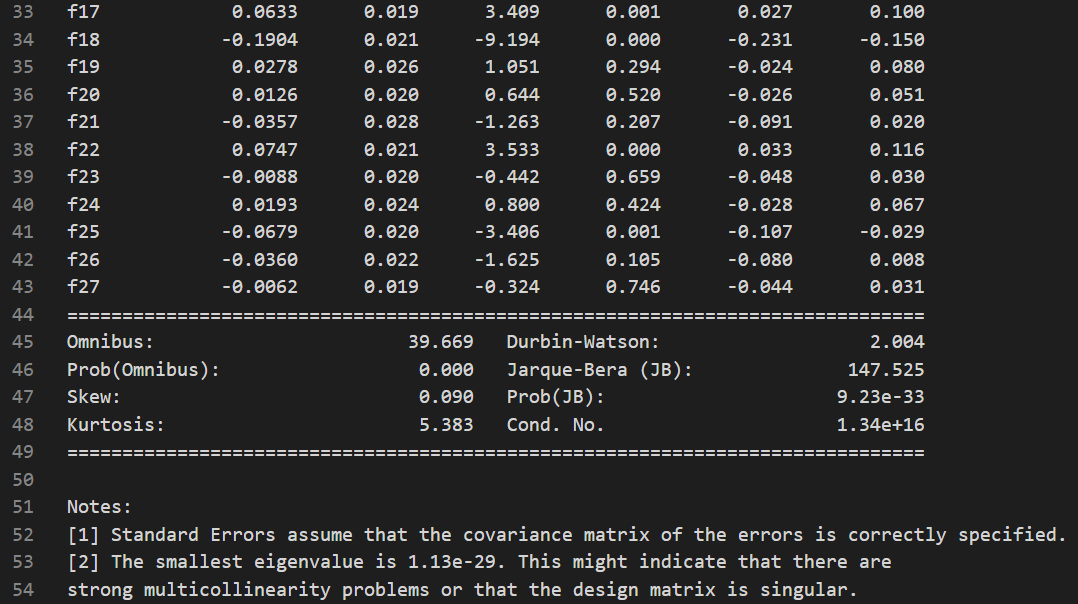
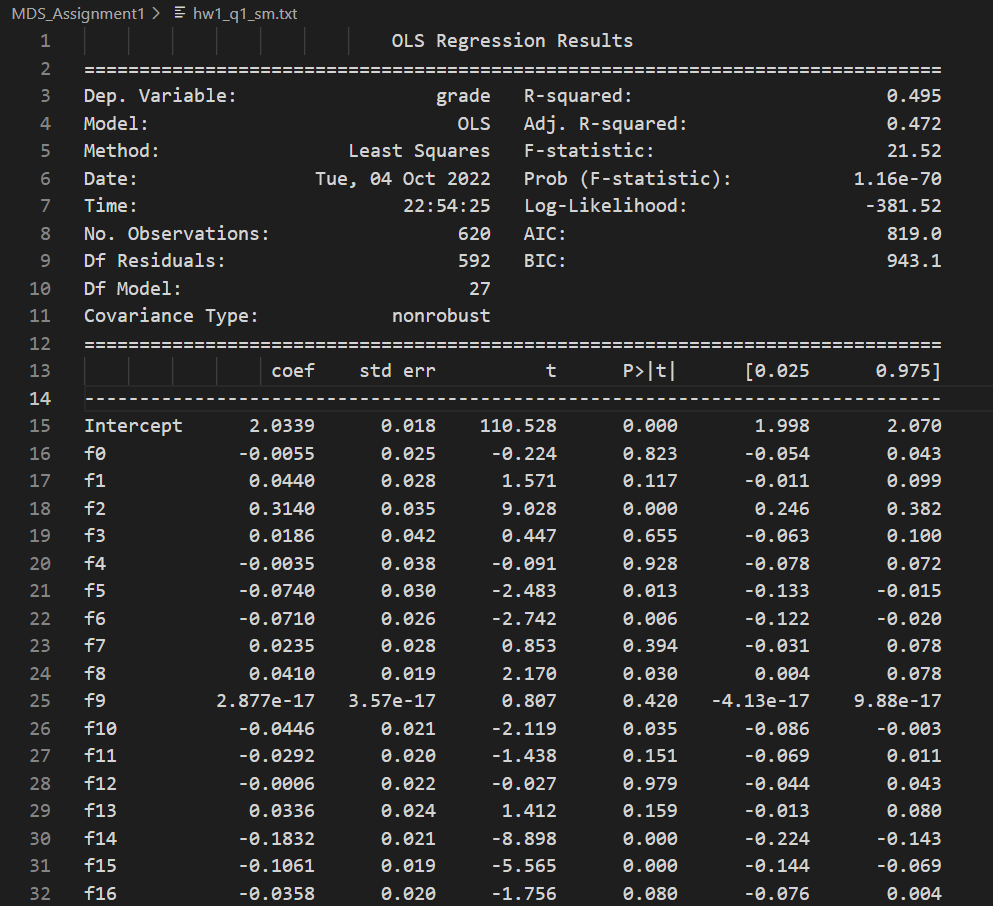
**作業一 資管碩二 呂晟維 r10725012**

1. **(35%) Linear Regression Analysis for Wine Quality**
2. (10%) Show the results of regression analysis as follows



1. (5%) The fitting of the linear regression is a good idea? If yes, why? If no, why? What’s the possible reason of poor fitting?

Ans:

使用以下兩指標解讀線性分析的結果。

R-squared: the measurement of how much of the independent variable is explained by changes in our dependent variables.

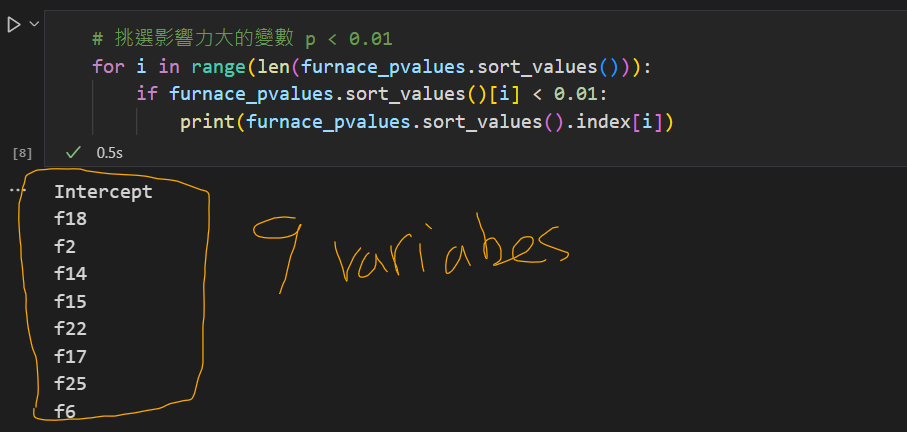
* + 統計結果的 R-squared 是 0.495，表示模型能解釋變數們 49.5% 的變量。R-squared < 0.5 指出此模型表現不佳。
  + Adj. R-squared 的結果雷同。

Prob (F-Statistic): this number to tell you the accuracy of the null hypothesis(H0), or whether it is accurate that your variables’ effect is 0.

* + H0 是殘差正常的假設。
  + 統計結果的 `Prob (F-Statistic)` 很小，為 1.16e-70，因此我們拒絕此假設，也表示此數據不太適合做 linear regression。

1. Based on the results, rank the independent variables by p values and which one are statistically significant variables with p values<0.01?

Ans: p越小影響力越大，如下程式運行的結果，依照小至大排序為f18 < f2 < … < f6。



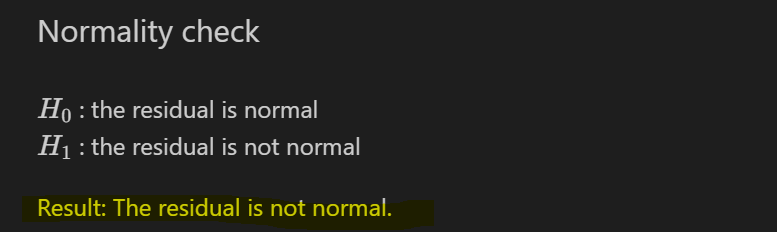
1. (15%) Testify the underlying assumptions of regression (1) Normality, (2) Independence, and (3) Homogeneity of Variance with respect to residual

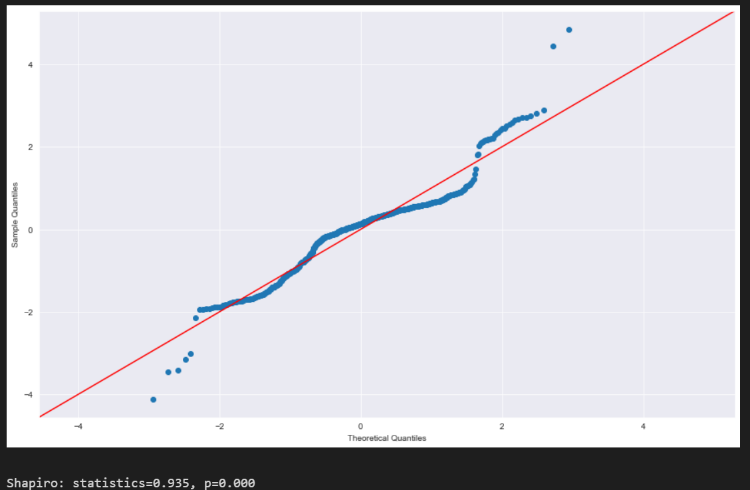
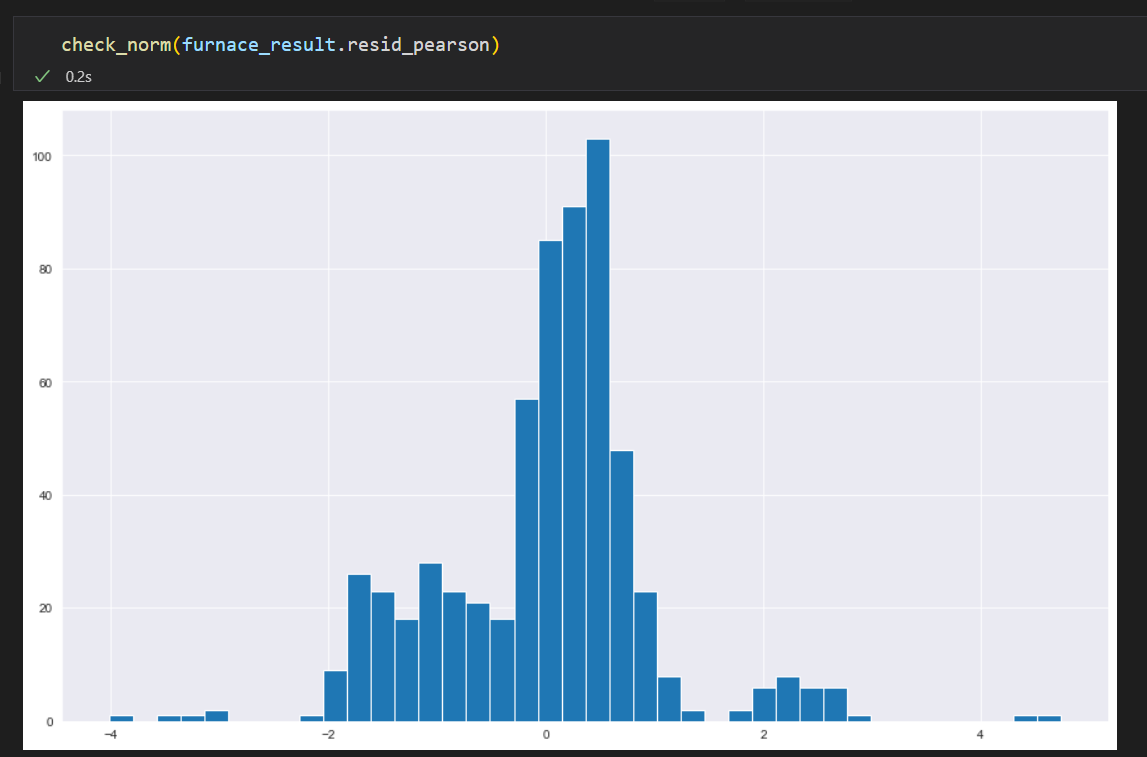
Ans:

* Normality: Shapiro-Wilk test 常態性檢定 => Not normal
* Independence: Durbin-Watson test 獨立性檢定 => Independent, features間關聯性低
* Homogeneity: Breusch-Pagan test 異質變異數檢定 => No Homogeneity

(一) Normality Check

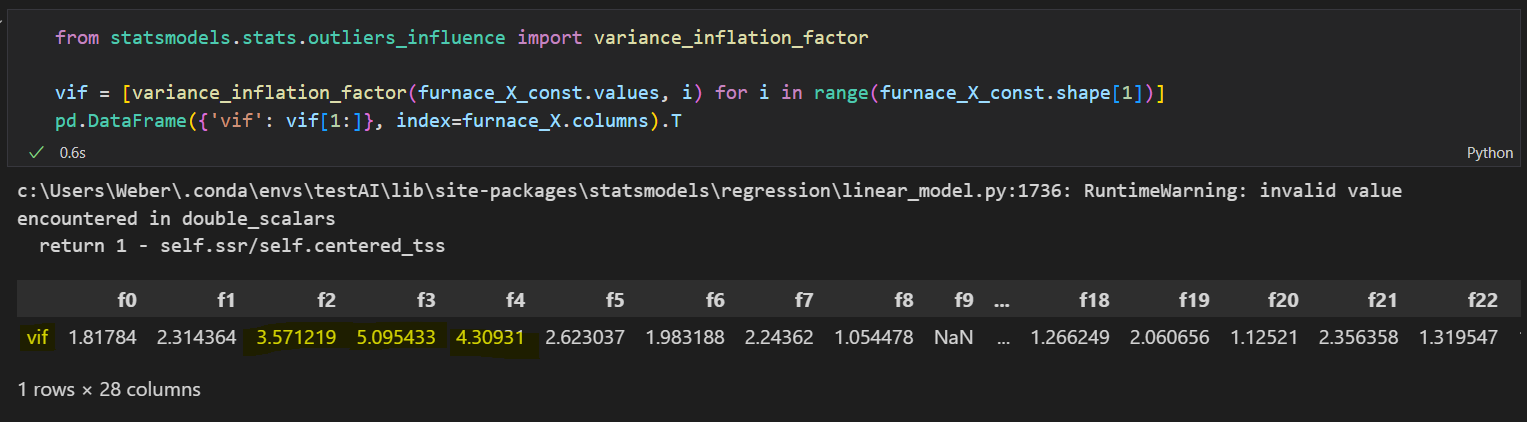
The residual is not normal. Because `Shapiro: statistics=0.935, p=0.000`. Result p-value < 0.05 so we reject H0.



****

(二) Independence Check (aka check multicollinearity)

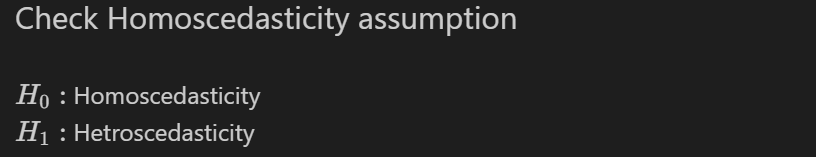
None of the VIF (Variance Inflation Factor) value of the features > 10, therefore, no significant multicolinearity within the variables. No features should be removed.

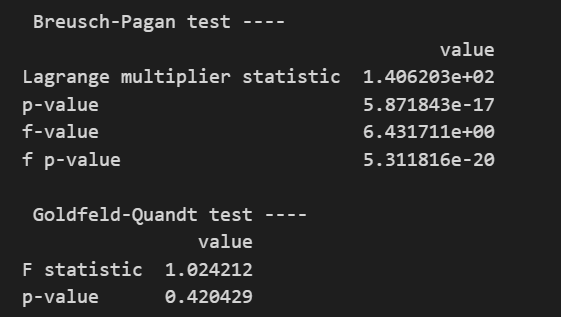


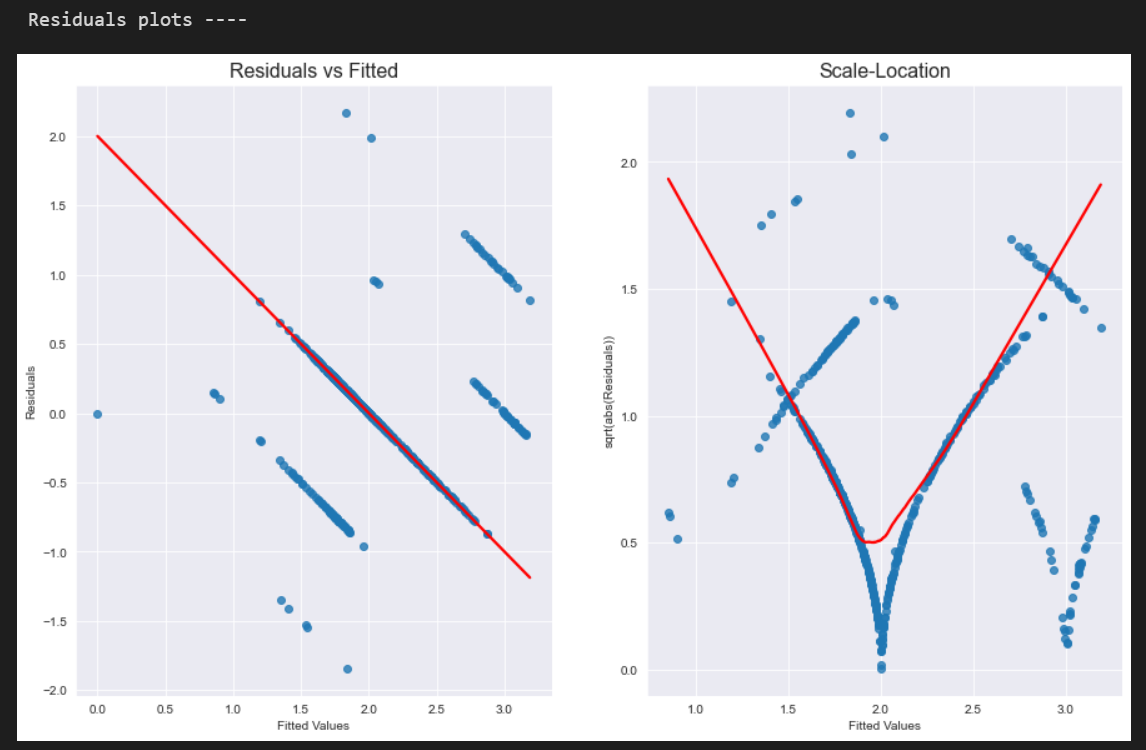
(三) Homogeneity Check (aka Homoscedasticity)

B-P test shows hetroscedasticity, G-Q test shows homoscedasticity.

B-P test reject null hypothesis with p-value = 5.3e-20 < 0.5. Which means residual doesn’t show homoscedasticity. G-Q test result a p-value = 0.42. Although it less than 0.5 a little bit. It is more acceptable than the result of B-P test.







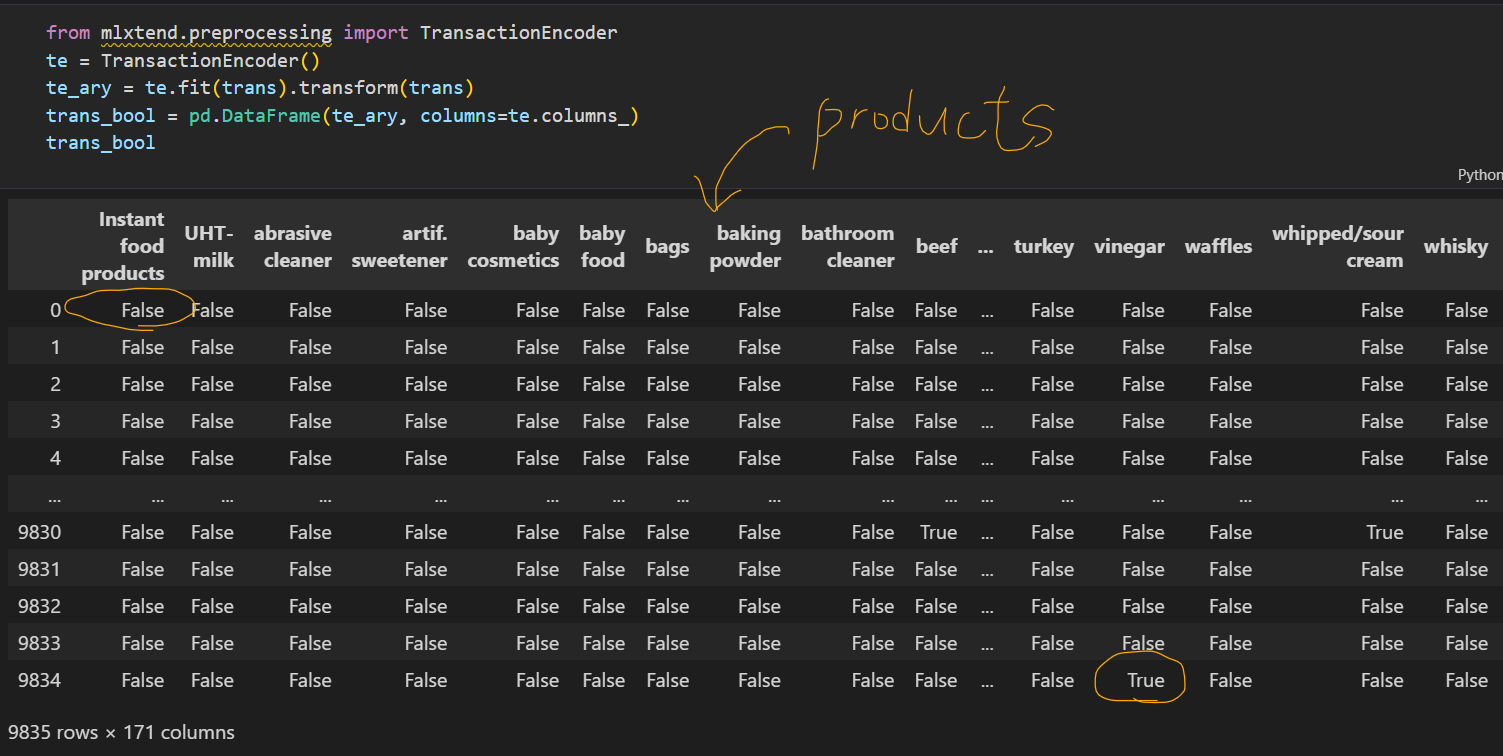
1. **(35%) Association Rule Market Basket Analysis**

* **Set the minimum support to 0.001**
* **Set the minimum confidence of 0.15**

1. (10%) How to handle the raw dataset via data preprocessing?

Ans: Make each transaction record in to Boolean vector, so the dataset is a Boolean matrix.

Item bought indicates the corresponding column is True, otherwise False.



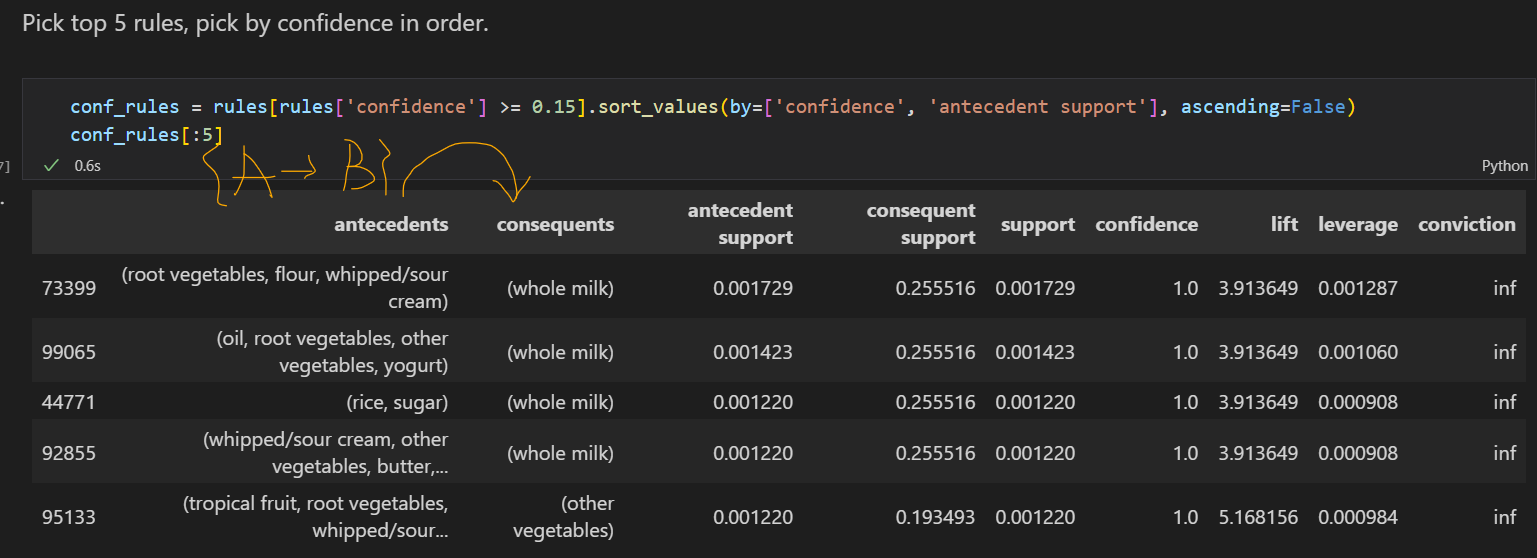
1. (10%) What’s the top 5 association rules? Show the support, confidence, and lift to each specific rule, respectively?

Ans:

All the 5 rules show that “whole milk” and “other vegetables” is a great consequent product. The most frequent rule is {root vegetables, flour, whipped/sour cream → whole milk}. Its support value is 0.001729, which means this itemset buying ratio.

All the confidence is 1.0 which means these are strong causal relationship.

Lift > 1 means these rules appear more frequent than normal case of consequent.



1. (5%) Please provide/guess the “story” to interpret one of top 5 rules you are interested in.

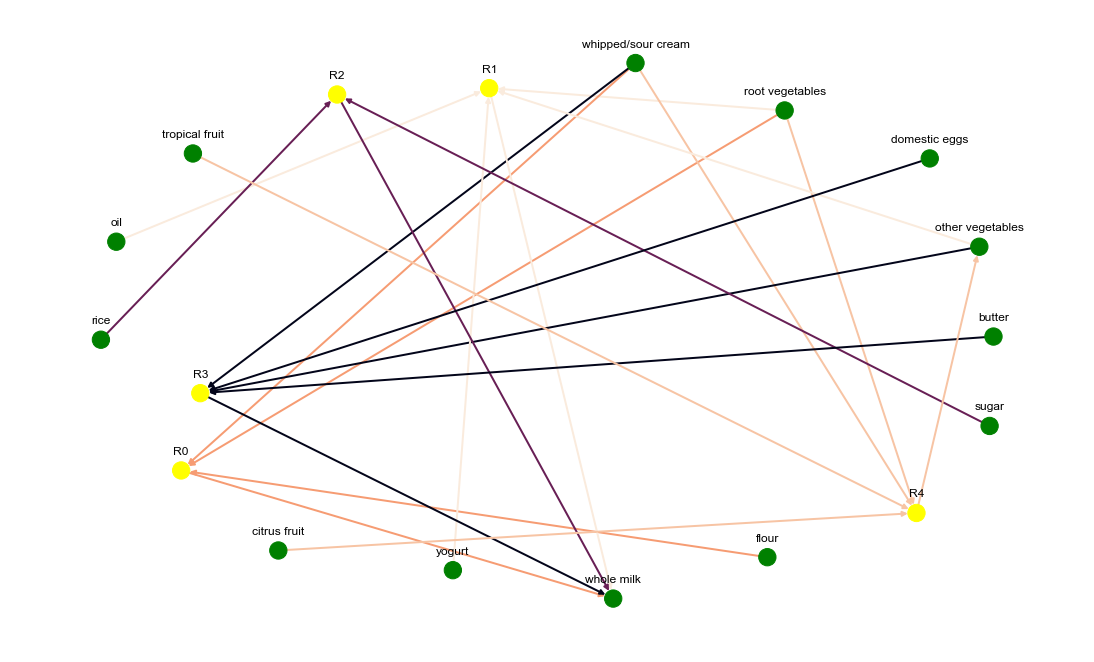
Ans:

The first rule is {root vegetables, flour, whipped/sour cream → whole milk} may tell us that people tend to cook bisque soup. As well as the vegetables and flour, milk is a necessary ingredient for bisque soup.

1. (10%) Give a visualization graph of your association rules.

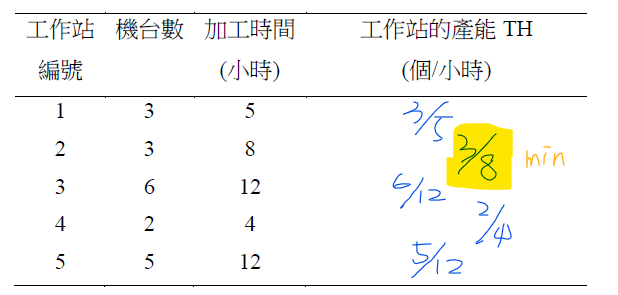
Ans:

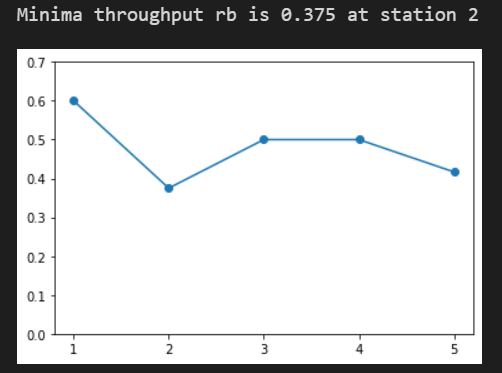
For the clearness, I show the graph from top 5 rules. We can find an interesting causal relationship: most edges point from “Rule node” to “whole milk” and “other vegetables”.



1. **(30%) Manufacturing System Analysis**
2. (10%)根據 Little’s Law 試 計算各工作站的產出率 TH於下表，試問瓶頸站的產出率 𝑟𝑏、最小生產週期時間 (總加工時間， 𝑇0)、關鍵在製品水準 (𝑊0)各為多少？

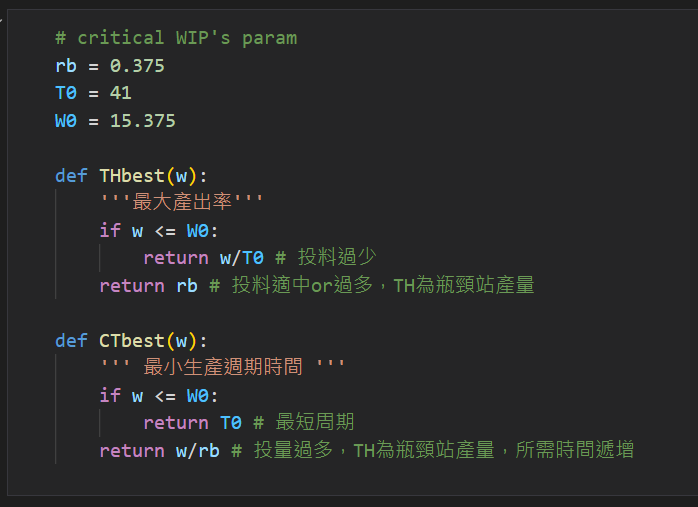
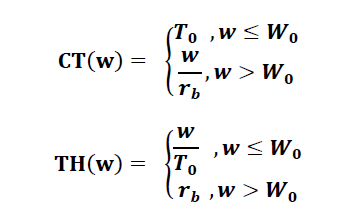
Ans: 𝑟𝑏 = 0.375, 𝑇0 = 41, 𝑊0 = 15.375.  
(𝑊0 = 𝑟𝑏 \* 𝑇0)





1. (10%)試給出最佳績效 (best case)下，最大的產出率 (THbest)與最小生產週期時間(CTbest)的計算公式 (提示講義22~29頁)

Ans: 如下方程式碼。



1. (10%)根據該問題的產線，試程式撰寫建立一模擬模型(或用套裝軟體、數值分析)來驗證，當在製品 WIP數量超過工廠產能時，其生產週期將嚴重惡化。也就是當產線的投料速度(投產量)大於產線的產出率，此時生產系統將處於非穩態的狀態(non steady state)。試用圖表呈現WIP、 CT與 TH之間惡化的關係。(提示講義22~29頁)

Ans: 如下圖，W0時WIP = 15.375是critical wip，為圖中的轉折點。

