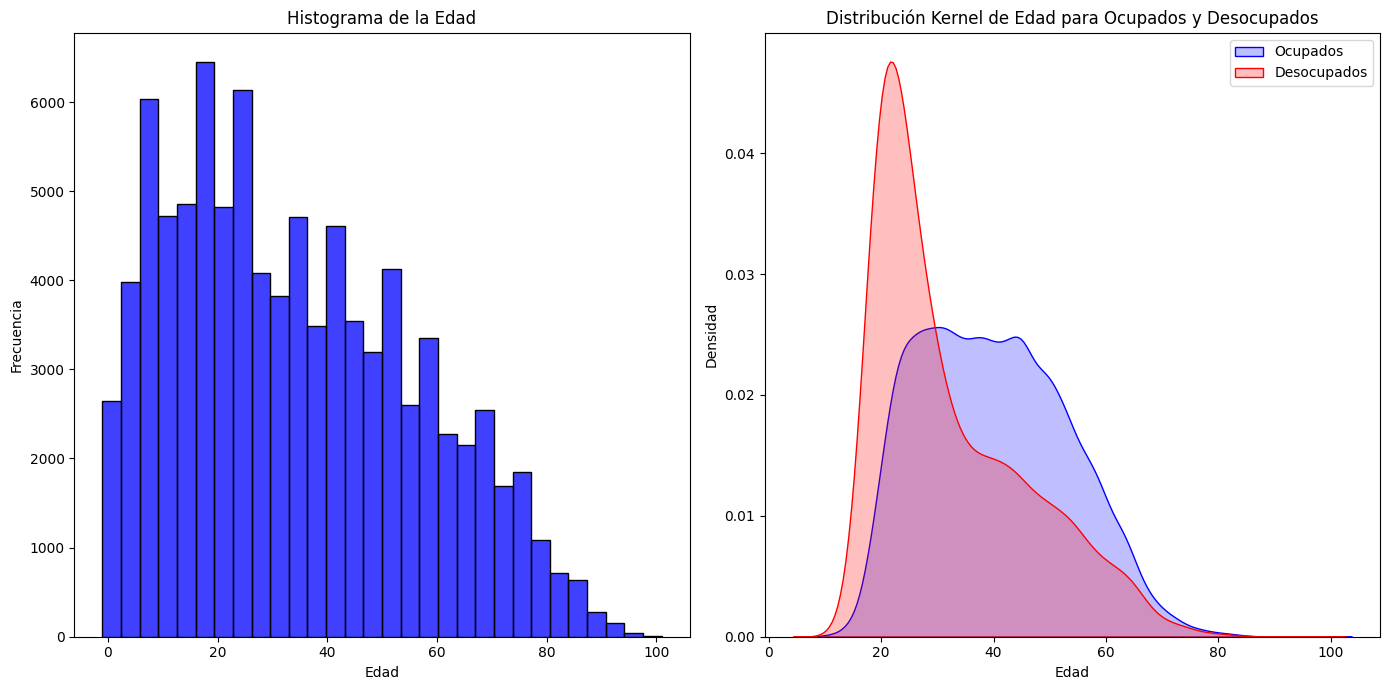
**REPORTE BigData\_TP3\_GRUPO#20**

**Parte I – Ejercicio 1**

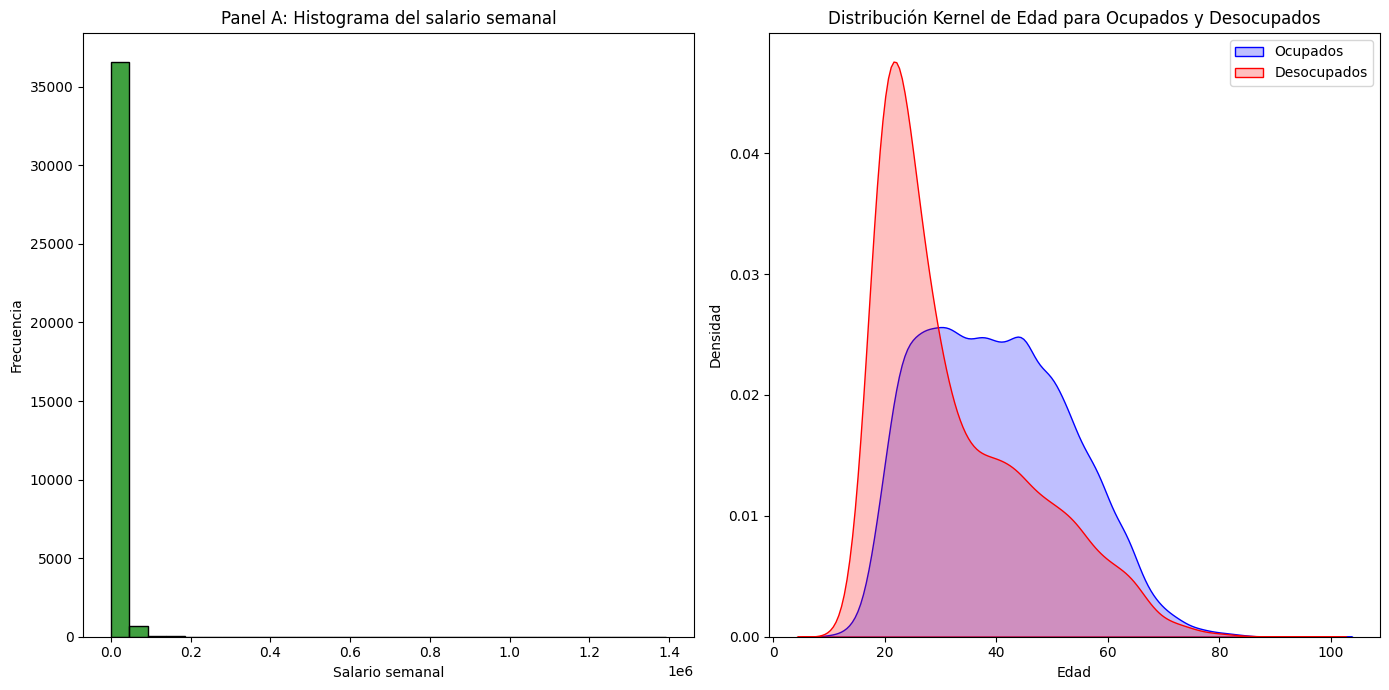


Podemos observar en el histograma y la distribución de kernel, que la mayoría de desocupados se concentran en las edades de alrededor de los veinte años, y la mayoría de ocupados se distribuyen de manera pareja entre las edades de treinta a cincuenta años.

**Parte I – Ejercicio 2**

Nota: Profesora, tuvimos inconvenientes con la estadística descriptiva, en un principio habíamos podido obtener los valores, pero cuando activamos al código nuevamente nos dio error.

**Parte I – Ejercicio 3**



El histograma y la distribución de kernel, nos indica que los salarios más bajos se concentran entre las edades más jóvenes, y los desocupados.

**Parte I – Ejercicio 4**

La estadística descriptiva muestra una media de 0, lo que podría indicar que más del 50% de las personas entrevistadas trabajan 0 horas, es decir son desocupados.

**Parte I – Ejercicio 5**

**¿Cuál es el tamaño de la de la base de datos para su región con las variables originales unificadas?**

Tamaño de la base unificada para GBA (sin NaNs): 2467 \*quiero aclarar que sólo se tomaron en cuenta 5 variables (edad, edad2, educ, salario\_semanal y horastrab) debido a que en la repartición del trabajo me guie para hacer este punto en base a los que siguen y no los anteriores.

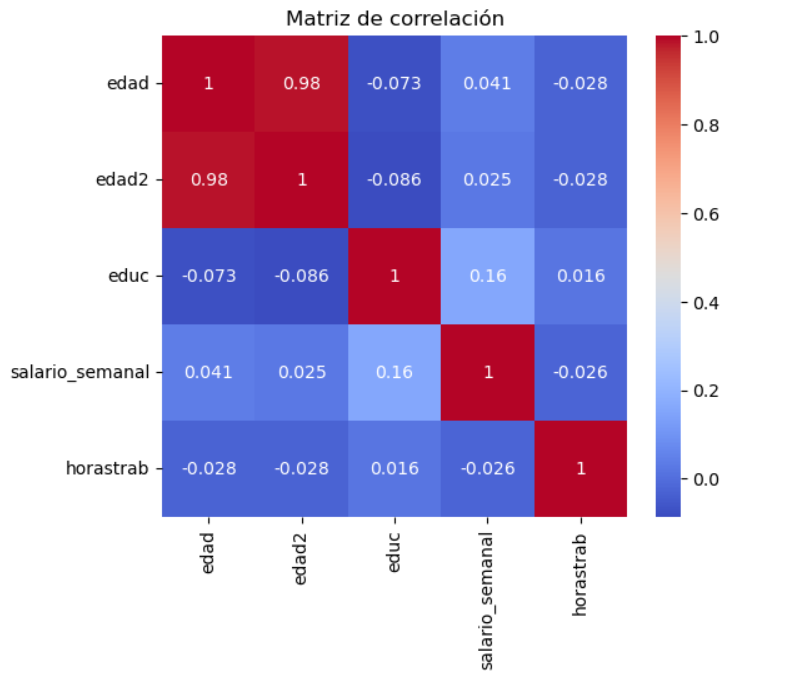
**Tabla 1. Resumen de la base final para la región YYY**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2004** | **2024** | **Total** |
| Cantidad observaciones | 0 | 5629 | 5629 |
| Cantidad de observaciones con Nas en la variable “Estado” | 0 | 0 | 0 |
| Cantidad de Ocupados | 0 | 0 | 0 |
| Cantidad de Desocupados | 0 | 0 | 0 |
| Cantidad de variables limpias y homogeneizadas | 6 | 6 | 6 |

Nota: Se calcula la “cantidad de Ocupados” como aquellos con la variable “Estado==Ocupado” y Cantidad de Desocupados como aquellos con la variable “Estado==Desocupado”.

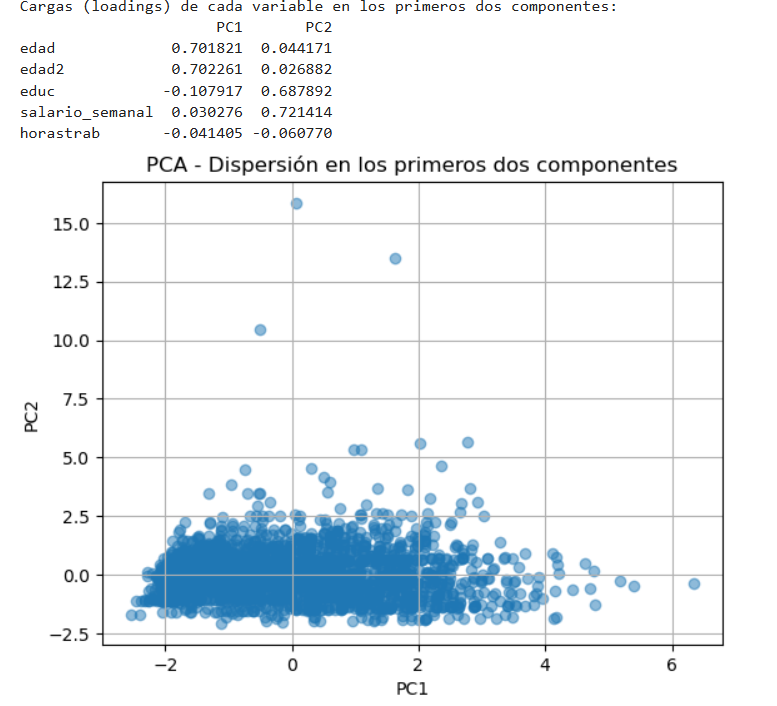
**Parte II – Ejercicio 1**

Podemos notar que hay una relación diagonal entre las cinco variables que elegimos. Existe una alta correlación entre edad y edad 2, también tenemos una correlación positiva entre educación y salario semanal, lo que nos dice que a mayor educación suele corresponder un mejor salario.

****

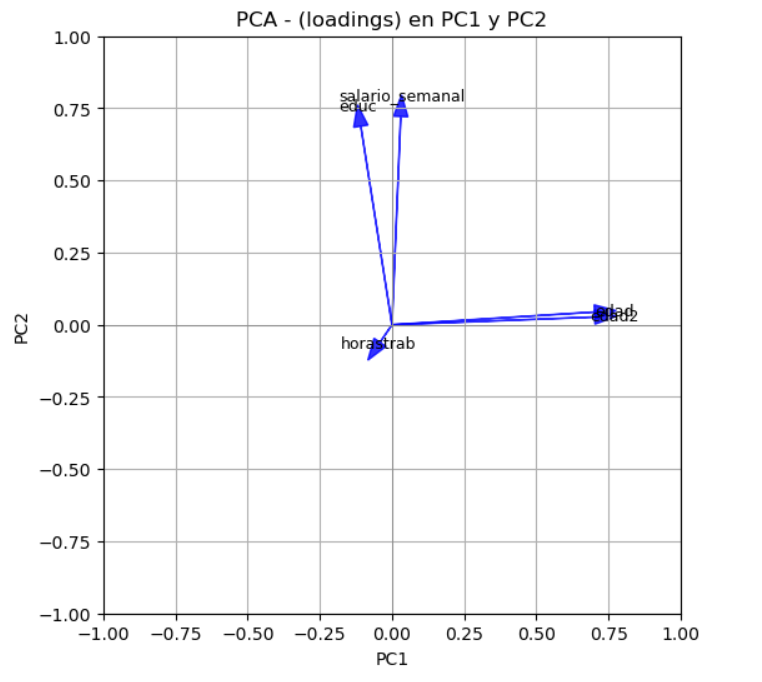
**Parte II – Ejercicio 2**

Las cinco variables fueron homogeneizadas y estandarizadas pata poderlas comparar adecuadamente. Según los loadings observados el PC1 está muy influenciado por “edad” y “edad2” donde se compara ¿n las diferencias en el ciclo de vida. Mientras que en PC 2, existe una influencia alta de educación y salario semanal, lo que nos lleva a acercarnos a la conclusión de la matriz de correlación.



**Parte II – Ejercicio 3**

El PCA logró separar claramente dos dimensiones: una relacionada con la edad y otra con las condiciones laborales o el capital humano. Esto ayuda a entender mejor cómo se organizan los datos cuando reducimos las opciones a sólo dos ejes principales.

****