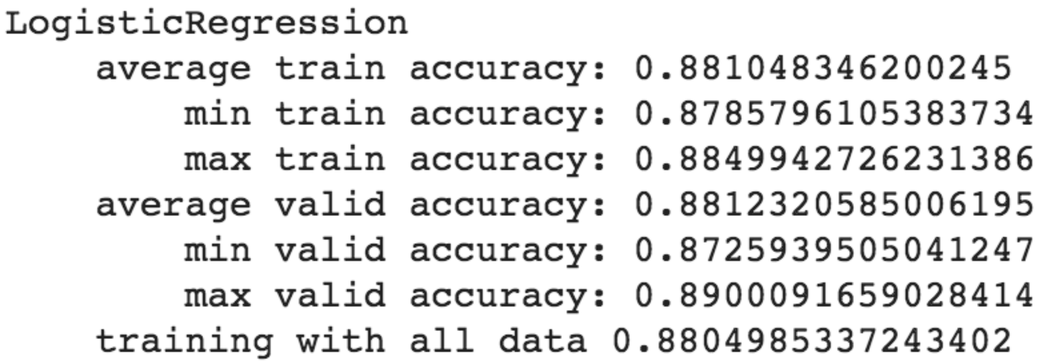
1. Dataset：Travel Review Ratings Data Set
2. Analyze the data：Category 11 以及 User 為 object，Category 12 以及 Category 24 有缺失，其餘皆為 5456 筆 float64 的資料。
3. Define problem：

   使用者對健身類的平均評價是否高於食物類的平均評價
4. Original result：



```
LogisticRegression
    average train accuracy: 0.881048346200245
        min train accuracy: 0.8785796105383734
        max train accuracy: 0.8849942726231386
    average valid accuracy: 0.8812320585006195
        min valid accuracy: 0.8725939505041247
        max valid accuracy: 0.8900091659028414
    training with all data 0.8804985337243402
```

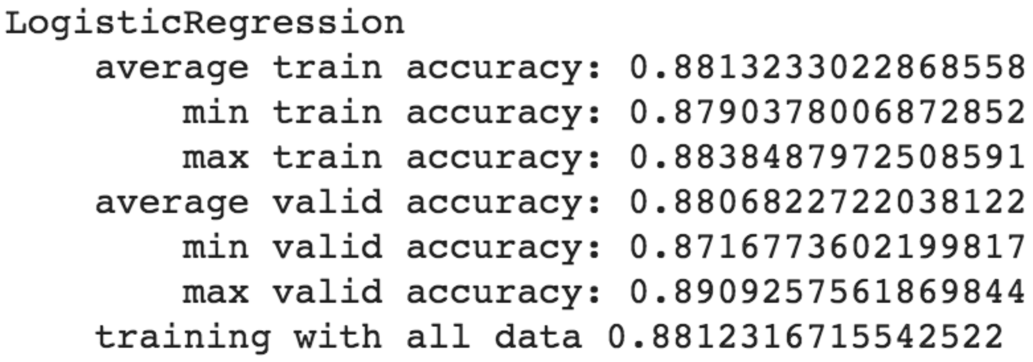
5. Reason：

使用 Logistic Regression，training data 將較靜態的活動地點，以及比較需要活動的地點，以及剩下的類別分成 Static 和 Dynamic 兩類以及其他類，缺失資料以平均值填補。

6. My approaches 1：

   training data 不要特別再分類，按照原本 dataset 給的分類，accuracy 從 0.88 提高 0.001 左右。

   Improvement 1：



```
LogisticRegression
    average train accuracy: 0.8813233022868558
        min train accuracy: 0.8790378006872852
        max train accuracy: 0.8838487972508591
    average valid accuracy: 0.8806822722038122
        min valid accuracy: 0.8716773602199817
        max valid accuracy: 0.8909257561869844
    training with all data 0.8812316715542522
```

   My approaches 2：

   改成使用 AdaBoost 分類，accuracy 從 0.881 提高到 0.93。

   Improvement 2：

```
AdaBoost
    average train accuracy: 0.9315890714719937
        min train accuracy: 0.9282932416953036
        max train accuracy: 0.9337915234822451
    average valid accuracy: 0.9213701513884832
        min valid accuracy: 0.9120073327222732
        max valid accuracy: 0.9349220898258478
    training with all data 0.9305351906158358
```