COFFEE QUALITY REPORT

About the Dataset:

The Coffee Quality Institute (CQI) is a non-profit organization focused on enhancing the global coffee industry's quality and value. Established in 1996, CQI conducts various activities, including research, training, and certification programs, to improve coffee quality standards and support sustainability initiatives.

The dataset provided by CQI contains comprehensive information on coffee production, processing, and sensory evaluation. It includes sensory attributes such as aroma, flavor, acidity, and body, along with data on defects, processing methods, origin regions, and more.

Objective:

The primary objective of this project is to utilize the CQI dataset to gain insights into the factors influencing coffee quality. Specifically, the project aims to address the following research questions:

1. Identify the key determinants of coffee quality, focusing on sensory attributes such as aroma, flavor, acidity, etc.

2. Investigate correlations between processing methods, origin regions, and coffee quality scores.

3. Analyze trends and patterns in defect occurrences and their impact on overall coffee quality.

4. Explore how different variables interact to influence Total Cup Points, representing an overall measure of coffee quality.

Tools Used:

Power BI was employed to analyze and visualize the dataset due to its robust data processing capabilities and interactive visualization features.

Shape of Data:

Columns: The dataset consists of 31 columns, including sensory attributes, defects, processing methods, and geographical information.

Numerical Columns: These include ratings for sensory attributes, defect counts, moisture percentage, altitude, and grading dates.

Categorical Columns: These encompass coffee variety, region, country of origin, processing methods, and color.

Features to be Considered:

Key features considered in the analysis include sensory attributes (aroma, flavor, acidity, etc.), defects (Category One and Category Two), coffee variety, region, country of origin, processing methods, altitude, grading date, and color.

Data Preprocessing:

1. Removed unnecessary columns and handled missing values.
2. Eliminated duplicates and standardized the color column values.
3. Translated non-English words in the region column to English.
4. Processed the altitude column by splitting ranges and calculating averages.
5. Conducted trimming and formatting operations for consistency.

Calculated Columns and Measures in Power BI Desktop:

Max Overall: Maximum overall coffee quality rating observed.

Sum Cup Points: Total cup points calculated from sensory attributes.

Total Defects: Combined count of Category One and Category Two defects.

Visualization:

Cards: Displayed key measures such as Max Overall, Sum Cup Points, and Total Defects alongside relevant coffee types and altitude.

Scatter Plots: Utilized scatter plots to visualize the relationship between sensory attributes and overall coffee quality. One scatter plot focused on attribute-wise overall scores, while another examined the correlation between attributes.

Line Chart Employed a line chart to depict variety-wise average overall ratings and total defects, providing insights into how different coffee varieties perform in terms of quality and defect occurrences.

Treemap: Utilized a treemap to present region-wise average overall ratings, allowing for easy comparison of coffee quality across different regions.

Donut Chart: Incorporated a donut chart to illustrate country-wise average overall ratings, providing a concise overview of coffee quality in various countries.

Column Chart: Included a column chart to showcase total cup points by processing method, highlighting the impact of different processing methods on coffee quality.

These visualizations offer a comprehensive understanding of the dataset, enabling stakeholders to identify patterns, trends, and correlations related to coffee quality and defects.

Formatting:

Formatted the report for clarity and aesthetics, including interactive slicers for parameter selection and dynamic updates in scatter plots.

Insights:

Overall Coffee Quality:

1. The maximum overall coffee quality rating observed in the dataset is 8.58.
2. All sensory attributes contribute positively to the overall coffee quality rating.
3. The sensory attribute ratings are consistently above 8, indicating high quality.
4. Higher overall quality ratings are associated with higher total cup points, with a maximum of 10.13k.

Defects and Coffee Quality:

1. Defects negatively impact coffee quality, with higher defect counts leading to lower quality ratings.
2. Category One defects, such as black beans and insect damage, have a visible impact on coffee quality.
3. Category Two defects, such as over-fermentation and staleness, are subtler but still affect overall quality.

Regional Quality Variation:

1. Colombia, specifically the Piendamo region, produces coffee with the highest quality rating of 8.58.

2. Regional variations exist, with certain regions consistently producing higher quality coffee.

Processing Methods and Quality:

1. Double Anaerobic processing method yields higher cup points, indicating better quality coffee.

2. Notable examples include the Pieandemo region in Colombia, which uses the double anaerobic method, producing coffee with a quality rating of 8.58

Analysis :

1. **What are the key determinants of coffee quality as evaluated through sensory attributes such as aroma, flavor, acidity, etc.?**

   Examining the correlation between key sensory attributes, including aroma, flavor, acidity, body, balance, and aftertaste, and the overall coffee quality score demonstrates that elevated ratings for these attributes align with higher overall quality scores for the coffee**.**

2. **Is there a correlation between processing methods, origin regions, and coffee quality scores?**

   When examining the correlation between processing methods and coffee quality scores, it becomes evident that certain processing methods have a notable impact on coffee quality scores, as some methods result in more defects, thereby lowering the overall quality of the coffee. Similarly, when considering coffee-growing regions, the highest coffee quality scores are observed in Piendamó, Cauca region, Colombia, followed by Laos and Papayan in Colombia. These findings suggest that coffee cultivation in these regions benefits from favorable weather conditions conducive to high-quality coffee production. Conversely, in Grezia, Brazil, the weather conditions are marginally conducive to coffee cultivation, which may contribute to lower coffee quality scores in this region.

3. **Can we identify any trends or patterns in defect occurrences and their impact on overall coffee quality?**

   Defects in coffee significantly impact its quality score. There are two types: visual (category 1) and taste (category 2) defects. Visual defects, like damaged beans, are seen, while taste defects, like off-flavors, are detected when tasting. Both types reduce the coffee's quality and satisfaction. It's crucial to identify and minimize defects for consistent, high-quality coffee production.

4. **How do different variables interact to influence the Total Cup Points, which represent an overall measure of coffee quality?**
   Various factors, including processing methods, regions, and sensory attributes, play pivotal roles in determining the overall quality assessment and cup points assigned to coffee. Processing methods dictate how beans are treated, directly impacting quality, while sensory attributes are fundamental to flavor and profile, directly influencing coffee quality. Additionally, regional weather conditions exert influence on both cup points and coffee quality. These interwoven elements collectively shape the evaluation of coffee quality and the accurate allocation of cup points.

Conclusion:

Through comprehensive analysis and visualization, the project provides valuable insights into the factors driving coffee quality, empowering stakeholders to make informed decisions and drive improvements in the coffee industry.