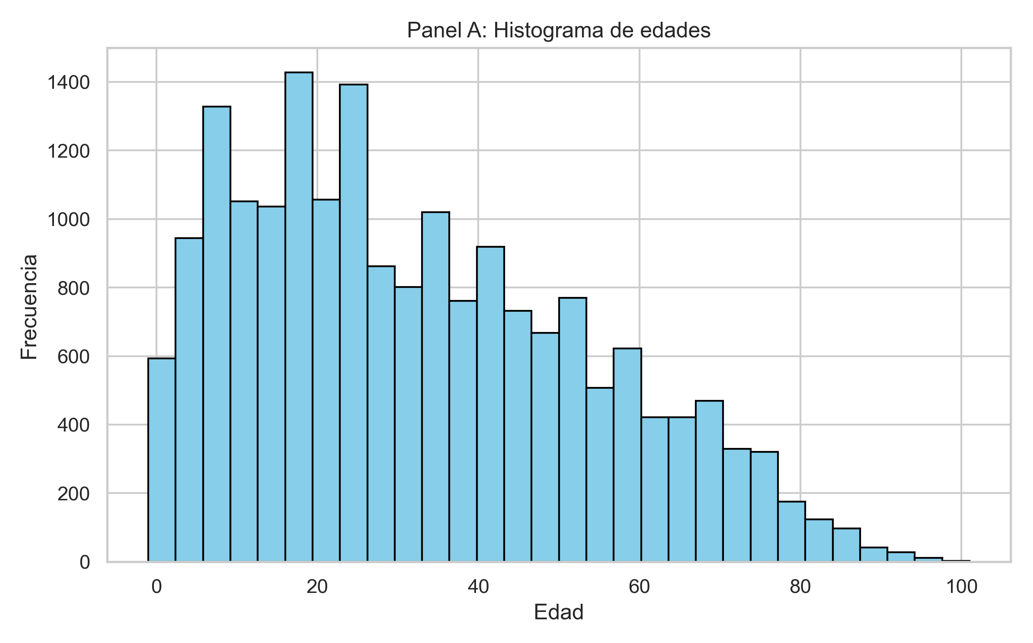
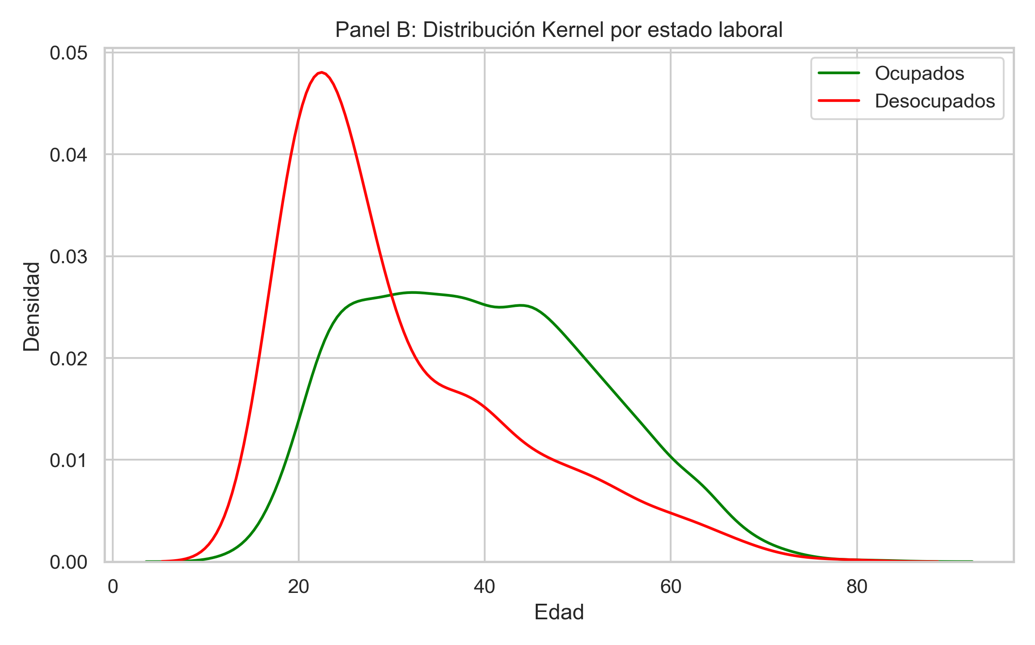
Gráficos y Descripciones

**PARTE 1**

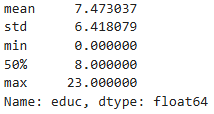
**Punto 1: Distribución de las edades**





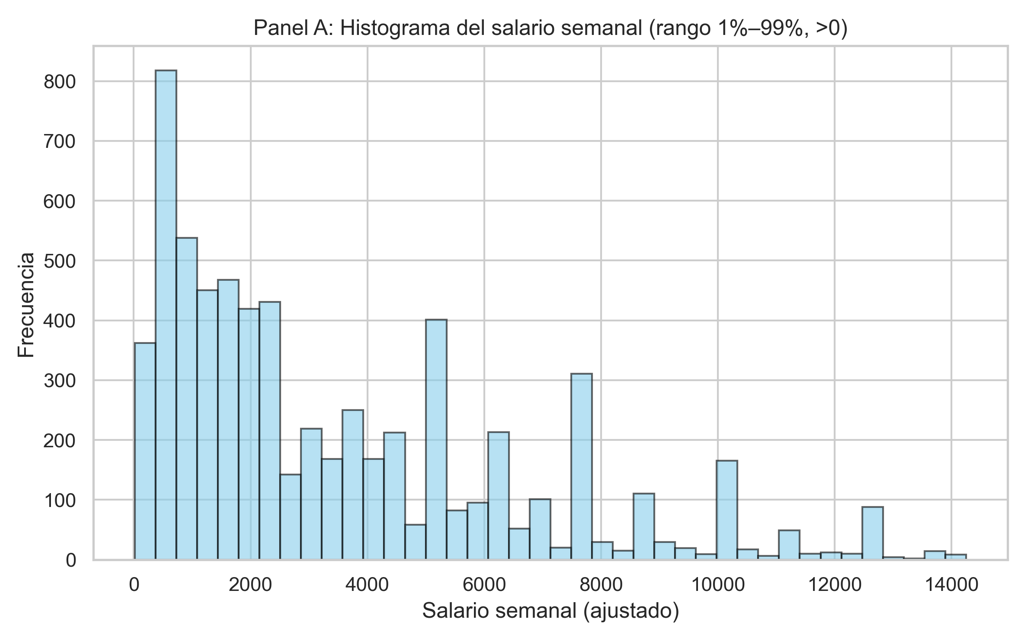
El histograma (**Panel A**) muestra una alta concentración de personas entre los 15 y 25 años, lo que refleja una base poblacional joven. En la distribución de Kernel, (**Panel B**) los ocupados se agrupan en edades intermedias, mientras que los desocupados presentan un pico más marcado entre los 20 y 40 años. Esto indica una mayor incidencia del desempleo en jóvenes adultos. Las curvas muestran perfiles laborales diferenciados según edad.

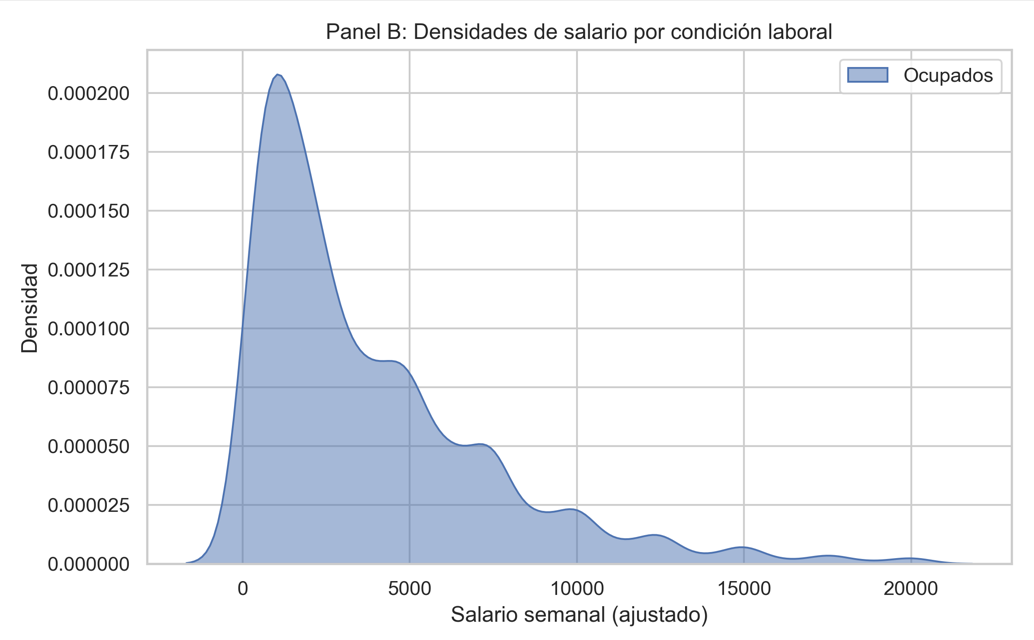
**Punto 2: Estadística descriptiva variable *“educ”***



La variable educ fue construida combinando CH12, CH13 y CH14, para estimar los años de educación formal completados. El promedio es de 7,47 años y la mediana es 8, lo que indica una leve concentración en niveles educativos bajos o incompletos. El valor máximo es 23 años, reflejando casos de estudios universitarios o posgrados. La distribución es amplia, con alta dispersión entre niveles educativos.

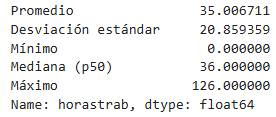
**Punto 3: Distribución del salario semanal**

****

****

Para facilitar el análisis de la parte “más rica en datos” del gráfico, se aplicó un recorte entre los percentiles 1 y 99 del salario semanal ajustado, excluyendo además los valores nulos o iguales a cero. Un total de 12.105 personas presentaron un ingreso semanal igual a 0, lo que justificó su exclusión para mejorar la legibilidad. En el **Panel A** se observa una fuerte concentración de ingresos bajos, con presencia decreciente de salarios más altos. El **Panel B** muestra una distribución asimétrica de los ocupados, con mayoría en rangos bajos y una larga cola hacia valores elevados. La curva de desocupados no se incluye por falta de varianza, ya que sus ingresos son nulos o residuales en casi todos los casos.

**Punto 4: Variable de horas trabajadas *“horastrab”***

****

La variable “***horastrab”*** fue construida como la suma de las horas trabajadas en la ocupación principal y en otras ocupaciones, considerando solo a personas ocupadas. El promedio semanal es de 35 horas y la mediana es de 36, lo que indica una distribución bastante centrada. Sin embargo, se observaron valores extremos de hasta 126 horas, posiblemente vinculados a pluriempleo o errores de carga. Esta heterogeneidad se refleja también en una desviación estándar elevada, cercana a 21 horas.

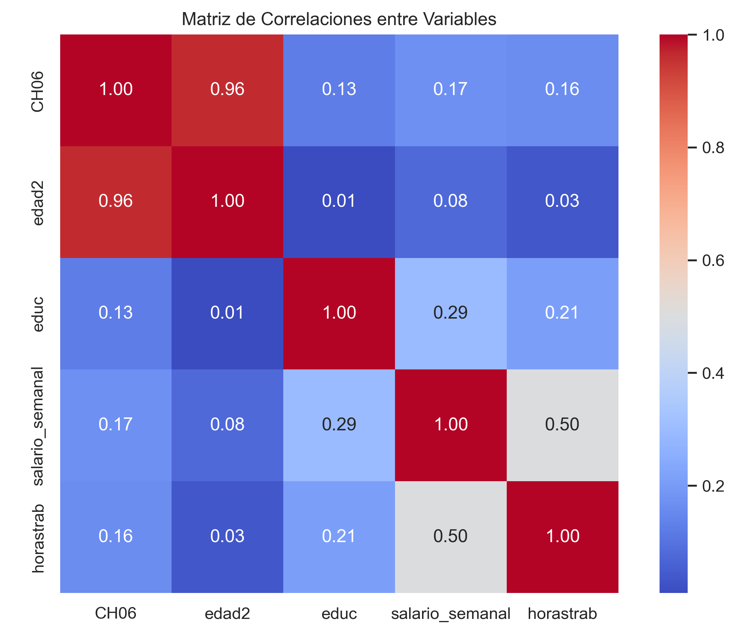
**Punto 5: Resumen de la base final para la región YYY**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2004** | **2024** | **Total** |
| **Cantidad de observaciones** | 9393 | 9699 | 19092 |
| **Cantidad de observaciones con Nas en la variable “Estado”** | 8 | 9 | 17 |
| **Cantidad de Ocupados** | 3337 | 4263 | 7600 |
| **Cantidad de Desocupados** | 597 | 256 | 853 |
| **Cantidad de variables limpias y homogeneizadas** | 185 | 185 | 185 |

La tabla resume las principales características de la base final. Se observa un total de 19.092 observaciones, distribuidas de forma equilibrada entre ambos años. La cantidad de personas ocupadas es considerablemente mayor que la de desocupadas, y la base contiene 185 variables limpias y unificadas, lo que garantiza consistencia para los análisis realizados.

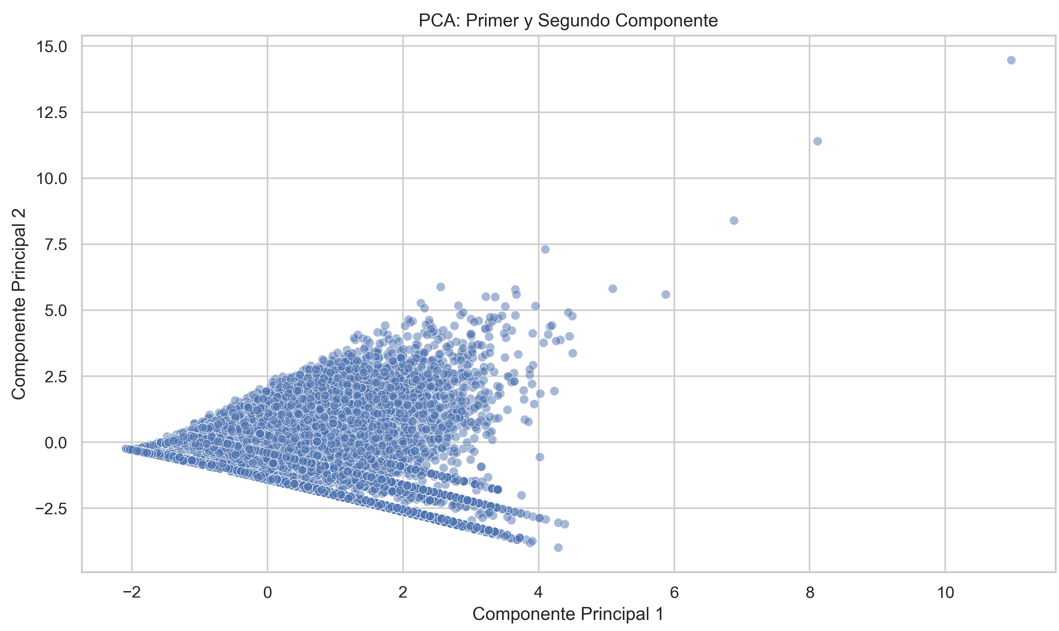
**PARTE 2**

**Punto 1: Matriz de correlaciones**



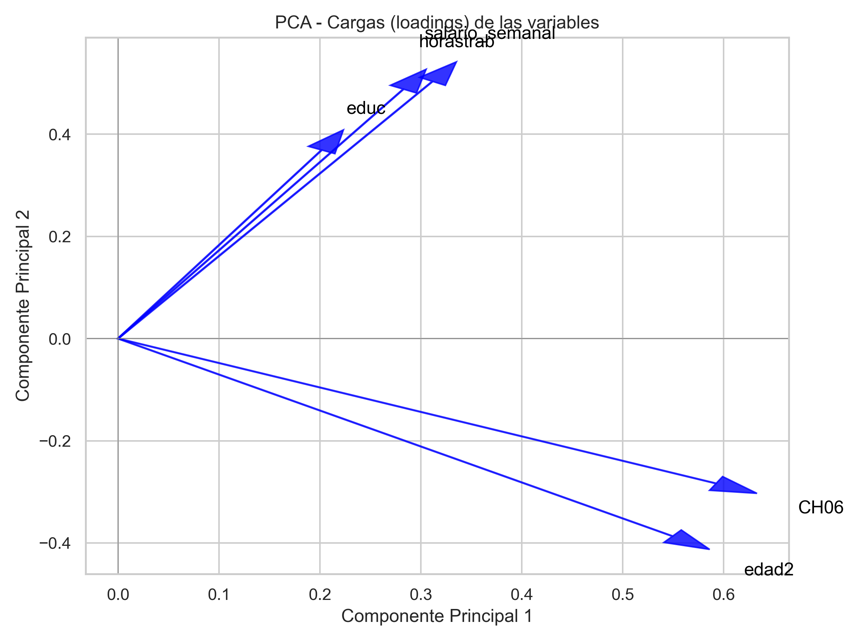
Se destaca la alta correlación entre ***“CH06”*** (edad) y ***“edad2”***, lo cual es esperable ya que una es el cuadrado de la otra. También se observa una correlación moderada entre ***“salario semanal”*** y ***“horastrab”*** (0.50), lo que indica que quienes trabajan más horas tienden a percibir mayores ingresos. En cambio, la variable ***“educ”*** presenta correlaciones débiles con las demás, lo que sugiere que su comportamiento es independiente, podría aportar información diferenciada en los análisis posteriores.

1. ***PCA***

**Punto 2: Indices (scores)**

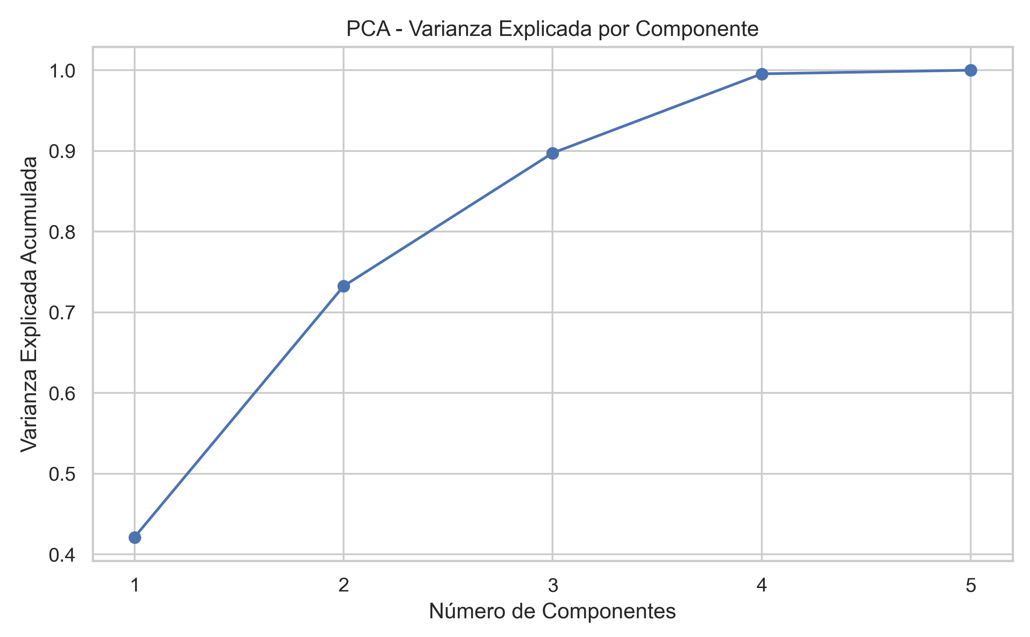
El gráfico muestra la distribución de los datos en el espacio de los dos primeros componentes principales. Se observa una alta concentración en valores bajos de ambos ejes, con algunos casos dispersos. Esto sugiere cierta estructura en los datos que podría ser aprovechada por técnicas de agrupamiento.

**Punto 3:**

****

El gráfico muestra la contribución de cada variable a los dos primeros componentes principales. ***“horastrab”*** y ***“salario\_semanal”*** tienen mayor peso en el componente 1, mientras que ***“edad”*** y ***“edad2”*** se alinean más con el componente 2. La variable ***“educ”*** aporta en menor medida a ambos componentes, reforzando su comportamiento independiente observado en la matriz de correlación.

**Punto 4:**

****

El gráfico muestra que los dos primeros componentes explican aproximadamente el 73% de la varianza total, mientras que con tres componentes se alcanza el 90%. A partir del cuarto componente, el aporte marginal disminuye notablemente. Esto sugiere que los primeros dos o tres componentes concentran la mayor parte de la información relevante.