

Optimización de campañas promocionales a través de la ciencia de datos



CODER HOUSE

Lorenzo Guimaraes

Proyecto Final para Data Science II: Machine Learning para la ciencia de datos

Análisis e interpretación de caso, visualización y disposición de datos
para el desarrollo de un modelo predictivo



[Proyecto en Github](#)

CODER HOUSE



[Link al Notebook](#)

Data Science II - Comisión
60895

Lorenzo
Guimaraes

Etapas del proyecto

- **Identificación de objetivo y desarrollo de hipótesis**
- **Adquisición de los datos**
- **Limpieza y ajuste (Data Wrangling)**
- **Análisis exploratorio**
- **Tratamiento de los datos: Transformación de variables, feature engineering**
- **Balanceo y escalado**
- **Principal Component Análisis**
- **Aplicación de modelos:**
 - DecisionTree
 - Regresión logística
 - RandomForestClassifier
- **Validación y optimización de modelos**
- **Comparación y selección de modelo**
- **Conclusión**

La empresa

Una empresa del sector de gastronomía. con más de 100 mil clientes se especializa en la venta de alimentos gourmet, carnes, vinos, frutas exóticas y demás productos, a través de tres medios diferentes, via web, en la tienda física, o catálogos.

En este caso analizaremos una campaña piloto realizada con el fin de realizar un muestreo del alcance de su campañas promocionales a fin de maximizar el impacto de campañas futuras junto al lanzamiento de sus nuevos productos.



La campaña

La campaña tuvo 2240 clientes seleccionados al azar a los cuales se ofreció una oferta promocional de sus nuevos productos, los que fueron aceptando en los meses siguientes fueron categorizados como "Si" a la respuesta.

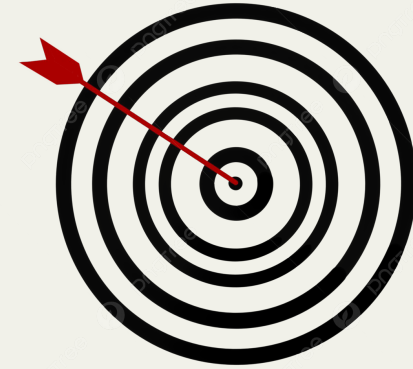


El costo total de la campaña superó casi por el doble a los ingresos generados con las ventas de la oferta promocional.

La empresa cree que los pronósticos para los próximos tres años no son muy prometedores... por lo que quieren hacer algo al respecto.

Descripción y objetivo

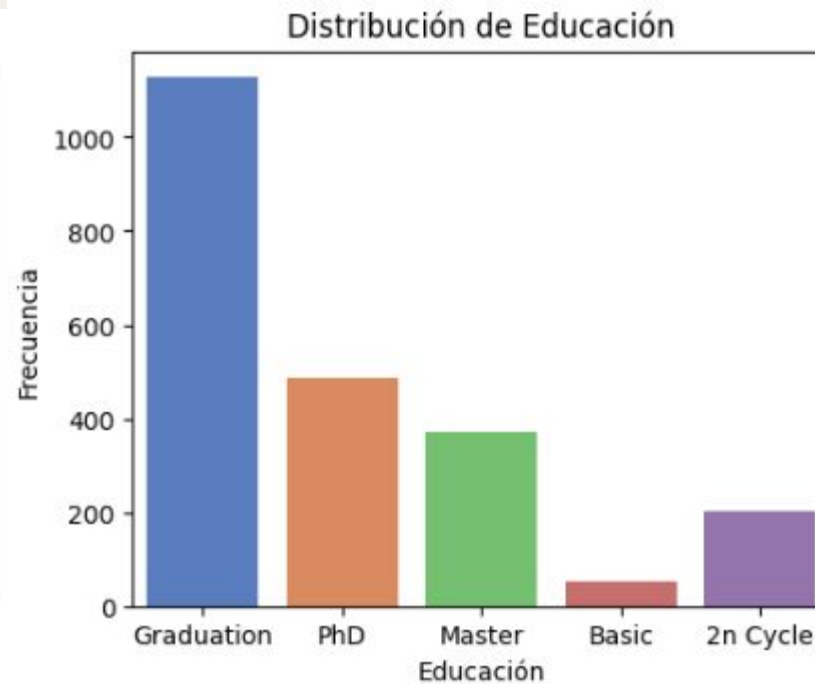
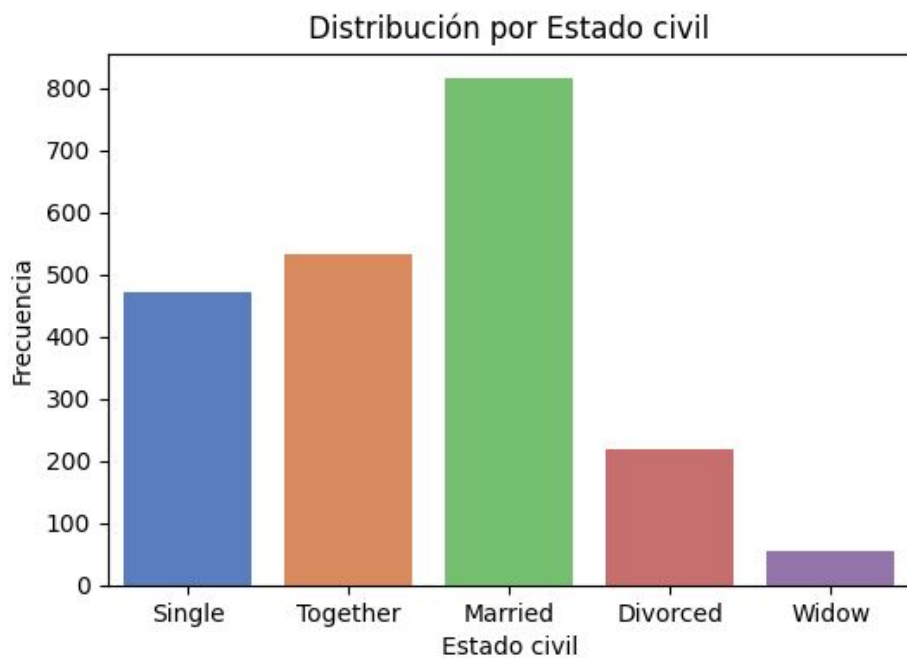
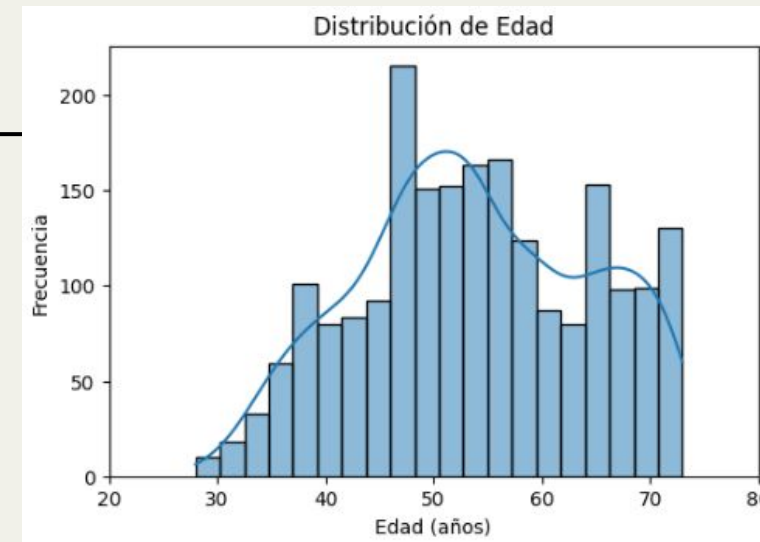
El objetivo de este proyecto es desarrollar un modelo predictivo el cual permita predecir el comportamiento de los clientes y aplicarlo al resto de la customer base.



Se busca maximizar el impacto de las futuras campañas y lograr ofrecer promociones seleccionando específicamente a los clientes que aceptarían la oferta, y evitar a los que la rechazarían, así hacer la campaña altamente rentable.

Algunas visualizaciones sobre los clientes

Algo que resultó no tener mucha correlación alguna.



Correlaciones

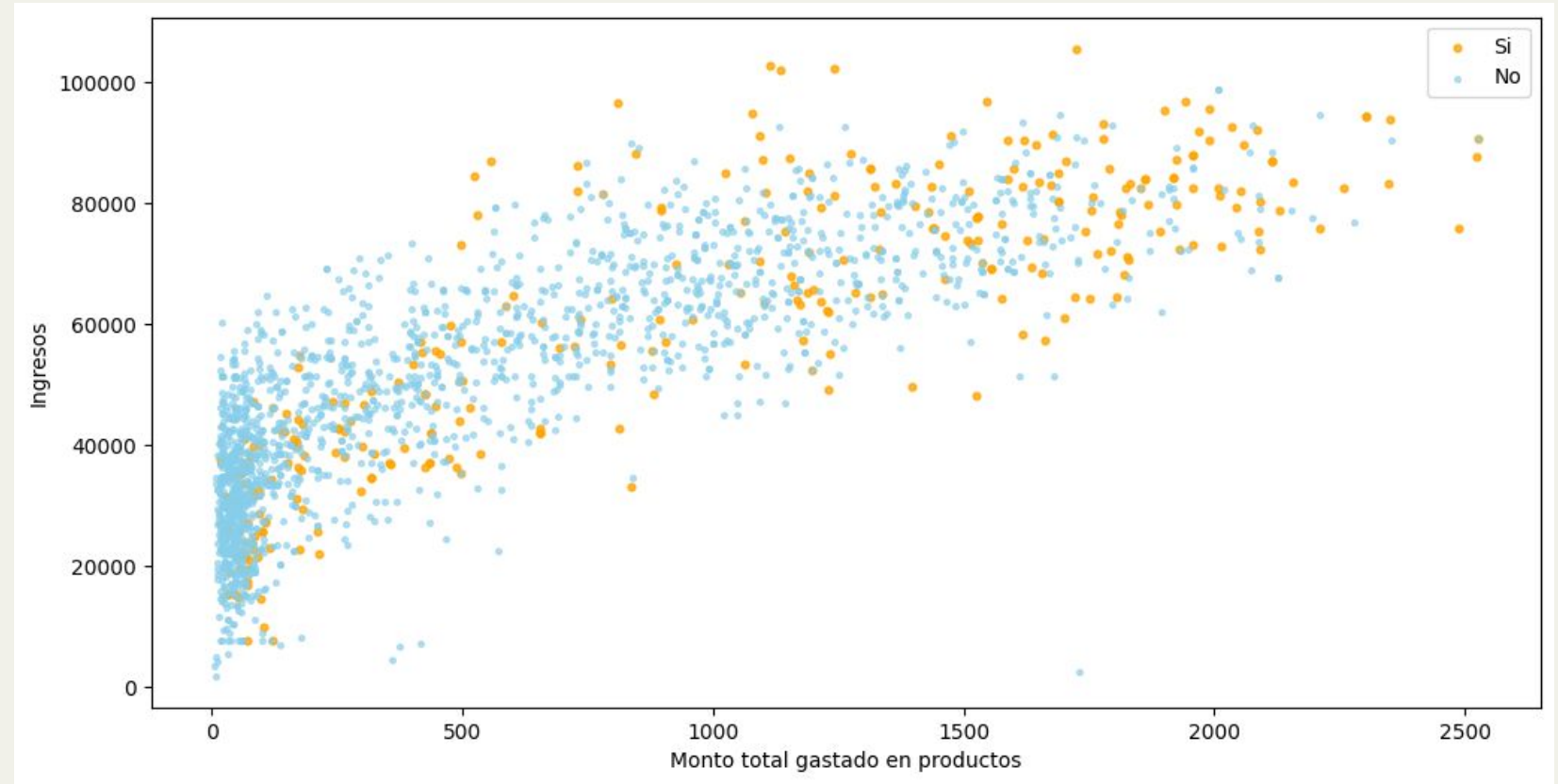
MntWines

MntMeatProducts

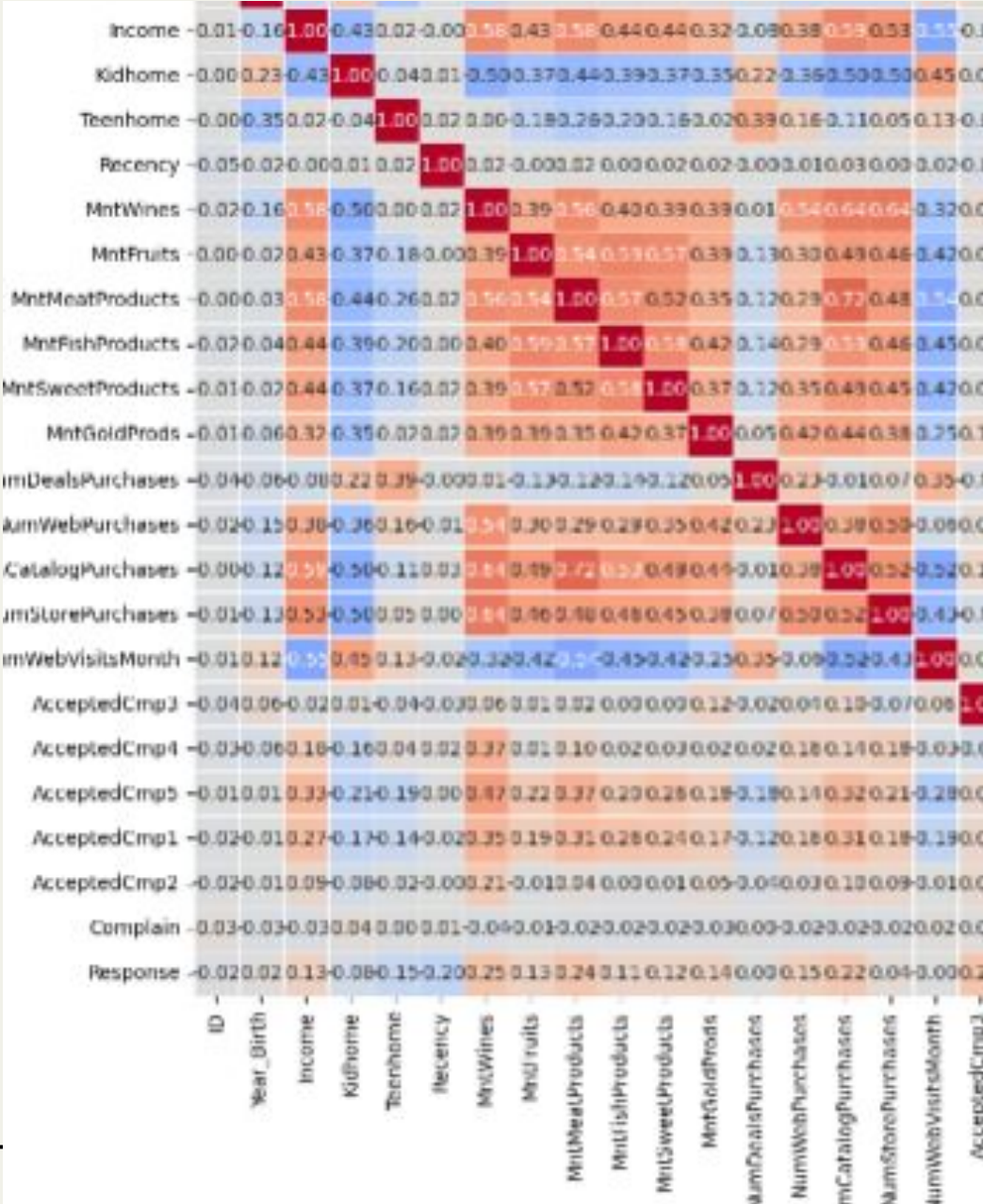
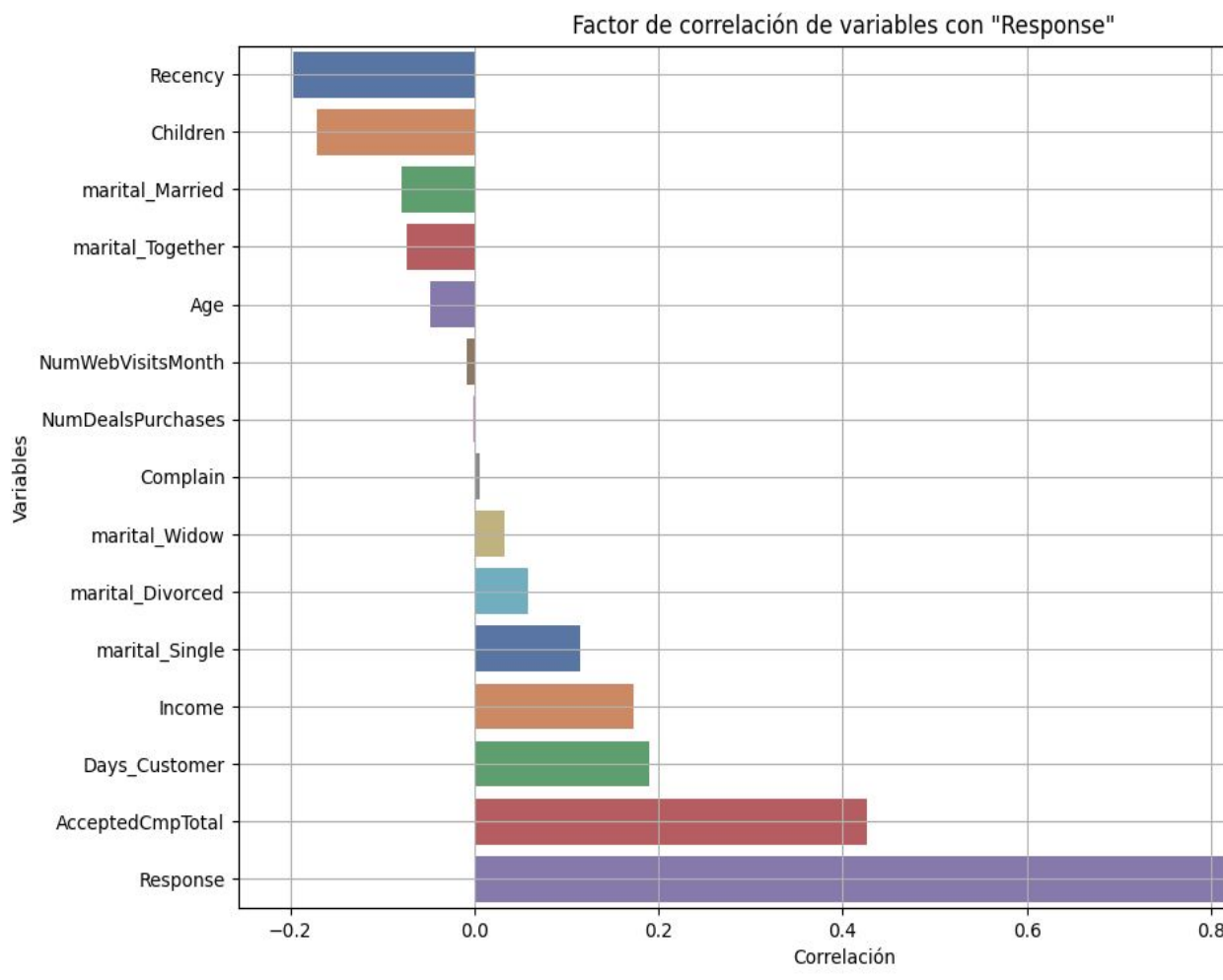
1	0.54	0.64	0.64
30.30	0.49	0.46	
20.29	0.72	0.48	
40.29	0.53	0.46	

NumCatalogPurchases

NumStorePurchases

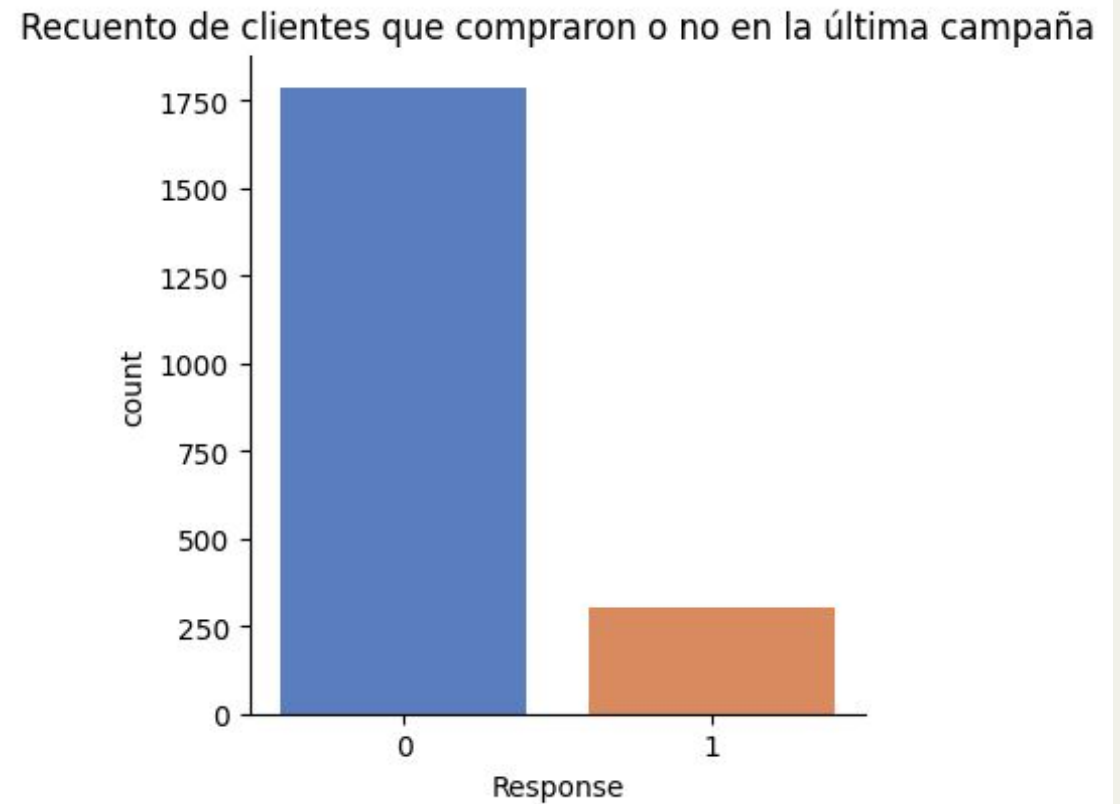
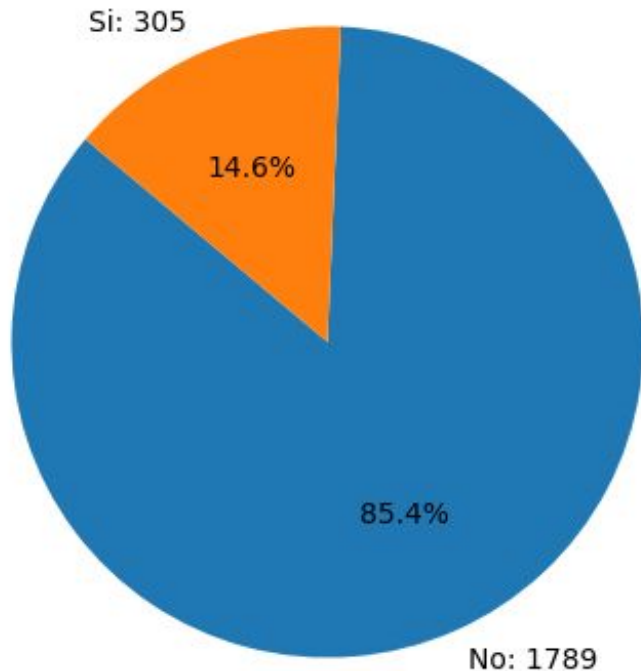


Visualización de correlaciones



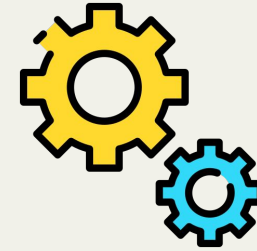
Balanceo de muestras

Al tener una distribución desigual en las muestras de la variable a predecir, puede traer un desbalance en el entrenamiento, lo cual genera dificultades para la predicción.



Se hizo un balanceo con la técnica de oversampling, lo cual produce muestras sintéticas en base a los datos existentes, para igualar la cantidad.

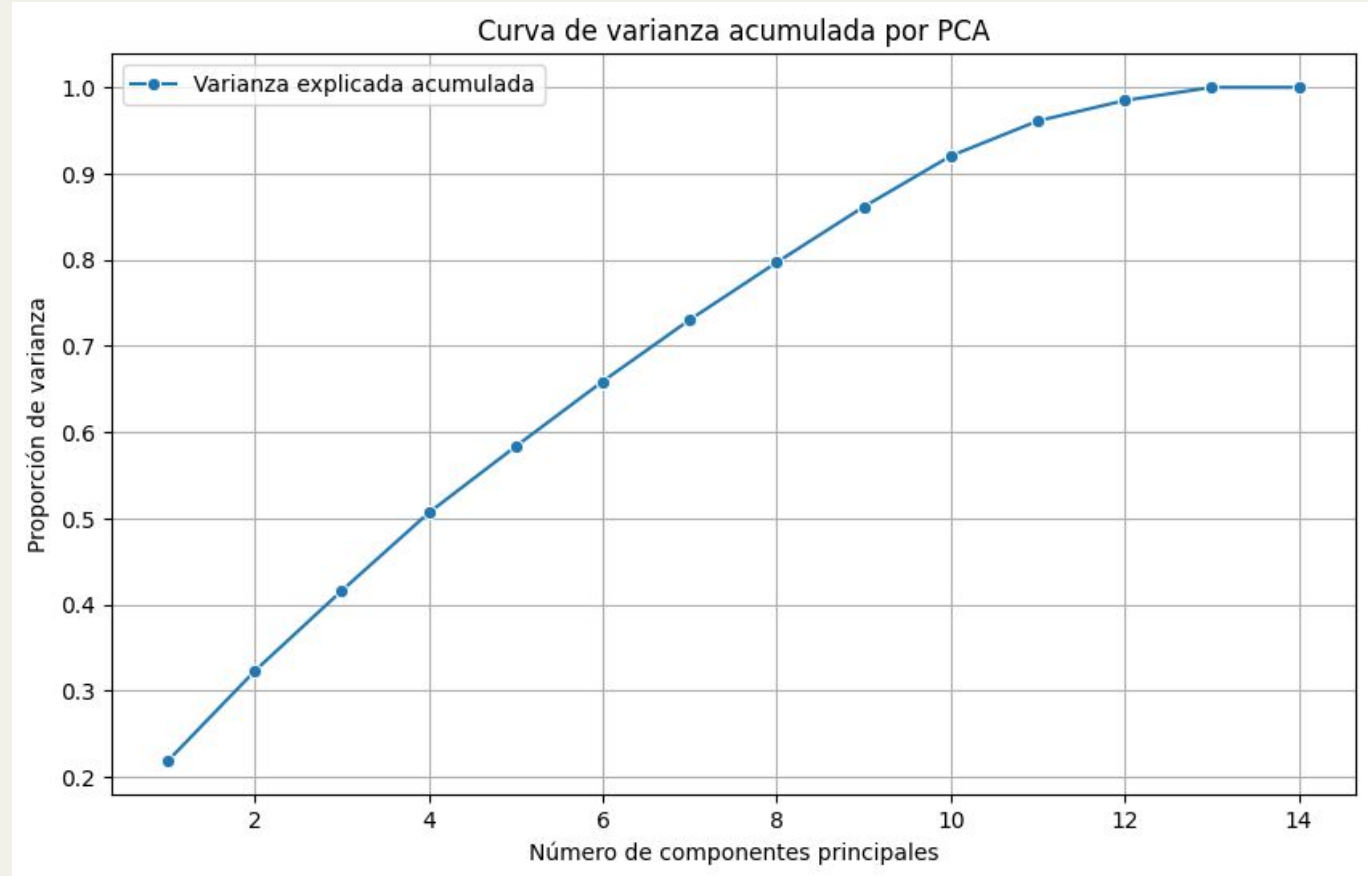
PCA - Análisis de componentes principales



El PCA es una técnica la cual mide la varianza que aporta la cantidad de componentes.

El punto ideal en mi caso está en los 12 componentes, donde ya no hay un aumento significativo.

Cuanto mayor la acumulación de varianza y menor la cantidad de componentes, mejor.





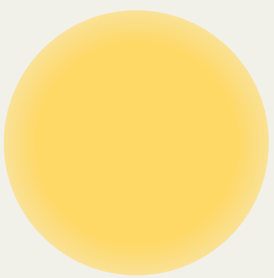
Modelos

Arbol de decisión

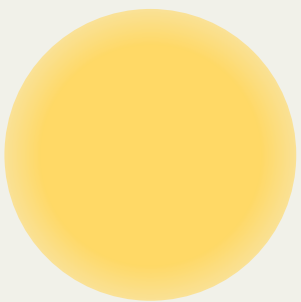
A

Regresión Logística

RandomForestClassifier









Thank you!

Do you have any questions?

hello@mail.com

555-111-222

mydomain.com



Editable Icons

