

Marcos Alcocer Gil

tutorizado por

Santiago Rojo Muñoz

ÁREA 1

TFM

Segmentación de clientes y optimización de su fidelización mediante aprendizaje computacional





1. Contexto



2. Segmentación



3. Transferencia



4. Fidelización



5. Resultados

Negocio



COLOMBIA

VENTA DE INFORMACIÓN COMERCIAL



**VENTA DE INFORMACIÓN
COMERCIAL**



evaluación crediticia



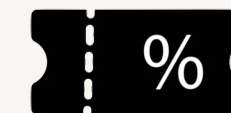
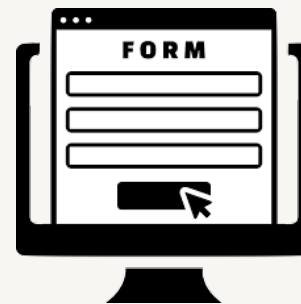
mercadotecnia



conocimiento de socios



analítica de datos



Herramientas



Python



Jupyter



Conda



NumPy



Pandas



Matplotlib



Seaborn



Plotly



Scikit-learn



XGBoost



TensorFlow



Keras



Statsmodels



Optuna

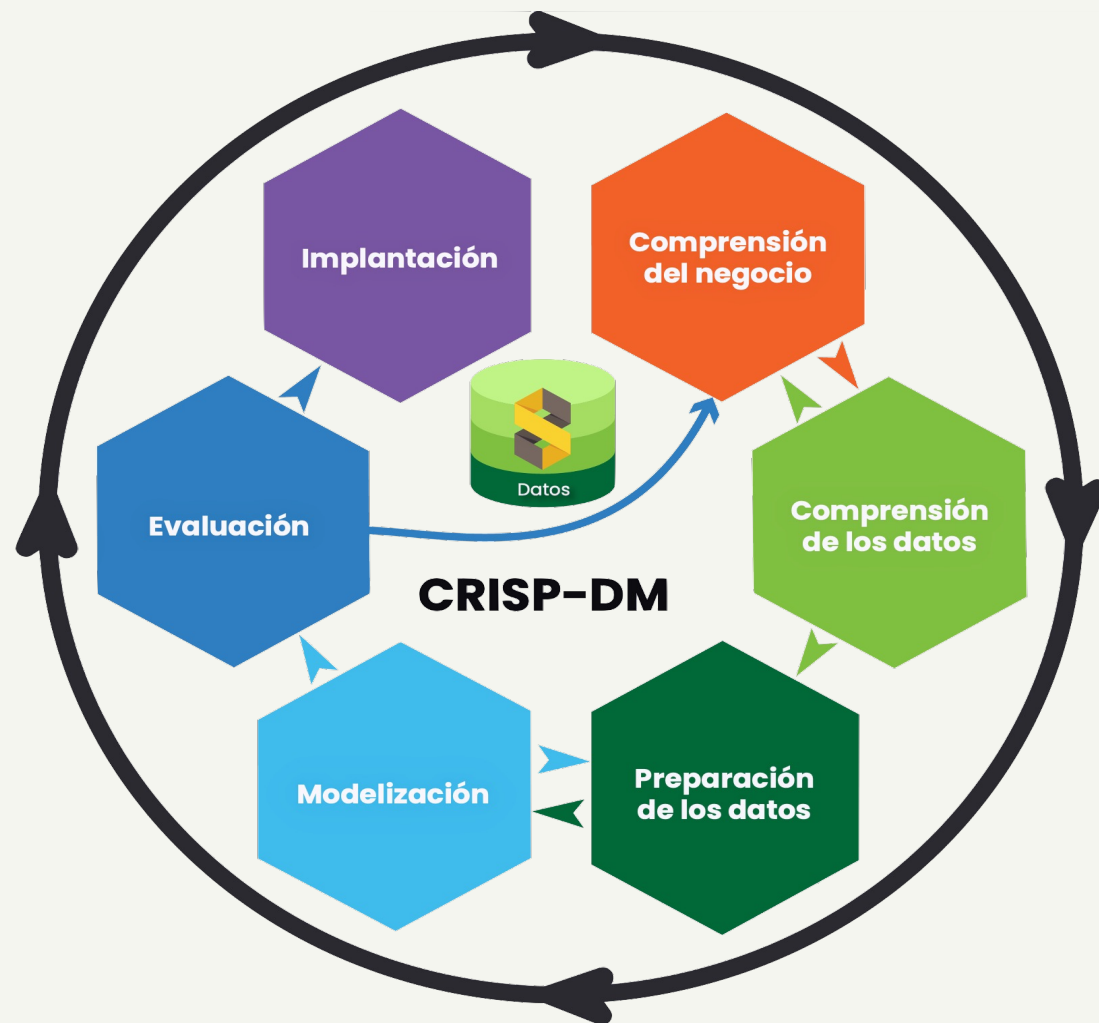


Imblearn



SciPy

Lean Startup





Concepto



Datos



CLIENTES.TXT



VENTAS.TXT



CONSULTAS.TXT

PERFIL EMPRESARIAL

FORMA JURÍDICA

SECTOR

ESTADO

DEPARTAMENTO

TAMAÑO

ONBOARDING

FECHA REGISTRO

FECHA PRIMERA COMPRA

DÍAS HASTA PRIMERA COMPRA

CANAL

CAMPAÑA

VENTAS

NÚMERO DE COMPRAS

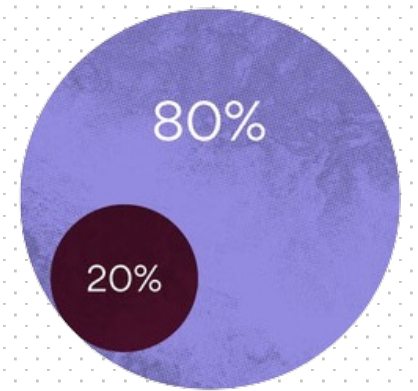
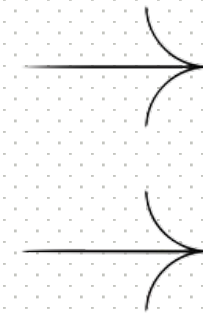
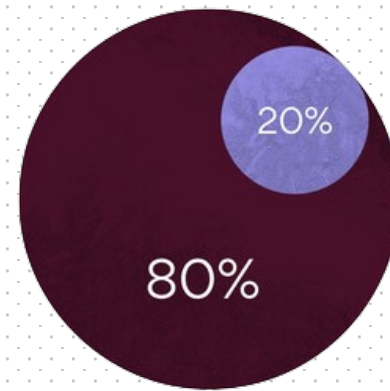
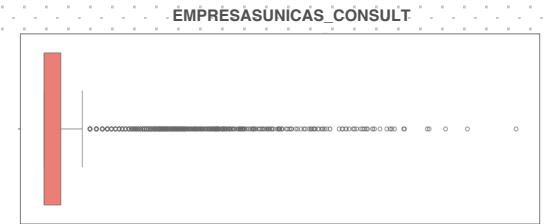
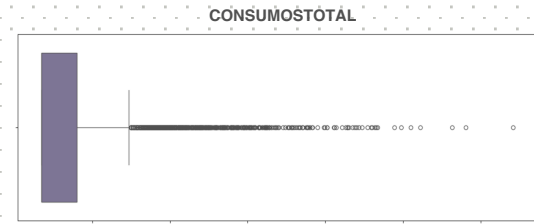
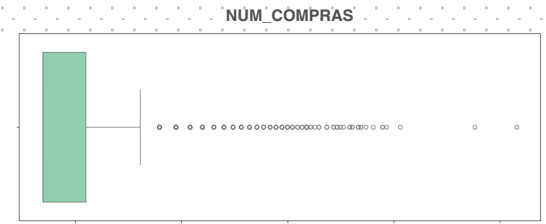
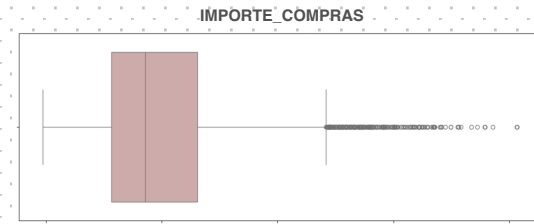
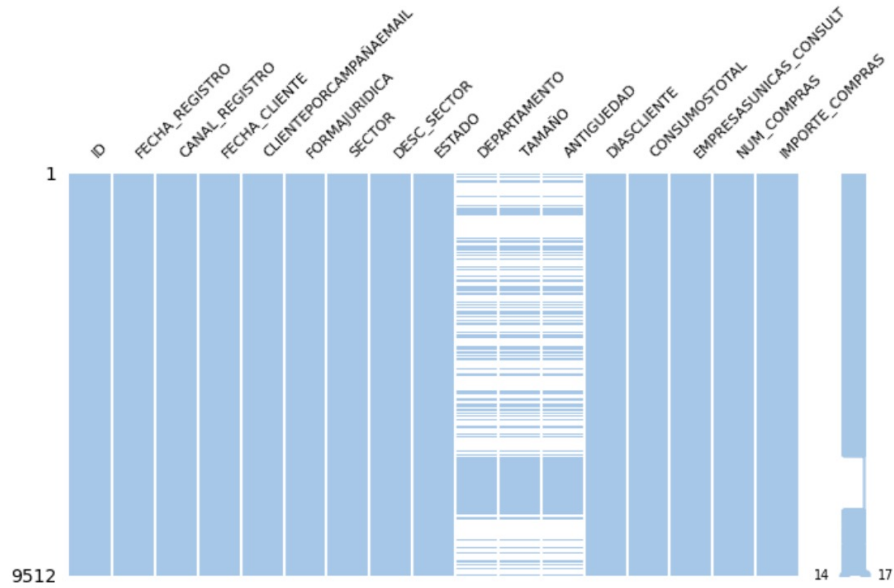
IMPORTE TOTAL

CONSULTAS

TOTAL CONSUMOS

EMPRESAS ÚNICAS CONSULTADAS

Inspección



Limpeza y codificación

1



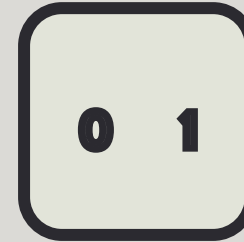
RECORTE
MODERADO DE
OUTLIERS

2



*FEATURE
ENGINEERING*

3



CODIFICACIÓN
NUMÉRICA ORDINAL
Y BINARIA

4



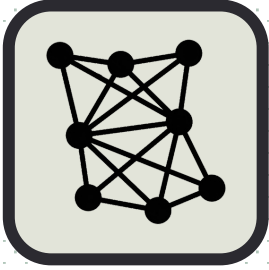
TRANSFORMACIÓN Y
ESTANDARIZACIÓN

Análisis de colinealidad

Variable 1	Variable 2	Correlation
SECTOR_NOSECTOR	ESTADO_VIVA	1.00
FORMAJURIDICA_PERSONA FISICA	ESTADO_VIVA	1.00
FORMAJURIDICA_PERSONA FISICA	SECTOR_NOSECTOR	1.00
FORMAJURIDICA_EMBAJADAS Y ORGANISMOS INTERNACI...	SECTOR_ACTIVIDADES DE ORGANIZACIONES Y ENTIDAD...	1.00
ANTIGUEDAD	FORMAJURIDICA_PERSONA FISICA	-0.99
ANTIGUEDAD	ESTADO_VIVA	-0.99
ANTIGUEDAD	SECTOR_NOSECTOR	-0.99
CONSUMOSTOTAL	EMPRESASUNICAS_CONSULT	0.98
TAMAÑO	ANTIGUEDAD	0.95
TAMAÑO	ESTADO_VIVA	-0.94
TAMAÑO	SECTOR_NOSECTOR	-0.94
TAMAÑO	FORMAJURIDICA_PERSONA FISICA	-0.94
TAMAÑO	FORMAJURIDICA_SOCIEDAD	0.81
ANTIGUEDAD	FORMAJURIDICA_SOCIEDAD	0.72
FORMAJURIDICA_PERSONA FISICA	FORMAJURIDICA_SOCIEDAD	-0.72
FORMAJURIDICA_SOCIEDAD	SECTOR_NOSECTOR	-0.72
FORMAJURIDICA_SOCIEDAD	ESTADO_VIVA	-0.72

PCA

PC	Main Feature	Explained Variance Ratio	Cumulative Explained Variance Ratio
PC1	ANTIGUEDAD	12.13%	12.13%
PC2	CONSUMOSTOTAL	6.77%	18.90%
PC3	FORMAJURIDICA_EMPRESARIO	4.59%	23.49%
PC4	FORMAJURIDICA_ESAL	3.71%	27.20%
PC5	DIASCLIENTE	3.51%	30.71%
PC6	CANAL_REGISTRO_SEM	3.08%	33.79%
PC7	FORMAJURIDICA_SOCIEDAD EXTRANJERA	2.98%	36.78%
PC8	SECTOR_ACTIVIDADES PROFESIONALES, CIENTÍFICAS ...	2.87%	39.64%
PC9	SECTOR_COMERCIO AL POR MAYOR Y AL POR MENOR; R...	2.82%	42.46%
PC10	ESTADO_EXTINGUIDA	2.71%	45.18%
PC11	SECTOR_INDUSTRIAS MANUFACTURERAS	2.69%	47.86%
PC12	SECTOR_INFORMACIÓN Y COMUNICACIONES	2.62%	50.48%
PC13	SECTOR_INFORMACIÓN Y COMUNICACIONES	2.60%	53.08%
PC14	SECTOR_ACTIVIDADES INMOBILIARIAS	2.58%	55.66%
PC15	SECTOR_ACTIVIDADES DE SERVICIOS ADMINISTRATIVO...	2.55%	58.20%
PC16	SECTOR_COMERCIAL / INDUSTRIAL NO DEFINIDA	2.54%	60.74%
PC17	SECTOR_AGRICULTURA, GANADERÍA, CAZA, SILVICULT...	2.53%	63.27%
PC18	SECTOR_EDUCACIÓN	2.53%	65.80%
PC19	SECTOR_DISTRIBUCIÓN DE AGUA; EVACUACIÓN Y TRAT...	2.52%	68.31%
PC20	SECTOR_SUMINISTRO DE ELECTRICIDAD, GAS, VAPOR ...	2.52%	70.83%



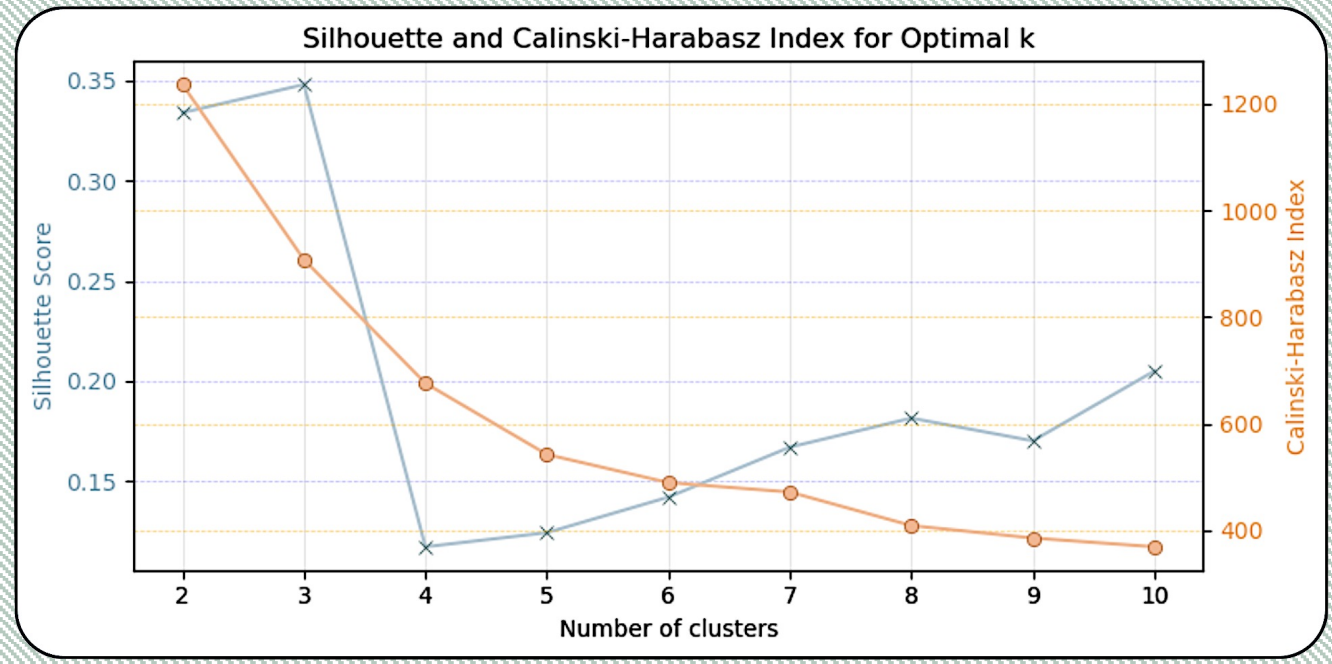
K-MEDOIDS

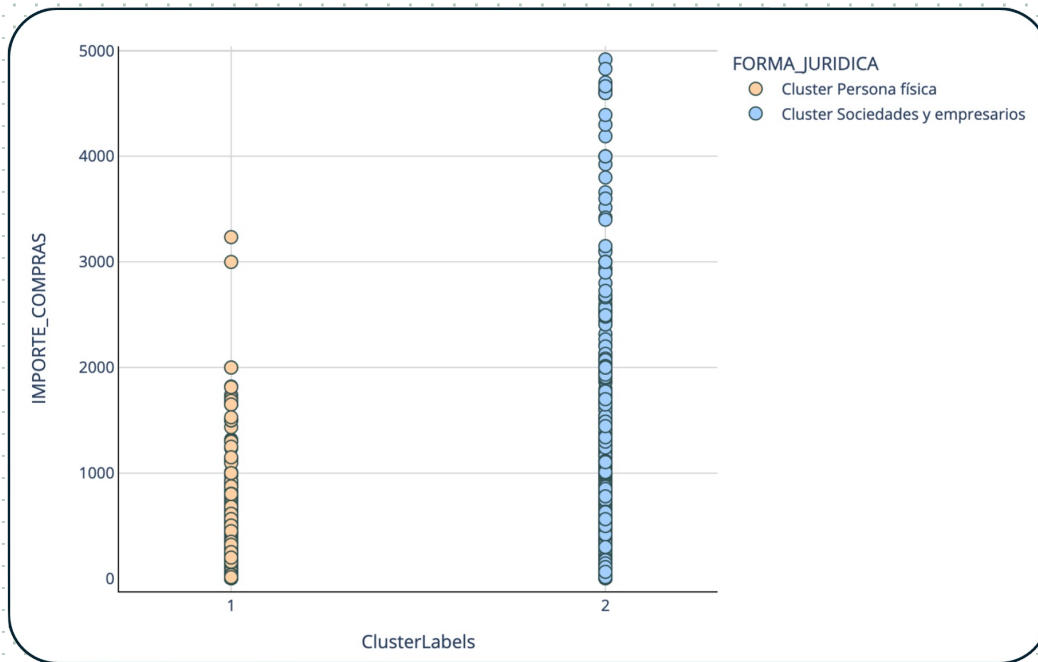
basado en distancias
robusto frente a *outliers*



MANHATTAN

robusta frente a *outliers*



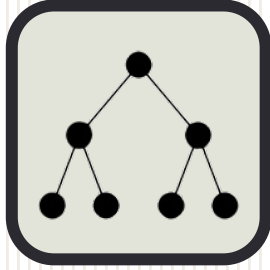


FORMA_JURIDICA	Statistic	1	2
Persona física	Counts	5591	0
	Distribution (%)	100.00	0.00
Sociedades y empresarios	Counts	0	3867
	Distribution (%)	0.00	100.00

Segmentación dominada por el perfil empresarial

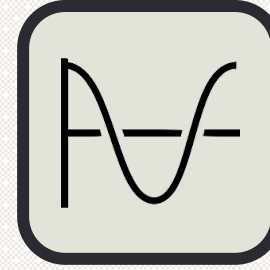
Variable	Importance
FORMAJURIDICA_SOCIEDAD	1.14
FORMAJURIDICA_PERSONA FISICA	1.02
ANTIGUEDAD	0.87
TAMAÑO	0.72
DEPARTAMENTO_SCORE	0.33
IMPORTE_COMPRAS	0.12

Feature	Baseline Score	Score Without	Score Change
FORMAJURIDICA_PERSONA FISICA	0.333848	0.313858	0.019990
ANTIGUEDAD	0.333848	0.318582	0.015266
TAMAÑO	0.333848	0.319829	0.014019
FORMAJURIDICA_SOCIEDAD	0.333848	0.324457	0.009391
SECTOR_COMERCIO AL POR MAYOR Y AL POR MENOR; R...	0.333848	0.326870	0.006978
SECTOR_ACTIVIDADES INMOBILIARIAS	0.333848	0.328171	0.005678
SECTOR_SUMINISTRO DE ELECTRICIDAD, GAS, VAPOR ...	0.333848	0.329183	0.004666
SECTOR_ACTIVIDADES PROFESIONALES, CIENTÍFICAS ...	0.333848	0.329330	0.004518
FORMAJURIDICA_EMPRESARIO	0.333848	0.329408	0.004440
SECTOR_INDUSTRIAS MANUFACTURERAS	0.333848	0.330213	0.003635
SECTOR_ALOJAMIENTO Y SERVICIOS DE COMIDA	0.333848	0.330378	0.003470
SECTOR_INFORMACIÓN Y COMUNICACIONES	0.333848	0.331379	0.002469
ESTADO_INACTIVA	0.333848	0.331487	0.002361
SECTOR_ACTIVIDADES DE SERVICIOS ADMINISTRATIVO...	0.333848	0.331891	0.001958
SECTOR_CONSTRUCCIÓN	0.333848	0.332237	0.001611
DEPARTAMENTO_SCORE	0.333848	0.332443	0.001406
SECTOR_TRANSPORTE Y ALMACENAMIENTO	0.333848	0.332722	0.001126
SECTOR_ACTIVIDADES FINANCIERAS Y DE SEGUROS	0.333848	0.332822	0.001027
SECTOR_OTRAS ACTIVIDADES DE SERVICIOS	0.333848	0.333463	0.000385
SECTOR_ACTIVIDADES DE ATENCIÓN DE LA SALUD HUM...	0.333848	0.333508	0.000340
SECTOR_COMERCIAL / INDUSTRIAL NO DEFINIDA	0.333848	0.333698	0.000150
SECTOR_AGRICULTURA, GANADERÍA, CAZA, SILVICULT...	0.333848	0.333799	0.000049



AGLOMERATIVO

calcula la máxima
distancia entre puntos



AFINIDAD DEL COSENO

comportamiento equivalente
independiente la magnitud

ATRIBUTOS DE VENTAS



importe total
de compras



número de
compras

SEGMENTACIÓN POR CLÚSTER



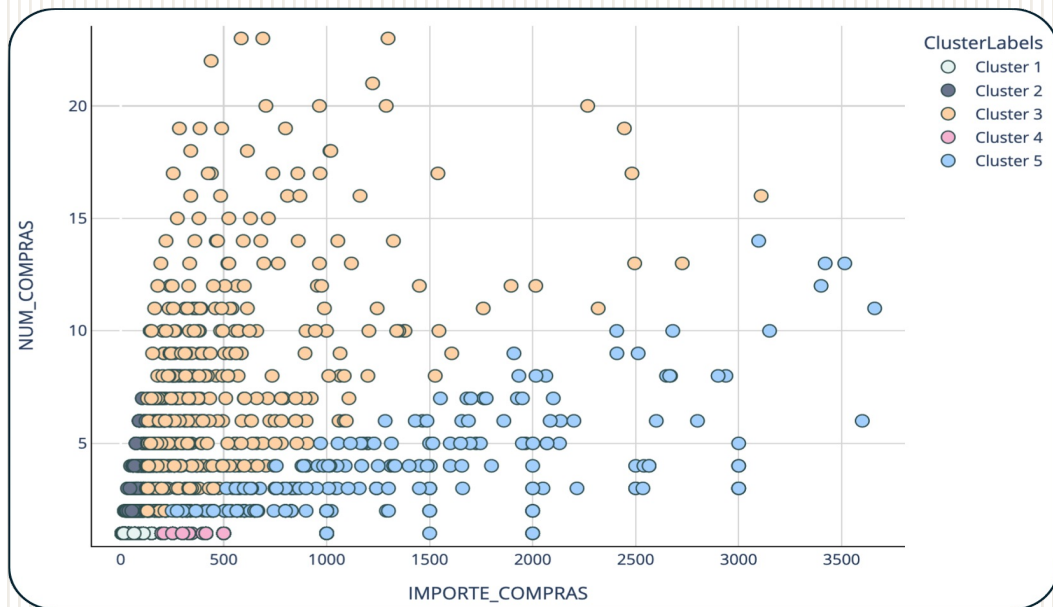
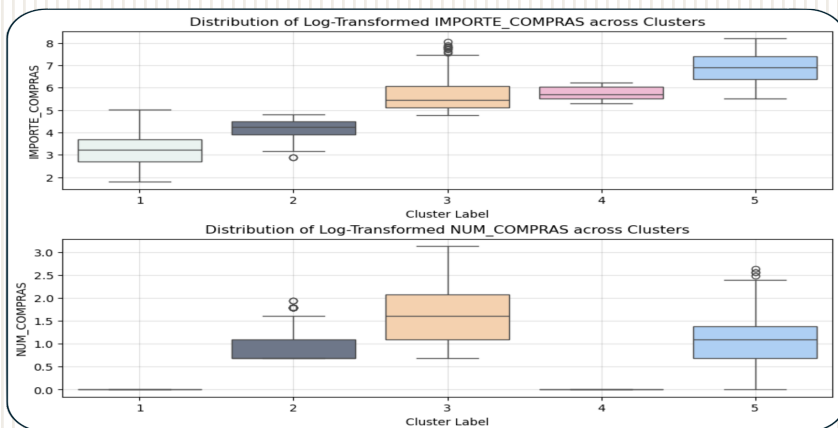
sociedades y
empresarios



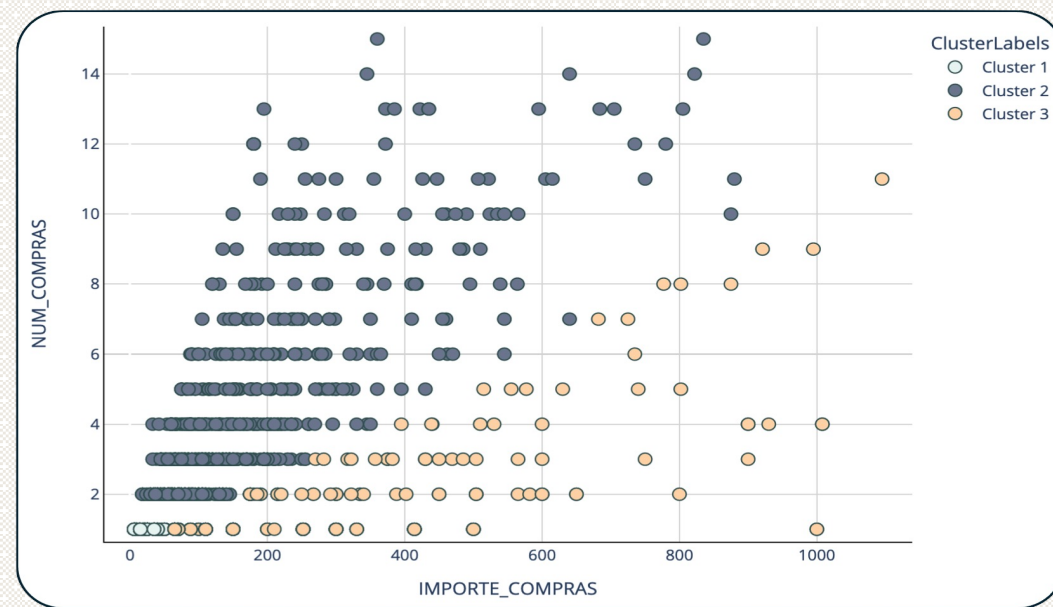
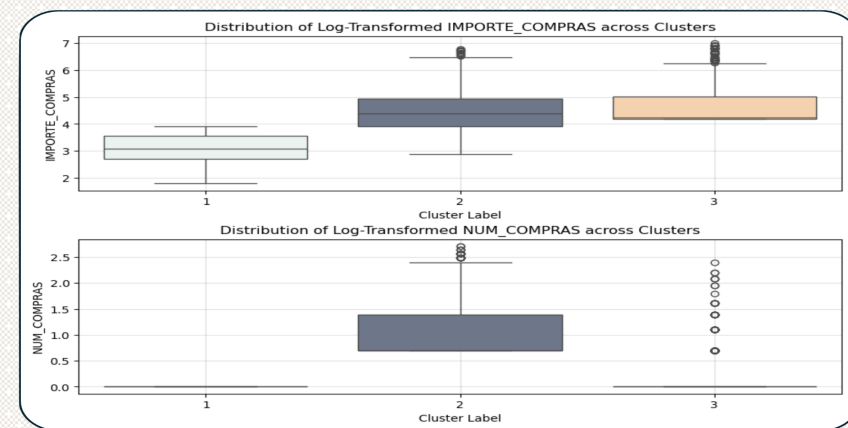
personas
físicas



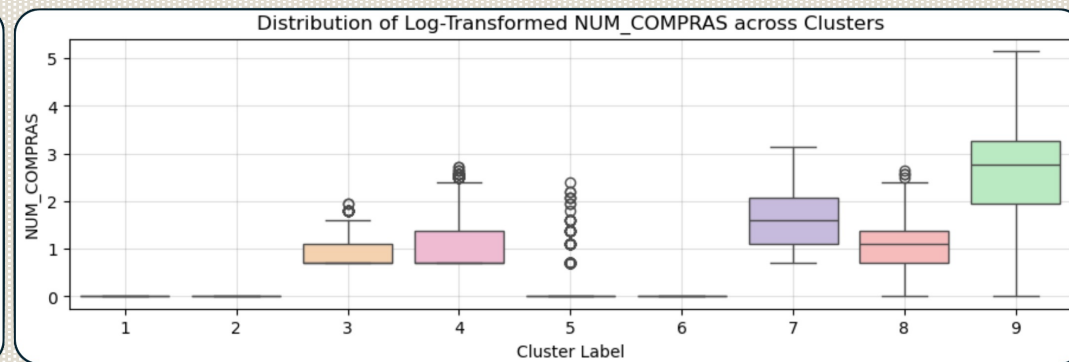
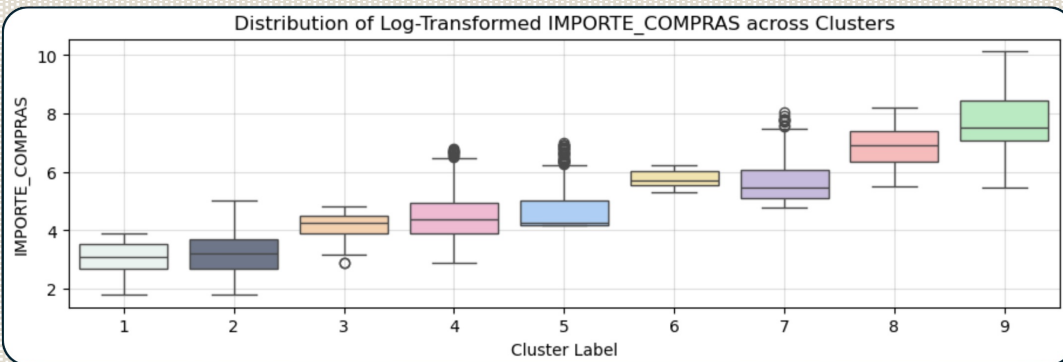
SOCIEDADES Y EMPRESARIOS



PERSONAS FÍSICAS



Resultados



	1	2	3	4	5	6	7	8	9
ANTIGUEDAD	No aplicable	Más de 10 Años	Más de 10 Años	No aplicable	No aplicable	Más de 10 Años	Más de 10 Años	Más de 10 Años	Más de 10 Años
CANAL	Directorios	Directorios	Directorios	Directorios	WEB	WEB	WEB	WEB	WEB
CLIENTEPORCAMPÑAEMAIL	no	no	no	no	no	no	no	no	no
CONSUMOTOTAL mean	1.42	1.62	3.83	7.05	37.90	288.51	32.62	593.47	5,493.93
DEPARTAMENTO	No aplicable	BOGOTA	BOGOTA	No aplicable	No aplicable	BOGOTA	BOGOTA	BOGOTA	BOGOTA
DIASCLIENTE mean	94.57	114.97	132.17	98.76	90.49	70.74	79.87	83.43	48.51
EMPRESASUNICAS_CONSULT mean	1.15	1.28	2.13	3.26	26.70	245.70	14.34	347.26	2,370.16
ESTADO	VIVA	ACTIVA	ACTIVA	VIVA	VIVA	ACTIVA	ACTIVA	ACTIVA	ACTIVA
FORMAJURIDICA	PERSONA FISICA	SOCIEDAD	SOCIEDAD	PERSONA FISICA	PERSONA FISICA	SOCIEDAD	SOCIEDAD	SOCIEDAD	SOCIEDAD
IMPORTE_COMPRAS mean	23.70	38.39	70.92	118.07	148.52	344.07	373.78	1,204.59	3,722.84
NUM_COMPRAS mean	1.00	1.00	2.59	3.37	1.26	1.00	6.19	3.46	19.27
SECTOR	NOSECTOR	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACI...	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACI...	NOSECTOR	NOSECTOR	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACI...	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACI...	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACI...	NOSECTOR
TAMAÑO	No aplicable	MICRO	MICRO	No aplicable	No aplicable	MICRO	PEQUEÑA	MEDIANA	GRANDE

Desplazamientos



Desplazamientos



Downgrading



Up-selling



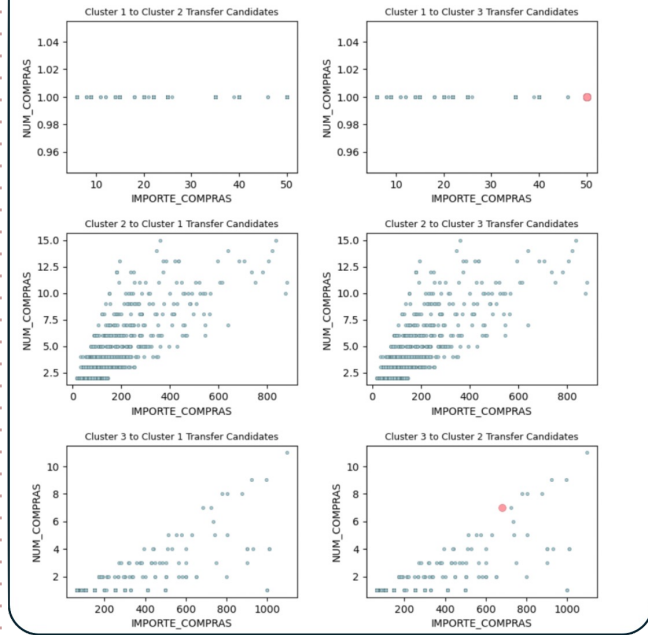
Cross-selling



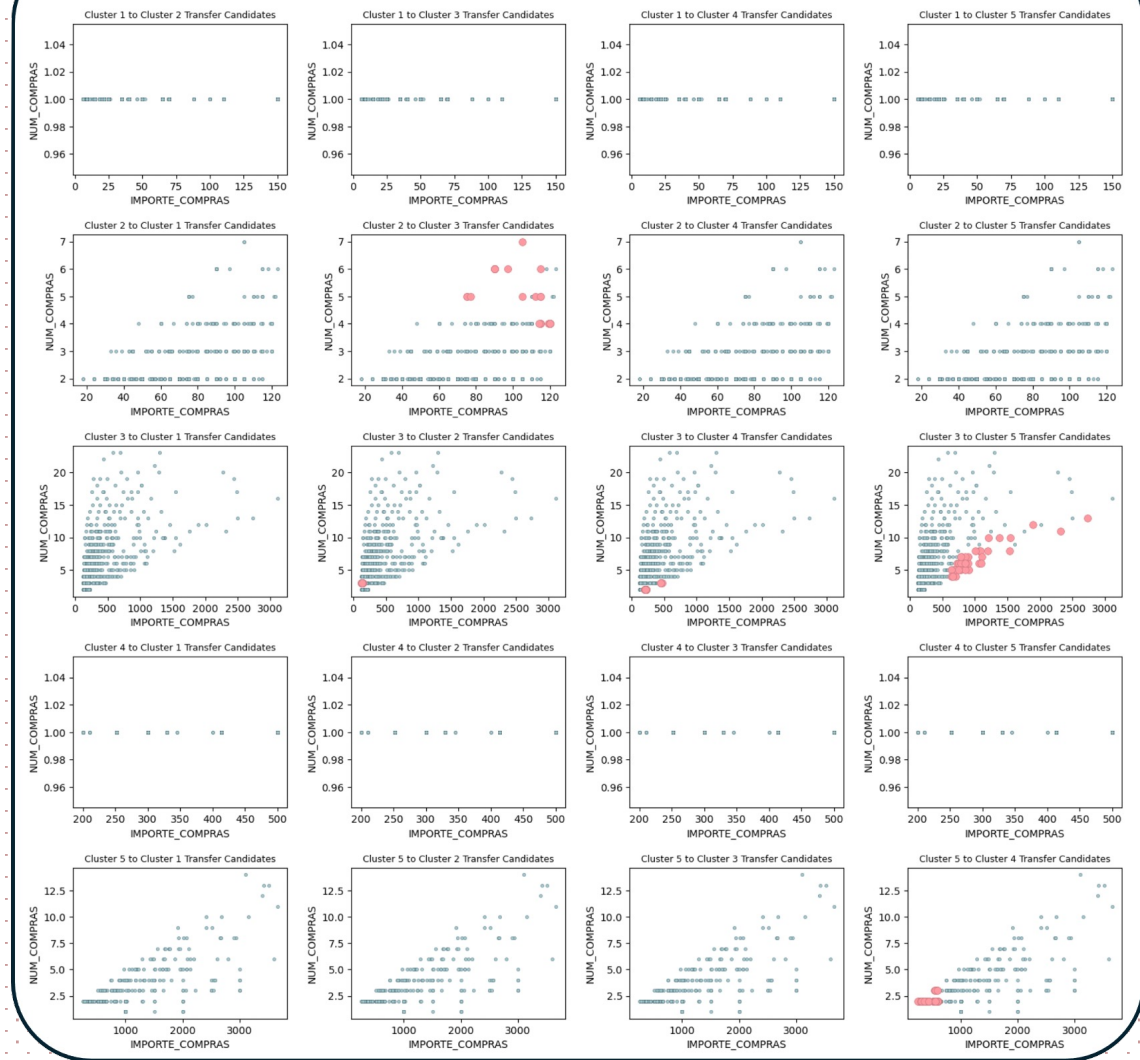


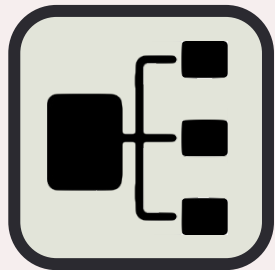
PREDICCIÓN DE TRANSFERENCIA POR DISTANCIA Y SIMILITUD A OTROS SEGMENTOS

Distance & Similarity-Based Cluster Transfer Candidates for Physical Persons



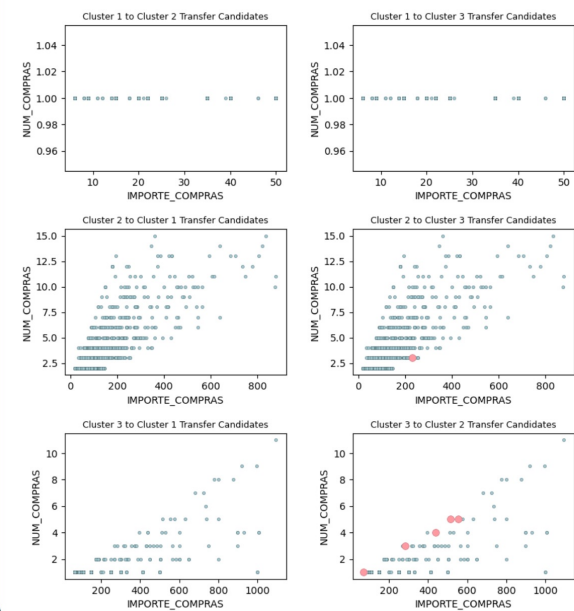
Distance & Similarity-Based Cluster Transfer Candidates for Companies and Entrepreneurs



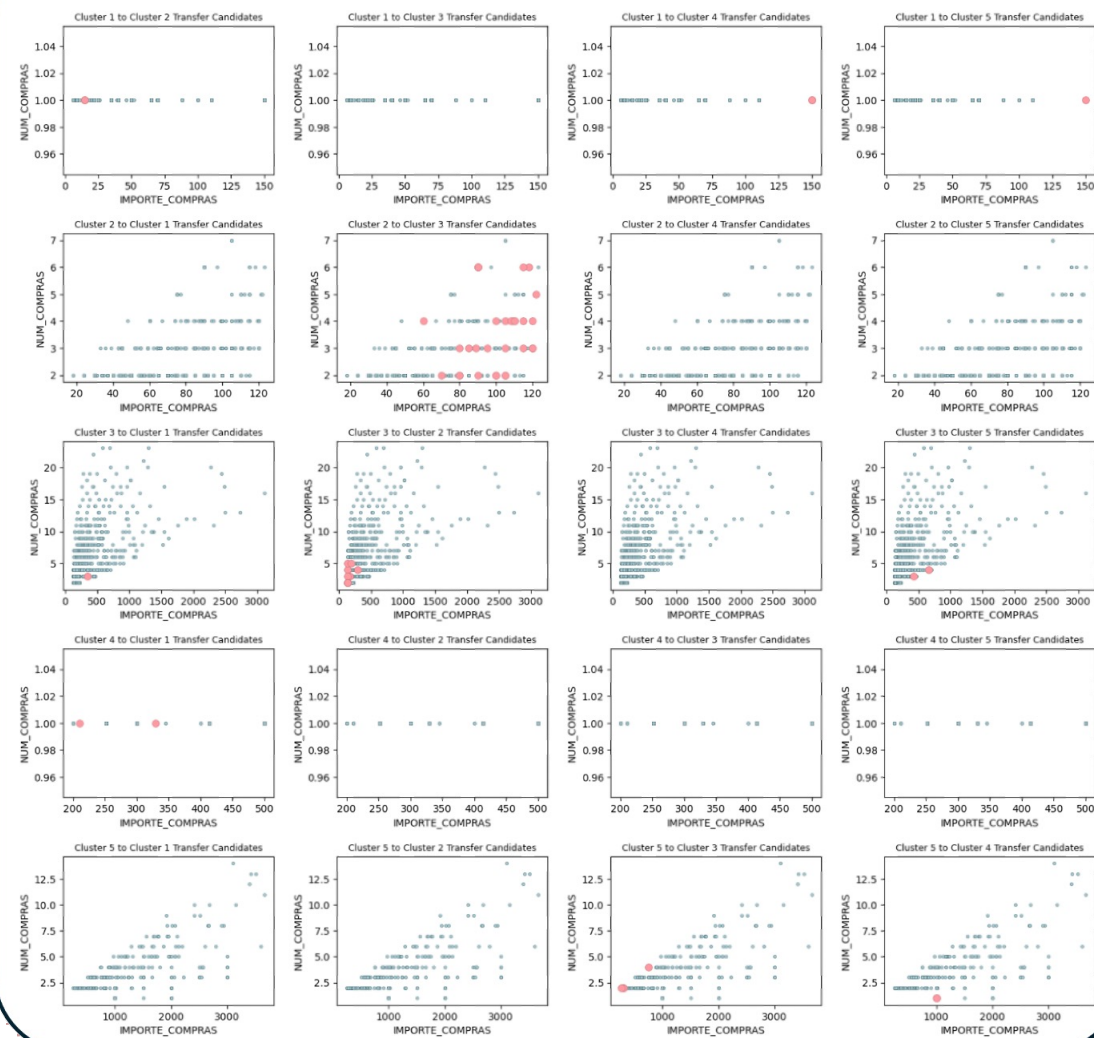


PREDICCIÓN POR CLASIFICACIÓN MULTINOMIAL MEDIANTE RED NEURONAL

Multinomial Prediction-Based Cluster Transfer Candidates for Physical Persons

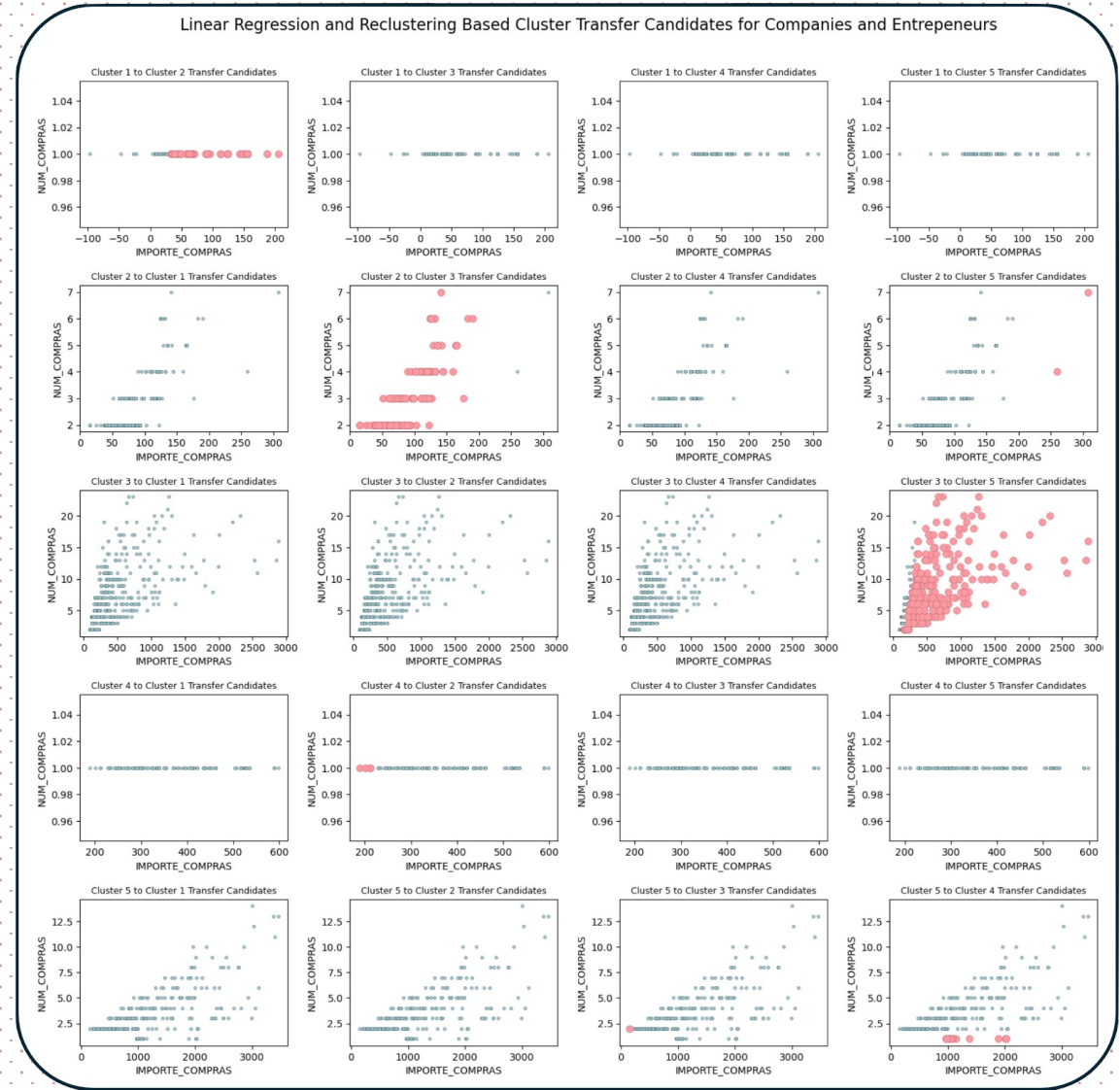
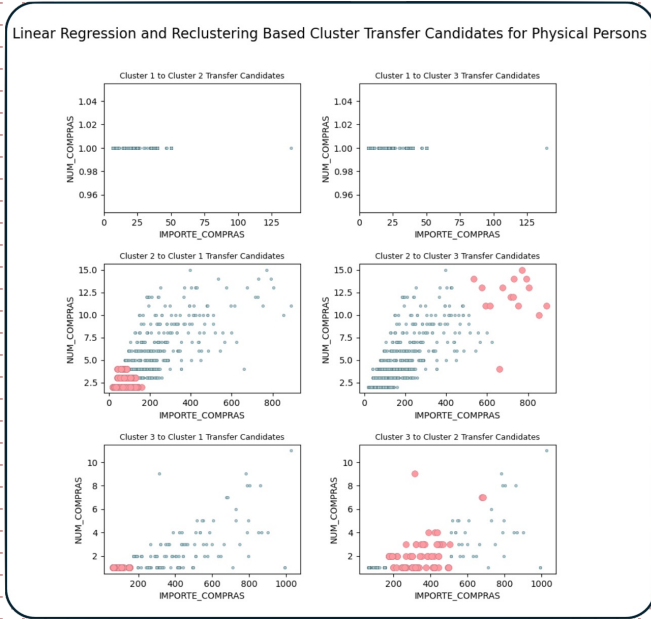


Multinomial Prediction-Based Cluster Transfer Candidates for Companies and Entrepreneurs





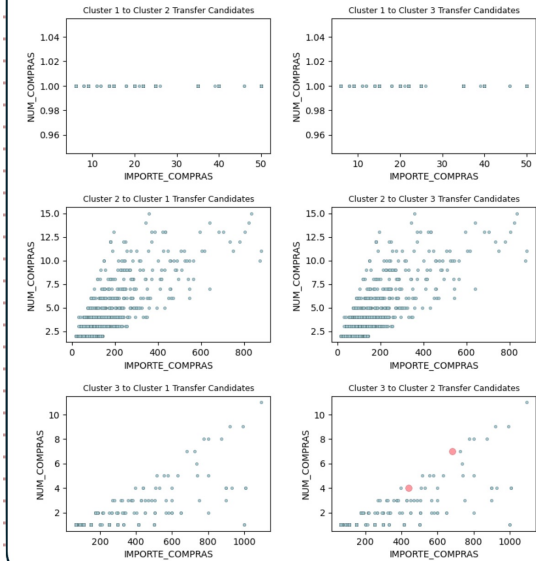
PREDICCIÓN POR REGRESIÓN DEL IMPORTE DE COMPRA Y REAGRUPAMIENTO



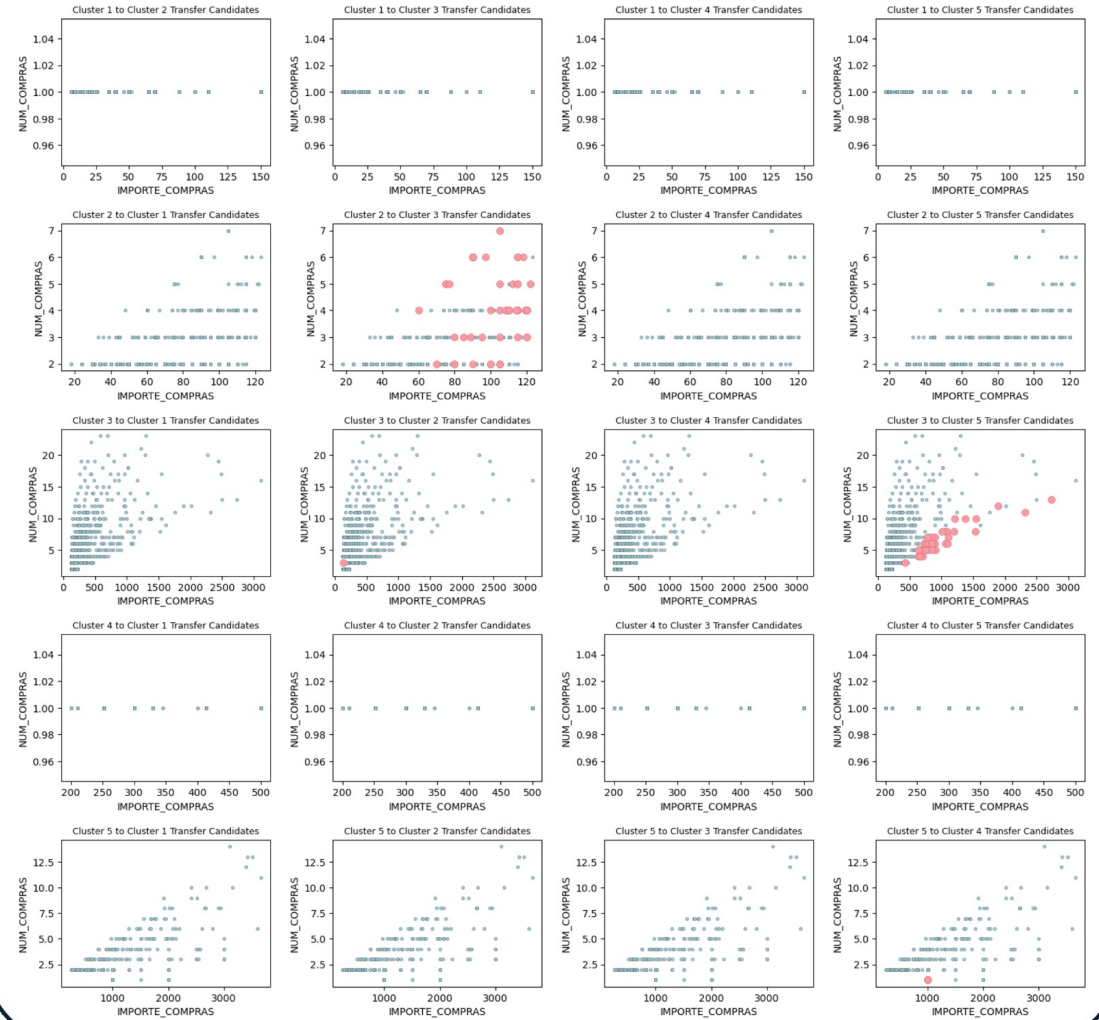


PREDICCIÓN POR CONSENSO

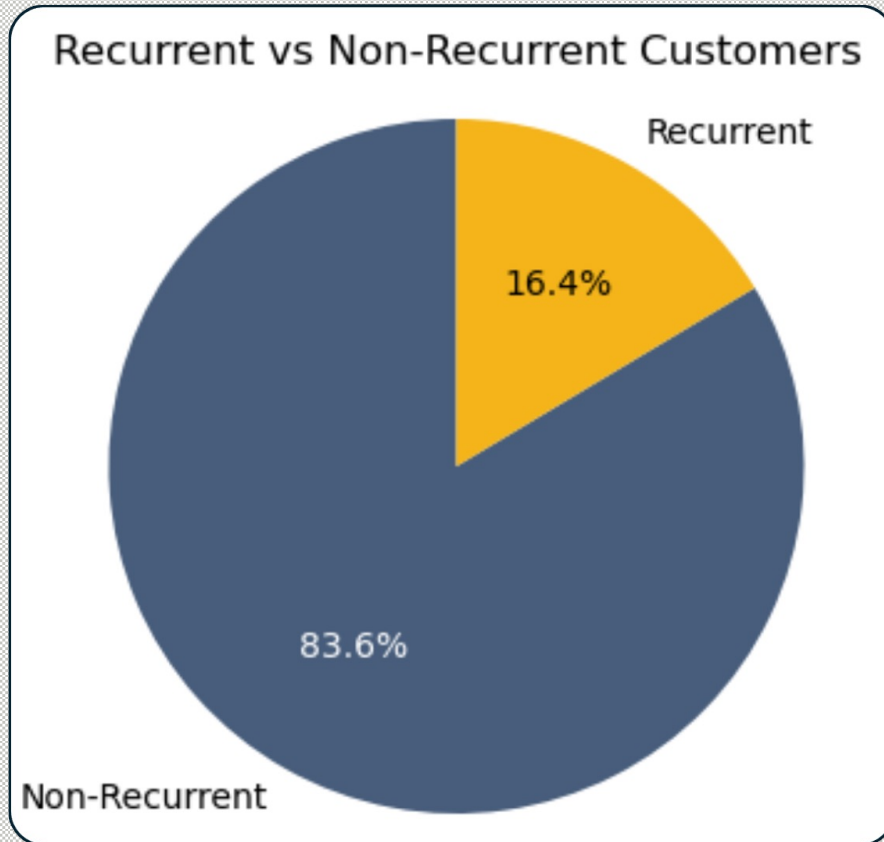
Consensus Multimetric-Based Cluster Transfer Candidates for Physical Persons



Consensus Multimetric-Based Cluster Transfer Candidates for Companies and Entrepreneurs



ETIQUETADO



No recurrente

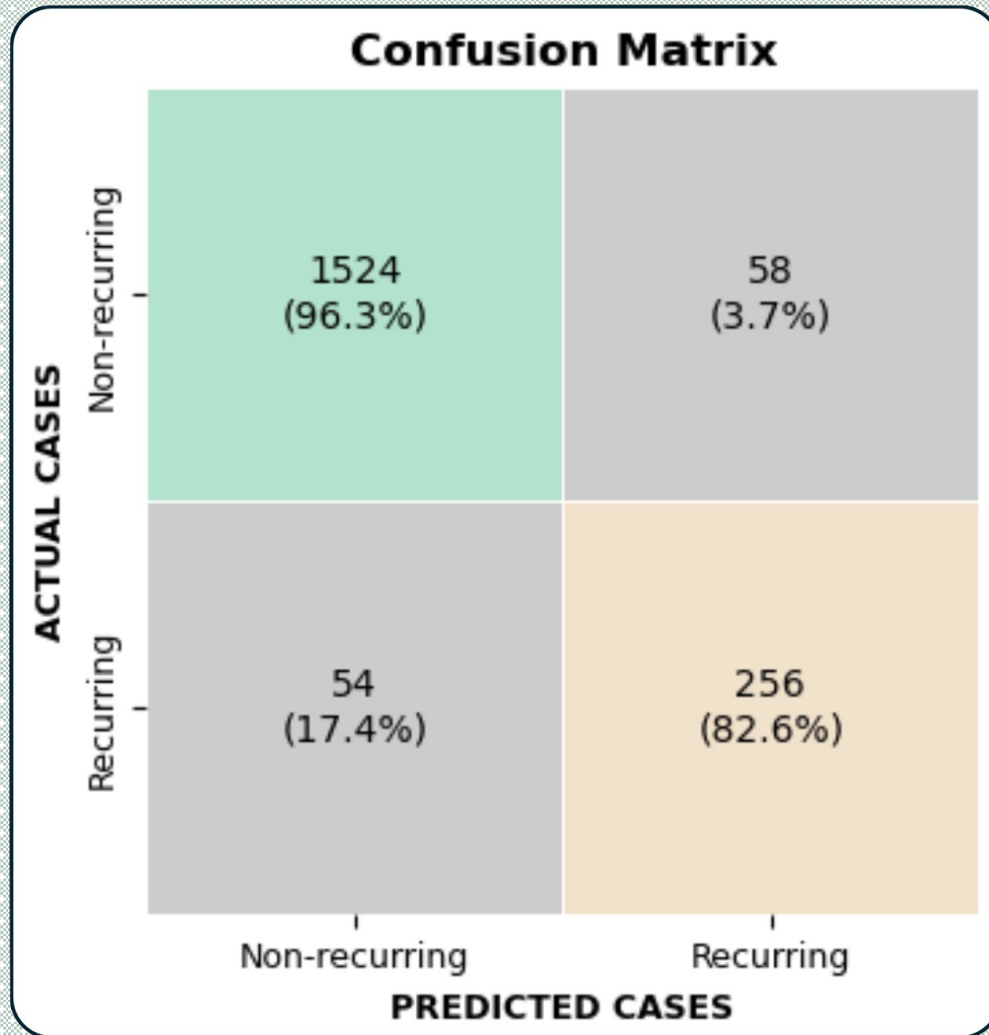
≤ 2 productos
o
compra única



Recurrente

resto

PREDICCIÓN



Classification Report:

	precision	recall	f1-score	support
0	0.97	0.96	0.96	1582
1	0.82	0.83	0.82	310
accuracy			0.94	1892
macro avg	0.89	0.89	0.89	1892
weighted avg	0.94	0.94	0.94	1892

Precisión **94 %**

Detección

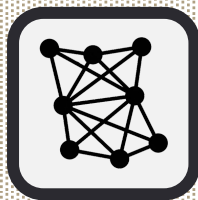


96,3 %



82,6 %

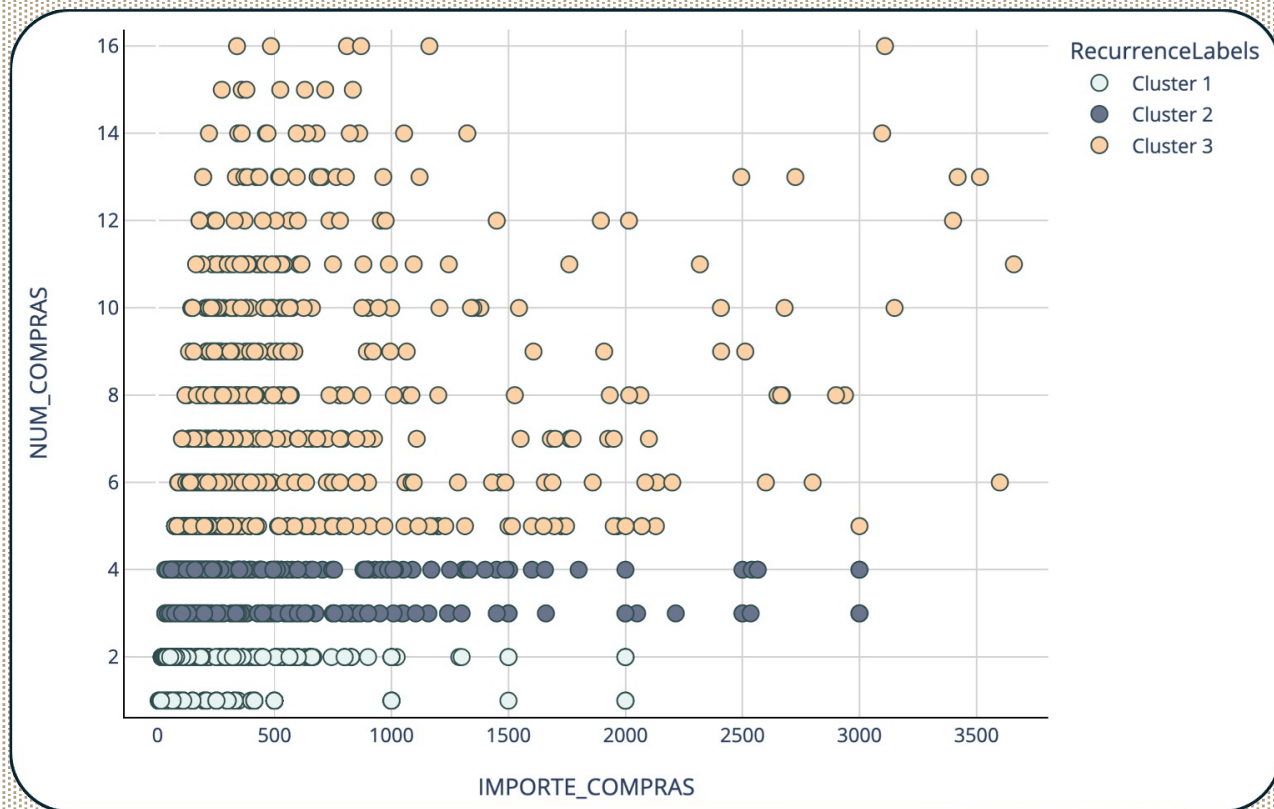
SEGMENTACIÓN



K-medoids



Manhattan



No recurrentes
 ≤ 2 compras

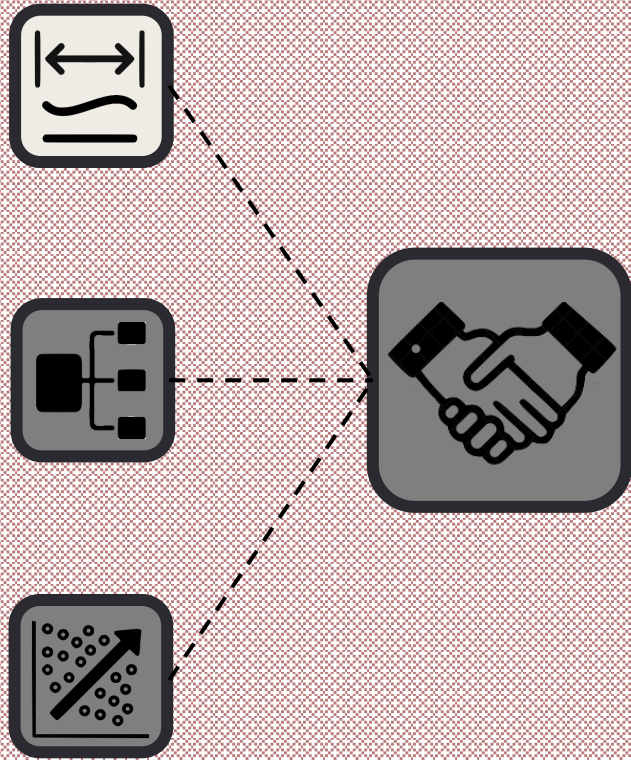


Recurrencia media
3-4 compras

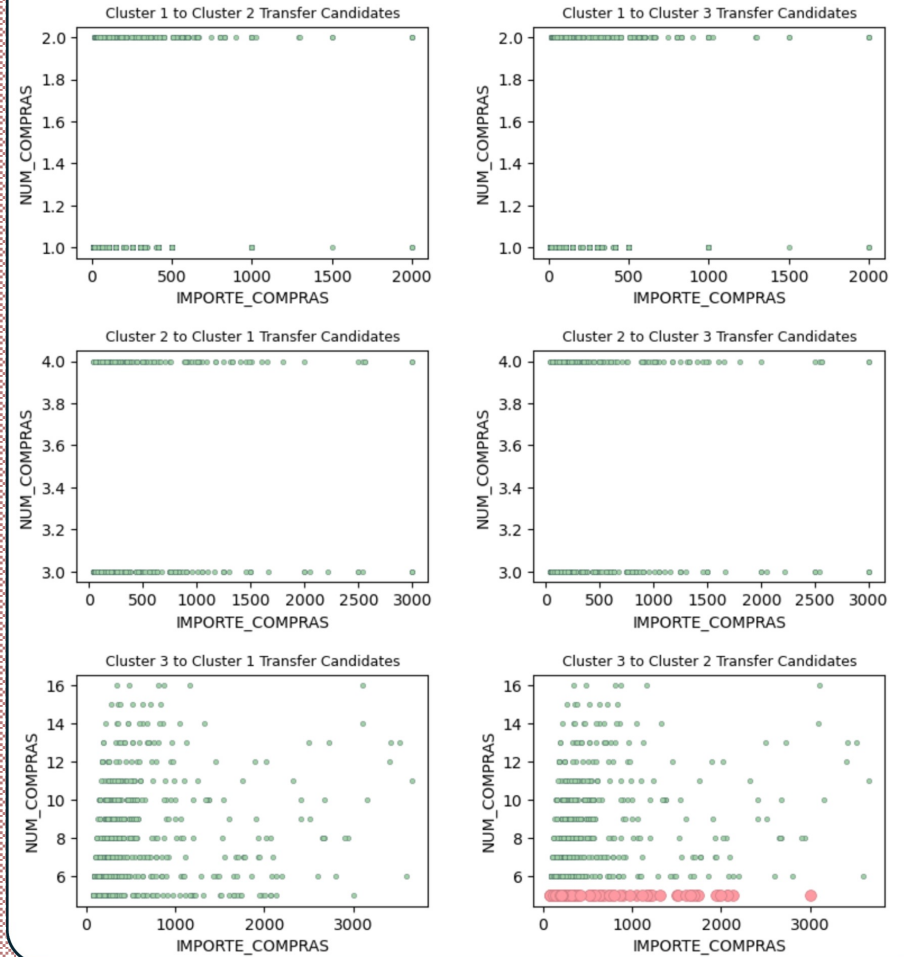


Recurrencia alta
 > 4 compras

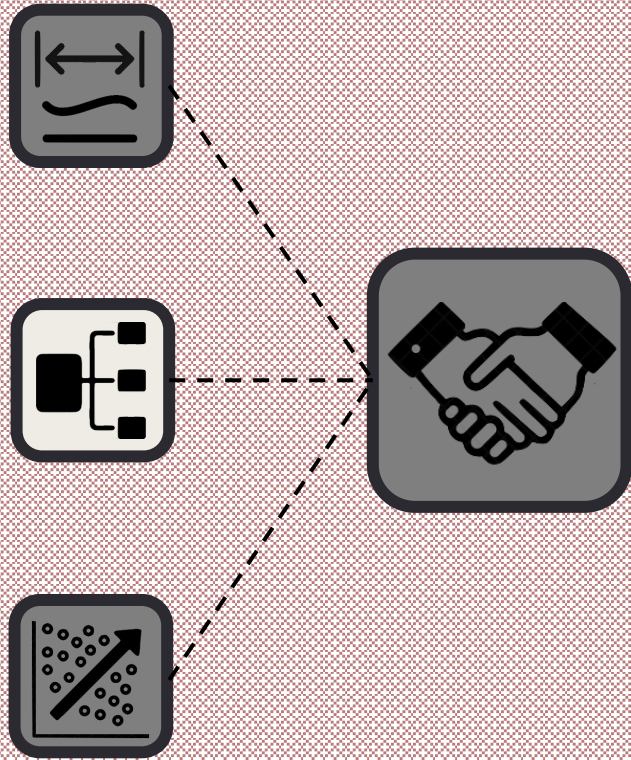
TRANSFERENCIA



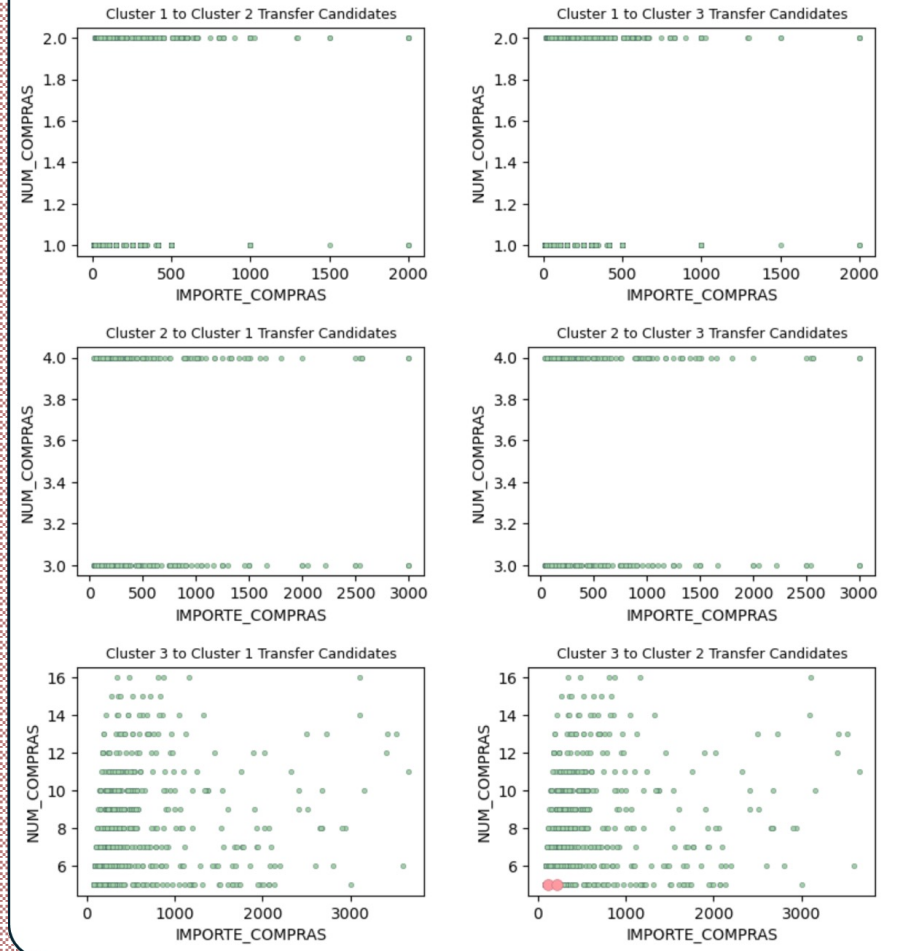
Distance & Similarity-Based Cluster Transfer Candidates for Recurrence



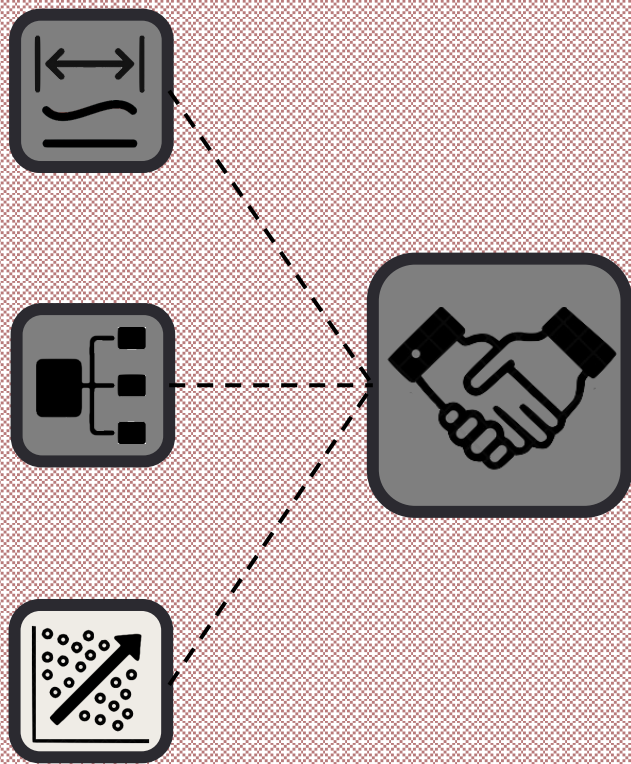
TRANSFERENCIA



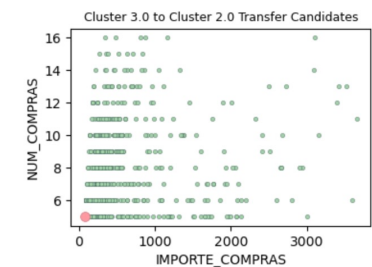
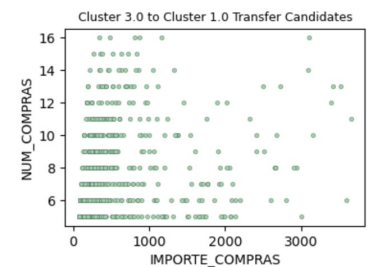
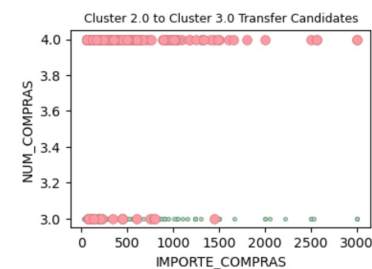
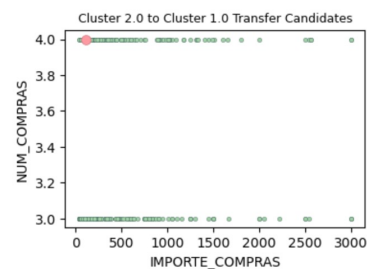
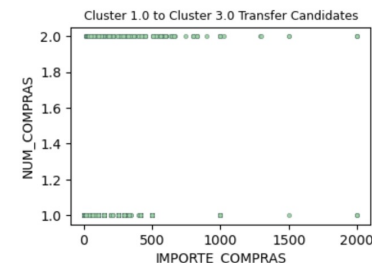
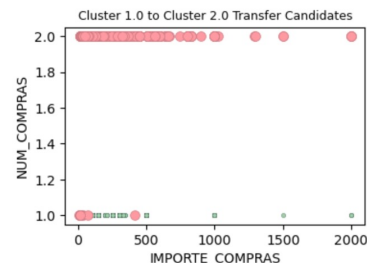
Multinomial Prediction-Based Cluster Transfer Candidates for Recurrence



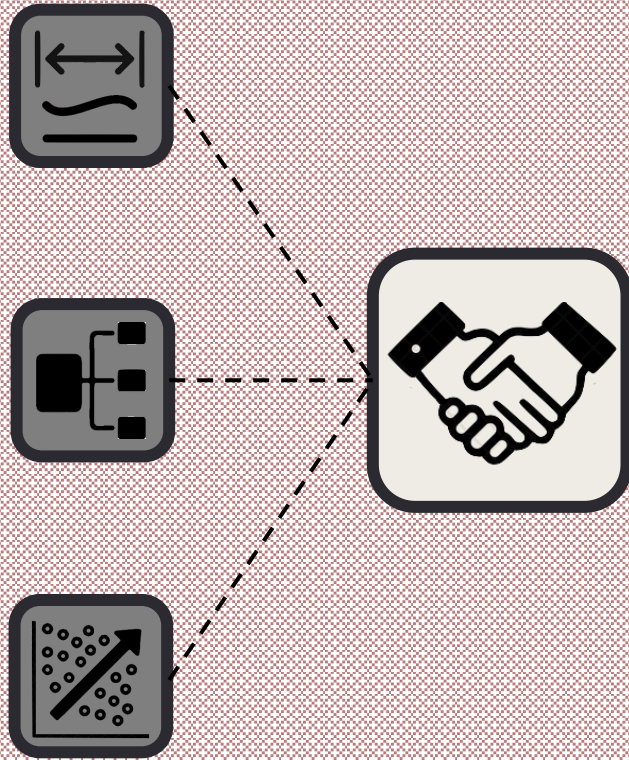
TRANSFERENCIA



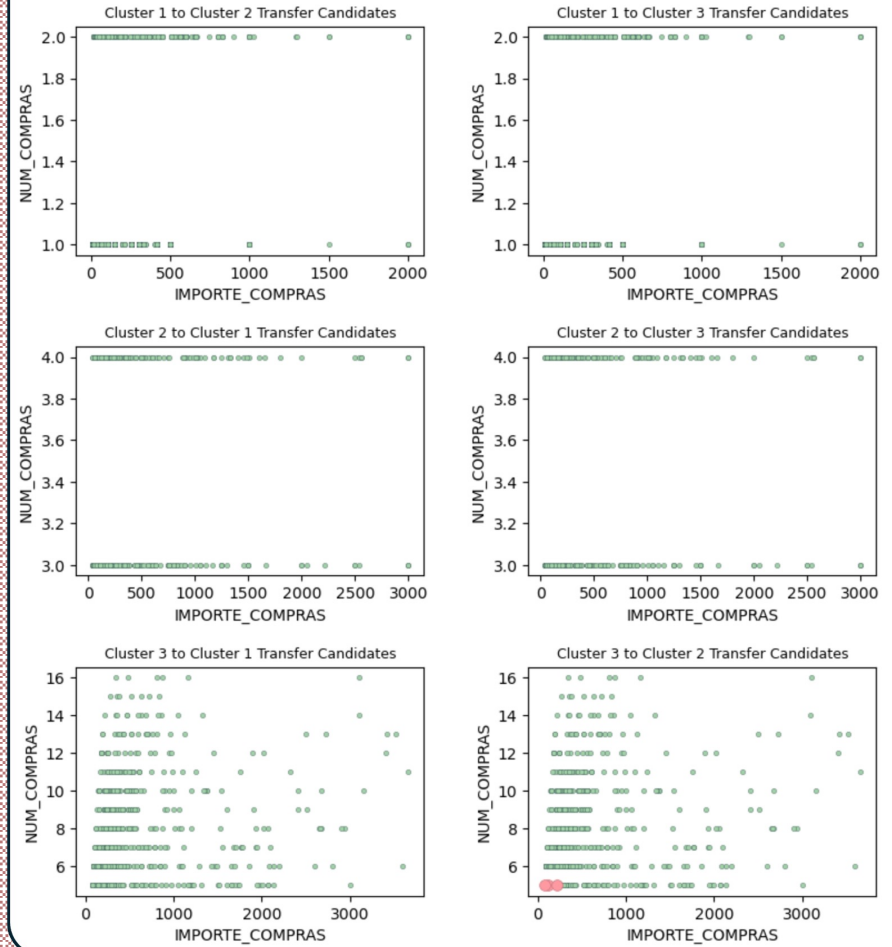
Linear Regression and Reclustering Based Cluster Transfer Candidates for Recurrence

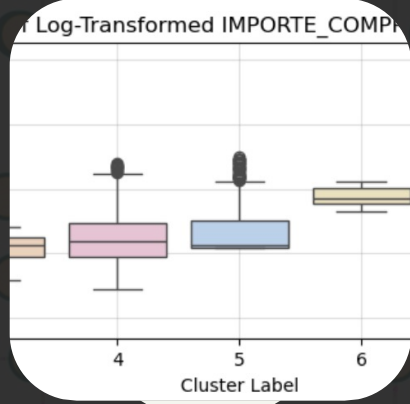


TRANSFERENCIA



Consensus Multimetric-Based Cluster Transfer Candidates for Recurrence





Segmentación

Agrupamiento en dos pasos según el perfil y hábitos de compra del cliente

0	0.98	0.91	0.95	31
1	0.68	0.92	0.78	31
accuracy			0.91	1892
macro avg	0.83	0.92	0.86	1892
weighted avg	0.93	0.91	0.92	1892

Confusion Matrix

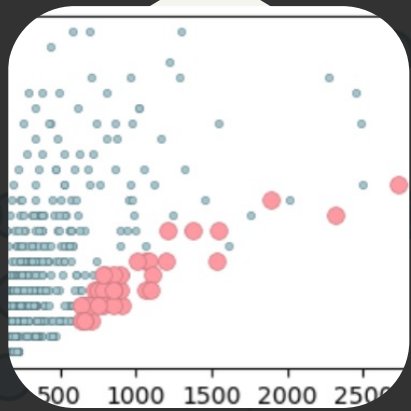
	Actual Non-recurring	Actual Recurring
Predicted Non-recurring	1446 (91.4%)	136 (8.6%)
Predicted Recurring	26 (8.4%)	284 (91.6%)

Fidelización

Modelo de predicción de recurrencia y herramientas de optimización de la fidelización

Transferencia

Metodología múltiple de predicción del cambio de segmento



Impacto

Exportación de datos y herramienta reutilizable





OBJETIVOS DE DESARROLLO SOSTENIBLE



8 TRABAJO DECENTE
Y CRECIMIENTO
ECONÓMICO



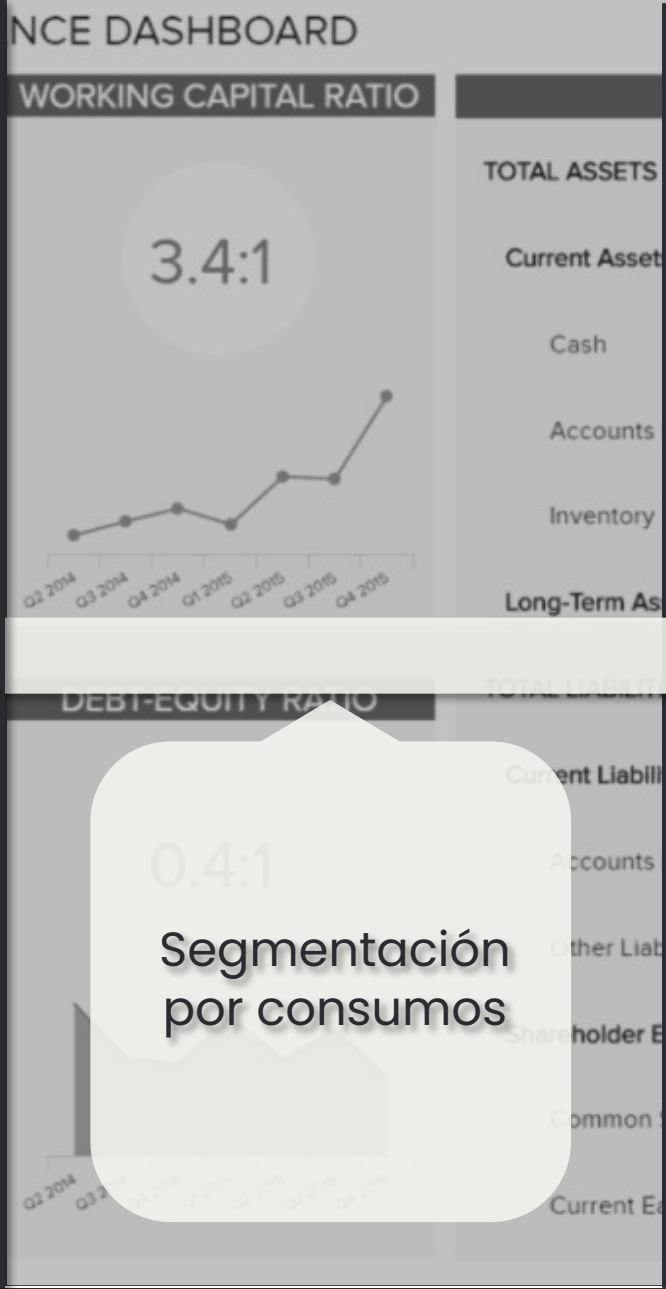
12 PRODUCCIÓN
Y CONSUMO
RESPONSABLES



13 ACCIÓN
POR EL CLIMA



Líneas futuras

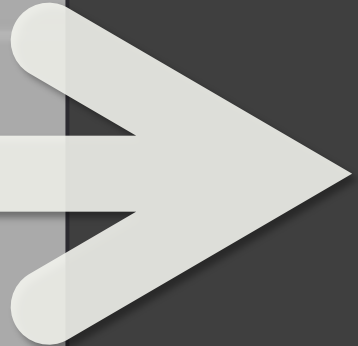


Análisis de personas físicas

Segmentación por consumos



Análisis de outliers



**MÁSTER UNIVERSITARIO
EN CIENCIA DE DATOS**



TRABAJO FINAL DE MÁSTER
JUNIO 2024