**Project Summary**

*This report is the definitive source of information for the project. It must contain all necessary information to recreate the project deliverables, including links to all source code, tools, libraries, together with access information (user names / passwords).*

| **Project Name** | Simulation of Electric Transit System |
| --- | --- |
| **Sponsor** | Professors Alberto Lamadrid, **Shalinee Kishore**, Shamim Pakzad, Larry Snyder |
| **Date** | 5/5/2023 |
| **Team Members** | Aarushi Singh, Matt Culbertson, Ryanne Cox |
| **Faculty Advisor** | Professor Lee-Urban |

**Project Description**

| *Provide a high-level description of what the completed project does, what it consists of, and any other information needed to understand it.*  Our project is split into two different parts – data analysis and simulation. In the completed project, the data analysis predicts the amount of energy the buses will produce. The data analysis will help us find the best route, conditions of the bus, charging times, and more to allow the buses to use the least amount of energy. In the completed project, the simulation models the amount of charge VTA buses use in their routes. The simulation also models how much charge can be acquired in a given period of time for these same VTA buses. The simulation produces terminal output and graphs that display the results of an individual run. |
| --- |

**Key Processes or Procedures**

| *List all necessary instructions for carrying out key processes, such as updating a web site, backing up a database, getting approval from the app store for a mobile application, building a compiled application, etc.*  Data analysis   * Understanding the Viriciti Data - compared various different variables to see the correlation. Examples include time-driving and energy-driven, average speed and energy driven. We created scatter plots to help us determine what variables were the most correlated. * Getting Rid of Outliers - to create more accurate plots, we looked at different ways to get rid of the outliers and created a gif to show our results * Comparing VTA and Viriciti data - making sure that the Viriciti and VTA provide similar results by comparing distance driven and energy used for both * Combine VTA and Viriciti data with weather - aggregated the Viriciti and VTA Data, with weather data we got from a weather API of locations close to the bus routes to get more variables that could affect the bus energy usage * Regression Models RMSE with KFold - created various different models to predict the percent of energy each bus would use. PGBM ended being one of the best * PGBM Model: Differentiating Weather Data - PGBM\_model - used the PGBM Model after being suggested by our sponsor and compared the various different types of data like only training the data on VTA data and testing it on only VTA with our model. Training and testing with VTA ended being the best model   Simulation   * Built a python file – driver\_analysis.py – that read all the entries in the clean\_drive\_data.xlsx file where the op\_id was neither 0 nor null. The file looks at the average kwh for each driver and creates a plot. This plot and file can be used to analyze driver efficiency. * Built a python file – finding\_dist.py – that looks at specified columns of the clean\_drive\_data.xlsx file (miles\_travelled and kwh) and find the best distribution to fit the data. The results of this file are then used to inform the rest of the simulation. * Built a python file – bat\_drain.py – that uses raw data and calculated distributions to create a simulation of the battery consumption for 10 independent buses. The 10 buses are created, assigned a route randomly, assigned a battery consumption (kwh) from a distribution, and calculates the amount of battery consumed based on these values. The file goes on to use the calculated battery remaining, the data from the charging station, and a time period to determine how much charge each bus has after charging for a given duration. This file creates figures to show calculated distributions and prints terminal code that details the steps and results of the code. |
| --- |

**Technologies/Tools**

| *Summarize the technical components of your project and what roles they play.  (Programming languages, IDEs, databases, data-access components, etc.)*  Data Analysis – The data analysis uses Python programming language. Within Python, we used packages like scilearn to help with the Regression Models. We used Juypter Notebook as it made working with graphs much easier and more visually appealing. In addition, it allows us to organize our code. We have uploaded all our Jupyter Notebooks to the Google Drive, which is also where the simulation lies. This overall makes us as a team more organized.  Simulation – The simulation uses the python programming language. Python comes with a lot of data analysis packages, so determining distribution fits and generating figures was relatively easy. Python is also really good for creating an environment, which can be used to run the simulation. Currently, and environment is not implemented, but adding one could be a useful extension in the future. The simulation was created using visual studio code. Visual studio code made importing and moving files around very easy. The simulation was uploaded to GitHub. The GitHub repository is home to both the simulation files and the data analysis work. The team used GitHub to have a common repository where we could share files and have version control. |
| --- |

**Skills**

| *Indicate CSB courses used on this project, and other required techniques/technologies and how you learned them*  Courses used for this project:  CSE262: Programming Languages. In this course, we learned how to use python as well as how to read excel files in a python script. These were critical skills for developing the simulation documents, as thy are all written in python and the data comes from excel sheets.  Github: We used a github repository for source control and consistency across team members. We learned this technology in several CSE courses as well as from programming experience.  CSE216: In CSE216, we learned about sprints. Throughout the semester, we used sprints to plan and track our progress. We also learned about teamwork in CSE216. We used the teamwork skills we gained in CSE216 to work together throughout the semester.  Youtube – used to learn discrete even simulation with python. I watched this video: <https://www.youtube.com/watch?v=jXDjrWKcu6w> for my seminar and learned a lot about modeling using python.  We also used Data Camp to get the basics of various python topics down like sklearn: <https://app.datacamp.com/learn/career-tracks/data-scientist-with-python> |
| --- |

**Challenges/Learnings**

| *List significant challenges encountered, and how you addressed them.. List other key project learnings, including a description of alternative technologies considered and a discussion on why you chose the tools you selected..*   * Lack of communication from sponsor: There were a few instances where we needed specific information from our sponsor’s graduate student and he either responded late of did not respond at all. Typically this would happen when we slacked him, so we followed up with email when this issue occurred. We chose to use slack because the simulation of electric transit project already had a slack channel, so we thought it would be the easiest way to communicate. * Weak python programming skills: At the beginning on the semester, the team did not have significant experience in python, which is the language we use the most. We addressed this proble by googling, watching youtube videos, and reading python package descriptions. We chose to use python because it was requested by the sponsor and we were confident we could learn python quickly. * Lots of paperwork: We found that we were spending a lot of our time tracking our progress rather than making progress. We felt that we had redundant strategies for keeping track of our backlog, so * Advanced plots using python matplotlib * Learned how to fit data with probability distributions |
| --- |

**Project Deliverables** *(include everything – source code, databases, documentation, physical hardware, etc.; add rows as needed. Note: everyone must include their project charter, so it is listed on the first line. Add other items as appropriate to your project)*

| **Deliverable** | **Location** | **Username/Password** |
| --- | --- | --- |
| **Project Charter** | **Capstone ‘23 Shared Drive :** [**drive**](https://docs.google.com/document/d/1IciMJR50TvjjxO9l79wT9lyKdbvzR6BJs9LwdxVSEBI/edit) | [**ebcapstone23@gmail.com**](mailto:ebcapstone23@gmail.com)  **Packard466** |
| **Project Files (source code)** | [**https://github.com/rec224/ElectricTransitSim**](https://github.com/rec224/ElectricTransitSim) | **N/A, the repository has been made public** |
| **Resoures** | [**https://drive.google.com/drive/u/0/folders/0AIjMo3TZvZ5rUk9PVA**](https://drive.google.com/drive/u/0/folders/0AIjMo3TZvZ5rUk9PVA) | [**ebcapstone23@gmail.com**](mailto:ebcapstone23@gmail.com)  **Packard466** |

**Future Recommendations**

| *What next steps should be taken for this project? What should be avoided? Indicate any recommendations you have for the sponsor and/or future project teams.*  Simulation: For the future, the next step that should be taken is creating a web page/dashboard where Californians can read and understand the data we have simulated. A visual deliverable would also help show the VTA what exactly the simulation is doing. I think returning to SUMO should be avoided. The terminal output/figures produced by the simulation files are easier to understand and manipulate than SUMO was. For future teams, I think there are modifications that could be made to make the simulation more accurate.  Data Analysis: For the future, the next steps that should be taken are to analyze our models and analyze them to find ways we can reduce the amount of energy buses can use. In the future, we will not have Nick, who has helped us with the majority of this project by giving us guidance. Thus, we will need to figure out long-term goals with Nick when we get a chance and set a schedule with steps on how we will fulfill them. |
| --- |