Investigating air quality impacts of vehicle electrification strategies for equity assessments

ESENG 503 project report

Submitted to:

Dr. Sita Syal

Mechanical Engineering

3556 GGB

Prepared by:

Marco Marcial

ISD Energy Systems Engineering

08/12/2025

Contents

Executive Summary

Amidst growing preoccupation with controlling climate change and assessing environmental damage, the electrification of the vehicle fleet proves to be a crucial measure when moving towards a carbon neutral future. Still, there are challenges to be overcome in order for this transition to be fruitful, one of them being the uneven distribution of electric vehicle adoption along varying socioeconomics. This project aims to access this distribution by constructing a model to analyze the distribution of vehicle emissions over the ZIP codes of the Santa Clara County, and, later, of the San Francisco Bay Area, allowing a better understanding of how emissions are distributed through different social groups, areas and the roles played by electric vehicles on this. These findings will then allow for social justice and healthcare measures to be taken in the realm of policymaking.

In order to better analyze this issue, the model integrates multiple datasets, including statistics on commute patterns for the Santa Clara County, database for vehicle emissions rates and fleet composition and spatial limits for ZIP code geographical boundaries. The model itself takes these datasets as inputs in order to calculate commute routes between two separate points using the Open-Source Routing Machine method and determine through which ZIPs these routes traveled through, as well as the pollutants emitted during them. After running for the entirety of the county, which spanned around 550 thousand origin and destination pairs and approximately 9.2 hours of processing time, the model outputted three separate datasets, which included aggregate emissions by ZIP code areas that received them, emissions for ZIP codes where trips that generated them began and emission for ZIP codes where trips generating them ended.

Analysis was conducted with these data outputs, which were focused on NOx as a representative pollutant, show that for all three datasets a small number of ZIPs account for a disproportional number of emissions, being that on the receiving end, as the origin of many commutes of the destination of several routes. The data was then cross-referenced with median household income data from the US Census Bureau and revealed that there is an overall tendency for lower-income ZIP codes to experience higher pollution exposure, while higher-income ZIPs are less susceptible to be in the receiving end of emissions but are more common destinations of commute routes. These findings corroborate with prior research that link vehicle traffic and emissions are more intense in lower income communities and also bring up the need for environmental justice analysis in the context of urban transportation.

The current stage of development of the project has produced a scalable and validated model that has the capacity of providing data to assist in equity-focused environmental analyses. Future steps and recommendations include making the overall statistic analysis conducted more robust by normalizing emissions outputs by population, area and employment, as well as increasing the number of socioeconomic indicators used in this study.

1.0 Introduction and current state

With the growing preoccupation with controlling climate change, attacking some of its main contributors has become a point of much attention. In this context, the electrification of transportation vehicle fleets through the incorporation of electric vehicles (EVs) has become a central measure of emission reduction and leading towards a carbon neutral future. Still, there are issues with the way this electrification is taking place, as the distribution of acquisition of EVs is not homogeneous between varying socio-economic and demographic populations (Hennessy, Syal 2023, Canepa et al 2019). This project aims to access this distribution by constructing a high-resolution proof of concept model to analyze the distribution of vehicle emissions over the different ZIP codes of the San Francisco Bay Area, which would allow, within other aspects, to better understand how emissions are distributed through different social groups and areas and the roles played by electric vehicles on this, allowing for social justice and healthcare measures to be taken on the realm of policymaking.

1.1 Overview of the Project

The transportation sector is one of the main contributors to air pollution in the United States, with internal combustion engine transportation vehicles being responsible for a great emission of various toxic gases and particulate matter (U.S. EPA 2023), which can bring serious consequences to both the environment as well as for the health of people. It is also valid to mention that due to the nature of such emissions, people of color are those mostly exposed to these pollutants (Tessum et al 2021), while increased vehicle traffic is specially linked to decreasing income and increasing non-white population (Rowangould 2013).

With these factors being of consideration, the reduction of emissions brought by the transportation sector has become an increasing priority around the world. One of the key answers to this issue lies in the electrification of the vehicle fleet. Even if in varying degrees of effectiveness in different states, especially due to different compositions of the energy matrix, there is a general reduction of emissions when looking at EVs in comparison to traditional internal combustion (IC) engine vehicles (Chroma et al 2020) as well as allowing for emissions not being concentrated along vehicle paths. It is also noted that, while CO2 is the primary and most abundant gas emitted by vehicles, its effects are global rather than local, contributing mainly to climate change rather than affecting the areas where it is being emitted. In contrast, other pollutants such as PM2.5, SO2, NOx, NH3 have significant impacts on local health, as well as the environment. These pollutants have direct contributions to respiratory illnesses as well as mainly affecting communities located near to high traffic areas.

Still, even with the electrification of the vehicle fleet being underway, the adoption of electric vehicles can be more strongly observed in higher income neighborhoods, while

disadvantaged communities do not follow this trend and suffer with a lack of proper infrastructure that can support the adoption of EVs (Hennessy, Syal 2023, Canepa et al 2019). These differences in EV adoption rate as well as the travel patterns of individuals living in distinct communities directly affect the emissions and the air quality of different locations of a geographical region.

In this context, the elaboration of policies that can promote the transition to EVs becomes an important topic for local governments. Moreover, in order to efficiently determine to which extent such policies must go and how efficient they would be in bringing environmental, health and societal impacts, tools that can evaluate health and equity outcomes make themselves necessary. With this in mind, the project aims to build a high-resolution fleet turnover model paired with a high-resolution traffic and emissions model (both in zip-code level) to assess air quality impacts and inequities resulting from fleet turnover and vehicle electrification policies. The project will be initially focused on the Santa Clara County for initial analysis and assessment of the proposed model and layer expanded to the San Francisco Bay Area and the entirety of the state of California to demonstrate its utility while allowing for even further expansion and future use in other projects.

Figure 1: Map of Santa Clara County

1.2 Review of Current State Design/Literature Review

The analysis being developed by this project is of great complexity due to incorporating many different aspects in order to allow for the proposed analysis to be efficiently conducted. In that light, the understanding of previous works in the areas related to this project is crucial to successfully move on with the project. This section will go into relevant literature that will serve as a basis to understanding future work and which will be built upon as the project advances and more robust analysis become possible.

In *An early look at plug-in electric vehicle adoption in disadvantaged communities in California* (Canepa et al 2019), the authors bring up the topic of plug-in electric vehicle (PEV) adoption and the benefits brought by it suffer differences through different income levels of society, which can be clearly seen since most early PEV adopters were wealthy consumers. Moreover, their research also pointed out how lower-income communities were the ones to suffer the most with the impact of environmental and transportation justice. The study also showed that even though there is a charger structure in lower income communities for the adoption of PEVs, many barriers also exist, such as the prohibitive price of the technology, lack of knowledge about or ease of accessing PEV incentives. It is also valid to point out that the method used for this study supports what is being done for the current project, as both utilize census tract data to look at the impacts of electric vehicles in different communities.

In *Assessing the health impacts of electric vehicles through air pollution in the United States* (Chroma et al 2020), it becomes possible to understand how vehicle emissions might have relevant health impacts in communities that are being exposed to them. For this, the study investigated mortality impacts per mile caused by fine particles from internal combustion engine vehicle (ICEV) tailpipe emissions of PM2.5, SO2, NOx, NH3, and volatile organic compounds, and power plant emissions of PM2.5, SO2, and NOx. The study then goes to calculate (in ¢/mile) the health benefits brought by EV adoption in different locations of the United States. These findings are especially relevant when looking at them through an optic of policy making, as they allow for an understanding of the economic benefit brought by an electric vehicle over its lifetime, which then allows for concrete and viable values for policy incentivizing the purchase of EVs to be made.

Prior analysis on the topic has also been done by Dr. Sita M. Syal and Dr. Eleanor Hennessy (both of whom conceptualized this project), in *Assessing justice in California's transition to electric vehicles* (Hennessy, Syal, 2023), which studied justice measures in three different areas: distribution of electric vehicles, allocation of state incentives, and the social and historical context of redlining. With this, it became possible to see how California's electric vehicle transition has not been just up to the current point.

1.3 Stakeholders

Considering the relevance of the model being developed by this study and taking its implications in the spheres of both health and equity, it is of great importance to clearly define the stakeholders related to it. Being a research development that is not being funded by any external organizations, the stakeholders for this project would be mainly comprised of the groups that are being most affected by its outcome. The target audience of the work that is being conducted would be policymakers that could use the findings obtained by this study as a basis to elaborate new policies in the areas of social justice, health and incentivizing a more equal level of electrification over different social groups. The second main stakeholder group would be comprised of groups that might be affected by policies that come from the findings of this study. These could be, but are not limited to, inhabitants of disadvantaged communities, individuals that are frequently exposed to vehicle pollution and potential new EV adopters that would benefit from stronger incentives.

1.4 Scope

This project will initially set its model to make an analysis of vehicle commute routes and emissions in the Santa Clara County, which will allow for a testing of the model in a region with good rate of EV adoption. Once this is done and the model is proved to successfully work on a smaller geographical area, it will be then ready for scaling, allowing for larger regions to be put into analysis, such as taking in data from the whole state of California.

1.5 Project deliverables

The project will aim to achieve a series of deliverables over its duration. Among them, we have a model to determine commute routes between two different census tract locations, emissions per mile traveled for passenger vehicles of varying ages and a model to estimate emissions for each of the commute routes previously determined. With these tasks being completed, it will then be possible to estimate which ZIP codes are the ones responsible for the greatest amounts of emissions and the project will be then finalized by adding demographic and socioeconomic data to the model and comparing ZIP codes that are causing and receiving emissions through different socio demographic groups. These deliverables put together would encompass a high-resolution model that could then be efficiently used to assess air quality impacts and inequities resulting from fleet turnover and vehicle electrification policies.

1.6 Project Timeline

Below is an overarching timeline for the project, with key due dates and tasks being organized as they are expected to be finalized and delivered:

- **Task 1:** Look into options for determining commute routes between two ZIP codes: possibilities include using Google Maps routing API, or using Dijkstra's shortest path algorithm (5/12)
    - Read proposal
    - Read 3-5 papers that are relevant
- **Task 2:** Download and familiarize yourself with commute data (LODES) (5/19)
- **Task 3:** Determine commute routes between each ZIP code origin-destination pair (OD pair) (6/2)
- **Task 4:** Extract a set of emissions factors per mile travelled for passenger vehicles for specific ages (PM2.5, SO2, NOX, VOCs, NH3) (EMFAC dataset) (6/9)
- **Task 5:** Develop a toy model estimate emissions in ZIP codes along commute routes from one ZIP code (6/23)
- **Task 6:** Scale up model and estimate emissions in each zip code caused by vehicles registered in each zip code for full set of zip codes in current year (7/14)
- **Task 7:** Analyze which zip codes cause the most emissions vs. which zip codes experience the most emissions (7/28)
- **Task 8:** Add demographic and socioeconomic data and compare zip codes causing vs receiving emissions across socio demographic groups (8/11)
- **Task 9:** Write up findings in final report (August 12)

2.0 Assumptions and methodology

Assumptions

It was necessary to make some assumptions to make the development of the model possible given the available data and limitations in the timeframe it was conducted in. Among the assumptions, due to limited geographical accuracy of some sets of data, the centroids of each geographical entity (either ZIP code or Geographical Identifiers) were used as coordinates whenever specific geographical points of these regions were needed, such as making estimates for the commute routes between an origin and a destination point, which are initially given only through a 15 digit GEOID (Geographic identifier) and assumed to be located in the centroid of such entity. Moreover, the model only considered emissions generated by commuters in the chosen area, fully excluding any other form of polluting sources from the scope of this project. Also, since it is not possible to determine the exact route that is taken by each commuter, the ideal route determined by the chosen routing method will be assumed to determine the distance traveled and through each ZIP codes each commuter went through when analyzing the impact, location and quantity of their emissions. A lack of accuracy was also noticed between the different chosen routing methods, something that will be further explained later in this section. Due to this, the initial model did not consider these differences and assumed that the routes generated by the OSRM (Open-Source Routing Machine) accounted for the most optimal paths. Since the baseline data used to calculate commutes not including precise mapping of what vehicle types were included in it, it was assumed that these were limited to passenger cars and light duty trucks, excluding collective public transportation, such as buses, for example, from the model's analysis. Finally, due to the analysis being limited to the Santa Clara County, commute routes were also limited to those whose starting and ending points were contained within the county.

2.1 Commute Route Mapping

As an initial step to work towards the fleet turnover model, it was necessary to find a way to analyze daily vehicle routes in order to trace their emissions profile effectively. For this end, the LODES (Longitudinal Employer-Household Dynamics Origin Destination Employment Statistics) dataset was used. This dataset was obtained from the United States Census Bureau and is a part of what is called the LEHD (Longitudinal Employer-Household Dynamics) program, which is an initiative of the Center for Economic Studies at the US Census and produces cost effective, public-use information that combines federal, state and Census Bureau data on employers and employees (U.S. Census Bureau, 2020). In this context, the LODES dataset is classified as Job-to-Job Flows (J2J) data, which is a set of statistics on job mobility on the United States (U.S. Census Bureau, 2020). The data itself comprises state level data on daily commutes for the citizens inhabiting a specific area and gives information for both the locations of the households of

commuters as well as the final destination of their commute. Said locations are given by very granular and precise GEOID data (15-digits). As a note, the greater the number of digits present on a GEOID, the smaller the area represented by it will be, with 15 digits being the maximum number of digits possible for this geographical delimiter. Moreover, for the sake of simplicity, these household and work location pairs will be called OD pairs from now on. Having established this, 2022 data for California was downloaded from the LODES database for JT00 (Job Type 00, total amount of jobs for designated region) with said data later being subjected to a filtering process to reduce it to commutes occurring within the Santa Clara County. With this done, it was possible to move on and begin working on a method to determine the commute routes between each of the county's OD pairs.

An algorithm for routing was constructed utilizing Python code and used three potential different methods to reach its desired goals: the Google Maps routing API, direct implementation of Dijkstra's algorithm in python and utilizing the OSRM method. More detail will be later given into each of these different routing tools, as well as their pros and cons, but they all worked, even if in different degrees of precision and efficiency, towards a similar goal, which was to establish the most optimal commute path being taken between each OD pair analyzed for the scope of the model. The code worked by obtaining the center point of the GEOIDs for both the origin and destination points of each OD pair, feeding the latitude and longitude coordinates of said points to the chosen routing methods in order to allow them to find the most efficient route that could be done between them, as well as through which ZIP codes these routes passed and the total number of miles driven in each of these ZIPs. Finally, the routes were recorded and passed on to other sections of the model in order to allow it to calculate emissions in each ZIP code. Below is an example of how one of these routes would be traced using Open Street Maps. Even though the map itself is not generated in the final model, the figure is a visual representation of the type of path that is being established and how the routes behave.

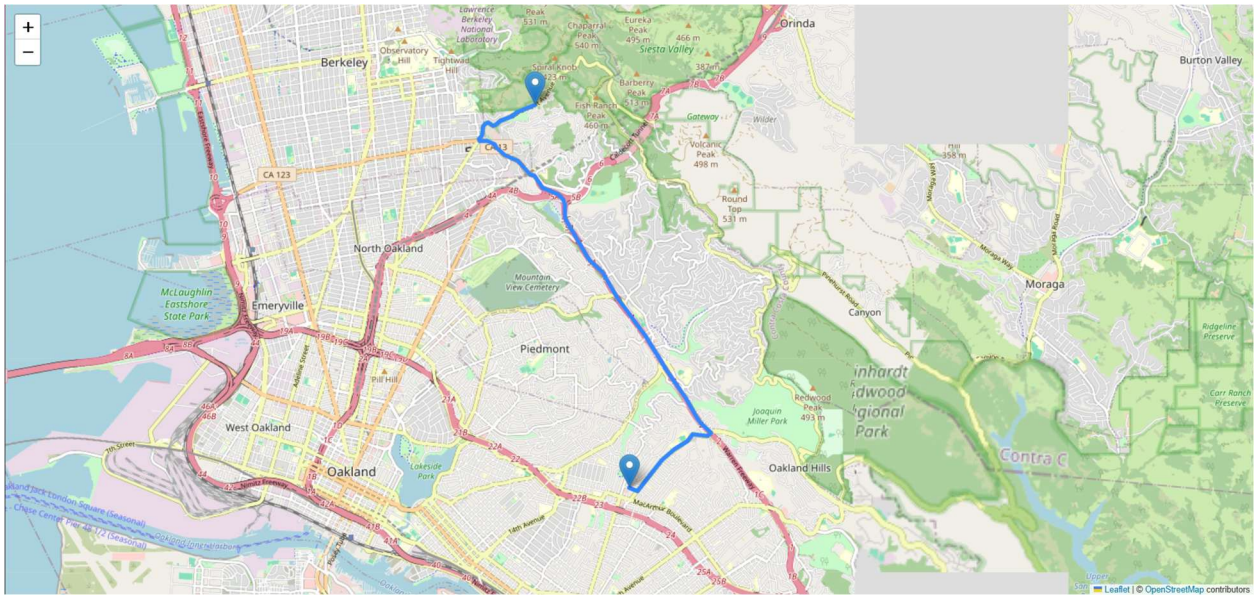


Figure 2: OD pair commute route in Open Street Maps

2.2 Rate of emissions factors

Moving on, it then became necessary to take a further look into what were the gases being analyzed by the project and what were their emissions profile for the state of California. For this, the following emissions were set as the object of this study:

- $PM2.5$ (Particulate Matter with a diameter of 2.5 micrometers or smaller)
- SO_2 (Sulfur Dioxide)
- NO_x (Nitrogen Oxides)
- VOC (Volatile Organic Compounds)
- NH_3 (Ammonia)
- CO_2 (Carbon Dioxide)

A quick note about this list is the fact that even though CO_2 is present, it was added later into the analyzed emissions and is mostly there due to being a commonly used metric for air pollution and being readily present in the datasets this study was based on. With that being said, all of the previous 5 emissions metrics have a direct impact on the health and wellbeing of the communities where they are being emitted and impose direct costs on public health (Chroma et al 2020).

To make a model that would determine the emissions (in a mass basis) being released into the atmosphere daily and in which ZIP codes of the county of Santa Clara, and later, in California, they were being focused on, as well as the ZIP codes which are mainly responsible for such emissions, the first step to be taken would be to understand the emissions per mile of the different vehicles commuting daily. For this end, data from the California Air Resources Board was

used, more specifically, the EMFAC (Emission Factors) dataset, which was developed to assess emissions from on-road vehicles including cars, trucks, and buses in California. This data is also approved by the US EPA for use in State Implementation Plan and transportation conformity analysis, which goes to show its reliability. For the purposes of this project, the "On Road Emissions and Emissions Rates" dataset was selected, outputting statewide data on road emissions for the calendar year of 2025. Vehicle category was limited to LDA and LDT1, which configure, respectively, passenger cars and light duty trucks (GVWR < 6000 lbs), fuel type comprised Gasoline, Diesel, Electric, Natural Gas, Hydrogen, Plug-in Hybrid and Fuel Cell Electric vehicles. Model years ranged from 1980 to 2026, and emissions data was given in tons per operation day. For each different vehicle model year and fuel type the dataset offered total vehicle miles traveled per day as well as the previously mentioned emissions data in tons. The appendix will contain the exact settings used as they were inputted into the EMFAC platform. With this, it was simply a matter of conducting the following operation to obtain emissions rates:

$$Eq\ 1\text{: } Emission\ rates\left(\frac{kg}{mile}\right) = \frac{Daily\ Total\ Emissions\ (Ton) * 1000}{Total\ Vehicle\ Miles\ Traveled\ (miles)}$$

Now, it was possible to obtain a new and cleaned dataset where the emission rates in kg/mile were established for each one of the different vehicle categories, model years and fuel types.

2.3 Emissions per Geographical Identification

With the emissions rates for different vehicle categories being determined, the next step in order to have a finalized model was to put together the rates with the routes for each OD pair that were previously determined. In order to do this, it would first be necessary to establish a link between the rates of emissions given by each vehicle and the vehicles that were responsible for making the daily commutes. The most straightforward way to deal with this issue was to go back to the EMFAC dataset, but this time utilizing their fleet database. The EMFAC fleet database consists of county level vehicle registrations which are organized in a similar way to what was seen in the emissions data, with vehicle weight categories being set as P (passenger vehicle) and T1 (Light-duty Trucks, GVWR < 6000), fuel type divided in Gasoline, Diesel, Electric, Natural Gas and Hydrogen, fuel technology including ICE (internal combustion engine, including conventional hybrid), BVE (battery electric vehicle), FCEV (fuel cell electric vehicle) and PHEV (plug-in hybrid electric vehicle). The data itself, as previously stated, gave vehicle registrations made to a specific Census Block Code (11 – digit GEOID), stating the previous information such as the weight class and fuel type of the registered vehicle. The appendix will contain the exact settings used as they

were inputted into the EMFAC platform. It is also valid to note that, considering that Census Block Codes are not the most granular type of GEOID, there are cases where for a same Block Code there are multiple vehicle registrations, something that will be of note for further considerations.

Considering that the data from both EMFAC datasets were very similar and used similar classifications for vehicle types, it became possible to make a cross between the two datasets in order to establish emission rates to each Census Block Code. Doing this would allow for an understanding of what vehicles are present in one area, making it so that the GEOID of an origin point from an OD pair could be matched to a vehicle registration, which would then allow for the emissions generated by a commute route to be efficiently calculated. Still, there was an issue with this method, which lay in the lack of granularity presented by Census Block Codes. Since in the case of some Codes there are multiple vehicle registrations related to the same GEOID, it was not possible to exactly pinpoint a single vehicle with a single emissions profile to a location. Also, considering how diverse emissions profiles may be, even a location with three different gasoline internal combustion engines could have different emissions profiles for each vehicle due to, for example, different model years. In order to solve this problem, it was decided that an average emissions profile would be taken for each location by using a weighted average method. This would allow for weights to be established for each vehicle based on their presence in a certain location and have these weights multiplied by the emissions profile of said vehicle, with the sum of all emissions being then divided by the vehicle population in said location. The calculation is done according to equation 2:

$$Eq2: E_{GEOID} = \frac{\sum_{i=1}^{n} E_i * N_i}{\sum_{i=1}^{n} N_i}$$

Where E_{GEOID}stands for the emissions profile for GEOID, E is the emissions profile of a vehicle type and N is the population of a determined vehicle inside of a GEOID. With this logic being established, calculations were made for the whole county of Santa Clara and a table directly relating GEOIDs to emission profiles was obtained.

2.4 ZIP code delimitation

The next step of setting up the model involves understanding which ZIP codes are being affected by each commute route. One of the crucial points of this analysis is to understand the impact of local vehicle emissions on ZIP codes of varying socioeconomical status, which makes it necessary that the routes that each commute is making can be directly related to different ZIP codes inside the area of one county. Moreover, it is also necessary to understand how many miles are being traveled inside of each ZIP code so that emissions profiles can be matched to them and

the precise emissions in kg/day can be established for all of the ZIP codes inside a county area, and, further down the development of the model, for the entirety of California.

As previously stated, each OD pair present in the LODES dataset went through Google Maps Routing API so that a route could be found between the two points. The detailed path output (polyline) contained several latitude and longitude coordinate data along the entirety of each commute route, which allowed for the precise geometry of each route to be obtained. Now, in order to relate this with the different ZIP codes crossed by each route, I utilized the TIGER/Line shapefiles. These shapefiles are published by the US Census Bureau and provide delimitations for a wide range of spatial entities, among which are ZIP Code Tabulation Areas, as well as Census Tracts and Block Groups. Incorporating the data contained in the shapefiles into the python script responsible by calling the Google Maps Directions API and calculating the routes, it was possible to overlay the geometry of each route to the delimitations of each ZIP code, which allows for a precise determination of through which ZIP codes each route is going through and how many miles are being traveled in each of them. Furthermore, this also allows for the starting point of the route to be effectively delimited as pertaining to a specific ZIP code, allowing for an analysis on which ZIP codes are being responsible for the most amount of vehicle emissions throughout the County and for further equity analysis to be made when crossing this with socioeconomical data.

2.5 Overall Model Logic

Considering all of the different steps detailed in the previous items of this section, it is also important to briefly establish a working flow of the model and how it will take the initial data in order to obtain the emissions being released into each ZIP code.

Firstly, the data present in the LODES dataset, containing the multiple OD pairs for the County of interest, will be taken and, for each one of its lines of data, the following process will be conducted:

- Google Maps Directions API will be called and will return a detailed geometrical route for each one of the OD pairs;
- The origin point of the commute, the origin location in the OD pair, will have its first 11 digits checked against the Census Block Codes of the emissions profile file, in order to establish an emissions profile to that specific commute;
- With that being done, the polyline generated previously will then be checked against the TIGER/Line shapefiles so that the commute route may be fitted inside the multiple ZIP codes it goes through, and the miles traveled in each ZIP code are obtained;

- With that knowledge, these miles can then be multiplied by the emissions profile of the origin destination of the commute, making it so that it is possible to know what are the emissions, in kg, being released in the air at each ZIP code.
- This information will be processed and recorded and will be used to output relevant data for further analysis in the form of .csv files

2.6 Routing Method Selection

Having the logic necessary to effectively run the model as well as all the python programming necessary for it finalized, the next matter of concern to finalize this phase of the project becomes getting to the point of running the model for the entirety of daily LODES commute data for the county of Santa Clara. This proved to be a challenge due to the sheer size of the dataset, which includes around 550 thousand OD pairs, which creates a constraint due to the computational power and time needed to have the model running.

Initially, there were two different possible methods that could be used to process emission data divided by ZIP code for Santa Clara, which were as follows:

- The first option was to utilize the Google Maps Routing API, which is a service that allows for an HTTPS request to be sent to the Google Maps servers and returns the ideal route between two different locations, providing directions with real time traffic for motorized vehicles and other means of transportation. This method has excellent granular precision due to the level of detail contained in the Google Maps database, allowing for precise routes to be traces between two different points while being very computationally efficient and having a low run time for each of the OD pairs, decreasing the total time to run the model. With that being said, the upsides of this method are weighted by the fact that this API is a paid service frequently used in industry, which translates to prices that may sometimes be excessive when looking at the amount of data that is to be analyzed. When looking at the chosen county, the total cost of running the model once would be $ 2200,00 (Google Maps Platform, 2025), making testing for the entirety of the model utilizing this method unfeasible.
- The second option was to utilize the data contained in OpenStreetMap (OSM) and combine it with an iteration of Dijkstra's algorithm directly coded in python to allow for routes to be found. OSM is a collaborative, open-source mapping project that provides free geographical data for the entire world. In the context of this project, the data provided by OSM is being used to build underlying road networks and nodes which can be used for route calculation using Dijkstra's algorithm. The algorithm itself consists of graph-based short-path search that identifies the minimum-cost route between two separate points. The graph is being generated by the OSM road network where road intersections are modeled and nodes, while travel time is the weight used in this case. The

algorithm then proceeds to iteratively select the nodes with the lowest cumulative cost from the origin point while updating the shortest path to its neighboring nodes and repeating this process until the destination is reached. This method is similar to the logic of the Google Maps API and has the benefit of being completely free but comes with some issues of itself. Firstly, as useful as OSM is, its data is not as detailed as the one found in Google Maps, which makes it so that there are relevant differences between the routes observed for the two methods, which were many times caused by mismatches between the actual start and end points of the route and their nearest nodes. Not only this, but this method is much more computationally intensive, especially since the routing algorithm is run locally on python, which decreases its efficiency when compared to an API hosted in Java or C++, such as the Google Maps one. This creates a scenario where running the model for Dijkstra's not only would lead to relevant inaccuracies in the established route when compared to the Google API, with mismatches in the ZIP codes traveled on each route as well as differences in the overall length of the routes being found, factors which compromise the output of the model and may lead to emissions being displayed for incorrect ZIPs, something that becomes even more relevant when considering that the outputs of the model are to be cross analyzed with socioeconomical census data, meaning that incorrect routing could lead to inconclusive conclusions later in the project. Moreover, there is also the issue of the time needed to compute the routes for each OD pair, which is significantly larger for Dijkstra's than it is for Google Maps. When considering the full scope of the data for Santa Clara, which comprises around 550 thousand OD pairs, these losses in computational time become more and more relevant, as the total processing time of the model is crucial for it to be feasible.

With these issues being put forth, comparing both methods became a matter of looking at a direct tradeoff between the main advantages and issues of each of the methods: the Google Maps API excels in the level of details of its database and the precision of its routing mechanism while also allowing for a large model using it to be executed effectively, but it does this at a steep cost that becomes specially relevant when considering the fact that the model is in its early stages of development and subject to the need of troubleshooting. Dijkstra's, on the other hand, offers a cost-free solution and allows for easier testing, but does this by sacrificing the accuracy of routes and has runtime issues that make the full run of the model something extremely difficult.

Considering the current situation and the difficulty of choosing a method that would better serve the current challenges proposed by the model, a third routing mechanism was found, which was that of Open-Source Routing Machine, or OSRM. OSRM is a high-performance routing engine that is designed to calculate optimal travel routes using road network from OpenStreetMap (OSM). It works by first processing internal raw data from the OSM utilizing an internal algorithm, which allows it to reduce the number of map nodes that will be considered

when doing the routing. Due to this preprocessing step, OSRM is able to achieve a high level of efficiency, outputting routes in times of milliseconds, something that even the Google Maps API could not do. It also supports multiple means of transportation in its set-up, but for the purpose of this project only the module pertaining to cars was used. Moreover, this engine is not run locally in python, it is set up as a local server that can receive requests through python code, similarly to what is observed in the case of Google Maps, being implemented in C++ and is able to handle large-scale repeated routing requests with relative ease.

Now, having in mind the capabilities of OSRM, its main advantage was to bridge the gap between Google Maps and Dijkstra's, allowing for a third option that would fix all issues previously found when deciding between both methods. When looking at the issue of routing precision, it was noticed that the routes traced by OSRM tended to match those of the Google Maps API much more consistently than what was being observed for Dijkstra's. The overall length of routes tended to be very similar for both methods and there was minimal ZIP code mismatch being found between the routes determined by the two. So, even if the routes had a level of difference, they would still at least be travelling through approximately the same areas within the county and would contribute to overall similar total emissions. Now, when looking at the question of timing, as previously stated, OSRM proved to not be computationally demanding and able to very efficiently find routes and emit the desired outputs in the span of milliseconds per route. Even when expanding the operations of this algorithm a bottleneck was not reached, which went to show in practice how reliable and able to withstand large data inputs OSRM was. Now, when comparing this with the test runs that were observed for Dijkstra's and even Google Maps, this can be seen as very relevant advantage. Dijkstra's main issue, even more glaring than its accuracy, was the fact that it was extremely inefficient and computationally intensive. As an algorithm that was ran locally from python using libraries from the programming language, it severely lacked optimization and would take upwards of tens of seconds when running each OD pair. While this worked decently well for individual testing, running the whole model utilizing it was basically impossible to do. This issue somewhat shows itself when looking at running a model only using the Google API. The time between each call of the API and the actual return of a precise route between an OD pair, with traveled ZIP codes and the total length driven in each of these ZIPs, would always be greater than a second. Even if this figure showed itself to be much more efficient than that observed for Dijkstra's, quick calculation shows that even at one second processing time per OD pair, the whole model, which included around 550 thousand OD pairs for the Santa Clara County, would take upwards of 6 days to run, a time figure that is still very significant albeit feasible. Having run times going as down as milliseconds, OSRM also provided a gain of efficiency even when compared with the Google API and allowed for the entire model to be run in a bit over 9 hours, which posed as a massive time reduction when compared to the two other methods. Finally, the fact that OSRM is free of cost also allowed for quick testing and troubleshooting of the

model, without the need to worry about potential issues that could arrive from eventually reaching the testing quota for the Google API, as well as allowing for tests including even thousands of OD pairs to be quickly executed. Below is a table comparing some of the metrics discussed in this section and showing the differences between the three routing methods.

Table 1: Routing methods comparison

Category	OSRM	Google Maps Routing API	Local Dijkstra's algorithm
OD pair time (seconds)	0.06	1.22	40
Total run time (hours)	9.2	186.39	6111
Cost	No cost	$2200	No cost
Precision	Very High Precision	Extremely Precise	Medium Precision

The time values for OSRM were given based on the final optimized run of the model, which allowed its time to go to extremely low levels. Running the other methods on the cluster, which will be discussed in the next section, could also allow for greater runtime efficiency and decreased overall times, although most likely not surpassing the OSRM levels.

2.7 Model Execution

With a final routing method being decided, the finalization of the model was a matter of finding an effective way to run the established algorithm for the entirety of Santa Clara County OD pairs. Now, considering a total of around 550 thousand OD pairs that comprised the dataset that was to be analyzed, as well as processing of the determined route alongside emissions profiles to yield the necessary emissions outputs, making it so that going over all OD pairs is very computationally intensive process that takes a prolonged time to complete.

Originally, it was expected for the model to be run locally utilizing my personal computer. In order to make this viable, some changes were made to the base python code to allow for greater runtime efficiency for it, mainly altering it to accommodate multithreading, which is a process that allows for multiple calculations to be run simultaneously by having them assigned to different threads within the same python process. Using this made it so that instead of having each OD pair go sequentially through OSRM routing and processing, multiple pairs could undergo this process simultaneously. Even though this process can eventually be bottlenecked by the memory of the machine running it, as well as the total number of requests that OSRM can process at once, it still allows significant run time reductions.

With these optimizations in place, a runtime of around 50 hours was obtained, with an average processing time of about 0.33 seconds per OD pair. Considering that there was not much

optimization that could be done from this point, it was decided that the model would instead use a virtual desktop in order to allow it to stay running for a long period of time and provide much more processing power than a personal computer would.

So, to also utilize the computational resources given upon Professor Sita Syal by the University of Michigan, the chosen virtual environment chosen was that of the Great Lakes Cluster, which is a campus-wide high performance computing cluster (HPC) that serves the needs of researchers across the university. The cluster primarily works utilizing a Linux environment and uses Slurm Workload Manager, which enables both interactive work as well as submitting batch jobs. After getting access to the cluster and setting up both OSRM and the necessary data input files in the virtual environments, it became possible to run the model in its entirety. The full final run took approximately 9.2 hours, with an average run time of about 0.06 seconds per OD pair, something that was both due to the implementation of multithreading as well as the overall superior computational power offered by the cluster. With that being finalized, the model outputted its data and more detail on that will be given in the following section of this report.

3.0 Results and Discussion

This section of the report will have the aim of discussing and analyzing the data that was outputted by the model, as well as making some crosses between it and socioeconomical data for the Santa Clara County, allowing for the initial steps to be taken towards the equity analysis this project aims to conduct in its future phases. As a side note, most of the development for the project that was done during this summer was done with the main overall intent of establishing the logic and a working code for the model to calculate routes between OD pairs and establish the emissions that were caused in each of these routes. Most of the details on the steps necessary to develop the model were detailed in the previous section of this manuscript, with the discussions contained in this section serving as initial interpretations of the data outputs obtained, which will be further explored and become more robust as the project moves forward onto its next steps and focuses on the socioeconomic analysis and clearly determines trends between electrification and the socioeconomic status of different ZIP codes. Also, it is important to note that all emissions data present in this section, as already previously determined, will be stated in kg.

With this being said, it is then of interest to move into the actual data. The final output of the model consisted of three .csv data files, which will be detailed below:

- receptor_zip_emissions.csv: This file contains the most important data for the analysis and gives the total aggregated emissions that were made in each ZIP, in kg, for the pollutants being studied by the model, that is, PM2.5, SOx, NOx, VOC, NH3 and CO2. The

dataset contains 72 different ZIP codes, ranging from ZIP 94020 to 9536, and contain all regions that were travelled through during the entirety of commute routes in the Santa Clara County. This data allows for a clear understanding of which regions of the County are being most affected by increased air pollution rates and would most probably benefit more from policies of EV incentive.

- origin_zip_emissions.csv: This file contains similar data to the previous, with emissions data, in kg, for the studied pollutants, but differs on the aggregation method. This file looks at which ZIP codes were responsible for emitting the most pollutants and does this by aggregating emissions for ZIP codes when they are the starting point of a commute route. This was, it is possible to see which regions of the county are the ones most responsible for contributing with the most emissions on the studied region. The final file had 64 ZIP codes, going from ZIP 94022 to 95148.
- destination_zip_emissions.csv: This file is almost identical to orign_zip_emissions.csv, with the key difference that this dataset is not aggregating based on the origin of commute routes, but in the destination of commute routes. The main purpose of this dataset was to understand if many routes were going to a specific ZIP code, making it so that one specific employer or area where certain companies are aggregated may be responsible for a relevant percentage of the pollution in the Santa Clara County.

It should also be added that a fourth output, a matrix clearly relating the pollution caused in each ZIP by routes starting in other ZIPs, similar to origin_zip_emissions.csv but with a considerably higher level of detail, but was ultimately scrapped due to it being extremely computationally demanding and raising processing time for the model from around 0.06 seconds per OD pair to about 5 seconds per pair. The table below exemplifies the way the final output is structured in the different .csv files:

Table 2: Example model output file organization

origin zip	PM25	SOx	NOx	VOC	NH3	CO2
94022	0.307301	0.051514	7.046099	12.88637	0.984761	12010.87
94024	0.362894	0.062206	8.486237	15.42435	1.179389	14506.26
94025	0.000336	6.61E-05	0.005964	0.011939	0.0012	15.40197
94028	0.003475	0.000554	0.097403	0.173546	0.010684	129.3862

Moving on, for the simplicity of the analysis, even if emissions will differ for different profiles, there is a tendency for all emissions to be proportional and more directly related to the distance traveled in each ZIP. Having this in mind, only NOx emissions will be analyzed, and it will be assumed that other pollutants follow a similar trend to it. In future iterations of the project this might be changed in order to allow for a more robust analysis but considering that the main

goal for now is to understand the overall behavior of the model's output, this simplification is not out of place.

Table 3: Descriptive statistics on NOx emissions for model output

	origin	destination	receiving
	NOX	NOX	NOX
mean	17.73639	18.01792	15.76568
median	12.49751	13.40076	13.44656
std	17.51019	17.0794	15.66798
min	0.005964	0.013812	4.52E-05
max	99.73293	71.52199	66.76822
range	99.72696	71.50817	66.76818
CV	0.987246	0.947912	0.993803
Q1	6.101821	5.393786	1.642155
Q3	26.29172	25.18756	24.43832
IQR	20.1899	19.79378	22.79616

Statistics contained in Table 3 include mean, median, standard deviation (std), minimum value (min), maximum value (max), range, coefficient of variation (CV), 25^{th} percentile (Q1), 75^{th} percentile (Q3) and Interquartile Range (IQR).

As shown in table 3, when looking at ZIP code emissions, the mean value is around 17.74 kg, with a median value of 12.5 kg, which indicates a right-skewed distribution where some ZIP codes have much higher emissions than the typical value. Moreover, the standard deviation is approximately the same as the mean, with a CV of around 0.99, which indicates significant variations between the samples of this dataset. There is also a considerably large range between the samples, which goes to show that some ZIP codes are major emitters while others end up being negligible as emission sources. Finally, it is possible to see through the IQR that a range of 20.19 kg is noticed for the middle 50% of ZIPs, which is around one fifth of the actual range of values, while the greatest concentration of this range is on the 75^{th} percentile upwards for this distribution, which corroborates with the previous analysis that this data is right skewed and a smaller number of ZIP codes strongly contribute to emissions.

Moving on, when looking at destination ZIP emissions, we have a mean of 18.02 kg and a median of 13.40 kg, which also indicates a right skewed distribution. When looking at standard deviation and CV, we have a coefficient of variation of 0.95, which is smaller than what was noticed for origin ZIP emissions but still indicates high variation between samples. Emissions range remains large, being 71.51 kg and showing once again that some ZIP codes are negligible as destinations while others are very common workplaces and are responsible for a large number of emissions in Santa Clara. Finally, the IQR of 20.19 kg with Q3 being of 25.19 kg confirms that

the data is right skewed and there is a concentration of high emission ZIP codes towards the upper quartile of the data.

Finally, when looking at statistics regarding the aggregate emissions received by each ZIP code, we can see that mean and median are closer together than in the other datasets, with a mean of 15.77 kg and a median of 13.45 kg. That being said, the standard deviation is very close to the mean and yields a CV of 0.99, meaning that there is still high variability within the data. Also, similarly to the other two datasets, range is basically equal to the maximum value present in the dataset, with the minimum being a ZIP code that has negligible emissions going through it. The IQR for this set of data is 22.80 kg, giving it the largest IQR for the smallest range over all datasets, which goes to show that this distribution is slightly more even than in the other two cases and shows more variation for the middle 50% ZIPs. That being said, lower value of the upper quartile is still way below the maximum value, which indicates the right skewness of the data.

In conclusion, what can be noticed from this analysis is how for all three datasets there is an uneven distribution in emissions in all cases and a small number of ZIP codes are either causing or receiving a greater number of emissions. In order to understand if there were any correlation between the three datasets, a Pearson correlation test was also run between the three datasets, and yielded the outputs seen in the table below:

Table 4: Pearsons correlations for model output data

metric	NOX origin	NOX destination	NOX receiving
NOX origin	1	0.115724	0.46816
NOX destination	0.115724	1	0.622826
NOX receiving	0.46816	0.622826	1

Looking at table 4, it is possible to see that there are varying correlations between the data sets. The correlation between origin and destination ZIP emissions is fairly weak, albeit positive, which is the most logical outcome when considering that ZIPs that have high origin emissions should have low destination emissions and vice versa. There is then a moderate positive correlation, of around 0.47, between origin and receiving emissions, meaning that there is a certain tendency for ZIP codes that originate routes to have somewhat higher emissions. Finally, there is a fairly stronger positive correlation between destination and receiving emissions, with a value of around 0.62, and showing that ZIP codes that are the destination of commute routes also tend to receive higher overall emissions.

With these correlations being determined, it is also valid to do some initial steps in the socioeconomical and equity analysis that characterize this project. For this end, median income data for households in the state of Santa Clara was retrieved from the United States Census Bureau and divided into ZIP code regions, allowing for an effective comparison of this information with the output data of the model. In order to conduct this analysis, the relation between income and emissions data of various ZIP codes was depicted in a series of scatter plots.

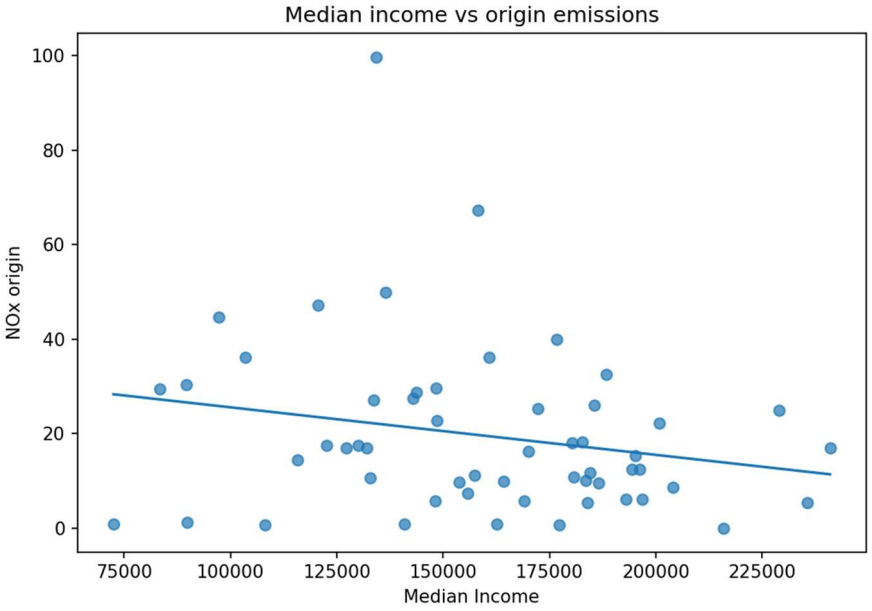


Figure 3: Median income and origin emissions scatter plot

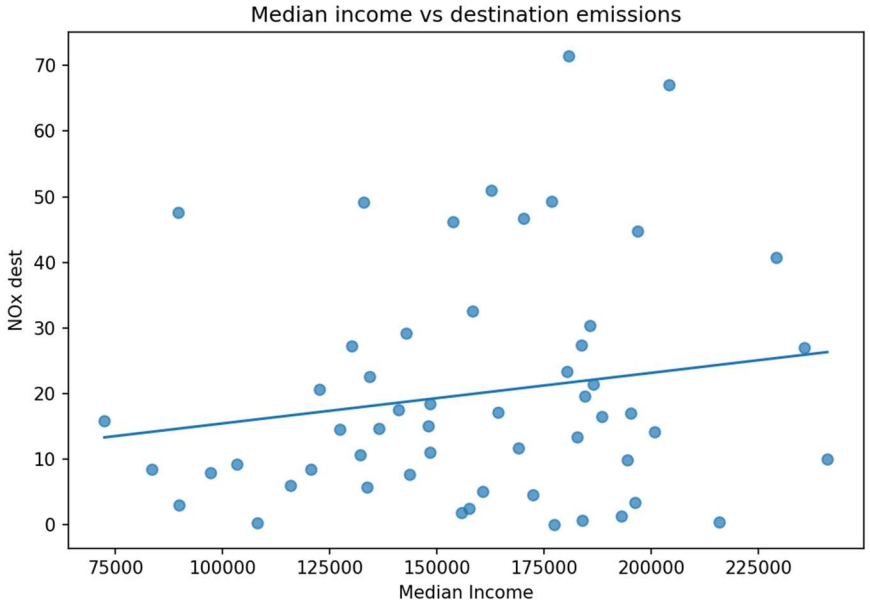


Figure 4: Median income and destination emissions scatter plot

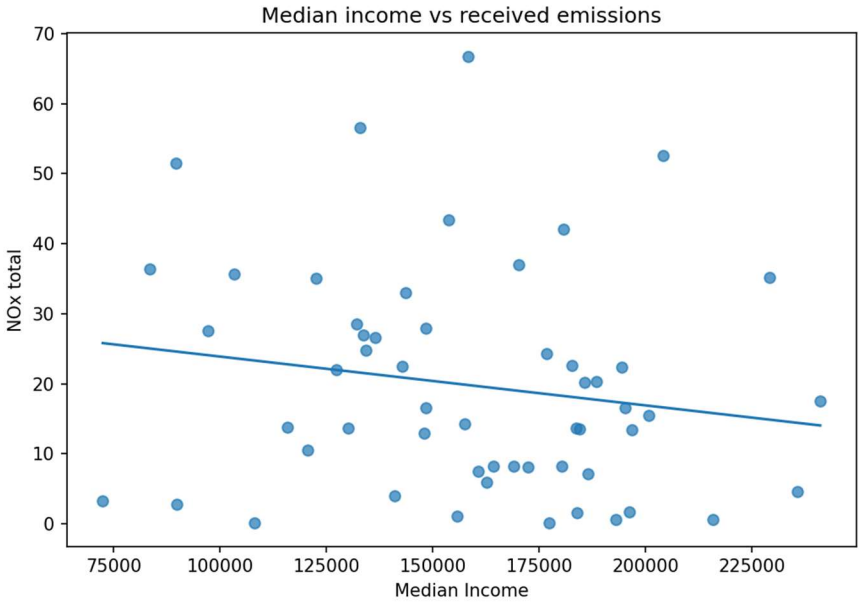


Figure 5: Median income and received emissions scatter plot

The three scatterplots also have trendlines added to them that show the relationship between the two variables in question, which will be further explored and discussed here. When looking at the relationship between median income per ZIP and the origin emissions, it is possible to see that there is a negative relationship, meaning that ZIPs with lower household median incomes tend to generate more emissions when being the origin points of commute routes. One possible explanation for this is the fact that these lower income ZIPs have a lesser amount of EV adoption, which makes it so that commute routes originating in them will most likely be transversed by IC vehicles, which have greater emissions rate per mile and might contribute to this trend of commutes originating at lower income regions generating more emissions. Moreover, the tendency of higher income ZIPs to have lower emissions when being the origin of commute routes could be explained by these regions having greater rates of EV adoption.

When looking at the relationship between median income per ZIP and destination emissions, an opposite trend is found from what was previously observed, where there is a tendency of higher income ZIPs to generate more emissions when being the destination of commute routes. This trend could be explained by the fact that many of the higher income ZIP codes present at the plot might configure regions that are not purely residential, being urban

areas with the presence of both households and work locations, with the latter most probably being the destination of many commute routes and making it so that the aggregate emissions for commute routes ending in these ZIP codes are inflated due to the fact of them having a tendency to being common destinations.

Finally, when looking at the relationship between median income per ZIP and the aggregate received emissions per ZIP, we have a negative correlation where higher emissions are observed in lower income ZIPs while higher income ZIPs have lower aggregate emissions. According to Rowangould (2013), as previously stated in the introduction, increased vehicle traffic is specially linked to decreasing income, which when added to the tendency of vehicles registered inside these ZIPs being IC vehicles (Hennessy, Syal 2023, Canepa et al 2019), would explain the higher emissions present in these regions.

4.0 Impact/Financial Benefits

Due to the nature of the project, it is difficult to currently access in an objective manner the impacts that were brought from its current stage of development. The work I have done for my Capstone 503 project has been in the context of a research project that will span a much longer development period than what was done up to this point, making it difficult to precisely judge the impacts that will be brought by the finalized analysis and model once the entirety of it is finalized. That being said, the project currently stands at a point where it has a fully functional model that can efficiently calculate commute routes between the different OD pairs of the region of a county in the state of California, as well as determine emissions profile for these routes and output aggregate emissions data for a series of different metrics that can be used for posterior analysis. Moreover, in the way that the model is currently set up, it has been fully executed for the County of Santa Clara and testing for the Santa Barbara County also showed promising results, with tests going up to 2000 OD pairs for the latter. With this, it is safe to say that the current model is fully fit for scalability and would be able to run for the entirety of the Bay Area or even the whole state of California at one point in case this is desired, albeit such a level of expansion would demand tremendous processing and computational power. Additionally, preliminary data analysis done during this phase of the project might allow for a better understanding of the data output and may serve as a basis for future iterations and phases of this development. Overall, the main impact of the development of the model as it has been done up to now is to set up a strong basis for the continuation of Professor Sita Syal's research and allow for future equity assessments on the realm of EV adoption and vehicle air pollution.

That being said, initial ideas for the model included the utilization of the Google Maps Routing API as the chosen routing mechanism for OD pair commutes. The efficient switch to an implementation of the OSRM method in this initial stage of the project allowed for savings related

to the use of the Google Maps Routing API, which would come to approximately $2200,00 for a total of 550000 requests, which would be the entirety of the OD pairs in the Santa Clara County.

5.0 Recommendations

With the current stage of the project finalized, the routing and emissions calculations model completed and the initial analysis of the data outputted by the model being done, it is then necessary to reflect upon the quality of the work that was done up to this point and suggest potential future recommendations that, if adopted, are sure to increase the overall accuracy of the model and allow for better and more precise analysis further down the development process.

The first point that must be touched here is that of the implementation of the Google Maps Routing API to the detriment of the OSRM routing method. As previously stated in this report, due to issues present in both Google Maps and Dijkstra's algorithm, it was decided that the preferred routing method chosen for the project would be OSRM, due to its high level of precision and low computing time per route. Still, even with its higher degree of precision, there are still mismatches between the routes calculated by the Google API and OSRM which, even if for now were considered to be negligible for the sake of having the model to give a final output, lead to some uncertainties that can eventually be a problem when considering that this study will need more precise outputs in order to effectively serve its stakeholders. Below are some statistics comparing the accuracy of the two routing methods.

Table 5: OSRM and Google Maps statistics comparison

	Mean	Max	Min	Standard Deviation
OSRM length (miles)	9.13	60.70	0.12	15.04
Google Maps length (miles)	9.21	60.67	0.13	15.01
ZIP mismatch	0.07	1	0	0.26
% total length difference	6.31	116.75	0.0026	13.32
Average % length difference for matching ZIPs	15.89	118.35	0.19	22.36

As seen in table 5, even though there is a tendency of the two routing methods to have very similar overall average lengths, as well as similar standard deviation for both of these metrics, as well as having a tendency to go over the same ZIP codes on every route, as seen with the ZIP mismatch figure having a maximum value of only 1, the average percent difference for matching ZIPs is where the main issue lies. This measure looks at the differences between routes inside of each ZIP code, taking the average difference between the length traveled for all ZIPs in both routes where they are matching. With this, it is possible to see that even if the total lengths of routes observed for both methods tend to be similar, there are still relevant differences in the routes traveled, as the paths taken within each ZIP code are different for OSRM and Google Maps. This proves to be an issue especially due to the fact that emissions profiles are directly related to traveled distance, meaning that these mismatches may lead to inaccuracies in the emissions received by each ZIP. Taking this into consideration, as well as the fact that one of the main stakeholders of this project would be policymakers, who would use the findings obtained through this study as a basis for equity policy on the topics of EV adoption and urban vehicle emissions, there is a need for very precise data that can precisely reflect reality. With this reasoning in mind, one of my recommendations for future phases of the project would be for the viability of obtaining funding towards the use of the Google Maps API to be studied, which would be something that would allow for more precise and accurate model outputs that would better serve stakeholders and increase the credibility of this study as a whole.

Furthermore, even if it was possible to go through the most important aspects of the output data of the model in this initial scenario, it would be of interest of those who continue the development of this project to conduct more robust and thorough data analysis on the data that was obtained through the model. Initially, a potential area of interest could be going over the statistical studies conducted in the results section of this report with an approach of normalizing emissions data by a number of factors, such as Area, Population and Job count of different ZIP codes. This would allow for a more thorough analysis of the problem and also a better understanding of the current relationships between socioeconomical factors and the emissions caused and received by different ZIP codes. It would also be interesting to look into more socioeconomic indicators for the analysis, such as studying the race of ZIP codes that are being affected by emissions and relating that to previously studied measures, such as median household income. It would also be interesting to expand the analysis for all pollutants to confirm that they all follow the same tendency as the analysis for NOx did.

Moreover, some of the analysis present in the results section was prone to the presence of noise, which could be attributed to a number of reasons, such as the presence of a small number of data points for statistical studies as well as the fact that the current analysis might be omitting other socioeconomical factors that are also relevant for the study of emissions. Due to

this, it is necessary to keep on working on the robustness of the data analysis in order to also allow for more precise and accurate predictions to be made based on the available data.

6.0 Summary/Conclusion

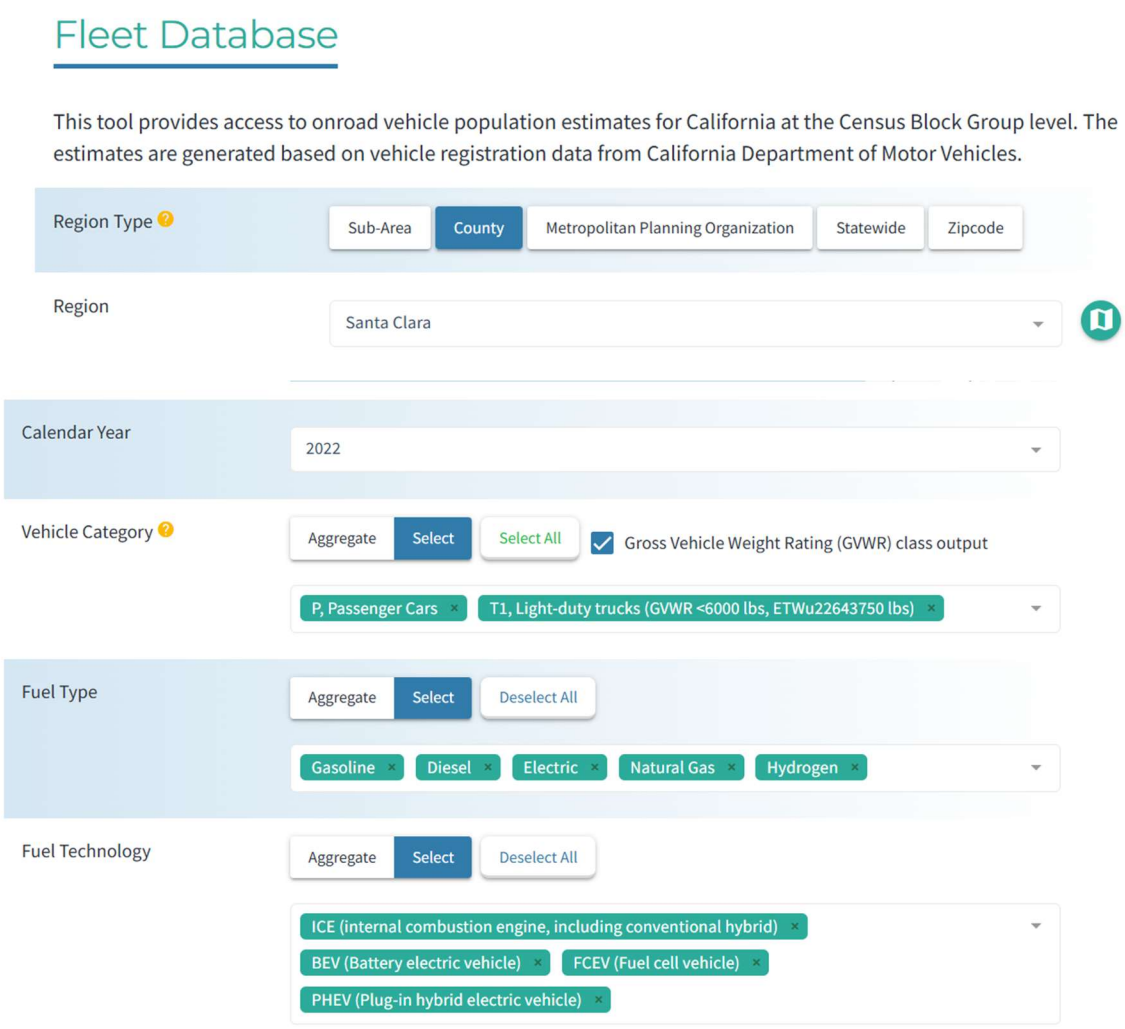
The problem to be tackled by the model at hand is extremely complex and demands time to be effectively understood and analyzed. Assessing air quality impacts and inequities resulting from fleet turnover and vehicle electrification policies requires a multifaceted approach and the lengthy implementation and troubleshooting of models that can take in the entirety of the input necessary to entirely comprise the problem and yield outputs that can represent it accurately and according to reality. Even though the development of this capstone project was mainly focused in starting and finalizing the development of a high resolution traffic and emissions turnover model that was set to run for data pertaining to the Santa Clara County and allow for some initial data analysis on the outputs of this model, it is of my understanding that the project will continue forth and require much more work in order to allow for it to fully capture the magnitude of the issue and present the desired outputs.

That being said, the overall development of the model proved to be successful, with a completely free OSRM server paired with python code being able to handle data for the entirety of Santa Clara and calculate ideal routes and emissions generated for all OD pairs contained in the LODES commute dataset for the selected region. Not only this, but at its current state the model and its code would be ready for expansion following minimal changes and adaptations to allow for the larger number of inputs and the processing power needed for said expansion.

Initial analysis also corroborated with many of the conclusions initially put forth in the introduction of this manuscript, agreeing with the authors that served as a basis for this study. This fact goes to validate the overall output of the model and attest for the analysis that was done up to this point, which sets the project to be built upon the work that has already been done. Not only this, but the conclusions derived from the data of the model and its cross analysis with socioeconomical data show clearly that there is a tendency for lower income ZIPs to be more affected by pollutant emissions when compared to higher income ZIPs, which damages the environment and the health of people that inhabit these regions. This only goes to show the need for equity and environmental analysis to be conducted in these spheres, which then in turn allow for incentives and policies to bring a change to these situations, being that by providing incentives for the purchase of EVs, the increase of infrastructure for adoption of these vehicles in lower income neighborhoods, or any measures focused in mitigating the unequal distribution of environmental burden.

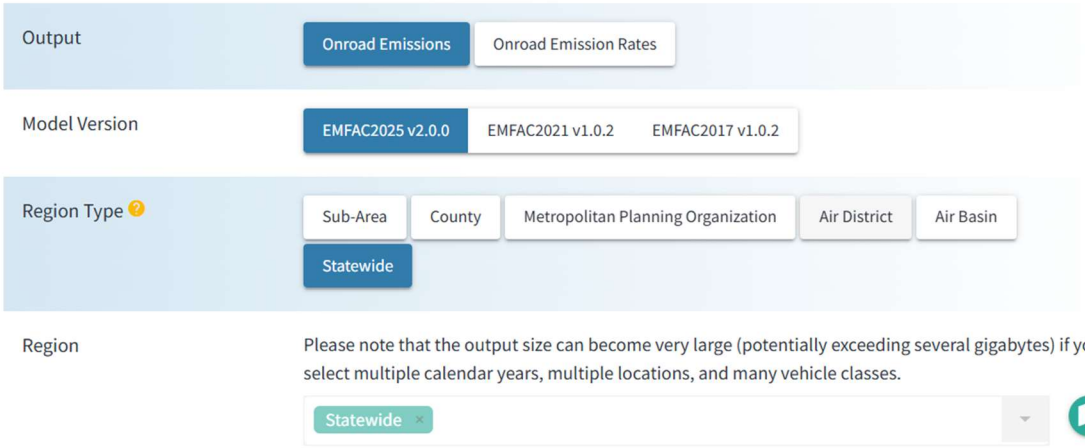
7.0 Appendices

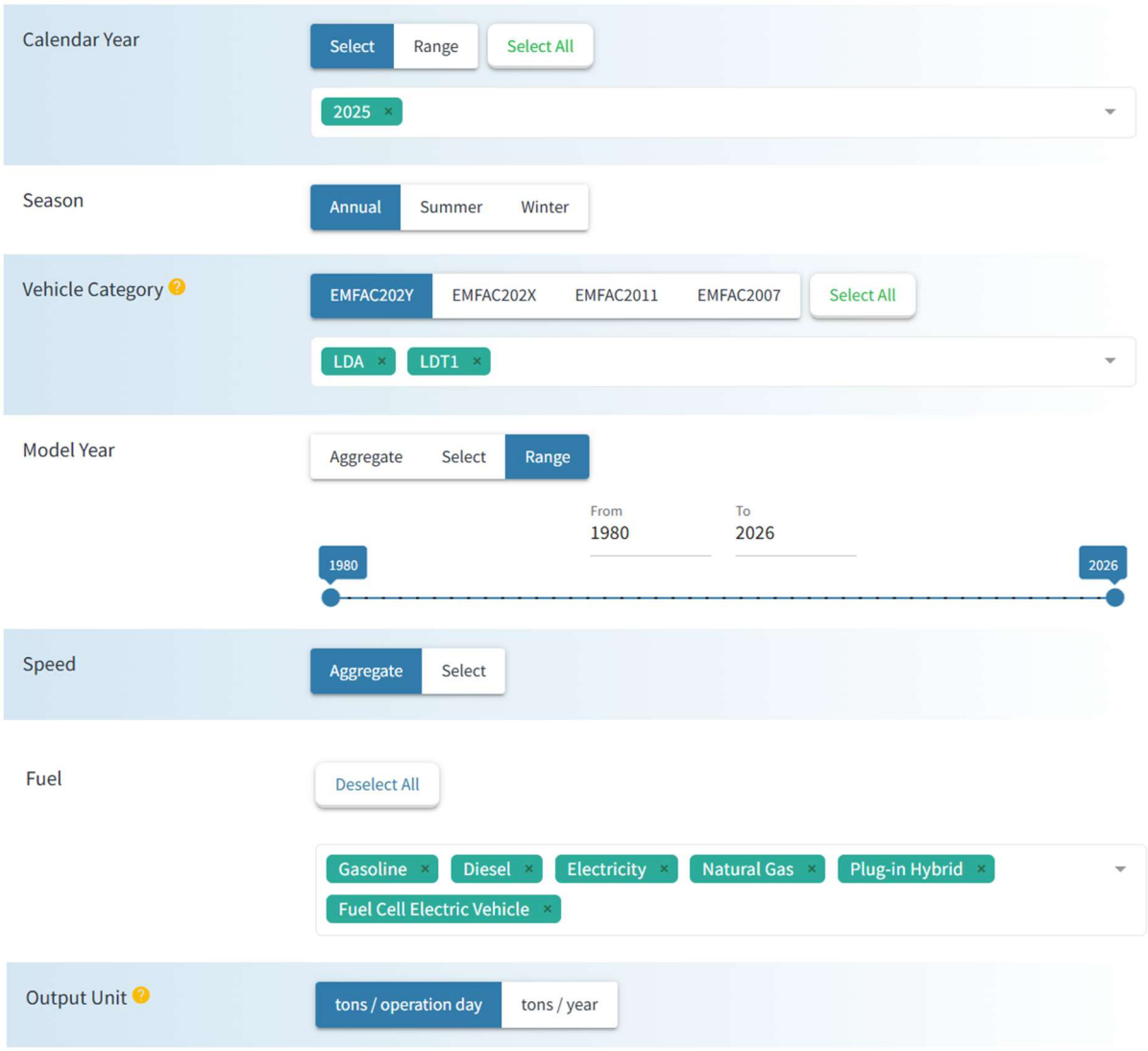
Appendix A – EMFAC platform data settings



Emissions Inventory

This tool provides emissions from onroad mobile sources in California. Please note that emissions extracted from this web tool are the same as those generated using EMFAC PC Application.




Calendar Year
Select
Range
Select All
2025
Season
Annual
Summer
Winter
Vehicle Category
EMFAC202Y
EMFAC202X
EMFAC2011
EMFAC2007
Select All
LDA
LDT1
Model Year
Aggregate
Select
Range
From
1980
To
2026
1980
2026
Speed
Aggregate
Select
Fuel
Deselect All
Gasoline
Diesel
Electricity
Natural Gas
Plug-in Hybrid
Fuel Cell Electric Vehicle
Output Unit
tons / operation day
tons / year

8.0 References

Hennessy, E.M., Syal, S.M. (2023). Assessing justice in California's transition to electric vehicles. *iScience*.

Canepa, K., Hardman, S., Tal, G. (2019). An early look at plug-in electric vehicle adoption in disadvantaged communities in California. *Transport Pol*.

U.S. EPA (2023). *Inventory of U.S. Greenhouse Gas Emissions and Sinks 19920-2021* (Available at: www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2021)

Tessum, C.W., Paloelella, D.A., Chambliss, S.E., Apte, J.S., Hill, J.D., Marshall, J.D. (2021). PM2.5 polluters disproportionately and systemically affect people of color in the United States. *Sci. Adv.*

Rowangould, G.M. (2013). A census of the US near-roadway population: public health and environmental justice considerations. *Transp. Res.*

Choma, E.F., Evans, J.S., Hammitt, J.K., Gómez-Ibáñez, J.A., Spengler, J.D. (2020). Assessing the health impacts of electric vehicles through air pollution in the United States. *Environ Int.*

US. "Data - Longitudinal Employer-Household Dynamics." *Census.gov*, 2020, lehd.ces.census.gov/data.

"On-Road (EMFAC) | California Air Resources Board." *Ca.gov*, 2025, ww2.arb.ca.gov/our-work/programs/msei/on-road-emfac.

"Platform Pricing & API Costs." *Google Maps Platform*, mapsplatform.google.com/pricing.