A1) Research Question

What medication is the most correlated, or purchased commonly, with the medication abilify?

A2) Objectives

The goal of this analysis is to utilize market basket analysis to determine what medications are most correlated with the medication abilfy. Creating a model that can determine medications that are commonly purchased together provides stakeholders with the ability to predict what medications patients may need and conduct further research for the cause of these correlations.

B1) Market Basket Analysis

Market basket analysis is a data analysis method utilized to determine patterns and relationships and is commonly used for transactional data. This technique analyzes data by calculating the support (frequency of items purchased together), confidence (strength of association), and lift (probability of items purchased together versus not) (Blattberg et al., 2008). Data is often filtered by association rules to make business decisions and interpretations. The expected outcome of this analysis is to identify the medication(s) that are most correlated with the medication Abilify based on the metrics previously mentioned in this section.

B2) Example Transaction

Figure 1

Example prescription purchase

```
In [36]: df.iloc[0]

Out[36]: Presc01                   amlodipine
         Presc02            albuterol aerosol
         Presc03                  allopurinol
         Presc04                 pantoprazole
         Presc05                    lorazepam
         Presc06                   omeprazole
         Presc07                   mometasone
         Presc08                  fluconozole
         Presc09                   gabapentin
         Presc10                  pravastatin
         Presc11                       cialis
         Presc12                     losartan
         Presc13     metoprolol succinate XL
         Presc14             sulfamethoxazole
         Presc15                      abilify
         Presc16               spironolactone
         Presc17                albuterol HFA
         Presc18                 levofloxacin
         Presc19                 promethazine
         Presc20                    glipizide
         Name: 0, dtype: object
```

B3) Assumption

An assumption of market basket analysis is the linearity of variables. Market basket analysis is based on the relationships that items have between each other. This technique works optimally when data correlates with each other, or baskets.

C1) Data Preparation

Before the analysis can be performed, our data needs to be cleaned and transformed as needed. Cleaning was conducted by detecting for missing values and duplicates. The data set provided also included many missing rows, these were removed. Missing values were also replaced with "false" and medications that were present resulted in "true." To transform the data, I utilized the package transaction encoder to transform the data set into a logical data frame with each column representing an item and each row representing a transaction.

A copy of the cleaned data set is attached.

Figure 2

Transaction Encoder to Transform Data

In [23]:
```
trans = []
for i in range (0, 7501):
    trans.append([str(df.values[i,j]) for j in range(0, 20)])
```

In [24]:
```
TE = TransactionEncoder()
array = TE.fit(trans).transform(trans)
```

In [25]:
```
cleaned_df = pd.DataFrame(array, columns = TE.columns_)
cleaned_df
```

Out[25]:

	Duloxetine	Premarin	Yaz	abilify	acetaminophen	actonel	albuterol HFA	albuterol aerosol	alendronate	allopurinol	...	trazodone HCl	triamcinolone Ace topical	triamterene	trime
0	False	False	False	True	False	False	True	True	False	True	...	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
3	False	False	False	False	False	False	False	False	False	True	...	False	False	False	
4	False	False	False	True	False	False	False	False	False	False	...	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
7496	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
7497	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
7498	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
7499	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
7500	False	False	False	False	False	False	False	False	False	False	...	False	False	False	

7501 rows × 120 columns

C2) Association Rules

Association rules are utilized to predict the probability of features being correlated to each other. To generate the association rules for our analysis, I utilized the apriori algorithm to calculate the frequencies of the items being purchased. I then utilized the association rules package to filter and prune for certain parameters.

Figure 3

Apriori Algorithm

In [49]:
```
a_rules = apriori(df_cleaned, min_support = 0.02, use_colnames = True)
a_rules.head()
```
Out[49]:

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

C3) Association Rules Table

Figure 4

Association Rules Table

In [50]:
```
ass_r = association_rules(a_rules, metric = 'lift', min_threshold = 1)
ass_r
```
Out[50]:

	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	(abilify)	(amphetamine salt combo)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
3	(amphetamine salt combo)	(abilify)	0.068391	0.238368	0.024397	0.356725	1.496530	0.008095	1.183991	0.356144
4	(amphetamine salt combo xr)	(abilify)	0.179709	0.238368	0.050927	0.283383	1.188845	0.008090	1.062815	0.193648
...	...	...	...	...	...	...	...	...	...	...
89	(diazepam)	(metoprolol)	0.163845	0.095321	0.022930	0.139951	1.468215	0.007312	1.051893	0.381390
90	(glyburide)	(doxycycline hyclate)	0.170911	0.095054	0.020131	0.117785	1.239135	0.003885	1.025766	0.232768
91	(doxycycline hyclate)	(glyburide)	0.095054	0.170911	0.020131	0.211781	1.239135	0.003885	1.051852	0.213256
92	(losartan)	(glyburide)	0.132116	0.170911	0.028530	0.215943	1.263488	0.005950	1.057436	0.240286
93	(glyburide)	(losartan)	0.170911	0.132116	0.028530	0.166927	1.263488	0.005950	1.041786	0.251529

94 rows × 10 columns

C4) Top Three Rules

Lift indicates the measure of the likelihood of the consequent to occur when the antecedent is present. The support value is the measure of the frequency an itemset occurs within the dataset. The confidence is the measure of likelihood an itemset will occur if another itemset is present. By filtering the association rules with a lift of at least 1.9, support level of at least 0.02, and a confidence level of at least 0.34, we can identify the top three rules of our association table.

Figure 5

Top Three Rules

```
In [66]: ass_r[ (ass_r['lift'] >= 1.9) &
             (ass_r['confidence'] >= 0.34) &
              (ass_r['support'] >=0.02)]
Out[66]:
```

	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
73	(glipizide)	(carvedilol)	0.065858	0.174110	0.022930	0.348178	1.999758	0.011464	1.267048	0.535186
75	(lisinopril)	(carvedilol)	0.098254	0.174110	0.039195	0.398915	2.291162	0.022088	1.373997	0.624943

D1) Results

To answer our research question of determining the most correlated medications with Abilify, we need to further filter our analysis. To accomplish this, I filtered for the medication Abilify as either the antecedent or consequent and sorted by lift. I chose lift as the main sorting measurement as values of lift greater than one have an increased correlation. Evaluating the medication with the highest lift value in relationship to Abilify, we can see that the medication is metformin. The lift of this itemset is 1.91, indicating a high correlation that a person will purchase metformin in addition to abilify. The support of this itemset is 0.023, indicating the itemset frequency within this dataset is 2.3 percent. The confidence of this itemset is the proportion of prescriptions that include the itemset divided by the proportion of just the antecedent. Analyzing abilify as the antecedent and the consequent, the support is higher at 0.46, or 46 percent.

Figure 6

Isolating Abilify for Association Rules (Ali, 2023)

In [63]:
```
ant_df = ass_r[ass_r['antecedents'] == {'abilify'}].sort_values(by=['lift'], ascending = False)
con_df = ass_r[ass_r['consequents'] == {'abilify'}].sort_values(by=['lift'], ascending = False)
ant_df
```

Out[63]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
30	(abilify)	(metformin)	0.238368	0.050527	0.023064	0.096756	1.914955	0.011020	1.051182	0.627330
24	(abilify)	(glipizide)	0.238368	0.065858	0.027596	0.115772	1.757904	0.011898	1.056449	0.566075
28	(abilify)	(lisinopril)	0.238368	0.098254	0.040928	0.171700	1.747522	0.017507	1.088672	0.561638
23	(abilify)	(fenofibrate)	0.238368	0.051060	0.020131	0.084452	1.653978	0.007960	1.036472	0.519145
15	(abilify)	(clopidogrel)	0.238368	0.059992	0.022797	0.095638	1.594172	0.008497	1.039415	0.489364
33	(abilify)	(metoprolol)	0.238368	0.095321	0.035729	0.149888	1.572463	0.013007	1.064189	0.477993
7	(abilify)	(atorvastatin)	0.238368	0.129583	0.047994	0.201342	1.553774	0.017105	1.089850	0.467950
2	(abilify)	(amphetamine salt combo)	0.238368	0.068391	0.024397	0.102349	1.496530	0.008095	1.037830	0.435627
20	(abilify)	(doxycycline hyclate)	0.238368	0.095054	0.033729	0.141499	1.488616	0.011071	1.054100	0.430963
34	(abilify)	(naproxen)	0.238368	0.058526	0.020131	0.084452	1.442993	0.006180	1.028318	0.403076
8	(abilify)	(carvedilol)	0.238368	0.174110	0.059725	0.250559	1.439085	0.018223	1.102008	0.400606
17	(abilify)	(dextroamphetamine XR)	0.238368	0.081056	0.027463	0.115213	1.421397	0.008142	1.038604	0.389252
1	(abilify)	(amlodipine)	0.238368	0.071457	0.023597	0.098993	1.385352	0.006564	1.030562	0.365218
18	(abilify)	(diazepam)	0.238368	0.163845	0.052660	0.220917	1.348332	0.013604	1.073256	0.339197
27	(abilify)	(levofloxacin)	0.238368	0.063325	0.020264	0.085011	1.342461	0.005169	1.023701	0.334938
11	(abilify)	(cialis)	0.238368	0.076523	0.023997	0.100671	1.315565	0.005756	1.026851	0.314943
5	(abilify)	(amphetamine salt combo xr)	0.238368	0.179709	0.050927	0.213647	1.188845	0.008090	1.043158	0.208562
13	(abilify)	(citalopram)	0.238368	0.087188	0.024397	0.102349	1.173883	0.003614	1.016889	0.194486

D2) Practical Significance

The practical significance of our analysis results is the highly correlated relationship between metformin and abilify. The significance of our results can lead stakeholders to further research possible causes for this relationship by determining the indications and pharmacology of these medications. For patients that utilize anti-psychotic medications such as abilify, weight gain may increase due to hyperglycemia. Metformin may be prescribed in conjunction to prevent the weight-inducing effects of abilify (Hakami et al., 2022).

D3) Recommendations

The results of our research question are that customers within this dataset that purchase abilify are likely to also purchase metformin. I would recommend stakeholders to further research the average costs and necessity for these medications to be prescribed together. I would recommend stakeholders to interview prescribing providers for their indication for these medications to be taken together. If the reasoning is to prevent weight gain from abilify, are there alternative methods of weight gain prevention that can be implemented for cost savings to the patient.

Third-Party Code References

Ali, M. (2023, January 23). *Association rule mining in Python tutorial*. DataCamp. https://www.datacamp.com/tutorial/association-rule-mining-python

References

Blattberg, R.C., Kim, BD., Neslin, S.A. (2008). Market Basket Analysis. In: Database Marketing. International Series in Quantitative Marketing, vol 18. Springer, New York, NY. https://doi.org/10.1007/978-0-387-72579-6_13

Hakami, A. Y., Felemban, R., Ahmad, R. G., Al-Samadani, A. H., Salamatullah, H. K., Baljoon, J. M., Alghamdi, L. J., Ramadani Sindi, M. H., & Ahmed, M. E. (2022). The Association Between Antipsychotics and Weight Gain and the Potential Role of Metformin Concomitant Use: A Retrospective Cohort Study. *Frontiers in psychiatry*, *13*, 914165. https://doi.org/10.3389/fpsyt.2022.914165