

Proyección de reservas nacionales utilizando ARIMA y Regresión múltiple en Python

Objetivo y método utilizado

En este trabajo buscamos estimar las reservas brutas de República Dominicana en los próximos 2 años utilizando 7 variables independientes. Estas variables son variables dadas por el profesor que tienen incidencia en las reservas internacionales del país. Usaremos un modelo ARIMA para estimar el comportamiento de estas variables en el futuro y luego usaremos regresión múltiple lineal para estimar las reservas utilizando estas variables.

Estas variables son: El precio del barril de petróleo, las importaciones y exportaciones del país, la cantidad de moneda local circulante en el país, el precio del dólar con relación al peso, la inflación, la tasa de referencia de la tesorería de los EE. UU. y la deuda pública nacional. Tomamos datos de cada una de estas variables cada mes desde enero del 2010 hasta septiembre de 2018. Algunos datos que son diarios los promediamos o tomamos el dato de final de mes, algunos datos que son trimestrales usamos regresión lineal para estimar los valores en cada mes. Al final obtuvimos una matriz con todas las variables por mes de 2010 a 2018, esto nos facilita proyectar sobre cada columna de la matriz.

En la siguiente sección explicaremos la base teórica de este método que usamos.

ARIMA

En el análisis de series temporales, el modelo autorregresivo integrado de promedio móvil o ARIMA (acrónimo del inglés autoregressive integrated moving average) es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. En este modelo las estimaciones futuras vienen explicadas por los valores pasados y no por variables independientes.

Una función ARIMA(p,d,q) cuenta con 3 componentes. El primero es el componente autorregresivo AR(p) se refiere al uso de valores pasados en la ecuación de la regresión para la serie temporal. La d representa el orden de diferenciación en el componente integrado I(d), esta diferenciación se usa para remover la tendencia de nuestra serie temporal y volverla estacionaria para así entender mejor los ciclos. El tercer componente es la media móvil MA(q) representa que en la ecuación sólo se usan los primeros q términos, todos los otros son nulos. El proceso de ajustar un modelo ARIMA también se conoce como el método Box-Jenkins.

Todo análisis de series temporales tiene 4 partes principales: la parte estacional, la parte tendencia, la parte ciclo y el componente aleatorio. El componente tendencia se puede definir como un cambio a largo plazo que se produce en la relación al nivel medio, se identifica con un movimiento suave de la serie a largo plazo. El componente estacional es la variación por períodos que tiene la serie temporal (semanal, mensual, trimestral, cuatrimestral, semestral, anual, bienal, trienal, cuadrinal...). Estos efectos se pueden medir explícitamente y se pueden eliminar para analizar la serie, este proceso se llama desestacionalización de la serie temporal.

La parte ciclo de una serie temporal corresponde a patrones decrecientes o crecientes que no son estacionales, usualmente la tendencia y el ciclo son agrupados juntos. El ciclo y la tendencia se pueden estudiar usando medias móviles. Por último, a la parte de la serie que no se puede explicar por estos últimos tres componentes se le llama error, residuo o parte aleatoria. El proceso de extraer estos componentes se llama descomposición.

Regresión Múltiple

La regresión múltiple se utiliza para estudiar la posible relación entre varias variables independientes o explicativas y otra variable dependiente o respuesta. La regresión múltiple se usa para la predicción de respuestas a partir de variables explicativas, aunque hay que tener claro que los modelos de regresión nos informan de la presencia de relaciones, pero no del mecanismo causal de la respuesta.

Criterio de Información de Akaike

“El criterio de información de Akaike (**AIC**, en inglés Akaike Information Criterion) es una medida de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC proporciona un medio para la selección del modelo.

AIC maneja un trade-off entre la bondad de ajuste del modelo y la complejidad del modelo. Se basa en la entropía de información: se ofrece una estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos.” (Wikipedia)

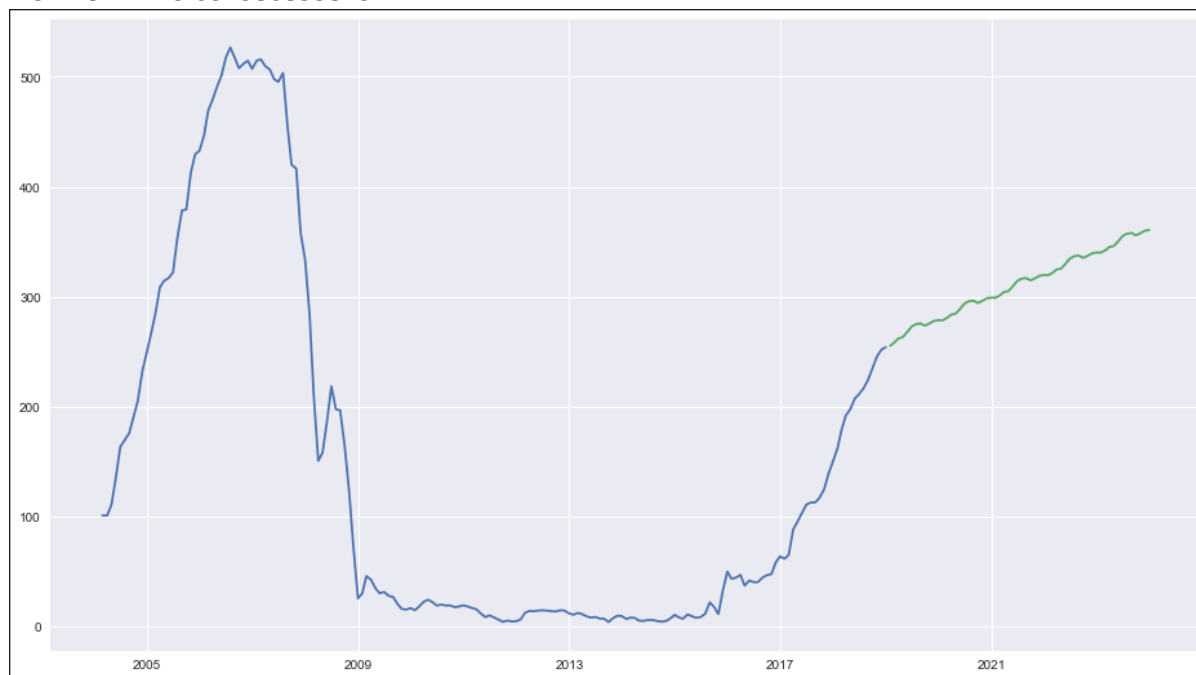
Hemos utilizado el AIC extensivamente para calcular nuestro mejor modelo ARIMA.

Proyecciones individuales con ARIMA: Información histórica en azul y proyección en verde,
estacionalidad expresada en meses (12=1 año, 24=2 años, 36=3 años)

Tasa de referencia a 6 meses del tesoro de EE. UU. (27-02-2004 hasta 31-12-2018)

Parámetros del modelo ARIMA utilizado: **(1, 1, 1), (0, 1, 1, 12)**

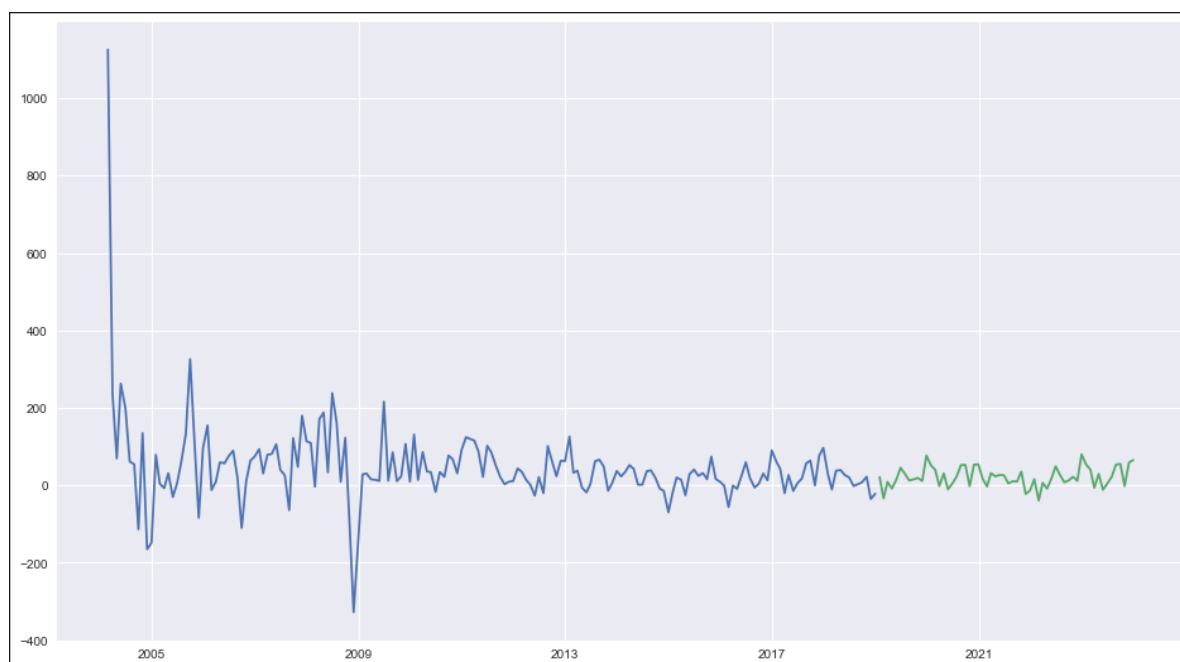
Y el **AIC = 1223.037580855845**



Inflación RD (27-02-2004 hasta 31-12-2018)

Parámetros del modelo ARIMA utilizado: **(1, 0, 1), (1, 1, 0, 36)**

Y el **AIC = 1166.7028147384585**



Precio del petróleo (2010-01-29 hasta 28-09-2018)

Parámetros del modelo ARIMA utilizado: **(1, 1, 0), (1, 1, 0, 24)**

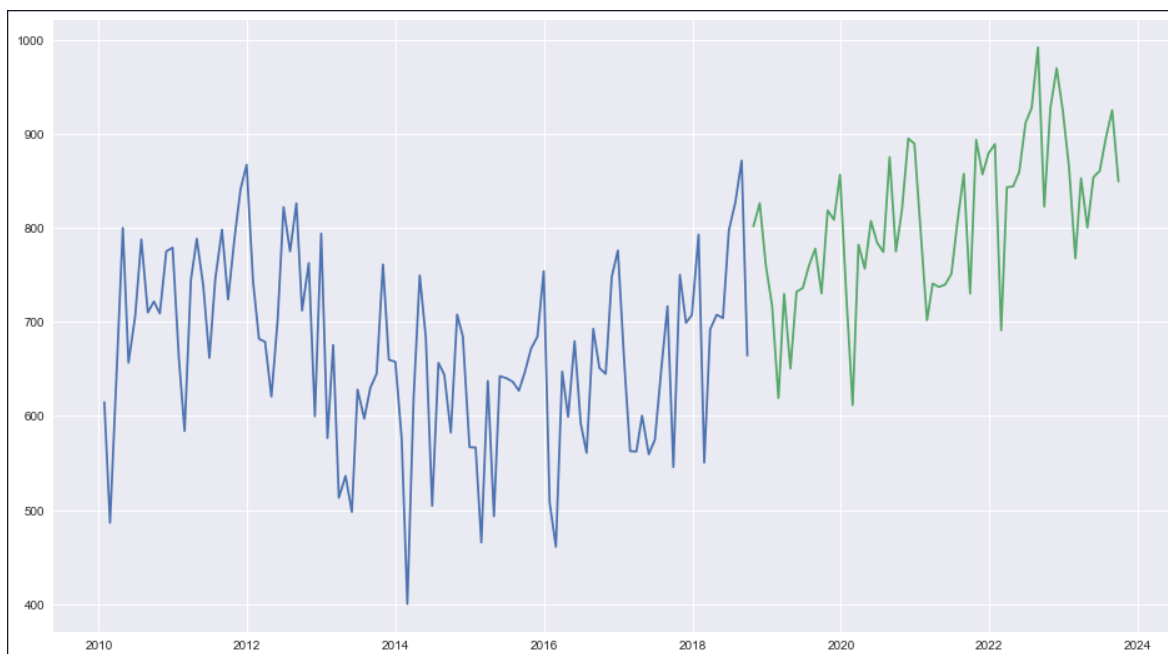
Y el AIC = **367.7885492152411**



Importaciones netas RD (2010-01-29 hasta 28-09-2018)

Parámetros del modelo ARIMA utilizado: **(1, 1, 1), (1, 1, 0, 24)**

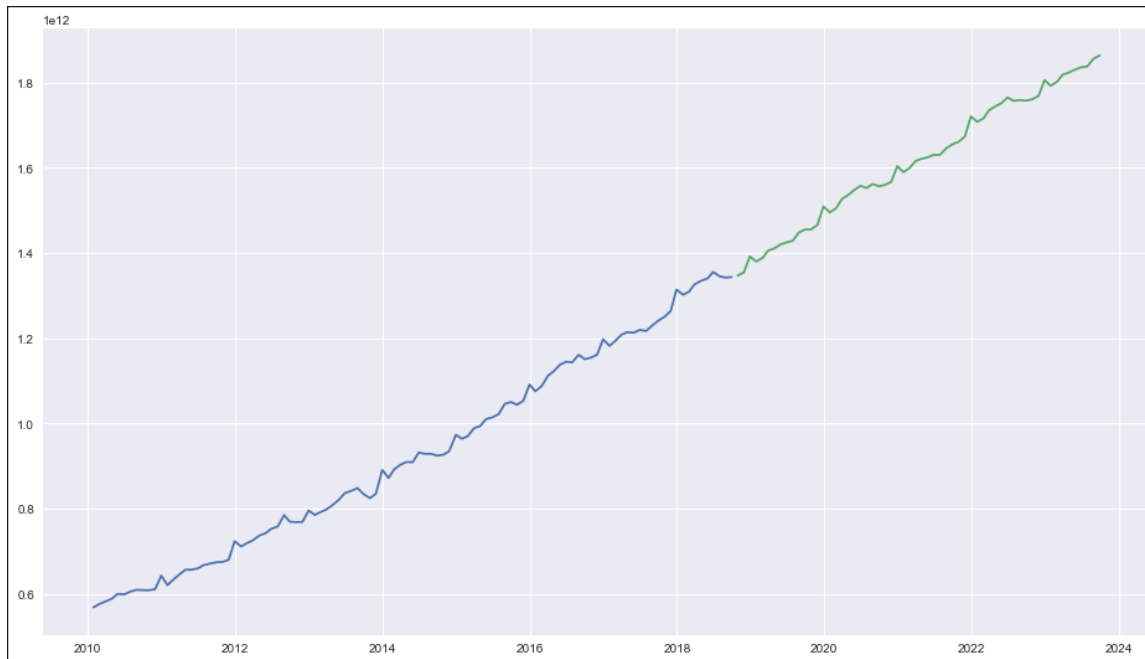
Y el AIC = **655.4795826129601**



Dinero en sentido amplio RD, dinero total circulando (2010-01-29 hasta 28-09-2018)

Parámetros del modelo ARIMA utilizado: **(1, 1, 0), (1, 1, 0, 24)**

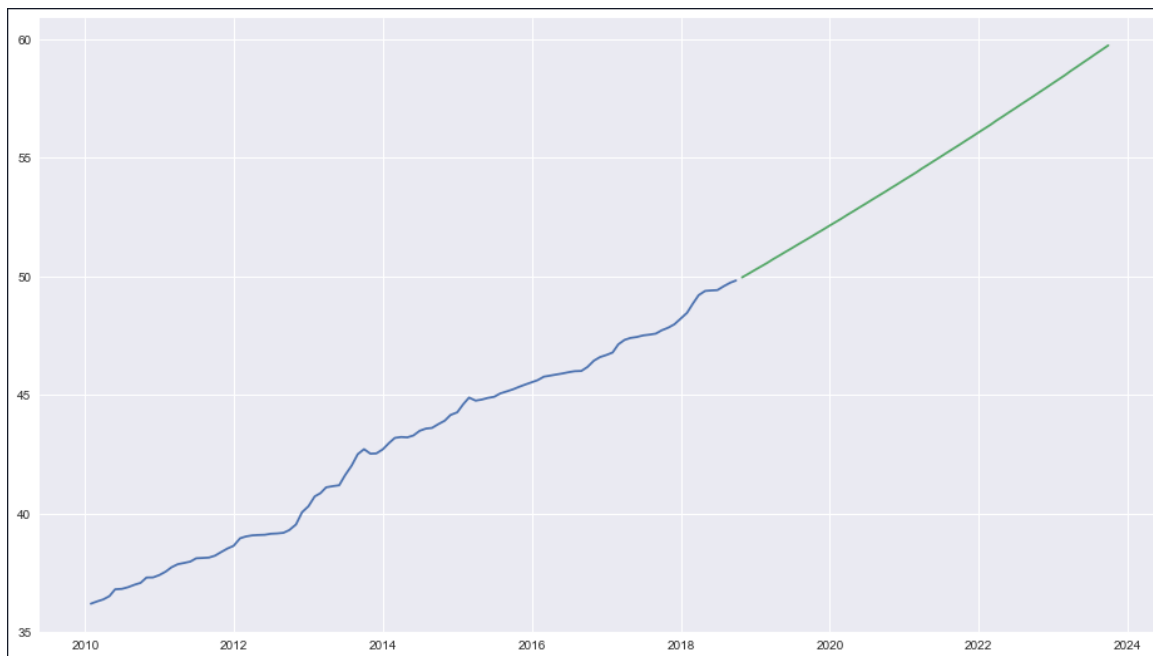
Y el AIC = **2677.8798047987025**



Tasa de cambio USD → DOP (2010-01-29 hasta 28-09-2018)

Parámetros del modelo ARIMA utilizado: **(1, 1, 0), (1, 1, 0, 24)**

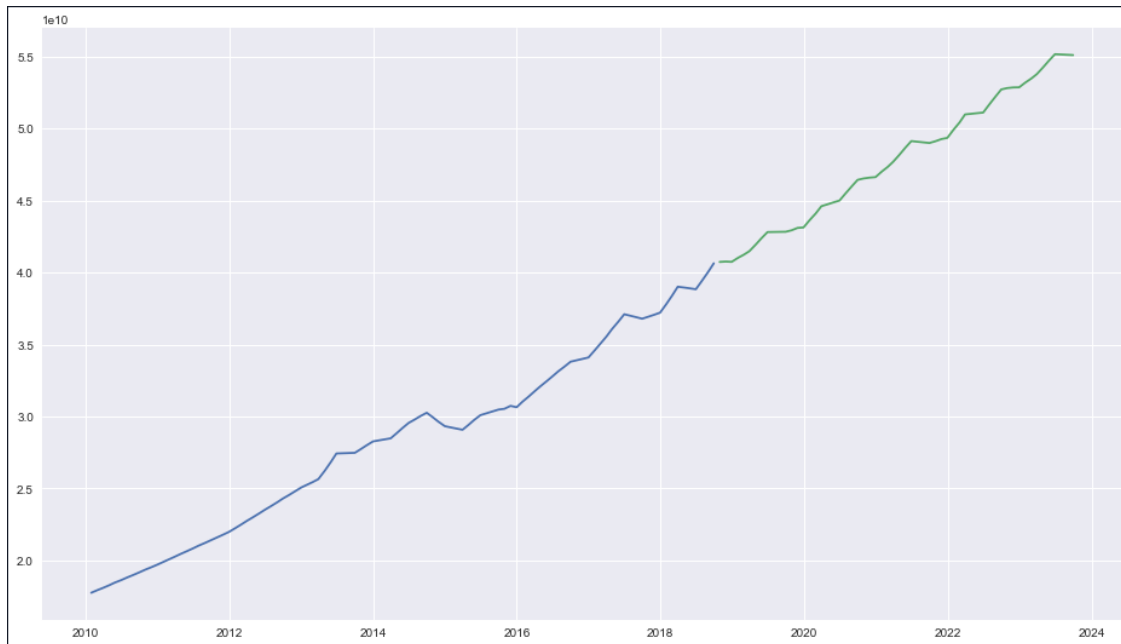
Y el AIC = **2263.192904339945**



Deuda pública consolidada RD (2010-01-29 hasta 28-09-2018)

Parámetros del modelo ARIMA utilizado: **(1, 1, 1), (0, 1, 1, 12)**

Y el AIC = **1223.037580855845**

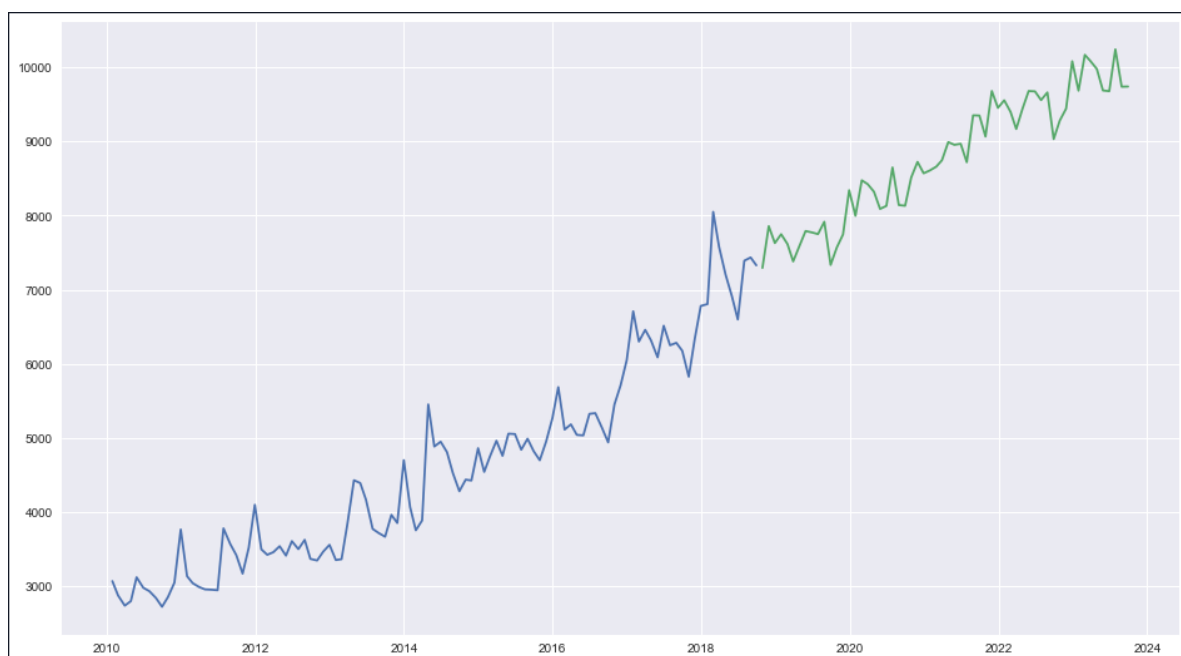


Análisis: Podemos claramente ver la tendencia creciente en las variables cuya información histórica refleja dicha tendencia. Las variables de Dinero en sentido amplio, deuda pública consolidada, precio del dólar y tasa de referencia a 6 meses del tesoro muestran una tendencia creciente.

La inflación se mantiene circulando entre 1% y -1%, bastante baja en general, el precio del petróleo lleva una estacionalidad bienal y las importaciones netas se ven creciendo en general, pero hay una clara estacionalidad anual a bienal.

Toda la información fue proyectada al 28 de septiembre del 2023 y fue evaluada y proyectada mensualmente para todas las variables.

Aquí tenemos la
regresión múltiple:



ANEXO: Información proyectada cruda mostrando el valor de las reservas brutas mensualmente desde octubre 2018 hasta septiembre 2023 **expresadas en miles de millones (*10^9):**

| | | | | | |
|-------------------|-------------|-------------------|-------------|-------------------|--------------|
| 2018-10-28 | 7294.227233 | 2020-07-28 | 8649.446351 | 2022-04-28 | 9443.344262 |
| 2018-11-28 | 7857.664922 | 2020-08-28 | 8142.421753 | 2022-05-28 | 9680.525791 |
| 2018-12-28 | 7629.123820 | 2020-09-28 | 8132.754286 | 2022-06-28 | 9675.493597 |
| 2019-01-28 | 7748.994199 | 2020-10-28 | 8514.779435 | 2022-07-28 | 9558.017757 |
| 2019-02-28 | 7615.806320 | 2020-11-28 | 8724.485901 | 2022-08-28 | 9661.463474 |
| 2019-03-28 | 7381.800301 | 2020-12-28 | 8570.594941 | 2022-09-28 | 9032.097088 |
| 2019-04-28 | 7591.397001 | 2021-01-28 | 8609.391976 | 2022-10-28 | 9282.652790 |
| 2019-05-28 | 7791.713779 | 2021-02-28 | 8659.889979 | 2022-11-28 | 9443.025823 |
| 2019-06-28 | 7772.774455 | 2021-03-28 | 8746.582113 | 2022-12-28 | 10081.955031 |
| 2019-07-28 | 7750.416506 | 2021-04-28 | 8991.373338 | 2023-01-28 | 9685.334464 |
| 2019-08-28 | 7916.121096 | 2021-05-28 | 8954.702842 | 2023-02-28 | 10168.544237 |
| 2019-09-28 | 7334.467806 | 2021-06-28 | 8968.760951 | 2023-03-28 | 10083.247858 |
| 2019-10-28 | 7566.964399 | 2021-07-28 | 8719.837306 | 2023-04-28 | 9977.996280 |
| 2019-11-28 | 7746.231899 | 2021-08-28 | 9352.201570 | 2023-05-28 | 9687.834772 |
| 2019-12-28 | 8342.250640 | 2021-09-28 | 9349.752700 | 2023-06-28 | 9678.442981 |
| 2020-01-28 | 7996.840517 | 2021-10-28 | 9066.035934 | 2023-07-28 | 10242.739073 |
| 2020-02-28 | 8476.258956 | 2021-11-28 | 9680.443921 | 2023-08-28 | 9738.458804 |
| 2020-03-28 | 8424.396478 | 2021-12-28 | 9453.494491 | 2023-09-28 | 9741.095746 |
| 2020-04-28 | 8321.731600 | 2022-01-28 | 9555.494069 | | |
| 2020-05-28 | 8089.914064 | 2022-02-28 | 9398.733821 | | |
| 2020-06-28 | 8130.723850 | 2022-03-28 | 9168.112233 | | |

Análisis de la proyección: Se proyectan 60 meses (5 años) desde la última fecha hasta la cual se tiene información de **todas** las variables. Las reservas claramente llevan una tendencia a crecer, se entiende que las caídas proyectadas que podríamos decir llevan un cierto patrón estacional anual muestran lo que probablemente serían inversiones y los picos y crecimientos retornos de inversiones.

Información general de la proyección:

Valor de las reservas en septiembre 2018: **7329.09648**

Valor proyectado de las reservas en septiembre 2023: **9741.095746**

Un crecimiento de un **32.90989689086476%**

¿Qué hace el código?:

Primero, la función ARIMA utilizada calcula todos los posibles parámetros de ARIMA para las estacionalidades mensual, bimensual, trimestral, cuatrimestral, semestral, anual, bienal, trienal y cuadrienal y luego itera sobre cada uno de ellos calculando el AIC para seleccionar el mejor modelo (el que tenga el AIC más alto).

Para acortar la longitud de ejecución y la complejidad del algoritmo, antes de correrlo se han reducido la cantidad posible de estacionalidades para cada variable de manera que no haya que calcular más estacionalidades de las cuales sabemos que son mejores. Por ejemplo, para importaciones netas, sabemos que la estacionalidad con mejor AIC es 24, entonces solo ejecutamos el algoritmo de manera que no haya que calcular el AIC para estacionalidades diferentes a 24 para esta variable en particular.

Luego de obtener cada modelo ARIMA lo adjuntamos a una tabla de datos como otra columna para cada variable. Al tener todas las columnas, creamos un modelo de regresión múltiple y designamos la columna de ReservasBrutas como la columna sobre la cual ejecutamos la proyección. Luego de la proyección, adjuntamos la información proyectada a un gráfico con la información histórica y podemos visualizar la proyección en el contexto de la data existente.