

FACULTAD DE INGENIERÍA
INGENIERÍA EN INFORMÁTICA Y SISTEMAS
ANÁLISIS DE DATOS I
CATEDRÁTICO STANLEY BOLAÑOS

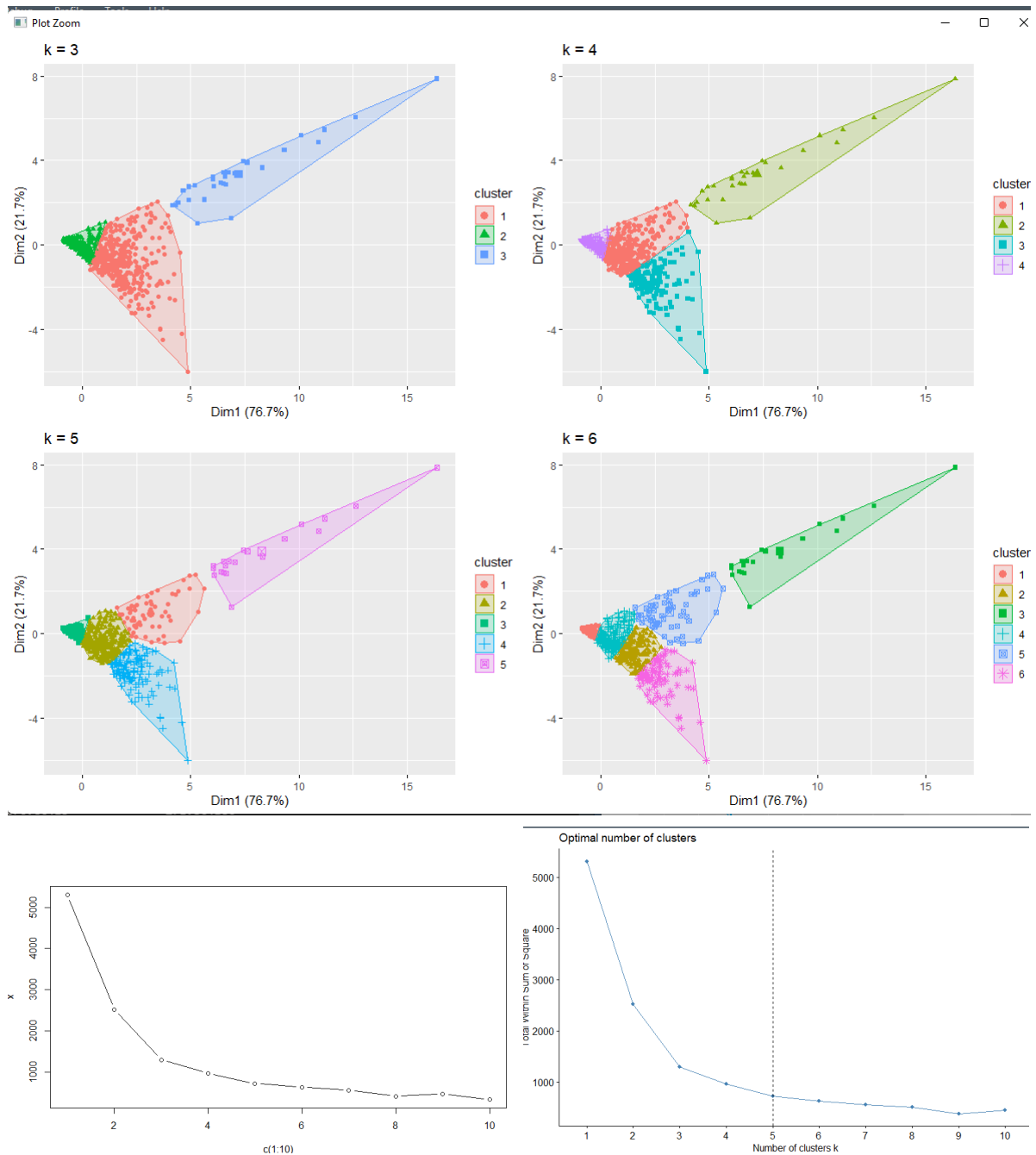
LABORATORIO #10

Clustering y Asociaciones

Lester Andrés García Aquino - 1003115
Mario Estuardo Gómez Orellana - 1020618
José Eduardo Tejeda León - 1097218
Mario Andrés Velasquez Banaay - 1092518

Guatemala, 07 de noviembre de 2021

PARTE I



1. ¿Cuál es el mejor número de clusters? ¿Por qué?

El mejor número de clusters considerado es de 5, ya que a pesar de la distancia tan reducida entre cada cluster, estos nunca llegan a superponerse entre sí por lo que se da una distribución balanceada entre los cluster para todos los registros de la vista que forma el data frame.

2. ¿Cuáles serían las etiquetas para cada cluster?

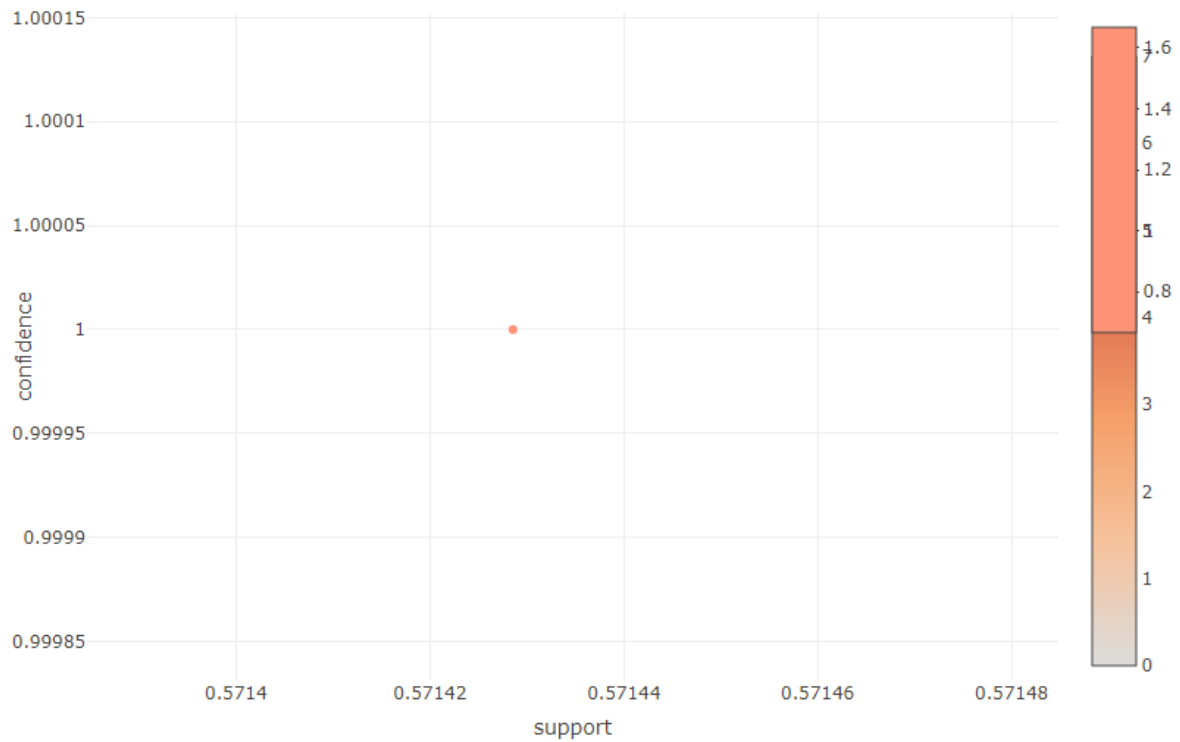
Grupo	Etiqueta	Descripción
1	Pro	Segundos que más cotizan y que generan ingresos
2	Básicos	Cuartos que más cotizan y que generan ingresos
3	Esencial	Menos partes cotizan, Menos ingreso genera
4	Standard	Terceros que más cotizan y que generan ingresos
5	Premium	Más ingresos generan y más cotizan

	Group.1	clustno	TotalPorParte	TotalPartesCotizadas	PromedioPartesCotizadas
1	1	1	490181.11	1641.33333	64.716667
2	2	2	149858.72	580.13898	72.196610
3	3	3	11274.06	42.62151	8.139442
4	4	4	231542.68	875.86429	172.364286
5	5	5	1186049.86	4292.15000	52.150000

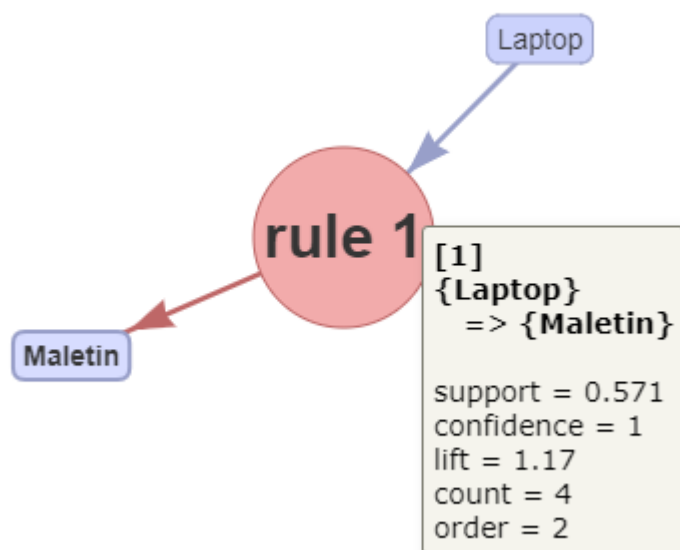


PARTE 2

1. Plot interactivo



Plot de grafos



1. ¿Existe alguna regla entre productos que tenga este nivel de confianza? ¿Cuáles?

Sí existen:

Laptop → Maletín

	lhs	rhs	support	confidence	lift	count
[1]	{Laptop}	=> {Maletín}	0.5714286	1	1.166667	4

Ya que tiene un lift cercano a 1, este es más probable que sea una coincidencia más que una asociación real.

2. ¿Cuál es el nivel de lift de cada regla encontrada?

1.17, ya que únicamente existe 1 regla.