

Medellín, 17 de octubre de 2023

PROYECTO FINAL DE MACHINE LEARNING NIVEL I TALENTO ESPECIALIZADO 2023

ENUNCIADO:

Para el reto final tiene las siguientes opciones.

1. Diseñar una encuesta en redes sociales o algún medio que le permita recolectar los datos.
2. Identificar las respuestas en la base de datos del DANE o alguna organización que evidencie información confiable.

Para alguna de las opciones mencionadas, identifique la respuesta a una pregunta de su interés, por ejemplo:

¿Cuál es su red social preferida?

¿Cuál es su equipo de fútbol preferido?

¿Cuál es el género musical de su preferencia?

¿Cuál es su serie o película favorita?

¿Tiene vehículo?, ¿declara renta?

¿Hace deporte?

¿Tiene diabetes?

¿Ha viajado al exterior?

¿Ha tenido multas de tránsito?

Esta pregunta debe ir acompañada de otras preguntas obligatorias que se presentan a continuación. Usted deberá preguntarle estos datos a cada una de las personas encuestadas o buscar la información en la base de datos seleccionada.

Preguntas	Tipo de datos	Opciones	Descripción
Nombre completo	String		
Género	String	Masculino – Femenino – Otro	
Edad	Int		
Nacionalidad	String		
Ciudad de nacimiento	String		
Ciudad de residencia	String		

Nivel académico	String	Primaria – Bachillerato – Técnico – Tecnológico – Profesional – Posgrado	
Estado civil	String	Soltero – Casado – Unión libre – Viudo	
Hijos	Int		Cantidad de hijos
Empleo	String	Formal – Informal – No posee	
Vivienda	String	Propia – Arrendada – Familiar	
Estrato socioeconómico	Int	1 – 2 – 3 – 4 – 5 - 6	Puede ser el registrado por el SISBEN para Colombia o el de la vivienda

Los datos de la tabla anterior, no pueden ser la pregunta de su interés, ya que dicha respuesta le permitirá con estos datos, determinar condiciones demográficas claras para analizar la respuesta a dicha pregunta inicial.

Procure que la respuesta a la pregunta de su interés tenga unas opciones definidas, esto le ayudará a filtrar mejor la información.

Tareas

1. Plantee un objetivo general que le permita identificar el por qué quiere conocer esta información, para su posterior análisis.
2. Diseñe la encuesta en un formulario de Google, Office o el que crea conveniente o, si seleccionó la opción 2, guarde la evidencia de los datos encontrados.
3. Encuentre un mínimo de 50 personas, procure siempre la veracidad de los datos.
4. Almacene la base de datos en un documento de Excel o similar.
5. Cree un archivo json o csv con la base de datos del archivo de Excel.
6. Diseñe una aplicación en Python donde pueda analizar los datos.
7. Desarrolle las siguientes tareas para cada una de las variables creadas con cada pregunta, la que usted planteó y las de la tabla:

1. Tabla de frecuencias
2. Medidas de tendencia central (Moda, Media y Mediana)
3. Medidas de dispersión (Varianza y Desviación estándar)
4. Medidas de posición (Cuartiles, quintiles y percentiles)
5. Gráficos (Barras, histogramas, tortas)

8. Desarrolle en su aplicación un análisis de variables de acuerdo con las siguientes preguntas, para identificar la pregunta que usted planteó usaremos la expresión {x}

1. ¿Cuántas personas {x} son mujeres?
2. ¿Cuántas personas {x} son hombres?
3. ¿Cuántas personas {x} son de otro género?

4. ¿Cuántas personas {x} tienen más de 18 años y menos de 30 años?
5. ¿Cuántas personas {x} tienen más de 30 años y menos de 50 años?
6. ¿Cuántas personas {x} tienen más de 50 años?
7. ¿Cuántas personas {x} son colombianas?
8. ¿Cuántas personas {x} son extranjeras?
9. ¿Cuántas personas {x} residen en su misma ciudad de nacimiento?
10. ¿Cuántas personas {x} tienen determinado nivel académico?
11. ¿Cuántas personas {x} tienen determinado nivel académico?
12. ¿Cuántas personas {x} tienen determinado estado civil?
13. ¿Cuántas personas {x} no tienen hijos?
14. ¿Cuántas personas {x} tienen 1 hijo?
15. ¿Cuántas personas {x} tienen 2 hijos o más?
16. ¿Cuántas personas {x} tienen empleos informales?
17. ¿Cuántas personas {x} tienen empleos formales?
18. ¿Cuántas personas {x} no tienen empleo?
19. ¿Cuántas personas {x} tienen casa propia?
20. ¿Cuántas personas {x} residen en casa arrendada?
21. ¿Cuántas personas {x} residen en casa familiar?
22. ¿Cuántas personas {x} son de determinado estrato socioeconómico?

Si tiene presente el objetivo inicialmente planteado para su pregunta de interés, será fácil sacar conclusiones al respecto del análisis en cada numeral anterior.

9. Diseñe gráficos en su aplicación de Python para presentar un informe de conclusiones al respecto del proyecto.

Utilice las librerías Numpy, pandas, Matplotlib y todas las que crea convenientes para la elaboración de las anteriores tareas. Use el paradigma funcional para simplificar su código y documéntelo para su fácil lectura e interpretación.

Criterios de evaluación

- Recopilación de los datos - 5%
- El código funciona - 30%
- Uso del paradigma funcional - 15%
- Uso de las librerías trabajadas en el curso - 35%
- Documentación del código - 5%
- Conclusiones del análisis - 10%