**作業1: 資料大掃除**

1. 開始之前,請開啟onlineRetail\_unclean.csv. 這是一個交易記錄的資料集。請進行必要之前處理前,並請存為ETL,也請稍加說明您做了那些處理,例如您所設計的資料格式(15%)

答：

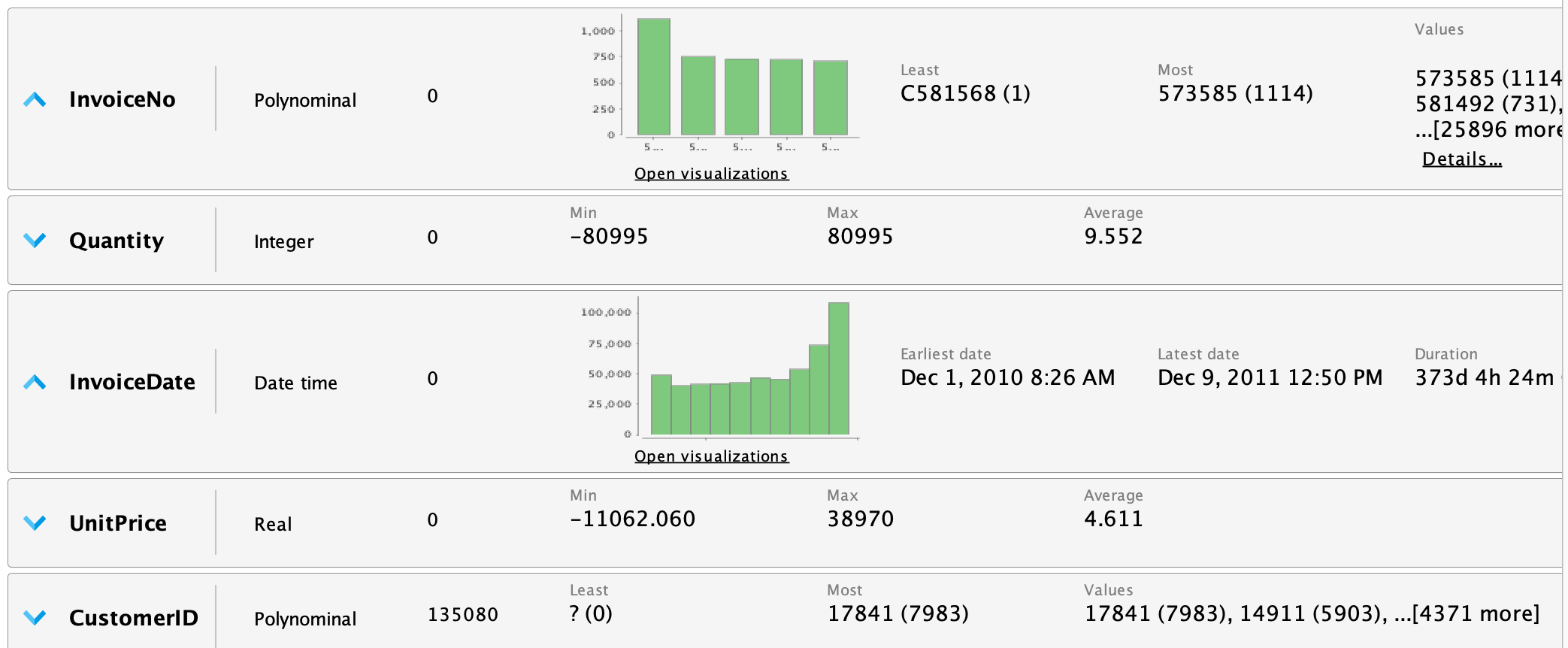
InvoiceNo, StockCode改為文字格式

InvoiceDate變更為date time格式

接著只挑選需用到的欄位進行處理CustomerID, InvoiceNo, InvoiceDate, Quantity, StockCode, UnitPrice

2. 開始之前,請進行必要之敍述統計,並請舉例說明在此資料集下,您可能可以進行之分析(10%).

答：哪個月份的購買交易次數最多



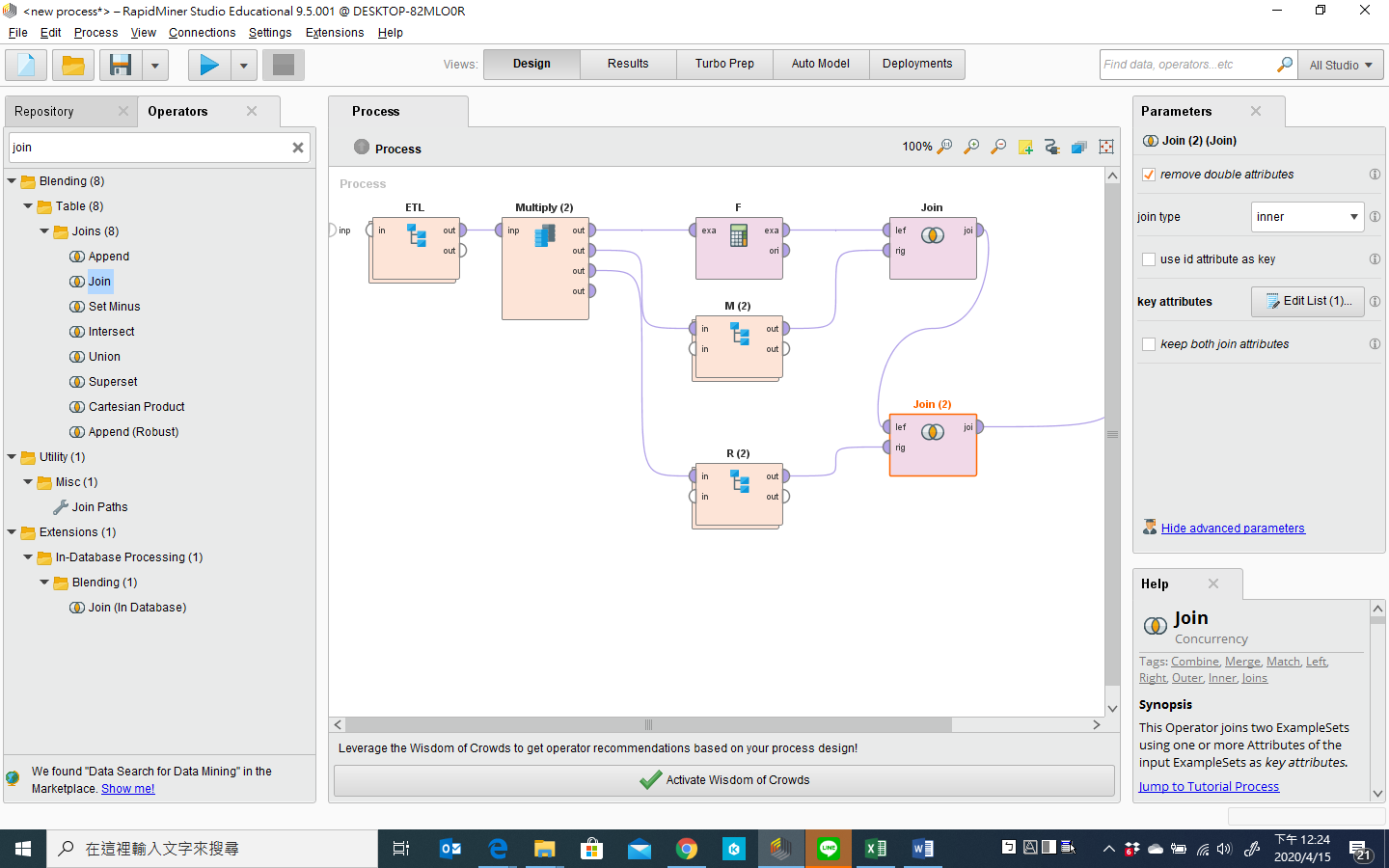
3. 接下來,我們要使用傳說中的RFM模型。但等等,目前的資料裡並沒有這3個欄位呀,那怎麼辦呢? 請產生這3個欄位唷,說明如下:

R (recentency): 距離今天的消費日期 用今天-交易天 (越靠近今天數字會越小)

F(frequency): 總共消費的次數 用id count

M(money): 總共消費的金額 單價＊數量再aggragate

第3題,我們是從消費者的角度出發唷! 請分別將R/F/M之資料產制方式存入Subprocess. (30%) 如下圖所示

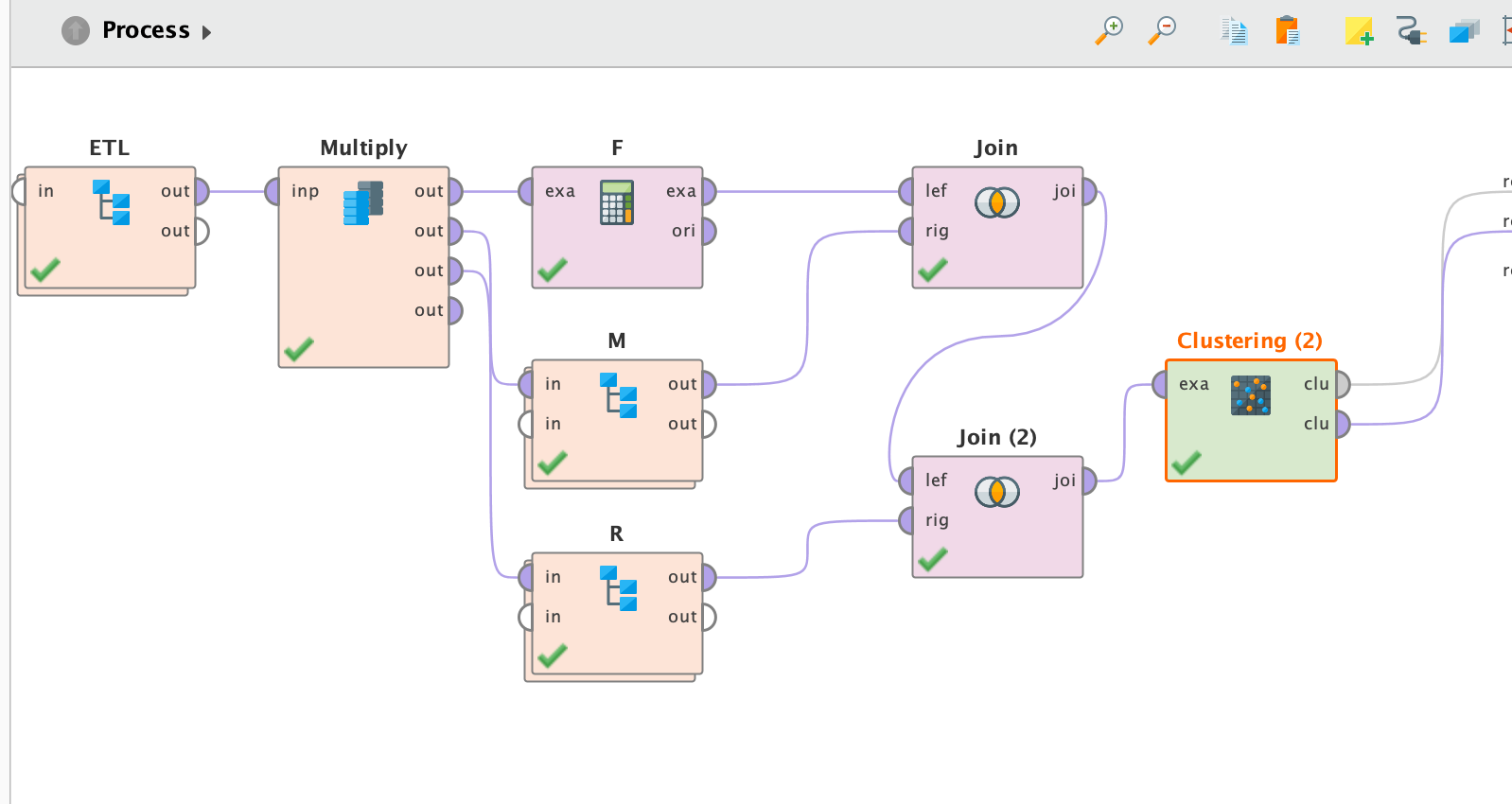


答：

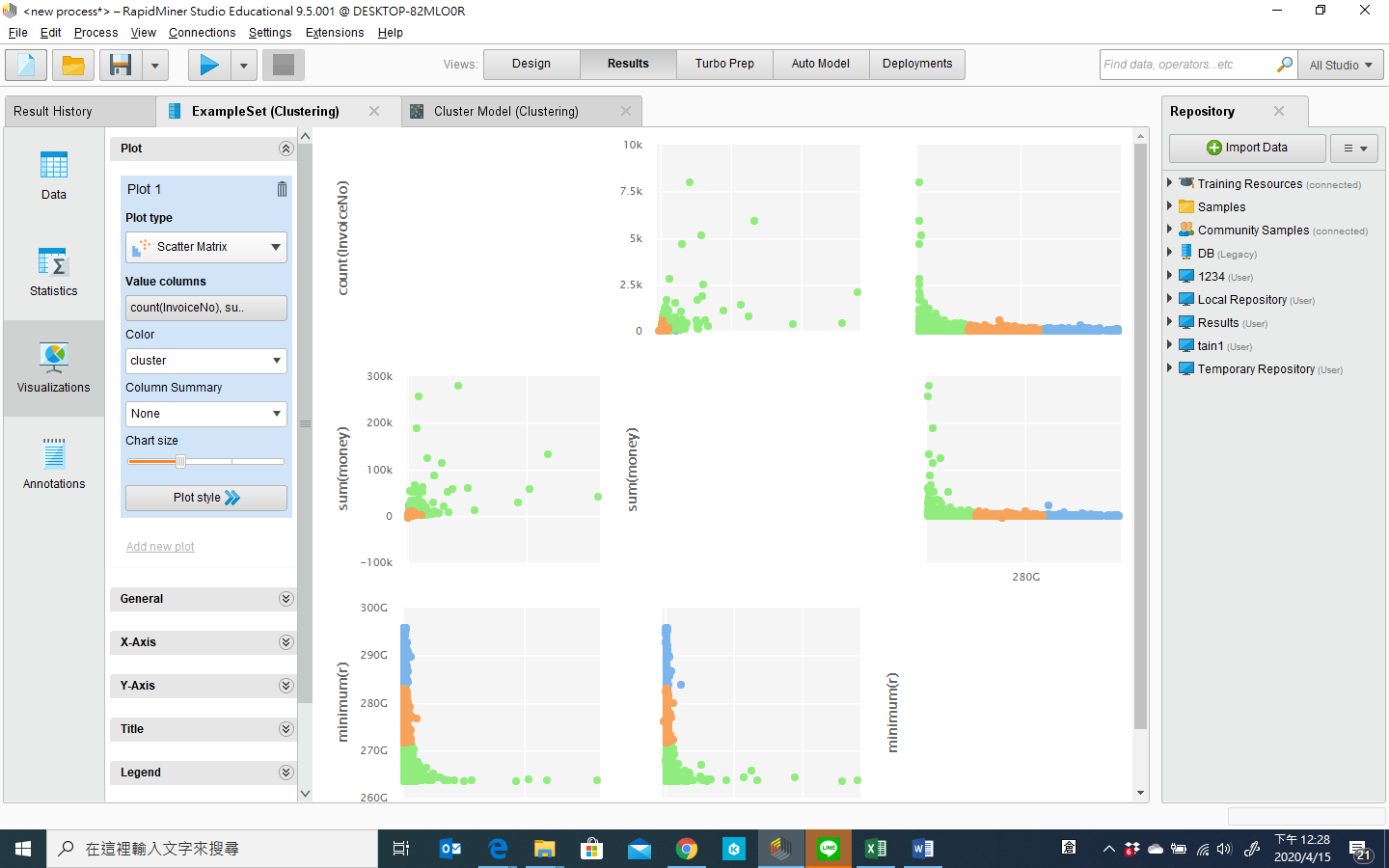
F: 使用aggregate先計算出invoiceNo的次數後，再用customerID進行group

M: 先產生一個attribute 用來計算單價成數量的金額，然後使用aggregate進行加總後，再用customerID進行group

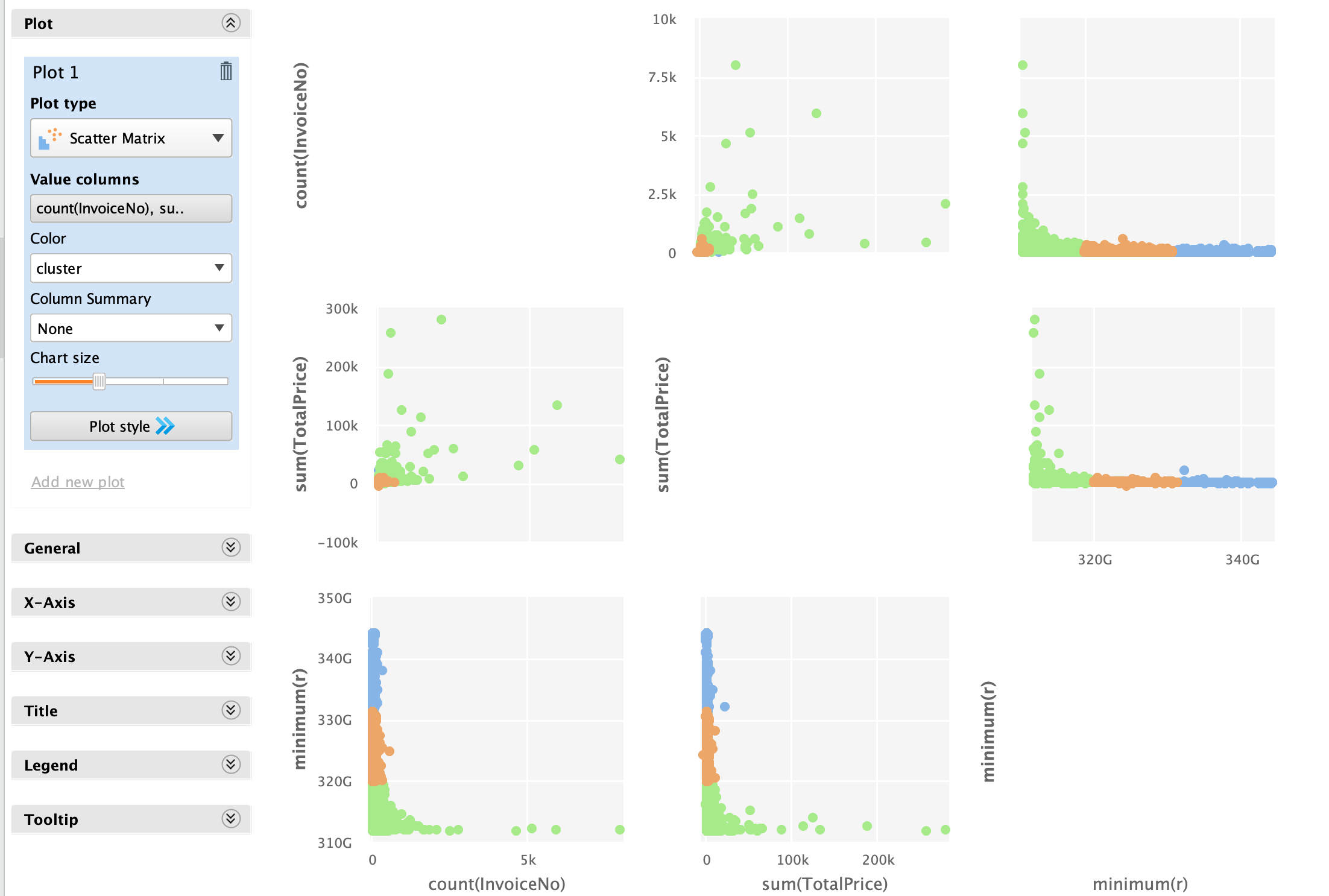
R: 先產生一個attribute 用來計算購買日到現在的日期差，然後使用aggregate找最小值，代表越靠近現在，再用customerID進行group



4.請繪制RFM之敍述統計圖表 (15%)。



答：



5. 接下來,我們要改變分析角度,變成是對產品面項使用傳說中的RFM模型,但等等,目前的資料裡還是沒有這3個欄位呀,所以請產生這3個欄位唷,說明如下:

R: 距離今天的最近一次被購買的時間

F: 總共被購買的次數

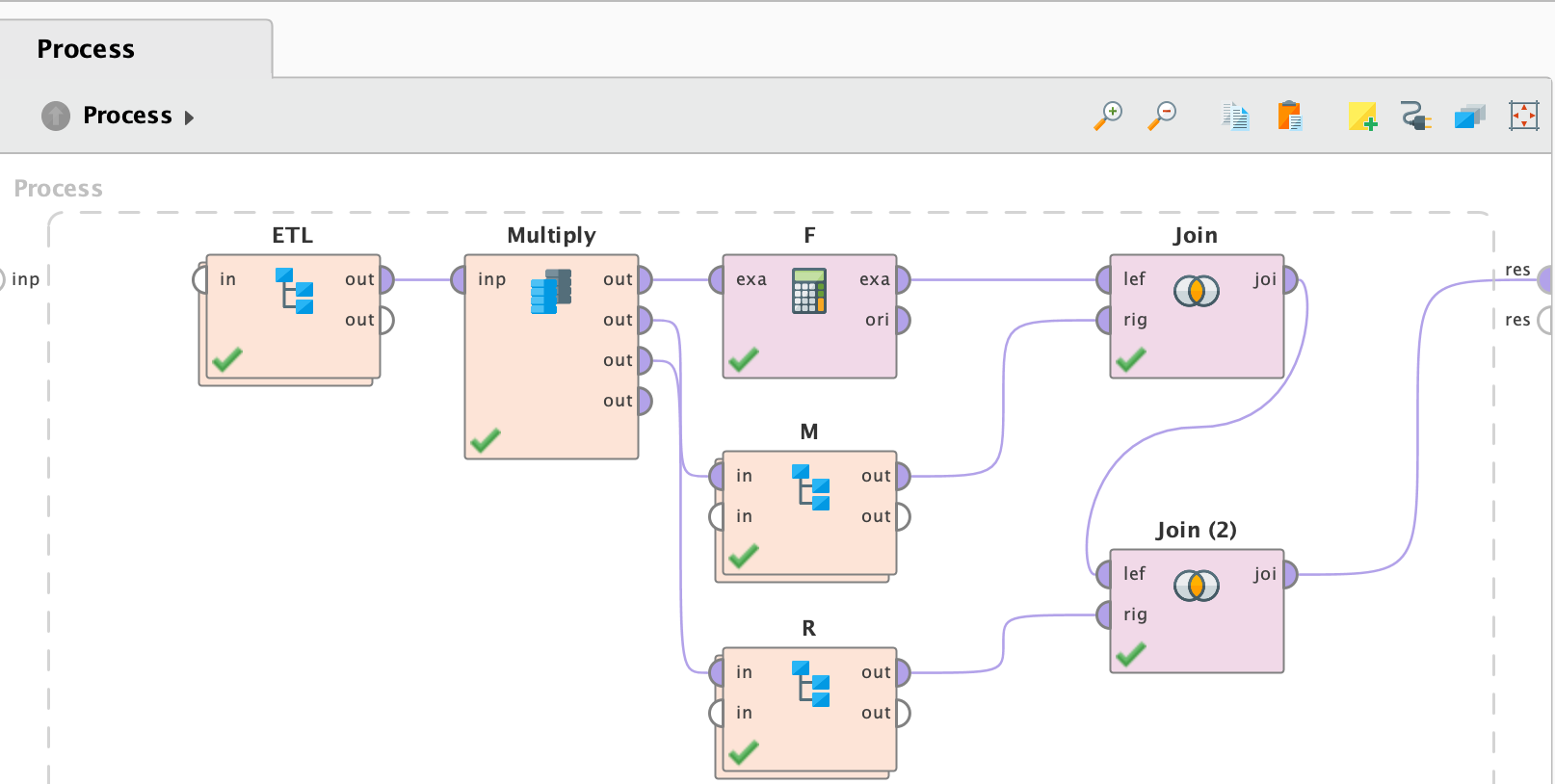
M: 總共被購買的金額

第5題,我們是以產品的角度出發唷! 請分別將R/F/M之資料產制方式存入Subprocess. (30%)

答: F(產品被購買次數): 使用aggregate先計算出invoiceNo的次數後，再用StockCode進行group

M(產品被購買的金額): 先產生一個attribute 用來計算單價成數量的金額，然後使用aggregate進行加總後，再用StockCode進行group

R(產品最近被購買日): 先產生一個attribute 用來計算購買日到現在的日期差，然後使用aggregate找最小值，代表越靠近現在，再用StockCode進行group



最後,請將相關結果,請將結果存為學號\_顧客.rmp以及學號\_產品.rmp。