作業1: 資料大掃除

請讀入openfood的檔案後,進行以下作業練習

1. openfood共有3個檔案,請合併讀入,貼下您的流程圖, (10%)並回答以下問題:

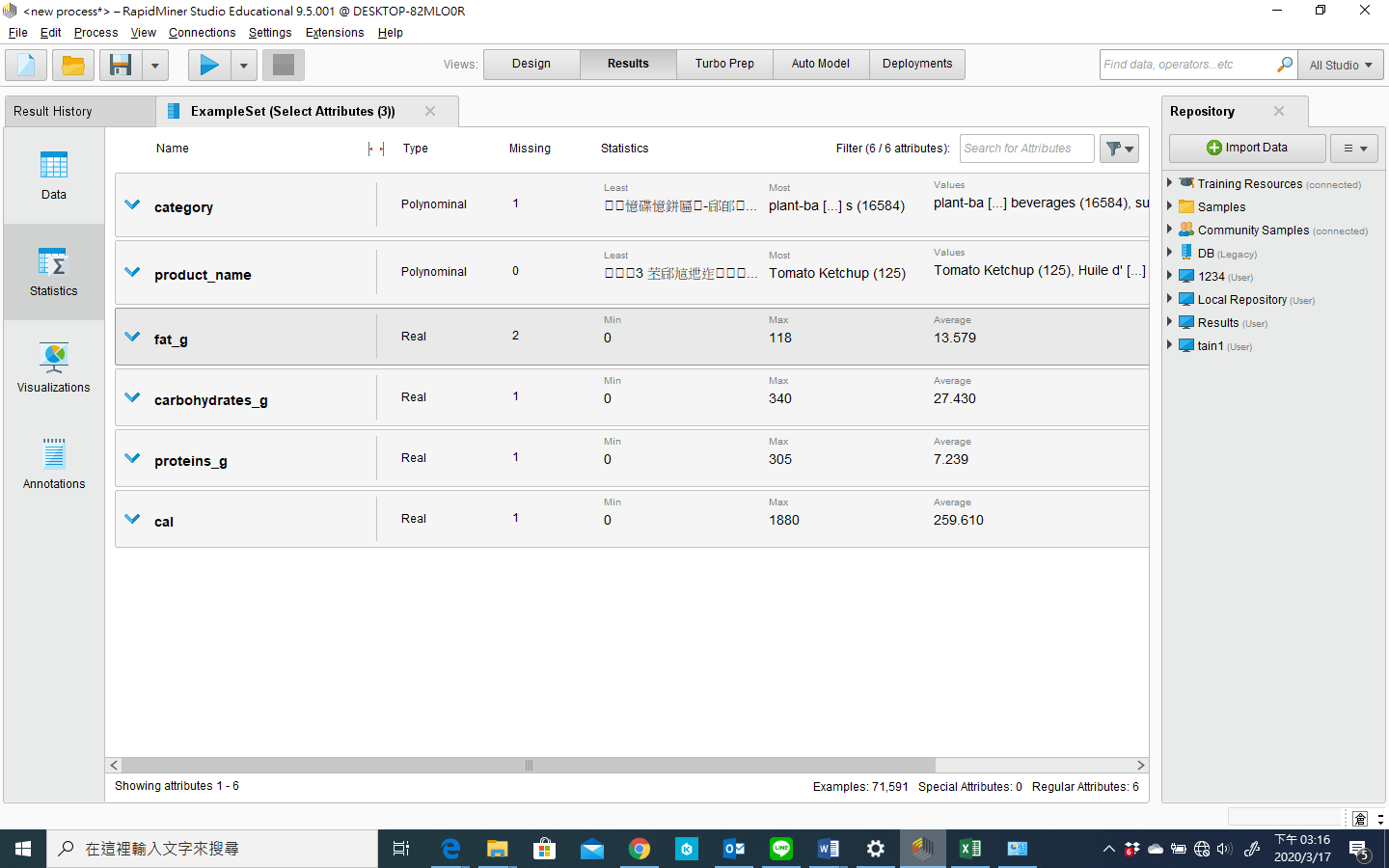
1-1 (5%)合併後,您的資料表共有( )欄位

1-2 (5%)那個(些)欄位的缺漏值最多:

1-3 (5%)那個(些欄位沒有缺漏值:

2. 因為資料來自不同的商店,各家連鎖店也都有不同命名方式，為了後續分析的需要，我們必需進行欄位的更名,請問您建議的做法,並請貼下您的流程圖及分析結果(10%)?

3. 為了說明方便,最後顧問建議留下以下屬性並進行必要之更名,請完成(10%)



4. 就在您終於完成檔案的取得與合併後,您以為你可以開始進行分析時,突然您發現在卡路里這個欄位有許多的缺漏值。若要刪除這些資料,您將損失許多資料,因此您要想辦法補值。關於卡路里的計算,請自行產生公式,每公克油脂提供9卡的熱量,每公克蛋白質及碳白質分別提供4卡的熱量。 (15%)

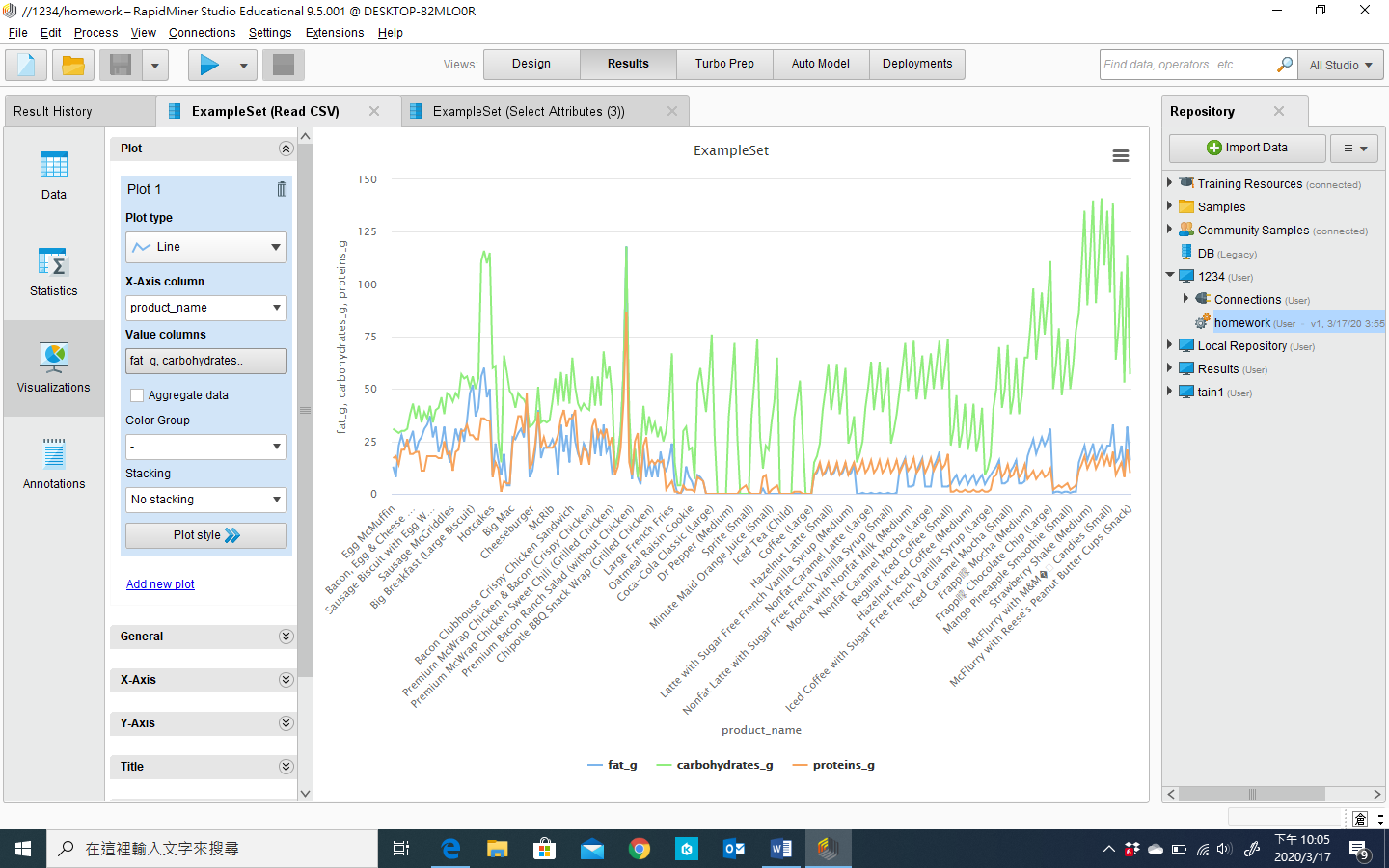
\*請貼出流程

5. 太棒了,現在我們終於可以開始進行一些敍述統計了,

5-1 (5%)請問您最後留下幾筆資料? 還有缺漏值嗎? 那些欄位?

5-3 (5%) 您建議怎麼處理這些缺漏值,請說明並進行必要之處置

6. 有圖有真相,請畫出2張圖,並加以說明。(10%)



6. 請增加一個欄位能夠識別不同的供應商(15%)

7. 請繪制出以下的圖(10%)

