Data Analysis Assignment

NOV 2024

Sales Data Analysis Assignment

In this assignment, you will analyze a dataset containing monthly sales records from a fictional store. This exercise will guide you through essential data manipulation, cleansing, and analysis steps. By the end, you should be able to answer key business questions based on the data provided.

Part 1: Data Parsing and Cleansing

The dataset is provided as a zipped file containing monthly sales data files for the year 2019. Each file follows the same structure, with columns:

- Order ID
- Product
- Quantity Ordered
- Price Each
- Order Date
- Purchase Address

Your first task is to parse, clean, and consolidate the data for analysis. Follow these steps:

1. Parsing and Merging:

- Write a function to load and parse each file.
- Consolidate all 12 files into a single CSV file containing the data for the entire year.

2. Data Cleansing:

- Remove duplicate headers if they appear as rows within the files.
- Cast each column to the appropriate data type. All columns are initially loaded as object types in pandas.
- Filter the data to include only records from 2019.

3. Data Transformation:

- Extract City and State information from the "Purchase Address" column and create two new columns for them.

- Identify the granularity of the data. Note that an order may span multiple rows if it contains different products. Consider each unique "Order ID" as a single order.

4. Data Enrichment:

- Calculate Order Counts: Count the number of orders by finding the unique "Order ID" values.

- Subtotal Calculation: Create a new column that calculates the subtotal for each item by multiplying "Quantity Ordered" by "Price Each".

- Category Mapping: Map each product to its respective category using the following dictionary:

```
product_category = {
    'Google Phone': 'Mobiles',
    'iPhone': 'Mobiles',
    'Vareebadd Phone': 'Mobiles',
    'USB-C Charging Cable': 'Accessories',
    'Lightning Charging Cable': 'Accessories',
    'AA Batteries (4-pack)': 'Accessories',
    'AAA Batteries (4-pack)': 'Accessories',
    'Bose SoundSport Headphones': 'Headphones',
    'Wired Headphones': 'Headphones',
    'Apple Airpods Headphones': 'Headphones',
    'Macbook Pro Laptop': 'Laptops',
    'ThinkPad Laptop': 'Laptops',
    '27in 4K Gaming Monitor': 'Monitors',
    '27in FHD Monitor': 'Monitors',
    '20in Monitor': 'Monitors',
    '34in Ultrawide Monitor': 'Monitors',
```

'Flatscreen TV': 'Home Appliances',

'LG Dryer': 'Home Appliances',

'LG Washing Machine': 'Home Appliances'

}

Part 2: Exploratory Data Analysis (EDA)

After preparing the data, perform Univariate and Bivariate Analysis on each column to understand the data distribution and relationships.

Part 3: Business Questions

Use the cleaned and merged dataset to answer the following business questions:

1. Product Analysis:

- Identify the best-selling product. Consider:

  - The product that appears in the most orders.

  - The product with the highest quantity ordered.

  - The product generating the highest revenue.

2. Trend Analysis:

- Analyze weekly trends to determine the number of orders placed during weekends.

- Identify preferred months for customer orders.

3. Hourly Analysis:

- Determine the peak hours for order placement. Use this information to recommend ideal staff scheduling times to handle peak customer demand.

4. Category Analysis:

- Analyze each product category over time. Identify trends and find out which category performs best in terms of sales and revenue.

5. Customer Loyalty:

- Identify the most loyal customers by analyzing:

  - The customer with the highest number of orders.

  - The customer generating the highest revenue.

6. Geographical Analysis:

- For each city and state, identify the most preferred product category.

Submission

- Include your code and output for each task in one Jupyter Notebook.

- Clearly state any assumptions you make and describe the approach you use to answer each question.

- you should also submit a report with all your findings and observations answering all the questions related to the business we encourage you to use visualization inside the report.