**Project 3 – Team 2 Proposal: Analyzing the Impact of Climate and Wine Quality**

**Problem Statement**

For our data engineering project, we chose to examine a large database that focuses on the quality of thousands of different wines from across the world. There are a multitude of factors that effect the quality of wine such a soil composition, technique, aging, topography and many more. We decided however that we wanted to use climate as the focal point of our study. Wine makers have often noted that temperature, precipitation and other climate variables can significantly affect grape cultivation and subsequently the quality of wine produced. This project aims to add value to that conversation by integrating wine quality data with detailed climate date to understand how climate conditions influence wine quality.

**Objectives:**

1. Data Extraction: Collecting wine data (region, quality, price, variety) and weather data (temperature, precipitation, other climate factors) in the forms of databases and API’s
2. Data Cleaning: Ensure that the datasets are free of null values, consistent across rows and free of errors.
3. Data Transformation: Structuring and converting the data so that it’s consistent and suitable for analysis later on.
4. Merging Data sets: Combining our cleaned dataset with regional climate information pulled from our API.
5. Data Analysis: Perform statistical analysis to identify correlations and trends between climate variables and wine quality
6. Data Loading: Load the cleaned data sets onto a database for storage
7. Present our findings: Using data visualizations, present our findings and explain our workflow process and analysis.

**Data Sources:**

* Wine Review dataset from Kaggle: <https://www.kaggle.com/datasets/zynicide/wine-reviews>
  + This data was scraped from reviews from Wine Enthusiast, a leading website and magazine on wine lifestyle products and content.
* Open WeatherMap API: <https://openweathermap.org/api>
  + An online service that gathers global weather data

**Sub Questions:**

1. Correlation Analysis:

* What are the correlations between average temperature, precipitation and wine quality scores?
* Are there any specific weather conditions which impact wine quality negatively?

1. Economic impacts:

* How do climate factors effect the price of wine?

1. Wine variety analysis:
   * Do different varieties tend to be produced more often under specific climate conditions?
   * Are certain wine varieties more frequently associated with higher scores?