

Trabajo Final de Máster - Data Science 2020-2021

**“Análisis de la inversión española en los
Estados Unidos Mexicanos”**



Introducción	2
Análisis individual de las variables	5
Balanza de pagos	5
Tasa de desempleo	6
Exportaciones	7
Importaciones	8
Inversión Extranjera Directa (IED)	9
Índice de precios del consumidor (IPC)	10
Producto Interior Bruto (PIB)	11
Deuda externa	12
Reservas	13
Tasa de interés	14
Inversión española en México	15
Análisis de la inversión española y las variables macroeconómicas en conjunto	25
Análisis de la inversión española en México en función de la Inversión Extranjera Directa de México (IED)	25
Modelo de Regresión Lineal (OLS)	28
Conclusión	35

1. Introducción

Cuatrecasas es una firma de abogados española con presencia nacional en diferentes territorios y con presencia internacional en Portugal (Lisboa y Oporto), Bogotá, Bruselas, Casablanca, Ciudad de México, Lima, Londres, Luanda, Maputo, Nueva York, Pekín, Santiago de Chile y Shanghai.

Una de las principales estrategias comerciales de la firma es asentar nuestra presencia en países latinoamericanos.

El principal target en las oficinas internacionales son las empresas españolas que inviertan o quieran invertir en países extranjeros.

Este estudio tiene por finalidad hacer un análisis de las variables macroeconómicas que mantienen una mayor correlación con la inversión española para que nos pueda ayudar en la toma de decisiones a la hora de buscar diferentes targets. En concreto nos centraremos en el país de México.

Las variables macroeconómicas que se utilizadas en el estudio son:

- PIB
- IPC
- Balanza de pagos
- Tasa de desempleo
- Exportaciones
- Importaciones
- Inversión Extranjera Directa (IED)
- IPC
- PIB
- Deuda pública
- Reservas
- Tasa de interés

Se han seleccionado estas variables porque son las que Bank of America (BoA) utiliza para medir el riesgo del negocio basado en razones económicas y financieras.

El estudio está dividido en diferentes fases:

1. Extracción del dato
2. Lectura y análisis de las variables
3. Estudio del conjunto de variables
4. Modelo de regresión
5. Análisis de la inversión española en México y de la inversión extranjera directa (IED) de México.

Para poder visualizar el estudio utilizaremos el entorno virtual de Jupyter Lab con Python 3.7.7 o en adelante con las siguientes librerías:

- matplotlib 3.0.2
- numpy 1.19.2
- pandas 1.2.4
- patsy 0.5.1
- scipy 1.6.3
- seaborn 0.11.1
- sinfo 0.3.4
- sklearn 0.24.2
- statsmodels 0.12.2

1.1. Guía de ejecución

El estudio se encuentra disponible en el repositorio de [GitHub](#). Encontraremos diferentes carpetas:

- **variables_economicas:** donde se encuentran los datos que han sido extraídos para el estudio en formato CSV o XLS.
- **Python:** Se encuentran todos los notebooks necesarios para realizar el análisis.

Pasos para desarrollar el estudio seguiremos el siguiente orden de ejecución de los notebooks de la carpeta Python de Github:

1. 1-función_variables.pynb --> en este notebook hay dos funciones que son necesarias para poder ejecutar el siguiente paso. Para poder ejecutarlo es importante guardar el fichero 'funciones_variables.py' que es el que ejecuta el script
2. Ejecutar todos los notebooks en los que el nombre empiece por 2. En total son 11 notebooks.
 - a. 2-Edit_Balanza de pagos.pynb
 - b. 2-Edit_Desempleo.pynb
 - c. 2-Edit_Exportaciones.pynb
 - d. 2-Edit_IED.pynb
 - e. 2-Edit_IPC.pynb
 - f. 2-Edit_importaciones.pynb
 - g. 2-Edit_PIB.pynb
 - h.
3. Inversion ES vs IED.pynb --> se analiza la inversión española en México respecto su total (IED).
4. 3-Variables_economicas.pynb --> limpieza y modelo de regresión (OLS)

1.2. Adquisición de datos

Los datos han sido obtenidos de diferentes fuentes:

- [Banco Mundial](#): PIB, desempleo, exportaciones, IED, IPC, importaciones, PIB, deuda publica, reservas y tasa de interés.
- [ICEX](#): seleccionamos la opción 'Inversión española en LATAM'. Accederemos al **modo avanzado** para que nos permita seleccionar el país destino y nos muestre también la información sectorial y el periodo. Seleccionaremos todos los datos que estén disponibles. Todos los sectores desde 1993 hasta 2020. El país destino seleccionaremos 'México'.

1.3. Estructura de los datos

Los datos macroeconómicos descargados del Banco Mundial se han tratado para poder analizarlos.

Estos datos podemos diferenciarlos en dos tipos:

- Porcentajes (tasa de desempleo, tipo de interés): El Banco Mundial nos devuelve este dato de todos los países del mundo. Para poder tratarlo se ha creado una función en la que se lee el archivo y se selecciona la fila que nos interesa, México. Se redondea para que nos devuelva el valor con dos decimales.
- Valores: Al igual que pasa con los porcentajes, el Banco Mundial nos devuelve este dato de todos los países y hacemos el mismo tratamiento. El valor nos viene dado en dólares. Para tenerlo todo en la misma unidad del Banco Mundial se extrajo la tasa de cambio de dólares a euros para el periodo de tiempo que vamos a analizar y poder realizar el cambio. El fichero con el tipo de cambio se encuentra en el repositorio dentro del apartado `Variables_economicas/Tipo de cambio- Dolar a Euro.csv`. Además, transformamos el valor en millones de euros para que los valores no queden muy grandes.

La variable de la inversión española en México (2-`Inversion ES.pynb`) viene dada en millones de euros. Esta variable cada columna corresponde al valor de la inversión de cada año por los diferentes sectores, un total de 90 sectores diferentes que posteriormente analizaremos. Para poder concatenar este dato con las variables macroeconómicas se utilizará la suma de los sectores por cada año.

Los datos macroeconómicos y el total de la inversión española se concatenan (3-`Variables_economicas.pynb`) quedando un dataset de 11 columnas (variables) y 27 filas (años, de 1993 al 2020).

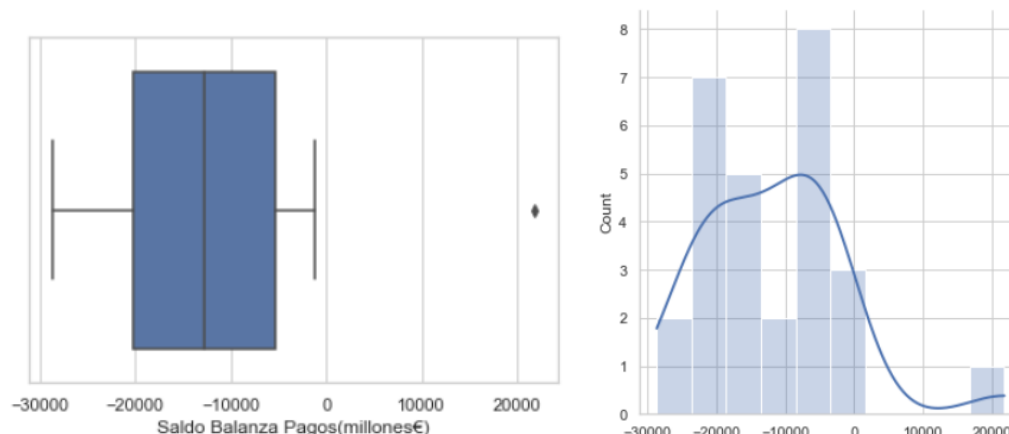
2. Análisis individual de las variables

En este apartado se analiza la distribución y evolución de cada variable junto con una breve explicación de cada una.

2.1. Balanza de pagos

La balanza de pagos es un indicador macroeconómico que proporciona información sobre la situación económica de un país de manera general. Es decir, permite conocer todos los ingresos que recibe un país procedentes del resto del mundo y los pagos que realiza tal país al resto del mundo debido a las importaciones y exportaciones de bienes, servicios, capital o transferencias en un período de tiempo.

En los siguientes gráficos destaca que la mayoría de los años la balanza de pagos se encontraba en déficit, con una media de 11.657,97 millones de euros. La media y la mediana (-12.828), se acercan llegando a tener casi una distribución simétrica de los datos. La mayoría de los datos se distribuyen entre -20.148 y -5.357. En ambos gráficos se puede apreciar un dato alejado del resto de valores, 21.748.



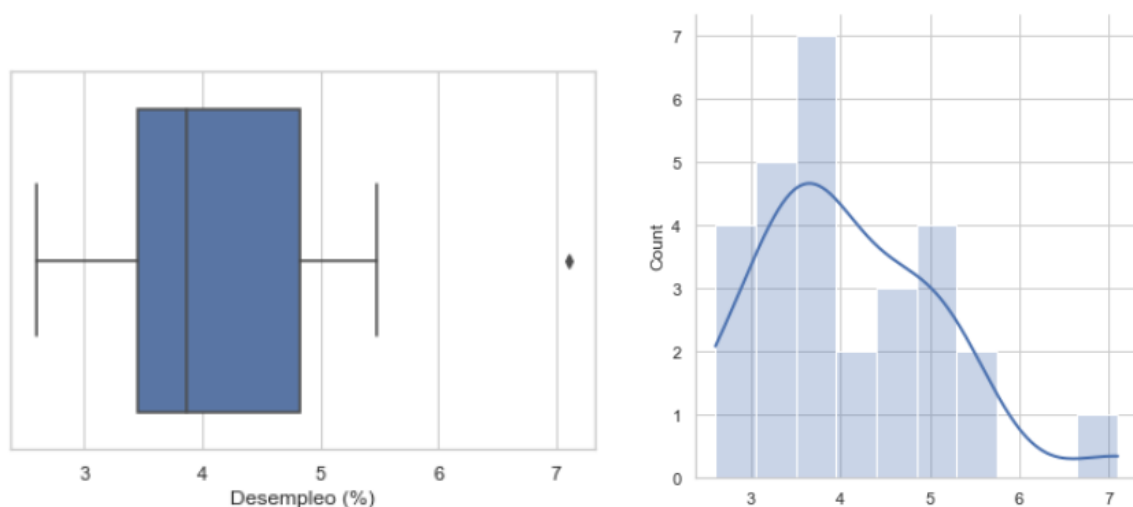
En la evolución de la balanza de pagos podemos apreciar que México ha estado siempre en déficit, excepto en 2020. En 1995 es la vez que más cerca estuvo con -1.177 millones de €. Luego sufrió una caída debido a una crisis financiera internacional y la devaluación del precio del petróleo. En 2001 empezó a crecer ligeramente. En 2010 por la crisis financiera mundial decreció hasta alcanzar el mínimo del análisis -28.594 millones €. Empezó a recuperarse hasta que en 2020 alcanzó el superávit (21.748 millones €), esto es debido a la importante ampliación del superávit de la balanza de mercancías no petroleras.



2.2. Tasa de desempleo

La tasa de desempleo es un porcentaje que expresa la proporción de parados que hay respecto al total de activos. Se considera cómo parado aquel trabajador que no tiene empleo y lo busca de forma activa.

En los siguientes gráficos podemos apreciar que la media de la tasa de desempleo es de 4.09%. La mayoría de los datos oscilan entre 3.45% y 4.83%. Destaca el máximo de 7.1% y el mínimo de 2.6%. La mediana (3.865%) es inferior a la media, por lo que los datos siguen una asimetría positiva. Los datos se concentran en la parte inferior de la distribución.



En el siguiente gráfico vemos la evolución de la tasa de desempleo a lo largo de los años. Uno de los grandes problemas en México durante muchos años ha sido la falta de empleo y la oleada migratoria. Destaca la subida de la tasa de desempleo en los 3 primeros años (1993-1996) debido a la crisis económica. Empezó a decrecer alcanzando la tasa de desempleo mínima del 2.6%. Se mantiene estable hasta 2001, donde empieza a subir la tasa de desempleo. Entre 2004 y 2008 la tasa de desempleo decreció como resultado del incremento del PIB. En 2008 se vuelve a apreciar otro aumento, debido a la crisis financiera mundial.

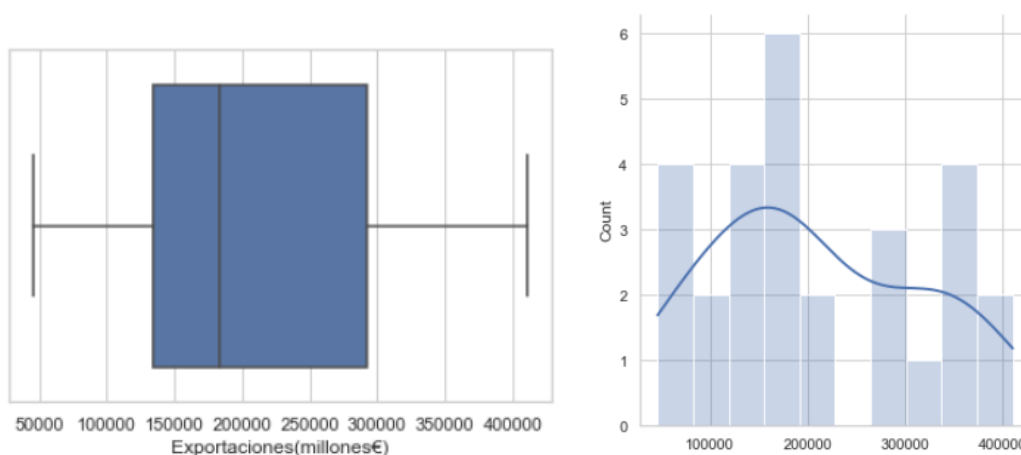
La tasa de desempleo empieza a descender ligeramente durante los 6 siguientes años, hasta 2014. A partir de este momento empieza a descender algo más rápido debido al turismo. Sufriendo otro aumento en 2020 por la crisis financiera provocada por la COVID.



2.3. Exportaciones

Las exportaciones son el conjunto de bienes y servicios vendidos por un país en territorio extranjero para su utilización.

La mayoría de los datos de las exportaciones se encuentran distribuidos entre 134.388 y 292.701 millones €. La mediana está sesgada a la derecha, es decir, los datos se concentran en la parte inferior de la distribución. La media (208.597) es mayor que la mediana (183.569). En ese caso no se aprecian datos atípicos.



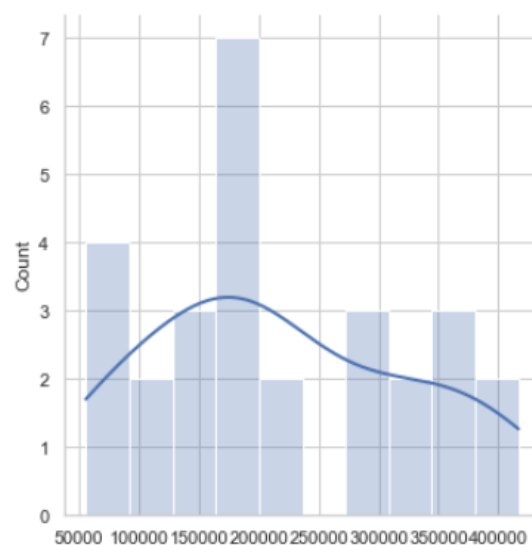
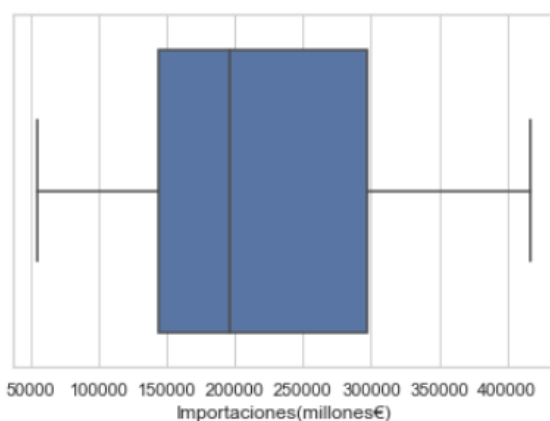
Las exportaciones en México siguen una tendencia creciente esto es debido a que México en 1992 entró en el Tratado de Libre Comercio de América del Norte (TLCAN) junto con Estados Unidos de América y Canadá. El tipo de cambio de Estados Unidos de 1993 a 2000 era muy estable lo que benefició a las exportaciones. Se aprecian en el gráfico unas ligeras caídas de las exportaciones en 2001 y 2009, la primera provocada por la crisis económica que vivió Estados Unidos tras el atentado y en 2009 por la crisis financiera mundial. En 2020 vuelven a caer las exportaciones por la crisis sanitaria que provocó el cierre de fronteras.



2.4. Importaciones

Las importaciones pueden ser cualquier producto o servicio recibido dentro de la frontera de un Estado nacional con propósitos comerciales.

Las importaciones en México siguen una distribución de los datos muy similar a las exportaciones. El 50% del valor de las importaciones se encuentran entre 196.585 y 296.683. La mediana tiene una asimetría positiva, es decir, los datos se concentran en la parte inferior de la distribución. La media (218.815) es mayor que la mediana (196.585). El valor mínimo es de 55.574 y el máximo de 416.728.



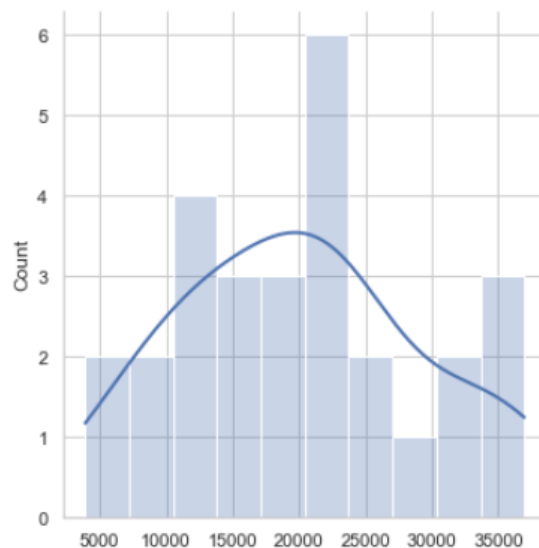
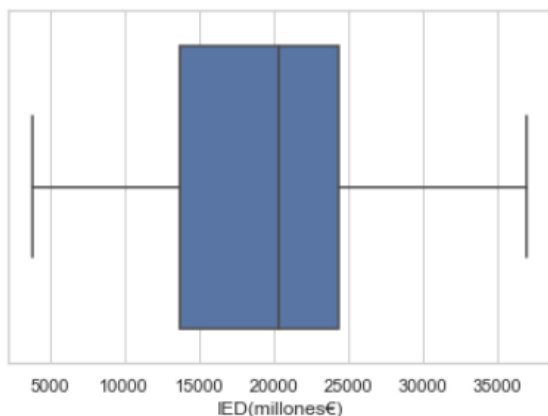
La tenencia de las importaciones es la misma que las exportaciones, tendencia creciente. Aumentan por la entrada de México en TLCAN. Se aprecia la caída de las importaciones en 2001, crisis económica de Estados Unidos, en 2009 por la crisis financiera mundial, y 2020 por la crisis sanitaria provocada por la COVID.



2.5. Inversión Extranjera Directa (IED)

La Inversión Extranjera Directa (IED) es aquella que tiene como propósito crear un vínculo duradero con fines económicos y empresariales de largo plazo, por parte de un inversionista extranjero en el país receptor.

Los datos se encuentran distribuidos entre 3.903 millones € y 36.945 millones €. La mayoría de los datos se encuentran entre 13.776 y 24.394 millones €. Siendo la media, 20.117 millones €, menor que la mediana, 20.402 millones €, por lo que los datos se encuentran sesgados a la izquierda. Es decir, los datos se concentran en la parte derecha de la distribución.



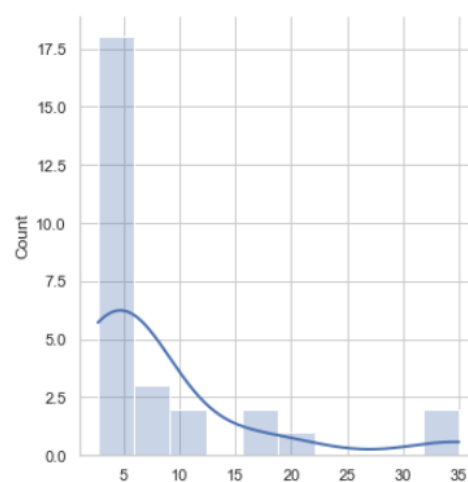
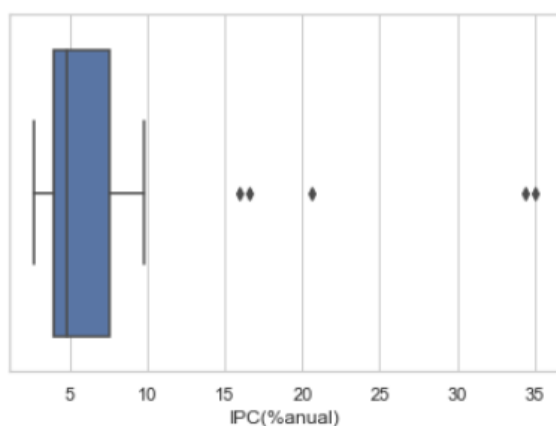
La IED sigue una tendencia creciente. En 2001 recibe un alto volumen de inversión, 33.742 millones de euros por el auge de la economía en Estados Unidos y el TLCAN. En 2002 caen las inversiones debido a que la actividad económica del país también está bajando y por la lenta recuperación de la economía estadounidense y por el cierre de un número importante número de empresas localizadas en la frontera norte. En 2009 se aprecia una caída provocada por la crisis mundial y el 2013 se alcanza el valor máximo de IED, 36.945 millones de euros, el valor más alto alcanzado desde que México registra este dato. Esto es debido a una operación de M&A, Grupo Modelo adquiere parte de AB Inbev, solo esta operación estaba valuada en 20 millones de euros.



2.6. Índice de precios del consumidor (IPC)

El índice de precios del consumidor (IPC) es un índice económico en el que se valoran los precios de un determinado conjunto de bienes y servicios (cesta familiar) determinado sobre la base de la encuesta continua de presupuestos familiares y la variación respecto del precio de cada uno.

La mayoría de los datos del IPC se distribuyen entre 3.93% y 7.6%, una distribución exponencial decreciente.. Los datos tienen una asimetría negativa, esto es que la media (8.3%) es menor que la mediana (4.79%). Los datos se concentran en la parte derecha de la distribución. Destacan 5 valores atípicos, 35%, 34.38%, 20.63%, 16.59 y 15.93%.



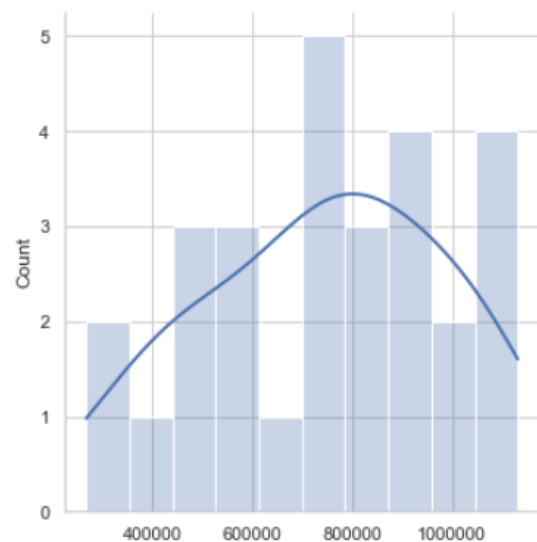
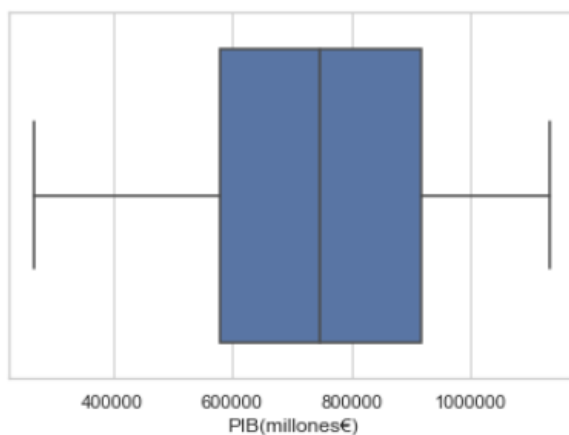
En la evolución podemos apreciar que los valores atípicos corresponden al periodo entre 1995 y 1999. Alcanzando el máximo de 35% en 1995, provocada por una subida del 4.1% en los precios de alimentación y bebidas no alcohólicas y transporte con un 3.1%. A partir de 2002 el IPC se estabiliza con una media del 5%.



2.7. Producto Interior Bruto (PIB)

El producto interior bruto (PIB) es una magnitud macroeconómica que expresa el valor monetario de la producción de bienes y servicios de demanda final de un país en un periodo determinado.

El 50% de los datos se encuentran distribuidos entre 578.834 millones de euros y 915.880 millones de euros. Los datos siguen casi una distribución simétrica dado que la media (744.851 millones€) está muy cercana a la mediana (745.962 millones €). Los datos están distribuidos hacia la derecha.



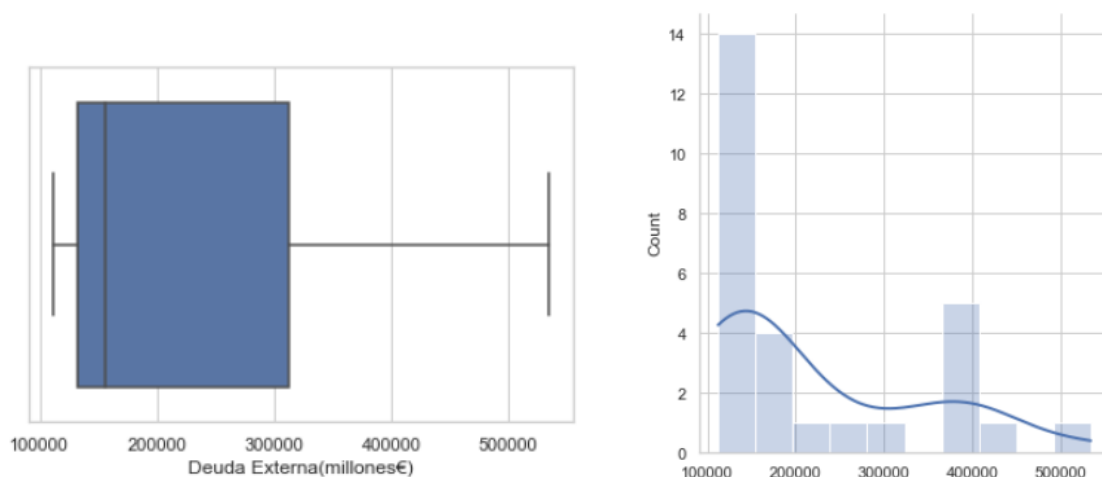
La evolución del PIB sigue una tendencia creciente. Destacando la caída en 1995 por la crisis internacional y la depreciación del valor del petróleo. A partir de 1995 hasta 2001 empieza a subir hasta que en 2001 la crisis de Estados Unidos hace que baje. En 2004 se empieza a recuperar hasta la crisis financiera mundial de 2009 donde vuelve a caer. En 2010 empieza una tendencia creciente que dura 10 años, alcanzando en 2019 el valor máximo de 1.131.451 millones de €. Con la crisis sanitaria de 2020 vuelve a caer a 880.836 millones de euros.



2.8. Deuda externa

La deuda externa es la suma de las deudas que tiene un país con entidades extranjeras. Se compone de deuda pública y deuda privada.

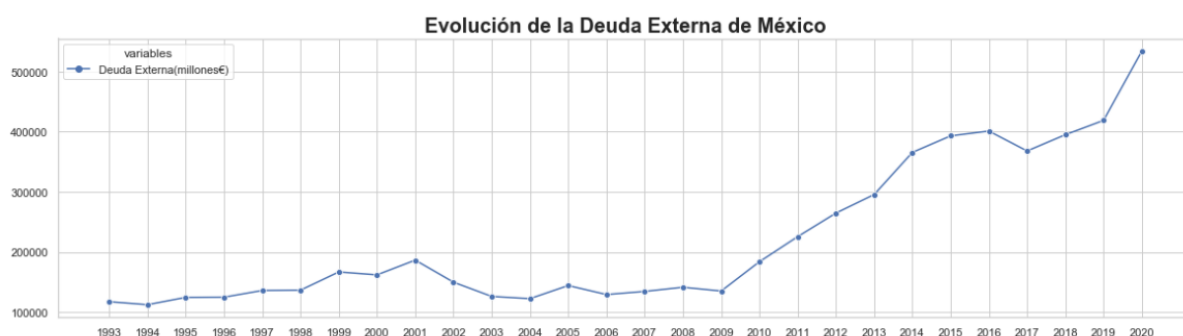
Los datos se encuentran distribuidos entre 112.309 millones de € y 534.121 millones de €. El 50% de los datos se concentran en la parte izquierda de la distribución, entre 113.104 millones de € y 318.028 millones de €. La mediana (155.833 millones de €) está sesgada a la derecha ya que es menor que la media (221.216 millones de €). Los datos se concentran en la parte inferior de la distribución.



De 1993 a 2008 la deuda externa era más o menos estable, entre 1998 y 2001 sufrió un ligero aumento. El aumento de la deuda externa a partir de 2009 se obtuvo principalmente a través de la emisión de Bonos de Desarrollo a tasa nominal fija y de Udibonos y a la incorporación de los bonos gubernamentales mexicanos a tasa fija al World Government Bond Index (WGBI) de Citigroup.

La solvencia de las finanzas públicas y la credibilidad de la política monetaria han permitido crear un entorno propicio para el desarrollo de los mercados financieros, facilitando con ello el flujo de capitales para financiar las inversiones en México.

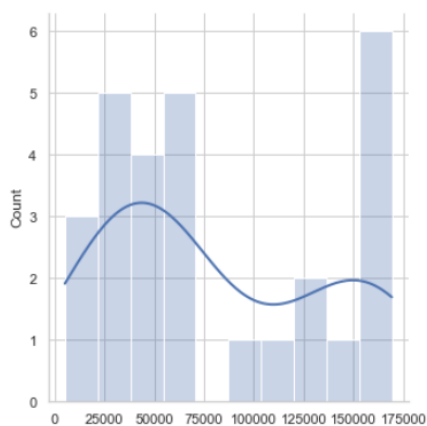
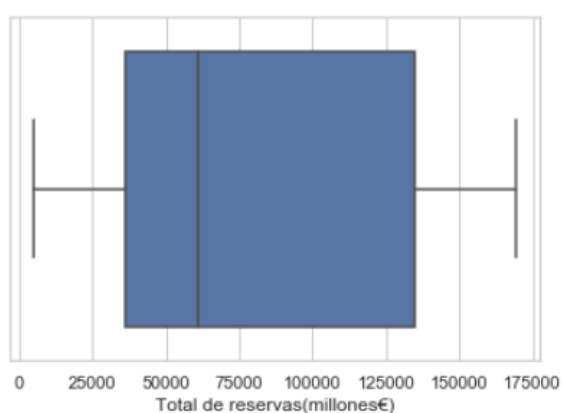
En 2020 se registra el valor más alto de deuda externa situándose en 534.121 millones de €.



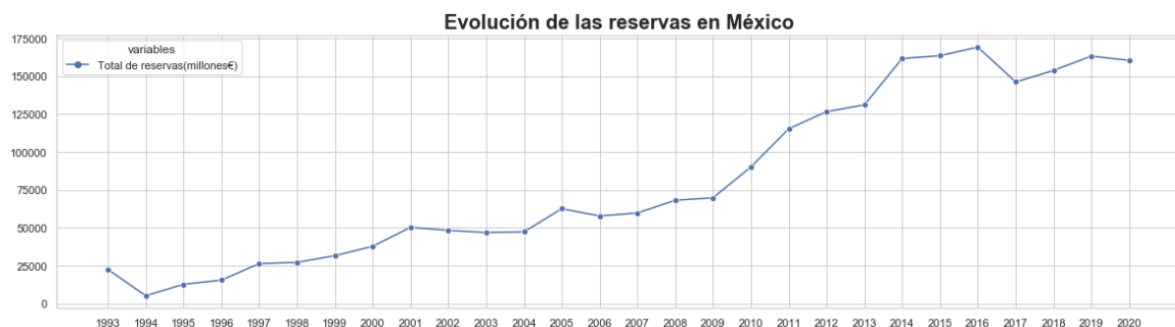
2.9. Reservas

Las reservas internacionales consisten en depósitos de moneda extranjera controlados por los bancos centrales y otras autoridades monetarias. Funcionan como indicador económico mostrando los recursos que dispone un país para hacer compras en el extranjero.

La mayoría de los datos se encuentran distribuidos entre 36.211 y 134.862. Siendo el valor máximo 169.199 y el valor mínimo 5.170. La media 81.098 es superior a la mediana 61.167, por lo que la mediana está sesgada a la derecha, es decir, los datos se concentran en la parte superior de la distribución.



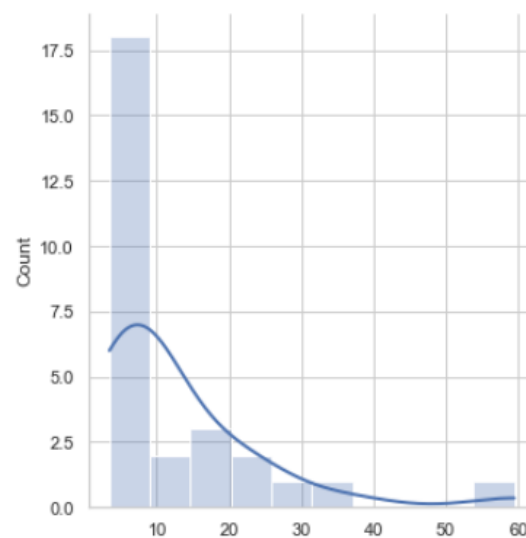
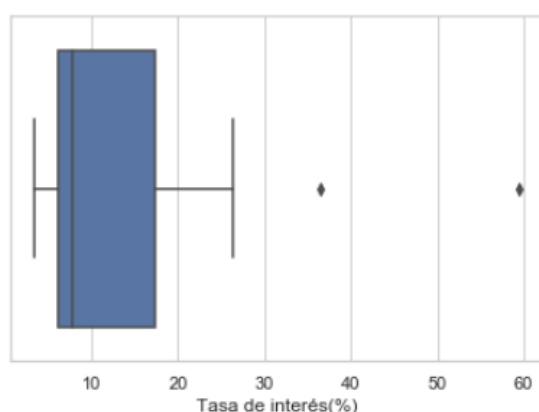
Las reservas internacionales en México siguen una tendencia creciente, alcanzando su menor valor en 1994 5.170 millones € y el máximo en 2016 169.199. Destaca el crecimiento de la reserva en 2014, debido a que el Gobierno federal vendió dólares a Banxico por un valor de 1.000 millones de dólares.



2.10. Tasa de interés

El tipo o tasa de interés es el porcentaje que el banco central carga a los bancos. Un cambio en el tipo o tasa de interés es normalmente un reflejo del cambio en la situación económica, y no una medida intencionada de política monetaria.

En la distribución se aprecia que los datos están distribuidos a la izquierda, entre 3.44% y 26.36%, excepto dos valores atípicos 36.39% y 59.43%. La mayoría de los datos se encuentran distribuidos entre 6.07% y 17.25%. Hay una asimetría positiva ya que la media (12.84%) es mayor que la mediana (7.8%), por lo tanto los datos se concentran en la parte inferior de la distribución.



En la evolución de la tasa de interés destaca el pico en 1995 alcanzando el valor máximo de 59.43%. Esto fue debido a la depreciación de la moneda nacional, el repunte de la inflación y las expectativas inflacionarias provocado por la crisis que estaba sufriendo el país. A partir de 1998 empieza a disminuir por un cambio en la política monetaria del país.

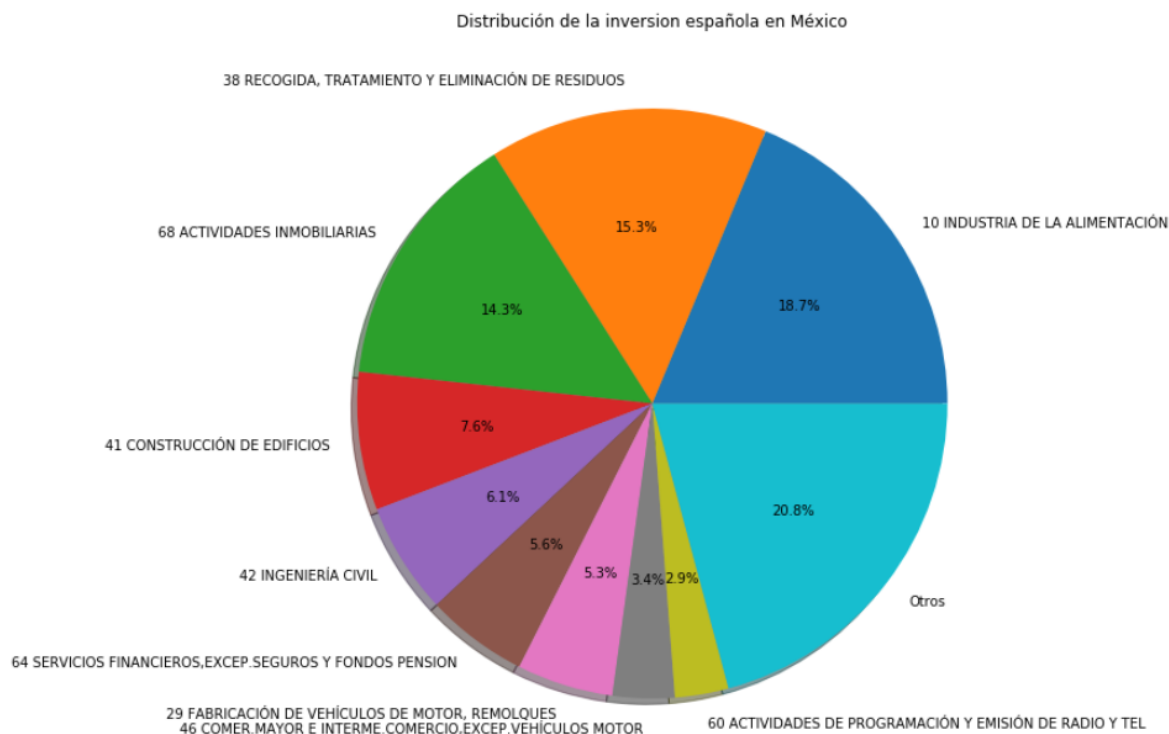


2.11. Inversión española en México

Esta variable nos indica el volumen de la inversión española en México. Por una parte estudiaremos los sectores a los cuales va dirigida dicha inversión y por otra el total de la inversión. Se trata de nuestra variable dependiente por lo que el análisis de esta variable será más profundo.

2.11.1. Inversión española en México por sectores

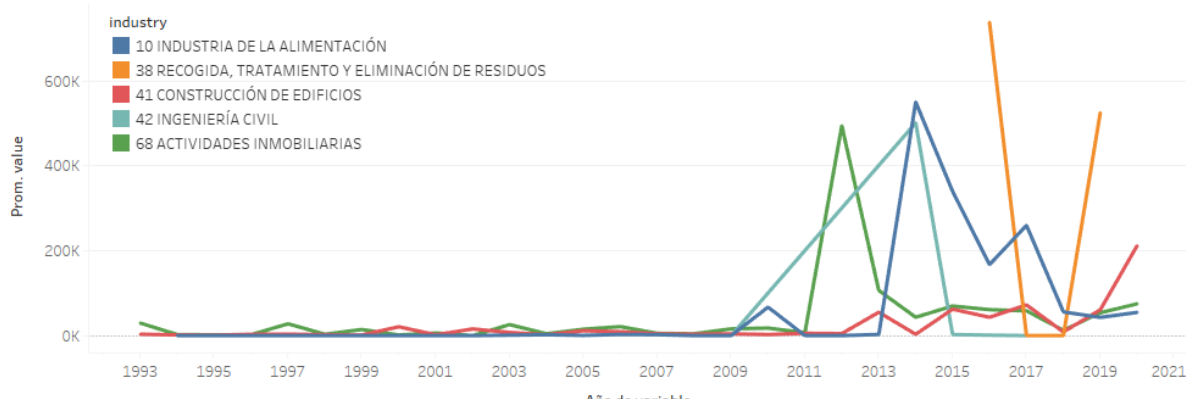
El análisis consta de un total de 90 sectores diferentes. En 10 sectores se reparte el 80% de la inversión.



El sector que más volumen de inversión capta por parte de España es la industria de la alimentación, seguido de 'recogida, tratamiento y eliminación de residuos', 'actividades inmobiliarias', 'construcción de edificios' e 'ingeniería civil'.

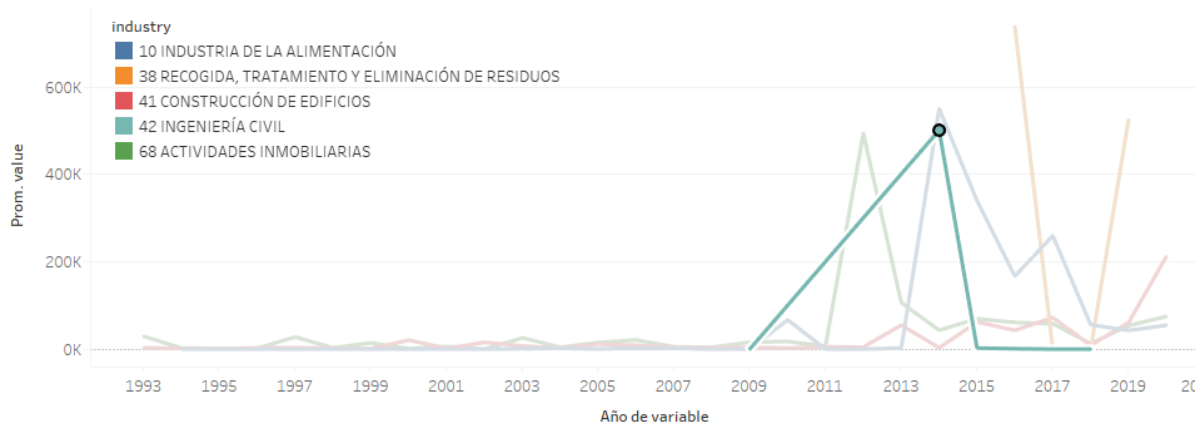
En el gráfico siguiente podemos ver la evolución de los principales sectores receptores de inversión.

Evolución de la inversión por sector

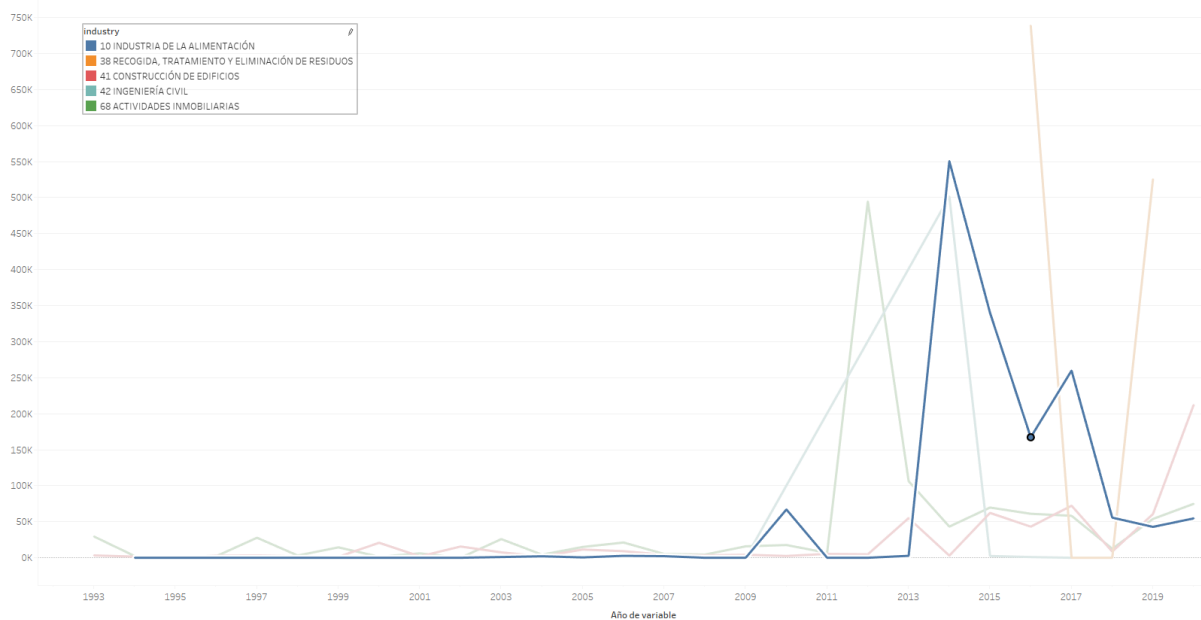


Destaca que la inversión española en estos sectores fue prácticamente nula hasta 2009. Antes de esta fecha la inversión era muy residual, provenía de las actividades inmobiliarias y de la construcción de edificios. A partir de 2010 hay un pico de inversión en el sector de la ingeniería civil y de la industria de la alimentación. Durante los 6 años siguientes, la ingeniería civil es uno de los sectores que más inversión recoge. En 2015 sufre una caída, esto se debe a la construcción de ACS de una planta de petróleo.

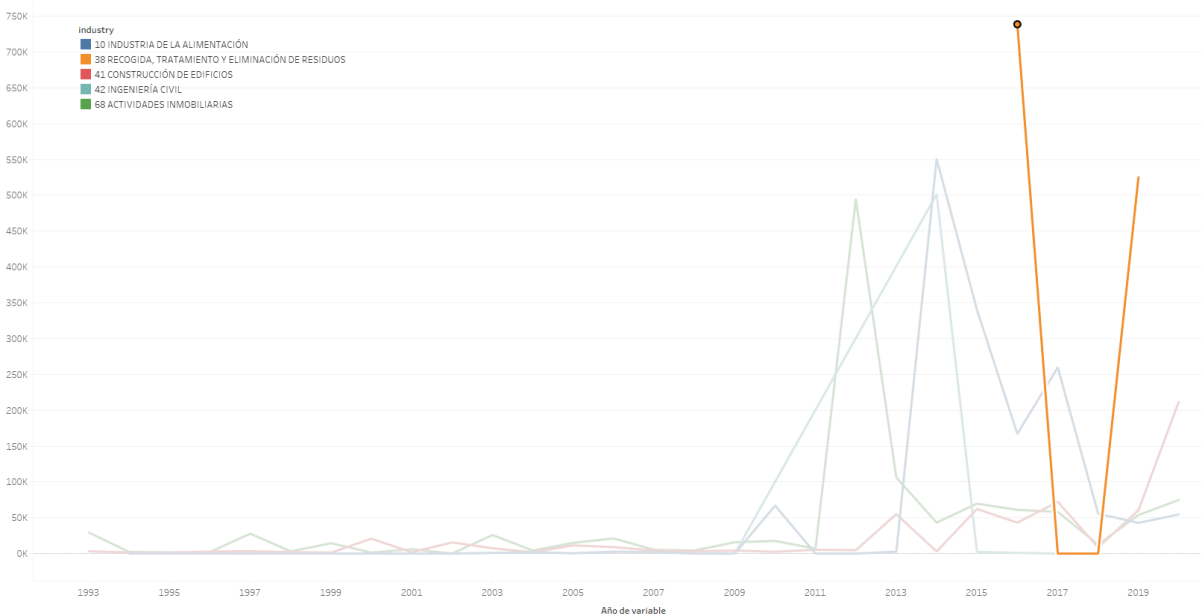
Evolución de la inversión por sector



En el sector de la alimentación empieza la inversión en 2010 con 10.000, pero es en 2014 cuando capta su mayor inversión, 550 millones de euros. La mayoría de esta inversión proviene de Gallina Blanca y de Nutrexpa con la creación de plantas en México.



El segundo sector donde más se ha invertido es en la recogida, tratamiento y eliminación de residuos. Cabe destacar que se invirtió en este sector en 2016 con 728 millones de euros y en 2018 con 525 millones de euros. Ambas debido a la planta de gestión de residuos de Naucalpan.



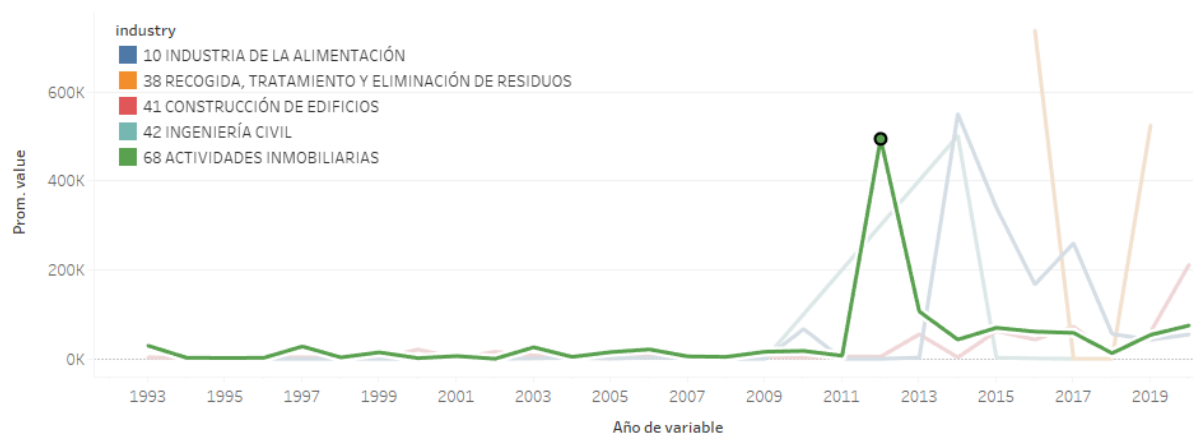
La construcción de edificios capta un total de 625 millones de euros. En este sector la inversión ha sido más moderada. Destacando la tendencia creciente desde 2018.

Evolución de la inversión por sector



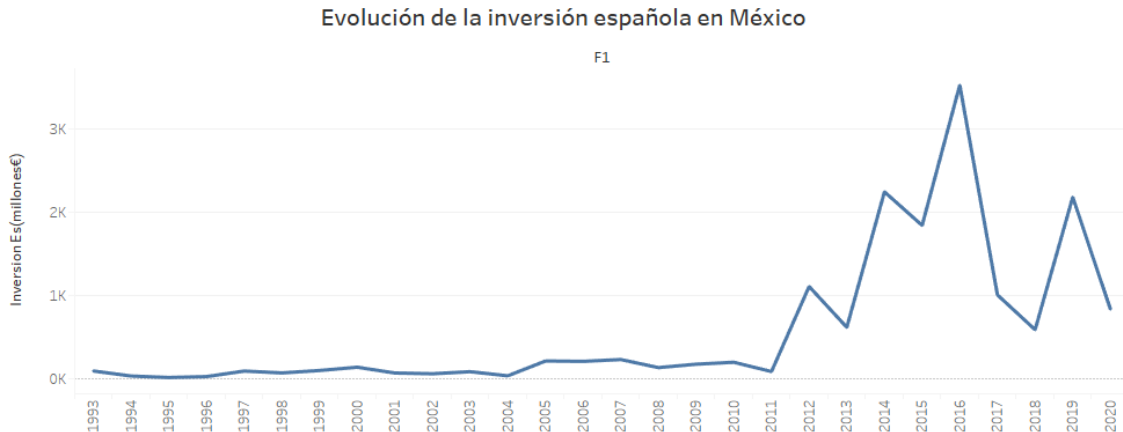
Las actividades inmobiliarias captaron un total de 1.179 millones de euros proveniente de la inversión española. La inversión ha sido moderada excepto en 2012 con una inversión de casi 500 millones de euros.

Evolución de la inversión por sector



2.11.2. Total de la inversión española en México

Durante los primeros 18 años la inversión española en México era muy baja, no fue hasta 2011 que empezó a tener una tendencia creciente. Alcanzando la mayor inversión en 2016 con 3.529 millones de euros. En 2017 la inversión cayó a situándose en 504 millones de euros y en 2018 en 295 millones de euros. En 2019 volvió a aumentar situándose en valores similares a 2014 en 1.091 millones de euros.

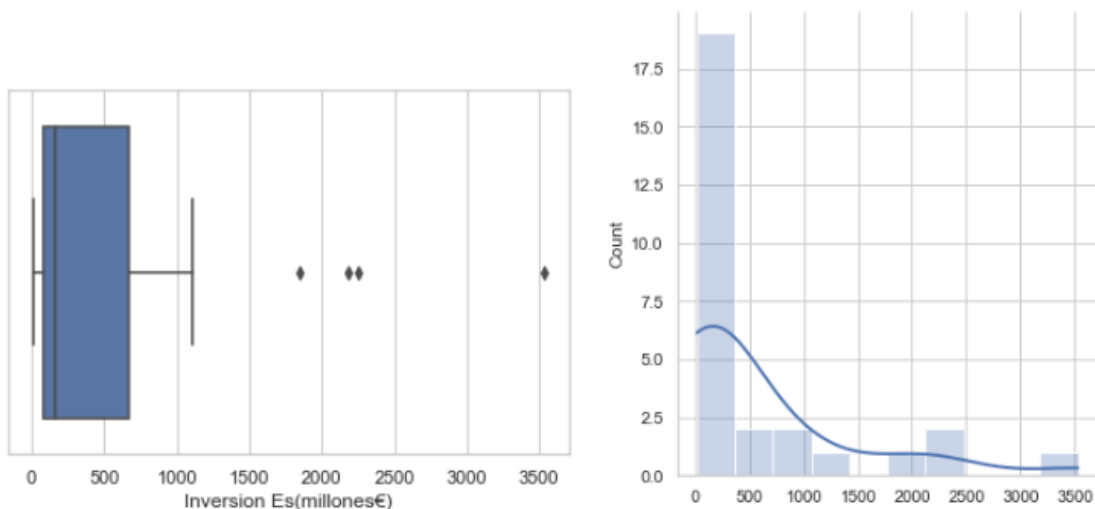


2.11.2.1. Análisis de la normalidad

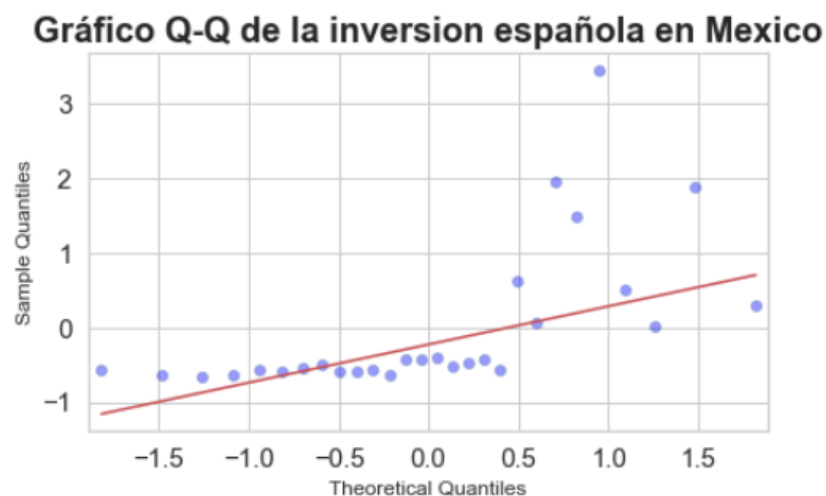
Vamos a analizar si los datos disponibles de la variable dependiente podrían seguir una distribución normal. Emplearemos tres estrategias: representaciones gráficas, métodos analíticos y test de hipótesis.

- **Representaciones gráficas**

En el siguiente gráfico podemos ver la distribución de los datos. La mayoría de los datos se encuentran entre 82 y 673 millones de euros. La media (573 mill €) es mayor que la mediana (158 mill €), lo que nos indica que hay una asimetría positiva, es decir, los datos se concentran en la parte inferior de la distribución, entre 82 y 158 millones de €. Además se puede apreciar 4 datos atípicos. Gráficamente podemos concluir con que no siguen una distribución normal.



El gráfico QQ también nos confirma que los datos no siguen una distribución normal. Se aprecia que los datos no se ajustan a la línea, sobre todo al final del dataframe.



- **Métodos analíticos**

En este caso se emplean los estadísticos de asimetría (Skewness) y Kurtosis. La asimetría nos da un valor de 2 y Kurtosis de 3.32, por lo tanto volvemos a rechazar la hipótesis de que la variable dependiente sigue una distribución normal.

- **Test de hipótesis**

Se emplea el test Shapiro-Wilk y D'Agostino K-squared test para analizar la normalidad. En ambos casos se considera cómo hipótesis nula que los datos proceden de una distribución normal.

```
# Shapiro-Wilk test
# =====
shapiro_test = stats.shapiro(df_mex)
shapiro_test

ShapiroResult(statistic=0.667131781578064, pvalue=1.0104060947924154e-06)

# D'Agostino's K-squared test
# =====
k2, p_value = stats.normaltest(df_mex)
print(f"Estadístico = {k2}, p-value = {p_value}")

Estadístico = [22.81205398], p-value = [1.11282127e-05]
```

En ambos casos el valor de probabilidad (p) es muy superior al nivel elegido (0.05) por lo que rechazamos la hipótesis nula.

2.11.2.2. Estacionariedad y estacionalidad

Al tratarse la variable dependiente (Inversion_ES) de una serie de tiempo analizaremos si es estacionaria y si sigue una estacionalidad.

- **Estacionariedad**

Una serie es estacionaria cuando su media y su varianza es constante en el tiempo. Para ello se aplica la prueba *Dicky-Fuller*. Los resultados del test son los siguientes:

- Test Statics: 4.257
- Critical Value:
 - 10%: -2.65
 - 5%: -3.03
 - 1%: -3.83

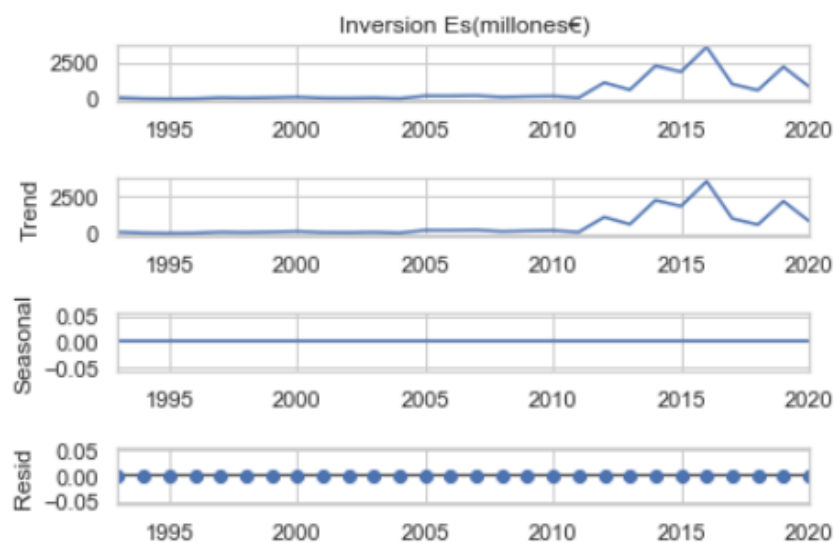
Test Statics es bastante alto y es muy superior a todos los valores críticos, por lo que no es una serie estacionaria.

- **Estacionalidad**

Comprobar si en nuestra serie hay ciertas tendencias que aparecen de forma cíclica.

Para comprobar si existe estacionalidad:

- Tendencia: la presencia de un patrón consistente
- Estacional: efectos cíclicos debido a la estacionalidad
- Residual: el error de predicción, la diferencia entre datos reales y los que estamos prediciendo.



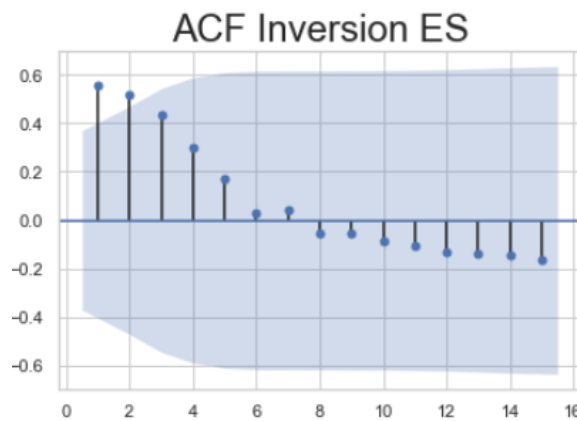
En la gráfica podemos ver que la tendencia (Trend) al principio es estacionaria, pero a partir de 2011 tiene una tendencia creciente, en 2018 decrece y luego vuelve a crecer ligeramente. La tendencia se aleja de la linealidad.

En la estacionalidad (Seasonal) es una línea recta, por lo que no se aprecia ningún ciclo en los datos.

Por lo tanto concluimos que no hay ni estacionalidad ni estacionariedad.

2.11.2.3. Autocorrelación

Utilizaremos la función de autocorrelación (ACF) y la función de autocorrelación parcial (PCAF) para determinar p y q . ACF describe la correlación entre la serie temporal con una versión rezagada de sí misma. PACF proporciona una correlación adicional explicada por cada término rezagado sucesivo.



Aplicando un retraso de 13 vemos que hay correlación positiva hasta el lag 7. Cuanto mayor sea el retraso, menor es la correlación. El área azul nos indica a mayor distancia en el tiempo más improbable es que esta correlación persista.

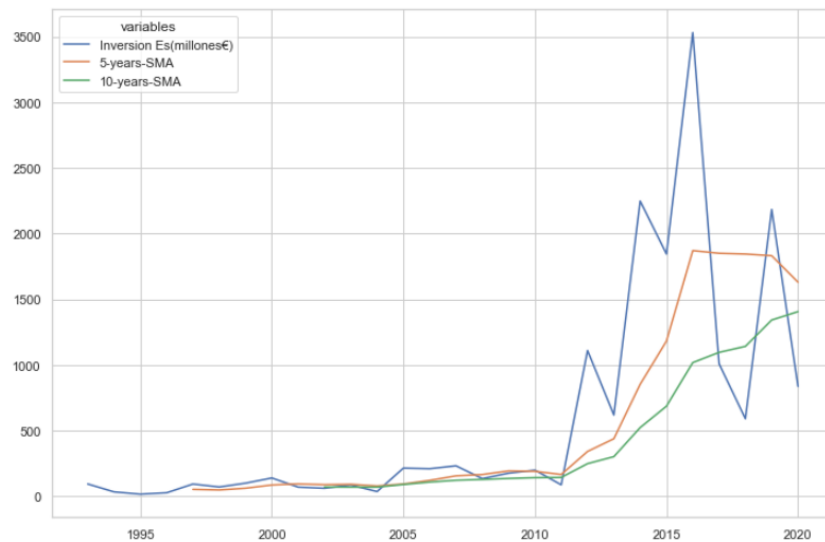
2.11.2.4. Media Móvil

Una media móvil se utiliza para analizar los datos de series de tiempo calculando una serie de medias de los diferentes subconjuntos del conjunto de datos completo. Las medias móviles también se utilizan para suavizar las fluctuaciones a corto plazo y poner de relieve las tendencias a largo plazo en los datos.

Hay muchos tipos diferentes de promedios móviles, utilizaremos promedios móviles simples (SMA) y los promedios móviles exponenciales (EWMA) ya que son los más populares entre analistas financieros.

- **Media móvil simple (SMA)**

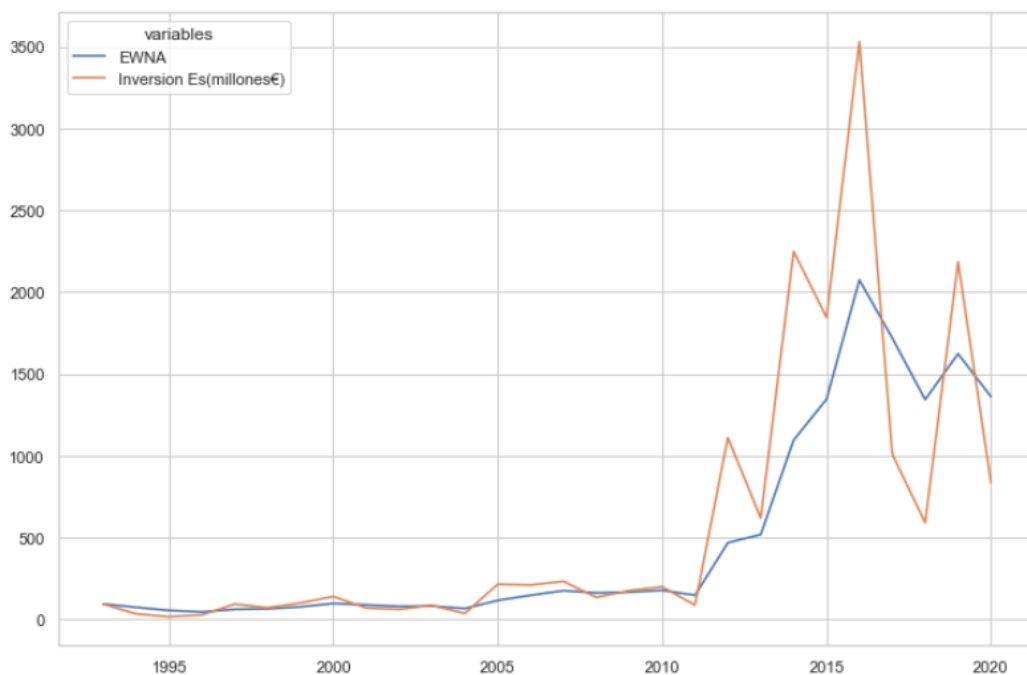
Es uno de los indicadores técnicos básicos que utilizan los traders o inversores para el análisis técnico de un índice o valores. Calculamos los promedios móviles de 5 y 10 años.



Aplicando el método SMA vemos que cuanto mayor sea el promedio que apliquemos más suavizadas están las fluctuaciones, pero no llegamos a conseguir la estacionariedad que buscamos en las series temporales.

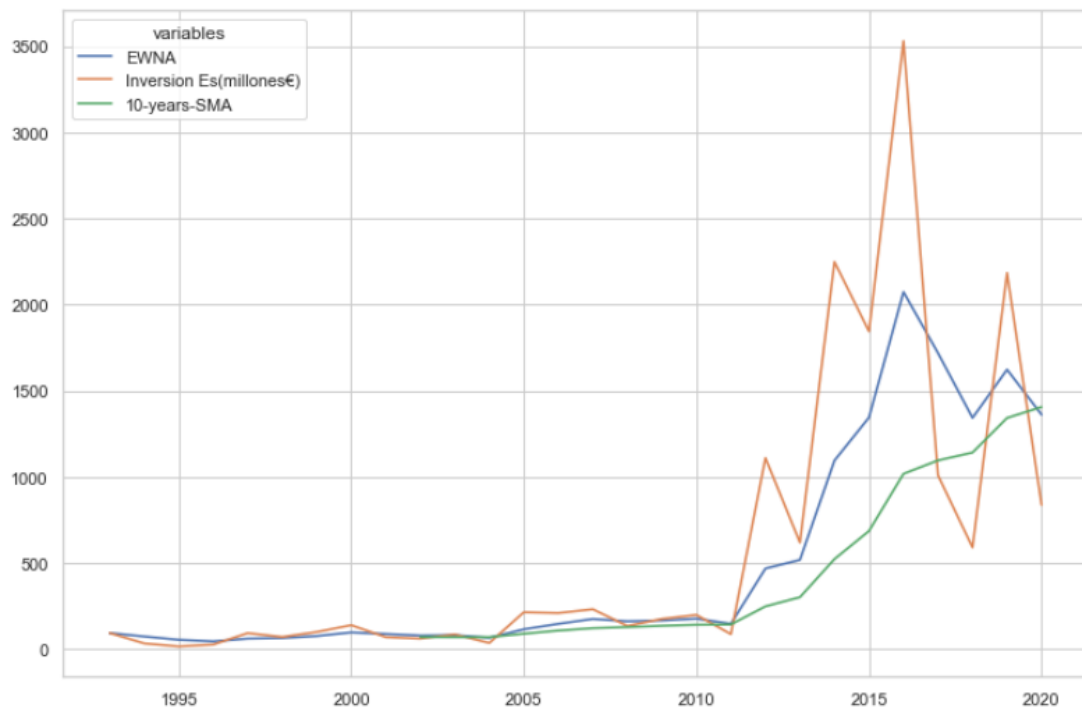
- **Media móvil exponencial (EWMA)**

Nos permitirá reducir el efecto de retraso de SMA y pondrá más peso en los valores que ocurrieron más recientemente.



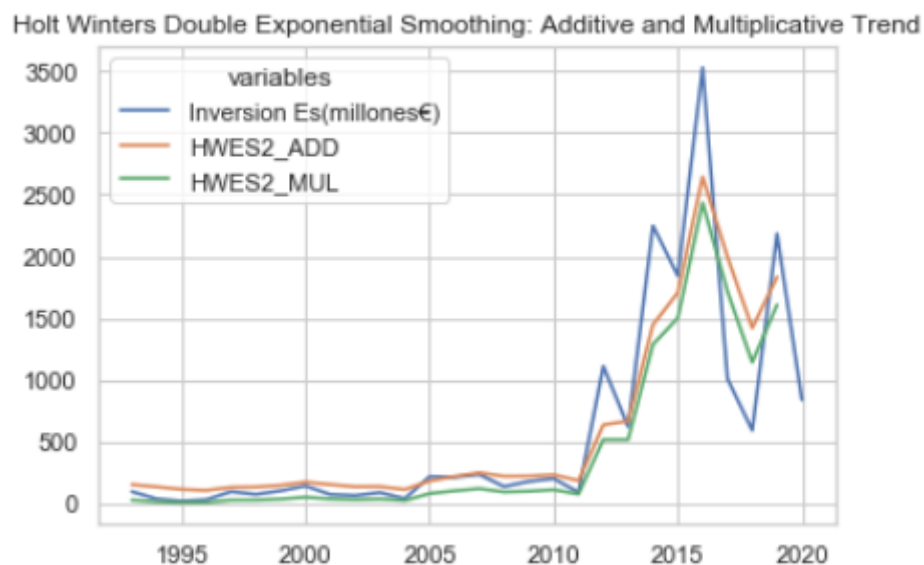
- **SMA vs EWMA**

SMA es mucho más suave. EWMA recoge mucho mejor la información que SMA. Las estimaciones de los periodos anteriores tienen más fuerza que los periodos antiguos.



- **Suavizado exponencial (Holt Winters)**

Con este método se le asigna menos peso a las observaciones pasadas usando un método exponencial. Utilizaremos el aditivo(ADD) y el multiplicativo (MUL), ambos son tripe suavizado exponencial.



El modelo Holt Winters se desvía un poco en los valles y bastante en los picos.

Los 3 modelos de suavizado son muy sensibles a los picos, y por lo general, no servirían para pronosticar un modelo prolongado.

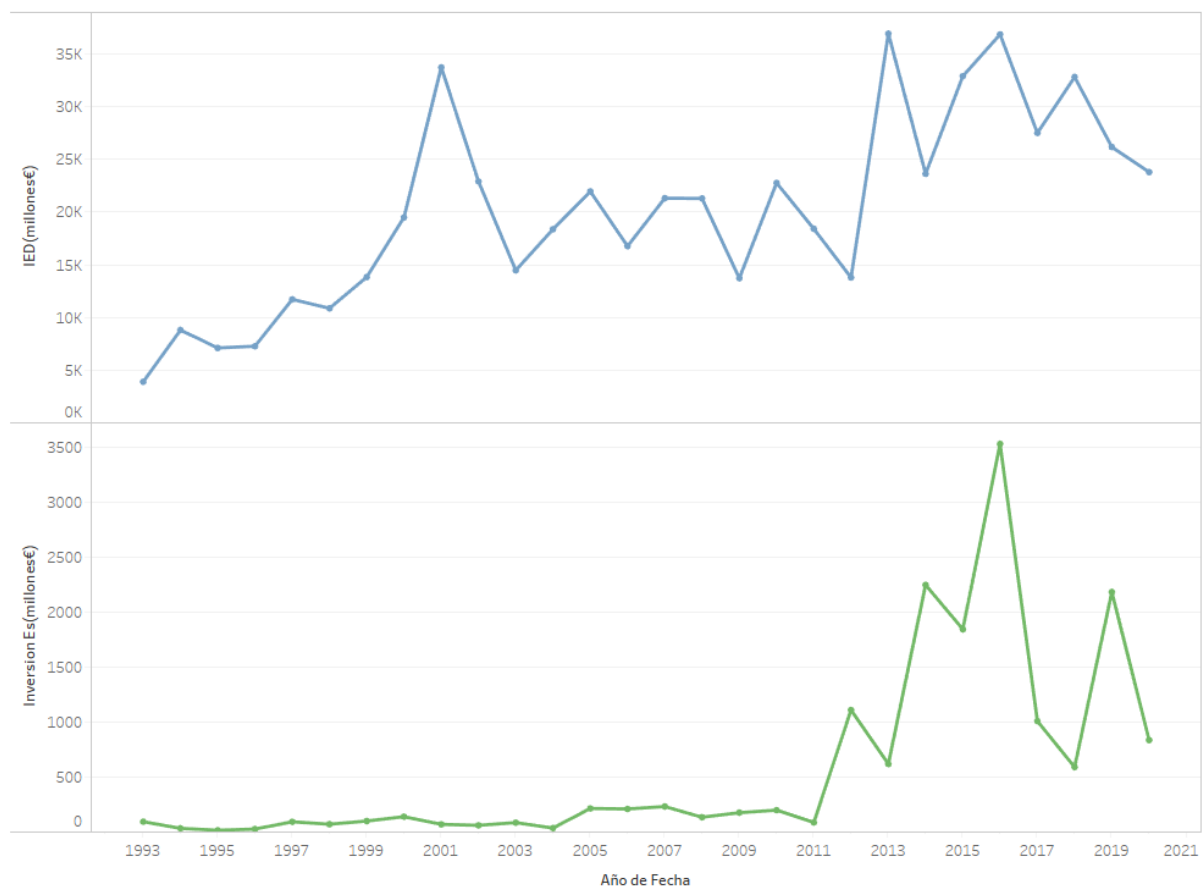
3. Análisis de la inversión española y las variables macroeconómicas en conjunto

3.1. Análisis de la inversión española en México en función de la Inversión Extranjera Directa de México (IED)

El objetivo de este apartado es comparar la inversión española en México con la IED.

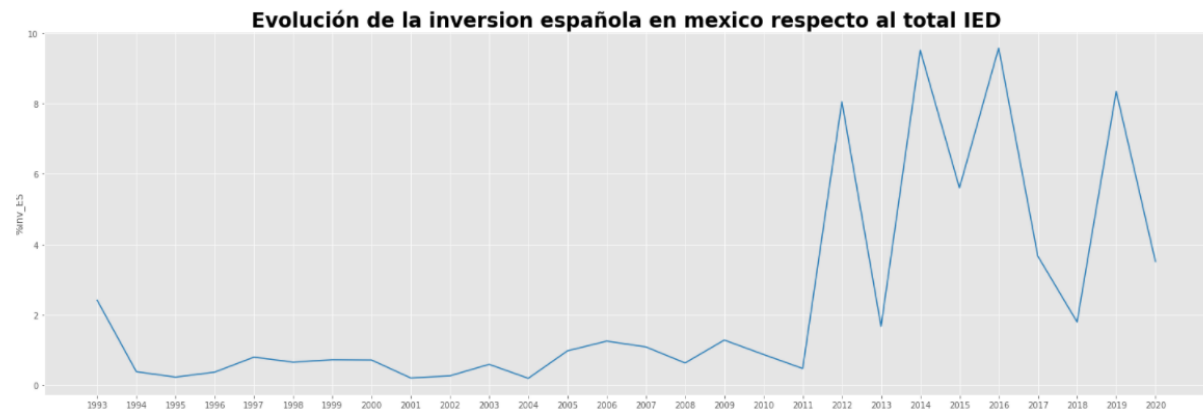
En el siguiente gráfico vemos la evolución de ambas variables.

Evolución de la IED y la inversión española en México

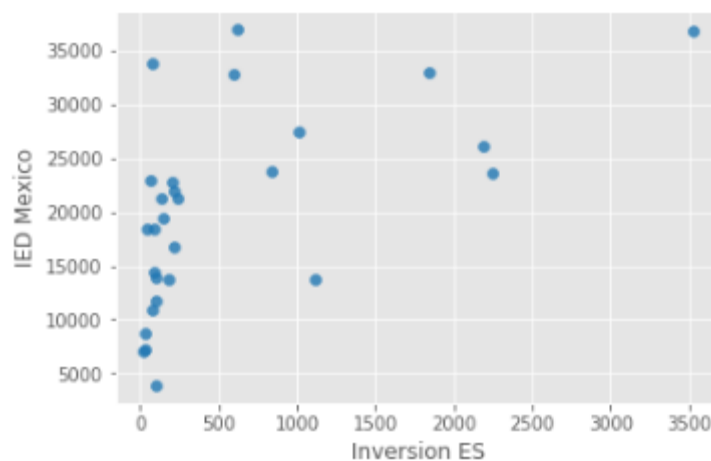


Mientras que la evolución de la IED tiene una tendencia creciente durante el periodo analizado, la inversión española hasta 2012 es prácticamente constante y el valor de las inversiones era mínimo. En la IED se aprecia el pico de 2001 como consecuencia del tratado de libre comercio con EEUU y Canadá, también se aprecia la caída en 2002 provocada por la crisis económica de EEUU provocada por el atentado de las torres gemelas. En 2008 también se ve otra caída de la IED por la crisis financiera mundial y en 2013 un incremento, siendo el año que más IED recibió México. La inversión de las empresas españolas fluctúa más porque se debe a operaciones concretas que realizan las empresas españolas. Es por ello que no siguen la misma tendencia.

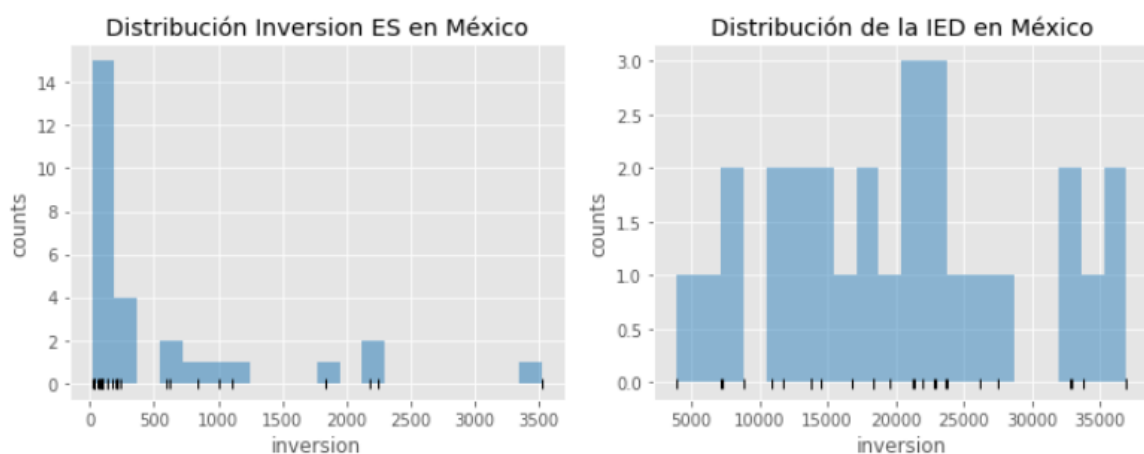
Durante los primeros años la inversión española no suponía más de 1.5% del IED recibido en México. Fue en 2012 cuando la inversión española supuso un 8.04% de la IED y en 2014 y 2016 un 9.5% del total de la IED de México. En 2018 apenas representó un 1.80% pero en 2019 volvió a subir a 8.34%. Siendo España y Estados Unidos los países que más invierten en México.



Con el diagrama de dispersión podemos apreciar que no existe una regresión lineal de datos.



En el siguiente gráfico vemos la distribución de ambas variables. Ninguna de las dos sigue una distribución normal, la que más se podría ajustar sería la IED.



Es importante analizar también la correlación entre variables. La correlación lineal es un método estadístico que permite cuantificar la relación lineal existente entre dos variables. Utilizaremos los estadísticos más empleados: Pearson, Spearman y Kendall. En nivel de correlación entre nuestras variables es el siguiente:

- Correlación Pearson: 0.56
- Correlación Spearman: 0.639
- Correlación Kendall: 0.497

Los 3 estadísticos marcan que la asociación entre la variable IED y la inversión española es moderada. con claras evidencias estadísticas de que la relación observada no se debe al azar ($p\text{-val} \approx 0$).

	n	r	CI95%	p-val	BF10	power
pearson	28	0.560493	[0.24, 0.77]	0.001921	22.76	0.895259

	n	r	CI95%	p-val	power
spearman	28	0.639299	[0.35, 0.82]	0.00025	0.96991

	n	r	CI95%	p-val	power
kendall	28	0.497354	[0.15, 0.73]	0.000122	0.791121

3.2. Modelo de Regresión Lineal (OLS)

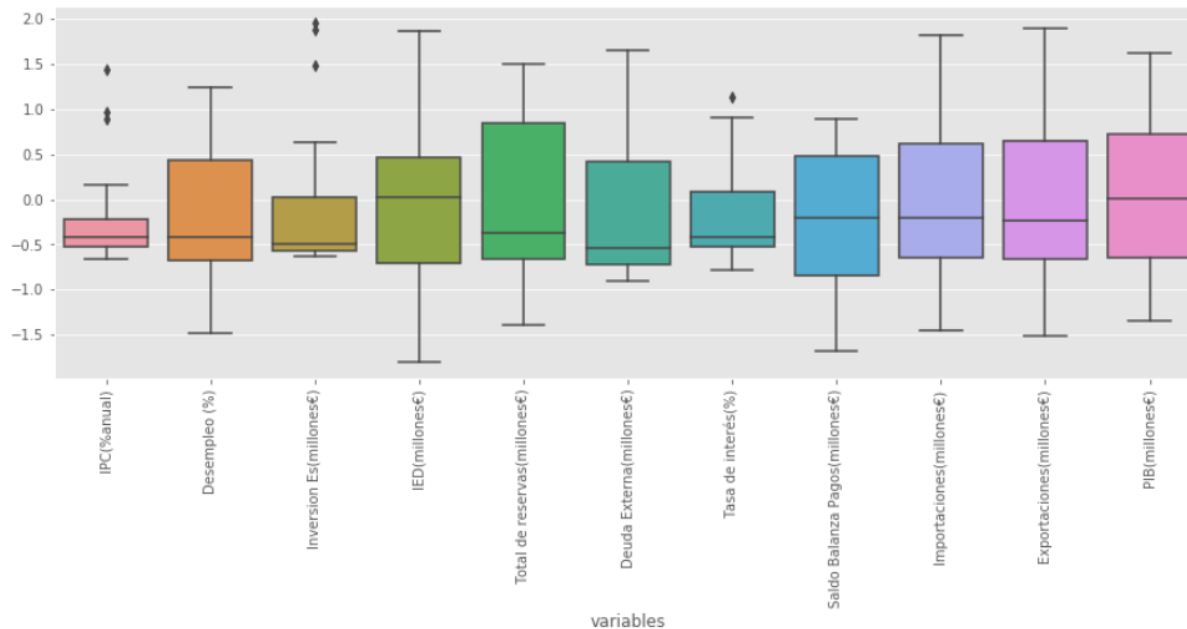
3.2.1. Análisis multivariante

En este apartado se analizan la variable dependiente (inversión española) y las variables independientes:

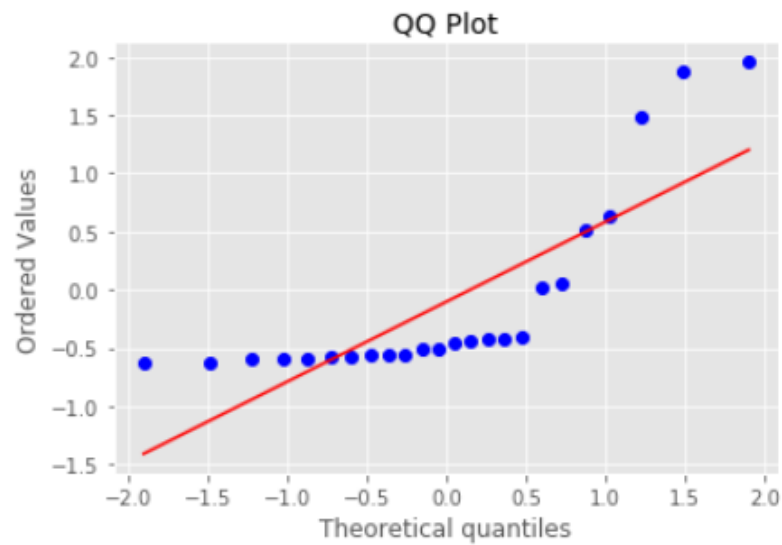
- PIB
- IPC
- Balanza de pagos
- Tasa de desempleo
- Exportaciones
- Importaciones
- Inversión Extranjera Directa (IED)
- IPC
- PIB
- Deuda pública
- Reservas
- Tasa de interés

Para analizar las variables las estandarizamos (standar scaler) para tener variables con el mismo rango y eliminamos los valores atípicos.

Para ver la distribución de los datos utilizaremos el gráfico boxplot. Las únicas variables que más se ajustan a una normal son la balanza de pagos y el PIB.



Con el gráfico QQ-Plot también podemos ver que el conjunto de datos no sigue una distribución normal ya que no se ajustan los puntos a la línea.



- **Matriz de correlación**

variables	IPC(%anual)	1	-0.28	-0.34	-0.49	-0.55	-0.34	0.89	-0.011	-0.53	-0.52	-0.6
	Desempleo (%)	-0.28	1	0.2	-0.059	0.33	0.12	-0.44	0.13	0.15	0.17	0.12
	Inversion Es(millones€)	-0.34	0.2	1	0.42	0.82	0.86	-0.44	-0.26	0.77	0.78	0.76
	IED(millones€)	-0.49	-0.059	0.42	1	0.67	0.67	-0.56	-0.31	0.74	0.73	0.79
	Total de reservas(millones€)	-0.55	0.33	0.82	0.67	1	0.94	-0.7	-0.21	0.96	0.97	0.93
	Deuda Externa(millones€)	-0.34	0.12	0.86	0.67	0.94	1	-0.46	-0.39	0.93	0.93	0.88
	Tasa de interés(%)	0.89	-0.44	-0.44	-0.56	-0.7	-0.46	1	-0.095	-0.65	-0.65	-0.7
	Saldo Balanza Pagos(millones€)	-0.011	0.13	-0.26	-0.31	-0.21	-0.39	-0.095	1	-0.19	-0.16	-0.23
	Importaciones(millones€)	-0.53	0.15	0.77	0.74	0.96	0.93	-0.65	-0.19	1	1	0.97
	Exportaciones(millones€)	-0.52	0.17	0.78	0.73	0.97	0.93	-0.65	-0.16	1	1	0.96
	PIB(millones€)	-0.6	0.12	0.76	0.79	0.93	0.88	-0.7	-0.23	0.97	0.96	1
	variables	IPC(%anual)	Desempleo (%)	Inversion Es(millones€)	IED(millones€)	Total de reservas(millones€)	Deuda Externa(millones€)	Tasa de interés(%)	Saldo Balanza Pagos(millones€)	Importaciones(millones€)	Exportaciones(millones€)	PIB(millones€)

Las variables 'total de reservas', 'deuda externa', 'importaciones', 'exportaciones' y 'PIB' están altamente correlacionadas positivamente con la inversión española, ya que su índice de correlación es mayor a 0.7.

En cambio las variables ‘saldo balanza pagos’, ‘IPC’ y ‘Tasa de interés’ tienen una correlación negativa con la inversión española.

variables	Inversion Es(millones€)
variables	
Inversion Es(millones€)	1.000000
Deuda Externa(millones€)	0.857031
Total de reservas(millones€)	0.824336
Exportaciones(millones€)	0.775435
Importaciones(millones€)	0.770454
PIB(millones€)	0.757251
IED(millones€)	0.423066
Desempleo (%)	0.201579
Saldo Balanza Pagos(millones€)	-0.262383
IPC(%anual)	-0.342866
Tasa de interés(%)	-0.440874

3.2.2. Regresión lineal múltiple (OLS)

Con el método de regresión lineal trataremos de modelar la relación entre la variable ‘Inversion ES’ y el resto de variables.

El objetivo es predecir el valor de la inversión en función del resto de variables.

Realizamos el modelo con todas las variables independientes de X. Obtenemos el siguiente resumen del modelo:

OLS Regression Results						
Dep. Variable:	Inversion Es(millones€)	R-squared:	0.938			
Model:	OLS	Adj. R-squared:	0.849			
Method:	Least Squares	F-statistic:	10.54			
Date:	Sat, 17 Jul 2021	Prob (F-statistic):	0.00249			
Time:	19:43:32	Log-Likelihood:	2.1255			
No. Observations:	18	AIC:	17.75			
Df Residuals:	7	BIC:	27.54			
Df Model:	10					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-0.1563	0.199	-0.787	0.457	-0.626	0.313
IPC(%annual)	0.1438	0.823	0.175	0.866	-1.802	2.090
Desempleo (%)	-0.0738	0.326	-0.226	0.827	-0.845	0.697
IED(millones€)	-0.5036	0.357	-1.411	0.201	-1.347	0.340
Total de reservas(millones€)	0.2152	1.871	0.115	0.912	-4.209	4.639
Deuda Externa(millones€)	1.4166	1.400	1.012	0.345	-1.893	4.726
Tasa de interés(%)	-0.4825	1.176	-0.410	0.694	-3.263	2.298
Saldo Balanza Pagos(millones€)	0.4303	0.436	0.988	0.356	-0.600	1.460
Importaciones(millones€)	1.3899	6.168	0.225	0.828	-13.194	15.974
Exportaciones(millones€)	-3.3780	6.541	-0.516	0.621	-18.845	12.089
PIB(millones€)	1.8070	0.838	2.155	0.068	-0.175	3.789
Omnibus:	1.105	Durbin-Watson:	2.750			
Prob(Omnibus):	0.576	Jarque-Bera (JB):	0.798			
Skew:	0.133	Prob(JB):	0.671			
Kurtosis:	2.003	Cond. No.	243.			

El modelo con todas las variables introducidas cómo predictores tiene un R2 alto (0.938), es decir, es capaz de explicar el 93.8% de la variabilidad observada en la inversión. El *p-value* (0.00249) no es significativo por lo que se puede aceptar que la relación observada no se debe al azar (*p-value* \approx 0).

Podemos observar que el coeficiente de determinación R2 es alto, pero ninguno de los predictores resulta significativo, lo que nos indica que hay un problema de multicolinealidad. Esto ocurre cuando un predictor está linealmente relacionado con uno o varios.

Para saber si existe colinealidad entre los predictores medimos el Factor de Inflación de la Varianza (VIF).

	VIF
IPC(%annual)	8.118633
Desempleo (%)	3.800923
IED(millones€)	3.849045
Total de reservas(millones€)	127.218260
Deuda Externa(millones€)	55.552659
Tasa de interés(%)	19.903107
Saldo Balanza Pagos(millones€)	7.024821
Importaciones(millones€)	1487.712844
Exportaciones(millones€)	1926.146284
PIB(millones€)	36.308710

VIF= 1 nos indica ausencia total de la colinealidad, por lo tanto nos indica que tenemos un problema de multicolinealidad. Para solucionarlo utilizamos dos métodos:

- Ir eliminando las variables que mayor VIF tengan hasta quedarnos con aquellas que tengan un VIF menor que 5.
- Análisis de Componentes Principales (PCA)

● VIF

Los predictores después de aplicar este método serían:

- IPC
- Desempleo
- IED
- Deuda Externa
- Saldo de la balanza de pagos

El resumen de aplicar el modelo de regresión lineal con estas variables sería:

OLS Regression Results						
Dep. Variable:	Inversion Es(millones€)	R-squared:	0.831			
Model:	OLS	Adj. R-squared:	0.761			
Method:	Least Squares	F-statistic:	11.82			
Date:	Sat, 17 Jul 2021	Prob (F-statistic):	0.000268			
Time:	19:43:42	Log-Likelihood:	-6.8480			
No. Observations:	18	AIC:	25.70			
Df Residuals:	12	BIC:	31.04			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-0.0463	0.191	-0.242	0.813	-0.463	0.370
IPC(%anual)	-0.4225	0.412	-1.025	0.326	-1.321	0.476
Desempleo (%)	0.0625	0.170	0.368	0.719	-0.308	0.433
IED(millones€)	-0.3171	0.191	-1.658	0.123	-0.734	0.100
Deuda Externa(millones€)	0.9429	0.164	5.759	0.000	0.586	1.300
Saldo Balanza Pagos(millones€)	-0.0135	0.170	-0.080	0.938	-0.383	0.356
Omnibus:	6.042	Durbin-Watson:	2.466			
Prob(Omnibus):	0.049	Jarque-Bera (JB):	3.480			
Skew:	-0.720	Prob(JB):	0.176			
Kurtosis:	4.602	Cond. No.	5.79			

$$\text{Inversion ES} = -0.0463 - 0.4225\text{IPC} + 0.0625\text{Desempleo} - 0.3171\text{IED} + 0.9429\text{Deuda Externa} - 0.0135\text{Saldo Balanza Pagos}$$

Es capaz de explicar el 83.1% de la varianza observada en la inversión española . El test F es ($p\text{-value} = 0.00000268$), por lo que se puede aceptar que la relación observada no se debe al azar ($p\text{-value} \approx 0$).

No se satisfacen las condiciones de normalidad, por lo que los intervalos de confianza estimados para los coeficientes y las predicciones no son fiables.

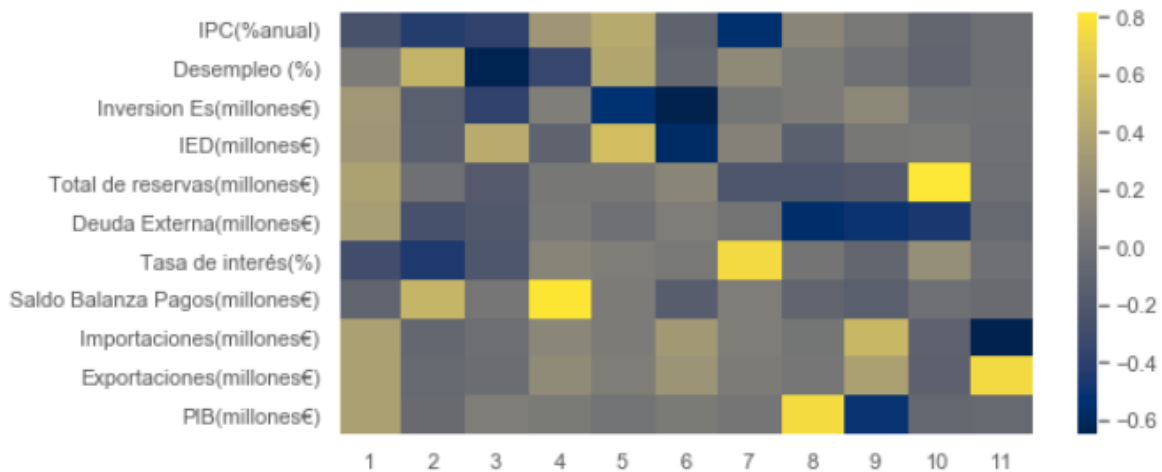
El error (rmse) de test es de 0.44. Las predicciones del modelo final se alejan en promedio 0.44 unidades del valor real.

- **PCA**

Aplicamos el método de reducción de dimensionalidad para simplificar la complejidad y poder solucionar el problema de multicolinealidad.

Una vez entrenado el objeto podemos ver toda la información de las componentes creadas.

La influencia de las variables en cada componente se pueden analizar con un *heatmap*.



Una vez calculadas las componentes principales, se puede conocer la varianza explicada por cada una de ellas, la proporción respecto al total y la proporción de varianza acumulada.

 Porcentaje de varianza explicada por cada componente

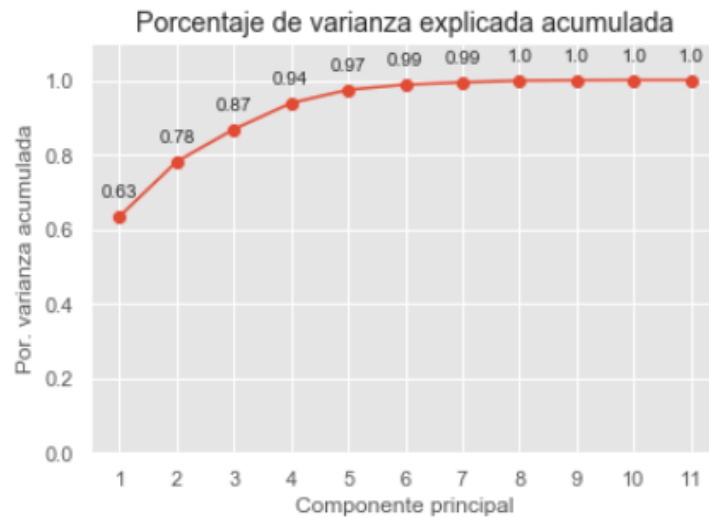
```
[6.34106897e-01 1.47033588e-01 8.68423062e-02 7.02552488e-02
3.56379665e-02 1.41645732e-02 5.87817631e-03 4.56148205e-03
9.88976581e-04 5.04665464e-04 2.61191774e-05]
```



En este caso, la primera componente explica el 63% de la varianza observada en los datos y la segunda el 15%. El resto de componentes no superan el 1% de la varianza explicada.

 Porcentaje de varianza explicada acumulada

[0.6341069 0.78114049 0.86798279 0.93823804 0.97387601 0.98804058
 0.99391876 0.99848024 0.99946922 0.99997388 1.]

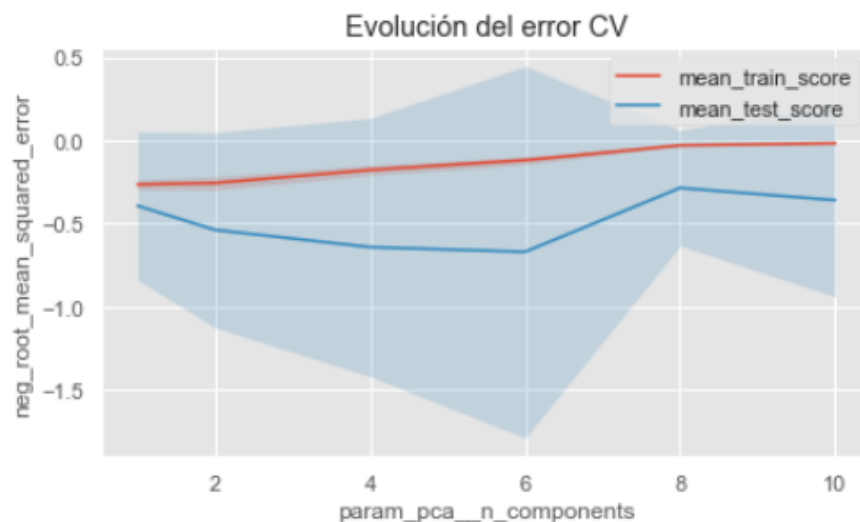


Si se emplean únicamente las 3 primeras componentes se conseguiría explicar el 87% de la varianza observada.

- **Principal Component Regression (PCR)**

El PCR combina el PCA con regresión lineal, por lo que este método será el óptimo para la reducción de dimensionalidad.

Primero evaluamos el modelo incluyendo todos los componentes.



```
-----  
Mejores hiperparámetros encontrados (cv)  
-----
```

```
{'pca__n_components': 8} : -0.2852991250730813 neg_mean_squared_error
```

El resultado de la validación cruzada muestra que, el mejor modelo se obtiene empleando las 8 primeras componentes.

Reentrenamos el modelo indicando esta configuración. Empleando las cinco primeras componentes del PCA como predictores en lugar de las variables originales, se consigue reducir el root mean squared error de 1.78 a 0.76.

4. Conclusión

Tras aplicar los diferentes métodos para solucionar la multicolinealidad, nos quedamos con el modelo que nos da el VIF ya que tiene un error menor (0.44) que el de PCA y PCR (0.76).

El modelo de regresión lineal múltiple es:

$$\text{Inversion ES} = -0.0463 - 0.4225\text{IPC} + 0.0625\text{Desempleo} - 0.3171\text{IED} + 0.9429\text{Deuda Externa} - 0.0135\text{Saldo Balanza Pagos}$$

Es capaz de explicar el 83.1% de la varianza observada en la inversión española . El test F es ($p\text{-value} = 0.00000268$), por lo que se puede aceptar que la relación observada no se debe al azar ($p\text{-value} \approx 0$).

Dado que no tenemos el valor de las variables utilizadas en el modelo para poder predecir la inversión española ya que se trata de datos macroeconómicos y que desconocemos su valor, no podemos realizar predicciones.

De todas formas, esta información nos puede ayudar a tomar decisiones sabiendo cuales son las variables que afectan a la inversión.

Listado de sectores

- 01 AGRICULTURA
GANADERÍA CAZA Y
SERV. RELACIONADOS
- 02 SILVICULTURA Y
EXPLOTACIÓN FORESTAL
- 03 PESCA Y
ACUICULTURA
- 05 EXTRACCIÓN DE
ANTRACITA HULLA Y
LIGNITO
- 06 EXTRACCIÓN DE
CRUDO DE PETRÓLEO Y
GAS NATURAL
- 07 EXTRACCIÓN DE
MINERALES METÁLICOS
- 08 OTRAS INDUSTRIAS
EXTRACTIVAS
- 09 ACTIVIDADES DE
APOYO A LAS
INDUSTRIAS
EXTRACTIVAS
- 10 INDUSTRIA DE LA
ALIMENTACIÓN
- 11 FABRICACIÓN DE
BEBIDAS
- 12 INDUSTRIA DEL
TABACO
- 13 INDUSTRIA TEXTIL
- 14 CONFECCIÓN DE
PRENDAS DE VESTIR
- 15 INDUSTRIA DEL
CUERO Y DEL CALZADO
- 16 INDUSTRIA MADERA
Y CORCHO EXCEPTO
MUEBLES CESTERÍA
- 17 INDUSTRIA DEL PAPEL
- 18 ARTES GRÁFICAS Y
REPRODUCCIÓN DE
SOPORTES GRABADOS
- 19 COQUERÍAS Y REFINO
DE PETRÓLEO
- 20 INDUSTRIA QUÍMICA
- 21 FABRICACIÓN DE
PRODUCTOS
FARMACÉUTICOS
- 22 FABRICACIÓN DE
PRODUCTOS DE CAUCHO
Y PLÁSTICOS
- 23 FABRICACIÓN DE
OTROS PRODUCTOS
MINERALES NO
METÁLICO
- 24 METALURGIA;
FABRICACIÓN
PRODUCTOS HIERRO
ACERO
- 25 FABRICACIÓN DE
PRODUCTOS METÁLICOS
EXCEPTO MAQUINAR
- 26 FABRICACIÓN DE
PRODUCTOS
INFORMÁTICOS
ELECTRÓNICOS
- 27 FABRICACIÓN DE
MATERIAL Y EQUIPO
ELÉCTRICO
- 28 FABRICACIÓN DE
MAQUINARIA Y EQUIPO
N.C.O.P.
- 29 FABRICACIÓN DE
VEHÍCULOS DE MOTOR
REMOLQUES
- 30 FABRICACIÓN DE
OTRO MATERIAL DE
TRANSPORTE
- 31 FABRICACIÓN DE
MUEBLES
- 32 OTRAS INDUSTRIAS
MANUFACTURERA
- 33 REPARACIÓN E
INSTALACIÓN DE
MAQUINARIA Y EQUIPO
- 35 SUMINISTRO DE
ENERGÍA ELÉCTRICA GAS
VAPOR Y AIRE
- 36 CAPTACIÓN
DEPURACIÓN Y
DISTRIBUCIÓN DE AGUA
- 37 RECOGIDA Y
TRATAMIENTO DE
AGUAS RESIDUALES
- 38 RECOGIDA
TRATAMIENTO Y
ELIMINACIÓN DE
RESIDUOS
- 39 ACTIVID. DE
DESCONTAMINACIÓN Y
OTROS SERVICIOS
- 41 CONSTRUCCIÓN DE
EDIFICIOS
- 42 INGENIERÍA CIVIL
- 43 ACTIVIDADES DE
CONSTRUCCIÓN
ESPECIALIZADA
- 45 VENTA Y REPARACIÓN
DE VEHÍCULOS DE
MOTOR Y MOTOCICL.
- 46 COMER.MAYOR E
INTERME.COMERCIOEXC
EP.VEHÍCULOS MOTOR
- 47 COMERCIO AL POR
MENOR EXCEPTO DE
VEHÍCULOS DE MOTOR
- 49 TRANSPORTE
TERRESTRE Y POR
TUBERÍA
- 50 TRANSPORTE
MARÍTIMO Y POR VÍAS
NAVEGABLES
INTERIORES
- 51 TRANSPORTE AÉREO
- 52 ALMACENAMIENTO Y
ACTIVIDADES ANEXAS AL
TRANSPORTE
- 53 ACTIVIDADES
POSTALES Y DE CORREOS
- 55 SERVICIOS DE
ALOJAMIENTO

- 56 SERVICIOS DE COMIDAS Y BEBIDAS
- 58 EDICIÓN
- 59 ACTIV. CINEMATOGRAF DE VIDEO Y PROGRA. DE TV GRAB.
- 60 ACTIVIDADES DE PROGRAMACIÓN Y EMISIÓN DE RADIO Y TEL
- 61 TELECOMUNICACIONES
- 62 PROGRAMACIÓNCONSULTORÍAOTRAS ACTIVIDADES RELAC.
- 63 SERVICIOS DE INFORMACIÓN
- 64 SERVICIOS FINANCIEROS EXCEP. SEGUROS Y FONDOS PENSION
- 65 SEGUROS REASEGURO. FONDOS PENSION EXCEPTO S. SOCIAL
- 66 ACTIVIDADES AUXILIARES A LOS SERVICIOS FINANCIEROS
- 68 ACTIVIDADES INMOBILIARIAS
- 69 ACTIVIDADES JURÍDICAS Y DE CONTABILIDAD
- 70 ACTIV. SEDES CENTRALES; ACTIV. CONSULTORÍA DE GESTIÓN
- 71 SERVICIOS TECNICOS ARQUITECTURA E INGENIERIA
- 72 INVESTIGACIÓN Y DESARROLLO
- 73 PUBLICIDAD Y ESTUDIOS DE MERCADO
- 74 OTRAS ACTIV. PROFESIONALES CIENTÍFICAS Y TÉCNICAS
- 75 ACTIVIDADES VETERINARIAS
- 77 ACTIVIDADES DE ALQUILER
- 78 ACTIVIDADES RELACIONADAS CON EL EMPLEO
- 79 ACTIV. AGENCIAS VIAJE OPERADOR TURÍST. RESERVAS
- 80 ACTIVIDADES DE SEGURIDAD E INVESTIGACIÓN
- 81 SERVICIOS A EDIFICIOS Y ACTIVIDADES DE JARDINERÍA
- 82 ACTIVIDADES ADMINISTRATIVAS DE OFICINA
- 84 ADMINISTRACIÓN PÚBLICA Y DEFENSA; SEGURIDAD SOCIAL
- 85 EDUCACIÓN
- 86 ACTIVIDADES SANITARIAS
- 87 ASISTENCIA EN ESTABLECIMIENTOS RESIDENCIALES
- 88 ACTIVIDADES DE SERVICIOS SOCIALES SIN ALOJAMIENTO
- 90 ACTIVIDADES DE CREACIÓN ARTÍSTICAS Y ESPECTÁCULOS
- 91 ACTIV. DE BIBLIOTECAS ARCHIVOS MUSEOS Y OTRAS
- 92 ACTIVIDADES DE JUEGOS DE AZAR Y APUESTAS
- 93 ACTIVIDADES DEPORTIVAS RECREATIVAS Y DE ENTRETENIM.
- 94 ACTIVIDADES ASOCIATIVAS
- 95 REPARACIÓN ORDENADORES EFECTOS PERSONALES
- 96 OTROS SERVICIOS PERSONALES
- 97 ACTIVIDADES DE LOS HOGARES COMO EMPLEADORES
- 98 ACTIV. HOGARES COMO PRODUC. BIENES Y SERV
- 99 ACTIV. ORGANIZACIONES Y ORGANISMOS EXTRATERRITORIA