Autonomous Data Warehouse
Oracle Machine Learning
Oracle Analytics Cloud

Market Basket Analysis (MBA) Revisited using SQL Pattern Matching

Shankar Somayajula
shankar.somayajula@oracle.com

Aug 22th, 2019
Agenda

- Typical MBA
- MBA Revisited
- SQL Pattern Matching
- Demo / Screenshots
- Summarize
- Q&A
Finding Patterns in Data

Typical use cases in today's world of fast exploration of data

- Financial Services
- Money Laundering
- Fraud Tracking Stock Market
- Utility Analysis
- Unusual Usage
- Law & Order Monitoring Suspicious Activities
- Retail Buying Patterns
- Session-ization Returns Fraud
- Telcos SIM Card Fraud
- Call Quality Money Laundering

Lots of Data
Typical MBA

- Transaction Data of type Master-Detail

<table>
<thead>
<tr>
<th>TRX_ID</th>
<th>LINEITEM_ID</th>
<th>PROD_ID</th>
<th>PROD_NM</th>
<th>LISEQ</th>
<th>LISEQ_DT</th>
<th>UNIT_PRICE</th>
<th>UNITS</th>
<th>QTY</th>
<th>EXTND_AMT</th>
<th>DISCOUNT_IND</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>112q</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>60</td>
<td>1</td>
<td>1</td>
<td>60N</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>105p</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>50</td>
<td>2</td>
<td>4</td>
<td>100N</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8a</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>10N</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>12b</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>100Y</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>15c</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>27</td>
<td>10</td>
<td>5</td>
<td>270Y</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>115r</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>70</td>
<td>4</td>
<td>1</td>
<td>280N</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>112q</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>60</td>
<td>6</td>
<td>2</td>
<td>360N</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>115r</td>
<td></td>
<td></td>
<td>1-05-OCT-18</td>
<td>63</td>
<td>4</td>
<td>1</td>
<td>252Y</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>21d</td>
<td></td>
<td></td>
<td>1-09-OCT-18</td>
<td>40</td>
<td>4</td>
<td>1</td>
<td>160N</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>112q</td>
<td></td>
<td></td>
<td>2-09-OCT-18</td>
<td>60</td>
<td>6</td>
<td>1</td>
<td>360N</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>12b</td>
<td></td>
<td></td>
<td>3-09-OCT-18</td>
<td>20</td>
<td>8</td>
<td>2</td>
<td>160N</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>115r</td>
<td></td>
<td></td>
<td>4-09-OCT-18</td>
<td>70</td>
<td>4</td>
<td>1</td>
<td>280N</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>12b</td>
<td></td>
<td></td>
<td>115-0CT-18</td>
<td>20</td>
<td>6</td>
<td>2</td>
<td>120N</td>
<td></td>
</tr>
</tbody>
</table>

- Rules of type "IF ... THEN ...

- with KPIs – Support, Confidence and Lift

<table>
<thead>
<tr>
<th>RULE_ID</th>
<th>ANTCOS</th>
<th>ANTCOS_ID</th>
<th>ANTCOS_IDfindByANTCOS</th>
<th>SUPPORT</th>
<th>CONFIDENCE</th>
<th>LIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100b</td>
<td>1000102</td>
<td>1000105</td>
<td>0.076923076923076927</td>
<td>1</td>
<td>4.3333333333333333</td>
</tr>
<tr>
<td>101</td>
<td>101b</td>
<td>1000102</td>
<td>1000105</td>
<td>0.076923076923076927</td>
<td>0.33333333333333333</td>
<td>4.3333333333333333</td>
</tr>
<tr>
<td>102</td>
<td>102r</td>
<td>1000102</td>
<td>1000105</td>
<td>0.33333333333333333</td>
<td>0.33333333333333333</td>
<td>4.3333333333333333</td>
</tr>
<tr>
<td>103</td>
<td>103b</td>
<td>1000102</td>
<td>1000105</td>
<td>0.33333333333333333</td>
<td>0.33333333333333333</td>
<td>4.3333333333333333</td>
</tr>
<tr>
<td>104</td>
<td>104b</td>
<td>1000102</td>
<td>1000105</td>
<td>0.33333333333333333</td>
<td>0.33333333333333333</td>
<td>4.3333333333333333</td>
</tr>
<tr>
<td>105</td>
<td>105g</td>
<td>1000102</td>
<td>1000105</td>
<td>0.33333333333333333</td>
<td>0.33333333333333333</td>
<td>4.3333333333333333</td>
</tr>
</tbody>
</table>
MBA Revisited

- Transaction Data augmented with "Tags"

- Rules of type "IF ... THEN ... "

- with more rules involving the added tags ...
MBA Revisited

- with all the standard KPIs …

---

<table>
<thead>
<tr>
<th>RULE_ID</th>
<th>ANTCONS</th>
<th>ANTCONS_ID</th>
<th>RULE_LENGTH</th>
<th>RULE_SUPPORT</th>
<th>RULE_CONFIDENCE</th>
<th>RULE_LIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Normal Spend: 10% to 90% (220G &lt; trx_amt &lt; 948G) =&gt; a</td>
<td>250002 =&gt; 1000008</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
<tr>
<td>73</td>
<td>Normal Spend: 10% to 90% (220G &lt; trx_amt &lt; 948G) =&gt; r</td>
<td>250002 =&gt; 1000115</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
<tr>
<td>41</td>
<td>Low Spend: Bottom 10% (trx_amt &lt;= 220G) =&gt; c</td>
<td>250003 =&gt; 1000015</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
<tr>
<td>27</td>
<td>Low Spend: Bottom 10% (trx_amt &lt;= 220G) =&gt; b</td>
<td>250003 =&gt; 1000012</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
<tr>
<td>75</td>
<td>High Spend: Top 10% (trx_amt &gt;= 948G) =&gt; z</td>
<td>250001 =&gt; 1000125</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
<tr>
<td>47</td>
<td>High Spend: Top 10% (trx_amt &gt;= 948G) =&gt; d</td>
<td>250001 =&gt; 1000021</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>RULE_ID</th>
<th>ANTCONS</th>
<th>ANTCONS_ID</th>
<th>RULE_LENGTH</th>
<th>RULE_SUPPORT</th>
<th>RULE_CONFIDENCE</th>
<th>RULE_LIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>SiestaTimeTrends =&gt; d</td>
<td>550004 =&gt; 1000021</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
<tr>
<td>51</td>
<td>LateNightTrx =&gt; d</td>
<td>550003 =&gt; 1000021</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
<tr>
<td>69</td>
<td>EveningHourBuys =&gt; q</td>
<td>550002 =&gt; 1000112</td>
<td>2.0076923076923076927</td>
<td>0.1425714257142571</td>
<td>1.057142571425714</td>
<td>1.857142571425714</td>
</tr>
</tbody>
</table>
MBA Revisited

- and a lot more ...

### Rule Table 1

<table>
<thead>
<tr>
<th>RULE_ID</th>
<th>ANTCONS_ID</th>
<th>ANTCONS_ID</th>
<th>RULE_LENGTH</th>
<th>ASOLD</th>
<th>BSOLD</th>
<th>ABSOLD</th>
<th>AOBSSOLD</th>
<th>ASEQ</th>
<th>BSEQQ</th>
<th>BSEQB</th>
<th>BASEQ</th>
<th>ASA</th>
<th>BSA</th>
<th>ABSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 b, z</td>
<td>p</td>
<td>1000012, 1000125</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>101 b, p</td>
<td>r</td>
<td>1000012, 1000015</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>102 q, z</td>
<td>b</td>
<td>1000112, 1000115</td>
<td>3</td>
<td>4</td>
<td>11</td>
<td>4</td>
<td>13</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>103 b, z</td>
<td>q</td>
<td>1000012, 100015</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>104 b, q</td>
<td>r</td>
<td>1000012, 1000115</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>105 q, z</td>
<td>b</td>
<td>1000112, 1000125</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>106 b, z</td>
<td>q</td>
<td>1000012, 1000125</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>107 b, q</td>
<td>r</td>
<td>1000012, 1000115</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>108 r, z</td>
<td>b</td>
<td>1000115, 10000125</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>109 b, z</td>
<td>r</td>
<td>1000012, 1000125</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>110 b, r</td>
<td>s</td>
<td>1000012, 1000115</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

### Rule Table 2

<table>
<thead>
<tr>
<th>RULE_ID</th>
<th>ANTCONS</th>
<th>ANTCONS_ID</th>
<th>RULE_LENGTH</th>
<th>ASOLD</th>
<th>BSOLD</th>
<th>ABSOLD</th>
<th>AOBSSOLD</th>
<th>ASEQ</th>
<th>BSEQQ</th>
<th>BSEQB</th>
<th>BASEQ</th>
<th>ASA</th>
<th>BSA</th>
<th>ABSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>High</td>
<td>Top 10% (TRX_AMT $/gt; 946)</td>
<td>250001</td>
<td>1000125</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>High</td>
<td>Top 10% (TRX_AMT $/gt; 946)</td>
<td>250001</td>
<td>1000021</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Normal</td>
<td>10% to 90% (2204 &lt; TRX_AMT $/lt; 946)</td>
<td>280002</td>
<td>1000008</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>Low</td>
<td>Bottom 10% (TRX_AMT $/leq; 2204)</td>
<td>250003</td>
<td>1000015</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>75</td>
<td>Normal</td>
<td>10% to 90% (2204 &lt; TRX_AMT $/leq; 946)</td>
<td>280002</td>
<td>1000115</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>Low</td>
<td>Bottom 10% (TRX_AMT $/leq; 2204)</td>
<td>250003</td>
<td>1000012</td>
<td>2</td>
<td>4</td>
<td>11</td>
<td>4</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>
MBA Revisited

- Extend MB Rules to include other types of MB Rule Components
Examples of MB Rules/Insights

• (diapers) => (beer)
• (peanutButter, jelly) => (bread)

• Many ways to improve traditional MB
  – Multiple levels of dimension ... SKU to Sub-Category to Category (ideally at same time)
  – Add additional dimensions – Trx/ Dimensional Attributes as tags

Multidimensional Rules with artificial/virtual products gives richer picture ...

• (Item=X, isOver18=TRUE, isNewCustomer=TRUE) => (Item=Y)
• (buyerAge >= 63, loyaltyAge>= 2) => (toothBrushBuy >=2)
• age(X,"20...29"), income(X,"52k...58k") => buys(X, "iPad")
Why Multiple Models?

- DB/Star Schema/Analysis Container (Host), MB Model (Context), MB Rules and MB KPIs
  - Lab like environment for multiple models being in play

Trx Dataset #1 (SS1, SS #1)  Trx Dataset #2 (SH2, SS #2)

Credits: 1. Photo by Markus Spiske on Unsplash, 2. Photo by Andrew Ridley on Unsplash
Why Multiple Models?

- DB/Star Schema/Analysis Container (Host), MB Model (Context), MB Rules and MB KPIs
  - Lab like environment for multiple models being in play

Model 1

Model 2

Model 3

Model 4

Credits: 1. Photo by Markus Spiske on Unsplash
Why Multiple Models?

- DB/Star Schema/Analysis Container (Host), MB Model (Context), MB Rules and MB KPIs
  - Lab like environment for multiple models being in play

Credits: 1. Photo by Markus Spiske on Unsplash
Why Multiple Models?

- DB/Star Schema/Analysis Container (Host), MB Model (Context), MB Rules and MB KPIs
  - Lab like environment for multiple models being in play

Model 1

Model 2

Model 3

MB Model for specific Time Period(s)

Model 4

Credits: 1. Photo by Markus Spiske on Unsplash
Why Multiple Models?

• DB/Star Schema/Analysis Container (Host), MB Model (Context), MB Rules and MB KPIs
  – Lab like environment for multiple models being in play

Model 1
Model 2
Model 3
Model 4

MB Model Partitioned by Country (say)

Credits: 1. Photo by Markus Spiske on Unsplash
MB Rules >> Patterns >> Insights ... #4

• A lot of MB Rules and Not all patterns are useful.

• Taking the MB Rule and analyzing it in different contexts is typically an offline exercise
  – Typically this would involve a lot of offline actions/modeling exercises to look at the Transactional dataset from different perspectives
  – From frinkiac :D

  – Well, There is a way ... and that’s where SQL Pattern Matching comes in.
A lot of MB Rules and Not all patterns are useful.
Taking the MB Rule and analyzing it in different contexts is typically an offline exercise.

Credits: 1. Photo by Zhifei Zhou on Unsplash, 2. Photo by Niklas Hamann on Unsplash
MB Rules >> Patterns >> Insights ... #4 (cont.)

- A lot of MB Rules and Not all patterns are useful.
- Taking the MB Rule and analyzing it in different contexts is typically an offline exercise.

Credits:
1. Photo by Zhifei Zhou on Unsplash
2. Photo by Niklas Hamann on Unsplash
• A lot of MB Rules and Not all patterns are useful.
• Taking the MB Rule and analyzing it in different contexts is typically an offline exercise

Credits: 1. Photo by Zhifei Zhou on Unsplash, 2. Photo by Niklas Hamann on Unsplash
MB Rules >> Patterns >> Insights ... #4 (cont.)

- A lot of MB Rules and Not all patterns are useful.
- Taking the MB Rule and analyzing it in different contexts is typically an offline exercise.

Credits: 1. Photo by Zhifei Zhou on Unsplash, 2. Photo by Niklas Hamann on Unsplash
MB Rules >> Patterns >> Insights ... #5

- Allow for What-if actions on MB Rules/Patterns
  - From frinkiac :D

- SQL Tools allow what-if ... Facilitate end users to perform what if actions via BI Tools.
Demo

• Autonomous Database Warehouse (ADW) … Oracle 18c database
  – ADW is optional but Oracle Database 12c+ is mandatory. ANSI SQL Feature Row Pattern Matching via keyword MATCH_RECOGNIZE should be available in the Db.
  – SQL scripts used for pre-processing/data preparation of the input data
  – SQL scripts also used for post-processing

• Oracle Machine Learning (OML) bundled/packaged with ADW
  – OML is optional but Oracle Data Mining (part of Oracle Advanced Analytics) is mandatory
  – ODM based pl/sql api is used to make a call to the Apriori Algorithm for performing Market Basket Analysis
  – SQL queries used to extract the patterns from the Association Rules (AR) model.

• Oracle Analytics Cloud (OAC)
  – Many advanced features of the solution leverage the rpd (data modeling layer) component of OAC
  – KPI Calculations and Deepdives on-demand need the modeling layer (rpd or equivalent)
Summary

• MBA
  – Pattern Discovery via OAA/ODM
  – Model used to extract Rules and core KPIs
  – No way to score Rules (need to rebuild)
  – Patterns of special interest (anomalous/obscure) cannot be found unless model settings are relaxed. We may get those and many many more.

• MBA Revisited
  – Pattern Discovery via OAA/ODM (same)
  – Rules/KPIs extracted into a Data Model allowing for BI/Adhoc analysis
  – Post – processing to setup the analysis context
  – SQL approach allows
    • New KPIs – KPIs of statistical nature as well as KPIs related to Business needs (as elaborate as needed)
    • Scoring against new data possible – patterns can degrade in performance
    • Score/Track Patterns against specific Trx subsets of interest
    • Adhoc BI/Exploratory Data Analysis of Patterns
    • Special Patterns of interest (Fraud use cases) with very low support can also be found as well as analyzed (what-if)
    • 2 independent ways to MB KPIs – ETL + DB/BI (faster) or DB View + DB/BI (slower, on demand)
• Useful?
• Very little shown of ADW/OML currently (end goal), using SQL Developer for most Db actions
• Need more details on Market Basket Analysis (MBA)? SQL Pattern Matching? ½ Hr talk precludes possibility of giving lot of introduction to the material.
Typical MBA

- **MB Rule: A => B [ s, c ]**
  - **Support:** denotes the frequency of the rule within transactions. A high value means that the rule involves a great part of database.
    
    \[
    \text{support}(A \Rightarrow B [ s, c ]) = p(A \cup B)
    \]
  
  - **Confidence:** denotes the percentage of transactions containing A which also contain B. It is an estimation of conditioned probability.
    
    \[
    \text{confidence}(A \Rightarrow B [ s, c ]) = p(B \mid A) = \frac{\sup(A,B)}{\sup(A)}
    \]
  
  - **Lift:** a measure of how much better a rule is at predicting the result than just assuming the result in the first place.
    
    \[
    \text{lift}(A \Rightarrow B [ s, c ]) = \frac{p(B \mid A)}{p(B)} = \frac{\sup(A,B)}{\sup(A) \cdot \sup(B)}
    \]
MBA Revisited

• **MB Rule: A => B [ s, c ]**
  
  — **Conviction:** measure of the number of times the rule would be incorrect if the association between A and B was purely random chance. Conviction is a measure of the implication and has value 1 if items are unrelated. Sensitive to the directionality of the rule (unlike lift)

    \[
    \text{conviction}(A \Rightarrow B [ s, c ]) = \frac{\sup(A).\sup(B')}{\sup(A,B')}
    \]

  — **Leverage or Piatetsky-Shapiro:** is the proportion of additional elements covered by both A and B above the expected if independent.

    \[
    \text{leverage}(A \Rightarrow B [ s, c ]) = \sup(A,B) - \sup(A) \cdot \sup(B)
    \]

  — **Max Confidence:** KPI max_conf is the maximum confidence of the two association rules related to A and B, namely, “p(A|B)” and “p(B|A)”.

    \[
    \text{max_conf}(A \Rightarrow B [ s, c ]) = \max(p(B|A), p(A|B))
    \]
MBA Revisited (Evolution)

• OAC/OBIEE Business Model

- Typical MB involves extraction of MB Rules/Pat terns from Trx Data.
- MB Rules are qualified with default MB KPIs
MBA Revisited

- OAC/OBIEE Business Model

- Typical MB involves extraction of MB Rules/Patterns from Trx Data.
- MB Rules are qualified with default MB KPIs
- BI schema for adhoc reporting/analysis can involve source Trx data analysis as well as pattern/MB Rule analysis (disjoint)
MBA Revisited

• OAC/OBIEE Business Model

• Typical MB involves extraction of MB Rules/Patterns from Trx Data.
• MB Rules are qualified with default MB KPIs
• BI schema for adhoc reporting/analysis can involve source Trx data analysis as well as pattern/MB Rule analysis (disjoint)
MBA Revisited

- OAC/OBIEE Business Model

- Typical MB involves extraction of MB Rules/Patterns from Trx Data.
- MB Rules are qualified with default MB KPIs
- BI schema for adhoc reporting/analysis can involve source Trx data analysis as well as pattern/MB Rule analysis
- Add Model Dimension for analysis context.
MBA Revisited

- OAC/OBIEE Business Model

- Typical MB involves extraction of MB Rules/Patterns from Trx Data.
- MB Rules are qualified with default MB KPIs
- Advanced BI schema to support adhoc reporting/analysis of MB Rules/Patterns across whole dataset or split by attribute fields as well against source Trx subset of interest.
- Model for analysis context.
MBA Revisited

- OAC/OBIEE Business Model
MBA Revisited

• OAC/OBIEE Business Model

Mining AR Model KPIs (Output)
MBA Revisited

- OAC/OBIEE Business Model
MBA Revisited

- OAC/OBIEE Business Model

MBKPIs (Model - Rule) – Dataset, All Trx
MBA Revisited

- OAC/OBIEE Business Model

MBKPIs (Model – Rule – Trx) – Data Subset, Partition, Deepdives
KPIs ... MB Rule and MB Rule-MB Trx levels

- (Rule KPIs) aSold – Count of Trx ... antecedents part of the MB Rule has been sold
- (Rule KPIs) bSold – Count of Trx .... consequent part of the MB Rule ...
- (Rule KPIs) abSold – Count of Trx ... both antecedents and consequent parts of the MB Rule ...
- (Rule KPIs) aobSold - Count of Trx ... either antecedents or consequent parts of the MB Rule ...
- (Sequential KPIs) aSeq - ... antecedents (as per MB Rule defn) have been sold in the same order in the Trx
- (Sequential KPIs) abSeq - ... antecedents (in any order, all of them) sold before consequent
- (Sequential KPIs) aSeqb - ... antecedents (in order, all of them ) sold before consequent
- (Sequential KPIs) bSeqa - ... consequent sold before antecedents (any order but all of them)
- (Sequential KPIs) baSeq - ... consequent sold before antecedents (in order as per MB Rule defn)
- (Share of wallet/Trx KPIs) aSA – a Sold Alone i.e. Trx is wholly composed of antecedents
- (Share of wallet/Trx KPIs) bSA – b Sold Alone i.e. Trx is wholly composed of consequent
- (Share of wallet/Trx KPIs) abSA – ab Sold Alone i.e. Trx is wholly composed of antecedents and consequent
KPIs ... MB Rule and MB Rule-MB Trx levels

- MB Rule KPIs stored in terms of
  - Counts (a, b, ab, aob, bnota, anotb, aSA, bSA etc.)
  - Units
  - Qty
  - Amt ($)

- MB Rule KPIs for any MB Rule can be derived/scored/calculated using Trx set at level:
  - Entire dataset
  - Partitioned subset based on Trx Header Attr/Trx Line Attr/Dim Attr/KPI values

Filtering of dataset (optional)

- Identification of existing MB Rules for Analysis occurs within the MB Model definition in terms of Min Supp, Min Conf, Max Rule Length etc. ... => Adhoc analysis of obscure MB Rules/Patterns not possible.

- What-If analysis .... Generate What-If MB Rule(s) either afresh or via Transformation(s) to an existing MB Rule to overcome above restrictions.
SQL Pattern Matching: Match Process Preparation

- Match Process Preparation (ETL and/or view)
SQL Pattern Matching: Match Process Preparation

- Match Process Preparation (ETL and/or view)
SQL Pattern Matching: Input, Processing, Output

1. Define input
2. Partition/order input
3. Process pattern
4. Using defined conditions
5. Output: rows per match
6. Output: columns per row
7. Go where after match?
8. Return Data at match level (select * ...) or higher levels (based on group bys)

```sql
select
  rule, -- get KPIs at rule level
  -- trx, -- uncomment to get KPIs at rule-trx level
  KPIs
from INPUT_TABLE_OR_QUERY
match_recognize(
  partition by "model", rule, trx order by
  rule_mbcomp_seq
  measures text_kpi_meas, kpi_meas,
  agg_kpi_meas, kpi_partial_meas etc
  one row per match
  after match skip past last row
  pattern (PERMUTE(apli*, bpli*, opli*))
  define apli as (mb_comp = 'a_part'), bpli as
  (mb_comp = 'b_part'), opli as (mb_comp = 'o_part')
) group by
  rule
-- , trx -- uncomment to get KPIs at rule-trx level
;
```

Slide Credit: Customized from Stew Ashton’s Advanced Pattern Matching pptx which was part of Oracle Code Paris 2018
Some Issues/Challenges

• Pattern Matching SQL needs a fixed pattern to use for matching.
  – We can write SQL for a single Rule ... to match against a dataset (many Trx)
    • For e.g. for rule “p, q, r => a” we use ....
      PATTERN ( permute(p,q,r) | a )
      DEFINE
      p as trxli_prod_nm = 'p',
      q as trxli_prod_nm = 'q',
      r as trxli_prod_nm = 'r',
      a as trxli_prod_nm = 'a'
      Data driven pattern, Dyn SQL
  – When we need to match many patterns (say, act on a whole AR model with 100+ rules of varying sizes) -- each against a trx dataset we should define the patterns via metadata/component structures.
    PATTERN (PERMUTE(apli*, bpli*, opli*))
    DEFINE
    apli as (mb_comp = 'a_part'),
    bpli as (mb_comp = 'b_part'),
    opli as (mb_comp = 'o_part')
    Metadata based pattern, SQL

• Same sql for any pattern => Allows integration into ETL or use in sql view to match dynamically via sql query (issued by BI Tools).
Some Issues/Challenges

• How should one handle the issue of merging multiple dimensions into a single analysis dimension? This problem comes up in Graphs too (All dimension members need to be combined into a set of nodes ... "node Id" identifier conflicts).

• Currently using offsets for each dimension to base their Ids. Product has the highest offset (unbounded).

<table>
<thead>
<tr>
<th>MB_PROD_TYPE_ID</th>
<th>MB_PROD_TYPE_DESC</th>
<th>MB_PROD_TYPE_OFFSET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 Product</td>
<td>1000000</td>
</tr>
<tr>
<td>2</td>
<td>2 CustSegAttr</td>
<td>200000</td>
</tr>
<tr>
<td>3</td>
<td>3 CustSpendAttr</td>
<td>250000</td>
</tr>
<tr>
<td>4</td>
<td>4 CustLoyaltyAttr</td>
<td>300000</td>
</tr>
<tr>
<td>5</td>
<td>5 FreqShopperAttr</td>
<td>350000</td>
</tr>
<tr>
<td>6</td>
<td>6 PromoAttr</td>
<td>400000</td>
</tr>
<tr>
<td>7</td>
<td>7 StoreAttr</td>
<td>450000</td>
</tr>
<tr>
<td>8</td>
<td>8 ProductAttr</td>
<td>500000</td>
</tr>
<tr>
<td>9</td>
<td>9 TrxAtr</td>
<td>550000</td>
</tr>
<tr>
<td>10</td>
<td>999999 NotApplicable</td>
<td>0</td>
</tr>
</tbody>
</table>
Sample Screenshots
Sample Screenshots

- MBA MB Rules – Types of Rule Linkages