MACHINE LEARNING:

un enfoque más amplio

María Lourdes Linares Barrera
Inteligencia Empresarial

SOBRE EL TRABAJO

AUTORA María Lourdes Linares Barrera

TRABAJO Machine Learning: un enfoque más amplio (más allá del Deep Learning y PowerBI)

OBJETIVOS

- Alternativas ecosistema R
- Mucho más que el “marketing”
- Mucho más de lo que se ve en la carrera
- Relación con el mundo laboral empresarial y académico

ÍNDICE DE CONTENIDOS

¿Y ESTO PARA QUÉ?

Motivaciones para aprender

PREDICCIÓN

Alternativas al Deep Learning (en el ecosistema R)

VISUALIZACIÓN

Alternativas al PowerBI (en el ecosistema R)

01

¿Y ESTO PARA QUÉ?

PERFILES Y ÁMBITOS

1

Análisis de datos
Descubrir patrones y tendencias
Construir modelos predictivos

DATA SCIENTIST

2

Aplicar el conocimiento extraído por Data Scientist a la toma de decisiones

DATA ANALYST

3

Bioinformática
Comportamiento social
Estudios de mercado
Estadística

INVESTIGACIÓN

4

Justificación de las decisiones tomadas por los modelos

XAI

¿Y ESTO PARA QUÉ?

Data Analyst Wealth - Data Science & BI (Openbank)

Santander

Madrid, Madrid provincia

To be successful in the role you must have experience on:

- +2 years of experience in a similar role.
- Statistical analysis: you know the basic statistical concepts, and know how to implement them in different programming languages (ideally R, Python or Spark). You know when to use different indicators and have experience summarizing data insights with them.
- Dealing with large datasets, and implementing scalable solutions. You know what data cleaning means and how to interact with a data engineer.
- How to solve problems using data: analyze the problem and decompose it into different analytical components, and recycle all the components you can from other projects.
- Software development: you know what git means, and how to develop your own code, test it and use it to produce actionable results. Open-source lovers get a plus if you show us your github!
- Data Visualization, and tools like ggplot, matplotlib, tableau, or similar. Bonus points for Microstrategy, or QuickSight.

Research Scientist

Bristol Myers Squibb

Sevilla, Sevilla provincia

Requirements:

- PhD in a relevant discipline, accompanied by original research publications
- Experience incorporating modern deep learning concepts (e.g. attention, graph-based, disentanglement) into models applied towards real-world challenges
- Strong grasp of scientific programming languages (e.g. Python, R) and relevant libraries and software (e.g. PyTorch, TensorFlow).

Data Scientist, Spain

TikTok (parte de ByteDance)

Madrid, Madrid provincia

Qualifications

- Bachelor's degree or above, majoring in statistics or data science is preferred;
- Proficiency in SQL/Excel/R/Python/Tableau. Familiar with common statistical methods and applications (A/B testing, probability, regression);
- At least 3 years of working experience in an analytical role involving e-commerce/user growth/product optimization/business analytics/performance marketing;
- Able to complete English reports and communicate with global staff independently in a diverse and cross-functional environment;

Junior Data Scientist

Deloitte

Madrid, Madrid provincia

Crea una cuenta de Indeed antes de continuar a la página web de la empresa

Solicitar en la página web

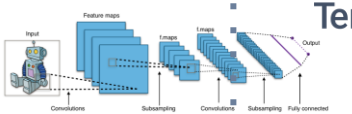
Requisitos:
Titulación superior orientada a Ingeniería. Valorable máster.
Experiencia de 0 a 3 años de experiencia
Valorable conocimiento lenguajes Python, Spark, R, y SAS sobre entornos Big Data y Cloud
Valorable conocimiento arquitectura Cloud (AWS, Azure, Google)

02

PREDICCIÓN

PREDICCIÓN DENTRO DEL ML

Machine Learning



TensorFlow

PyTorch

Predicción & Co

Deep Learning
Machine Learning Clásico
Series Temporales
Inteligencia Artificial Explicable

Visualización

EDA
PowerBI
Tableau
RShiny

Otros

Sistemas de recomendación
Tratamiento de redes/grafos
...

PREDICCIÓN

DATOS

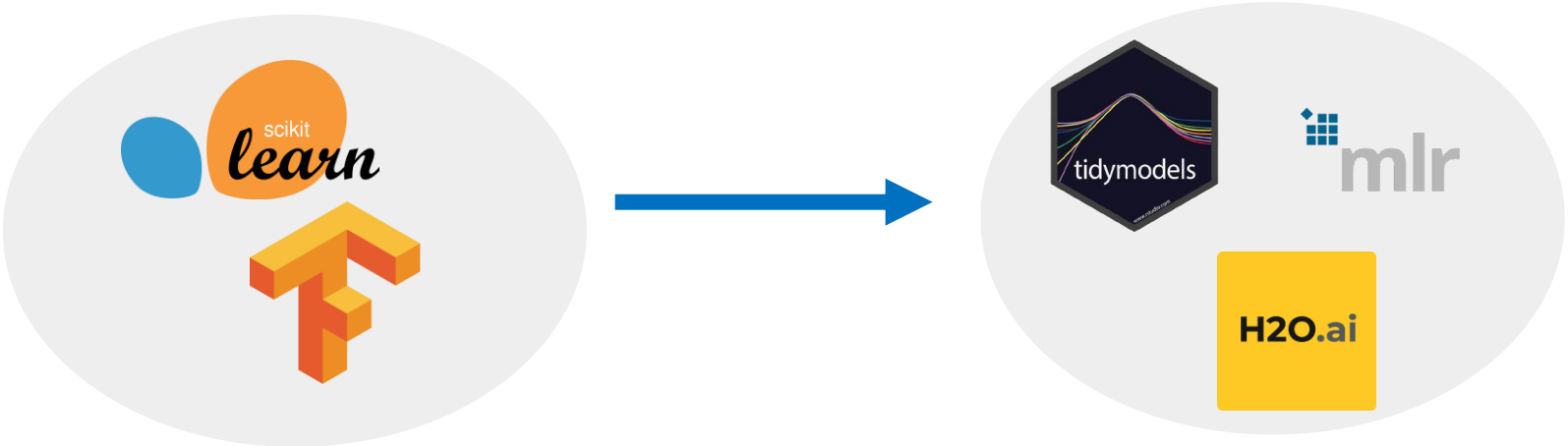
Librerías de estructura de datos y tratamiento de dataframes

pandas

NumPy

PREDICCIÓN

ALGORITMOS Algoritmos de Machine Learning y Deep Learning



PREDICCIÓN

XAI **Explicabilidad de los modelos de Machine Learning**

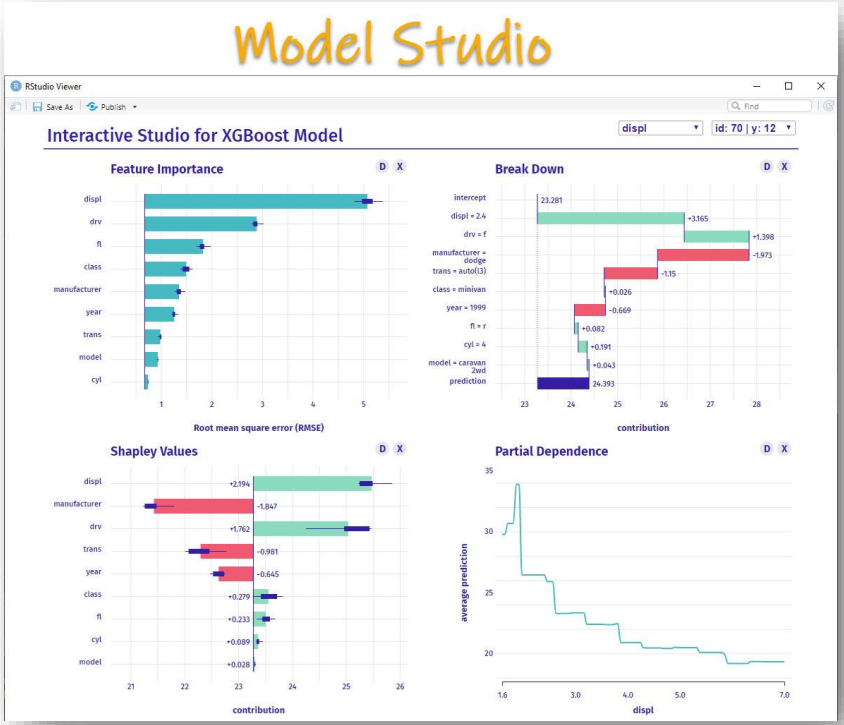
Los modelos básicos ya incluyen test estadísticos y estudio ANOVA más precisos

```
Call:
lm(formula = reventa ~ ., data = coches3)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8760 -1.6292 -0.0984  1.3597  7.2517

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       15.4588203  8.8347913   1.750 0.083080 .
ventas            -0.0004783  0.0042859  -0.112 0.911349
tipoCamión         0.6494654  1.1091668   0.586 0.559439
precio             0.8500987  0.0445224  19.094  < 2e-16 ***
motor_s           -1.1708220  0.6842818  -1.711 0.090030 .
caballos           0.0124367  0.0150276   0.828 0.409781
batalla            0.0680193  0.0847377   0.803 0.423959
anchura            0.1061208  0.1368104   0.776 0.439683
longitud          -0.0914974  0.0516296  -1.772 0.079264 .
peso_revestimiento -4.9036692  1.2878525  -3.808 0.000236 ***
tapon_combustible   0.2738897  0.1623872   1.687 0.094640 .
kpl               -0.1898848  0.1306856  -1.453 0.149209
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.802 on 105 degrees of freedom
Multiple R-squared:  0.9472,    Adjusted R-squared:  0.9417
F-statistic: 171.4 on 11 and 105 DF,  p-value: < 2.2e-16
```

PREDICCIÓN

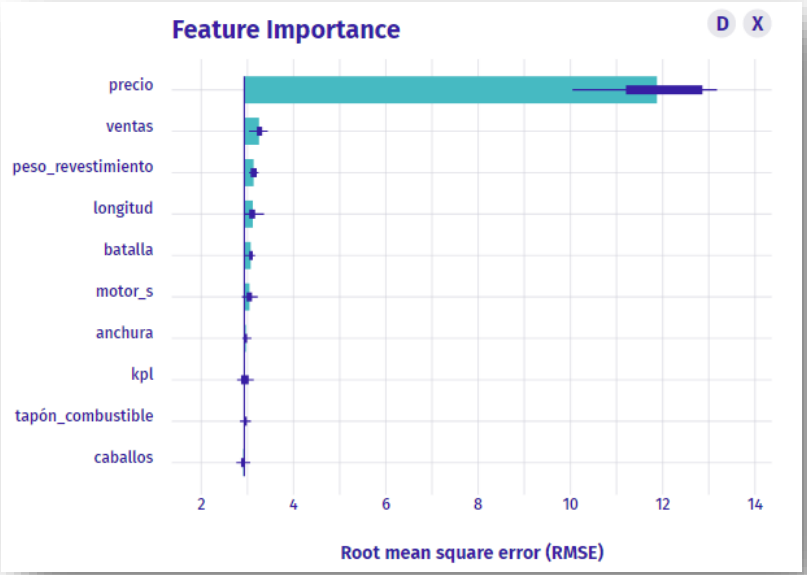
XAI — **Explicabilidad de los modelos de Machine Learning**

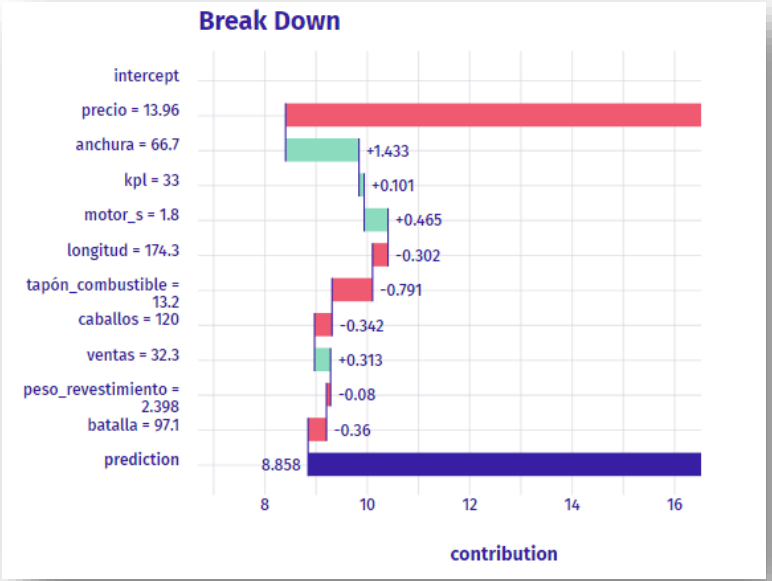
Módulos de XAI especializados:
- ModelStudio
- LIME



PREDICCIÓN

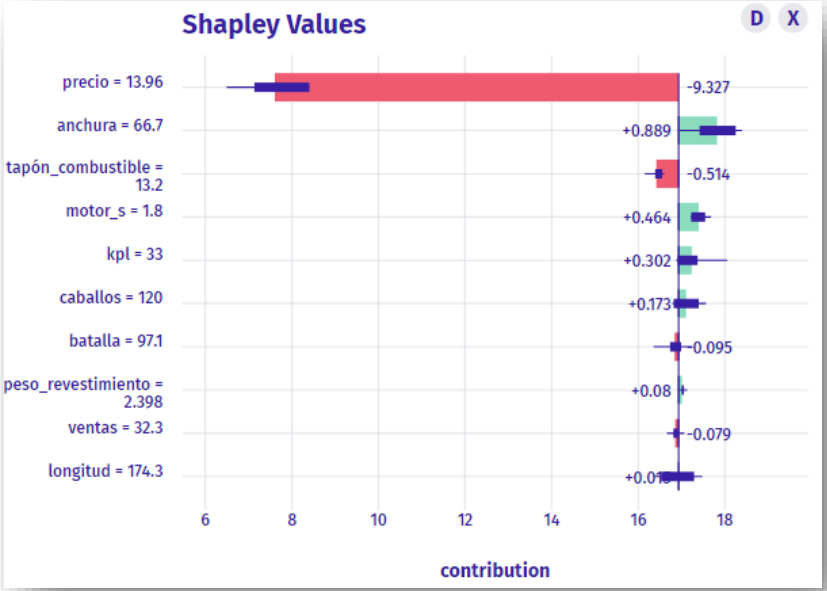
XAI Explicabilidad de los modelos de Machine Learning

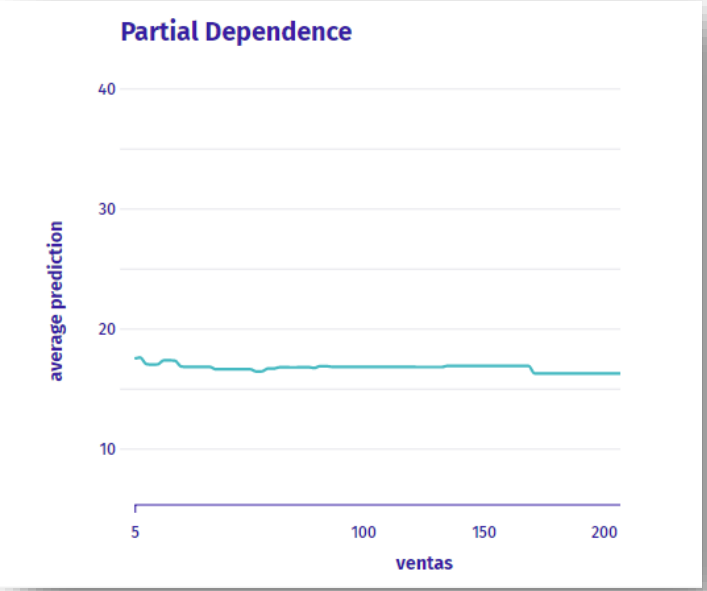




PREDICCIÓN

XAI Explicabilidad de los modelos de Machine Learning

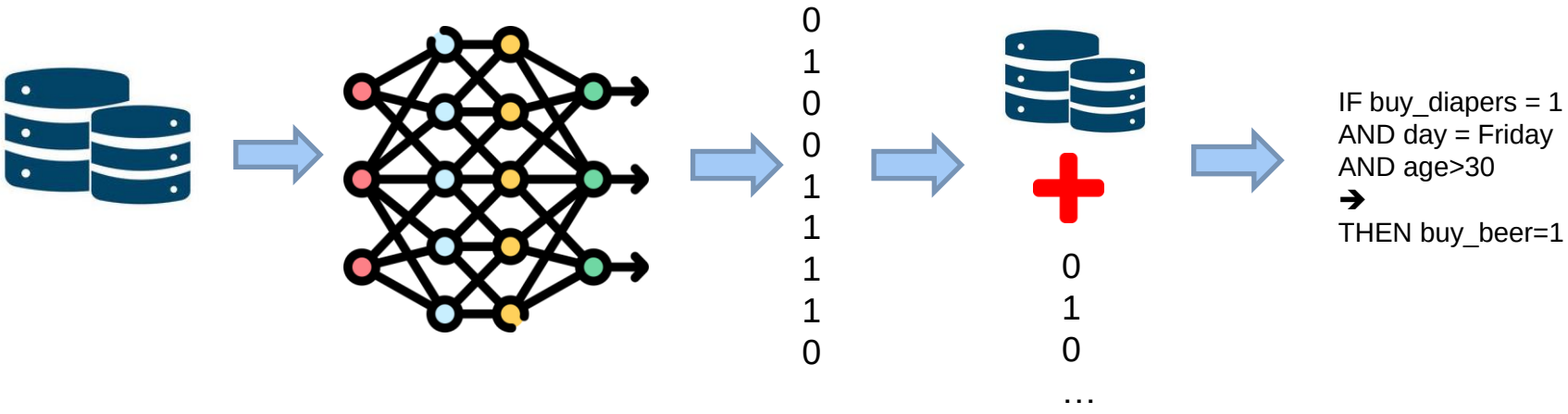




PREDICCIÓN

XAI **Explicabilidad de los modelos de Machine Learning**

Construcción de modelos de explicabilidad sobre modelos existentes:
- Reglas de asociación: ARULES que aprendan los patrones de decisión



SCRIPT DE EJEMPLO

```r
# ------------------------------------------------------------------------------
# Lectura del dataset y preprocesado

install.packages("tidyverse")
library(tidyverse)
library(readr)

coches <- read.table("Car_sales.txt", encoding = "UTF-8")

coches <- coches %>%
  select(c("reventa", everything()))
View(coches)

coches$fabricante = factor(coches$fabricante)
coches$modelo = factor(coches$modelo)
coches$tipo = factor(coches$tipo)

coches2 <- na.omit(coches)
summary(coches2)

coches2 <- coches2[,-c(2,3,5)]
View(coches2)
```

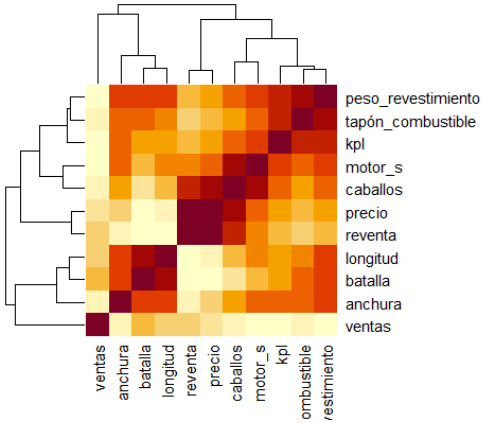
	reventa	ventas	precio	motor_s	caballos	batalla
1	16.360	16.919	21.500	1.8	140	101.2
2	19.875	39.384	28.400	3.2	225	108.1
4	29.725	8.588	42.000	3.5	210	114.6
5	22.255	20.397	23.990	1.8	150	102.6
6	23.555	18.780	33.950	2.8	200	108.7
7	39.000	1.380	62.000	4.2	310	113.0
9	28.675	9.231	33.400	2.8	193	107.3

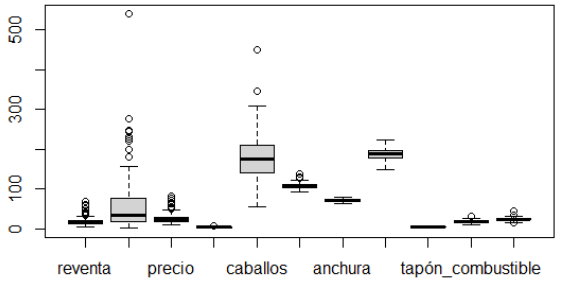
anchura	longitud	peso_revestimiento	tapón_combustible	kpl
67.3	172.4	2.639	13.2	28.0
70.3	192.9	3.517	17.2	25.0
71.4	196.6	3.850	18.0	22.0
68.2	178.0	2.998	16.4	27.0
76.1	192.0	3.561	18.5	22.0
74.0	198.2	3.902	23.7	21.0
68.5	176.0	3.197	16.6	24.0

Acceso al script
https://github.com/lourdesLB/rsample-predicion-visualization-xai

SCRIPT DE EJEMPLO

```
# ------------------------------------------------------------------------
# Representaciones

boxplot(coches2)
heatmap(abs( cor(coches2) ), scale="none")
```





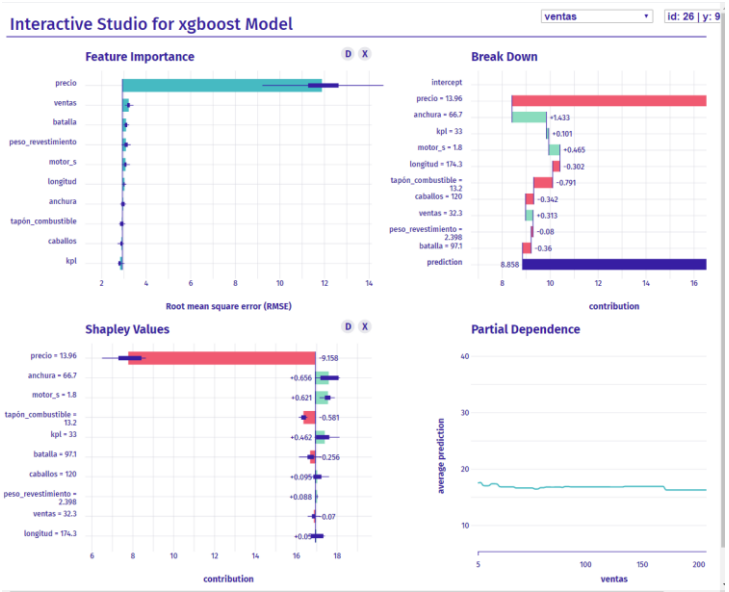
SCRIPT DE EJEMPLO

```r
# ------------------------------------------------------------------------
# Modelo de regresion lineal multiple con todas las variables

index <- sample(1:nrow(coches2), 0.8*nrow(coches2))
train <- coches2[index,]
test <- coches2[-index,]

reglineal = lm(reventa ~., data=train)
summary(reglineal)

predict(reglineal, test)
```

```
Call:
lm(formula = reventa ~ ., data = coches3)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8760 -1.6292 -0.0984  1.3597  7.2517

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       15.4588203  8.8347913   1.750 0.083080 .
ventas            -0.0004783  0.0042859  -0.112 0.911349
tipoCamión         0.6494654  1.1091668   0.586 0.559439
precio             0.8500987  0.0445224  19.094  < 2e-16 ***
motor_s           -1.1708220  0.6842818  -1.711 0.090030 .
caballos           0.0124367  0.0150276   0.828 0.409781
batalla            0.0680193  0.0847377   0.803 0.423959
anchura            0.1061208  0.1368104   0.776 0.439683
longitud          -0.0914974  0.0516296  -1.772 0.079264 .
peso_revestimiento -4.9036692  1.2878525  -3.808 0.000236 ***
tapon_combustible  0.2738897  0.1623872   1.687 0.094640 .
kpl               -0.1898848  0.1306856  -1.453 0.149209
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.802 on 105 degrees of freedom
Multiple R-squared:  0.9472,    Adjusted R-squared:  0.9417
F-statistic: 171.4 on 11 and 105 DF,  p-value: < 2.2e-16
```

SCRIPT DE EJEMPLO

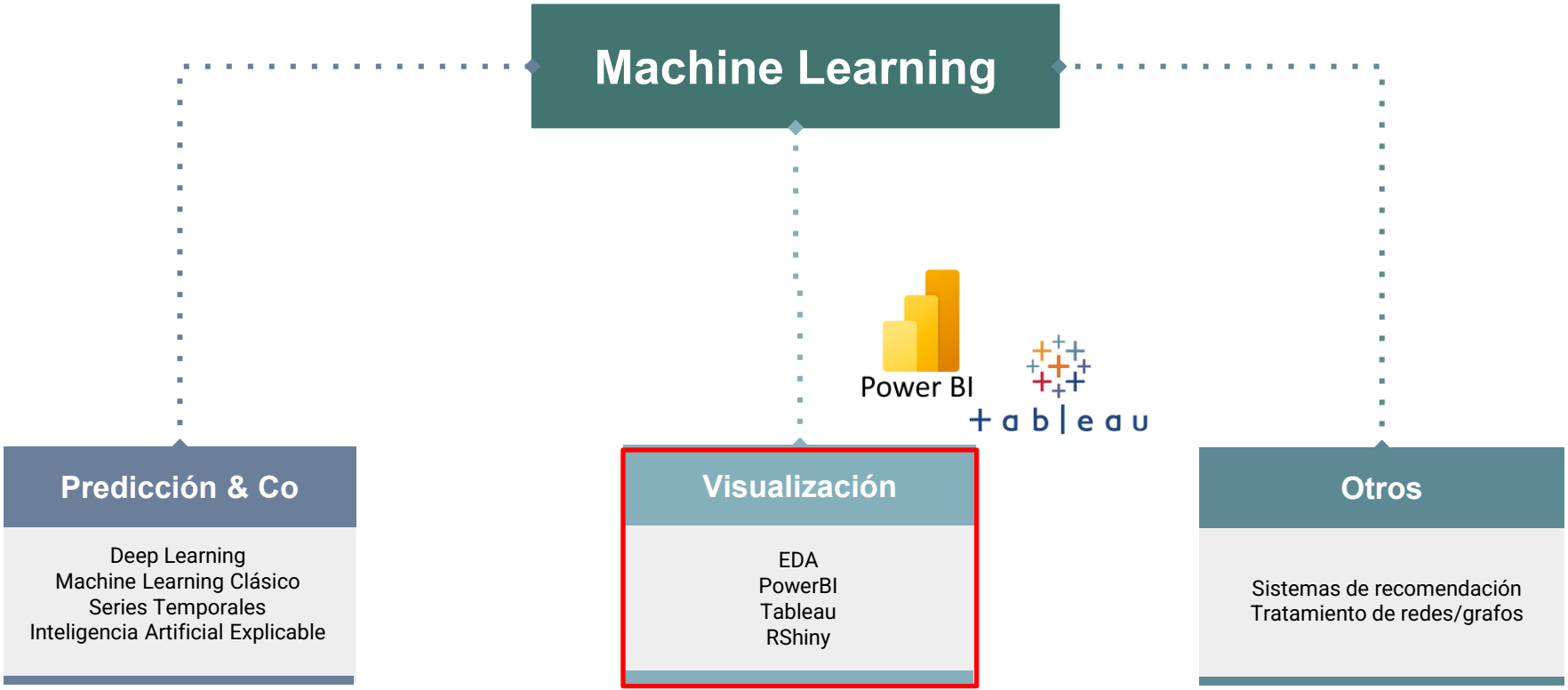
```r
# ------------------------------------------------------------------------
# ModelStudio para explicabilidad

install.packages("DALEX")
install.packages("DALEXtra")
install.packages("mlr")
install.packages("xgboost")
library(modelStudio)
library(xgboost)
library(DALEX)

train_matrix <- model.matrix(reventa ~.-1, train)
test_matrix <- model.matrix(reventa ~.-1, test)

xgb_matrix <- xgb.DMatrix(train_matrix, label = train$reventa)
params <- list(max_depth = 3,
               objective = "reg:linear",
               eval_metric = "rmse")
model <- xgb.train(params, xgb_matrix, nrounds = 500)

explainer <- explain(model,
                     data = test_matrix,
                     y = test$reventa,
                     type = "regression",
                     label = "xgboost")

modelStudio::modelStudio(explainer)
```



03

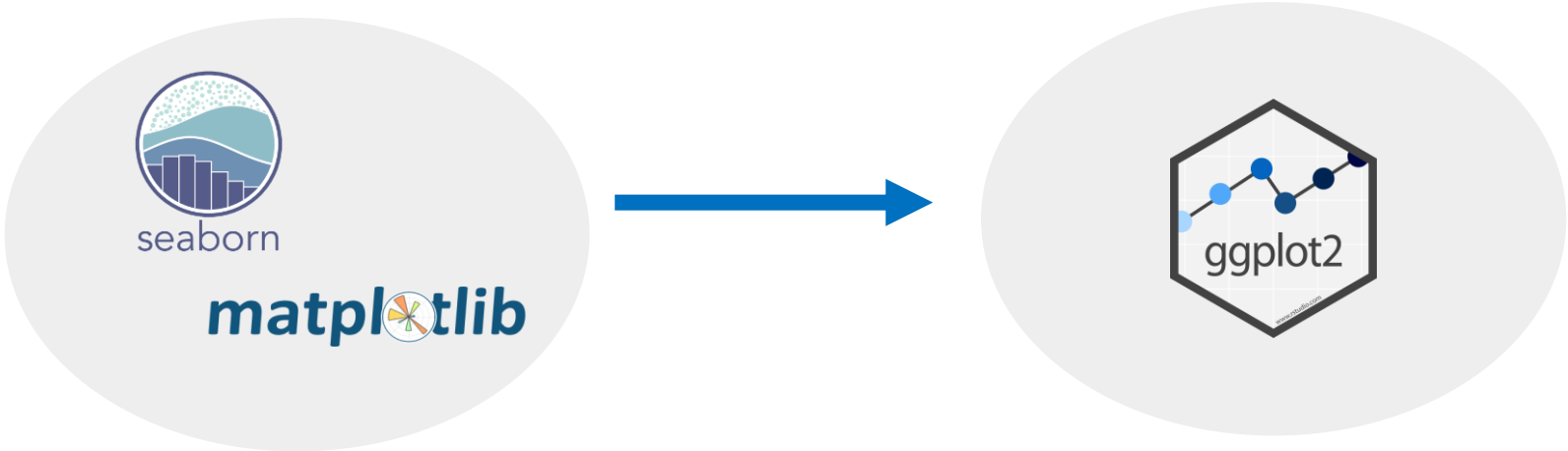
VISUALIZACIÓN

VISUALIZACIÓN DENTRO DEL ML



VISUALIZACIÓN

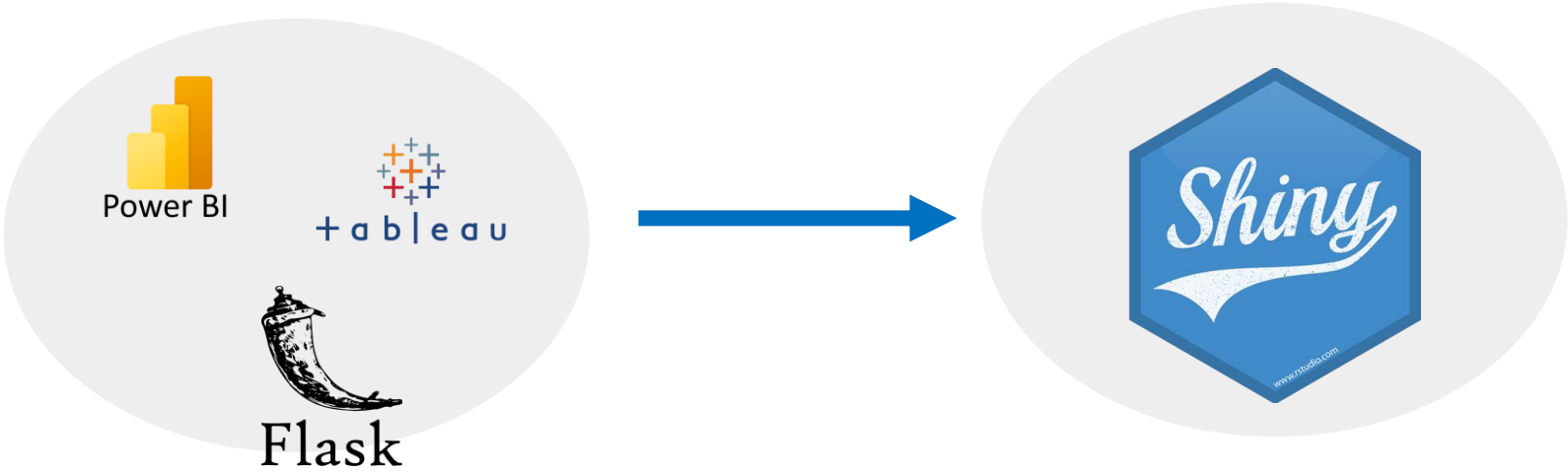
EDA Análisis exploratorio de datos



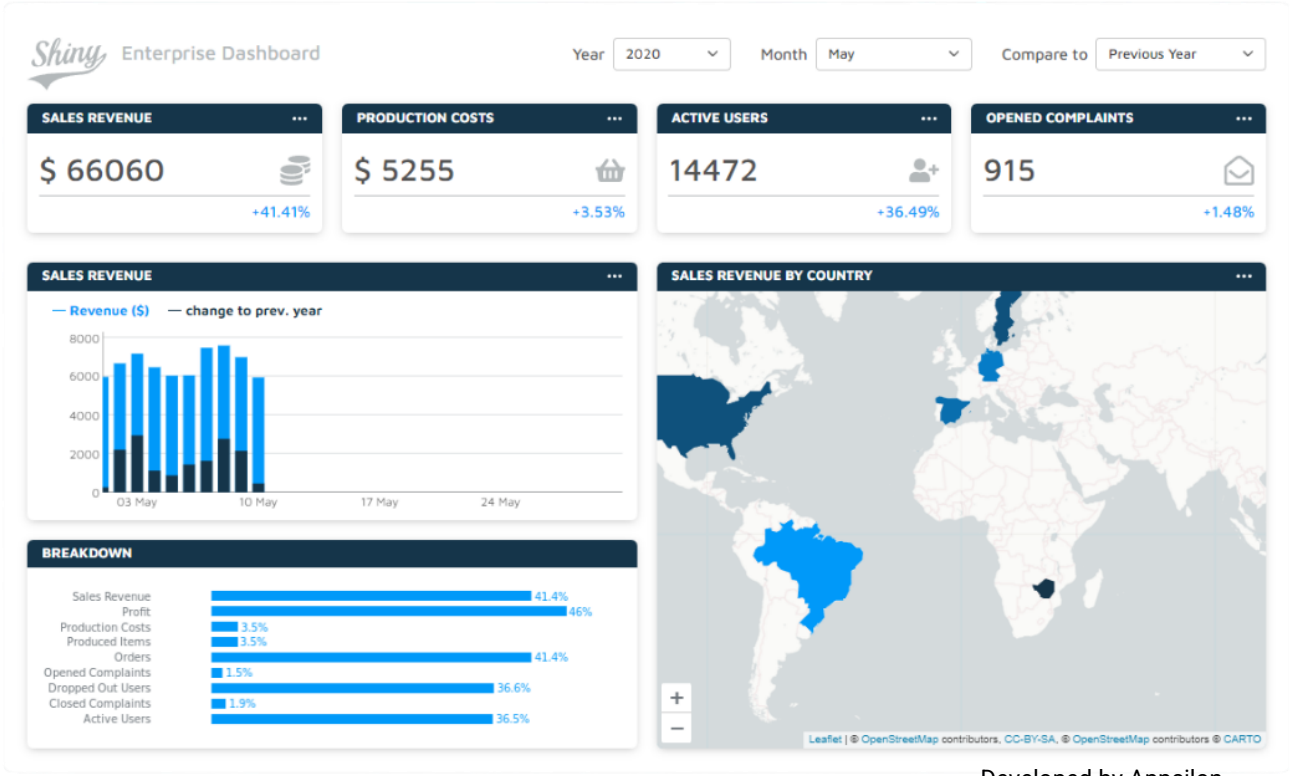
VISUALIZACIÓN

DASHBOARDS

Dashboards e informes. Interfaces web.



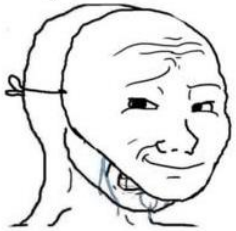
RSHINY



Developed by Appsilon

RSHINY

Python Data Scientists:



Make a web app with R Shiny.

No R can't be used in production. I demand PyTorch, sklearn, flask, Django, apis, IT, infrastructure, mlops, AWS so my app can scale…

R Data Scientists:

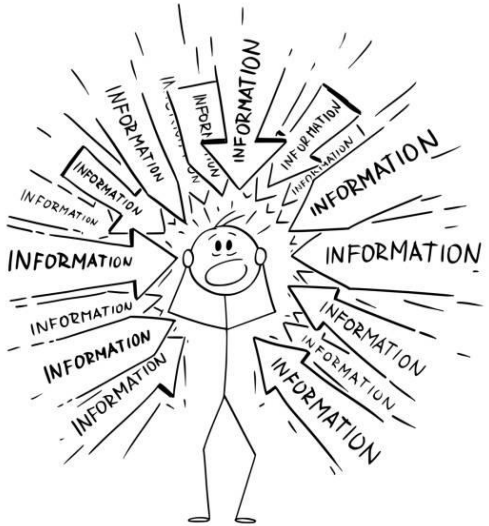
Make a web app with R Shiny.

Yeah let's just make a shiny app.

04

HAY MÁS...

(pero eso para otro día)



MUCHAS GRACIAS
POR SU ATENCIÓN