DATA ANALYTICS - Practical Assingment-2

1. **Create the following dataset in python.**

TID	*Items*
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

Convert the categorical values into numeric format.
Apply the apriori algorithm on the above dataset to generate the frequent itemsets and association rules. Repeat the process with different min_sup values.

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules

# Create the dataset
dataset = [
    ['Bread', 'Milk'],
    ['Bread', 'Diaper', 'Beer', 'Eggs'],
    ['Milk', 'Diaper', 'Beer', 'Coke'],
    ['Bread', 'Milk', 'Diaper', 'Beer'],
    ['Bread', 'Milk', 'Diaper', 'Coke']
]

# Convert the dataset into a DataFrame with one-hot encoding
te = TransactionEncoder()
te_ary = te.fit(dataset).transform(dataset)
df = pd.DataFrame(te_ary, columns=te.columns_)
```

```
print("One-hot encoded DataFrame:")
print(df)

frequent_itemsets = apriori(df, min_support=0.6, use_colnames=True)
print("\nFrequent itemsets with min_support=0.6:")
print(frequent_itemsets)


rules = association_rules(frequent_itemsets, metric="confidence",
min_threshold=0.7, num_itemsets = len (dataset))
print("\nAssociation rules with min_confidence=0.7:")
print(rules)
```

Create your own transactions dataset and apply the above process on your dataset.

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules


transactions = [
    ['Apple', 'Banana', 'Milk'],
    ['Apple', 'Diaper', 'Beer', 'Eggs'],
    ['Milk', 'Diaper', 'Beer', 'Coke'],
    ['Apple', 'Milk', 'Diaper', 'Beer'],
    ['Apple', 'Milk', 'Diaper', 'Coke']
]


te = TransactionEncoder()
te_ary = te.fit(transactions).transform(transactions)
df = pd.DataFrame(te_ary, columns=te.columns_)
```

```
print("One-hot encoded DataFrame:")
print(df)


frequent_itemsets = apriori(df, min_support=0.6, use_colnames=True)
print("\nFrequent itemsets with min_support=0.6:")
print(frequent_itemsets)


rules = association_rules(frequent_itemsets, metric="confidence",
min_threshold=0.7, num_itemsets = len (transactions))
print("\nAssociation rules with min_confidence=0.7:")
print(rules)
```