Assumptions:

After analyzing the data, I understand that the requirements is to provide coffee quality measures.

With that understanding, I came up with the following dimensions and facts to suffice the requirement.

**Dimensions**

1. Product: Sample Species Such as Arabica and Robusta
2. Customers: Customers are those who ask to grade their harvested coffee beans by a certifier
3. Locations: Country and Region where these Coffee beans come from
4. Certifiers: Registered entities who can certify the coffee by trained reviewers
5. Time: Grading Date on which day Coffee sample is certified

**Facts**

1. QualityMeasures: Coffee quality measures referenced per each sample in the certified data

Transformations That are applied:

1. Fuzzy String Matching Owner into Customer.Owner
2. Transformed ‘Category One Defects’ into a Number
3. Transformed ‘Category two Defects’ into a Number
4. Moisture% is converted into decimal
5. Separated ‘Certification Contact’ into name and contact

Further Scope for Transformations

Fuzzy String Matching: Most of the name columns come into this category

Ex: ‘Klem Organics’ and ‘KlemOrganics’

‘Certification Contact’ has alternate phone number. This can be further separated

Ex: Clare Rwakatogoro - 256-41-256940/233073

Grading\_date and Expiration Can be converted into DATE format instead of String