***Power BI Capstone project Documentation on Coffee Quality Data***

#### Business Problem: The Coffee Quality Institute (CQI) is a non-profit organization that works to improve the quality and value of coffee worldwide. It was founded in 1996 and has its headquarters in California, USA. CQI's mission is to promote coffee quality through a range of activities that include research, training, and certification programs.

#### The organization works with coffee growers, processors, roasters, and other stakeholders to improve coffee quality standards, promote sustainability, and support the development of the specialty coffee industry. By leveraging this dataset, coffee producers, suppliers, and retailers can gain valuable insights into the factors that determine coffee quality and make informed decisions to improve their products.

#### Objective: The primary goal of this project is to leverage the rich dataset provided by CQI to understand the factors that contribute to coffee quality.

**Tools Used: Power BI**

**About Data**: The data includes a range of information on coffee production, processing, and sensory evaluation. It also contains data on coffee genetics, soil types, and other factors that can affect coffee quality.

**Sensory evaluations (coffee quality scores)**

* **Aroma**: Refers to the scent or fragrance of the coffee.
* **Flavor:** The flavor of coffee is evaluated based on the taste, including any sweetness, bitterness, acidity, and other flavor notes.
* **Aftertaste:** Refers to the lingering taste that remains in the mouth after swallowing the coffee.
* **Acidity:** Acidity in coffee refers to the brightness or liveliness of the taste.
* **Body:** The body of coffee refers to the thickness or viscosity of the coffee in the mouth**.**
* **Balance:** Balance refers to how well the different flavor components of the coffee work together**.**
* **Uniformity:** Uniformity refers to the consistency of the coffee from cup to cup.
* **Clean Cup:** A clean cup refers to a coffee that is free of any off-flavors or defects, such as sourness, mustiness, or staleness.
* **Sweetness:** It can be described as caramel-like, fruity, or floral, and is a desirable quality in coffee.

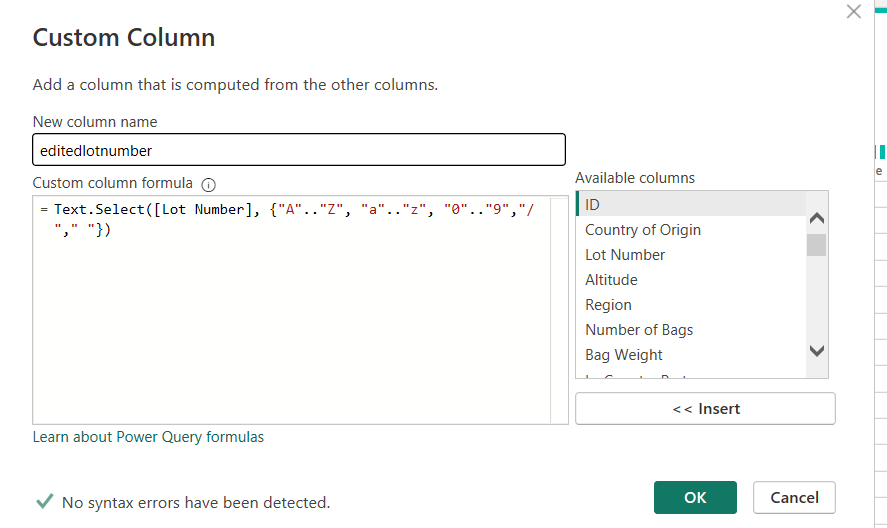
**“Total cup points”** is the total of all features given above**.**

**Data Preprocessing:** This coffee dataset consist of 207 rows. In this dataset few of the initial observations are there which are given below:

* **Consist of unwanted characters**
* **Consist of null values**
* **Presence of ~ symbol in altitude column**
* **Data transformation is required for Lot number, Region and In-country partner column.**

**Data transformation:**

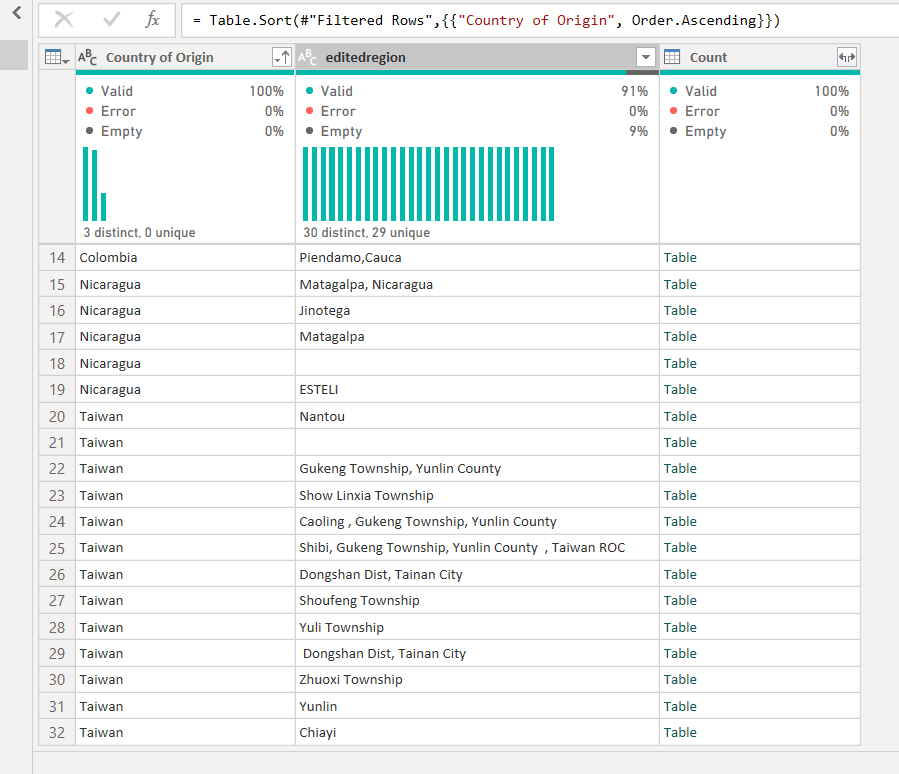
* The unwanted characters are removed by using the custom column feature in power BI tool.

****

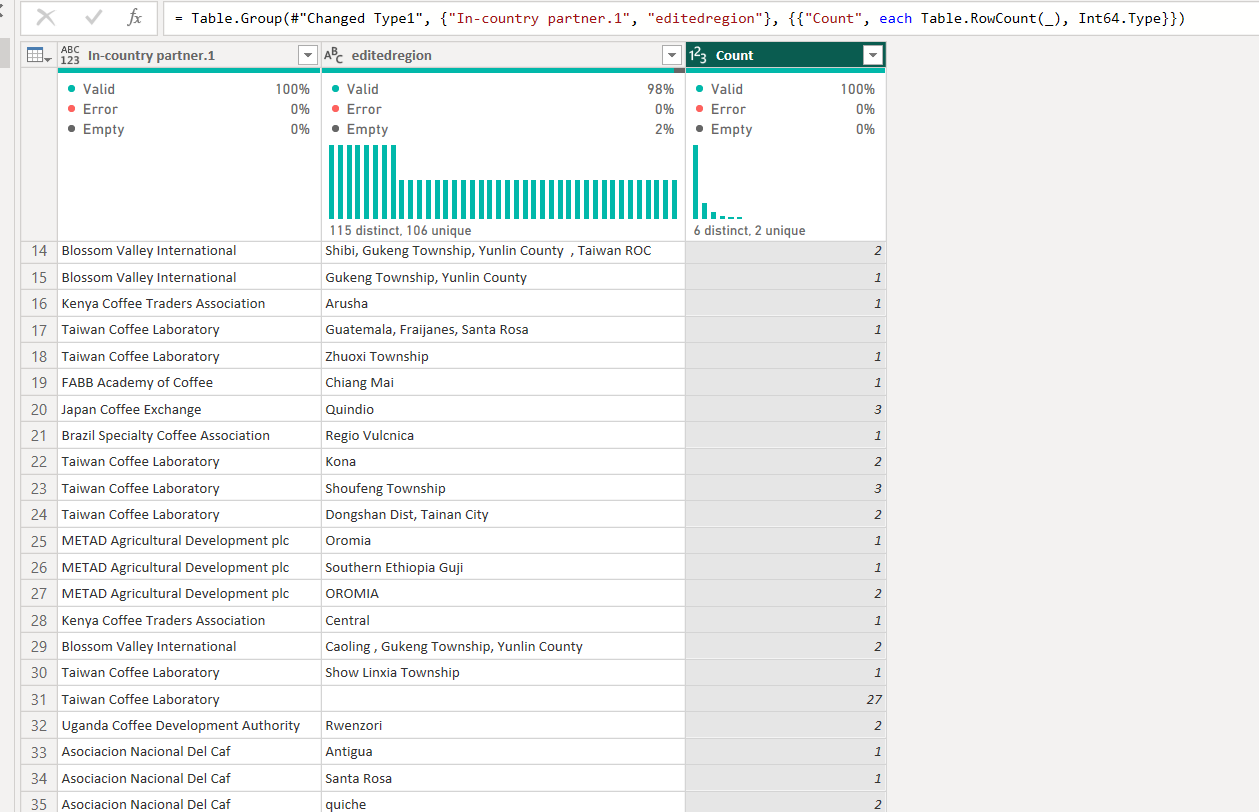
**Fig.1**

as shown in the Fig.1. In the similar fashion the other columns such as Region and In-country partner are customed.

* Presence of ~ value in one of cell from altitude column has been replaced by using replace value feature.
* Most of the null values are present in Region column which might be the useful for our analysis. Hence to fill the appropriate values in the missing Region cell grouping is incorporated to identify the changes.

****

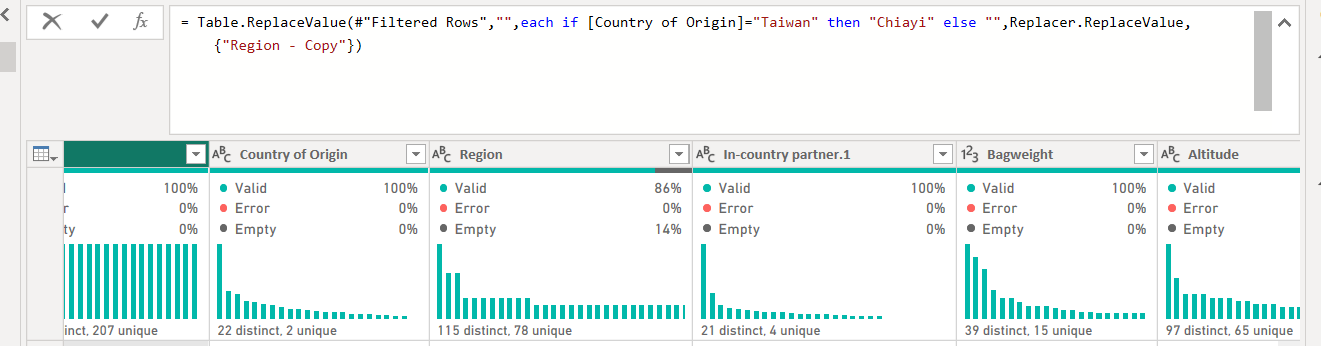
**Fig.2**



**Fig.3**

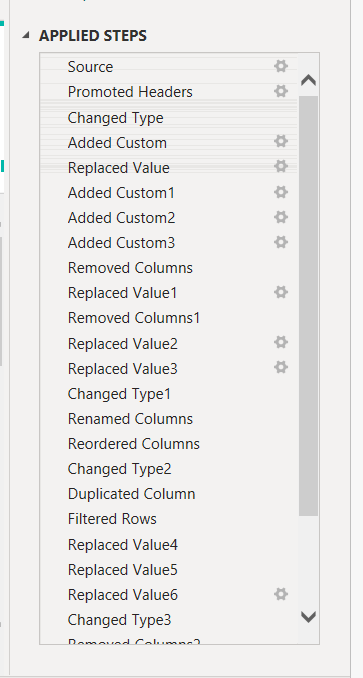
* Challenges faced on filling the missing values in Region column:

As per the observation made from above Fig.2 and Fig.3 mode value is used to fill the missing values. However, Region needs to be filled correctly on the basis of country of origin. For that the query is written to replace the values.

****

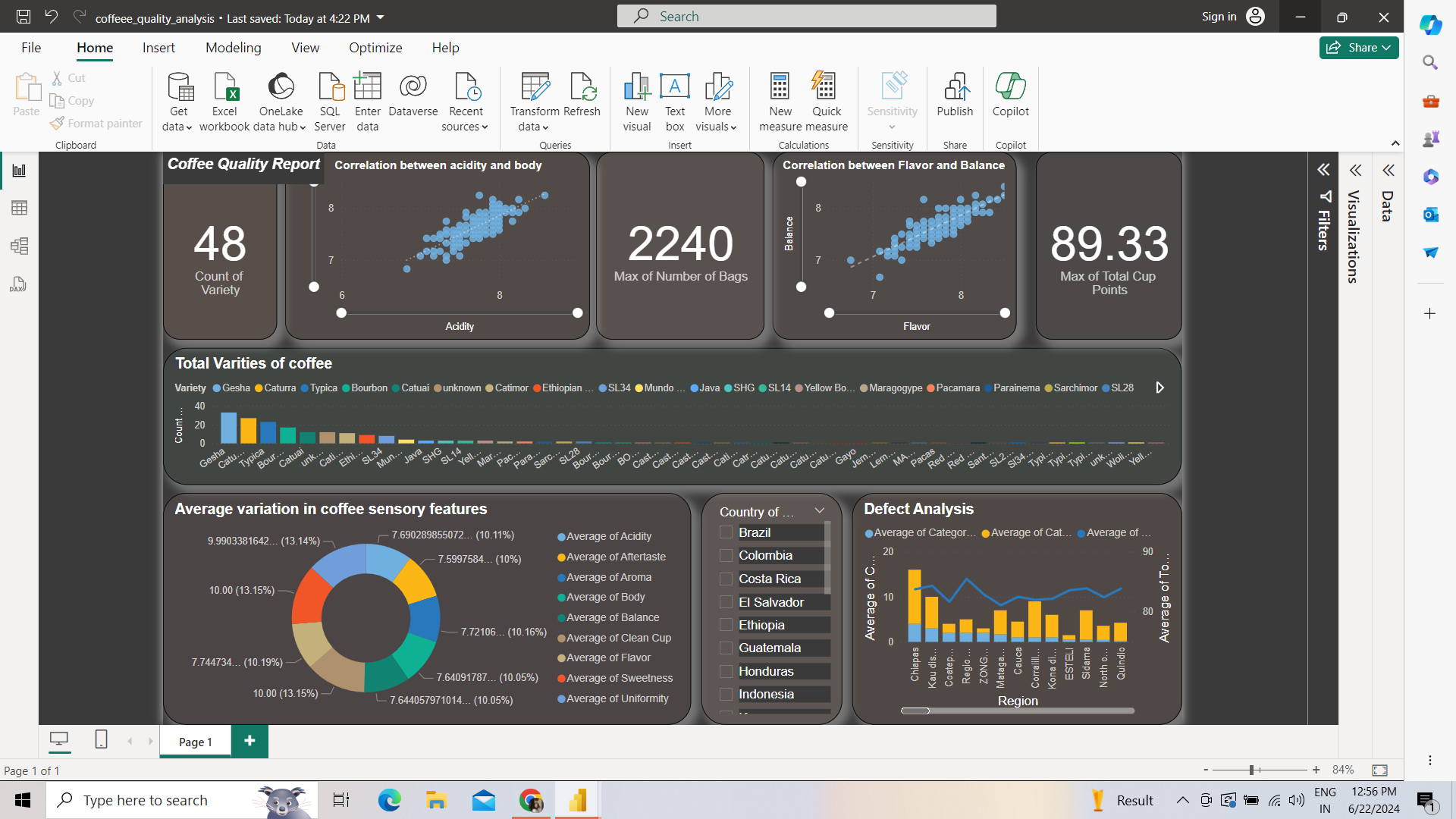
**Fig.4**

* The above steps are only the highlighted steps rest of the preprocessing is done to clean the data is shown in Applied steps screenshot given in Fig.5.



**Fig.5**

* **Power BI visualization:**

****

* **Key Insights Obtained from Power BI report:**
* **In the Brazil country of origin Mundo Novo variety of coffee is present more with the maximum of sweetness and uniformity as a leading quality factor.**
* **Maximum number which is 2240 of coffee bags are from Brazil country of origin.**
* **In the Taiwan country of origin acidity and body in coffee quality are correlated with each other.**
* **From the Indonesia country of origin category two defects are more from the Aceh Tengah Region.**
* **Recommendations:**
* **On the basis of coffee quality in the various countries one can identify the features responsible for the coffee quality such as climate condition, maintenance, government schemes etc. Hence the strategies can be built to implement the same.**