

Nombre del Programa: Magíster en Ciencia de Datos

Curso: Python para Ciencia de Datos

Profesor: Alejandro Cataldo

Tarea 4

Objetivo

El objetivo de esta tarea es que apliquen algunos de los conceptos aprendidos durante su revisión de las últimas semanas de clases sobre asociación de datos, regresiones y métodos de agrupación, y con esto, puedan aplicar algunas de las herramientas de Python que ya han aprendido. Para esto, deberán utilizar el archivo “Tarea4.xlsx”, que contiene datos sobre características más relevantes de celulares y sobre una encuesta realizada a diferentes compradores. A estos compradores se les pidió que indicaran la probabilidad de comprar cada uno de los teléfonos celulares al precio dado, en una escala de 1 (muy poco probable) a 10 (muy probable). También se les pidió información a cada participante respecto a: edad, sexo y ocupación.

Preguntas que deben responder

Tomando el archivo entregado, y cargándolo como Dataframe en Python, deben entregar un archivo en Jupyter Notebook con el código y la respuesta/resultado a cada una de las siguientes 5 preguntas.

- (1) Considere solamente los datos de: internal memory; RAM; performance; main camera; selfie camer; y battery size, de los celulares. Utilice el método de agrupación “K-Means” y construya 3 agrupaciones de celulares. ¿Qué puede decir sobre los precios de venta de los celulares de cada grupo?
- (2) Calcule la correlación (bivariada) entre el precio y todas las variables cuantitativas con información sobre los celulares. ¿Cuáles son las 3 variables que presentan una mayor correlación con el precio?
- (3) Construya una regresión multivariada entre precio y las 3 variables determinadas en la parte (2). Utilice un 20% de los datos para testear.
- (4) Construya una agrupación, utilizando K-Means, considerando ranking promedio (puesto por el comprador), edad y sexo. Escoja, con el método del codo, la cantidad de grupos. ¿Qué puede decir de las similitudes o diferencias intra e inter grupos?