* 月消費是否呈現常態分佈
* 利用相關係數探討月消費與各個變數有高度相關(r>0.5)
* 月消費與非連續變數是否有相關?(職業、性別、城市……) ，有的話呈現該相關性的散佈圖或合鬚圖
* 將客戶的性別、已婚未婚改成0或1編碼 ，將教育程度轉換成有序編碼.(分別為1~5)
* 將年收入做平均值和標準差
* 取五種數據分別為”客戶擁有的汽車數量、客戶居住在家裡孩子的數量、客戶擁有孩子的數量、年收入、客戶的性別”做特徵標準化，將畫圖從高到低去畫出特徵重要性。
* 挑出連續型的特徵篩選True的變數與特徵工程做變數，並且分拆數據集去創建線性回歸，再使用DateFrame方式進行排序。
* 此同時依照選擇的五種型別做MSE以及R^2，之後再來分拆數據集，利用20%驗證來分析，進行模型評估

透過殘差圖、判定係數等等來評估模型，並且做K-fold，之 後去修正，最後得出最佳化模型，並預測微軟每月消費預測