling are:

1. Improving health and safety (by lowering the number of bicycle crashes and encouraging citizens to be more physically active)
2. Improving infrastructure and allowing for easier, greener commute
3. Allowing people from low-income and no-income backgrounds to access more job opportunities through improved infrastructure

Analyzing future change in those aspects based on lane infrastructure is very difficult, since the health and income data is temporally sparse, and the data for cycling infrastructure changes is sparse spatially. This prohibits an impact-based area model. Instead, we developed a model that evaluates how much each area needs improvement in each of those aspects.

A key challenge for determining an optimal location is finding a method to split Manhattan into areas for analysis. Methods that were considered included utilizing major thoroughfares as boundaries or overlaying a grid over the city. However, these methods divided the city into units that we did not have demographic data for. Therefore, we chose to divide the city by zip code.

We had to drop certain zip codes from our model because the census bureau has not collected necessary data there. The dropped zip codes constitute 0.5% of New York City by population and 1% of New York City by area and have a total population of 39,000. For the model, these people do not exist.

We assume that generally static or slowly changing data (e.g. number of subway stations, obesity rates, and population) will not be significantly different in the timeframe of one or two years. Thus, the values we determine from this data will still be significant even if collection timeframes differed by a few years.

2.2 Assumptions

Given the inherent complexity of modeling within an environment as dynamic as a city, it would be unrealistic to consider every single potential variable that may be weakly or very indirectly linked to bike infrastructure development. By ignoring these possibly relevant but insignificant factors, we are able to minimize noise in our model.

2.3 Trivariate Analysis

In order find the most optimal zip code, we need to construct a value system to compare different areas. We identified three "axis of value" which are indices we believe are most relevant in determining the best location with an emphasis on low income communities which currently lack access to affordable transportation options.

1. Infrastructure Index I_i: The amount of transportation infrastructure residents in the zip code have access to. Infrastructure includes current bike lanes, subway stations, and bus stops.
2. Socioeconomic Index S_i: Takes into account average income and those with no income.
3. Health Index H_i: Takes into account accident rates and physical inactivity rates

For each of these three values, the index value is determined as a ratio between the value of the zip code and the NYC average . This means the final index values of the zip code is a measurement how much "better" a zip code's index is compared to NYC as a whole (e.g. an infrastructure index of 2 means based on our model the overall accessibility to transportation is twice as good as the city average). This makes it easier to compare to other zip codes that are similarly standardized.

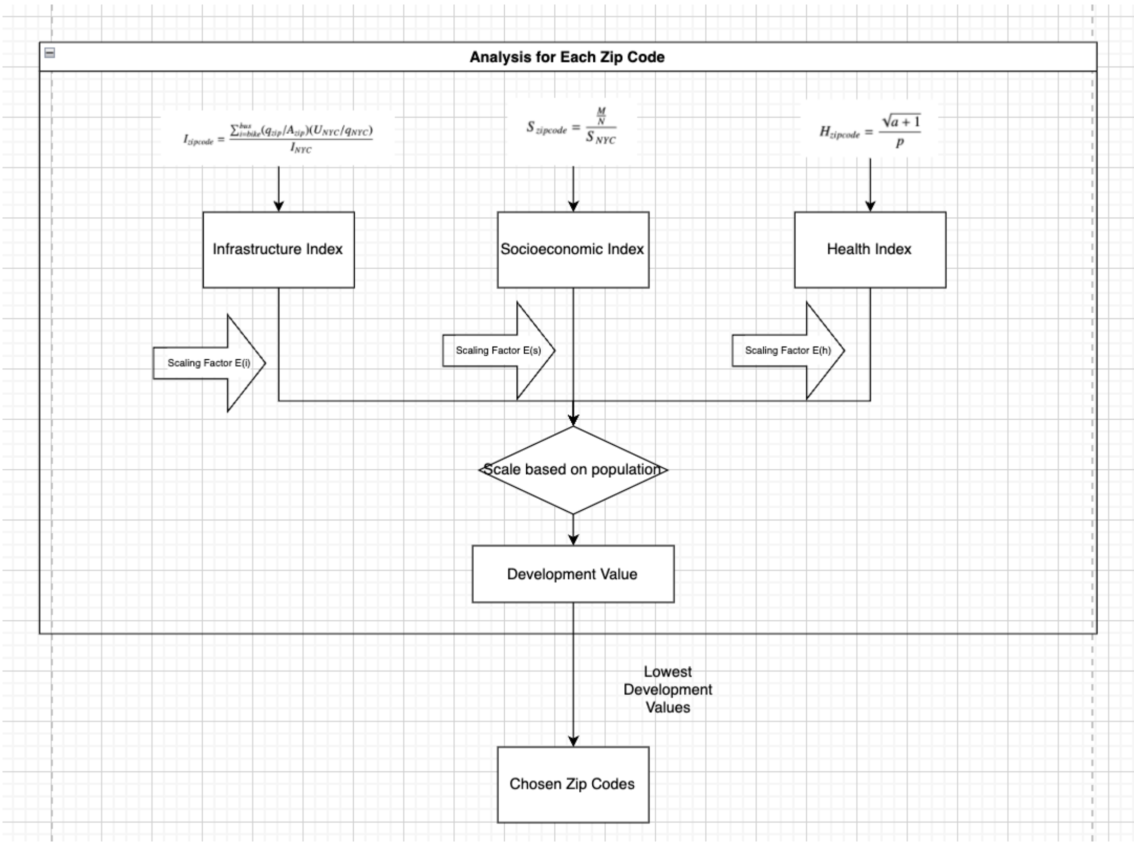
Then, we apply an emphasis scaling parameter (how much we value each of the three indices) which returns an overall **development value (D_v)**.

$$(E_I I_i + E_S S_i + E_H H_i) = D_v$$

where E represents the respective emphasis scaling factor for each index, and D_v represents each the development value. Finally, we determine priority for development as $1/D_v$ (how big the need in the zipcode is compared to others) times population the improvement will affect.

$$\text{priority} = \text{population}/D_v$$

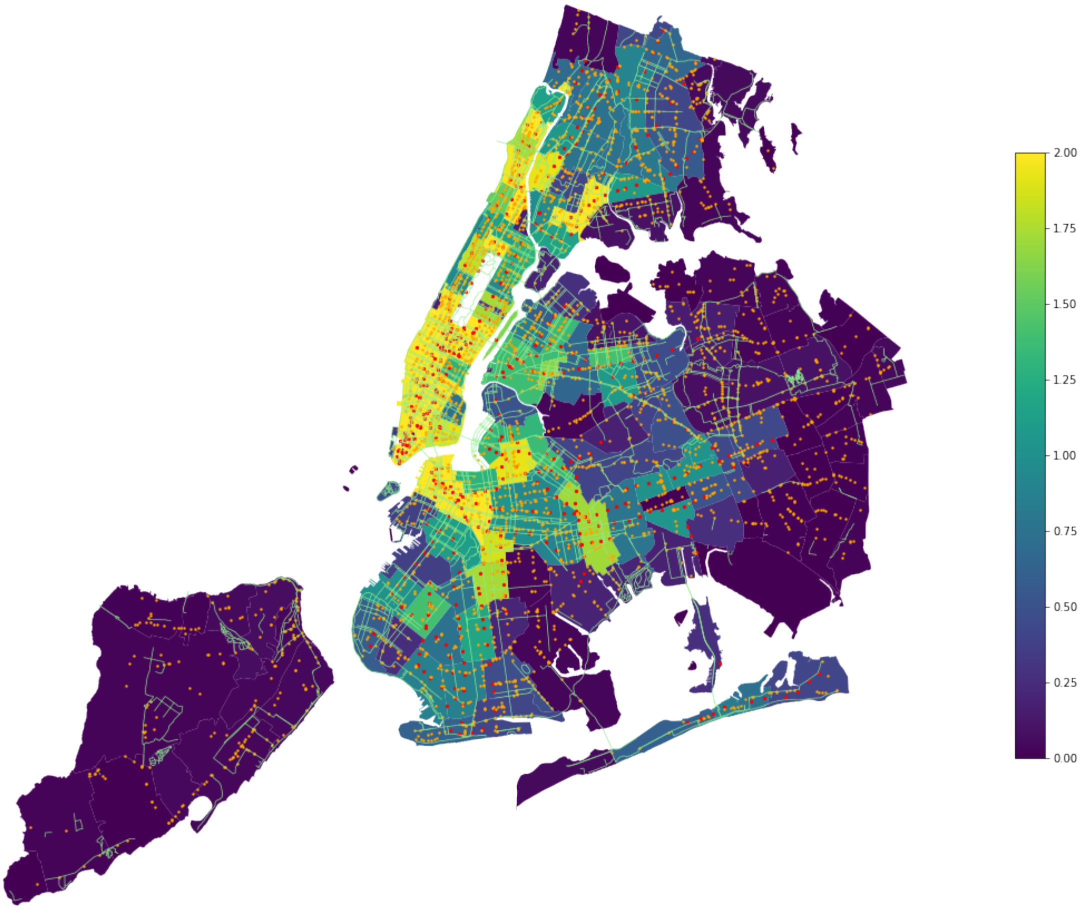
The zip code with the highest priority value thus has the largest need for cycling infrastructure improvement and the highest number of people who will benefit from it



A simplified diagram of our Locational Optimization Model

Below is a breakdown of the calculations of each of the indices

2.3.1 Infrastructure Index



Heatmap of transportation index value for each zipcode. Red dots represent subway station, orange dots represent bus stops and green lines are bikelanes. Made with geopandas

The infrastructure index is a measure of access to public transportation (bicycle, bus, subway) by zip code residents. The lower the index, the more bike-lanes are needed to supplement existing transport. We calculated it as:

$$I_{zipcode} = \frac{A_{bicycle}C_{bicycle} + A_{bus}C_{bus} + A_{subway}C_{sybway}}{I_{NYC}}$$

Where I_{NYC} is the mean value for the transportation index to normalize values, A denotes availability of each transport, and C denotes capacity of each transport.

We defined availability as units of transport (subway stations, bus stops, bicycle lane length) inside a zipcode divided by its area. Looking at only transportation inside zipcode is somewhat inaccurate, since transportation can be on the boundaries of zip codes, but it still is a good average measure for availability. Dividing by area represents differences in walking distance between big zip codes and small zipcodes. This definition makes sense, because having 1 subway station in big zipcode is less availability then 1 subway station in a small zipcode. Availability is calculated for each zipcode using geopandas intersection functionality.

$$A_{transport} = \frac{\text{units}_{transport}}{\text{area}}$$

However, different units of transports are not comparable. 1 subway station is much better than 1 bus stop, and comparing 1 subway station to 1 mile of bike lane is non-trivial. Here the capacity measure comes in. We defined capacity in terms of daily NYC trips by transport type divided by total number of respective units in NYC. For example, there are 2.4 millions people who use subway, and 472 stations, which gives us an approximate capacity of each station. Capacity is taken to be the same for all zipcodes.

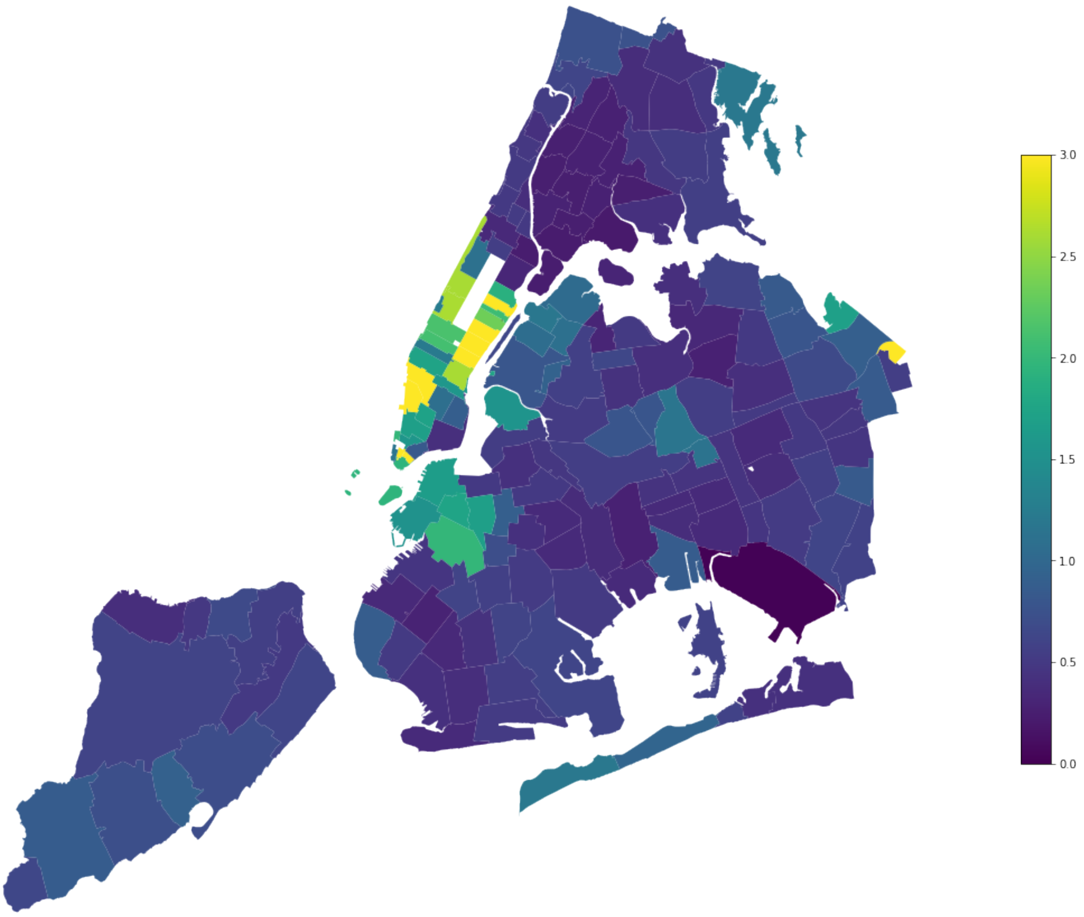
$$C_{transport} = \frac{\text{daily trips}_{transport}}{\text{units}_{transport}}$$

We found that each mile of bike lane accounts for around 3200 trips, each subway station for 5000 trips and each bus station for 2500 trips.

Our model guarantees that $A_{transport}C_{transport}$ for each transport is in the same units (trips/area) which allows summing them up into one index.

The image scale has been adjusted to be readable, but in general we found that some areas have 14 times as much transportation available (downtown Manhattan) than average. We also observed that areas on outskirts with no subway stations have significantly lower transportation index, which corresponds to what we wanted the index to capture.

2.3.2 Social Index



Heatmap of social index for each zipcode

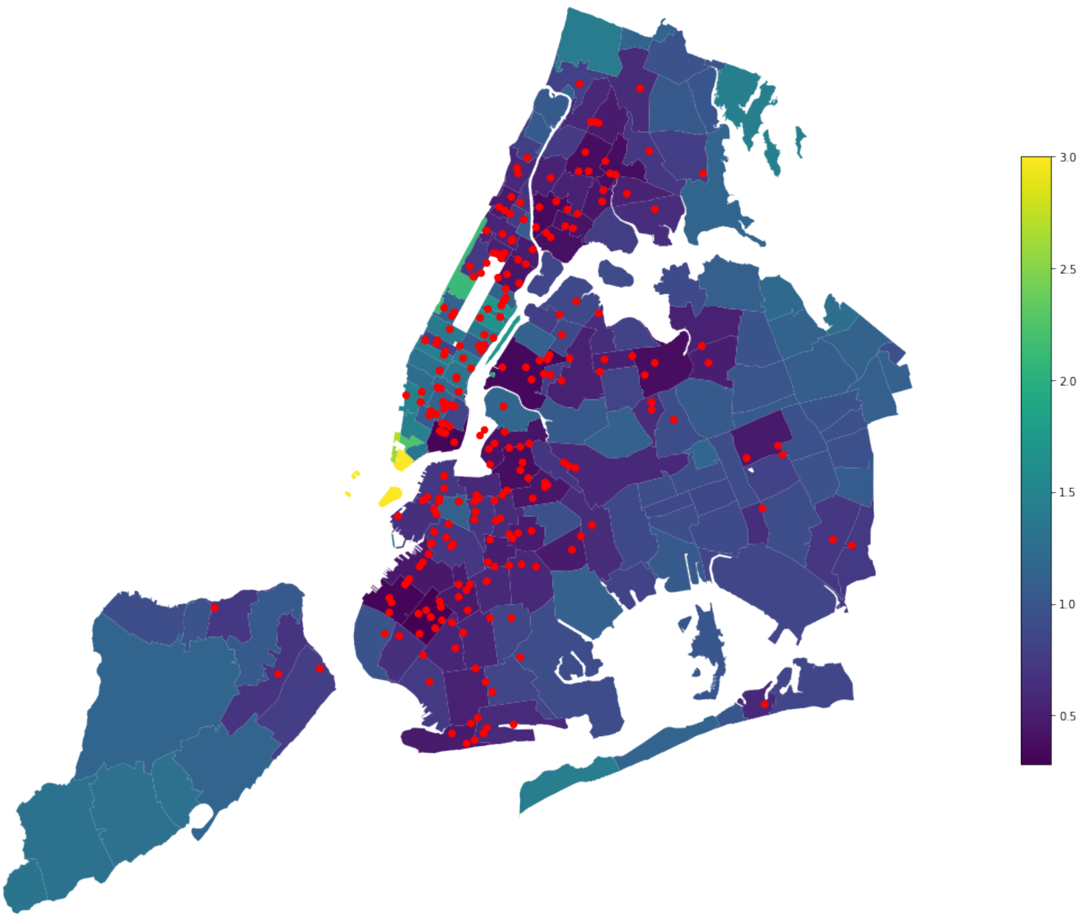
To address the problem's stated focus to improve bike lanes in socioeconomically disadvantaged communities as stated in the problem, we create a social index. This index captures how rich and employed inhabitants of different zipcodes are, which allows us to figure out what areas to focus on. This index is based on the median income and no-income percentage within each zip code and should shift the final development value lower and thus in favor of these communities. We use the expression:

$$S_{zipcode} = \frac{\frac{M}{N}}{S_{NYC}}$$

where N is the percentage of individuals who have no-income within the zip code and M the median income of the zip code. We chose to use a no income rate instead of unemployment as it provides a better metric which is able to take into account discouraged workers, students, and retirees. By dividing the median income by the proportion of individuals with no income, we establish an inverse relationship between the socioeconomic index and the overall socioeconomic index. As always, we end by dividing the overall value by the average socioeconomic index in NYC to normalize values.

The heatmap shows that Manhattan as the most affluent area, which corresponds to intuition. There are few outliers such as JFK airport area which has index of 0 since no-one lives there, and a bright yellow patch on the right end of the chart. That zipcode only includes part of a park and a hospital with very few people which results in anomaly. These anomalies do not impact final analysis.

2.3.3 Health Index



Heatmap of health index for each zipcode. Red dots indicate bicycle accidents in the past few years

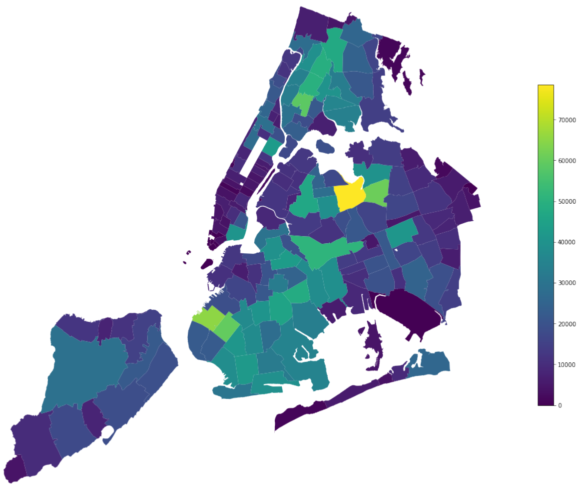
Our final index is the health index, which takes into account both the number of bicycle accidents and also the physical inactivity rate. The lower the index, the bigger need there is for bicycling infrastructure as a means to decrease rate of physical inactivity by encouraging people to bike or by mitigating number of accidents.

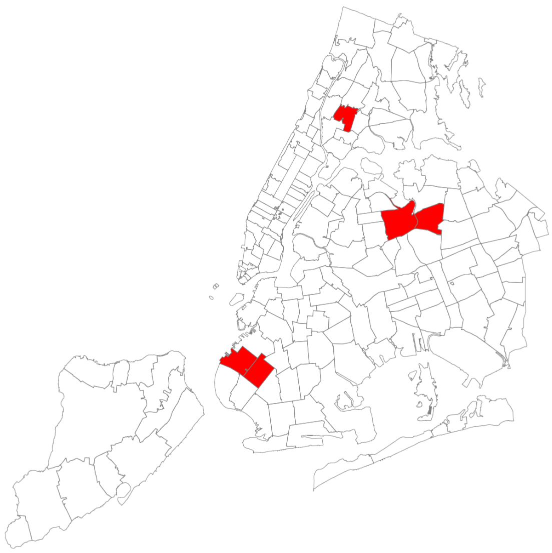
$$H_{zipcode} = \frac{\sqrt{a+1}}{p}$$

where p is the physical inactivity rate and a the number of accidents over the past 5 years. A square root function was applied to the number of accidents because of the potentially wide variation in the number of accidents alongside the sensitivity of a square root function at low crash numbers (i.e. there is greater difference between a zip code with 2 crashes and a one with 10 than one with 40 another with 48.)

The heatmap includes red dots which are places where cycling incidents occured in the past few years. The heatmpap corresponds to our intution that those regions need more work.

2.4 Applying the Model and Determining a Zip Code





Heatmap of priority index for each zipcode Top 5 zip codes that need biking infrastructure

We chose to weight transportation index by half, and health index by two, since transportation seems like it would be improved least by new bicycle infrastructure, and health the most. In the image below you can see the indexes (unweighted) for all top 5 zip codes that we suggest developing.

	ZIPCODE	priority	population	Ti	Hi	Ei	Bi
173	11368	78553.709043	109069.0	0.463886	0.374044	0.408433	1.298231
139	11220	65161.138386	97257.0	1.003751	0.322929	0.344828	1.504137
160	11355	60903.799113	82809.0	0.138269	0.509240	0.272054	1.404699
138	11219	59719.205634	92561.0	1.400030	0.280533	0.288856	0.473279
89	10456	58870.086284	89390.0	0.407698	0.533111	0.248357	1.385611

Indexes for top 5 zip codes. Ti is transportation access, Hi is health and safety, Ei is socioeconomic index, and Bi is the bike lane density for existing bike lanes. All values are relative to NYC average

Let's take a closer look at our first zip code, which corresponds to parts of Corona and Wellets points districts in Queens. As you can see from indexes, this region has access to transportation that is twice as bad as the average, health index that is almost threee times lower than average, and significantly lower than average socioeconomic index. Therefore, this district needs cycling infrastructure.

3 Comparing Protected and Conventional Bike Paths

Not all bike lanes are the same-there's a wide range of classifications based on how they are marked and/or protected. As mentioned in the assumptions section, this myriad of different bike path types will be simplified into two types:

1. **Conventional** bike lane infrastructure, which are portions of roadway that are designated to be used by cyclists by markings. The cost per mile based on past construction is approximate $ 180,000
2. **Protected** bike lane infrastructure, which are bike lanes physically separated from the road by a curb, car park space, or plants. In other words, it is like a sidewalk for bikes. The cost per mile based on past construction is approximate $ 600,000

Given this description, protected bike lanes are a safer option compared to conventional bike infrastructure. According to data collected by the NYC Department of Transportation, past implementations of protected bike lanes have reduced total injuries by 20%. At the same time, again considering network effects, the perceived safety and comfort derived from biking in a protected lane might further encourage increased usage of the lanes by current bikers and even convince new bikers who previously might have not been comfortable biking. This would broaden the demographics of bikers who regularly user bike lanes.

Our model aims to prioritizes safety. We do this through a two part model which takes in zip code specific inputs and returns a ratio for the mix between protected and conventional bike paths.

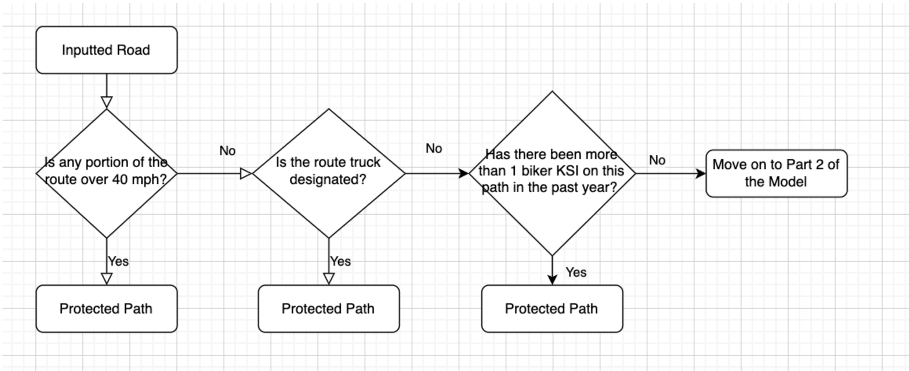
3.1 Assumptions

We assume that there is a binary choice of either protected or conventional bike paths for our model. Although technically there is a spectrum, assuming the binary state of paths allows for a clearer model without having to consider the myriad potential intermediate and varied bike paths.

Similarly, we assume that the costs to build protected and conventional bike paths remains fixed regardless of location in NYC. Small differences individual to each road resulting are negligible.

3.2 Part 1: Decision Tree

The first part of our model is a decision tree which determines on a road-by-road basis whether a protected lane is absolutely necessary for safety reasons.



We justify this tree as data from the Department of Transportation collected since 2014 have highlighted the increased risk bikers face on faster lanes. For example, over 53% of those KSI (Killed or Seriously Injured) in 2019 were due to a truck. At the same time, a road is designated necessary to be protected if more than a single KSI has occurred in the past year in order to account to any other potential hidden factors that are not included this simplified decision tree.

3.3 Part 2: A Sliding Scale

After applying a portion of funds to necessary investments for safety, we now use a sliding scale to determine the mix of protected and conventional infrastructure. It is important to note that although protected lanes are safer than conventional, over 90% of accidents actually occur on streets without bike lanes. Therefore, in the case of insufficient bike lane coverage within a zip code, the opportunity cost in investing in safer but fewer protected bike lanes might not be worth it.

We can quantify "insufficiency" in the length of bike lanes with a bike lane development metric b_{dev}:

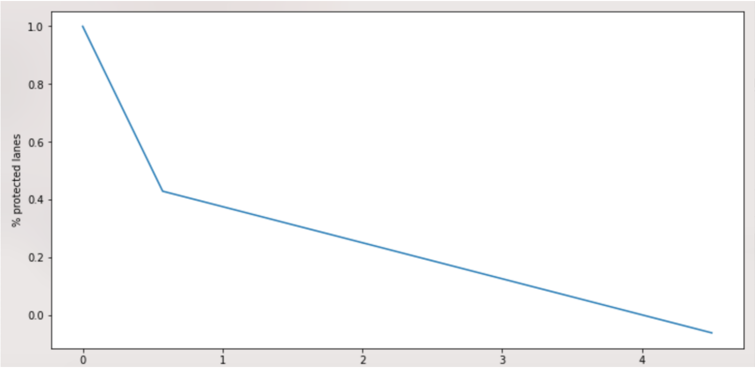
$$b_{dev} = \frac{B_l}{A} \frac{D_{NYC}}{D_{zipcode}}$$

where B_1 represents miles of bike lane within the zip code, A as the area of the zip code, $D_{zipcode}$ as the population density of the zip code and D_{NYC} as the population density of NYC. Taken together, the first fraction represents bike lane density, while the second fraction is a scaling factor which adjusts the first based on population.

By dividing this value by the bike lane development metric of NYC as a whole, we can obtain a ratio which represents the necessity of constructing new bike lanes. We can now split this into two cases:

1. Insufficient Coverage: If the ratio is less than or equal to 1, meaning that the bike lane coverage is less than or equal to the citywide average, we implement a linearly sliding scale where the percentage of new conventional lanes is 100% at a ratio of 0 and 50% to a ratio of 1.

2. On-par Coverage: If the ratio is greater than 1, we know that the bike lane coverage is above average. Therefore, we implement different linearly sliding scale where the percentage of new conventional lanes is 50% starting at a ratio of 0 and 0% at the maximum ratio of any zip code.



Sliding scale, where the x axis is the ratio of the zip code's bike lane development metric to the city's

3.4 Protected and Conventional Lanes Mix for Our Plan

We will aim to determine the specific mix of protected and conventional lanes for our first zip code 11368 near Corona, Queens. Beginning with the first step, we run through our decision tree. A main truck-designated road cutting through the zip code represents about 1.2 miles of needed protected lanes. At the same time, there are no streets with more than a single fatality. Thus, a total of 1.2 miles of protected bike lane at our assumed cost of $600,000 per mile of protected bike lane approximates to $720,000 in necessary safety spending.

Assuming that we proportion our budget equally among the five zip codes, we therefore have $730,000 left to utilize for the sliding scale in Part 2. Calculating the bike lane development metric for this zip code, and dividing by the ratio for NYC as a whole, we obtain a ratio of 2.72. Applying this ratio to the insufficient coverage sliding scale where the max ratio of all the zip codes is 4.5, we notice that the breakdown resulting percentage of new conventional lanes is 24.7% of the total funding left (or about 1 miles), meaning that the percentage of protected lanes is 75.3% of the funding (or approximately 0.91 miles).

4 Predicting Future Bike Ridership and Our Plan's Impact on Traffic

Given the third problem statement, determining the relationship between our plan's construction of new bike lanes and the resulting impact on traffic is essential. Therefore, in this section, we attempt to develop a general model that determines how a given amount of investment into bike infrastructure will affect the proportion of the population that uses a car versus rides a bike. In the model, we will attempt to consider the key factors that our investment plan will target.

4.1 Compartmental Modeling

Compartmental modeling is a popular technique used to analyze population dynamics, particularly in a fixed population. In compartmental modeling, all individuals in the total sample size is classified into a given category. Here, we will classify all our commuters of interest as either a member of C, reflecting that they travel via car for their daily commuter, or a member of B, indicating that they commute via bike.

As the simulation progresses, individuals can move between C and B, but they cannot leave the model altogether (hence the assumption of fixed population size).

4.2 Assumptions

As with all forms of dynamical modeling, we will establish some fundamental assumptions that ensure the model's validity, as well as provide some justifications for their usage:

- Target population: the total number of people who either commute by car or by bike. We only consider these people because we assume that people with other forms of transport like public transport will still stick to them as well long as they have access.
- Fixed population: we can ignore births and deaths as they are negligible in the scale of this problem.

4.3 Deriving the Model

In this model, we will model the transition rate between the two categories, C and B, using parameterized ordinary differential equations with time t as the dependent variable.

To determine the rate $\frac{dC}{dt}$, we must examine the relevant factors that may influence an individual's decision to transition from using a car for their commute to biking. In doing so, we hope to analyze and parameterize details of the system that will have a notable impact on ridership without considering too many variables that may complicate the model and generate excess noise. Therefore, the four factors that we propose must be considered are:

- Amount of conventional bike lanes
- Amount of protected bike lanes
- Social pressure to increase bike ridership
- Injury frequency among cyclists

Though other factors like quality of bike lanes, availability of bike stops, and the development of other public transport systems may certainly impact bike ridership, we argue that they need not be considered because they are either irrelevant to this case study, impossible to predict, or have too much of a negligible impact relative to the four parameters proposed above.

4.3.1 Conventional Bike Lanes

First, we introduce a parameter μ that models the impact that conventional lanes have on ridership. Specifically, μ represents the impact on ridership that one additional mile of conventional bike lane has on ridership. As the amount of conventional bike lanes increases, car users will gradually transition to biking at a rate proportional to the current number of car users:

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -\mu C$$
$$\frac{\mathrm{d}B}{\mathrm{d}t} = \mu C$$

4.3.2 Protected Bike Lanes

Next, we introduce a parameter ρ that models the impact that protected bike lanes have on ridership, with a similar definition as the one given for μ above. As the amount of protected bike lanes increases, car users will again transition to biking at a rate proportional to the current number of car users. However, literature suggests that protected bike lanes are much more popular than their conventional counterparts; protected bike lanes also occupy distinct spaces on roads, reducing the total number of commuters possible on a given road. Therefore, protected bike lanes are much more likely to be crowded than conventional lanes, particularly in areas of dense traffic, meaning that the total number of bikers should also be considered when modeling the transition rate due to the expansion of protected bike lanes, with higher transition rates occurring when there are fewer bikers.

$$\begin{aligned}\frac{\mathrm{d}C}{\mathrm{d}t} &= -\mu C - \rho C \\ \frac{\mathrm{d}B}{\mathrm{d}t} &= \mu C + \rho C\end{aligned}$$

Because protected lines are more attractive for bikers, both intuitively and as supported by existing research, we would suspect that $\rho > \mu$.

4.3.3 Social Pressure

Research also suggests that there is a social element involved in bike ridership [1]. When commuters are encouraged to use bikes, whether for social, economic, or environmental reasons, there is a notable shift to bike ridership. Therefore, we introduce a parameter ϕ that governs the impact that social pressure has on transition from car to bike ridership. Generally, more people will feel pressured to begin biking when they know or observe more bikers on the roads. We propose that this relationship is exponential (i.e. when people observe that there are lots of bikers on the road, there are even more inclined to start biking).

$$\begin{aligned}\frac{\mathrm{d}C}{\mathrm{d}t} &= -\mu C - \rho C - \phi e^{C} \\ \frac{\mathrm{d}B}{\mathrm{d}t} &= \mu C + \rho C + \phi e^{C}\end{aligned}$$

4.3.4 Injury Rate Among Cyclists

The last element that we consider is the injury rate among cyclists. As expected, extensive research shows that higher frequencies of biking-related injuries and fatalities significantly reduce willingness to bike [1]. Therefore, we introduce a parameter γ that represents the impact that cycling-related injuries and deaths have on discouraging ridership. As the number of such incidents increases, γ increases correspondingly. The amount of people who decide to transition back to driving cars because of biking-related incidents is dependent on the current number of bikers, meaning that γ and B must be directly related. This gives way to the final form of the model:

$$\begin{aligned}\frac{\mathrm{d}C}{\mathrm{d}t} &= -\mu C - \rho C - \phi e^{C} + \gamma B \\ \frac{\mathrm{d}B}{\mathrm{d}t} &= \mu C + \rho C + \phi e^{C} - \gamma B\end{aligned} \tag{1}$$

Note that the term γB is the only factor in this model that can cause bikers to revert to commuting by car.

4.4 Model Summary and Definitions

The compartmentalized system of differential equations given by (1) is the model for bike ridership that will be used in this study.

Observe that from system (1), we see that

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{\mathrm{d}C}{\mathrm{d}t} + \frac{\mathrm{d}B}{\mathrm{d}t} = 0$$

which verifies that N is a constant, as given by the second assumption outlined above. This means there should be leaks or entries of individuals in our simulations, and the sum of C and B should remain the same throughout.

To summarize, we have

- C: the proportion of the total population that commutes via car.
- B: the proportion of the total population that commutes via car.
- t: time in days.
- μ: the impact that conventional bike lanes have on bike ridership, as measured by the increase in ridership from one additional mile of bike lane.
- ρ: the impact that protected bike lanes have on bike ridership, as measured by the increase in ridership from one additional mile of bike lane.
- ϕ: the social pressure to transition to bike ridership.
- ρ: the impact that cycling-related injuries and deaths have on discouraging bike ridership.

4.5 Determining Parameter Values

Determining appropriate parameter values using preexisting research and real-world data is crucial for accurate modeling.

4.5.1 μ and ρ

pμ and ρ govern the impact of conventional and protected bike lanes on car usage. To compute their default values, we will analyze historical data regarding bike lanes in New York City and their corresponding ridership levels. Gu, Mohit, and Muennig (2015) find that on average, 45.5 miles of new bike lanes constructed in New York City increase ridership by a factor of approximately 0.0932 [2]. Xu and Chow (2019) find that approximately one-fifth of such lanes are protected and the remaining four-fifths are conventional, while research from New York City's Department of Transportation finds that developments in protected bike lanes and conventional bike lanes were each responsible for approximately 50% of the corresponding increased ridership [3].

Therefore, we approximate μ and ρ as follows:

$$\begin{aligned} \mu &= \frac{0.5(0.0932)}{0.8(45.5)} = 0.00128 \\ \rho &= \frac{0.5(0.0932)}{0.2(45.5)} = 0.00518 \end{aligned} \tag{2}$$

4.5.2 γ and Risk of Injury

γ represents the impact of cycling-related injuries on discouraging individuals from commuting by bike. We attempt to determine γ before ϕ because, as will be examined in the next section, the value of ϕ cannot be deterministically derived. As the perceived risk of an accident rises correspondingly to the actual number of bike-related accidents, we will define ϕ to be given by the annual incident rate of cycling-related accidents multiplied by the extent of the impact that crash/injury risk has on discouraging commuters [4]:

$$\gamma = \frac{\text{number of bike-related accidents}}{\text{total number of rides}} = \frac{18,718}{530,000} = 0.035 \tag{3}$$

4.5.3 ϕ and the Impact of Social Pressure Using Sensitivity Analysis

ϕ represents the social factor associated with transitioning from commuting by car to commuting by bike. Extensive literature justifies the need for this parameter by outlining the existence of social pressure in relation to commuting via bike, in that individuals are more likely to transition to cycling if they observe more people around them doing in the same.

However, it is difficult to quantify the impact of this parameter, as its value is significantly more arbitrary than the ones above. Therefore, in order to approximate ϕ, we perform a sensitivity analysis where ϕ is the independent variable and $B(1100)$ is the dependent variable, where $t = 0$ represents the system at the beginning of 2015. Given this initial condition, the state of B at $t = 1100$ will represent the proportion of the population that commutes via bike at approximately the beginning of 2018, which we have ridership data available for. From this simulation, we can then select the value of ϕ that best corresponds with the actual data [4].

Data from the New York City Department of Transporation states that in 2018, approximately 8.9% of commuters biked, compared to 5% in 2015. Therefore, we will conduct the simulation with $B(0) = 0.05, C(0) = N - B(0) = 0.65$. We search for the value of ϕ such that $B(1000) = 0.089$.

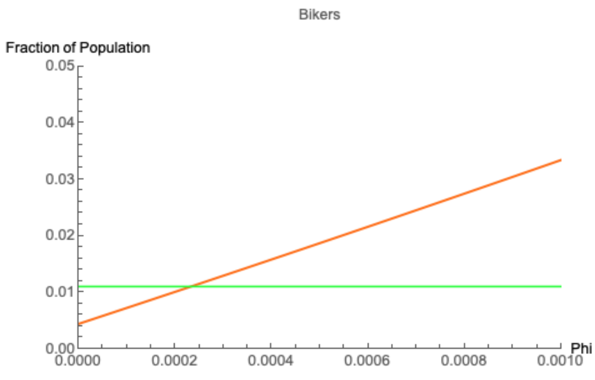


Figure 1: Sensitivity Analysis of ϕ

The Solve function in Mathematica computes that specifically, $\phi = 0.000233$.

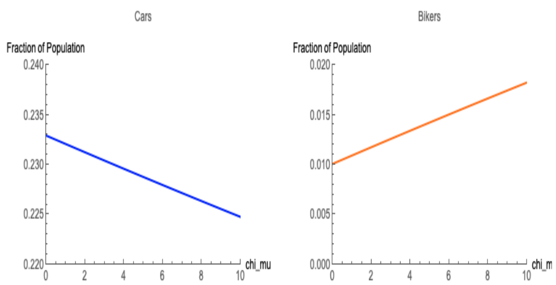
4.6 Predicting Future Ridership

Now with a complete model and accurately determined parameter values, we can tune individual parameter values in our model to study the impact of our developments on car ridership.

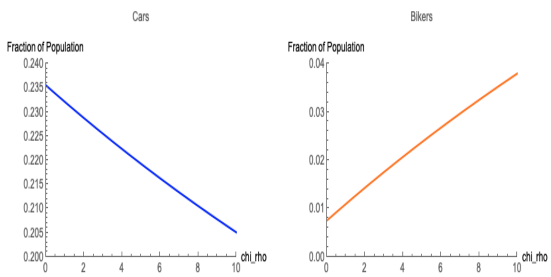
Specifically, we are interested in varying the values of μ, ρ, and γ. We argue that the policies and developments we enact will not affect ϕ directly, but instead only implicitly through changes in C. Therefore, we introduce tuning parameters $\chi_\mu, \chi_\rho, \chi_\gamma$ that scale the values of their corresponding subscripted parameters.

$$\begin{aligned}\frac{\mathrm{d}C}{\mathrm{d}t} &= -\chi_\mu \mu C - \chi_\rho \rho - \phi e^B + \chi_\gamma \gamma B \\ \frac{\mathrm{d}B}{\mathrm{d}t} &= \chi_\mu \mu C + \chi_\rho \rho C + \phi(1-C)B - \chi_\gamma \gamma B\end{aligned} \tag{4}$$

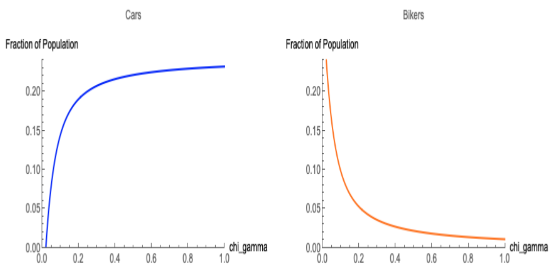
Of course, the default value of $\chi_\mu, \chi_\rho, \chi_\gamma$ is 1. However, we can now tune the values of χ to scale each tuning parameter's value to investigate the impact that the corresponding parameter. In the following groups, the independent variable is a tuning parameter and the dependent variable is the state of the system at $t = 1000$, where $t = 0$ represents the start of 2018. All other parameters besides the tuning parameter are held constant.



(a) **Commute Breakdown vs.** χ_μ



(b) **Commute Breakdown vs.** χ_ρ



(c) **Commute Breakdown vs.** χ_γ

Figure 2: Car Ridership and Bike Ridership vs. $\chi_\mu, \chi_\rho, \chi_\gamma$. Note the different scales used on both axes throughout all three subfigures.

4.7 Discussion

Figure 2 provides a visual representation of our results. Figure 2a and 2b show that increasing the amount of conventional and protected roads both contribute to decreasing ridership, with strong effects coming from protected roads, which agrees with what we would expect. However, Figure 2c shows perhaps the most interesting result, which is that decreasing injury risk has the most significant impact on increasing ridership.

For instance, increasing the amount of conventional bike lanes by ten times increases bike ridership by twofold, while increasing the amount of protected bike lanes by ten times increases ridership by a factor of four. However, reducing the likelihood by injury by five times can increase ridership by over fivefold, with greater and greater effects occurring as injury risk decreases beyond that. Therefore, this model provides strong evidence in favor of protected bike lanes, which generally offer increased safety for pedestrians relative to conventional lanes.

5 Strengths and weaknesses

5.1 Strengths

A key strength of our locational model is the fact that it doesn't just take in inputs regarding current ridership, but also data on access to other forms of public transportation. It also uniquely incorporates median income and accident rate data from the zip code. This combination of accessibility, socioeconomic and public health data and weighting provides a holistic perspective on our goal of improving equitable access to transportation. We can see the benefit of this approach crafted specifically to address the prompt in the identification of unique zip codes, the majority of which had not been identified as key areas needing investment in the most recent NYC.gov ridership reports.

Our parameters are also adjustable, allowing for a level of model flexibility specifically catered towards policymakers who can adjust the plan should there be a political reevaluation of priorities (e.g. emphasizing prioritizing accessibility to transportation by applying an emphasis scaling parameter) instead of scrapping the plan altogether.

5.2 Weaknesses

A potential weakness may lie in our separation of the problem into three models. Compartmentalizing the three models might make it difficult to integrate into a single plan, as different models may consider variated inputs and could potentially unknowingly place emphasis differently on certain values. A single model might have consolidated our assumptions and distilled our zip code findings more clearly. However, our team believed that there would be too many parameters for such a model which would have greatly increased the chance of overfitting

Our model evaluates the need for biking infrastructure, rather than its future impact. Evaluating future impact is possible by using historical data about biking lane expansion and relevant metrics. A lot of the data needed is inaccessible or temporally and spatially sparse, which makes implementation of an impact estimation model much more difficult

Our model utilizes zipcodes as the fundamental building block instead of specific streets. As a result, our model is discrete in terms of where to build new bike lanes. It is possible to evaluate the specific streets to build bike lanes with access to more spatially dense ridership data, such as Strava data.

Our model calculates bike lane length, subway station number and bus station number contained inside the zipcode. This representation is tabular and does not account for bike lanes that run on the boundaries of zip codes, subways that are in neighbor zip codes, etc. This is a by product of doing zipcode-level analysis and could only be avoided by considering more granular spatial units.

6 Conclusions

We had three tasks for this challenge:

1. Identify an area for investment
2. Determine the mix of protected and conventional bike lanes
3. Determine the impact of the plan on traffic

We took a unique approach in delegating a model to address each of the three challenges. For the first problem, with a focus on equitable access, we used a zip code finding model built upon a trivariate sum based on currently accessible infrastructure index, socioeconomic index, and health. We determined 5 locations, with **11368** by Corona, Queens given the highest priority.

We then developed a model heavily based on safety as the key value to determine our distribution between protected and unprotected lanes. Protected lanes are safer, but bring with it an opportunity cost of fewer miles given a limited budget. Therefore, our model took in the amount of existing biking infrastructure to conduct a cost-benefit analysis to determine the most optimal mix of lanes.

Finally, using differential equations, we were able to roughly predict the impact of our plan on traffic and commuter patterns.

7 Letter to DOT

DATE: February 2022

Dear Mr. Ydanis Rodriguez,

Our team has been tasked with preparing a plan which most effectively utilizes the recent $ 7.25 million grant to expand current biking infrastructure. Using historical government census data for each zip code, we developed a series of models which suggested a handful of locations to invest in biking infrastructure, the breakdown of the type of bike lane, and potential impacts on traffic and future bike ridership

Our first model focused on identifying a handful of zip codes for investment with the goals of improving equitable transportation access for low-income community and improving biker safety. We did this by integrating data from each zip code on the accessibility of current transportation options, socioeconomic status, and the number of accidents into a single development value. The lower the development value, the more in need of investment the zip code. Ultimately, we identified the following locations in order of priority for new bike lane investment:

1. **11368**, near Corona, Queens
2. **11220**, near Sunset Park, Brooklyn
3. **11355**, near Flushing, Queens
4. **11219**, near Borough Park, Brooklyn
5. **10456**, near Throgs Neck, The Bronx

With these potential locations in mind, we then determined an optimal split between protected bike lanes and conventional lanes with a two-step model which emphasized safety. The model first required potentially dangerous roads to have protected bike lanes rather than conventional ones. Then, with the leftover money if the current number of bike lanes was deemed insufficient compared to NYC average, the model prioritized conventional lanes. This decision was in light of the fact that according to the NYC DOT, over 90% of bike related fatalities occurred on roads without any bike lanes, so lane coverage is still key in maximizing biker safety.

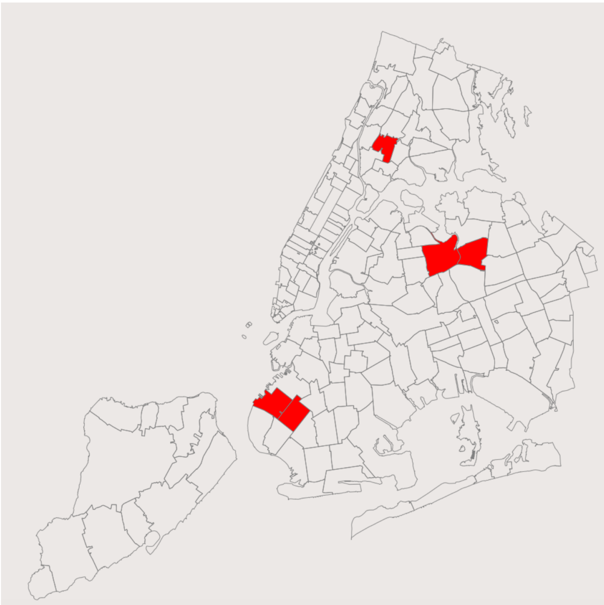


Figure 3: Map of priority zip codes

We then utilized **11368**, our most optimal location, as a sample to determine the split between investments in protected and unprotected lanes, obtaining a roughly 50-50 split in terms of distance after applying a necessary 1.2 mile protected bike line adjacent to a high speed corridor.

Our final model determined the potential impact our aforementioned proposed model has on traffic. Using a parameterized compartmental model analyzing the transition rate between car ridership and bike ridership, we studied the impact that new bike lanes, social pressure, and risk of injury would have on traffic flow. We concluded that although bike lanes, both conventional and protected, were important for reducing traffic, the key influential factor was the improvement of safety standards that can minimize cyclist injuries and fatalities. We also concluded using a sensitivity analysis that there is a significant impact of social pressure that encourages commuters to start biking once they observe those around them doing the same.

Although we are confident in our models, we believe that it is important to disclose potential limitations of our model. Our locations and the mix of protected and conventional bike lanes is based on past data, meaning that it's relevance could be rendered insignificant in the case of unseen changes in the current environment.
At the same time, our model focuses improving equitable access focused on communities rather than purely optimizing riders. As a result, sections of area like zip codes were analyzed rather than individual roads. This is further challenged by the fact that we focused exclusively on zip codes as locational units due to the uniformity of data available.

However, we highlight this model's strength in its adaptability. We suggest multiple specific viable zip codes as optimal locations for investment. Our model determining priority investments also has built-in adjustable parameters for the three aforementioned inputs into the development value. Depending on a policymaker's preference or focus on either accessibility of current transportation options, socioeconomic status, and/or the number of accidents for potential future plans, the policymaker will be able to adjust accordingly and obtain appropriate zip codes.

We hope that you find our plan amenable, and look forward to seeing its implementation
Sincerely,

Our Team

References

[1] Sergio A Useche, Luis Montoro, Jaime Sanmartin, and Francisco Alonso. “Healthy but risky: A descriptive study on cyclists’ encouraging and discouraging factors for using bicycles, habits and safety outcomes”. In: *Transportation research part F: traffic psychology and behaviour* 62 (2019), pp. 587–598.

[2] Hannah Rebentisch, Rania Wasfi, Daniel P Piatkowski, and Kevin Manaugh. “Safe streets for all? Analyzing infrastructural response to pedestrian and cyclist crashes in New York City, 2009–2018”. In: *Transportation research record* 2673.2 (2019), pp. 672–685.

[3] Susan Jia Xu and Joseph Y. J. Chow. “A longitudinal study of bike infrastructure impact on bikesharing system performance in New York City”. In: *International Journal of Sustainable Transportation* 14.11 (2020), pp. 886–902. DOI: `10.1080/15568318.2019.1645921`. eprint: `https://doi.org/10.1080/15568318.2019.1645921`. URL: `https://doi.org/10.1080/15568318.2019.1645921`.

[4] PPM. “NYC Bicycle Safety Overview: Infrastructure & Crash Stats”. In: (2020).