

## INDIVIDUAL CHALLENGE 1

### -DATA SCIENCE-

### CORRELATION



#### CHALLENGE STATEMENT

Recks is a company that designs urban shoes, it has spent approximately 4 years dedicating a part of its income to marketing through television, radio and newspapers, but Jhon, the CEO wants to optimize spending by allocating all this investment to a single medium of the three. To do this, John has sent you an Excel in which the expenditure in each of the media is shown together with the income in euros for that week. From these data, Jhon wants you to find the correlation between spending on each medium and income and decide which medium is the one that generates the greatest economic return, for this:

#### SECTIONS

- Obtain a graph that shows the correlation between the expenses in each media with the sales. To do so, you must use Matplotlib library. -30 points-
- What is the correlation between investment between Marketing spending in each medium and income? (ex: Periodic-income correlation = 0.75986...) -40 points-
- In which of the three media would you recommend John to maintain or increase his investment? why? -30 points-

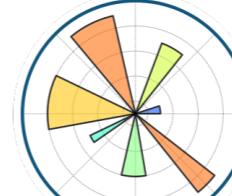
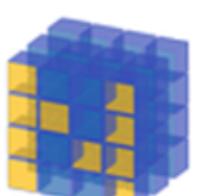
#### RESOURCES

The excel with the historical data will be provided by discord group with the name "dsc12.csv"

#### Delivery

The deadline to deliver your resolution is at 19:00.

#### STACK



## INDIVIDUAL CHALLENGE 1

### -DATA SCIENCE-

### CORRELATION



#### Background

Racks es una empresa que diseña zapatillas urbanas, esta lleva aproximadamente 4 años destinando una parte de sus ingresos al marketing a través de la televisión, radio, periódico, redes sociales y email, pero Juan, el CEO quiere optimizar el gasto destinando toda esta inversión a dos medios de los cinco. Para hacer esto Juan te ha enviado un Excel en el que se muestra el gasto en cada uno de los medios de marketing junto a los ingresos en euros referente a esa semana. Por ello Juan quiere saber cuales son los dos medios que mayor retorno económico genera, para ello:

#### Reto

- Realiza un análisis de este dataset y explica de que dos medios eliminarías el gasto en marketing y a que dos medios aumentarías el gasto en marketing con tal de maximizar los Ingresos. 250ptos
- Realiza una representación gráfica de los datos que le permitan a Juan entender mejor tu análisis. 250ptos

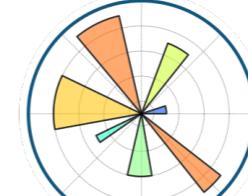
#### Dataset

El excel con los valores de gastos históricos de la empresa se adjuntara en el canal de Discord "Marketing.csv"

#### Entrega

The deadline to deliver your resolution is at 19:00.

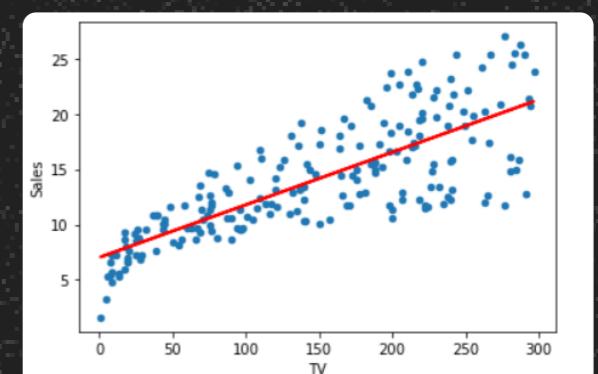
#### STACK



## INDIVIDUAL CHALLENGE 2

### -DATA SCIENCE-

### LINEAR REGRESSION



#### CHALLENGE STATEMENT

From the excel with the weekly Marketing expenses in the different media and sales, define a linear regression taking as predictive variables ['TV', 'Radio'] and variable to predict ['Sales']. The use of the Scikit-Learn library is recommended.

#### SECTIONS

- Being  $y(w,x) = w_0 + w_1 * x_1$  the equation that defines the lineal regression, define the value of the coefficients  $w_0$ ,  $w_1$  and  $w_2$ .
- What will be the prediction in sales if € 500 is allocated to television and € 80 to radio?

#### RESOURCES

The excel needed is the same needed for the challenge 1, the one provided in the discord group: "dsc12.csv"

#### DEADLINE

The deadline to deliver the results is 2 hours after the publication of the challenge

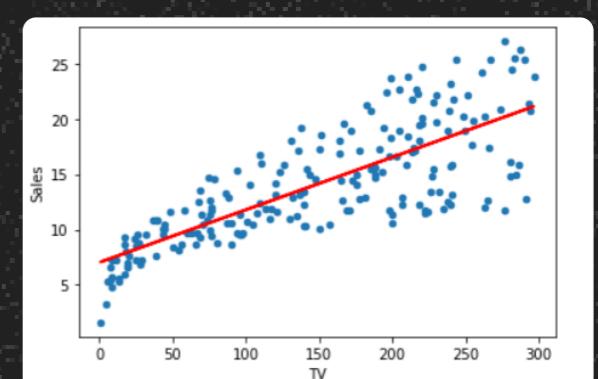
#### SKILLS



## INDIVIDUAL CHALLENGE 2

### -DATA SCIENCE-

### LINEAR REGRESSION



#### CHALLENGE STATEMENT

From the excel with the weekly Marketing expenses in the different media and sales, define a linear regression taking as predictive variables ['TV', 'Radio'] and variable to predict ['Sales']. The use of the Scikit-Learn library is recommended.

#### SECTIONS

- Being  $y(w,x) = w_0 + w_1 * x_1$  the equation that defines the lineal regression, define the value of the coefficients  $w_0$ ,  $w_1$  and  $w_2$ .
- What will be the prediction in sales if € 500 is allocated to television and € 80 to radio?

#### RESOURCES

The excel needed is the same needed for the challenge 1, the one provided in the discord group: "dsc12.csv"

#### DEADLINE

The deadline to deliver the results is 2 hours after the publication of the challenge

#### SKILLS



## INDIVIDUAL CHALLENGE 3 -DATA SCIENCE-

NPL

### CHALLENGE STATEMENT

The main objective of this challenge is to apply a text classification algorithm to the scikit-learn dataset **fetch\_20newsgroups**. Which is a collection of approx 20.000 newsgroup documents. You have more info about the dataset in this link:

[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch\\_20newsgroups.html](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_20newsgroups.html)

### SECTIONS

- Preprocess all the text, normalize it, vectorize it and try as much predictive algorithm as you want to get the best performance.
- Write down in the last cell of your challenge section the best score you got from your models.

### RESOURCES

To import the Scikit-learn dataset you just have to write the following line into your google colab notebook:

```
from sklearn.datasets import fetch_20newsgroups
```

You can get more info about this dataset here:

[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch\\_20newsgroups.html](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_20newsgroups.html)

### SKILLS



## INDIVIDUAL CHALLENGE 3 -DATA SCIENCE-

NPL

### CHALLENGE STATEMENT

The main objective of this challenge is to apply a text classification algorithm to the scikit-learn dataset **fetch\_20newsgroups**. Which is a collection of approx 20.000 newsgroup documents. You have more info about the dataset in this link:

[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch\\_20newsgroups.html](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_20newsgroups.html)

### SECTIONS

- Preprocess all the text, normalize it, vectorize it and try as much predictive algorithm as you want to get the best performance.
- Write down in the last cell of your challenge section the best score you got from your models.

### RESOURCES

To import the Scikit-learn dataset you just have to write the following line into your google colab notebook:

```
from sklearn.datasets import fetch_20newsgroups
```

You can get more info about this dataset here:

[https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch\\_20newsgroups.html](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_20newsgroups.html)

### SKILLS

