

TRABAJO PRÁCTICO N°3 - GRUPO N°2

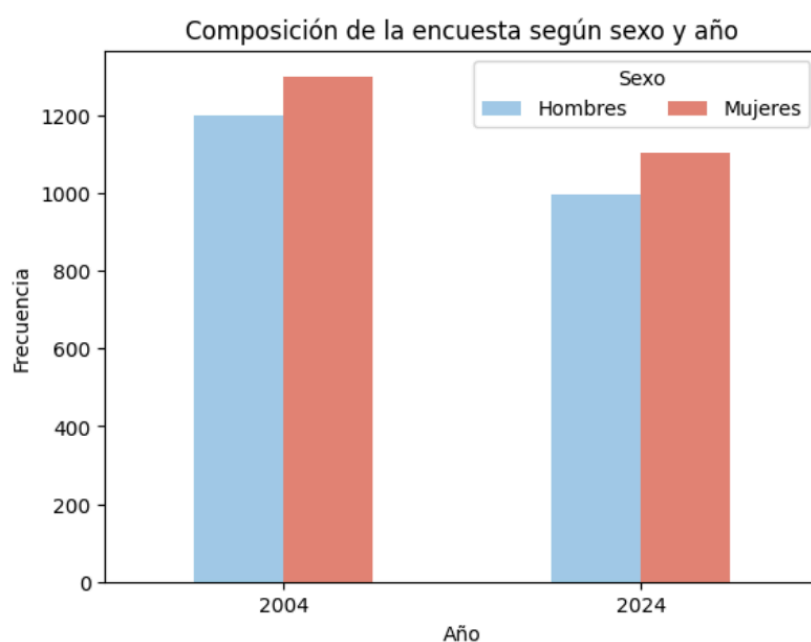
ANÁLISIS DESCRIPTIVO Y PREDICCIÓN DE DESOCUPACIÓN

Big Data y Aprendizaje Automático para Economistas
Guadalupe Gorostiaga y Valentina Fuentes Mortensen

Parte I

1. El Instituto Nacional de Estadísticas y Censos (INDEC) identifica a las personas desocupadas como aquellas que, no teniendo ocupación, están buscando activamente trabajo y están disponibles para trabajar. No distingue entre formas de precariedad laboral tales como personas que realizan trabajos transitorios mientras buscan activamente otro trabajo, aquellas que trabajan jornadas involuntariamente por debajo de lo normal, otras que han suspendido la búsqueda por falta de oportunidades visibles, ocupadas en puestos por debajo de la remuneración mínima o en puestos por debajo de su calificación, etc. Estas distinciones si son relevadas aunque como indicadores separados.
 2. Este reporte tiene como objetivo realizar un análisis comparativo y predictivo de la desocupación en el aglomerado de Gran Tucumán - Tafí Viejo para el primer trimestre de los años 2004 y 2024 usando como insumo la Encuesta Permanente de Hogares (EPH) relevada por el INDEC.
- c) En el Gráfico 1 se observa una disminución en la cantidad total de encuestados para ambos sexos en 2024 en comparación con 2004, y las mujeres continúan siendo la mayoría en ambos años.

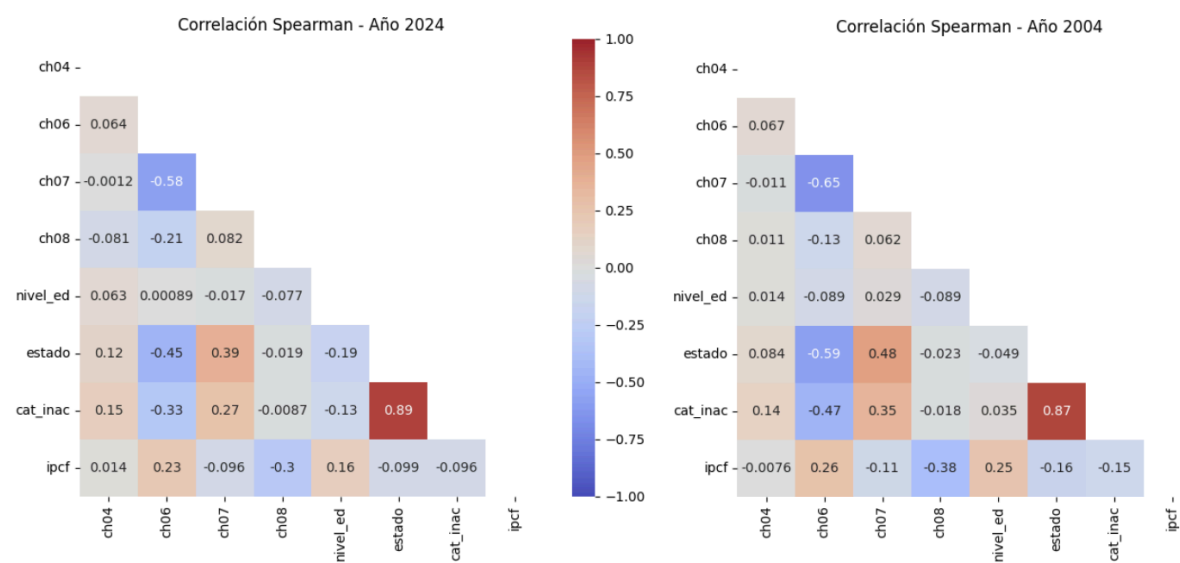
Gráfico 1



d) Al observar las correlaciones entre las variables seleccionadas que se presenta en el Gráfico 2, la correlación fuerte es entre el estado y la categoría de inactividad, aunque esta es resultado de cómo se construyen las variables. También se identifica una correlación positiva entre la edad y el estado civil, así como entre el estado y el estado civil; sin embargo, ambas relaciones carecen de sentido debido a que la variable de estado civil no sigue un orden lógico.

Por otro lado, hay una correlación positiva significativa entre el ingreso per capita familiar y el nivel educativo alcanzado, lo cual es consistente con lo esperado. En contraste, se observa una correlación negativa entre el ingreso per capita familiar con el estado y con la categoría de inactividad, lo cual tiene lógica considerando que, a mayor ingreso, es menos probable estar desocupado o inactivo. Además, hay una relación negativa entre el estado y la edad, lo que implica que a mayor edad, se incrementa la probabilidad de estar en situaciones como ocupado o desocupado. Es interesante notar que la relación entre las variables no ha cambiado considerablemente entre 2004 y 2024, aunque podemos destacar la correlación más fuerte entre el ingreso per capita familiar y el nivel educativo para 2024.

Gráfico 2



e) La proporción de desocupados en la muestra de 2024 es casi dos veces menos que la de 2004 , mientras que la proporción de inactivos aumentó en alrededor de 2 puntos porcentuales (ver Tabla 1).

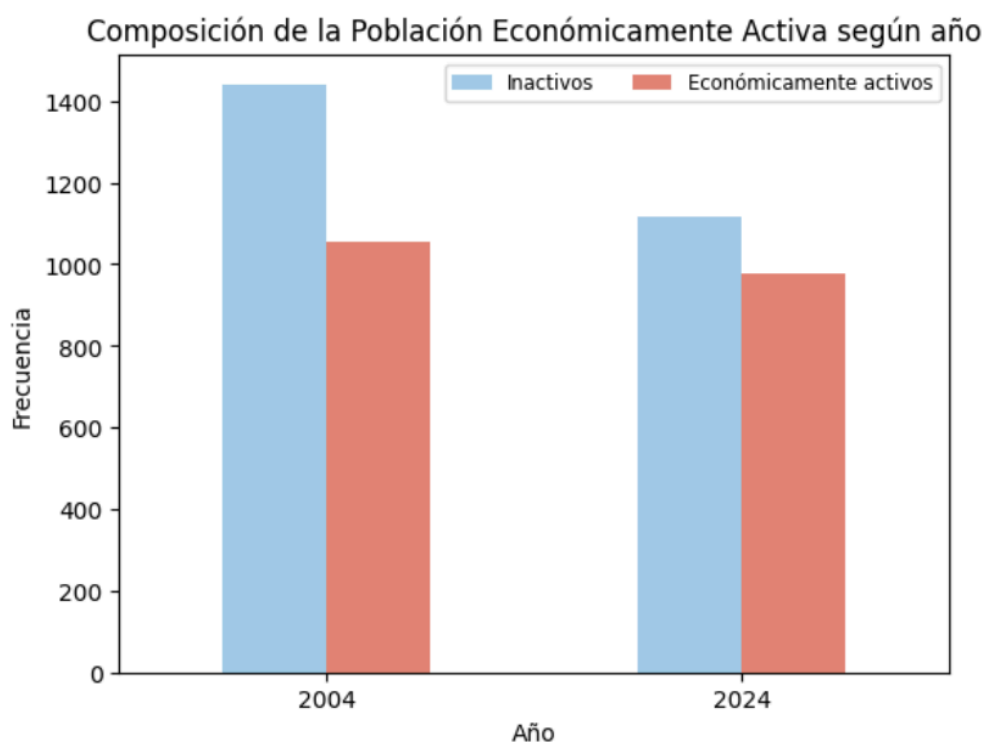
Tabla 1

	2004	2024
Proporción de desocupados	7.12%	3.86%
Proporción de inactivos	38.90%	41.42%

IPCF promedio de ocupados	\$ 254.76	\$ 152,318.01
IPCF promedio de desocupados	\$ 142.57	\$ 101,478.01
IPCF promedio de inactivos	\$ 204.55	\$ 127,528.96

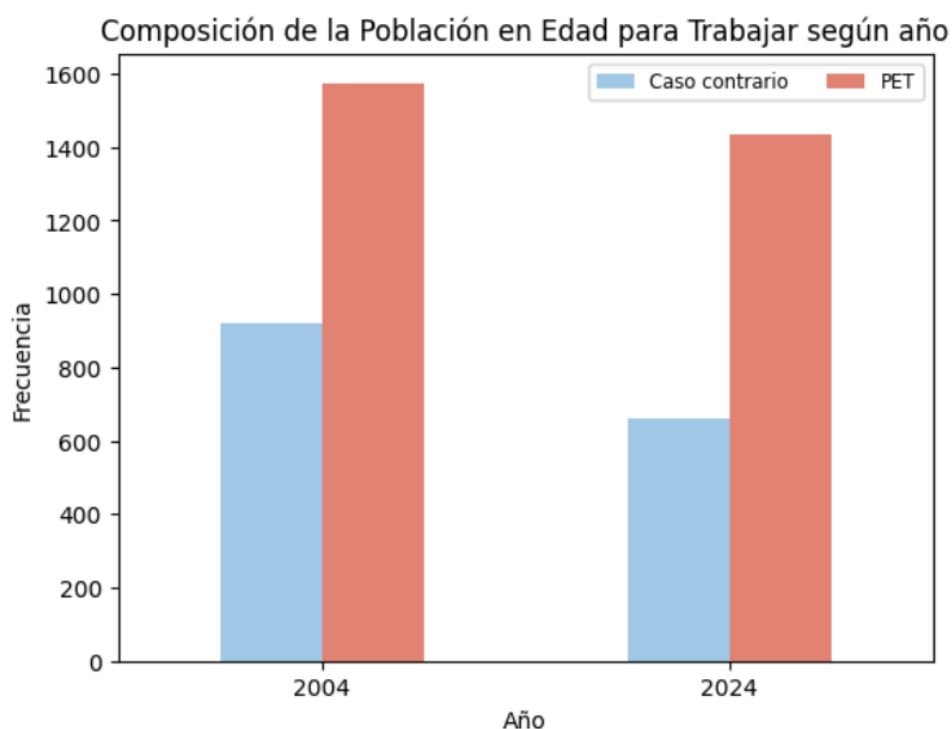
3. La cantidad de personas en el total de la muestra que no respondieron cuál es su condición de actividad fueron 10.
4. El siguiente gráfico muestra que, en ambos años, los inactivos son más numerosos que los económicamente activos. Sin embargo, la diferencia entre estos grupos se reduce en 2024 en comparación con 2004, lo que indica un aumento relativo de la población activa.

Gráfico 3



5. La Población en Edad para Trabajar (PET) representa la mayoría de la población en ambos años. La población que no pertenece a este grupo (como niños y ancianos) parece haber disminuido más que la PET entre 2004 y 2024. Esta disminución podría explicarse por cambios en las decisiones de los hogares con respecto a tener hijos, lo que ha lleva a una reducción en la natalidad y, por ende, a la disminución en el grupo de niños.

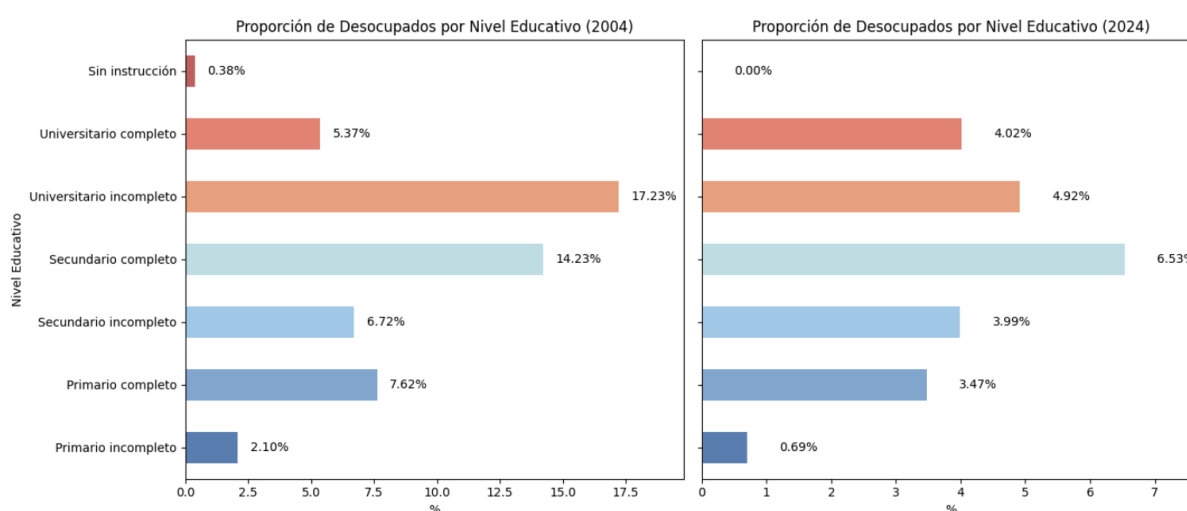
Gráfico 4



6. En 2004, 178 personas se encontraban desocupados. En 2024, la cantidad de desocupados en la muestra es de 81. Nuevamente, esto no significa que la cantidad haya sido disminuido, sino por tamaño de la muestra.

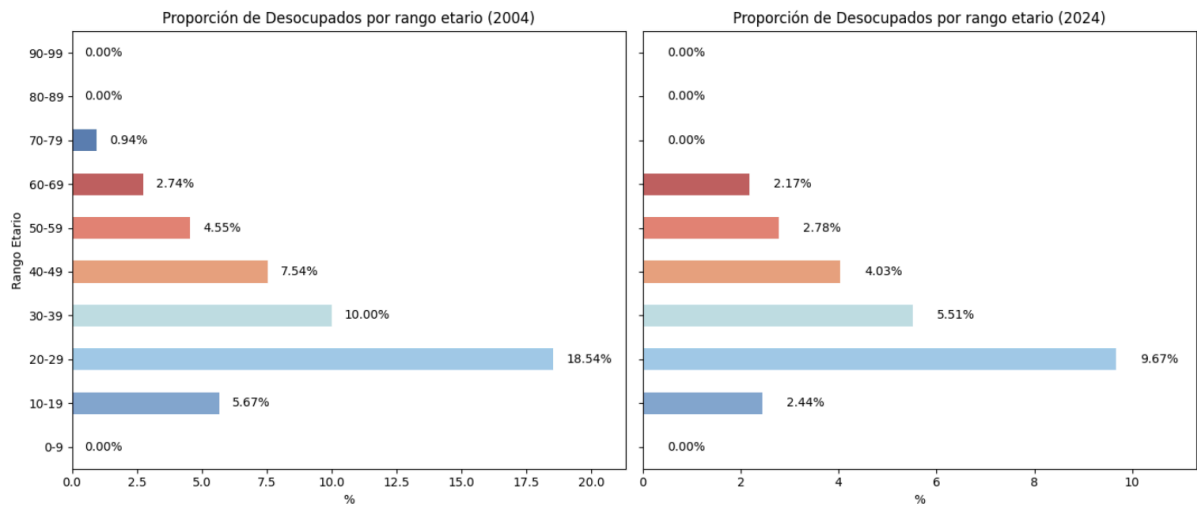
a) Respecto a la proporción de desocupados según el nivel educativo, en 2004, la mayoría de las personas desempleadas contaban con educación universitaria incompleta, mientras que un 0,38% no tenía ningún tipo de instrucción. Este último porcentaje se redujo a 0% en 2024, y la mayoría de los desocupados en 2024 cuenta con el nivel de secundaria completa.

Gráfico 5



b) A grandes rasgos, se observa que la proporción de desocupados en los grupos de 30 a 39, 20 a 29 y 10 a 19 años se redujo casi a la mitad entre 2004 y 2024.

Gráfico 6



7. La tasa de desocupación que reporta el INDEC se calcula como la proporción de desocupados en relación con la PEA. Sin embargo, una crítica frecuente en la economía laboral es que la participación en el mercado laboral es una decisión endógena de los individuos, lo que puede afectar el indicador. Para evitar que la tasa de desocupación varíe debido a esta decisión, sería más apropiado ampliar la base de comparación y considerar a toda la PET, en lugar de solo la PEA. La ventaja de usar la medida de Economía Laboral es que tendremos una medida menos volátil en el tiempo porque la base depende de la estructura demográfica y no de las decisiones.

Tabla 2

Tasa de desocupación	2004	2024
INDEC	8.76%	3.99%
Economía Laboral	5.92%	2.69%

Parte II

El objetivo de esta sección es intentar predecir si una persona está desocupada o no.

3. El método que mejor predice la desocupación en ambos años es la Regresión Logística, la decisión se basó en el mejor desempeño balaceando los siguientes criterios: Acurracy, AUC y ROC curve.

Tabla 3

Accuracy

	2004	2024
--	------	------

Regresión Logística	0.945	0.955
Análisis discriminante lineal	0.94	0.96
KNN con k=3	0.923	0.949
Naive Bayes	0.658	0.954

Tabla 3

AUC

	2004	2024
Regresión Logística	0.898	0.759
Análisis discriminante lineal	0.500	0.518
KNN con k=3	0.595	0.531
Naive Bayes	0.808	0.498

4. La proporción predicha de desocupados dentro de los no respondientes es 0 (de 10 sujetos que no contestaron).