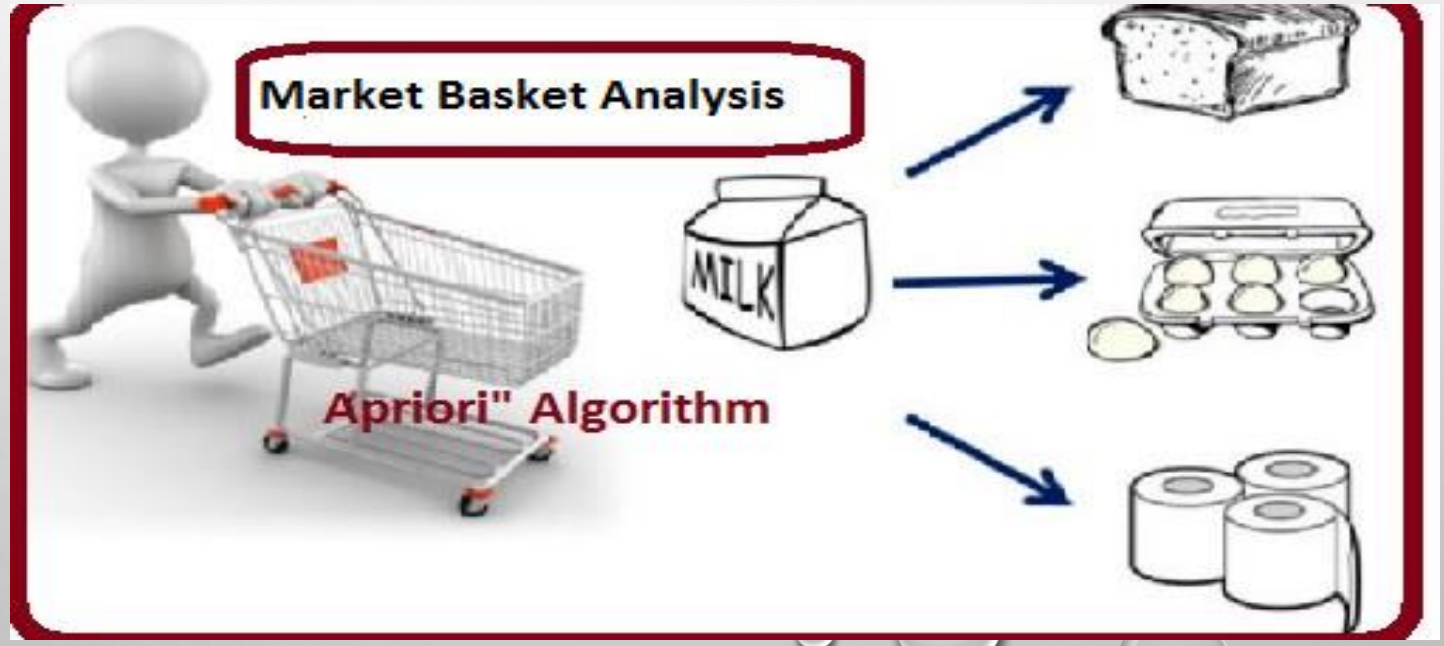
MARKET BASKET INSIGHTS

Phase 2 Submission Document

PROJECT **:** MARKET BASKET INSIGHTS



***INTRODUCTION*:**

THIS MODULE INTRODUCES MARKET BASKET ANALYSIS.

THE NAME OF THIS TECHNIQUES, MARKET BASKET ANALYSIS, PROBABLY DERIVES FROM ITS FREQUENT APPLICATION TO THE ANALYZING CONSUMER BUYING BEHAVIOR OF CONSUMERS IN SUPERMARKETS. UPON CHECK-OUT, THE ITEMS IN THE CART OR BASKET ARE RECORDED. ESPECIALLY WHEN LINKED TO CUSTOMER OR MEMBERSHIP CARDS, THE DATA PROVIDES VALUABLE INSIGHT IN BUYING PATTERNS.

SOME EXAMPLES:

PURCHASES IN SUPERMARKETS

BEHAVIORAL TRAITS OF CRIMINALS

DNA PATTERNS IN DISEASES

OR THE TITANIC DATA …

CONTENT FOR PROJECT PHASE 2:

CONSIDER EXPLORING ADVANCED ASSOCIATION ANALYSIS TECHNIQUES OR USING VISUALIZATION TOOLS FOR ENHANCED INSIGHTS PRESENTATION

Data Source

A good data source for market basket insights should be accurate, complete, cover the business.

Dataset
Link: https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis

STRATEGY

- DATA IMPORT
- DATA UNDERSTANDING AND EXPLORATION
- TRANSFORMATION OF THE DATA – SO THAT IS READY TO BE CONSUMED BY THE ASSOCIATION RULES ALGORITHM
- RUNNING ASSOCIATION RULES
- EXPLORING THE RULES GENERATED
- FILTERING THE GENERATED RULES
- VISUALIZATION OF RULE

EXPLORATORY DATA ANALYSIS (EDA):

- VISUALIZE AND ANALYZE THE DATASET TO GAIN INSIGHTS INTO THE RELATIONSHIPS BETWEEN VARIABLES.
- IDENTIFY CORRELATIONS AND PATTERNS THAT CAN INFORM FEATURE SELECTION AND ENGINEERING.
- PRESENT VARIOUS DATA VISUALIZATIONS TO GAIN INSIGHTS INTO THE DATASET.
- EXPLORE CORRELATIONS BETWEEN FEATURES AND THE TARGET VARIABLE.
- DISCUSS ANY SIGNIFICANT FINDINGS FROM THE EDA PHASE THAT INFORM FEATURE SELECTION.

FEATURE ENGINEERING:

- CREATE NEW FEATURES OR TRANSFORM EXISTING ONES TO CAPTURE VALUABLE INFORMATION.

. EXPLAIN THE PROCESS OF CREATING NEW FEATURES OR TRANSFORMING EXISTING ONES.

- SHOWCASE DOMAIN-SPECIFIC FEATURE ENGINEERING, SUCH AS PROXIMITY SCORES OR COMPOSITE

INDICATORS.

- EMPHASIZE THE IMPACT OF ENGINEERED FEATURES ON MODEL PERFORMANCE.

APRIORI ALGORITHM USED IN MARKET BASKET ANALYSIS

ALGORITHMS THAT USE ASSOCIATION RULES INCLUDE AIS, SETM AND APRIORI. THE APRIORI ALGORITHM IS COMMONLY CITED BY DATA SCIENTISTS IN RESEARCH ARTICLES ABOUT MARKET BASKET ANALYSIS. IT IDENTIFIES FREQUENT ITEMS IN THE DATABASE AND THEN EVALUATES THEIR FREQUENCY AS THE DATASETS ARE EXPANDED TO LARGER SIZES.

DATA VISUALIZATION

DATA VISUALIZATION IS THE PRACTICE OF TRANSLATING INFORMATION INTO A VISUAL CONTEXT, SUCH AS A MAP OR GRAPH, TO MAKE DATA EASIER FOR THE HUMAN BRAIN TO UNDERSTAND AND PULL INSIGHTS FROM. THE MAIN GOAL OF DATA VISUALIZATION IS TO MAKE IT EASIER TO IDENTIFY PATTERNS, TRENDS AND OUTLIERS IN LARGE DATASET_. THE TERM IS OFTEN USED INTERCHANGEABLY WITH OTHERS, INCLUDING INFORMATION GRAPHICS, INFORMATION VISUALIZATION AND STATISTICAL GRAPHICS.

INSIGHTS ARE COMPUTED AND DISPLAYED AT EACH CATEGORY OF A SINGLE CATEGORICAL EXPLANATORY FIELD, OR EACH COMBINATION OF CATEGORIES OF A PAIR OF CATEGORICAL EXPLANATORY FIELDS IN THE VISUALIZATION.

MARKET BASKET ANALYSIS IMPLEMENTATION WITHIN PYTHON

. WITH THE HELP OF THE APRIORI PACKAGE, WE WILL BE IMPLEMENTING THE APRIORI ALGORITHM IN ORDER TO HELP THE MANAGER IN MARKET BASKET ANALYSIS .

STEP-1: WE IMPORT THE NECESSARY LIBRARIES REQUIRED FOR THE IMPLEMENTATION.

```
IMPORT NUMPY AS NP
IMPORT MATPLOTLIB.PYPLOT AS PLT
IMPORT PANDAS AS PD
```

STEP-2: LOAD THE DATASET

NOW WE HAVE TO PROCEED BY READING THE DATASET WE HAVE, THAT IS IN A CSV FORMAT. WE DO THAT USING PANDAS MODULE'S READ_CSV FUNCTION

```
DATASET = PD.READ_CSV("../INPUT/MARKET-
BASKET/MARKET_BASKET_OPTIMISATION.CSV", HEADER = NONE)

TRANSACTIONS = []

FOR I IN RANGE(0, 7501):

        TRANSACTIONS.APPEND([STR(DATASET.VALUES[I,J]) FOR J IN
RANGE(0,20)])
```

Step-3: Take a glance at the records

```
Input

   dataset
```

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
0	shrimp	almonds	avocado	vegetables mix	greengrapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	greentea	honey	salad	mineral water	salmon	antioxydant juice	frozensmoothie	spinach	olive oil
1	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	mineral water	milk	energy bar	whole wheat rice	greentea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7496	butter	light mayo	freshbread	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7497	burgers	frozenvegetables	eggs	frenchfries	magazines	greentea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7498	chicken	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7499	escalope	greentea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7500	eggs	frozensmoothie	yogurtcake	low fat yogurt	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

STEP-4: LOOK AT THE SHAPE

INPUT DATASET

OUTPUT

```
(7501, 20)
```

STEP-5: CONVERT PANDAS DATA FRAME INTO A LIST OF LISTS

```
FOR I IN RANGE(0, 7501): TRANSACTIONS.APPEND([STR(DATASET.VALUES[I,J]) FOR J IN RANGE(0,20)])
```

```
!PIP INSTALL APYORI
```

```
COLLECTING APYORI
  DOWNLOADING APYORI-1.1.2.TAR.GZ (8.6 KB)
  PREPARING METADATA (SETUP.PY) ... - DONE
BUILDING WHEELS FOR COLLECTED PACKAGES: APYORI
BUILDING WHEEL FOR APYORI (SETUP.PY) ... - \ DONE
  CREATED WHEEL FOR APYORI: FILENAME=APYORI-1.1.2-PY3-NONE-ANY.WHL SIZE=5974
SHA256=DED9511119EE58B9D5E1562B3A7DB41AC2D1CD3913B6E7AE7D7039E68B7465AA
STORED IN DIRECTORY: /ROOT/.CACHE/PIP/WHEELS/CB/F6/E1/57973C631D27EFD1A2F375BD6A83B2A616C4021F24AAB84080
. SUCCESSFULLY BUILT APYORI INSTALLING COLLECTED PACKAGES: APYORI
  SUCCESSFULLY INSTALLED APYORI-1.1.2
```

STEP-6: BUILD THE APRIORI MODEL

WE IMPORT THE APRIORI FUNCTION FROM THE APRIORI MODULE. WE STORE THE RESULTING OUTPUT FROM APRIORI FUNCTION IN THE 'RULES' VARIABLE. TO THE APRIORI FUNCTION, WE PASS 6 PARAMETERS:

1. THE TRANSACTIONS LIST AS THE MAIN INPUTS

*2. MINIMUM SUPPORT, WHICH WE SET AS 0.003 WE GET THAT VALUE BY CONSIDERING THAT A PRODUCT SHOULD APPEAR AT LEAST IN 3 TRANSACTIONS IN A DAY. OUR DATA IS COLLECTED OVER A WEEK. HENCE, THE SUPPORT VALUE SHOULD BE 3*7/7500 = 0.0028*

3. MINIMUM CONFIDENCE, WHICH WE CHOOSE TO BE 0.2 (OBTAINED OVER-ANALYZING VARIOUS RESULTS)

4. MINIMUM LIFT, WHICH WE'VE SET TO 3

5. MINIMUM LENGTH IS SET TO 2, AS WE ARE CALCULATING THE LIFT VALUES FOR BUYING AN ITEM B GIVEN ANOTHER ITEM A IS BOUGHT, SO WE TAKE 2 ITEMS INTO CONSIDERATION.

6. MINIMUM LENGTH IS SET TO 2 USING THE SAME LOGIC.

```
FROM APYORI IMPORT APRIORI

RULES = APRIORI(TRANSACTIONS = TRANSACTIONS, MIN_SUPPORT = 0.003, MIN_CINFIDENCE = 0.2, MIN_LIFT = 3,
MIN_LENGTH = 2, MAX_LENGTH = 2)
```

STEP-7: PRINT OUT THE NUMBER OF RULES AS LIST

```
RESULTS = LIST(RULES)
```

STEP-8: HAVE A GLANCE AT THE RULES

INPUT

```
RESULTS
```

OUTPUT

```
. [RELATION RECORD(ITEMS=FROZENSET({'BROWNIES', 'COTTAGE CHEESE'}),
SUPPORT=0.0034662045060658577, ORDERED_STATISTICS=[ORDERED STATISTIC(ITEMS_BASE=FROZEN
SET({'BROWNIES'})

 ITEMS_ADD=FROZEN SET({'COTTAGE CHEESE'}), CONFIDENCE=0.10276679841897232, LIFT=3.225329518580382),
ORDERED STATISTIC(ITEMS_BASE=FROZENSET({'COTTAGE CHEESE'}), ITEMS_ADD=FROZEN SET({'BROWNIES'}),
CONFIDENCE=0.10878661087866107, LIFT=3.2253295185803816)]),

  RELATION RECORD(ITEMS=FROZENSET({'LIGHT CREAM', 'CHICKEN'}), SUPPORT=0.004532728969470737, ORDERED
STATISTICS=[ORDERED STATISTIC(ITEMS_BASE=FROZEN SET({'CHICKEN'}), ITEMS_ADD=FROZEN SET({'LIGHT CREAM'}),
CONFIDENCE=0.07555555555555556, LIFT=4.843950617283951), ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'LIGHT
CREAM'}), ITEMS_ADD=FROZENSET({'CHICKEN'}), CONFIDENCE=0.29059829059829057, LIFT=4.84395061728395)]),

   RELATION RECORD(ITEMS=FROZENSET({'ESCALOPE', 'MUSHROOM CREAM SAUCE'}), SUPPORT=0.005732568990801226,
ORDERED STATISTICS=[ORDERED STATISTIC(ITEMS_BASE=FROZENSET({'ESCALOPE'}), ITEMS_ADD=FROZEN SET({'MUSHROOM
CREAM SAUCE'}), CONFIDENCE=0.0722689075630252, LIFT=3.7908326967150496), ORDERED
STATISTIC(ITEMS_BASE=FROZEN SET({'MUSHROOM CREAM SAUCE'}), ITEMS_ADD=FROZENSET({'ESCALOPE'}),
CONFIDENCE=0.3006993006993007, LIFT=3.790832696715049)]),
```

RELATION RECORD(ITEMS=FROZEN SET({'ESCALOPE', 'PASTA'}), SUPPORT=0.005865884548726837, ORDERED_STATISTICS=[ORDERED STATISTIC(ITEMS_BASE=FROZENSET({'ESCALOPE'}), ITEMS_ADD=FROZEN SET({'PASTA'}), CONFIDENCE=0.07394957983193277, LIFT=4.700811850163794), ORDERED STATISTIC(ITEMS_BASE=FROZEN SET({'PASTA'}), ITEMS_ADD=FROZEN SET({'ESCALOPE'}), CONFIDENCE=0.3728813559322034, LIFT=4.700811850163794)]),

RELATION RECORD(ITEMS=FROZENSET({'FRESH BREAD', 'TOMATO JUICE'}), SUPPORT=0.004266097853619517, ORDERED_STATISTICS=[ORDERED STATISTIC(ITEMS_BASE=FROZEN SET({'FRESH BREAD'}), ITEMS_ADD=FROZEN SET({'TOMATO JUICE'}), CONFIDENCE=0.09907120743034055, LIFT=3.2593558198902826), ORDERED STATISTIC(ITEMS_BASE=FROZEN SET({'TOMATO JUICE'}), ITEMS_ADD=FROZEN SET({'FRESH BREAD'}), CONFIDENCE=0.14035087719298245, LIFT=3.2593558198902826)]),

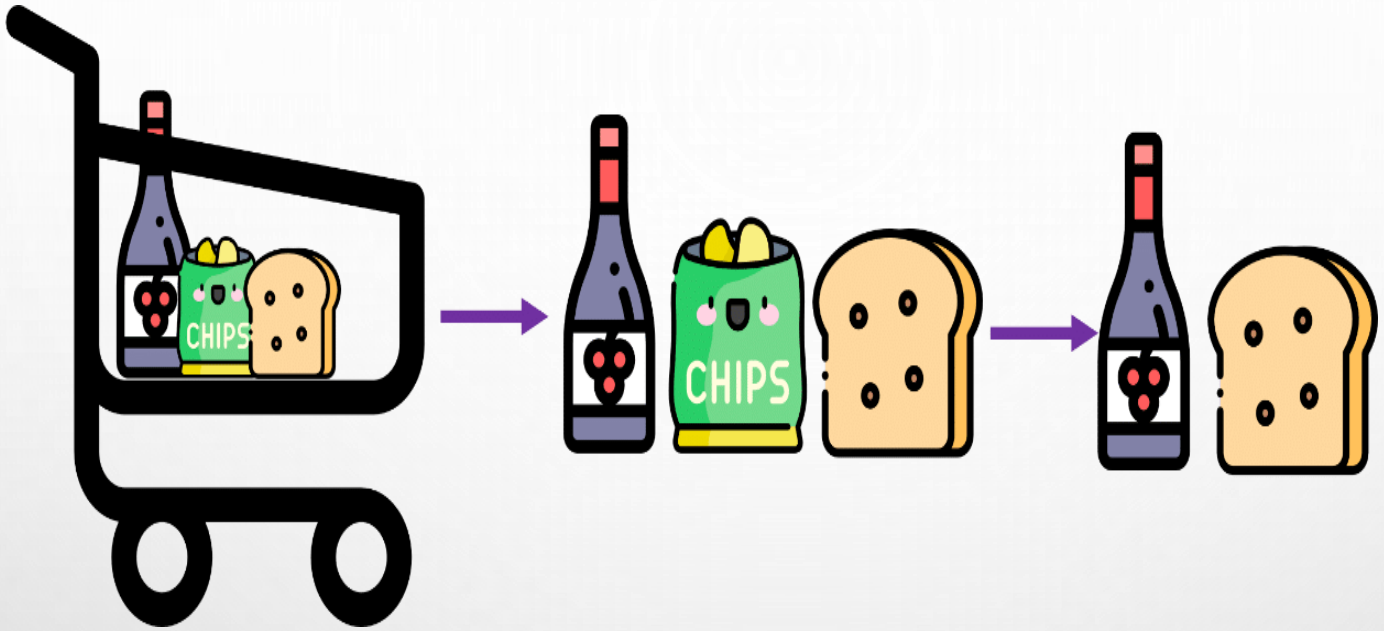
RELATION RECORD(ITEMS=FROZEN SET({'HONEY', 'FRESH TUNA'}), SUPPORT=0.003999466737768298, ORDERED _STATISTICS=[ORDERED STATISTIC(ITEMS_BASE=FROZENSET({'FRESH TUNA'}), ITEMS_ADD=FROZEN SET({'HONEY'}), CONFIDENCE=0.17964071856287428, LIFT=3.785070308820561 3), ORDERED STATISTIC(ITEMS_BASE=FROZEN SET({'HONEY'}), ITEMS_ADD=FROZENSET({'FRESH TUNA'}), CONFIDENCE=0.08426966292134831, LIFT=3.785070308820561 3)]), RELATIONRECORD(ITEMS=FROZEN SET({'HONEY', 'FROMAGE BLANC'}), SUPPORT=0.00333288894814024 8, ORDERED_STATISTICS=[ORDERED STATISTIC(ITEMS_BASE=FROZEN SET({'FROMAGE BLANC'}), ITEMS_ADD=FROZEN SET({'HONEY'}), CONFIDENCE=0.2450980392156863, LIFT=5.164270764485569), ORDERED STATISTIC(ITEMS_BASE=FROZEN SET({'HONEY'}), ITEMS_ADD=FROZEN SET({'FROMAGE BLANC'}), CONFIDENCE=0.0702247191011236, LIFT=5.16427076448557)]),

Relation Record(items=frozenset({'ground beef', 'herb & pepper'}), support=0.01599786695107319 2, ordered_statistics=[OrderedStatistic(items_base=frozenset({'ground beef'}), items_add=frozenset({'herb & pepper'}), confidence=0.16282252374491 2, lift=3.291993841134928), OrderedStatistic(items_base=frozenset({'herb & pepper'}), items_add=frozenset({'ground beef'}), confidence=0.3234501347708895, lift=3.291993841134928 5)]),

RELATION RECORD(ITEMS=FROZENSET({'GROUND BEEF', 'TOMATO SAUCE'}), SUPPORT=0.005332622317024397, ORDERED_STATISTICS=[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'GROUND BEEF'}), ITEMS_ADD=FROZENSET({'TOMATO SAUCE'}), CONFIDENCE=0.05427408412483039 6, LIFT=3.840659481324083), ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'TOMATO SAUCE'}), ITEMS_ADD=FROZENSET({'GROUND BEEF'}), CONFIDENCE=0.377358490566037 7, LIFT=3.840659481324083)]),

RELATION RECORD(ITEMS=FROZENSET({'LIGHT CREAM', 'OLIVE OIL'}), SUPPORT=0.003199573390214638, ORDERED_STATISTICS=[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'LIGHT CREAM'}), ITEMS_ADD=FROZENSET({'OLIVE OIL'}), CONFIDENCE=0.20512820512820515, LIFT=3.114709851551957 3), ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'OLIVE OIL'}), ITEMS_ADD=FROZENSET({'LIGHT CREAM'}), CONFIDENCE=0.04858299595141700 5, LIFT=3.114709851551957)]),

RELATION RECORD(ITEMS=FROZENSET({'OLIVE OIL', 'WHOLE WHEAT PASTA'}), SUPPORT=0.007998933475536596, ORDERED_STATISTICS=[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'OLIVE OIL'}), ITEMS_ADD=FROZENSET({'WHOLE WHEAT PASTA'}), CONFIDENCE=0.12145748987854252, LIFT=4.122410097642296), ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'WHOLE WHEAT PASTA'}), ITEMS_ADD=FROZENSET({'OLIVE OIL'}), CONFIDENCE=0.2714932126696833, LIFT=4.122410097642296)]),. RELATION RECORD(ITEMS=FROZENSET({'PASTA', 'SHRIMP'}), SUPPORT=0.005065991201173177, ORDERED_STATISTICS=[ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'PASTA'}), ITEMS_ADD=FROZENSET({'SHRIMP'}), CONFIDENCE=0.3220338983050847, LIFT=4.506672147735896), ORDEREDSTATISTIC(ITEMS_BASE=FROZENSET({'SHRIMP'}), ITEMS_ADD=FROZENSET({'PASTA'}), CONFIDENCE=0.07089552238805 97, LIFT=4.506672147735896)])]



Step-9: Visualizing the results

. *In the LHS variable, we store the first item from all the results, from which we obtain the second item that is bought after that item is already bought, which is now stored in the RHS variable. The supports, confidences and lifts store all the support, confidence and lift values from the results .*

```
DEF INSPECT(RESULTS):
  LHS =[TUPLE(RESULT[2][0][0])[0] FOR RESULT IN RESULTS]
   RHS =[TUPLE(RESULT[2][0][1])[0] FOR RESULT IN RESULTS]
  SUPPORTS =[RESULT[1] FOR RESULT IN RESULTS]
    CONFIDENCES =[RESULT[2][0][2] FOR RESULT IN RESULTS]
   LIFTS =[RESULT[2][0][3] FOR RESULT IN RESULTS]
   RETURN LIST (ZIP(LHS, RHS, SUPPORTS, CONFIDENCES, LIFTS))    B     RESULTSINDATAFRAME
= PD.DATAFRAME(INSPECT(RESULTS), COLUMNS = ["LEFT HAND SIDE", "RIGHT HAND SIDE",
"SUPPORT", "CONFIDENCE", "LIFT"])
```

CONCLUSION (PHASE 2):

PROJECT CONCLUSION :

- IN THE PHASE 2 CONCLUSION, WE WILL SUMMARIZE THE KEY FINDINGS AND INSIGHTS FROM THE ADVANCED ASSOCIATION ANALYSIS TECHNIQUES. WE WILL REITERATE THE IMPACT OF THESE TECHNIQUES ON IMPROVING THE MARKET BASKET INSIGHTS.
- BASED ON THE RESULTS OF THESE CALCULATIONS CAN BE USED AS A RECOMMENDATION FOR RETAIL OWNERS TO ARRANGE THE ARRANGEMENT OF PRODUCT CATALOGS AND TAKE STRATEGIC STEPS TO IMPROVE PRODUCT MARKETING.. BY UTILIZING THE ASSOCIATION RULES WHICH ARE DISCOVERED AS A RESULT OF THE ANALYSES, THE RETAILER CAN APPLY EFFECTIVE MARKETING AND SALES PROMOTION STRATEGIES, HE WILL BE ABLE INCREASE CUSTOMER ENGAGEMENT AND IMPROVE CUSTOMER EXPERIENCE AND IDENTIFY CUSTOMER BEHAVIOR.

THANK YOU