

## **Churn prediction using ML**

### **Español:**

#### **Introduction**

Se analizarán los datos de los usuarios de la plataforma de streaming de música para identificar grupos de clientes “similares”, aplicando modelos de clústering.

Tal vez la empresa pueda aplicar distintas estrategias de marketing para estos diferentes grupos identificados.

#### **Pasos:**

##### **1. Preparación de los datos para aplicar modelos de clústering.**

Evaluación visual e intuitiva de a dos variables por vez: visualizaciones de cómo se comportan las variables presentes en el dataset mediante el gráfico “pairplot”.

Buscaremos responder:

- a. ¿Es posible en alguna de ellas visualizar “grupos” separados? ¿Qué variables presentan dicha separación?
- b. ¿Se puede visualizar algún impacto de la variable target (Churn) en los posibles grupos?

##### **2. Creación de modelos de clustering: K-means**

- a. Buscamos el número de clusters óptimo observando la inercia de cada uno.
- b. Correr K-means utilizando como parámetro el número de clusters obtenido en el punto anterior.
- c. Analizar la cantidad de casos en cada clúster.
- d. Grafique de a dos variables incluyendo el label obtenido en el clustering. ¿Es

posible determinar alguna separación de grupos?

Documentación utilizada:

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

### 3. Creación de modelos de clustering: Algoritmo Jerárquico

- Aglomeraciones utilizando el algoritmo “AgglomerativeClustering”.
- Analizar la cantidad de casos en cada clúster.
- Gráfico de a dos variables incluyendo el label obtenido en el clustering. ¿Es posible determinar alguna separación de grupos?
- Conclusión final de los métodos de clustering.

Documentación:

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>

## English:

### Introduction

The data of users from the music streaming platform will be analyzed to identify "similar" customer groups using clustering models.

Perhaps the company can apply different marketing strategies for these identified groups.

### Steps:

#### 1. Data preparation for applying clustering models.

Visual and intuitive evaluation of two variables at a time: examining how the variables present in the dataset behave through the "pairplot" graph.

We will seek to answer:

- a. Is it possible to visualize separate "groups" in any of them? Which variables show such separation?
- b. Can any impact of the target variable (Churn) be visualized in the potential groups?

## **2. Clustering Model Creation: K-means**

- a. Determine the optimal number of clusters by observing the inertia of each one.
- b. Run K-means using the number of clusters obtained in the previous step as a parameter.
- c. Analyze the number of cases in each cluster.
- d. Plot two variables including the label obtained from clustering. Is it possible to determine any separation of groups?

Documentation:

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

## **3. Clustering Model Creation: Hierarchical Algorithm**

- Perform agglomerations using the "AgglomerativeClustering" algorithm.
- Analyze the number of cases in each cluster.
- Plot two variables including the label obtained from clustering. Is it possible to determine any separation of groups?
- Final conclusion of the clustering methods

Documentation:

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>