**Project: Analyze Supermarket Data Across the Country - Company XYZ**

****

**Company XYZ** owns a supermarket chain across the country. Each major branch located in 3 cities across the country recorded sales information for 3 months, to help the company understand sales trends and determine its growth, as the rise of supermarkets competition is seen to increase.

The data folder contains datasets from three different branches; Lagos, Abuja and Port Harcourt. Each data file from the branches contains the same attribute information and see below the attribute description.

**- Description**

**Invoice ID:** Customer Identification number

**Branch:** Supermarket Branch across the country (A, B, C)

A - Lagos Branch

B - Abuja Branch

C - Port Harcourt Branch

**City**: Supermarket Location

**Customer Type:** Type of customers, Members - Returning customer with membership card, Normal - Customer without membership (could be returning, first-time or walk-in customer)

**Gender**: Customer Gender Information

**Product line:** Product categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel

**Unit Price**: Price of each product in Naira

**Quantity**: Number of products purchased by customer

**Tax**: 5% tax fee for customer buying

**Total**: Total price including tax

**Date**: Date of purchase (Supermarket Record available from January 2019 to March 2019)

**Time**: Purchase time (Supermarket Hours - 10am to 9pm)

**Payment**: Payment used by customer for purchase (3 methods are available – Cash, Card and Epay)

**COGS**: Cost of goods sold

**Gross margin percentage**: Gross margin percentage

**Gross income:** Gross income

**Rating**: Customer Satisfaction rating on their overall shopping experience (On a scale of 1 to 10)

**- Projects Requirements**

**Step 1 - Loading Datasets**

* Correct use of pathname pattern - glob
* Combine all the files generated in a list and export to a CSV.

**Step 2 - Data Exploration**

* Use the head() method to view first few rows of the dataset
* Check the number of rows and columns present in the data using the shape attribute.
* Generate the names of the columns using the columns attribute.
* Use describe function to generate the statistical summary of the dataframe
* Use meaningful sentences to describe findings from the data statistical summary
* Use of correct method to check for Missing values
* Check the information of the DataFrame using the info method.

**Step 3 - Dealing with DateTime Features**

* Use to\_datetime() to convert the date column to datetime
* Check the datatype to confirm if it's in datetime
* Accurate conversion of the time column & prints appropriate data type
* Accurate extraction of the Day, Month, Year & Hour features
* The numbers of unique hours of sales in the supermarket are accurately determined.
* Result that shows an array that contains the unique sales hours.

**Step 4 - Unique Values in Columns**

* Appropriate method to generate the unique values in the categorical columns (apart from the example - Branch column).
* Generated the count figure of the categorical values using the value\_counts() method.

**Step 5 - Aggregation with GroupBy**

* A groupby object with the "City Column", and aggregation function of sum and mean.
* A table that shows the gross income of each city, and determines the city with the highest total gross income.
* Optional - Use of appropriate methods & descriptions to explore other columns such as Unit Price, Quantity.

**Step 6 - Data Visualization**

* Appropriate use of countplot to determine the branch with the highest sales record.
* Optional - Appropriate use of countplot to determine the most used payment method & city with the most sales.
* Appropriate use of countplot to determine the highest & lowest sold product line.
* Result that shows the Payment channel used by most customers to pay for each product line. Chart should also show the "product line" column on the Y-axis, and the "hue" parameter for the "Payment" column.

**- Getting Started**

To start the project, navigate to this [GitHub repository](https://github.com/bleso-a/Data-Analysis-Project) and fork the project. The instructions and steps to complete the project can be found in the `Readme` file and `starter\_notebook.ipynb` file.