

```
In [2]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules
```

```
In [3]: # Example dataset: Keywords extracted from news articles or social media posts
data = [
    ["Bitcoin", "Cryptocurrency", "Investment"],
    ["Stocks", "Market", "Trading"],
    ["Economy", "Recession", "Stimulus"],
    ["Technology", "Innovation", "Startups"],
    ["Real Estate", "Property", "Housing"]
]
```

```
In [4]: # Transaction Encoder to convert data into a one-hot encoded format
te = TransactionEncoder()
te_ary = te.fit(data).transform(data)
df = pd.DataFrame(te_ary, columns=te.columns_)

# Applying Apriori algorithm
frequent_itemsets = apriori(df, min_support=0.2, use_colnames=True)

# Generating association rules
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.7)
```

```
In [5]: # Displaying frequent itemsets
print("Frequent Itemsets:")
print(frequent_itemsets)
```

Frequent Itemsets:

| | support | itemsets |
|----|---------|---------------------------------------|
| 0 | 0.2 | (Bitcoin) |
| 1 | 0.2 | (Cryptocurrency) |
| 2 | 0.2 | (Economy) |
| 3 | 0.2 | (Housing) |
| 4 | 0.2 | (Innovation) |
| 5 | 0.2 | (Investment) |
| 6 | 0.2 | (Market) |
| 7 | 0.2 | (Property) |
| 8 | 0.2 | (Real Estate) |
| 9 | 0.2 | (Recession) |
| 10 | 0.2 | (Startups) |
| 11 | 0.2 | (Stimulus) |
| 12 | 0.2 | (Stocks) |
| 13 | 0.2 | (Technology) |
| 14 | 0.2 | (Trading) |
| 15 | 0.2 | (Bitcoin, Cryptocurrency) |
| 16 | 0.2 | (Bitcoin, Investment) |
| 17 | 0.2 | (Cryptocurrency, Investment) |
| 18 | 0.2 | (Recession, Economy) |
| 19 | 0.2 | (Stimulus, Economy) |
| 20 | 0.2 | (Property, Housing) |
| 21 | 0.2 | (Real Estate, Housing) |
| 22 | 0.2 | (Startups, Innovation) |
| 23 | 0.2 | (Innovation, Technology) |
| 24 | 0.2 | (Stocks, Market) |
| 25 | 0.2 | (Market, Trading) |
| 26 | 0.2 | (Property, Real Estate) |
| 27 | 0.2 | (Recession, Stimulus) |
| 28 | 0.2 | (Startups, Technology) |
| 29 | 0.2 | (Stocks, Trading) |
| 30 | 0.2 | (Bitcoin, Investment, Cryptocurrency) |
| 31 | 0.2 | (Recession, Stimulus, Economy) |
| 32 | 0.2 | (Property, Real Estate, Housing) |
| 33 | 0.2 | (Startups, Innovation, Technology) |
| 34 | 0.2 | (Stocks, Market, Trading) |

```
In [6]: # Displaying association rules
print("\nAssociation Rules:")
print(rules[['antecedents', 'consequents', 'support', 'confidence']])
```

Association Rules:

| | antecedents | consequents | support \ |
|----|------------------------------|------------------------------|-----------|
| 0 | (Bitcoin) | (Cryptocurrency) | 0.2 |
| 1 | (Cryptocurrency) | (Bitcoin) | 0.2 |
| 2 | (Bitcoin) | (Investment) | 0.2 |
| 3 | (Investment) | (Bitcoin) | 0.2 |
| 4 | (Cryptocurrency) | (Investment) | 0.2 |
| 5 | (Investment) | (Cryptocurrency) | 0.2 |
| 6 | (Recession) | (Economy) | 0.2 |
| 7 | (Economy) | (Recession) | 0.2 |
| 8 | (Stimulus) | (Economy) | 0.2 |
| 9 | (Economy) | (Stimulus) | 0.2 |
| 10 | (Property) | (Housing) | 0.2 |
| 11 | (Housing) | (Property) | 0.2 |
| 12 | (Real Estate) | (Housing) | 0.2 |
| 13 | (Housing) | (Real Estate) | 0.2 |
| 14 | (Startups) | (Innovation) | 0.2 |
| 15 | (Innovation) | (Startups) | 0.2 |
| 16 | (Innovation) | (Technology) | 0.2 |
| 17 | (Technology) | (Innovation) | 0.2 |
| 18 | (Stocks) | (Market) | 0.2 |
| 19 | (Market) | (Stocks) | 0.2 |
| 20 | (Market) | (Trading) | 0.2 |
| 21 | (Trading) | (Market) | 0.2 |
| 22 | (Property) | (Real Estate) | 0.2 |
| 23 | (Real Estate) | (Property) | 0.2 |
| 24 | (Recession) | (Stimulus) | 0.2 |
| 25 | (Stimulus) | (Recession) | 0.2 |
| 26 | (Startups) | (Technology) | 0.2 |
| 27 | (Technology) | (Startups) | 0.2 |
| 28 | (Stocks) | (Trading) | 0.2 |
| 29 | (Trading) | (Stocks) | 0.2 |
| 30 | (Bitcoin, Investment) | (Cryptocurrency) | 0.2 |
| 31 | (Bitcoin, Cryptocurrency) | (Investment) | 0.2 |
| 32 | (Cryptocurrency, Investment) | (Bitcoin) | 0.2 |
| 33 | (Bitcoin) | (Cryptocurrency, Investment) | 0.2 |
| 34 | (Investment) | (Bitcoin, Cryptocurrency) | 0.2 |
| 35 | (Cryptocurrency) | (Bitcoin, Investment) | 0.2 |
| 36 | (Recession, Stimulus) | (Economy) | 0.2 |
| 37 | (Recession, Economy) | (Stimulus) | 0.2 |
| 38 | (Stimulus, Economy) | (Recession) | 0.2 |
| 39 | (Recession) | (Stimulus, Economy) | 0.2 |
| 40 | (Stimulus) | (Recession, Economy) | 0.2 |
| 41 | (Economy) | (Recession, Stimulus) | 0.2 |
| 42 | (Property, Real Estate) | (Housing) | 0.2 |
| 43 | (Property, Housing) | (Real Estate) | 0.2 |
| 44 | (Real Estate, Housing) | (Property) | 0.2 |
| 45 | (Property) | (Real Estate, Housing) | 0.2 |
| 46 | (Real Estate) | (Property, Housing) | 0.2 |
| 47 | (Housing) | (Property, Real Estate) | 0.2 |
| 48 | (Startups, Innovation) | (Technology) | 0.2 |
| 49 | (Startups, Technology) | (Innovation) | 0.2 |
| 50 | (Innovation, Technology) | (Startups) | 0.2 |
| 51 | (Startups) | (Innovation, Technology) | 0.2 |
| 52 | (Innovation) | (Startups, Technology) | 0.2 |
| 53 | (Technology) | (Startups, Innovation) | 0.2 |

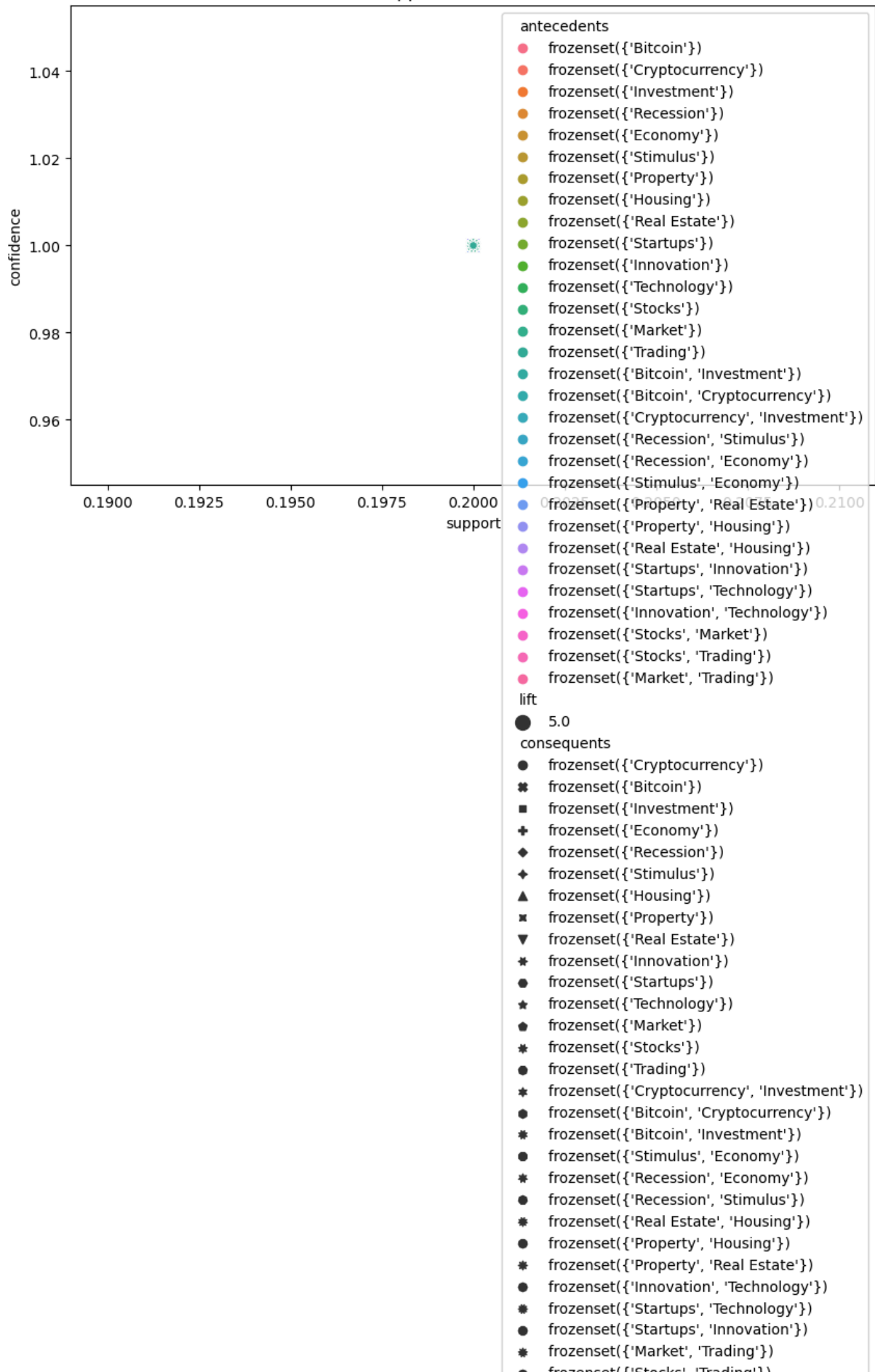
| | | | |
|----|-------------------|-------------------|-----|
| 54 | (Stocks, Market) | (Trading) | 0.2 |
| 55 | (Stocks, Trading) | (Market) | 0.2 |
| 56 | (Market, Trading) | (Stocks) | 0.2 |
| 57 | (Stocks) | (Market, Trading) | 0.2 |
| 58 | (Market) | (Stocks, Trading) | 0.2 |
| 59 | (Trading) | (Stocks, Market) | 0.2 |

| | |
|----|------------|
| | confidence |
| 0 | 1.0 |
| 1 | 1.0 |
| 2 | 1.0 |
| 3 | 1.0 |
| 4 | 1.0 |
| 5 | 1.0 |
| 6 | 1.0 |
| 7 | 1.0 |
| 8 | 1.0 |
| 9 | 1.0 |
| 10 | 1.0 |
| 11 | 1.0 |
| 12 | 1.0 |
| 13 | 1.0 |
| 14 | 1.0 |
| 15 | 1.0 |
| 16 | 1.0 |
| 17 | 1.0 |
| 18 | 1.0 |
| 19 | 1.0 |
| 20 | 1.0 |
| 21 | 1.0 |
| 22 | 1.0 |
| 23 | 1.0 |
| 24 | 1.0 |
| 25 | 1.0 |
| 26 | 1.0 |
| 27 | 1.0 |
| 28 | 1.0 |
| 29 | 1.0 |
| 30 | 1.0 |
| 31 | 1.0 |
| 32 | 1.0 |
| 33 | 1.0 |
| 34 | 1.0 |
| 35 | 1.0 |
| 36 | 1.0 |
| 37 | 1.0 |
| 38 | 1.0 |
| 39 | 1.0 |
| 40 | 1.0 |
| 41 | 1.0 |
| 42 | 1.0 |
| 43 | 1.0 |
| 44 | 1.0 |
| 45 | 1.0 |
| 46 | 1.0 |
| 47 | 1.0 |

| | |
|----|-----|
| 48 | 1.0 |
| 49 | 1.0 |
| 50 | 1.0 |
| 51 | 1.0 |
| 52 | 1.0 |
| 53 | 1.0 |
| 54 | 1.0 |
| 55 | 1.0 |
| 56 | 1.0 |
| 57 | 1.0 |
| 58 | 1.0 |
| 59 | 1.0 |

```
In [7]: # Visualizing the Support vs Confidence
plt.figure(figsize=(10, 6))
sns.scatterplot(x="support", y="confidence", size="lift", data=rules, hue="antecedent")
plt.title('Support vs Confidence')
plt.show()
```

Support vs Confidence



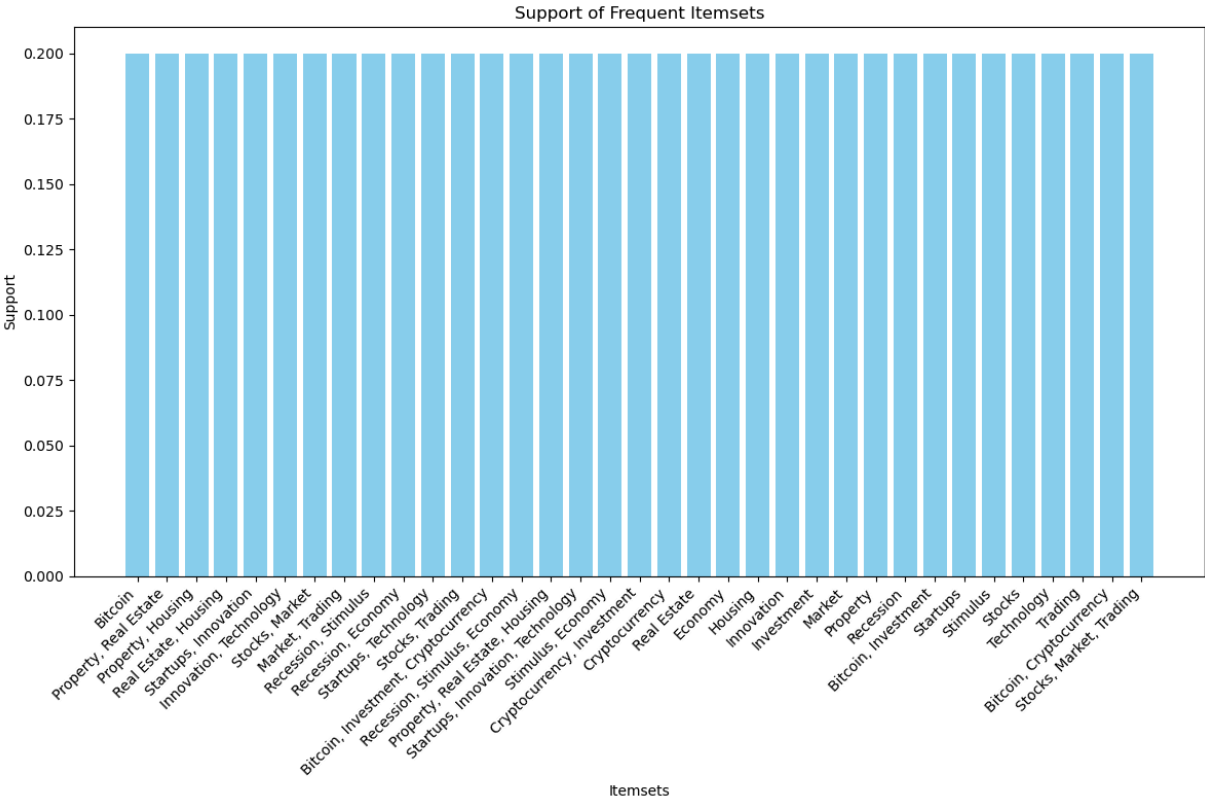
```

● frozenset({ 'Stocks', 'Trading' })
● frozenset({'Stocks', 'Market'})

```

```
# Plotting the frequent itemsets
frequent_itemsets['itemsets'] = frequent_itemsets['itemsets'].apply(lambda x: ', '.join(x))
frequent_itemsets = frequent_itemsets.sort_values(by='support', ascending=False)

plt.figure(figsize=(12, 8))
plt.bar(frequent_itemsets['itemsets'], frequent_itemsets['support'], color='skyblue')
plt.xlabel('Itemsets')
plt.ylabel('Support')
plt.title('Support of Frequent Itemsets')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



```
In [11]: import numpy as np

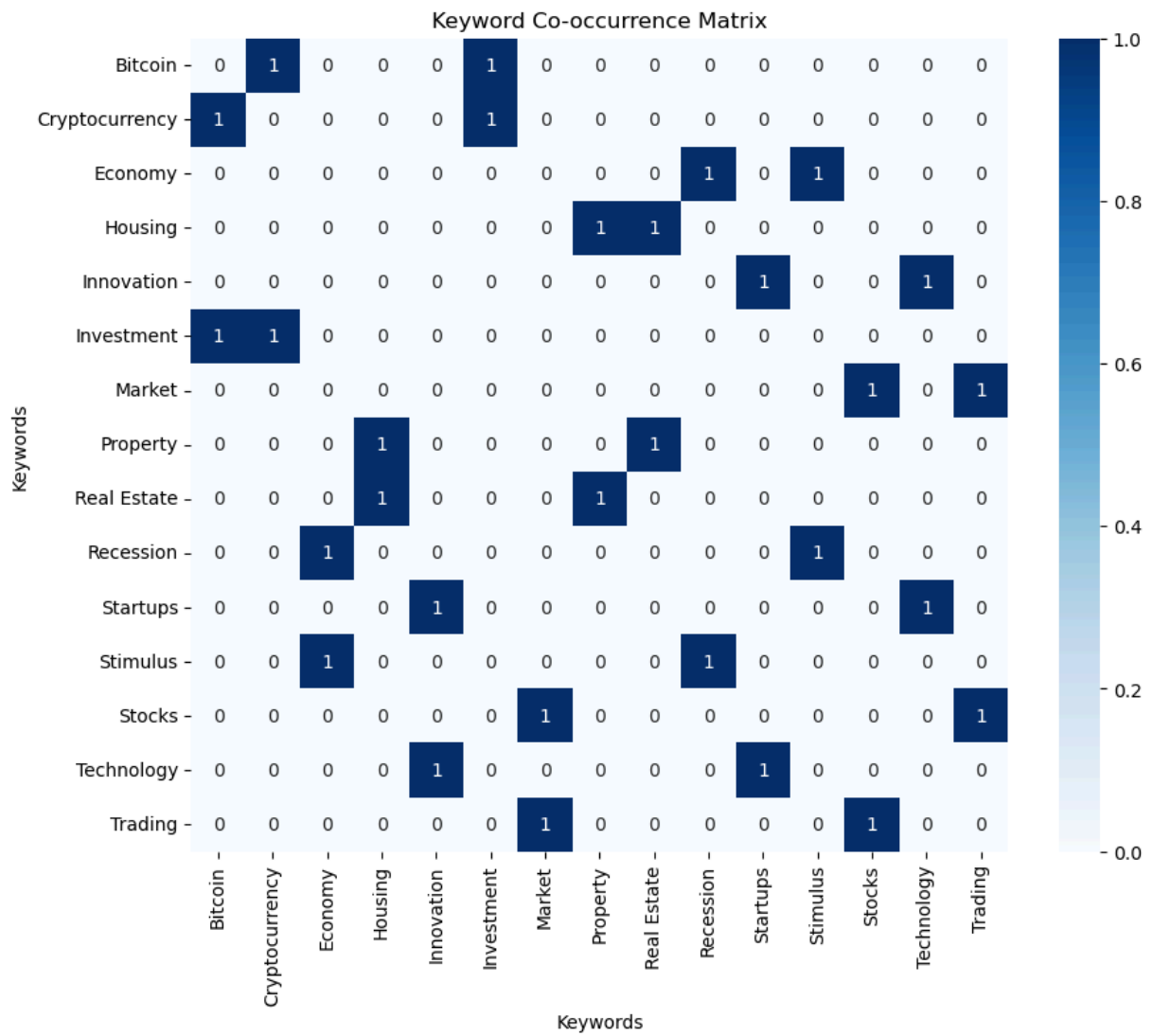
# Transaction Encoder to convert data into a one-hot encoded format
te = TransactionEncoder()
te_ary = te.fit(data).transform(data)
df = pd.DataFrame(te_ary, columns=te.columns_)

# Create a co-occurrence matrix
co_occurrence_matrix = np.dot(df.T, df)
np.fill_diagonal(co_occurrence_matrix, 0) # Optional: Set diagonal to 0 to ignore

# Convert matrix to DataFrame for better handling
co_occurrence_df = pd.DataFrame(co_occurrence_matrix, index=df.columns, columns=df.columns)

# Plotting the co-occurrence matrix as a heatmap
```

```
plt.figure(figsize=(10, 8))
sns.heatmap(co_occurrence_df, annot=True, cmap='Blues', fmt='g')
plt.title('Keyword Co-occurrence Matrix')
plt.xlabel('Keywords')
plt.ylabel('Keywords')
plt.show()
```



In []: