

MARKET BASKET INSIGHTS

USING MACHINE LEARNING

Project Title: Market Basket Insights

Phase 3: Development part 1

Topic: *Start building the Market Basket Insights model by loading and pre-processing the dataset*

Market Basket Insights

Introduction:

Market Basket Analysis (MBA) is a powerful data mining technique that has revolutionized the way businesses understand and leverage customer behavior. It is a method used to unearth hidden patterns and associations within vast transactional datasets, particularly in the retail and e-commerce sectors. By analyzing customer purchase histories, MBA helps businesses identify relationships between products that are frequently purchased together.

In essence, MBA provides valuable insights into the question of "What do customers buy together?" This knowledge is instrumental in enhancing various aspects of business operations, such as inventory management, product recommendations, and marketing strategies.

This introduction to market basket analysis will delve deeper into the key concepts and methodologies, its real-world applications, and the benefits it brings to organizations looking to optimize their product offerings and customer experiences. It's a tool that empowers businesses to make data-driven decisions and strengthen their competitiveness in today's data-centric marketplace.

Necessary step to follow:

1.Import Libraries:

Start by importing the necessary libraries.

Program:

```
import pandas as pd

# Load the dataset

dataset_path = '/kaggle/input/market-basket-analysis/Assignment-1_Data.xlsx'

df = pd.read_excel(dataset_path)
```

2.Initial Exploration:

We'll perform an initial exploration of the dataset to understand its structure and characteristics.

Program:

```
print("Number of rows and columns:", df.shape)

print("\nData Types and Missing Values:")

print(df.info())

print("\nFirst few rows of the dataset:")
```

```
print(df.head())
```

Output:

Number of rows and columns: (522064, 7)

Data Types and Missing Values:

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 522064 entries, 0 to 522063

Data columns (total 7 columns):

#	Column	Non-Null Count	Dtype
0	BillNo	522064 non-null	object
1	Itemname	520609 non-null	object
2	Quantity	522064 non-null	int64
3	Date	522064 non-null	datetime64[ns]
4	Price	522064 non-null	float64
5	CustomerID	388023 non-null	float64
6	Country	522064 non-null	object

dtypes: datetime64[ns](1), float64(2), int64(1), object(3)

memory usage: 27.9+ MB

None

First few rows of the dataset:

	BillNo	Itemname	Quantity	Date \
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00
1	536365	WHITE METAL LANTERN	6	2010-12-01 08:26:00
2	536365	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01

08:26:00

4 536365 RED WOOLLY HOTTIE WHITE HEART. 6 2010-12-01

08:26:00

	Price	CustomerID	Country
0	2.55	17850.0	United Kingdom
1	3.39	17850.0	United Kingdom
2	2.75	17850.0	United Kingdom
3	3.39	17850.0	United Kingdom

Preprocessing:

We'll preprocess the data to ensure it's ready for analysis.

```
#Check Missing Values
```

```
print("Missing Values:")
```

```
print(df.isnull().sum())
```

```
#Drop Rows with Missing Values
```

```
df.dropna(inplace=True)
```

Output:

Missing Values:

BillNo 0

Itemname 1455

Quantity 0

Date 0

Price 0

CustomerID 134041

Country 0

dtype: int64n

DataFrame:

Convert dataframe into transaction data

```
transaction_data = df.groupby(['BillNo', 'Date'])['Itemname'].apply(lambda  
x: ', '.join(x)).reset_index()
```

#Drop Unnecessary Columns

```
columns_to_drop = ['BillNo', 'Date']
```

```
transaction_data.drop(columns=columns_to_drop, inplace=True)
```

Save the transaction data to a CSV file

```
transaction_data_path = '/kaggle/working/transaction_data.csv'
```

```
transaction_data.to_csv(transaction_data_path, index=False)
```

Display the first few rows of the transaction data

```
print("\nTransaction Data for Association Rule Mining:")
```

```
print(transaction_data.head())
```

```
transaction_data.shape
```

Output:

Transaction Data for Association Rule Mining:

Itemname

0 WHITE HANGING HEART T-LIGHT HOLDER, WHITE META...

1 HAND WARMER UNION JACK, HAND WARMER RED POLKA DOT

2 ASSORTED COLOUR BIRD ORNAMENT, POPPY'S PLAYHOU...

3 JAM MAKING SET WITH JARS, RED COAT RACK PARIS ...

4 BATH BUILDING BLOCK WORD

(18192, 1)

Conclusion:

Market Basket Analysis is a valuable technique that offers profound insights into customer behavior and product associations, making it an indispensable tool for businesses across various industries. In conclusion, here are some key takeaways:

Enhanced Customer Understanding: MBA provides a deeper understanding of customer preferences and purchase patterns. By identifying which products are frequently bought together, businesses can tailor their strategies to meet customer demands more effectively.

Improved Inventory Management: MBA helps businesses optimize inventory levels by stocking products that are commonly bought together. This not only reduces the risk of overstocking but also ensures that popular items are consistently available.

Personalized Recommendations: Through the insights gained from market basket analysis, businesses can offer personalized product recommendations to customers. This enhances the shopping experience and increases the likelihood of cross-

selling and upselling.

Targeted Marketing: MBA enables more precise and efficient marketing efforts. Businesses can create targeted promotions and campaigns based on the associations they discover, increasing the chances of attracting and retaining customers.