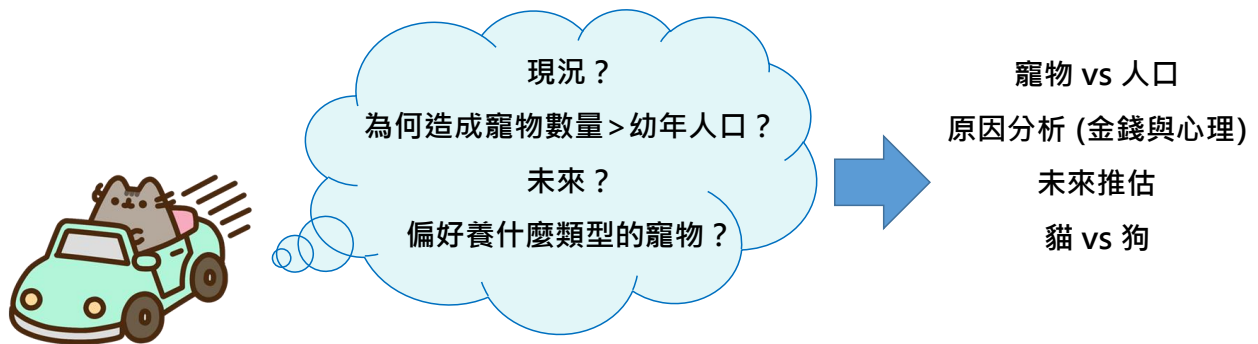


台灣寵物數量分析

李旻恩 2022/08/25

一、研究動機：

- ◆ 自己家裡本身就養了幾隻寵物，周圍養寵物的人數也不少，衍生想了解台灣飼養寵物的夥伴數量，在網路上進行相關資訊的查詢。
- ◆ 根據 2018 年內政部、農委會與時勢公司整理分析，0~14 歲幼年人口與全台犬貓數量於 2020 年達到黃金交叉，估計 2021 年將首度發生寵物數量比幼年人口數還多的情況。
- ◆ 因以上原因而進行台灣寵物數量的分析，並產生了以下疑惑…

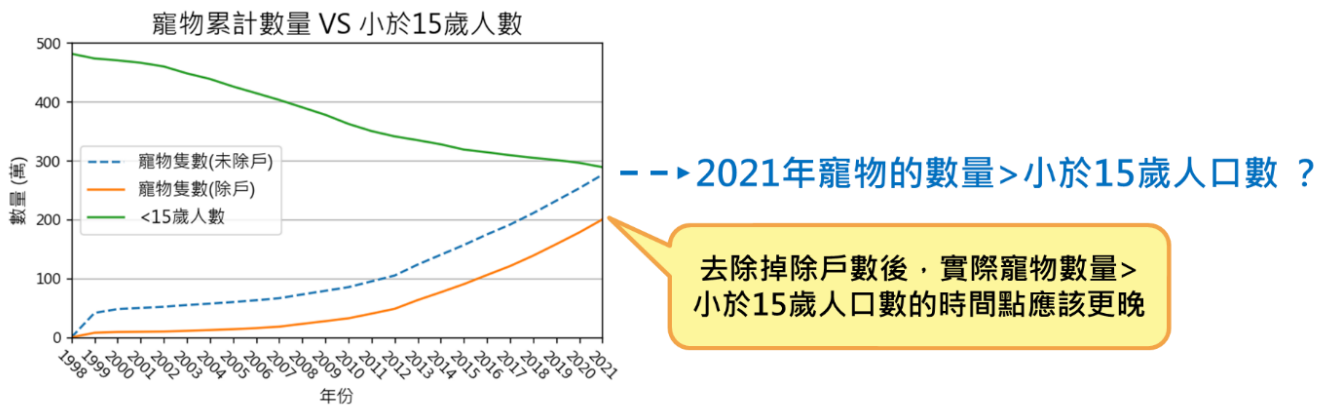
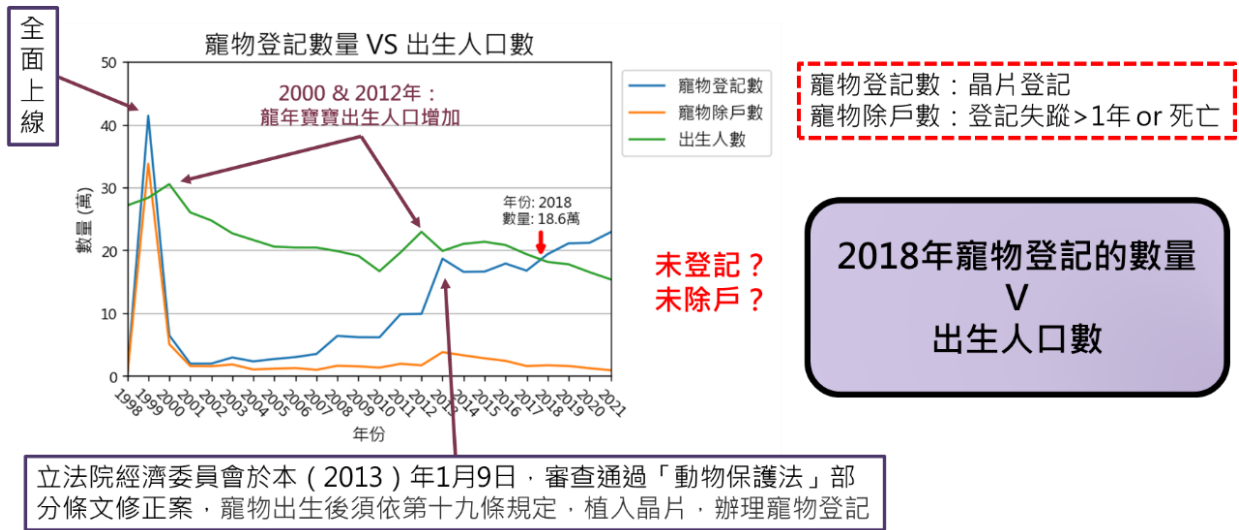


二、資料概述：

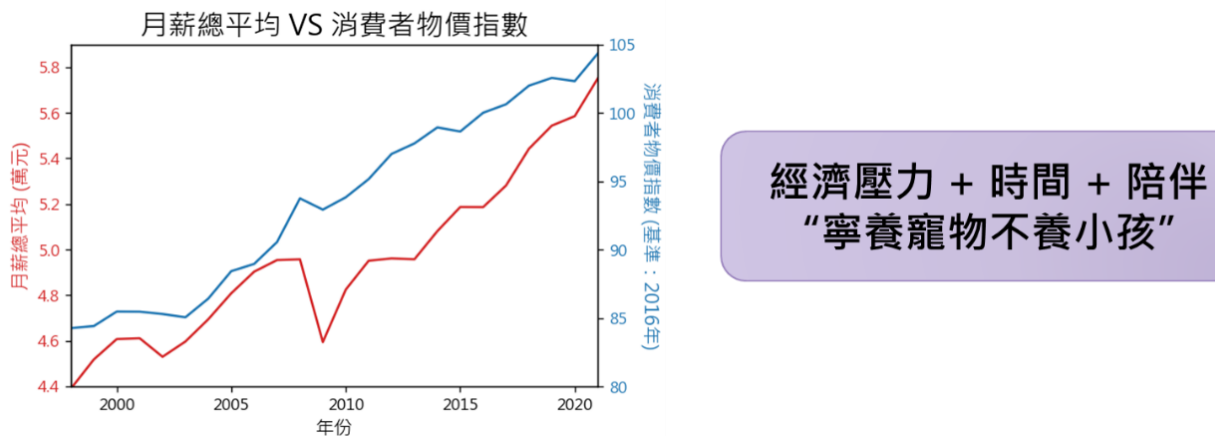
- 寵物登記數與除戶數 (1998~2021)
 - 寵物登記管理資訊網
- 出生人口 & 0~14 歲人口 (1998~2070)
 - 國家發展委員會 人口推估查詢系統
- 平均薪資 (1998~2021)
 - 行政院主計總處 薪情平台
- 消費者物價指數 (1998~2021)
 - 中華民國統計資訊網
- 銷售額與家數 (版本六：2008~2012，版本七：2013~2018)
 - 中華民國財政部 財政統計資料庫查詢
- 抗憂鬱藥物使用人數 (2012~2020)
 - 政府資料開放平台
- 家貓家犬數量 (2013、2015、2017、2019、2021)
 - 動物保護資訊網

三、圖表分析：

(一) 寵物與人口



(二) 原因分析



2017年

全球教育支出排行

排名	國家地區	花費(元)
1	香港	4,070,559
2	阿聯酋	3,060,842
3	新加坡	2,184,921
4	美國	1,800,691
5	台灣	1,737,859
6	中國	1,321,074
7	澳洲	1,121,182
8	馬來西亞	784,753
9	英國	765,750
10	墨西哥	702,610

備註：每位子女由小學到大學的平均花費
資料來源：HSBC，時勢公司整理。

光教育費用
~170萬



2019年
(富邦產險整理)

貓狗花費表格

各種費用	養貓花費	養狗花費
晶片費用	植入費：300元 + 登記費用	
伙食費	一年大約5,000~15,000元	
施打疫苗	三合一疫苗500~900元	幼犬：三劑五合一疫苗+狂犬病疫苗3,000~3,900元 成犬：五合一疫苗+狂犬病疫苗3,000~3,900元 心絲蟲預防200~300元/月
其他開銷	3,000~5,000元/年	2,000~5,000元/年
醫療費	不一定，500~上萬元不等	
喪葬費	2,500~50,000元	

不含生病、意外、物品損毀
50~60萬 (壽命10年)



回歸分析：(2012~2020年)

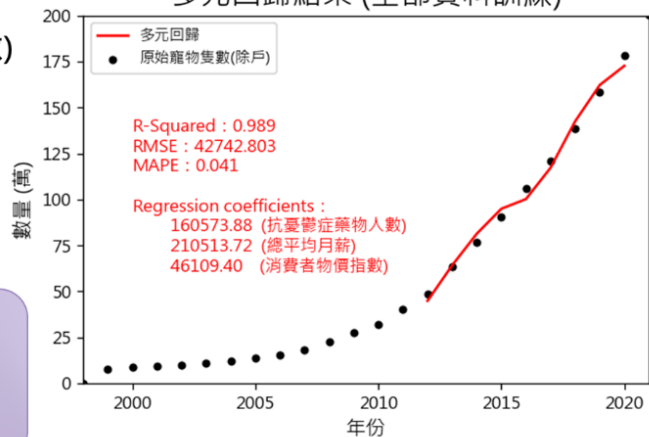
金錢 (總平均月薪，消費者物價指數)
心理 (抗憂鬱藥物人數)

1. 抗憂鬱藥物人數
2. 總平均月薪
3. 消費者物價指數

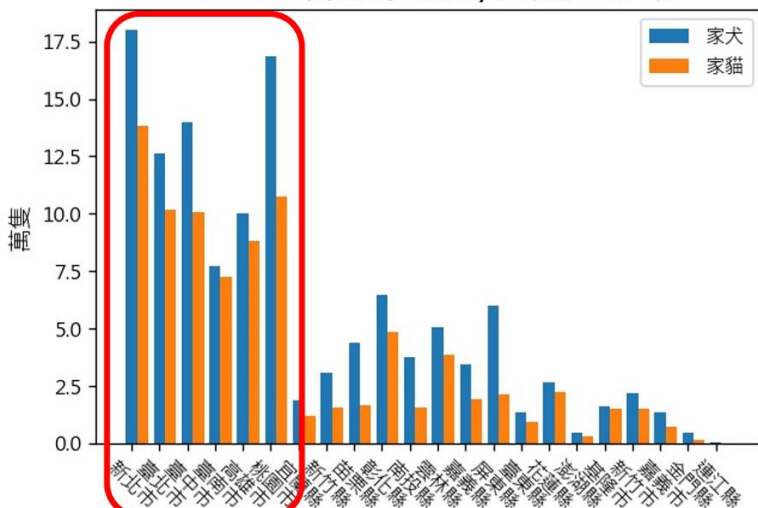
總平均月薪為最主要影響寵物數量的因素，服用
抗憂鬱藥物人數次之。

這三種自變數無未來推估資訊無法進行多元回歸預測...

多元回歸結果 (全部資料訓練)

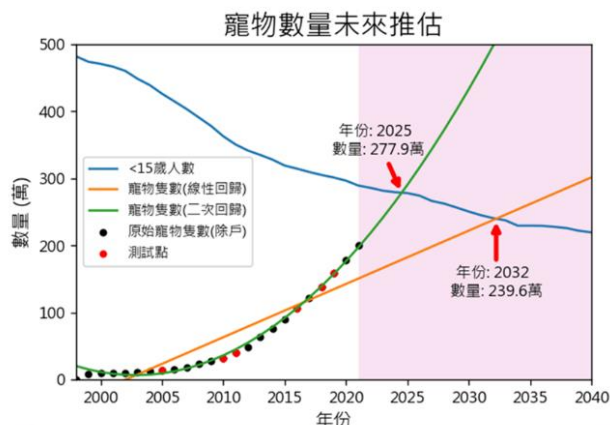


2021年各縣市家貓/犬數量分布



都市化程度較高的區域，
寵物的數量較多

(三) 未來預測



回歸預測：1999~2021年 → 預測2021年以後
(同以 20% 測試資料計算得分結果)

線性回歸：(自變數為年份)

R-Squared(解釋力)：0.853

RMSE(誤差)：212703.528(隻)

MAPE(誤差)：0.449

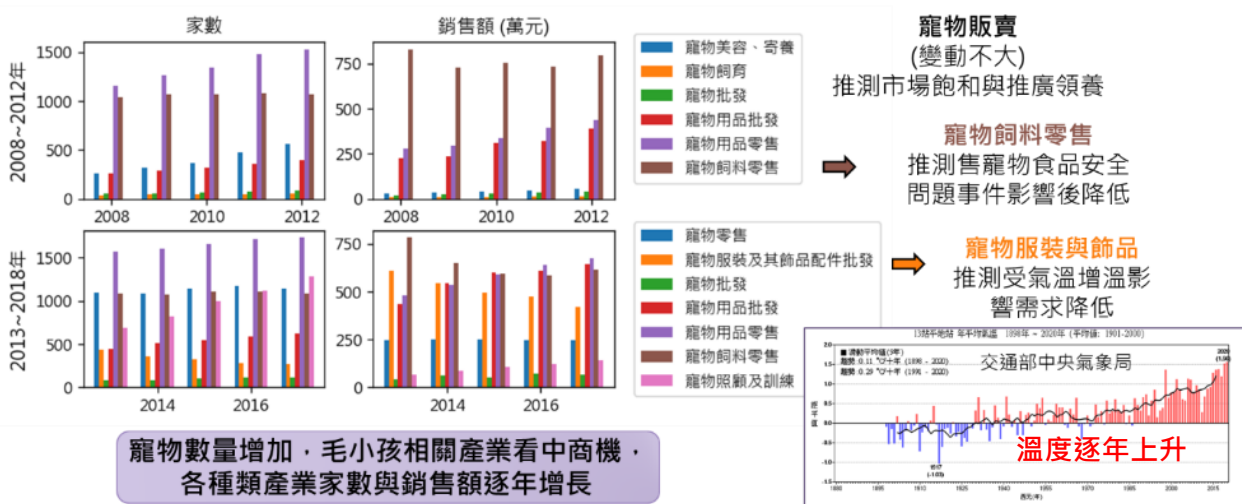
二次回歸：(自變數為年份)

R-Squared(解釋力)：0.997

RMSE(誤差)：31292.816(隻)

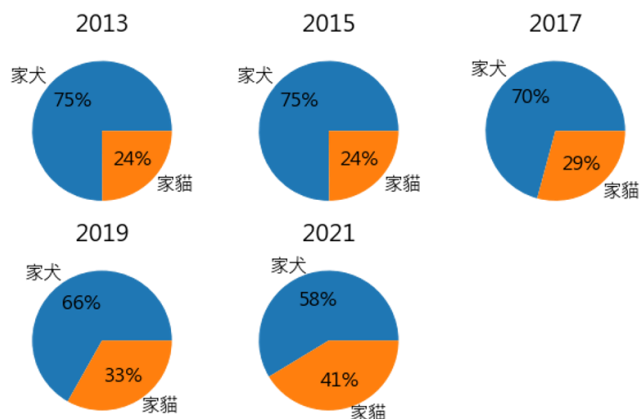
MAPE(誤差)：0.097

(四) 延伸



中央氣象局臺灣長期氣候變化全書：https://www.cwb.gov.tw/V8/C/K/Encyclopedia/climate/climate7_all.html

每年飼養貓與狗數量分布



家犬的數量 > 家貓
但飼養貓的佔筆有逐年
增長的趨勢

四、 結論：

- ◆ 登記寵物的數量隨著時間日益增長，隨著生育小孩的意願降低，在 2018 年時出生人口數就已經低於登記寵物的數量。
- ◆ 造成寧願養寵物不養小孩的原因，與經濟上的壓力高度相關，在此情況下心理上的陪伴需求增加，寵物數量逐漸升高，相關行業看準商機家數與銷售額逐年增長。
- ◆ 假設其他未考慮的因素變化率固定，以年份進行多項式回歸推估，推測未來 2025 年小於 15 歲人口將低於寵物的數量。
- ◆ 都市化程度較高的區域，寵物的數量較多，且大多數人偏好養狗，但飼養貓的佔比有逐年增長的趨勢。

五、 問題與解決：

(一) 資料蒐集 (已判斷為公開資料可取用)

Try 1：爬蟲 (失敗)



```
from bs4 import BeautifulSoup as bs
from selenium import webdriver
from selenium.webdriver.common.by import By
# from selenium.webdriver.support.ui import import time

pet_list = []

driver = webdriver.Chrome("chromedriver")
driver.get("https://www.pet.gov.tw/Web/0302.aspx?PG=1")
# 日期定位
select_start = driver.find_element(By.ID,"txtSDATE")
select_end = driver.find_element(By.ID,"txtEDATE")
year = [2020]
# 選擇日期
select_start.send_keys(str(year[0])+"/01/01")
select_end.send_keys(str(year[0])+"/12/31")
# 選擇選項"全部"
select_type = driver.find_element(By.CLASS_NAME,"circle")
select_type.click()
# 點擊搜尋
driver.find_element(By.ID,"aSearch").click()
driver.implicitly_wait(5)
# 資料解析
soup = bs(driver.page_source, "lxml")
table = soup.select_one("#CountTown")
print(table.prettify())
```

```
<bound method Tag.prettify of <table class="table table-bordered table-g13" id="CountTown">
<thead>
<tr>
<th>縣市</th>
<th>鄉鎮</th>
<th>登記單位數</th>
<th>登記數(A)</th>
<th>除戶數(B)</th>
<th>轉讓數(C)</th>
<th>變更數(D)</th>
<th>絕育數(E)</th>
<th>絕育除戶數(F)</th>
<th>免絕育數(G)</th>
<th>免絕育除戶數(H)</th>
<th>絕育率</th>
<th>繁殖管理率</th>
</tr>
</thead>
<tbody>
</tbody>
</table>>
```

無法直接檢視所以存取網頁原始碼

```
42 with open("test.html","w", encoding="utf8") as file:
43     file.write(soup.prettify())
```

```
for (var key in json){
    dal.Ini(json[key]);
    S("#CountTown").append("<tr id='"+key+"'>")
    + "<td data-th='縣市'><a href='void(0)'; class='CountySpan' id='"+dal.ArealD + "'> " + dal.AreaName + "</a></td>"
    + "<td data-th='鄉鎮'> " + dal.AreaName + "</td>"
    + "<td data-th='登記單位數'> " + dal.fld01 + "</td>"
    + "<td data-th='登記數(A)'> " + dal.fld02 + "</td>"
    + "<td data-th='除戶數(B)'> " + dal.fld03 + "</td>"
    + "<td data-th='轉讓數(C)'> " + dal.fld04 + "</td>"
    + "<td data-th='變更數(D)'> " + dal.fld05 + "</td>"
    + "<td data-th='絕育數(E)'> " + dal.fld06 + "</td>"
    + "<td data-th='絕育除戶數(F)'> " + dal.fld07 + "</td>"
    + "<td data-th='免絕育數(G)'> " + dal.fld08 + "</td>"
    + "<td data-th='免絕育除戶數(H)'> " + dal.fld09 + "</td>"
    + "<td data-th='絕育率(E-F)/(A-B)'> " + (((dal.fld04 - dal.fld03) / (dal.fld02 - dal.fld03)) * 100.0).toFixed(2) + "</td>"
    + "<td data-th='繁殖管理率(E-F)/(G-H)/(A-B)'> " + (((dal.fld04 - dal.fld03) / (dal.fld07 - dal.fld08)) * 100.0).toFixed(2) + "</td>"
    + "</tr>";
    dal.fld01 += parseInt(dal.fld01);
    dal.fld02 += parseInt(dal.fld02);
    dal.fld03 += parseInt(dal.fld03);
    dal.fld04 += parseInt(dal.fld04);
    dal.fld05 += parseInt(dal.fld05);
    dal.fld06 += parseInt(dal.fld06);
    dal.fld07 += parseInt(dal.fld07);
    dal.fld08 += parseInt(dal.fld08);
    dal.fld09 += parseInt(dal.fld09);
}
```

但還是包在script中無解

Try 2：截圖轉表格（失敗）

```
from selenium import webdriver
from PIL import Image
import time
from selenium.webdriver.common.by import By

driver = webdriver.Chrome("chromedriver")
driver.get("https://www.pet.gov.tw/Web/O302.aspx?PG=1")

year = list(range(1998, 2022))
# year = [1998]
def get_pic(year, path):
    for i in range(len(year)):
        print(year[i])
        filename = path+str(year[i])+".png"
        select_start = driver.find_element(By.ID, "txtSDATE")
        select_end = driver.find_element(By.ID, "txtEDATE")
        date = str(year[i])+"/01/01"
        select_start.send_keys(date)
        date = str(year[i])+"/12/31"
        select_end.send_keys(date)

        driver.find_element(By.ID, "aSearch").click()
        time.sleep(10)
        # 調整式窗大小與比例
        driver.execute_script("document.body.style.zoom='0.5';")
        driver.execute_script("var action=document.documentElement.scrollTop=350")
        # 截圖
        # driver.get_screenshot_as_file("2330.png")
        driver.get_screenshot_as_file(filename)
        webpage=Image.open(filename)
        # image_crop=webpage.crop(box=(left, upper, right, bottom))
        # 截圖範圍擷取
        image_crop=webpage.crop(box=(15, 90, 1025, 750))
        # 存截圖結果
        image_crop.save(filename)
        image_crop=Image.open(filename)
        # 調回式窗比例（避免定位查詢會出現問題）
        driver.execute_script("document.body.style.zoom='1';")
        # 清空時間選項
        select_start.clear()
        select_end.clear()
        time.sleep(10)
```

```
for j in driver.find_elements(By.CLASS_NAME, "form-check-label"):
    # print(i.text)
    if j.text == "全部":
        print(format(" 全部 ", "^30"))
        j.click()
        path = "./data/pic/Pet_All_"
        get_pic(year, path)
    elif j.text == "狗":
        print(format(" 狗 ", "^30"))
        j.click()
        path = "./data/pic/Pet_Dog_"
        get_pic(year, path)
    elif j.text == "貓":
        print(format(" 貓 ", "^30"))
        j.click()
        path = "./data/pic/Pet_Cat_"
        get_pic(year, path)

driver.quit()
```

截圖結果

縣市	全部	狗	貓	全部	狗	貓	全部	狗	貓
臺北市	352	30242	1141	23945	45942	15641	777	192	9
臺中市	239	18428	517	14545	45881	11528	163	151	4
臺南市	256	32184	1138	19382	32191	23088	422	804	12
高雄市	184	18221	468	11227	10259	9791	95	685	14
基隆市	243	29743	1448	19570	22129	17979	814	687	16
台北市	168	19187	533	11348	28807	10337	110	70	4
宜蘭縣	25	4077	123	2648	5977	2191	39	214	8
新竹縣	30	5129	161	3791	6477	3084	31	130	4
苗栗縣	31	5880	388	3874	7356	3247	44	29	0
彰化縣	78	14638	724	8491	8718	10016	468	168	2
南投縣	20	7981	262	2648	3442	6318	83	176	6
雲林縣	41	8334	507	3410	7269	6233	416	186	5
嘉義縣	14	5444	349	3171	2706	3132	32	54	0
屏東縣	93	7812	272	5786	5332	4302	78	41	2
嘉義市	7	4919	601	1191	2405	3270	824	46	2
花蓮縣	33	5510	62	1685	2451	3474	34	16	0
台東縣	10	1055	35	730	1289	488	12	89	6
澎湖縣	17	3816	139	2412	4059	2418	5	17	1
金門縣	37	3847	154	2644	2782	1611	18	41	1
連江縣	33	1829	92	1630	1790	976	67	10	1
合計	1	1177	46	834	2233	706	22	268	14
總計	1	108	6	75	281	50	3	0	0
合計	1546	228961	9286	150887	252173	139880	3655	4074	111

先嘗試用現有圖片轉表格方式，結果差(數字->英文)
手動輸入每年合計資料，截圖方便確認資料...

未來可嘗試用深度學習模型來判斷表格內容...

(二) 資料清理

營利事業家數及銷售額 - 第6次修訂.csv

營利事業家數及銷售額 - 第7次修訂.csv

統計期	行業別	家數總計	銷售額總計
99年	4511-1 購物用品批發	84	402368
99年	4552-1 購物用品批發及零售品配件批發	436	6078567
99年	4699-2 購物用品批發	445	4390445
99年	4852-2 購物用品零售	1096	2463810
99年	4852-2 購物用品零售	1569	4761100
99年	4852-2 購物用品零售	1081	7821744
99年	9630-1 購物用品批發	96	553480
99年	9690-2 購物用品批發及零售	86	598709
99年	4511-1 購物用品批發	86	598709
99年	4552-1 購物用品批發及零售品配件批發	356	5431603
99年	4699-2 購物用品批發	507	5464693
99年	4852-2 購物用品零售	1085	2521793
99年	4852-2 購物用品零售	1597	5365566
99年	4852-2 購物用品零售	1069	6509259
99年	9630-1 購物用品批發	816	844957
99年	9690-2 購物用品批發及零售	99	530988
99年	4511-1 購物用品批發	99	530988
99年	4552-1 購物用品批發及零售品配件批發	723	4951777
99年	4699-2 購物用品批發	547	5585186
99年	4852-2 購物用品零售	1141	2490745
99年	4852-2 購物用品零售	1653	5914829
99年	4852-2 購物用品零售	1105	5985121
99年	9630-1 購物用品批發	3 (D)	
99年	9690-2 購物用品批發及零售	990	1055152
99年	4511-1 購物用品批發	110	779768
99年	4552-1 購物用品批發及零售品配件批發	281	4746615
99年	4699-2 購物用品批發	587	6080885
99年	4852-2 購物用品零售	1172	2462470
99年	4852-2 購物用品零售	1715	6404629
99年	4852-2 購物用品零售	1102	5847288
99年	9630-1 購物用品批發	3 (D)	
99年	9690-2 購物用品批發及零售	1119	1216653
99年	4511-1 購物用品批發	110	689198
99年	4552-1 購物用品批發及零售品配件批發	265	4195136
99年	4699-2 購物用品批發	615	6470263
99年	4852-2 購物用品零售	1135	2498506
99年	4852-2 購物用品零售	1728	6753887
99年	4852-2 購物用品零售	1086	6138028
99年	9630-1 購物用品批發	5 (D)	
99年	9690-2 購物用品批發及零售	1284	1437153

兩個版本分開處理 (年份，排序，清理，重排，型態)

```
##% 6. 銷售額與家數 (六) (七)
sale6 = pd.read_csv(r"./data/營利事業家數及銷售額-第6次修訂.csv", encoding=("ANSI"))
sale6.dropna(inplace=True)
sale6["統計期"] = sale6["統計期"].str.replace("年", "").astype(int)+1911
sale6.sort_values(["行業別", "統計期"], inplace=True)
sale6.replace("D", np.nan, inplace=True)
sale6.dropna(inplace=True)
sale6.reset_index(drop=True, inplace=True)
sale6["<b>家數總計</b>"] = sale6["<b>家數總計</b>"].astype(int)
sale6["<b>銷售額總計</b>"] = sale6["<b>銷售額總計</b>"].astype(int)
sale7 = pd.read_csv(r"./data/營利事業家數及銷售額-第7次修訂.csv", encoding=("ANSI"))
sale7.dropna(inplace=True)
sale7["統計期"] = sale7["統計期"].str.replace("年", "").astype(int)+1911
sale7.sort_values(["行業別", "統計期"], inplace=True)
sale7.replace("D", np.nan, inplace=True)
sale7.replace("-", np.nan, inplace=True)
sale7.dropna(inplace=True)
sale7.reset_index(drop=True, inplace=True)
sale7["<b>家數總計</b>"] = sale7["<b>家數總計</b>"].astype(int)
sale7["<b>銷售額總計</b>"] = sale7["<b>銷售額總計</b>"].astype(int)
```


(三) 回歸處理 (多元, 一次, 二次)

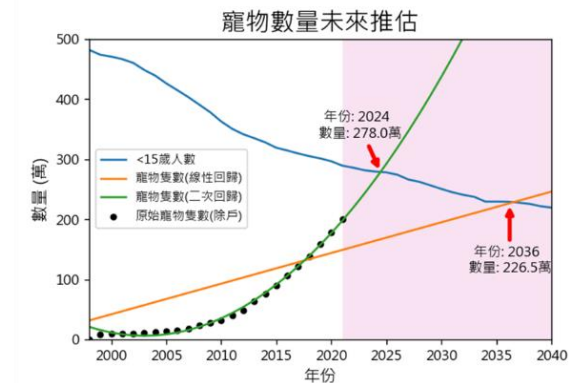
自變數(X) →	抗憂鬱症藥物人數	總平均月薪	消費者物價指數
回歸係數 (未標準化)	1.10	99.84	22858.53
回歸係數 (標準化)	160573.88	210513.72	46109.40

[illegible]

(四)繪圖問題 - 交點、方形區域、文字顯示

```
# %% function
# 兩兩鏈的交點計算 (應有更好的方法)
def cross_point(line1,line2):
    # 取四點座標
    x1=line1[0]; y1=line1[1]; x2=line1[2]; y2=line1[3]
    x3=line2[0]; y3=line2[1]; x4=line2[2]; y4=line2[3]

    k1=(y2-y1)*1.0/(x2-x1) # 計算k1, 避免點均為整數, 需要進行浮點數轉化
    b1=y1*1.0-x1*k1*1.0 # 整型轉浮點型是關鍵
    if (x4-x3)==0: # L2 直線斜率不存在在操作
        k2=None
        b2=0
    else:
        k2=(y4-y3)*1.0/(x4-x3)#斜率存在在操作
        b2=y3*1.0-x3*k2*1.0
    if k2==None:
        x=x3
    else:
        x=(b2-b1)*1.0/(k1-k2)
    y=k1*x*1.0+b1*1.0
    return [x,y]
```

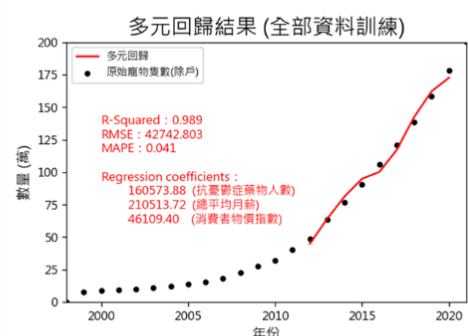


```
plt.title("寵物數量未來推估", fontsize=18)
plt.broken_barh([(2021, 2040)], (0, 500), facecolors='tab:pink', alpha=0.2)
plt.plot(range(1998, 1998+len(population.iloc[0])-1), population.iloc[0,1]*1000/10000, label="＜15歲人數")
plt.plot(year, pet_lm_predict_tt/10000, label="寵物隻數(線性回歸)")
plt.plot(year, pet_lm2_predict_tt/10000, label="寵物隻數(二次回歸)")
```

```
t = f'''
R-Squared: {metrics.r2_score(y.loc[14-dy:22-dy], lm3_tt.predict(x_new2)):.3f}
RMSE: {(metrics.mean_squared_error(y.loc[14-dy:22-dy], lm3_tt.predict(x_new2)))**0.5:.3f}
MAPE: {metrics.mean_absolute_percentage_error(y.loc[14-dy:22-dy], lm3_tt.predict(x_new2)):.3f}

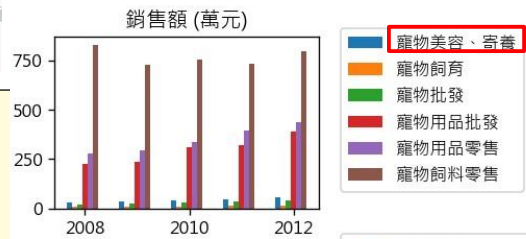
Regression coefficients:
    {lm3_tt.coef_[0][0]:.2f}    (抗憂鬱症藥物人數)
    {lm3_tt.coef_[0][1]:.2f}    (總平均月薪)
    {lm3_tt.coef_[0][2]:.2f}    (消費者物價指數)
'''

# print(t)
fig, ax = plt.subplots()
plt.title("多元回歸結果 (全部資料訓練)", fontsize=18)
plt.plot(range(2012,2021), pet_lm3_predict_tt/10000, label="多元回歸", color="red")
plt.scatter(pet["年份"], np.array(pet_sum)/10000, label="原始寵物隻數(除戶)", s=15, c="k")
plt.axis([1998,2021,0,200])
ax.text(2000, 50, t, color="red")
# plt.grid()
plt.xlabel("年份", fontsize=font)
plt.ylabel("數量 (萬)", fontsize=font)
plt.legend(loc="upper left", fontsize=9)
plt.savefig("./figure/Pet_plot_11.png", dpi=120)
plt.show()
```



Index	西元(年)	統計期	行業別	> 家數總計 </>	> 銷售額總計 </>
0	2008	97年	9690-21寵物美容、寄養	259	295292

```
import re
width = 0.1
plt.suptitle("")
plt.subplot(2,2,1)
x = np.array(list(range(2008,2013)))-0.3
a = 0
for i in range(0,30,5):
    plt.bar(x+a, sale6["<b>家數總計</b>"][i:i+5], width=width, label=re.findall("[^0-9-]+",sale6["行業別"][i])[0])
    a += width
plt.ylabel("2008~2012年", fontsize=font)
plt.title("家數", fontsize=font)
```



正則取字串

六、Python 認證合格證明

測驗日期	考試名稱	考生姓名	英文姓名	身分證統一編號	考場名稱	學科成績	錯誤率	術科成績	錯誤率	總分	錯誤率	合格
2022/7/26	程式語言-專業級(Python 3)	李旻恩	LEE, MIN-EN		聯成新莊認證中心	0	無	100	無	100	無	專業級
2022/8/4	軟體開發知識-專業級	李旻恩	LEE, MIN-EN		聯成新莊認證中心	98	0	無	無	98	無	專業級
2022/8/4	網頁資料擷取與分析-專業級(Python 3)	李旻恩	LEE, MIN-EN		聯成新莊認證中心	0	無	100	無	100	無	專業級
2022/8/16	人工智慧：機器學習-專業級(Python 3)	李旻恩	LEE, MIN-EN		聯成新莊認證中心	0	無	100	無	100	無	專業級
測驗日期	考試名稱	考生姓名	英文姓名	身分證統一編號	考場名稱	學科成績	錯誤率	術科成績	錯誤率	總分	錯誤率	合格

