abdulrab2101 done 4da8cb5 · 16 minutes ago

567 lines (567 loc) · 24.4 KB

Preview Code Blame Raw

Retail Data Project

The data I obtained from the webpage is found here***.***

Q: What information is provided about the total sales for different product categories?

This bar chart displays the total sales for different product categories. The categories and their respective total sales are as follows:

- **Electronics**: Just above $90,000,000.
- **Grocery**: Just below $90,000,000.
- **Clothing**: Approximately $70,000,000.
- **Books**: Slightly above $70,000,000.
- **Home Decor**: Just below $70,000,000.

Each bar is colored in pink, with the length of the bar corresponding to the total sales amount in each product category. The y-axis represents the total sales, while the x-axis represents the different product categories.

In [1]:
```
SELECT Product_Category, SUM(Total_Amount) as Total_Sales
FROM Sales
WHERE Product_Category IS NOT NULL AND Total_Amount IS NOT NULL
GROUP BY Product_Category
ORDER BY Total_Sales DESC;
```

(5 rows affected)
Total execution time: 00:00:00.646

Out[1]:

Product_Category	Total_Sales
Electronics	97398318.24399948
Grocery	91076809.20081711
Clothing	74830333.54456043
Books	74588468.58282375
Home Decor	74291146.9127779

Q: What does the data tell us about the number of orders for different order statuses?

This bar chart displays the number of orders for different order statuses. The order statuses and their respective number of orders are as follows:

- **Delivered**: The highest number of orders, around 130,000.
- **Shipped**: The second highest, approximately 80,000 orders.
- **Processing**: Around 60,000 orders.
- **Pending**: About 40,000 orders.
- **NULL**: Very few or no orders.

Each bar is colored in pink, with the length of the bar corresponding to the number of orders in each status category. The x-axis represents the number of orders, while the y-axis represents the different order statuses.

In [5]:
```
SELECT Order_Status, COUNT(*) as Number_of_Orders
```

```
SELECT Order_Status, COUNT(*) as Number_of_Orders
FROM Sales
GROUP BY Order_Status
ORDER BY Number_of_Orders DESC;
```

(5 rows affected)
Total execution time: 00:00:00.637

Out[5]:

Order_Status	Number_of_Orders
Delivered	130449
Shipped	65024
Processing	57199
Pending	49103
NULL	235

Q: Find the number of orders from each country, excluding those with a null value for the country, and sort the results by the number of orders in descending order?

- **Selecting Columns:**
  - We want to select the `Country` column to identify each country.
  - We use `COUNT(*)` to count the number of orders for each country. We alias this count as `Number_of_Orders` for clarity in the output.
- **Filtering Rows:**
  - We need to exclude rows where the `Country` is null. This is achieved using the `WHERE Country IS NOT NULL` condition.
- **Grouping Results:**
  - To count the number of orders per country, we use the `GROUP BY Country` clause. This groups the results by each unique country value.
- **Ordering Results:**
  - We want to sort the results by the number of orders in descending order. This is done using the `ORDER BY Number_of_Orders DESC` clause.

In [2]:

```
SELECT Country, COUNT(*) as Number_of_Orders
FROM Sales
WHERE Country IS NOT NULL
GROUP BY Country
ORDER BY Number_of_Orders DESC;
```

(5 rows affected)
Total execution time: 00:00:00.695

Out[2]:

Country	Number_of_Orders
USA	95223
UK	63066
Germany	52830
Australia	45319
Canada	45301

Q: How do the total sales compare across the USA, UK, Germany, Australia, and Canada based on the given data?

A: The graph illustrates the total sales in millions for five countries: USA, UK, Germany, Australia, and Canada. The USA leads with the highest sales, followed by the UK, Germany, Australia, and Canada.

In [3]:
```
SELECT Country, SUM(Total_Amount) as Total_Sales
FROM Sales
WHERE Country IS NOT NULL AND Total_Amount IS NOT NULL
GROUP BY Country
ORDER BY Total_Sales DESC;
```

(5 rows affected)
Total execution time: 00:00:00.486

Out[3]:

Country	Total_Sales
USA	129722153.95488453
UK	86667468.17741013
Germany	72129649.2223692
Australia	61994472.37411213
Canada	61703209.44601822

Q: What information does the data provide about the distribution of payment methods such as Credit Card, Debit Card, Cash, and PayPal?

A: The doughnut chart shows the proportion of different payment methods used. The chart includes Credit Card, Debit Card, Cash, and PayPal, with each segment representing the share of each payment method.

In [8]:
```
SELECT Payment_Method, COUNT(*) as Number_of_Transactions
FROM Sales
WHERE Country = 'USA' AND Payment_Method IS NOT NULL
GROUP BY Payment_Method
ORDER BY Number_of_Transactions DESC;
```

(4 rows affected)
Total execution time: 00:00:00.533

Out[8]:

Payment_Method	Number_of_Transactions
Credit Card	26877