

EXAMEN PYTHON BÁSICO

Indicaciones Generales:

Entrega : hasta LAS 12:00 p.m. del sábado 02 de noviembre.

Materiales para utilizar: Está permitido el uso de apuntes. Debe usar Python para resolver los ejercicios planteados (incluir el código y programas usados para resolver las preguntas).

Al término de este examen el alumno deberá enviar la solución al correo msiguenas@socialdata-peru.com. El archivo Word debe ser guardado con su apellido paterno y materno sin espacios.

Puntaje Total : 20 puntos

CASO 1:

Teniendo en cuenta la base de datos TRAIN_FUGA.csv que corresponde a los datos de entrenamiento de una data que analiza la fuga de clientes de una entidad bancaria realizar las siguientes tareas:

1. Análisis exploratorio de las variables: medidas y visualización. (5 puntos)
2. Realizar un análisis exploratorio sobre presencia de outliers. (2 puntos)
3. Realizar una discretización de las variables : INGRESO_BRUTO_M1 y EDAD teniendo en cuenta al menos dos técnicas de discretización no supervisada y agregar las variables discretizadas a nuestro conjunto de datos original (2 puntos)
4. Aplicar dos técnicas de balanceo de datos a nuestra variable TARGET (objetivo) y agregarlas a nuestro conjunto de datos original (3 puntos)

Nota: revisar el diccionario de variables del caso.

CASO 2:

La presente aplicación captura datos socioeconómicos a nivel distrital para la realización de un ejemplo de reducción de dimensiones haciendo uso del análisis de componentes principales y factorial.

Las variables a reducir son: porcentaje de hogares sin medios de comunicación (porc_hogares_sin_medios), porcentaje de alfabetismo (alfabetismo), porcentaje de hogares con 2 o más necesidades básicas incubiertas (porc_2_nbi), índice de desarrollo humano (IDH) y el coeficiente de desigualdad de GINI (GINI).

1. Realizar un análisis de componentes principales para reducción de la dimensionalidad (4 puntos)
2. Realizar un análisis factorial para reducción de la dimensionalidad (4 puntos)