

Analyze Supermarket Data Across the Country - Company XYZ

Executive Summary

Introduction

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.

- 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

Conclusion

This is a data analysis performed on the recorded data of the three major branches of the supermarket XYZ across the country. The analysis was successfully carried out through the technological tools employed which include Python, Pandas, Seaborn among others. The data from all the branches were

combined in one file for easy exploration and analysis. This process aid the successful completion of the tasks completed which include:

1. Generation of important statiscal summaries
2. Conversion of date and time to datetime and extraction of specific date features.
3. Data visualizations to uncover important insights.

Insights

The insights I was able to uncover from analysing the dataset are as follows;

1. the total number of rows and and columns being 1000,17 respectively.
2. the city with the highes gross income is Port Harcourt.
3. the Branch with the highest Sales record is Branch A, lagos.
4. the most used payment method is epay
5. the Highest and Lowest Sold product lines, highest in Fashion Accesories and lowest in health and beauty
6. the city with most sales is Lago
7. Most used payment channel for each branch is different
8. Branch B has the lowest rating among the three branches
9. The closest thing both genders buy is in Electronic Appliances
- 10.The farthest thing both genders buy is in Home and Lifestyle