

HW05 Mid-term

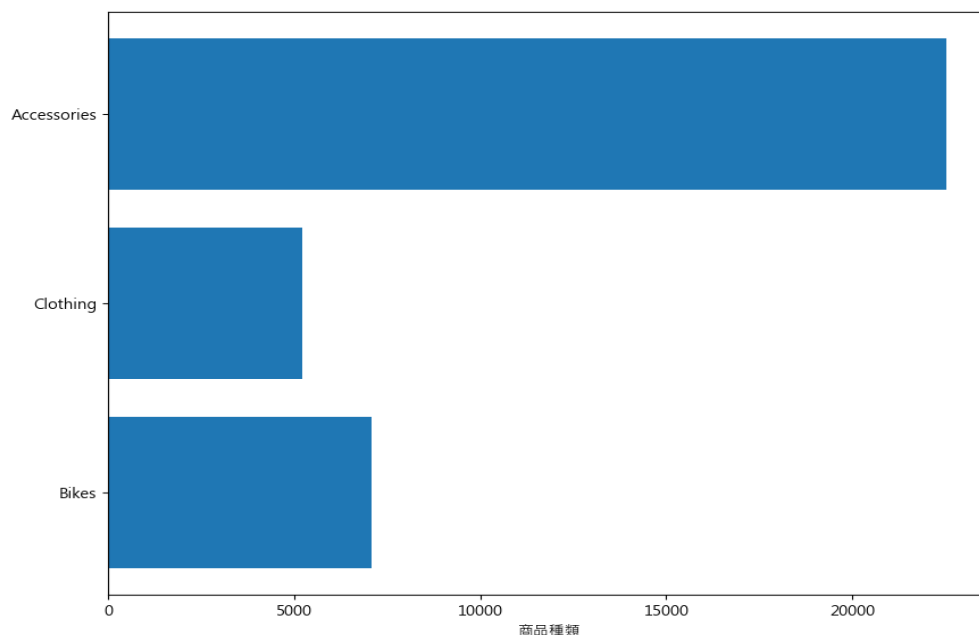
C109152304 許智程

請利用資料集 OnLine_Sale.xlsx 進行以下資料預處理及分析:

1. 新增 3 個欄位，建立每筆資料的銷售成本(單位成本 x 銷售數量)、銷售收入(單位售價 x 銷售數量)及收益(銷售收入 - 銷售成本)。
 - 注：結果以 csv 檔輸出，採 UTF-8 編碼，直接用 excel 開啟會導致亂碼，於是採用 excel 導入 csv 檔並截圖的方式呈現

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	日期	年	月	年紀	性別	國別	商品種類	商品次分類	銷售數量	單位成本	單位售價	回購	銷售成本	銷售收入	收益				
2	2015/8/31	2015	August	17	M	France	Accessories	Bottles and Cages	3	6.67	6.666666667	Y	20.01	20	-0.01				
3	2015/8/31	2015	August	17	M	France	Accessories	Bottles and Cages	3	63.33	50	Y	189.99	150	-39.99				
4	2015/8/5	2015	August	17	M	Germany	Accessories	Bottles and Cages	3	66.67	95.66666667	Y	200.01	287	86.99				
5	2015/8/5	2015	August	17	M	Germany	Accessories	Bottles and Cages	1	20	28	N	20	28	8				
6	2015/8/18	2015	August	17	M	United States	Accessories	Bottles and Cages	1	125	149	N	125	149	24				
7	2015/9/29	2015	September	17	M	United States	Accessories	Bottles and Cages	3	10	11	Y	30	33	3				
8	2015/9/29	2015	September	17	M	United States	Accessories	Bottles and Cages	1	270	266	N	270	266	-4				
9	2015/10/23	2015	October	17	M	United States	Accessories	Bottles and Cages	2	7.5	8	Y	15	16	1				
10	2015/10/23	2015	October	17	M	United States	Accessories	Bottles and Cages	1	171	165	N	171	165	-6				
11	2015/10/15	2015	October	17	M	United States	Accessories	Bottles and Cages	1	50	51	N	50	51	1				
12	2015/10/26	2015	October	17	M	United States	Accessories	Bottles and Cages	2	40	43.5	N	80	87	7				
13	2015/10/26	2015	October	17	M	United States	Accessories	Bottles and Cages	3	6.67	7.666666667	Y	20.01	23	2.99				
14	2015/8/28	2015	August	17	M	United Kingdom	Accessories	Bottles and Cages	2	62.5	71.5	N	125	143	18				
15	2015/10/4	2015	October	17	M	United Kingdom	Accessories	Bottles and Cages	3	42	46	Y	126	138	12				
16	2015/10/4	2015	October	17	M	United Kingdom	Accessories	Bottles and Cages	3	41.67	47.66666667	Y	125.01	143	17.99				
17	2015/10/22	2015	October	17	M	United Kingdom	Accessories	Bottles and Cages	3	60	71.66666667	Y	180	215	35				
18	2015/10/22	2015	October	17	M	United Kingdom	Accessories	Bottles and Cages	3	5	6	Y	15	18	3				
19	2015/10/27	2015	October	17	M	United Kingdom	Accessories	Bottles and Cages	3	66	71.66666667	Y	198	215	17				
20	2015/10/27	2015	October	17	M	United Kingdom	Accessories	Bottles and Cages	3	41.67	48.33333333	Y	125.01	145	19.99				
21	2015/12/18	2015	December	17	M	United Kingdom	Accessories	Bottles and Cages	1	170	199	N	170	199	29				
22	2015/12/18	2015	December	17	M	United Kingdom	Accessories	Bottles and Cages	2	72.5	79.5	N	145	159	14				
23	2015/11/13	2015	November	17	M	United States	Clothing	Caps	3	30	35.66666667	Y	90	107	17				
24	2015/10/23	2015	October	17	M	United States	Clothing	Caps	3	33	35.66666667	Y	99	107	8				
25	2015/12/25	2015	December	17	M	United Kingdom	Clothing	Caps	3	57	68.66666667	Y	171	206	35				
26	2015/8/28	2015	August	17	M	United States	Accessories	Bottles and Cages	1	60	70	N	60	70	10				

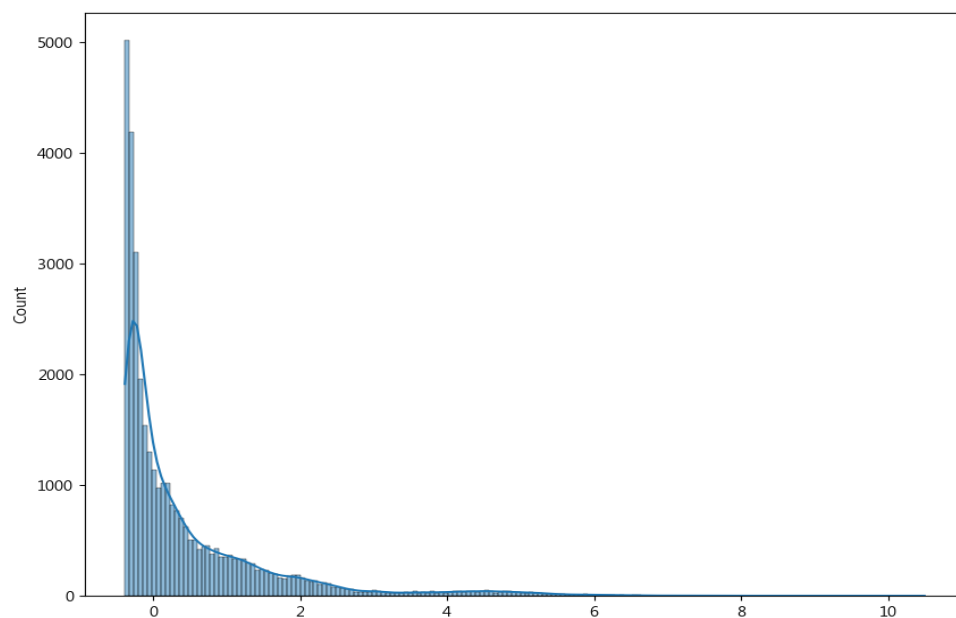
2. 分析商品種類的分佈、特徵、及資料變異等視覺化分析。



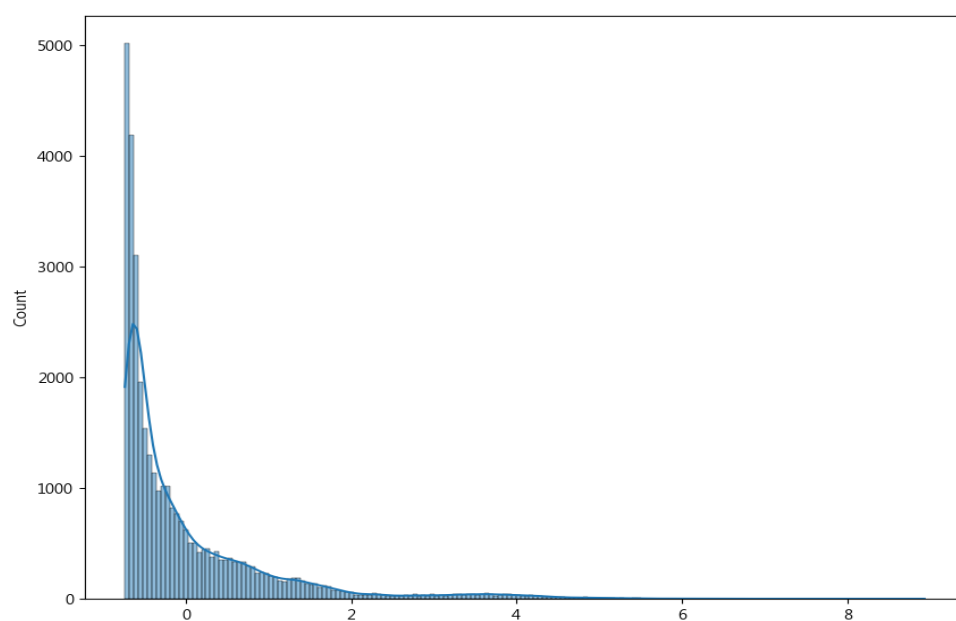
3. 針對”單位售價”與”收益”此兩個欄位，執行下列任意兩種正規化的方法：RobustScaler()、StandardScaler()、MinMaxScaler()、MaxAbsScaler()、Normalizer()。

1. 單位售價

- RobustScaler

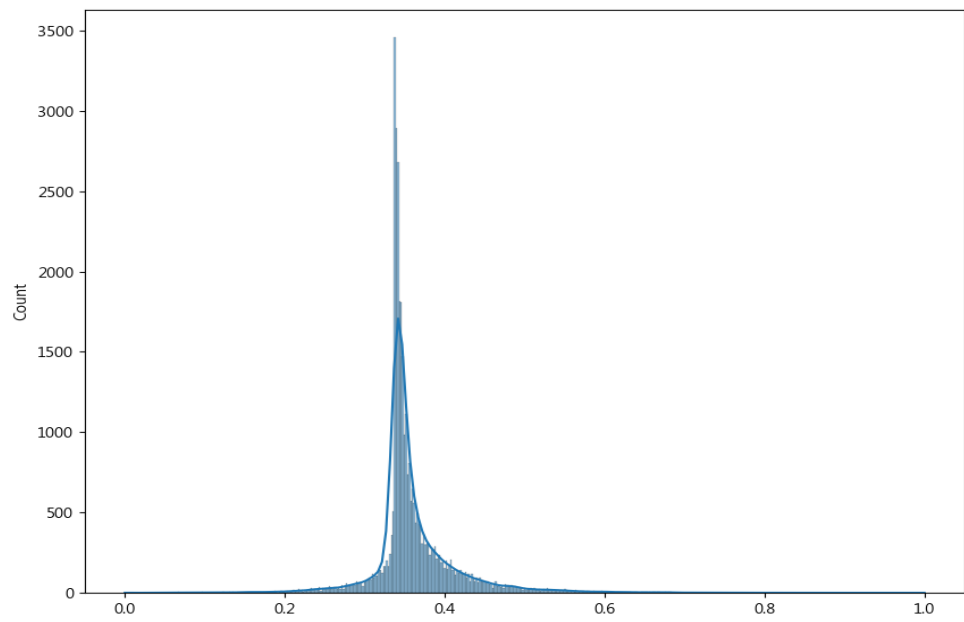


- StandardScaler

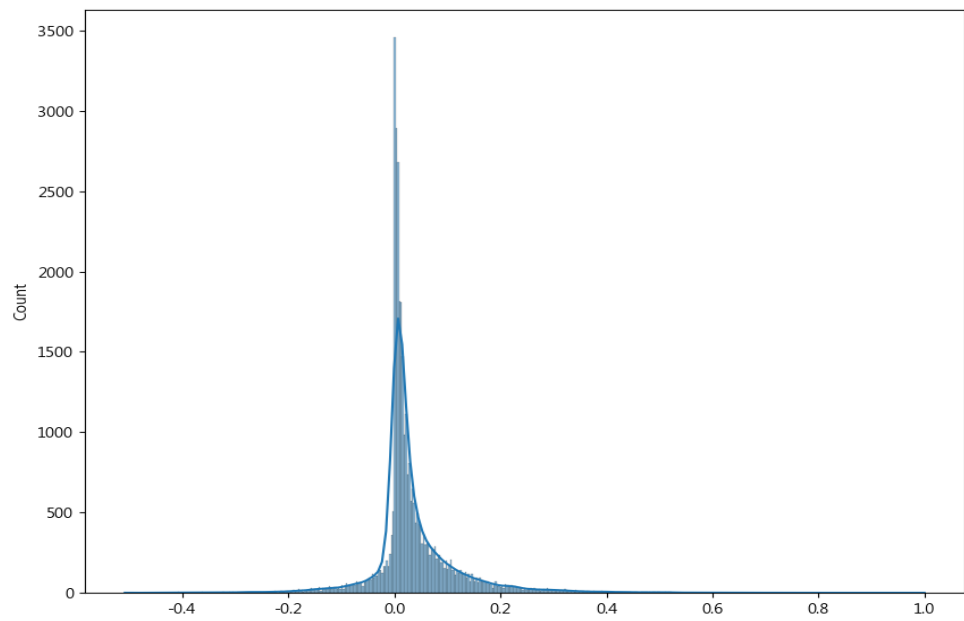


2. 收益

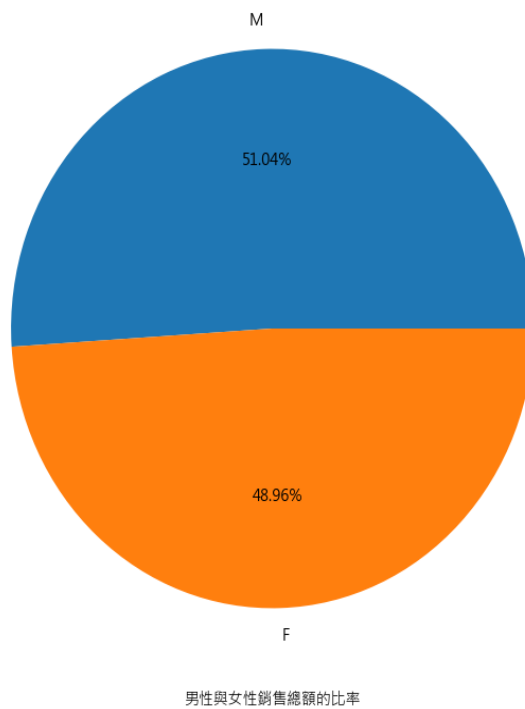
- MinMaxScaler



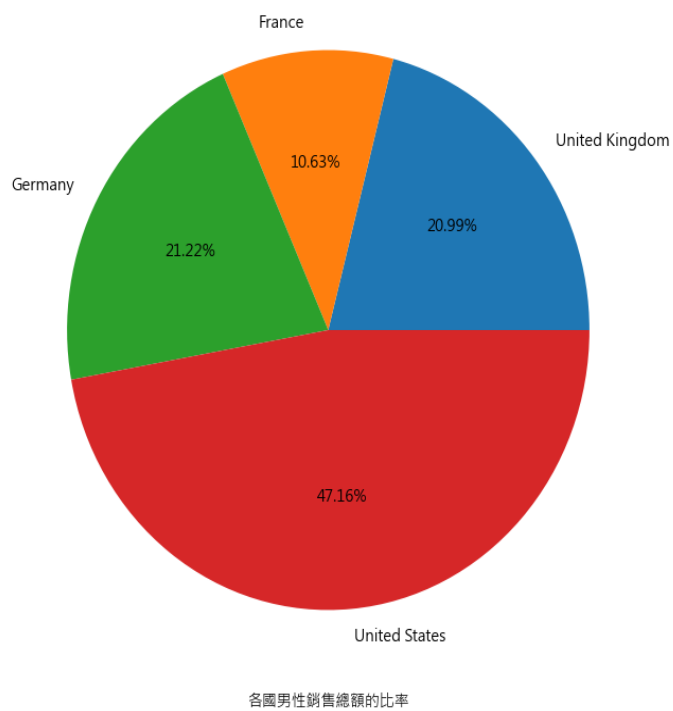
- MaxAbsScaler



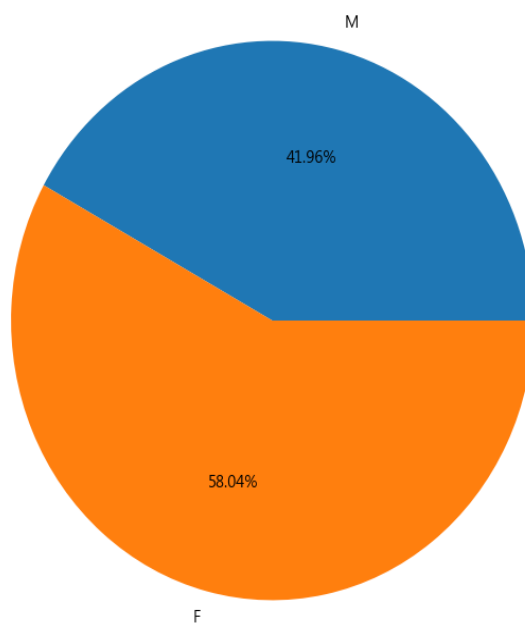
4. 針對 a~d，採用圓餅圖呈現
- a. 男性與女性銷售總額的比率為何？



- b. 哪一個國家的男性總銷售金額最高？

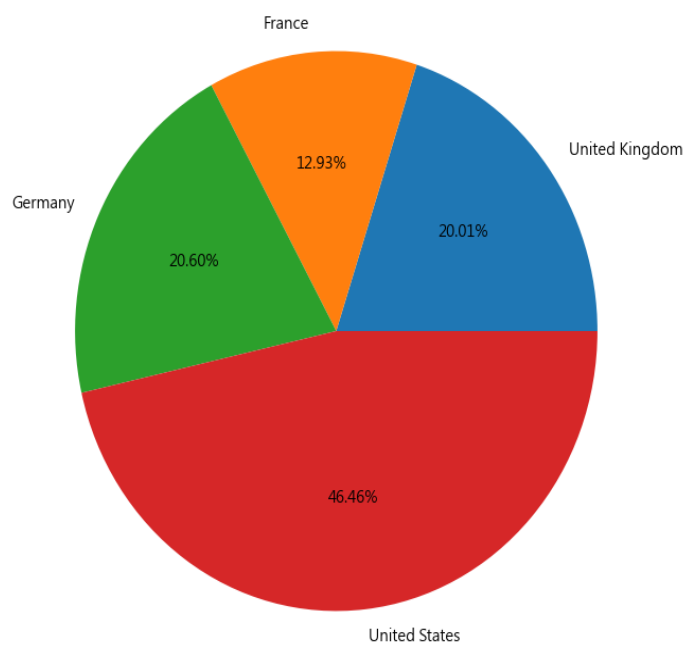


c. 法國男性與女性銷售總額的比率為何?



法國男性與女性銷售總額的比率

d. 男性與女性銷售總額的比率為何?



各國銷售總額的比率

5. 根據顧客是否會回購，使用 DecisionTreeClassifier 建立預測分類器，取得其預測的結果 (評估)，在此以 ROC(Receiver Operating Characteristics) 呈現。

