

Guía de Ejercicios I – ACD104

Indicaciones Generales:

- 1) Resolver los siguientes enunciados, utilizando Python, Matplotlib, Numpy o Pandas. **La solución debe ser registrada en el Aula Digital UDB, no se calificarán soluciones enviados via email.**
- 2) Se debe redactar un archivo en **formato PDF** con la siguiente **estructura y contenido**:
 - i. Portada
 - ii. Por cada problema y su respectivo enunciado se debe proporcionar listado de código fuente, y una o más capturas de pantalla que demuestren el correcto funcionamiento de la solución del enunciado respectivo.
- 3) **Por cada enunciado se debe entregar el archivo de código fuente respectivo.**
- 4) **Fecha y hora última de entrega: Lunes 8 de Septiembre a las 23:59 horas.**
- 5) **Cualquier evidencia de copia total o parcial, se asignatura la nota de 0.0.**
- 6) **En caso de no cumplir con los items 1), 2), 3), 4) y 5), la solución del enunciado respectivo no será calificado.**

Problema 1. Leer el archivo de datos marketing.csv, almacenar el contenido en un DataFrame, **y proceder a resolver los siguientes enunciados:**

- a) Cargado el archivo marketing.csv en el Dataframe, mostrar, mediante DataFrame.describe(), las características resumen de las columnas que no son numéricas **(5%)**.
- b) Cargado el archivo marketing.csv en el Dataframe, mostrar, mediante DataFrame.describe(), las características resumen de las columnas que sí son numéricas. Mostrar los cuartiles, es decir, percentiles 25%, 50% y 75% **(5%)**.
- c) Cambiar los nombres de las columnas Year_Birth, Education y Marital_Status del DataFrame a Año_de_nacimiento, Educación y Estado_civil, respectivamente. Devolver una copia del DataFrame, de modo que el DataFrame original se mantenga inalterado **(5%)**.
- d) Determinar cuantas columnas del Dataframe tienen valores perdidos (valores faltantes) **(5%)**.
- e) Cambiar el tipo de datos de las columnas MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts y MntGoldProds del DataFrame, realizar al cambio al tipo de dato float64 **(5%)**.
- f) Obtener con una sola orden, la media, desviación y suma de la columna Income. Indexar los elementos resultantes utilizando las etiquetas Media, Desviación y Suma respectivamente **(5%)**.
- g) Utilizando una sola orden, obtener la media y desviación estándar para todas las columnas “Mnt” en **mnt_columns** **(5%)**.
- h) Mostrar los 5 clientes **(5 filas)** cuya suma de gastos (columnas en mnt_coluns) es mayor que el resto de clientes **(5%)**.
- i) Calcular el porcentaje de dinero que gasta cada cliente (columnas en mnt_columns) con respecto a sus ingresos **(5%)**.
- j) Calcular la media y desviación del 50% de los clientes que tienen un mayor salario **(5%)**.

Problema 2. Cargar el siguiente dataset con URL según se indica, almacenar el contenido en un DataFrame, y **proceder a resolver los siguientes enunciados:**

- a) Cargar el dataset con URL indicada, en un Dataframe, mostrar cuantas filas y columnas tiene el dataset, también mostrar todas las columnas y por cada columna mostrar el nombre de la columna según el siguiente formato, cambiar el tipo de dato con un identificador en Español: int64 por entero64, object por “contenedor general”, y remover la columna “Non-Null Count” **(5%)**.

No.	Columna	Tipo
---	-----	-----
0	age	int64
1	workclass	object
.....		
.....		

URL: <https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data>
<https://archive.ics.uci.edu/dataset/2/adult>
Utilizar como Nombre de las Columnas:
 columnas = [
 "age", "workclass", "fnlwgt", "education", "education-num", "marital-status",
 "occupation", "relationship", "race", "gender", "capital-gain", "capital-loss",
 "hours-per-week", "native-country", "income"

- b) Utilizando una sola orden, obtener información de cuantiles de las siguiente columnas **(5%)**:
 age hours-per-week capital-gain capital-loss
- c) Utilizando una sola orden, obtener información de Porcentaje de ingresos >50K por país.
 Utilizando una sola orden, obtener información de Porcentaje de ingresos >50K por nivel educativo **(5%)**.
- d) Utilizando una sola orden, obtener información de estadísticas de columnas no numéricas
 Utilizando una sola orden, obtener información de estadísticas de columnas numéricas **(5%)**.
- e) Transformar la columna edad (age), dado los siguientes rangos de edades, se debe mostrar las cadenas respectivas siguientes en la columna edad modificada (age) **(5%)**:

Edad (rango)	>= 0 y <25	>= 25 y <45	>= 45 y <65	>= 65 y <100
Edad (cadena)	Joven	Adulto	Adulto Maduro	Adulto Mayor

- f) Utilizando una sola orden, obtener información de Ocupaciones con mayor proporción de ingresos >50K **(5%)**.
- g) Mostrar un gráfico de frecuencias por edad **(5%)**.
- h) Mostrar en un solo gráfico, los dos gráficos de box-plot por edad e ingresos <=50 y >50 **(5%)**.
- i) Mostrar un gráfico de barras por ingresos según nivel educativo <=50 y >50 **(5%)**.
- j) Mostrar una matriz de gráficos de dispersión entre todas las variables numéricas del DataFrame del Dataset, en función de los ingresos <=50 y >50, se debe colorear los puntos de los gráficos según la categoría de esta columna ingresos **(5%)**.