SYMBIOSIS INSTITUTE OF COMPUTER STUDIES AND RESEARCH

SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

MBA – IT

Fundamentals of Data Science

Data Mining Project on

"Predictive Maintenance Data of Agency Procurement (Textiles Mills)"

Submitted By :

Dnyaneshwar Barapatre (23030141021)

Mannat Shah (23030141039)

Pulkit Chhajed (23030141046)

Rahul Pandey (23030141047)

Priyank Rathore (23030141049)

Submitted To :

Dr. Priti Kulkarni

Acknowledgement

We would like to fully appreciate the all-important guidance, staunch support, and in-depth feedback provided by Dr. Priti Kulkarni over the course of this Data Mining Project related to 'Predictive Maintenance Data on Inventory Arrays (Textiles Mills).' Her expertises and total dedication helped us to become successful in this project.

Using this opportunity, we would like to express my deepest gratitude to the teaching faculty who made creation of the ideal learning environment possible and supplied the proper tools for the implementation of the project. From the very first days of self-study till now, their insistence on high academic performance has always stirred me within.

Besides, we owe a bighearted gratitude to the organization that gave us their data for analysis, including Agency Procurement Predictive Maintenance Data in Textiles Mills, for their efforts in providing the necessary data to conduct our analysis. As a result, our partnership with such experts has been extremely valuable in the effective construction of our machine learning models.

Besides that, we would like to highlight my teammates, who have been active listeners in the discussions, offered ideas and played the role of a friendly critic to keep us and our project on track. Their concerted nature is what has actually emboldened the general learning toward meaningful.

Furthermore, we show our extreme thankfulness to our family and friends for the constant support and cheerfulness which they have given me all this time. The fact that they believed in me and my capabilities made me move forward with all strength and courage during the project.

Sincere thanks going to all who participated in the collation and perfection of this machine-learning model. We could not thank you enough for the weight support you've been kind enough to give me; We am so happy to have been given the chance to work on the exceptional and eye-opening project.

Index

Abstract

The main role of the agents/brokers is bringing together buyers of textiles with sellers through assisting with the transactions. Here, the agent will be doing the job of dealing with suppliers, sponsoring samples for customers, and making orders. Although, government departments have low market conditions in textiles procurement, volatility in terms of demand and suppliers remain a challenge to be dealt with until a certain level of certainty is reached. A comprehensive dedication of this paper is the procurement process optimization for apparel operations by focusing on expenses reduction, inventory control and this way contributing to performance enhancement.

At the beginning of the project, we collect the data from the variety of datasets on the procurement transactions, suppliers and the related environmental indicators. We process data using state-of-the- data preprocessing approaches such as data cleaning, transformation, and integration to check that it has the correct form and completeness for analysis.

The key in the implementation of machine learning algorithms is to reveal patterns, trends, and possible correlations between items or services that might lead to reshaping the procurement system. The mainly exploratory data analytics will will be engaged to visualize the data, enabling the user to comprehended the textile supply chain from mining operations point of view, as well.

Introduction

Agent/broker is the same to a person who does all the logistical bookings, facilitates trade transactions and collects a commission whenever the deal they work on is done. Agent communicates with the supplier and asks for the samples, after which he shows the samples to the buyer, eventually the buyer transmits the order requirement and the agent forwards the information to the supplier, the supplier packs the orders and the buyer performs the quality assurance tests and then the two parties agree on the terms of payment.

Problem Statement

The Gender equity office is at a point where there is an urgent need to carry out an in-depth study of its financial (2021–2023) data. The goal is to acquire the valuable information on the prior sales period, reveal the trends, and uncover the influencing factors which will help us boosting sales performance in the given period. The very business that seeks to benefit from owning this historical data already may use it in their decision-making and projections into future sales.

Objectives

The purpose of machining learning in the segment of Agency Procurement System for the textiles is to maximize the effectiveness and efficiency of procurement process in the textile industry via data driven innovative techniques. The primary objectives include:

• **Efficiency Improvement:** Speeding up and deploying automation in purchasing process which serves to decrease the number of people´s involvement and delays in decisions. The aim is to streamline the overall performance of procurement to shorten the time that is required for obtaining the required parts.

• **Cost Optimization:** For instance, emulation of the machine learning technology helps to condense reams of historical procurement data and draw cost-cutting trends. Through decisions on procurement options, selecting the supplier is followed mechanisms of ensuring the best value for money in the qualitative matters, quantity and performance from the supplier.

• **Supplier Relationship Management:** Exploiting the machine learning algorithms to assess and anticipate suppliers' performance, resulting in the agencies have opportunity to make a rational decision about supplier choice and partnership.

• **Data-Driven Decision Making:** Convert machine learning models' strength into a tool that helps analyze mass data and give cognized the systems in the procurement strategy for decision-making purposes. Give agencies room to use data as a basis for their decision-making which is in tune with the requirements and objectives of business operations.

Dataset Information

The DataSet of a Textile Agency Company is selected, who are fronted end for Procurement of Materials from suppliers end to the Buyer end, receiving Payment for the Invoice raised and receiving Commission for rendered services.

The Data Source is for a Firm Bearing a Name “Hira Textiles” , Where He Deals with the Agency business of suiting and shirting fabric from all across India Country.

The data is originally in .csv format and consists of 1148 Instances and 16 Attributes.

Dataset: Data Used: See the attachment (Please be informed this dataset shall only be available using sicsr email.

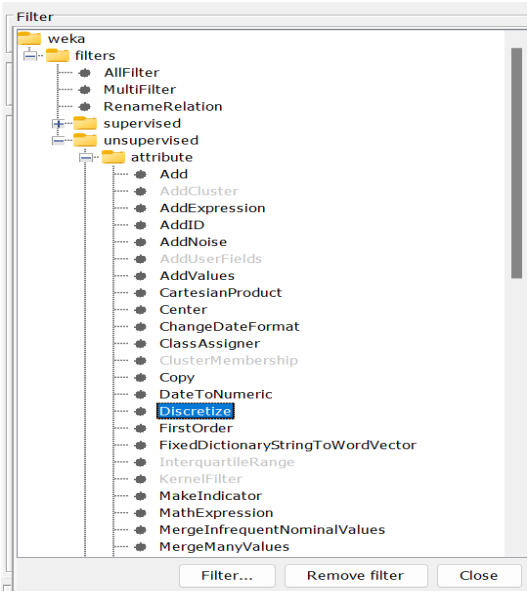
https://drive.google.com/file/d/18I_wh86QXLEldmY8CcjY0the6GIK9X_E/view?usp=drive_link

Data Preprocessing

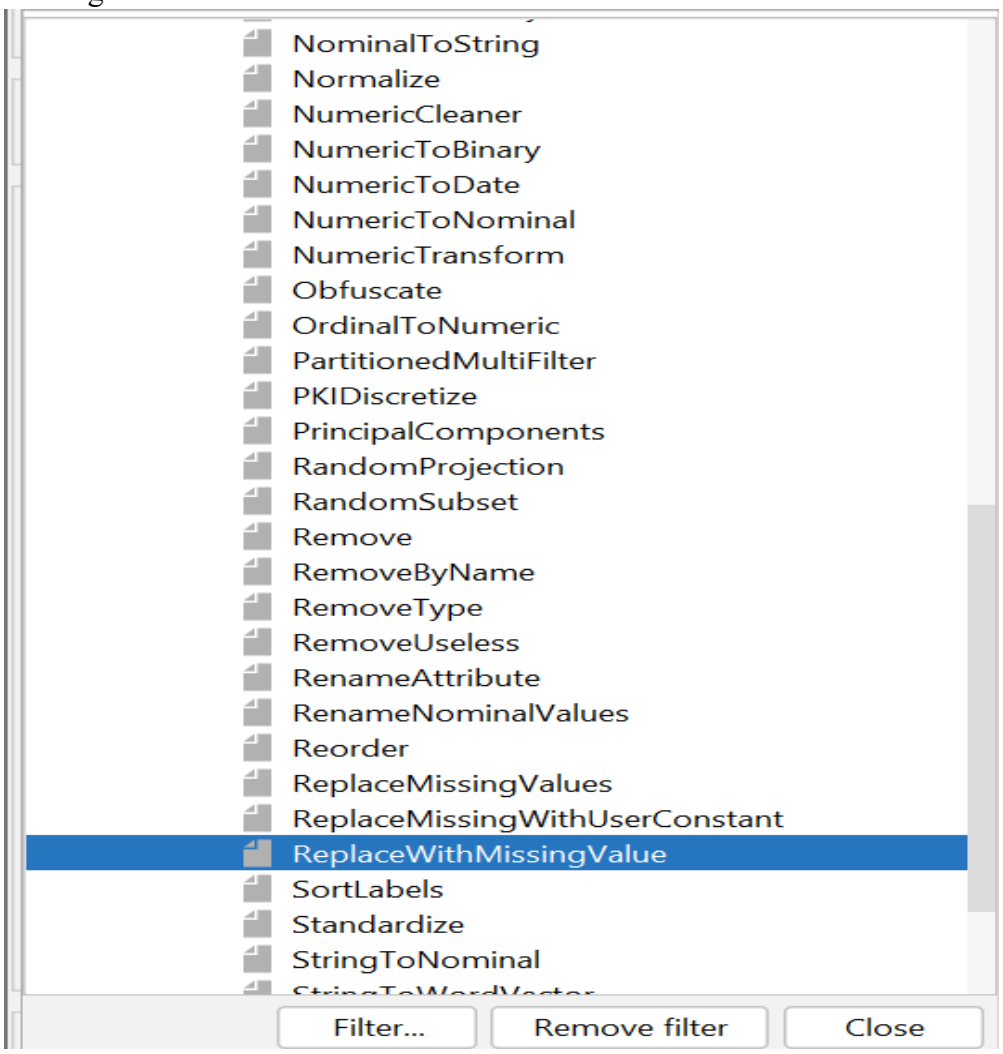
For Data Preprocessing following functions were used to preprocess the dataset:For Data Preprocessing following functions were used to preprocess the dataset:

• **Remove the Balance Amount Attribute:** To delete the Balance Amount attribute, select the check box the remove button you can find should be highlighted by pressing the button you remove the Balance Amount attribute.

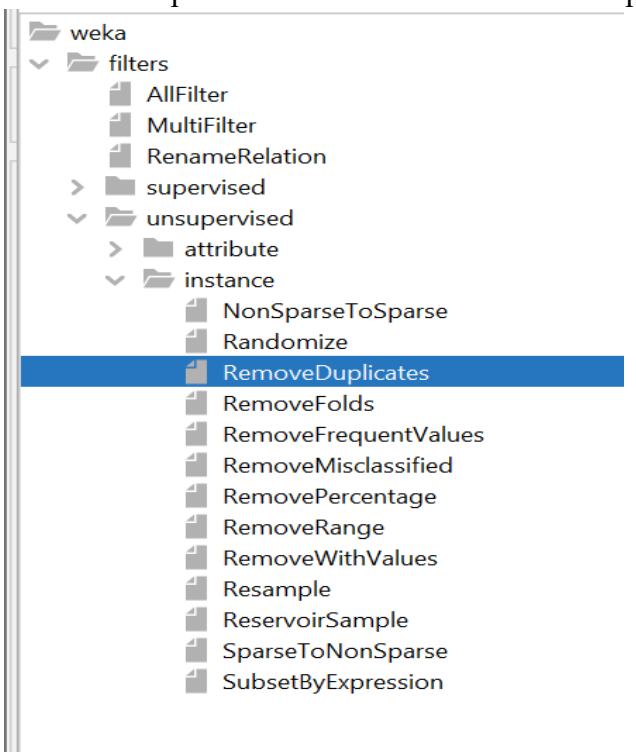
• **Discretize the Dataset:** Attribute Type is Dataset in which Nominal Chart is changed into Numeric by applying discretize to Filters, Open Filters -> Unsupervised ->Attribute -> discretize and press Apply. In this step we look for for applying the Algorithms of machine learning



• **Replace Missing Values:** So For Processing The Missing Instances to All and Then Under Filter -> Unsupervised -> Attribute -> ReplaceMissingValues, press Apply. Define an attribute array for missing values and index, and then use the add filter with restricting instances with missing values.



- **Remove Duplicate Values:** For Removing the Duplicate Instances select All and then under filter -> unsupervised ->Instances ->Remove Duplicates and press Apply.



- **Attribute Evaluator and Search Method:** In Weka, most of the feature selection chores are implemented through an Attribute Evaluator concept which subsequently gets combined with a Search Method. Polytrope Quality Attributor evaluates traits of details. The Search Method is an experiment that helps to discover the best subset of features in the search space.

Exploratory Data Analysis

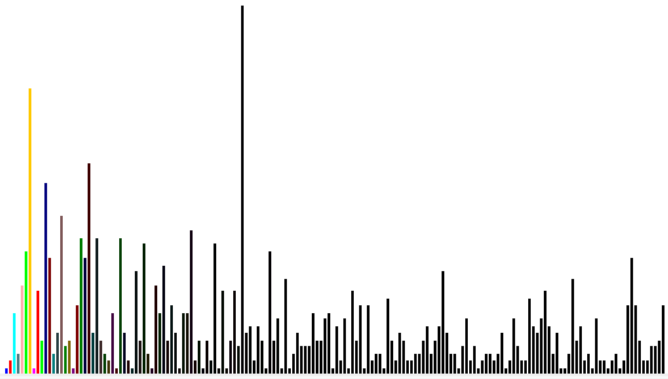
- **Entry date :** The Date on which the Bill is being entered on the agents System

Selected attribute

Name: Entry Date Type: Nominal
Missing: 1 (0%) Distinct: 168 Unique: 25 (2%)

No.	Label	Count	Weight
1	01-04-2023	1	1
2	04-04-2023	2	2
3	07-04-2023	9	9
4	08-04-2023	3	3
5	10-04-2023	13	13
6	11-04-2023	18	18
7	15-04-2023	42	42
8	17-04-2023	1	1
9	18-04-2023	12	12
10	19-04-2023	5	5

Class: Entry Date (Nom) Visualize All

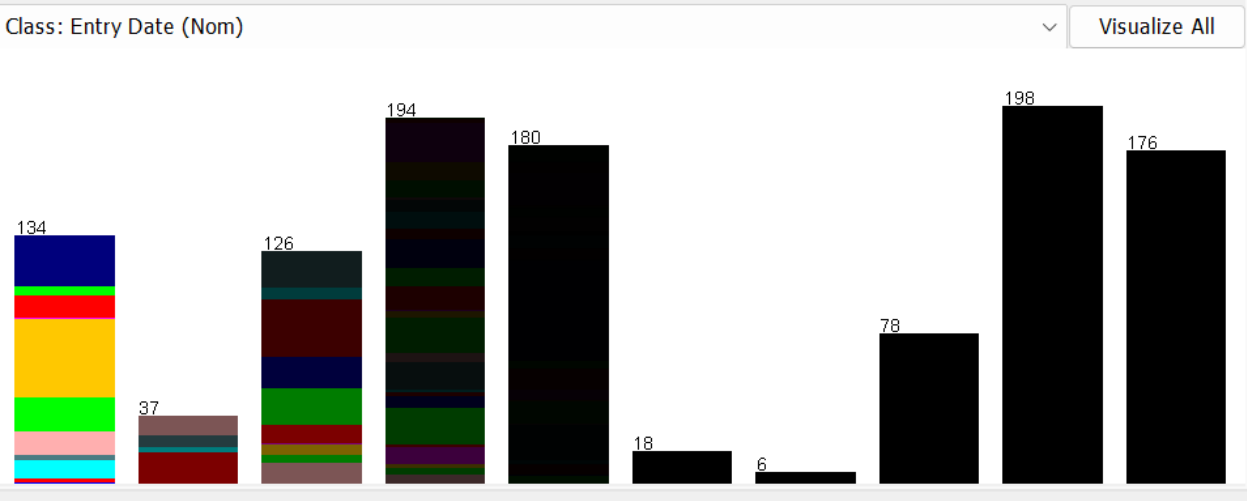


- **Bill Entry Id:** It is System generated number for all the bill entries.

Selected attribute

Name: Bill Entry id Type: Nominal
Missing: 1 (0%) Distinct: 10 Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-4481]'	134	134
2	'(4481-4667]'	37	37
3	'(4667-4853]'	126	126
4	'(4853-5039]'	194	194
5	'(5039-5225]'	180	180
6	'(5225-5411]'	18	18
7	'(5411-5597]'	6	6
8	'(5597-5783]'	78	78
9	'(5783-5969]'	198	198
10	'(5969-inf)'	176	176



- **Bill Number:** It is Supplier Bill Number.

Selected attribute

Name: Bill Number
Missing: 1 (0%)
Distinct: 856
Type: Nominal
Unique: 670 (58%)

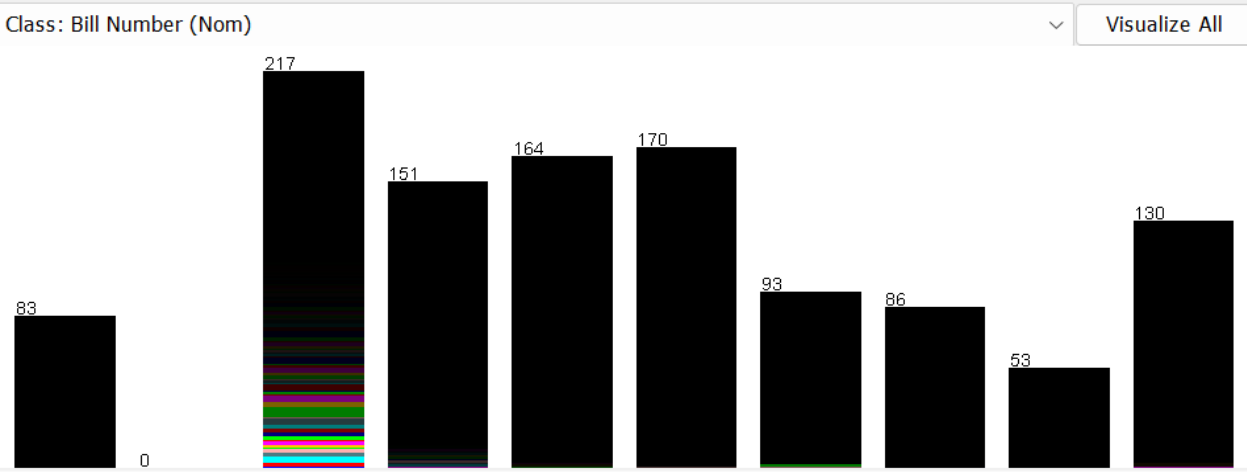
No.	Label	Count	Weight
1	late diff	1	1
2	2	2	2
3	3	4	4
4	4	2	2
5	43	2	2
6	40024	1	1
7	40025	1	1
8	505	2	2
9	40012	1	1
10	15	2	2

- **Bill Month:** It is the Month in which the Bill is being received

Selected attribute

Name: Bill Month
Missing: 1 (0%)
Distinct: 9
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	'(-inf-2.1]'	83	83
2	'(2.1-3.2]'	0	0
3	'(3.2-4.3]'	217	217
4	'(4.3-5.4]'	151	151
5	'(5.4-6.5]'	164	164
6	'(6.5-7.6]'	170	170
7	'(7.6-8.7]'	93	93
8	'(8.7-9.8]'	86	86
9	'(9.8-10.9]'	53	53
10	'(10.9-inf)'	130	130

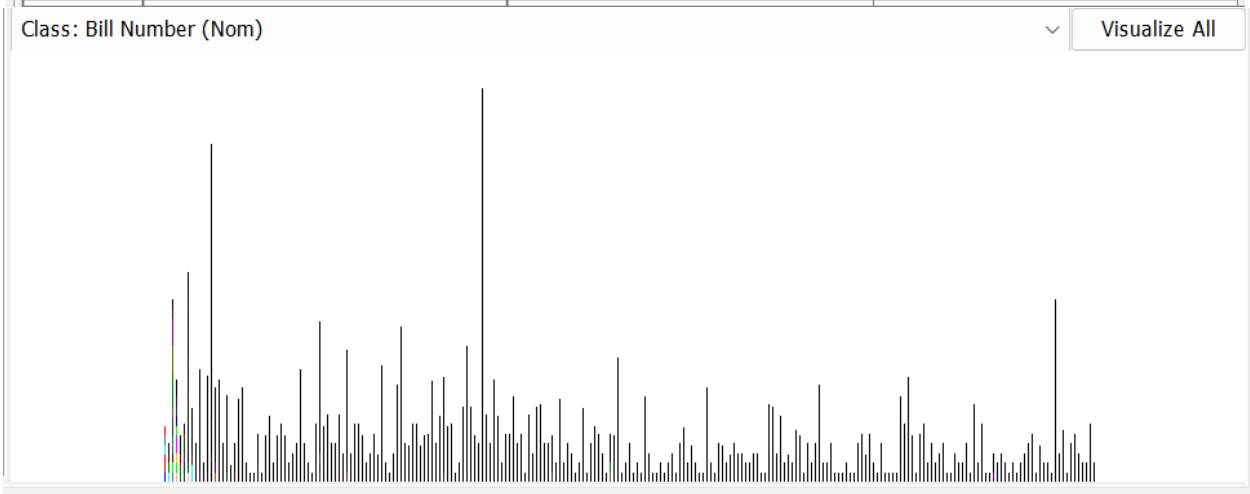


- **Bill Date:** It is the bill date mentioned on the Invoice

Selected attribute

Name: Bill Date
Missing: 1 (0%)
Distinct: 241
Type: Nominal
Unique: 44 (4%)

No.	Label	Count	Weight
1	01-04-2023	6	6
2	05-04-2023	4	4
3	06-04-2023	20	20
4	03-04-2023	11	11
5	04-04-2023	5	5
6	07-04-2023	6	6
7	12-04-2023	22	22
8	11-04-2023	8	8
9	13-04-2023	4	4
10	08-04-2023	12	12



- **Supplier City:** It is city in which the supplier works or has his presence

Selected attribute

Name: Supplier City
Missing: 2 (0%)
Distinct: 9
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Bhilwara	710	710
2	Burhanpur	42	42
3	surat	6	6
4	bhilwara	99	99
5	Ahmedabad	9	9
6	Ichalkarnaji	131	131
7	Balotra	111	111
8	BHILWARA	34	34
9	Ichalkaranji	4	4

Class: Bill Number (Nom)

Visualize All

- **Supplier Group:** It is the group made by Agent according to its features

Selected attribute

Name: Supplier Group
Missing: 2 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

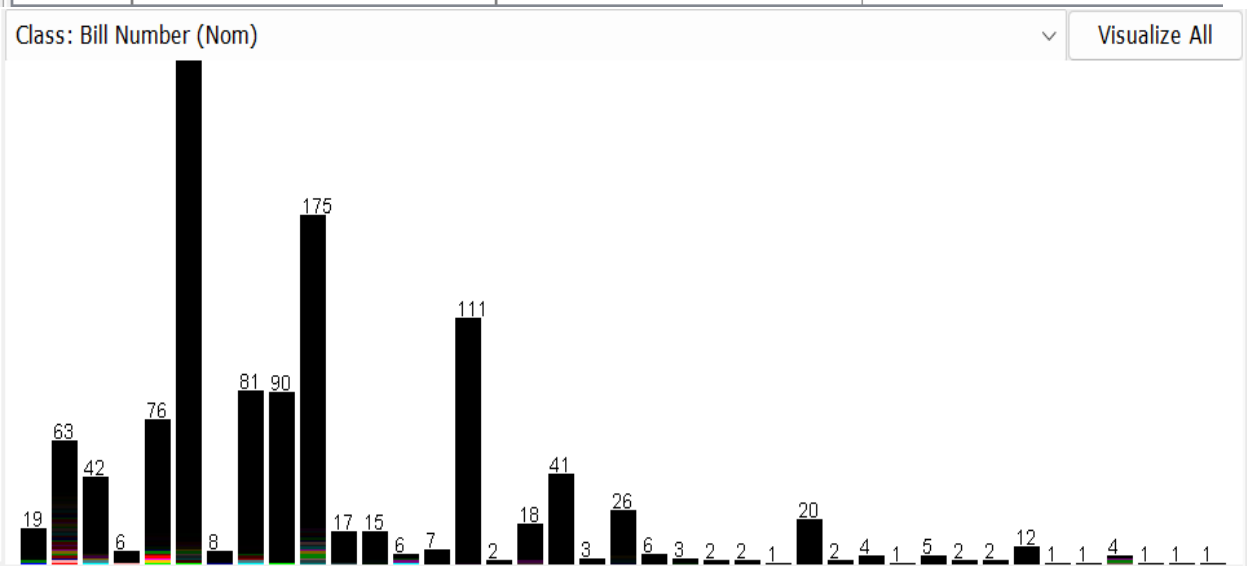
No.	Label	Count	Weight
1	Commission	1137	1137
2	Ahmedabad	9	9

- **Supplier Name:** It is the Name of the Supplier

Selected attribute

Name: Supplier Name
Missing: 1 (0%)
Distinct: 39
Type: Nominal
Unique: 7 (1%)

No.	Label	Count	Weight
1	SAIKRIPA TEXTILES	19	19
2	SANKALP SUITINGS	63	63
3	AKSHIT TEXTILES	42	42
4	SHREE SALASAR TEXTILES	6	6
5	SUNCITY TEXTILE PVT LTD	76	76
6	CHHABRA SYNCOTEX PVT.LTD	270	270
7	GANESHMAL MAHENDRAKUMAR	8	8
8	JYOTI SUITINGS	81	81
9	AARVI INTERNATIONAL	90	90
10	SANYAM TEX FAB	175	175

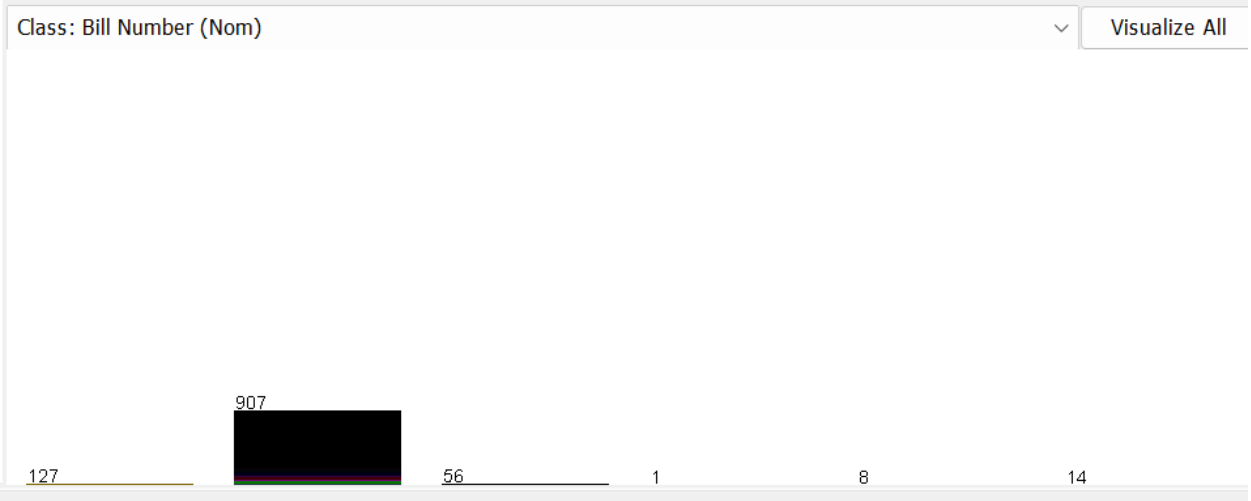


- **Buyer Group:** It is the group made by Agent according to its features.

Selected attribute

Name: Buyer Group
Missing: 35 (3%)
Distinct: 6
Type: Nominal
Unique: 1 (0%)

No.	Label	Count	Weight
1	PJ	127	127
2	Buyer	907	907
3	Laurel	56	56
4	Group 1	1	1
5	Suruchi Creations	8	8
6	Shri Anand Textiles	14	14

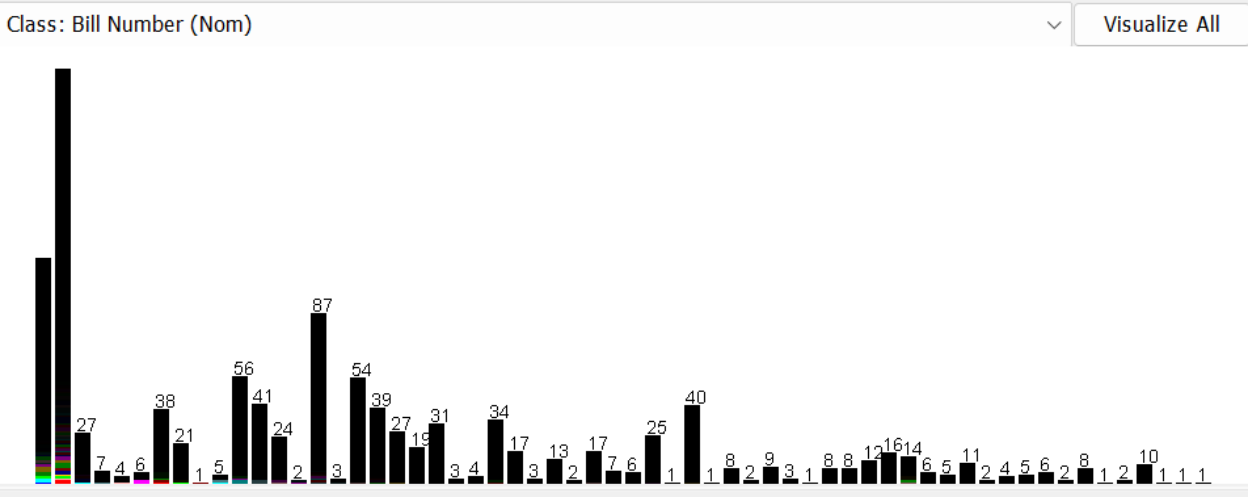


- **Buyer Name:** It is the Name of the Buyer.

Selected attribute

Name: Buyer Name
Missing: 1 (0%)
Distinct: 60
Type: Nominal
Unique: 8 (1%)

No.	Label	Count	Weight
1	PJ DRESSES	115	115
2	SHRI BALAJI ENTERPRISES	221	221
3	GOLDY CREATION	27	27
4	SRISHTI GARMENTS	7	7
5	YASH ENTERPRISES	4	4
6	KHETAN & SONS	6	6
7	KAMAL CUTPIECE CENTER	38	38
8	KHATRIJI COLLECTION	21	21
9	HARSH ENTERPRISES	1	1
10	AADI CREATIONS	5	5

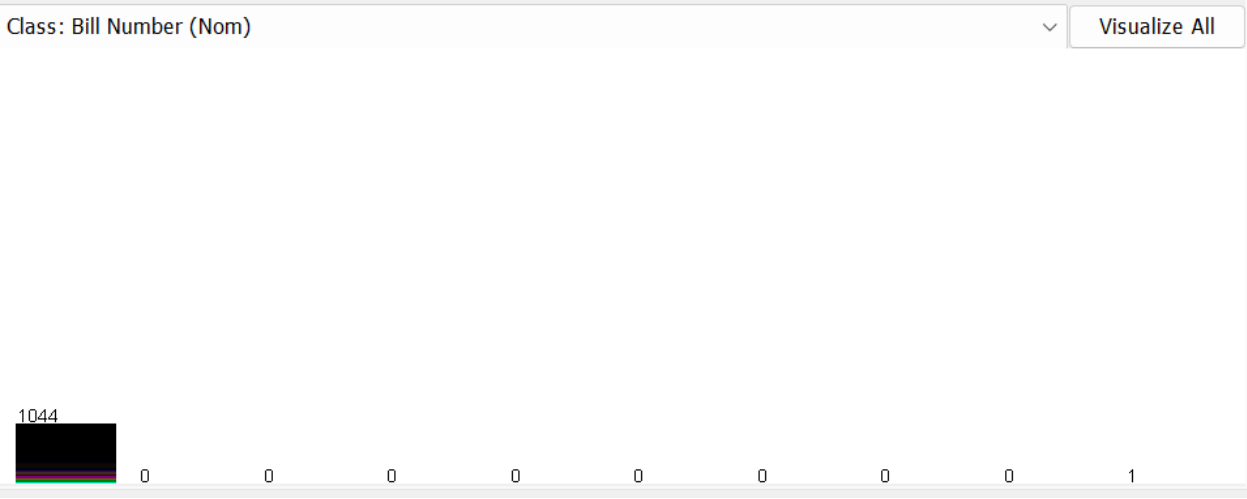


- **Bill Amount:** It is the Total bill amount for goods sold prepared by the supplier

Selected attribute

Name: Bill Amount
Missing: 103 (9%)
Distinct: 2
Type: Nominal
Unique: 1 (0%)

No.	Label	Count	Weight
1	'(-inf-6006194.171]'	1044	1044
2	'(6006194.171-12012246.342]'	0	0
3	'(12012246.342-18018298.513]'	0	0
4	'(18018298.513-24024350.684]'	0	0
5	'(24024350.684-30030402.855]'	0	0
6	'(30030402.855-36036455.026]'	0	0
7	'(36036455.026-42042507.197]'	0	0
8	'(42042507.197-48048559.368]'	0	0
9	'(48048559.368-54054611.539]'	0	0
10	'(54054611.539-inf)'	1	1



- **Payment Date:** It is the date on which the buyer has made payment.

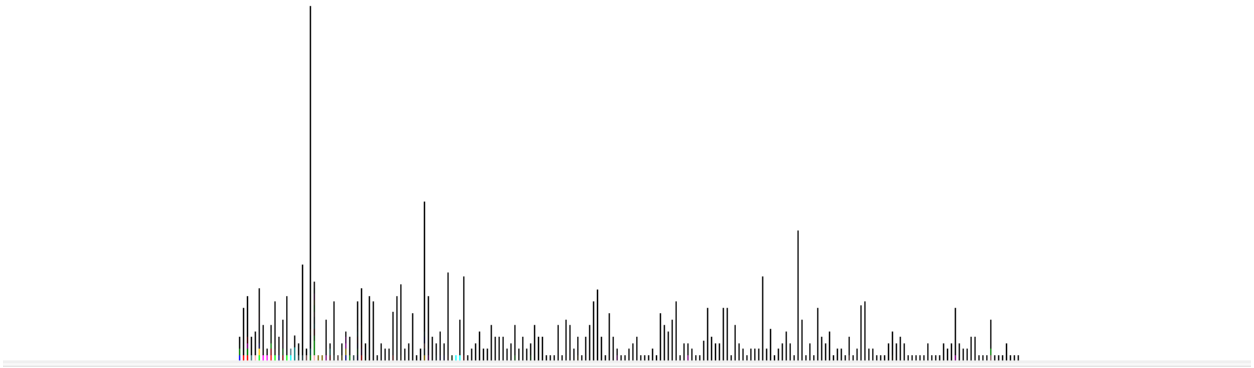
Selected attribute

Name: Payment Date
Missing: 250 (22%)
Distinct: 199
Type: Nominal
Unique: 53 (5%)

No.	Label	Count	Weight
1	01-04-2023	4	4
2	21-09-2023	9	9
3	14-12-2023	11	11
4	28-07-2023	4	4
5	24-07-2023	5	5
6	08-02-2024	12	12
7	27-05-2023	6	6
8	07-06-2023	2	2
9	17-05-2023	6	6
10	10-07-2023	10	10

Class: Bill Number (Nom)

Visualize All

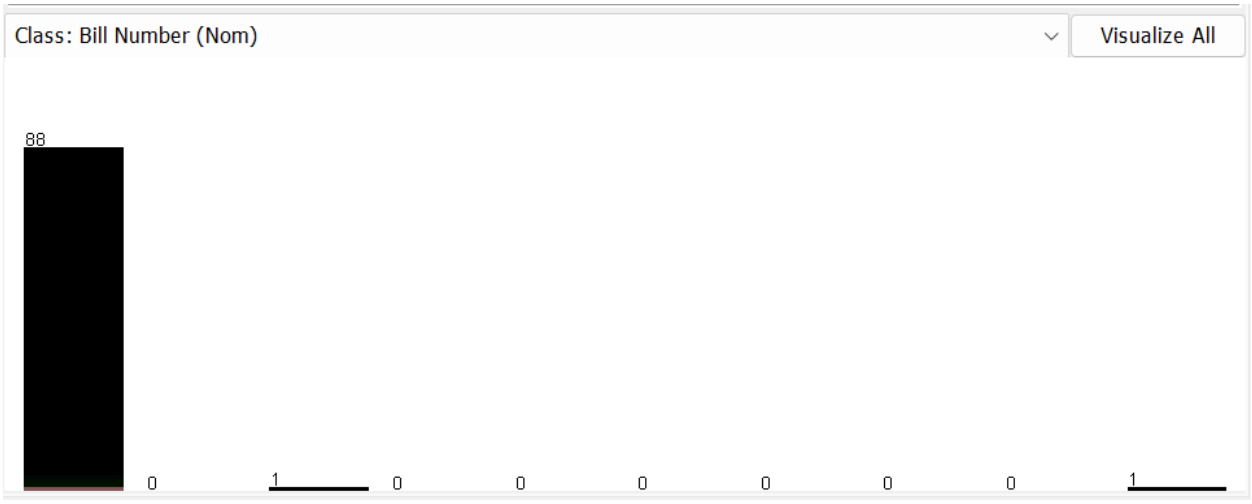


- **Discount:** It is amount of any discount provided to the buyer.

Selected attribute

Name: Discount
Missing: 1058 (92%)
Distinct: 3
Type: Nominal
Unique: 2 (0%)

No.	Label	Count	Weight
1	'(-inf-16438.127]'	88	88
2	'(16438.127-32876.004]'	0	0
3	'(32876.004-49313.881]'	1	1
4	'(49313.881-65751.758]'	0	0
5	'(65751.758-82189.635]'	0	0
6	'(82189.635-98627.512]'	0	0
7	'(98627.512-115065.389]'	0	0
8	'(115065.389-131503.266]'	0	0
9	'(131503.266-147941.143]'	0	0
10	'(147941.143-inf)'	1	1

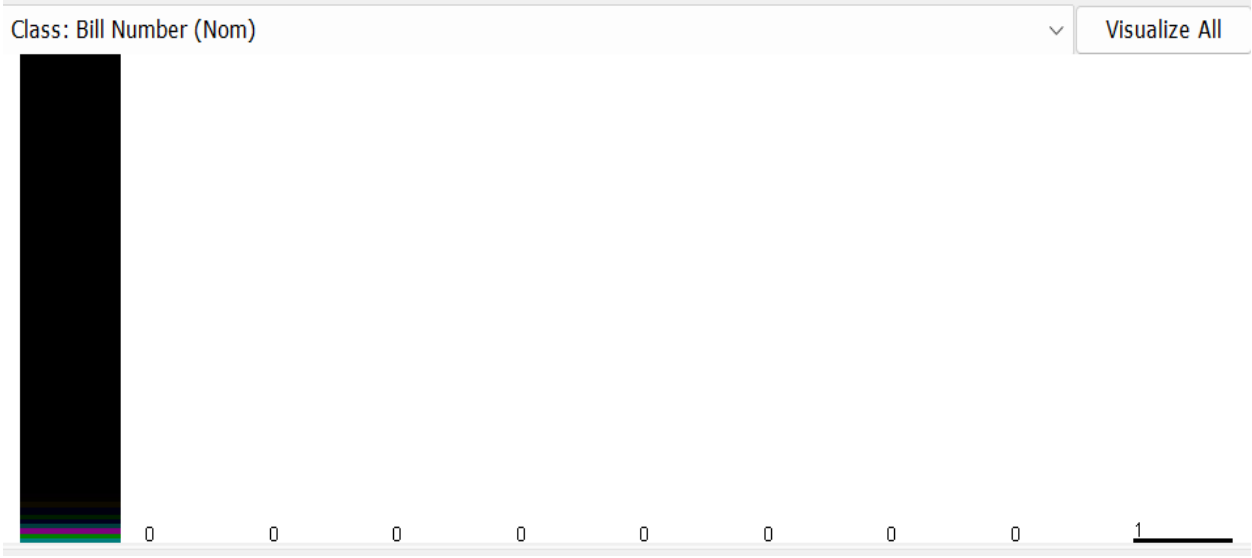


- **Goods Return:** It is amount of any Goods Returned by the buyer.

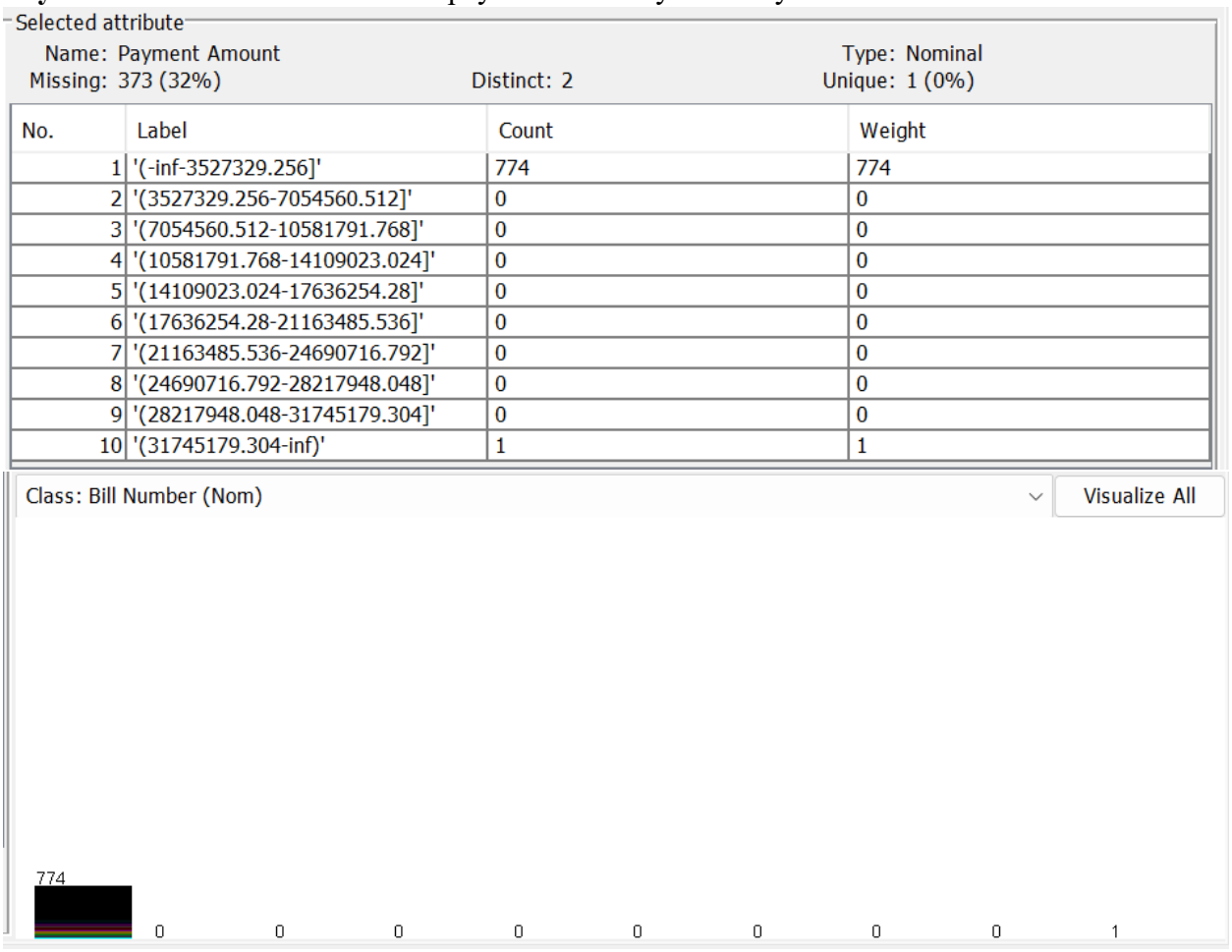
Selected attribute

Name: Goods Return
Missing: 1026 (89%)
Distinct: 2
Type: Nominal
Unique: 1 (0%)

No.	Label	Count	Weight
1	'(-inf-571303.2]'	121	121
2	'(571303.2-1141839.4]'	0	0
3	'(1141839.4-1712375.6]'	0	0
4	'(1712375.6-2282911.8]'	0	0
5	'(2282911.8-2853448]'	0	0
6	'(2853448-3423984.2]'	0	0
7	'(3423984.2-3994520.4]'	0	0
8	'(3994520.4-4565056.6]'	0	0
9	'(4565056.6-5135592.8]'	0	0
10	'(5135592.8-inf)'	1	1



- **Payment Amount:** It is amount of payment made by with buyer.



Model Selection

The Machine Learning Algorithms used in this project are :The Machine Learning Algorithms used in this project are :

• Cluster:

1. **SimpleKMeans:** As for the SimpleKMeans function that is also the clustering algorithm, this is implemented in Weka. It often behaves like a whip that separates the data into particular clusters based on similarities. The decision as to how to implement SimpleKMeans or any other clustering technique is dependent on the type of your data and the aims of your analysis.

Why Choose SimpleKMeans in Weka:Why Choose SimpleKMeans in Weka:

- **Simplicity:** SimpleKMeans algorithm is quite simple and basic, very intuitive, and also has a good user-friendly interface. It is providing a partitioning approach that assigns instances to a group based on their distance to group centres.

- **Speed:** SimpleKMeans is an algorithm that is practically fast and is capable of managing large data sets at a moderate computational cost.
- **Scalability:** It can operate a high number of instances and attributes, so that its application can be spun across a large range of uses.
- **Ease of Interpretation:** The interpretation of the results of SimpleKMeans are not that difficult, particularly since the cluster characteristics and centroids can also be easily understood.

• Associate:

2. **Apriori:** Apriori algorithms are widely used for mining association rules that are information rich patterns originating from huge data sets. Apriori makes frequent itemset discovery and the association rule thereof through the process oparated by this algorithm.

Why Choose Apriori Algorithm

- **Association Rule Mining:** Apriori is something that has been specifically woven for game for discovering association rules, that are patterns that expose relations among items in a dataset. It enables us to make a narrative out of data by observing the co-occurrence patterns.
- **Market Basket Analysis:** Also, Apriori has increasingly been used in market basket analysis, which main task is to find associations between the products that clients buy most typically together.
- **Simple and Intuitive:** Apriori algorithm is comparatively more easy and is conceptually quick. Hence, this algorithm is preferred for beginners in the association rule mining.

Model Training

• Cluster

SimpleKMeans : In Weka, the SimpleKMeans class is used to cluster based on k-means. Below is a step-by-step guide on how to train a k-means model using the `SimpleKMeans` algorithm in Weka. Below is a step-by-step guide on how to train a k-means model using the `SimpleKMeans` algorithm in Weka:

a) **Load Data:**
   a. a.Open Weka Explorer.
   b. Tap the "Open file" option and then it will open your dataset.

b) **Choose Clusterer:**
   a. Go to the "Cluster" tab in "Weka Explorer" by selecting that tab.
   b. For the `clusterer`, choose `SimpleKMeans` from the list of available clusterers.

c) **Set Options:**

a. The next step is to click on `SimpleKmeans` and hit the "Choose" button to its options.
b. Set number of number paramaters such as the `numClusters` ,other options like initialization method and seed.

d) Optional: Normalize Data:

a. The data processing that you will need to do may differ on the basis of the data characteristic you have. You may need to normalize or standardize the attributes as per the data feature you have. Access to this operators can easily be done in the area named 'Preprocess', which is situated in the tab.

e) Build Clusterer:

a. Upon the options have been considered, launch button to make clustering k-means.

f) View Results:

a. The pitfall is you can also see the clustered results in the "Cluster Assignments" section in the "Cluster" tab.
b. The panel will display the cluster assignments for the data instances in your dataset that you provide.

g) Evaluate the Model:

a. You can inspect the proportions of the distribution and use graphical and numeric metrics to perform the analysis. For evaluating the clustering quality, you need to see how the clustering assignment fits into the dataset and perform a comparison using clustering evaluation metrics.
b. Weka features silhouette score, davies-bouldin number, and other metrics used in clustering result evaluation.

- **Associate**

Apriori

In Weka programming, training a model is done in three steps: parameter settings, data set specification, then the launching of the program through Apriori algorithm. Here's a step-by-step guide on how to train an Apriori model in Weka:Through this process, the blueprint of the procedure will be defined:

a) Open Weka:

a. Launch the aviation simulator and set it up to study the likelihood of final products acceptance for human-machine interaction.

b) Load Dataset:

a. By clicking on the "Open File" button, which is a folder icon located in the top left corner, you will be directed to the file location.
b. Use dataset for selection, here association rules mining.

c) Choose Apriori Algorithm:

a. The first step for the user is to go to the "Profile tab" from the "Processing Panel" then choose "Associate" as the panel.

b. Therefore, the "Apriori" algorithm will be picked out from the list as it is among the association rule mining algorithms.

d) Set Algorithm Parameters:

a. Input a configuration of Apriori parameters.Common parameters include:Input a configuration of Apriori parameters.Common parameters include:
b. Minimum Support (`-N`): The minimum support frequency for a measured combination of things is a minimum amount that this combination of items must have been purchased at least to be needed by consumers.
c. Minimum Confidence (`-C`): The minimum confidence level allowed to produce associations that is computed by the algorithm known as the confidence threshold.
d. Maximum Number of Items (`-I`): As the set grows in size, the largest fragment number among its terms.

e) Run Apriori:

a. Pick “Start” button to be the Apriori algorithm launched.

f) Review Results:

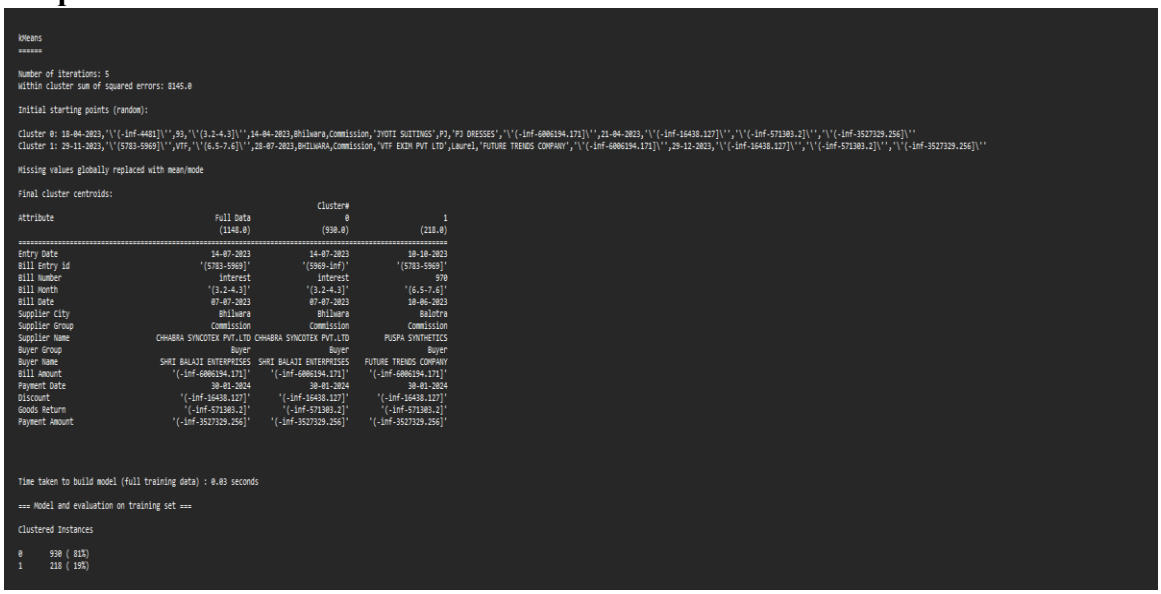
a. Algorithm will be tested until the "Classifier output" panel will appear and the result will be displayed.
b. Today I have cited just one instance of these successful digital marketing campaigns.

g) Evaluate Model:

a. If now you have completed preparations and you are going to test your model, you can take "Associating" option where you will have to establish more evaluations methods. On the other hand you can employ "Classify" option in order to apply your methods on novel examples again for another test.

Model Evaluation

- **Cluster:**
  - **I. SimpleKMeans:**



```
kMeans
======

Number of iterations: 5
Within cluster sum of squared errors: 8145.0

Initial starting points (random):

Cluster 0: 18-04-2023,'\'(-inf-4481]\'',93,'\'(3.2-4.3]\'',14-04-2023,Bhilwara,Commission,'JYOTI SUITINGS',P3,'P3 DRESSES','\'(-inf-6006194.171]\'',21-04-2023,'\'(-inf-16438.127]\'','\'(-inf-571383.2]\'','\'(-inf-3527329.256]\''
Cluster 1: 29-11-2023,'\'(5783-5969]\'',VTF,'\'(6.5-7.6]\'',28-07-2023,BHILWARA,Commission,'VTF EXIM PVT LTD',Laurel,'FUTURE TRENDS COMPANY','\'(-inf-6006194.171]\'',29-12-2023,'\'(-inf-16438.127]\'','\'(-inf-571383.2]\'','\'(-inf-3527329.256]\''

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                    Cluster#
Attribute                                Full Data          0                1
                                          (1148.0)     (930.0)          (218.0)
=============================================================================
Entry Date                              14-07-2023  14-07-2023       10-10-2023
Bill Entry Id                      '(5783-5969]'  '(5969-inf)'  '(5783-5969]'
Bill Number                             interest    interest              970
Bill Month                           '(3.2-4.3]'  '(3.2-4.3]'    '(6.5-7.6]'
Bill Date                             07-07-2023  07-07-2023       10-06-2023
Supplier City                            Bhilwara    Bhilwara          Balotra
Supplier Group                         Commission  Commission       Commission
Supplier Name          CHHABRA SYNCOTEX PVT.LTD CHHABRA SYNCOTEX PVT.LTD  PUSPA SYNTHETICS
Buyer Group                              Buyer       Buyer            Buyer
Buyer Name             SHRI BALAJI ENTERPRISES  SHRI BALAJI ENTERPRISES  FUTURE TRENDS COMPANY
Bill Amount              '(-inf-6006194.171]'  '(-inf-6006194.171]'  '(-inf-6006194.171]'
Payment Date                       30-01-2024  30-01-2024       30-01-2024
Discount                  '(-inf-16438.127]'  '(-inf-16438.127]'  '(-inf-16438.127]'
Goods Return               '(-inf-571383.2]'  '(-inf-571383.2]'  '(-inf-571383.2]'
Payment Amount           '(-inf-3527329.256]'  '(-inf-3527329.256]'  '(-inf-3527329.256]'



Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      930 ( 81%)
1      218 ( 19%)
```

Result

The output sample has been created from k-Means clustering applying to the dataset with 15 attributes and 1148 samples. The dataset actually is a valuable information source on bills which encloses the following attributes for example Entry Date, Bill Number, Supplier City, Buyer Name, Bill Amount, Payment Date, Discount, Goods Return, and Payment Amount.

Here is an interpretation of the key sections of the output:

Cluster Model:

- In the given set up, the k-Means clustering technique was implemented with the specified parameters.
- The algorithm has carried out 5 iterations, and the judgment is a wss of 8145. No changes necessary.
- The study revealed two distinguishable segments in the population (Cluster 0 and Cluster 1).

Initial Starting Points (Random):

- The centers of the clusters are initially marked, alignment being represented by those attributes that characterize the nearest centroid.
- Consider, these features comprise Cluster 0: entries date is on 14-07-2023, bill number is interest, the supplier city is Bhilwara etc.

Missing Values Globally Replaced:

- The ML defects were globally covered places, with mean/mode.

Final Cluster Centroids:

- The final centroids of the clusters as well as their attribute values are demonstrated with the centroid of each cluster's values that are highlighted in each cluster.
- All the factors are displayed with range for the entire process, Cluster zero and Cluster one.

Clustered Instances:

- Satisfyingly, the clusters have relatively been formed, with 930 instances (81%) and 218 instances (19%) belonging to Cluster 0 and Cluster 1 respectively.

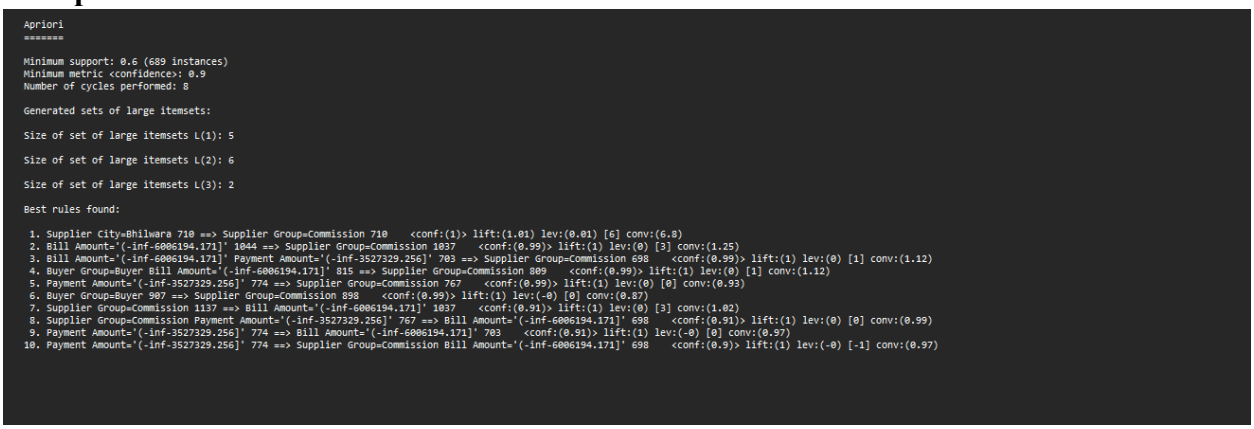
Interpretation:

- The clustering has grouped all instances having the same feature values into the group of records.
- Cluster 0 seems to include a higher number of instances than the later clusters, and the mean values do show the features of the instances from that cluster.
- In a similar fashion, Cluster 1 operates under values which are distinctly different to those of Cluster 2.

Time Taken:

- Time to build up the model with full training data is the one of the tracking item which is 0.03 seconds.

- **Associate**
  - **I. Apriori**



Result

Utilized here is the general purpose machine learning algorithm known as the Apriori algorithm, which is usually used for association rule mining. Relationship discovery, or finding novel patterns, is the main point of an association rule mining algorithm. Here the algorithm is being used for a particular 15 attribute dataset where the final dataset contains 1148 instances.

Here are some key parameters and results from the Apriori algorithm:Here are some key parameters and results from the Apriori algorithm:

- Minimum Support: lower than 0.6 (60%), which only describes the cases in at least 60% of instances.
- Minimum Confidence: The p-value they arrived at was 0.9, so the discovered correlations should be considered as only associative.
- Number of Cycles: In order to get a better understanding of the algorithm, I will dive deeper into this variable and list the number of iterations it went through: 8.

Generated sets of large itemsets:
- Size of set of large itemsets L(1): A vacancy of two years or more should be a mandatory condition to avoid conflicts of interest and ensure the proper running of the state's affairs
- Size of set of large itemsets L(2): All the proven technologies related to reduce air pollutions, leaks and emissions, have been improved in the past decades, so there is nothing that prevent us from enforcing regulation aiming at controlling pollutant emissions.
- Size of set of large itemsets L(3): Through cross-disciplinary studies such as biology, ecology, and even mathematics, it seeks answers to questions regarding the harmony of ecosystems, the spread and control of infectious diseases, and the evolutionary adaptations that allow species to thrive on our planet.

Best Rules Found:

- City provider, which is Bhilwara, is part of the Commission group that have a confidence of 1 (100%).
- The discontinuous value of both sides of the blade (-inf to -6006194.171] indicate that the Supplier Group will be 'Commission' and the certainty of accuracy will be at 0.99.
- If Bill Amount fell under the group '[(-inf,-6006194.171]' and Payment Amount fell under the cluster '[(-inf,-3527329.256]' them Supplier Group was 'Commission' with Confidence of 0.99.
- In case, if for Buyer group and range of Bill amount is less than or equal to ('-inf-6006194.171), then the Supplier group will be 'Commission' and the probability will be 80%.
- According to the IPL, if Payment Amount is in the range with '(-inf-3527329.256]', then the Supplier Group is 'Commission' with a confidence level of 0.99.
- Therefore the Supplier Group is 'Commission' with a confidence (c) of 0.99 if the Buyer Group is 'buyer'.
- If the attribute group 'Commission' has the value of 'Bill Amount' which lies in the possible range of '(-inf-6006194.171]' then resultant confidence value will be 0.91.
- For the Commission excepted supplier with a Payment Amount variable within '(-inf,3527329.256]', the Bill Amount variable will be within '(-inf,6006194.171)' range with the confidence 0.91.

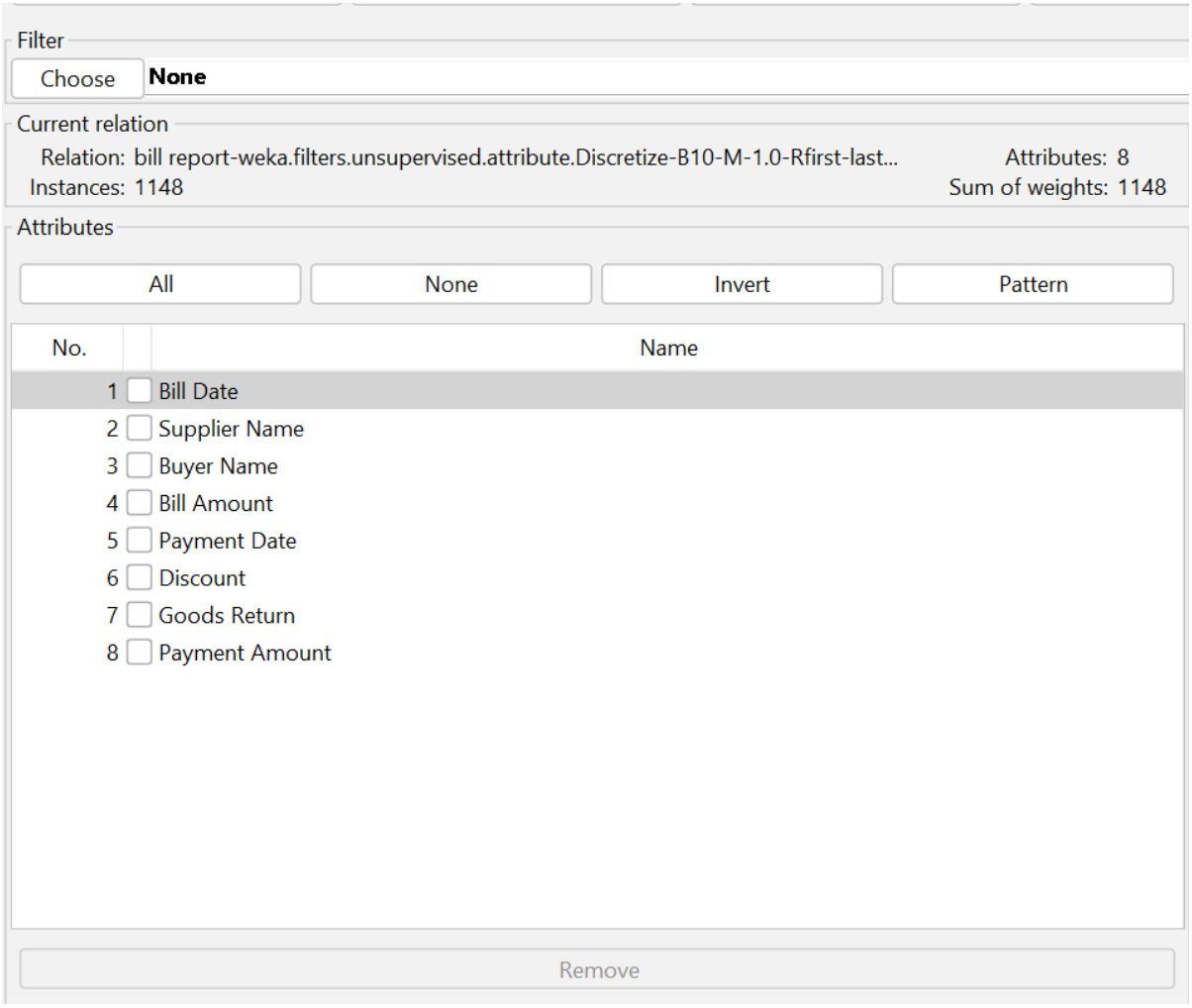
- If the Payment Amount is from a range '(-inf-3527329.256]', then Bill Amount is from '(-inf-6006194.171]' with a probability of 0.91.
- 10. If there are '-inf' to 3527329.256 in the Payment Amount, and the value of Supplier Group is 'Commission' then the Bill Amount is undefined minus 6006194.171 with confidence 0.9.

  These rules in the dataset are meant to show the close connections and dependencies between several different features giving an idea of the relationships between three things: Buyer Group, Payment Amount, Bill Amount, and the Supplier Group.

Analysis after Feature Selection

Introduction

In the Dataset, For Knowing the Accuracy that leads more value to the column various instances that did not lead more value to the dataset have been removed.



By applying the machine learning Algorithms, for knowing the accuracy of the dataset.

1. **Cluster**

- **SimpleKMeans:**

```
kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 3670.0

Initial starting points (random):

Cluster 0: 14-04-2023,'JYOTI SUITINGS','PJ DRESSES','\'(-inf-6006194.171]\'',21-04-2023,'\'(-inf-16438.127]\'','\'(-inf-571
Cluster 1: 28-07-2023,'VTF EXIM PVT LTD','FUTURE TRENDS COMPANY','\'(-inf-6006194.171]\'',29-12-2023,'\'(-inf-16438.127]\''

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                                  Cluster#
Attribute                                      Full Data                  0                     1
                                                (1148.0)             (1083.0)              (65.0)
=============================================================================================================
Bill Date                                     07-07-2023           07-07-2023            04-12-2023
Supplier Name                      CHHABRA SYNCOTEX PVT.LTD CHHABRA SYNCOTEX PVT.LTD      VTF EXIM PVT LTD
Buyer Name                          SHRI BALAJI ENTERPRISES  SHRI BALAJI ENTERPRISES  FUTURE TRENDS COMPANY
Bill Amount                        '(-inf-6006194.171]'     '(-inf-6006194.171]'     '(-inf-6006194.171]'
Payment Date                                30-01-2024           30-01-2024            30-01-2024
Discount                             '(-inf-16438.127]'       '(-inf-16438.127]'       '(-inf-16438.127]'
Goods Return                          '(-inf-571303.2]'        '(-inf-571303.2]'        '(-inf-571303.2]'
Payment Amount                     '(-inf-3527329.256]'     '(-inf-3527329.256]'     '(-inf-3527329.256]'
```

```
Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      1083 ( 94%)
1        65 (  6%)
```

Result

Thus, with the provision of information, the data has been divided (Group 0 and Group 1) into two groups (based on kMeans algorithm with preview parameters. Here is a summary of the clustering results: Here is a summary of the clustering results:

Cluster 0:

- Instances: 83% of the data is a reflection of the year 2023.

- Centroid:

- Bill Date: 07-07-2023

- Supplier Name: CHHABRA (CHHABRA) has produced fabrics for big-name brands such as Nike, Levi's, Marks & Spencer, H&M, and Mango, among others.

- Buyer Name: SHRI BALAJI TRADING, a name which stands for Business Acumen, Credibility and Sympathy towards our customers.

- Bill Amount: '(-inf-6006194.171]'

- Payment Date: 30-01-2024

- Discount: '(-inf-16438.127]'

- Goods Return: '(-inf-571303.2]'

- Payment Amount: '(-inf-3527329.256]'

Cluster 1:

- Instances: While only 65 (6% of the data set)

- Centroid:

- Bill Date: 04-12-2023

- Supplier Name: VTF EXIM PVT will also create employment opportunities for local residents and contribute to a more diverse economic base, attracting additional businesses and investment to the region.

- Buyer Name: FUTURE TRENDS COMPANY – FUTEX. We live in an era of constant innovation and technological advancement. As time unfolds, individuals' interests and needs change, creating niches to target different customer segments.

- Bill Amount: '(-inf-6006194.171]'

- Payment Date: 30-01-2024

- Discount: '(-inf-16438.127]'

- Goods Return: '(-inf-571303.2]'

- Payment Amount: '(-inf-3527329.256]'

There are some clusters of classes here which have similar attributes and features in common due to that feature. The model was constructed with a randomized starting point in the first iteration and successive algorithms.

You should keep in mind that this information is specific to the chosen clustering algorithm and on the particular dataset that you will use during this practice. The actual interpretation of these

clusters may change when you do an analysis that is related to your data, as well as the goals that you will set.

2. Associate

- **Apriori**

```
Apriori
=======

Minimum support: 0.1 (115 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 1

Best rules found:

 1. Supplier Name=CHHABRA SYNCOTEX PVT.LTD 270 ==> Bill Amount='(-inf-6006194.171]' 260    <conf:(0.96)> lift:(1.06) lev:(0.01) [14] conv:(2.22)
 2. Supplier Name=CHHABRA SYNCOTEX PVT.LTD Payment Amount='(-inf-3527329.256]' 231 ==> Bill Amount='(-inf-6006194.171]' 221    <conf:(0.96)> lift:(1.
 3. Payment Amount='(-inf-3527329.256]' 774 ==> Bill Amount='(-inf-6006194.171]' 703    <conf:(0.91)> lift:(1) lev:(-0) [0] conv:(0.97)
```

Result

The Apriori algorithm identified relationship pattern through the given data to generate the association rules. Here are the three best rules found:Here are the three best rules found:

Rule 1:
- Antecedent (IF): Company & Address=CHHABRA SYNCOTEX PVT. LTD.,224/742, Gandhi Nagar Delhi-110003
- Consequent (THEN): CThe bill amount='(-infinity-60'06194.17)'.
- Support: 270 instances
- Confidence: 0.96
- Lift: 1.06
- Leverage: 0.01
- Conviction: 2.22

Rule 2:
- Antecedent (IF): CHHABRA SYNCOTEX PVT.LTD is the supplier's name and the payment is –Lah ft3527329.256.
- Consequent (THEN): Bill('(-inf/'6013n294.171]');
- Support: 231 instances
- Confidence: 0.96
- Lift: 1.05
- Leverage: 0.01
- Conviction: 1.9

Rule 3:
- Antecedent (IF): Payment Amount(0, '(-inf-3527329.256]')
- Consequent (THEN): Bill Amount= '(–inf-6006194.171]'
- Support: 774 instances
- Confidence: 0.91
- Lift: 1
- Leverage: -0
- Conviction: 0.97

These rules, in effect, link the attributes from different records in the data-set. Such as, according to Rule 1 United States' prediction for range of unknown Bill Amount at CHHABRA SYNCOTEX PVT. LTD. company is "(-inf-6006194.171]" with 96% confidence. Also, Rule 2 is the case such that when the Supplier Name is "CHHABRA SYNCOTEX PVT.LTD" and the Payment Amount is within range "(-inf-3527329.256]," the result was 96% confident the Bill Amount was in range.

Conclusion

Both analyses draws on the use of different machine learning models that give the analysis an original outlook, thereby unveiling the patterns and trends within the data. Let's compare the key aspects of these Analyses. Let's compare the key aspects of these Analyses:

- **Clustering Results:**

• **Analyses 1 (Clustering):** The used SimpleKMeans algorithm brought about the detection of two clusters (Cluster 0 and Cluster 1), that had distinct centroid features. Cluster 0, which was the crowd in 94% of cases.

• **Analyses 2 (SimpleKMeans):** Using the K-means algorithm the results obtained were two clusters (one was Cluster 0 and the other was Cluster 1). In Cluster 0 there were 81% of cases and all the data about a centroid aggregate describes the centroids attributes.
Comparison: Clustering of the provided data was accomplished with the use of SimpleKMeans and is comprised of two clusters, by the same name chosen to portray the prevalence and the nature of instances within each cluster.

- **Association Rule Mining:**

• **Analyses 1 (Apriori):** The set of three most significant association rules is indicated, which feature the very strong associations between Supplier. Name, Payment amount and Bill amount.

• **Analyses 2 (Apriori):** The Apriori algorithm unearthed the top 10 rules where the attributes were involved in processes such as the bill amount - payment amount, supplier city - supplier group, and buyer group-supplier group.
Comparison: The Analytics is not only focused on Clustering and Association Rule Mining which both provides different insights but also making the decision-making more effective.

Synthesis
Despite similarities, Analysis 2 goes beyond Analysis 1 to simplify the process by using the Kmeans algorithm and permitting the user to discover more association rules. The dilemma in the division of the two frameworks might come from the analysis purpose and the dept of the insights. Combining the synergies between both these techniques could very well make the data comprehension better. Moreover, it will address the inherent bias that is caused by either approach alone.

References

- Gomes, L., Vale, Z., & Corchado, J. M. (2020). Microgrid management system based on a multi-agent approach: An office building pilot. Measurement, 154, 107427
- Das, R., Kephart, J. O., Lefurgy, C., Tesauro, G., Levine, D. W., & Chan, H. (2008, May). Autonomic multi-agent management of power and performance in data centers. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: industrial track (pp. 107-114).
- Zimmermann, R., & Paschke, A. (2003). PAMAS: Agent-based supply chain event management system. AMCIS 2003 Proceedings, 244.
- Heinrich, S., Durr, H., Hanel, T., & Lassig, J. (2005). An agent-based manufacturing management system for production and logistics within cross-company regional and national production networks. International Journal of Advanced Robotic Systems, 2(1), 2.
- Gazafroudi, A. S., Pinto, T., Prieto-Castrillo, F., Prieto, J., Corchado, J. M., Jozi, A., ... & Venayagamoorthy, G. K. (2017, June). Organization-based multi-agent structure of the smart home electricity system. In 2017 IEEE Congress on Evolutionary Computation (CEC) (pp. 1327-1334). IEEE.
- O'BRIEN, P. D., & Wiegand, M. E. (1998). Agent based process management: applying intelligent agents to workflow. The Knowledge Engineering Review, 13(2), 161-174.