PART I: RESEARCH QUESTION

PROPOSAL OF QUESTION

Which medications are frequently prescribed together?

DEFINED GOAL

The primary goal is to identify patterns of medication co-prescription to optimize treatment regimens and improve patient outcomes.

PART II: MARKET BASKET JUSTIFICATION

EXPLANATION OF MARKET BASKET

Market Basket Analysis (MBA) is a data analysis technique to understand the relationship between items in a dataset. When applied to the medicine prescription data, MBA helps identify patterns in which medications are frequently prescribed together.

MBA works by first having data where each row represents a transaction, and each column represents a specific medication. The values are typically True if the medication was prescribed and False if it was not. Then, MBA identifies combinations of medications that frequently appear together. From these medications, MBA generates rules that describe the likelihood of one medication being prescribed, given that another medication has been prescribed.

The expected outcomes are identifying the commonly co-prescribed medications and optimizing inventory management. Select medications with a high lift value to identify common co-prescribed medications. Moreover, pharmacies can use this analysis to manage their inventory better to ensure they stock enough of the medications prescribed together.

TRANSACTION EXAMPLE

Presc01	Presc02	Presc03	Presc04	Presc05	Presc06	Presc07	Prec08
abilify	atorvastatin	folic acid	naproxen	losartan			

Presc 09	Presc 10	Presc 11	Presc 12	Presc 13	Presc 14	Presc 15	Presc 16	Presc 17	Presc 18	Presc 19	Presc 20

Presc06 to Presc20 are empty in the original dataset.

MARKET BASKET ASSUMPTION

MBA assumes that each transaction is independent of other transactions. This means that the presence or absence of items in one transaction does not affect the presence or absence of items in another transaction. If the assumption of independence is disregarded, the association rules generated might not accurately reflect true prescribing patterns and could lead to biased conclusions.

PART III: DATA PREPARATION & ANALYSIS

TRANSFORMING THE DATA SET

1. Initialize all libraries and packages to be used

```
# Data Handling
import csv
import pandas as pd

# Ignore Warnings
import warnings
warnings.filterwarnings("ignore")

# Visualizations
import matplotlib.pyplot as plt

# Machine Learning
from mlxtend.frequent_patterns import apriori, association_rules
```

2. Load the dataset into Pandas dataframe

```
# Specifiy CSV file path
file_path = r'C:\Users\kolgi\OneDrive - Western Governors University\D212\medical_market_basket.csv'

# Open the CSV file and read it using DictReader
with open(file_path, 'r') as csvfile:
    csvreader = csv.DictReader(csvfile)

# Read the CSV file into a pandas DataFrame then open head
df = pd.read_csv(file_path)
```

3. View the first 5 rows of the DataFrame
   Df.head()
4. View the index, column names, non-null count and data types
   Df.info()
5. Check for null/missing values in the dataset, then count the null values for each column
   Df.isnull().sum()
6. Check for duplicates in the data
   Df.duplicated()
7. Keep only the rows where the Presc01 column has non-null values, then display the shape.

```
# Retain only the rows where the values in the column named 'Presc01' are not null (or not NaN).
df = df[df['Presc01'].notna()]
df.shape
```

```
(7501, 20)
```

8. Initialize an empty list and then iterate over each row from index 0 to 7500. For each row, create a list of string representations of the values in the first 20 columns.

```
# Initialize an empty list to store lists (From Dr. Kamara, n.d.)
rows = []
# Iterate over rows
for i in range(0, 7501):
    rows.append([str(df.values[i,j])
for j in range(0, 20)])
```

9. Use TransactionEncoder() to convert a list of lists representing transactions into a Boolean array

```
# Initialize TransactionEncoder
encoder = TransactionEncoder()

# Fit the encoder to the list of lists and transform the data into a boolean array
array = encoder.fit_transform(rows)

# Convert the boolean array into a DataFrame for better visualization
df_encoded = pd.DataFrame(array, columns=encoder.columns_)

# Display the boolean array
df_encoded
```

10. Print all column names from df_encoded

```
# View all column names (from Dr. Kamara, n.d.)
for col in df_encoded.columns:
    print(col)
```

11. Remove any column named nan from df_encoded then print the shape

```
# Drop the 'nan' column from df_encoded
df_encoded = df_encoded.drop(columns=['nan'], errors='ignore')

# Print the DataFrame after dropping empty columns
df_encoded.shape
```

```
(7501, 119)
```

12. Save the filtered DataFrame to a CSV file

```
# Save the filtered DataFrame to a CSV file
df_encoded.to_csv('task3_transformed.csv', index=False)
```

Cleaned dataset attached as *task3_transformed.csv*

CODE EXECUTION

1. Use the extracted and cleaned csv

```
# Specifiy CSV file path
file_path = r'C:\Users\kolgi\OneDrive - Western Governors University\D212\task 3\task3_transformed.csv'

# Open the CSV file and read it using DictReader
with open(file_path, 'r') as csvfile:
    csvreader = csv.DictReader(csvfile)

# Read the CSV file into a pandas DataFrame then open head
df = pd.read_csv(file_path)
```

2. Use the Apriori algorithm with a minimum support of 0.02

```
# Create frequent itemsets using Apriori algorithm
rules = apriori(df, min_support=0.02, use_colnames=True)
rules
```

	support	itemsets
0	0.046794	(Premarin)
1	0.238368	(abilify)
2	0.020397	(albuterol aerosol)
3	0.033329	(allopurinol)
4	0.079323	(alprazolam)

ASSOCIATION RULES TABLE

```python
# Generate association rules table
rules_table = association_rules(rules, metric='lift', min_threshold=1)
rules_table.head(10)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(amlodipine)	(abilify)	0.071457	0.238368	0.023597	0.330224	1.385352	0.006564	1.137144	0.299568
1	(abilify)	(amlodipine)	0.238368	0.071457	0.023597	0.098993	1.385352	0.006564	1.030562	0.365218
2	(amphetamine salt combo)	(abilify)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144
3	(abilify)	(amphetamine salt combo)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
4	(amphetamine salt combo xr)	(abilify)	0.179709	0.238368	0.050927	0.283383	1.188845	0.008090	1.062815	0.193648
5	(abilify)	(amphetamine salt combo xr)	0.238368	0.179709	0.050927	0.213647	1.188845	0.008090	1.043158	0.208562
6	(atorvastatin)	(abilify)	0.129583	0.238368	0.047994	0.370370	1.553774	0.017105	1.209650	0.409465
7	(abilify)	(atorvastatin)	0.238368	0.129583	0.047994	0.201342	1.553774	0.017105	1.089850	0.467950
8	(carvedilol)	(abilify)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314	0.369437
9	(abilify)	(carvedilol)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606

TOP THREE RULES

Top 3 rules for support

```python
# Sort the association rules by support
sorted_support = rules_table.sort_values(by='support', ascending=False)
sorted_support.head(3)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
8	(carvedilol)	(abilify)	0.174110	0.238368	0.059725	0.343032	1.439085	0.018223	1.159314	0.369437
9	(abilify)	(carvedilol)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606
19	(diazepam)	(abilify)	0.163845	0.238368	0.052660	0.321400	1.348332	0.013604	1.122357	0.308965

Rule 1 indicates that 5.97% of all transactions contain carvedilol and abilify. Rule 2 is similar to Rule 1 but in reverse, showing that 5.97% of all transactions contain both abilify and carvedilol. Rule 3 indicates that 5.27% of all transactions contain diazepam and abilify.

Top 3 rules for confidence

```python
# Sort the association rules by confidence
sorted_confidence = rules_table.sort_values(by='confidence', ascending=False)
sorted_confidence.head(3)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
31	(metformin)	(abilify)	0.050527	0.238368	0.023064	0.456464	1.914955	0.011020	1.401255	0.503221
25	(glipizide)	(abilify)	0.065858	0.238368	0.027596	0.419028	1.757904	0.011898	1.310962	0.461536
29	(lisinopril)	(abilify)	0.098254	0.238368	0.040928	0.416554	1.747522	0.017507	1.305401	0.474369

Rule 1 indicates that 45.65% of transactions containing metformin also contain abilify. Rule 2 indicates that 41.9% of transactions containing glipizide also have abilify. Rule 3 indicates that 41.7% of transactions containing lisinopril also contain abilify.

Top 3 rules for lift

```
# Sort the association rules by lift
sorted_lift = rules_table.sort_values(by='lift', ascending=False)
sorted_lift.head(3)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
74	(carvedilol)	(lisinopril)	0.174110	0.098254	0.039195	0.225115	2.291162	0.022088	1.163716	0.682343
75	(lisinopril)	(carvedilol)	0.098254	0.174110	0.039195	0.398915	2.291162	0.022088	1.373997	0.624943
73	(glipizide)	(carvedilol)	0.065858	0.174110	0.022930	0.348178	1.999758	0.011464	1.267048	0.535186

Rule 1 suggests that patients prescribed carvedilol are about 2.29 times more likely to be prescribed lisinopril. Rule 2 is the same as Rule 1, but in reverse, patients prescribed lisinopril are about 2.29 times more likely to be prescribed carvedilol. Rule 3 indicates that patients prescribed with glipizide are about 1.99 times more likely also to be prescribed with carvedilol.

PART IV: DATA SUMMARY & IMPLICATIONS

SIGNIFICANCE OF SUPPORT, LIFT, AND CONFIDENCE SUMMARY

Support measures how frequently an itemset appears in the dataset. The analysis for top support highlights that carvedilol and abilify are prescribed together in approximately 5.97% of all transactions in the dataset. The high support value indicates that the combination of these two medications is relatively common in the dataset.

Confidence measures the likelihood that a consequent item is prescribed when the antecedent item is prescribed. High confidence means that when a patient is prescribed one medication, there is a high probability that they will also be prescribed another specific medication. The analysis for top confidence indicates that 45.65% of the time, when metformin is prescribed, abilify is also prescribed. This high confidence value suggests a strong likelihood that patients on metformin are also being prescribed abilify.

Lift measures the strength of an association between two medications beyond their individual occurrence rates. A lift greater than 1 indicates a positive association. The lift value of 2.29 for the top lift itemset means that the prescription of lisinopril is twice as likely when carvedilol is prescribed compared to the prescription of lisinopril independently. This high lift value indicates a strong positive association that is not by chance alone.

PRACTICAL SIGNIFICANCE OF FINDINGS

The findings from the analysis provide valuable insights that can help optimize inventory management, identify potential drug interactions, and cost management and efficiency.

Knowing which medications are most frequently prescribed together allows pharmacies to optimize their inventory. Medications with high support values should be stocked in higher quantities to meet patient demand. Moreover, lift values greater than 1 highlight strong associations between medications that are prescribed together more frequently than by chance. This can alert healthcare providers to potential drug interactions. Finally, by understanding the most common medication combinations, healthcare providers can reduce waste by ordering the right quantities of these drugs and avoiding overstocking less frequently prescribed medications.

COURSE OF ACTION

Integrate medication management for high-support itemsets. Since carvedilol and abilify are prescribed together in approximately 5.97% of transactions, healthcare providers should ensure coordinated care for patients receiving these medications. This involves monitoring for potential interactions and providing comprehensive patient education on the effects and proper use of these medications.

Enhance patient monitoring for high-confidence associations. A confidence value of 45.65% indicates that nearly half of the patients on metformin are also prescribed abilify. Providers should be aware of the reasons behind this co-prescription and manage any possible side effects.

The strong association of 2.29 between carvedilol and lisinopril suggests that pharmacies should ensure both medications are available together. Pharmacies can utilize stock management and inventory optimization with high-lift associations. This can help in reducing wait times for prescriptions and improving patient satisfaction.