

Análisis de DataSet en R

Orden de Compras y Ventas de Productos



ANALISIS DEL DATASET

Resumimos el Excel original, en un dataset agrupado por fecha, articulo, sumando las variables como cantidades y monto

1	A	B	D	E	G	H	K	M	O	P	Q	R	S	T	U	V	W
2	Fch.C	Nro de OC	Li	FCH_VEN	Orden de Venta	Vender A	Vendedor 1	Tipo de articulo	Articulo	Descripcion	UM	Qty O	Qty Factu	FILL E	Qty No V	Pcr Fill	VENT
39257	20/09/2022	W0010693		3	01/10/2022	OV216980	15090640	0027	002	500372	CHOCOLATE 60% CACAO 18X5	CJ	1	1,00	100,00	0,00	100,00
39258	20/09/2022	W0010693		4	01/10/2022	OV216980	15090640	0027	008	500916	M.ENLI MORA DPACK 12X350	CJ	1	1,00	100,00	0,00	100,00
39259	20/09/2022	W0010693		5	01/10/2022	OV216980	15090640	0027	004	500076	DUME ECK POTE S/A 16X330	CJ	1	1,00	100,00	0,00	100,00
39260	20/09/2022	W0010693		6	01/10/2022	OV216980	15090640	0027	001	500797	MUESLI FRUTAL 10X700	CJ	1	1,00	100,00	0,00	100,00
39261	20/09/2022	W0010693		7	01/10/2022	OV216980	15090640	0027	002	500353	CHOCOLA 85% CACAO 18X50G	CJ	1	1,00	100,00	0,00	100,00
39262	20/09/2022	W0010693		8	01/10/2022	OV216980	15090640	0027	017	500971	GALL CACAO NARAN 15X100G	CJ	1	1,00	100,00	0,00	100,00
39263	20/09/2022	W0010693		9	01/10/2022	OV216980	15090640	0027	002	500379	CHOCOLATE NUSS 18X50	CJ	1	1,00	100,00	0,00	100,00
39264	20/09/2022	W0010693		10	01/10/2022	OV216980	15090640	0027	008	500913	M.ENLI DZNO DPACK 12X350	CJ	1	0,00	0,00	1,00	0,00
39265	20/09/2022	W0010693		11	01/10/2022	OV216980	15090640	0027	017	500963	GALLETA COCO EN 15X120GR	CJ	1	1,00	100,00	0,00	100,00
39266	20/09/2022	W0010693		12	01/10/2022	OV216980	15090640	0027	008	500914	M.ENLI FRUT DPACK 12X350	CJ	1	1,00	100,00	0,00	100,00
39267	20/09/2022	W0010693		13	01/10/2022	OV216980	15090640	0027	014	500870	KETCHUP EN LINEA 14X350G	CJ	1	1,00	100,00	0,00	100,00
39268	20/09/2022	W0010693		14	01/10/2022	OV216980	15090640	0027	001	500799	MUESLI MULTI SEM10X700	CJ	1	1,00	100,00	0,00	100,00
39269	20/09/2022	W0010693		15	01/10/2022	OV216980	15090640	0027	017	500967	GALLE CHIPS CHOC 15X120G	CJ	1	1,00	100,00	0,00	100,00
39270	20/09/2022	W0010692		1	01/10/2022	OV216979	76756447	0027	008	500238	MERM PAST ECK FRAM 1XSKG	UN	10	9,00	90,00	1,00	90,00

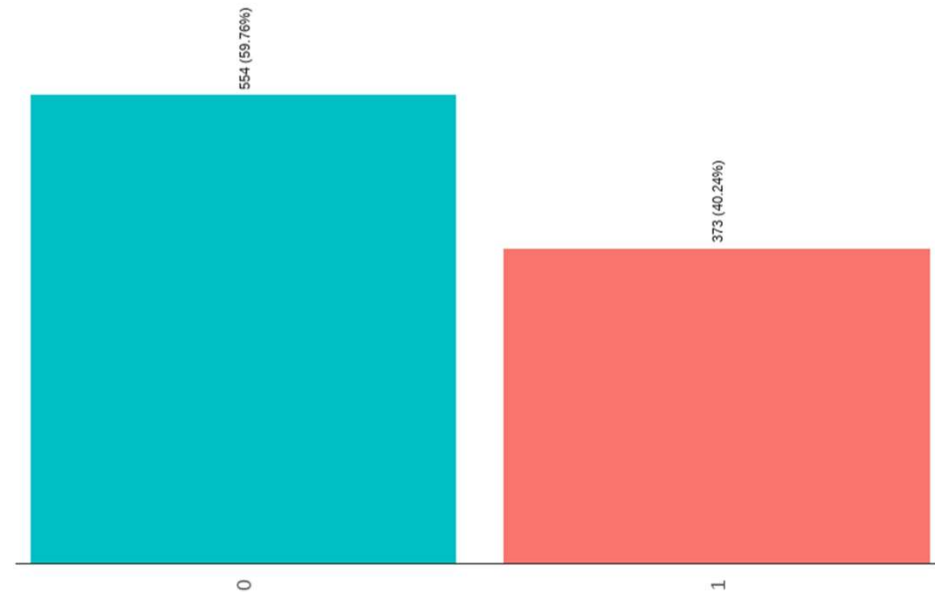


1	A	B	C	D	E	F	G	H	I	J
2	dia_oc	mes_oc	id_articulo	articulo	sum_Qty_Orden	sum_Qty_Facturada	mg_unit	stock	fill_Rate	
2	1	10	1	CEREALES	1321	1075	13569,72	1075	81,38	
3	1	10	2	CHOCOLATES	41	41	13370,11	67	100	
4	1	10	4	DULCE DE FRUTA	1056	736	12390,14	981	69,7	
5	1	10	5	DULCE DE LECHE	272	272	14377,85	563	100	
6	1	10	6	ENDULZANTE	48	38	19880,71	52	79,17	
7	1	10	8	MERMELADA	2387	1995	11078,77	1995	83,58	
8	1	10	11	SIN TIPO	8	8	28417,83	12	100	
9	1	10	14	ADEREZOS	41	41	11186,26	67	100	
10	1	10	17	GALLETAS	4	4	5777	6	100	
11	2	10	1	CEREALES	633	462	16520,01	616	72,99	
12	2	10	2	CHOCOLATES	74	74	15443,11	159	100	
13	2	10	4	DULCE DE FRUTA	235	233	14255,79	310	99,15	
14	2	10	5	DULCE DE LECHE	223	222	19214,81	296	99,55	
15	2	10	6	ENDULZANTE	24	18	39350,8	24	75	
16	2	10	8	MERMELADA	996	734	12813,58	978	73,69	
17	2	10	11	SIN TIPO	10	10	20034,59	15	100	
18	2	10	14	ADEREZOS	6	3	12955,16	3	50	

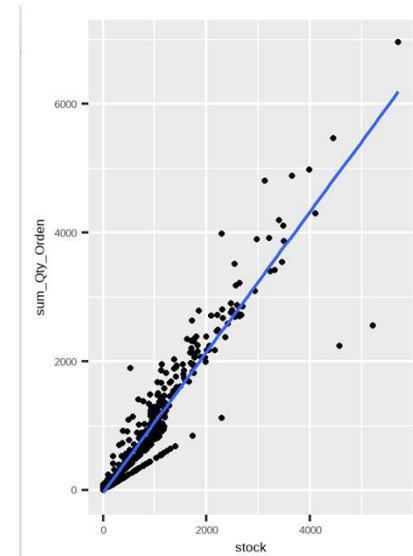
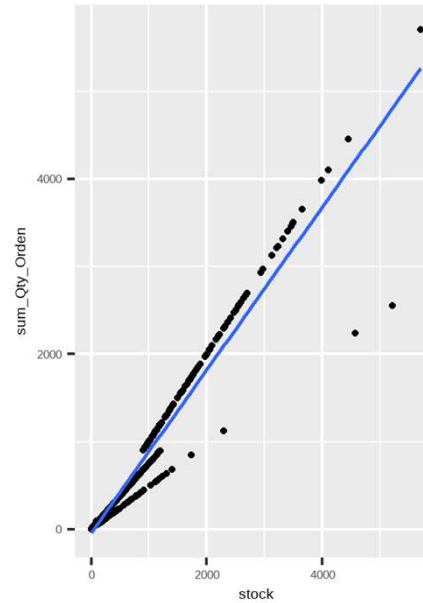
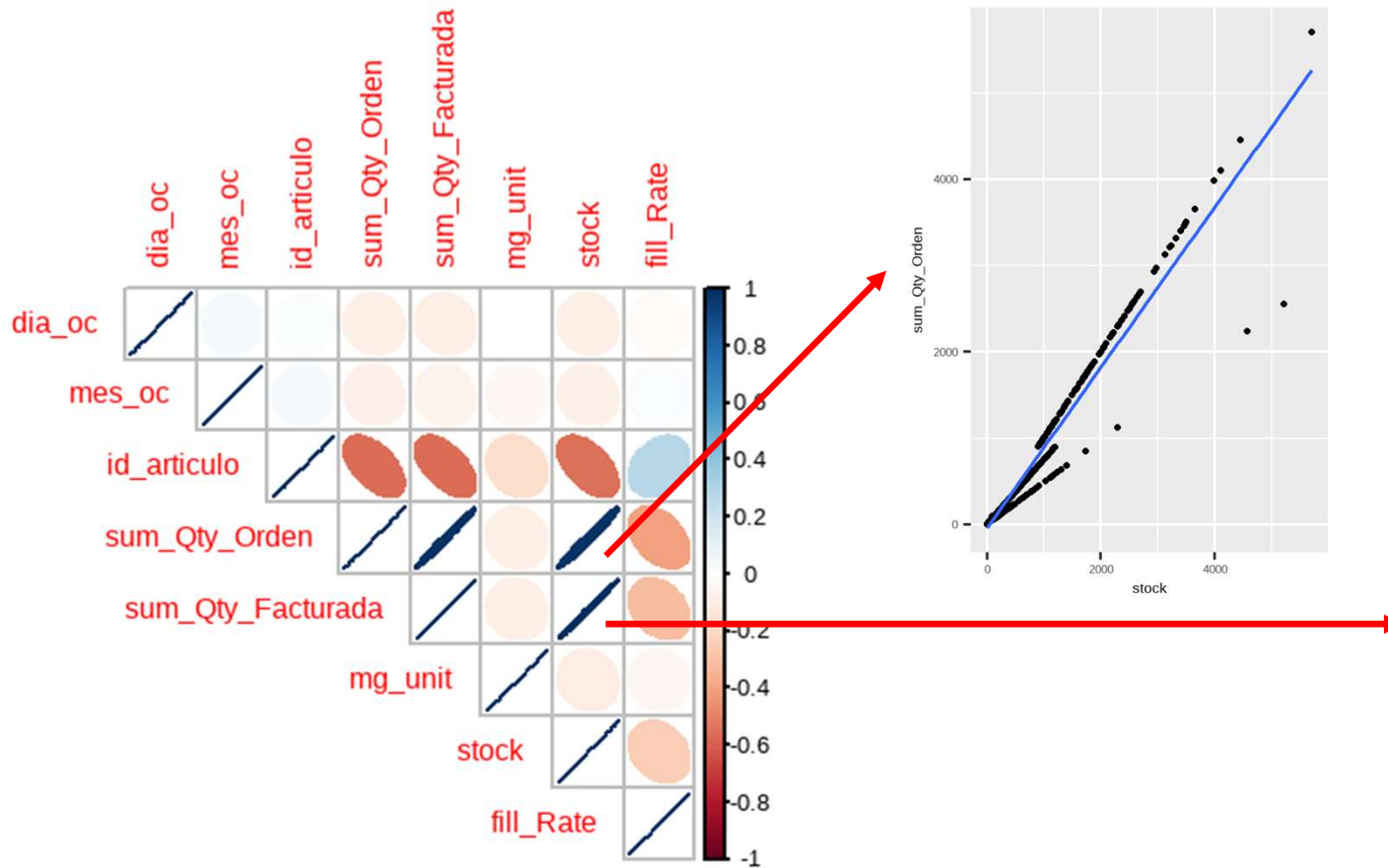
CORRELACION DE VARIABLES (Cumple o No cumple)

	dia_oc	mes_oc	id_articulo	sum_Qty_Orden	sum_Qty_Facturada	mg_unit	stock	fill_Rate
dia_oc	1.0000	0.032	0.015	-0.083	-0.085	-0.0047	-0.088	-0.021
mes_oc	0.0324	1.000	0.034	-0.080	-0.070	-0.0339	-0.077	0.023
id_articulo	0.0145	0.034	1.000	-0.554	-0.558	-0.1724	-0.543	0.272
sum_Qty_Orden	-0.0834	-0.080	-0.554	1.000	0.988	-0.0865	0.979	-0.407
sum_Qty_Facturada	-0.0849	-0.070	-0.558	0.988	1.000	-0.0834	0.996	-0.307
mg_unit	-0.0047	-0.034	-0.172	-0.087	-0.083	1.0000	-0.097	-0.043
stock	-0.0882	-0.077	-0.543	0.979	0.996	-0.0969	1.000	-0.247
fill_Rate	-0.0210	0.023	0.272	-0.407	-0.307	-0.0432	-0.247	1.000

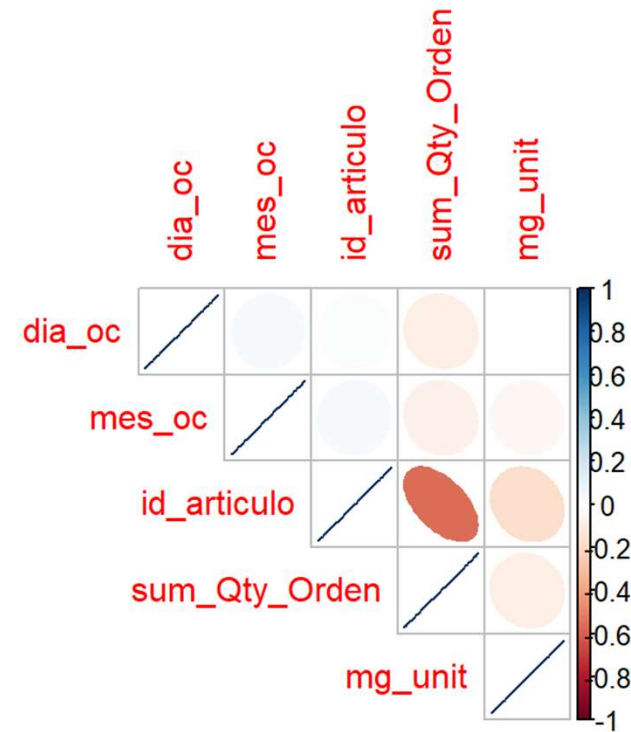
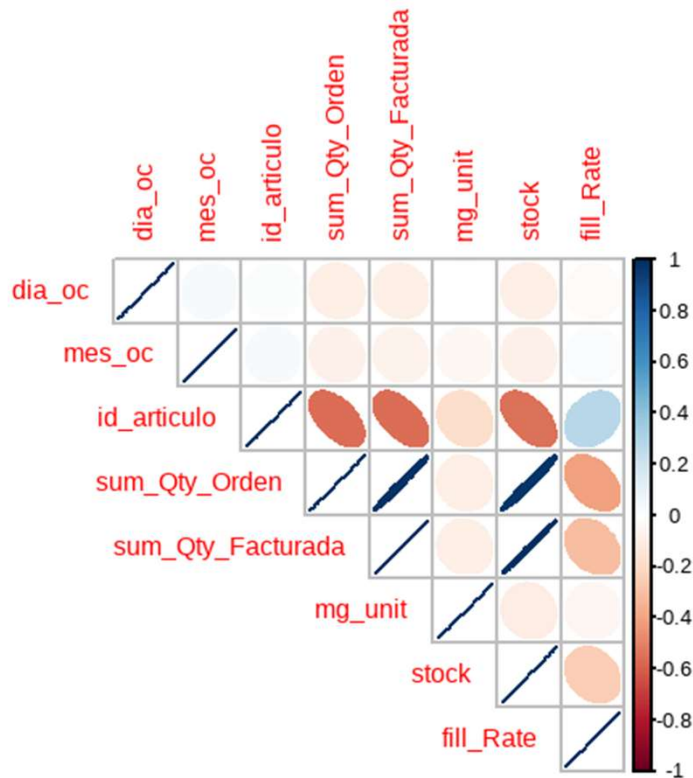
```
> dim(midata)
[1] 927  10
```



Correlación fuerte



Correlación, se eliminan variables



Inferencia de coeficientes modelo 1

Variables	Estimación	Std.Error	Z.value	P-valor
(Intercept)	0.8155	1.2325	0.6616	0.5082
dia_oc	-0.0151	0.0099	-1.5169	0.1293
mes_oc	-0.0681	0.1073	-0.6351	0.5254
id_articulo	0.0841	0.0174	4.8237	0.0000
sum_Qty_Orden	-0.0029	0.0005	-6.4586	0.0000
mg_unit	0.0000	0.0000	-2.2984	0.0215

##Inferencia de coeficientes modelo 1 y modelo 2

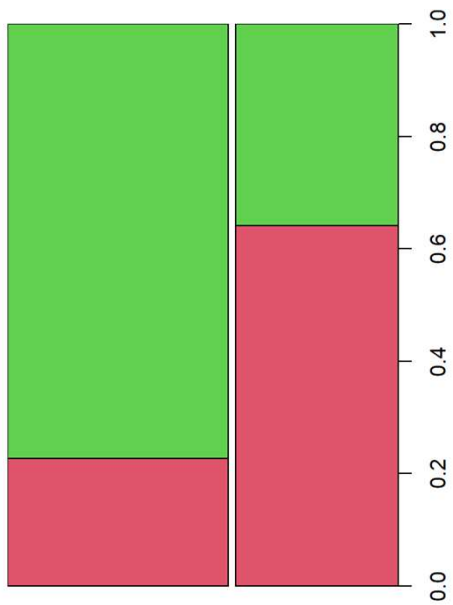
```
> table(pred=pred.clas,obs=test.data$cumplimiento)
      obs
pred  0  1
  0 82 24
  1 28 50
```

Variables	Estimación	Std.Error	Z.value	P-valor
(Intercept)	0.0622	0.3352	0.1855	0.8529
dia_oc	-0.0153	0.0099	-1.5386	0.1239
id_articulo	0.0844	0.0174	4.8425	0.0000
sum_Qty_Orden	-0.0029	0.0005	-6.4595	0.0000
mg_unit	0.0000	0.0000	-2.2987	0.0215

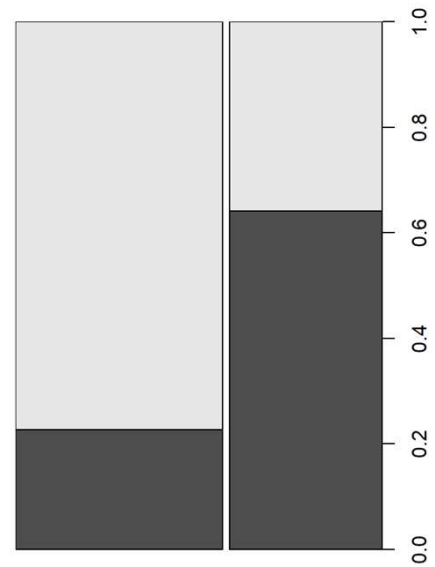
```
> table(pred=pred.clas2,obs=test.data$cumplimiento)
      obs
pred  0  1
  0 82 24
  1 28 50
```

Análisis de variables

```
plot(pred.clas,test.data$cumplimiento,col=c(2,3))
```

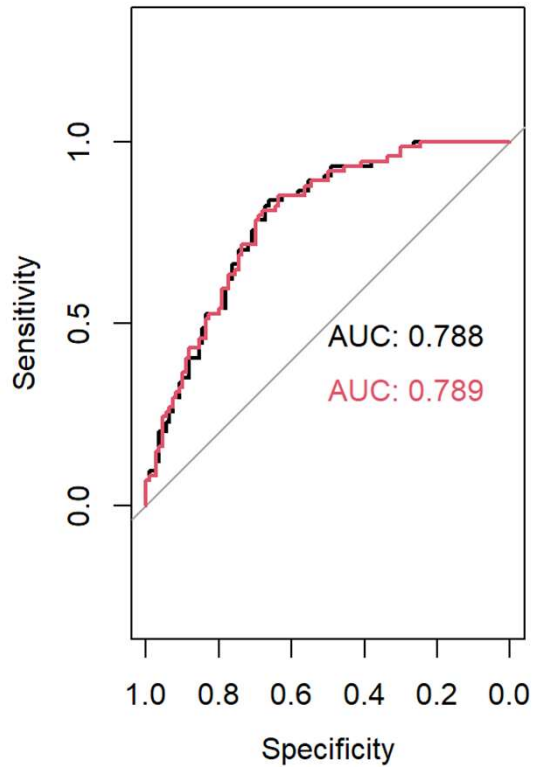


```
plot(pred.clas,test.data$cumplimiento,)
```



Análisis de variables

#área bajo la curva y medidas muestran
modelos idénticos



Medidas	Modelo_1	Modelo_2
Accuracy	0.7174	0.7174
Sensitivity	0.7455	0.7455
Specificity	0.6757	0.6757
Precision	0.7736	0.7736
Recall	0.7455	0.7455
F1	0.7593	0.7593

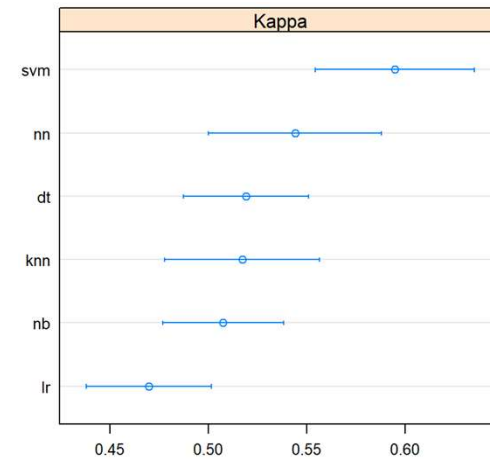
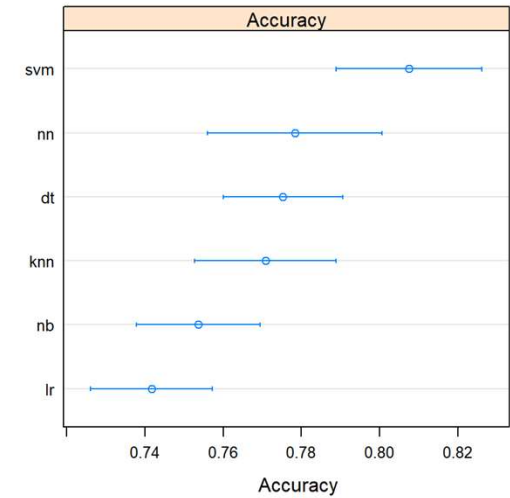
Análisis de variables

Accuracy

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
lr	0.6533333	0.7297297	0.7432432	0.7416941	0.7591892	0.8513514	0
dt	0.6933333	0.7466667	0.7702703	0.7753033	0.8074324	0.8648649	0
nb	0.6756757	0.7224324	0.7567568	0.7536358	0.7905405	0.8378378	0
nn	0.6351351	0.7358108	0.7702703	0.7783150	0.8133333	0.8800000	0
svm	0.7027027	0.7866667	0.8108108	0.8074961	0.8378378	0.9189189	0
knn	0.6666667	0.7432432	0.7702703	0.7708362	0.7993243	0.8918919	0

Kappa

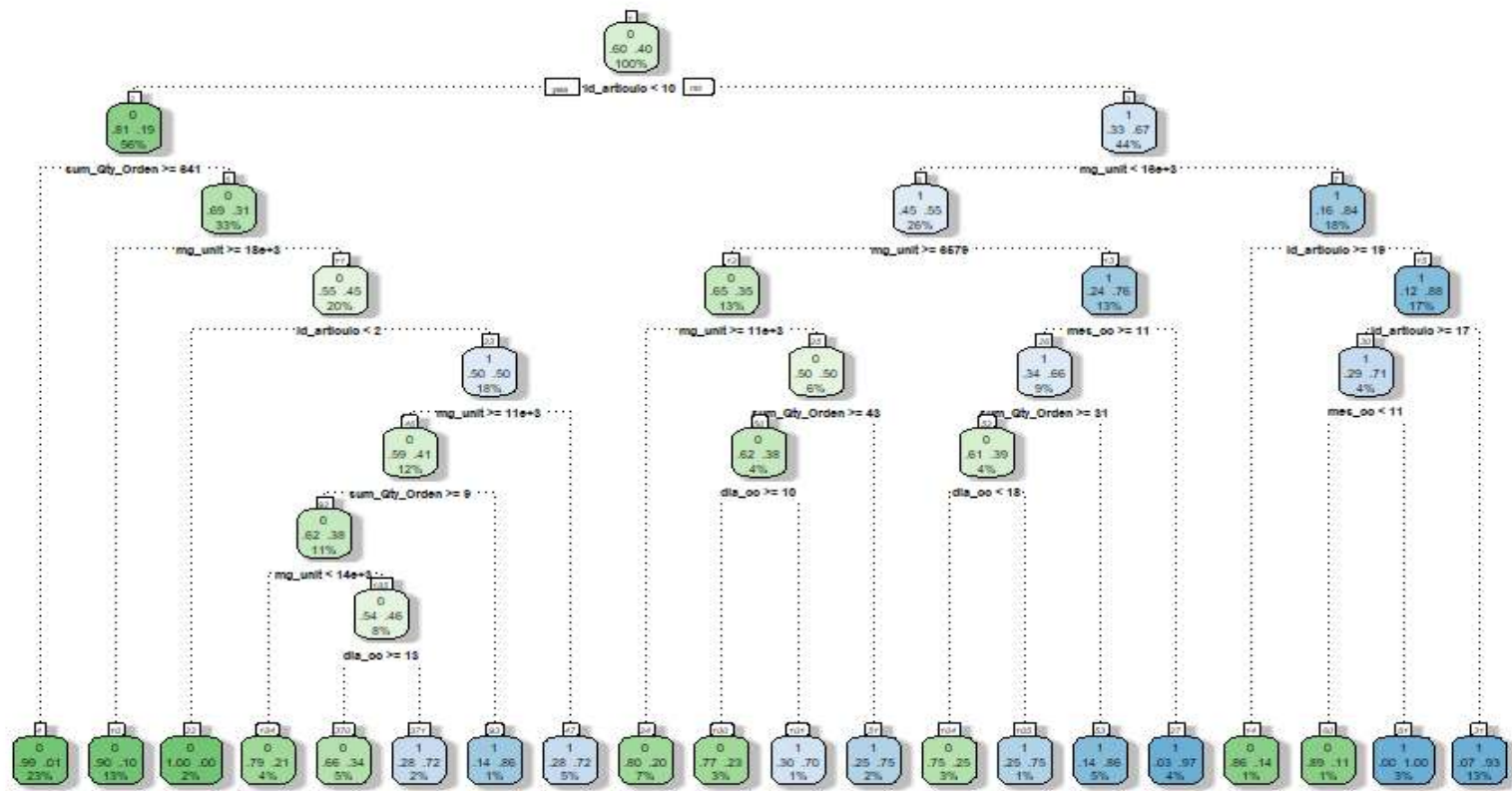
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
lr	0.2696629	0.4397628	0.4701746	0.4699814	0.4954545	0.6964952	0
dt	0.3850267	0.4586764	0.5072359	0.5190604	0.5777030	0.7072785	0
nb	0.3480176	0.4532877	0.5148410	0.5075778	0.5669985	0.6706231	0
nn	0.2702703	0.4700049	0.5404925	0.5440986	0.6186057	0.7513812	0
svm	0.3630673	0.5467314	0.6093434	0.5946496	0.6584049	0.8281734	0
knn	0.2896000	0.4480370	0.5178228	0.5171956	0.5839233	0.7733538	0



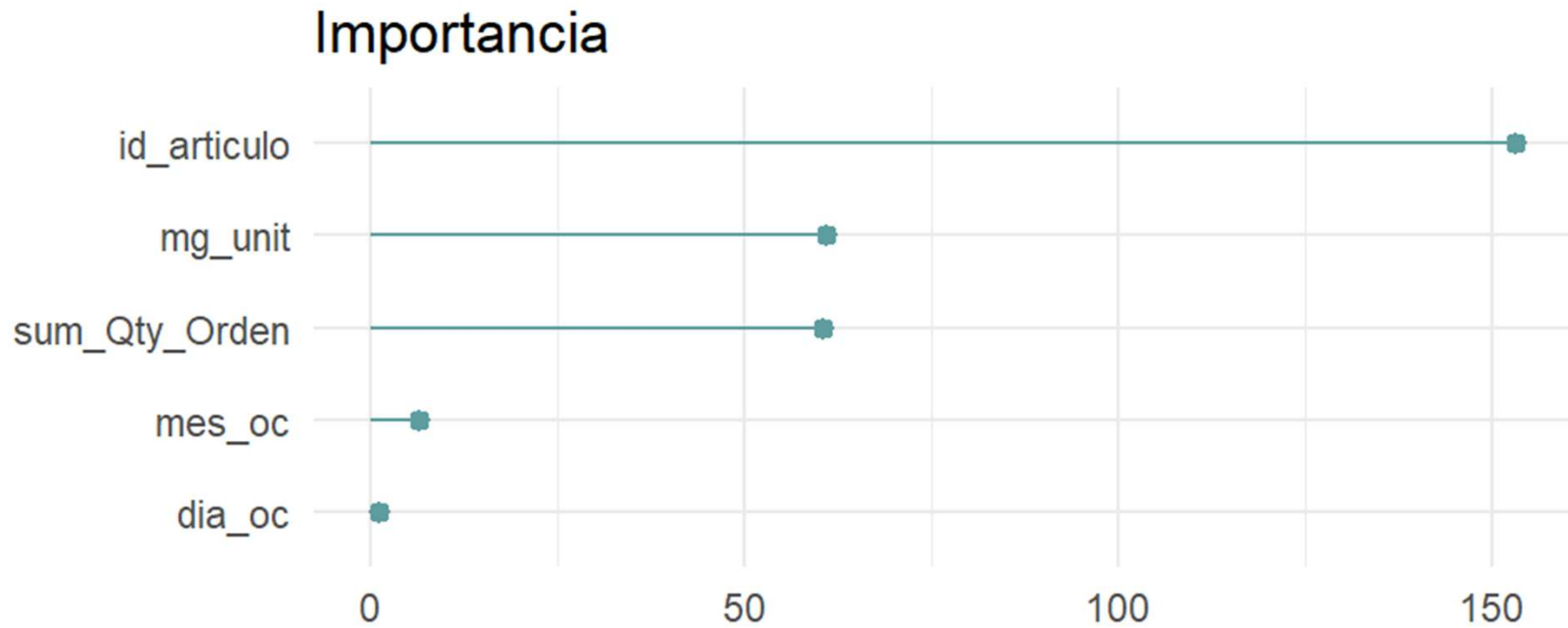
Análisis de variables

Medidas	LR	DT	NN	NB	SVM	KNN
Accuracy	0.7283	0.7880	0.7935	0.7663	0.8207	0.7989
Sensitivity	0.7545	0.9091	0.8273	0.7091	0.8455	0.8455
Specificity	0.6892	0.6081	0.7432	0.8514	0.7838	0.7297
Precision	0.7830	0.7752	0.8273	0.8764	0.8532	0.8230
Recall	0.7545	0.9091	0.8273	0.7091	0.8455	0.8455
F1	0.7685	0.8368	0.8273	0.7839	0.8493	0.8341

Árbol de decisión



Importancia de variables



Análisis de variables

```
> confusionMatrix(pred.diab1,test.data$cumplimiento)
```

Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	91	16
1	19	58

Accuracy : 0.81

95% CI : (0.746, 0.864)

No Information Rate : 0.598

P-Value [Acc > NIR] : 6.28e-10

Kappa : 0.607

Mcnemar's Test P-Value : 0.735

Sensitivity : 0.827

Specificity : 0.784

Pos Pred Value : 0.850

Neg Pred Value : 0.753

Prevalence : 0.598

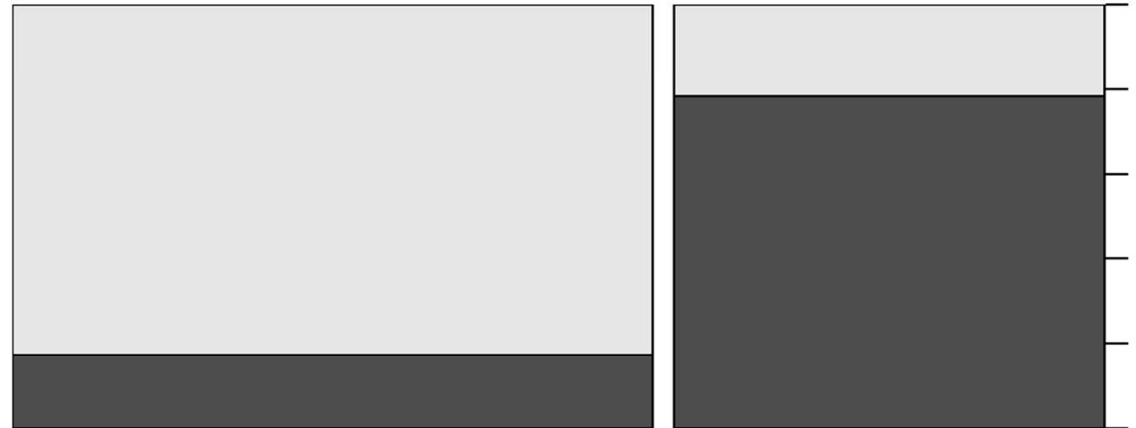
Detection Rate : 0.495

Detection Prevalence : 0.582

Balanced Accuracy : 0.806

'Positive' Class : 0

#Predecir#cp=0.001



Análisis de variables

Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	99	23
1	11	51

Accuracy : 0.815
95% CI : (0.751, 0.869)
No Information Rate : 0.598
P-Value [Acc > NIR] : 2.14e-10

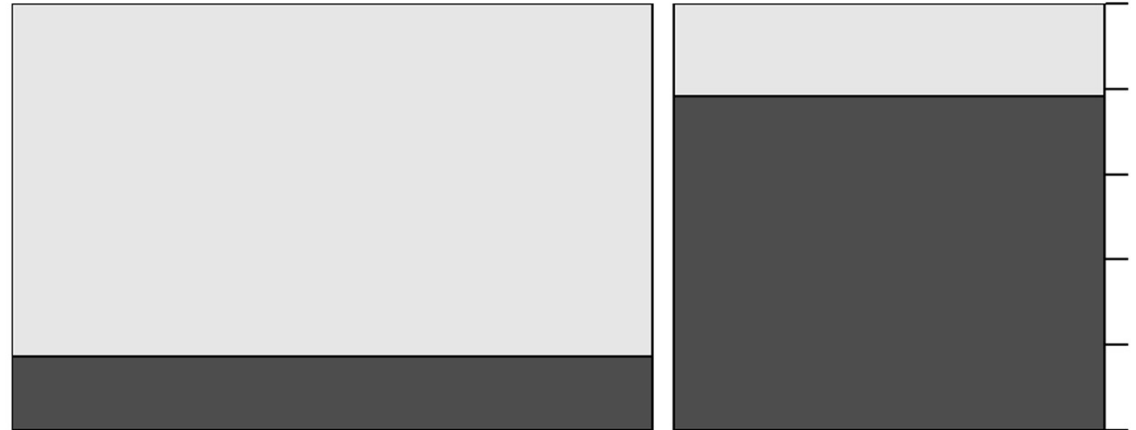
Kappa : 0.605

McNemar's Test P-Value : 0.0592

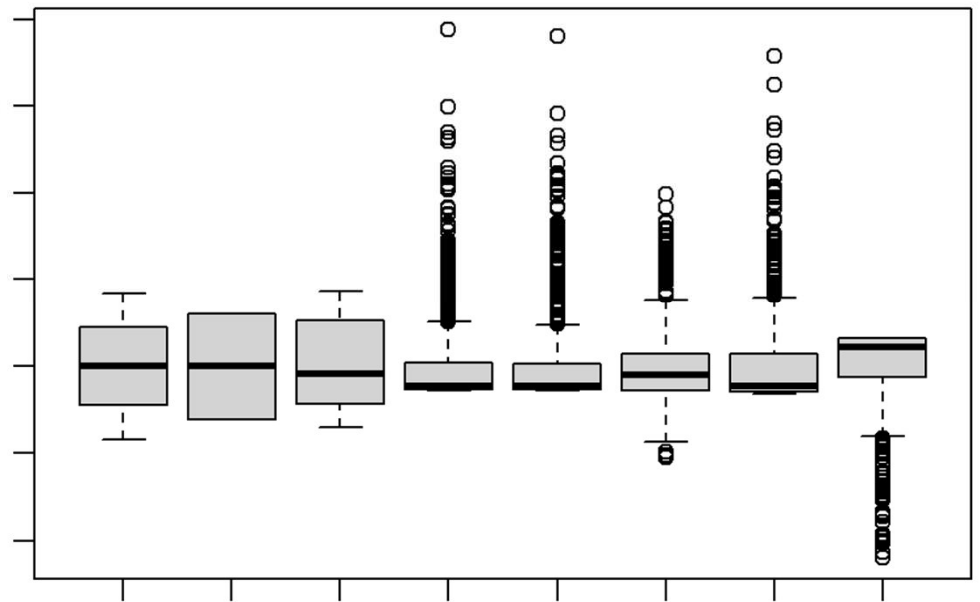
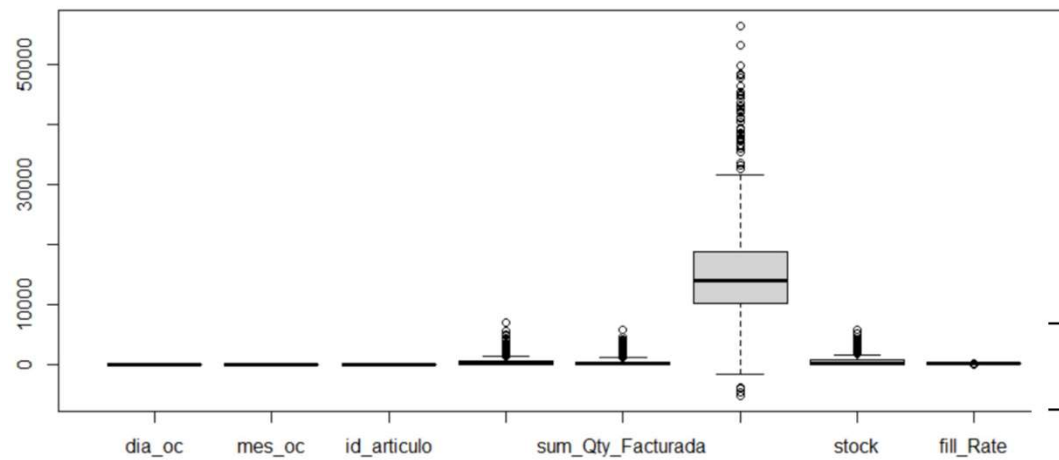
Sensitivity : 0.900
Specificity : 0.689
Pos Pred Value : 0.811
Neg Pred Value : 0.823
Prevalence : 0.598
Detection Rate : 0.538
Detection Prevalence : 0.663
Balanced Accuracy : 0.795

'Positive' Class : 0

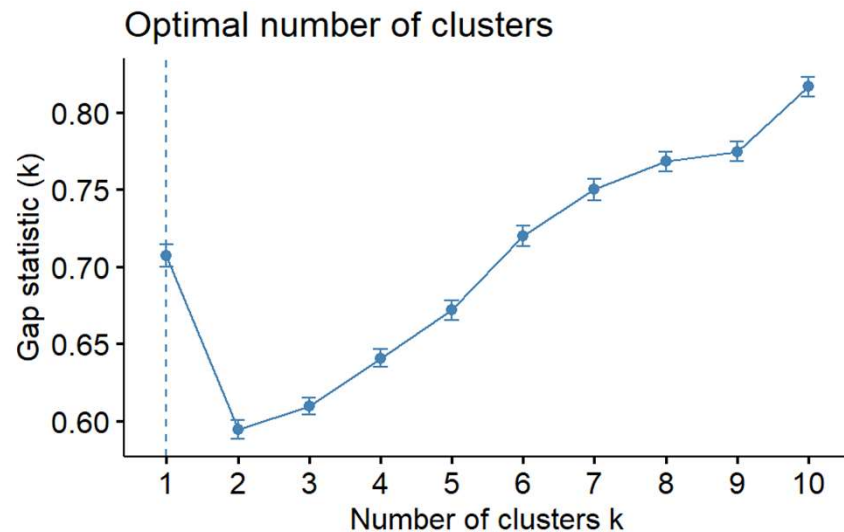
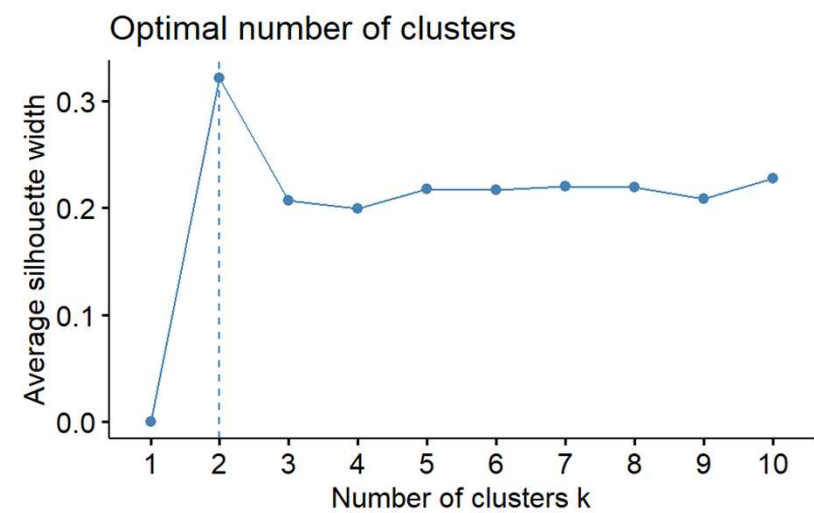
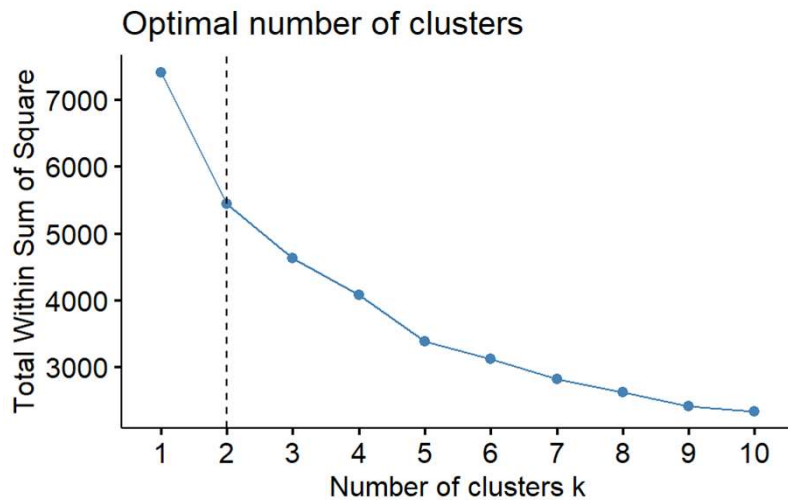
#modelo para el mejor cp



Análisis de variables

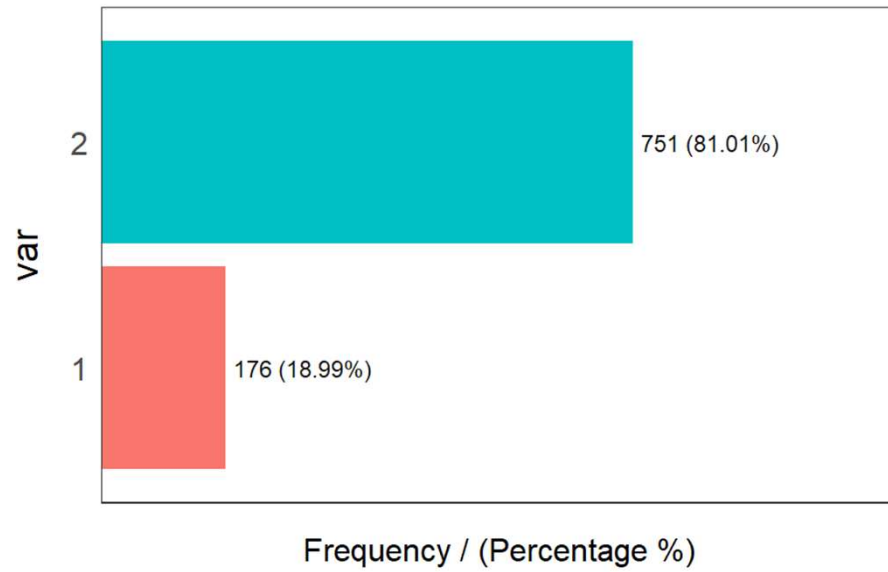


Análisis de variables

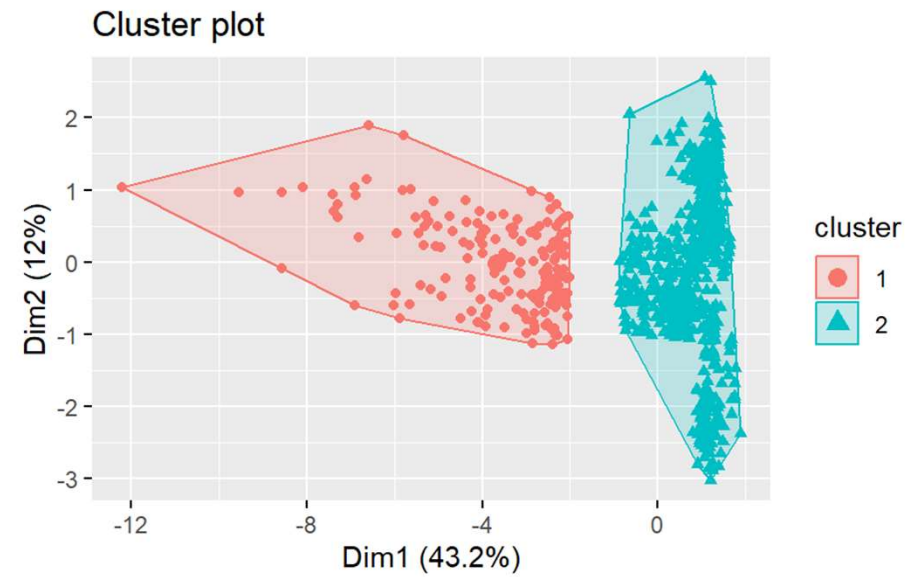


```
> fviz_nbclust(dej1, kmeans, method = "gap_stat")
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 100) [one "." per sample]:
..... 50
..... 100
Warning messages:
1: did not converge in 10 iterations
2: did not converge in 10 iterations
3: did not converge in 10 iterations
4: did not converge in 10 iterations
5: did not converge in 10 iterations
6: did not converge in 10 iterations
> |
```

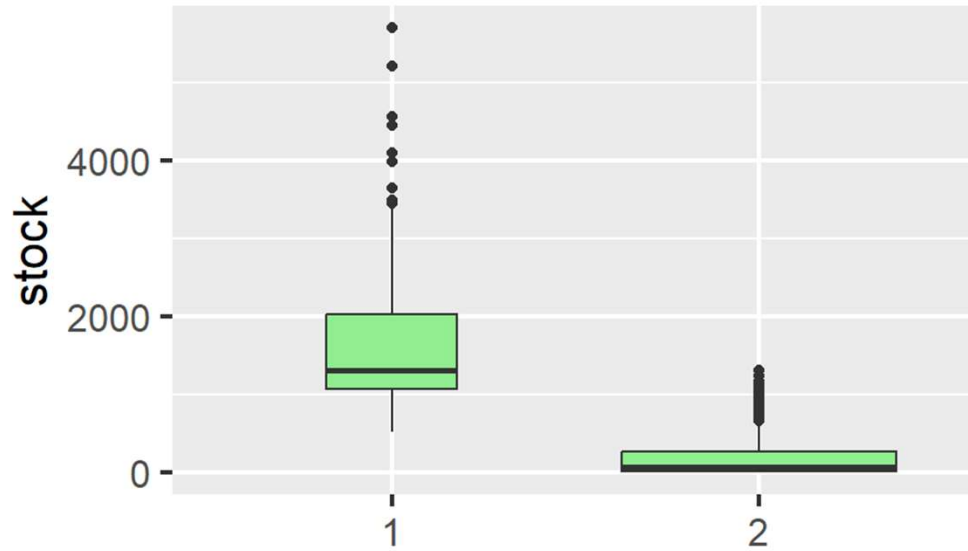

Análisis de variables



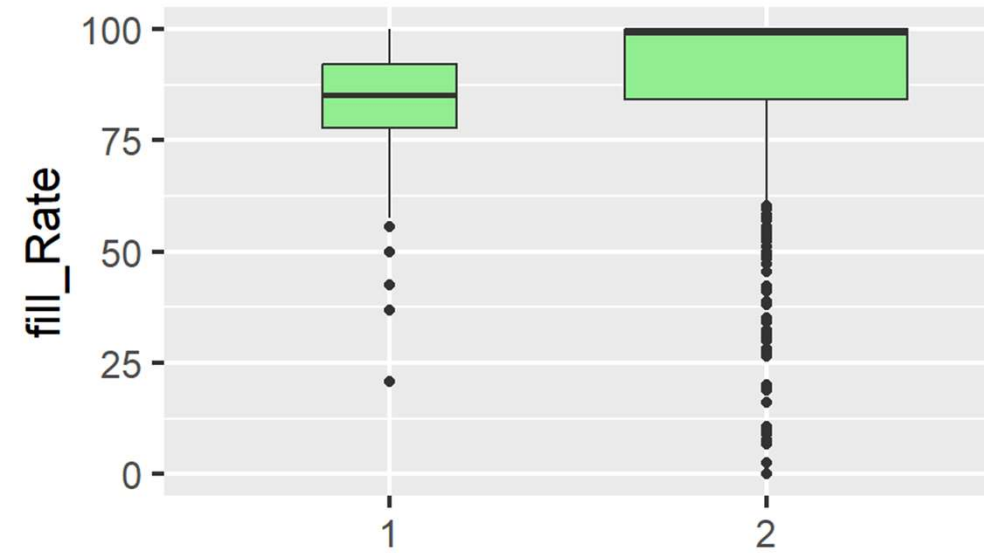
1. No cumple
2. Cumple



Análisis de variables

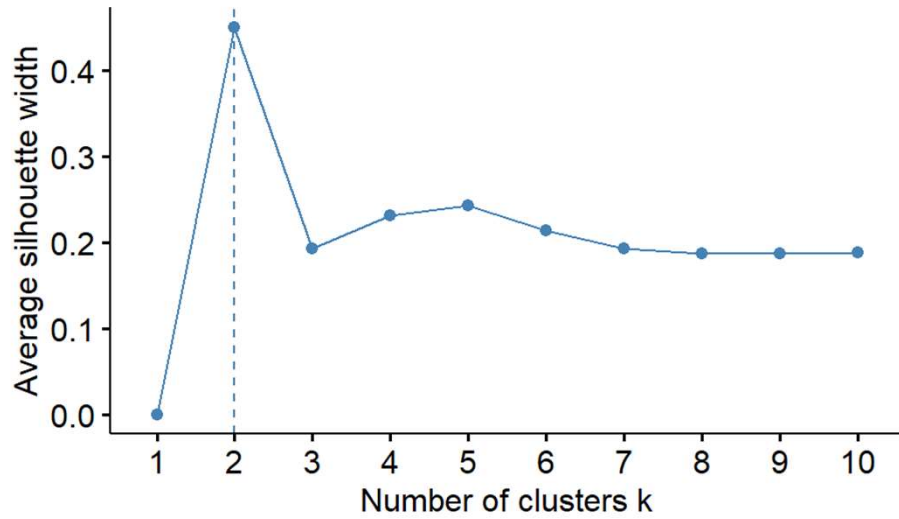


1. No cumple
2. Cumple



Análisis de variables

Optimal number of clusters



Cluster Dendrogram

