

MEMORIA

Estudio del impacto de la tasa de variación del PIB en la tasa de paro

Exploratory Data Analysis (EDA)

Por Jose Antonio Moreno De Castro

Mayo 2022



Índice:

1. Objetivo.....	Pág. 3
2. Recursos utilizados.	Pág. 3-4
3. BBDD.....	Pág. 4-5
4. Análisis.....	Pág. 6-11
4.1 Exploración y descripción del database.....	Pág. 6
4.2 Limpieza y preparación de datos.....	Pág. 6
4.3 Análisis y visualización gráfica.....	Pág. 7-11
5. Conclusiones.....	Pág. 11-12
5.1 Grupo más afectado por el paro.....	Pág. 11
5.2 Grupo menos afectado por el paro.....	Pág. 11
5.3 Grupo más sensible a las variaciones del PIB.....	Pág. 12
5.4 Grupo menos sensible a las variaciones del PIB.....	Pág. 12
5.5 Comparación con las impresiones iniciales.....	Pág. 12

1. Objetivo

El objetivo de este Análisis exploratorio de datos o EDA (Exploratory Data Analysis) es analizar el impacto que tiene la tasa de variación del PIB sobre la tasa de paro en España tanto a nivel general como en diferentes grupos de población, estos grupos han sido divididos en 4:

- Menores de 25 años (Tasa de paro juvenil).
- Mayores de 25 años.
- Hombres
- Mujeres

Es decir, primero estudiaremos la relación entre la tasa de variación del PIB y la tasa de paro a nivel general, después compararemos el impacto que tiene dicha relación entre la tasa de variación del PIB y la tasa de paro atendiendo a la división por edades (mayores y menores de 25) y por último estudiaremos como afecta esa relación a la población en función de su sexo (hombre o mujer).

La finalidad sería poder llegar a las siguientes conclusiones:

- Relación entre la tasa de variación del PIB y la tasa de paro
- Grupo mas afectado por el paro
- Grupo menos afectado por el paro
- Grupo mas sensible a las variaciones del PIB (menor estabilidad laboral)
- Grupo menos sensible a las variaciones del PIB (mayor estabilidad laboral)

2. Recursos utilizados

Para poder realizar este análisis he utilizado diferentes herramientas entre las que se encuentran:

Lenguaje de Programación:

- Python 3.9.5

Librerías:

- Numpy: especializada en el cálculo numérico y el análisis de datos.
- Pandas: manejo y análisis de estructuras de datos.
- Matplotlib: creación de gráficos

Jupyter Notebooks con Visual Studio Code:

- Aquí es donde desarrollamos el lenguaje de programación elegido de manera que se vea mas clara su interpretación que en un script.

Microsoft Word:

- Programa para elaborar la memoria.

Adobe Acrobat Reader DC:

- Programa para presentar la memoria en PDF

Microsoft Power Point:

- Programa para elaborar la presentación del proyecto.

3. BBDD

Para llevar a cabo mi análisis he tenido que descargar dos bases de datos, ambas sacadas de la página oficial del Instituto Nacional de Estadística (INE) <https://ine.es/index.htm> un organismo autónomo encargado de la coordinación general de los servicios estadísticos de la Administración General del Estado y la vigilancia, control y supervisión de los procedimientos técnicos de los mismos.

Entre los trabajos que realiza, destacan las estadísticas sobre la demografía, economía, y sociedad españolas. A través de la página web oficial se pueden seguir todas las actualizaciones de los distintos trabajos y estudios.

1. Tasas de Paro:

Este dataset contiene los datos de las tasas de paro en España de los últimos 20 años (desde el año 2002 hasta el año 2021) divididas en trimestres, en el dataset podemos encontrar las tasas de paro trimestrales de cada una de las comunidades autónomas además de la tasa de paro a nivel nacional.

Los datos están separados en 7 columnas, 5 de ellas son categorías referidas a un grupo poblacional concreto:

1. Periodo: Esto se refiere al año y trimestre de este.
2. Comunidad: Datos referidos a cada comunidad autónoma o al total nacional.
3. Total: Tasa de paro total de cada región indicada.
4. Menores_25: Se refiere a la tasa de paro juvenil, que es la tasa de paro del conjunto de la población menor de 25 años.
5. Mayores_25: Se refiere a la tasa de paro del conjunto de la población con una edad igual o mayor a 25 años.
6. Parados_hombres: Tasa de paro del conjunto de la población masculina.
7. Parados_mujeres: Tasa de paro del conjunto de la población femenina.

2. Variación trimestral del PIB:

Este dataset contiene los datos de la tasa de variación del PIB de los últimos 20 años en España dividido en trimestres al igual que el primer dataset de las tasas de paro, la diferencia es que en esta ocasión los datos son a nivel nacional por lo que tomaremos el dato nacional para cada una de las diferentes Comunidades Autónomas.

Los datos están separados en tan solo 2 columnas:

1. Periodo: Esto se refiere al año y trimestre de este. Esta columna es exactamente idéntica a la del anterior dataset ya que se refiere a los datos comprendidos entre el primer trimestre de 2002 y el cuarto trimestre de 2021.
2. Var_PIB: En esta columna tenemos las tasas de las variaciones trimestrales del PIB para cada trimestre de cada año desde el 2002 hasta el 2021.

4. Análisis

4.1 Exploración y descripción del Database

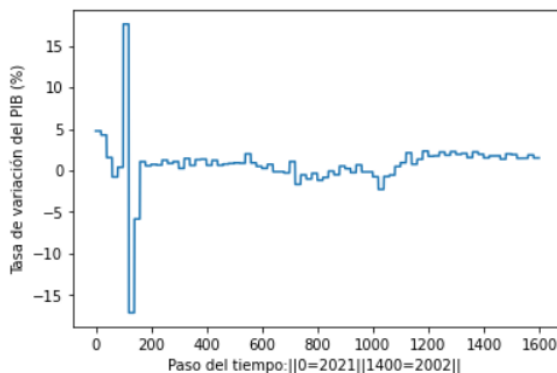
Observo los dataset para echar un vistazo a su contenido y poder detectar posibles errores que me puedan causar problemas futuros.

4.2 Limpieza y preparación de datos

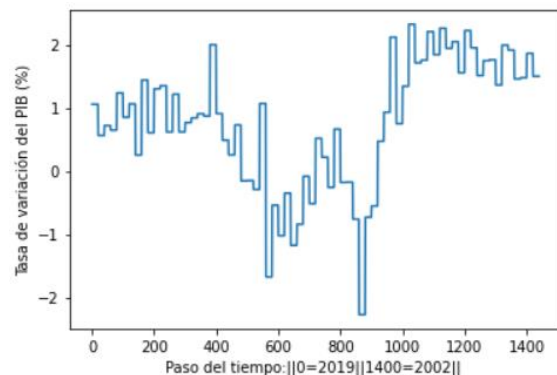
Lo primero que hice fue juntar ambos dataset para crear uno solo con todos los datos y eliminé la columna periodo ya que aparecía duplicada y era exactamente idéntica en ambas bases de datos y la columna Unnamed: 0 que se generó al juntarlos.

Por otro lado dividí la columna Periodo en dos columnas adicionales, una en la que se indicaba el año y otra en la que se indicaba el trimestre de dicho año por si me facilitaba el análisis en el futuro.

Al empezar a analizar los datos mediante gráficas me di cuenta de que los años del COVID eran atípicos, alterando gravemente las gráficas por lo que tuve que eliminar los datos referidos a los años 2020 y 2021 lo que causó cambios bastante significativos:



Con 2021 y 2020



Sin 2021 y 2020

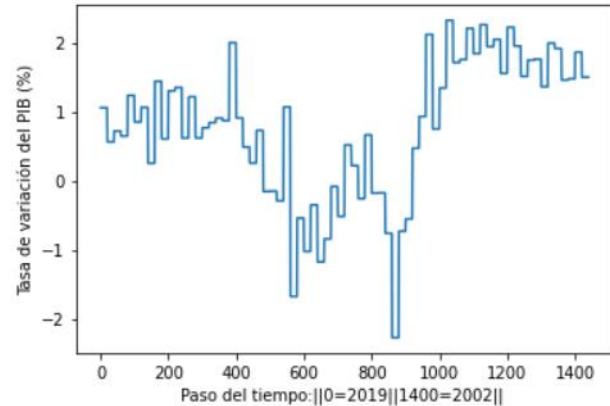
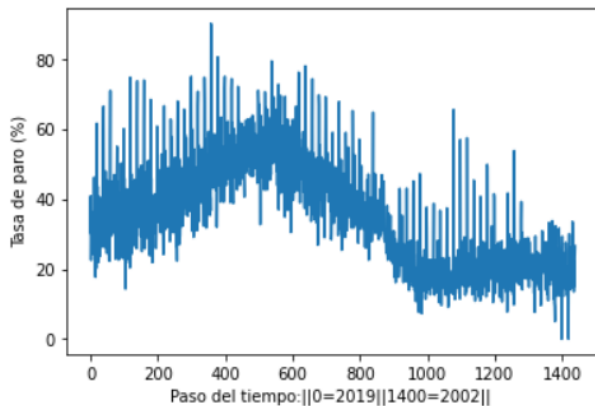
4.3 Análisis y visualización gráfica

Una vez tengo ya listo mi dataset final me pongo a analizar los datos de los que dispongo para que me den una idea inicial de lo que me va a salir como resultado:

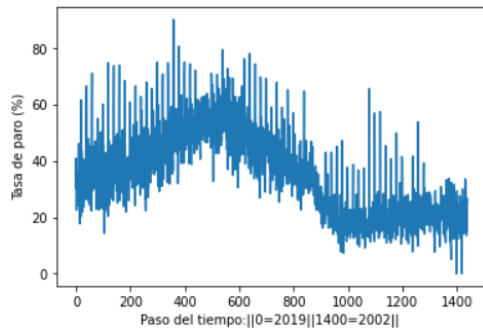
	Total	Menores_25	Mayores_25	Parados_hombres	Parados_mujeres	Var_PIB
count	1440.000000	1440.000000	1440.000000	1440.000000	1440.000000	1440.000000
mean	16.050646	35.158062	14.299319	14.219944	18.641660	0.785196
std	7.518449	15.271719	6.993594	7.524689	8.192455	1.015171
min	0.410000	0.000000	0.460000	0.000000	1.210000	-2.267600
25%	10.050000	22.212500	8.637500	7.907500	12.537500	0.150125
50%	14.795000	33.675000	13.230000	12.980000	16.810000	0.867850
75%	21.005000	46.625000	18.975000	19.560000	23.715000	1.529675
max	39.190000	90.350000	37.210000	39.890000	55.280000	2.330400

Aquí ya podemos intuir que el grupo poblacional que mas sufre el desempleo son los menores de 25 años, los cuales tienen una media de tasa de paro superior al doble de la tasa de paro media total, por el contrario, el grupo poblacional de los hombres es el que menos sufre el desempleo con la tasa de paro media mas baja de todos los grupos, pero esto es a priori, debemos analizar mas los datos antes de sacar cualquier conclusión.

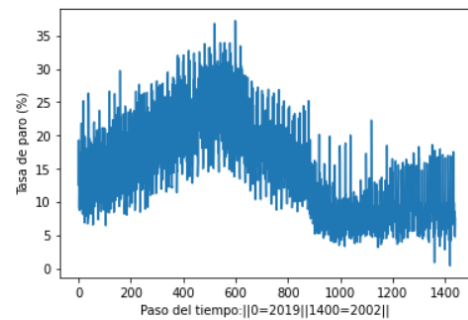
Para continuar, calculo la covarianza y el coeficiente de correlación entre el paro total y la tasa de variación del PIB, la cual, como era de esperar, es negativa, lo que quiere decir que están relacionadas de manera inversa, y de hecho así puede verse en los gráficos, ya que el gráfico de la tasa de paro es cóncavo y el de la tasa de variación es convexo, cuando una variable sube, la otra baja:



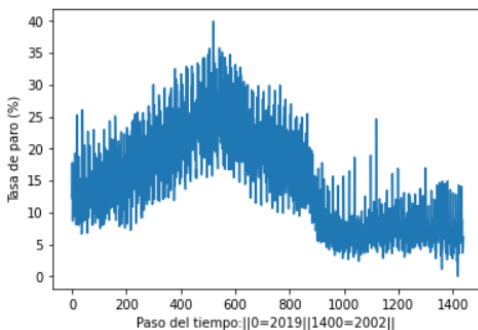
A parte de ese gráfico se sacaron los gráficos de todos los grupos poblacionales, y todos tenían la misma forma:



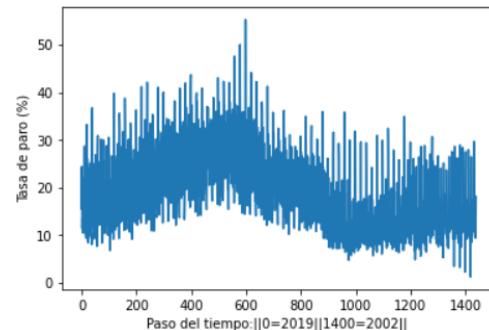
Menores de 25



Mayores de 25



Hombres

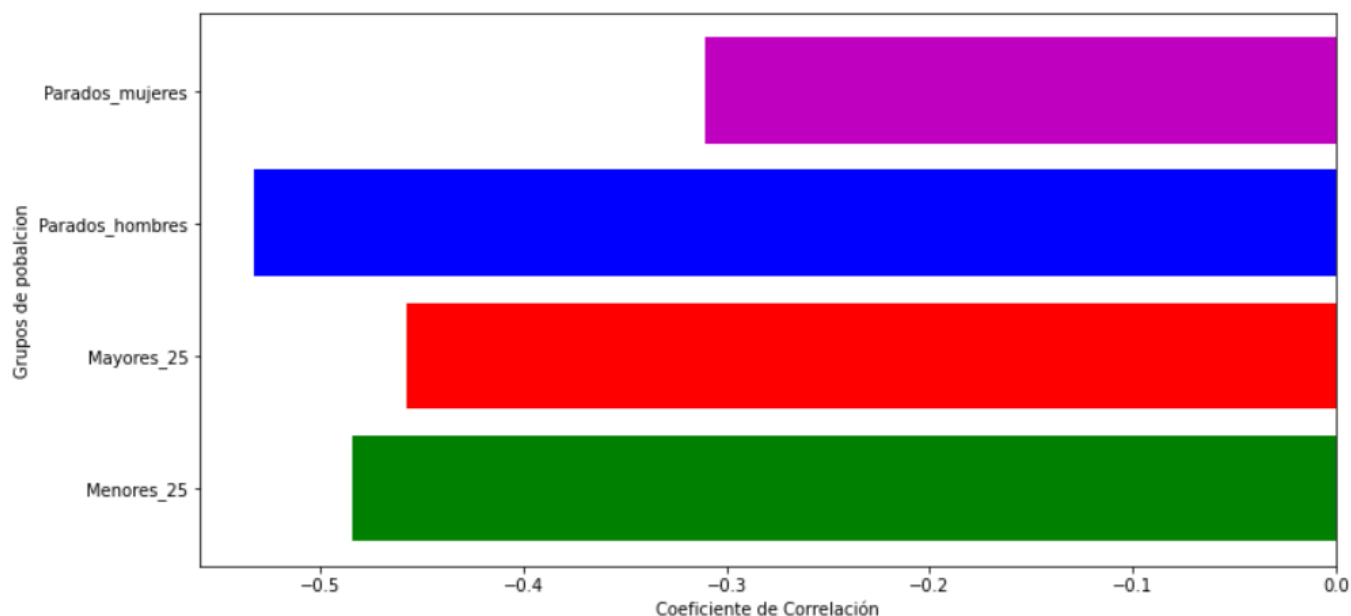


Mujeres

Para continuar el análisis lo que hice fue calcular los coeficientes de correlación de cada uno de los grupos, ya que como podemos observar cada uno tiene una concavidad más o menos acentuada, lo que quiere decir que no todos los grupos responden igual ante una bajada del PIB:

- Para el grupo Menores_25 el coeficiente de correlación entre la variación del PIB y la tasa de paro es de -0.4844075817740141
- Para el grupo Mayores_25 el coeficiente de correlación entre la variación del PIB y la tasa de paro es de -0.45786141247544593
- Para el grupo Parados_hombres el coeficiente de correlación entre la variación del PIB y la tasa de paro es de -0.5328564800617882
- Para el grupo Parados_mujeres el coeficiente de correlación entre la variación del PIB y la tasa de paro es de -0.3108996031365061

Tras hacer los cálculos los transformé en un gráfico para poder ilustrar esas desigualdades:



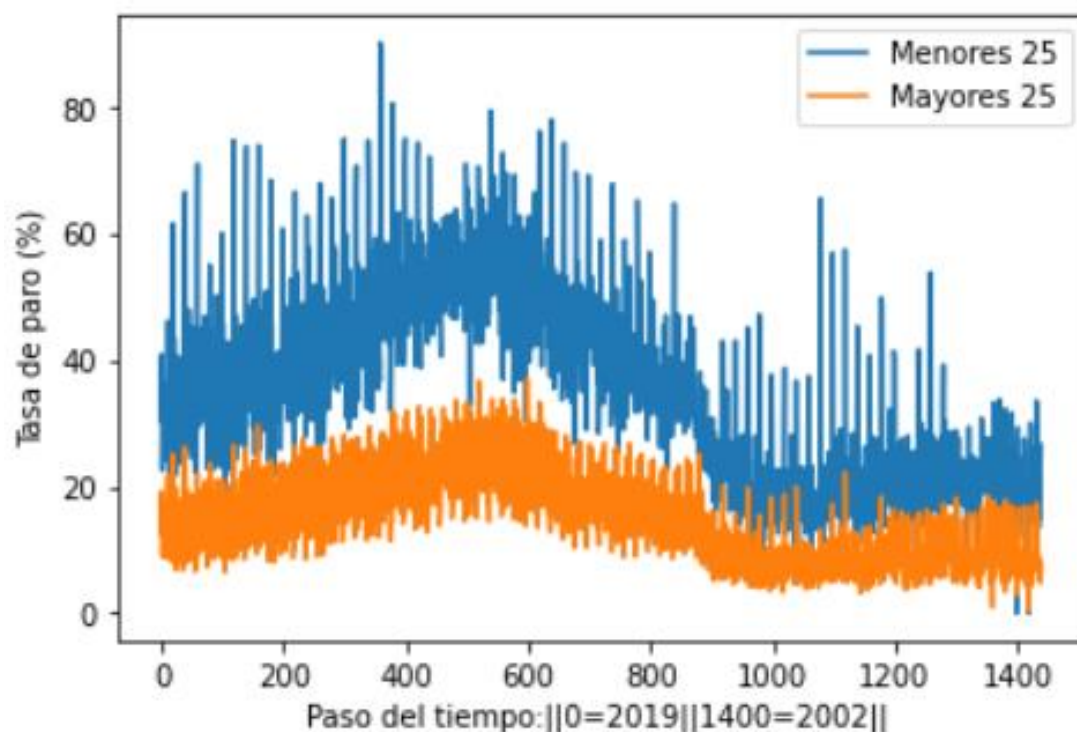
Aquí se puede apreciar como los hombres presentan mayor sensibilidad ante las variaciones del PIB que cualquier otro grupo, y las mujeres el que menos. Lo que nos deja claro este gráfico es que a nivel de sensibilidad con respecto a la tasa de variación del PIB es mas importante si perteneces al grupo de hombres o mujeres que si eres mayor o menor de 25 ya que si bien los menores de 25 presentan una mayor correlación, la diferencia con los mayores de 25 no es tan grande como la que hay entre hombres y mujeres.

El siguiente paso fue comparar cada grupo poblacional con su opuesto:

1. Comparación mayores y menores de 25:

Si comparamos el coeficiente de correlación de los menores de 25 con el de los mayores de 25 podemos apreciar que es un poco mayor el de los menores de 25, lo que quiere decir que ante una bajada de la tasa de variación del PIB, la tasa de paro de los menores de 25 aumentará más que la de los mayores de 25.

Gráfico de la tasa de paro de los Mayores y menores de 25:



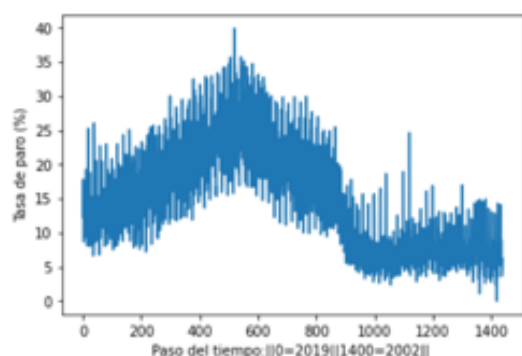
En los gráficos también se puede apreciar esa forma cóncava, contraria a la gráfica de variación del PIB, siendo más acentuada la concavidad de los menores de 25, indicando una mayor sensibilidad ante las variaciones del PIB, tal y como se veía con el coeficiente de correlación.:

2. Comparación por sexos:

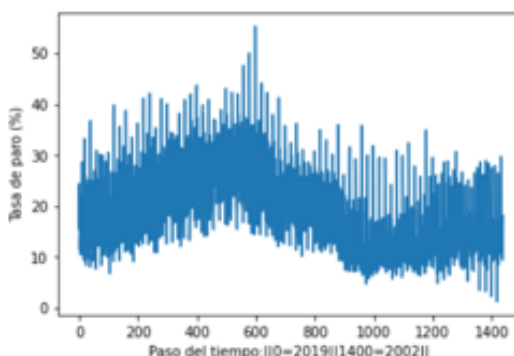
Si comparamos el coeficiente de correlación de los hombres parados con el de las mujeres paradas podemos apreciar que es significativamente mayor el de los hombres parados, lo que quiere decir que ante una bajada de la tasa de variación del PIB, la tasa de paro de los hombres aumentará más que la de las mujeres.

En los gráficos también se puede apreciar esa forma cóncava, contraria a la gráfica de variación del PIB, siendo más acentuada la concavidad de los hombres parados, indicando

una mayor sensibilidad ante las variaciones del PIB, tal y como se veía con el coeficiente de correlación.



Hombres



Mujeres

5. Conclusiones

5.1 Grupo más afectado por el paro:

Los más afectados por el paro son los menores de 25, independientemente de si el PIB aumenta o disminuye ya que de media hay muchos más parados menores de 25 que cualquier otro conjunto poblacional (35%), llegando a ser este casi el doble del segundo grupo con más tasa media de paro, las mujeres (18%).

5.2 Grupo menos afectado por el paro:

El grupo menos afectado por el paro es sin duda el conjunto poblacional formado por personas mayores de 25 años, ya que a pesar de tener de media un 1% más de paro que el conjunto de los hombres, el coeficiente de correlación de los hombres es mucho más cercano a -1, es decir, que ante una bajada del PIB, el paro del conjunto de los hombres aumentará más que el conjunto de los mayores de 25 años.

5.3 Grupo más sensible a las variaciones del PIB:

Sin duda alguna, el grupo al que más afectan las variaciones del PIB es el conjunto de los hombres, ya que aunque es el grupo que menos paro presenta de todos también es el que tiene una correlación más cercana a -1 de todos, es decir, que ante una bajada del PIB son el grupo donde más trabajo se va a destruir, y en cambio ante una época de prosperidad donde el PIB aumenta, es el conjunto de la población donde más trabajo se va a generar.

5.4 Grupo menos sensible a las variaciones del PIB:

Sin duda alguna, en este ámbito el caballo ganador son las mujeres, a pesar de tener una tasa de paro media ligeramente superior a la media total, masculina y de mayores de 25, su coeficiente de correlación es de tan solo un -0.3, es decir, el más alejado de -1 de todos, por lo que ante una bajada del PIB (recesión) las mujeres que estén empleadas tendrán más posibilidades de mantener sus empleos ya que la tasa de paro de las mujeres no aumenta tanto ante las caídas del PIB. Por el contrario, ante una situación de expansión o aumento del PIB no se generará tanto empleo como en los otros grupos.

5.5 Comparación con las impresiones iniciales:

En un inicio pensábamos de manera acertada que el grupo mas afectado por el paro eran los menores de 25 ya que su tasa de paro media es muy superior a la del resto de grupos, en cambio, en un inicio parecía que el conjunto poblacional masculino era el mejor de todos laboralmente y no ha resultado ser del todo cierto ya que a pesar de tener la tasa de paro media mas baja es el conjunto poblacional mas sensible ante las variaciones del PIB, siendo en este donde más empleo se destruirá ante una caída del PIB, al contrario que las mujeres que parecen ser las que mas estabilidad tienen. En cualquier caso parece que lo importante para asegurarte no estar en paro es ser mayor de 25 años ya que parece ser el conjunto poblacional que mejores condiciones tiene, una tasa de paro media baja y una sensibilidad ante las caídas del PIB no muy grande.