



法律案件諮詢系統

110354001吳宥群 110354003王辰安
110354013陳韋豫 110354014李其軒
110354015潘禹翔 110354024林韋成
110354029陳槐廷

目錄頁

Contents Page



第一部分

委託介紹

第二部分

法院判決書介紹

第三部分

資料處理

第四部分

模型建立

第五部分

系統介面



委託介紹

► 目的:

建立一個法律案件諮詢系統，方便律師根據每個案件不同的情況，回答系統內的重要問題後，得到該案件勝訴或敗訴的機率，以協助律師選擇是否受理該案件。

► 現況:

由於民事、刑事訴訟案件中有許多種類，因此我們縮小案件的範圍，僅針對交通案件作分析，找出影響判決的關鍵因素。



判決書介紹

- ▶ 在司法院裁判書系統中，我們可以查詢到歷年交通案件的判決書。
- ▶ 判決書主文可以快速了解案件的判決結果。
- ▶ 判決書事實及理由中詳細記錄事實經過以及違反的法規。



判決書介紹

臺灣新竹地方法院刑事簡易判決

111年度竹交簡字第634號

聲 請 人 臺灣新竹地方檢察署檢察官
被 告 [REDACTED]

上列被告因公共危險案件，經檢察官聲請以簡易判決處刑（111年度偵字第16383號），本院判決如下：

主 文

[REDACTED] 吐氣所含酒精濃度達每公升零點二五毫克以上而駕駛動力交通工具，累犯，處有期徒刑肆月，併科罰金新臺幣壹萬元，徒刑如易科罰金，罰金如易服勞役，均以新臺幣壹仟元折算壹日。

事 實 及 理 由

- 一、本件犯罪事實及證據，除犯罪事實欄關於為警攔查時間之記載應更正為「民國111年11月5日凌晨1時49分許」外，餘均引用檢察官聲請簡易判決處刑書之記載（如附件）。
- 二、核被告所為，係犯刑法第185條之3第1項第1款之吐氣所含酒精濃度達每公升0.25毫克以上而駕駛動力交通工具罪。又被告有附件犯罪事實欄所示有期徒刑執行完畢之事實，業經檢察官具體記載於附件犯罪事實欄內，並有檢察官提出之刑案



判決書介紹

► 從主文中可以快速判斷此案件判決的類型為：

1.有罪 2.無罪

► 從主文、全文中找出關鍵變數：

1.酒駕 2.駕照 3.過失傷害 4.過失致死 5.肇事逃逸 6.非法藥物 7.自首
8.累犯 9.減速 10.超速 11.精神疾病 12.證據 13.車輛類型



資料處理

爬蟲抓取司法院裁判書系統



資料爬取

篩選關鍵變數並做EDA



選擇變數

資料處理



正則表達式以及結巴分詞切分判
決書



資料處理

利用 Python 到司法院裁判書系統抓取判決書資料。

- ❖ 時間範圍

有罪判決書：110年11月1日到111年10月31日

無罪判決書：108年11月1日到111年10月31日

- ❖ 判決地點：台灣各地方法院

- ❖ 司法院裁判書系統: <https://judgment.judicial.gov.tw/FJUD/default.aspx>

- ❖ 樣本數:19073 筆(有罪、無罪)

- ❖ 使用套件: re, requests, BeautifulSoup, Comment, selenium



資料處理

❖ 資料處理目的：

- 輸入判決書主文，並將判決書內容轉換成我們所考慮會與判決結果有關的關鍵變數值。
- 藉由EDA及Logistic Regression判斷該關鍵變數是否重要或顯著，以進行決策樹的建立。

❖ 資料轉換及文字處理方式：

- 正則表達式
- 結巴分詞



正則表達式 (Regression Expression, regex)

- ❖ 用「字串」來描述文中符合指定格式的「字串」
 - ex. regex = 'a|b'
 - string = ['ab' , 'a' , 'b' , 'c' , 'erb']
 - 匹配 => ^ ^ ^ ^ ^
- ❖ 應用在文本上，可以用來匹配、替代、分組取出等等
 - ex. regex = '(.+)nihao'
 - string = ['panpannihao' , 'rogernihao' ,
 'gogorohi']
 - 分組 => panpan roger None



正則表達式 (Regular Expression, regex)

- ❖ 對於各個變數，「觀察」是否有某些特定的模式
- ❖ 以變數「有無過失傷害」作為範例
- ❖ 在主文中，不同年代，不同的紀錄員會有不一樣的寫法，好比過失傷害、過失傷害人、過失致人傷害、過失致人重傷害.....
- ❖ 利用正則表達式網站(Regex101)一次匹配大量判決書主文



資料處理

REGULAR EXPRESSION

6 matches (3 730 steps, 6.3ms)

`[:r" (?:(?:過失)?致?人?受?於?重?傷害?罪?人?`

" gm



TEST STRING

- 1 黃威哲犯過失傷害罪，處拘役肆拾日，如易科罰金，以新臺幣壹仟元折算壹日。
- 2 李崑境駕駛動力交通工具而吐氣所含酒精濃度達每公升零點二五毫克以上，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日；又犯過失傷害罪，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日。應執行有期徒刑肆月，如易科罰金，以新臺幣壹仟元折算壹日。犯罪
- 3 鍾招智犯過失傷害罪，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日。
- 4 葉益榮犯不能安全駕駛動力交通工具罪，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日。
- 5 許瑞祥駕駛動力交通工具而吐氣所含酒精濃度達每公升零點二五毫克以上，處有期徒刑參月，如易科罰金，以新臺幣壹仟元折算壹日。
- 6 潘韻心犯肇事致人傷害逃逸罪，處有期徒刑陸月，如易科罰金，以新臺幣壹仟元折算壹日。
- 7 陳子偉駕駛動力交通工具而血液中酒精濃度達百分之零點零五以上，處有期徒刑肆月，如易科罰金，以新臺幣壹仟元折算壹日。又犯過失傷害罪，處有期徒刑肆月，如易科罰金，以新臺幣壹仟元折算壹日。應執行有期徒刑柒月，如易科罰金，以新臺幣壹仟元折算壹日。
- 8 陳金榮汽車駕駛人，無駕駛執照駕車，因過失傷害人，處拘役伍拾玖日，如易科罰金，以新臺幣壹仟元折算壹日。



正則表達式 (Regression Expression, regex)

變數	正則表達式
有無過失傷害	(?:過失)?致?人?受?於?重?傷害?罪?人?
有無過失致死	過失致人?於?死?罪?
有無酒駕	(?:動力交通工具而有?)(?:吐氣 血液中)?(?:所含)?酒精濃度達(?:百分之 每公升)?零點(?:.+?)(?:毫克)?以上罪(?:之情形)? (?:不能安全駕駛)?(?:而駕駛)?動力交通工具罪(?:而駕駛)?者?
有無駕照	無(?:駕駛執)?照



正則表達式 (Regression Expression, regex)

變數	正則表達式
是否自首	自首 坦承(?:犯行 不諱)?
是否服用非法藥物	(?:第.級)毒品?
是否肇事逃逸	逃逸
是否為累犯	累犯



正則表達式 (Regression Expression, regex)

判決結果	正則表達式
有罪	(?:受刑人)?(?:否准 不准)(?:受刑人)?(?:.+)?易科罰金(?:.+)(?:應予)? 撤銷
	原(?:判決 處分)(?:.*)?撤銷前?上?開撤銷部分
	(?:緩刑)(?:[一二三四五六七八九十壹貳參肆伍陸柒捌玖拾廿卅貳叁 叁陆1234567890年個月日\.] +)
	原(?:判決 處分)(?:.*)?撤銷(?:.+)(?:未經許可 過失(?:傷害)?致?人 於(?:於死)? 竊盜 施用 持有 肇事 使 意圖 駕駛 犯 因 服用酒類 吐 氣 於公務員依法執行職務時)(?:.+)



正則表達式 (Regression Expression, regex)

判決結果	正則表達式
無罪	(?:原(?:判決 處分)撤銷)?(?:.+?)無罪(?:.+)?
	(?:原(?:判決 處分)撤銷)?(?:.+?)不罰(?:.+)?
	(?:原(?:.+?)撤銷)?(?:.+?)訴訟費用新(?:.+?)幣(?:.+?)元由被告負擔(?:.+)?



結巴分詞

❖ 原理：

- 基於**預設的辭典**（基於人民日報語料等資源訓練）進行高效率的詞彙掃描並生成多個有向無環圖。
- 以動態規劃的方式計算該有向無環圖的概率，每個圖的概率則是依據詞頻而來，最後選取**有著最大概率的圖做為分詞的結果**，作為最後分詞的結果。

❖ 對於未於預設詞典裡卻有實際意義的詞：

- 藉由結疤分詞裡所使用到的演算法（Hidden Markov model + Viterbi）分出未登錄詞。
- 直接增設自訂義辭典。



結巴分詞

❖ 增設自訂義辭典：

- 針對我們考慮的屬性進行定義，如：自用小客車、證據不足等。

❖ 分詞結果的模糊處理：

- 藉由分出某些關鍵詞以判斷該關鍵屬性的值。如若分出「自用小客車」，則「是否有小客車」這個關鍵屬性則定義成1，若無則定義成0。
- 雖分出關鍵詞，但無法判定對象。不過該詞仍為該判決書之特性，故我們仍考慮它。
- 是否採用則須以EDA與Logistic Regression決定。



資料處理

結巴分詞

參考屬性	屬性值 = 1	屬性值 = 0
有無超速	超速	無提及
有無精神疾病	精神疾病	無提及
有無減速	無提及	無減速、沒有減速、未減速慢行、沒有減速慢行
有無證據	無提及	證據不足、證據尚不足、無證據不得認定犯罪事實、無證據



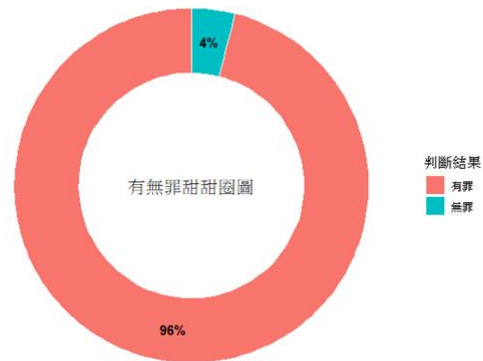
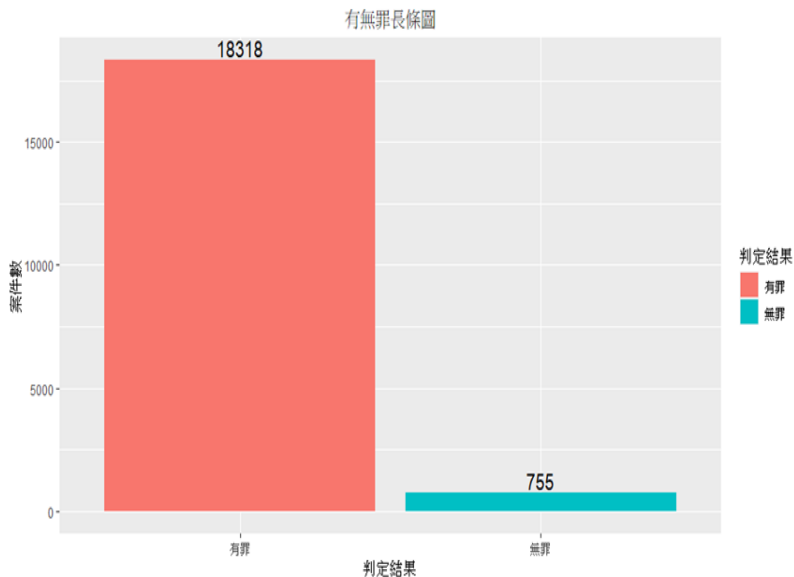
資料處理

結巴分詞

參考屬性	屬性值 = 1	屬性值 = 0
是否有小客車	小客車、汽車、轎車、休旅車、自用小客車	無提及
是否有貨車或大客車	貨車、大客車、公車	無提及
是否有機車	輕型機車、重型機車	無提及
是否有自行車	自行車、腳踏車	無提及



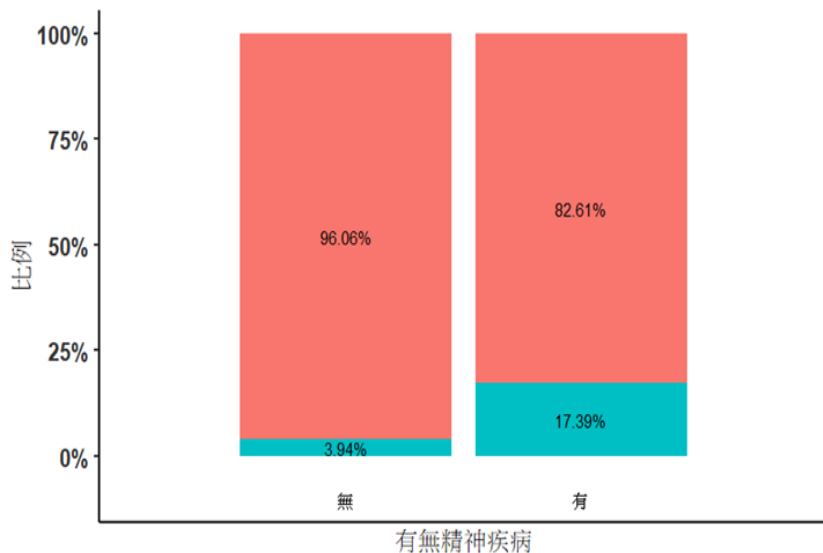
EDA-判決結果



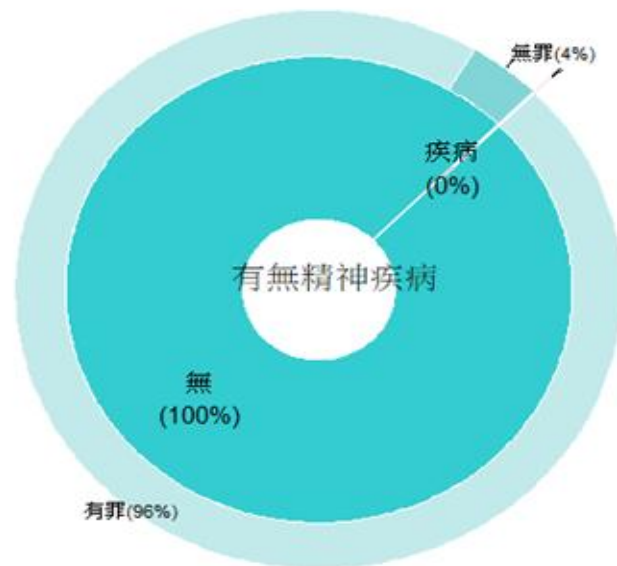


EDA-有無精神疾病

在有無精神疾病下有無罪相對比例長條圖



有無罪 vs. 有無精神疾病(依有無罪比例)

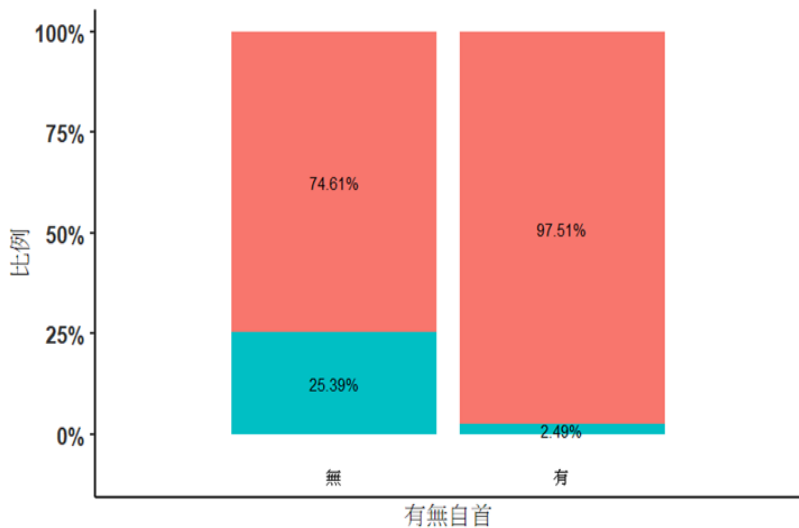




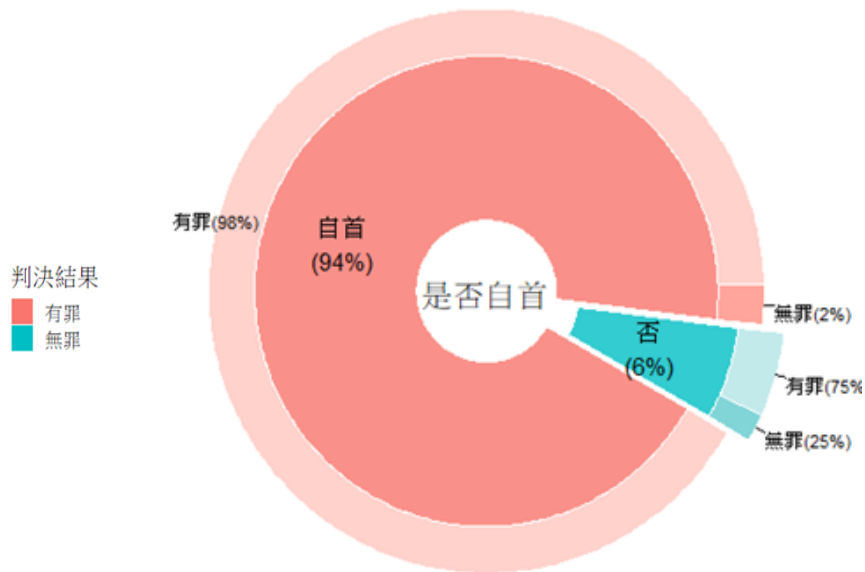
資料處理

EDA- 有無自首

在有無自首下有無罪相對比例長條圖



有無罪 vs. 是否自首(依有無罪比例)

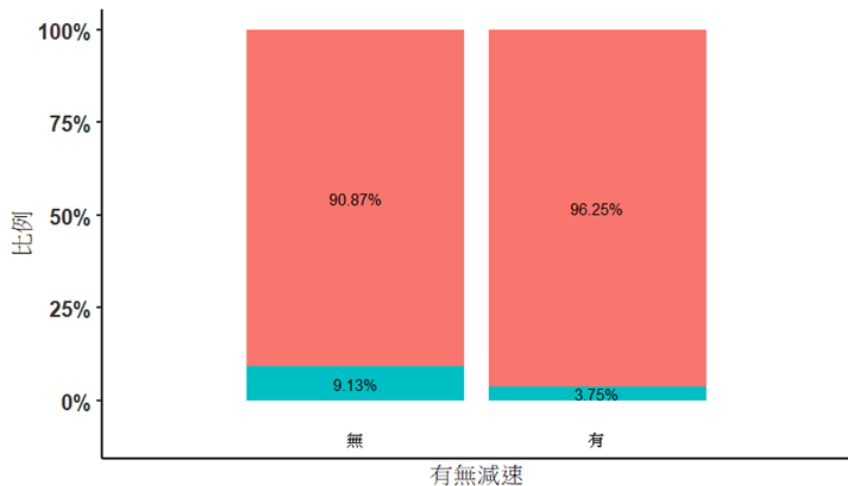




資料處理

EDA-有無減速

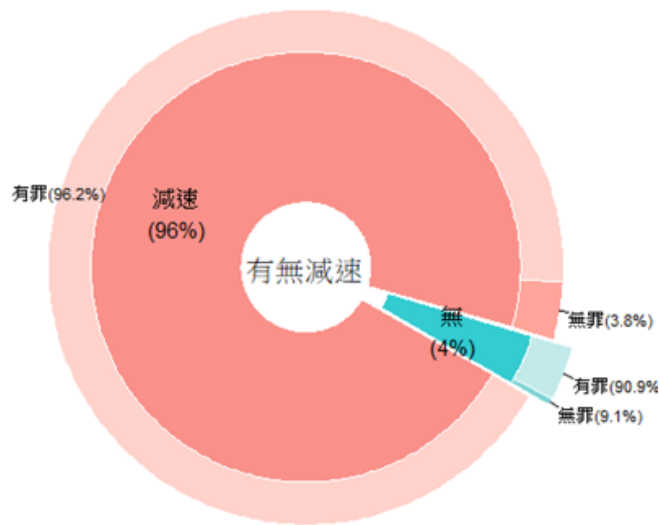
在有無減速下有無罪相對比例長條圖



有無罪 vs. 有無減速(依有無罪比例)

判決結果

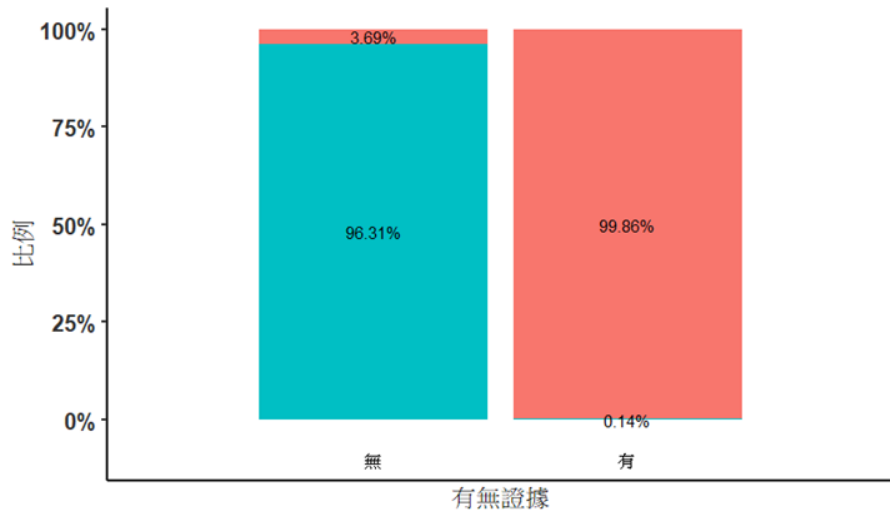
有罪
無罪



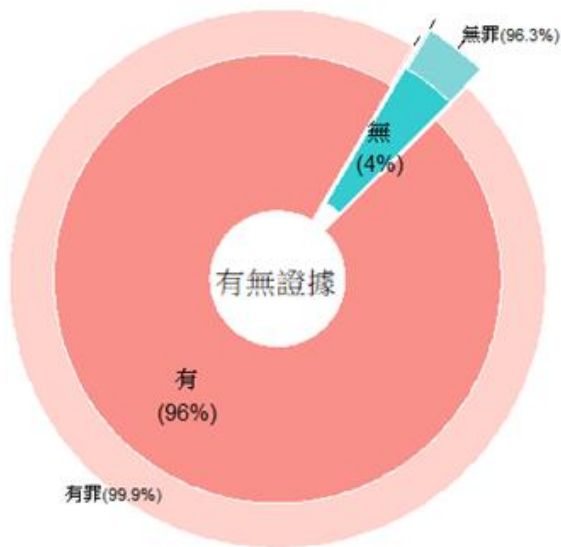


EDA-有無證據

在有無證據下有無罪相對比例長條圖



有無罪 vs. 有無證據(依有無罪比例)

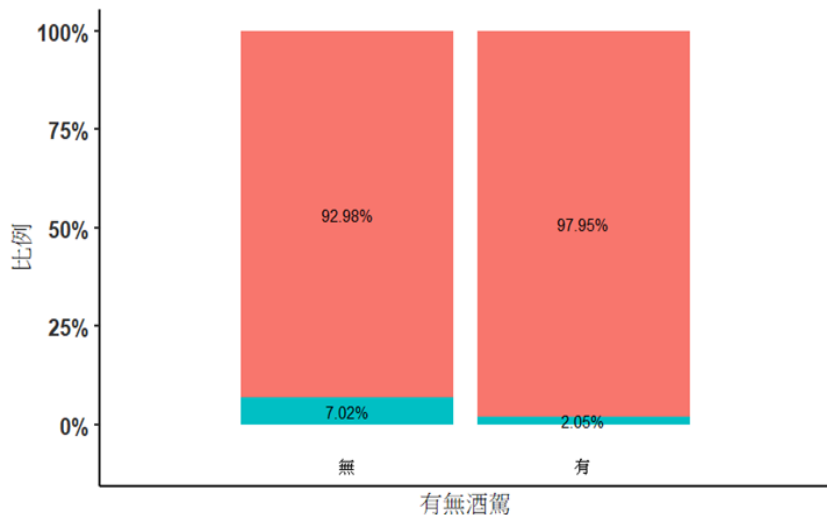




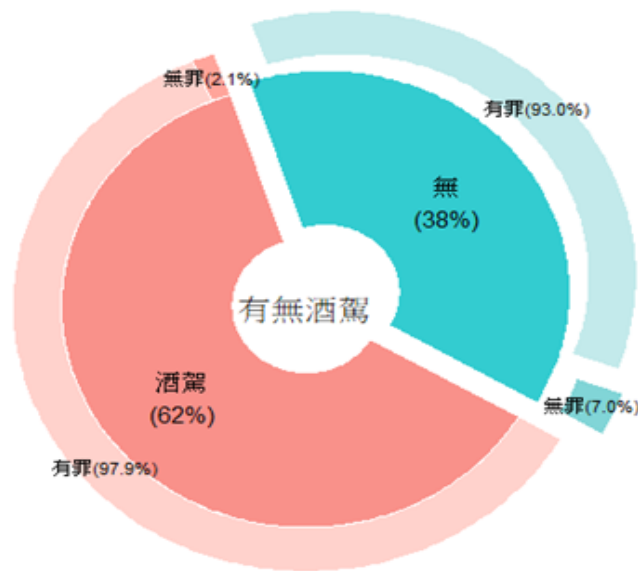
資料處理

EDA-有無酒駕

在有無酒駕下有無罪相對比例長條圖



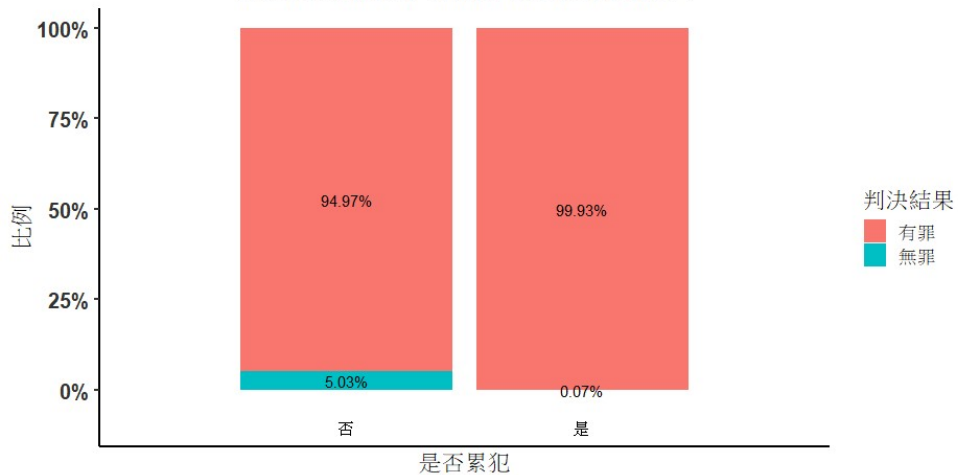
有無罪 vs. 有無酒駕(依有無罪比例)



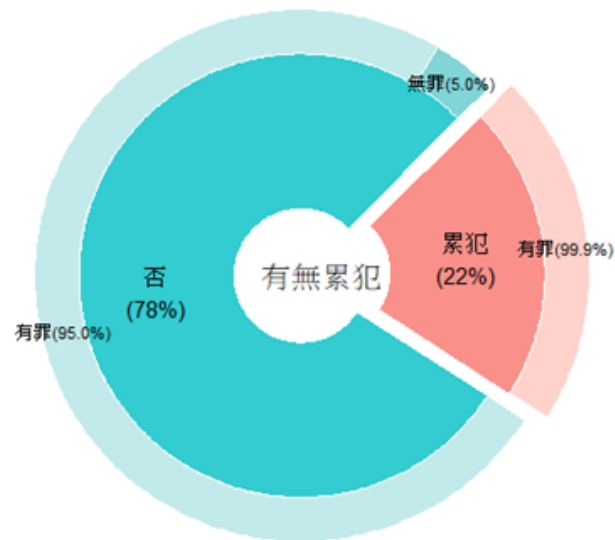


EDA-是否為累犯

在是否為累犯下有無罪相對比例長條圖



有無罪 vs. 是否累犯(依有無罪比例)

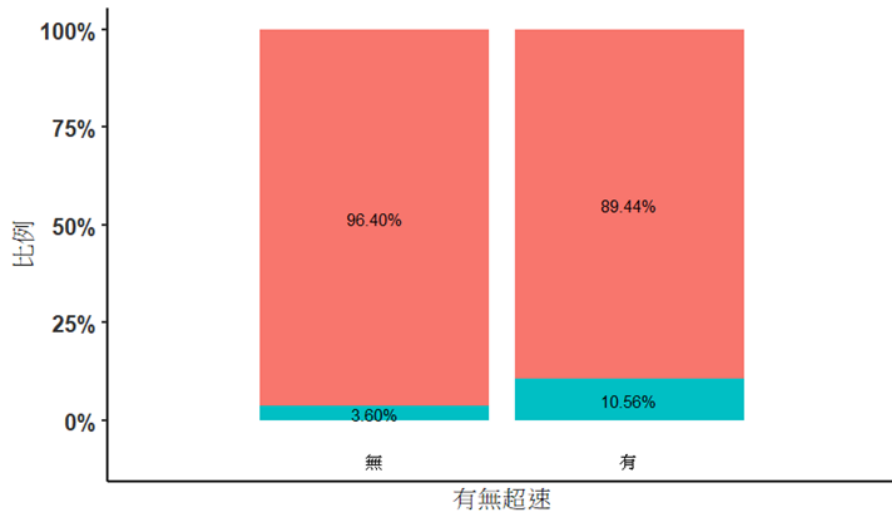




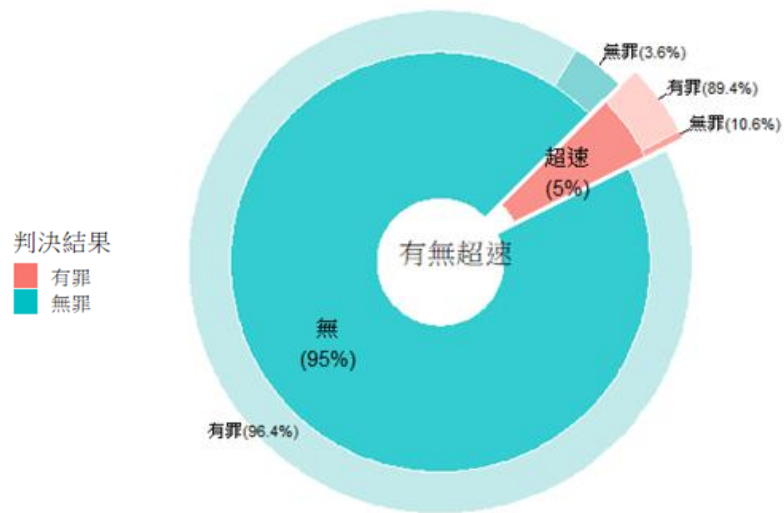
資料處理

EDA-有無超速

在有無超速下有無罪相對比例長條圖



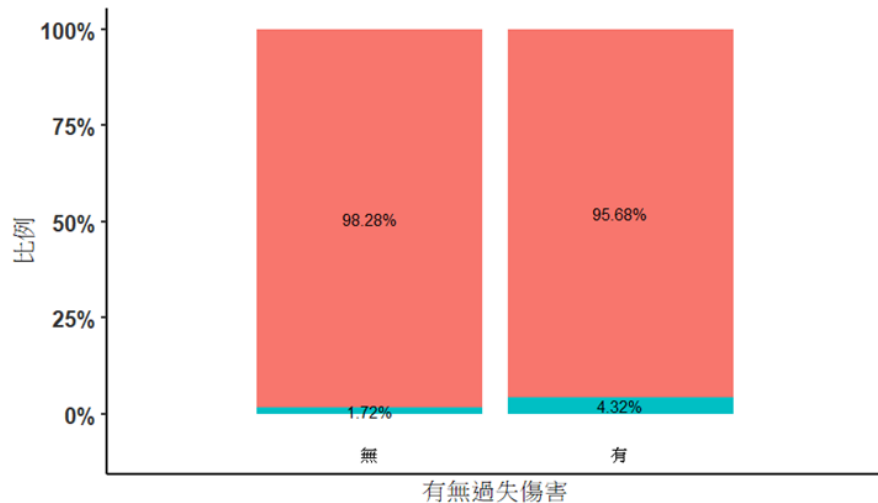
有無罪 vs. 有無超速(依有無罪比例)



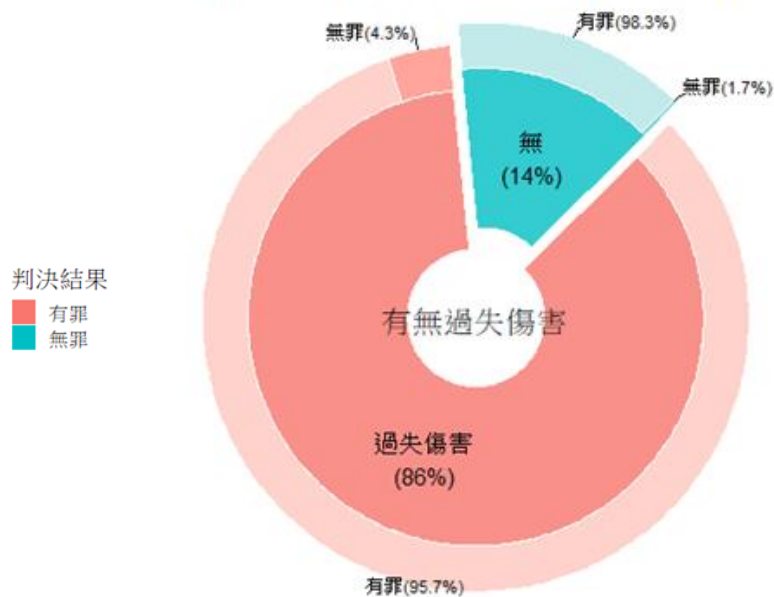


EDA-有無過失傷害

在有無過失傷害下有無罪相對比例長條圖



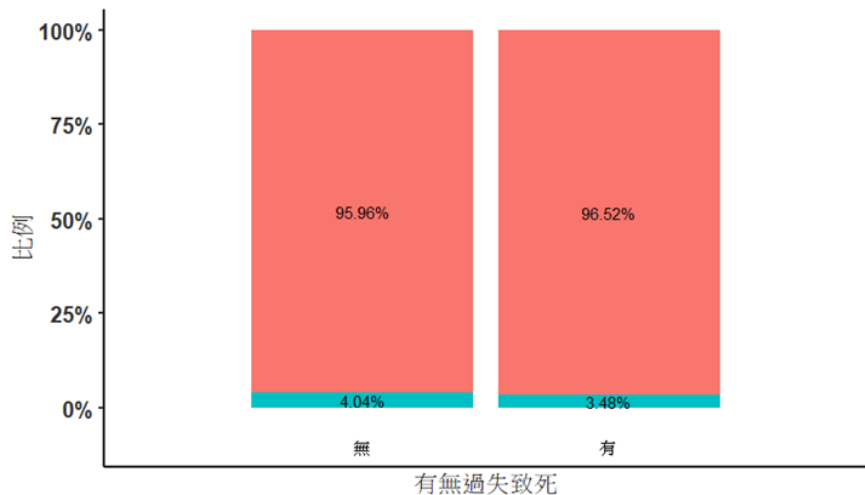
有無罪 vs. 有無過失傷害(依有無罪比例)



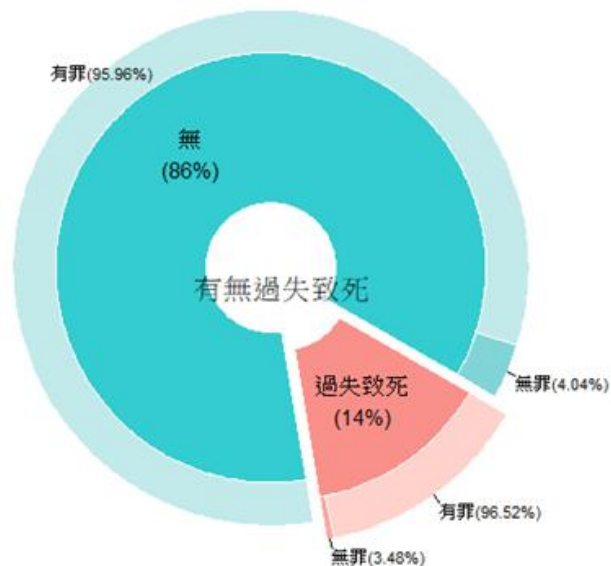


EDA-有無過失致死

在有無過失致死下有無罪相對比例長條圖



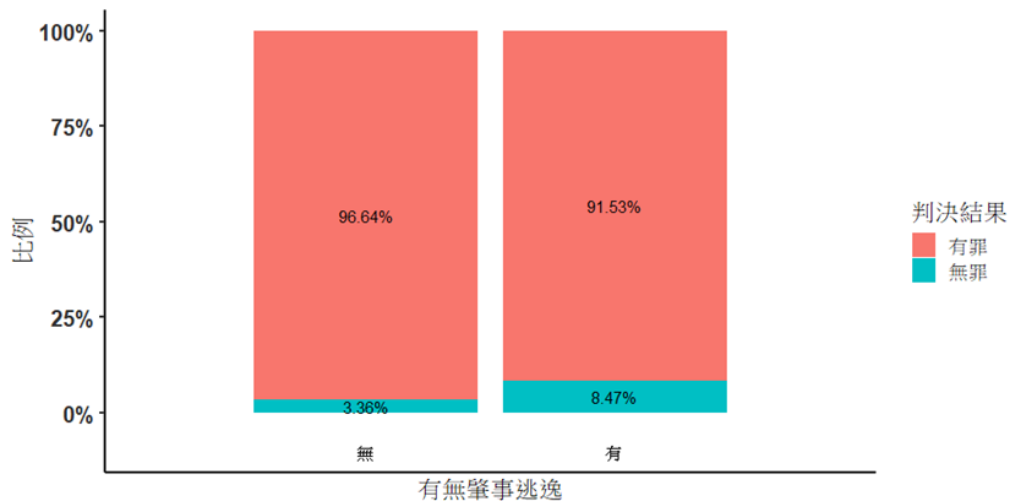
有無罪 vs. 有無過失致死(依有無罪比例)



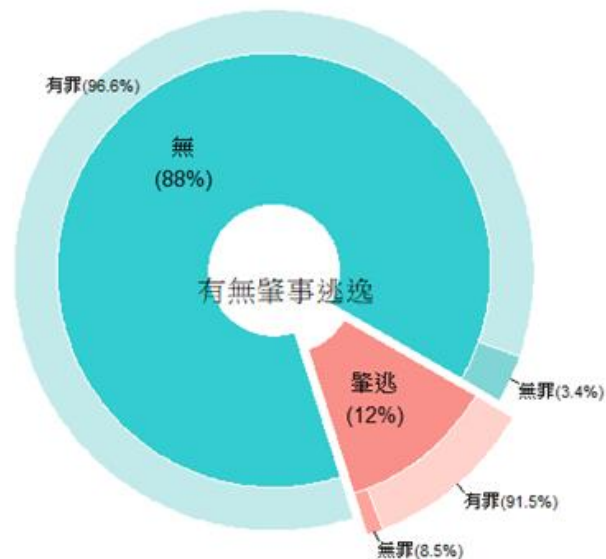


EDA-有無肇事逃逸

在有無肇事逃逸下有無罪相對比例長條圖



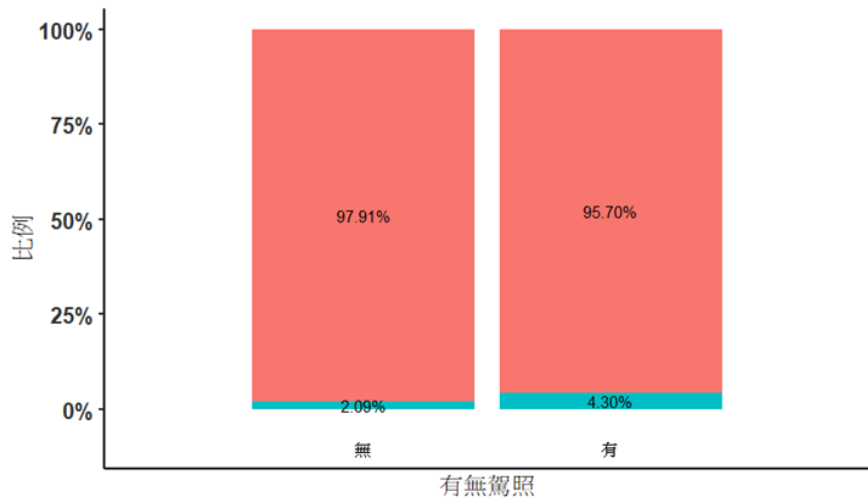
有無罪 vs. 有無肇事逃逸(依有無罪比例)



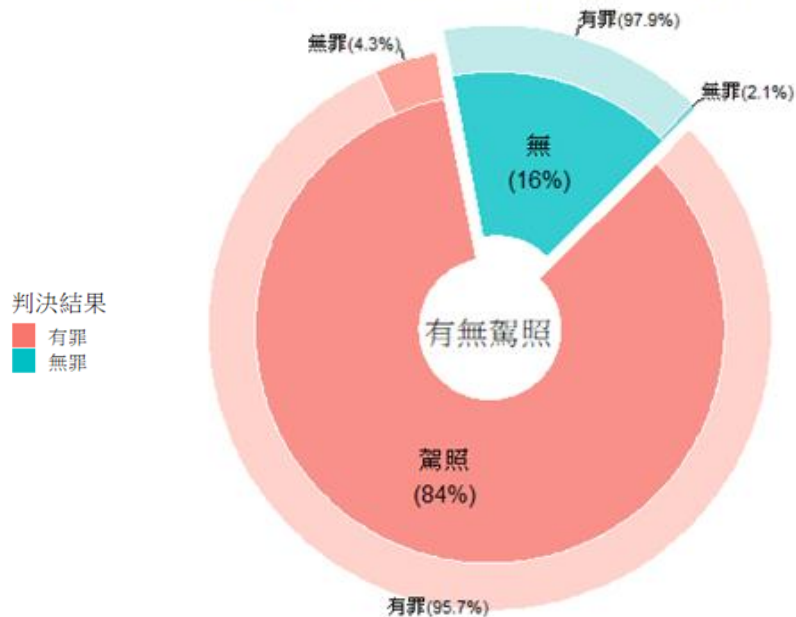


EDA-有無駕照

在有無駕照下有無罪相對比例長條圖



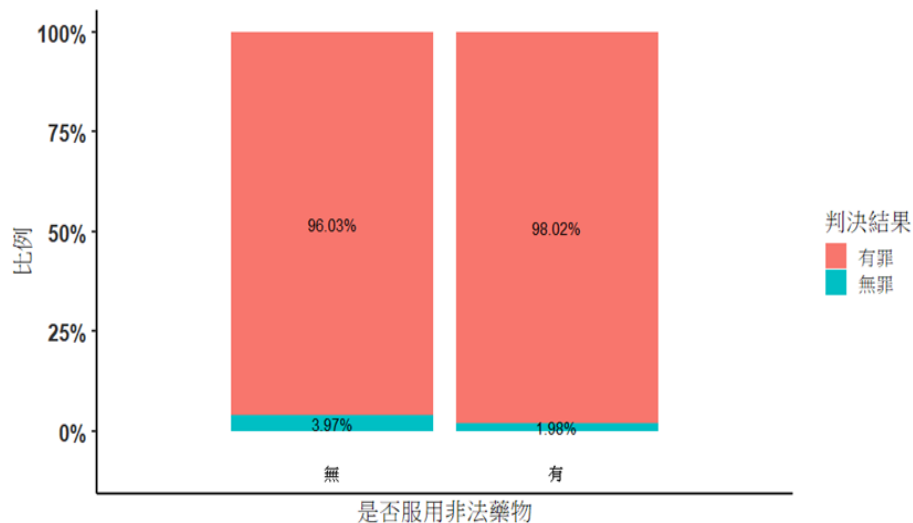
有無罪 vs. 有無駕照(依有無罪比例)



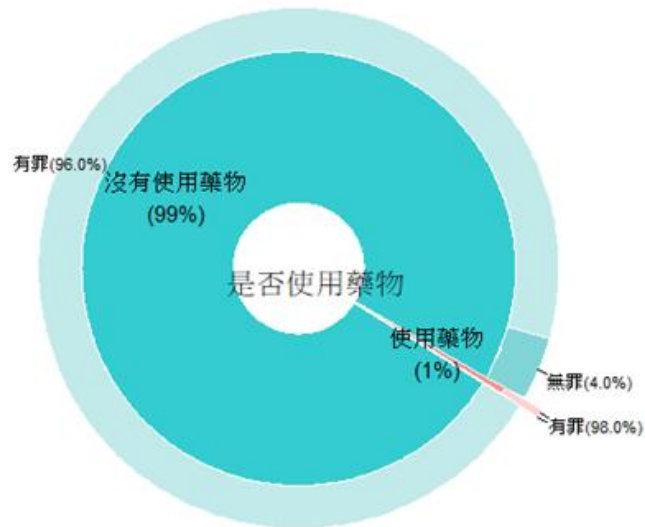


EDA-是否有服用非法藥物

在是否服用非法藥物下有無罪相對比例長條圖



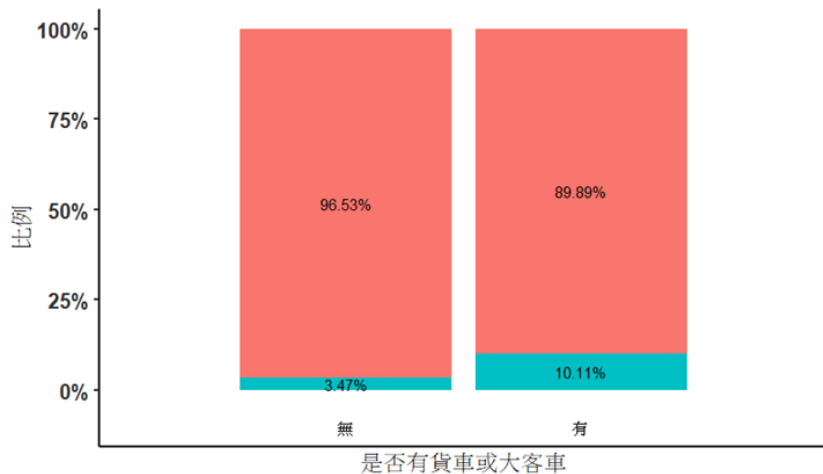
有無罪 vs. 是否服用非法藥物(依有無罪比例)



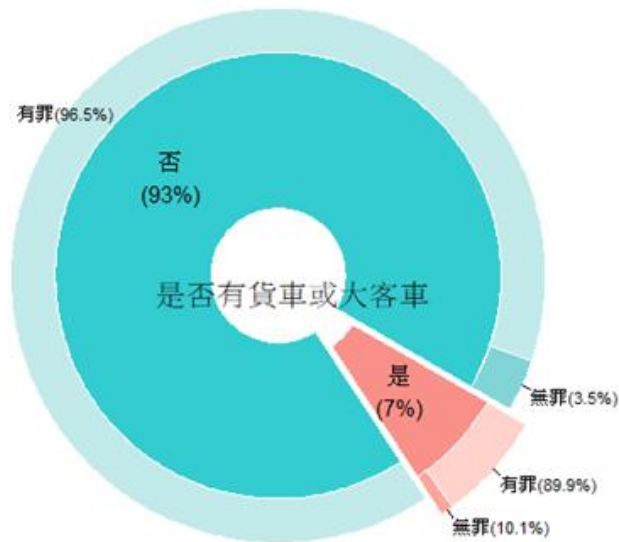


EDA-是否有貨車或大客車

在是否有貨車或大客車下有無罪相對比例長條圖



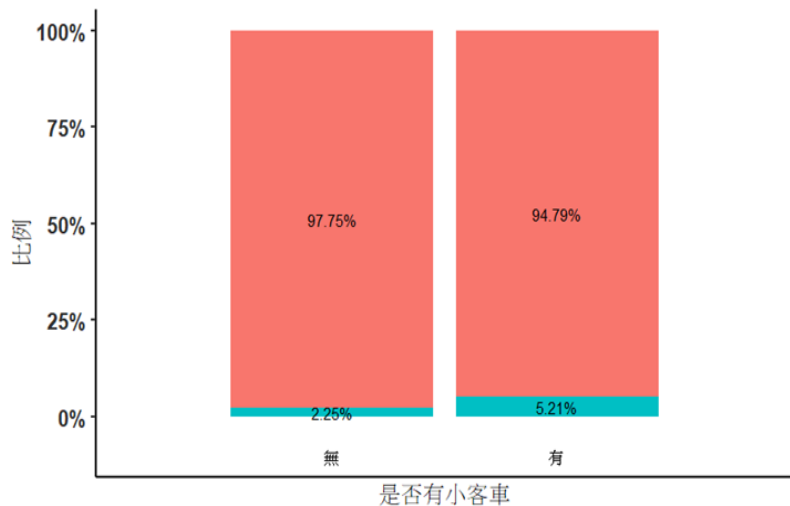
有無罪 vs. 是否有貨車或大客車(依有無罪比例)



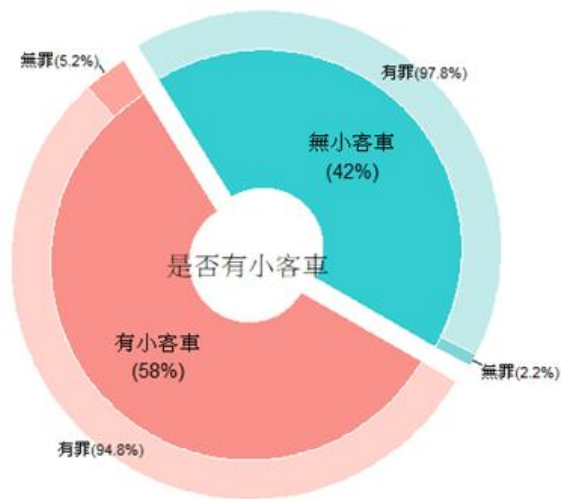


EDA-是否有小客車

在是否有小客車下有無罪相對比例長條圖



有無罪 vs. 是否有小客車(依有無罪比例)

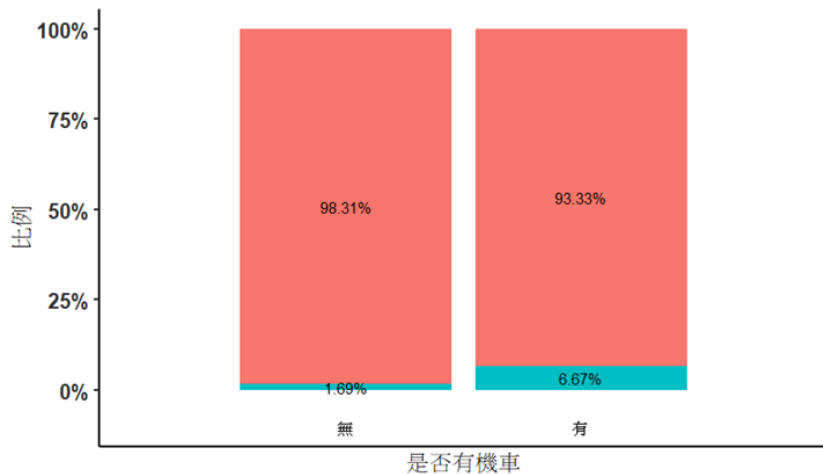




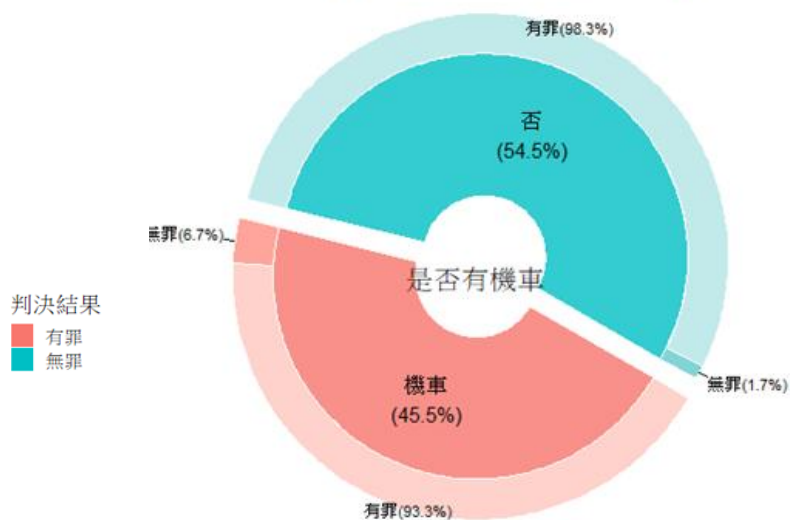
資料處理

EDA-是否有機車

在是否有機車下有無罪相對比例長條圖



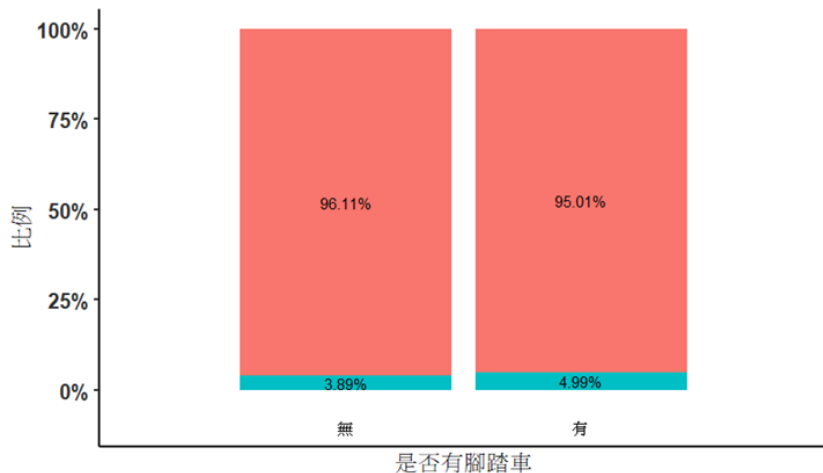
有無罪 vs. 是否有機車(依有無罪比例)



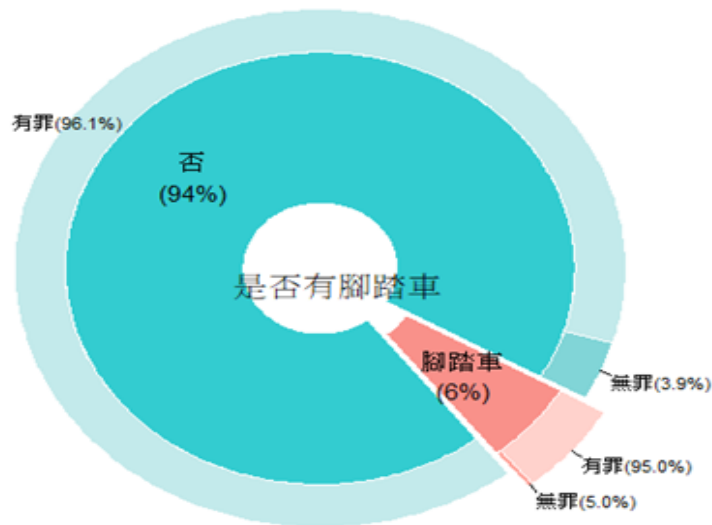


EDA-是否有腳踏車

在是否有腳踏車下有無罪相對比例長條圖



有無罪 vs. 是否有腳踏車(依有無罪比例)





模型建立

- ❖ 變數介紹
- ❖ Logistic Regression
 - 羅吉斯回歸模型
- ❖ Decision Tree
 - 決策樹模型
 - 重要變數
 - 混淆矩陣



模型建立

變數介紹:

自變數	應變數
<div>有無酒駕</div> <div>有無小客車</div> <div>有無駕照</div> <div>有無肇事逃逸</div> <div>有無減速</div> <div>有無超速</div> <div>有無精神疾病</div> <div>有無證據</div> <div>有無過失傷害</div> <div>決結果</div> <div>有無過失致死</div> <div>是否累犯</div> <div>是否自首</div> <div>是否使用藥物</div> <div>是否有貨車或大客車</div> <div>是否有機車</div> <div>是否有腳踏車</div> <div>1,是</div> <div>0,否</div>	<div>1,有罪</div> <div>0,無罪</div> <div>判</div>



模型建立

Logistic Regression model

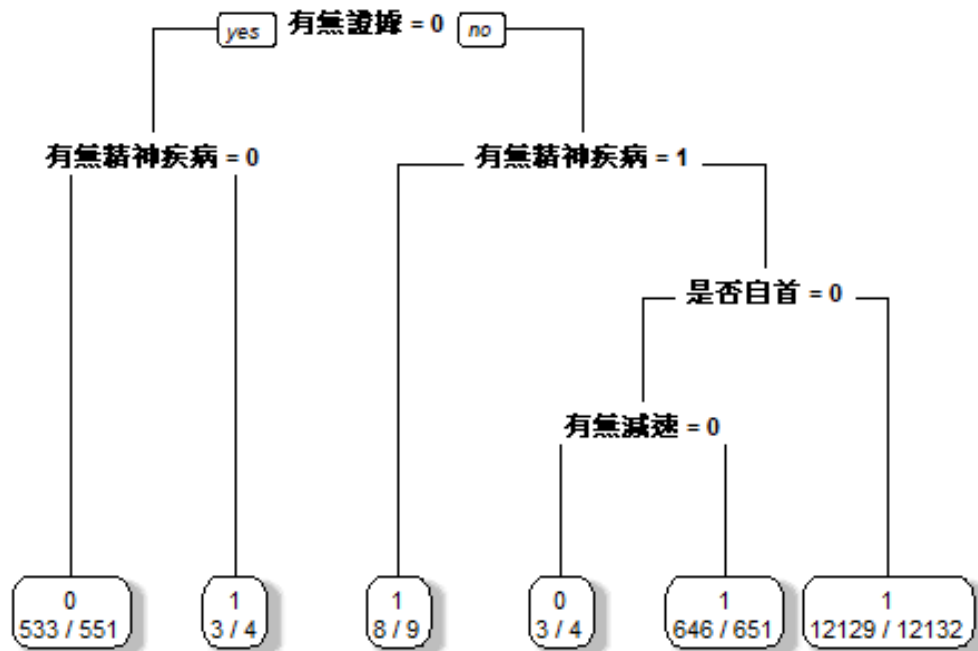
原變數有16個，經過向後篩選變數後，刪除了其中的7個，分別為有無酒駕、有無駕照、有無超速、有無過失傷害、有無過失致死、有無使用藥物，是否有小客車，並且得到以下模型。

$$\begin{aligned}\log(\text{odds}) = & -7.06 - 1.04X_{\text{肇事逃逸}} + 1.32X_{\text{減速}} + 4.3X_{\text{精神疾病}} + 11.36X_{\text{證據}} \\ & + 4.48X_{\text{累犯}} + 4.06X_{\text{自首}} - 1.83X_{\text{貨車大客車}} - 1.22X_{\text{機車}} - 2.42X_{\text{腳踏車}}\end{aligned}$$



模型建立

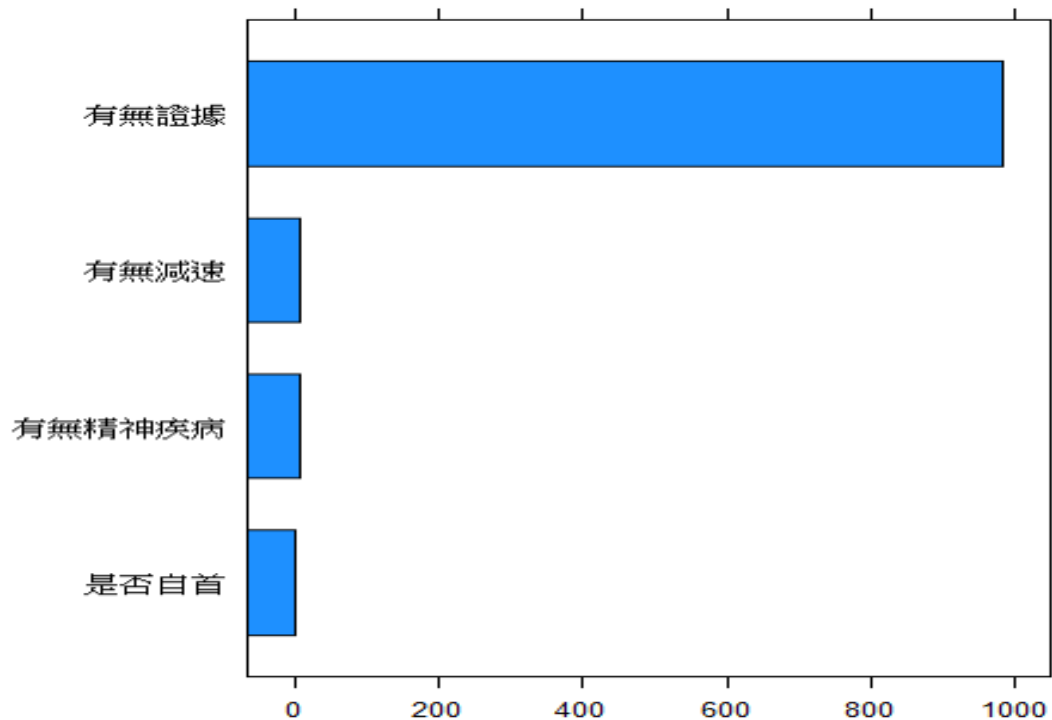
Decision Tree





模型建立

Decision Tree (variable importance)

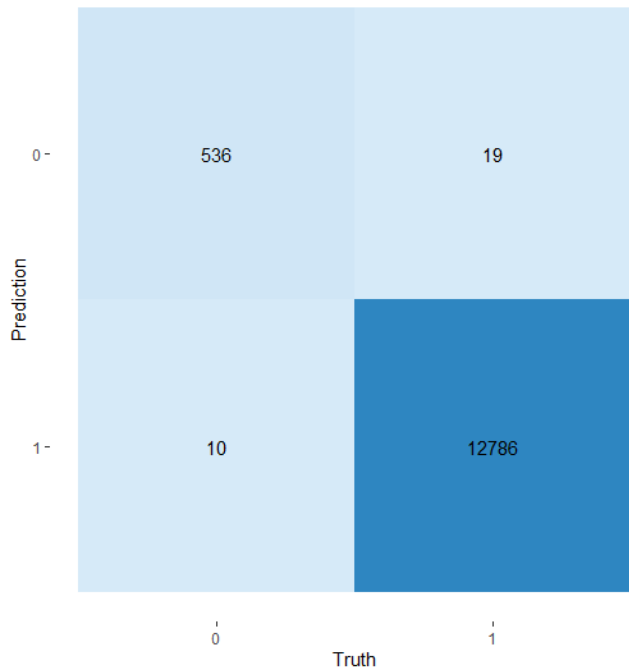




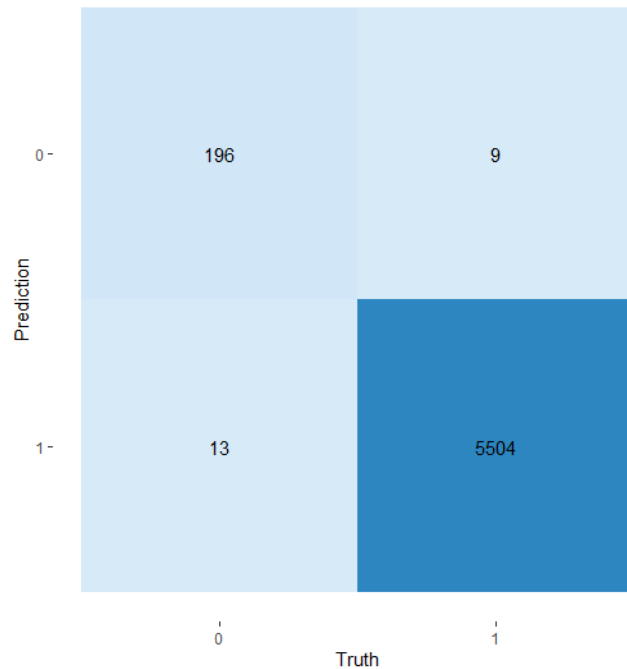
模型建立

Decision Tree (confusion matrix)

In-Sample



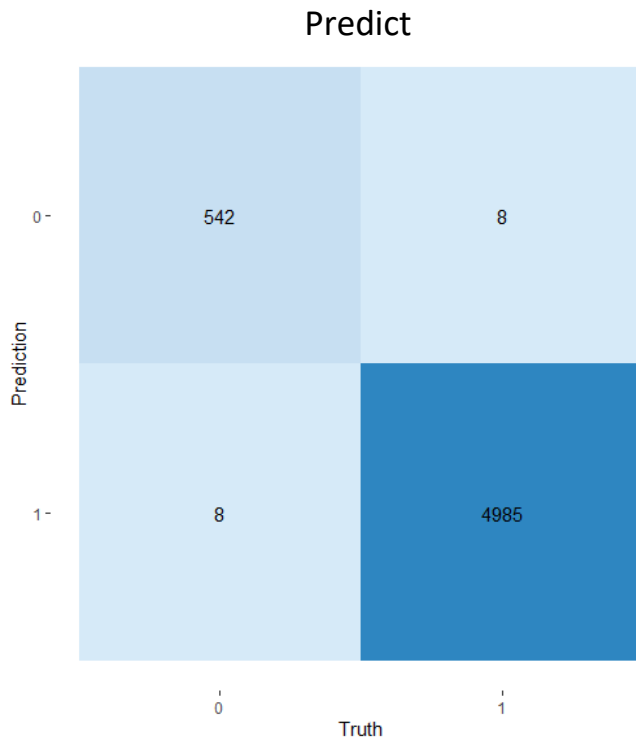
Out-Sample





模型建立

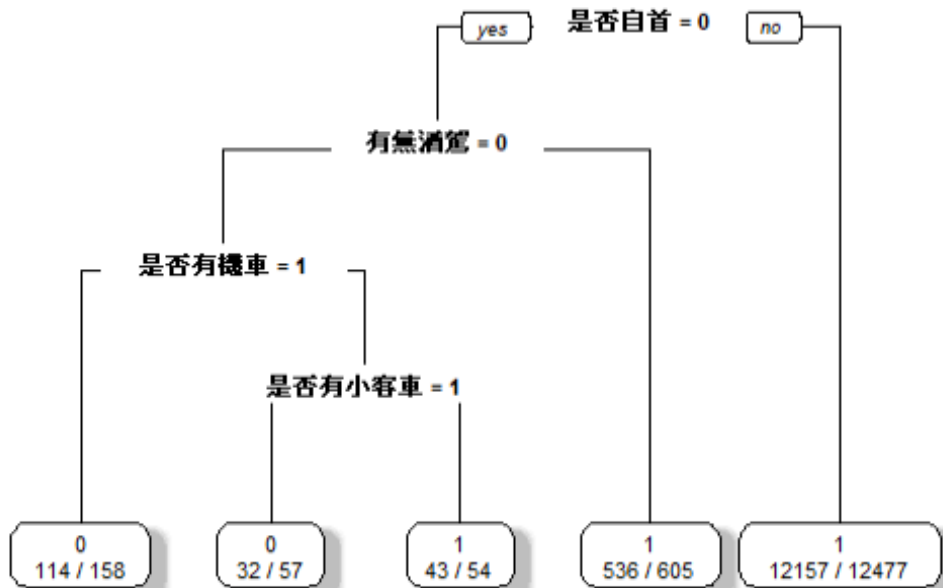
Decision Tree (confusion matrix)





模型建立

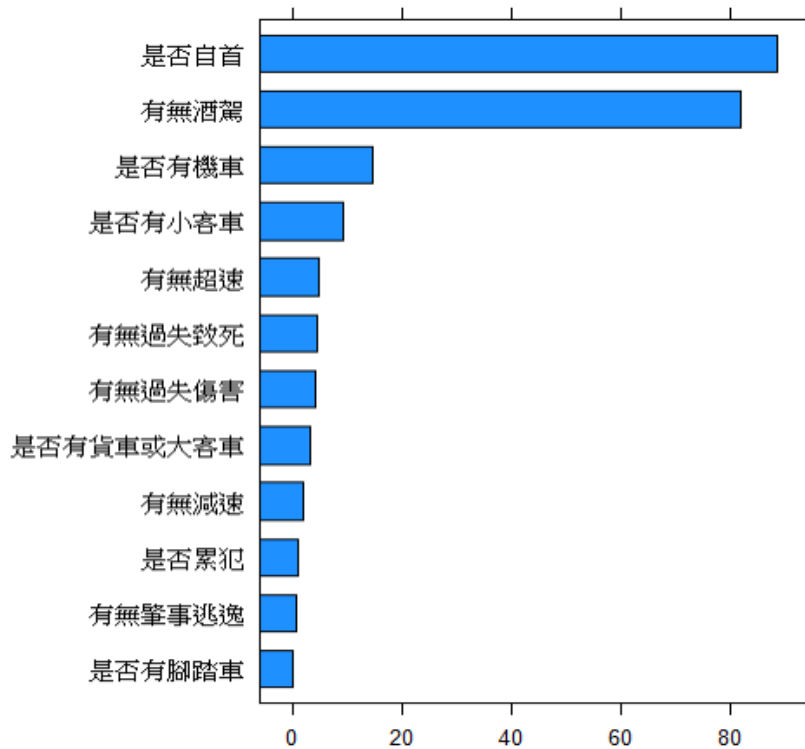
刪除有無證據變數後的 Decision Tree





模型建立

刪除有無證據變數後的 Decision Tree (variable importance)

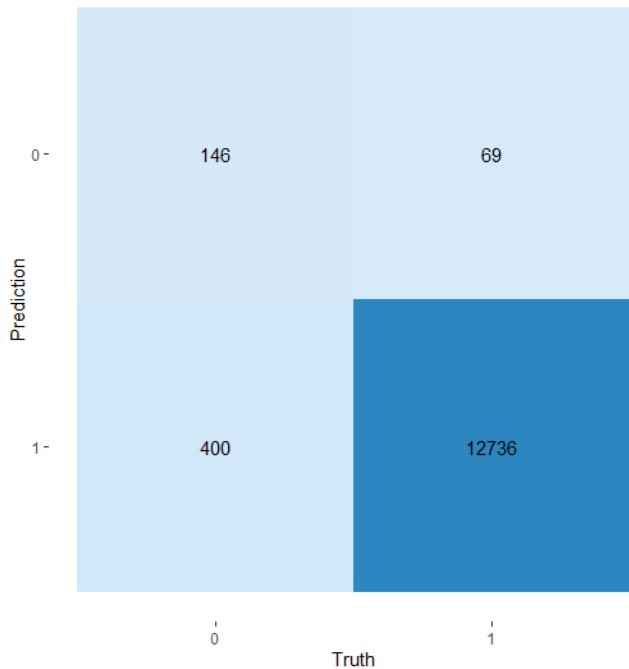




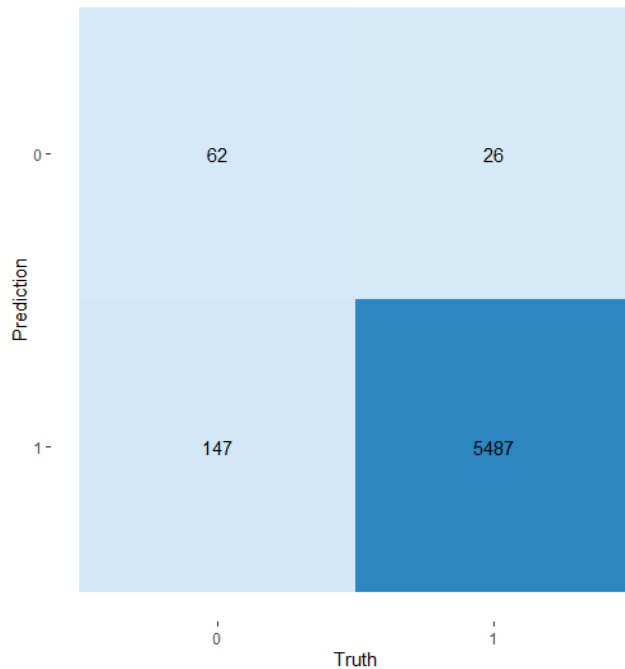
模型建立

刪除有無證據變數後的 Decision Tree

In-Sample



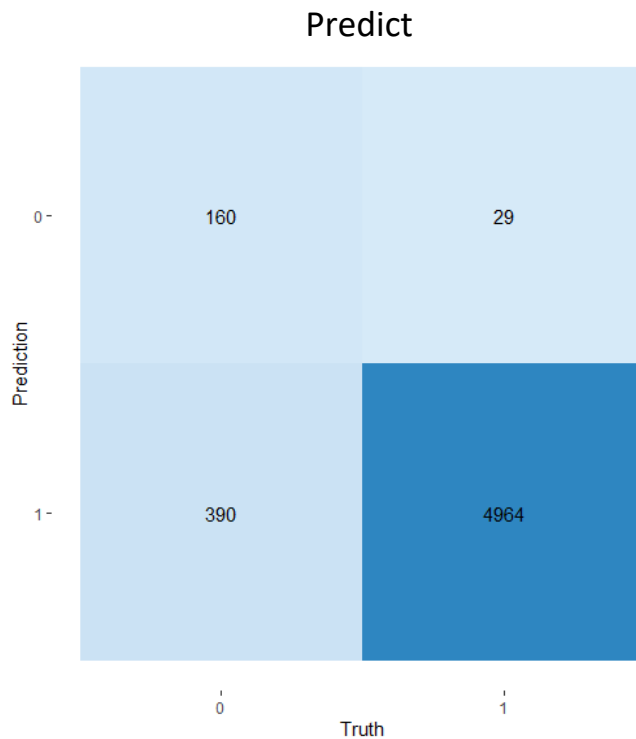
Out-Sample





模型建立

刪除有無證據變數後的 Decision Tree





系統介面



- ❖ Shiny簡介：
 - 它是一個以R語言為主，建立互動式網頁的套件。
 - 使用者於使用者介面(UI)輸入參數後，這些參數會輸入到伺服器(Server)去做運算，再把計算後的結果回傳到使用者介面給使用者看。
- ❖ 判決預測系統使用：
 - 律師根據客戶狀況，點選系統中的表單，即能預測此判決被告有無罪和其機率，以協助律師決定是否接案。
- ❖ Decision tree：
 - model1：自變數為前面提及的所有自變數(共15個)
 - model2：自變數為前面提及的所有變數但不包含有無證據(共14個)



系統介面

交通事故判決書結果預測

model1(包含有無證據選項)

model2(不包含有無證據選項)

預測結果為：

88.89% 有罪

有無酒駕

☐ 無

☒ 有

有無駕照

☐ 無

☒ 有

有無肇事逃逸

☒ 無

☐ 有

有無減速

☒ 無

☐ 有

有無超速

☒ 無

☐ 有

有無精神疾病

☐ 無

☒ 有

有無證據

☐ 無

☒ 有

有無過失傷害

☒ 無

☐ 有

有無過失致死

☒ 無

☐ 有

是否累犯

☐ 否

☒ 是

是否自首

☐ 否

☒ 是

是否使用藥物

☒ 否

☐ 是

是否有小客車

☒ 否

☐ 是

是否有貨車或大客車

☒ 否

☐ 是

是否有機車

☐ 否

☒ 是

是否有腳踏車

☒ 否

☐ 是

❖ https://yoyoman.shinyapps.io/final_report_ver4/

A wooden gavel with a light-colored head and a darker handle rests on a matching wooden block. The background is a blurred image of a scale of justice and a stack of papers, suggesting a legal or judicial setting. The entire scene is overlaid with a semi-transparent grey rectangle.

Thanks !