

```
!pip install mlxtend
```

```
import pandas as pd

cols = [
    "InvoiceNo",
    "Description",
    "Quantity",
    "CustomerID",
    "Country"
]

df = pd.read_excel(
    "Online Retail.xlsx",
    usecols=cols,
    dtype={
        "InvoiceNo": str,
        "Description": str,
        "Quantity": "int32",
        "CustomerID": "float32",
        "Country": str
    }
)

df.head()
```

	InvoiceNo	Description	Quantity	CustomerID	Country	
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	17850.0	United Kingdom	
1	536365	WHITE METAL LANTERN	6	17850.0	United Kingdom	
2	536365	CREAM CUPID HEARTS COAT HANGER	8	17850.0	United Kingdom	
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	17850.0	United Kingdom	
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	17850.0	United Kingdom	

```
# Remove rows without product description
df = df.dropna(subset=["Description"])

# Remove rows without customer ID
df = df.dropna(subset=["CustomerID"])

# Remove returns / negative quantities
df = df[df["Quantity"] > 0]

# Remove cancelled invoices (start with C)
df = df[~df["InvoiceNo"].str.startswith("C")]

df.shape
```

```
(397924, 5)
```

```
df = df[df["Country"] == "United Kingdom"]
```

```
top_products = (
    df["Description"]
    .value_counts()
    .head(200) # safe size
    .index
)

df = df[df["Description"].isin(top_products)]
```

```
basket = (
    df.groupby(["InvoiceNo", "Description"])["Quantity"]
    .sum()
    .unstack()
    .fillna(0)
)
```

```
basket = (basket > 0).astype("int8")
```

```
basket = basket.sample(
    n=min(5000, len(basket)),
```

```
from mlxtend.frequent_patterns import apriori

frequent_items = apriori(
    basket,
    min_support=0.01,
    use_colnames=True
)

frequent_items.head()
```

[illegible]

	support	itemsets
0	0.0224	(3 STRIPEY MICE FELTCRAFT)
1	0.0442	(6 RIBBONS RUSTIC CHARM)
2	0.0292	(60 CAKE CASES VINTAGE CHRISTMAS)
3	0.0392	(60 TEATIME FAIRY CAKE CASES)
4	0.0302	(72 SWEETHEART FAIRY CAKE CASES)

```

/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated
    return datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated
    return datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated
    return datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated
    return datetime.utcnow().replace(tzinfo=utc)

```

Next steps: [Generate code with frequent items](#) [New interactive sheet](#)

```
from mlxtend.frequent_patterns import association_rules

rules = association_rules(
    frequent_items,
    metric="confidence",
    min_threshold=0.3
)

rules.head()
```

```

/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)

```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction
0	(60 CAKE CASES VINTAGE CHRISTMAS)	(PAPER CHAIN KIT 50'S CHRISTMAS )	0.0292	0.0664	0.0118	0.404110	6.085988	1.0	0.009861	1.566731
1	(60 CAKE CASES VINTAGE CHRISTMAS)	(PAPER CHAIN KIT VINTAGE CHRISTMAS)	0.0292	0.0472	0.0120	0.410959	8.706756	1.0	0.010622	1.617544
2	(SET OF 20 VINTAGE CHRISTMAS NAPKINS)	(60 CAKE CASES VINTAGE CHRISTMAS)	0.0292	0.0292	0.0102	0.349315	11.962845	1.0	0.009347	1.491966
3	(60 CAKE CASES VINTAGE CHRISTMAS)	(SET OF 20 VINTAGE CHRISTMAS NAPKINS)	0.0292	0.0292	0.0102	0.349315	11.962845	1.0	0.009347	1.491966
4	(72 SWEETHEART FAIRY CAKE CASES)	(60 TEATIME FAIRY CAKE CASES)	0.0302	0.0392	0.0128	0.423841	10.812272	1.0	0.011616	1.667595

```

/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated,
return datetime.datetime.utcnow().replace(tzinfo=utc)

```

Next steps: [Generate code with rules](#) [New interactive sheet](#)

```

rules = rules.sort_values(
    by="lift",
    ascending=False
)

rules[["antecedents", "consequents", "support", "confidence", "lift"]].head(10)

```

```
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
return datetime.datetime.utcnow().replace(tzinfo=utc)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
return datetime.datetime.utcnow().replace(tzinfo=utc)
```

	antecedents	consequents	support	confidence	lift	
76	(FELTCRAFT PRINCESS LOLA DOLL)	(FELTCRAFT PRINCESS CHARLOTTE DOLL)	0.0162	0.692308	27.256208	
77	(FELTCRAFT PRINCESS CHARLOTTE DOLL)	(FELTCRAFT PRINCESS LOLA DOLL)	0.0162	0.637795	27.256208	
474	(JUMBO BAG APPLES, JUMBO BAG VINTAGE LEAF)	(JUMBO BAG PEARS)	0.0110	0.714286	25.693731	
477	(JUMBO BAG PEARS)	(JUMBO BAG APPLES, JUMBO BAG VINTAGE LEAF)	0.0110	0.395683	25.693731	
388	(WOODEN HEART CHRISTMAS SCANDINAVIAN)	(WOODEN STAR CHRISTMAS SCANDINAVIAN)	0.0214	0.713333	25.295508	

```
def recommend(product):
    results = rules[
        rules["antecedents"].apply(lambda x: product in x)
    ]

    return results.sort_values(
        by="lift",
        ascending=False
    )["consequents", "confidence", "lift"].head(5)
```

ROSES REGENC... REGENCY CAKES...

```
import warnings
warnings.filterwarnings("ignore")
```

```
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated, datetime.datetime.now() is preferred
recommend("WHITE HANGING HEART T-LIGHT HOLDER")
```

	consequents	confidence	lift	
647	(WOODEN PICTURE FRAME WHITE FINISH)	0.700000	11.784512	
646	(WOODEN FRAME ANTIQUE WHITE )	0.543103	10.286050	
453	(HEART OF WICKER LARGE)	0.548387	10.080645	
451	(HEART OF WICKER SMALL)	0.525773	8.911410	

Start coding or [generate](#) with AI.