Coffee Quality Database

The Coffee Quality Database presents a rich repository of information encompassing various aspects of coffee production and quality evaluation. The objective of this project is to conduct a comprehensive data exploration, cleaning, and analysis to derive meaningful insights from the dataset.

This database contains the following information:

Quality Measures

- Aroma
- Flavour
- Aftertaste
- Acidity
- Body
- Balance
- Uniformity
- Cup Cleanliness
- Sweetness
- Moisture
- Defects

Bean Metadata

- Processing Method
- Colour
- Species (arabica / robusta)

Farm Metadata

- Owner
- Country of Origin
- Farm Name
- Lot Number
- Mill
- Company
- Altitude
- Region

Tasks to be performed include:

1. Data Cleaning:
    - Handle missing values: Identify and handle missing values in the dataset, employing techniques such as imputation or removal based on the nature and extent of missingness.
    - Address inconsistencies: Check for inconsistencies or anomalies in the data, such as typos or irregular formatting, and rectify them to ensure data integrity.
2. Exploratory Data Analysis (EDA):
    - Univariate Analysis: Explore distributions and summary statistics of individual variables such as quality measures (aroma, flavor, etc.), bean metadata, and farm metadata.
    - Bivariate Analysis: Investigate relationships between pairs of variables, examining correlations or associations between quality measures, bean characteristics, and farm attributes.
    - Multivariate Analysis: Explore interactions and dependencies among multiple variables using techniques like clustering or dimensionality reduction.
3. Visualization:
    - Utilize appropriate visualization techniques (histograms, box plots, scatter plots, etc.) to visually represent the distribution and relationships within the dataset.
    - Create insightful visualizations to highlight patterns, trends, and anomalies in the data, facilitating better understanding and interpretation.
4. Analysis and Interpretation:
    - Interpret findings from the data exploration and visualization, providing insights into factors influencing coffee quality.
    - Identify key trends, patterns, or outliers that may impact quality assessment, production processes, or sourcing decisions.
    - Offer recommendations or suggestions based on the analysis, such as potential areas for improvement in production methods, selection of bean varieties, or optimization of farming practices.
5. Documentation:
    - Document the entire data exploration and analysis process, including methodologies, findings, and interpretations.
    - Present the analysis report in a clear and concise manner, catering to both technical and non-technical audiences.