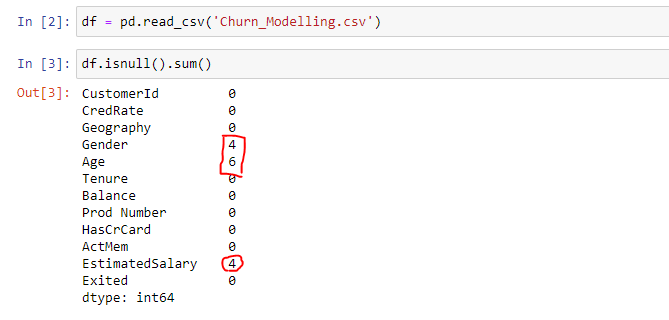
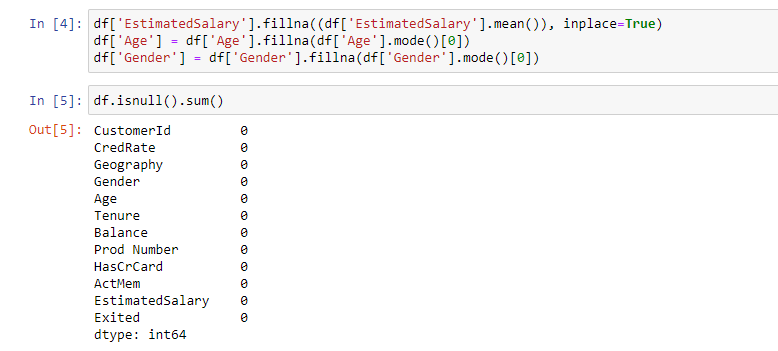
Python部分

1. 載入 Churn\_Modelling.csv 資料集，並印出哪些欄位含有遺漏值

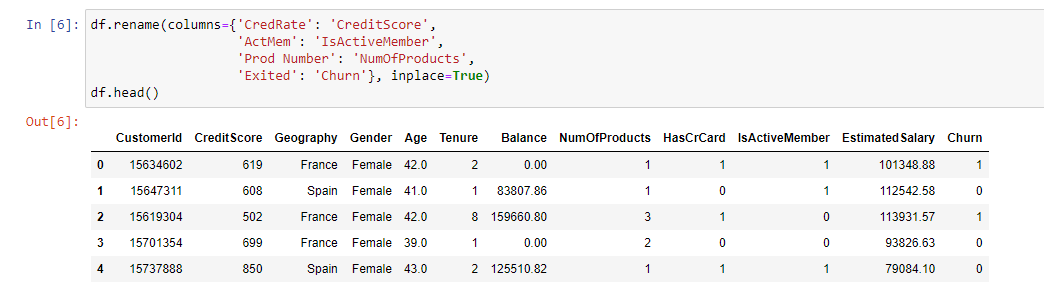
(missing value)。 (5%)



2. 以平均值填入 EstimatedSalary 的遺漏值，以眾數填入 Age 與 Gender 的遺漏值。(10%)

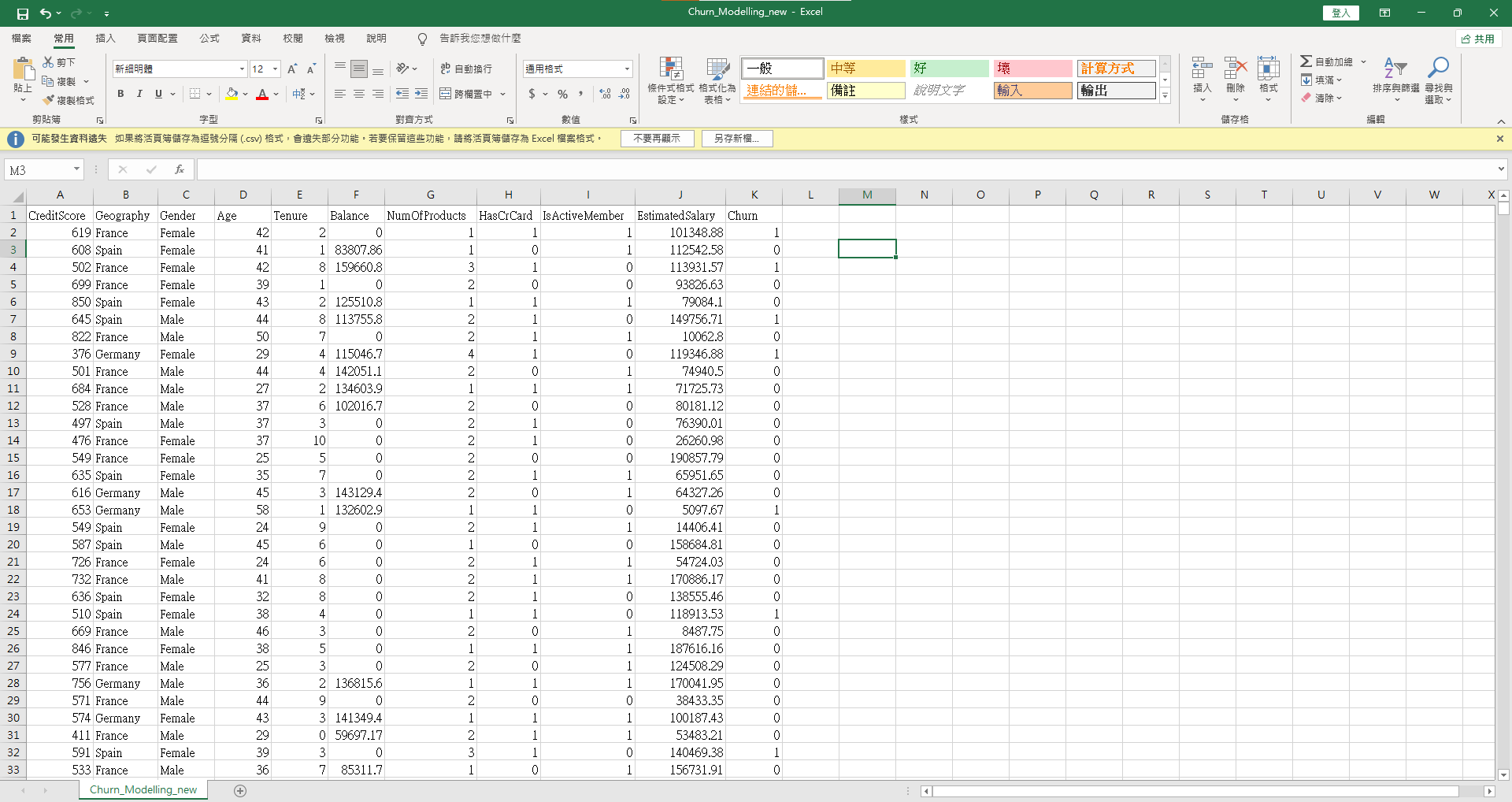


3. 修改欄位名稱，將 CredRate 改成 CreditScore、ActMem 改成IsActiveMember、Prod Number 改成 NumOfProducts、Exited 改成Churn，以利後續分析資料。 (5%)

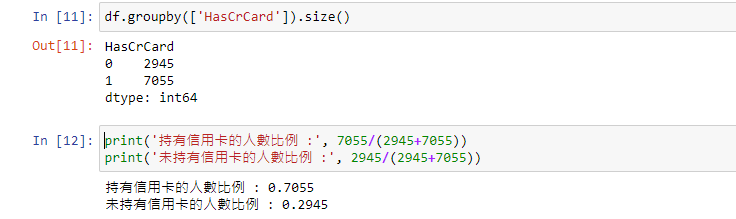
4. 去除 CustomerId,欄位，並將Geography、Gender、HasCrCard、 Churn、IsActiveMember 修改資料型態為 category，印出所有欄位的資料型態，並存成新的 CSV 檔 (設定index=False)。(5%)

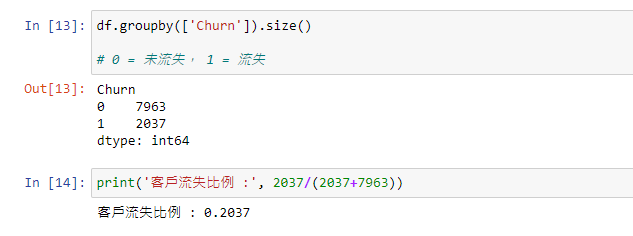


新的excel檔案：

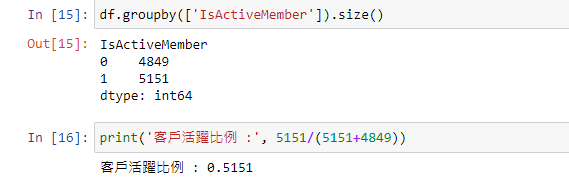
5. 對各個欄位進行分析，了解目前銀行客戶的概況：

(1) 對 HasCrCard 欄位進行分析，說明有多少比例的人持有信用卡，多少比例的人不持有信用卡。(3%)

(2) 對 Churn 欄位進行分析，說明有多少比例的客戶流失。(3%)

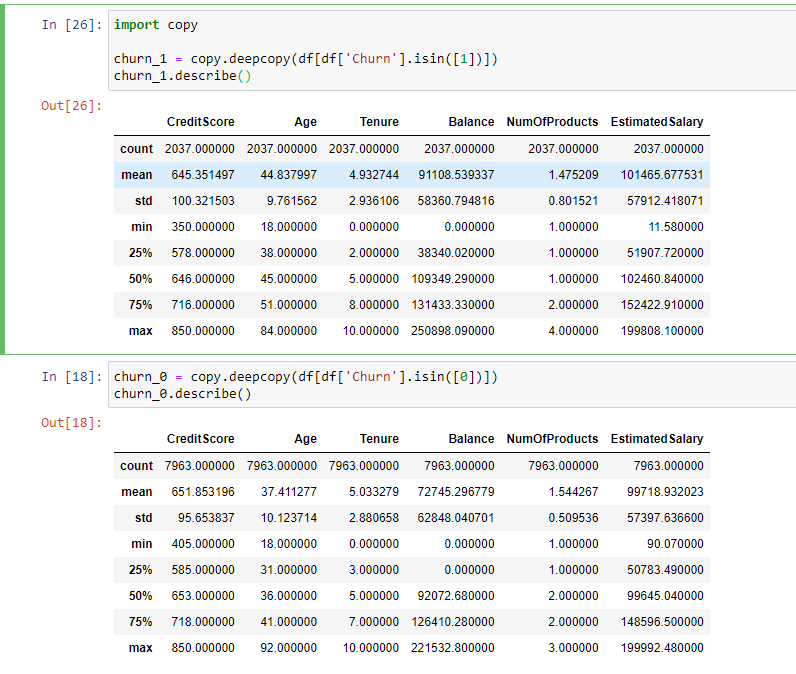


(3) 對 IsActiveMember 欄位進行分析，說明有多少比例的客戶仍是活躍狀態。

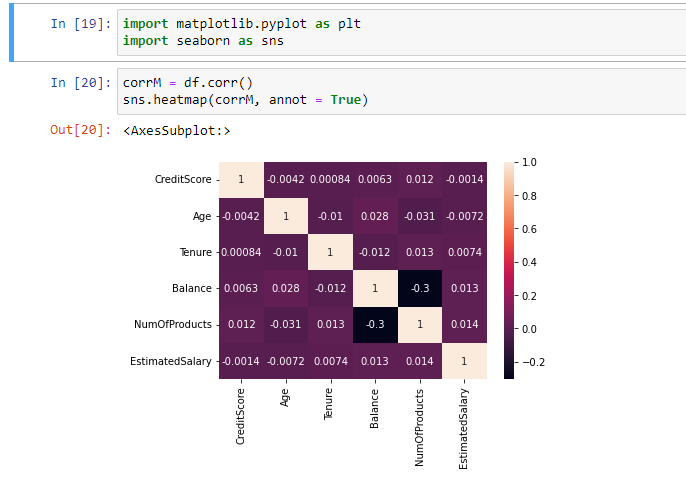


(4)對Churn 進行分析，觀察流失客戶跟未流失客戶的資料平均值

將Churn = 0 與 1的狀況分別用不同dataframe儲存，並使用describe()查看分別的統計數據

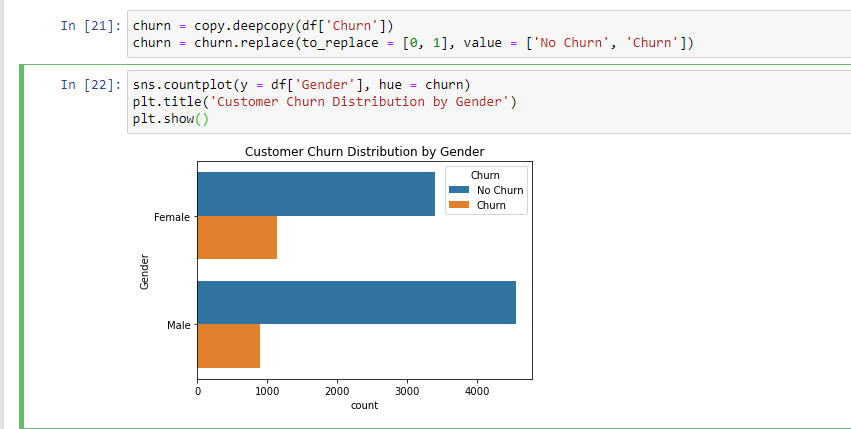


(5)計算屬性間的相關係數，並用seaborn繪製出熱力圖(heatmap) (8%)



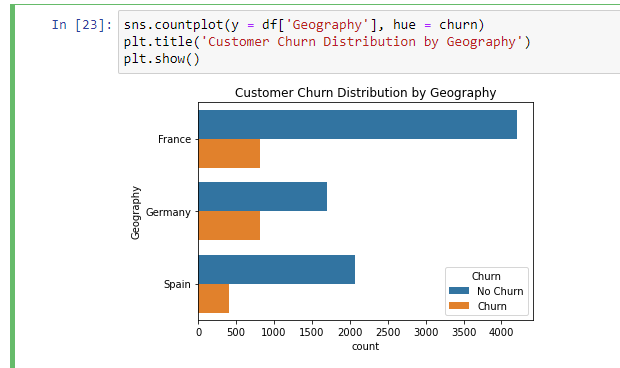
6.運用資料視覺化來幫助分析：

(1)繪出Gender與Churn 的數量關係，分析不同性別於客戶流失的關係，如下圖所示。(Hint: seaborn.countplot())(10%)

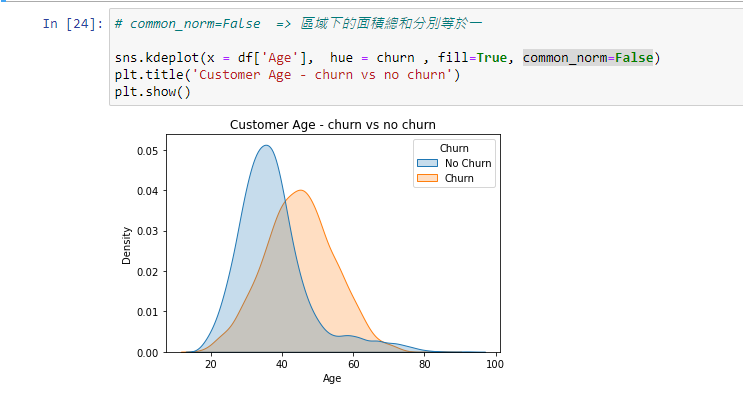


將Churn的欄位0與1分別以Churn和No Churn替代，並丟進hue中

(2)繪出Geography與Churn 的數量關係，分析不同地區於客戶流失的關係。(Hint: seaborn.countplot())(5%)

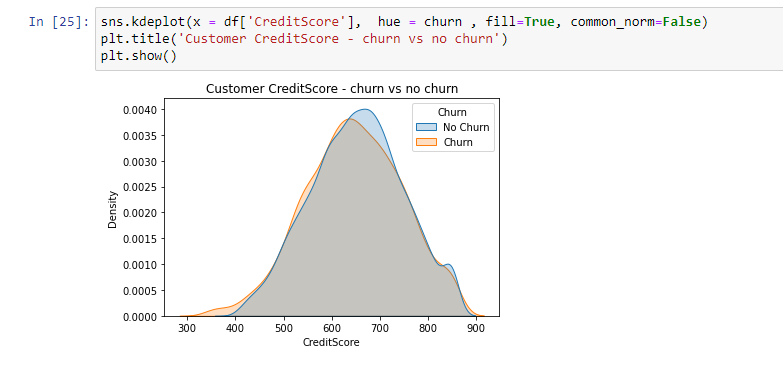


(3) 繪出 Age 分布與 Churn 的關係，分析不同年齡於客戶流失率的關係，如下圖所示。 (Hint: seaborn.kdeplot()) (10%)



可見年齡較低(約20~40歲)的顧客相較於年齡較高(約45歲左右)的顧客不易流失

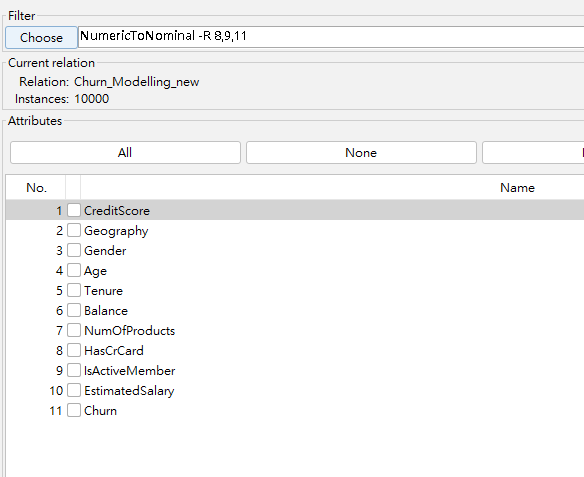
(4) 繪出 CreditScore 與 Churn 的關係，分析客戶信用分數於客戶流失率的關係，(Hint: seaborn.kdeplot()) (7%)



可見客戶流失與否與客戶信用分數無關聯。

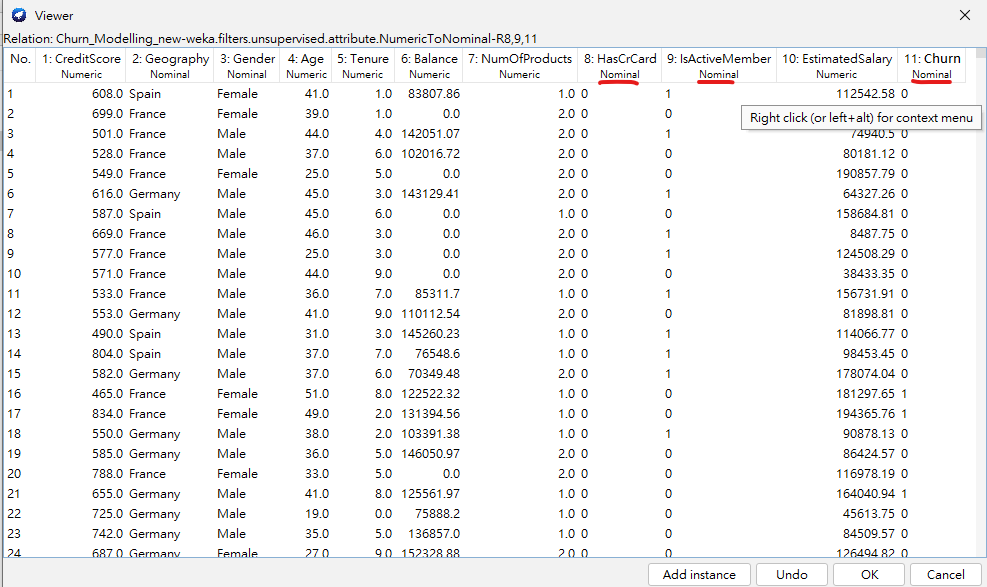
WEKA部分

(1) 將 HasCrCard, IsActiveMember, Churn 轉成 Nominal 屬性。(10%)



針對8, 9, 11欄(HasCrCard, IsActiveMember, Churn)使用NumericToNominal

結果如下：



(2) 使用 Attribute Selection，以 CfsSubsetEval 及 BestFirst 來篩選屬性，並說明屬性篩選結果。 (10%)

