

# 臺灣近年交通概況分析

Yi-Hong Wang

2025-08

## 1 介紹

### 1.1 前言

近年自從行人地獄的名詞成為臺灣交通狀況的代名詞後，越來越多人民投入到關心交通現況以及政策的環境當中，而剛好我在網路上發現警政署統計網中記錄著多年來臺灣發生的交通事故件數，因此在此我想分析這些資料去更細緻的瞭解臺灣的交通狀況及演變

### 1.2 資料簡介

在此我們蒐集了 2 份資料，分別是：

1. 臺灣近年交通事故件數，資料涵蓋時間為 2003 年 1 月至 2022 年 3 月
2. 臺灣近年酒駕件數與傷亡人數，資料涵蓋時間為 2014 年 1 月至 2025 年 4 月

這兩筆資料都記錄每月事故發生的件數，我將會分析這兩筆資料的趨勢並建立模型嘗試預測未來的情況。

### 1.3 模型簡介

細想後可以發現上述 2 筆資料與當下時間相關，因此明顯為統計學當中的時間序列 (time series) 資料，因此在此使用的模型為 SARIMAX 模型，全名為 Seasonal Autoregressive Integrated Moving Average with eXogenous Regressors，中文翻譯為季節性自迴歸整合移動平均與外生變數模型。

將模型內容拆開來理解

- S (Seasonal，季節性)：SARIMAX 模型能夠捕捉資料中週期性重複的模式之關鍵，例如每年重複出現的銷售高峰、每週的某一天特別忙碌等。
- AR (Autoregressive，自迴歸)：這部分表示模型會使用過去時間點的觀測值來預測未來。白話來說，就是「今天的狀況，會受到過去某幾天狀況的影響」。
- I (Integrated，整合)：這部分表示模型會對資料進行差分 (Differencing)。差分的目的是為了讓時間序列變得穩定 (Stationary)。一個穩定的時間序列其統計特性 (如平均值和變異數) 不會隨時間而變化，這對於大多數預測模型來說是個重要的假設。
- MA (Moving Average，移動平均)：這部分表示模型會使用過去時間點的預測誤差來預測未來。白話來說，就是「如果模型上次預測錯了，這次會根據錯誤來調整」。
- X (eXogenous Regressors，外生變數)：這是 SARIMAX 模型最大的靈活性所在。它允許你將外部因素 (例如天氣、節假日、促銷活動、經濟指標等) 納入模型中，作為預測目標變數的額外解釋變數。這能讓模型更全面地考慮各種影響因素，通常能提高預測的準確性。

## 2 資料分析

### 2.1 交通事故件數

繪製出交通事故件數與時間的分布圖

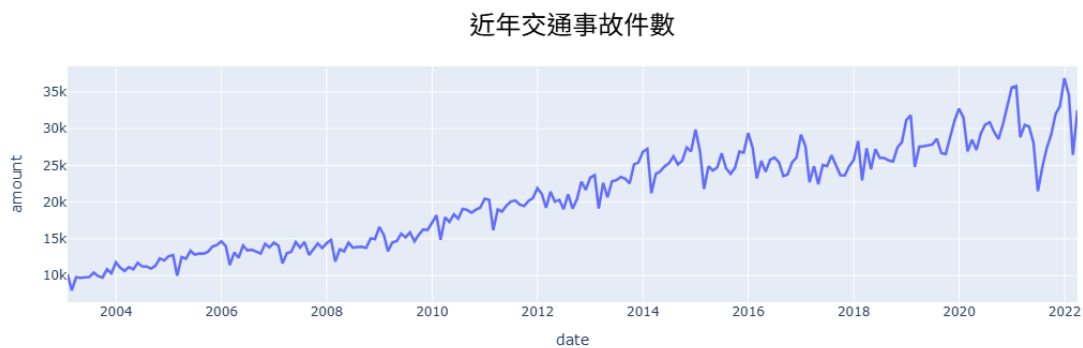


Figure 1

可以很清楚地知道事故件數隨著時間有明顯的提升，具有增加的趨勢。再更詳細地觀察可以看出似乎有季節性影響。

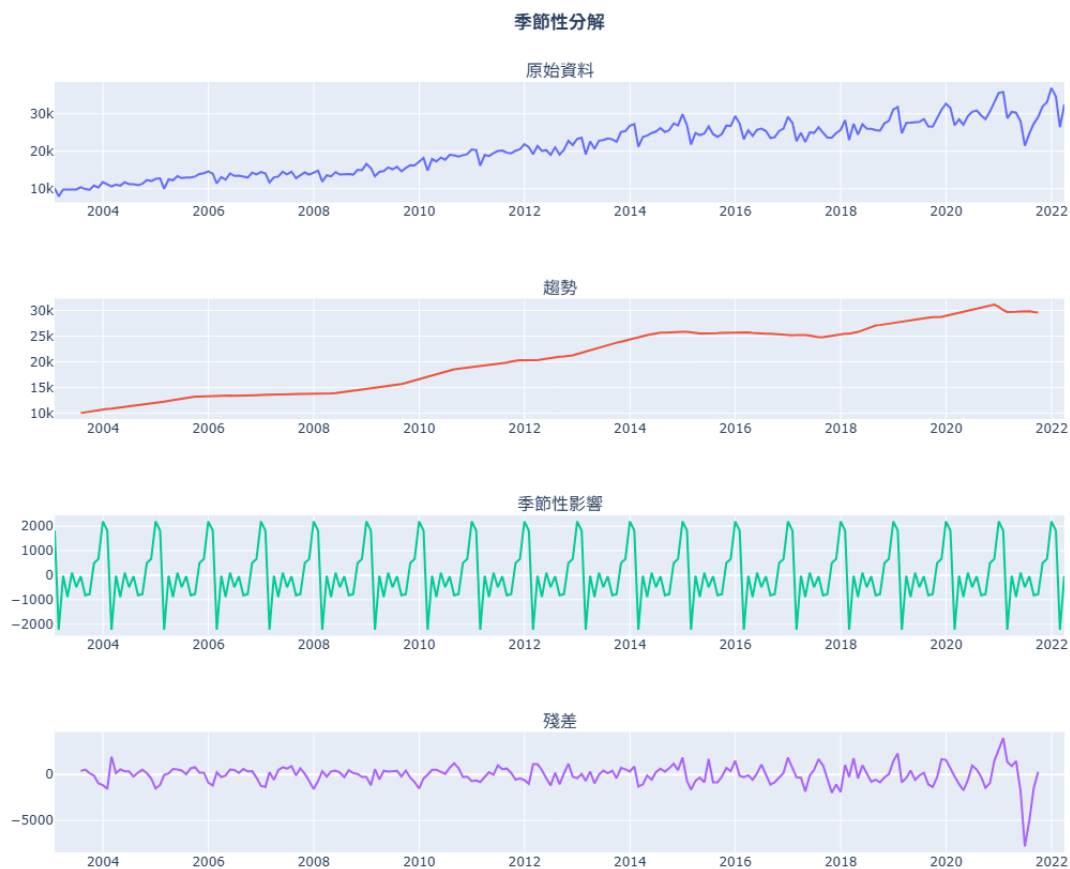


Figure 2

資料具有增加趨勢，需要進行差分，並且為了辨明是否需要季節性差分，將資料進行 1 次性差分，季節性差分，以及 2 種差分皆執行後，繪製其 acf 圖以及 pacf 圖並進行 Augmented Dickey-Fuller 檢定是否為可建立模型的穩定資料。

首先檢視原本資料的 acf 及 pacf 圖

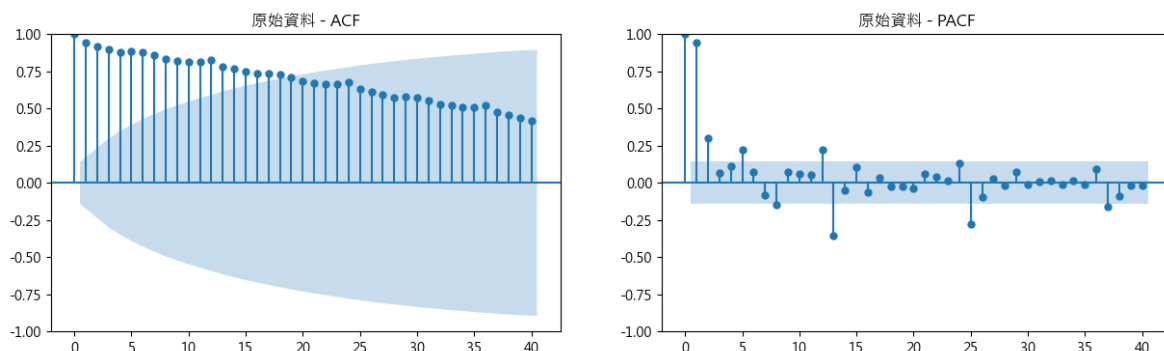


Figure 3

這 2 張圖可用來判斷目前資料與先前資料的相關性 (correlation)，原始資料 acf 圖展現出與先前資料的強烈相關，因此必定不是穩定態的資料。再來檢視經過 1 次差分資料的 acf 及 pacf 圖。

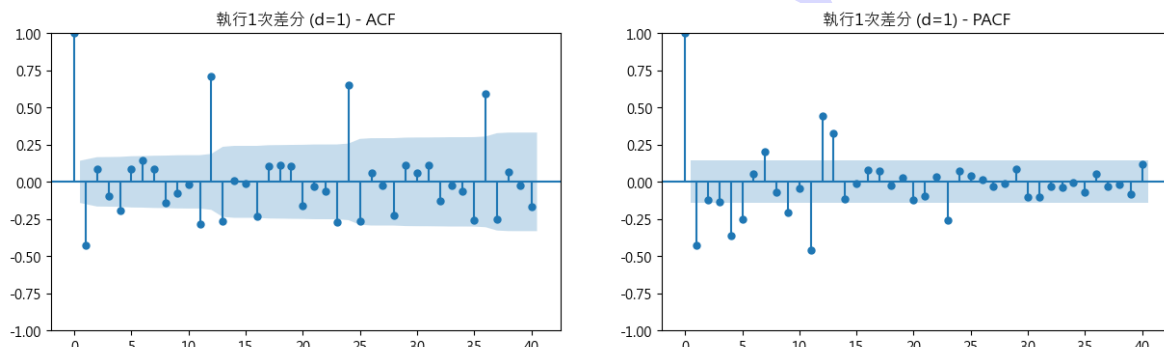


Figure 4

經過 1 次差分後，acf 圖中如同原始資料的強烈相關性質漸消失，但卻有週期性的強相關性，觀察後可以發現週期 (m) 間隔為 12，展現資料具有季節性影響，因此再觀察資料進行季節性差分後的結果，週期 (m) 挑選為 12。

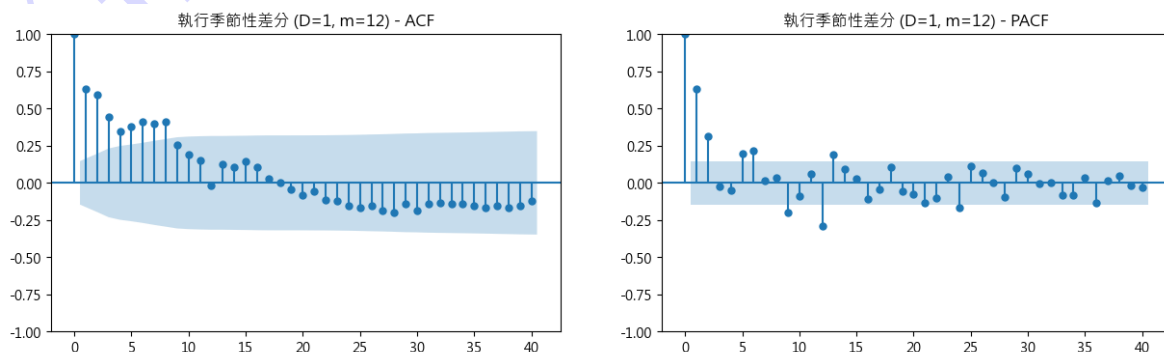


Figure 5

只經過季節性差分的結果可以看出，雖然週期性的強相關性已經消除，但其餘相關性仍然存在，因此

最後我們結合 2 種差分繪製其結果。

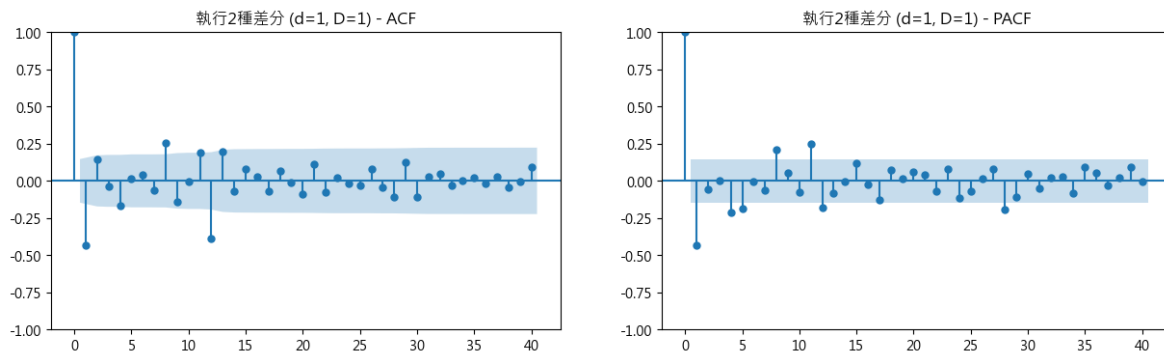


Figure 6

相較於前面的結果，2 種差分皆執行的資料展現出較穩定，與先前資料較無相關，趨近於穩定的結果。除了圖片判斷之外，我也對差分後的資料進行 Augmented Dickey-Fuller 檢定，若是檢定的 p-value 小於 0.05，便有足夠信心確定該資料為穩定可進行模型建立。只經過 1 次差分的資料檢定 p-value 為 0.0302，經過季節性差分的資料檢定 p-value 為 0.0937，2 種差分皆執行的資料檢定 p-value 為 0.0192，因此最後確定建立模型的資料需要執行 2 種差分。

綜合上述的結果，建立最佳的 SARIMAX 模型，利用了 python 當中的 library pmdarima 協助進行模型建立，挑選出  $SARIMAX(p,d,q) \times (P,D,Q,m) = SARIMAX(1,1,1) \times (0,1,1,12)$  的最佳模型，稍微說明當中參數的意義。

非季節性部分: 這部分描述了時間序列的非季節性趨勢和短期波動  $(p, d, q) = (1, 1, 1)$

- $p = 1$  (AR 階數): 表示模型使用過去 1 個時間點的數值來預測當前值。這代表當前的數值與前一期數值存在直接的線性關係 (在經過差分後)。
- $d = 1$  (差分階數): 表示資料需要進行一次差分才能變得平穩 (Stationary)。這通常意味著原始時間序列具有長期趨勢 (例如持續上升或下降)，通過差分可以消除這個趨勢。
- $q = 1$  (MA 階數): 表示模型使用過去 1 個時間點的預測誤差 (殘差) 來預測當前值。這意味著當前的數值會受到前一期預測誤差的影響。

季節性部分: 這部分描述了時間序列中週期性重複的模式  $(P, D, Q, m) = (0, 1, 1, 12)$

- $P = 0$  (季節性 AR 階數): 表示模型沒有使用過去季節性週期的數值作為季節性自迴歸因子。
- $D = 1$  (季節性差分階數): 表示資料需要進行一次季節性差分才能使其季節性部分變得平穩。這通常意味著資料的季節性波動的平均值或方差會隨時間變化，通過季節性差分可以穩定這種季節性模式。
- $Q = 1$  (季節性 MA 階數): 表示模型使用過去 1 個季節性週期的預測誤差來預測當前值。這意味著當前的數值會受到 12 個月前預測誤差的影響 ( $m=12$ )
- $m = 12$  (季節性週期): 這表示資料的季節性週期是 12 個時間點。在大多數情況下，這代表資料是月度數據，季節性模式每 12 個月重複一次 (例如年度季節性)。

## 2.2 酒駕事故件數

繪製出酒駕事故件數與時間的分布圖

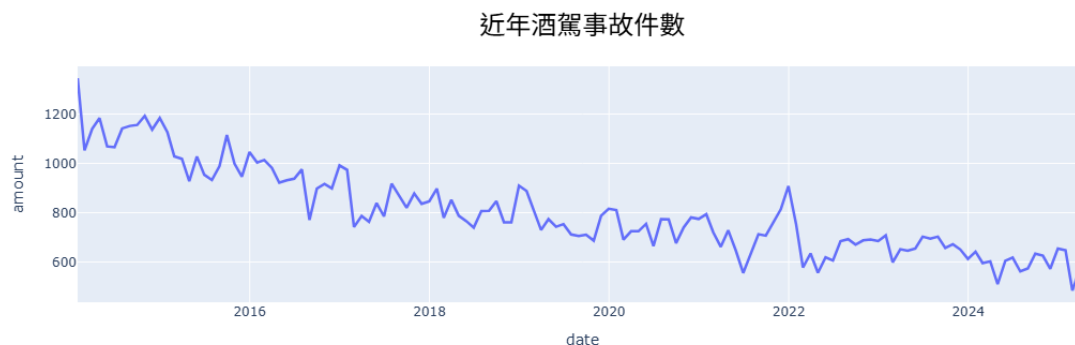


Figure 7

可以很清楚地知道事故件數隨著時間有明顯地減少，具有減少的趨勢。再更詳細地觀察可以看出似乎有季節性影響。

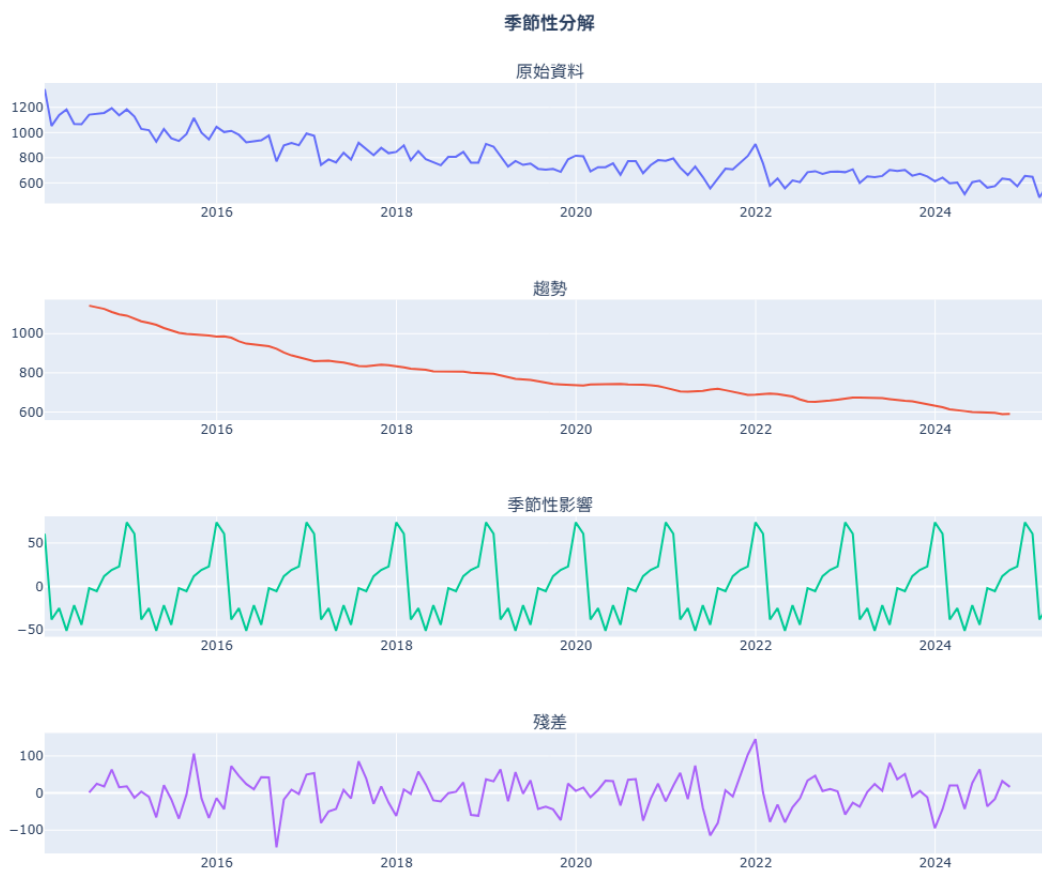


Figure 8

資料具有減少趨勢，需要進行差分，並且為了辨明是否需要季節性差分，將資料進行 1 次性差分，季節性差分，以及 2 種差分皆執行後，繪製其 acf 圖以及 pacf 圖並進行 Augmented Dickey-Fuller 檢定

是否為可建立模型的穩定資料。在此不像前 1 筆資料，我不再畫出原始資料的 acf 圖以及 pacf 圖，而是直接畫出差分後的結果進行比較。

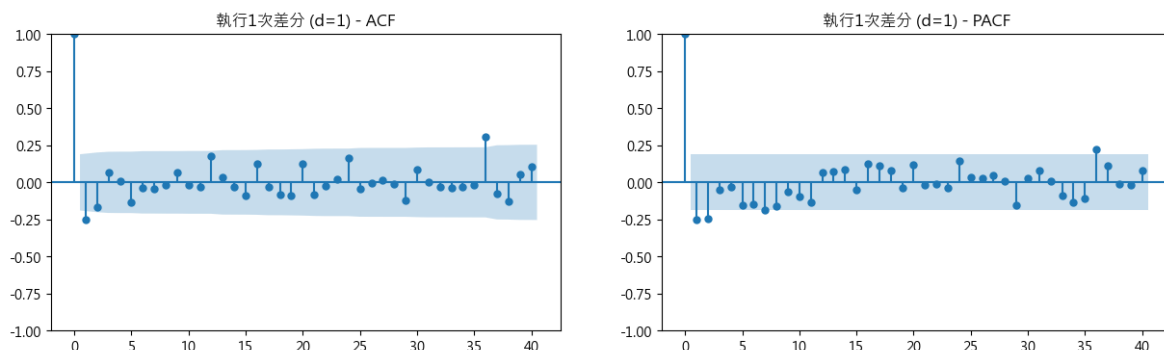


Figure 9

經過 1 次差分後，acf 圖中強烈相關性皆消失，季節性相關也不明顯，再觀察資料進行季節性差分的結果，週期 (m) 挑選為 12。

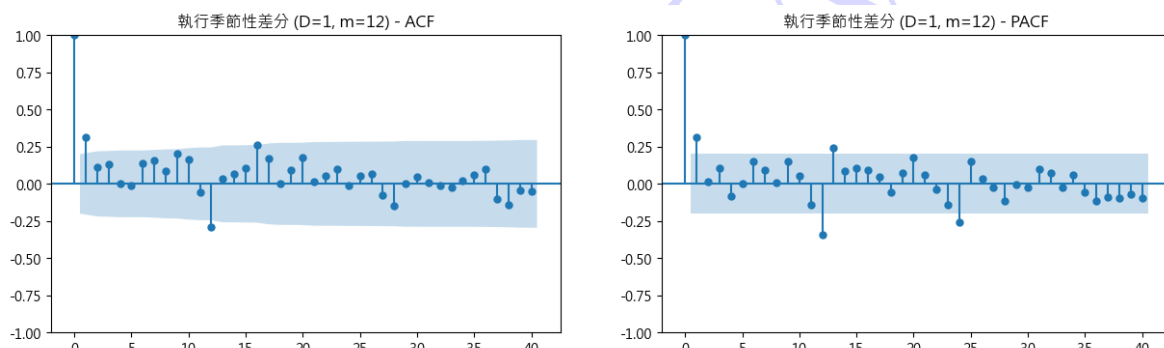


Figure 10

只經過季節性差分的結果可以看出，同前 1 張圖強烈相關性與季節性相關也不明顯。最後我們結合 2 種差分繪製其結果。

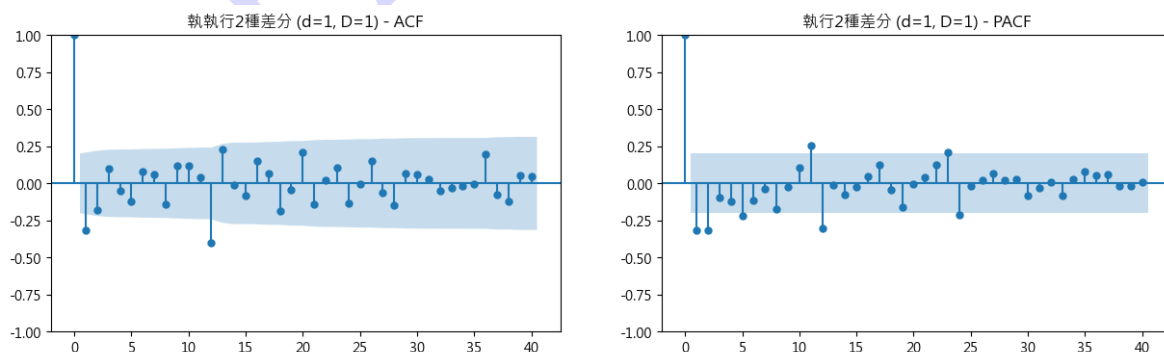


Figure 11

綜合於前面的結果，3 種不同的差分結果 acf 圖以及 pacf 圖皆差異不大，因此為了更準確的判斷，對差分後的資料進行 Augmented Dickey-Fuller 檢定。只經過 1 次差分的資料檢定 p-value  $< 0.00001$ ，經過季節性差分的資料檢定 p-value 為 0.0578，2 種差分皆執行的資料檢定 p-value 為 0.0028。由於資料經過季節性差分並無明顯助益，因此決定資料只需進行 1 次差分。

綜合上述的結果，建立最佳的 SARIMAX 模型，同樣利用了 python 當中的 library pmdarima 協助進行模型建立，挑選出  $\text{SARIMAX}(p,d,q)\times(P,D,Q,m) = \text{SARIMAX}(0, 1, 2)\times(3, 0, 0, 12)$  的最佳模型。

非季節性部分: 這部分描述了時間序列的非季節性趨勢和短期波動  $(p, d, q) = (0, 1, 2)$

- $p = 0$  (AR 階數): 模型沒有使用過去的值作為自迴歸因子。這表示今天的數值與昨天、前天的數值沒有直接的線性關係，而是與它們的預測誤差有關。
- $d = 1$  (差分階數): 資料需要進行一次差分來移除趨勢，使其變得平穩 (Stationary)。這通常意味著原始時間序列具有長期趨勢 (例如持續上升或下降)，通過差分可以消除這個趨勢。
- $q = 2$  (MA 階數): 模型使用過去 2 個時間點的預測誤差 (殘差) 來預測當前值。這意味著當前的數值會受到前一期和前兩期預測誤差的影響。

季節性部分: 這部分描述了時間序列中週期性重複的模式  $(P, D, Q, m) = (3, 0, 0, 12)$

- $P = 3$  (季節性 AR 階數): 模型使用過去 3 個季節性週期的數值來預測當前值。例如，如果週期  $m = 12$ ，模型會看 12 個月前、24 個月前和 36 個月前的數值來進行預測，以捕捉季節性關係。
- $D = 0$  (季節性差分階數): 表示資料不需要進行季節性差分。這通常意味著資料的季節性模式是穩定的，不會隨時間改變幅度或方差。
- $Q = 0$  (季節性 MA 階數): 表示沒有任何特定的滯後項被納入季節性移動平均部分
- $m = 12$  (季節性週期): 這表示資料的季節性週期是 12 個時間點。在大多數情況下，這代表資料是月度數據，季節性模式每 12 個月重複一次 (例如年度季節性)。

### 3 預測結果與結論

#### 3.1 交通事故件數

建模完成後，預測未來 36 個月的事故件數

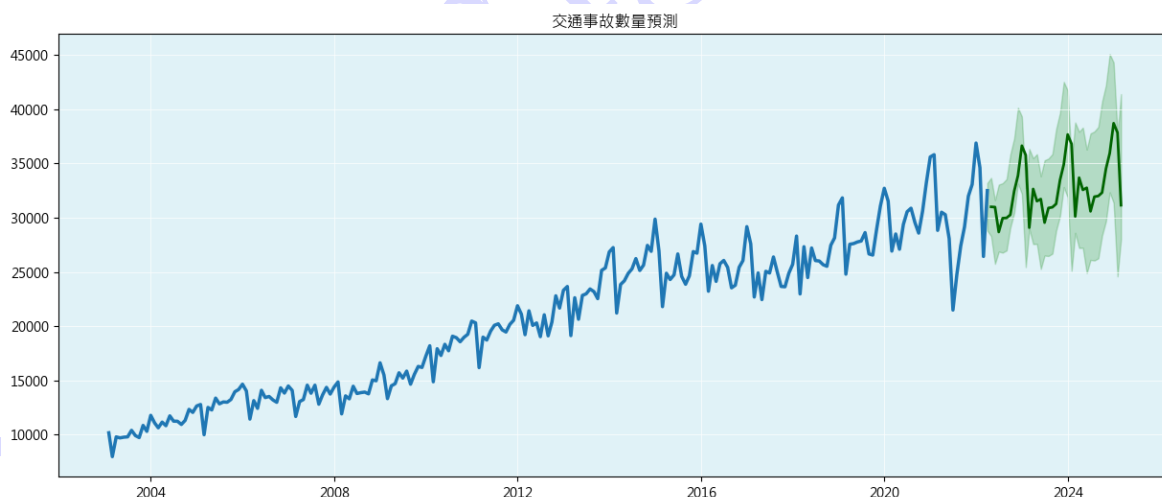


Figure 12

若沒有進行政策調整，未來事故件數將會繼續維持成長的趨勢。

#### 3.2 酒駕事故件數

建模完成後，同樣預測未來 36 個月的事故件數

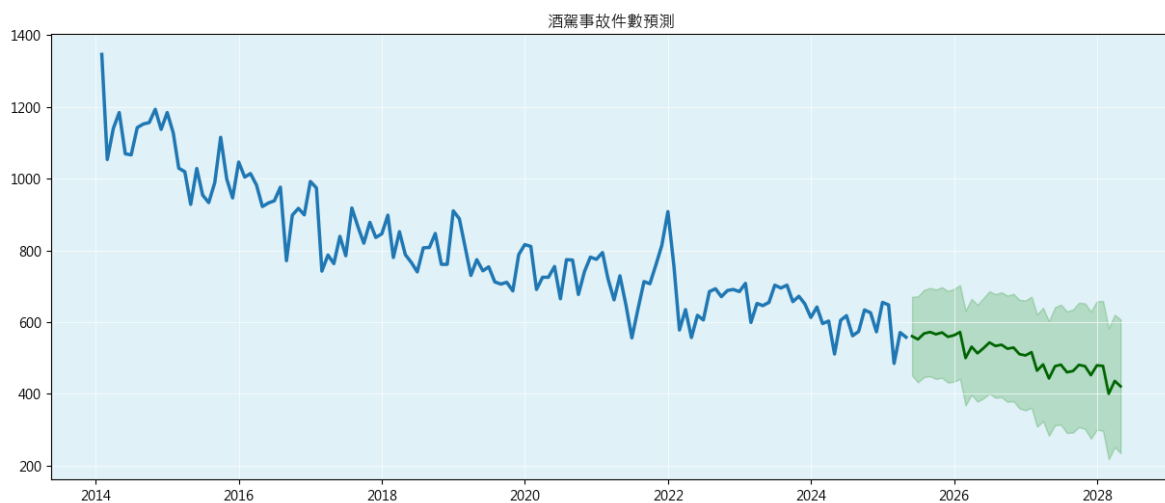


Figure 13

未來酒駕件數將會維持減少的趨勢，2028 年月總量將會減少至 400 件以下。

在 2 筆資料重疊的時間段，事故件數不斷增加，酒駕件數卻是不斷減少，顯示有酒駕以外的事故肇因不只填補了酒駕事件帶來的短缺，甚至帶動總體事故數的成長，這是主管機關需要找出並制訂應對方法的關鍵。