

Análisis de datos en tiendas de comercio electrónico para segmentación de clientes

Por:

Diego Alejandro Cifuentes Garcia
Juan David Arcila Moreno
Sebastian Patiño Barriento

Justificación 01

Objetivos 02

Metodología 03

Contenidos

04 Desarrollo y técnicas

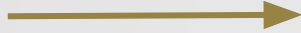
05 Resultados

06 Conclusiones

Justificación



**¿Qué clientes tengo?
¿Qué características
comparten?**



¿Cómo los fidelizo?



**Diseño e implemento
campañas de
Marketing**

Objetivos



Objetivo principal

Segmentar los clientes del retail partiendo de su comportamiento de adquisición de artículos con el fin de facilitar la implementación de estrategias de fidelización.



Dataset #1

Segmentar a partir del gasto que ha hecho el cliente en las compras.



Dataset #2

Segmentar a partir del comportamiento y características de los artículos registrados en las compras

Metodología: CRISP-DM



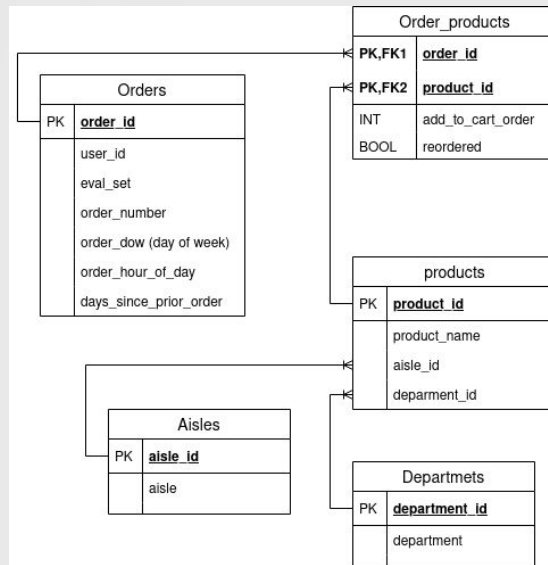
1. Entendimiento

- Tienda electrónica localizada en **Reino Unido**.
- Se enfoca en la venta al **detal** de **regalos para toda ocasión**.
- Se presentan las ventas realizadas entre el 1/12/2009 y el 09/12/2011
- Dataset mantenido por *UCI Machine Learning Repository*.

Los campos incluyen, entre otros:

- Descripción del producto.
- **Precio Unitario** del producto.
- Cantidad adquirida.
- Fecha de la transacción.
- ID del cliente.
- País de origen de la transacción.

540K Registros



~3.5M Registros

2.1. Preparación: E-Commerce

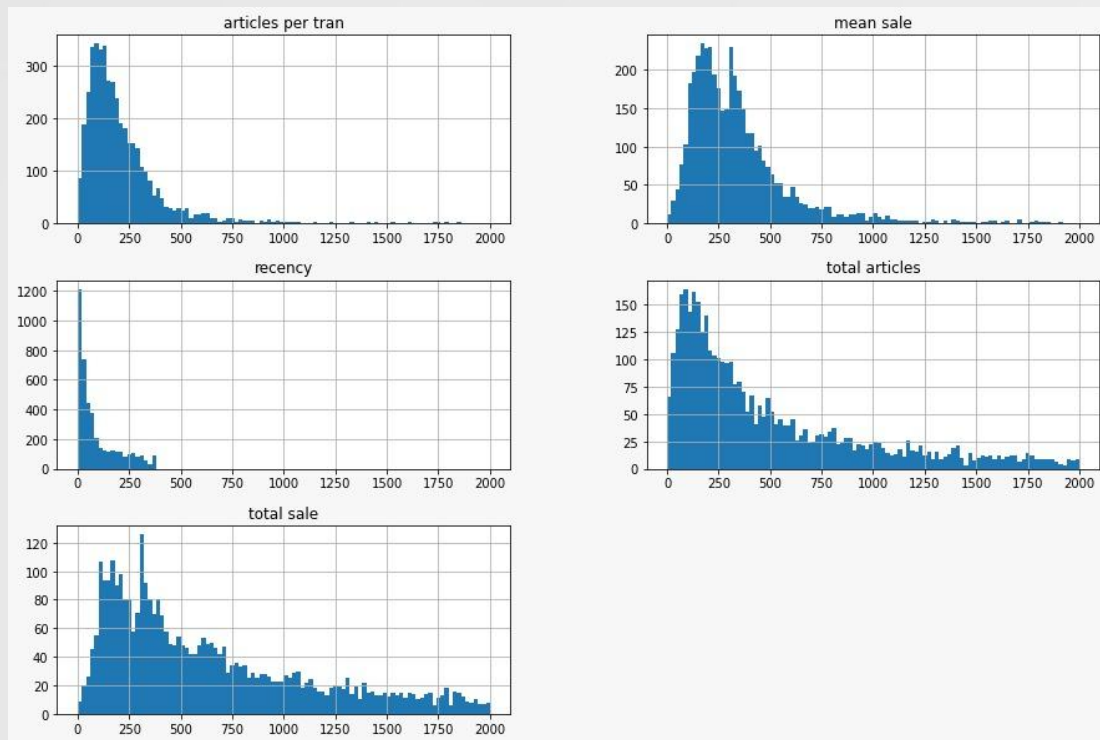
Se generaron variables adicionales para ayudar a la clasificación:

- Artículos totales por cada factura.
- Media de los precios en las transacciones del cliente.
- Número de días desde la última compra.
- Total de artículos adquiridos.
- Costo total de los artículos que ha adquirido.
- Número de transacciones hecha por el cliente.
- Cantidad de artículos únicos comprados por transacción.

Eliminación de Outliers

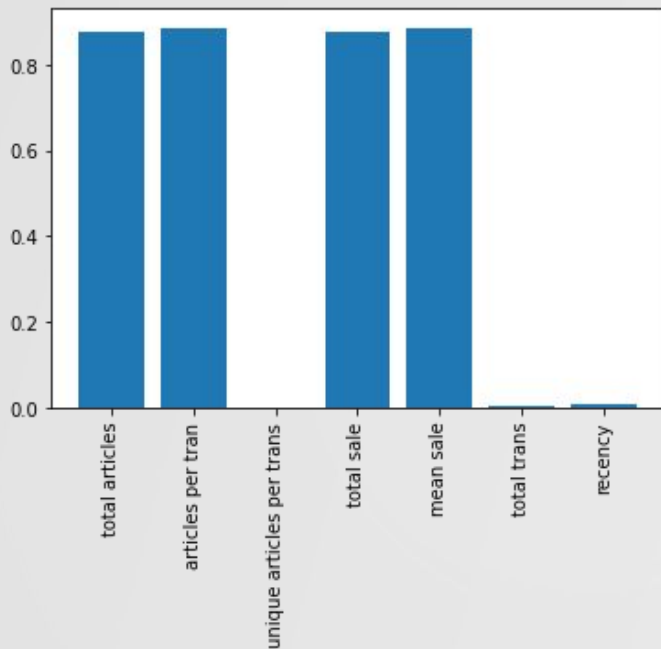
- Clientes con una sola transacción
- Cálculo de Z-score (variables alejadas 3 desviaciones estándar de su media)

Histogramas de variables luego de eliminación de Outliers

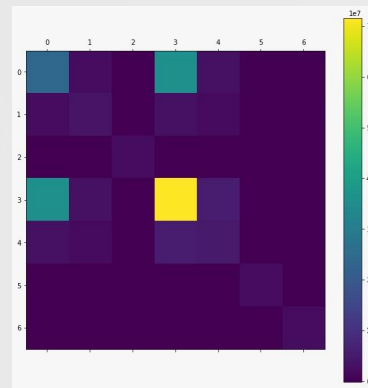


2.1. Preparación: E-Commerce

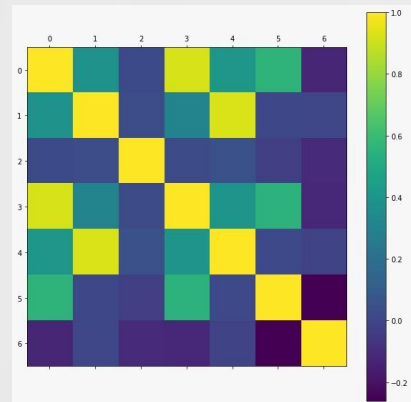
Análisis de dependencias (R^2)



Covarianza



Correlaciones



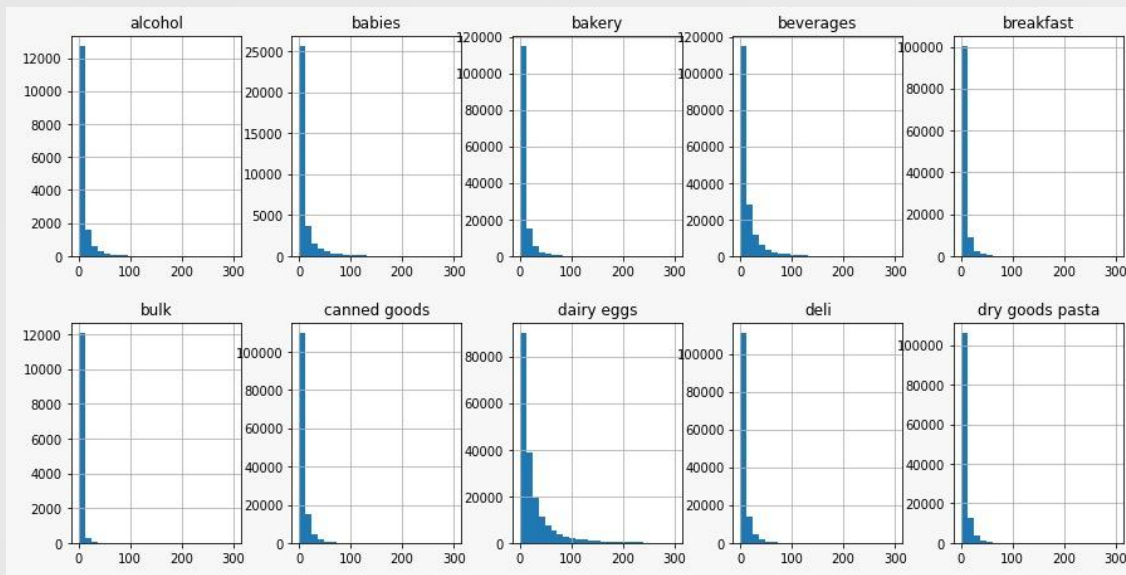
2.2. Preparación: Instacart

Se realizó un proceso *ETL* para obtener los campos de interés:

- Artículos facturados por departamento
- Demás columnas desnormalizadas

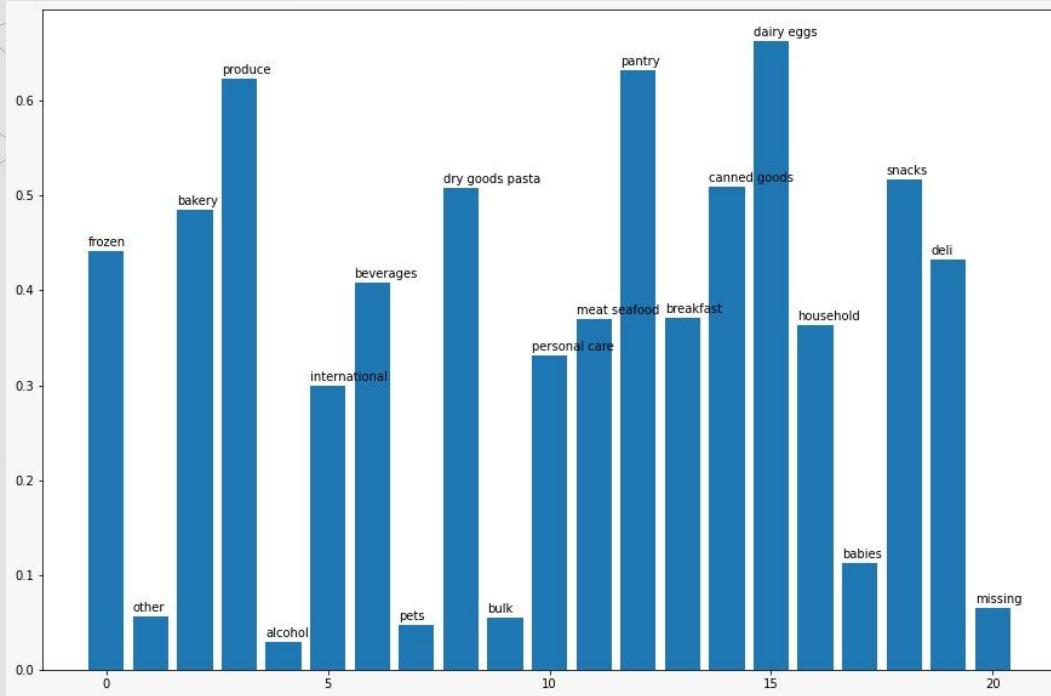


Histogramas de valores por cada departamento



2.2. Preparación: Instacart

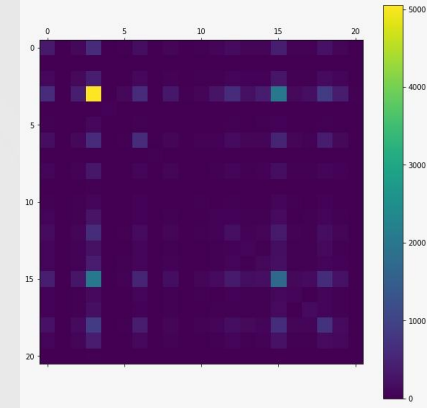
Análisis de dependencias entre departamentos (R^2)



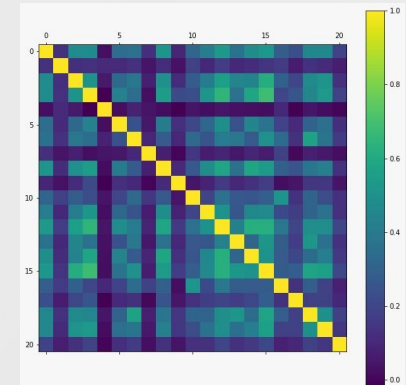
- Otros Analisis:

- $VG=3.64 \times 10^{35}$, Coef de dependencia=31%

Covarianza departamentos

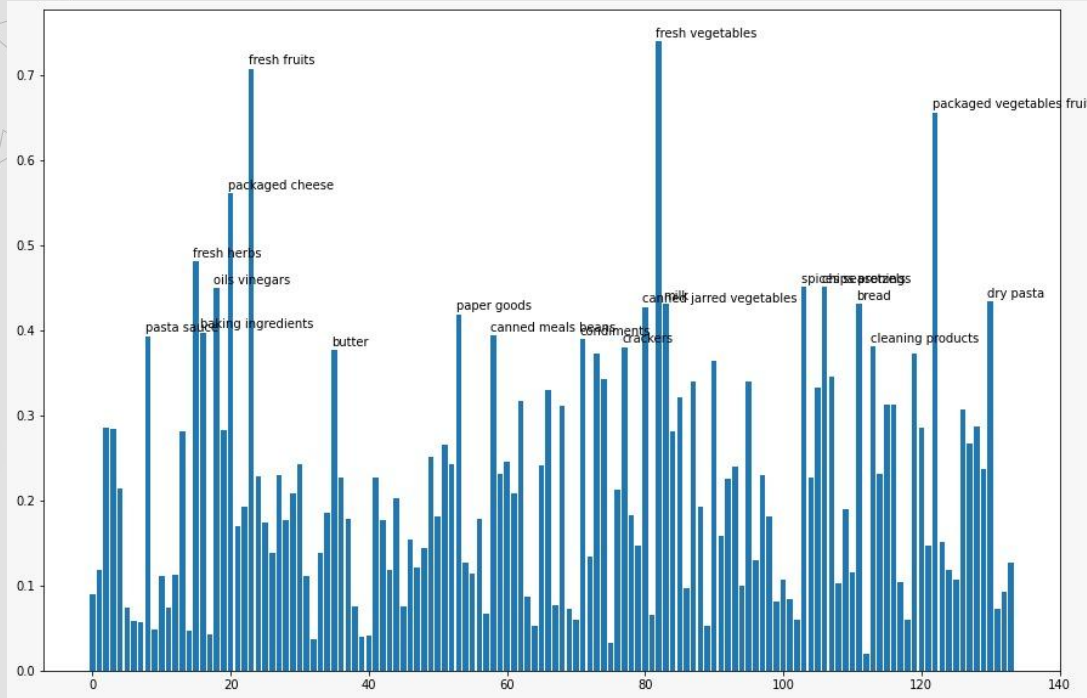


Correlaciones departamentos



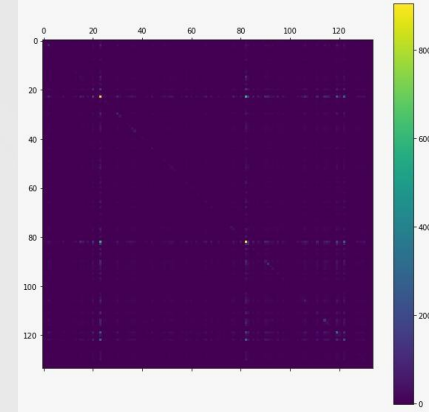
2.2. Preparación: Instacart

Análisis de dependencias entre pasillos (R²)

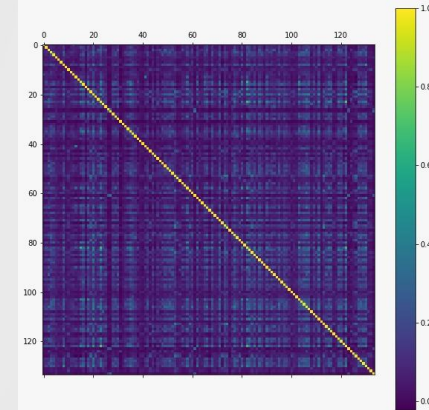


- Otros Analisis:
 - $VG = 3.73 \times 10^{73}$, Coef de dependencia = 18%

Covarianza pasillos



Correlaciones pasillos



3. Modelado



K-MEANS

Segmentar los clientes. Clientes que comparten características pueden pertenecer a una misma campaña de marketing.

Usando métrica **Euclidiana** y de **Mahalanobis**

Análisis de Siluetas



Modelo de regresión lineal en Instacart

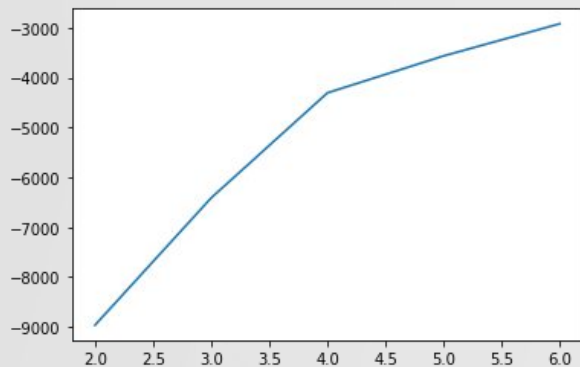
Predice compra de vegetales



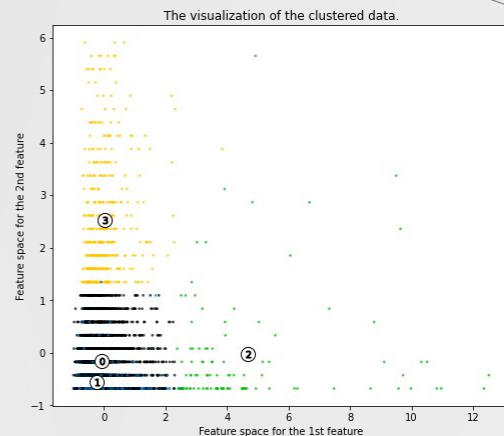
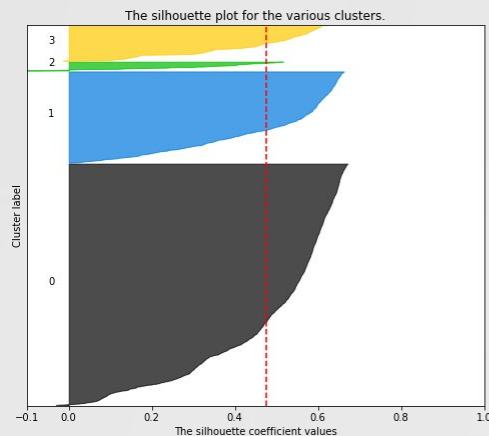
3.1. Modelado de E-Commerce

K	Puntuación silueta promedio
2	0.37
3	0.44
4	0.47
5	0.46
6	0.43

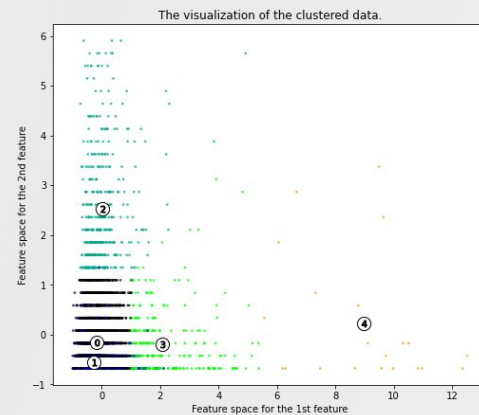
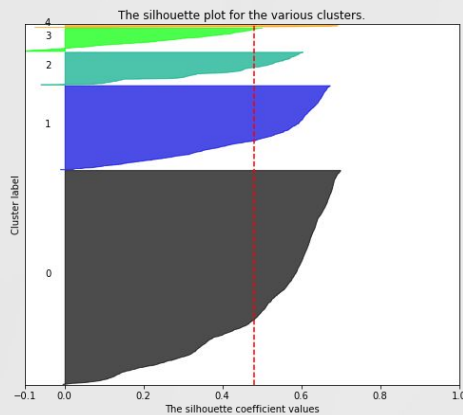
Suma de todas las distancias para cada k



Silhouette analysis for KMeans clustering on sample data with $n_clusters = 4$



Silhouette analysis for KMeans clustering on sample data with $n_clusters = 5$

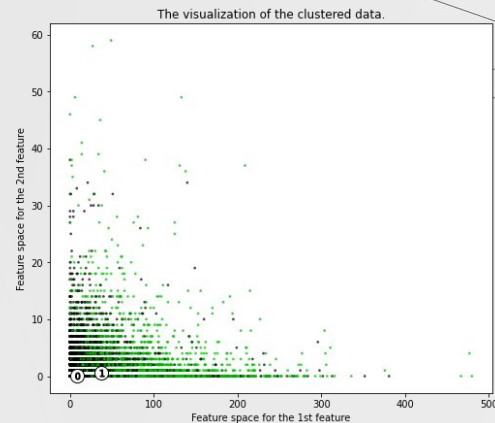
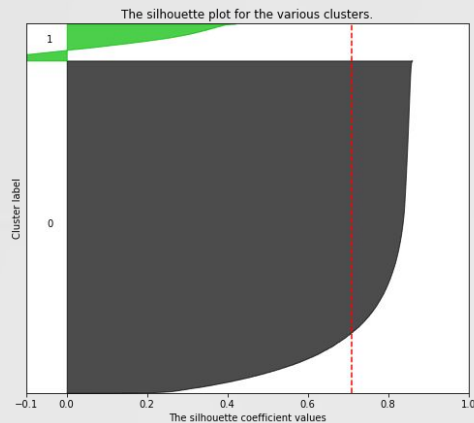




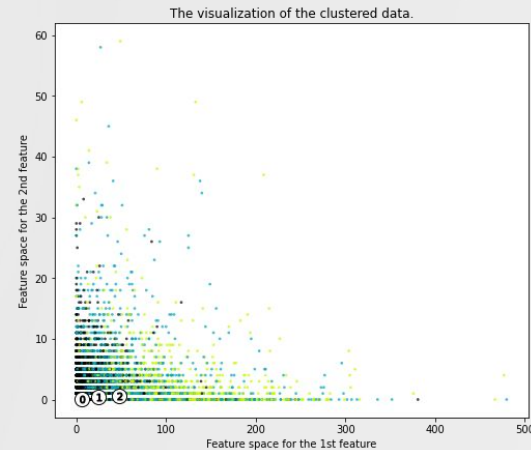
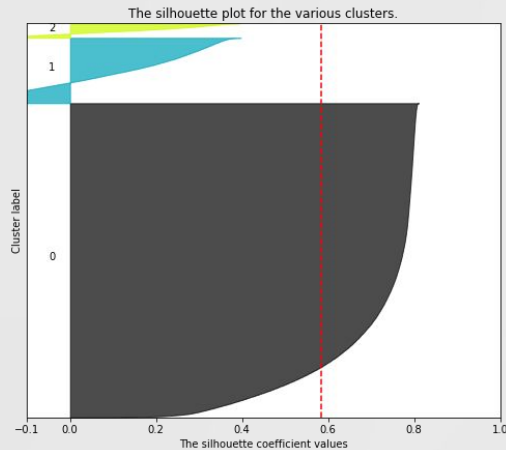
3.2. Modelado de Instacart

K	Puntuación silueta promedio
2	0.71
3	0.58
4	0.48
5	0.47
6	0.41

Silhouette analysis for KMeans clustering on sample data with $n_clusters = 2$



Silhouette analysis for KMeans clustering on sample data with $n_clusters = 3$





3.2. Modelado de Instacart

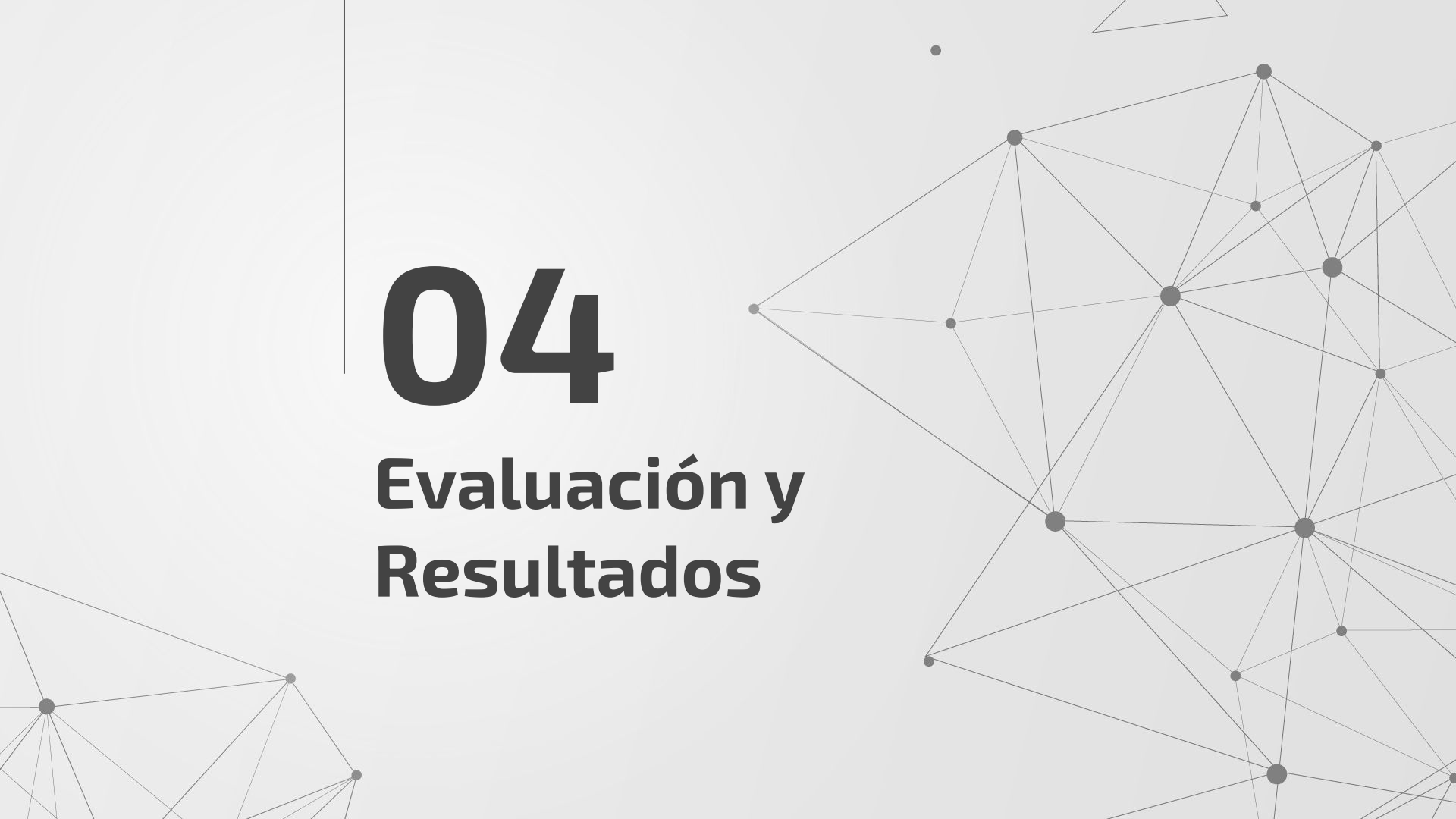
Modelo de regresión lineal en Instacart

Modelo	Train (R^2)	Test (R^2)	# Parametros
Lineal	0.740932	0.731464	133
Ridge	0.740932	0.731464	133
Lasso	0.733220	0.723949	54
Elastic	0.736440	0.727093	71

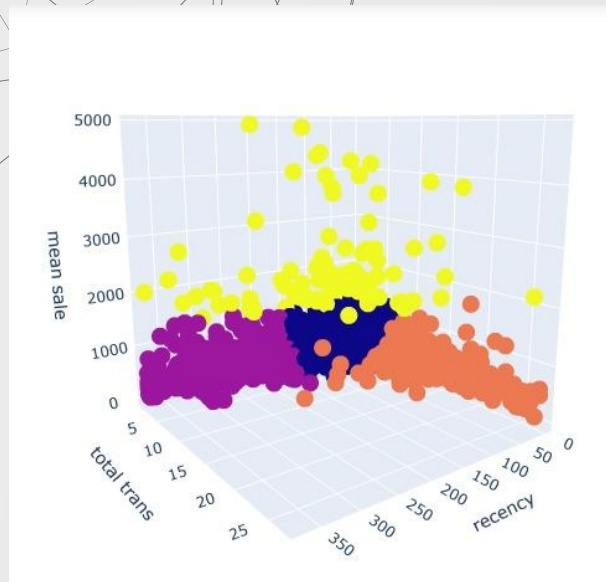
Artículos comprados por el cliente en cada pasillo para predecir el número de vegetales que compra.

04

Evaluación y Resultados



4.1. Evaluación y Resultados E-Commerce



- **Morado:** *Clientes que no volvieron a comprar.* (Recencia y transacciones altas).
- **Anaranjados:** *Clientes fieles a la compañía.* (Recencia baja, transacciones altas).



- **Azules:** *Clientes nuevos en la compañía.* (Recencia baja y nivel de transacciones media-baja).
- **Amarillos:** *Generan ganancias a la compañía* (Tienen una venta promedio alta, sin importar su recencia).

4.2. Evaluación y Resultados Instacart

```
{'fresh herbs': 1.5761808631578627,
'grains rice dried goods': 0.9345221362255276,
'oils vinegars': 0.7208791380913167,
'canned jarred vegetables': 0.6785465498334853,
'meat counter': 0.6760754741004801,
'poultry counter': 0.6026518813374828,
'seafood counter': 0.54458384783037,
'spices seasonings': 0.47635715244316745,
'tofu meat alternatives': 0.42310927110006086,
'canned meals beans': 0.3553327304667675,
'packaged vegetables fruits': 0.3366186098861134,
'fresh fruits': 0.3012762893706071,
'butter': 0.29944498924734664,
'eggs': 0.281273824768218,
'asian foods': 0.2773922997589251,
'soup broth bouillon': 0.22929119556610264,
'dry pasta': 0.195121036495169,
'pickled goods olives': 0.16879789901691927,
'condiments': 0.16086051443218397,
'pasta sauce': 0.15748508990899088,
'canned meat seafood': 0.13133616408409407,
'bulk grains rice dried goods': 0.08915945681909532,
'tortillas flat bread': 0.0844460752927546,
'frozen produce': 0.07109769202879822,
'other creams cheeses': 0.07019498986539723,
'packaged meat': 0.05657110954382472,
'hot dogs bacon sausage': 0.0501504930157616,
'packaged poultry': 0.048638845739511515,
'baking ingredients': 0.03479658473624968,
'soy lactosefree': 0.027360728256831366,
'specialty cheeses': 0.011196023441348079,
..
}
```

```
'fruit vegetable snacks': -0.0971045387270356,
'breakfast bars pastries': -0.10109383858700476,
'granola': -0.10461798143266768,
'cereal': -0.10878544072949561,
'breakfast bakery': -0.11331304439533242,
'instant foods': -0.12698918457974467,
'doughs gelatins bake mixes': -0.1501667238718087,
'packaged produce': -0.15688240580517462,
'canned fruit applesauce': -0.18365311506446091}
```

- Resultados de Clustering **no** fueron concluyentes (haciendo uso los departamentos, pasillos y otras columnas).
- La regresión de “Fresh Vegetables” tiene dió **buenos** resultados.
- Existen relación con los pesos.

Conclusiones



Resultados de clustering

Son altamente interpretables y puede aplicarse a distintas campañas de marketing

Depende de la articulación con otros componentes de la arquitectura de negocio

Despliegue de la solución



Aplicables a otras organizaciones

Creemos que los análisis realizados pueden ser útiles para otra organizaciones de comercio electrónico.

The background of the slide features a complex, abstract geometric pattern. It consists of numerous thin, light gray lines that intersect at various points, creating a web-like or network structure. Some of these intersection points are marked with small, solid dark gray circles. The overall effect is a modern, technical, and interconnected aesthetic. The text is centered in the middle of the slide, overlaid on this pattern.

Muchas gracias por su atención

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.
