

INTERPRETACIONES TP

Punto 1

Interpretación Histograma

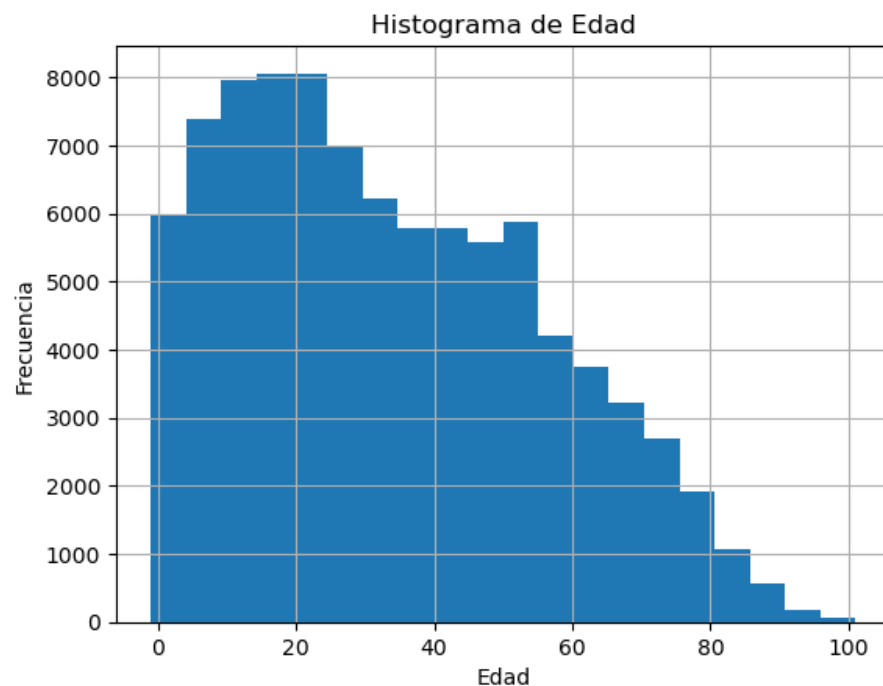
El histograma representa la **distribución de los salarios semanales** en la muestra de datos. Cada barra muestra la **frecuencia** de un rango de valores de salario, permitiendo visualizar cómo están distribuidos los ingresos de los trabajadores.

◆ Aspectos clave:

- **Alta concentración de valores bajos:** Se observa que la mayoría de los trabajadores tienen salarios bajos, lo que sugiere que los ingresos en la región no están distribuidos equitativamente.
- **Sesgo a la derecha:** La distribución presenta una **cola larga hacia la derecha**, lo que significa que hay una pequeña cantidad de individuos con salarios significativamente altos.
- **Posibles valores extremos:** Si hay barras alejadas del resto, estos podrían ser **casos atípicos** o trabajadores con ingresos fuera de lo común.

✦ Conclusión:

Este histograma refleja una distribución típica de los salarios en economías con alta disparidad de ingresos. La mayoría de los trabajadores tienen **salarios bajos**, mientras que unos pocos cuentan con ingresos elevados, generando una distribución sesgada.



Interpretación Distribución de Kernel

El gráfico de **distribución Kernel del salario semanal** muestra una representación suavizada de la dispersión de los ingresos en la población analizada. A diferencia del histograma, la distribución Kernel permite visualizar la **forma de la distribución** sin depender del número de intervalos o bins.

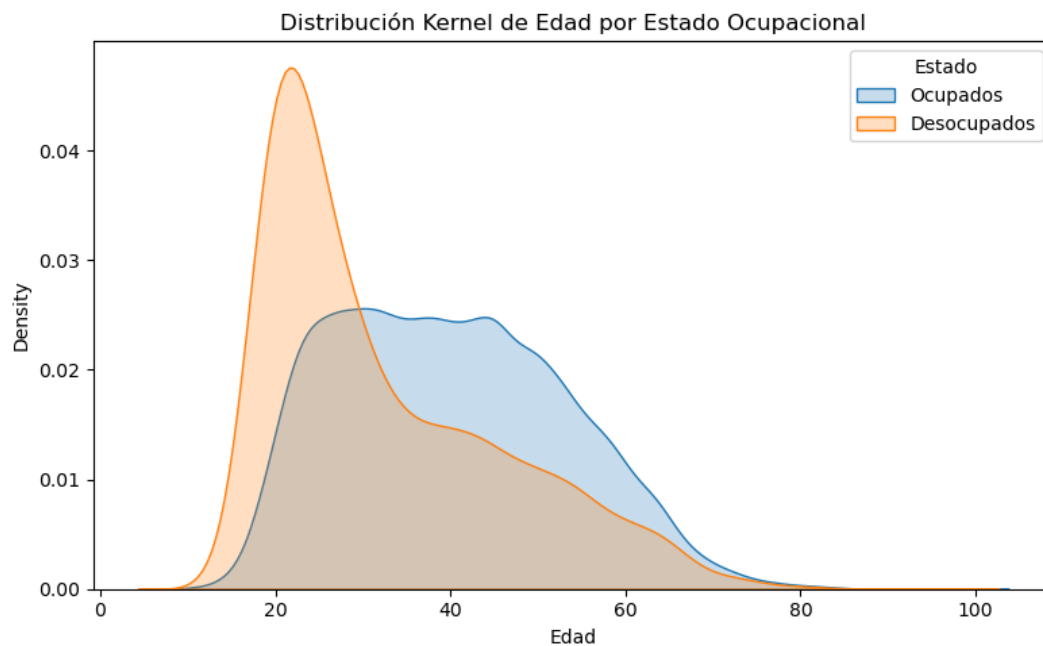
◆ Aspectos clave:

- **Alta concentración inicial:** La curva muestra una gran cantidad de trabajadores con salarios bajos, reflejando una desigualdad en la distribución de ingresos.
- **Sesgo a la derecha:** Se observa una cola larga hacia valores más altos, indicando que hay pocos trabajadores con salarios significativamente superiores.
- **Menos variabilidad visual que el histograma:** Al ser un método suavizado, la curva facilita la interpretación de la distribución general sin afectar la precisión de los datos.

📌 Conclusión:

El gráfico confirma que **la mayoría de los trabajadores tienen salarios bajos**, mientras que una minoría percibe ingresos más altos. Esta asimetría en la distribución es común en estudios de mercado laboral y refleja la desigualdad

salarial dentro de la población analizada.



PUNTO 2

Interpretación Distribución de Educación

Hemos limpiado los datos de educación, eliminando valores fuera del rango lógico (0 a 25 años). Ahora los nuevos resultados reflejan mejor la realidad de la muestra.

◆ Aspectos clave en la distribución:

- **Promedio (mean = 9.28 años)** → En promedio, las personas tienen **9.28 años de educación**, lo que sugiere que muchos no completaron la educación secundaria.
- **Desviación estándar (std = 5.63 años)** → Existe **alta variabilidad**, indicando que hay individuos con niveles educativos muy diferentes.
- **Mínimo (min = 0 años)** → Algunos no tienen educación formal, reflejando posibles **barreras educativas** o acceso limitado.
- **Mediana (p50 = 12 años)** → La mitad de la muestra tiene **12 años o menos de educación**, lo que indica que **la mayoría finalizó la secundaria**.
- **Máximo (max = 21 años)** → Existen individuos con **hasta 21 años de educación**, alcanzando estudios universitarios avanzados.

✍ Conclusión:

La educación presenta **gran diversidad** en la muestra, con una fuerte

concentración en el rango de secundaria, pero también casos sin acceso educativo formal. La limpieza de los datos nos permite ver una distribución más coherente, destacando las diferencias entre los distintos niveles educativos en la población.

Promedio: 9.28

Desviación estándar: 5.63

Mínimo: 0.0

Mediana (p50): 12.0

Máximo: 21.0

PUNTO 3

Interpretación Histograma del Salario Semanal (Panel A)

El histograma del salario semanal nos permite visualizar la frecuencia con la que se presentan diferentes niveles de ingreso en la población analizada.

◆ Aspectos clave:

Alta concentración de valores bajos: La mayoría de los trabajadores tiene salarios bajos, indicando una distribución desigual de ingresos.

Sesgo a la derecha: La distribución presenta una cola larga, reflejando que pocos individuos tienen salarios muy altos.

Posibles valores extremos: Si aparecen barras separadas del resto, podrían ser casos atípicos o registros con ingresos fuera de lo común.

✦ Conclusión:

Este histograma sugiere que el mercado laboral tiene una alta desigualdad salarial, donde la mayoría recibe ingresos bajos mientras que unos pocos logran salarios significativamente altos.

Interpretación Distribución de Kernel (Panel B)

El gráfico de distribución Kernel muestra de forma suavizada la dispersión de salarios, permitiendo visualizar diferencias entre ocupados (Estado == 1) y desocupados (Estado == 2).

◆ Aspectos clave:

Diferencias entre ocupados y desocupados:

Los ocupados tienen una mayor concentración de salarios en valores medios y altos.

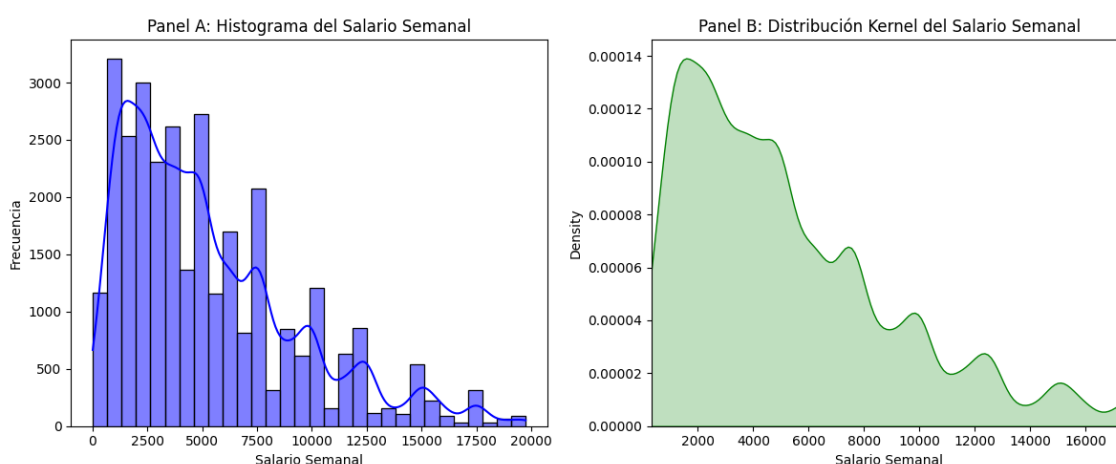
Los desocupados presentan una distribución más baja, reflejando ingresos limitados o ausencia de salario.

Sesgo a la derecha en ambas curvas: Se confirma que la mayoría de los trabajadores ganan poco, mientras que unos pocos reciben salarios elevados.

Variabilidad entre ambos grupos: La separación entre las curvas sugiere que la condición laboral afecta significativamente la estructura salarial.

✚ Conclusión:

La distribución Kernel refuerza la desigualdad observada en el histograma, mostrando que los desocupados tienen menores ingresos, y los ocupados presentan una curva más extendida hacia valores altos, indicando mayores oportunidades de salario dentro del empleo formal.



PUNTO 4

Interpretación de los resultados de horas trabajadas

La distribución de horastrab muestra una gran heterogeneidad en el mercado laboral.

◆ **Los trabajadores con pocas horas predominan:**

La mediana es de 6 horas semanales, lo que indica que la mitad de la población trabaja menos de lo que se consideraría una jornada laboral completa. Esto puede reflejar empleo informal, trabajos de medio tiempo o situaciones de subempleo.

◆ **Variabilidad extrema en la carga laboral:**

La desviación estándar es alta, lo que significa que las diferencias en las horas trabajadas entre individuos son muy marcadas. Mientras unos trabajan muy poco, otros tienen jornadas extendidas, con un máximo de 83 horas semanales, lo cual sugiere condiciones laborales exigentes en ciertos sectores.

◆ **Polarización en la distribución:**

Si bien hay una concentración de trabajadores con pocas horas, también existen individuos con una alta carga laboral, generando una estructura de trabajo polarizada, donde unos apenas trabajan y otros tienen jornadas prolongadas.

✂ **Conclusión:**

Los datos sugieren que el mercado laboral no ofrece un esquema uniforme de empleo. La disparidad en las horas trabajadas puede deberse a diferencias en el acceso a empleos formales, condiciones laborales específicas de ciertos sectores o situaciones económicas que limitan la estabilidad del empleo.

PUNTO 5

Tabla 1. Resumen de la base final para la región Gran Buenos Aires

	2004	2024	Total
Cantidad observaciones	45.289	46.050	91.339
Cantidad de observaciones con Nan en la variable "Estado"	0	0	0
Cantidad de Ocupados	17.022	20.325	37.347
Cantidad de Desocupados	2.717	1.362	4.079
Cantidad de variables limpias y homogeneizadas	183	183	183

PARTE 2

Punto 1:

El análisis de la matriz de correlación muestra las relaciones entre variables clave como edad, edad al cuadrado (edad2), educación (educ), salario semanal (salario_semanal) y horas trabajadas (horastrab).

◆ Interpretación de los resultados:

Edad y Edad² (edad y edad2) → Como era de esperar, tienen una correlación casi perfecta (0.96), lo que confirma que edad2 simplemente amplifica la variabilidad de edad.

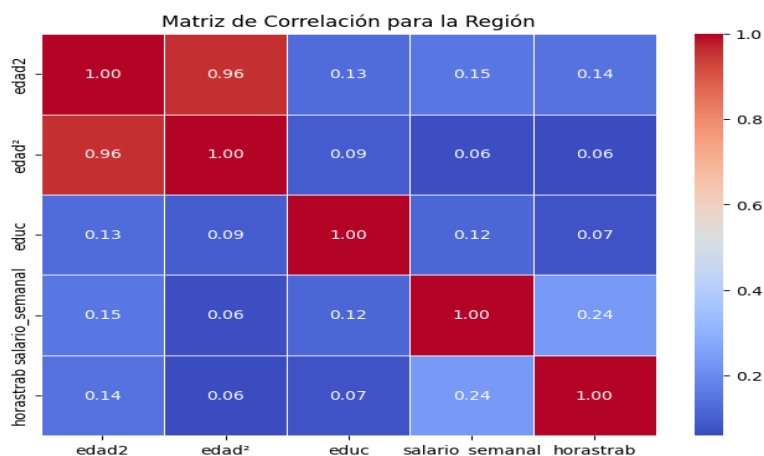
Educación (educ) y Salario Semanal (salario_semanal) → Hay una correlación positiva (0.12), lo que sugiere que más educación está asociada con salarios más altos, aunque la relación no es extremadamente fuerte.

Horas trabajadas (horastrab) y Salario Semanal (salario_semanal) → Presentan una correlación moderada (0.24). Esto significa que quienes trabajan más horas tienden a recibir un salario mayor, pero no es la única variable influyente.

Edad y Horas Trabajadas → La correlación es muy baja (0.06), indicando que la edad no influye significativamente en la cantidad de horas trabajadas, al menos en este conjunto de datos.

✚ Conclusión:

El análisis confirma algunas relaciones esperadas, como la influencia de la educación en el salario y la relación entre edad y edad². Sin embargo, las correlaciones no son extremadamente fuertes, lo que indica que otros factores también juegan un papel importante en los ingresos y horas de trabajo.

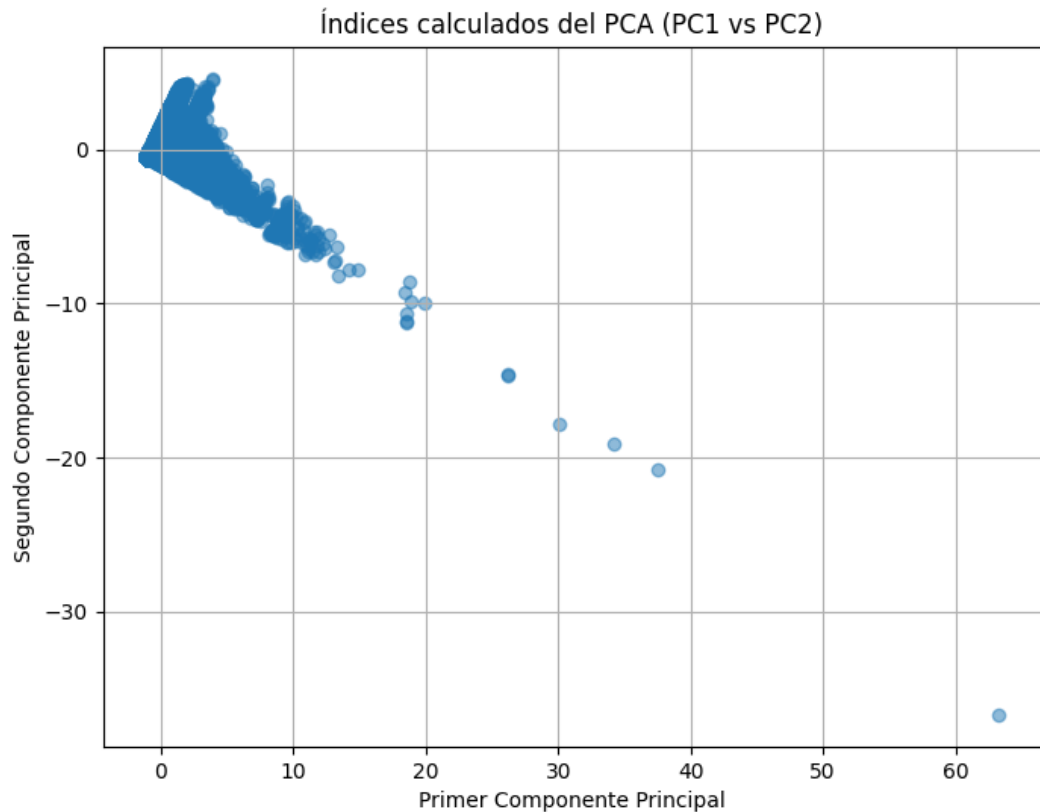


Punto 2:

El gráfico muestra los **índices calculados del PCA**, comparando el **Primer Componente Principal (PC1)** en el eje X y el **Segundo Componente Principal (PC2)** en el eje Y.

✦ **Análisis de la distribución:**

- La mayoría de los puntos están concentrados cerca del origen, lo que sugiere que muchas observaciones tienen valores bajos en ambos componentes.
- Algunos puntos se extienden hacia la derecha (valores altos en PC1) y hacia abajo (valores negativos en PC2), lo que podría indicar la presencia de **grupos diferenciados** o **outliers** en los datos.
- La amplitud en los ejes sugiere que PC1 tiene una mayor variabilidad en comparación con PC2.

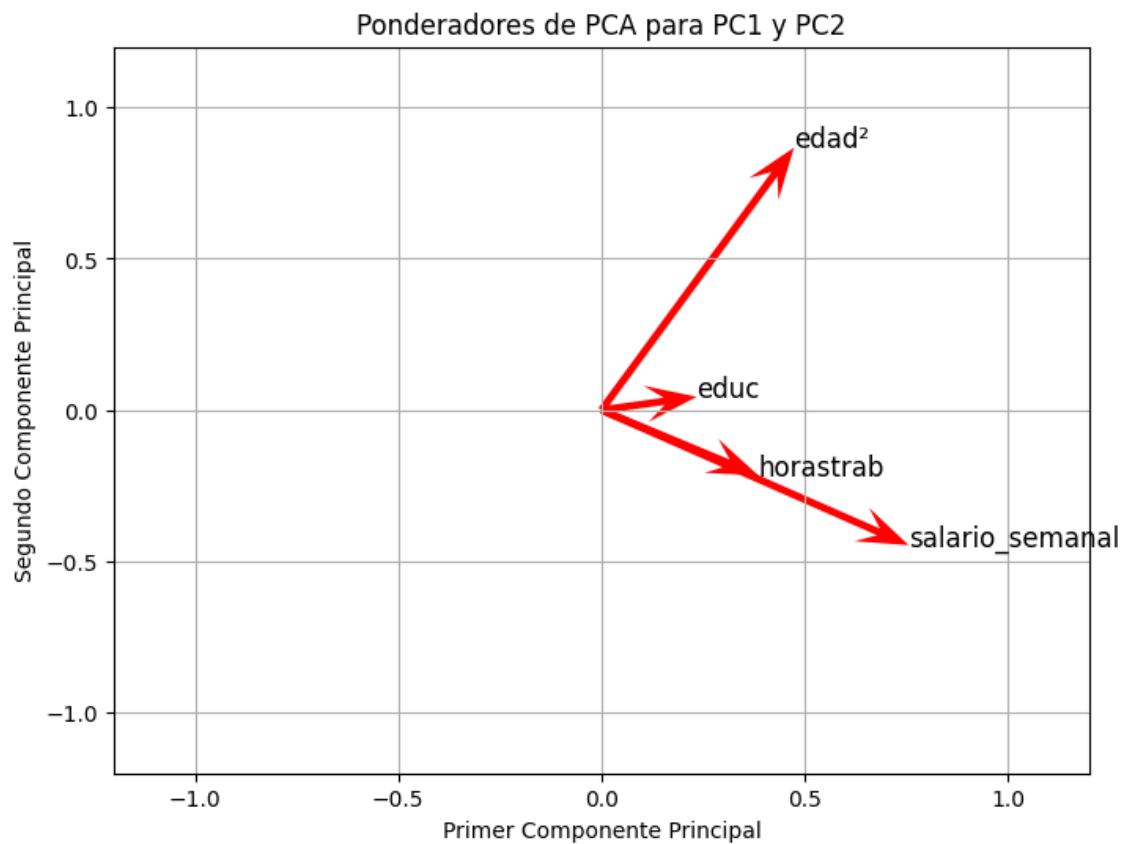


PUNTO 3:

El gráfico de **Componentes Principales (PCA)** ilustra la relación entre las variables edad^2 , educ, horastrab y salario_semanal con los dos primeros componentes principales (PC1 y PC2).

Relación entre variables →

- educ y salario_semanal tienen una orientación similar, lo que sugiere que están **correlacionadas en este espacio de PCA**.
- edad^2 apunta en una dirección diferente a horastrab, lo que podría indicar que la edad cuadrática tiene un comportamiento distinto respecto al tiempo trabajado.



PUNTO 4:

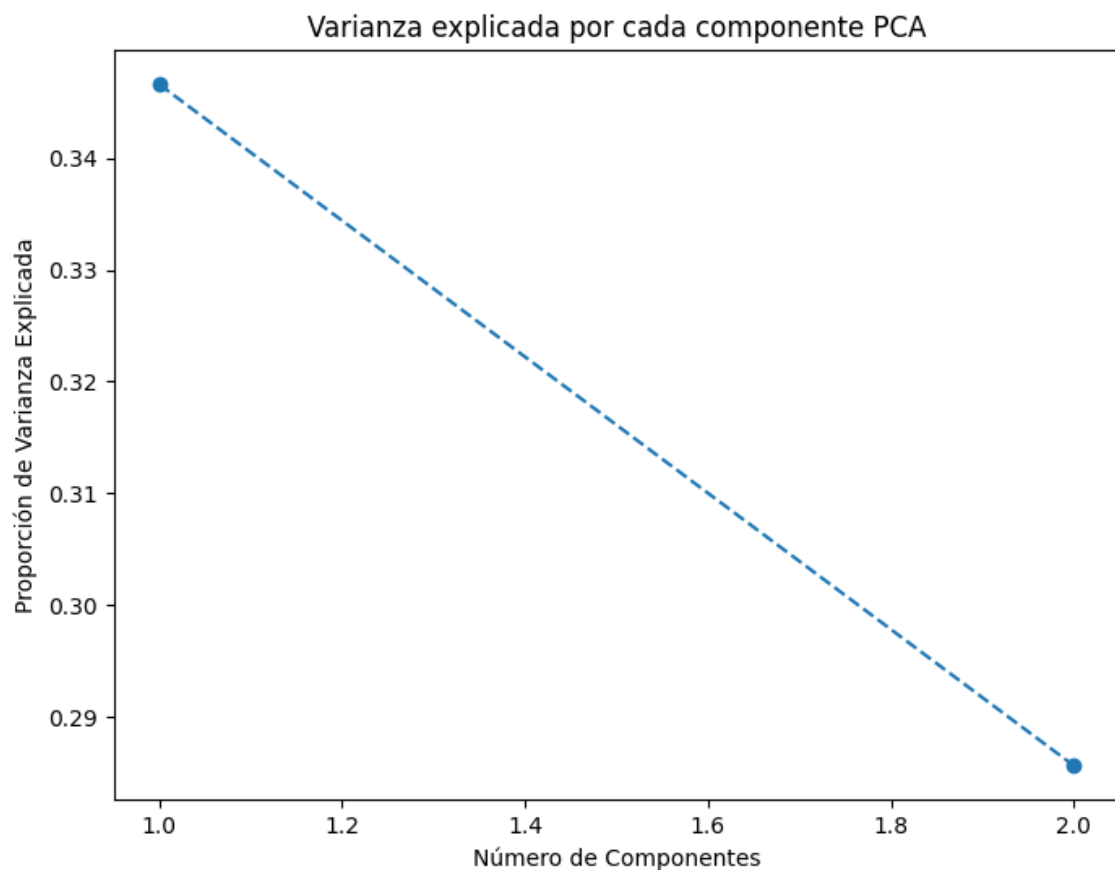
El gráfico muestra la **proporción de varianza explicada** por cada componente principal en el **PCA (Análisis de Componentes Principales)**.

Interpretación de los resultados:

Componente 1: Explica aproximadamente **34%** de la variabilidad total de los datos. Esto indica que este primer componente captura la mayor parte de la información presente en las variables originales.

Componente 2: Explica aproximadamente **29%** de la variabilidad. Aunque tiene menos peso que el primer componente, sigue aportando una cantidad significativa de información.

Disminución de la varianza explicada: Como se observa en la línea discontinua, **PC2 explica menos varianza que PC1**, lo que es común en PCA, ya que los componentes se ordenan de mayor a menor contribución a la variabilidad total.



PUNTO 5 Y 6:

El gráfico muestra los resultados de **K-Means clustering** con diferentes valores de **k** (número de grupos). Se observa cómo los puntos de datos se agrupan en función de **Edad** (x-axis) y **Educación** (y-axis).

Análisis de los resultados:

K=2: Se generan **dos clusters**, lo que sugiere una separación más general entre grupos. Es posible que los datos se dividan en **baja educación vs. alta educación** o **jóvenes vs. mayores**.

K=4: Se observan **cuatro clusters**, lo que permite una mayor diferenciación entre los grupos. Esto podría reflejar niveles educativos más específicos o distintas etapas laborales.

K=10: La segmentación es mucho más detallada, con **clusters más pequeños**, lo que podría capturar diferencias en educación o edad con mayor precisión. Sin embargo, demasiados clusters pueden generar ruido en los datos.

Un **dendrograma** es un gráfico que representa la estructura de un **clustering jerárquico**, mostrando cómo los datos se agrupan de manera progresiva.

Características principales:

Estructura de árbol → Cada unión en el gráfico representa la fusión de dos grupos en un nivel superior.

Distancia entre nodos → La altura de las conexiones indica qué tan similares son los grupos antes de combinarse.

Uso en clustering → Se usa para decidir el número óptimo de clusters en un análisis jerárquico.

