بسم الله الرحمن الرحيم

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا

ماجستير علوم الحاسوب

العنوان

تنقيب بيانات المحلات التجارية بواسطة قاعدة الارتباط (خوارزمية البداهة)
دراسة حالة: محلات تجارية بمدينة ام درمان

Shops Mining Using Association Rule(A priori Algorithm)

Case Study: Omdurman City Shops

هذه الاطروحة مقدمة لنيل درجة الماجستير في علوم الحاسوب

اعداد الطالب :

الجيلي كباشي إبراهيم سليمان

المشرف:

د. محمد الحافظ مصطفي موسي زين العابدين

يوليو 2017م

الإستهلال

(آيـــــــــة)

(اللّهُ لاَ إِلَـهَ إِلاَّ هُوَ الْحَيُّ الْقَيُّومُ لاَ تَأْخُذُهُ سِنَةٌ وَلاَ نَوْمٌ لَّهُ مَا فِي السَّمَاوَاتِ وَمَا فِي الأَرْضِ مَن ذَا الَّذِي يَشْفَعُ عِنْدَهُ إِلاَّ بِإِذْنِهِ يَعْلَمُ مَا بَيْنَ أَيْدِيهِمْ وَمَا خَلْفَهُمْ وَلاَ يُحِيطُونَ بِشَيْءٍ مِّنْ عِلْمِهِ إِلاَّ بِمَا شَاء وَسِعَ كُرْسِيُّهُ السَّمَاوَاتِ وَالأَرْضَ وَلاَ يَؤُودُهُ حِفْظُهُمَا وَهُوَ الْعَلِيُّ الْعَظِيمُ)

الإهداء

إلى من تجرعوا الكاس فارغا ليسقياني قطرة حبر

الي من كلت انـاملهم ليقدموا لنـا لحظة سعادة

الي من حصدوا الاشواك عن دربي ليمـهدوا لي طريق العلم

(امي ،ابي،خالتي)

الي من احببتهم حبا لو مر علي ارض قاحلة لتفجرت منـها ينـابيع المحبة

(إخوتي واخواتي)

إلى من هم أقرب إلي من روحي ، إلى من أستمد منـهم عزيمتي وإصراري

(أصدقائي)

إلى من تحلو بالآخاء وتميزوا بالوفاء والعطاء الي من عرفت كيف اجدهم وعلمونـي ألا أضيعهم

(زملاء العمل بكلية الصديق أحمد البشير)

اهدي ثمرة جهدي المتواضع

الشكر والتقدير

الحمد لله من قبل ومن بعد، حمداً كما ينبغي لجلال وجه وعظيم سلطانه، والصلاة والسلام على أشرف خلق الله سيدنا وحبيبنا ورسولنا محمد صل الله عليه وسلم.

أتقدم بالشكر والتقدير لكل من مد يد العون لي، وساعدني في إنجاز هذا البحث المتواضع وأخص بالشكر والتقدير إلى المشرف د. محمد الحافظ على ما بذله من جهد وأشرفه على هذه الرسالة والذي لم يبخل على بوقته وعلمه لإتمام هذا العمل.

ولا أنسى أصحاب المحلات التجارية (هبة، الدومة، شقلبان، الأمين، البخيت) الذين أخذت منهم قاعدة البيانات اللازمة للبحث.

وكل الشكر والتقدير لجامعة السودان للعلوم والتكنولوجيا.

أسأل الله أن يجزي الجميع عني خير الجزاء وأن يوفقهم لما يحب ويرضى

مستخلص

الشركات في الوقت الحالي غنية بكميات هائلة من البيانات ولكنها ضعيفة في المعلومات المستخرجة من البيانات. وتعتبر هذه البيانات الضخمة مصدرا قيما، ورغم أن مفهوم تنقيب البيانات لا يزال جديدا، فإن العديد من الشركات الدولية أصبحت تعتمد عليها لاتخاذ قرارات استراتيجية. قواعد الارتباط هي واحدة من مجالات التطبيق الرئيسية لتعدين البيانات، ونظرا لمجموعة من معاملات العملاء على المنتجات، والهدف من ذلك هو العثور على الارتباطات بين مبيعات السلع. باستخدام تحليل سلة الزبائن يمكننا معرفة المنتجات الاساسية التي يتم شراؤها معا من قبل الزبائن، والتي لا يتم شراها معا لا في بعض الاحيان، ويمكن الاستفادة منها في تخطيطات مخزن أفضل، وتحديد أنماط شراء العملاء. الهدف الرئيسي من هذه الرسالة هو تحديد أنماط شراء العملاء (من محلات السوبر ماركت) في السودان، ولاية الخرطوم ، محلية أم درمان.حيث تتمثل مشكلة البحث في كيف يتم معرفة انواع السلع التي يشتريها المستهلك معاً للاستفادة منها في تحقيق رغبات الزبائن ولتحقيق أهداف البحث استخدمت قواعد الارتباط (خوارزمية a priori) على بيانات المعاملات تم جمعها خصيصا لهذه الدراسة، وهذه البيانات التي تم جمعها من المحلات التجارية تستخدم أنظمة محوسبة والمحلات التجارية الأخرى التي تستخدم نظام البيع والشراء اليدوي. وتزودنا بيانات المعاملات المجمعة هذه بمعلومات قيمة عن الوقائع المشتركة والمشتريات المشتركة للمنتجات. ويمكن استخدام هذه المعلومات كأساس لاتخاذ قرارات بشأن نشاط التسويق مثل الدعم الترويجي ومراقبة المخزون والحملات عبر البيع. وفي نهاية هذا البحث قام الباحث باكتشاف وتحديد أفضل

24 نمط شراء سوداني عند الشراء من محلات السوبر ماركت. ومن النتائج المتحصل عليها وجد الباحث ان اللحمة والدقيق والخبز اكثر السلع مباعا مع بعضها ومن السلع الأساسية لزبائن المنطقة.

Abstract

Companies are currently rich in huge amounts of data but are weak in information extracted from data. This massive data is a valuable resource. Although the concept of data mining is still new, many international companies are relying on it to make strategic decisions. Association rules are one of the main application areas for data mining, given a number of customer transactions on products, the goal is to find association between commodity sales. Using basket analysis customers I can find out the basic products that are purchased together by customers and identifying customer buying patterns. The main objective of this research is to identify customer purchase patterns (from supermarkets) in Sudan, Khartoum state, Omdurman locality. Where the problem of research is how to know the types of goods that the consumer buys together to take advantage of them in the fulfillment of the wishes of customers To achieve the research objectives, the a priori algorithm has been used to extract buying patterns from the collected data. These data collected from shops which has computerized systems as well as from others which use manual systems. These combined transaction data provide us with valuable information on common facts and joint product purchases. This information can be used as a basis for making decisions on marketing activity such as promotional support, inventory control, and cross-selling campaigns. At the end of this research, the researcher discovered and identified the best 24 Sudanese purchase patterns when buying from supermarkets. Among the results obtained, the researcher found that meat, flour and bread are the most items which purchased together.

فهرس المحتويات

الاستهلال..... ب
الاهداء..... ت
الشكر والتقدير..... ث
مستخلص..... ج
Abstract..... خ
فهرس المحتويات..... د
فهرس الجداول..... ش
فهرس الاشكال..... ص

الباب الأول : المقدمة

1.1 المقدمة 1
2.1 مشكلة البحث..... 1
3.1 منهجية البحث
.....2 4.1 أهمية
البحث..... 2
5.1أهداف البحث 3
6.1 مكان البحث وأدواته اللازمة 3
7.1 أداة جمع البيانات 3
8.1 الأدوات اللازمة 3
9.1 هيكلية البحث 4

الباب الثاني : ادبيات البحث والدراسات السابقة

الفصل الأول :ادبيات البحث

.....5

1.1.2 المقدمة..... 5

2.1.2 تنقيب البيانات..... 5

3.1.2 أسباب تنقيب البيانات 6

4.1.2 خطوات استكشاف المعرفة من قواعد البيانات..... 7

1.4.1.2 اختيار البيانات 8

2.4.1.2 تهيئة البيانات 8

3.4.1.2 نقل البيانات 8

4.4.1.2 تنقيب البيانات..... 8

5.4.1.2 تقييم النموذج..... 8

6.4.1.2 تمثيل المعرفة..... 8

5.1.2 انوع تنقيب البيانات..... 9

1.5.1.2 التنقيب الوصفي 9

2.5.1.2 التنقيب التنبؤي..... 9

6.1.2 مهام تنقيب البيانات..... 10

1 .6.1.2 العنقدة..... 10

2.6.1.2 اكتشاف التسلسل..... 10

3.6.1.2 التصنيف.....,..... 10

4.6.1.2 الانحدار..... 11

5.6.1.2 التلخيص..... 11

6.6.1.2 قاعدة الارتباط..... 11

7.1.2 مقاييس قواعد الارتباط..... 11

1.7.1.2 Support..... 12

2.7.1.2 Confidence 13

8.1.2 مجتمع وعينة البحث..... 14

9.1.2 مقدمة عن weka..... 14

1.9.1.2 صيغة ARFF..... 14

2.9.1.2 طريقة استخدام weka..... 16

10.1.2 خوارزمية a priori..... 17

11.1.2 مثال يوضح طريقة عمل خوارزمية a priori..... 19

الفصل الثاني:الدراسات السابقة..... 24

1.2.2 دراسة: 2012**Rashmi Jha**..... 24

2.2.2 دراسة :2Phani Prasad J1, Murlidher Mourya..... 26

3.2.2 دراسة : Mehmet Aydin ulas..... 26

4.2.2 دراسة: Roshan Gangurde1, Binod Kumar2, S. D. Gore 3……28

5.2.2 دراسة: Prof. Kavitha Venkatachari……29

الباب الثالث: المنهجية ومجموعة البيانات والأدوات المستخدمة

1.3 مقدمة 30

2.3 اختيار البيانات..... 30

3.3 تهيئة البيانات..... 31

4.3 تنقيب البيانات 31

5.3 وصف البيانات والعمليات التي تمت عليها..... 32

1.5.3 وصف المحلات التجارية..... 32

2.5.3 العمليات علي البيانات.....37

1.2.5.3 الخطوة الاولي دمج البيانات 37

2.2.5.3 الخطوة الثانية انشاء جدول.....46

3.2.5.3 الخطوة الثالثة إضافة أعمدة للجدول..... 47

4.2.5.3 الخطوة الرابعة تحويل البيانات.....48

5.2.5.3 الخطوة الخامسة جمع السلع..... 53

6.2.5.3 الخطوة السادسة عملية تحويل البيانات الي 0،1..... 53

7.2.5.3 الخطوة السابعة عملية الاستبدال..... 55

8.2.5.3 الخطوة الثامنة تحويل البيانات الي صيغة csv..... 56

6.3 احصائيات البيانات..... 57

7.3 طريقة تجهيز البيانات لتعمل علي برنامج weka..... 63

1.7.3 تهيئة البيانات (weka) 42

2.7.3 تحميل البيانات في المستكشف..... 64

3.7.3 الخطوة الثانية..... 65

الباب الرابع: النتائج ومناقشتها

1.4 مقدمة..... 67

2.4 النتائج 68

3.4 مناقشة النتائج..... 72

الباب الخامس: التوصيات والخاتمة

1.5 التوصيات..... 74

2.5 الخاتمة..... 74

المراجع..... 75

فهرس الجداول

جدول (1.2) : مثال لحساب support..... 12

جدول (2.2) : مثال لحساب confidence..... 13

جدول (1.4) : تجارب support و confidence..... 68

جدول (2.4) : جزء من النتائج المتحصل عليها..... 72

فهرس الاشكال

شكل (1.2) يضح خطوات استكشاف المعرفة من قواعد البيانات.....7

شكل (2.2) يوضح أنواع تنقيب البيانات.....9

شكل (3.2) يوضح طريقة عمل خوارزمية a priori.....18

شكل (1.3) يوضح بيانات سوبر ماركت هبة في المرحلة الاولي.....34

شكل (2.3) يوضح اكثر سلعة 24 مباعا.....35

شكل (3.3) يوضح قوة العلاقة بين السلع.....36

شكل (4.3) يوضح طريقة جمع السلع المتشابه في اسم واحد.....45

شكل (5.3) يوضح عملية اشاء جدول جديد ready data.....46

شكل (6.3) يوضح عملية إضافة 24 عمود للجدول الجديد.....47

شكل (7.3) يوضح طريقة تحويل البيانات.....52

شكل (8.3) يوضح جمع السلع من كل الصفوف ووضعها في صف واحد.....53

شكل (9.3) يوضح عملية تحويل البيانات الي 1،0 54

شكل (10.3) يوضح عملية تحويل 0الي not buy و 1 الي buy.....55

شكل (11.3) يوضح شكل البيانات بعد تحويلها الي صيغة csv.....56

شكل (12.3) يوضح مخططات البيع واحصاء البيانات.....63

شكل (13.3) يوضح واجهة الاختيار الرسومية ل weka.....64

شكل (14.3) يوضح واجهة المستكشف في weka.....65

شكل (15.3) يوضح شكل النتائج في weka.....66

الباب الأول

مقدمة

1.1المقدمة

ازدادت حاجة مؤسسات الأعمال في عالم اليوم إلي المعلومات الضرورية لصنع القرارات بفعالية مما فرض علي تلك المؤسسات التوجه لتخزين بياناتها بأمان للتمكن من استرجاعها ومن ثم معالجتها للحصول علي المعلومة المطلوبة لاستخدامها في دعم القرارات المطلوبة في جميع انشطة المؤسسة وعلي اختلاف المستويات الادارية فيها اذ تعد المعلومات التي تمتلكها المؤسسات ثروة ويجب الاستفادة منها وتنقبها للوصل للمعرفة ومن خوارزميات تنقيب البيانات والتي سوف يعتمد عليها الباحث في هذا البحث خوارزمية قواعد الارتباط Associations Rule هي احد الواجهات الواعدة من Data Mining كأداة من ادوات اكتشاف المعرفة knowledge discovery in databases ولديها القدرة على تصفح كميات هائلة من البيانات ، وهي تسمح باستخراج القوانين الممكنة التي تشرح بعض الصفات الموجودة اعتمادا على وجود الصفات الاخرى. بمعنى اخر تساعد هذه الأدوات في توليد وإيجاد قواعد ارتباطية معينة بين مجموعة من البيانات المخزنة مسبقاً.

2.1 مشكلة البحث

أن الواقع الاقتصادي للمتاجر جعلها تواجه حقيقة تتمحور في اهمية البحث عن ادوات تسهل عملية الحصول الفعال على تحليل ونتائج تزيد من ارباحها للكشف عن معرفة جديدة وذلك عن طريق

تحليل الكميات الهائلة من البيانات الموجودة في قاعدة البيانات مسبقاً ويمكن تحديد مشكلة البحث من الاتي: كيف يتم معرفة انواع السلع التي يشتريها المستهلك معاً للاستفادة منها في تحقيق رغبات الزبائن بطريقة أفضل وتحليل لسلة المستهلك.

3.1 منهجية البحث

جمع البيانات اللازمة من مصادرها الثانوية والاولية، وذلك على النحو التالي:

أولا: البيانات الثانوية (المنشورة) : تم جمعها من الكتب، والدراسات السابقة في المجلات المتخصصة ،والدراسات العليا ،والدراسات والأبحاث المنتشرة علي شبكة الانترنت.

ثانيا: البيانات الأولية: قام الباحث بجمعها من قاعدة بيانات اثنين سوبرماركت وباقي البيانات من كراسات ديون ثلاث بقالات.

وتمثل منهجية تحليل البيانات في استخدام المنهج التنبؤي حيث تم استخدام برنامج weka لاختبار التجارب والوصول للنتائج.

4.1 أهمية البحث

تساعد مثل هذه الدراسة إدارات المتاجر على الاطلاع على التقنيات الحديثة التي تستخدمها مؤسسات الأعمال في ترشيد القرارات التسويقية ودعمها لأغرا ض تحقيق المزايا التنافسية في الأسواق المحلية.

5.1 أهداف البحث

يمكن تحديد أهم الأهداف التي يسعى البحث الى تحقيقها بالآتي :

1- فهم سلوك المستهلك في جزء من منطقة ام درمان .

2- معرفة أنواع السلع الذي يوجد بينها ارتباط في المبيعات والتي تباع معا.

6.1 مكان البحث والادوات اللازمة

مكان البحث السودان أم درمان الثورة مرفق خريطة في الملاحق

7.1 اداة جمع البيانات

اعتماد اً على أهداف البحث تم الحصول على البيانات من قاعدة بيانات النظام المصمم لمتجرين وثلاثة متاجر من كراسات الديون

8.1 الادوات اللازمة

1.برنامج(WEKA)

2.بيانات(data)من محلين تجاريين وكراسات ديون لثلاثة متاجر.

9.1 هيكلية البحث

يتكون البحث من أربعة أبواب، الباب الأول المقدمة تناول الاطار لعام للبحث وتكون من المقدمة، مشكلة الدراسة، أهمية الدراسة، اهداف الدراسة، منهجية الدراسة، هيكل الدراسة.

اما الباب الثاني ادبيات الدراسة والدراسات السابقة فقد اشتمل على فصلين الفصل الأول تحدث عن مقدمة عن تنقيب البيانات، أسباب تنقيب البيانات، استكشاف المعرفة من قواعد البيانات، انواع تنقيب البيات، مجتمع وعينة البحث، خوارزمية (a priori) . وعرض الفصل الثاني الدراسات السابقة.

وتناول الباب الثالث المنهجية ومجموعة البيانات والبرامج المستخدمة والعمليات على البيانات وتهيئتها وتناول الباب الرابع الذي عرض النتائج ومناقشتها وأخيرا الباب الخامس تناول التوصيات والخاتمة والمراجع والملاحق.

الباب الثاني

ادبيات البحث والدراسات السابقة

الفصل الأول: أدبيات البحث

1.1.2 المقدمة

اشتمل هذا الباب على فصلين الفصل الأول تحدث عن مقدمة عن تنقيب البيانات، أسباب تنقيب البيانات، استكشاف المعرفة من قواعد البيانات، انواع تنقيب البيات، مجتمع وعينة البحث، خوارزمية (a priori) . وعرض الفصل الثاني الدراسات السابقة.

مع كثرة البيانات الموجودة والمخزونة فيما يسمي بقواعد البيانات ، اصبحت موضوع تساؤل عديد من الباحثين للاستفادة منها، ومع زيادة انتشار مستودعات التخزين الضخمة اصبح من الضروري إيجاد تقنيات وطرق ووسائل لاستخلاص المعلومات والمعرفة من مثل هذه البيانات المكدسة واستغلالها في حل المشاكل واتخاذ القرارات، باستخدام تطبيقات الحاسوب الحديثة والتي تعتبر تكنلوجيا حديثة ذكية قائمة علي جعل الحاسوب يفكر كما يفكر الانسان ويفعل الانسان " وهو ما يعرف الذكاء الاصطناعي، جاءت فكرة الكشف والتنقيب علي هذه البيانات بطرق ذكية للمساعدة في حل المشاكل واتخاذ القرارات . تنقيب البيانات هو استخراج المعرفة من البيانات، وتحليل البيانات، وتعتبر خطوة من خطوات استكشاف المعرفة من قواعد البيانات. [1]

2.1.2 تنقيب البيانات

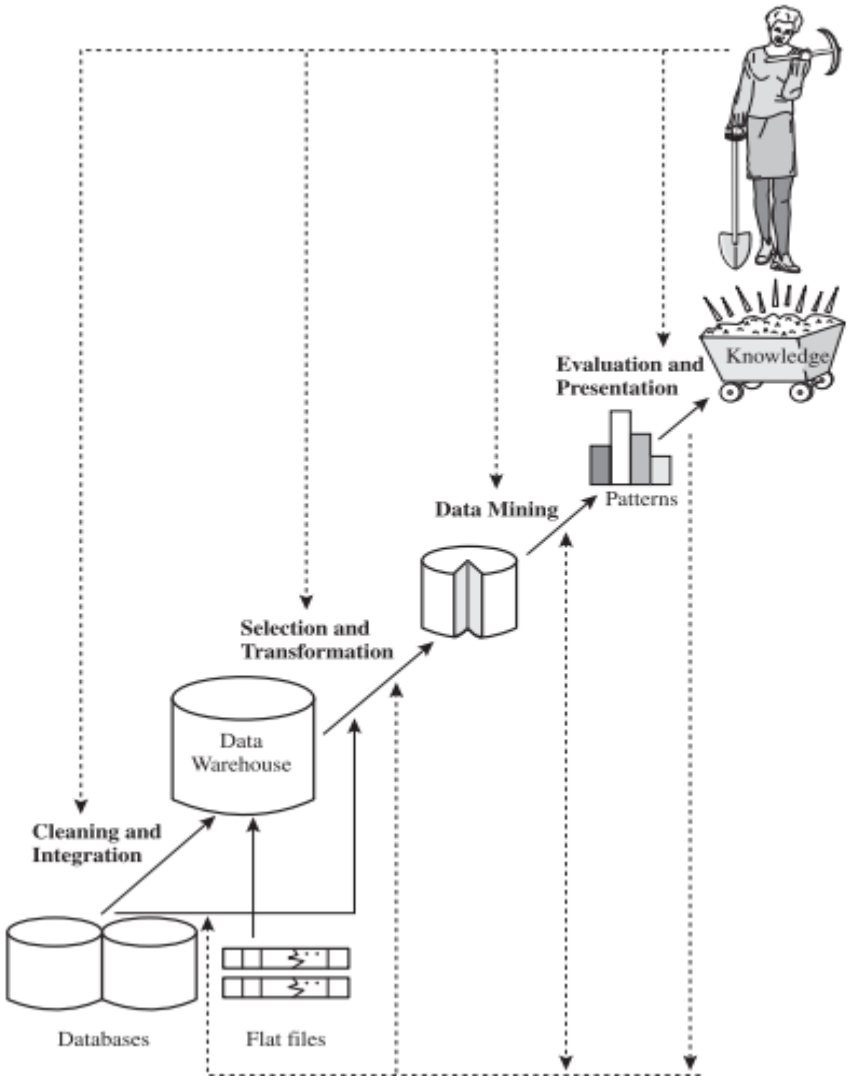
هي عملية بحث عن معرفة من البيانات دون فرضيات مسبقة. وكما يعرف التنقيب في البيانات على أنه عملية تحليل كمية بيانات عادة ما تكون كمية كبيرة لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات. يطبق اسم (نموذج) على العلاقات والبيانات

الملخصة التي يتم الحصول عليها من التنقيب في البيانات. يتعامل تنقيب البيانات عادة مع بيانات يكون قد تم الحصول عليها بغرض غير غرض التنقيب في البيانات (مثلاً قاعدة بيانات التعاملات في مصرف ما) مما يعني أن طريقة التنقيب في البيانات لا تؤثر مطلقاً على طريقة تجميع البيانات ذاتها. هذه هي أحد النواحي التي يختلف فيها التنقيب في البيانات عن الإحصاء، ولهذا يشار إلى عملية التنقيب في البيانات على أنها عملية إحصائية ثانوية. يشير التعريف أيضاً إلى أن كمية البيانات تكون عادة كبيرة، أما في حال كون كمية البيانات غيرة فيفضل استخدام الطرق الإحصائية العادية في تحليلها، وتعتبر عملية تنقيب البيانات عملية تحليله والبحث في بيانات ضخمة وهائلة لاستخراج انماط مفيدة وإيجاد العلاقات ومدى الارتباط بين عناصرها . [1] .

3.1.2 أسباب تنقيب البيانات

إن التنقيب في البيانات اصبح تكنولوجيا مفيدة و مهمة في ظل التطوير السريع وانتشار واستخدام قواعد البيانات كما أن استخدامها يوفر للمؤسسات وأجهزة الامن في جميع المجالات القدرة علي الاستكشاف والتركيز علي أهم المعلومات في قواعد البيانات. تركز تقنيات التنقيب علي بناء تنبؤات مستقبلة واستكشاف السلوك والاتجاهات، مما يسمح بتقدير قرارات صحيحه واتخاذها في الوقت المناسب، غير انها تجيب علي العديد من الأسئلة في وقت قياسي خاصة تلك النوعية من الاسئلة التي يصعب الإجابة عليها، ان لم يكن مستحيلا، باستخدام تقنيات الاحصاء الكلاسيكية والتي كانت قد وجدت تستغرق وقتا طويلا والعديد من إجراءات التحليل.

4.1.2 خطوات استكشاف المعرفة من قواعد البيانات



شكل(1.2) يوضح خطوات اكتشاف المعرفة من قواعد البيانات[1]

1.4.1.2 إختيار البيانات (Data Selection)

في هذه المرحلة يتم تحديد واسترجاع البيانات الملائمة علي حسب نوع المعلومات المطلوب استخراجها من قاعدة البيانات

2.4.1.2 تهيئة البيانات (Data Preprocessing)

وهي مرحلة تجهيز وعزل البيانات المبهمة أو المفقودة أو التي تحتوي على شوائب من بقية البيانات كإلغاء المعلومات المتكررة، التصحيح الشكلي، معالجة البيانات الناقصة وجعلها جاهزة للتطبيق وتشمل هذه المرحلة (تنظيف البيانات ،إزالة البيانات المفقودة، اشتقاق البيانات، دمج البيانات)

3.4.1.2 نقل البيانات (Data Transformation)

وهي عملية نقل البيانات التي تم اختيارها الى شكل ملائم لإجراءات البحث

4.4.1.2 تنقيب البيانات (Data Mining)

في هذه المرحلة سيتم تطبيق أسلوب ذكي لاستخراج نماذج مفيدة بقدر الإمكان

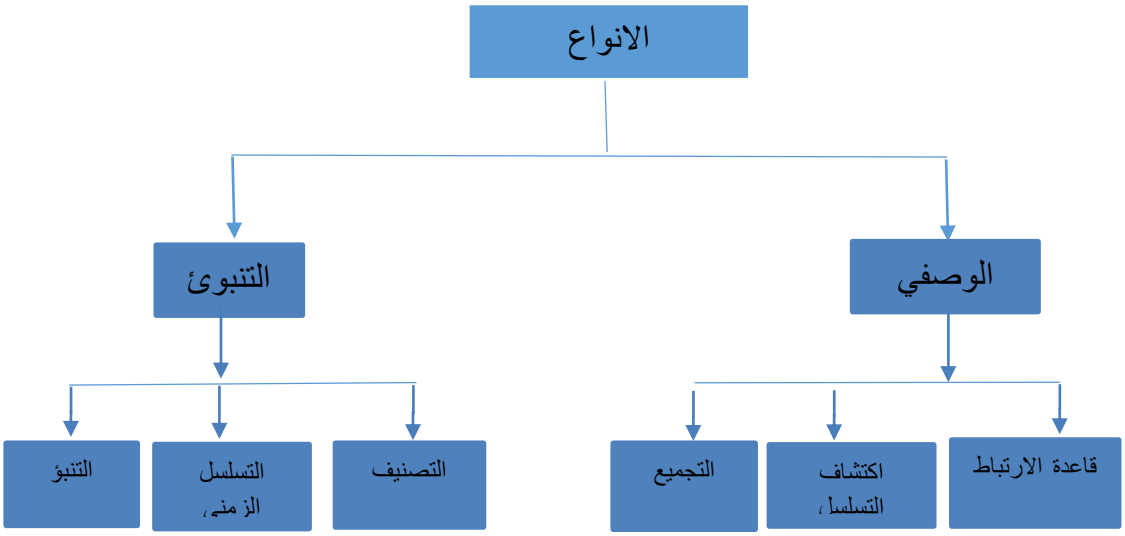
5.4.1.2 تقييم النموذج (Pattern Evaluation)

بعد استخراج النماذج المهمة التي تمثل المعرفة يتم تقييم هذه النماذج

6.4.1.2 تمثيل المعرفة (Knowledge Representation)

هي المرحلة الأخيرة من مراحل اكتشاف المعرفة في قواعد البيانات والتي يراها المستفيد، وهي المرحلة الأساسية والتي تستخدم أسلوب مرئي لمساعدة المستفيد في فهم وتفسير النتائج المستخرجة. [1]

5.1.2 أنواع تنقيب البيانات



شكل(2.2) يوضح انواع تنقيب البيانات [7]

1.5.1.2التنقيب الوصفي

يعتمد علي إعادة تنظيم البيانات، والتنقيب في أعماقها لاستخراج النماذج الموجودة فيها كتشابه الزبائن الذي يسمح لك بإنشاء وصف بسيط عن مجموعه زبائن متشابهين ولا يستوجب وجود هدف لمثل ذلك التشابه.

2.5.1.2 التنقيب التنبؤي

يحاول إيجاد أفضل التنبؤات اعتمادا من المعطيات كمعرفه المنتج الأفضل لزبون معين ، باختصار يعتمد هذا التنقيب على استخدام معلومات قديمة لتوقع ما سيحدث في المستقبل وتكون لدى مثل هذه البيانات هدف.

6.1.2 مهام تنقيب البيانات

من مهام تنقيب البيانات العنقدة واكتشاف التسلسل والتصنيف والانحدار والتلخيص وقاعدة الارتباط وسوف نشرح كل مهمة علي حدة.**[7]**

1.6.1.2 العنقدة

الهدف منها هو تحديد الاتجاهات داخل البيانات ، ويحاول هذا الاسلوب العثور على مجموعات من العناصر التي توجد عادة معا. وهي عمليه تقسيم البيانات إلى مجموعه من الأصناف إعتماداً على تشاركها في الخواص المتشابهة وأن العنقدة هي تقسيم غير موجه للبيانات ، وهي عكس التصنيف ، كما أنها تساعد المستفيد على فهم التركيب الطبيعي للمجموعات من البيانات.

2.6.1.2 اكتشاف التسلسل

تقنيات التسلسل تستخدم التحميل القائم على الوقت لانتزاع معلومات مفيدة ، وهي مماثلة للتجميع في انها تستخدم لتحديد العناصر التي تحدث معا ، ولكن الأهم من ذلك انها تستخدم لتحديد أي من العناصر يحدث أولاً .

3.6.1.2التصنيف

يستخدم التصنيف بشكل واسع في حل كثير من المشكلات من خلال تحليل مجموعه من البيانات ووضعها في شكل أصناف او أقسام يمكن استخدامها فيما بعد لتصنيف البيانات المستقبلية وهناك عدد

من الطرق التي يمكن إستخدامها في تصنيف البيانات بإستخدام الخوارزميات مثل الخوارزمية الإحصائية ، وخوارزميات الشبكات العصبية ، والعديد من خوارزميات تعلم الالة.

4.6.1.2الإنحدار

هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة ولتنبؤ قيم مستقبلية،ويستخدم للعثور على دالة تستخرج البيانات بأقل خطأ.[6]

5.6.1.2 التلخيص

توفر عرضا محكما للبيانات يحتوي علي التصور الكامل وتوليد التقرير.

6.6.1.2 قاعده الإرتباط

هي أحد التقنيات الواعدة من تنقيب البيانات كأداة من أدوات إكتشاف المعرفة ولديها القدرة على تحميل كميات هائلة من البيانات، وهي تسعي لإيجاد القوانين والعلاقات التي تشرح بعض الصفات الموجودة إعتماداً علي وجود الصفات الأخرى، بمعني آخر أنها قواعد ارتباطيه معينه بين مجموعه من البيانات في قواعد البيانات.

7.1.2 مقاييس قواعد الإرتباط

هذا التنقيب يكتشف مجموعة من القواعد التي يتم مشاركتها بين نسبة كبيرة من القواعد .وتميل الى إنتاج عدد كبير من القواعد والهدف هو العثور على القواعد المفيدة للمستخدمين ، ويمكن ان تقاس

الجدوى بطريقتين : موضوعية ، ذاتي.

القياس الموضوعي يشمل التحميل الاحصائي للبيانات مثلsupport ،confidence.

قاعدة الارتباط عبارة عن تعبير ضمني يوضح إرتباط X و Yبالصورة التالية:Y→X

حيث ان كل منها تمثل مجموعة (item set)

Support 1.7.1.2

يتم تعيين الحد الأدنى من الدعم والحد الأدنى من الثقة من قبل المستخدمين، وهي معلمات خوارزمية أبريوري لتوليد قاعدة الارتباط. وتستخدم هذه المعلمات لاستبعاد القواعد في النتيجة التي لها دعم أو الثقة أقل من الحد الأدنى من الدعم والحد الأدنى من الثقة على التوالي.

Support هو مؤشر على عدد مرات ظهور مجموعة من العناصر في مجموعة البيانات كنسبة المعاملات

TID	**ITEMSET**	**SUPPORT= OCCURRENCET/TOTAL SUPPORT**
1	**ABC**	TOTAL SUPPORT=5
2	**ABD**	SUPPORT{AB}=2/5=40%
3	**BC**	SUPPORT{BC}=3/5=60%
4	**AC**	SUPPORT{ABC}=1/5=20%
5	**BCD**	

جدول(1.2): مثال لحساب Support [7]

Confidence 2.7.1.2

تقوم بقياس عدد مرات ظهور العنصر Yفي المعاملات التي تحتوي على العنصر X.

TID	ITEMSET	Given X→Y Confidence=Occurrence{Y}/ Occurrence{x}
1	ABC	Confidence{A→B}=2/3=66%
2	ABD	Confidence{B→C}=3/4=75%
3	BC	Confidence{A B→C}=1/2=50%
4	AC	
5	BCD	

جدول (2.2) : يوضح مثال لحساب Confidence [7]

حيثTIDتمثل transaction

توجد هنالك العديد من المصطلحات لابد من التعرف عليها لشرح الخوارزمية:

1-مجموعة اصناف السلع: عندنا يقال K- item set يعني ان تكون مجموعة مكونة من kعنصر.

2-مجموعة عناصر السلع المكررة :هي التي لديها الحد الادنى من support.

8.1.2 مجتمع وعينة البحث

تم اختيار القطاع التجاري كمجتمع البحث، وذلك على اعتبار ان هذا القطاع يتعامل مع رأس المال والمتمثل بالتجار، وتم الحصول على البيانات من خمسة محلات تجارية كميدان للتطبيق، اذ تألفت عينة البحث من عدد من السلع المتوفرة بالمتجر جزء منها حصلت عليه من قاعدة البيانات لمتجرين والاخر من كراسات ديون لثلاثة متاجر.

9.1.2 .1 مقدمة عن weka

هي برنامج يستخدم لتنقيب البيانات وهي تحتوي على الأدوات المستخدمة في حل المشاكل الرئيسية في تنقيب البيانات مثل : الإنحدار والتصنيف والتجميع ، جميع الخوارزميات تأخذ مدخلاتها في شكل جدول أحادي العلاقة في صورة(ARFF) على أي حال الطريقة الرئيسة للتخزين في (Weka) هي في صورة(ARFF) ويمكن بكل سهولة التحويل من صورة برنامج الجدولة إلي صورة(ARFF)معظم ملفات(ARFF)تتكون من قائمة من الحالات وقيم وخواص لكل من الحالات التي تم يفصل بينها بالفاصلة. [2]

2.9.1.2 صيغةARFF

الصفوف التي تبدأ بعلامة (%) هي تعليقات ، يلي التعليقات في بداية الملف هو إسم العلاقة ،الخواص الإسمية يجب ان يتبعها مجموعة من القيم التي يمكن ان تأخذها وتكون هذه القيم متضمنة

في الأقواس المعرجة. القيم يمكن أن تحتوي على فراغات في هذه الحالة لابد من ان تكون مضمنة بين علامات تقدير، الخواص الرقمية لابد ان تليها الكملة المحجوزة(Numeric)كما هو موضح بالشكل التالي الذي يوضح مثال لملف(ARFF) للعلاقة(project):

@relation project
@attribute sn numeric

@attribute sex numeric

@attribute Arc numeric

@attribute ENC numeric

@attribute physics numeric

@attribute chemistry numeric

@attribute biology numeric

@attribute compute numeric

@attribute engineering numeric

أحد الطرق لإستخدام (Weka) هي تطبيق طريقة التعلم لمجموعة البيانات وتحميل مخرجاتها لمعرفة المزيد عن البيانات، وايضاً يمكن إستخدام النماذج التي تم تعليمها مسبقا لتوليد التوقعات على

حالات جديدة، وكذلك يمكن إستخدام مجموعة مختلفة من النماذج التي تم تعليمها مسبقاً ومقارنة اداؤها من أجل إختيار احدها للتنبؤ. في واجهة (Weka) التفاعلية يمكن تحديد طريقة التعلم المراد إستخدامها من القائمة ، والكثير من الطرق لديها مجموعة من المعلمات الإنضباطية التي يمكنك الوصول اليها من خلال ورقة الملكية أو من خلال محرر كائن، ويتم إستخدام وحدة التقييم المشتركة لقياس أداء جميع المصنفات. تطبيقات نظم التعلم الفعلي هي من المصادر الأكثر قيمة التي توفرها،(weka) ولكن الأدوات اللازمة لتجهيز البيانات التي تسمى المرشحات تأتي في المرتبة الثانية. مثل المصنفات يمكنك تحديد المرشحات من القائمة وتكييفها مع متطلباتك.

3.9.1.2 طريقة إستخدام (Weka)

أسهل الطرق لإستخدام (Weka) هو من خلال إستخدام واجهة المستخدم الرسومية وتسمى المستكشف، (Explorer) وهذا الأخير يتيح الوصول الى جميع منشأت (Weka) بإستخدام إختيار القائمة وملأ الإستمارة. مثلا يمكن بسرعة كبيرة قراءة قاعدة البيانات من ملف (ARFF) وتكوين شجرة القرارات منها. لكن شجرة القرارات تعتبر فقط البداية لأنه يوجد هنالك الكثير من الخوارزميات لابد من إكتشافها، واجهة المستكشف تساعدك على فعل هذا ، يمكنها توجيهك بعرض الخيارات في شكل قوائم. هنالك نوعين آخرين من واجهات المستخدم (Weka) بالإضافة للمستكشف، واجهة تدفق المعلومات (Knowledge Flow)وهي تسمح بتصميم الترتيبات لمعالجة البيانات المجرية (Streamed data) والواجهة الثالثة (Weka)هي المجرب (Experimental)وهي مصممة للإجابة على الأسئلة الأساسية للعملية عند الطلب على تقنيات الإرتداد والتصنيف . بالإضافة الى هذه الواجهات التفاعلية توجد

الوظيفة **الأساسية لل**(Weka.) وهي التي يمكن الوصول اليها في شكل صف بإدخال الأوامر النصية ، وهي بدورها تسمح بالوصول لكل المميزات بالنظام. عندما يعمل (Weka)لابد من الإختيار من بين هذه الأربع واجهات.

9.1.2 خوارزمية a prior

هي خوارزمية للتنقيب عن العناصر المتكررة وتطبيق قوانين الإرتباط على(transactional databases).وتعمل عن طريق تحديد العناصر الفردية المتكررة في قاعده البيانات وتوسيعها لتشمل مجموعات أكبر. وأكبر صنف هو الصنف الأكثر ظهوراً في قاعدة البيانات، ومجموعات السلع المتكررة(The frequent item sets) يتم تحديدها عن طريق خوارزمية(a priori)التي تستخدم لتحديد قوانين الإرتباط. هذه الخوارزمية مصممة للعمل على قواعد البيانات التي تحتوي على (transactions)، وتم تصميم الخوارزميات الأخرى لإيجاد قوانين الإرتباط في حالة عدم وجود (transaction)أو عدم وجود(timestamps)وينظر إلى كل(transaction)علي أنها مجموعة من الوحدات (item set) ، تستخدم الخوارزمية منهج"bottom up" حيث يتم تمديد مجموعات فرعيه للعنصر الواحد في اللحظة المعينة وهي خطوة(candidate generation)ويتم إختبار هذه المجموعات المرشحة وتنتهي هذة الخوارزمية عندما يتم العثور على أي تمديدات أخرى ناجحة. وتستخدم هذة الخوارزمية كل من (breadth first search) و(tree Hash) لحساب المجموعات المرشحة بكفاءة، حيث يتم توليد المجموعات المرشحة ذات الطول k من المجموعات المرشحة ذات الطول $K-1$.

الخوارزمية كالآتي:

Algorithm: Apriori. Find frequent itemsets using an iterative level-wise approach based on candidate generation.

Input:

- D, a database of transactions;
- *min_sup*, the minimum support count threshold.

Output: L, frequent itemsets in D.

Method:

```
(1)   L1 = find_frequent_1-itemsets(D);
(2)   for (k = 2; Lk−1 ≠ ϕ; k++) {
(3)       Ck = apriori_gen(Lk−1);
(4)       for each transaction t ∈ D { // scan D for counts
(5)           Ct = subset(Ck, t); // get the subsets of t that are candidates
(6)           for each candidate c ∈ Ct
(7)               c.count++;
(8)       }
(9)       Lk = {c ∈ Ck | c.count ≥ min_sup}
(10)  }
(11)  return L = ∪k Lk;

procedure apriori_gen(Lk−1:frequent (k − 1)-itemsets)
(1)   for each itemset l1 ∈ Lk−1
(2)       for each itemset l2 ∈ Lk−1
(3)           if (l1[1] = l2[1]) ∧ (l1[2] = l2[2]) ∧ ... ∧ (l1[k − 2] = l2[k − 2]) ∧ (l1[k − 1] < l2[k − 1]) then {
(4)               c = l1 ⋈ l2; // join step: generate candidates
(5)               if has_infrequent_subset(c, Lk−1) then
(6)                   delete c; // prune step: remove unfruitful candidate
(7)               else add c to Ck;
(8)           }
(9)   return Ck;

procedure has_infrequent_subset(c: candidate k-itemset;
          Lk−1: frequent (k − 1)-itemsets); // use prior knowledge
(1)   for each (k − 1)-subset s of c
(2)       if s ∉ Lk−1 then
(3)           return TRUE;
(4)   return FALSE;
```

شكل(3.2) : خوارزمية a priori [1]

توضيح الرموز

1. Tتوضح (Transaction) .للسلع

2.(€) توضح support.

3.(C_K) توضح المجموعة المرشحة في المستوى k.

4.(Count[c])توضح حقل هيكلة البيانات التي تمثل المجموعات المرشحة ، الذي من المفترض

أن تكون قيمتة الإبتدائية تساوي صفرا

5. (L)توضح اكبر صنف (Large item set).

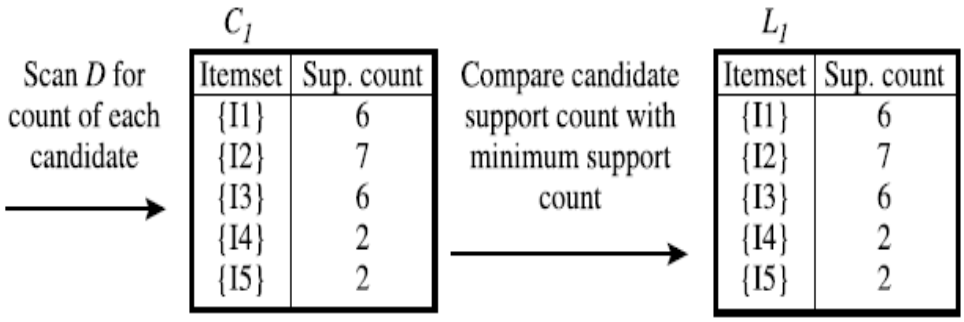
10.1.2 مثال اخوارزمية a prior

يوضح طريقة عمل الخوارزمية: لمجوعة من المعاملات في مركز فرعي لبيع الالكترونيات[1]

TID	*List of item_IDs*
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

1- Generate L1 with min-support:

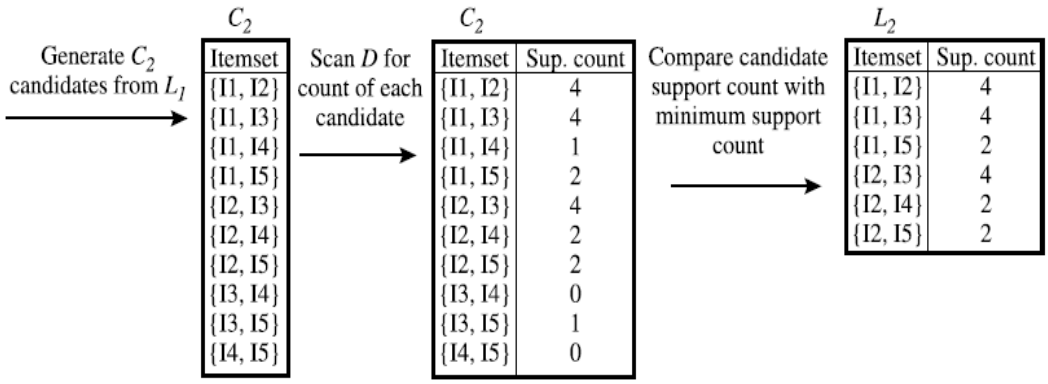
إيجاد قائمة السلع التي تحقق الحد الأدنى



Min support:2

2- Generate L2 with min-support:

إيجاد القائمة الثانية بعد عملية الاتحاد للسلع في المرحلة الاولي التي تحقق الحد الأدنى:



3- Generate L3 with min-support:

إيجاد القائمة الثالثة بعد عملية الاتحاد للسلع في المرحلة الثانية التي تحقق الحد الأدنى:

$$\text{Join: } C_3 = L_2 \bowtie L_2 = \{\{I1, I2\}, \{I1, I3\}, \{I1, I5\}, \{I2, I3\}, \{I2, I4\}, \{I2, I5\}\} \bowtie \{\{I1, I2\}, \{I1, I3\}, \{I1, I5\}, \{I2, I3\}, \{I2, I4\}, \{I2, I5\}\}$$
$$= \{\{I1, I2, I3\}, \{I1, I2, I5\}, \{I1, I3, I5\}, \{I2, I3, I4\}, \{I2, I3, I5\}, \{I2, I4, I5\}\}.$$

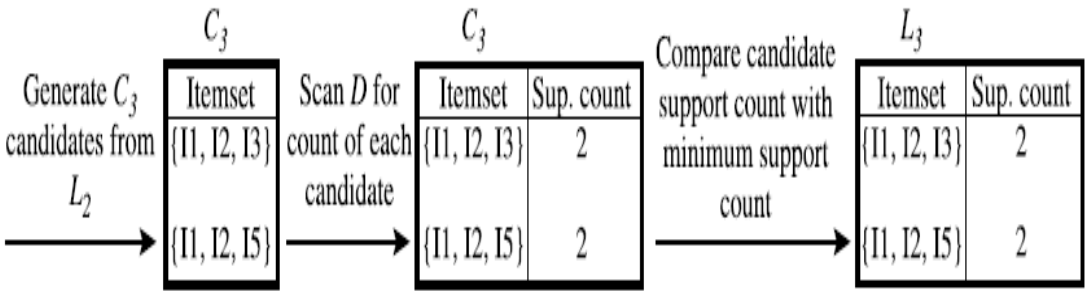
4- Generate L3 with pruning:

عملية التقليم يتم فيها استخدام خاصية خوارزمية a prior والتي تنص علي ان أي مجموعة جزئية غير خالية من مجموعة العناصر المتكررة يجب ان تكون متكررة.

Prune using the Apriori property: All nonempty subsets of a frequent itemset must also be frequent. Do any of the candidates have a subset that is not frequent?

- The 2-item subsets of {I1, I2, I3} are {I1, I2}, {I1, I3}, and {I2, I3}. All 2-item subsets of {I1, I2, I3} are members of L_2. Therefore, keep {I1, I2, I3} in C_3.
- The 2-item subsets of {I1, I2, I5} are {I1, I2}, {I1, I5}, and {I2, I5}. All 2-item subsets of {I1, I2, I5} are members of L_2. Therefore, keep {I1, I2, I5} in C_3.
- The 2-item subsets of {I1, I3, I5} are {I1, I3}, {I1, I5}, and {I3, I5}. {I3, I5} is not a member of L_2, and so it is not frequent. Therefore, remove {I1, I3, I5} from C_3.
- The 2-item subsets of {I2, I3, I4} are {I2, I3}, {I2, I4}, and {I3, I4}. {I3, I4} is not a member of L_2, and so it is not frequent. Therefore, remove {I2, I3, I4} from C_3.
- The 2-item subsets of {I2, I3, I5} are {I2, I3}, {I2, I5}, and {I3, I5}. {I3, I5} is not a member of L_2, and so it is not frequent. Therefore, remove {I2, I3, I5} from C_3.
- The 2-item subsets of {I2, I4, I5} are {I2, I4}, {I2, I5}, and {I4, I5}. {I4, I5} is not a member of L_2, and so it is not frequent. Therefore, remove {I2, I4, I5} from C_3.

5- Generate L3 with pruning:



Generate L3 with pruning:

Therefore, $C_3 = \{\{I1, I2, I3\}, \{I1, I2, I5\}\}$ after pruning.

6- Generate L4:

لإيجاد العناصر المرشحة في المرحلة الرابعة :يتم ذلك بعملية الاتحاد بين عناصر القائمة الثالثة ($L_3 \times L_3$) ويكون الناتج كالاتي: $\{I_1,I_2,I_3,I_5\}$ وسوف نقوم بتقليم هذه المجموعة لأن هنالك مجموعة جزئية منها غير مكررة وهي $\{I_2,I_3,I_5\}$ ولذلك سوف تكون المجموعة المرشحة C4مجموعة خالية وسوف تنتهي الخوارزمية بإيجاد العناصر المكررة التي تم الحصول عليها.

The algorithm uses $L_3 \bowtie L_3$ to generate a candidate set of 4-itemsets, C_4. Although the join results in {{I1, I2, I3, I5}}, this itemset is pruned because its subset {{I2, I3, I5}} is not frequent. Thus, $C_4 = \phi$, and the algorithm terminates, having found all of the frequent itemsets. ∎

7- Generate strong rules:

I1^I2 =>I5, Confidence = 2/4 =50%

I1^I5=>I2, Confidence = 2/2 =100%

I2^I5=>I1, Confidence = 2/2 =100%

I1^I2=>I5, Confidence = 2/6 =33%

I2^I1=>I5, Confidence = 2/7 =29%

I5^I1=>I2, Confidence = 2/2 =100%

الفصل الثاني: الدراسات السابقة

1.2.2 دراسة: Rashmi Jha2012

Association Rules Mining for Business Intelligence [3]

تحدث الباحث عن ذكاء الأعمال Business Intelligence وعرفها هي أي معلومات مستمدة من تحليلات البيانات الموجودة والتي يمكن الاستفادة منها في تطوير المنظمة. وتعدين البيانات هي مجموعة فرعية من ذكاء الاعمال أو وسيلة عملية لاستخلاص ذكاء الاعمال من البيانات باستخدام النمذجة الإحصائية للبيانات. ويمكن استخدامها للعثور على العلاقات المتبادلة بين عناصر البيانات المختلفة القرض والتي يمكن استخدامها لتحسين أداء الأعمال أو على الأقل فهم ما يحدث على نحو أفضل. ثم تحدث عن النمو المتسارع السريع في حجم وعدد من قواعد البيانات المتاحة في التطبيقات التجارية والصناعية والإدارية وغيرها، وهو مهم لدراسة كيفية استخراج المعرفة من البيانات الضخمة، في قواعد البيانات العلائقية، المعاملات قد استقطبت في الآونة الأخيرة الكثير من الاهتمام في مجتمعات قاعدة البيانات. وتتمثل المهمة في استخلاص مجموعة من القواعد يوجد بينها علاقة قوية.

– ثم تحدث الباحث عن المعاملات التجارية وذكر أن التنقيب بواسطة قاعدة الارتباط يتم تخزين كمية هائلة من البيانات إلكترونيا في معظم المؤسسات. التوازن لجميع السلع المباعة. على سبيل المثال، مجموعة Wal–Mart ، مع أكثر من 4000 متجر، تجمع نحو 20 مليون نقطة من بيع بيانات المعاملات كل يوم. ونظرا لهذه للكميات من البيانات، انه امر جيد من الناحية التجارية في محاولة لتحليلها للعثور على معلومات. وهذا مثال للاهتمام من تحليل قاعدة بيانات كبيرة من المعاملات للمحلات التجارية التي تهدف لإيجاد قاعدة الارتباط هذا ما يسمى بقوانين قاعدة الارتباط أو تحليل لسلة الزبائن. أنها تنطوي على البحث عن عادات العملاء مثيرة للاهتمام من خلال النظر للارتباط. التنقيب بواسطة، ثم تحدث عن قاعد الارتباط لديها العديد من التطبيقات الأخرى من تحليل سلة

الزبائن، بما في ذلك التسوق، وتجزئة العملاء، والطب، والتجارة الإلكترونية، والتصنيف والتجميع والتنقيب على شبكة الإنترنت، المعلوماتية الحيوية والمالية.

– ثم تحدث عن قاعدة الارتباط بمثال بسيط من المعاملات صاحب المتجر يريد أن يعرف العناصر التي تباع معا في كثير من الأحيان. ونحن نفترض أن عدد العناصر في المحل هو N وتتمثل هذه العناصر التي كتبها **I{i1, i2….in}** وT دلالة على الصفقة من قبل **T{t1, t2….tn}** مع كل معرف وحيد , (Tرقم المعاملة) ويحدد كل مجموعة فرعية من العناصر من الصنف مجموعة اشتريت من قبل عميل واحد. كل معاملة من الاصناف M تكون {I1،I2 I، ... ...، IN} معm<=n . الآن نبحث عن علاقة الارتباط، إعطاء رقم كبير. من المعاملات، ويتم تحديد هذا الي أن العناصر التي تميل إلى أن تحدث معا. وتجدر الإشارة إلى أن التنقيب بقاعدة الارتباط لا يأخذ بعين الاعتبار كميات من مواد تم شراؤها. غالبا ما تكون مكتوبة قاعدة الارتباط،X وY. X → Yغالبا ما يشار اليها على انها سابقة ظهور Y كما يترتب القاعدة. فإنه يشير إلى أن فقط X وY تم العثور معا في كثير من الأحيان في البيانات الواردة ولا تظهر وجود علاقة سببية مما يعني أن شراء X من قبل العملاء يسبب لها لشراء Y. افترض X وY السلع تظهر معا في 10٪ فقط من الصفقة ولكن كلما ظهر X يظهر أن هناك 80٪ نسبة أن Y تشتري معها. وجود 10٪ من X وY معا ما يسمى الدعم أو انتشار القاعدة 80٪ فرصة يسمى الثقة أو القدرة على التنبؤ من القاعدة.

2.2.2 دراسة: Phani Prasad J1, Murlidher Mourya2

A Study on Market Basket Analysis Using a Data Mining Algorithm [4]

تقدم هذه الورقة الكثير من دراسات الحالة حول قواعد الارتباط وخوارزميات تنقيب البيانات الموجودة والمستخدمة لتحليل سلة السوق، كما ذكر الباحث بوضوح حول خوارزمية a priori القائمة وتنفيذها بشكل واضح وأيضا عن مشاكلها وحلولها .في المستقبل نفس الخوارزمية يمكن تعديلها ويمكن تمديدها في العمل المستقبلي الذي يقلل أيضا من تعقيد الوقت. ثم تحدث عن قاعدة الارتباط وعرفها هي أداة قوية جدا حاليا في تنقيب البيانات. وهي تحدد الارتباط بين السلع في قواعد البيانات الكبيرة . مثال نموذجي لقاعدة الارتباط هو تحليل سلة السوق .في هذا طريقة أو نهج يفحص عادات الشراء من العملاء من خلال تحديد الارتباطات بين السلع التي تم شراؤها من قبل العملاء في سلتهم .وهذا يساعد على زيادة في مبيعات منتج معين من خلال تحديد السلع المتكررة التي تم شراؤها من قبل العملاء . في هذه الورقة ركز الباحث أساسا على دراسة استخراج البيانات الموجودة في مجموعة البيانات على خوارزمية سلة التسوق.

3.2.2 دراسة: Mehmet Aydin Ulas

MARKET BASKET ANALYSIS FOR DATA MINING [5]

تحدث الباحث عن تراكم البيانات في معظم الشركات القائمة على قاعدة من البيانات من العملاء والعقود. مع تطبيقات التجارة الإلكترونية تنمو قاعدة البيانات بسرعة، فإن الشركات يكون لديها كمية كبيرة من البيانات في أشهر فقط وليس في سنوات. وتنقيب البيانات، والمعروفة أيضا باسم اكتشاف المعرفة في قواعد البيانات (knowledge discovery in databases)، هو العثور على

الاتجاهات، والأنماط، والارتباطات، والشذوذ في قواعد البيانات هذه التي يمكن أن تساعدنا على تحقيق مستقبل دقيق واتخاذ قرارات مهمة. قواعد الارتباط هي واحدة من مجالات التطبيق الرئيسية لتنقيب البيانات. نظرا لمجموعة من معاملات العملاء على المنتجات، والهدف من ذلك هو العثور على الارتباطات بين مبيعات السلع. ثم إعطاء وصف خوارزمية A priori وعرض النتائج التي تؤخذ منGIMA (Tururk A.S) سلسلة سوبرماركت تركية كبيرة. أيضا استخدم طريقتين إحصائيتين : تحليل المكونات الرئيسية و كيفية الكشف عن الترابط بين مجموعات من العناصر.

وصف مجموعة البيانات المستخدمة لاختبار A priori هي من GIMA (Tururk A.S). الاختبار يتم علي جميع مجموعات البيانات في يونيو، يوليو، أغسطس 2000 في واحدة من أسواق السلسلة. تحتوي مجموعات البيانات على 756،868 معاملة. ويحتوي على 140،610 item set. وهناك7،237 أنواع من المنتجات. يظهر كل عنصر في المتوسط في 105 item set. وهناك5،909 منتج أقل من المتوسط و 1،328 في المتوسط. هناك 9،985 عميل في قاعدة البيانات. ثم تحدث عن خوارزمية أبريوري وعرفها هي واحدة من أسرع وأقرب الأدوات لرابطة التعدين. استخدم الباحث خوارزمية أبريوري لقواعد الارتباط في قاعدة البيانات الكبيرة من Tururk A.S gimaكانت البيانات المستخدمة من مجموعة سوبرماركت gima. واخذ الباحث البيانات في فصل الصيف وكان لدينا النتائج التالية. السلعة التي تباع اكثر هي الطماطم. وكانت الطماطم موجودة في 14 في الhمائة من المعاملات. ثاني أكثر سلعة بيعا كانت الخبز ، ثم جاء الخيار. ونتيجة لذلك استنتج الباحث أن شعب تركيا يأكل السلطة كثيرا في الصيف. ومن النتائج التي توصل لها الباحث. انا اذا اشترى الناس الخيار سوف يشتروا الطماطم مع نسبة 67%.,وايضا قال إن نسبة 67 % من السكان تعني أن 67 % من

الأشخاص الذين اشتروا الخيار اشتروا أيضا الطماطم (البندورة). وهناك قاعدة أخرى من هذا القبيل قد تكون 66 %في المائة من الأشخاص الذين اشتروا الفلفل الأخضر اشترى أيضا الطماطم. إذا أعطينا أمثلة للقواعد مع ثلاثة سلع وجد لديهم 55 % من الناس الذين اشتروا الشبت والطماطم اشتروا أيضا البقدونس.

4.2.2 دراسة :Roshan Gangurde1, Dr. Binod Kumar2, Dr. S. D. Gore3

Building Prediction Model using Market Basket Analysis [8]

تحدث الباحث عن تحول تحليل سلال التسوق إلى أن تكون جذابة جدا لتجار التجزئة. والتطور جعل من الممكن لهم جمع المعلومات من عملائهم وما يشترونه.وايضا تحدث عن نقطة البيع الإلكترونية وانها توسيع استخدام وتطبيق بيانات المعاملات في سلة السوق وتحليلها. في مجال تجارة التجزئة، وتحليل هذه المعلومات قيمة للغاية لفهم عملية الشراء وسلوك المستهلك . وأنماط شراء التعدين يسمح لتجار التجزئة بضبط الترقيات، وتخزين إعدادات وخدمة المستهلكين بشكل أفضل. تستخدم التحليلات التنبؤية تقنيات استخراج البيانات من أجل وضع تنبؤات حول الأحداث المستقبلية، وتقديم توصيات بناء على هذه التوقعات.

5.2.2 دراسة : Prof. Kavitha Venkatachari

MARKET BASKET ANALYSIS USING FP GROWTH AND APRIORI ALGORITHM: A CASE STUDY OF MUMBAI RETAIL STORE [9]

تحدث الباحث عن ان الشركات غنية بكميات هائلة من البيانات ولكنها ضعيفة في استخرج المعلومات من تلك البيانات. وتعتبر البيانات الكبيرة مورد قيم ، فالشركات في مجموعة متنوعة من الصناعات تعتمد عليها لاتخاذ القرارات الاستراتيجية. تحليل سلة المستهلك هي تقنية مفيدة جدا لمعرفة أنواع السلع التي تباع في سلة المستهلك معا. الهدف الرئيسي من الورقة هو أن نرى كيف تختلف المنتجات المختلفة في متاجر متنوعة وكيفية استغلال هذه العلاقات من قبل نشاطات تسويقية. ويمكن استخدام هذه المعلومات كأساس لاتخاذ قرارات بشأن التسويق مثل الدعم الترويجي، ومراقبة المخزون، والحملات عبر البيع. وللعثور على قواعد الجمعيات استخدم الباحث (خوارزمية البداهة) .للتعرف على مجموعات العناصر المتكررة. ولقد جمع البيانات من متجر التجزئة في مومباي .

الباب الثالث

المنهجية ومجموعة البيانات والبرامج المستخدمة

1.3مقدمة

اشتمل هذا الباب على اختيار البيانات وتهيئة البيانات وتنقيب البيانات ووصف البيانات وعمليات المعالجة على البيانات واحصائيات البيانات وطريقة تجهيز البيانات لتعمل على برنامج Weka بالتعامل مع المحلات التجارية نجد أنه تتوفر الكثير من المعلومات التجارية المتعلقة بالسلع وهي غير مستقلة، وفي هذا البحث تم تحديد بعض هذه البيانات كالتالي مصدر البيانات كل البيانات

1 – (هبة سوبر ماركت) تاريخ البيانات التي تم جمعها: فبراير من2015 وحتي ديسمبر 2015 . عدد كل سجلات المعاملات في قاعدة البيانات: 2321254سجل. عدد الصفقات: 386250 معاملة. عدد اصناف السلع في المتجر: 37062 صنف.

2– مصدر البيانات: (الدومة سوبر ماركت) تاريخ البيانات التي تم جمعها: من ديسمبر 2014.وحتي أكتوبر 2015عدد كل سجلات المعاملات في قاعدة البيانات: 1592200 سجل.

عدد الصفقات: 318440 معاملة. عدد اصناف السلع في المتجر: 112،05 صنف

3– مصدر البيانات: (بقالة الامين) تاريخ البيانات التي تم جمعها: من مارس 2014 وحتي أغسطس 2014. عدد الصفقات: 3826معاملة

4– مصدر البيانات 4: (بقالة البخيت) تاريخ البيانات التي تم جمعها: من يونيو 2015 الي أغسطس 2015.: عدد الصفقات:1651معاملة

5– مصدر البيانات: 5 (بقالة شقلبان) تاريخ البيانات التي تم جمعها: من أبريل 2015 وحتي أغسطس 2015. عدد الصفقات:5246 معاملة. وقد تمت معالجتها وفقا لخطوات تنقيب البيانات وهي كالتالي:

2.3 أولا إختيار البيانات

تم الحصول على البيانات من قبل قاعدة بيانات اثنين سوبر ماركت وثلاثة بقالات من كراسات الديون وكانت كل معاملة لمجوعة من السلع موضحة بالسعر والتاريخ ، وتتمثل هذه البيانات في أصناف السلع التي يتم بيعها في السوبر ماركت والبقالات خلال فترات زمنية محددة. وقد تحصلنا علي 276548عمليات تجارية بعد المعالجات التي تمت عليها .هذا العدد تم التعامل مع اكثر 24سلعه تم بيعها في السوبر ماركت الاثنين، والثلاثة بقالات وعند إدخال البيانات ظهرت بعض المشاكل التي تخص أنواع السلع المتشابه كثيرة مثل المشروبات الغازية نجد انها كثيرة حتي النوع الواحد نجد انه به احجام مختلفة في قاعدة البيانات ولقد قام الباحث بدمج كل أنواع المشروبات الغازية باسم واحد وهو gase water وكذلك كل أنواع البسكويت باسم واحد وكذلك الجبنة والكيك والبيض والطحينية والأرز والعدس والمربي والسكر والشعرية والمكرونة والاندومي والتانك واللحوم والصلصة والشبس والحليب والشاي والزيوت والخبز باسم واحد فقط كما مبين اسفل في العمليات علي قاعدة البيانات .

3.3 ثانيا تهيئة البيانات

بعد إكتمال جمع البيانات من مصادرها واستبعاد البيانات الغير مكتملة من مجموعة البيانات لأن هذه البيانات لابد ان تكون مكتملة لأنها تحتوي على أصناف السلع المباعة ولان كمية البيانات

كبيرة جدا ولا يمكن إدخالها للبرنامج لقد قمت باختيار اكثر أربعة وعشرين سلعة مباعا والبيانات كانت موجدة في قاعدة بيانات اوراكل في السوبر ماركت وستة كراسات ديون من ثلاثة بقالات والكمية موضحة بالتفصيل وكل صنف بعد المعالجة وتحليليه. تم تحديد كم مرة تم بيعه بالتفصيل لكل الاربعة وعشرين صنف حيث تم تفصيل نسبة بيعها بالتفصل.

ومن ثم تم تخزين أكثر أربعة وعشيرين صنف تم بيعها في جدول منفصل، بالإضافة الي تهيئة الشكل العام للملف بحيث يناسب أداة تنقيب البيانات المستخدمة وهي (WEKA)وهي تقبل البيانات في صورة ملف(ARFF) أو قاعدة بيانات، وتحويلها الي (Excel sheet) أيضا كما تم الحصول على المعلومات الخاصة بالأصناف مثل (النوع، السعر، تاريخ البيع) من قاعد البيانات وكراسات الديون ودمج كل البيانات معا.

4.3 ثالثا تنقيب البيانات

في هذه المرحلة تم إستخدام تقنية قاعدة الإرتباط وتطبيق خوارزمية(A priori) باستخدام برنامج (Weka).

5.3 وصف البيانات وعمليات المعالجة التي تمت عليها

قمت بإختيار بيانات من قاعدة بيانات اثنين سوبرماركت والبيانات الأخرى من كراسات ديون ثلاث بقالات وبعد ذلك تحولت لمرحلة تهيئة البيانات وفيها قمت بتجهيز وعزل البيانات المبهمة او المفقودة وهي مرحلة تنظيف البيانات وإزالة البيانات المفقودة ومن ثم قمت بدمج جميع البيانات في

ملف اكسل ومن ثم نقل البيانات الي شكل ملائم لاجراءات البحث وهي جعلها في صورة ARFF ومن ثم ادخلتها لبرنامج WEKA وبعد ذلك حصلت علي النتائج

1.5.3 وصف المحلات التجارية

وصف موقع البيانات قام الباحــث بجمع البيانات من مدينة ام درمان من سوبر ماركت هبة، وسوبر ماركات الدومة، وبقالة الأمين، وبقالة بخيت ، وبقالة شقلبان وسوف أوضح صورة البيانات في المرحلة الاولي في سوبر ماركت هبة كما موضح في الشكل (1.3) ، وترتيب السلع من الأكثر مباعا الي الأقل كما هو موضح في الشكل (2.3) ومدى قوة العلاقة بين كل سلعتين ويوجد ثلاثة جدال في الشكل (3.3) الأول موضح به العلاقات القوية المتوسطة والثالث يوضح العلاقات الضعيفة. وبقية وصف البيانات الأخرى موضح في الملاحق.

مصدر البيانات 1: (هبة سوبر ماركت) تاريخ البيانات التي تم جمعها: فبراير 2016.
عدد كل سجلات المعاملات في قاعدة البيانات: 2321254سجل.
عدد الصفقات: 386250 معاملة.
عدد اصناف السلع في المتجر: 37062 صنف.

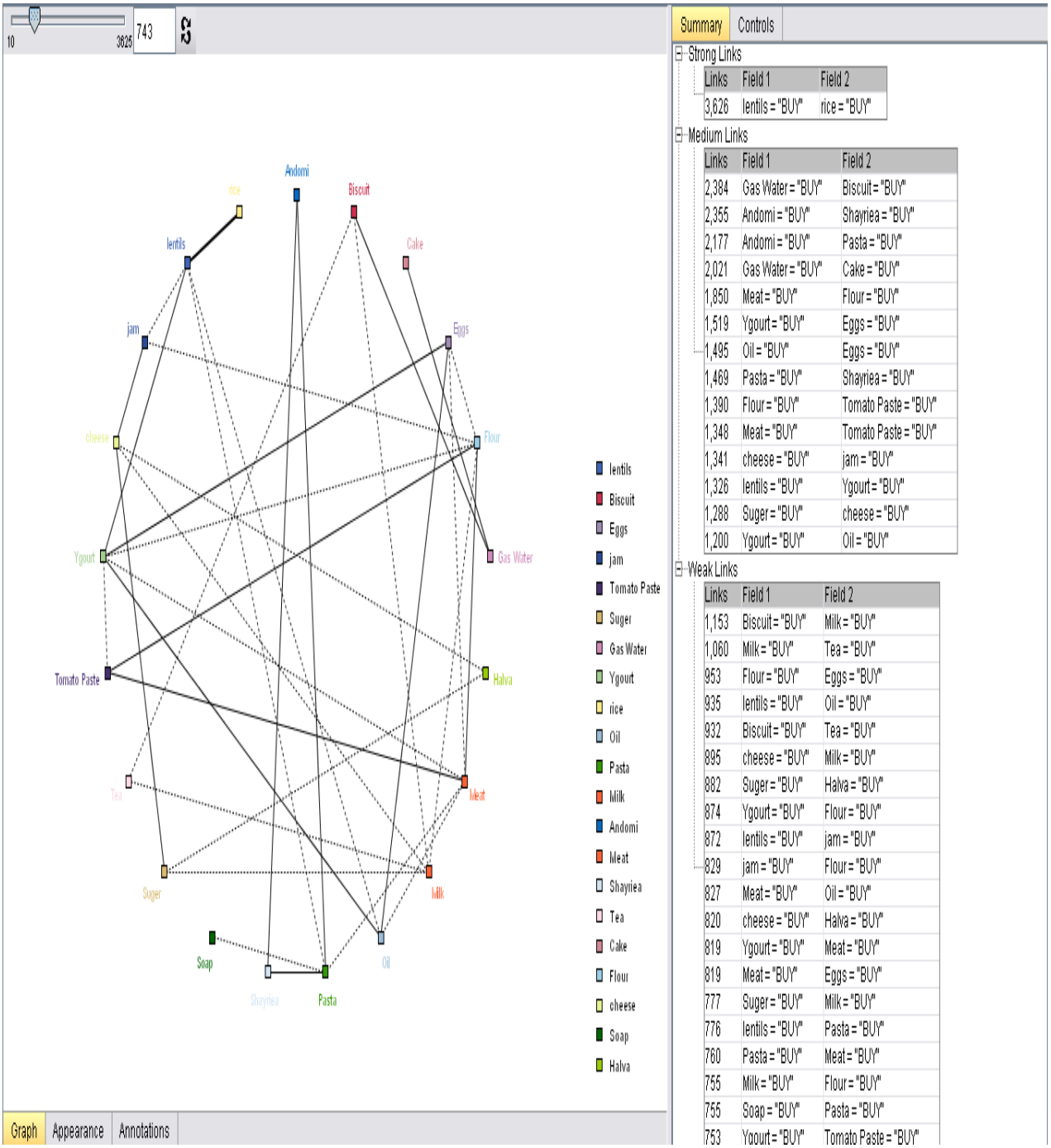
	A	B
1	BARCODE NUMBER	ITEM
37039	6281036247407	شيتوز جبنه وسطه
37040	6251001112108	مناديل فالين ٧٠منديل
37041	5000159407236	شوكلاتة مارس
37042	922522922014	ارياف ١كيلو
37043	922522922038	ارياف جبنه موزاريلا
37044	922522922038	ارياف جبنه موزاريلا
37045	6294001815174	جالكسى شوكولاتة
37046	12000002489	بيبسي علبة
37047	2305102000117	زبادي دايمة
37048	6223000410653	ليون بطعم الكاتشب
37049	1120102001077	بيبسى موبايل ٥٠٠ مل
37050	6223000410226	ليون بطاطس
37051	6281036567208	بطاطس دوريتوس ١٨٥ جرام
37052	6289174237557	كباب صيني
37053	2496466111080	كفتة بقري ميمو ٥٠٠ جرام
37054	جداد	جداد
37055	1120102001077	بيبسى موبايل ٥٠٠ مل
37056	6223000410301	ليون بطاطس
37057	6221007015147	مرقة ماجي
37058	6221007015147	مرقة ماجي
37059	6281036567208	بطاطس دوريتوس ١٨٥ جرام
37060	8212633832558	شاي الغزالتين ١٠٠غ
37061	6281034900366	فيمتو عصير الفااكهاا الطبيعي
37062	24000012306	شرائح اناناسdel monte

شكل (1.3) يوضح بيانات سوبر ماركت هبة في صورتها الاولي

Field	Proportion True	%	Count
Gas Water		9.14	10905
Biscuit		8.92	10650
Suger		8.84	10555
cheese		8.67	10348
lentils		8.07	9632
Cake		7.8	9313
Halva		7.34	8759
jam		7.28	8692
Milk		7.17	8557
rice		6.47	7717
Soap		6.42	7667
Tea		6.04	7206
Andomi		5.93	7074
Pasta		5.24	6248
Ygourt		4.89	5833
Flour		4.76	5686
Shayriea		4.63	5522
Meat		4.53	5410
Oil		4.07	4852
Tang		4.06	4841
Eggs		3.9	4659
Tomato Paste		3.6	4298
chips		3.44	4111
Bread		3.39	4044

شكل (2.3) يوضح بيانات سوبر ماركت هبة

اول عمود به اسم الصنف والثاني به نسبة بيع الصنف والثالث به كم مرة تم بيع الصنف



شكل (3.3) يوضح قوة العلاق بين كل عنصرين ومقسم الي ثلاثة علاقة قوية ومتوسطة وضعيفة

2.5.3 العمليات على البيانات

الخطوات التالية تمت على (الدومة سوبر ماركت وهبة سوبر ماركت)

1.2.5.3 الخطوة الاولي: دمج كل أنواع السلع المتشابه في فئة واحدة مثل: بيبسي ltr2، بيبسي ltr1.5، بيبسي 500 مل، بيبسي ML350، ltr2 mernda، ltr1.5mernda ... الخ، والماء: كما هو موضح في الشكل (4.3) .

ITEM (1) : GAS WATER ITEMS

```
Update car_stock_trans
SET
ITEM_CoDE = 'GAS WATER'
WHERE
ITEM_CoDE = '012000020742' or ITEM_CoDE = '012000005459' or
ITEM_CoDE = '1120102001190'or ITEM_CoDE = '1120102001855'or
ITEM_CoDE = '012000002496'or ITEM_CoDE = '01200800030'or
ITEM_CoDE = '01200800153' …. Etc
```

ITEM (2) : Biscuit

```
Update car_stock_trans
SET
ITEM_CoDE = 'BISCUIT'
WHERE
ITEM_CoDE = '8000380007219'or ITEM_CoDE = '6291007700275'or
ITEM_CoDE = '6291007700015'or ITEM_CoDE = '8690705025112'or
ITEM_CoDE = '5000183007457'or ITEM_CoDE = '8000350003036'or
ITEM_CoDE = '6281033213023' …. Etc
```

ITEM (3) : CAKE

```
Update car_stock_trans
SET
ITEM_CoDE = 'CAKE'
WHERE
ITEM_CoDE = '6291100600724'or ITEM_CoDE = '1971072601350'or
ITEM_CoDE = '5201002001292' or ITEM_CoDE = '6291100601776'or
ITEM_CoDE = '1971072601374'or ITEM_CoDE = '6291100602889'or
ITEM_CoDE = '6251034118108' …. Etc
```

ITEM (4) : SUGER

```
Update car_stock_trans
SET
ITEM_CoDE = ‘SUGER’
WHERE
ITEM_CoDE = ‘سكر 50 كيلو’or ITEM_CoDE = ‘سكر 10 كيلو’or
ITEM_CoDE = ‘سكر كيلو’or ITEM_CoDE = 90125003or ITEM_CoDE =
‘6091011010212’or ITEM_CoDE = ‘سكر مسحون ابو شامه’or ITEM_CoDE =
‘04021000010100277’ …. Etc
```

ITEM (5) : CHEESE

```
Update car_stock_trans
SET
ITEM_CoDE = ‘CHEESE’
WHERE
ITEM_CoDE = ‘جبنة’ or ITEM_CoDE = ‘جبنه مضفره’ or ITEM_CoDE =
‘2305102001060’or ITEM_CoDE = ‘2305102001022’or ITEM_CoDE =
‘6223000752098’or ITEM_CoDE = ‘6223000756164’or ITEM_CoDE =
‘6223000756447’ …. Etc
```

ITEM (6) : LENTILS

```
Update car_stock_trans
SET
ITEM_CoDE = ‘LENTILS’
WHERE
ITEM_CoDE = ‘10000076’ or ITEM_CoDE = ‘10002179’ or
ITEM_CoDE = ‘10003046’ or ITEM_CoDE = ‘10003053’ or
ITEM_CoDE = ‘8697669202122’ or ITEM_CoDE = ‘8697669207615’ or
ITEM_CoDE = ‘2961658030147’ …. Etc
```

ITEM (7) : RICE

```
Update car_stock_trans
SET
ITEM_CoDE = ‘RICE’
WHERE
ITEM_CoDE = ‘6281100201137’ or ITEM_CoDE = ‘2961658030017’ or
ITEM_CoDE = ‘2961658030161’ or ITEM_CoDE = ‘2961658030062’ or
ITEM_CoDE = ‘2961658030079’ or ITEM_CoDE = ‘690225104654’ or
ITEM_CoDE = ‘690225105002’ …. Etc
```

ITEM (8) : JAM
Update car_stock_trans SET ITEM_CoDE = ‘JAM’ WHERE ITEM_CoDE = ‘5027876065259’ or ITEM_CoDE = ‘6281041228316’ or ITEM_CoDE = ‘10004296’ or ITEM_CoDE = ‘6221029620121’ or ITEM_CoDE = ‘6221029600123’ or ITEM_CoDE = ‘6221029670126’ or ITEM_CoDE = ‘6221029650135’ …. Etc

ITEM (9) : HALVA
Update car_stock_trans SET ITEM_CoDE = ‘HALVA’ WHERE ITEM_CoDE = ‘6221029560267’ or ITEM_CoDE = ‘6224000719067’ or ITEM_CoDE = ‘6224000719098’ or ITEM_CoDE = ‘6221029520131’ or ITEM_CoDE = ‘6223000052631’ or ITEM_CoDE = ‘6223000052655’ or ITEM_CoDE = ‘6223000052648’ …. Etc
ITEM (10) : MILK
Update car_stock_trans SET ITEM_CoDE = ‘MILK’ WHERE ITEM_CoDE = ‘8716200478175’ or ITEM_CoDE = ‘8692971031569’ or ITEM_CoDE = ‘1003’ or ITEM_CoDE = ‘8717405009225’ or ITEM_CoDE = ‘6281022118421’ or ITEM_CoDE = ‘8212634924092’ or ITEM_CoDE = ‘016229001711’ …. Etc
ITEM (11) : SoAP
Update car_stock_trans SET ITEM_CoDE = ‘SoAP’ WHERE ITEM_CoDE = ‘5413149997592’ or ITEM_CoDE = ‘6281119000370’ or ITEM_CoDE = ‘501132167143’ or ITEM_CoDE = ‘5011321397613’ or ITEM_CoDE = ‘6001106204617’ or ITEM_CoDE = ‘4901780728619’ or ITEM_CoDE = ‘6318148288681’ …. Etc

ITEM (12) : TEA
Update car_stock_trans SET ITEM_CoDE = ‘TEA’ WHERE ITEM_CoDE = ‘6221048700873’ or ITEM_CoDE = ‘6281013221000’ or ITEM_CoDE = ‘6281006854901’ or ITEM_CoDE = ‘8718114850085’ or ITEM_CoDE = ‘8212633832626’ or ITEM_CoDE = ‘070177085315’ or ITEM_CoDE = ‘8718114784540’ …. Etc
ITEM (13) : ANDoMI
Update car_stock_trans SET ITEM_CoDE = ‘ANDoMI’ WHERE ITEM_CoDE = ‘089686060362’ or ITEM_CoDE = ‘089686060027’ or ITEM_CoDE = ‘089686060461’ or ITEM_CoDE = ‘5285000395164’ or ITEM_CoDE = ‘089686120141’ or ITEM_CoDE = ‘089686170313’ or ITEM_CoDE = ‘089686120073’ …. Etc
ITEM (14) : PASTA
Update car_stock_trans SET ITEM_CoDE = ‘PASTA’ WHERE ITEM_CoDE = ‘048400439006’ or ITEM_CoDE = ‘8005391108436’ or ITEM_CoDE = ‘8005391190431’ or ITEM_CoDE = ‘041190022243’ or ITEM_CoDE = ‘6161106640034’ or ITEM_CoDE = ‘8005391155430’ or ITEM_CoDE = ‘8004323933696’ …. Etc

ITEM (15) : SHARIEA
Update car_stock_trans SET ITEM_CoDE = ‘SHARIEA’ WHERE ITEM_CoDE = ‘8004323212654’ or ITEM_CoDE = ‘8003740130084’ or ITEM_CoDE = ‘5285000392057’ or ITEM_CoDE = ‘896861209054’ or ITEM_CoDE = ‘089686120110’ or ITEM_CoDE = ‘5285000392064’ or ITEM_CoDE = ‘8004149000916’ …. Etc

ITEM (16) : YGoURT
Update car_stock_trans SET ITEM_CoDE = ‘YGoURT’ WHERE ITEM_CoDE = ‘735400365104’ or ITEM_CoDE = ‘2305102000254’ or ITEM_CoDE = ‘2305102001220’ or ITEM_CoDE = ‘2305102001237’ or ITEM_CoDE = ‘2305102000322’ or ITEM_CoDE = ‘2305102000360’ or ITEM_CoDE = ‘024911000027’ …. Etc
ITEM (17) : MEAT
Update car_stock_trans SET ITEM_CoDE = ‘MEAT’ WHERE ITEM_CoDE = ‘10000045’ or ITEM_CoDE = ‘10000014’ or ITEM_CoDE = ‘10000021’ or ITEM_CoDE = ‘10000038’ or ITEM_CoDE = ‘10000120’ or ITEM_CoDE = ‘10004982’ or ITEM_CoDE = ‘10008’…. Etc

ITEM (18) : FLoUR
Update car_stock_trans SET ITEM_CoDE = ‘FLoUR’ WHERE ITEM_CoDE = ‘1114102000062’ or ITEM_CoDE = ‘888970696995’ or ITEM_CoDE = ‘02491301473’ or ITEM_CoDE = ‘1114102000079’ or ITEM_CoDE = ‘5027876048641’ or ITEM_CoDE = ‘10000441’ or ITEM_CoDE = ‘10000458’ …. Etc
ITEM (19) : TANG
Update car_stock_trans SET ITEM_CoDE = ‘TANG’ WHERE ITEM_CoDE = ‘043000306055’ or ITEM_CoDE = ‘043000303863’ or ITEM_CoDE = ‘043000303368’ or ITEM_CoDE = ‘043000303375’ or ITEM_CoDE = ‘043000303870’ or ITEM_CoDE = ‘043000303313’ or ITEM_CoDE = ‘043000303887’ …. Etc

ITEM (20) : Oil
Update car_stock_trans SET ITEM_CoDE = ‘oIL’ WHERE ITEM_CoDE = ‘72000’ or ITEM_CoDE = ‘22000’ or ITEM_CoDE = ‘8691313010064’ or ITEM_CoDE = ‘10006184’ or ITEM_CoDE = ‘10006191’ or ITEM_CoDE = ‘631583104627’ or ITEM_CoDE = ‘8428483200168’ …. Etc

ITEM (21) : EGGS
Update car_stock_trans SET ITEM_CoDE = ‘EGGS’ WHERE ITEM_CoDE = ‘2279151269200’ or ITEM_CoDE = ‘819052000322’ or ITEM_CoDE = ‘819052000308’ or ITEM_CoDE = ’00.’ or ITEM_CoDE = ‘360’ or ITEM_CoDE = ‘بيض’ or ITEM_CoDE = ‘2626’ …. Etc

ITEM (22) : ToMATo PASTE
Update car_stock_trans SET ITEM_CoDE = ‘ToMATo PASTE’ WHERE ITEM_CoDE = ‘717273501881’ or ITEM_CoDE = ‘8851978601018’ or ITEM_CoDE = ‘6291105241441’ or ITEM_CoDE =’ 6221048824524’ or ITEM_CoDE = ‘6917878007489’ or ITEM_CoDE = ‘8690575011918’ or ITEM_CoDE = ‘8690714133051’ …. Etc

ITEM (23) : CHIPS
Update car_stock_trans SET ITEM_CoDE = ‘CHIPS’ WHERE ITEM_CoDE = ‘6281036170101’ or ITEM_CoDE = ‘5033876045085’ or ITEM_CoDE = ‘6223000410424’ or ITEM_CoDE =’ 6223000410585’ or ITEM_CoDE = ‘6281036170200’ or ITEM_CoDE = ‘6281036071231’ or ITEM_CoDE = ‘6281036071132’ …. Etc

ITEM (24) : BREAD
Update car_stock_trans SET ITEM_CoDE = ‘BREAD’ WHERE ITEM_CoDE = ‘عيش’ or ITEM_CoDE = ‘0001’ or ITEM_CoDE = ‘0002’ or ITEM_CoDE =’ 0004’

الشكل (4.3) : يوضح طريقة جمع السلع المتشابهة في اسم واحد

2.2.5.3 الخطوة الثانية: بعد تصنيف السلع، قمت با نشاء جدول جديد الذي يحتوي على أكثر 24 سلع مباعا فقط في قاعدة البيانات كما موضح بالشكل (5.3)

Create Table
Creat table Ready_data AS select TRAN_No , ITEM_CoDE from CAR_SToCK_TRAN where ITEM_CoDE = ‘GAS WATER’ oR ITEM_CoDE = ‘BISCUIT’ oR ITEM_CoDE = ‘CAKE’ oR ITEM_CoDE = ‘SUGAR’ oR ITEM_CoDE = ‘CHEESE’ oR ITEM_CoDE = ‘LENTILS’ oR ITEM_CoDE = ‘RICE’ oR ITEM_CoDE = ‘JAM’ oR ITEM_CoDE = ‘HALVA’ oR ITEM_CoDE = ‘MILK’ oR ITEM_CoDE = ‘SoAP’ oR ITEM_CoDE = ‘TEA’ oR ITEM_CoDE = ‘ANDoMI’ oR ITEM_CoDE = ‘PASTA’ oR ITEM_CoDE = ‘SHARIEA’ oR ITEM_CoDE = ‘YGoURT’ oR ITEM_CoDE = ‘MEAT’ oR ITEM_CoDE = ‘FLoUR’ oR ITEM_CoDE = ‘TANG’ oR ITEM_CoDE = ‘oIL’ oR ITEM_CoDE = ‘EGGS’ oR ITEM_CoDE = ‘ToMATo PASTE’ oR ITEM_CoDE = ‘CHIPS’ oR ITEM_CoDE = ‘BREAD’

الشكل (5.3) : يوضح انشاء جدول جديد (Ready_data)

3.2.5.3 الخطوة الثالثة: إضافة 24 عمود إلي الجدول Ready_data (اكثر اربعة وعشرين صنف تم بيعها) كما هو موضح في الشكل (6.3)

Add Column
Alter Table Ready_data ADD (GAS WATER VARCHAR2(10), BISCUIT VARCHAR2(10), CAKE VARCHAR2(10), SUGAR VARCHAR2(10), CHEESE VARCHAR2(10), LENTILS VARCHAR2(10), RICE VARCHAR2(10), JAM VARCHAR2(10), HALVA VARCHAR2(10), MILK VARCHAR2(10), SoAP VARCHAR2(10), TEA VARCHAR2(10), ANDoMI VARCHAR2(10), PASTA VARCHAR2(10), SHARIEA VARCHAR2(10), YGoURT VARCHAR2(10), MEAT VARCHAR2(10), FLoUR VARCHAR2(10), TANG VARCHAR2(10), oIL VARCHAR2(10), EGGS VARCHAR2(10), ToMATo PASTE VARCHAR2(10), CHIPS VARCHAR2(10), BREAD VARCHAR2(10))

الشكل (6.3) : يوضح إضافة 24 عمود لجدول Ready_data

4.2.5.3 الخطوة الرابعة: تحويل البيانات من صفوف لأعمدة واستخدام 1 للسلعة التي تم شراها واستخدام 0 للسلع التي لم يتم شراها.

ITEM (1) : GAS WATER ITEMS

```
UPDATE Ready_data
SET
GAS WATER = ‘1’
Where
ITEM_CoDE = ‘GAS WATER’
```

ITEM (2) : Biscuit

```
UPDATE Ready_data
SET
BISCUIT = ‘1’
Where
ITEM_CoDE = ‘BISCUIT’
```

ITEM (3) : CAKE

```
UPDATE Ready_data
SET
CAKE = ‘1’
Where
ITEM_CoDE = ‘CAKE’
```

ITEM (4) : SUGER

```
UPDATE Ready_data
SET
SUGER = ‘1’
Where
ITEM_CoDE = ‘SUGER’
```

ITEM (5) : CHEESE

```
UPDATE Ready_data
SET
CHEESE = ‘1’
Where
ITEM_CoDE = ‘CHEESE’
```

ITEM (6) : LENTILS
UPDATE Ready_data SET LENTILS = ‘1’ Where ITEM_CoDE = ‘LENTILS’

ITEM (7) : RICE
UPDATE Ready_data SET RICE = ‘1’ Where ITEM_CoDE = ‘RICE’

ITEM (8) : JAM
UPDATE Ready_data SET JAM = ‘1’ Where ITEM_CoDE = ‘JAM’

ITEM (9) : HALVA
UPDATE Ready_data SET HALVA = ‘1’ Where ITEM_CoDE = ‘HALVA’

ITEM (10) : MILK

UPDATE Ready_data
SET
MILK = ‘1’
Where
ITEM_CoDE = ‘MILK

ITEM (11) : SoAP
UPDATE Ready_data SET SoAP = ‘1’ Where ITEM_CoDE = ‘SoAP’
ITEM (12) : TEA
UPDATE Ready_data SET TEA = ‘1’ Where ITEM_CoDE = ‘TEA’

ITEM (13) : ANDoMI
UPDATE Ready_data SET ANDoMI = ‘1’ Where ITEM_CoDE = ‘ANDoMI

ITEM (14) : PASTA
UPDATE Ready_data SET PASTA = ‘1’ Where ITEM_CoDE = ‘PASTA

ITEM (15) : SHARIEA
UPDATE Ready_data SET SHAERIA = ‘1’ Where ITEM_CoDE = ‘SHAERIA’

ITEM (16) : YGoURT
UPDATE Ready_data SET YGoURT = ‘1’ Where ITEM_CoDE = ‘YGoURT’

ITEM (17) : MEAT
UPDATE Ready_data SET MEAT = ‘1’ Where ITEM_CoDE = ‘MEAT’

ITEM (18) : FLoUR
UPDATE Ready_data SET FLoUR = ‘1’ Where ITEM_CoDE = ‘FLoUR’

ITEM (19) : TANG
UPDATE Ready_data SET TANG = ‘1’ Where ITEM_CoDE = ‘TANG

ITEM (20) : OIL

UPDATE Ready_data
SET
oIL = ‘1’
Where
ITEM_CoDE = ‘oIL’

ITEM (21) : EGGS
UPDATE Ready_data SET EGGS = ‘1’ Where ITEM_CoDE = ‘EGGS’

ITEM (22) : ToMATo PASTE
UPDATE Ready_data SET ToMATo PASTE = ‘1’ Where ITEM_CoDE = ‘ToMATo PASTE’

ITEM (23) : CHIPS
UPDATE Ready_data SET CHIPS = ‘1’ Where ITEM_CoDE = ‘CHIPS’

ITEM (24) : BREAD
UPDATE Ready_data SET BREAD = ‘1’ Where ITEM_CoDE = ‘BREAD’

شكل (7.3) يوضح طريقة تحويل البيانات

5.2.5.3 الخطوة الخامسة: جمع كل السلع في الصفوف المتعددة للعمليات المختلفة لصف واحد، كما هو موضح في الشكل (8.3)

Assembly step
Create table basket_analysis AS SELECT TRAN_No , SUM(GAS WATER) AS 'GAS WATER' , SUM(BISCUIT) AS 'BISCUIT' , SUM(CAKE) AS 'CAKE' , SUM(SUGAR) AS 'SUGAR' , SUM(CHEESE) AS 'CHEESE' , SUM(LENTILS) AS 'LENTILS' , SUM(RICE) AS 'RICE' , SUM(JAM) AS 'JAM' , SUM(HALVA) AS 'HALVA' , SUM(MILK) AS 'MILK' , SUM(SoAP) AS 'SoAP' , SUM(TEA) AS 'TEA' , SUM(ANDoMI) AS 'ANDoMI' , SUM(PASTA) AS 'PASTA' , SUM(SHARIEA) AS 'SHAERIA' , SUM(YGoURT) AS 'YGoURT' , SUM(MEAT) AS 'MEAT' , SUM(FLoUR) AS 'FLoUR' , SUM(TANG) AS 'TANG' , SUM(oIL) AS 'oIL' , SUM(EGGS) AS 'EGGS' , SUM(ToMATo PASTE) AS 'ToMATo PASTE' , SUM(CHIPS) AS 'CHIPS' , SUM(BREAD) AS 'BREAD' FRoM Ready_data GRoUP BY TRAN_No

شكل (8.3) :يوضح جمع كل السلع في صف واحد

6.2.5.3 الخطوة السادسة: النتيجة النهائة لل DATA بعد كتابة 1 لكل سلعة تم بيعها و 0 للسعة التي لم يتم بيعها في كل عملية شراء كما هو موضح في الشكل (9.3)

TID	Gas	Water	Biscuit	Cake	Suger	cheese	lentils	rice	jam	Halva	Milk	Soap	Tea	Andomi	Pasta	Shayriea	Ygourt	Meat	Flour	Tang	Oil	Eggs	Tomato Paste	chips	Bread
1		0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	
2		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	
3		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
4		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5		0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
6		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7		0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
8		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9		1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
10		0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11		0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12		0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13		1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14		0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
15		0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16		1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	
19		0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
20		0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
21		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

شكل (9.3) :يوضح عملية تحويل البيانات الي 0،1

7.2.5.3 الخطو ةالسابعة: يتم فيها استبدال 0بكلمة "NoT BUY"لم يشتري الصنف واستبدال و 1بكلمة " buy"تم شراء الصنف كما هو موضح في الشكل (10.3)

	Gas Wa	Biscuit	Cake	Suger	cheese	lentils	rice	jam	Halva	Milk	Soap	Tea	Andomi	Pasta	Shayrie	Ygourt	Meat	Flour	Tang
2	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
3	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
4	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
5	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
6	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	BUY	NOT BUY
7	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY
8	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
9	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
10	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
11	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
12	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
13	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
14	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
15	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
16	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
17	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
18	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
19	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
20	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
21	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY
22	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
23	NOT BUY	BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
24	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY
25	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY
26	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY	NOT BUY

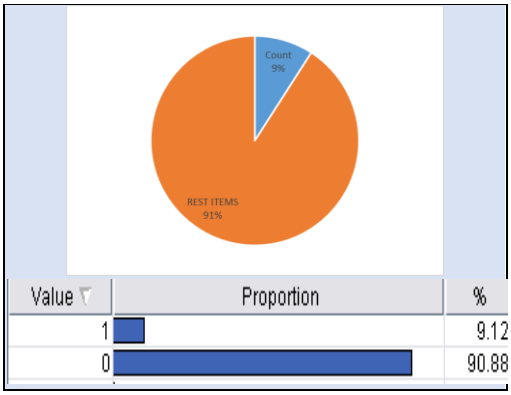
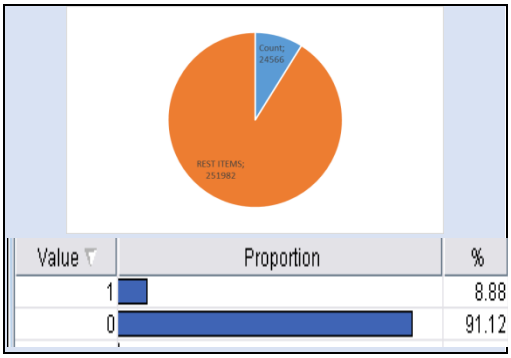
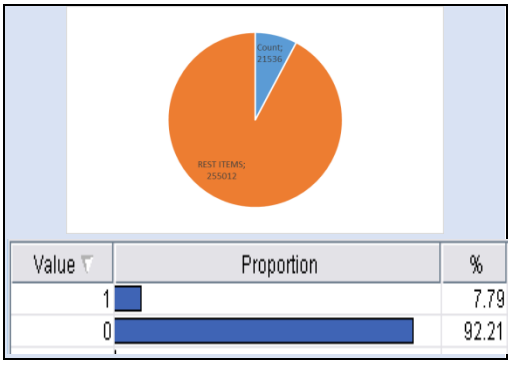
شكل (10.3): يوضح عملية تحويل 0،1 الي BUY, NOT BUY

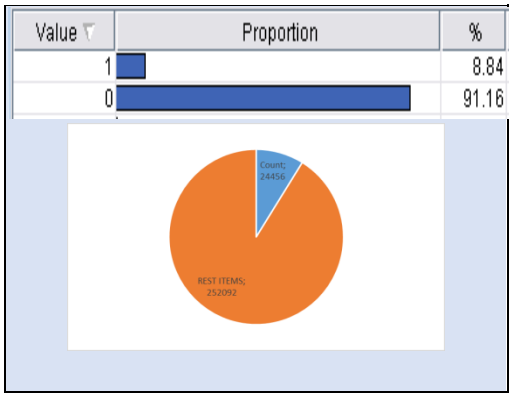
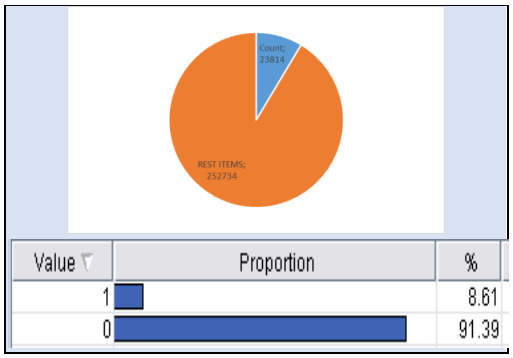
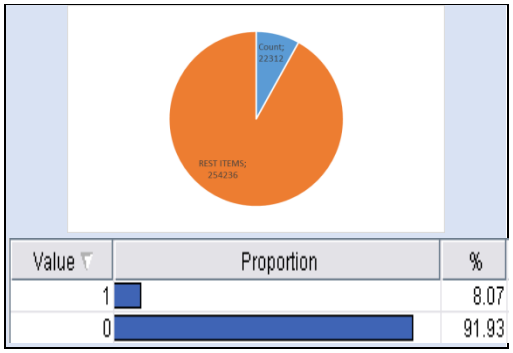
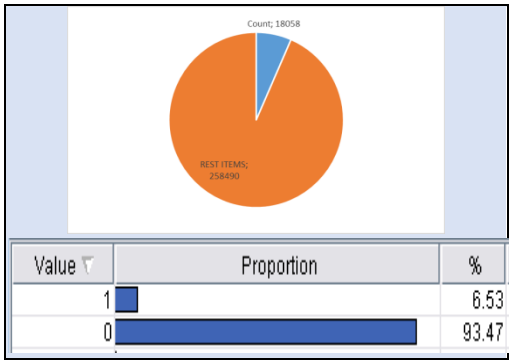
8.2.5.3 الخطوة الثامنة : جمع كل بيانات الخمسة محلات التجارية في ملف واحد csv كما هو موضح في الشكل (11.3) .

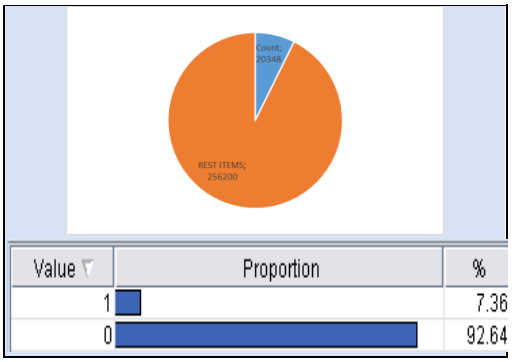
شكل (11.3) : يوضح شكل البيانات بعد تحويلها الي صيغة csv

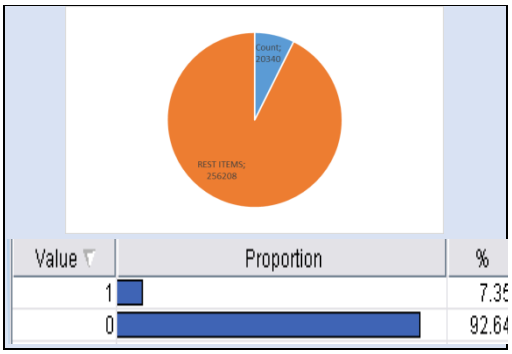
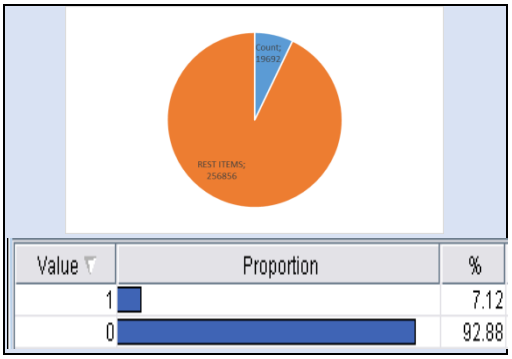
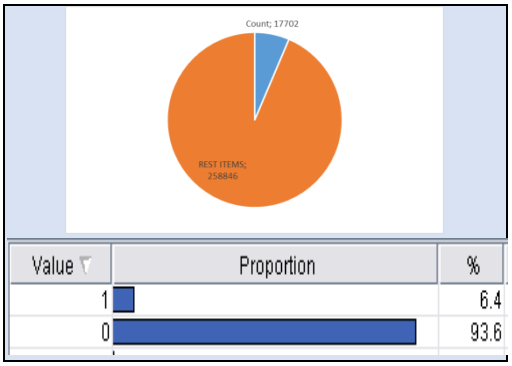
6.3 إحصائيات البيانات

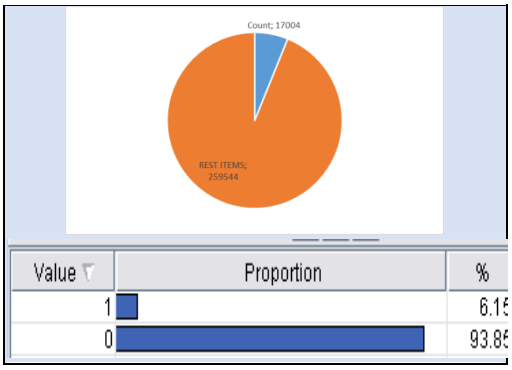
قمت بعمل مجموعة من المخططات بناً ء علي البيانات المتحصل عليها وهي توضح مباع السلع واحصائيات المنتجات التي تم بيعها في الفترة المحددة مسبقا. كما هو موضح بالشكل (12.3)

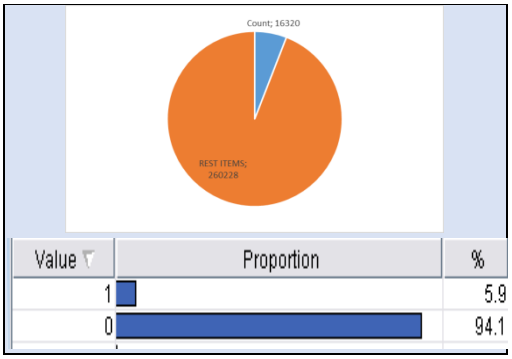
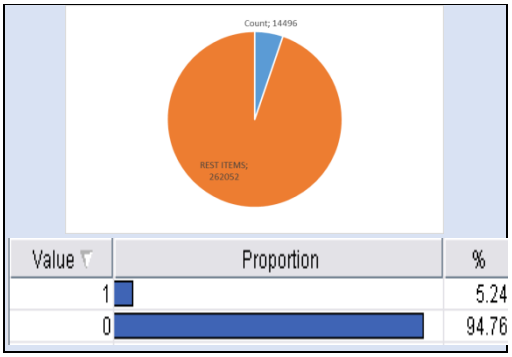
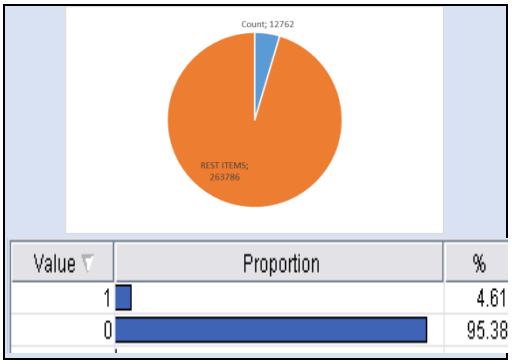
الرقم	السلعة	العدد	النسبة	الرسم البياني
1	مياه غازية	25212	9.116681372	
2	بسكويت	24566	8.883087	
3	كيك	21536	7.7874	

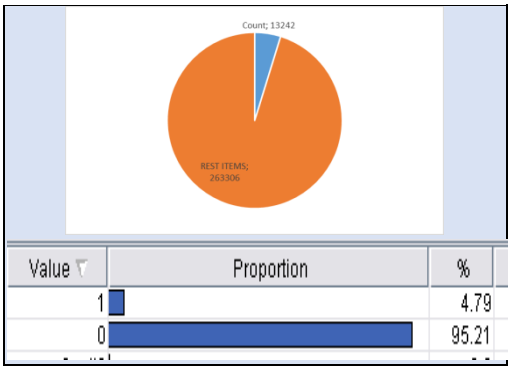
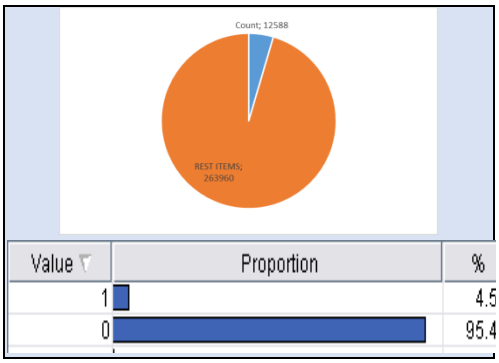
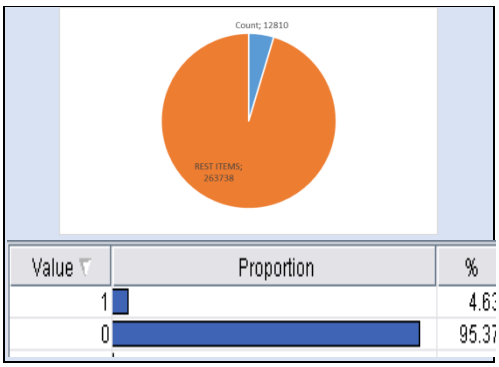
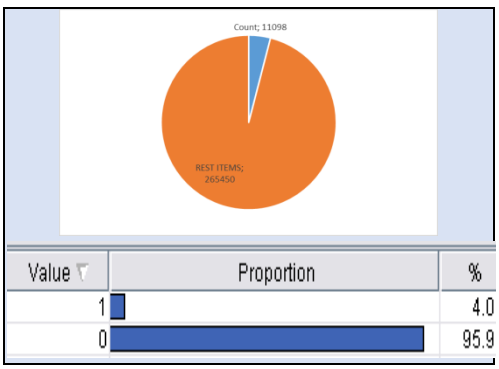
4	السكر	24456	8.84331	
No	**Item**	**Count**	**Percent**	**CHART**
5	جبنة	23814	8.611163	
6	العدس	22312	8.06804	
7	الأرز	18058	6.53	

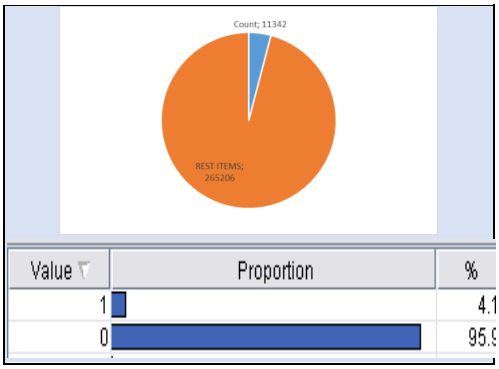
8	المربة	20348	7.358	

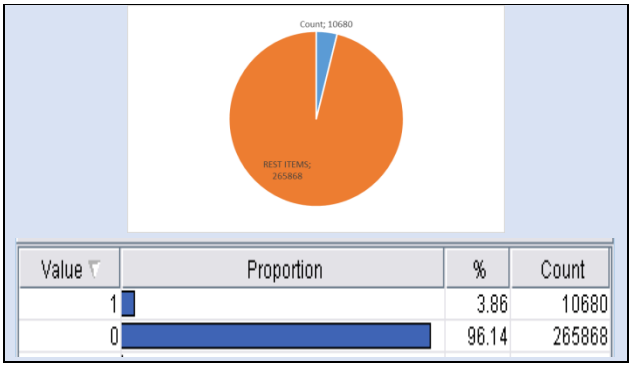
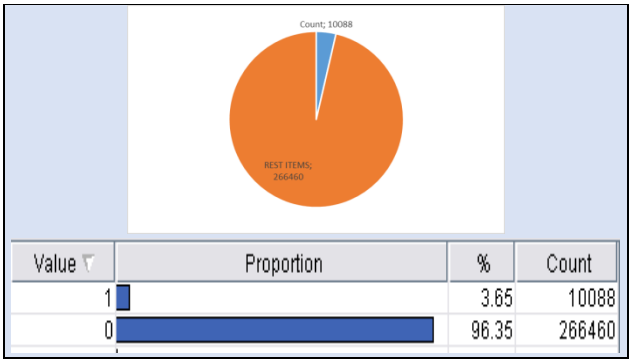
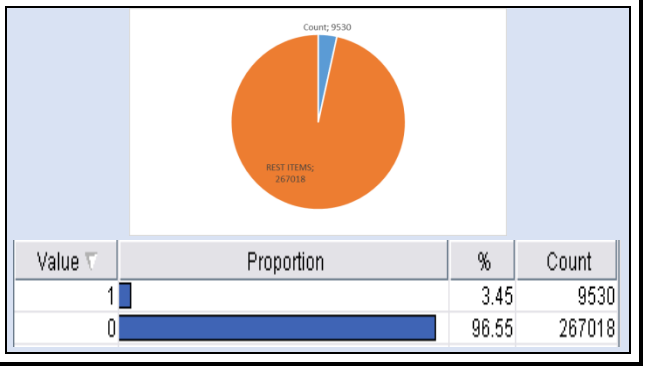
No	Item	Count	Percent	CHART
9	الطحنية	20340	7.355	
10	الحليب	19692	7.121	
11	الصابون	17702	6.4011	

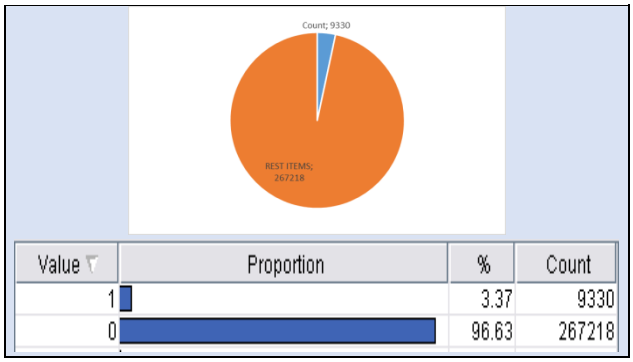
12	الشاي	17004	6.149	

No	Item	Count	Percent	CHART
13	اندومي	16320	5.9013264	
14	المكرونة	14496	5.2418	
15	الشعرية	12762	4.61475042	

16	الزبادي	13242	4.788319	
17	اللحمة	12588	4.5518	
18	الدقيق	12810	4.6321	
19	التانك	11098	4.013	

20	الزيت	11342	4.101	

No	Item	Count	Percent	CHART
21	بيض	10680	3.862	
22	الصلصة	10088	3.647829672	
23	الشبس	9530	3.4461	

24	الخبز	9330	3.37374	

الشكل (12.3) يوضح مخططات بيع واحصاء السلع

7.3 طريقة تجهيز البيانات لتعمل علي برنامجWeka

افترض أن لديك مجموعة من البيانات وتريد بناء شجرة القرارات منها ، أولا لابد من تهيئة هذه البيانات واطلاق المستكشف والقيام بتحميل البيانات فيه ، ثانيا لابد من إختيار طريقة بناء شجرة القرار وبناء الشجرة وتفسير المخرجات. في المستكشف يمكنك التنقل ذهاباً وايابا بين النتائج التي تحصلت عليها، تقييم النماذج التي تم بناءها على مختلف قواعد البيانات، تصور بشكل تخطيطي كلتا النماذج وقواعد البيانات نفسها والتي تتضمن أخطاء التجميع التي يمكن أن يصنعها النموذج.

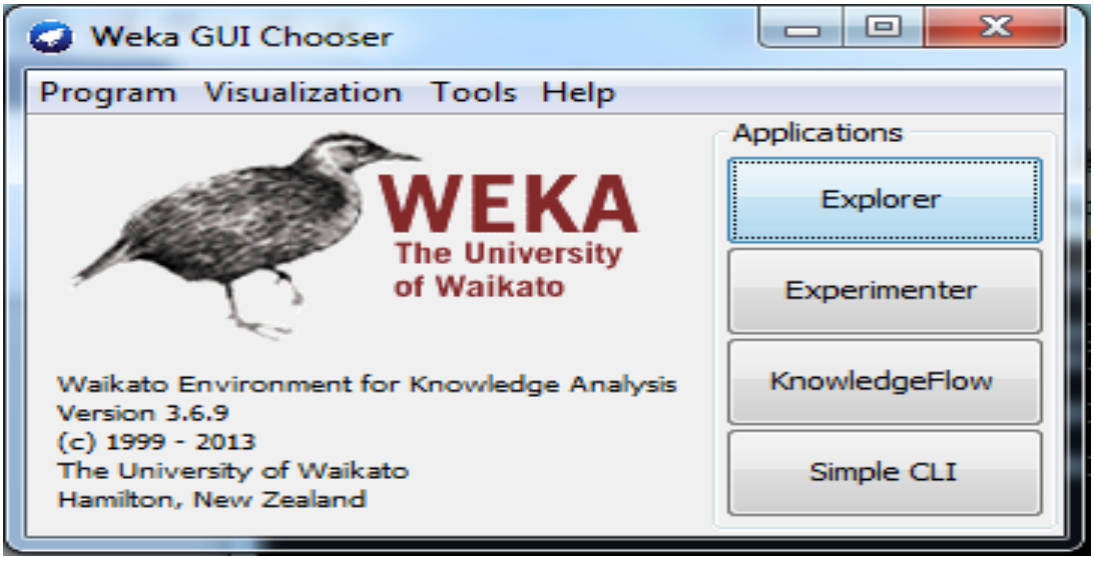
1.7.3 تهيئة البيانات (Weka)

فقد قام الباحث في هذة المرحلة بتحويل البيانات من ملف EXCEL الي صورة (Comma Separated Value CSV) لكي تعمل علي واجهة WEKA غالبا ما يتم عرض البيانات في صورة (Excel Sheet) أو قاعدة بيانات ، على أي حال الطريقة الرئيسة للتخزين في

(Weka) هي في صورة(ARFF) ويمكن بكل سهولة التحويل من صورة برنامج الجدولة إلي صورة(ARFF)معظم ملفات(ARFF)تتكون من قائمة من الحالات وقيم وخواص لكل من الحالات التي تم الفصل بينها بالفاصلة كما ذكرنا سابقا. معظم برامج الجدولة وبرامج قواعد البيانات تسمح بتصدير البيانات في ملف (Comma Separated Value CSV) في شكل مجموعة من السجلات مع وجود الفاصلة بين القيم .

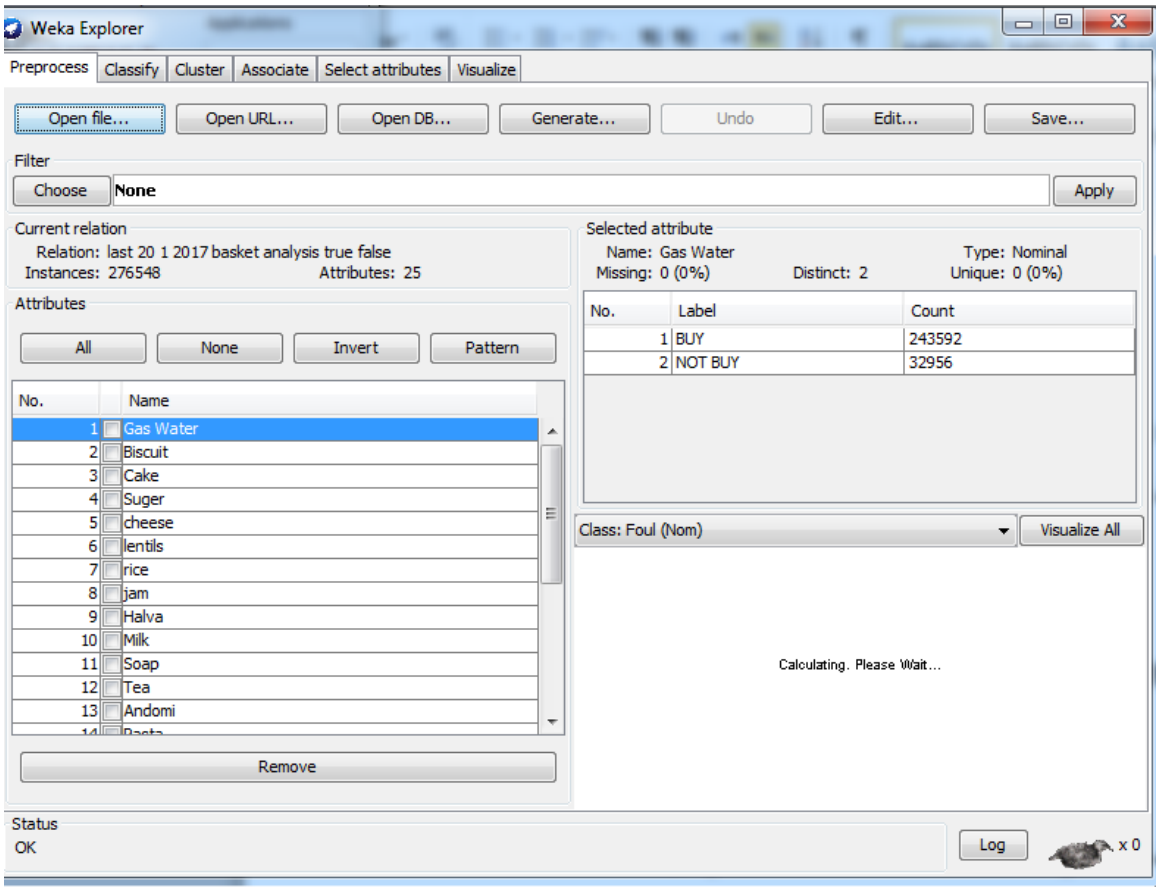
2.7.3 تحميل البيانات في المستكشف

نبدأ بتحميل البيانات في المستكشف والبداية في تحميلها ، في البدء نقوم بتشغيل برنامج (Weka) للحصول على واجهة الإختيار الرسومية كما هو موضح بالشكل **(13.3)**



الشكل (13.3) :واجهة الاختيار الرسومية. (Weka)

3.7.3 الخطوة الثانية : نقوم بإختيار المستكشف Explore من الأربع خيارات المتاحة على الجهة اليمنى من الشكل وتظهر لنا واجهة المستكشف كما هو موضح في الشكل **الشكل (14.3)**

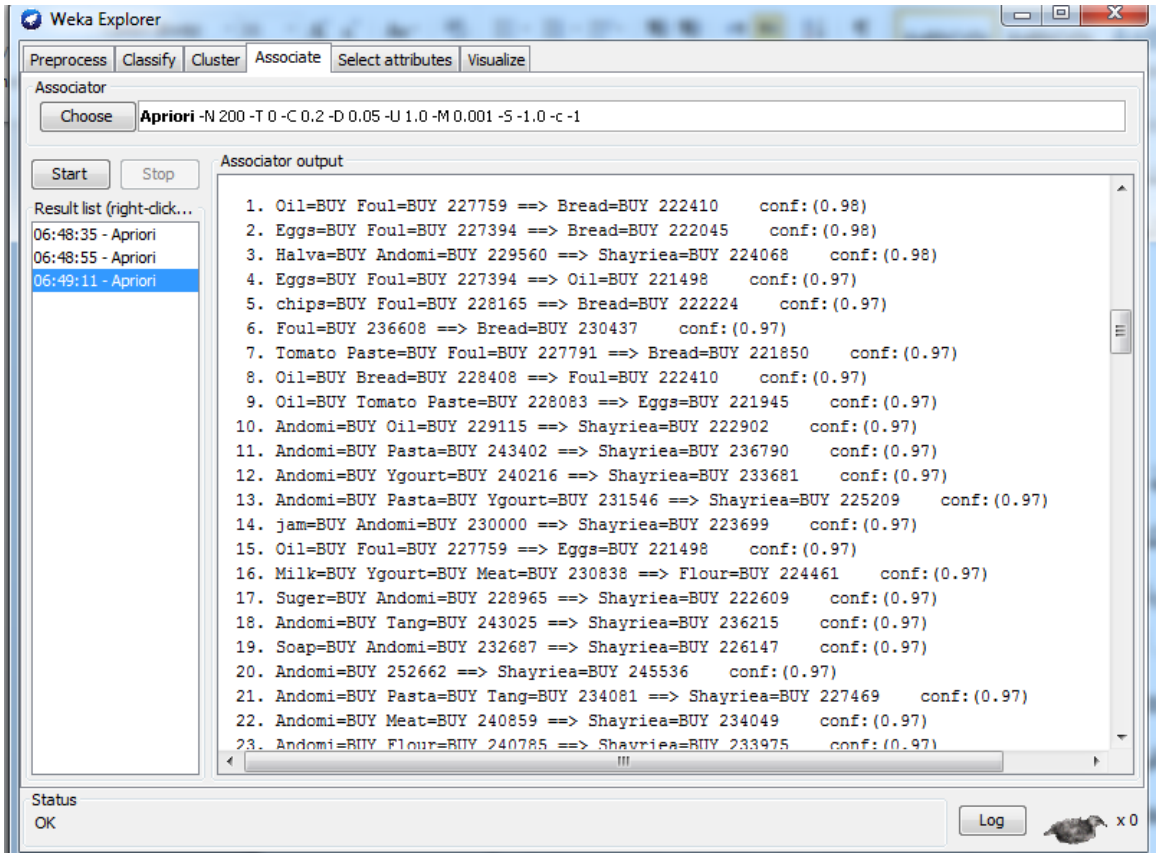


الشكل (14.3) : واجهة المستكشف فيweka.

في الواقع الشكل يوضح ما يتم عرضه بعد القيام بتحميل البيانات. الست قوائم الأولى هي عبارة عن العلميات الأساسية التي يدعمها المستكشف، حاليا نحن في مرحلة المعالجة المبدئية، اضغط زر "افتح ملف" الذي من خلاله يمكن ان تختار الملف الذي تريد تحميله. إختيار الملف ذو الامتداد (.ARFF)بعد ان يتم تحميل البيانات نقوم بتحديد الصفات (attributes) التي نريد إجراء عملية التجميع عليها ويوجد أيضا خيار إختيارها جميعها واجراء التجارب عليها. ومن ثم بالرجوع للست

قوائم الأولى نقوم بإختيار associate ثم ، start بعدها يقوم البرنامج بعمل توليد للنتائج

والمخرجات كما هو موضح في الشكل(15.3) :



الشكل (15.3) يوضح النتائج في برنامج weka

الباب الرابع

النتائج ومناقشتها

1.4 مقدمة

قام الباحث بإجراء العديد من التجارب وتوصل الي مجموعة من النتائج. وقام باختيار افضل النتائج للمناقشة وهي التي تتكون من الأربعة سلع التي تباع معا بعد إجراء العديد من التجارب علي برنامج(Weka) وذلك بتغيير قيم(Confidence)و (Support) ،وتغيير الصفات التي يتم إختيارها في مرحلة التهيئة للبيانات (Preprocessing) تم الحصول على كمية كبيرة من القواعد(Rules) وبعد تحميل تلك القواعد وجد أن بعضها يصنف على أنه بديهي (Trivial) والبعض الاخر تتباين درجة منطقيته علي أساس قيمة ال (Confidence).بعد تحميل النتائج المتحصل عليها كافة، قمنا بإختيار القواعد التي توضح قوة إرتباط الاصناف ببعضها من حيث التأثير علي مستوي اكثر الأصناف التي تباع مع بعض وتوصلت الي مجموعة من النتائج. وقام الباحث باختيار افضل النتائج للمناقشة وهي التي تتكون من الأربعة سلع التي تباع معا مما يساعد في إتخاذ القرارات الملائمة لتطوير وتحسين المبيعات لكل صنف عن طريق إتباع منهجية معينة (كزيادة في توفير بعضي الأصناف التي تباع سويا وتوفرها جميعها بصورة دائمة). ووجد الباحث افضل خواص تعطي افضل نتائج وقوانين هي الموجودة في الصف رقم ثلاثة وقام الباحث بمناقشة جزء منها كما مبين في الجدول (1.4) يوضح مجموعة التجارب التي قام الباحث بإجراها علي الخواص في خوارزمية (a priori)

No	n.Rules	Support	Confidence
1	10	0.1	0.9
2	20	0.01	0.2
3	200	0.001	0.2

جدول (1.4) يوضح مجموعة التجارب في الخواص Support& Confidence

ومن النتائج (لاحظ الباحث انه كلما زادت نسبة(support) يقل عدد(item set) وكلما كانت عدد(item set) الناتجة قليل كلما كانت العلاقات بينهم منطقية اكثر، التغيير في كلاً من قيم (confidence)و (support)بزيادة أو نقصان يؤدي إلى ظهور قواعد مختلفة ذات درجات مختلفة من المنطقية . كما ظهرت كمية كبيرة من النتائج البديهية والنتائج الغير منطقية عند تحديد جميع الصفات لإيجاد العلاقات بينهم مما جعلنا نقوم بتحديد صفات معينة واجراء العلميات عليها لاكتشاف علاقات اكثر منطقية بين الأصناف. كما ظهر في جدول (2.5): جزء من النتائج المتحصل عليها.

2.4 وكانت النتائج كالاتي:

مجموعة النتائج التي تتكون من قواعد ذات أربعة سلع وتحدد قوة الارتباط بينها كما موضحة في جدول (2.4)

الرقم	مجموعة العناصر	العنصر المرتبط بها	confidence
1	Oil=BUY Pasta=BUY Ygourt=BUY 231546 ==>	Bread=BUY 225209	(0.97)
2	Milk=BUY Ygourt=BUY Meat=BUY 230838 ==>	Flour=BUY 224461	(0.97)

3	Oil=BUY Pasta=BUY Sugar=BUY 234081 ==>	Bread=BUY 227469	(0.97)
4	Oil=BUY Ygourt=BUY Sugar=BUY 230975 ==>	Bread=BUY 224440	(0.97)
5	Oil=BUY Pasta=BUY Meat=BUY 232960 ==>	Bread=BUY 226348	(0.97)
6	Oil=BUY Ygourt=BUY Flour=BUY 229821 ==>	Bread=BUY 22328	(0.97)
7	Oil=BUY Ygourt=BUY Meat=BUY 229809 ==>	Bread=BUY 223274	(0.97)
8	Oil=BUY Pasta=BUY Flour=BUY 232164 ==>	Bread=BUY 225552	(0.97)
9	Pasta=BUY Ygourt=BUY Flour=BUY 232514 ==>	Meat=BUY 225874	(0.97)
10	jam=BUY Ygourt=BUY Meat=BUY 229902 ==>	Flour=BUY 223327	(0.97)
11	Oil=BUY Meat=BUY Flour=BUY 232691 ==>	Bread=BUY 225881	(0.97)
12	Oil=BUY Flour=BUY Sugar=BUY 231544 ==>	Bread=BUY 224734	(0.97)
13	Oil=BUY Meat=BUY Sugar=BUY 231538 ==>	Bread=BUY 224728	(0.97)
14	Ygourt=BUY Meat=BUY Andomi=BUY 231178 ==>	Flour=BUY 224361	(0.97)
15	Ygourt=BUY Flour=BUY Andomi=BUY 231207 ==>	Meat=BUY 224361	(0.97)
16	Flour=BUY Sugar=BUY Andomi=BUY 231100 ==>	Meat=BUY 224254	(0.97)
17	Milk=BUY Meat=BUY Sugar=BUY 231303 ==>	Flour=BUY 224386	(0.97)
18	Tang=BUY Ygourt=BUY Meat=BUY 228574 ==>	Flour=BUY 221700	(0.97)
19	Milk=BUY Flour=BUY Sugar=BUY 232355 ==>	Meat=BUY 224386	(0.97)
20	Bread=BUY Ygourt=BUY Meat=BUY 233823 ==>	Flour=BUY 226473	(0.97)
21	Ygourt=BUY Flour=BUY Sugar=BUY 242314 ==>	Meat=BUY 234692	(0.97)
22	Bread=BUY Ygourt=BUY Flour=BUY 233835 ==>	Meat=BUY 226473	(0.97)
23	Pasta=BUY Bread=BUY Flour=BUY 234056 ==>	Meat=BUY 226684	(0.97)
24	Pasta=BUY Ygourt=BUY Meat=BUY 233224 ==>	Flour=BUY 225874	(0.97)
25	Pasta=BUY Flour=BUY Sugar=BUY 233434 ==>	Meat=BUY 226062	(0.97)

26	Meat=BUY Sugar=BUY Andomi=BUY 231597 ==>	Flour=BUY 224254	(0.97)
27	jam=BUY Meat=BUY Sugar=BUY 230445 ==>	Flour=BUY 223128	(0.97)
28	Oil=BUY Pasta=BUY Flour=BUY 232164 ==>	Meat=BUY 224792	(0.97)
29	Soap=BUY Ygourt=BUY Flour=BUY 232841 ==>	Meat=BUY 225425	(0.97)
30	Soap=BUY Ygourt=BUY Meat=BUY 232842 ==>	Flour=BUY 225425	(0.97)
31	Oil=BUY Ygourt=BUY Meat=BUY 229809 ==>	Flour=BUY 222459	(0.97)
32	Oil=BUY Ygourt=BUY Flour=BUY 229821 ==>	Meat=BUY 222459	(0.97)
33	Tea=BUY Ygourt=BUY Flour=BUY 229752 ==>	Meat=BUY 222345	(0.97)
34	Tang=BUY Meat=BUY Sugar=BUY 229714 ==>	Flour=BUY 222286	(0.97)
35	jam=BUY Ygourt=BUY Flour=BUY 230803 ==>	Meat=BUY 223327	(0.97)
36	Halva=BUY Ygourt=BUY Flour=BUY 230244 ==>	Meat=BUY 222732	(0.97)
37	Oil=BUY Bread=BUY Meat=BUY 234049 ==>	Pasta=BUY 226348	(0.97)
38	Tea=BUY Ygourt=BUY Meat=BUY 229914 ==>	Flour=BUY 222345	(0.97)
39	Cake=BUY Ygourt=BUY Flour=BUY 229495 ==>	Meat=BUY 221916	(0.97)
40	Oil=BUY Ygourt=BUY Meat=BUY 229809 ==>	Pasta=BUY 222184	(0.97)
41	Tang=BUY Ygourt=BUY Flour=BUY 229322 ==>	Meat=BUY 221700	(0.97)
42	Halva=BUY Ygourt=BUY Meat=BUY 230392 ==>	Flour=BUY 222732	(0.97)
43	Cake=BUY Ygourt=BUY Meat=BUY 229647 ==>	Flour=BUY 221916	(0.97)
44	Pasta=BUY Bread=BUY Sugar=BUY 235444 ==>	Oil=BUY 227469	(0.97)
45	Oil=BUY Meat=BUY Flour=BUY 232691 ==>	Pasta=BUY 224792	(0.97)
46	Oil=BUY Meat=BUY Sugar=BUY 231538 ==>	Pasta=BUY 223639	(0.97)
47	Milk=BUY Flour=BUY Sugar=BUY 232355 ==>	Meat=BUY 224386	(0.97)
48	Bread=BUY Meat=BUY Sugar=BUY 234748 ==>	Flour=BUY 226660	(0.97)

49	Bread=BUY Meat=BUY Sugar=BUY 234748 ==>	Flour=BUY 226660	(0.97)
50	Pasta=BUY Meat=BUY Sugar=BUY 234150 ==>	Flour=BUY 226062	(0.97)
51	Oil=BUY Bread=BUY Flour=BUY 233975 ==>	Meat=BUY 225881	(0.97)
52	Pasta=BUY Bread=BUY Meat=BUY 234852 ==>	Flour=BUY 226684	(0.97)
53	Soap=BUY Meat=BUY Sugar=BUY 233637 ==>	Flour=BUY 225484	(0.97)
54	Oil=BUY Bread=BUY Meat=BUY 234049 ==>	Flour=BUY 225881	(0.97)
55	.Oil=BUY Meat=BUY Sugar=BUY 231538 ==>	Flour=BUY 223450	(0.97)
56	.Soap=BUY Flour=BUY Sugar=BUY 233651 ==>	Meat=BUY 225484	(0.97)
57	Oil=BUY Flour=BUY Sugar=BUY 231544 ==>	Meat=BUY 223450	(0.97)
58	Oil=BUY Pasta=BUY Meat=BUY 232960 ==>	Flour=BUY 224792	(0.96)
59	Tea=BUY Meat=BUY Sugar=BUY 230985 ==>	Flour=BUY 222878	(0.96)
60	Tea=BUY Flour=BUY Sugar=BUY 231028 ==>	Meat=BUY 222878	(0.96)
61	jam=BUY Flour=BUY Sugar=BUY 231347 ==>	Meat=BUY 223128	(0.96)
62	Halva=BUY Meat=BUY Sugar=BUY 230018 ==>	Flour=BUY 221818	(0.96)
63	Cake=BUY Meat=BUY Sugar=BUY 230647 ==>	Flour=BUY 222367	(0.96)
64	Halva=BUY Flour=BUY Sugar=BUY 230081 ==>	Meat=BUY 221818	(0.96)
65	Oil=BUY Bread=BUY Flour=BUY 233975 ==>	Pasta=BUY 225552	(0.96)
66	Pasta=BUY Bread=BUY Meat=BUY 234852 ==>	Oil=BUY 226348	(0.96)
67	Oil=BUY Ygourt=BUY Flour=BUY 229821 ==>	Pasta=BUY 221474	(0.96)
68	Pasta=BUY Bread=BUY Flour=BUY 234056 ==>	Oil=BUY 225552	(0.96)
69	Tang=BUY Flour=BUY Sugar=BUY 230668 ==>	Meat=BUY 222286	(0.96)
70	Pasta=BUY Bread=BUY Ygourt=BUY 233713 ==>	Oil=BUY 225209	(0.96)
71	Oil=BUY Bread=BUY Sugar=BUY 236215 ==>	Pasta=BUY 227469	(0.96)

72	Oil=BUY Flour=BUY Sugar=BUY 231544 ==>	Pasta=BUY 222923	(0.96)
73	Oil=BUY Ygourt=BUY Sugar=BUY 230975 ==>	Pasta=BUY 222305	(0.96)
74	Pasta=BUY Ygourt=BUY Meat=BUY 233224 ==>	Bread=BUY 224351	(0.96)
75	Pasta=BUY Ygourt=BUY Sugar=BUY 232816 ==>	Bread=BUY 223943	(0.96)
76	Pasta=BUY Ygourt=BUY Flour=BUY 232514 ==>	Bread=BUY 223641	(0.96)
77	Meat=BUY Flour=BUY Andomi=BUY 233275 ==>	Ygourt=BUY 224361	(0.96)

جدول (2.4) يوضح جزء من النتائج المتحصل عليها

3.4 مناقشة النتائج

1- أظهرت النتائج ان اكثر سلعة أساسية هي الدقيق حيث ظهر في61 قانون من اصل 77قانون في المعاملات حيث ظهر الدقيق 26مرة كعنصر مرتبط

2- كما أظهرت النتائج أيضا ان اللحمة من السلع الأساسية ويأتي في المرتبة الثانية من حيث الظهور حيث ظهرت في 61 قانون من اصل 77 قانون حيث ظهرت اللحمة **24**مرة كعنصر مرتبط

3- كما أظهرت النتائج أيضا ان الخبز من السلع متوسطة الشراء من حيث عدد مرات الظهور حيث ظهر في 28 قانون من اصل 77 قانون حيث ظهر الخبز 14مرة كعنصر مرتبط

4- كما أظهرت النتائج أيضا ان الزيت من السلع المتوسطة الشراء من حيث الظهور حيث ظهر في 31 قانون من اصل 77 قانون حيث ظهر الزيت4 مرات كعنصر مرتبط.

5- كما أظهرت النتائج أيضا ان المعكرونة من السلع المتوسطة الشراء من حيث الظهور حيث ظهر في 28 قانون من اصل 77 قانون حيث ظهرت المعكرونة7 مرات كعنصر مرتبط.

6- كما أظهرت النتائج أيضا ان الزبادي من السلع المتوسطة الشراء من حيث الظهور حيث ظهر في 31 قانون من اصل 77 قانون حيث ظهرت الزبادي مرة واحدة كعنصر مرتبط.

7- كما أظهرت النتائج أيضا ان الصابون والمربة والطحنية والشاي والحليب والكيك والأرز والعدس والتانك من اقل السلع من حيث الظهور حيث ظهرت بنسب قليلة.

8- النتائج التي توصل لها الباحث إذا اشتري الناس الزيت والمعكرونة والزبادي سوف يشتروا الخبز بنسبة 97% ، وأيضا ان 97% من الأشخاص الذين اشتروا الخبز والزبادي والدقيق اشتروا أيضا لحمة، وهنالك قاعدة اخري من هذا القبيل قد يكون 96% من الأشخاص الذين اشتروا زيت ومعكرونة ولحمة اشتروا والدقيق، وأيضا بنسبة 96% من الأشخاص الذين اشتروا شاي ولحمة وسكر اشتروا دقيق.

الباب الخامس

الخاتمة والتوصيات

1.5 الخاتمة

بعد الإطلاع على نتائج البحث وتحليلها ظهرت أهمية قاعدة البيانات بالنسبة للمتاجر ،وبالتالي ظهرت الحوجة إلي تحميل بيانات المتاجر للإستفادة منها في صنع القرارات، حيث أن إجراء العديد من التجارب في هذا المجال يمكن المؤسسات التجارية من تطوير الخطة المتبعة وزيادة الكفاء التسويقية ، وكما توضح من البحث والأبحاث السابقة أن تنقيب البيانات واستخدام قوعد الارتباط من الطرق الحديثة ذات الكفاءة العالية فتحميل وتجميع وتصنيف البيانات والتوصل إلى علاقات تزيد من إمكانية التحكم في النظام وتقديم أفضل ما يمكن للوصول للأهداف المرجوة ونتيجة لذلك استنتج الباحث ان اهل منطقة الثورة كثيرا ما يشترون اللحم والخبز والزيت والدقيق والسكر والزبادي. ومن النتائج التي توصل لها الباحث إذا اشتري الناس الزيت والمعكرونة والزبادي سوف يشتروا الخبز بنسبة 97% ، وأيضا ان 97% من الأشخاص الذين اشتروا الخبز والزبادي والدقيق اشتروا أيضا لحمة، وهنالك قاعدة اخري من هذا القبيل قد يكون 96% من الأشخاص الذين اشتروا زيت ومعكرونة ولحمة اشتروا والدقيق، وأيضا بنسبة 96% من الأشخاص الذين اشتروا شاي ولحمة وسكر اشتروا دقيق.

2.5 التوصيات

1- تطبيق الخوارزمية علي بيانات كل سوبرماركت منفصلة عن الأخرى ومن ثم مقارنة النتائج.

2- لزيادة كفاءة النتائج المتحصل عليها من هذا البحث اوصي بإضافة السلع التي تعزر علي اضافتها ومن ثم تطبيق الخوارزمية عليها وقياس مدي الارتباط بين جميع السلع.

3- تطبيق خوارزميات اخري علي هذه البيانات ومقارنتها مع النتائج التي توصل الباحث اليها.

المراجع:.

[1] Jiawei Han and Micheline Kamber, data mining concepts and techniques third edition, ISBN 978-0-12-381479-1; 2012.

[2] Ian H.written&E.frank. data mining practical machine learning tool and techniques second edition, ISBN: 0-12-088407-0;2005.

[3] Rashmi Jha , Association Rules Mining for Business Intelligence, International Journal of Scientific and Research Publications, Volume 4, Issue 5, May 2014 1 ISSN 2250-3153, 2014.

[4] Phani Prasad J1and Murlidher Mourya2, A Study on Market Basket Analysis Using a Data Mining Algorithm ,International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001,2008. Certified Journal, Volume 3, Issue 6, 2013.

[5] Mehmet Aydin Ulas ,MARKET BASKET ANALYSIS FOR DATA MINING, Submitted to the Institute for Graduate Studies in Science and Engineering in partial fulfillment of the requirements for the degree of Master of Science in Computer Engineering, 2001.

[6] stackoverflow.com,10/7/2017,8:30 am.

[7] Alia abul Aziz Bashir, Amin Ali, moshtaha Ibrahim Mohammed,

Exploration the performance of the academic student using data mining

Techniques, a project submitted to obtain a bachelor of honors in computer science, university of Sudan, august 2014

[8] Roshan Gangurde1, Dr. Binod Kumar2, Dr. S. D. Gore3, Building Prediction Model using Market Basket Analysis, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 2, February 2017, Website: www.ijircce.com

[9] Prof. Kavitha Venkatachari, MARKET BASKET ANALYSIS USING FP GROWTH AND APRIORI ALGORITHM: A CASE STUDY OF MUMBAI RETAIL STORE, BVIMSR's Journal of Management Research, Vol. 8 Issue – 1 : April : 2016

الملاحق

مصدر البيانات 2: (الدومة سوبر ماركت)

تاريخ البيانات التي تم جمعها:

ديسمبر 2014.وحتي أكتوبر 2015

عدد كل سجلات المعاملات في قاعدة البيانات:

1592200 سجل.

عدد الصفقات:

318440 معاملة.

عدد اصناف السلع في المتجر:

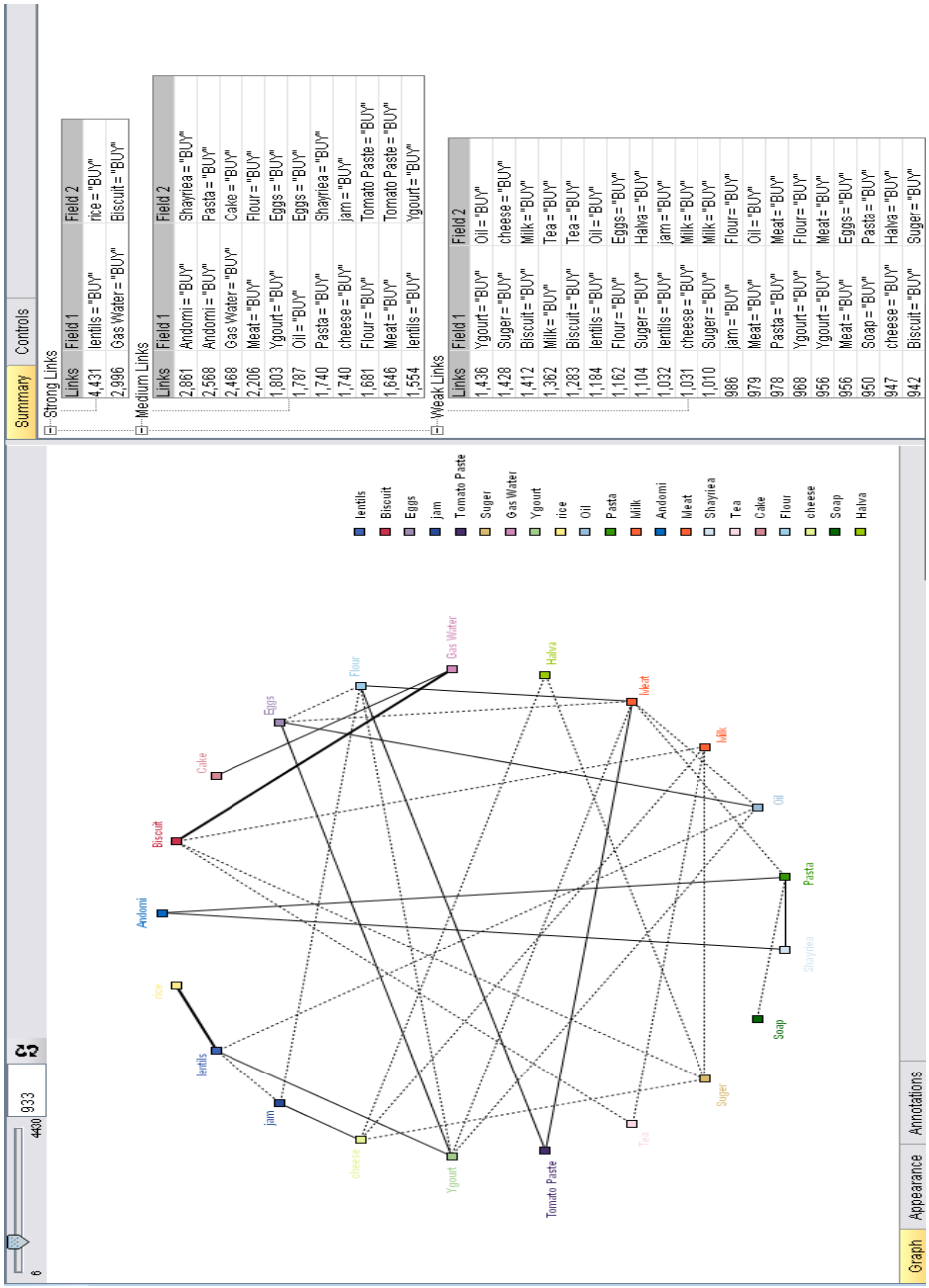
12،051 صنف

	A	B	C	D	E
123754	85673	SHAIREA			
123755	85674	HALVA			
123756	85674	YGOURT			
123757	85675	MEAT			
123758	85676	YGOURT			
123759	85677	WATER GAS			
123760	85678	TOMATO			
123761	85679	RICE			
123762	85680	BISCUIT			
123763	85680	MILK			
123764	85681	WATER GAS			
123765	85682	SOAP			
123766	85683	CHEESE			
123767	85684	ANDOMI			
123768	85685	SOAP			
123769	85686	PASTA			
123770	85687	TEA			
123771	85688	LENTILS			
123772	85688	RICE			
123773	85688	TOMATO			
123774	85689	JAM			
123775	85690	JAM			
123776	85690	SHAIREA			
123777	85690	YGOURT			
123778	85691	SUGER			

شكل يوضح يوضح بيانات سوبر ماركت الدومة في صورتها الاولي

Field	Proportion True	%	Count
Gas Water		9.13	13380
Biscuit		8.91	13054
Suger		8.85	12959
cheese		8.55	12528
lentils		8.07	11827
Cake		7.76	11371
jam		7.42	10869
Halva		7.34	10757
Milk		7.11	10420
rice		6.58	9638
Soap		6.39	9359
Tea		6.26	9163
Andomi		5.9	8636
Pasta		5.25	7694
Ygourt		4.71	6903
Shayriea		4.61	6755
Meat		4.56	6676
Flour		4.53	6629
Oil		4.1	6011
Tang		3.96	5804
Eggs		3.81	5586
Tomato Paste		3.69	5402
chips		3.45	5054
Bread		3.36	4923

شكل يوضح بيانات سوبرماركت الدومة اول عمود به اسم الصنف والثاني به نسبة بيع الصنف والثالث به كم مرة تم بيع الصنف



شكل يوضح قوة العلاق بين كل عنصرين ومقسم الي ثلاثة علاقة قوية ومتوسطة وضعيفة

1-SOPPORT 0.01 CONFIDENCE 0.9 نتائج التجربة الاولي

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: last 20 1 2017 basket analysis true false

Instances: 276548

Atangributes: 24

Gas Water

Biscuit

Cake

sugar

cheese

lentils

rice

jam

Halva

Milk

Soap

Tea

Oil

Pasta

bread

Ygourt

Meat

Flour

Sugar

Andomi

Eggs

Tomato Paste

chips

shayriea

=== Associator model (full training set) ===

Apriori

=======

Minimum support: 0.9 (248893 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 2

Generated sets of large itemsets:

Size of set of large itemsets L(1): 16

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

Best rules found:

1. Flour=BUY 263738 ==> Meat=BUY 255356 <conf:(0.97)
2. Meat=BUY 263960 ==> Flour=BUY 255356 <conf:(0.97)
3. Flour=BUY 263738 ==> Sugar=BUY 253640 <conf:(0.96)
4. Ygourt=BUY 263306 ==> Sugar=BUY 253008 <conf:(0.96)
5. Meat=BUY 263960 ==> Sugar=BUY 253582 <conf:(0.96)
6. Ygourt=BUY 263306 ==> Meat=BUY 252560 <conf:(0.96)
7. Ygourt=BUY 263306 ==> Flour=BUY 252412 <conf:(0.96)
8. Flour=BUY 263738 ==> Ygourt=BUY 252412 <conf:(0.96)
9. Meat=BUY 263960 ==> Ygourt=BUY 252560 <conf:(0.96)
10. Sugar=BUY 265450 ==> Flour=BUY 253640 <conf:(0.96)

2-SOPPORT 0.001 CONFIDENCE 0.2 نتائج التجربة الثانية

=== Run information ===

Scheme: weka.associations.Apriori -N 20 -T 0 -C 0.2 -D 0.05 -U 1.0 -M 0.01 -S -1.0 -c -1

Relation: last 20 1 2017 basket analysis true false

Instances: 276548

Atangributes: 24

Gas Water

Biscuit

Cake

sugar

cheese

lentils

rice

jam

Halva

Milk

Soap

Tea

Oil

Pasta

bread

Ygourt

Meat

Flour

tang

Andomi

Eggs

Tomato Paste

chips

shayriea

=== Associator model (full training set) ===

Apriori

=======

Minimum support: 0.85 (235066 instances)

Minimum metric <confidence>: 0.2

Number of cycles performed: 3

Generated sets of large itemsets:

Size of set of large itemsets L(1): 24

Size of set of large itemsets L(2): 66

Size of set of large itemsets L(3): 10

Best rules found:

1. Oil=BUY Pasta=BUY 243402 ==> bread=BUY 236790 <conf:(0.97)
2. Oil=BUY Sugar=BUY 243025 ==> bread=BUY 236215 <conf:(0.97)
3. Oil=BUY 252662 ==> bread=BUY 245536 <conf:(0.97)
4. Ygourt=BUY Flour=BUY 252412 ==> Meat=BUY 244790 <conf:(0.97)
5. Pasta=BUY Flour=BUY 243204 ==> Meat=BUY 235832 <conf:(0.97)
6. Ygourt=BUY Meat=BUY 252560 ==> Flour=BUY 244790 <conf:(0.97)
7. Flour=BUY 263738 ==> Meat=BUY 255356 <conf:(0.97)> lift:(1.01)
8. Meat=BUY 263960 ==> Flour=BUY 255356 <conf:(0.97)

9. Meat=BUY Sugar=BUY 253582 ==> Flour=BUY 245258 <conf:(0.97)

10. Flour=BUY Sugar=BUY 253640 ==> Meat=BUY 245258 <conf:(0.97)

11. bread=BUY Flour=BUY 244524 ==> Meat=BUY 236430 <conf:(0.97)

12. bread=BUY Meat=BUY 244598 ==> Flour=BUY 236430 <conf:(0.97)

13. Pasta=BUY Meat=BUY 244000 ==> Flour=BUY 235832 <conf:(0.97)

14. Soap=BUY Flour=BUY 243550 ==> Meat=BUY 235383 <conf:(0.97)

15. Soap=BUY Meat=BUY 243616 ==> Flour=BUY 235383 <conf:(0.97)

16. Pasta=BUY bread=BUY 245294 ==> Oil=BUY 236790 <conf:(0.97)

17. Oil=BUY bread=BUY 245536 ==> Pasta=BUY 236790 <conf:(0.96)

18. Pasta=BUY 254442 ==> bread=BUY 245294 <conf:(0.96)

19. Oil=BUY 252662 ==> Pasta=BUY 243402 <conf:(0.96)

20. Andomi=BUY 250953 ==> Sugar=BUY 241739 <conf:(0.96)

3-SOPPORT 0.0001 CONFIDENCE 0.2 نتائج التجربة الثالثة

3-=== Run information ===

Scheme: weka.associations.Apriori -N 200 -T 0 -C 0.2 -D 0.05 -U 1.0 -M 0.001 -S -1.0 -c -1

Relation: last 20 1 2017 basket analysis true false

Instances: 276548

Atangributes: 24

Gas Water

Biscuit

Cake

sugar

cheese

lentils

rice

jam

Halva

Milk

Soap

Tea

Oil

Pasta

Bread

Ygourt

Meat

Flour

tang

Andomi

Eggs

Tomato Paste

chips

Shayriea

=== Associator model (full training set) ===

Apriori

=======

Minimum support: 0.8 (221238 instances)

Minimum metric <confidence>: 0.2

Number of cycles performed: 4

Generated sets of large itemsets:

Size of set of large itemsets L(1): 24

Size of set of large itemsets L(2): 199

Size of set of large itemsets L(3): 225

Size of set of large itemsets L(4): 54

Best rules found:

1. Halva=BUY Oil=BUY 229560 ==> Bread=BUY 224068 conf:(0.98)
2. Andomi=BUY Tomato Paste=BUY 228083 ==> Eggs=BUY 221945 conf:(0.97)
3. Oil=BUY Andomi=BUY 229115 ==> Bread=BUY 222902 conf:(0.97)
4. Oil=BUY Pasta=BUY 243402 ==> Bread=BUY 236790 conf:(0.97)
5. Oil=BUY Ygourt=BUY 240216 ==> Bread=BUY 233681 conf:(0.97)
6. Oil=BUY Pasta=BUY Ygourt=BUY 231546 ==> Bread=BUY 225209 conf:(0.97)
7. jam=BUY Oil=BUY 230000 ==> Bread=BUY 223699 conf:(0.97)
8. Milk=BUY Ygourt=BUY Meat=BUY 230838 ==> Flour=BUY 224461 conf:(0.97)
9. Tang=BUY Oil=BUY 228965 ==> Bread=BUY 222609 conf:(0.97)
10. Oil=BUY Sugar=BUY 243025 ==> Bread=BUY 236215 conf:(0.97)
11. Soap=BUY Oil=BUY 232687 ==> Bread=BUY 226147 conf:(0.97)
12. Oil=BUY 252662 ==> Bread=BUY 245536 conf:(0.97)
13. Oil=BUY Pasta=BUY Sugar=BUY 234081 ==> Bread=BUY 227469 conf:(0.97)

14. Oil=BUY Meat=BUY 240859 ==> Bread=BUY 234049 conf:(0.97)

15. Oil=BUY Flour=BUY 240785 ==> Bread=BUY 233975 conf:(0.97)

16. Andomi=BUY Shayriea=BUY 228408 ==> Eggs=BUY 221947 conf:(0.97)

17. Oil=BUY Ygourt=BUY Sugar=BUY 230975 ==> Bread=BUY 224440 conf:(0.97)

18. Andomi=BUY chips=BUY 228150 ==> Eggs=BUY 221689 conf:(0.97)

19. Oil=BUY Pasta=BUY Meat=BUY 232960 ==> Bread=BUY 226348 conf:(0.97)

20. Oil=BUY Ygourt=BUY Flour=BUY 229821 ==> Bread=BUY 223286 conf:(0.97)

21. Oil=BUY Ygourt=BUY Meat=BUY 229809 ==> Bread=BUY 223274 conf:(0.97)

22. Oil=BUY Pasta=BUY Flour=BUY 232164 ==> Bread=BUY 225552 conf:(0.97)

23. Flour=BUY Andomi=BUY 240121 ==> Meat=BUY 233275 conf:(0.97)

24. Pasta=BUY Ygourt=BUY Flour=BUY 232514 ==> Meat=BUY 225874
conf:(0.97)

25. jam=BUY Ygourt=BUY Meat=BUY 229902 ==> Flour=BUY 223327 conf:(0.97)

26. Milk=BUY Oil=BUY 230549 ==> Bread=BUY 223897 conf:(0.97)

27. Oil=BUY Meat=BUY Flour=BUY 232691 ==> Bread=BUY 225881 conf:(0.97)

28. Oil=BUY Flour=BUY Sugar=BUY 231544 ==> Bread=BUY 224734 conf:(0.97)

29. Oil=BUY Meat=BUY Sugar=BUY 231538 ==> Bread=BUY 224728 conf:(0.97)

30. Ygourt=BUY Meat=BUY Andomi=BUY 231178 ==> Flour=BUY 224361
conf:(0.97)

31. Ygourt=BUY Flour=BUY Andomi=BUY 231207 ==> Meat=BUY 224361
conf:(0.97)

32. Flour=BUY Sugar=BUY Andomi=BUY 231100 ==> Meat=BUY 224254 conf:(0.97)

33. Milk=BUY Meat=BUY 241528 ==> Flour=BUY 234335 conf:(0.97)

34. rice=BUY Flour=BUY 237593 ==> Meat=BUY 230516 conf:(0.97)

35. Eggs=BUY Tomato Paste=BUY 228779 ==> Andomi=BUY 221945 conf:(0.97)

36. Milk=BUY Meat=BUY Sugar=BUY 231303 ==> Flour=BUY 224386 conf:(0.97)

37. Tea=BUY Oil=BUY 232962 ==> Bread=BUY 225994 conf:(0.97)

38. Tang=BUY Ygourt=BUY Meat=BUY 228574 ==> Flour=BUY 221700 conf:(0.97)

39. Ygourt=BUY Flour=BUY 252412 ==> Meat=BUY 244790 conf:(0.97)

40. Eggs=BUY 236570 ==> Andomi=BUY 229426 conf:(0.97)

41. Pasta=BUY Flour=BUY 243204 ==> Meat=BUY 235832 conf:(0.97)

42. Eggs=BUY Shayriea=BUY 228957 ==> Andomi=BUY 221947 conf:(0.97)

43. Cake=BUY Oil=BUY 230332 ==> Bread=BUY 223259 conf:(0.97)

44. Ygourt=BUY Meat=BUY 252560 ==> Flour=BUY 244790 conf:(0.97)

45. jam=BUY Meat=BUY 240244 ==> Flour=BUY 232847 conf:(0.97)

46. Milk=BUY Ygourt=BUY Flour=BUY 231679 ==> Meat=BUY 224461 conf:(0.97)

47. Eggs=BUY chips=BUY 228833 ==> Andomi=BUY 221689 conf:(0.97)

48. Ygourt=BUY Meat=BUY Sugar=BUY 242262 ==> Flour=BUY 234692 conf:(0.97)

49. Meat=BUY Andomi=BUY 240811 ==> Flour=BUY 233275 conf:(0.97)

50. Bread=BUY Ygourt=BUY Meat=BUY 233823 ==> Flour=BUY 226473 conf:(0.97)

51. Ygourt=BUY Flour=BUY Sugar=BUY 242314 ==> Meat=BUY 234692 conf:(0.97)

52. Bread=BUY Ygourt=BUY Flour=BUY 233835 ==> Meat=BUY 226473 conf:(0.97)

53. Pasta=BUY Bread=BUY Flour=BUY 234056 ==> Meat=BUY 226684 conf:(0.97)

54. Pasta=BUY Ygourt=BUY Meat=BUY 233224 ==> Flour=BUY 225874 conf:(0.97)

55. Pasta=BUY Flour=BUY Sugar=BUY 233434 ==> Meat=BUY 226062 conf:(0.97)

56. Meat=BUY Sugar=BUY Andomi=BUY 231597 ==> Flour=BUY 224254 conf:(0.97)

57. jam=BUY Meat=BUY Sugar=BUY 230445 ==> Flour=BUY 223128 conf:(0.97)

58. Oil=BUY Pasta=BUY Flour=BUY 232164 ==> Meat=BUY 224792 conf:(0.97)

59. Flour=BUY 263738 ==> Meat=BUY 255356 conf:(0.97)

60. Soap=BUY Ygourt=BUY Flour=BUY 232841 ==> Meat=BUY 225425 conf:(0.97)

61. Soap=BUY Ygourt=BUY Meat=BUY 232842 ==> Flour=BUY 225425 conf:(0.97)

62. Oil=BUY Ygourt=BUY Meat=BUY 229809 ==> Flour=BUY 222459 conf:(0.97)

63. Oil=BUY Ygourt=BUY Flour=BUY 229821 ==> Meat=BUY 222459 conf:(0.97)

64. Eggs=BUY 236570 ==> Shayriea=BUY 228957 conf:(0.97)

65. Tang=BUY Meat=BUY 239504 ==> Flour=BUY 231796 conf:(0.97)

66. lentils=BUY Meat=BUY 234257 ==> Flour=BUY 226707 conf:(0.97)

67. Tea=BUY Ygourt=BUY Flour=BUY 229752 ==> Meat=BUY 222345 conf:(0.97)

68. Tang=BUY Meat=BUY Sugar=BUY 229714 ==> Flour=BUY 222286 conf:(0.97)

69. Halva=BUY Pasta=BUY 231329 ==> Bread=BUY 223842 conf:(0.97)

70. jam=BUY Ygourt=BUY Flour=BUY 230803 ==> Meat=BUY 223327 conf:(0.97)

71. Meat=BUY 263960 ==> Flour=BUY 255356 conf:(0.97)

72. Andomi=BUY Eggs=BUY 229426 ==> Shayriea=BUY 221947 conf:(0.97)

73. Andomi=BUY Eggs=BUY 229426 ==> Tomato Paste=BUY 221945 conf:(0.97)

74. Halva=BUY Ygourt=BUY Flour=BUY 230244 ==> Meat=BUY 222732
conf:(0.97)

75. Eggs=BUY 236570 ==> chips=BUY 228833 conf:(0.97)

76. Oil=BUY Meat=BUY 240859 ==> Pasta=BUY 232960 conf:(0.97)

77. Meat=BUY Sugar=BUY 253582 ==> Flour=BUY 245258 conf:(0.97)

78. Milk=BUY Flour=BUY 242304 ==> Meat=BUY 234335 conf:(0.97)

79. Oil=BUY Bread=BUY Meat=BUY 234049 ==> Pasta=BUY 226348 conf:(0.97)

80. Tea=BUY Ygourt=BUY Meat=BUY 229914 ==> Flour=BUY 222345 conf:(0.97)

81. Eggs=BUY 236570 ==> Tomato Paste=BUY 228779 conf:(0.97)

82. lentils=BUY Flour=BUY 234433 ==> Meat=BUY 226707 conf:(0.97)

83. Cake=BUY Ygourt=BUY Flour=BUY 229495 ==> Meat=BUY 221916 conf:(0.97)

84. Flour=BUY Sugar=BUY 253640 ==> Meat=BUY 245258 conf:(0.97)

85. Bread=BUY Flour=BUY 244524 ==> Meat=BUY 236430 conf:(0.97)

86. Oil=BUY Ygourt=BUY Meat=BUY 229809 ==> Pasta=BUY 222184 conf:(0.97)

87. Tang=BUY Ygourt=BUY Flour=BUY 229322 ==> Meat=BUY 221700 conf:(0.97)

88. Halva=BUY Ygourt=BUY Meat=BUY 230392 ==> Flour=BUY 222732 conf:(0.97)

89. Bread=BUY Meat=BUY 244598 ==> Flour=BUY 236430 conf:(0.97)

90. Pasta=BUY Meat=BUY 244000 ==> Flour=BUY 235832 conf:(0.97)

91. Soap=BUY Flour=BUY 243550 ==> Meat=BUY 235383 conf:(0.97)

92. cheese=BUY Flour=BUY 232013 ==> Meat=BUY 224221 conf:(0.97)

93. Oil=BUY Flour=BUY 240785 ==> Meat=BUY 232691 conf:(0.97)

94. Cake=BUY Ygourt=BUY Meat=BUY 229647 ==> Flour=BUY 221916 conf:(0.97)

95. Andomi=BUY Eggs=BUY 229426 ==> chips=BUY 221689 conf:(0.97)

96. cheese=BUY Meat=BUY 232052 ==> Flour=BUY 224221 conf:(0.97)

97. Soap=BUY Meat=BUY 243616 ==> Flour=BUY 235383 conf:(0.97)

98. Tea=BUY Flour=BUY 240694 ==> Meat=BUY 232544 conf:(0.97)

99. Pasta=BUY Bread=BUY Sugar=BUY 235444 ==> Oil=BUY 227469 conf:(0.97)

100. Oil=BUY Meat=BUY 240859 ==> Flour=BUY 232691 conf:(0.97)

101. Oil=BUY Meat=BUY Flour=BUY 232691 ==> Pasta=BUY 224792 conf:(0.97)

102. jam=BUY Flour=BUY 241066 ==> Meat=BUY 232847 conf:(0.97)

103. Oil=BUY Meat=BUY Sugar=BUY 231538 ==> Pasta=BUY 223639 conf:(0.97)

104. Milk=BUY Flour=BUY Sugar=BUY 232355 ==> Meat=BUY 224386 conf:(0.97)

105. Soap=BUY Oil=BUY 232687 ==> Pasta=BUY 224697 conf:(0.97)

106. Halva=BUY Flour=BUY 240030 ==> Meat=BUY 231767 conf:(0.97)

107. Bread=BUY Meat=BUY Sugar=BUY 234748 ==> Flour=BUY 226660
conf:(0.97)

108. Tea=BUY Meat=BUY 240843 ==> Flour=BUY 232544 conf:(0.97)

109. Bread=BUY Flour=BUY Sugar=BUY 234754 ==> Meat=BUY 226660
conf:(0.97)

110. Pasta=BUY Meat=BUY Sugar=BUY 234150 ==> Flour=BUY 226062
conf:(0.97)

111. Shayriea=BUY 237549 ==> chips=BUY 229336 conf:(0.97)

112. Oil=BUY Bread=BUY Flour=BUY 233975 ==> Meat=BUY 225881 conf:(0.97)

113. Tomato Paste=BUY 237167 ==> Shayriea=BUY 228962 conf:(0.97)

114. Cake=BUY Flour=BUY 240740 ==> Meat=BUY 232404 conf:(0.97)

115. Tomato Paste=BUY 237167 ==> chips=BUY 228954 conf:(0.97)

116. Pasta=BUY Bread=BUY 245294 ==> Oil=BUY 236790 conf:(0.97)

117. rice=BUY Meat=BUY 238818 ==> Flour=BUY 230516 conf:(0.97)

118. Pasta=BUY Bread=BUY Meat=BUY 234852 ==> Flour=BUY 226684
conf:(0.97)

119. chips=BUY 237621 ==> Shayriea=BUY 229336 conf:(0.97)

120. Soap=BUY Meat=BUY Sugar=BUY 233637 ==> Flour=BUY 225484 conf:(0.97)

121. Tang=BUY Flour=BUY 240178 ==> Meat=BUY 231796 conf:(0.97)

122. Oil=BUY Bread=BUY Meat=BUY 234049 ==> Flour=BUY 225881 conf:(0.97)

123. Oil=BUY Meat=BUY Sugar=BUY 231538 ==> Flour=BUY 223450 conf:(0.97)

124. Biscuit=BUY Flour=BUY 230514 ==> Meat=BUY 222457 conf:(0.97)

125. Soap=BUY Flour=BUY Sugar=BUY 233651 ==> Meat=BUY 225484 conf:(0.97)

126. Oil=BUY Flour=BUY Sugar=BUY 231544 ==> Meat=BUY 223450 conf:(0.97)

127. lentils=BUY Flour=BUY 234433 ==> Ygourt=BUY 226218 conf:(0.96)

128. Gas Water=BUY Flour=BUY 231212 ==> Meat=BUY 223109 conf:(0.96)

129. Oil=BUY Pasta=BUY Meat=BUY 232960 ==> Flour=BUY 224792 conf:(0.96)

130. Tea=BUY Meat=BUY Sugar=BUY 230985 ==> Flour=BUY 222878 conf:(0.96)

131. Tea=BUY Flour=BUY Sugar=BUY 231028 ==> Meat=BUY 222878 conf:(0.96)

132. Halva=BUY Meat=BUY 240243 ==> Flour=BUY 231767 conf:(0.96)

133. lentils=BUY Meat=BUY 234257 ==> Ygourt=BUY 225973 conf:(0.96)

134. Tomato Paste=BUY 237167 ==> Eggs=BUY 228779 conf:(0.96)

135. Cake=BUY Meat=BUY 240963 ==> Flour=BUY 232404 conf:(0.96)

136. jam=BUY Flour=BUY Sugar=BUY 231347 ==> Meat=BUY 223128 conf:(0.96)

137. Oil=BUY Bread=BUY 245536 ==> Pasta=BUY 236790 conf:(0.96)

138. Halva=BUY Meat=BUY Sugar=BUY 230018 ==> Flour=BUY 221818
conf:(0.96)

139. Tomato Paste=BUY 237167 ==> Meat=BUY 228708 conf:(0.96)

140. jam=BUY Pasta=BUY 232035 ==> Bread=BUY 223757 conf:(0.96)

141. Oil=BUY Flour=BUY 240785 ==> Pasta=BUY 232164 conf:(0.96)

142. Biscuit=BUY Meat=BUY 230718 ==> Flour=BUY 222457 conf:(0.96)

143. Eggs=BUY 236570 ==> Sugar=BUY 228096 conf:(0.96)

144. Gas Water=BUY Meat=BUY 231415 ==> Flour=BUY 223109 conf:(0.96)

145. Cake=BUY Meat=BUY Sugar=BUY 230647 ==> Flour=BUY 222367 conf:(0.96)

146. Halva=BUY Flour=BUY Sugar=BUY 230081 ==> Meat=BUY 221818
conf:(0.96)

147. Pasta=BUY 254442 ==> Bread=BUY 245294 conf:(0.96)

148. Oil=BUY Bread=BUY Flour=BUY 233975 ==> Pasta=BUY 225552 conf:(0.96)

149. Oil=BUY Ygourt=BUY 240216 ==> Pasta=BUY 231546 conf:(0.96)

150. Cake=BUY Flour=BUY Sugar=BUY 230703 ==> Meat=BUY 222367 conf:(0.96)

151. Shayriea=BUY 237549 ==> Tomato Paste=BUY 228962 conf:(0.96)

152. Shayriea=BUY 237549 ==> Eggs=BUY 228957 conf:(0.96)

153. Pasta=BUY Bread=BUY Meat=BUY 234852 ==> Oil=BUY 226348 conf:(0.96)

154. Oil=BUY Bread=BUY Ygourt=BUY 233681 ==> Pasta=BUY 225209 conf:(0.96)

155. Tang=BUY Pasta=BUY 230745 ==> Bread=BUY 222367 conf:(0.96)

156. Oil=BUY Ygourt=BUY Flour=BUY 229821 ==> Pasta=BUY 221474 conf:(0.96)

157. Pasta=BUY Bread=BUY Flour=BUY 234056 ==> Oil=BUY 225552 conf:(0.96)

158. Tang=BUY Flour=BUY Sugar=BUY 230668 ==> Meat=BUY 222286 conf:(0.96)

159. Tomato Paste=BUY 237167 ==> Flour=BUY 228545 conf:(0.96)

160. lentils=BUY Meat=BUY 234257 ==> rice=BUY 225739 conf:(0.96)

161. Pasta=BUY Bread=BUY Ygourt=BUY 233713 ==> Oil=BUY 225209 conf:(0.96)

162. chips=BUY 237621 ==> Tomato Paste=BUY 228954 conf:(0.96)

163. Pasta=BUY Andomi=BUY 230904 ==> Bread=BUY 222470 conf:(0.96)

164. Pasta=BUY Ygourt=BUY 242586 ==> Bread=BUY 233713 conf:(0.96)

165. Oil=BUY 252662 ==> Pasta=BUY 243402 conf:(0.96)

166. Andomi=BUY 250953 ==> Sugar=BUY 241739 conf:(0.96)

167. Oil=BUY Sugar=BUY 243025 ==> Pasta=BUY 234081 conf:(0.96)

168. Eggs=BUY 236570 ==> Ygourt=BUY 227832 conf:(0.96)

169. chips=BUY 237621 ==> Eggs=BUY 228833 conf:(0.96)

170. Oil=BUY Bread=BUY Sugar=BUY 236215 ==> Pasta=BUY 227469 conf:(0.96)

171. Flour=BUY Andomi=BUY 240121 ==> Ygourt=BUY 231207 conf:(0.96)

172. Milk=BUY Pasta=BUY 232318 ==> Bread=BUY 223671 conf:(0.96)

173. Oil=BUY Flour=BUY Sugar=BUY 231544 ==> Pasta=BUY 222923 conf:(0.96)

174. jam=BUY Oil=BUY 230000 ==> Pasta=BUY 221406 conf:(0.96)

175. Pasta=BUY Sugar=BUY 244592 ==> Bread=BUY 235444 conf:(0.96)

176. Pasta=BUY Meat=BUY 244000 ==> Bread=BUY 234852 conf:(0.96)

177. Bread=BUY Andomi=BUY 232329 ==> Sugar=BUY 223617 conf:(0.96)

178. Oil=BUY Ygourt=BUY Sugar=BUY 230975 ==> Pasta=BUY 222305 conf:(0.96)

179. Soap=BUY Pasta=BUY 235546 ==> Bread=BUY 226698 conf:(0.96)

180. Flour=BUY Andomi=BUY 240121 ==> Sugar=BUY 231100 conf:(0.96)

181. Pasta=BUY Flour=BUY 243204 ==> Bread=BUY 234056 conf:(0.96)

182. Pasta=BUY Andomi=BUY 230904 ==> Sugar=BUY 222192 conf:(0.96)

183. Oil=BUY Bread=BUY 245536 ==> Sugar=BUY 236215 conf:(0.96)

184. Pasta=BUY Ygourt=BUY Meat=BUY 233224 ==> Bread=BUY 224351 conf:(0.96)

185. Pasta=BUY Ygourt=BUY Sugar=BUY 232816 ==> Bread=BUY 223943 conf:(0.96)

186. Oil=BUY 252662 ==> Sugar=BUY 243025 conf:(0.96)

187. Pasta=BUY Ygourt=BUY Flour=BUY 232514 ==> Bread=BUY 223641 conf:(0.96)

188. Pasta=BUY Andomi=BUY 230904 ==> Meat=BUY 222092 conf:(0.96)

189. Soap=BUY Andomi=BUY 231485 ==> Sugar=BUY 222648 conf:(0.96)

190. lentils=BUY 245448 ==> rice=BUY 236077 conf:(0.96)

191. Tea=BUY Pasta=BUY 234731 ==> Bread=BUY 225768 conf:(0.96)

192. Meat=BUY Flour=BUY Andomi=BUY 233275 ==> Ygourt=BUY 224361 conf:(0.96)

193. Ygourt=BUY Andomi=BUY 240978 ==> Sugar=BUY 231764 conf:(0.96)

194. Meat=BUY Andomi=BUY 240811 ==> Sugar=BUY 231597 conf:(0.96)

195. Flour=BUY 263738 ==> Sugar=BUY 253640 conf:(0.96)

196. Oil=BUY Pasta=BUY 243402 ==> Sugar=BUY 234081 conf:(0.96)

197. Tomato Paste=BUY 237167 ==> Andomi=BUY 228083 conf:(0.96)

198. Oil=BUY Flour=BUY 240785 ==> Sugar=BUY 231544 conf:(0.96)

199. jam=BUY Ygourt=BUY 240017 ==> Flour=BUY 230803 conf:(0.96)

200. Bread=BUY 256085 ==> Sugar=BUY 246235 conf:(0.96)