

## “PROGRAMA ESPECIALIZADO CIENCIA DE DATOS CON PYTHON”

### PROYECTO N° 02



#### Ejercicio N° 1

Fichero **autos.csv**

Generar una función y utilizar esta para recodificar la variable “**país de origen**” (1, 2, 3) en el dataframe a través de la función **apply**.

#### Ejercicio N° 2

Fichero **autos.csv**

Generar un Notebook Colab con código y resultados, que contiene información de autos de un periodo de tiempo. Resolver los siguientes:

- Limpieza y tratamiento de las variables (Según corresponda) - Generar 5 nuevas variables aplicando los siguientes criterios con las variables numéricas:

1.  $\text{Log}_{10}(\text{var1}/\text{var2})$

2.  $\text{Sqrt}(\text{var1}) * \exp(\text{var2}) / 200$
3. Si  $\text{var 1} > \text{var 2} \rightarrow 5$ , caso contrario 3
4.  $1 / \log(\text{var1} / \text{var2}) * 100$
5.  $\text{var2}^{**2} / \text{var1}$
6.  $\text{var1} / \text{var2}^{**2}$

### Ejercicio N° 3

Fichero [autos.csv](#)

- **Crear una función** que permita generar 3 variables cualitativas ordinales a partir de las variables numéricas con los siguientes criterios:

- ★ Si es mayor a la “media + 1\*desviación estándar” → Alto ★
- Si está entre “media - 1\*desviación estándar” y “media + 1\*desviación estándar” → Medio
- ★ Si es menor a la “media - 1\*desviación estándar” → Bajo

Nota: Mostrar con las nuevas variables generadas.

Elegir con libertad las variables a trabajar.

*La función debe contar con su respectiva documentación.*

### Ejercicio N° 4

Fichero: [datos\\_seguros.csv](#)

- **Crear una función en Python** que muestre los siguientes gráficos (simultáneamente) para las variables numéricas, según una variable cualitativa:

1. Cajas y bigotes
2. Barras de la media
3. Histograma con densidad (Curva)
4. Gráfico de mapa de calor de la media de una tabla de contingencia (doble entrada)

*Debe graficar los 4 gráficos.*

- *La función debe contar con su respectiva documentación.*

## Ejercicio N° 5

- **Crear una función en Python** que devuelva un mapa de calor con la correlación de todas las variables numéricas (incluye las generadas), según país.

- *La función debe contar con su respectiva documentación.*

## Ejercicio N° 6

Fichero: [datos\\_seguros.csv](#)

- **Crear una función en Python** que devuelva las siguientes estadísticas

- *Media*
- *Mediana*
- *Moda*
- *Trimedia*
- *Varianza*
- *Media Armónica*
- *Coeficiente de Variación*
- *Desviación estándar*
- *Percentiles: 10, 25, 35, 50, 65, 75, 90 y 95*
- *Mínimo*
- *Máximo*

**Proyecto N°2:** Se adjunta Data:

1. Resolver los ejercicios

2. Entregable: generar un informe en Python Notebook, debidamente explicado, paso a paso; Opcional: subir a GitHub o un repositorio similar (enviar el link). 3. Subir un video en Youtube sobre la solución de los casos, explicando; compartir link.



**Entrega:** Subir al Google Classroom “Proyecto N° 02”

\*. Entrega del proyecto: personal.

Fecha de entrega máxima: 10 de diciembre de 2024

Instructora: Msc (e) Ing. Est. E Inf. Yaneth Ccahuantico Mendoza