 給三個表格如下:

Customer

|  |  |  |
| --- | --- | --- |
| C\_id | C\_name | City |
| M1 | John | Chicago |
| M2 | Mary | NY |
| M3 | Ann | NY |

Purchase

|  |  |  |
| --- | --- | --- |
| C\_id | P\_Id | Amount |
| M1 | P55 | 500 |
| M1 | P36 | 700 |
| M2 | P55 | 400 |
| M2 | P47 | 800 |
| M3 | P55 | 300 |

Product

|  |  |  |
| --- | --- | --- |
| P\_Id | P\_name | Price |
| P36 | orange | 2 |
| P47 | Apple | 4 |
| P55 | banana | 8 |

1. 請找出住在NY的客戶的編號、姓名和其購買產品數量。
2. 請用SQL指令寫出。

|  |
| --- |
| Select c.C\_id,c.C\_name,p.Amount  From Customer c  Join Purchase p  On c.C\_id = p.C\_id  WHERE c.City = ‘NY’; |

1. 請用R指令寫出。(p. 139 %in%; p.160 merge; p.172, subset)

|  |
| --- |
| #一. 請找出住在NY的客戶的編號、姓名和其購買產品數量。  #setwd("path") # path 設定資料夾放csv檔案的位置  setwd("C:/R")  Customer <- read.csv(file = 'Customer.csv')  Purchase <- read.csv(file = 'Purchase.csv')  Product <- read.csv(file = 'Product.csv')  x = merge(Customer,Purchase,all.x = T)  subset(x,City='NY',select = c("C\_id","C\_name","Amount")) |
|  |

二. 請找出住在NY的客戶的編號、姓名和其購買產品名稱。

1. 請用SQL指令寫出。

|  |
| --- |
| SELECT c.C\_id, c.C\_name,o.P\_id  FROM  Customer c  JOIN Purchase o  ON (c.C\_id = o.C\_id)  JOIN Product p  ON (p.P\_id = o.P\_id)  WHERE c.City Like ‘NY’; |

1. 請用R指令寫出。(p. 139 %in%; p.160 merge; p.172, subset)

|  |
| --- |
| #二. 請找出住在NY的客戶的編號、姓名和其購買產品名稱。  #install.packages('dplyr') #安裝套件  library(dplyr) #使用套件  #setwd("path") path 設定資料夾放csv檔案的位置  setwd("C:/R")  Customer <- read.csv(file = 'Customer.csv')  Purchase <- read.csv(file = 'Purchase.csv')  Product <- read.csv(file = 'Product.csv')  x <- merge(Customer,Purchase,all.x = T)  y <- merge(x,Product,all.x = T)  y  y %>% filter(City == 'NY') %>% select(C\_id,C\_name,P\_name) |
|  |

三 . 請找出住在NY的客戶的編號和其平均購買量(以客戶編號排序)。

1. 請用SQL指令寫出。

|  |
| --- |
| SELECT  c.C\_id, AVG(p.Amount)  FROM  Customer c  JOIN Product p  ON (c.C\_id = p.C\_id)  WHERE c.City Like ‘NY’; |

1. 請用R指令寫出。(p. 139 %in%; p.160 merge; p.172 subset; p.176 arrange; p.178 summarise, group\_by).

|  |
| --- |
| #三 . 請找出住在NY的客戶的編號和其平均購買量(以客戶編號排序)。  #install.packages('dplyr') #安裝套件  library(dplyr) #使用套件  #setwd("path") path 設定資料夾放csv檔案的位置  setwd("C:/R")  Customer <- read.csv(file = 'Customer.csv')  Purchase <- read.csv(file = 'Purchase.csv')  table <- merge(Customer,Purchase , all.x = T)  table %>% select(C\_id,City,Amount) %>% filter(City == 'NY') %>% group\_by(C\_id) %>% summarise("BUY Avg" = mean(Amount)) %>% arrange(C\_id)  # 先畫範圍 , 過濾 filter 找我要的 NY , 群組是 c\_id , summarise 算 , arrange 排序 ADD desc(C\_id) 則是大到小 |
|  |

四.  請找出住在NY且平均購買量大於400個產品的客戶的編號和其購買總金額(以客戶編號排序)。

1. 請用SQL指令寫出。

|  |
| --- |
| SELECT c.C\_id , AVG(o.Amount) , SUM(p.Price \* o.Amount)  FROM  Customer c  JOIN Purchase o  ON (c.C\_id = o.C\_id)  JOIN Product p  ON (p.P\_id = o.P\_id)  WHERE c.City Like ‘NY’  GROUP BY c.C\_id  HAVING AVG(o.Amount) > 400  ORDER BY c.C\_id; |

(2)請用R指令寫出。(p. 139 %in%; p.160 merge; p.172 subset; p.176 arrange; p.178 summarise, group\_by)

|  |
| --- |
| #四. 請找出住在NY且平均購買量大於400個產品的客戶的編號和其購買總金額(以客戶編號排序)。  #install.packages('dplyr') #安裝套件  library(dplyr)  #setwd("path") path 設定資料夾放csv檔案的位置  setwd("C:/R")  Customer <- read.csv(file = 'Customer.csv')  Purchase <- read.csv(file = 'Purchase.csv')  Product <- read.csv(file = 'Product.csv')  table <- merge(Customer,Purchase , all.x = T)  table <- merge(table,Product,all.x = T)  table  table %>% filter(City == 'NY' , mean(Amount) > 400) %>% group\_by(C\_id) %>% summarise("BUY\_AVG" = mean(Amount),"TOTAL\_BUY" = sum(Price \* Amount)) |
|  |