Modelo ARIMA

El modelo ARIMA es un modelo autoregresivo integrado de promedio movil. Utiliza variaciones y regresiones de datos para poder predecir en el futuro.

Hay tres terminos importantes en este modelo, los terminos p, d y q.

P: Orden del termino autoregresivo (AR) D: Orden de diferenciación (I) Q: Orden del termino de media movil (MA)

Es importante saber que se ha de tratar de hacer la serie estacionaria, para ello hay que conocer el termino de diferenciación. El termino de diferenciacion sera 0 (d=0), cuando la serie sea totalmente estacionaria. En caso contrario, se ha de saber el orden para poder hacerla estacionaria.

¿Como se sabe el orden?

In [11]: plt.figure(figsize=(15, 6))

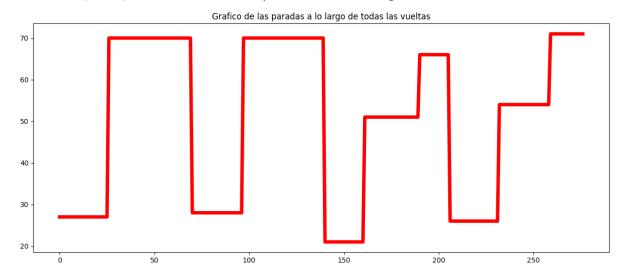
El orden se sabe usando la prueba ADF (Augmented Dickey Fuller). Esta prueba tiene como hipotesis nula -> Serie no estacionaria, por lo que si PValor > 0.05, entonces seguimos adelante con la diferenciación.

Para ello vamos a probarlo en nuestro dataset real, el dataset de las vueltas de Checo Perez en Austria en los GP 2018-2021.

```
In [2]: import warnings
         warnings.filterwarnings('ignore')
In [3]:
         import numpy as np, pandas as pd
         from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
         import matplotlib.pyplot as plt
In [4]: vueltas = pd.read csv("../datasets completos/laps Austria PER todas vueltas.
         vueltas.head()
Out[4]:
            LapTime LapNumber
                                 Compound TyreLife FreshTyre TrackStatus Stint AirTemp
         0
                  0
                                       NaN
                                                NaN
                                                          NaN
                                                                           NaN
                                                                                    22.7
         1 01:13.062
                              2 SUPERSOFT
                                                 1.0
                                                          True
                                                                        2
                                                                            1.0
                                                                                    22.7
           01:11.715
                              3 SUPERSOFT
                                                 2.0
                                                          True
                                                                            1.0
                                                                                    22.7
                                                                        1
            01:11.123
                              4 SUPERSOFT
                                                 3.0
                                                          True
                                                                            1.0
                                                                                    22.8
                                                 4.0
            01:11.171
                              5 SUPERSOFT
                                                          True
                                                                        1
                                                                            1.0
                                                                                    22.9
```

```
plt.plot(vueltas['nextPit'], lw=5, color="r")
plt.title("Grafico de las paradas a lo largo de todas las vueltas")
```

Out[11]: Text(0.5, 1.0, 'Grafico de las paradas a lo largo de todas las vueltas')



Ahora procedemos a ver la estacionalidad de nuestro modelo

```
In [21]: from statsmodels.tsa.stattools import adfuller
    estacionalidad = adfuller(vueltas['nextPit'])
    print("ADF estadístico: ", estacionalidad[0])
    print("P-valor: ", estacionalidad[1])
```

ADF estadístico: -2.796833353501025

P-valor: 0.0587275266358281

Dado que P-Valor > 0.05, se ha de diferenciar la serie

¿Pero no es muy poca la diferencia como para poder obviar esa diferencia que parece poco significativa?

```
In []:
```