



## Apriori Algorithm using Python

### Market Basket Analysis using Apriori Algorithm in Machine Learning



AMAN KHARWAL / ⏰ NOVEMBER 16, 2020 / 📁 MACHINE LEARNING / 2

In Machine Learning, the Apriori algorithm is used for data mining association rules. In this article, I will take you through Market Basket Analysis using the Apriori algorithm in Machine Learning by using the Python programming language.

PRICE DROP	PRICE DROP	PRICE DROP	Buy E Com Quartz
------------	------------	------------	------------------

## What is Association Mining?

Association mining is typically performed on transaction data from a retail marketplace or online e-commerce store. Since most transaction data is large, the a priori algorithm makes it easy to find these patterns or rules quickly.

[Also, Read – 100+ Machine Learning Projects solved and explained.](#)

Association rules are used to analyze retail or transactional data and are intended to identify strong rules mainly found in transactional data using measures of interest, based on the concept of strong principals.

# Study in US without GRE

iSchoolConnect

## How does the Apriori Algorithm Work?

The Apriori algorithm is the most popular algorithm for mining association rules. It finds the most frequent combinations in a database and identifies the rules of association between elements, based on 3 important factors:

1. Support: the probability that X and Y meet
2. Confidence: the conditional probability that Y knows x. In other words, how often does Y occur when X came first.
3. Lift: the relationship between support and confidence. An increase of 2 means that the probability of buying X and Y together is twice as high as the probability of simply buying Y.

Apriori uses a “bottom-up” approach, in which frequent subsets are extended one item at a time (one step is called candidate generation) and groups of candidates are tested against the data. The algorithm ends when no other successful extension is found.

Now, I will take you through the task of Market Basket analysis using the Apriori Algorithm using Python and Machine Learning.

PRICE DROP

PRICE DROP

PRICE DROP

Buy E  
Com  
Quartz

## Market Basket Analysis with Apriori Algorithm using Python

Market basket analysis, also known as association rule learning or affinity analysis, is a data mining technique that can be used in various fields, such as marketing, bioinformatics, the field of marketing, education, nuclear science, etc.

The main goal of market basket analysis in marketing is to provide the retailer with the information necessary to understand the buyer's purchasing behaviour, which can help the retailer make incorrect decisions.

There are different algorithms for performing market basket analysis. Existing algorithms operate on static data and do not capture data changes over time. But the Apriori algorithm not only leverages static data but also provides a new way to account for changes that occur in the data.

**Study in US  
without GRE**

iSchoolConnect

I will start this task of Market Basket Analysis with Apriori Algorithm by importing the necessary Python libraries:

```

1 import numpy as np # linear algebra
2 import pandas as pd # data processing
3 import plotly.express as px
4 import apyori
5 from apyori import apriori

```

apriori.py hosted with ❤ by GitHub

[view raw](#)

Now let's load the dataset. The dataset that I am using in this task can be downloaded from [here](#):

```

1 data = pd.read_csv("Groceries_dataset.csv")
2 data.head()

```

	Member_number	Date	itemDescription
0	1808	21-07-2015	tropical fruit
1	2552	05-01-2015	whole milk
2	2300	19-09-2015	pip fruit
3	1187	12-12-2015	other vegetables
4	3037	01-02-2015	whole milk

## Data Exploration

Let's first have a look at the top 10 most selling products:

```

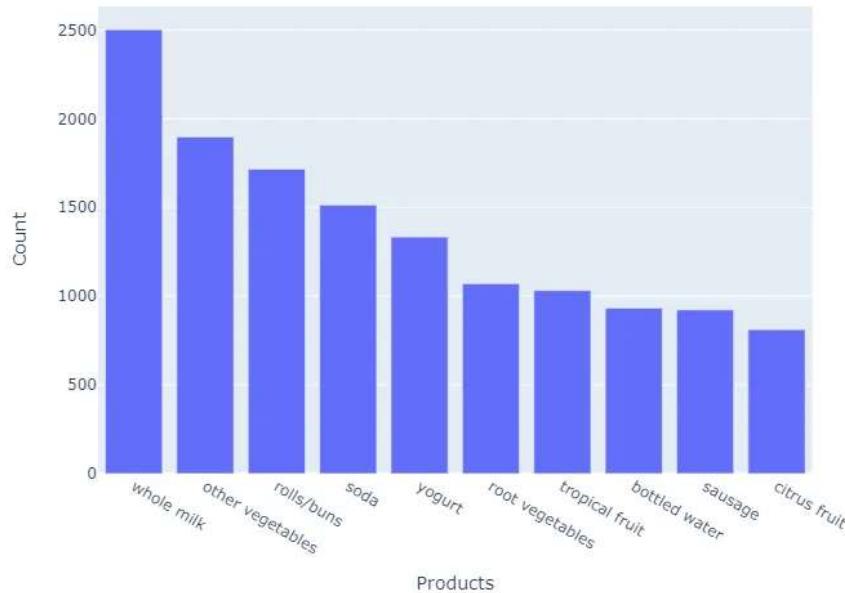
1 print("Top 10 frequently sold products(Tabular Representation)")
2 x = data['itemDescription'].value_counts().sort_values(ascending=False)[:10]
3 fig = px.bar(x=x.index, y=x.values)
4 fig.update_layout(title_text= "Top 10 frequently sold products (Graphical Representation)")
5 fig.show()

```

apriori.py hosted with ❤ by GitHub

[view raw](#)

## Top 10 frequently sold products (Graphical Representation)



Now let's explore the higher sales:

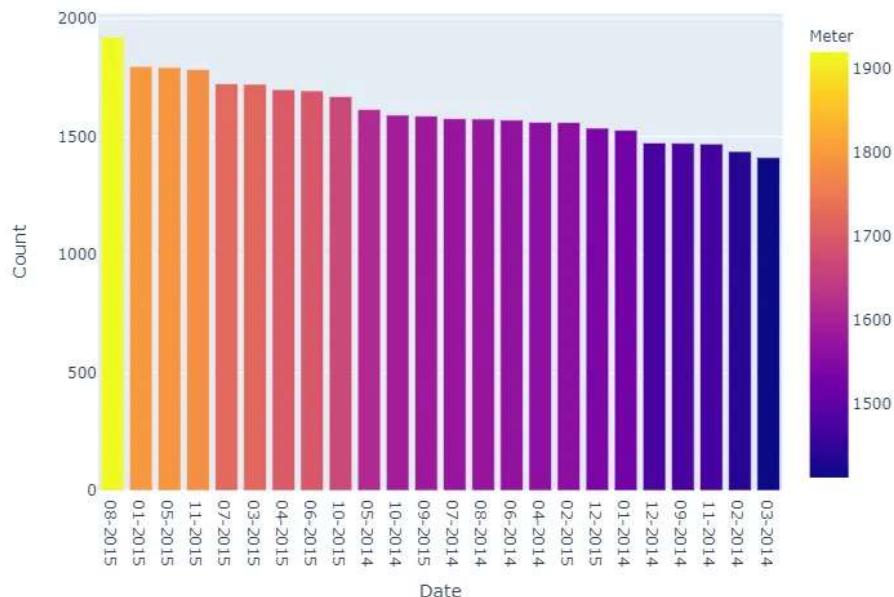
```

1  data["Year"] = data['Date'].str.split("-").str[-1]
2  data["Month-Year"] = data['Date'].str.split("-").str[1] + "-" + data['Date'].str.split("-").str[0]
3  fig1 = px.bar(data["Month-Year"].value_counts(ascending=False),
4                  orientation= "v",
5                  color = data["Month-Year"].value_counts(ascending=False),
6                  labels={'value':'Count', 'index':'Date','color':'Meter'})
7
8  fig1.update_layout(title_text="Exploring higher sales by the date")
9
10 fig1.show()

```

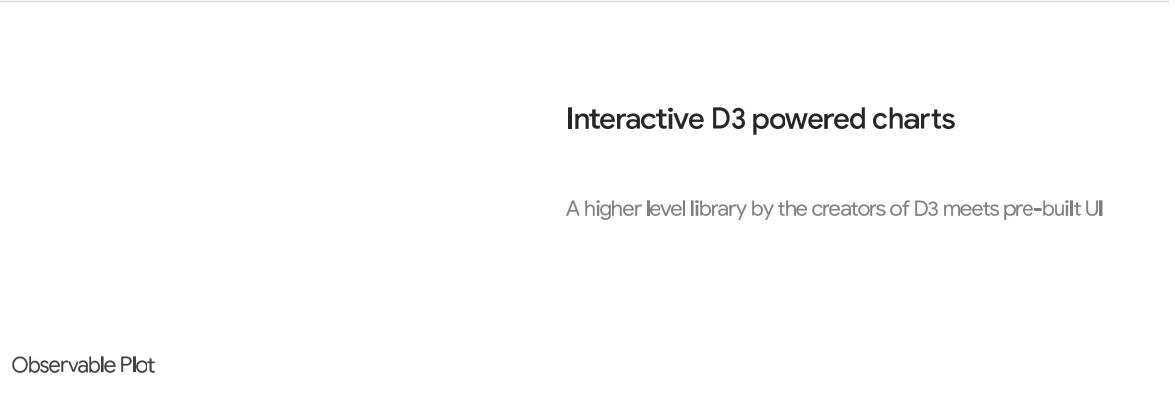
apriori.py hosted with ❤ by GitHub [view raw](#)

## Exploring higher sales by the date



## Observations:

From the above visualizations we can observe that:



1. Milk is bought the most, followed by vegetables.
2. Most shopping takes place in August / September, while February / March is the least demanding.

## Implementation of Apriori Algorithm using Python

Now, I will implement the Apriori algorithm in machine learning by using the Python programming language for the tasks of market basket analysis:

```
1 rules = apriori(transactions, min_support = 0.00030, min_confidence = 0.05, min_lift = 3,
2 association_results = list(rules)
3 print(association_results[0])
```

[apriori.py hosted with ❤ by GitHub](#) [view raw](#)

```
RelationRecord(items=frozenset({'liver loaf', 'fruit/vegetable
juice'}), support=0.00040098910646260775, ordered_statistics=
[OrderedStatistic(items_base=frozenset({'liver loaf'}),
items_add=frozenset({'fruit/vegetable juice'}), confidence=0.12,
lift=3.5276227897838903)])
```

```
1 for item in association_results:
2
3     pair = item[0]
4     items = [x for x in pair]
5
6     print("Rule : ", items[0], " -> " + items[1])
7     print("Support : ", str(item[1]))
8     print("Confidence : ", str(item[2][0][2]))
9     print("Lift : ", str(item[2][0][3]))
10
11    print("=====")
```

[apriori.py hosted with ❤ by GitHub](#) [view raw](#)

```
Rule : liver loaf -> fruit/vegetable juice
```

```
Support : 0.00040098910646260775
```

```
Confidence : 0.12
```

```
Lift : 3.5276227897838903
```

```
=====
```

```
Rule : ham -> pickled vegetables
```

```
Support : 0.0005346521419501437
```

```
Confidence : 0.05970149253731344
```

```
Lift : 3.4895055970149254
```

```
=====
```

```
Rule : roll products -> meat
```

```
Support : 0.0003341575887188398
```

```
Confidence : 0.06097560975609757
```

```
Lift : 3.620547812620984
```

```
=====
```

```
Rule : misc. beverages -> salt
```

```
Support : 0.0003341575887188398
```

```
Confidence : 0.05617977528089888
```

```
Lift : 3.5619405827461437
```

```
=====
```

```
Rule : spread cheese -> misc. beverages
```

```
Support : 0.0003341575887188398
```

```
Confidence : 0.05
```

```
Lift : 3.170127118644068
```

```
=====
```

```
Rule : soups -> seasonal products
```

```
Support : 0.0003341575887188398
```

```
Confidence : 0.10416666666666667
```

```
Lift : 14.704205974842768
```

```
=====
```

```
Rule : spread cheese -> sugar
```

```
Support : 0.00040098910646260775
```

```
Confidence : 0.06
```

```
Lift : 3.3878490566037733
```

```
=====
```

I hope you liked this article on the Apriori algorithm in Machine Learning by using the Python programming language. Feel free to ask your valuable questions in the comments section below.



Aman Kharwal

Coder with the ❤ of a Writer || Data Scientist | Solopreneur |  
Founder

---

ARTICLES: 1202



PREVIOUS



NEXT

## Recommended For You

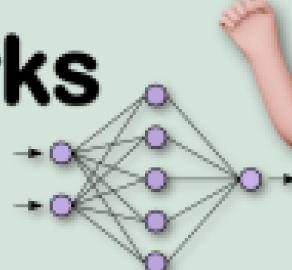
# The Best Approaches for Time Series Analysis



[Best Approaches for Time Series Analysis](#)

January 8, 2022

# Understand How Neural Networks Work



[How Neural Network Works](#)

January 6, 2022



### Visualize a Machine Learning Algorithm using Python

December 29, 2021



### Activation Functions in Neural Networks

December 23, 2021

## 2 Comments



Fabio Bianchi

DEC 26, 2020 / 4:17 PM

REPLY

I Aman. Sorry for the question. The code return this error:

```
rules = apriori(transactions, min_support = 0.00030,  
min_confidence = 0.05, min_lift = 3, max_length = 2, target =  
"rules")
```

NameError: name 'transactions' is not defined

The code:

```
import numpy as np
import pandas as pd
import plotly.express as px
import apyori
from apyori import apriori

rules = apriori(transactions, min_support = 0.00030,
min_confidence = 0.05, min_lift = 3, max_length = 2, target =
"rules")
association_results = list(rules)
print(association_results[0])

RelationRecord(items=frozenset({'liver loaf', 'fruit/vegetable
juice'}), \
support=0.00040098910646260775, \
ordered_statistics=
[OrderedStatistic(items_base=frozenset({'liver loaf'}),\
items_add=frozenset({'fruit/vegetable juice'}), \
confidence=0.12, lift=3.5276227897838903)])
```



**Aman Kharwal**

DEC 26, 2020 / 5:28 PM

REPLY

Complete code can be found here:

[https://github.com/amankharwal/Website-data/blob/master/association\\_rule\\_market\\_basket\\_analysis.ipynb](https://github.com/amankharwal/Website-data/blob/master/association_rule_market_basket_analysis.ipynb)

### Leave a Reply

Enter your comment here...