**Reporte de Análisis de Market Basket**

**Vendedor**: Samtec

**Período de tiempo**: 4 meses (mayo, junio, julio, agosto)

**Cantidad de registros**: 333038

**Cantidad de partes distintas**: 78421

Detalle:

El conjunto de datos utilizado se obtiene mediante varias consultas a la base de datos. El origen y objetivo de todos los datos es que sean utilizados posteriormente en un análisis de patrones de compra más detallado, sin embargo, para el caso de estudio actual se utiliza la información para un análisis de asociación de productos (market basket analysis). El conjunto de datos obtiene información de diferentes tipos de clientes, así como el detalle de las ventas con el detalle de número de orden, número de línea e información de productos.

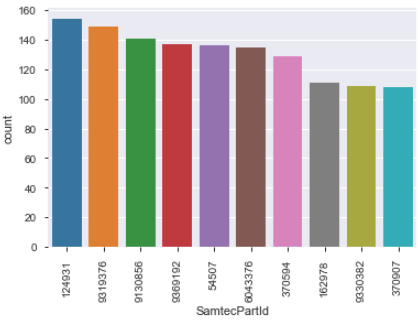
**Resultados**:

A continuación, se detallan las principales 10 reglas de asociación de productos encontradas para cada uno de los tipos de cliente distintos. A su vez se especifica el nivel de soporte mínimo utilizado en cada caso.

**Tipo de Cliente 1: OEM**

min\_support=0.00035 (equivalente a un 0.035%)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Antecedents | Consequents | Antecedent support | Consequent support | Support | Confidence | Lift | Leverage | Conviction |
| 17604 | 264623 | 0.00038593 | 0.000413497 | 0.000358 | 0.928571429 | 2245.657 | 0.000358204 | 13.99421105 |
| 264623 | 17604}) | 0.000413497 | 0.00038593 | 0.000358 | 0.866666667 | 2245.657 | 0.000358204 | 7.497105524 |
| 3128271 | 8127975 | 0.000413497 | 0.000799427 | 0.000358 | 0.866666667 | 1084.11 | 0.000358033 | 7.4940043 |
| 4298438 | 9130856 | 0.001957217 | 0.003280406 | 0.001654 | 0.845070423 | 257.6116 | 0.001647566 | 6.433371926 |
| 7016828 | 7016784 | 0.000441063 | 0.000441063 | 0.000358 | 0.8125 | 1842.141 | 0.000358169 | 5.330980998 |
| 7016784 | 7016828 | 0.000441063 | 0.000441063 | 0.000358 | 0.8125 | 1842.141 | 0.000358169 | 5.330980998 |
| 8245787 | 8245782 | 0.000496196 | 0.000578895 | 0.000386 | 0.777777778 | 1343.556 | 0.000385643 | 4.497394972 |
| 9389570 | 8770898 | 0.000744294 | 0.001433455 | 0.000579 | 0.777777778 | 542.5897 | 0.000577828 | 4.493549454 |
| 284789 | 266138 | 0.001047525 | 0.001488587 | 0.000772 | 0.736842105 | 494.9942 | 0.000770301 | 3.794343368 |
| 8132608 | 8127975 | 0.000606462 | 0.000799427 | 0.000441 | 0.727272727 | 909.7429 | 0.000440578 | 3.663735436 |

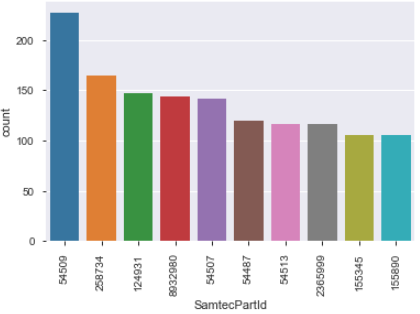


*Gráfico con partes de mayor rotación para tipo de cliente 1.*

**Tipo de Cliente 2: Distribuidor**

min\_support=0.00039 (equivalente a un 0.039%)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Antecedents | Consequents | Antecedent support | Consequent support | Support | Confidence | Lift | Leverage | Conviction |
| 5303641, 4889578 | 3634029 | 0.000732 | 0.001119 | 0.000689 | 0.941176 | 841.1584 | 0.000688 | 16.98098 |
| 8114060 | 8134959 | 0.000602 | 0.001033 | 0.000559 | 0.928571 | 899.0506 | 0.000559 | 13.98554 |
| 132361 | 86861 | 0.000559 | 0.000689 | 0.000516 | 0.923077 | 1340.596 | 0.000516 | 12.99105 |
| 8353505 | 3349346 | 0.000559 | 0.000904 | 0.000516 | 0.923077 | 1021.407 | 0.000516 | 12.98825 |
| 517332 | 164260 | 0.000516 | 0.001463 | 0.000473 | 0.916667 | 626.4877 | 0.000473 | 11.98244 |
| 4889578 | 3634029 | 0.000947 | 0.001119 | 0.000861 | 0.909091 | 812.4825 | 0.00086 | 10.98769 |
| 3349346, 393622 | 8353505 | 0.000516 | 0.000559 | 0.00043 | 0.833333 | 1489.551 | 0.00043 | 5.996643 |
| 517332 | 86861 | 0.000516 | 0.000689 | 0.00043 | 0.833333 | 1210.26 | 0.00043 | 5.995869 |
| 8353505, 3349346 | 393622 | 0.000516 | 0.000689 | 0.00043 | 0.833333 | 1210.26 | 0.00043 | 5.995869 |
| 4889578, 3634029 | 5303641 | 0.000861 | 0.001248 | 0.000689 | 0.8 | 641.0207 | 0.000687 | 4.99376 |

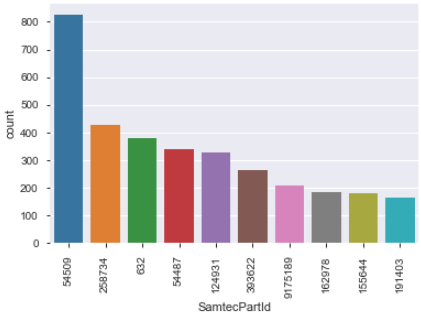


*Gráfico con partes de mayor rotación para tipo de cliente 2.*

**Tipo de Cliente 3: Fabricante por contrato**

min\_support=0.0003 (equivalente a un 0.03%)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Antecedents | Consequents | Antecedent support | Consequent support | Support | Confidence | Lift | Leverage | Conviction |
| 4342039 | 8931943 | 0.000578 | 0.000578 | 0.000542 | 0.9375 | 1621.934 | 0.000542 | 15.99075 |
| 359107 | 211145, 309294 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |
| 309294 | 211145 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |
| 359107, 309294 | 211145 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |
| 359107 | 211145 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |
| 8931943 | 4342039 | 0.000578 | 0.000578 | 0.000542 | 0.9375 | 1621.934 | 0.000542 | 15.99075 |
| 309294 | 211145, 359107 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |
| 359107 | 309294, 199054 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |
| 359107, 309294 | 211145, 193724 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |
| 359107, 309294 | 193724 | 0.000578 | 0.000542 | 0.000542 | 0.9375 | 1730.063 | 0.000542 | 15.99133 |



*Gráfico con partes de mayor rotación para tipo de cliente 3.*

**Plan de ataque**:

Se realiza un análisis de los datos agrupando las ordenes por tres tipos de cliente ya que cada uno tiene diferentes patrones de compra. Las compras de un fabricante de manufactura son distintas a las de un distribuidor de producto, por ejemplo.

Se identifica la información a continuación para cada tipo de cliente:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| IdTipoCliente | Cantidad líneas | Cantidad de ordenes | Cantidad de productos distintos | Cantidad de ordenes con más de 1 artículo | Cantidad de partes distintas en ordenes con más de 1 artículo |
| 1 | 90316 | 42891 | 35024 | 15529 | 25884 |
| 2 | 134706 | 39442 | 41215 | 14407 | 35544 |
| 3 | 108016 | 64050 | 24768 | 13417 | 16611 |

Se identifican cuáles son los clientes con mayor cantidad de ordenes y el porcentaje que abarcan para excluirlos del análisis de mercado.

Se realiza el análisis de la información utilizando el algoritmo Apriori en Python, ejecutado en Amazon SageMaker, utilizando una máquina con las siguientes características:

Nombre de la instancia: ml.m5.24xlarge

vCPU: 96, Mem (GiB): 384, Network Performance: 25 Gigabit

**Conclusiones**:

1. Si bien la cantidad de ordenes es grande, la cantidad de artículos en el inventario es tan grande que se hace necesario realizar una segmentación por tipo de cliente.
2. La segmentación por tipo de cliente nos enseña las diferentes necesidades de compra de acuerdo con el tipo de cliente. Se puede observar cómo un cliente de manufactura realiza compras de productos específicos debido a la actividad comercial que realiza, dichas listas de productos comprados se realizan para el consumo propio y su variación de productos no es tan grande, al contrario de un distribuidor que compra productos sin una relación específica ya que suple la demanda de diferentes clientes de manufactura.
3. En cuanto a la calidad de los datos, por el tipo de análisis realizado fueron utilizadas pocas columnas (id del cliente, número de orden, nombre del producto). Fue de gran ayuda contar con información completa en su mayoría.
4. Se podría realizar una mayor segmentación de la información si se contara con información de la industria a la que pertenece cada cliente, esto permitiría encontrar reglas de asociación más específica para cada tipo de cliente y tipos de producto.
5. Una de las opciones de análisis para realizar a futuro es un análisis de time series para identificar la frecuencia con la que se solicitan algunas partes.