

Kelompok 2

Groceries (Market Basket Analysis)





Project by :

1. Denggan Aulia Lubis (winning eleven)
2. Ismi Oktaviani (winning eleven)
3. putri andriani (winning eleven)
4. Miftah khuljanah (Winning Eleven)
5. Rizkiyatul Komariyah (Jumatec)
6. MOKHAMMAD FATIKHUL IKHSAN (jumatec)
7. Rafli Saviola (Jumatec)
8. Jaya Hartina Harahap (Jumatec)

Workflow

✓ Business Understanding

✓ Tujuan

✓ Manfaat

✓ Data Understanding

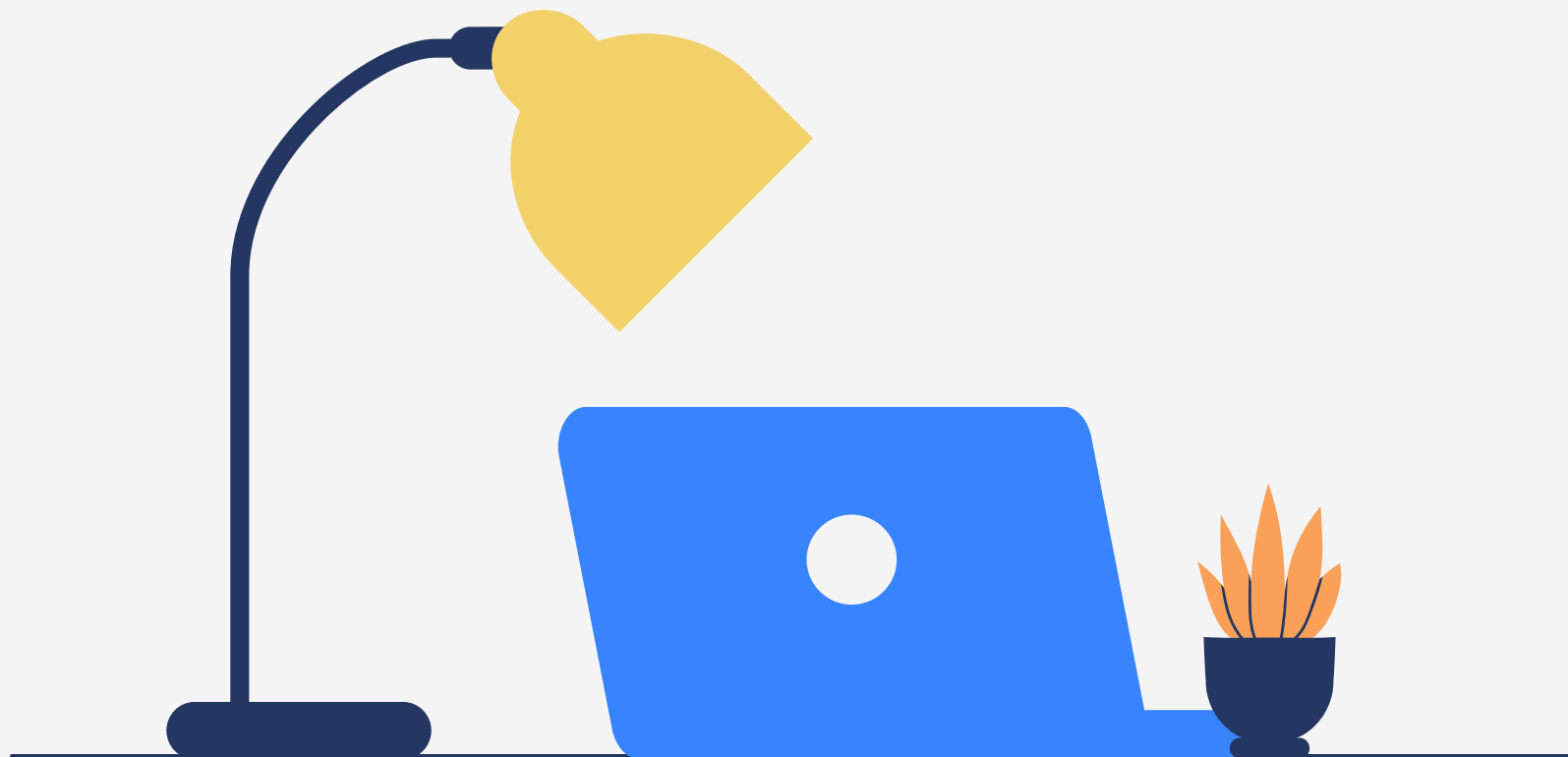
✓ Data Preparation

✓ Modelling

✓ visualisasi dan memahami data

✓ Market basket analysis

✓ Conclusion



Business Understanding

Groceries adalah toko bahan makanan yang menyediakan pesanan pembelian orang-orang dari toko kelontong. Groceries memiliki penjualan yang cukup tinggi dan juga ingin memiliki peningkatan inovasi dalam mencapai profit penjualan yang maksimal.

Tim proyek 2 akan membantu Groceries dalam membuat rekomendasi produk yang dapat membantu meningkatkan profit penjualan dengan menggunakan data penjualan.



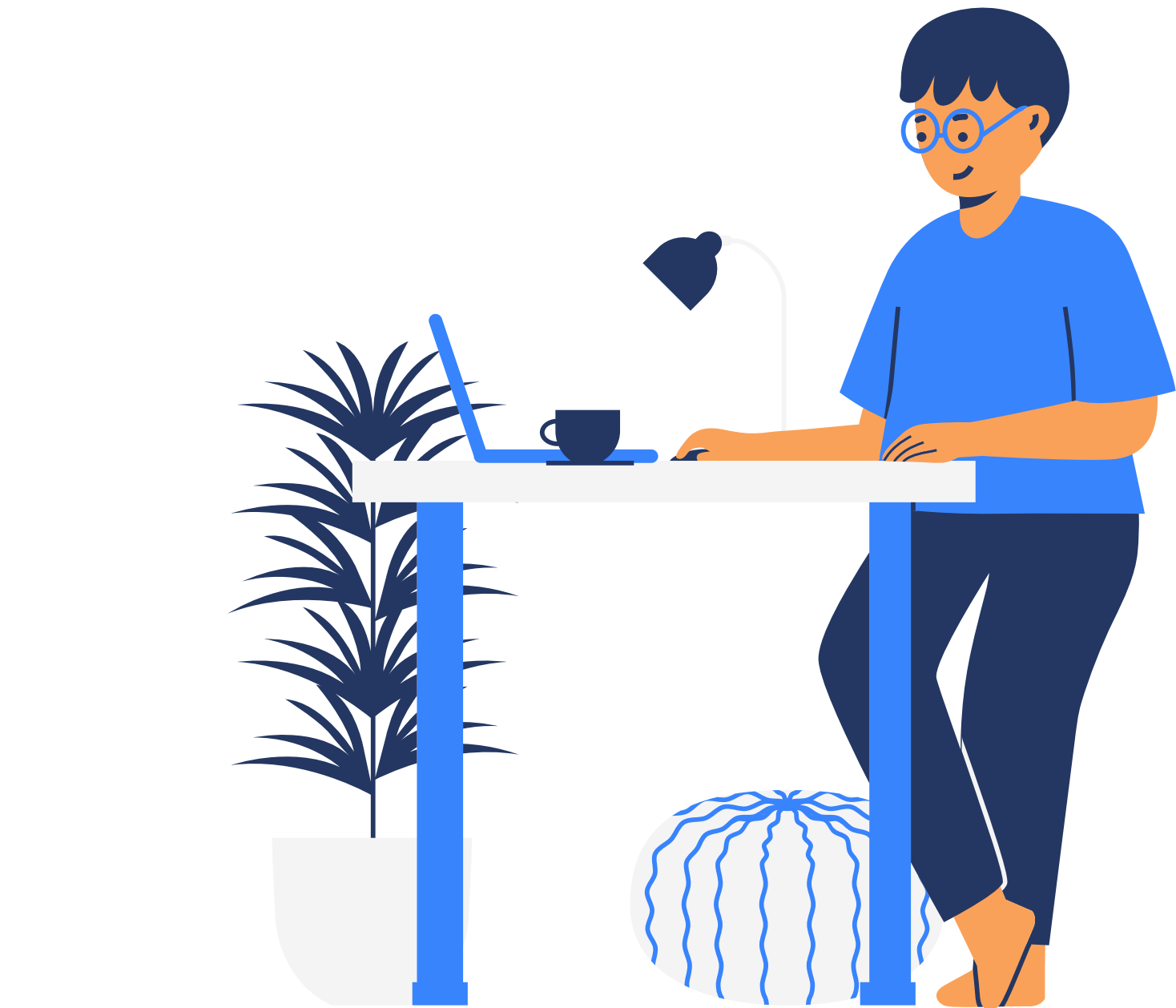
Tujuan

Dengan Market Basket Analysis yaitu untuk melakukan analisis buying habit konsumen dengan menemukan asosiasi antar beberapa item yang berbeda, yang diletakkan konsumen dalam shopping basket (keranjang belanja) yang dibeli pada suatu transaksi tertentu. Tujuan dari Market Basket Analysis ini adalah untuk menentukan produk manakah yang pelanggan beli dalam waktu bersamaan. Hasil dari metode MBA kemudian akan dianalisis untuk merekomendasikan paket produk guna meningkatkan profit penjualan



Manfaat

Dengan memanfaatkan market basket analysis, akan mendapatkan pengetahuan tentang produk apa yang dibeli pelanggan (what), produk apa saja (which) yang sering dibeli secara bersama-sama dan berpeluang untuk dipromosikan, siapakah mereka (who) dan mengapa mereka melakukan suatu pembelian (why). Serta dengan menggunakan Market basket analysis dapat meningkatkan penjualan dan efisiensi promosi berdasarkan rekomendasi paket produk.





Data understading

Member number = nomer member

Date = Tanggal

Transaction = Transaksi

Item description = deskripsi barang

Data Understanding



	Member_number	Date	Transaction	itemDescription
0	1808	21/07/2015	1	tropical fruit
1	2552	05/01/2015	2	whole milk
2	2300	19/09/2015	2	pip fruit
3	1187	12/12/2015	3	other vegetables
4	3037	01/02/2015	3	whole milk
...
19995	1721	15/10/2015	9111	whole milk
19996	2877	10/04/2015	9112	waffles
19997	3848	11/10/2015	9112	frozen vegetables
19998	4344	01/03/2015	9113	long life bakery product
19999	3475	10/06/2015	9113	whole milk

20000 rows × 4 columns

Data Preparation



Member Number



Item Description

Hanya dibutuhkan 2 atribut yaitu Member number dan item Description dalam Tahap Data Preparation karena secara umum Association Rule (MBA) hanya berfokus pada data transaksi sehingga hanya melibatkan atribut no transaksi (Member number) dan item yang dibeli (item Description)

Modelling

Penggunaan Market Basket Analysis dalam melihat kecenderungan pelanggan terhadap pembelian produk secara bersamaan menggunakan algoritma Apriori



Dalam melakukan MBA, kami menggunakan bahasa pemrograman Python

Source code untuk membaca Dataset

```
[ ] df = pd.read_csv('/content/Groceries_dataset.csv')  
df
```

berikut adalah tampilan dari dataset yang kami gunakan



	Member_number	Date	Transaction	itemDescription
0	1808	21/07/2015	1	tropical fruit
1	2552	05/01/2015	2	whole milk
2	2300	19/09/2015	2	pip fruit
3	1187	12/12/2015	3	other vegetables
4	3037	01/02/2015	3	whole milk
...
19995	1721	15/10/2015	9111	whole milk
19996	2877	10/04/2015	9112	waffles
19997	3848	11/10/2015	9112	frozen vegetables
19998	4344	01/03/2015	9113	long life bakery product
19999	3475	10/06/2015	9113	whole milk

20000 rows × 4 columns

Visualisasi dan memahami data

Kita tahu bahwa dataset ini direkam dari 01/01/2014 hingga 30/12/2015. Sebelum kita masuk dalam pemodelan, kita harus mengeksplorasi dan memvisualisasikan penjualan dalam periode waktu ini. Barang apa yang paling banyak dibeli pelanggan

```
[ ] # Mencetak 20 data penjualan item terbanyak
    most_sold = df['itemDescription'].value_counts().head(20)

    print('Most Sold Items: \n')
    print(most_sold)
```

Most Sold Items:

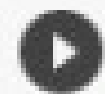
whole milk	1607
other vegetables	1186
sausage	924
tropical fruit	898
rolls/buns	843
root vegetables	749
citrus fruit	736
soda	672
yogurt	596
frankfurter	580
pip fruit	574
pork	531
beef	475
canned beer	456
chicken	406
bottled water	385
bottled beer	348
pastry	315
hamburger meat	293
ham	252

Name: itemDescription, dtype: int64

```
[ ] transaction_count = df.groupby(by='itemDescription')[['Transaction']].count().sort_values(by='Transaction', ascending=False)
def convert_to_percentage(x):
    return 100 * x / float(x.sum())

transaction_percentage = transaction_count.apply(convert_to_percentage)
transaction_percentage.head()
```

Transaction	
itemDescription	
whole milk	8.035
other vegetables	5.930
sausage	4.620
tropical fruit	4.490
rolls/buns	4.215

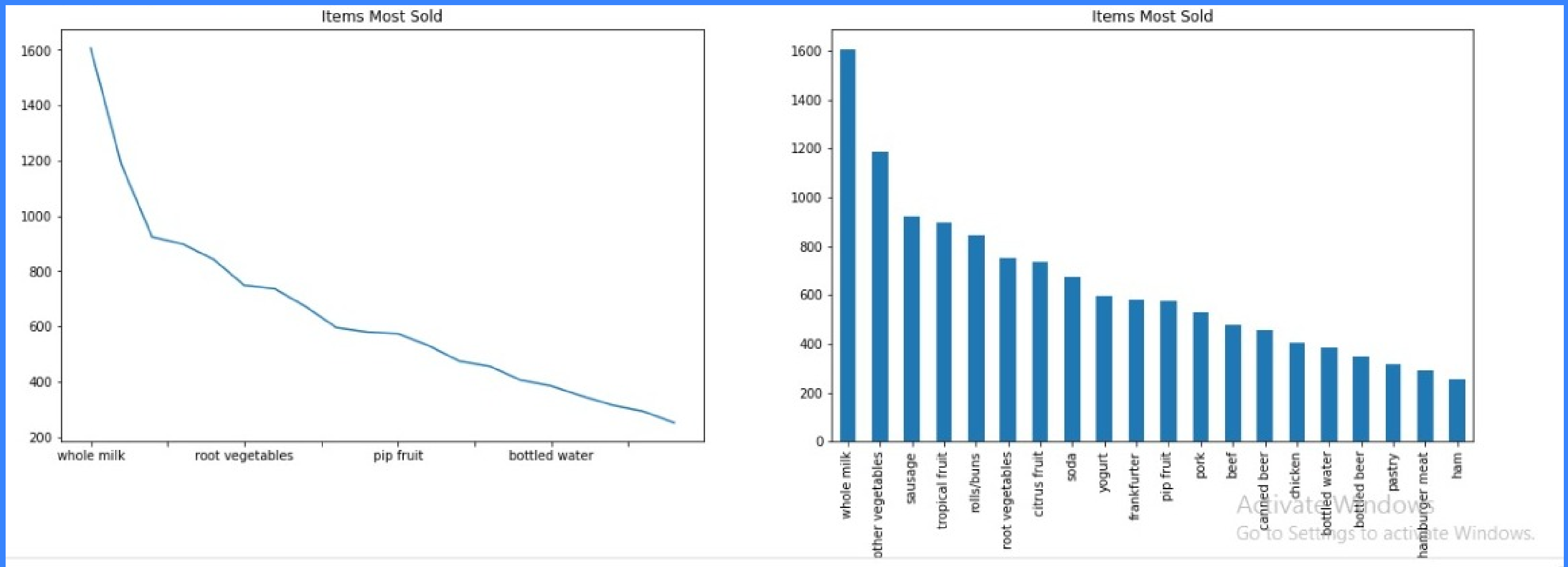


```
# visualisasikan data penjualan item terbanyak ke line chart dan bar chart  
plt.figure(figsize=(20,6))
```

```
plt.subplot(1,2,1)  
#plt.plot(most_sold)  
most_sold.plot(kind='line')  
plt.title('Items Most Sold')
```

```
plt.subplot(1,2,2)  
most_sold.plot(kind='bar')  
plt.title('Items Most Sold')
```


berikut adalah tampilan hasil tampilan line chart dan bar chart



Market Basket Analysis

Pertama kita mengubah item list menjadi data transaksi dengan menggunakan TransactionEncoder. Kita akan menggunakan frequent item-sets untuk menghasilkan aturan asosiasi dengan menggunakan algoritma Apriori.

```
[ ] from mlxtend.preprocessing import TransactionEncoder  
    from mlxtend.frequent_patterns import association_rules, apriori
```

kami membuat list transaction_list() yang unik sehingga kita dapat mengubah data kita ke dalam format yang benar

```
[ ] transaction_list = []

# For loop to create a list of the unique transactions throughout the dataset:
for i in df['Transaction'].unique():
    tlist = list(set(df[df['Transaction']==i]['itemDescription']))
    if len(tlist)>0:
        transaction_list.append(tlist)
print(len(transaction_list))
```

8976

kita menggunakan TransactionEncoder untuk merubah bentuk transaction_list()

```
[ ] te = TransactionEncoder()  
    te_ary = te.fit(transaction_list).transform(transaction_list)  
    df2 = pd.DataFrame(te_ary, columns=te.columns_)
```

sekarang kita terapkan Apriori. Kita akan menggunakan parameter `min_threshold` (nilai ambang batas yang ditentukan) dalam aturan asosiasi untuk metrik lift menjadi 1,0 karena jika kurang dari satu, maka kedua item tersebut kemungkinan tidak akan dibeli bersama. Kita akan mengurutkan nilai berdasarkan keyakinan untuk melihat kemungkinan suatu barang dibeli jika pendahulunya dibeli.

```
[ ] frequent_itemsets = apriori(df2, min_support=0.01, use_colnames=True)
rules = association_rules(frequent_itemsets, metric='lift', min_threshold=1.0)
metrix = pd.DataFrame(rules)
metrix[['antecedents', 'consequents', 'support', 'confidence', 'lift']].sort_values('lift', ascending=False)
```

	antecedents	consequents	support	confidence	lift
0	(whole milk)	(rolls/buns)	0.014483	0.088737	1.001893
1	(rolls/buns)	(whole milk)	0.014483	0.163522	1.001893

Conclusion

Berdasarkan pembuatan rekomendasi paket dari hasil analisis MBA, disimpulkan bahwa pembuatan rekomendasi pakatnya mendapatkan hasil yang sesuai dengan tujuan. Produk whole milk menduduki produk yang populer atau yang paling laris di groceries.



😊 THANK YOU



REFERENSI

<https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset>

