



Universidad de Buenos Aires

Facultad de Ciencias Económicas

Maestría en Economía Aplicada

Big data y Aprendizaje Automático Aplicado

Trabajo Práctico 3

Alumnos:

Oscar Cuentas Sandy

Jhonatan Saúl Elguera Quispe

Rodrigo Hermoso

Profesora:

María Noelia Romero

3° Trimestre 2024

Índice general

Introducción.....	3
Parte I. Análisis descriptivo	3
<i>Población desocupada: concepto y medición</i>	3
<i>Análisis descriptivo del mercado laboral argentino 2004 y 2024</i>	4
<i>Composición de la población económicamente activa y en edad de trabajar</i>	6
<i>Población desocupada: segmentos y metodología alternativa</i>	7
Parte II. Clasificación	9

Introducción

El cálculo de las personas empleadas y desempleadas represente un indicador de gran importancia para todos los estados, en base a este se realizan las políticas públicas orientadas a mejorar el bienestar de la sociedad.

En la actualidad, los países latinoamericanos y especialmente de la región enfrentan diversos desafíos en el mercado laboral, con una cada vez mayor proporción de personas que trabajan en el sector informal y no cuentan con las condiciones mínimas para tener una vida estable como lo es una remuneración constante, seguro de salud, seguro de desempleo, entre otros que si se les brinda a las personas activas en el sector formal. En el caso de Argentina, esta situación ha sido objeto de diferentes discursos políticos en procura de mejorar esta realidad en los ciudadanos, pero sin lograr grandes cambios a lo largo de al menos, las dos últimas décadas.

En el presente trabajo se hace hincapié en los datos que nos proporciona la Encuesta Permanente de Hogares (EPH) para dos puntos en el tiempo, el primero ubicado en el primer trimestre de 2004, una época donde se terminaba de superar la crisis institucionales del 2001 y comenzaba a emerger una situación altamente favorable para el país en relación a la economía, por otro lado, se observa el mismo trimestre veinte años después, en 2024, un periodo donde las políticas de los primeros meses del actual gobierno fueron muy significativas para todos los mercados de la economía, especialmente para el laboral.

En la primera parte del trabajo explicaremos como realiza el INDEC la definición para personas ocupadas y desocupadas, observaremos diferentes variables en los dos periodos de tiempo y analizaremos las diferencias o similitudes encontradas, además exploraremos la problemática de contar con información incompleta por parte de los hogares que no reportan sus ingresos en las encuestas.

En la segunda parte del trabajo haremos un trabajo de clasificación donde buscaremos predecir si una persona está o no está desocupada utilizando distintas variables de características individuales.

Parte I. Análisis descriptivo

Población desocupada: concepto y medición

Según la información presentada en los documentos metodológicos de la EPH, se define como desocupados a todas aquellas personas que sin tener trabajo se encuentren disponibles para trabajar y han buscado activamente una ocupación en un período de referencia determinado.

Para la medición de este concepto en la EPH se considera como desocupados abiertos a todas las personas de 10 años y más que no trabajaron en la semana de referencia, estaban

disponibles y buscaron activamente trabajo (enviar currículos, responder a avisos de diarios/internet, consultar a parientes o amigos, etc.) en algún momento de los últimos treinta días. También, se incluyen dentro de este grupo a las personas disponibles que no buscaron trabajo, porque ya tenían trabajo asegurado y a los suspendidos sin pago que no buscaron porque esperaban ser reincorporados.

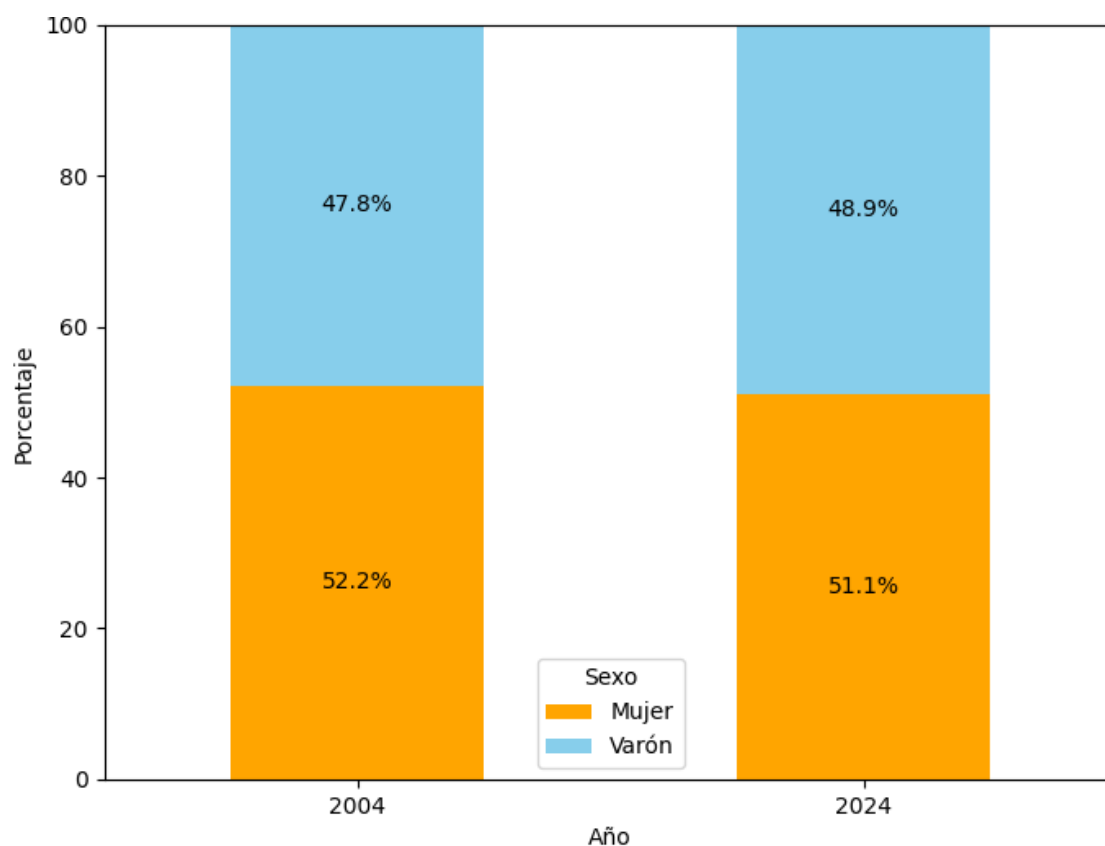
El concepto de desocupación no incluye otras formas de empleo inadecuado tales como personas que realizan trabajos transitorios mientras buscan activamente una ocupación, aquellas que trabajan jornadas involuntariamente por debajo de lo normal, los ocupados en puestos por debajo de la remuneración mínima o en puestos por debajo de su calificación, ni a los desocupados que han suspendido la búsqueda por falta de oportunidades visibles de empleo, etc. Estas modalidades son también relevadas en la EPH como indicadores separados relacionados a la calidad del empleo.

Análisis descriptivo del mercado laboral argentino 2004 y 2024

Utilizando las bases proporcionadas por el INDEC, en primer lugar, hacemos un análisis descriptivo de las diferencias entre algunas variables para los dos periodos de tiempo. Por ejemplo, se puede observar en el Gráfico 1 que, al diferenciar por género, la población no experimenta una variación significativa en el tiempo.

Gráfico 1

Distribución de las muestras según género, 2004 y 2024

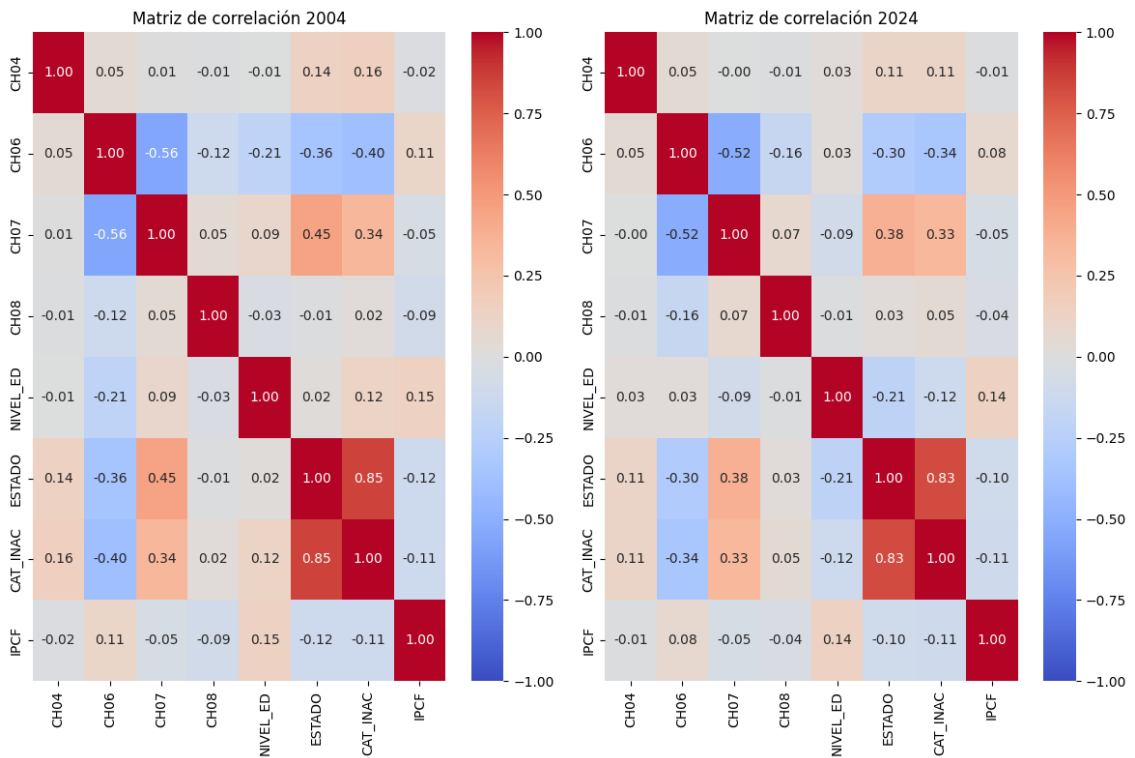


Ahora, como segundo ejercicio, se realiza un cruce entre las diferentes variables para construir una matriz de correlación. Al analizar los resultados, se observa que tanto en

2004 como en 2024 se presenta una estructura similar, por lo que la descripción realizada será válida para ambos años. En primer lugar, se identifica una correlación directa entre el estado de actividad y la categoría de inactividad. Asimismo, estas dos variables muestran una correlación directa con el estado civil, mientras que, en contraste, exhiben una correlación indirecta con la edad. Estos hallazgos sugieren que las personas mayores y aquellas que están en una unión tienden a ubicarse en la categoría de ocupados.

Gráfico 2

Matriz de correlación entre medidas de interés, 2004 y 2024



Por otro lado, al estimar algunos indicadores de interés, encontramos resultados interesantes. Como era de esperarse, según la Tabla 1, los ingresos en promedio de las personas ocupadas son significativamente superior de aquellos que no lo están, el mismo triplica el valor de las personas que no encuentran un trabajo activo en el mercado laboral, sin embargo, también se observa que el ingreso promedio de los inactivos duplica al ingreso de los desocupados, lo cual hace referencia a que probablemente aquellas personas sean en su mayoría obtengan ingresos por motivos no laborales (rentas, alquileres, remesas, etc.) o sean dependientes de algún familiar o pareja.

Asimismo, uno de los problemas de la robustez de los indicadores es el creciente número de hogares que no reportan ingresos, esto puede generar subestimaciones en las estadísticas laborales. Entre las muestras de ambos años se registra un total de 51 casos que no respondieron su condición de actividad y, por lo tanto, no reportan información de ingresos.

Tabla 1

Datos descriptivos del mercado laboral

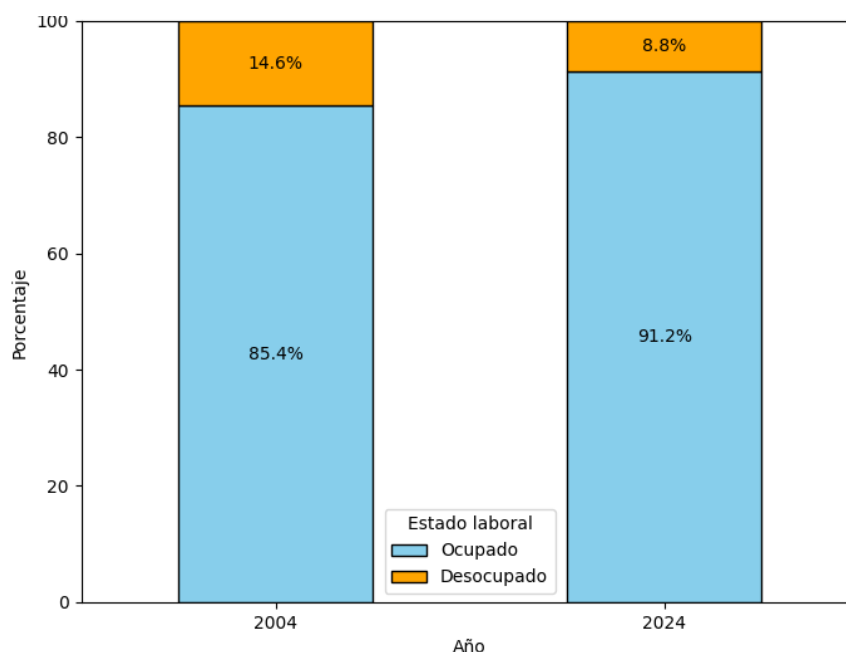
Indicador	
Cantidad de Desocupados	839
Cantidad de Inactivos	5,462
Media de IPCF Ocupados	106,443.4
Media de IPCF Desocupados	31,656.0
Media de IPCF Inactivos	63,863.1

Composición de la población económicamente activa y en edad de trabajar

Como se puede observar en el Gráfico 3, si observamos únicamente la diferencia entre la proporción de ocupados y desocupados de la PEA en los dos periodos de tiempo, se aprecia que, pese a la complicada situación socioeconómica del país en el primer trimestre de 2024, una mayor proporción de la ciudadanía se encuentra ocupada según los resultados en comparación con el primer trimestre de 2004. En específico, entre los periodos de análisis, el porcentaje de ocupados aumentó en 6 puntos porcentuales.

Gráfico 3

Distribución de la población económicamente activa según estado de actividad, 1t-2004 y 1t-2024



En relación con la población en edad de trabajar, se registra que entre los años de estudio se ha incrementado su magnitud. Lo cual se representa como la mayor expansión poblacional existente en los últimos veinte años. Al estimar la proporción de personas ocupadas y desocupadas

Al desagregar esta población se observa un incremento de la población económicamente activa y un sostenimiento de la población en condición inactiva. En específico, la PEA se incrementó en un punto porcentual, con una principal contribución de los ocupados. Esto

va en línea con el observado en el gráfico anterior. Ambos resultados muestran respaldan la idea que en el largo del tiempo se ha incrementado la proporción de gente ocupada.

Población desocupada: segmentos y metodología alternativa

Por otro lado, con la finalidad de analizar más a profundidad el grupo de los desocupados se estiman la tasa de desempleo según nivel educativo y grupo de edad. Con respecto al nivel educativo, entre 2004 y 2024, se observa una mejoría para la población argentina, ya que los mejores grados de instrucción académica están asociadas a mayores probabilidades de encontrar trabajo en la actualidad. Muestra de ello se observa en el Gráfico 5 como los que tienen mayores niveles de educación cuentan con una probabilidad baja de estar desocupados, lo cual también representa como el mercado laboral se ha vuelto más exigente con las aptitudes que busca en los profesionales, claramente hoy en día, una mejor educación es sinónimo de contar con un trabajo seguro.

Gráfico 4

Distribución de la población en edad de trabajar según estado de actividad, 2004 y 2024

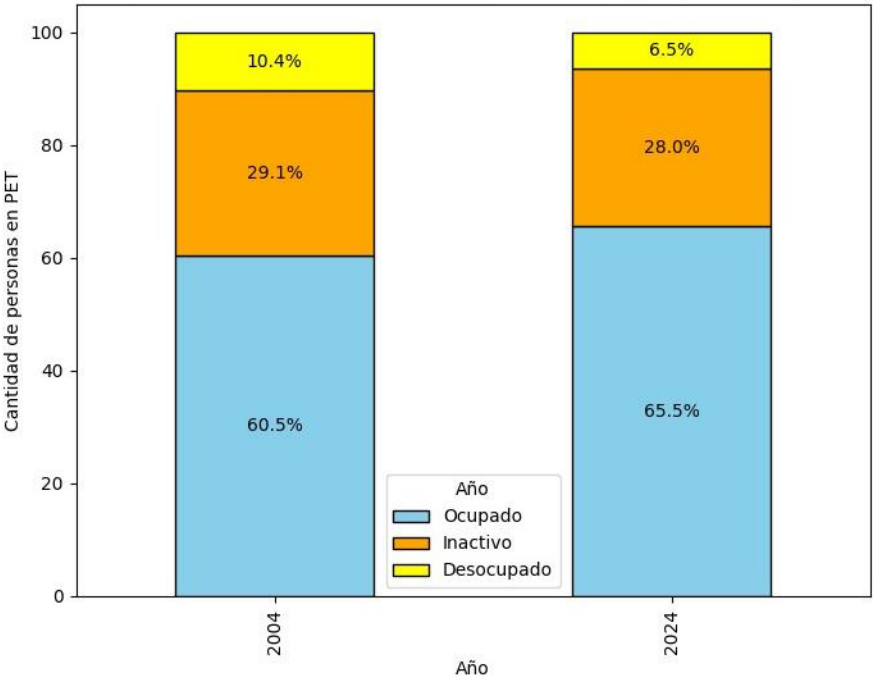
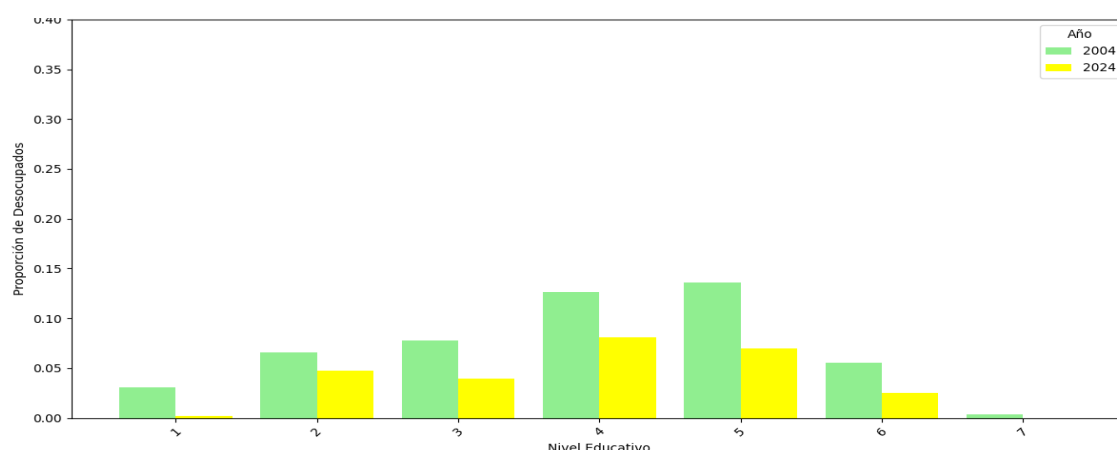


Gráfico 5

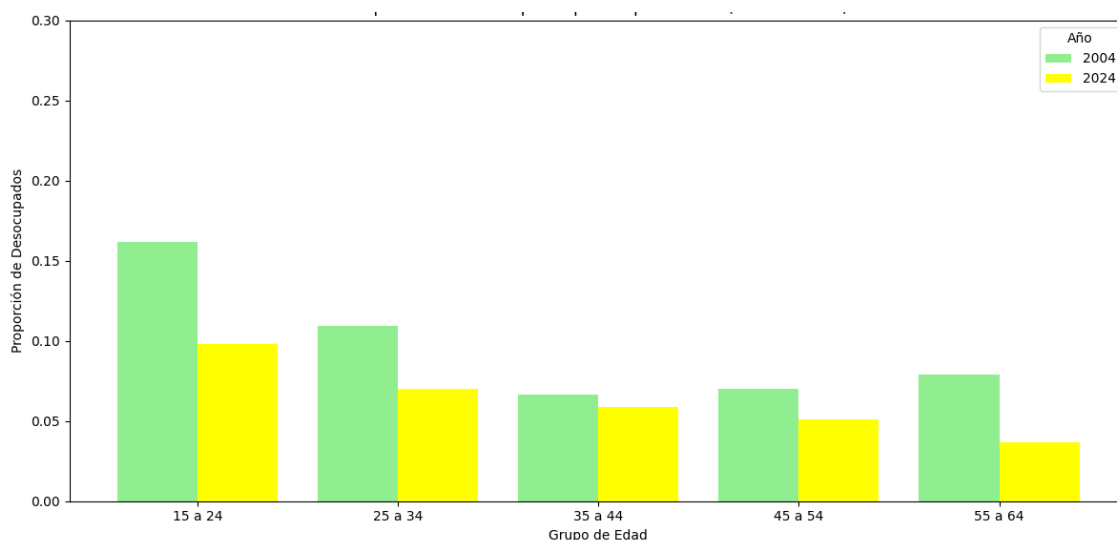
Distribución de población desocupada según nivel educativo, 1t-2004 y 1t-2024



De igual manera, si se compara la proporción de desempleados por grupo de edad entre los años 2004 y 2024, se observa una disminución generalizada en todos los grupos etarios. El grupo de 15-24 años sigue siendo el más afectado, aunque presenta una reducción significativa en 2024, lo que sugiere mejoras en las oportunidades laborales para los jóvenes. No obstante, siguen siendo un grupo vulnerable. Los demás grupos, desde el de 25 a 34 años hasta el de 55-64 años, también evidencian una reducción en las tasas de desempleo, aunque las diferencias son menos marcadas, indicando una mejora más uniforme.

Gráfico 6

Distribución de población desocupada según grupo de edad, 1t-2004 y 1t-2024



Por último, se muestra una estimación alternativa de la tasa de desempleo considerando en el denominador la población en edad de trabajar, en vez, de la población económicamente activa. Una de las ventajas de esta estimación alternativa es que puede capturar el efecto de los desalentados, es decir, de aquella población que esta fuera de la PEA, porque consideran que no obtendrán trabajo debido a las condiciones laborales o barreras estructurales. Sin embargo, el efecto de este grupo puede no observarse directamente en este método, ya que también se incluye a aquellos que no están laborando

por decisiones personales que no están vinculadas al desaliento. Según la Tabla 2, las estimaciones por el método alternativa resultan menores. Con respecto a su evolución temporal, la tasa de desempleo que considera el INDEC registra mayor caída que el obtenida por el método alternativo.

Tabla 2

Tasa de desempleo metodología INDE y alternativa, 1t-2004 y 1t-2024

	Tasa de desempleo INDEC	Tasa de desempleo alternativa
2004	14.6 %	10.8 %
2024	8.8 %	6. 7 %
Diferencia (puntos porcentuales)	5.8	4.1

Parte II. Clasificación

Con la finalidad de clasificar a los desocupados se realiza un conjunto de ejercicios de clasificación para las muestras del 2004 y 2024. Los métodos considerados son: regresión logística, análisis discriminante lineal, vecinos más cercanos y Naives Bayes. Los resultados para la muestra del 2004 y 2024 se obtienen las siguientes métricas de precisión presentadas en la Tabla 3 y Tabla 4. Se encuentra en ambas muestras que el método de “vecinos más cercanos” es un método de clasificación más adecuado para identificar a la población desocupada. Cabe precisar para la muestra del 2004, si bien la metodología de la regresión logística y Naives Bayes genera métricas de precisión más elevadas, están son muy elevadas y sugieren cierto sesgo que sobreestimar los resultados de estos métodos de clasificación.

Tabla 3

Métricas de precisión de métodos de clasificación para 1t-2004

	Regresión logística	Análisis discriminante lineal	Vecinos más cercanos	Naives Bayes
Accuracy	1.00	0.93	0.93	0.99
AUC	1.00	0.55	0.60	0.99

Tabla 4

Métricas de precisión de métodos de clasificación para 1t-2024

	Regresión logística	Análisis discriminante lineal	Vecinos más cercanos	Naives Bayes
Accuracy	0.95	0.95	0.95	0.95
AUC	0.51	0.50	0.57	0.50

Al emplear ambos métodos de clasificación para estimar la proporción de desocupados en la muestra de aquellos que no reportaron una información para su condición de actividad se encuentra que para la muestra del 2004 el 40% (4 de los 10 casos no reportados) se encontraría en condición de desocupado. Mientras que, para la muestra del 2024 el 37% (15 de los 41 casos no reportados) se encontraría en condición de desocupado.

Tabla 5

Estimación de condición de actividad empleando el método de vecinos más cercanos, 1t - 2004

	Número de casos	%
Ocupado	6	60%
Desocupado	4	40%
Total	10	100%

Tabla 4

Estimación de condición de actividad empleando el método de vecinos más cercanos, 1t - 2024

	Número de casos	%
Ocupado	26	63%
Desocupado	15	37%
Total	41	100%