



Noviembre 17, 2022

Machine Learning y Economía

Emilio Sandoval Palomino
Humberto Acevedo Cruz Manjarrez

“The world is its own best model”

- Rodney Brooks

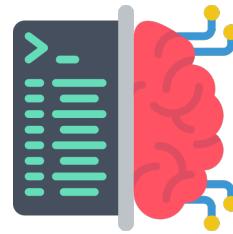
“All models are wrong”

- George Box

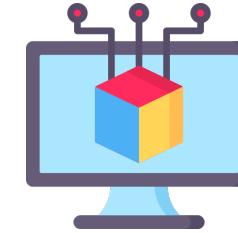
Índice



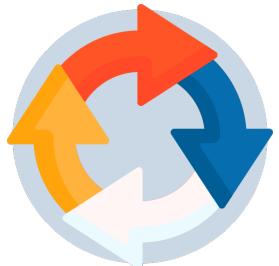
DL vs ML vs AI



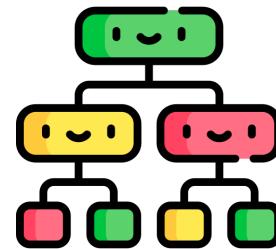
Tipos de
aprendizaje



Principales
herramientas



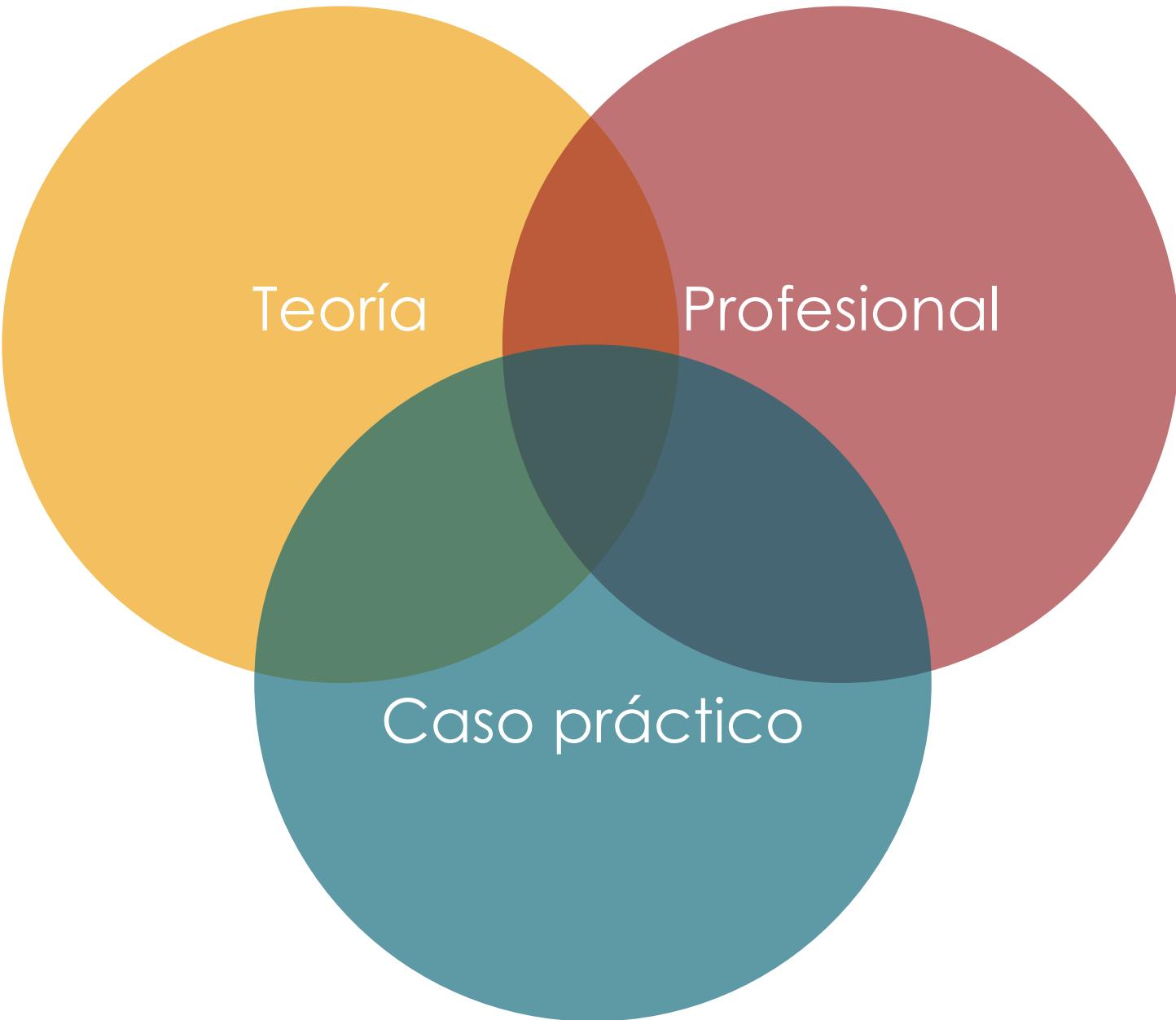
Círculo de vida de
un modelo



Caso práctico



Desentrañando
un modelo

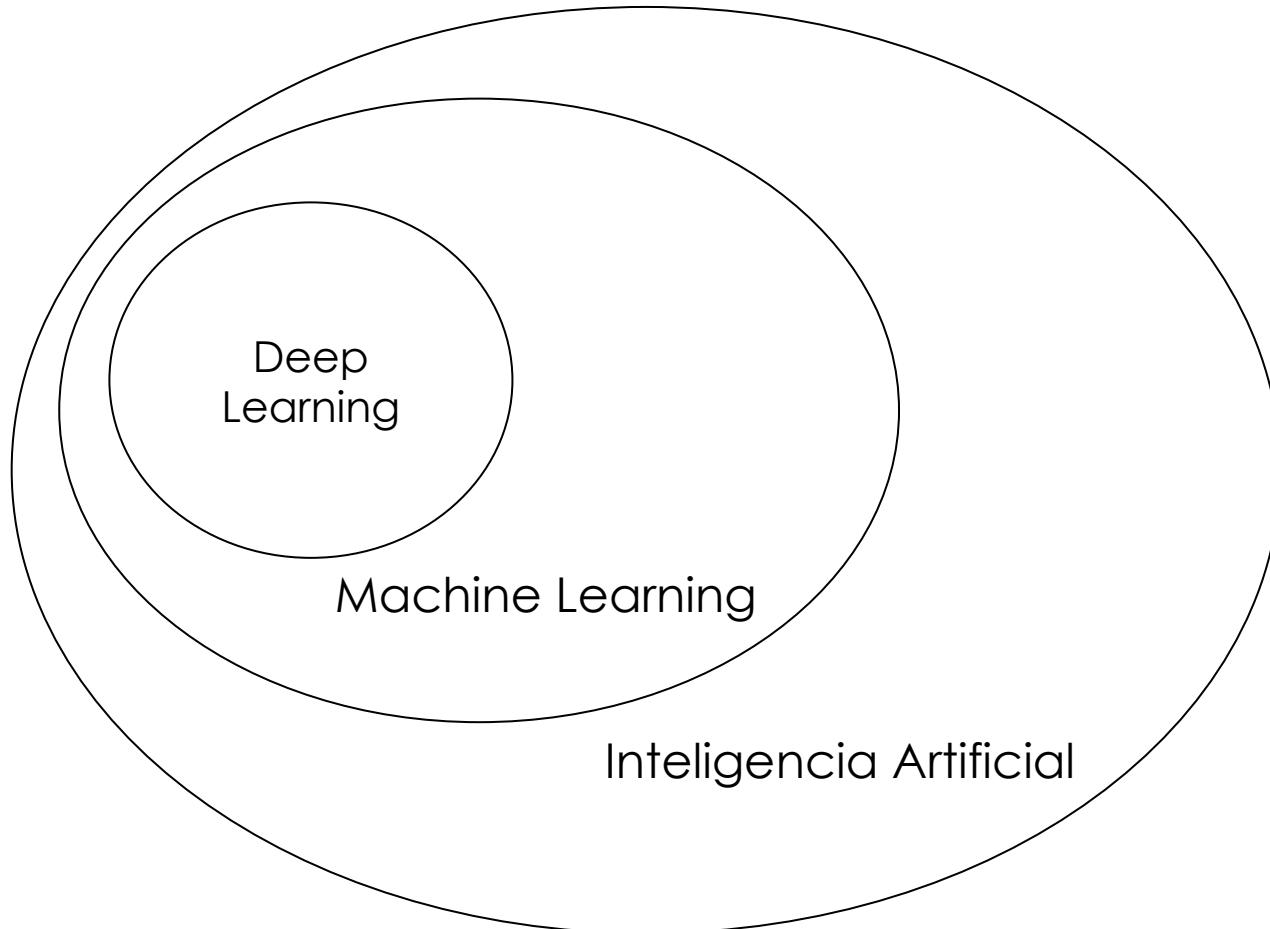


Problemas lineales

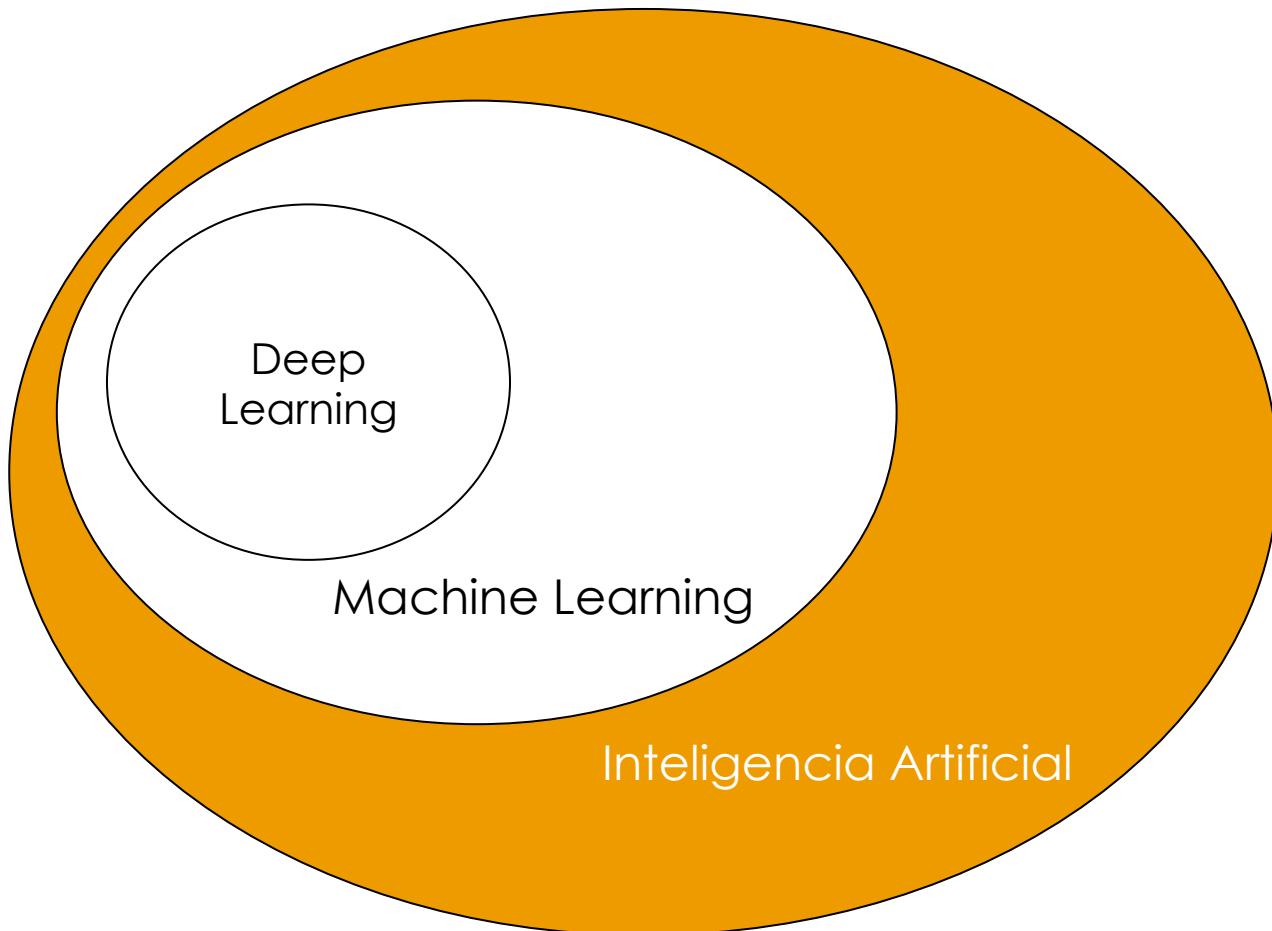
Problemas no lineales

Problemas no lineales

DL vs ML vs AI



DL vs ML vs AI



Multidisciplina que conjuga diferentes campos
bajo técnicas matemáticas y de computación
y da aplicaciones a través de diferentes
algoritmos

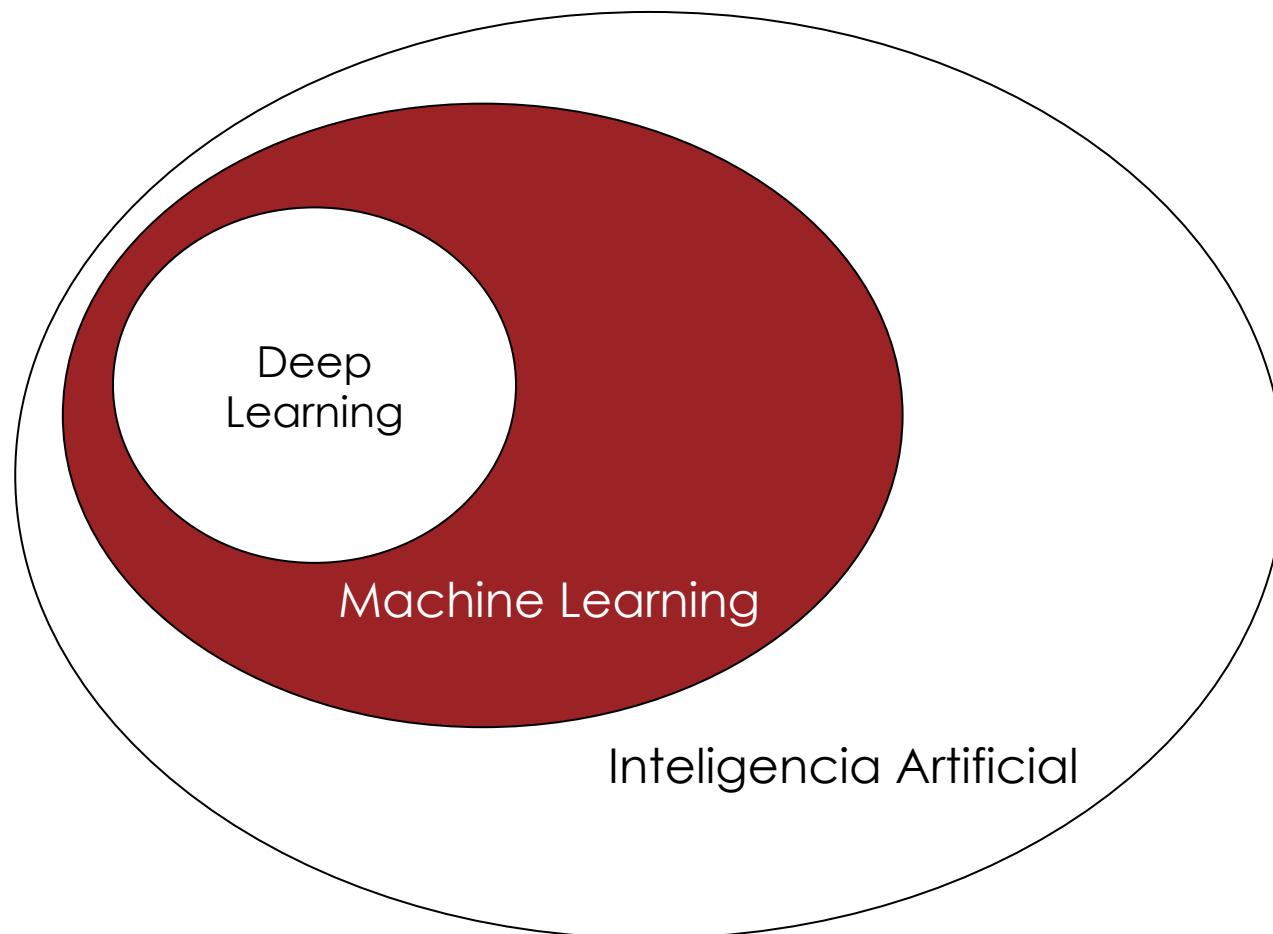


Self driving
vehicles

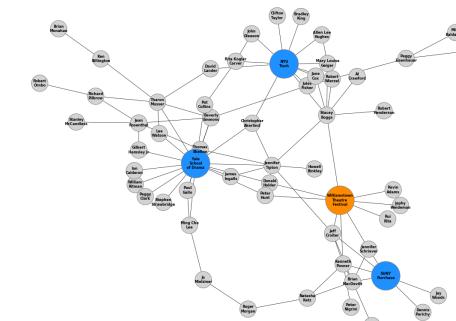


Robotics

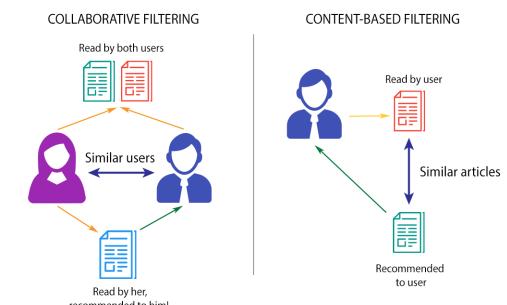
DL vs ML vs AI



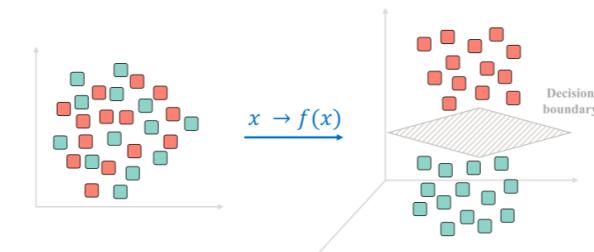
Redes Neuronales Artificiales las cuales se basan en el Teorema de Aproximación Universal. Difíciles de explicar, aunque con resultados que sobrepasan a otros modelos.



Pattern Recognition

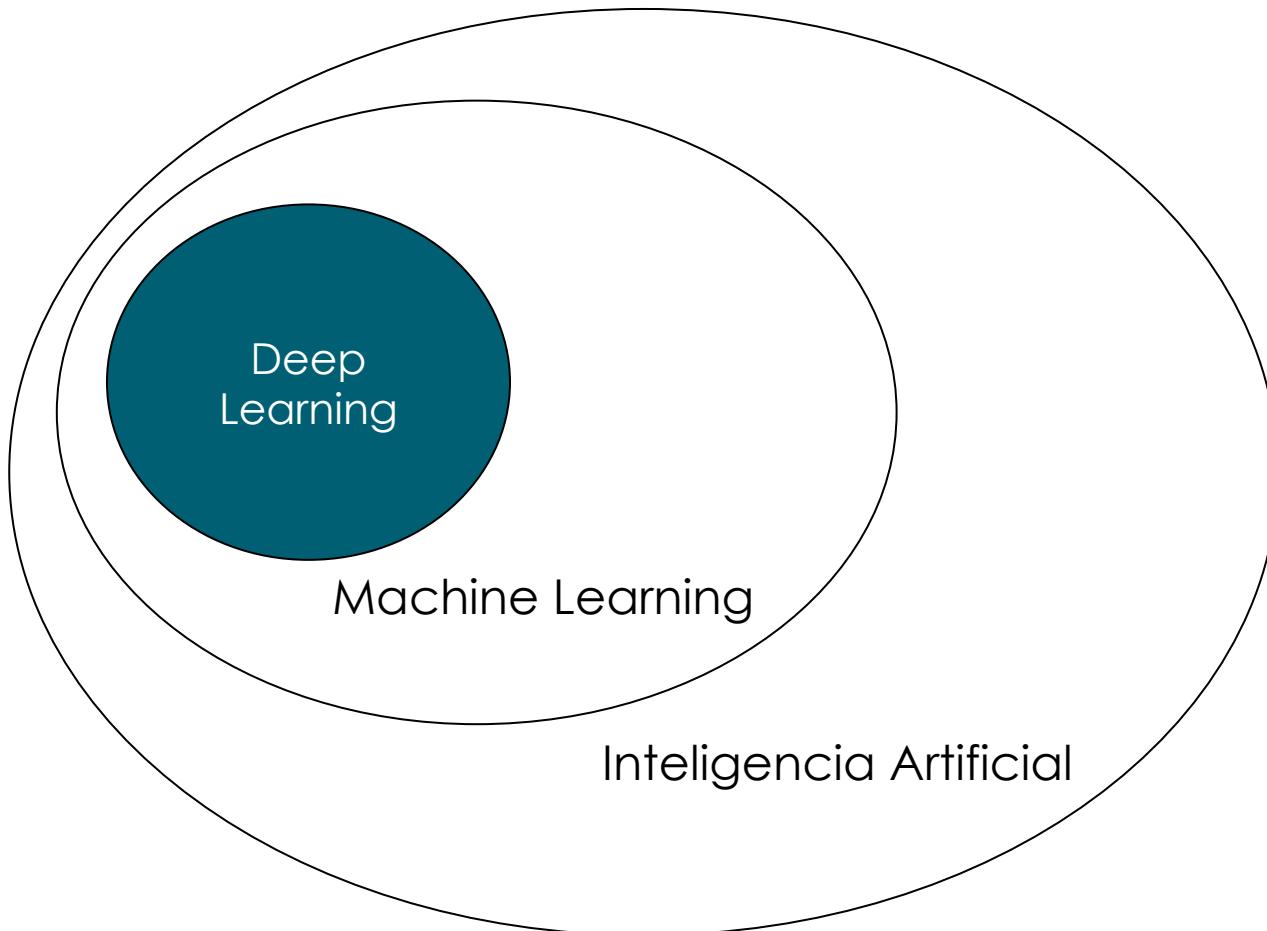


Recommender systems



Decision boundaries

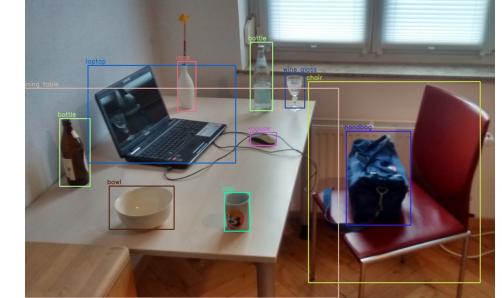
DL vs ML vs AI



Redes Neuronales Artificiales las cuales se basan en el Teorema de Aproximación Universal. Difíciles de explicar, aunque con resultados que sobrepasan a otros modelos.



Difussion
models



Object
Detection

This is easily the most underrated film inn the Brooks cannon. Sure, its flawed. It does not give a realistic view of homelessness (unlike, say, how Citizen Kane gave a realistic view of lounge singers, or Titanic gave a realistic view of Italians YOU IDIOTS). Many of the jokes fall flat. But still, this film is very lovable in a way many comedies are not, and to pull that off in a story about some of the most traditionally reviled members of society is truly impressive. Its not The Fisher King, but its not crap, either. My only complaint is that Brooks should have cast someone else in the lead (I love Mel as a Director and Writer, not so much as a lead).

Natural Language Processing

Empezando

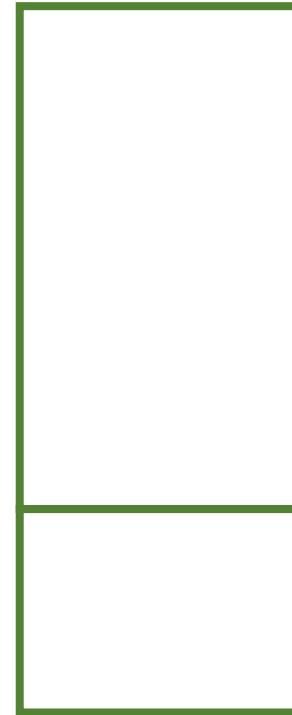
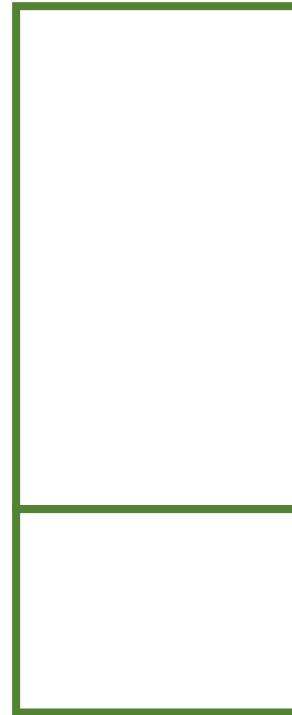
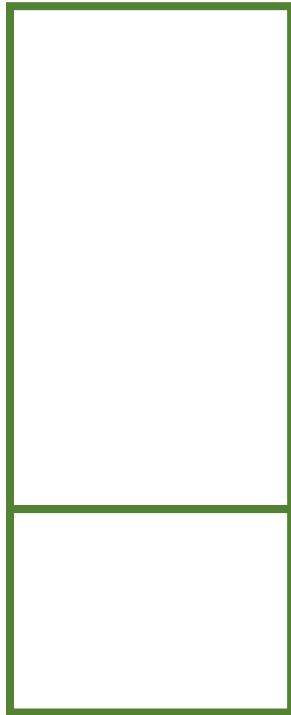
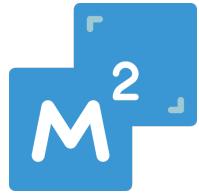
Programación

Base de Datos

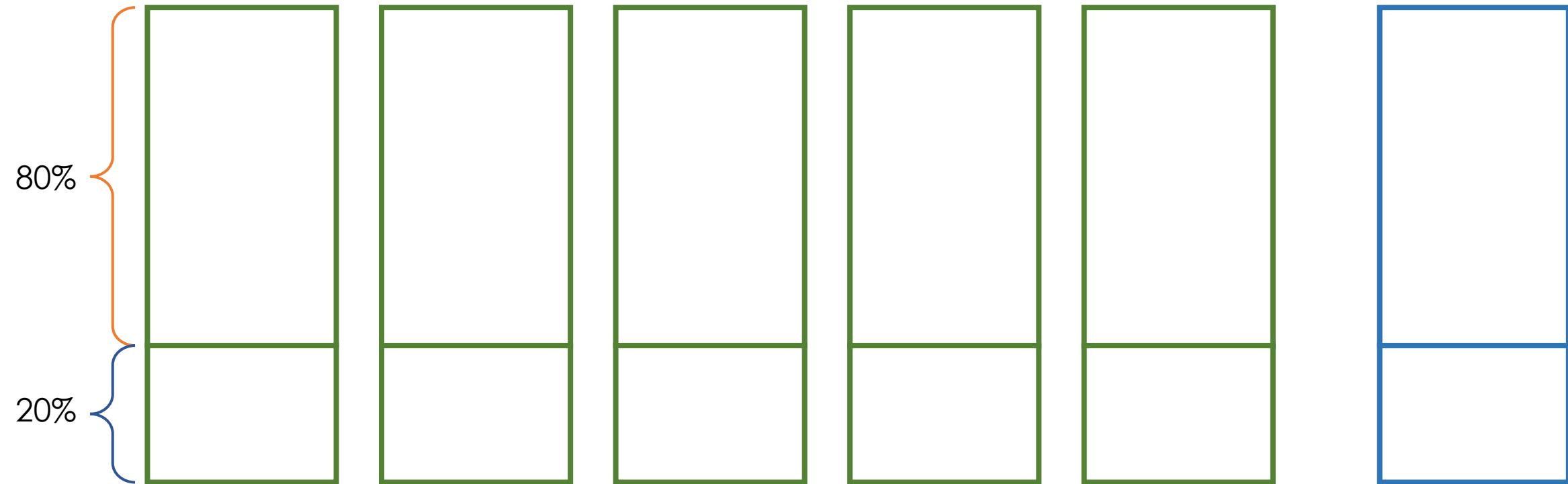
Matemáticas

Storytelling

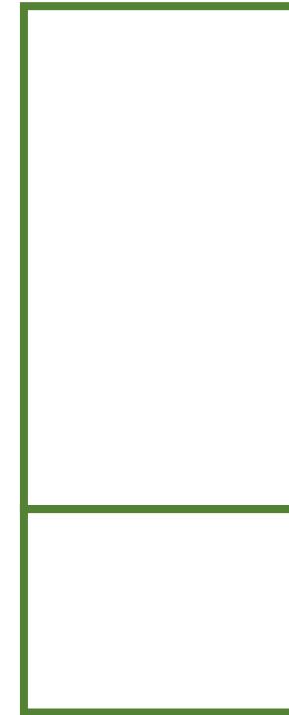
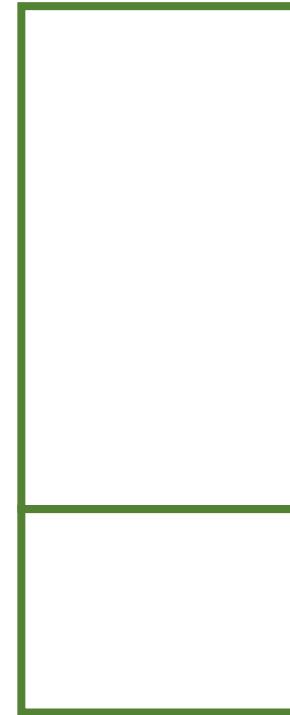
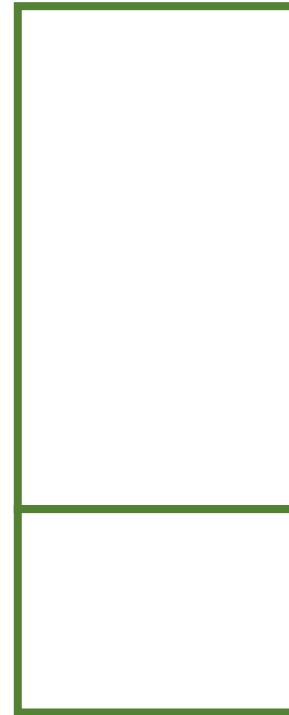
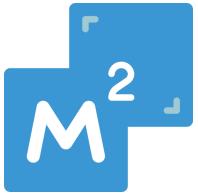
Aprendizaje Supervisado



Aprendizaje Supervisado

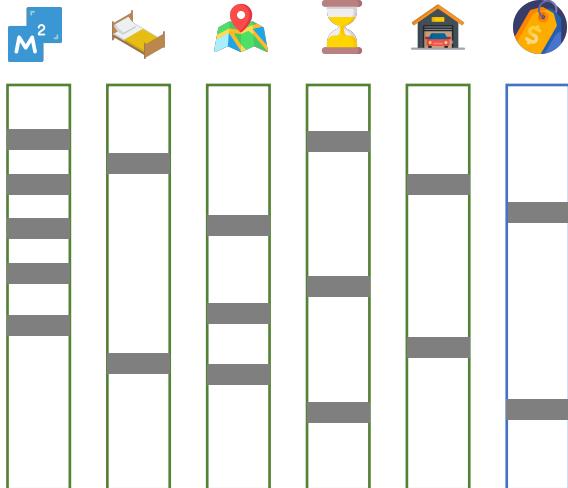


Aprendizaje no Supervisado



Otros aprendizajes

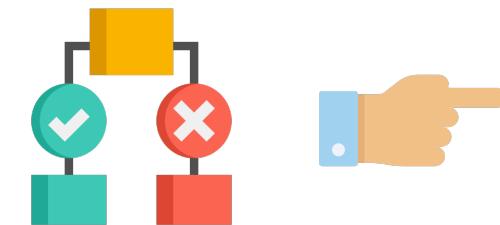
Semisupervisado



Por refuerzo

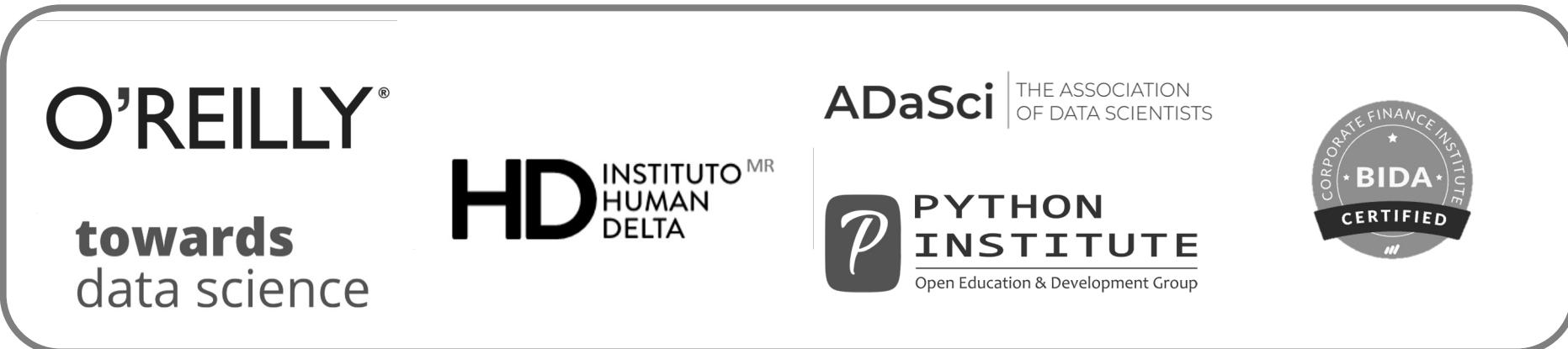
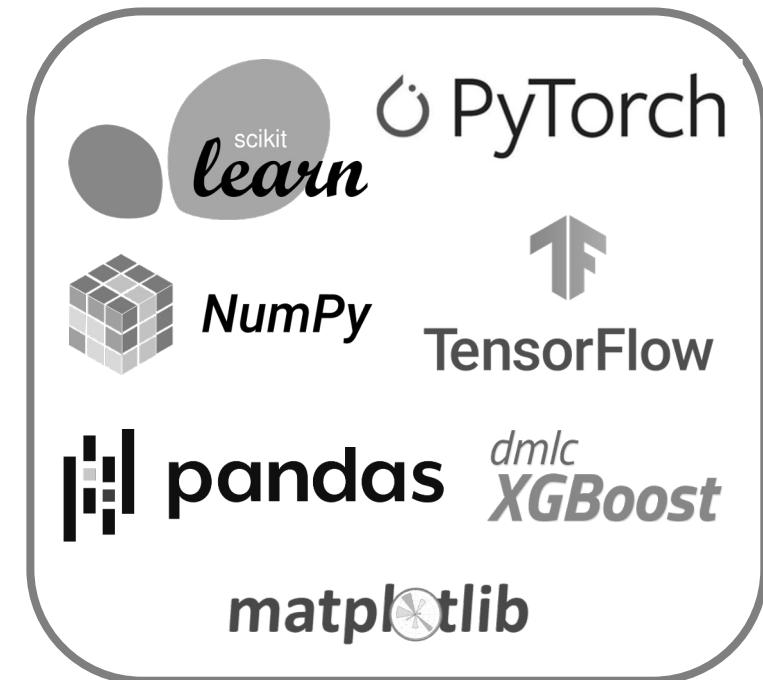
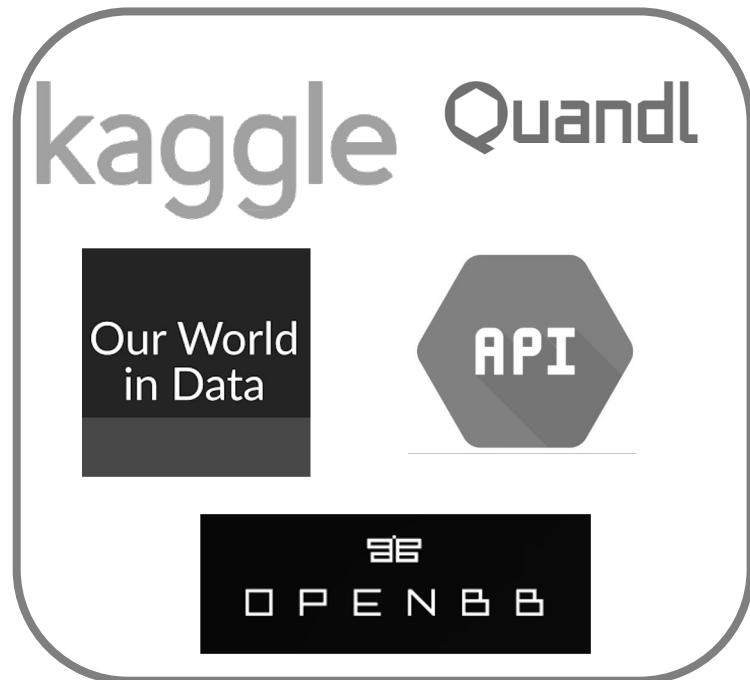


Activo

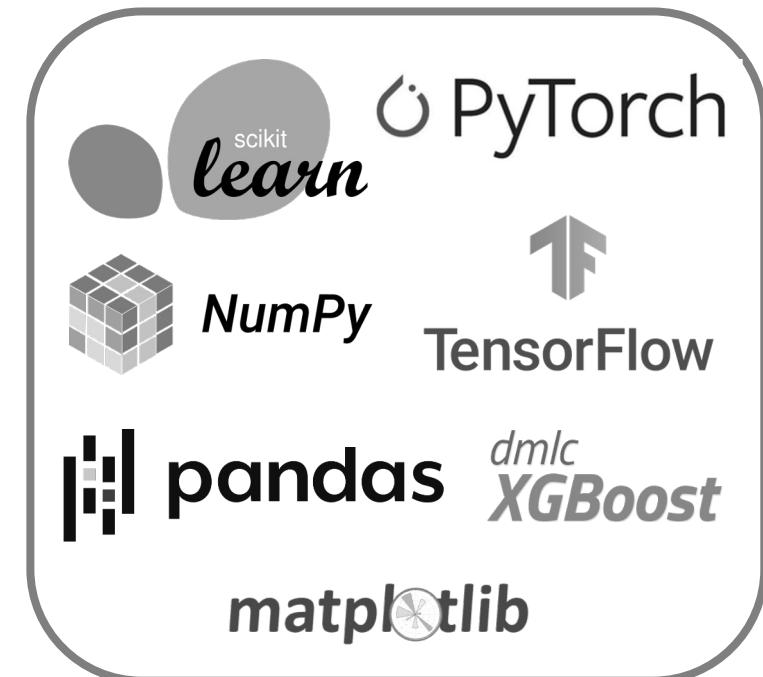
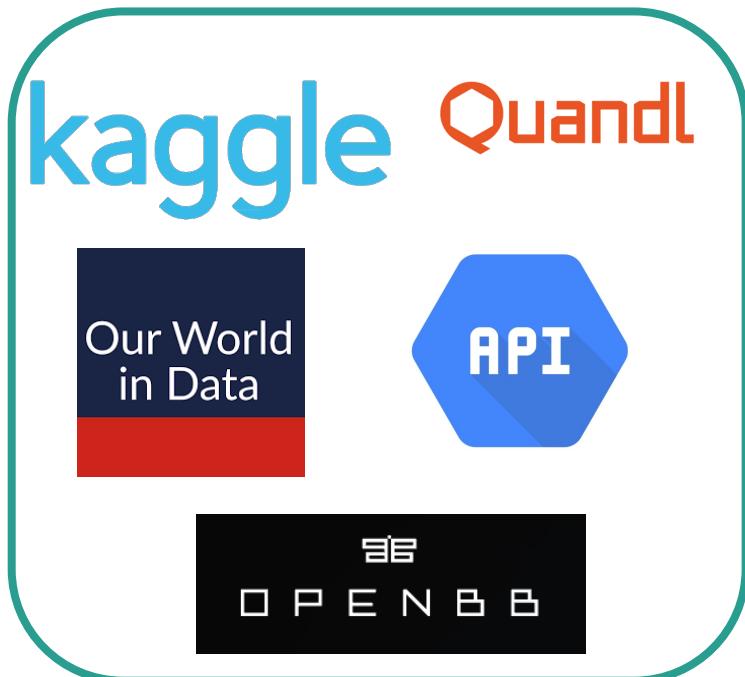


\$290
\$140
\$210
\$210
\$135
\$123
\$110

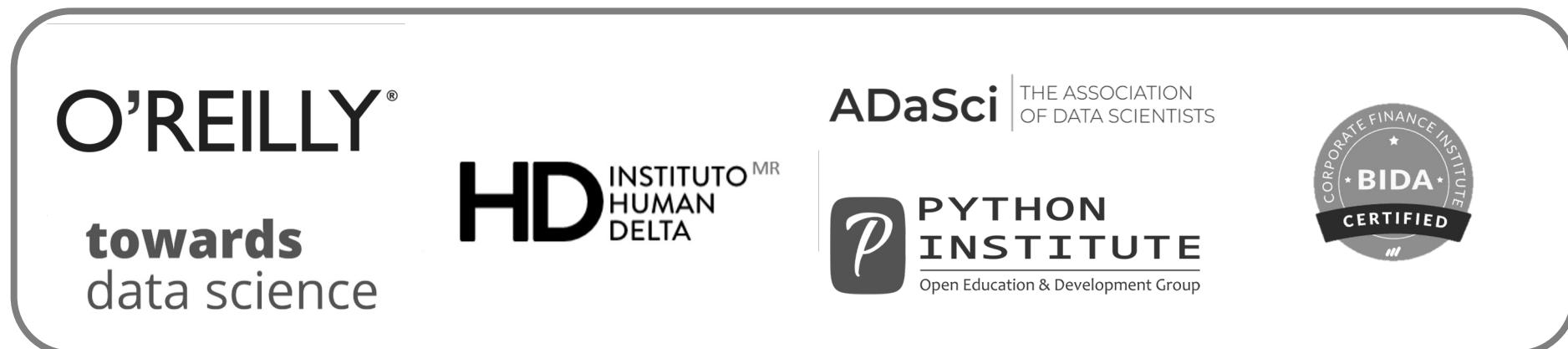
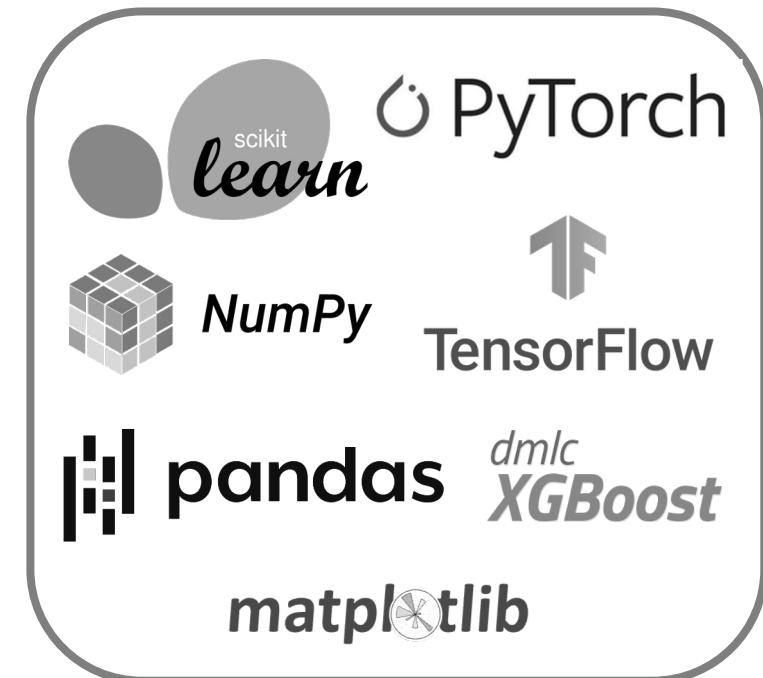
From Zero to Hero



From Zero to Hero



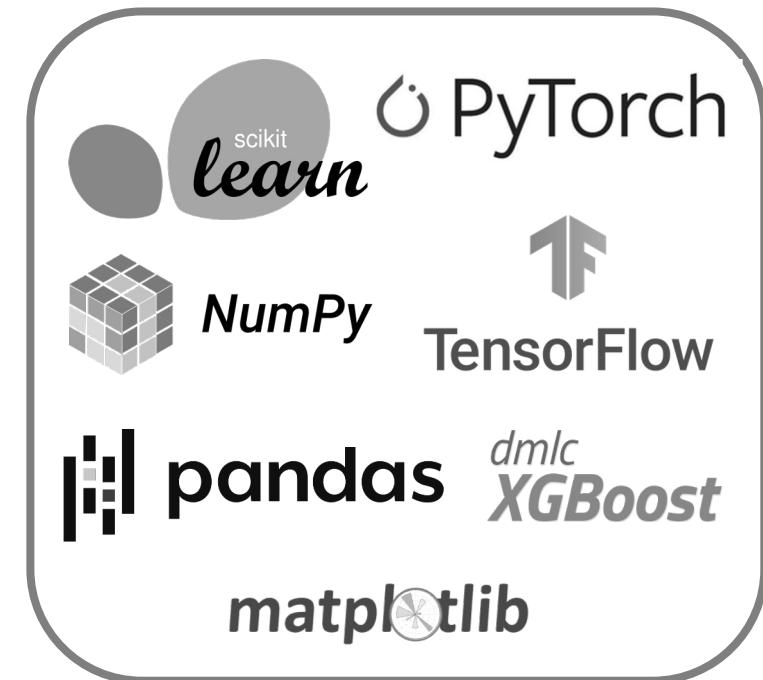
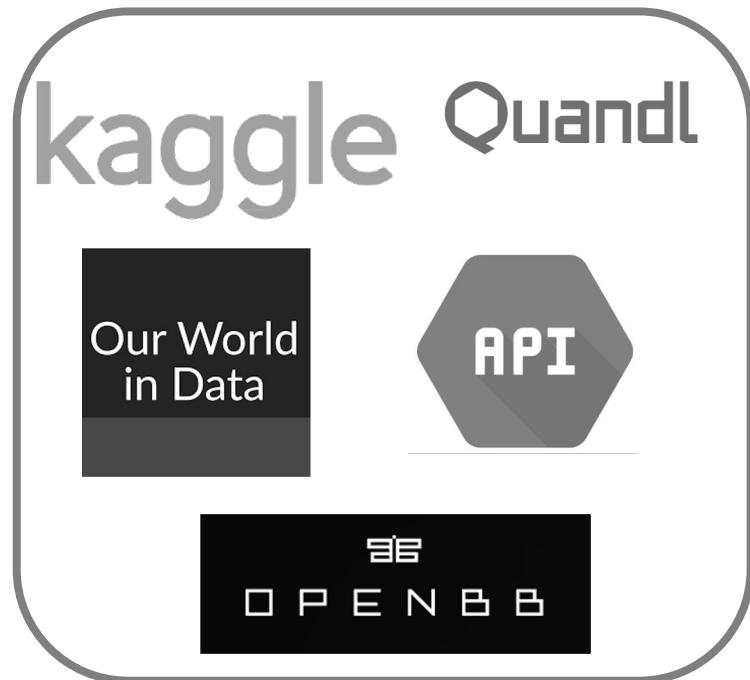
From Zero to Hero



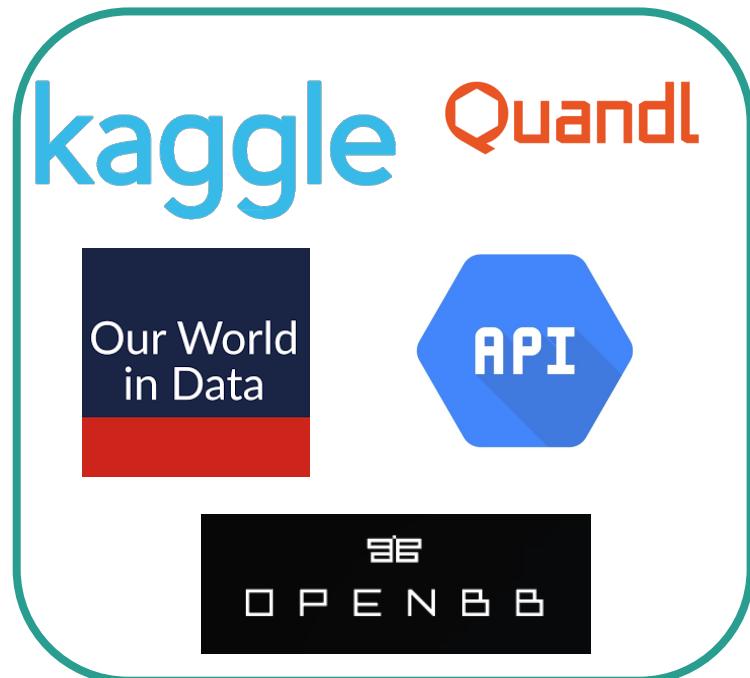
From Zero to Hero



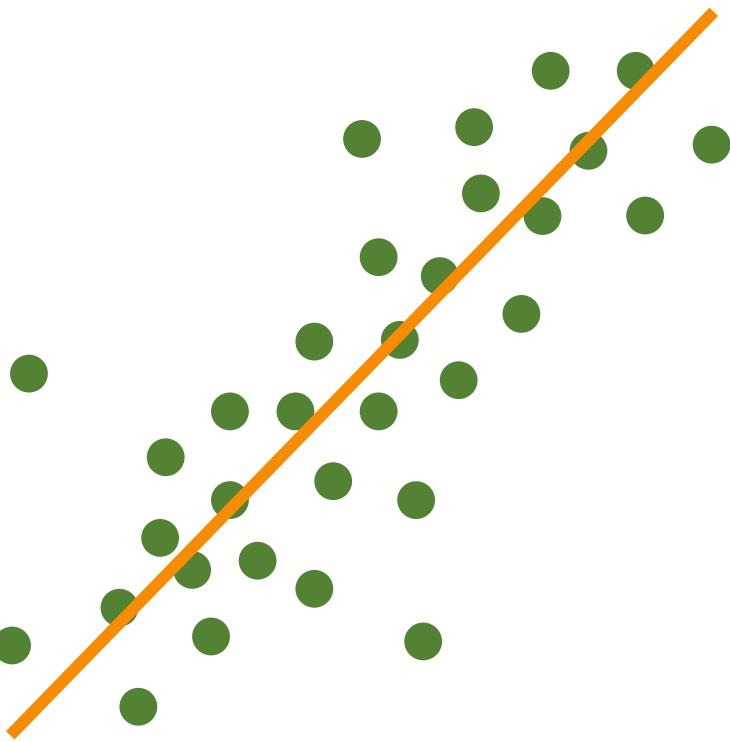
From Zero to Hero



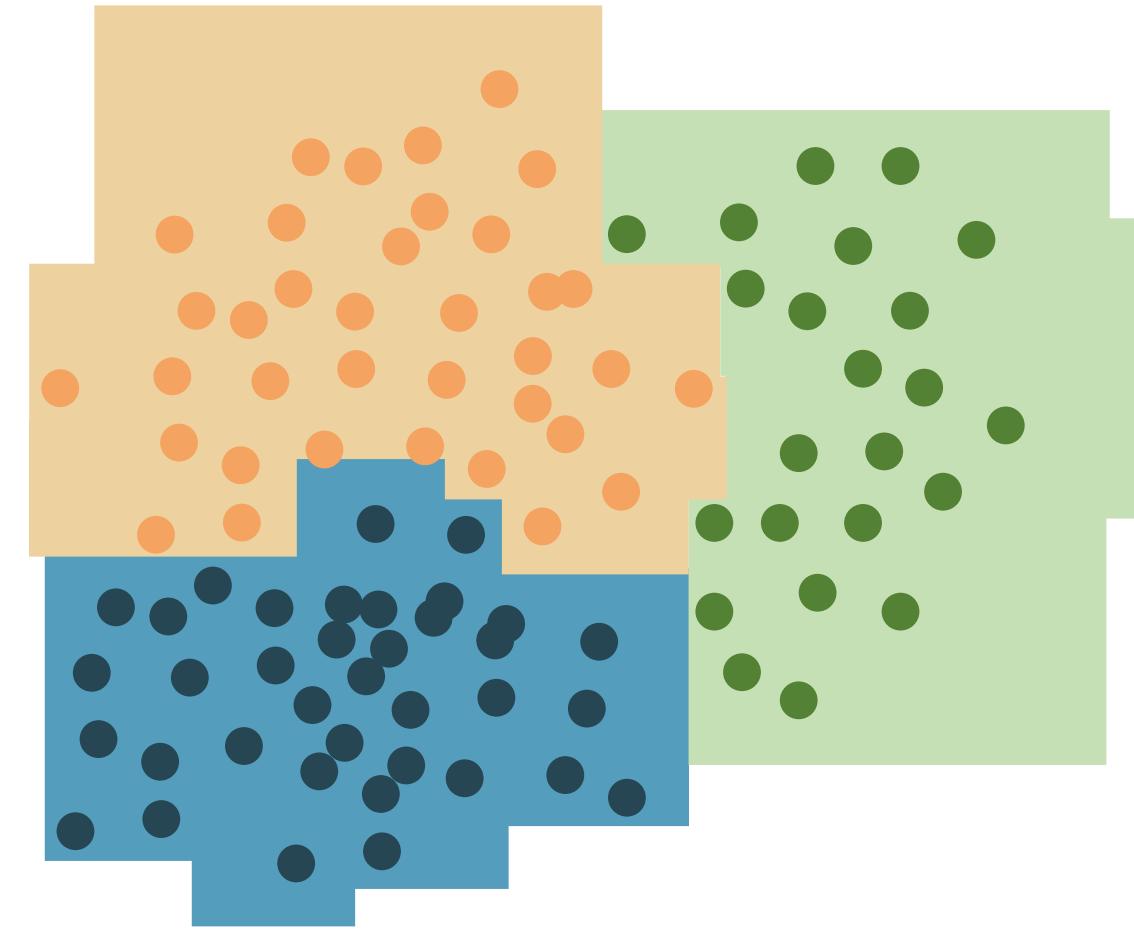
From Zero to Hero



Regresión vs Clasificación

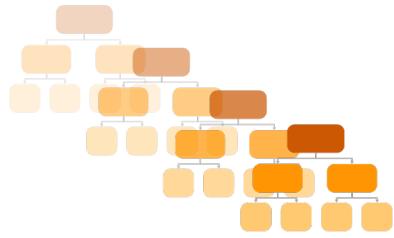


- Buscamos una recta que ajuste a una dispersión
 - y es numérica

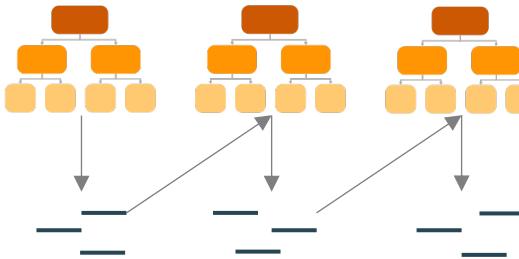


- Buscamos dividir a través de un límite
 - y es categórica

Universo de modelos



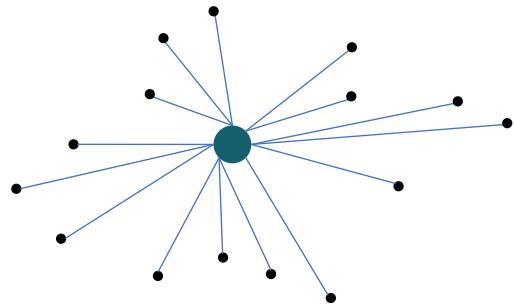
Random Forest Classifier



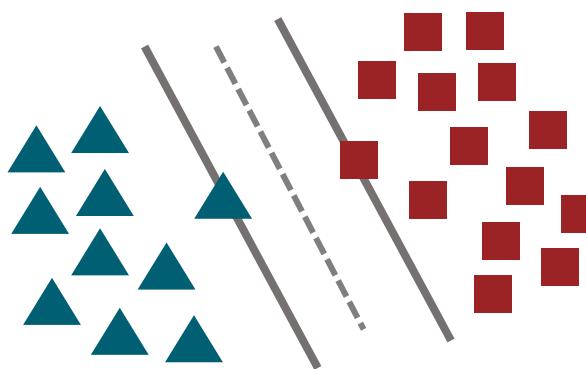
XGBoost



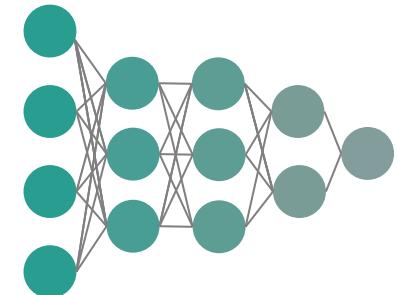
Ada Boost



K-Means



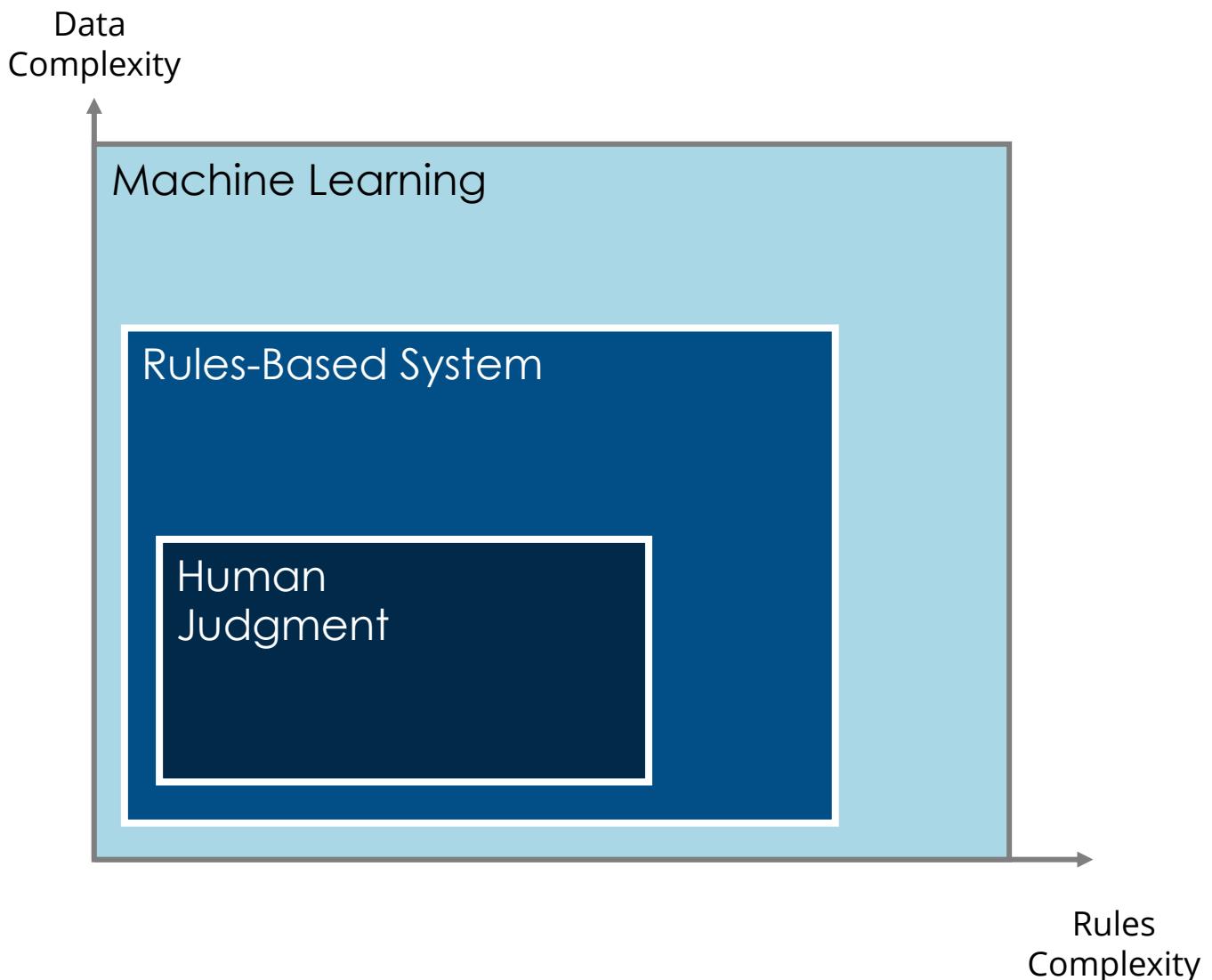
SVM



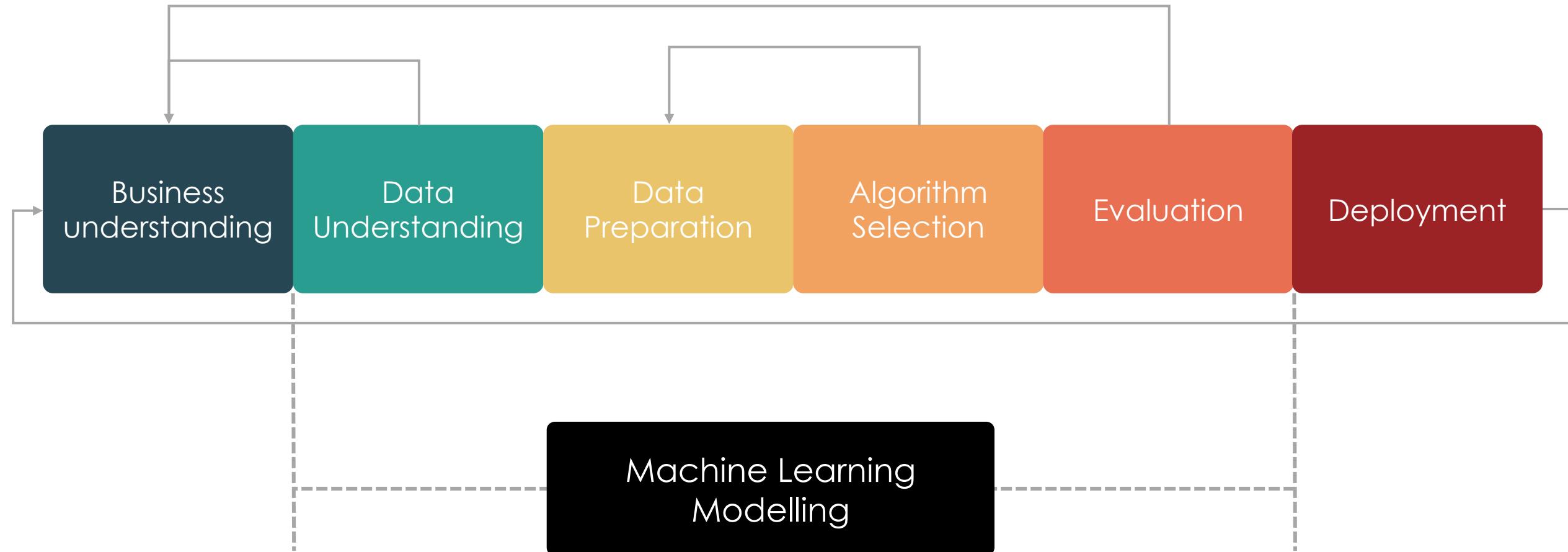
Dense Model

y muchos más...

¿Cuándo ocupar ML?



Círculo de vida de un modelo

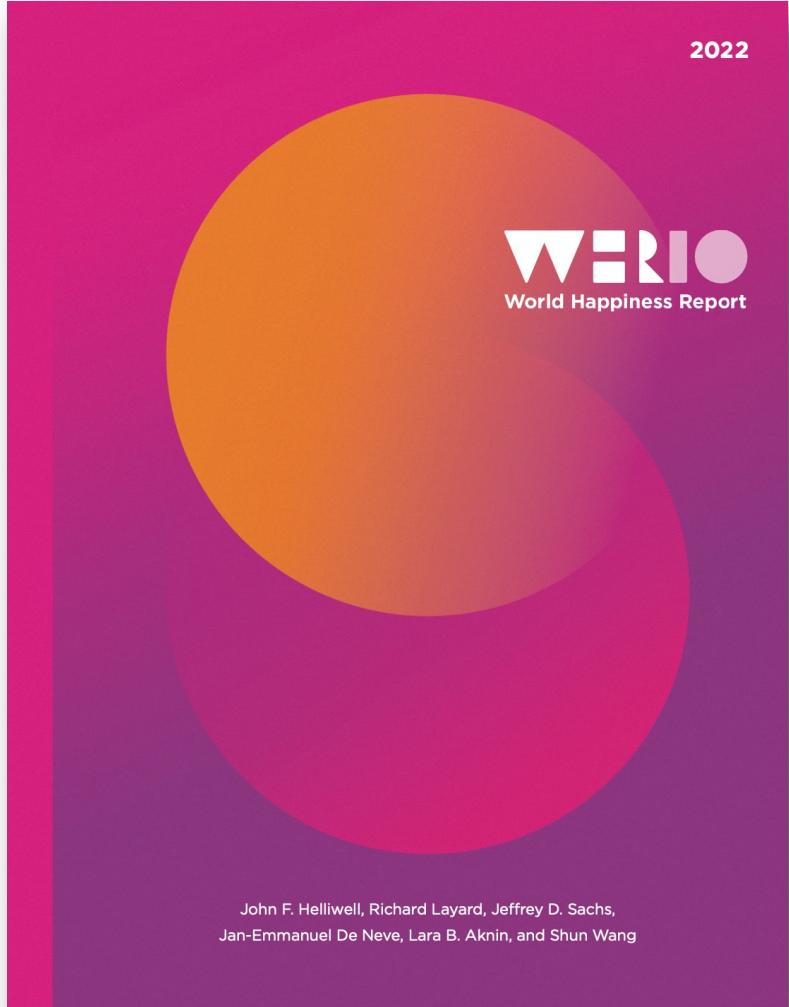


Círculo de vida de un modelo



¿El dinero **conlleva** a la felicidad?

Business Understanding



John F. Helliwell, Richard Layard, Jeffrey D. Sachs,
Jan-Emmanuel De Neve, Lara B. Akinin, and Shun Wang

Tras 10 años de publicación, este reporte pretende tomar una **visión holística** para medir la felicidad de cerca de 150 países.

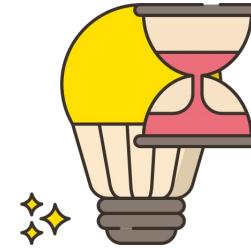
El índice de felicidad se lleva a cabo con encuestas **económicas, psicológicas y nacionales**.

Data Understanding



GDP

En términos de Paridad de Poder Adquisitivo Ajustado a 2017



Healthy Life Expectancy

Serie de tiempo de esperanza de vida al nacer



Social Support

Si estuvieras en problemas, ¿Tienes familiares o amigos con los cuales contar para pedir ayuda en cualquier momento?

1 = sí

0 = no



Freedom to make life choices

¿Estás satisfecho con la libertad de elegir qué hacer con tu vida?

1 = sí

0 = no

Data Understanding



Generosity

¿Has donado dinero a la caridad en el último mes?

1 = sí

0 = no



Perceptions of corruption

¿La corrupción se ha extendido en el gobierno de tu país?

¿La corrupción se ha extendido en las empresas de tu país?

1 = sí

0 = no



Positive affect

Promedio de medidas de risa, disfrute, o de hacer o aprender algo interesante

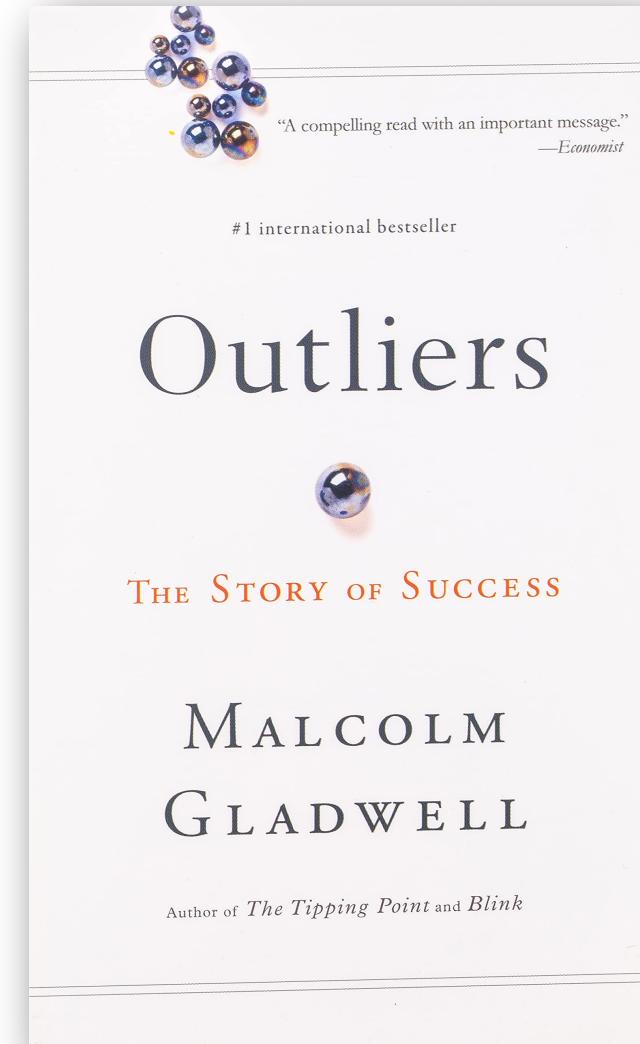


Negative affect

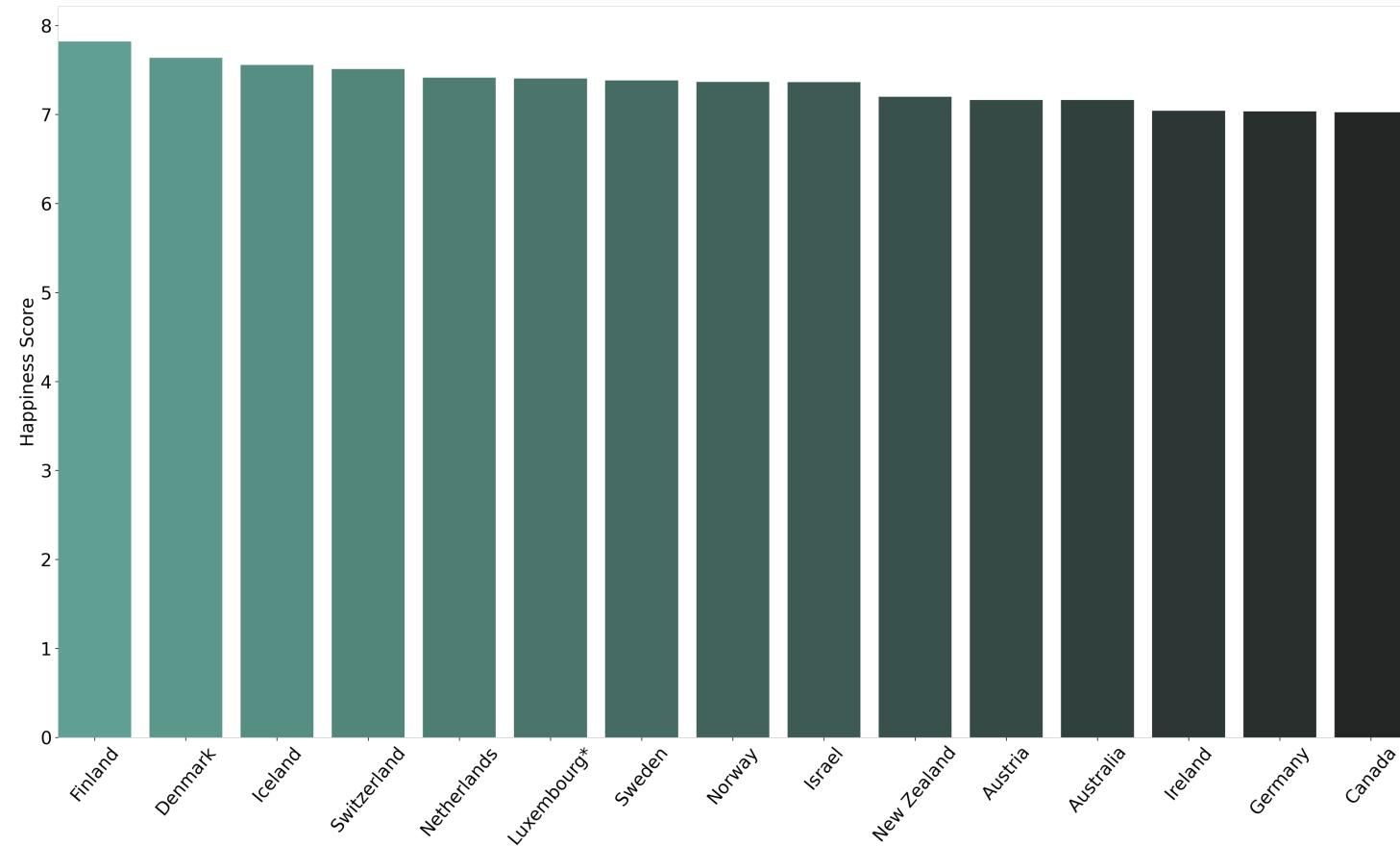
Promedio de medidas de preocupación, tristeza y enojo

Opcional: Outliers

¿Qué variables impactarían más en un *happiness score* alto?



Insights



146
países

5.55
Promedio de
happiness score

2.4 - 7.8
Rango

Insights

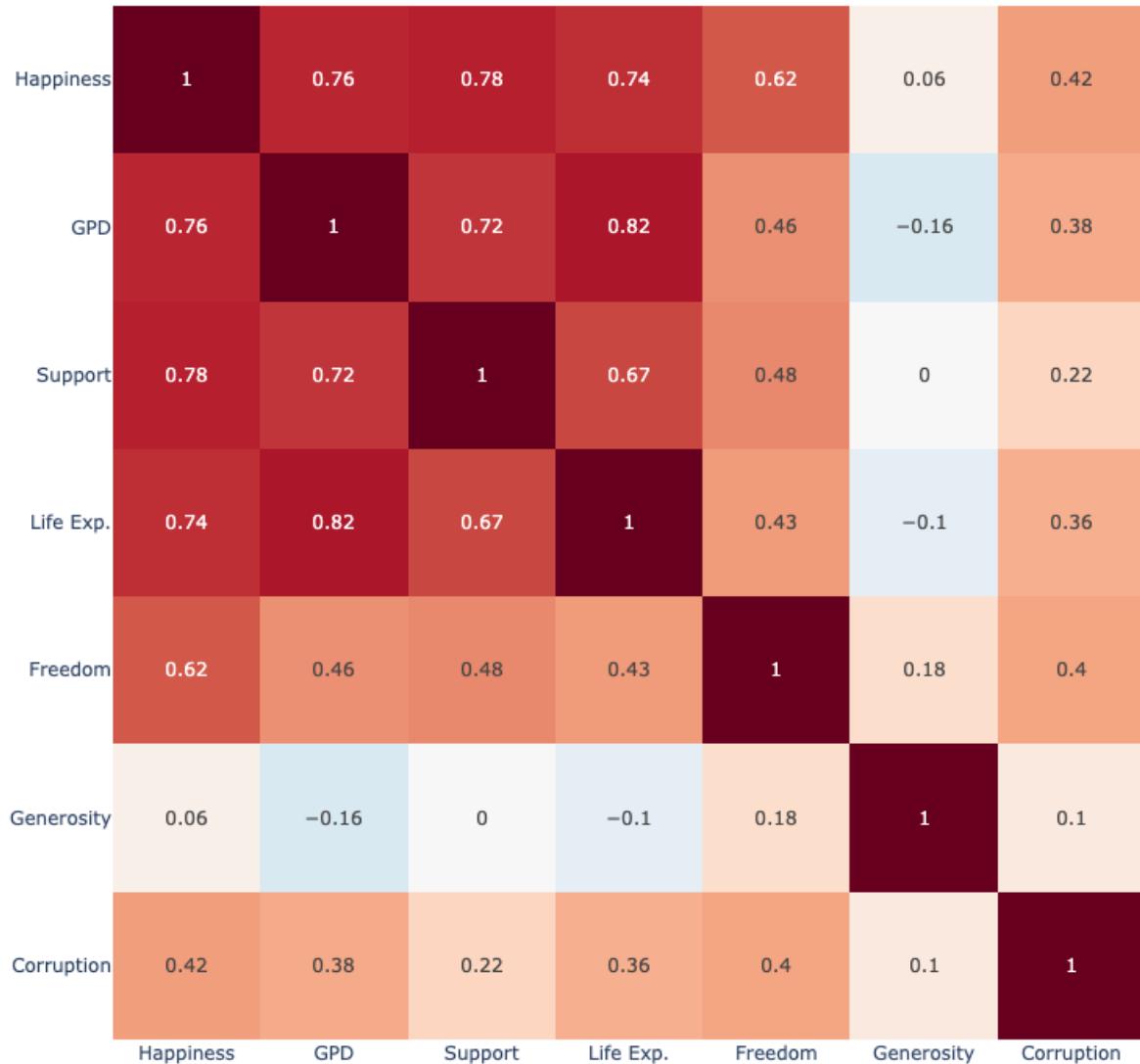
Linealmente, si el PIB crece, también lo hará la felicidad, y de forma inversa. Lo mismo ocurre con poblaciones con un alto soporte emocional.

Donar dinero a la caridad y tener un soporte emocional no están correlacionadas en lo absoluto.

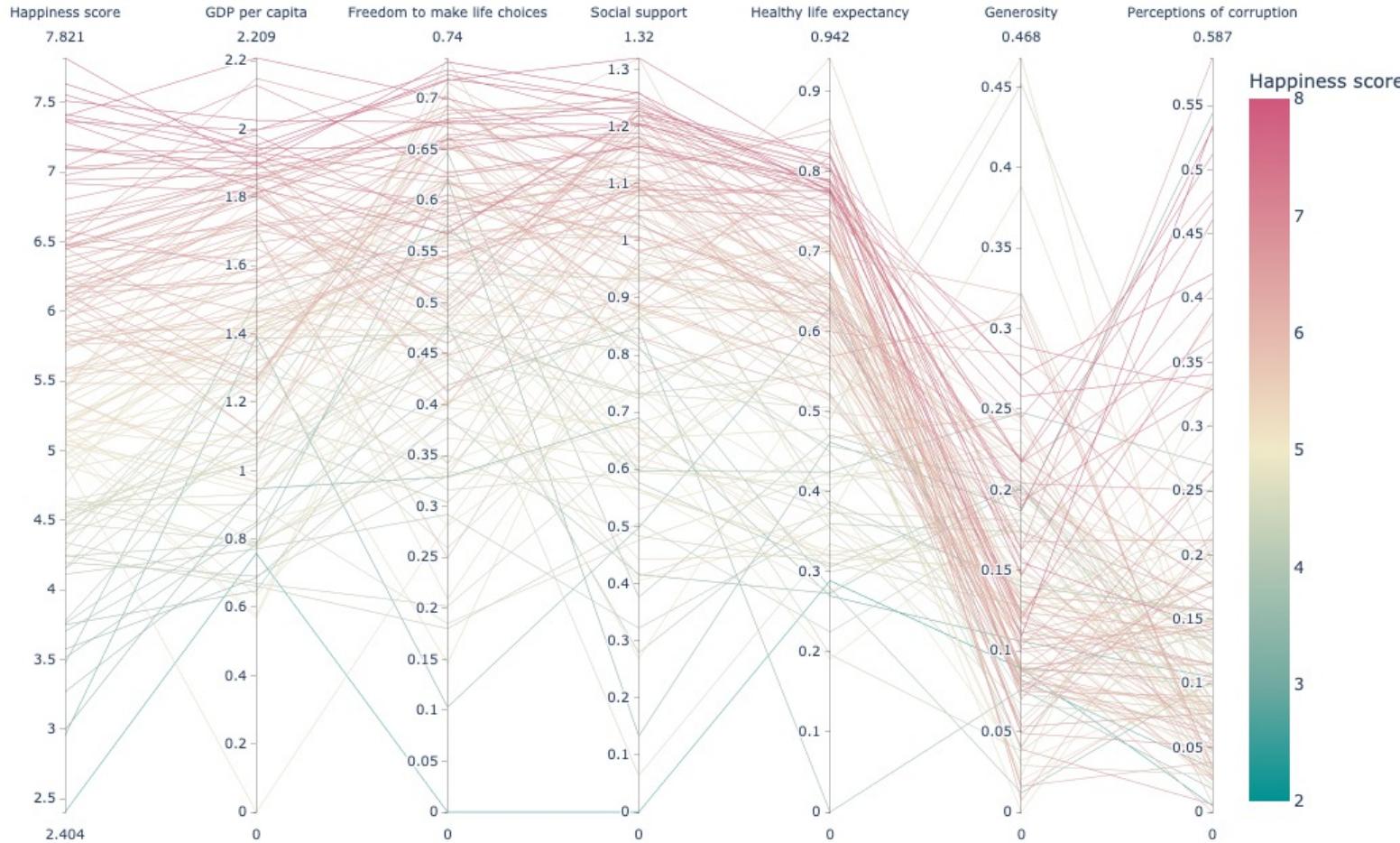
Que el país/empresas sean corrupto no se relaciona con donar.

Cuando el PIB crece, se espera que la generosidad disminuya.

Nota
Correlación ≠ Causalidad



Insights

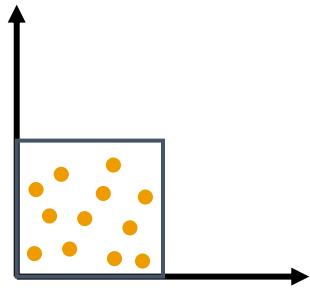


Un plot de coordenadas paralelas nos ayuda a graficar **más de tres dimensiones**.

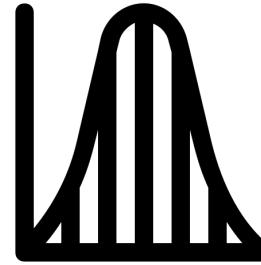
El **cruce** de valores de felicidad bajos con valores altos nos indica **linealidad** entre ciertas variables.

Esperanza de vida conlleva a valores bajos de **generosidad**, mientras que este conlleva a valores altos de **corrupción**.

Data Preparation



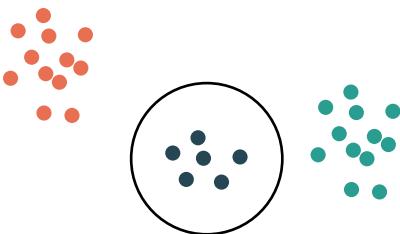
Normalización



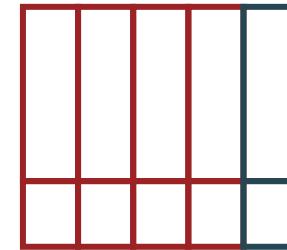
Estandarización



Concatenación

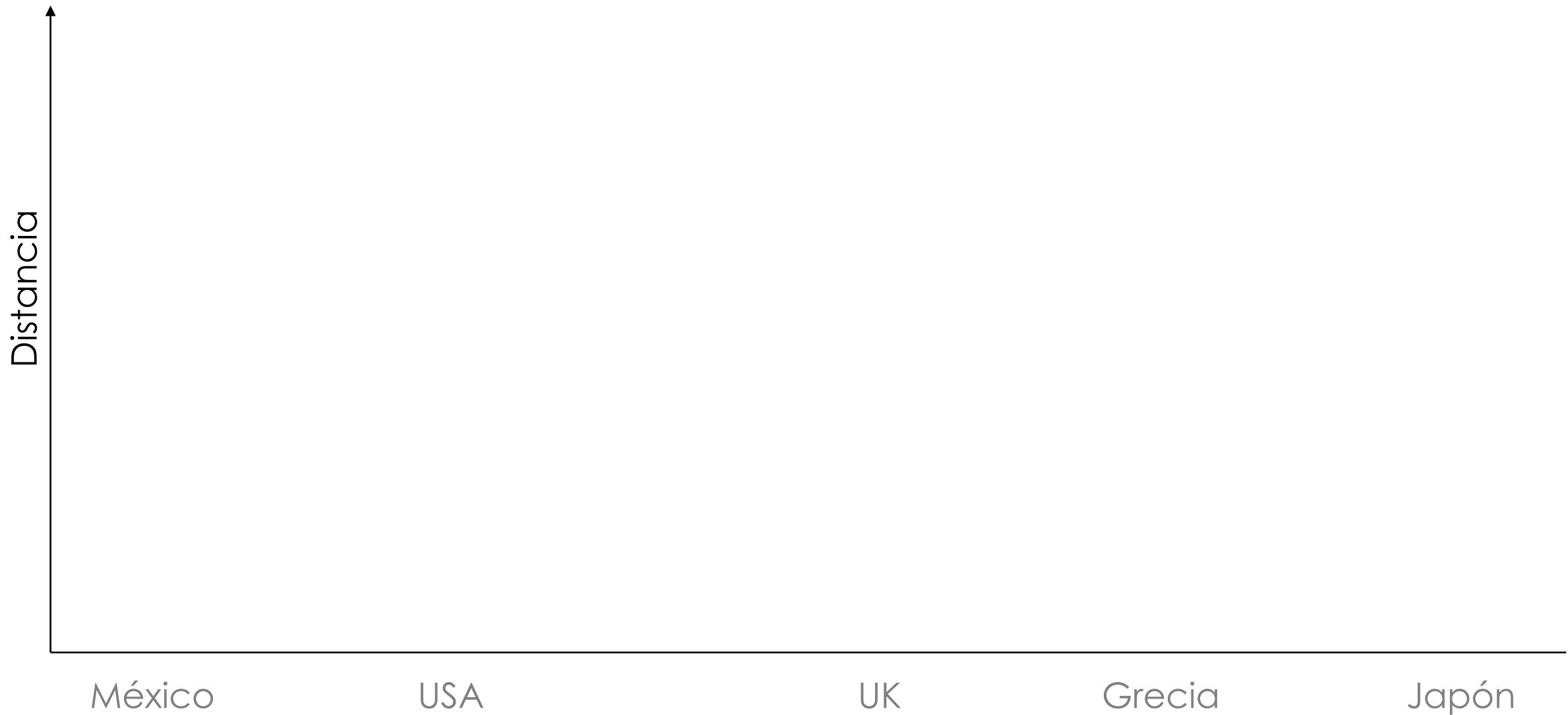


Datos sintéticos

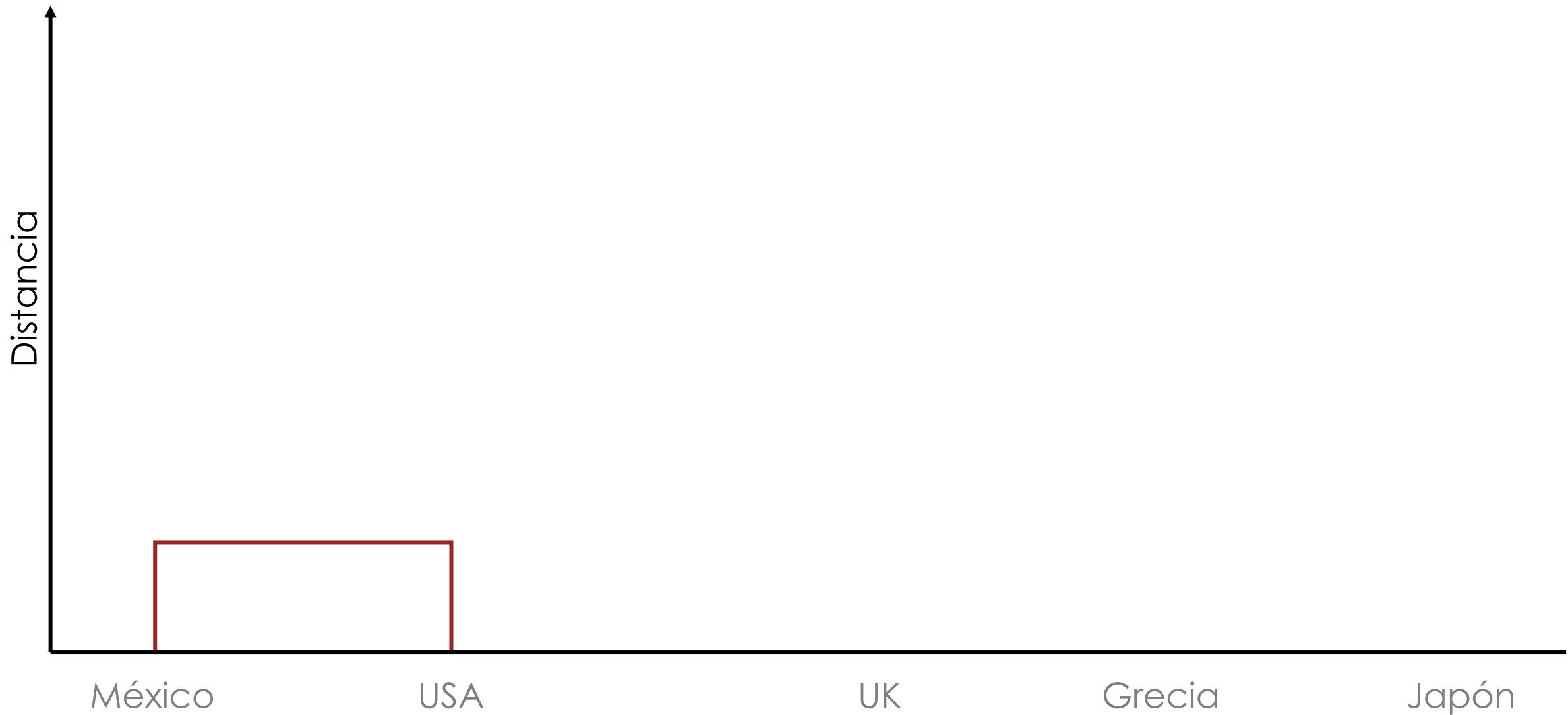


Particiones

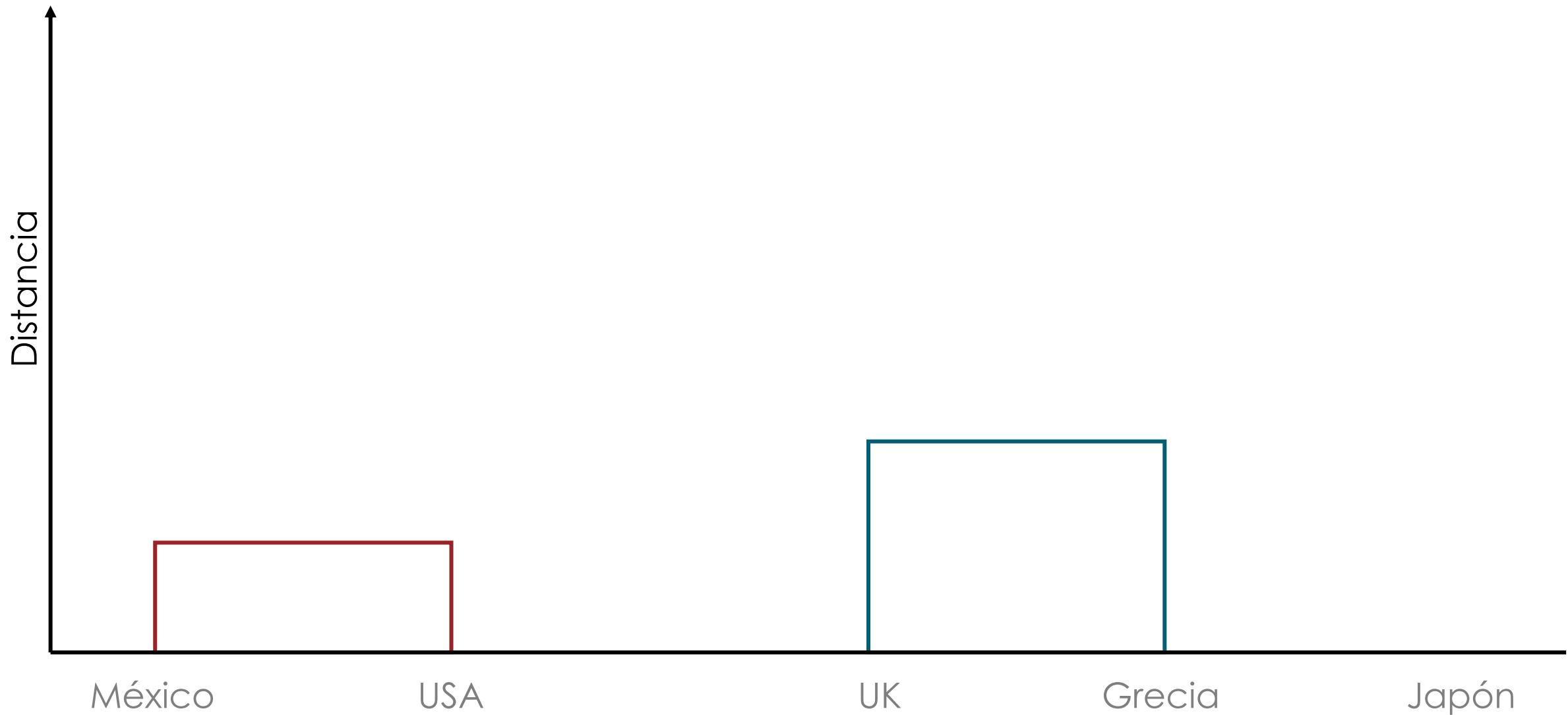
Hierarchical Clustering



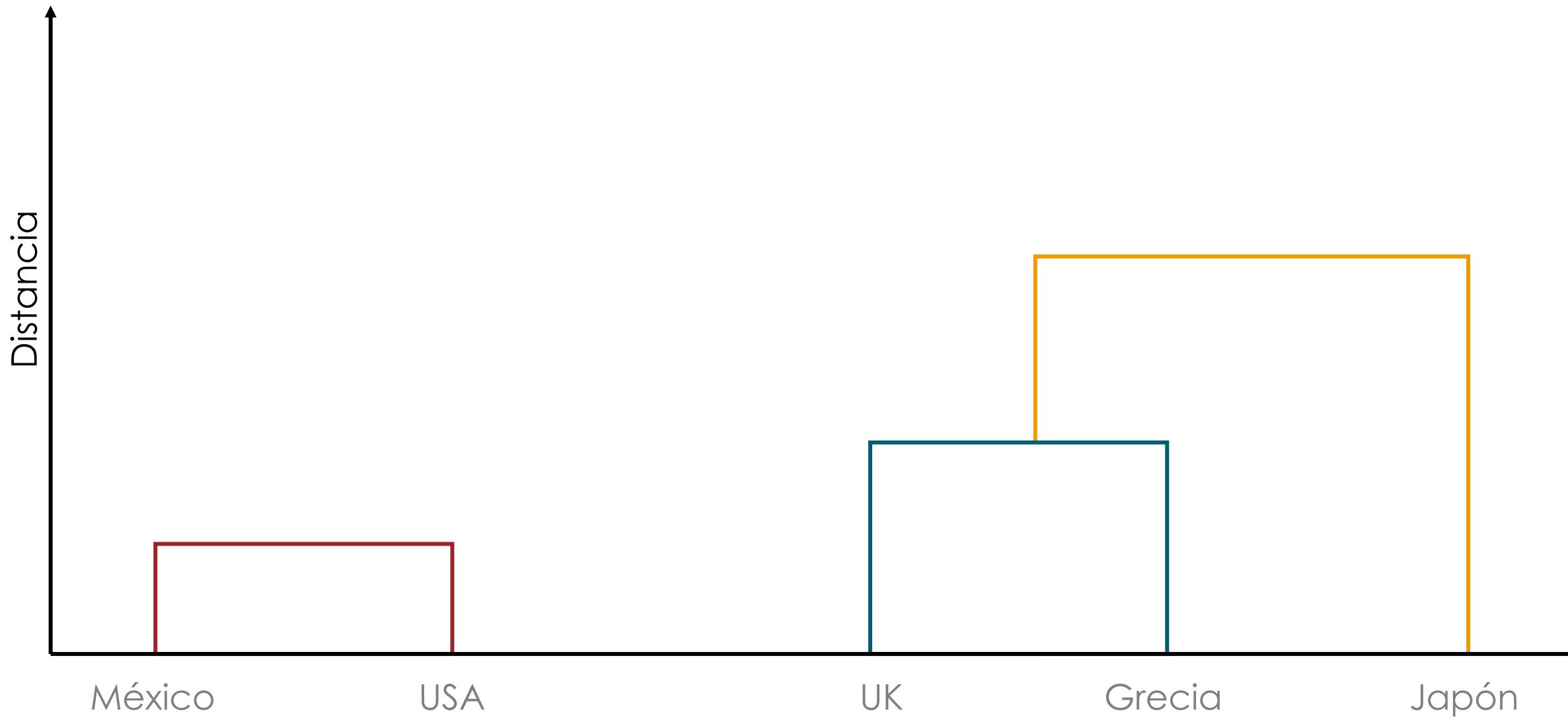
Hierarchical Clustering



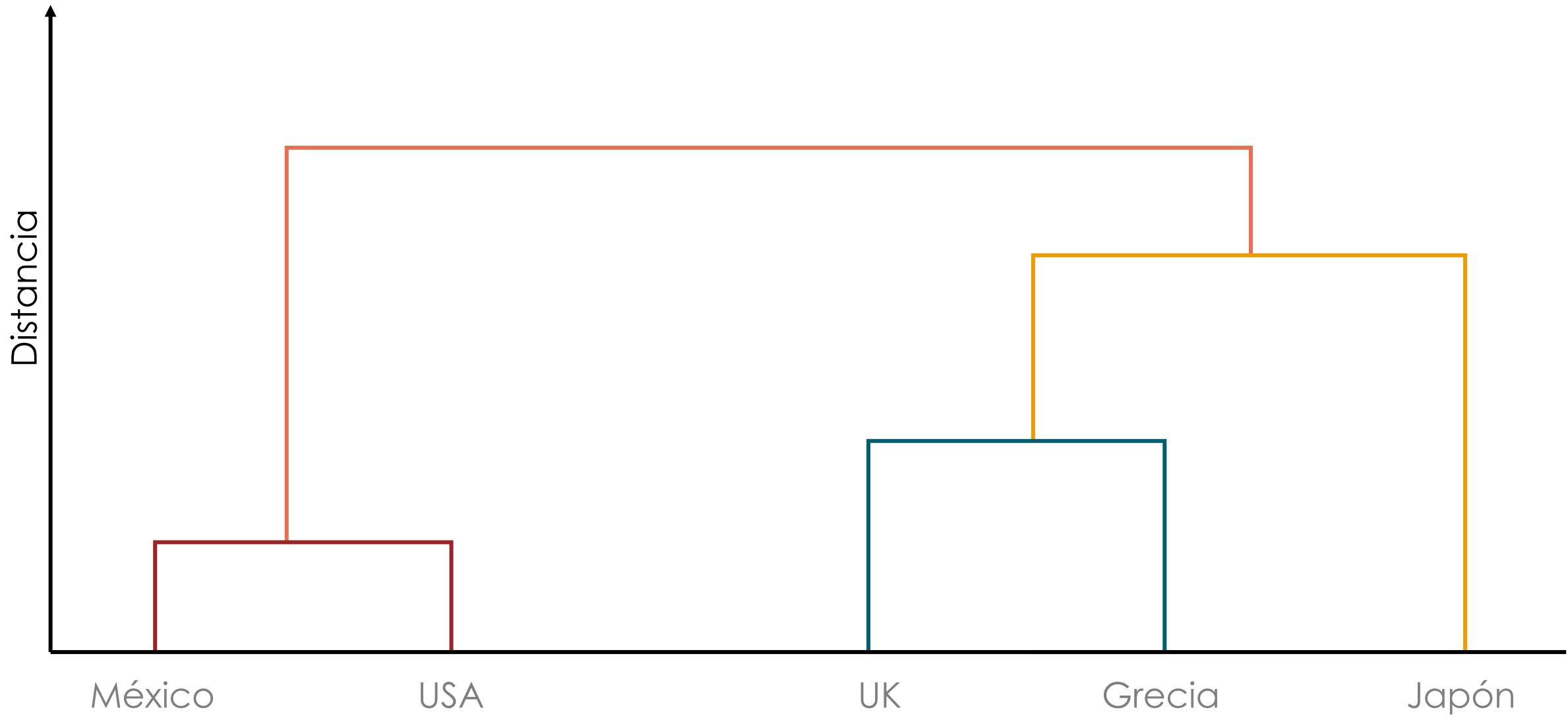
Hierarchical Clustering



Hierarchical Clustering



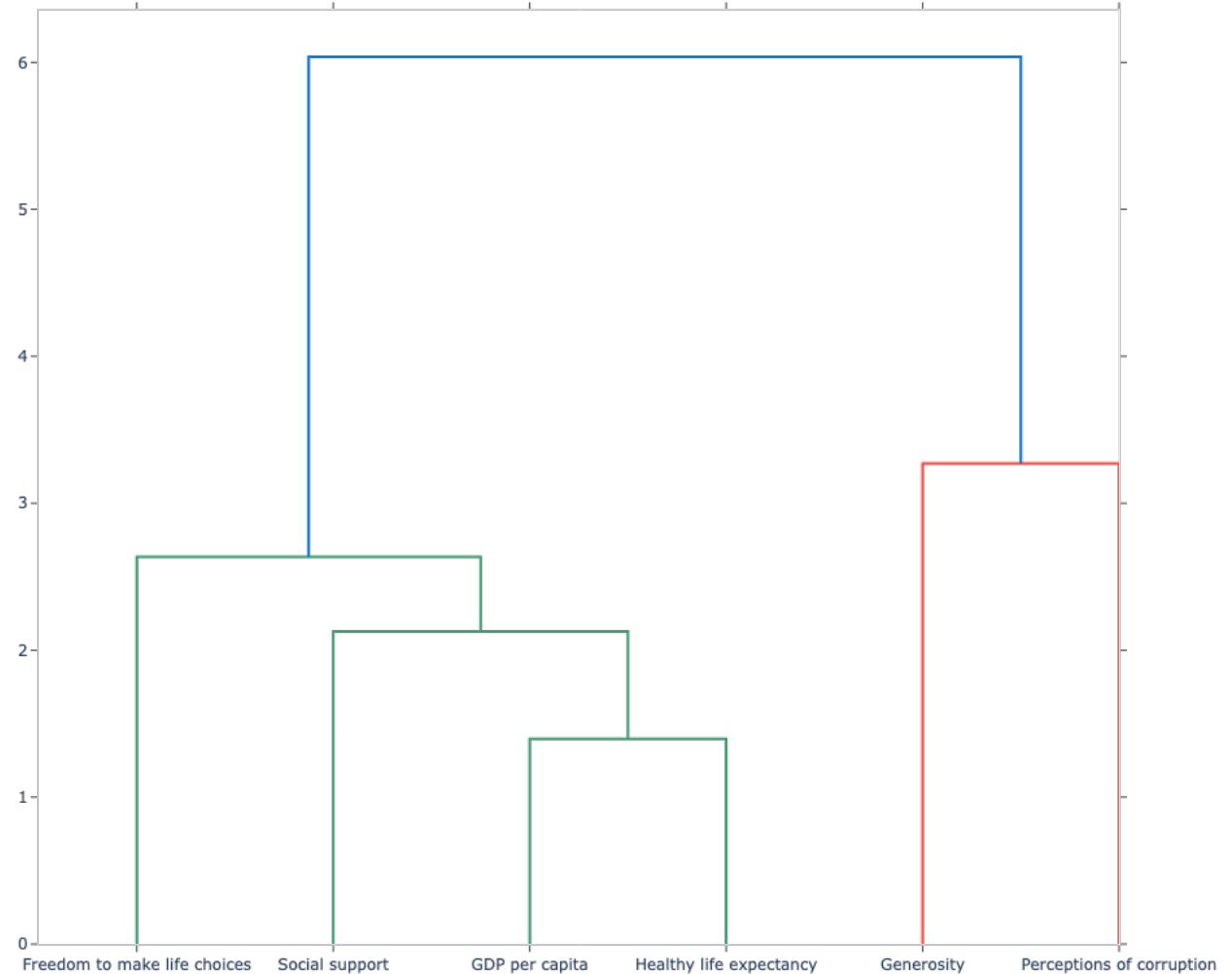
Hierarchical Clustering



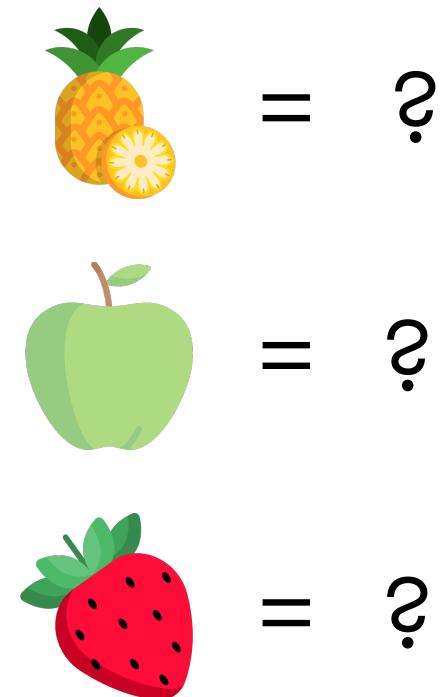
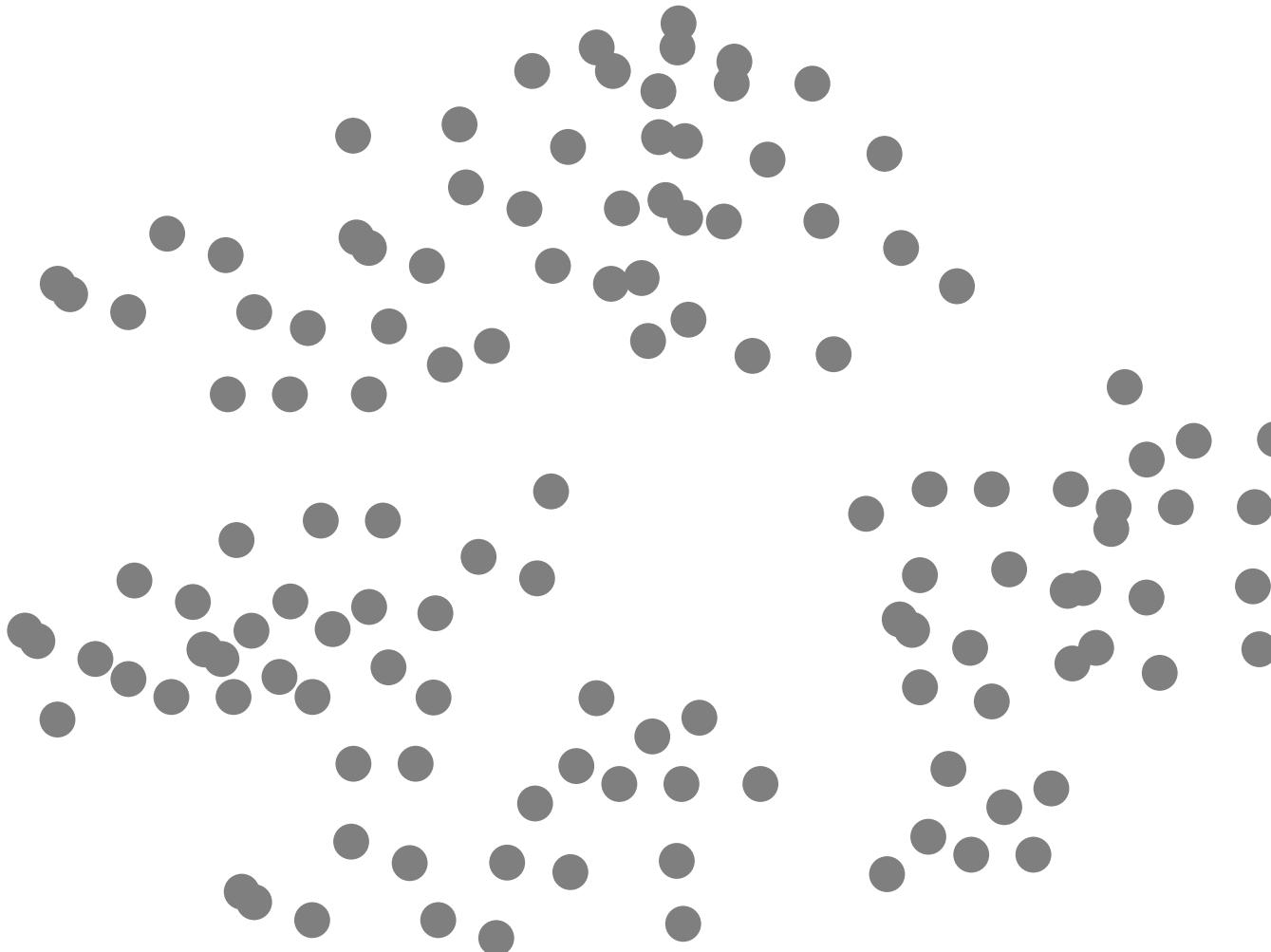
Hierarchical Clustering

El PIB y la esperanza de vida conforman un **primer cluster**, representando una relación entre variables.

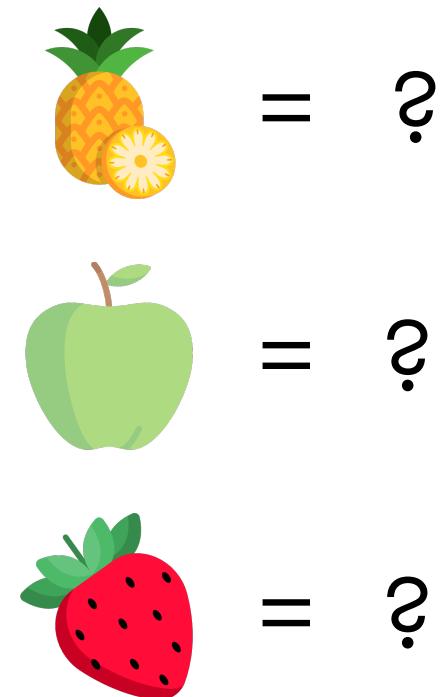
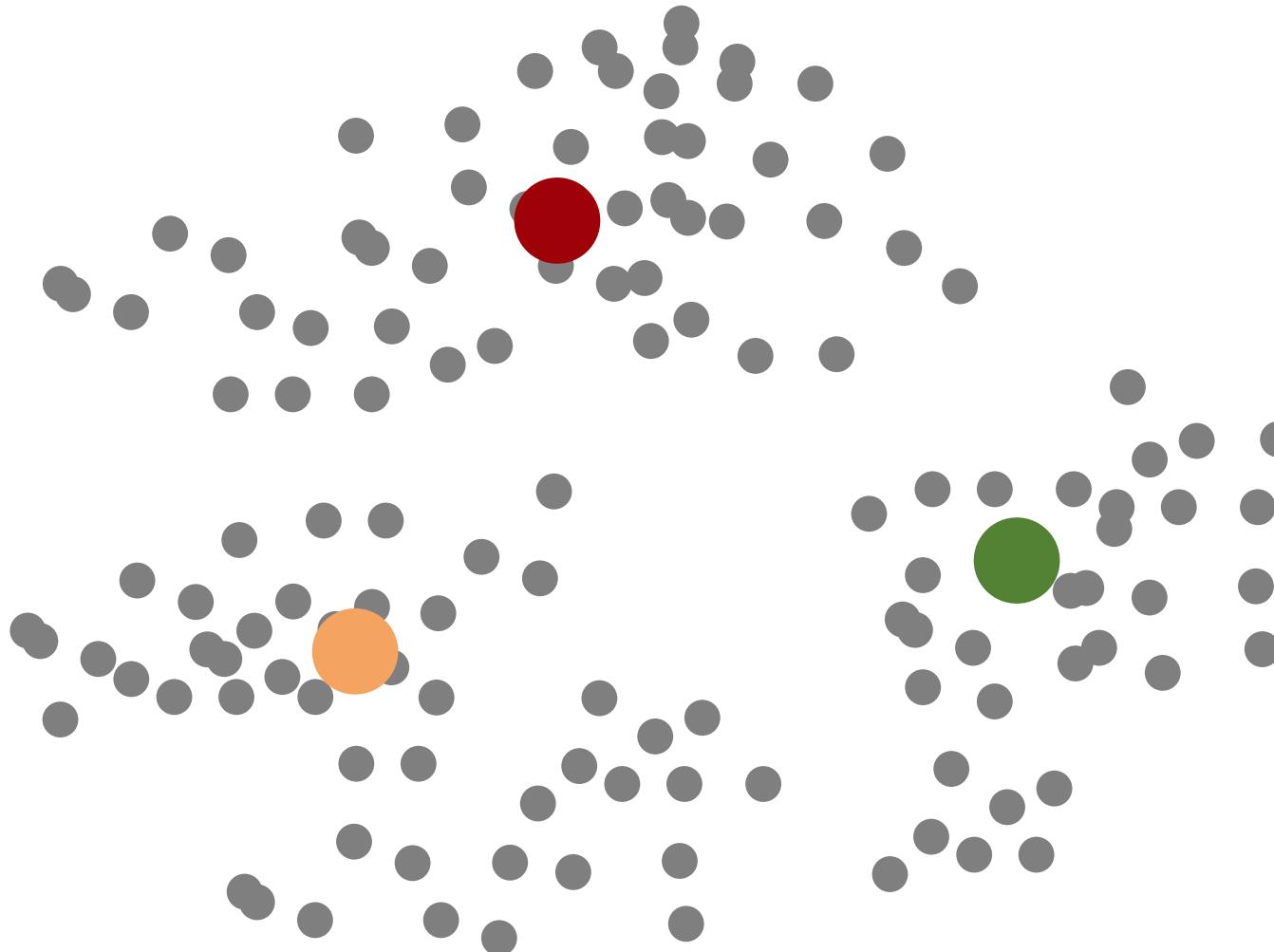
Generosidad y corrupción se encuentran en un mismo cluster pero **alejados** del resto de variables.



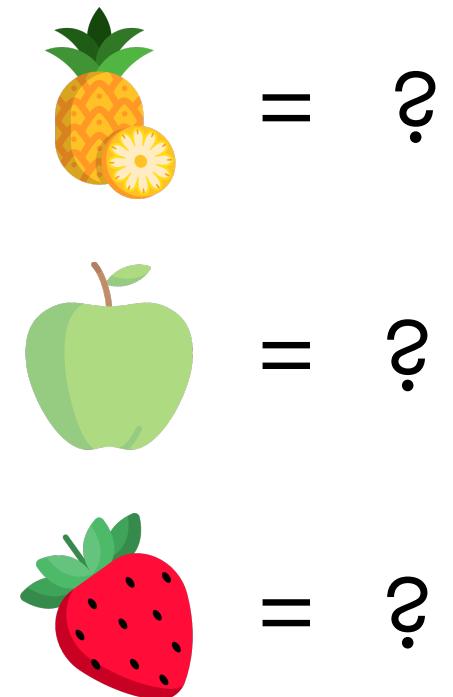
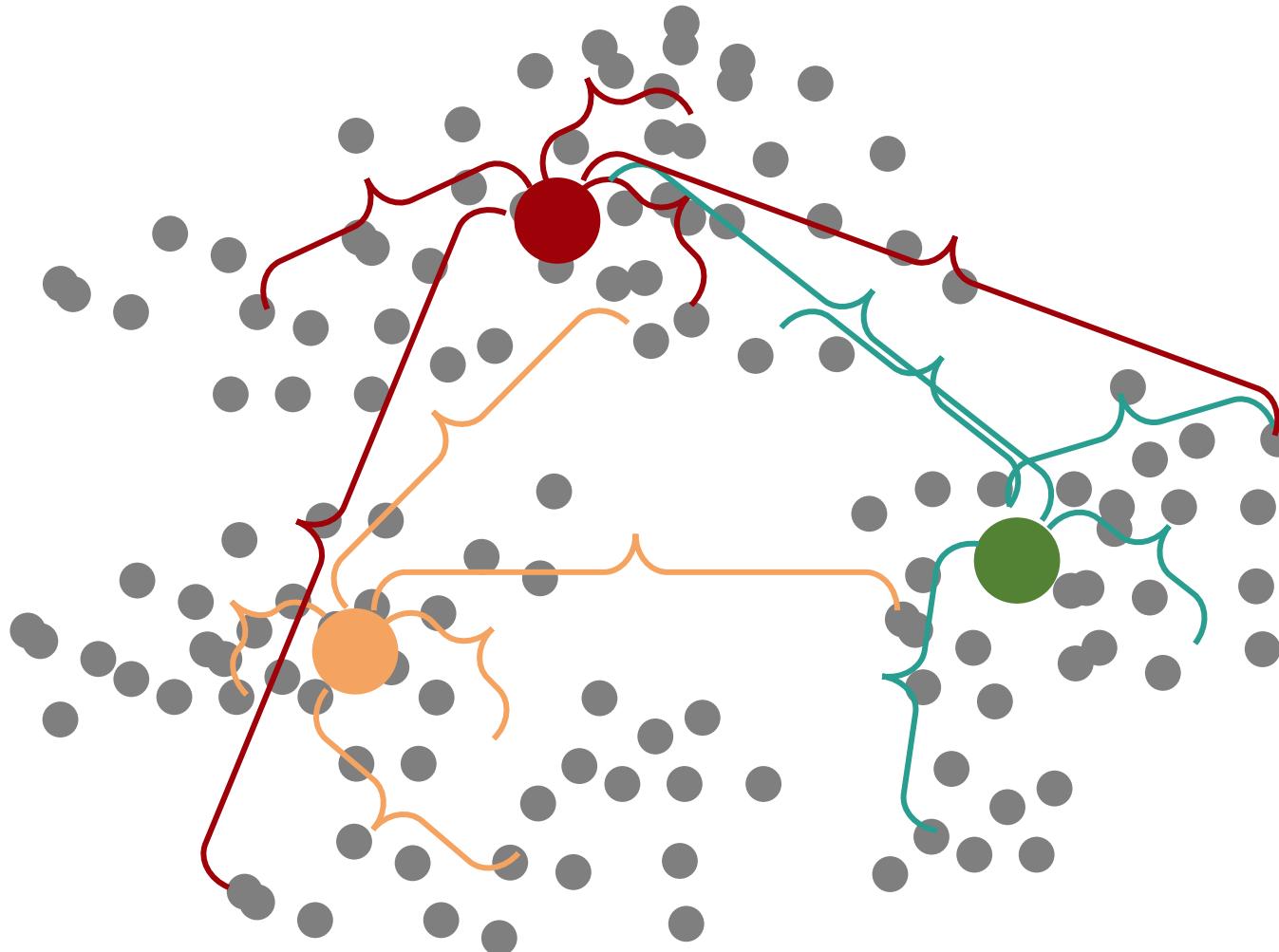
K-Means



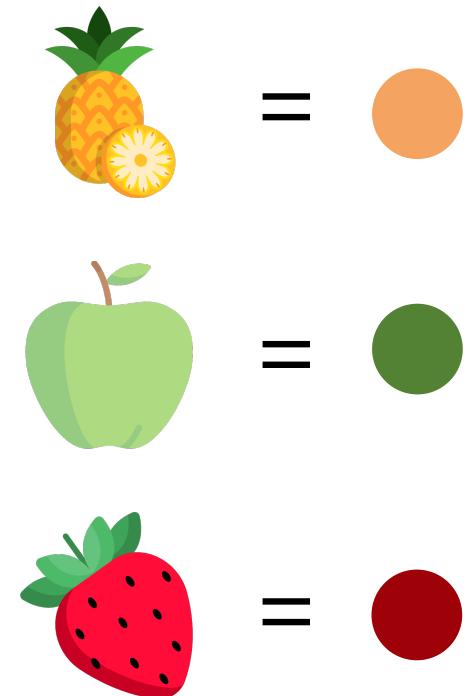
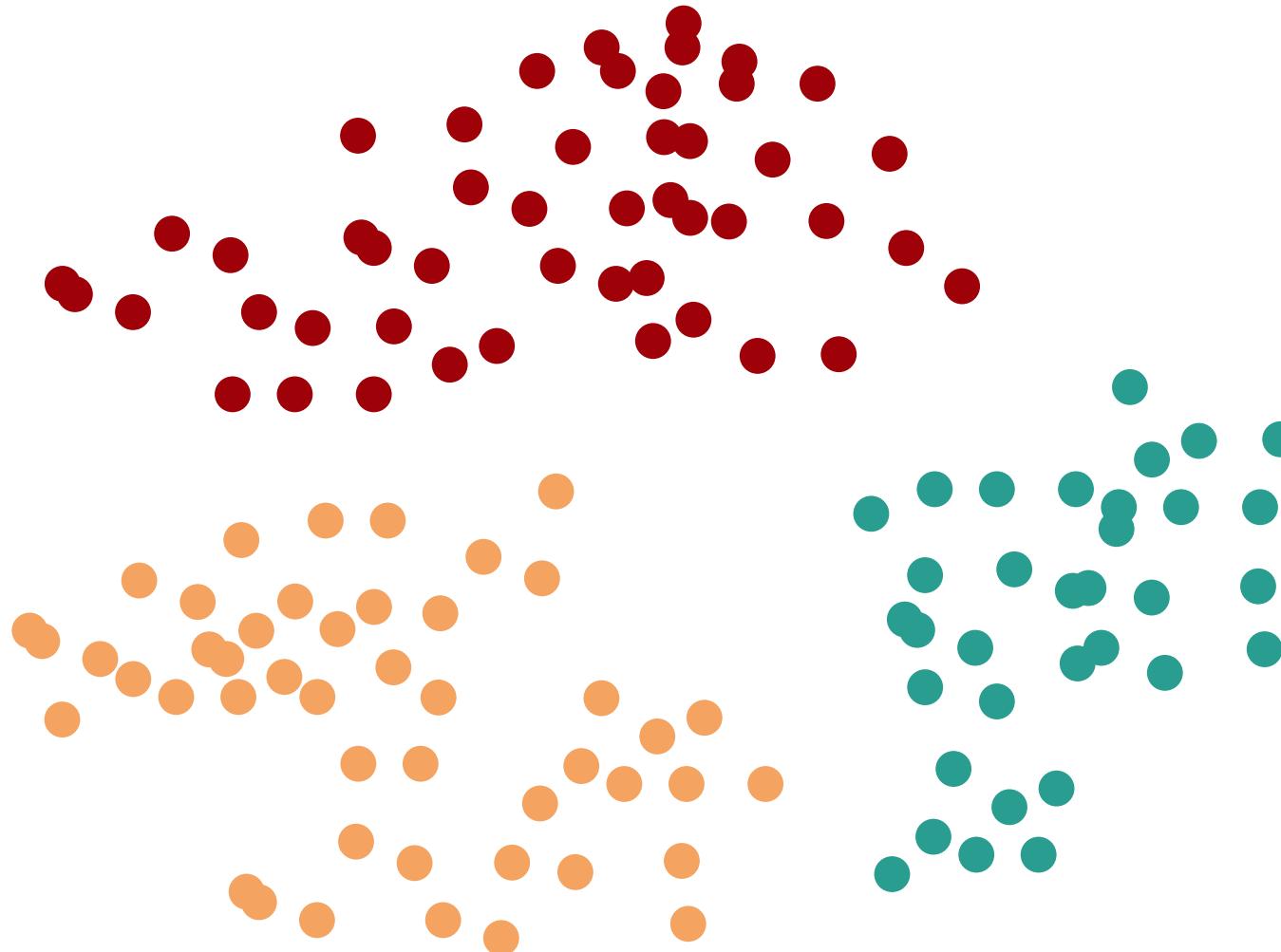
K-Means



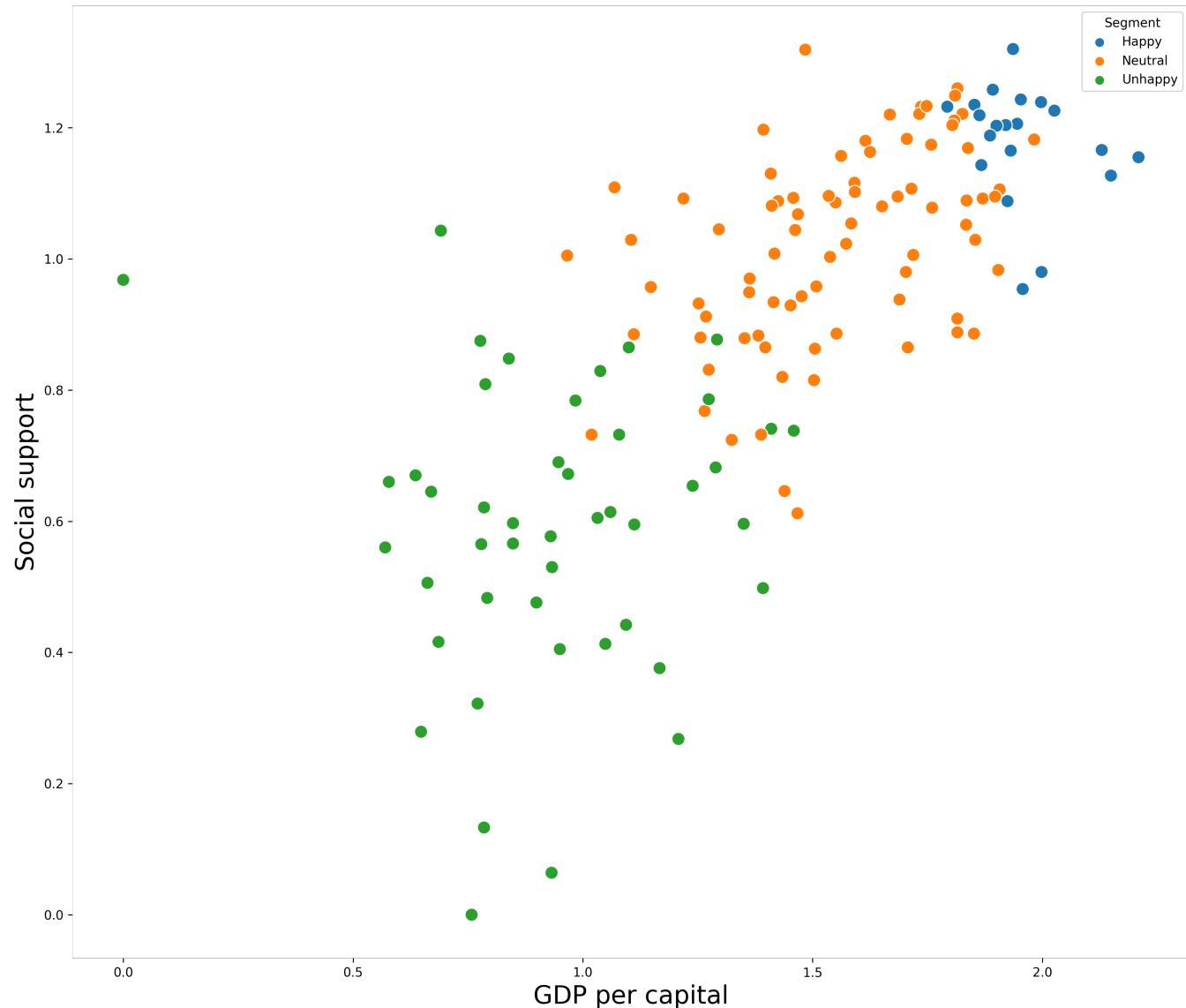
K-Means



K-Means



K-Means



Este modelo responde a la pregunta de

¿Qué pasaría si tuvieramos todos los datos, excepto el que tratamos de predecir (Happiness score)?

Con k=3, el modelo logra entender de forma eficiente que existen tres clusters y que a mayor valor de las variables, mayor felicidad

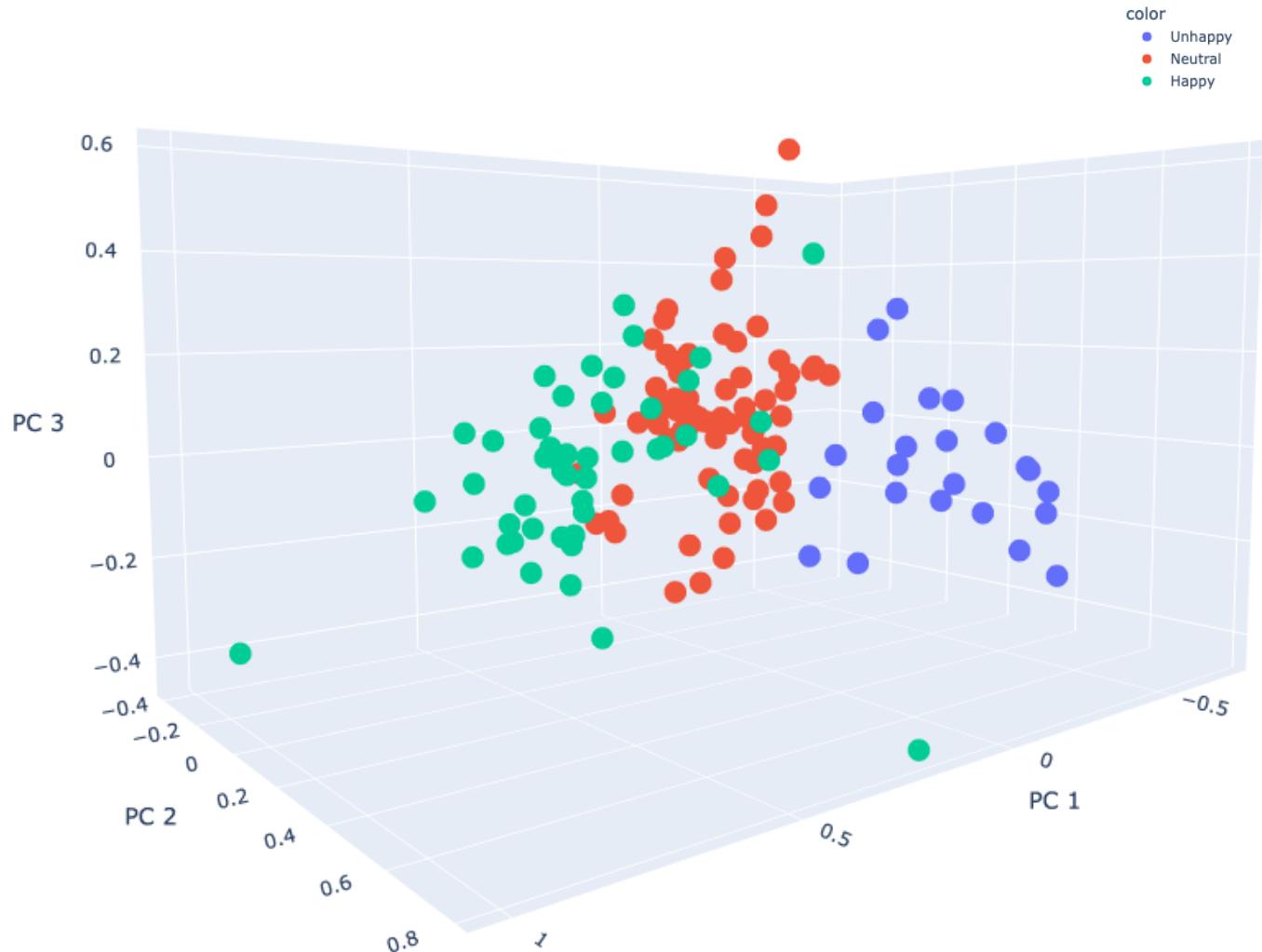
PCA

Q: ¿Qué puedo hacer si tengo 8 variables significativas y no puedo meterlas todas en un modelo?

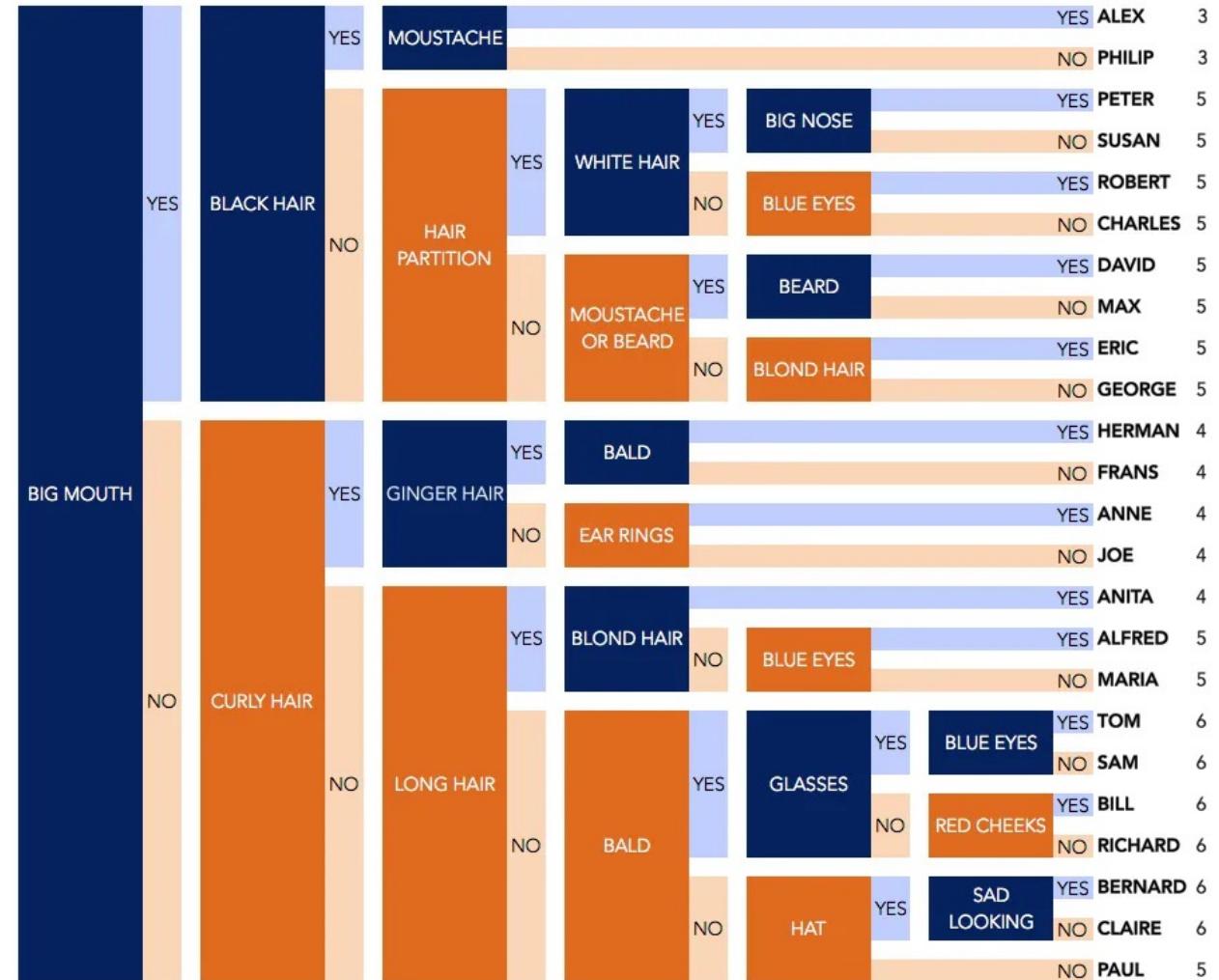
A: Reducción de la dimensionalidad

Un *Principal Component Analysis (PCA)* enfatiza sobre la variación de datos para encontrar patrones, reduciendo las 8 variables en n número menor de componentes.

Básicamente, contrae la información en su punto más esencial.



Random Forest y XGBoost



Random Forest y XGBoost



Sam



Robert



Anne



Sam

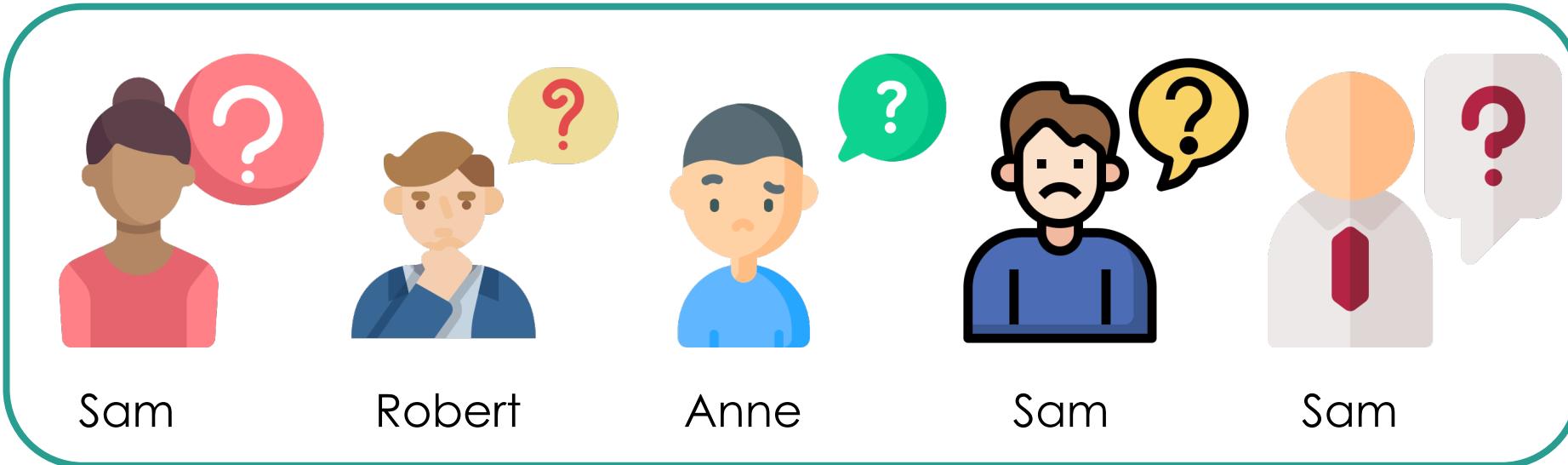


Sam

Random Forest y XGBoost

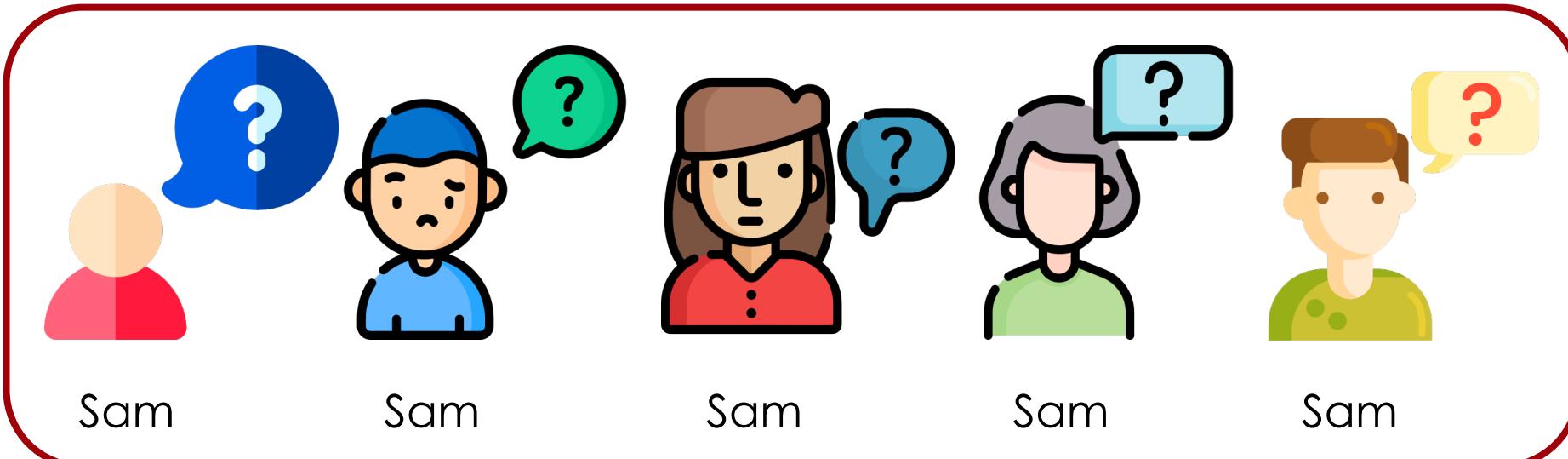
Random Forest

- Sobreajuste
- Desbalanceo
- Preferencia en hiperparámetros

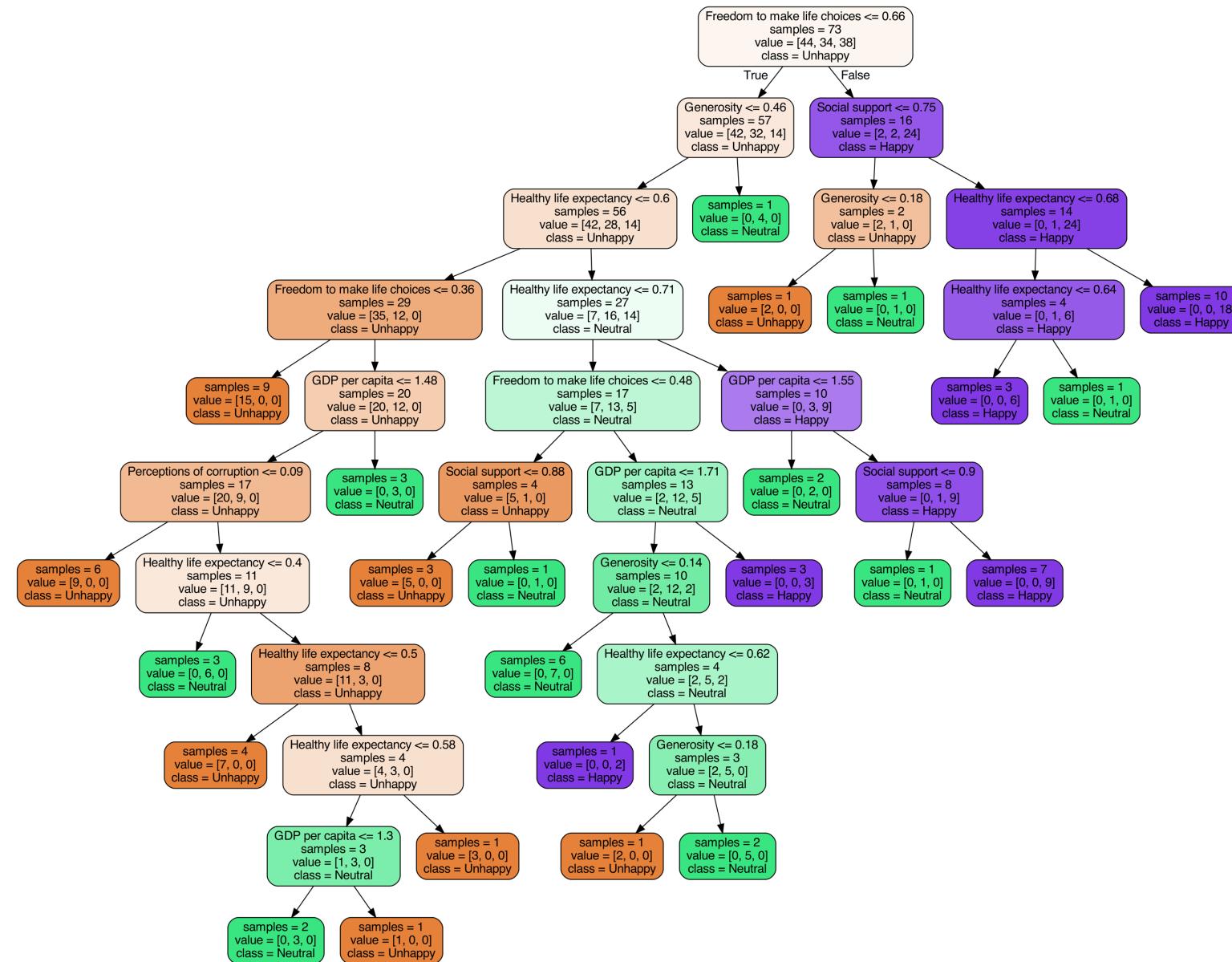


XGBoost

- Limpieza
- Preferencia en espacios funcionales



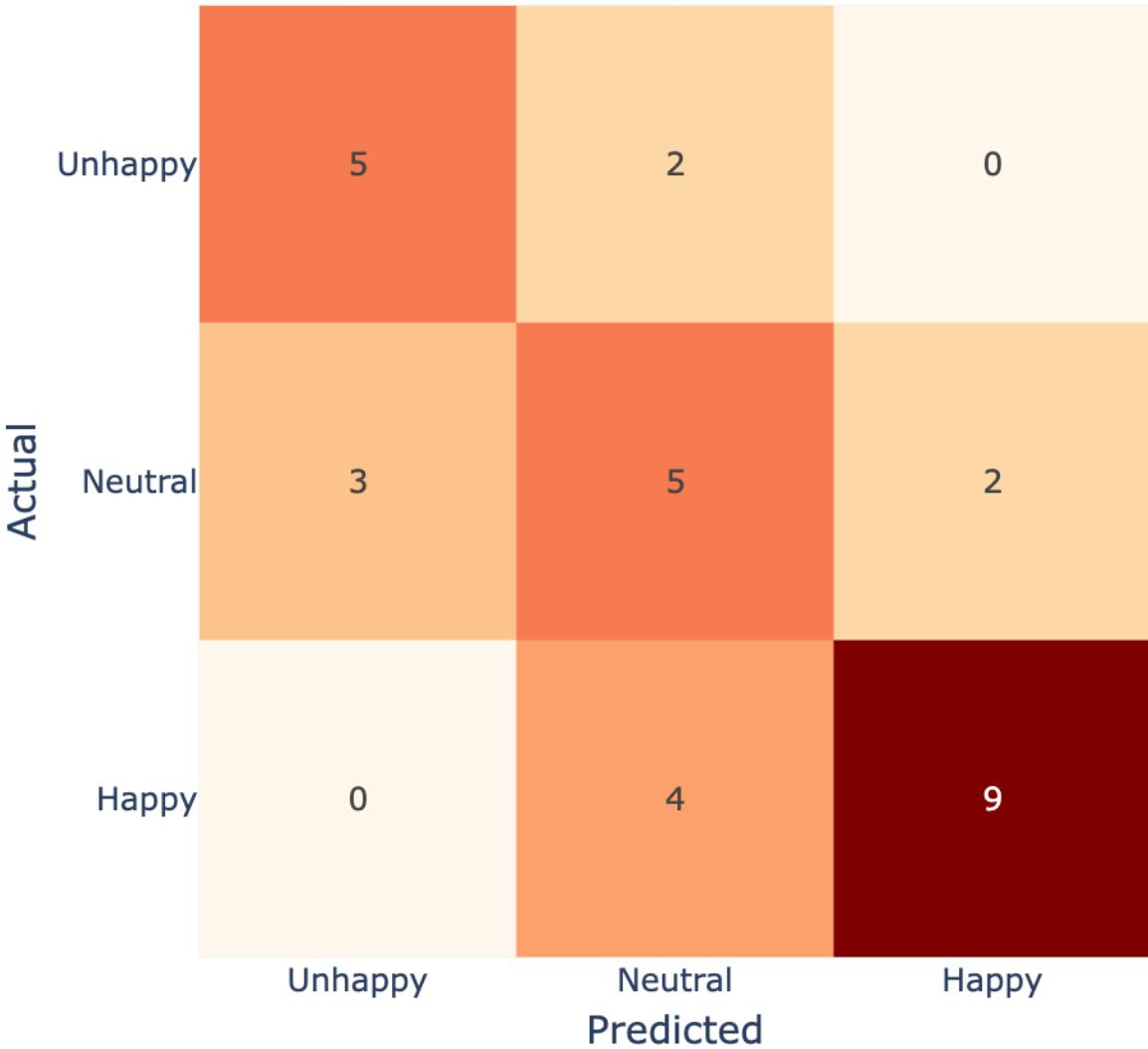
Random Forest y XGBoost



Lo primero que uno se debe de preguntar, es qué tan **libre** de **decidir** sobre tu vida eres.

El **secreto de la felicidad** (según uno de nuestros árboles) es tener un soporte emocional y una esperanza de vida...el **dinero sobra**.

¿Qué tanto se equivoca mi modelo?

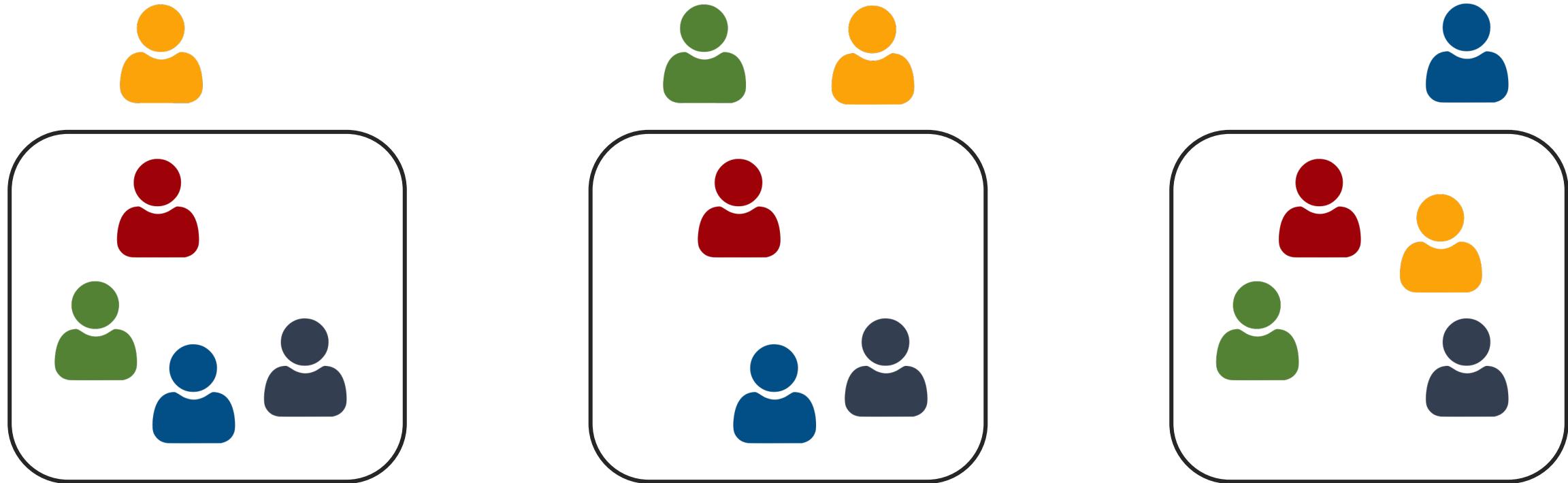


¿Qué preferimos?

¿Qué nuestro modelo se equivoque con un falso positivo?

¿O con un falso negativo?

Explainable Artificial Intelligence (XAI)



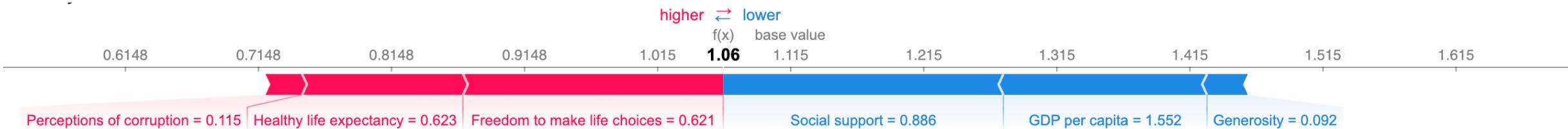
Teoría de juegos cooperativa

Responde a la pregunta de qué pasa si quito o añado una variable

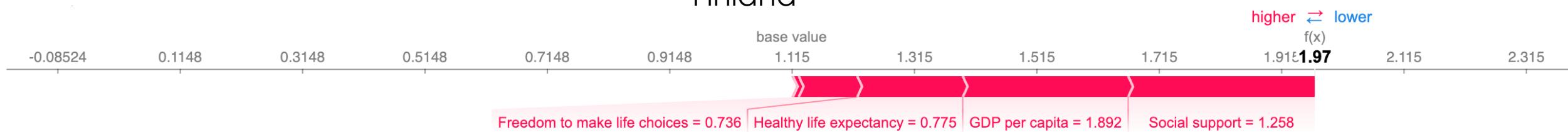
Se calculan efectos marginales de quitar una variable

Explainable Artificial Intelligence (XAI)

Mexico



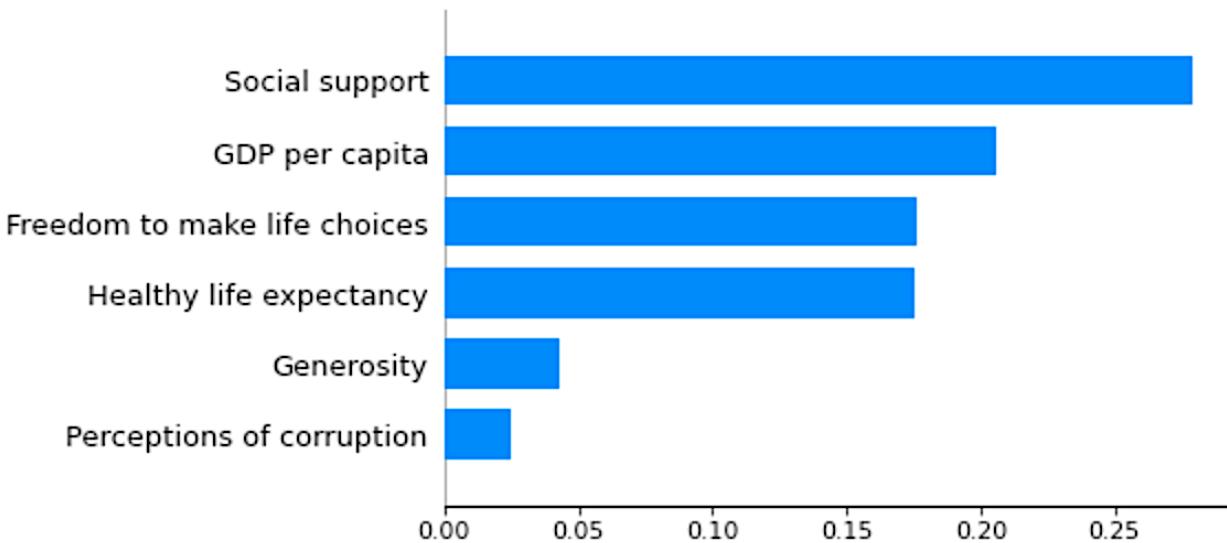
Finland



Afghanistan



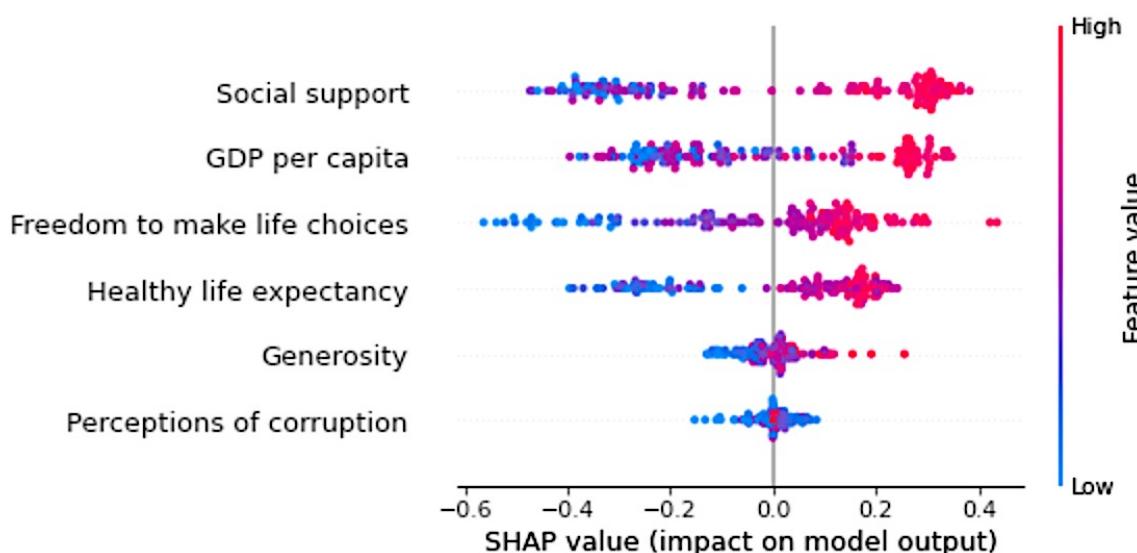
Explainable Artificial Intelligence (XAI)



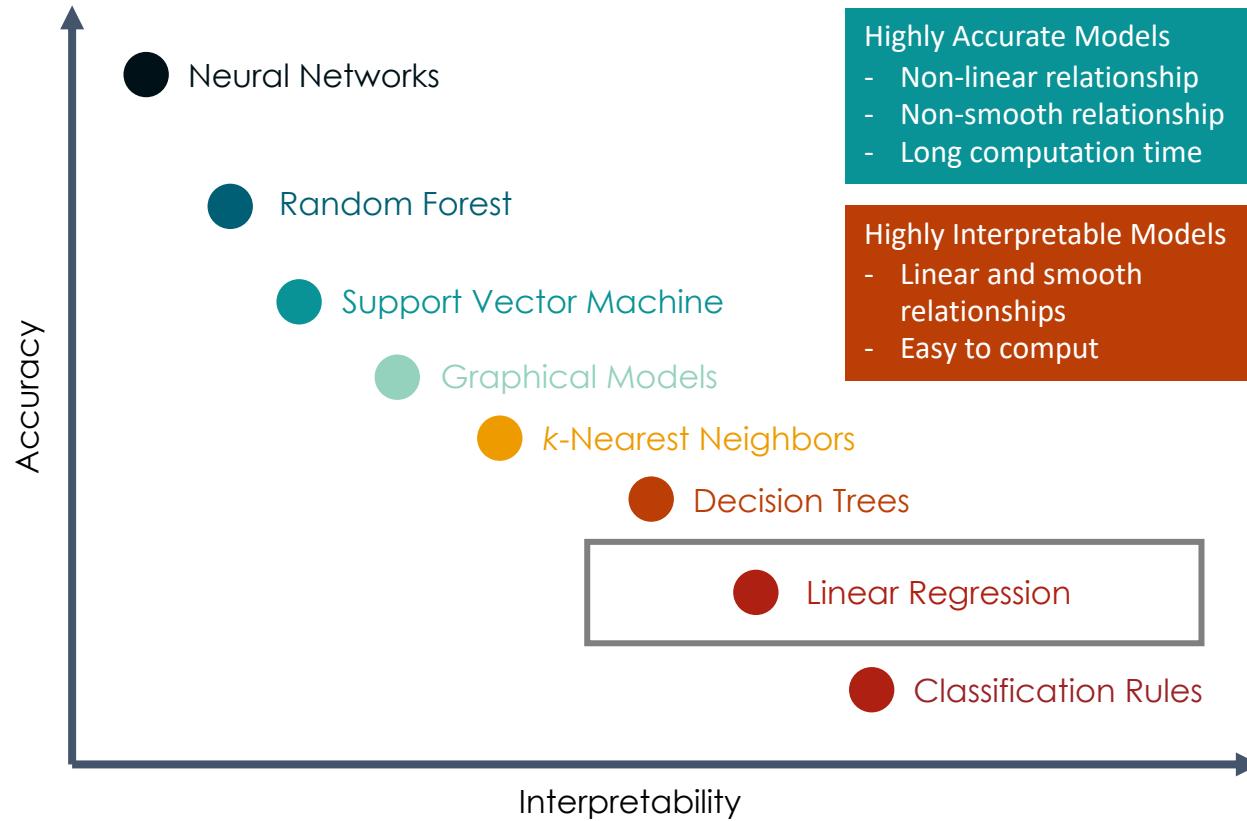
Un país que **aspira a tener un nivel alto de felicidad**, debería de enfocarse en mecanismos de **soporte social o emocional**.

El PIB, la libertad y la salud son tan **importantes** como el primero.

Que una población sea o no generosa, así como su percepción de la corrupción, **no impacta** en gran medida.



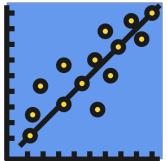
Tradeoffs



“If Francis Galton, Karl Pearson, Ronald Fisher, and Jerzy Neyman had had access to computers, they may have created an entirely different field”

- Marcos M. López de Prado

Clásico vs ML



Recae en supuestos como linealidad y dependencia



Son los más estudiados / usados en la academia



Se entiende qué hace el modelo por dentro

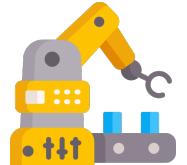


Softwares facilitan la aplicación de modelos

Métodos como Ensemble estimators o Bootstrapping, tienen mejores resultados



Aplicaciones a la industria debido a su flexibilidad



Las aplicaciones profesionales superan a las académicas



Difícil de comprender qué hacen por dentro



Se requiere programar



Gran número de modelos



Comentarios Finales

Cada **problema** se resuelve con un **aprendizaje** distinto

Existe una gran **variedad** de modelos

Machine Learning **no es magia** y los modelos **no son cajas negras**

Falta una **unificación** entre ML y Economía

Con el conocimiento de **Estadística y Probabilidad** de la facultad es más que suficiente para **comenzar** en ML

Repositorio de GitHub



<https://github.com/milioe/Machine-Learning-y-Economia>

GRACIAS

Emilio Sandoval Palomino

 emilio.sandpal@gmail.com

 milioe

Humberto Acevedo Cruz Manjarrez

 hac106@hotmail.com

BetoACE