**Python Project**

Name of the Project:

Quality Assessment Analytics

**Done By : Tanmay Saha**

Introduction

This dataset contains relevant information about the patterns and associations in beverage manufacturing companies like Coke, Pepsi, etc.

1. **Id:** Product Id (this is system generated while collecting sample)
2. **Assembly line:** There are 2 different assembly lines A and B used for production of beverages.
3. **Quantity (lts.):** The amount of beverage filled in the bottle.
4. **CO2 dissolved:** Amount of CO2 for carbonation in gms.
5. **Time limit Crossed:** If the time limit has crossed to process the bottle (fill & packaging).

Objective

One of the primary objectives here is to understand the distributions and patterns within the data. We want to see what impact if our assembly lines are working fine or not. What are possible issues that could occur within the assembly lines?

Find the Correlations between numerical variables and Treatment of outliers.

Building the code

* Data Loading and Overview
* Getting a Summary of the Dataset
* Descriptive Statistics
* Count Number of Unique Values
* Checking for Unique Values
* Check for Duplicate Values
* Replace "\_" with NaN in the entire DataFrame
* Handling Missing Values
* Data can come in different styles or units. We need to make everything the same, like changing date formats or converting measurements
* Data can come in different styles or units. We need to make everything the same, like changing date formats or converting measurements
* Check if there are any rows with missing data. Substitute these values with the mean of numerical or mode for non-numerical data with respect to individual assembly lines
* A histogram will give a peek into how your distribution looks like. You can define the number of peaks which means it can be a unimodal, bimodal or multimodal distribution depending on the peaks involved.
* A box plot on the other hand helps you understand min, max and interquartile range where most of the data is centred around.

Analysis

In our analysis, we have witnessed that assembly line B has higher chances of introducing latency in filling the bottles as compared to assembly line A. This definitely has a negative impact on our overall production as it slows down the production rate and therefore can affect the final shipment.

The recommendation over here is to investigate the root cause asap. This may include fixing or upgrading the components involved in the production line, updating the softwares involved, etc.

In our analysis, we have encountered certain problems with CO2 dissolved or carbonation analysis. If you conduct further analysis, you will find that the carbonation values are very high (>15 gms per litre). Furthermore, there are chances of getting high carbonation which may lead to higher spillage upon opening the bottle.

This issue needs to be addressed immediately. Understand that it could be a data issue as well since there were outliers registered in CO2 dissolved and therefore, we may need to collect other samples to conclude things later on.

By addressing these challenges and capitalising on the insights drawn from the data, we believe the company can significantly enhance the production lines and therefore the final experiences of the end customers by timely delivery & less spillage issues.