

PROJECT DETAILS

Coders Cave Objective: Analyze customer behaviour based on transaction data

Develop a data analysis project that focuses on understanding and analyzing customer behaviour based on transactional data. The goal is to perform Exploratory Data Analysis (EDA) to derive valuable insights into customer preferences, purchasing patterns, and overall behaviour.

Details

Context:

Nielsen Holdings plc (NYSE: NLSN) is a global measurement and data analytics company that provides the most complete and trusted view available of consumers and markets worldwide.

Nielsen receives transaction level scanning data (POS Data) from its partner stores on a regular basis. Stores sharing POS data include bigger format store types such as supermarkets, hypermarkets as well as smaller traditional trade grocery stores (Kirana stores), medical stores etc. using a POS machine.

While in a bigger format store, all items for all transactions are scanned using a POS machine, smaller and more localized shops do not have a 100% compliance rate in terms of scanning and inputting information into the POS machine for all transactions.

A transaction involving a single packet of chips or a single piece of candy may not be scanned and recorded to spare customer the inconvenience or during rush hours when the store is crowded with customers.

Thus, the data received from such stores is often incomplete and lacks complete information of all transactions completed within a day.

Additionally, apart from incomplete transaction data in a day, it is observed that certain stores do not share data for all active days. Stores share data ranging from 2 to 28 days in a month. While it is possible to impute/extrapolate data for 2 days of a month using 28 days of actual historical data, the vice versa is not recommended.

Nielsen encourages you to create a model which can help impute/extrapolate data to fill in the missing data gaps in the store level POS data currently received.

Project Portfolio:

1. https://github.com/vishnuvardhankunsoth/Analyze_customer_behavior_based_on_transaction_data/blob/main/Normal%20Task.ipynb
2. https://github.com/vishnuvardhankunsoth/Analyze_customer_behavior_based_on_transaction_data/blob/main/README.md

Data set:

The dataset that contains store level data by brands and categories for select stores. The file contains brand level data for 10 stores for the last 3 months. This can be referred to as the ideal data.

CSV Files:

1. https://github.com/vishnuvardhankunsoth/Store_Transaction_Data_Insights/files/14276082/Hackathon_Ideal_Data.csv
2. https://github.com/vishnuvardhankunsoth/Store_Transaction_Data_Insights/files/14276116/Hackathon_Mapping_File.csv
3. https://github.com/vishnuvardhankunsoth/Store_Transaction_Data_Insights/files/14276142/Data.set.zip

Tasks:

1. Importing python libraries.
2. Importing CSV file.
3. Performing EDA (Exploratory Data Analysis).
4. Exploring individual columns.
5. Creating Visualizations.
6. Insights.

Insights from analysis:

1. No Null Values: There are no missing values in the dataset.
2. Rows = 14260, Columns = 10: There are 14260 rows and 10 columns in the dataset.
3. The dataset covers three months of data.(i.e., M1,M2,M3)
4. There are a total of ten stores represented in the dataset.
5. The mean quantity sold is 16.354488.
6. The standard deviation of quantity sold is 34.365583.
7. The mean purchase amount is 294.455330.
8. The standard deviation of purchase amounts is 760.129558.
9. The median purchase amount is 99.0
10. The dataset contains 80 unique products (GRP)
11. The month with the highest sales is Month 2 (M2)
12. The fourth store achieved the maximum quantity sold.
13. The eighth store generated the highest sales revenue.
14. Plotted graphs by the top 5 categories sold from each store.
15. Plotted graphs by the top 5 categories sold by each month.