

Nombre del Programa: Magíster en Ciencia de Datos

Curso: Python para Ciencia de Datos

Profesor: Alejandro Cataldo

Tarea 3

Objetivo

El objetivo de esta tarea es que apliquen algunos de los conceptos aprendidos durante su revisión de las últimas semanas de clases sobre visualización y asociación de datos, y con esto, puedan aplicar algunas de las herramientas de Python que ya han aprendido. Para esto, deberán utilizar el archivo “Tarea3.xlsx”, que contiene datos sobre estudiantes (base de datos basada en la base de datos student_data, obtenida de la página Kaggle).

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) ¿Cuál es la correlación entre ausencias (absences) a clases de un estudiante y las demás variables numéricas de las bases de datos? ¿Alguna de estas correlaciones es importante?
- (2) ¿El promedio de ausencias es diferente según al tipo de escuela en la que estudia el estudiante? ¿y con respecto al género del estudiante?
- (3) Considere que el indicador que quiere observar es si un estudiante asistió al colegio todos los días de clases. ¿Cómo queda la tabla de frecuencia de este indicador con respecto a los años de educación de la mamá (Medu) y los años de educación del papá del estudiante (Fedu)? ¿Hay alguna asociación que parezca importante? (Lo que se debe contar como frecuencia es la cantidad de estudiantes que no faltaron ningún día a clases).
- (4) A través de un gráfico boxplot indique qué tan distinta es la calificación en primer grado (G1 que va de 0 a 20) de los estudiantes dependiendo de quién es su tutor (guardian).
- (5) Haga un gráfico de dispersión que tenga en el eje “y” la calificación en el segundo grado (G2), en el eje “x” esté el tiempo de viaje (traveltime) y en el gráfico el tamaño del punto (x,y) esté dado por la cantidad de estudiantes con esa combinación. Al observar el resultado de este gráfico, ¿ve alguna asociación interesante de explicar?