Grocery Store Sales - YesFood

Overview

YesFood is a grocery store offering a variety of items, including produce, meat, dairy, baked goods, snacks, and other essential household foods. As food prices increase, YesFood aims to ensure they stock products across all categories that cater to a wide range of price points, making sure they have options for all types of customers.

Problem Statement

YesFood, a grocery store chain, wants to analyze their product data from 2022 to understand pricing variations and identify opportunities to cater to a wider range of customer budgets. This is particularly important as food prices are increasing, and they want to ensure they offer affordable options across all categories.

Goals

01. Clean the Data

Address missing values in various columns (brand, weight, price, year_added, stock_location).

02. Analyze Price Variations

Identify the minimum and maximum prices for each product category.

03. Focus on High-Selling Categories

Analyze meat and dairy products with high average unit sales (more than 10) to understand their price points and potential for catering to different budgets.

Dataset

The dataset is a CSV file named "**products.csv**" containing customer purchase records from YesFood's loyalty program for the year 2022. It includes information about each product, such as:

- **product_id:** Unique product identifier
- **product_type:** Category (Produce, Meat, Dairy, Bakery, Snacks)
- **brand:** Brand name
- **weight:** Weight in grams
- **price:** Price in US dollars
- **average_units_sold:** Average number of units sold per month
- **year_added:** Year the product was added to YesFood's inventory
- **stock_location:** Warehouse location (A, B, C, D)

Data Findings

Table 1.

During 2022, a bug in the product system caused the **`year_added`** value to be missing for some products introduced that year. Since the year a product was added could affect its price, having this information is crucial. The table identifies how many products have a missing **`year_added`** value. The output should be a single column, **`missing_year`**, with a single row giving the number of missing values.

- Fields : 1
- Records : 1

Table 2.

Clean the missing values

- Fields : 8
- Records : 1700

Table 3.

To analyze price variations, this query identifies the minimum and maximum prices for each product category. The query returns the **`product_type`**, **`min_price`** and **`max_price`** columns.

- Fields : 3
- Records : 5

Table 4.

Additionally, the table focuses on meat and dairy products where the average units sold exceeded ten, allowing for a more detailed analysis. The query returns the **`product_id`**, **`price`** and **`average_units_sold`** columns.

- Fields : 3
- Records : 698

Data Findings

	product_id	product_type	brand	weight	price	average_units_sold	year_added	stock_location
0	1	Bakery	TopBrand	602.61 grams	11.00	15	NaN	C
1	2	Produce	SilverLake	478.26	8.08	22	2022.0	C
2	3	Produce	TastyTreat	532.38 grams	6.16	21	2018.0	B
3	4	Bakery	StandardYums	453.43 grams	7.26	21	2021.0	d
4	5	Produce	GoldTree	588.63	7.88	21	2020.0	a
...	...	...	...	...	...	...	...	...
1695	1696	Meat	TastyTreat	503.99 grams	14.08	25	2017.0	A
1696	1697	Meat	GoldTree	526.89	16.13	25	2016.0	D
1697	1698	Bakery	YumMie	583.85	7.05	16	2021.0	a
1698	1699	Produce	TopBrand	441.64	8.10	19	2019.0	A
1699	1700	Meat	TopBrand	518.6	15.89	24	2021.0	A

1700 rows × 8 columns

Table 1.

During 2022, a bug in the product system caused the **`year_added`** value to be missing for some products introduced that year. Since the year a product was added could affect its price, having this information is crucial. The table identifies how many products have a missing **`year_added`** value. The output should be a single column, **`missing_year`**, with a single row giving the number of missing values.

- Fields : 1
- Records : 1

Data Findings

 # Table 2.

Clean the missing values

- Fields : 8
- Records : 1700

	product_id	product_type	brand	weight	price	average_units_sold	year_added	stock_location
0	1	Bakery	TopBrand	602.61	11.00	15	2022	C
1	2	Produce	SilverLake	478.26	8.08	22	2022	C
2	3	Produce	TastyTreat	532.38	6.16	21	2018	B
3	4	Bakery	StandardYums	453.43	7.26	21	2021	D
4	5	Produce	GoldTree	588.63	7.88	21	2020	A
...	...	...	...	...	...	...	...	...
1695	1696	Meat	TastyTreat	503.99	14.08	25	2017	A
1696	1697	Meat	GoldTree	526.89	16.13	25	2016	D
1697	1698	Bakery	YumMie	583.85	7.05	16	2021	A
1698	1699	Produce	TopBrand	441.64	8.10	19	2019	A
1699	1700	Meat	TopBrand	518.60	15.89	24	2021	A

1700 rows × 8 columns

Data Findings

	product_type	min_price	max_price
0	Snacks	5.20	10.72
1	Produce	3.46	8.78
2	Dairy	8.33	13.97
3	Bakery	6.26	11.88
4	Meat	11.48	16.98

Table 3.

Key Insights:

- **Price Range Variation:** The table reveals significant price variations across different product categories. For instance, meat products tend to have higher prices compared to produce or snacks.
- **Category-Specific Pricing:** Each product category has its own price range, influenced by factors such as production costs, supply chain, and perceived value.
- **Potential for Targeted Pricing Strategies:** YesFood can leverage this information to implement targeted pricing strategies for different product categories and customer segments.

Possible Interpretations:

- **Product Differentiation:** Higher-priced products within a category may offer premium quality, organic ingredients, or unique features.
- **Competitive Pricing:** YesFood may be aligning its prices with competitors or industry standards.
- **Customer Segmentation:** The variation in prices can cater to different customer segments with varying budgets and preferences.

Data Findings

	product_id	price	average_units_sold
0	6	16.20	24
1	8	15.77	28
2	9	11.57	30
3	10	13.94	27
4	11	9.26	26
...	...	...	...
693	1694	16.00	25
694	1695	12.88	20
695	1696	14.08	25
696	1697	16.13	25
697	1700	15.89	24

698 rows × 3 columns

Table 4.

Potential Insights:
By analyzing this data, we can gain insights into the relationship between product price and sales volume. Here are some potential observations:

- **Price and Sales Correlation:**
  - There might be a correlation between price and average units sold. A higher price might lead to fewer units sold, or vice versa.
  - However, other factors like product quality, brand reputation, and marketing efforts can also influence sales.
- **Product Popularity:**
  - Products with higher average unit sales might be more popular or in higher demand.
  - Identifying these products can help inform inventory management and marketing strategies.
- **Pricing Strategy:**
  - Analyzing the price distribution can help assess the effectiveness of the pricing strategy.
  - Are there opportunities to optimize prices to increase sales or profit margins?

Insights

01.

The analysis highlights the importance of clean data for accurate insights.

02.

YesFood offers products across various price points within each category, potentially catering to a diverse customer base.

03.

Analyzing high-selling products provides detailed information for strategic pricing decisions and ensuring affordability for specific customer segments.