Project Proposal: Coffee Review Analysis & Rating Prediction

Introduction

The coffee industry is vast, and consumer preferences can vary significantly based on sensory characteristics such as aroma, flavor, and acidity. However, many coffee producers lack a systematic way to understand how these factors influence consumer ratings. This project aims to bridge that gap by developing a predictive model that analyzes various sensory and categorical features to forecast overall coffee ratings.

Problem Statement

The primary problem we aim to solve is the lack of reliable methods for coffee producers and consumers to assess and predict coffee quality based on sensory characteristics and origin. By predicting coffee ratings, we can help stakeholders make informed decisions about sourcing, roasting, and selecting coffee products that align with consumer preferences.

Target Clients

The target clients for this project are coffee producers, distributors, and enthusiasts. Producers can leverage the insights gained from our analysis to optimize their sourcing and roasting processes, ultimately improving product quality. Consumers, on the other hand, will benefit from a predictive model that can recommend coffee products tailored to their taste preferences.

Data Sources

1. **Coffee Review Dataset**
    - Source: Kaggle Coffee Review Dataset
    - Description: This dataset contains detailed reviews of various coffees, including sensory attributes like aroma, flavor, acidity, and overall rating.

2. **Geocoding API**
    - Source: Geopy Documentation
    - Description: This API will be used to convert coffee origin and roasting origin addresses into geographic coordinates, enabling a deeper analysis based on location.
3. **Weather API**
    - Source: Open Weather API
    - Description: Weather data will be collected for coffee-growing regions to analyze how environmental factors influence coffee quality and ratings.

Potential Outcomes

1. **Feature Insights**: Identify which sensory features (e.g., aroma, aftertaste) have the most significant impact on overall coffee ratings.
2. **Predictive Model**: Develop a robust model that can be utilized by coffee producers and consumers to estimate product ratings based on sensory characteristics.
3. **Geographic Analysis**: Quantify how the coffee's geographic and roasting origin influences quality and ratings, providing actionable insights for producers to enhance their sourcing and roasting strategies.