

Big Data - Analisis Exploratorio de Datos

Utilizando la herramienta [Orange Data Mining](#), importe el archivo de datos hogares.csv y elabore un informe con las respuestas a las preguntas indicadas en esta guía.

Hogares.csv contiene datos ficticios de 100 hogares encuestados.

Condiciones de entrega

- Entrega Grupal: Deben entregarlo por Github con el nombre del grupo
- Plazo máximo: hasta la siguiente clase.

Diccionario de Datos

HogarID: Identificador único del hogar, entero que va de 1 a 100.

IngresoMensual: Ingresos totales del hogar por mes, en pesos argentinos (ARS). Representa la suma de todos los ingresos de los miembros del hogar.

GastoMensual: Gastos totales del hogar por mes, en pesos argentinos (ARS). Incluye gastos fijos (como alquiler o hipoteca, servicios, etc.) y gastos variables (como alimentación, transporte, ocio).

NumPersonas: Número de personas que conforman el hogar. Valor entero que indica la cantidad de individuos viviendo en el hogar.

TipoVivienda: Tipo de estructura habitacional donde reside el hogar. Puede ser "Casa" o "Departamento", reflejando las diferentes modalidades de vivienda.

ZonaVivienda: Zona geográfica donde se ubica el hogar. Las categorías incluyen "Urbana", "Suburbana" y "Rural", indicando el contexto geográfico y posiblemente socioeconómico de la vivienda.

EdadPromedioHogar: Edad promedio de los miembros del hogar, en años. Esta variable es un promedio de las edades de todos los individuos que conforman el hogar.

Guía de Preguntas

1. Análisis de Tendencia Central y Dispersión:

Calcule las medidas de tendencia central (media, mediana) y dispersión (varianza, desviación estándar) para las variables IngresoMensual, GastoMensual, y EdadPromedioHogar. ¿Los valores encontrados son consistentes con lo que esperarías en un contexto económico real?

2. Verificación de Máximos y Mínimos:

Identifique los valores máximos y mínimos para IngresoMensual, GastoMensual, y EdadPromedioHogar. ¿Estos valores son realistas? Discuta cualquier valor que parezca inusual o fuera de lo esperado.

3. Relación entre Variables:

Explore la relación entre el EdadPromedioHogar y otras variables numéricas como IngresoMensual y GastoMensual. ¿Hay alguna tendencia visible que indique cómo la edad promedio del hogar afecta los ingresos y gastos?

4. Análisis de Outliers:

Utilizando técnicas gráficas, identificar datos atípicos en las variables numéricas. Describa por qué estos puntos podrían considerarse atípicos y qué factores podrían explicar estos valores.

5. Comparación entre Grupos:

Analice cómo varía la EdadPromedioHogar entre diferentes tipos de vivienda (Casa, Departamento) y zonas de vivienda (Urbana, Suburbana, Rural). ¿Existen diferencias significativas que puedan sugerir patrones demográficos?

6. Correlaciones:

Investigue la correlación entre las variables numéricas. ¿Qué relaciones se pueden identificar como relevantes y cómo se interpretan?