**Project Objectives/Outcomes:**

* At the end of the project, we want to get a better sense of how the air quality in the City of Berkeley is affected by the traffic on the highway.
* The goal is to create a model that predicts the air quality given a number of factors.
* …

**Project Overview**

* <Mini Summary of the Project>

**Planning**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week/Date** | **Description** | **Phase** | **Notes** |
| 1 - 9/13 | Getting to know the team | Orientation |  |
| 2 - 9/20 | Onboarding/Getting access to data | Orientation |  |
| 3 - 9/27 | Data Cleaning & Exploratory Analysis: Looking for patterns | Cleaning/EDA | - Might take longer because data are difficult to access |
| 4 – 10/4 | Exploratory Analysis: Performing some statistical analysis | EDA |  |
| 5 – 10/11 | Exploratory Analysis: Creating useful graphs and giving data-driven recommendations | EDA/BI | - Midterm Season |
| 6 – 10/18 | Midsemester Presentation | Presentation/Mid-Point Check-In |  |
| 7 – 10/25 | Feature Engineering: Find useful features for model\* | Modeling |  |
| 8 – 11/1 | Train Model | Modeling |  |
| 9 – 11/8 | Evaluate Model | Modeling |  |
| 10 – 11/15 | Finetune Model | Modeling |  |
| 11 – 11/22 | Create Dashboard | DevOps |  |
| 12 – 11/29 | Final Presentation | Presentation |  |
| 13 – 12/6 | Evaluate Project | Reflection |  |

***Requirements/Tools***

Since we will be performing a lot of EDA and data analysis, we expect to be using the following libraries:

- Pandas

- Matplotlib

- Numpy

For the modeling, we aren’t entirely sure what to use, but we could potentially use a Linear Regression model and if that doesn’t seem to be powerful enough, we can scale it up to other models such as Tree Regressors.

***Time Commitment***

The project is a little bit on the heavier-side, so we need to be a little bit more critical of the things we can get done. Allocating 3 weeks for modeling might be tight, but creating the visualizations and EDA could be potentially accomplished in a shorter time window. It will also be interesting to see what happens during the midterms season where we will have less time at hand.

***Stakeholders***

We are working with the company, our teammates. However, our data comes from the City of Berkeley which might be a bottleneck. Additionally, our data will be stored on a Cloud instance, so in a sense, we are also collaborating with Google Cloud Platforms. Mainly, our workflow will be to report to the company we are working with in a weekly meeting. Then we (our team) work on the project. In turn, our mentor will be assisting us and do the communication with the company and City of Berkeley.

***Challenges***

- It’s uncertain whether we will be able to finish everything at the end of the semester

- Many people on our team haven’t worked with more advanced models yet

- The partner doesn’t seem to be responsive

- The problem is to broadly scoped, so we might need to do some narrowing down

**Reflection: What were your motivations behind this planning?**

<In a few sentences, reflect on why you created the planning as is>