

Proto  
type

BEDU: Machine Learning

# Proyecto Esfera de Cristal

---

Dic 2022.



B+EDU

Santander



# Equipo 12

“Estamos contigo”



Francisco Valerio López



Daniel Gutiérrez Ruiz



Diego Zacarías de la Rosa

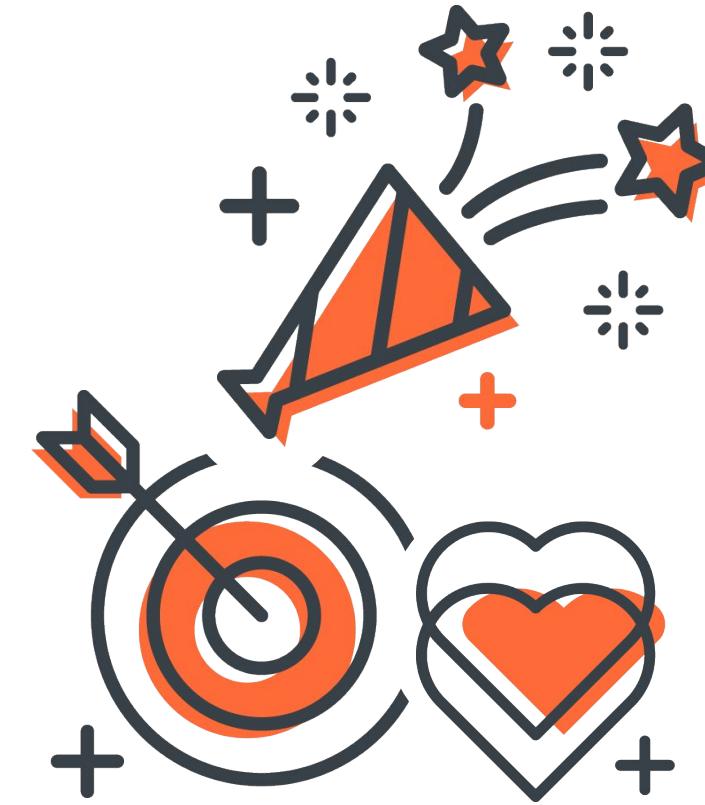


Baltasar Vez de la Garza



Proto  
type

# Y tú sabes ¿cuánto se gasta en PROMOCIÓN? · · ·



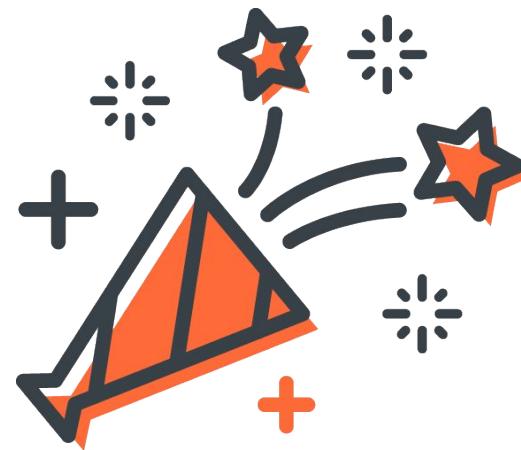
+ proto  
+ type + + +



Mínimo  
**10%**  
Del Ingreso  
Bruto

**“la estrategia  
de promoción  
es  
fundamental”**

- # ▪ ▪ ▪ ▪ Promoción, Promoción, Promoción



# ¿Cómo medir una promoción?

The diagram illustrates the formula for calculating ROI. It features the acronym **ROI** in large purple letters followed by an equals sign (=). To the right of the equals sign is a horizontal blue line. Above the line, the word **Ganancias** (Profits) is positioned above a bar chart consisting of four colored bars (blue, green, orange, pink). Below the line, the word **Inversión** (Investment) is positioned above two gold coins, one of which has a dollar sign (\$) on it. Another gold coin with a dollar sign (\$) is shown below the line, under the label **Inversión**. A minus sign (-) is placed between the bar chart and the bottom gold coin.

**NO**

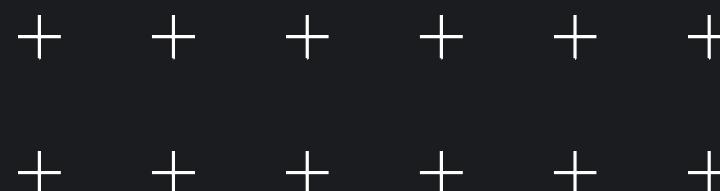
Existen métodos para predecir el éxito de una promoción



## Predicir



*Grandes y medianas empresas con al menos 1 año de existencia, que cuenten con la capacidad de recolectar y almacenar datos.*



# Data Scientist

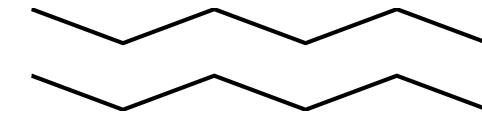


# Objetivo



Demostrar la efectividad de los modelos de Machine learning como herramienta para la predicción del éxito de campañas de promoción

▪ ▪ ▪ ▪ ▪ Data Set



El Banco Bank & Gold quiere repetir su  
última promoción de oferta de créditos  
personales

Las características que tenemos de cada cliente son:

**Edad | Experiencia Laboral | Ingresos | Código Postal | Dependientes | Gastos | Nivel de  
Educación | Hipoteca | Crédito Personal | Activos Financieros | Cuenta Ahorros  
Banca Online | Tarjeta de Crédito |**

+

+

+

+

+

+

+

+

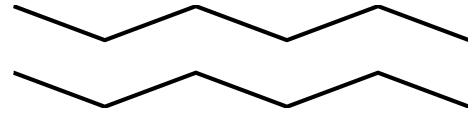
+

+

+

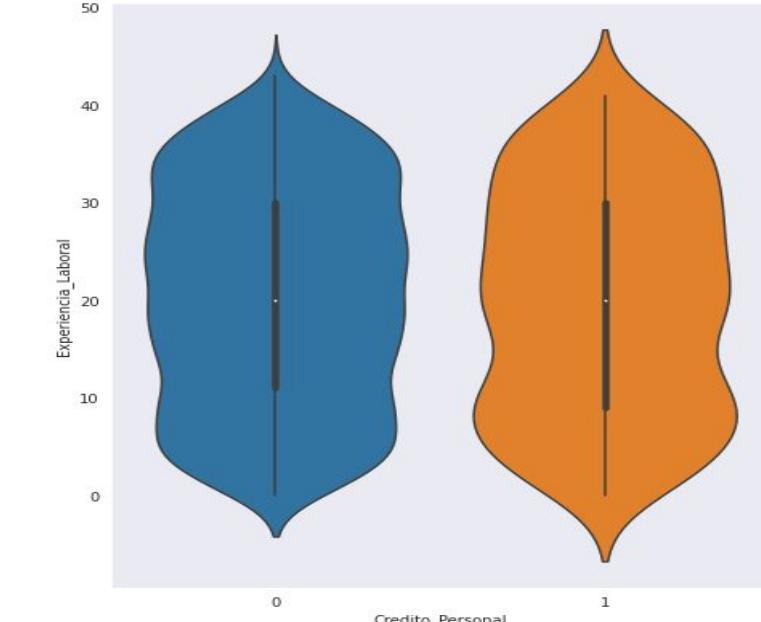
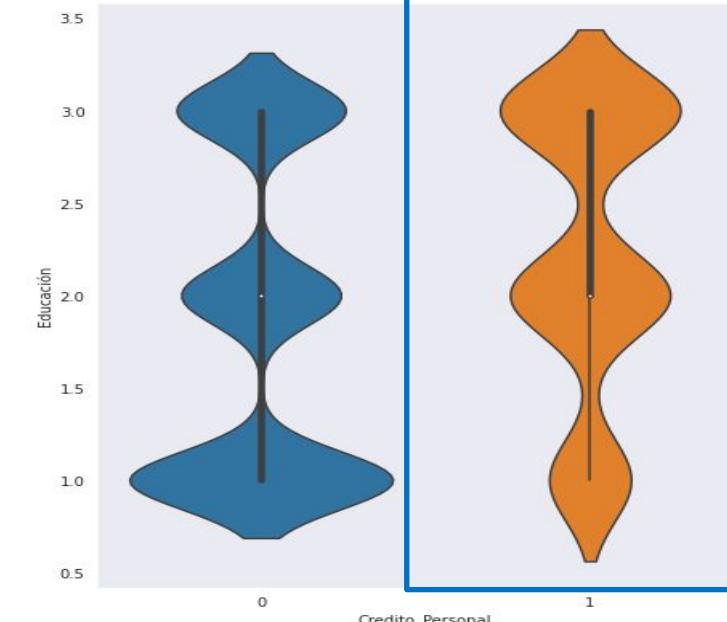
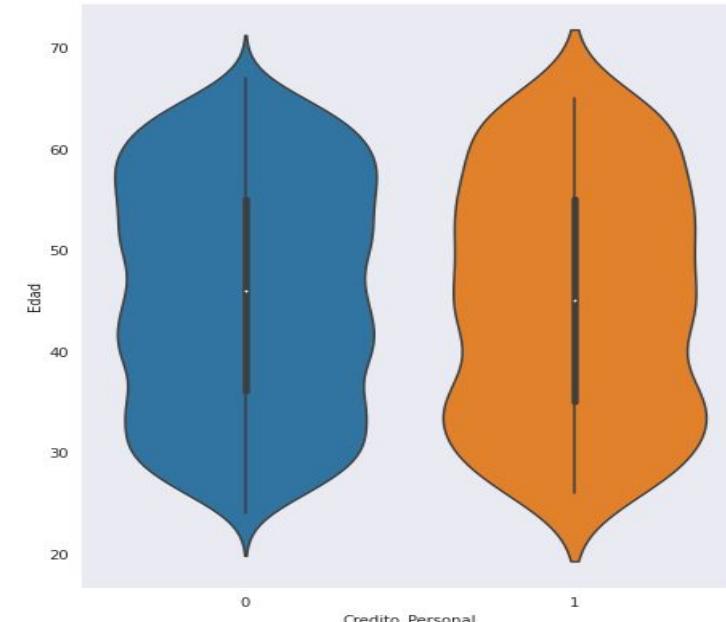
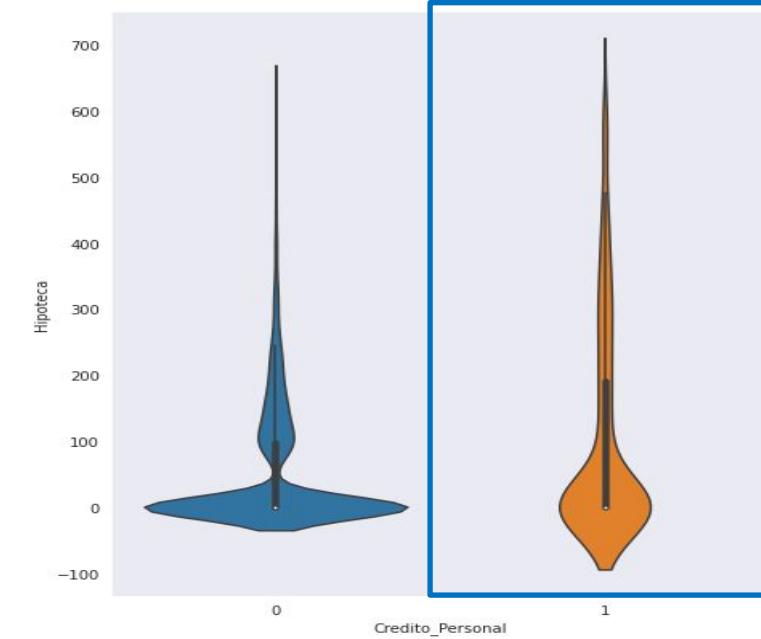
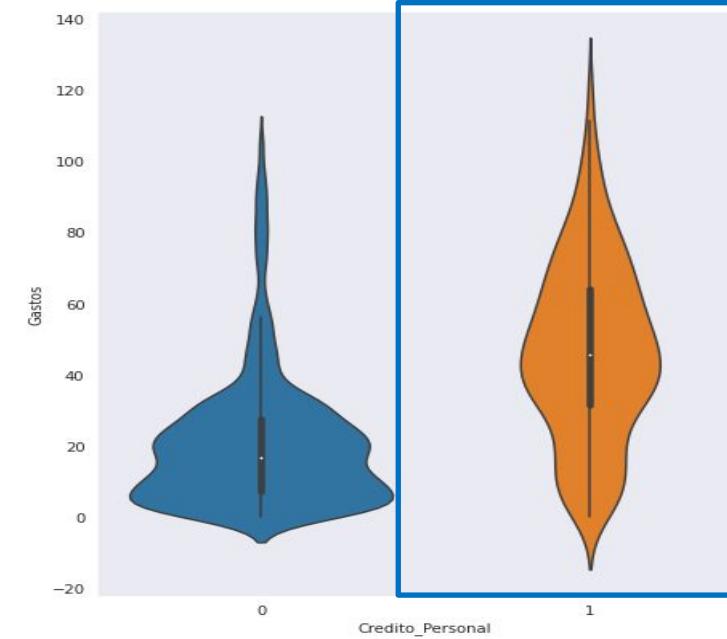
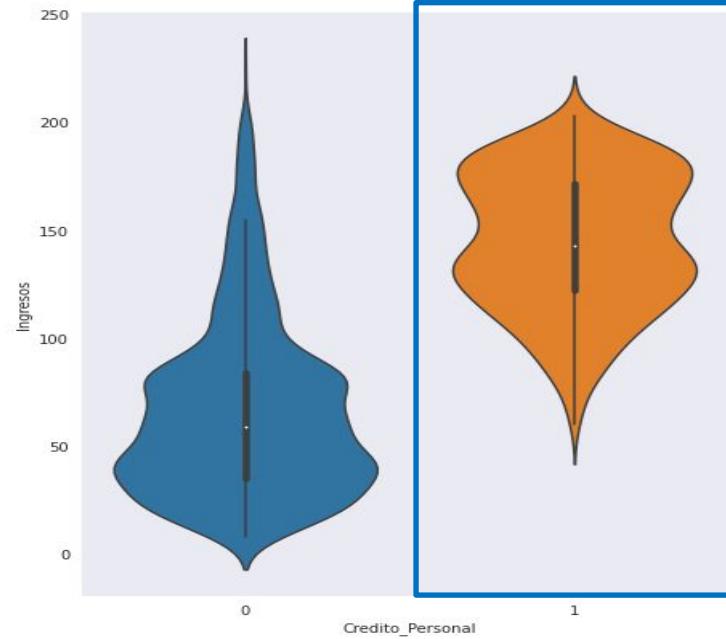
+

# Exploración de Locación y Variabilidad Segmentado por dato de Interés

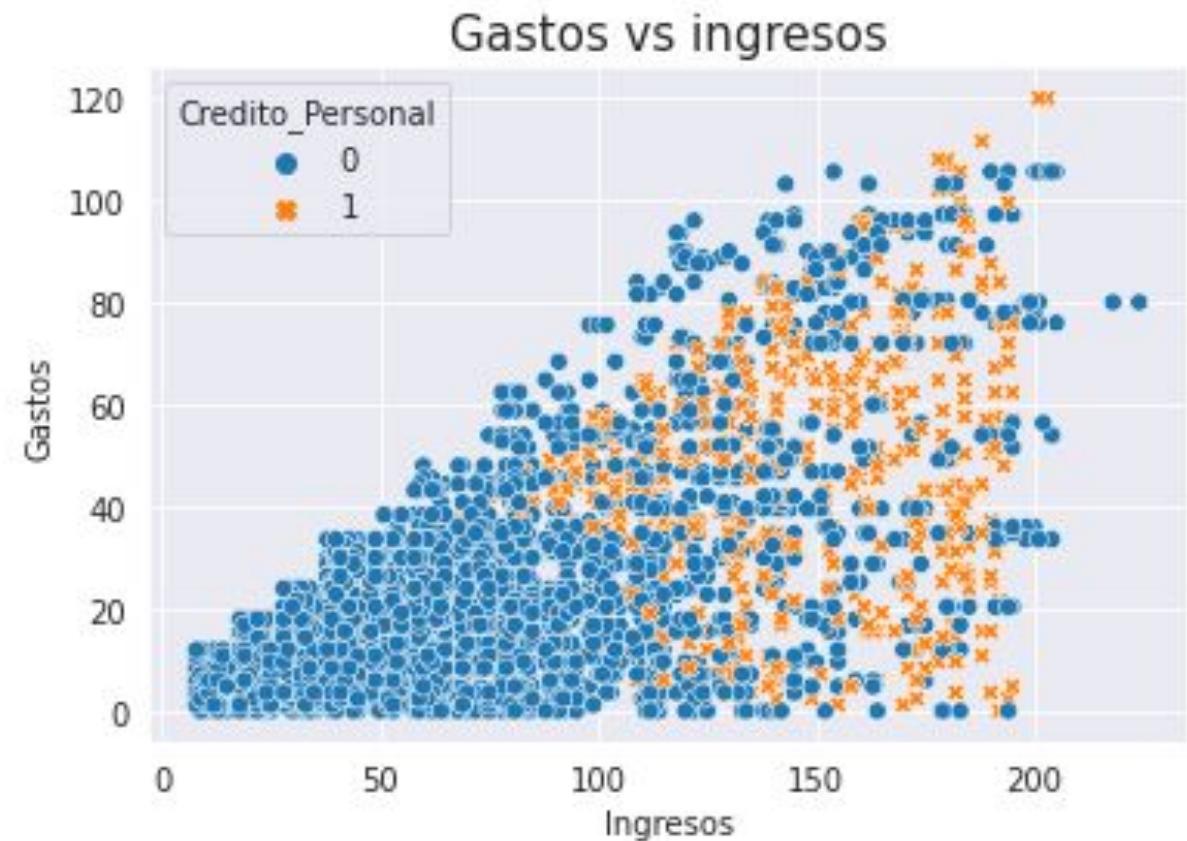
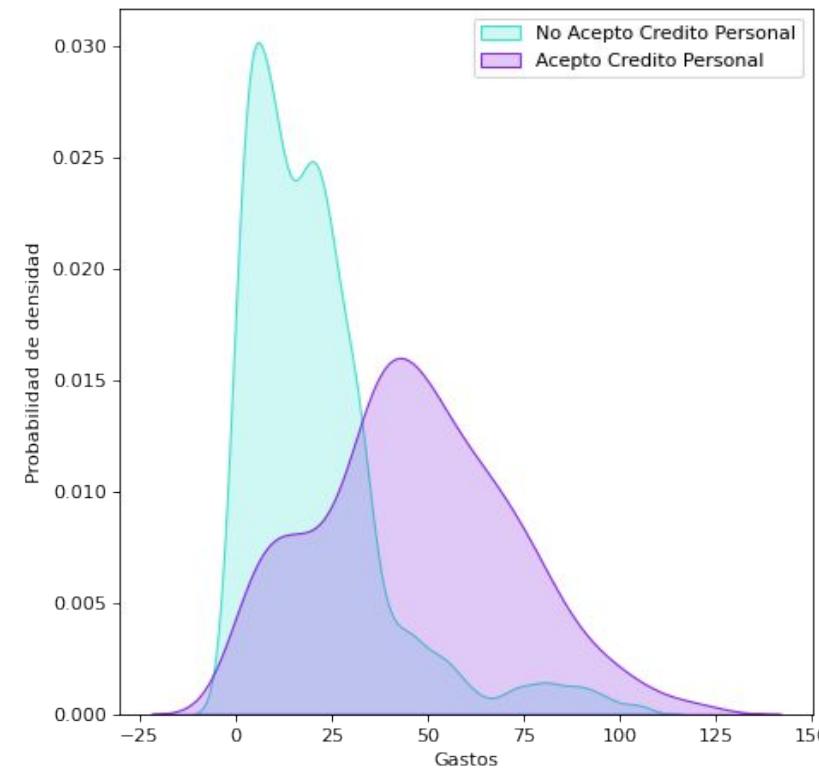
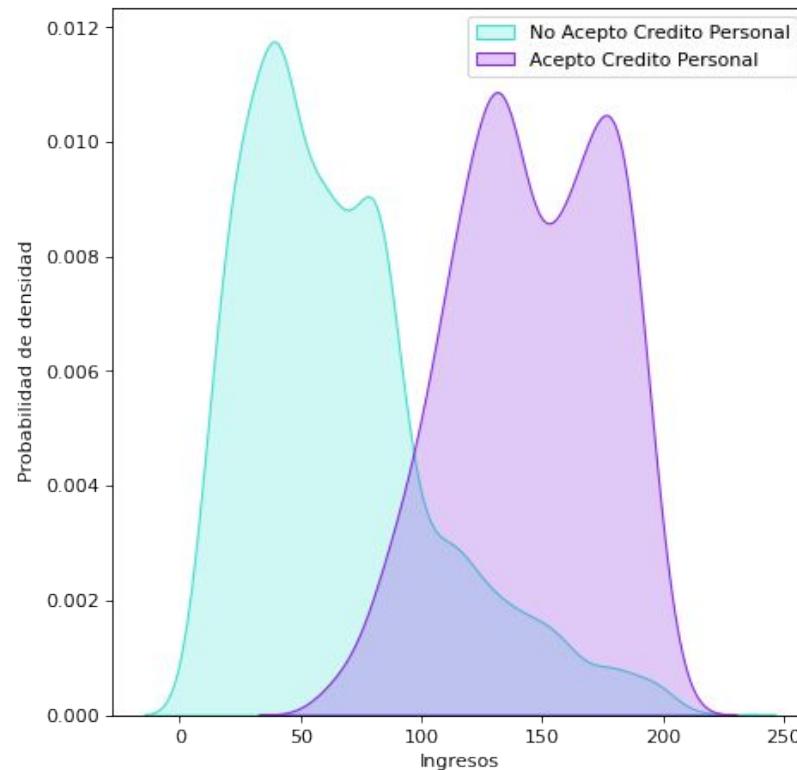
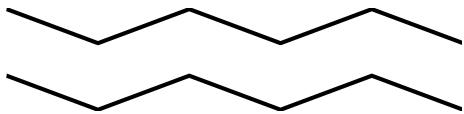


- < Ingreso
- < Gasto
- < Hipoteca
- < Educación

■ Aceptó el Crédito Personal  
■ No Aceptó el Crédito Personal



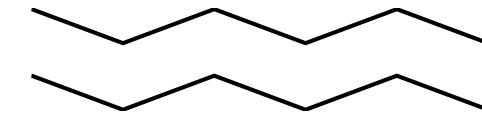
# • • • • Visualización de variables



+

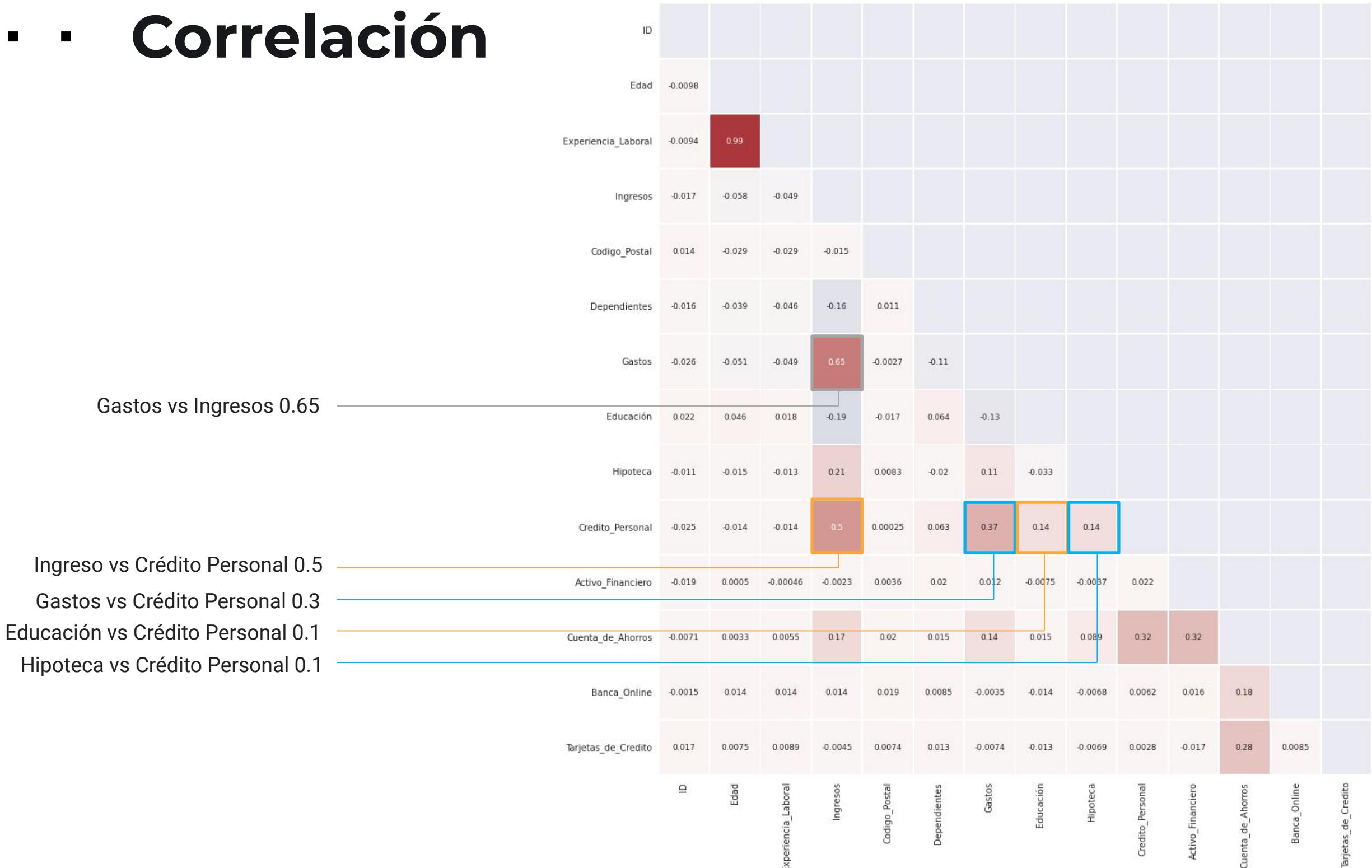
+

- • • •
- Visualización Avanzada

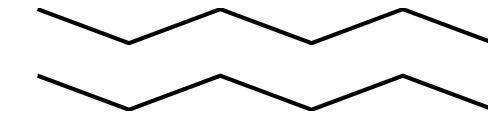


Podemos notar que las personas que tienen mayores ingresos y por ende que tienen más gastos son las que tienden a aceptar el crédito personal

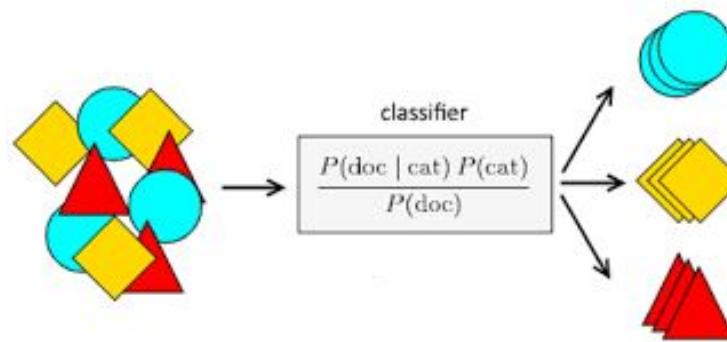
# Correlación



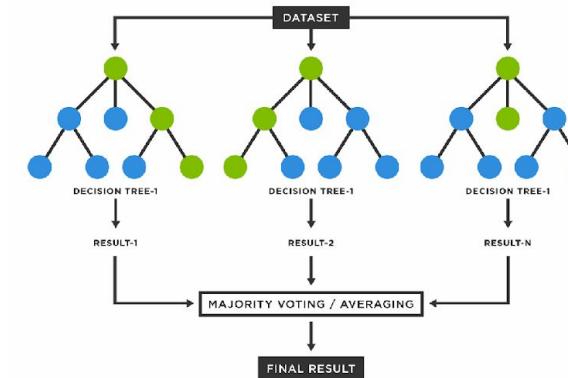
# Uso de Machine Learning



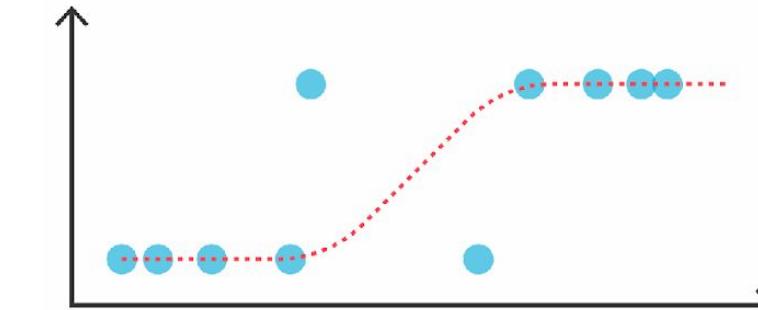
## Naive Bayes



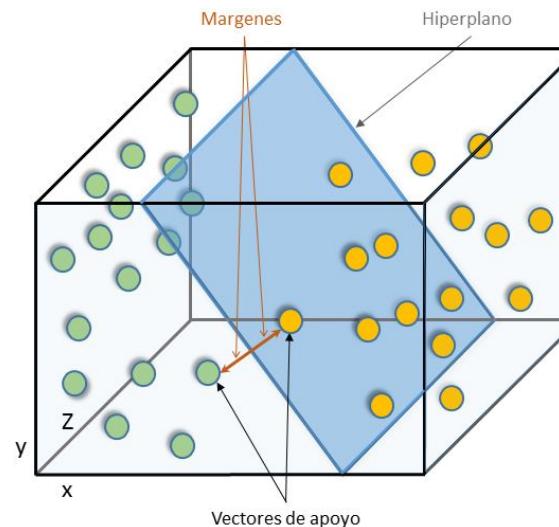
## Random Forest



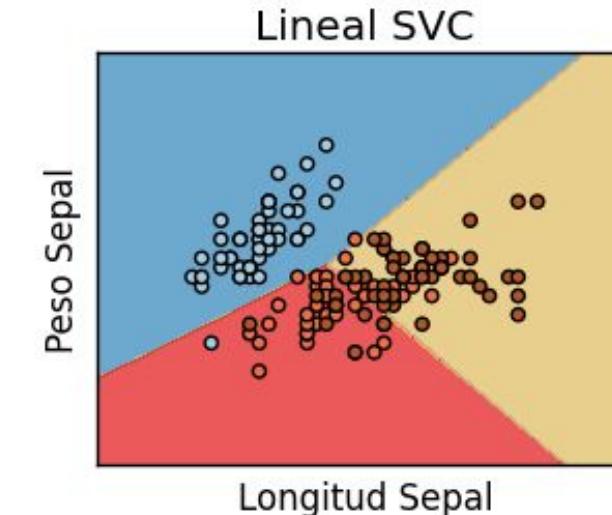
## Regresión Logística



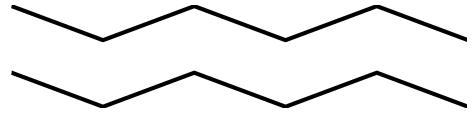
## Máquina de soporte vectorial



## Máquina de Soporte vectorial Lineal



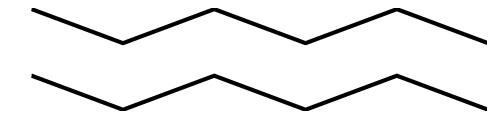
# Esferas de cristal a tu medida (Guerra de Clasificadores)



	Regresión Logística	Naive Bayes	Máquina de soporte Vectorial (SVC)	Máquina de soporte Vectorial lineal (Linear SVC)	Random Forest
Precisión	94%	89.1 %	90.0 %	95.4 %	98.6 %
Sensibilidad	60.8%	60.8 %	0.0 %	62.8 %	89.5 %
Especificidad	97.9	92.2 %	100.0 %	99.0 %	99.6 %



# • • • • Esferas de cristal a tu medida (Guerra de Clasificadores)



## Regresión Logística

Matriz de confusión

1310 Verdaderos Negativos	27 Falso Negativos
58 Falsos Positivos	90 Verdaderos Positivos

## Naive Bayes

Matriz de confusión

1234 Verdaderos Negativos	103 Falso Negativos
58 Falsos Positivos	90 Verdaderos Positivos

## SVC

Matriz de confusión

1337 Verdaderos Negativos	0 Falso Negativos
148 Falsos Positivos	0 Verdaderos Positivos

## Linear SVC

Matriz de confusión

1324 Verdaderos Negativos	13 Falso Negativos
55 Falsos Positivos	93 Falsos Positivos

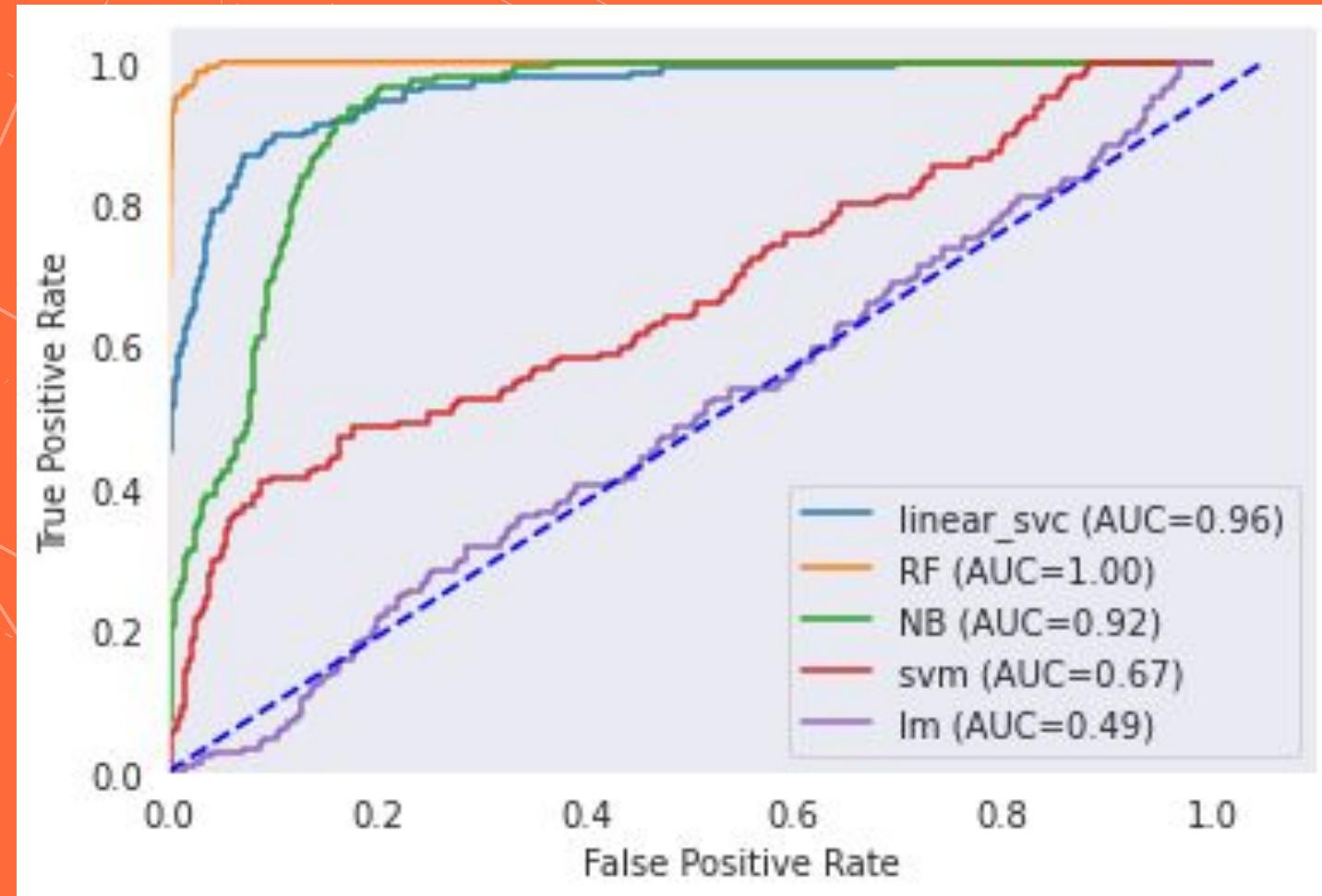
## Random Forest

Matriz de confusión

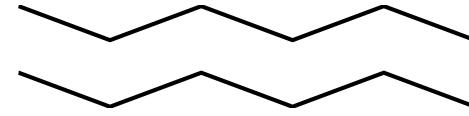
1336 Verdaderos Negativos	5 Falso Negativos
15 Verdaderos Positivos	129 Verdaderos Positivos



# Comparativo de modelos de curvas ROC/AUC



# Conclusión



## Características de Importancia Random Forest

Características	importancia
Ingresos	0.327941
Educación	0.179373
Gastos	0.166443
Dependientes	0.1041131
Cuenta_de_Ahorros	0.056672
Hipoteca	0.039185
Codigo_Postal	0.037731
Experiencia_Laboral	0.034857
Edad	0.034472
Tarjetas_de_Credito	0.008187
Banca_Online	0.007251
Activo_Financiero	0.003758

Random Forest también es capaz de aportarnos información complementaria como las características más relevantes

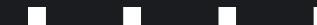
**Los modelos de Machine Learning tienen la capacidad de realizar predicción sobre una gran variedad de problemas científicos y de negocio.**

**Su uso para la predicción del éxito de campañas de promoción es completamente viable solucionando la carencia de herramientas de marketing actuales**



*De parte del Equipo 12  
Gracias.*

+++  
+++



# ▪ ▪ ▪ ▪ ▪ Referencias

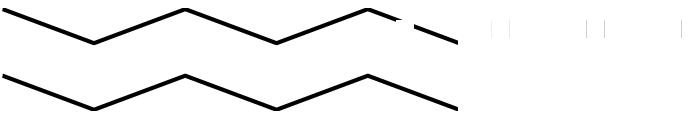


Imagen de Naive Bayes

<https://hands-on.cloud/implementing-naive-bayes-classification-using-python/>

Imagen Random Forest

<https://www.tibco.com/reference-center/what-is-a-random-forest>

Imagen Maquina de Soporte Vectorial

<https://numerentur.org/svm/>

Imagen Maquina de Soporte Vectorial

<https://dlegorreta.wordpress.com/2015/08/14/suppor-vector-machine-maquina-de-soporte-de-vectores-en-python/>

Imagen Regresión Logística

<https://www.tibco.com/reference-center/what-is-logistic-regression>

Modelo para curvas ROC

<https://github.com/vappiah/Machine-Learning-Tutorials/blob/main/notebooks/evaluation/Compare%20Classifiers%20using%20ROC%20Curves.ipynb>

++ + + + + + +

+ + + + + + +