

## 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?

IngresoMensual:

- Media: 91719,8818
- Mediana: 91055,67
- Dispersión: 0,3550

GastoMensual:

- Media: 62151,3
- Mediana: 62352,3
- Dispersión: 0,401912

EdadPromedioHogar:

- Media: 40,192
- Mediana: 38
- Dispersión: 0,411913

Los valores encontrados no son consistentes con un contexto económico realista ya que en cuanto al ingreso y gasto esperaríamos que sea más elevado y la edad teniendo en cuenta que si en la casa hay hijos debería ser más baja

### 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.

IngresoMensual:

- Máximo: 148557,56
- Mínimo: 40607,43

GastoMensual:

- Máximo: 127691
- Mínimo: 20623,9

EdadPromedioHogar:

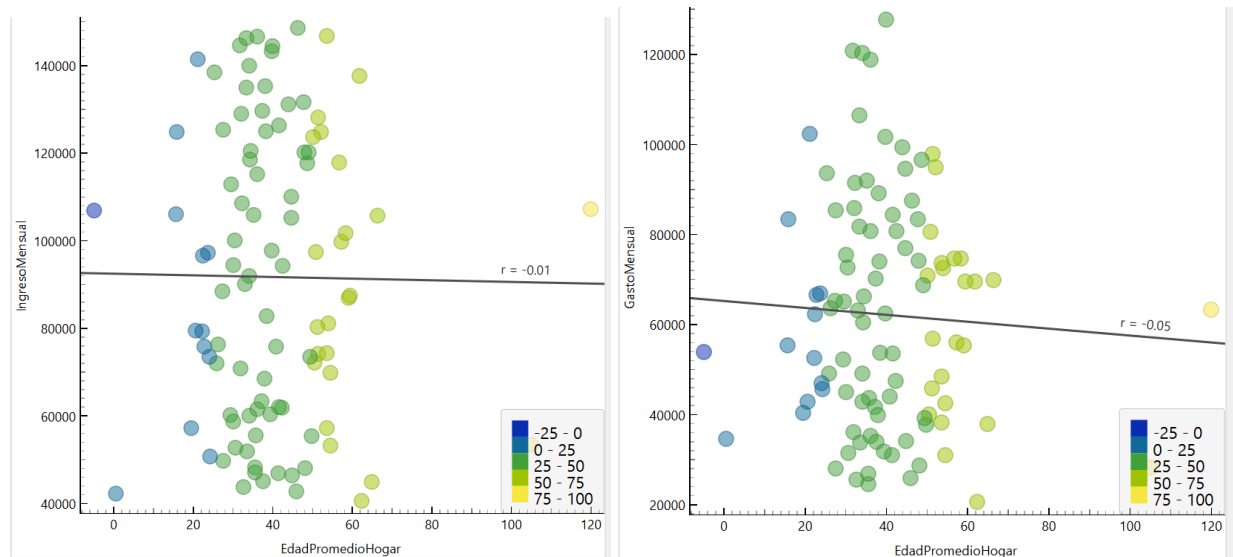
- Máximo: 120
- Mínimo: -5

No consideramos que se trate de valores realistas. La edad no puede ser un valor negativo, como es el caso del mínimo de esta variable, y tampoco es realista que el promedio del hogar sea superior a los 100 años. A su vez, en el contexto socioeconómico en el que se encuentra el país actualmente, los valores tanto de gastos como de ingresos, mínimos y máximos, no van

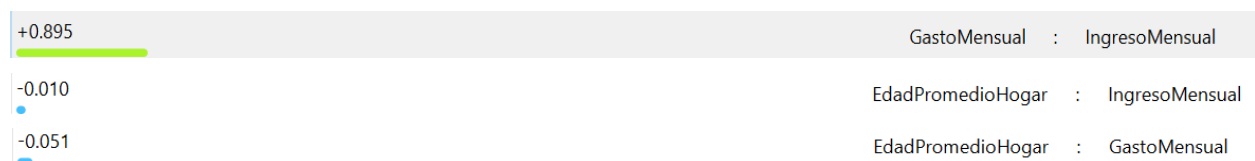
de la mano con la realidad, sabiendo que el salario mínimo en la Argentina ronda los 300 dólares.

### 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?



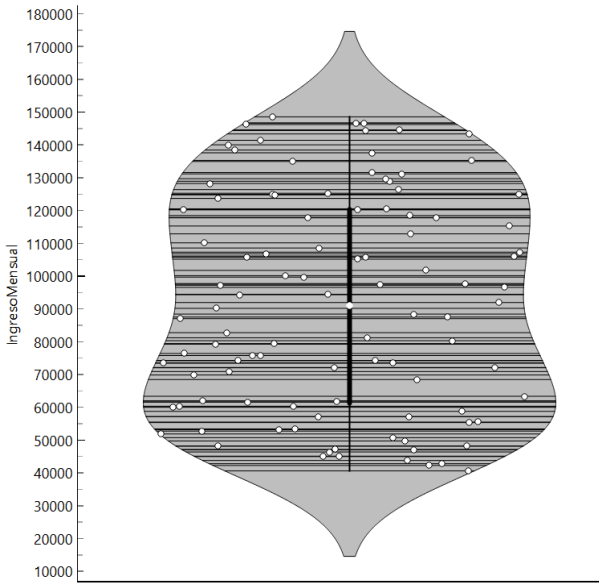
Utilizamos el Scatter Plot para analizar la correlación entre las variables. Como podemos ver en los gráficos, no existe una relación fuerte entre las variables de la Edad con las de los Ingresos y los Gastos. No necesariamente alguien mayor cobra más que alguien menor, o viceversa. El data base presenta mayoría de personas en un rango etario entre 25-50, y este grupo tiene desde los mínimos ingresos y gastos hasta los valores máximos de estas variables. A su vez, el coeficiente de correlación R es negativo y muy bajo, cercano a 0, en ambos casos.



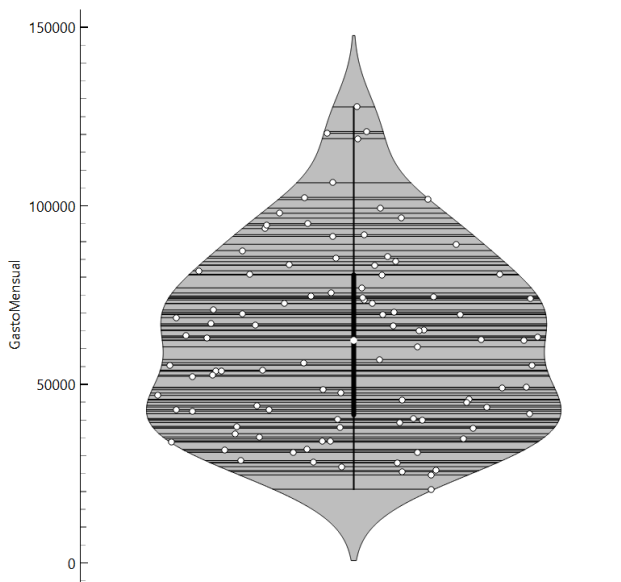
A su vez, utilizamos Correlations para analizar la correlación entre variables. Lógicamente, Gastos e Ingresos están fuertemente correlacionadas. Sin embargo, estas variables en relación a la edad de las personas no tienen una correlación fuerte, por ende concluimos que la edad promedio del hogar no influye en los ingresos y gastos de los hogares.

### 4. Análisis de Outliers:

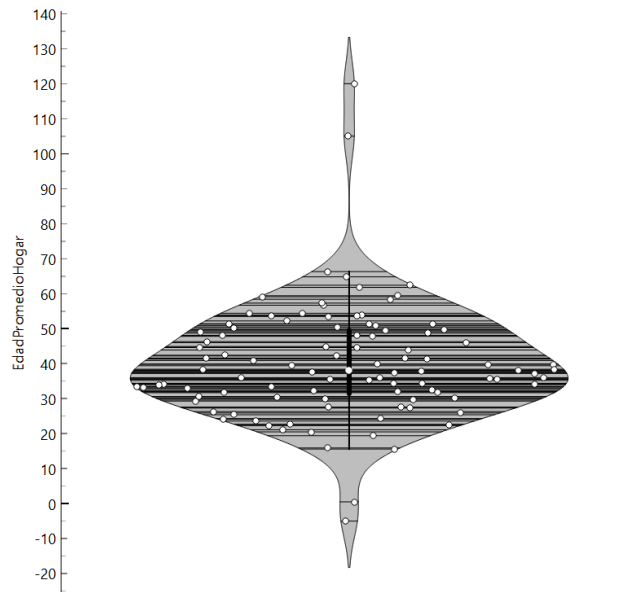
Utilizando técnicas gráficas, identificar datos atípicos en las variables numéricas. Describa por qué estos puntos podrían considerarse atípicos y qué factores podrían explicar estos valores.



*Para analizar la existencia de outliers, utilizamos el Violin Plot sobre la variable Ingreso Mensual, donde los puntos blancos representan registros individuales. En este caso, vemos cómo en algunas líneas hay puntos únicos, como es en los casos de los valores máximos y mínimos del ingreso, y representan hogares poco comunes (ej. ingresos extraordinariamente bajos o altos). La distribución de ingresos está concentrada entre 40.000 y 140.000. Los hogares con ingresos más altos (a la derecha) son menos frecuentes, lo que genera una cola más fina del violín.*



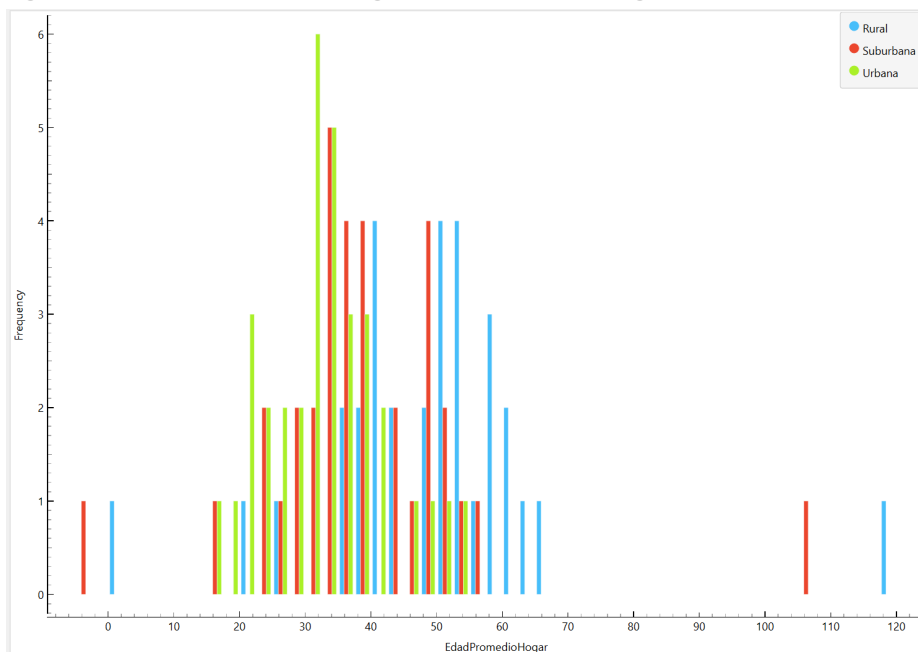
*Utilizamos la misma herramienta para analizar la variable Gasto Mensual y sus outliers. Se notan puntos aislados por encima de 120.000, que representan hogares con gastos muy altos. También existen algunos valores muy bajos, menores a 20000 aproximadamente, que se alejan de los valores normales de nuestro data base.*



*Por último, decidimos utilizar la herramienta con la variable EdadPromedioHogar, y obtuvimos que en nuestra base de datos, tenemos edades negativas (-5 años), lo cual es imposible, e indica posible error de registro. Del mismo modo existen edades demasiado altas, mayores a 100 años, cerca de los 130. Resultan poco realistas como promedio de un hogar, siendo datos mal cargados o outliers.*

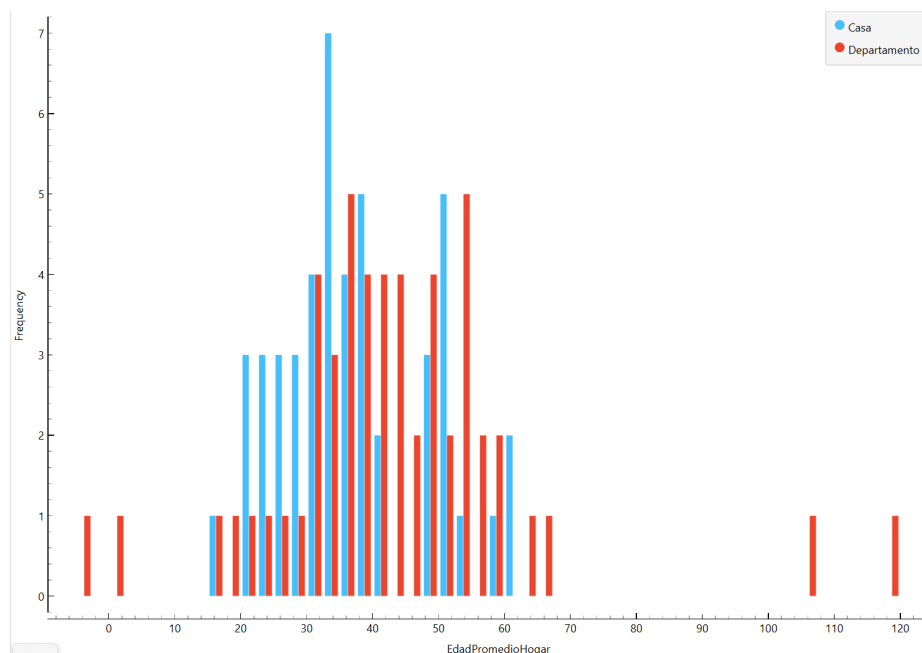
## 5. Comparación entre Grupos:

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?



Utilizamos la herramienta de Distribution para analizar cómo varía la EdadPromedioHogar entre las diferentes zonas de las viviendas. Vimos cómo las tres zonas urbanas se superponen, pero existen ciertos casos a mencionar. La zona urbana se ve más concentrada en el rango de edad entre 20 y 45 años, con una alta predominancia en los 35 años. La zona rural se ve más concentrada entre edades de entre 40 y 60 años. Por último, la zona suburbana se ve más concentrada entre 35 y 50 años.

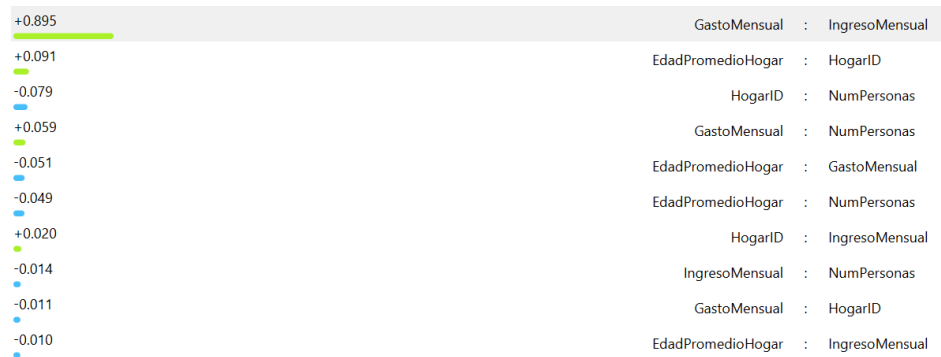
Además, los valores extremos (0 y mayores a 100) los interpretamos como outliers, y no casos realistas.



Utilizamos la misma herramienta para analizar la relación entre la EdadPromedioHogar y el tipo de vivienda, ya sea una casa o un departamento. Las casas se concentran más en los 30 y 40 años, con un pico fuerte alrededor de los 35 años. Por otro lado, los departamentos tienen más presencia en el rango 30 y 50 años, con varios picos en los 35 y los 55 años. Al igual que en el caso anterior, consideramos como outliers aquellos datos que concentran a los departamentos entre edades menores a 0 y mayores a 100.

## 6. Correlaciones:

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



La única relación fuerte y relevante es entre IngresoMensual y GastoMensual (+0.895), como mencionamos en el punto 3. Demuestran un patrón esperable, donde los ingresos determinan la capacidad de gasto de los hogares.

Las demás correlaciones son muy bajas (cercanas a 0), lo que indica que ni la edad promedio del hogar ni el número de personas tienen un impacto directo claro sobre los ingresos o los gastos en este dataset.