

EVALUACIÓN N°1 (ABPRO)

Plan Formativo	Nivel de Dificultad:
Ciencia de Datos	Medio
Nombre del proyecto: Evaluación Módulo 3	Tema: Obtención y Preparación de datos
Objetivo del proyecto: (Competencias del Módulo):	Aplicar técnicas de obtención, limpieza y preparación de datos, utilizando criterios de imputación y manipulando las estructuras de datos a conveniencia para satisfacer las necesidades de información acorde al lenguaje Python.
<p>Ejecución: Grupal</p> <p>Usted cuenta con 3 archivos csv con los siguientes nombres:</p> <p>state-abbrevs.csv state-area.csv state-population.csv</p> <p>Con estos archivos realice los siguientes ejercicios:</p> <ol style="list-style-type: none">1. Cargue los tres datasets usando pandas.2. Revise los datos cargados y realice las modificaciones necesarias para empezar a trabajar3. Elimine las columnas que no aportan información4. Agrupe por state/region y ages, para luego agregar usando promedio, mínimo, máximo y contar5. Forme un nuevo dataset llamado df que tenga la información de los 3 sets importados. Use las llaves necesarias.6. Calcule la población por área para cada estado. Para esto agrupe por state/region y calcule el promedio de la población. Guarde dicha información en un data frame con las columnas 'state' y 'pop/area'	

(hint: puede extraer los valores de un objeto groupby, finalizando su sentencia con .values, o extraer sus etiquetas con .index)		
7. Finalmente, cree un dataframe que posea: El estado, el área, la población, la población por área y como índice la abreviación del nombre del estado.		
Requerimientos de los participantes		
Conocimientos previos <ul style="list-style-type: none">• Estructuras Algorítmicas• Estructura de Control• Condicionalidad• Expresiones Lógicas	Actitudes para el trabajo <ul style="list-style-type: none">• Cumplimiento de plazos• Buenas prácticas de codificación• Diseño y Estructura• Trabajo en equipo• Optimización del tiempo	Valores <p>Tiempo de resolución.</p> <p>Enfoque al requerimiento.</p> <p>Estructura de Solución.</p>
Objetivo General de Aprendizaje	El participante al finalizar el proyecto será capaz de: <p>Aplicar técnicas de obtención, limpieza y preparación de datos, utilizando criterios de imputación y manipulando las estructuras de datos a conveniencia para satisfacer las necesidades de información acorde al lenguaje Python.</p>	
Objetivos particulares	<ul style="list-style-type: none">- Desarrollo de algoritmos en pseudocódigo- Trabajo en equipo	
Duración del proyecto	1 jornada de clases	
Tips o listado de Preguntas Guía		
<ul style="list-style-type: none">- ¿Cuáles son los datos de entrada y de salida de mi algoritmo?- ¿Qué variables debo declarar?		

- ¿Qué pasos deben ser considerados en el algoritmo?
- ¿Cómo manipular datos perdidos?

Productos para obtener durante la realización del proyecto

Algoritmo desarrollado en Python según requerimientos planteados.

Especificaciones de desempeño

Deberá realizar la actividad según requerimientos técnicos y en un plazo máximo de 1 clase máx 3 personas por grupo; el resultado deberá ser un único archivo entregado en formato de Python con extensión .py o .ipynb

Sugerencias bibliográficas para la investigación

Limpieza de datos en python

<https://aprendeia.com/limpieza-y-procesamiento-de-datos-con-codigo-en-python/>

numpy

<https://aprendeia.com/introduccion-a-numpy-python-1/>

Operaciones con arreglos

<https://claudiovz.github.io/scipy-lecture-notes-ES/intro/numpy/operations.html>

valores perdidos

<https://www.youtube.com/watch?v=i-c80qYqbs&t=22s>