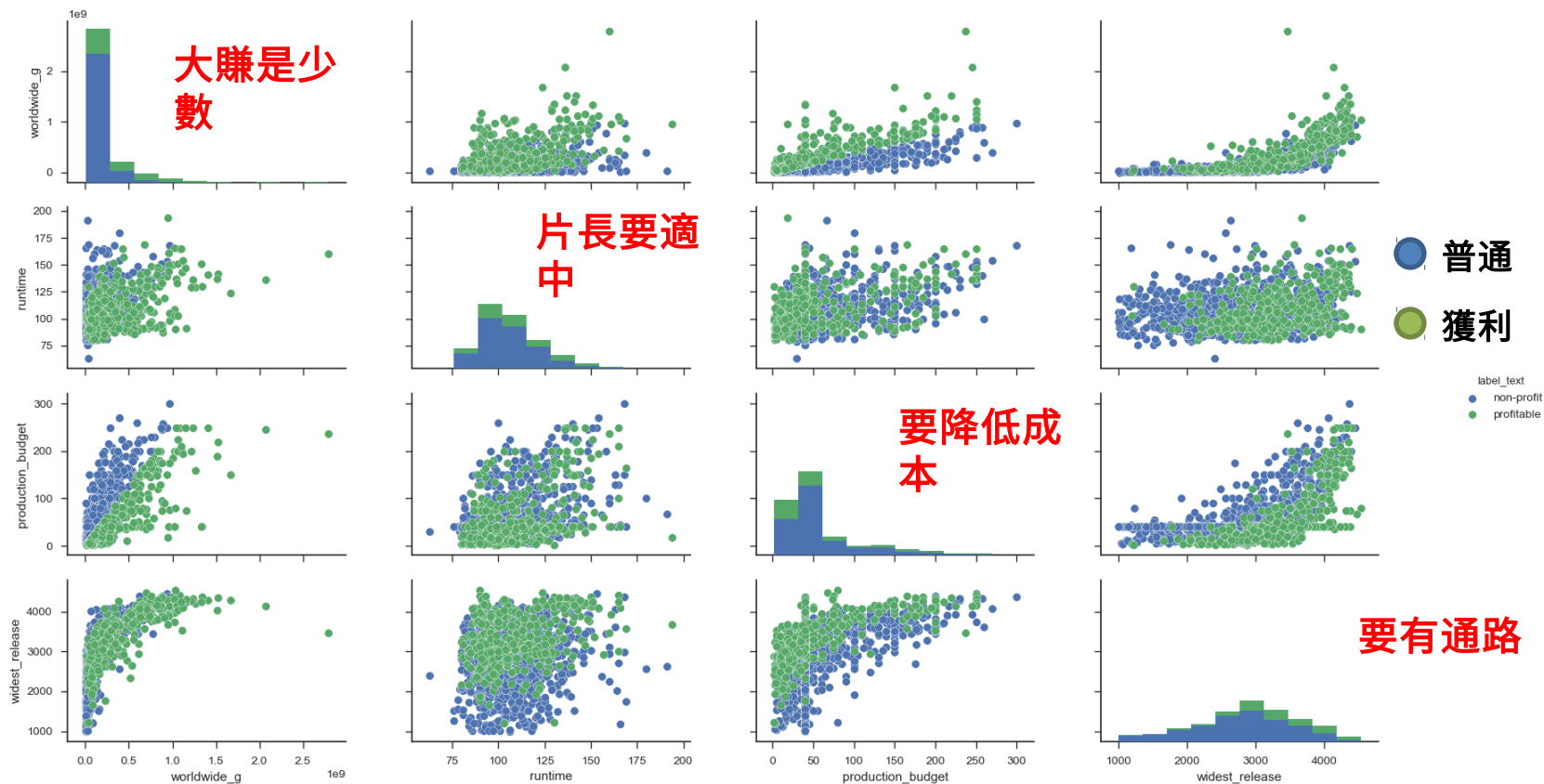


電影票房預測系統

Exploratory data analysis

(探索性數據分析)



全球票房

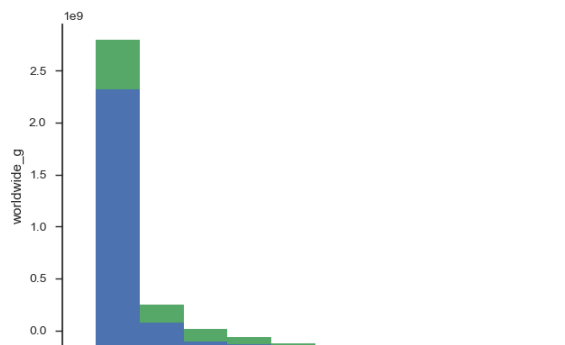
電影片長

拍片預算

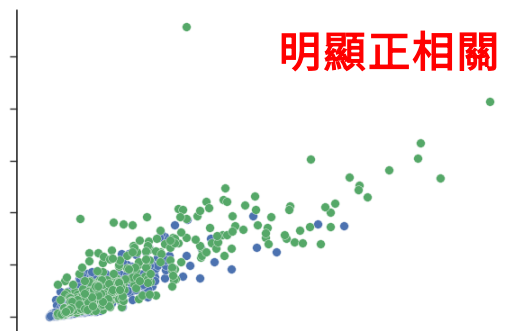
上映戲院
數

Exploratory data analysis

全球票房

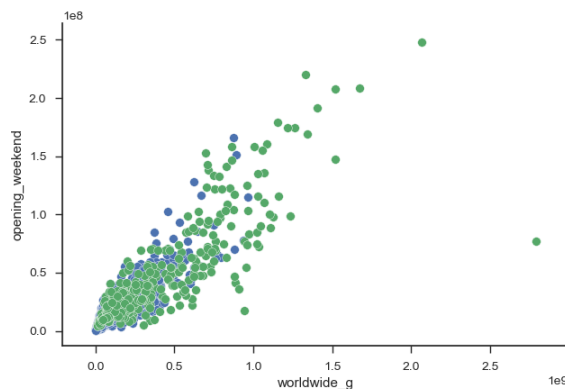


明顯正相關

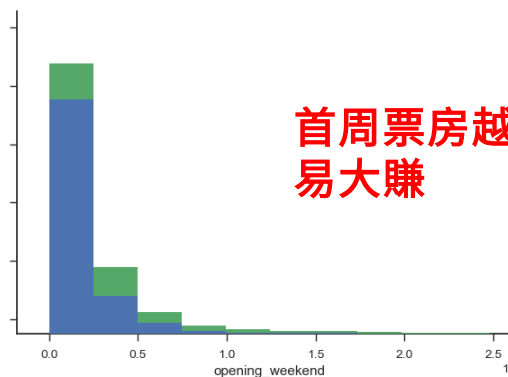


● 普通
● 獲利

首周票房



首周票房越高越容易大賺



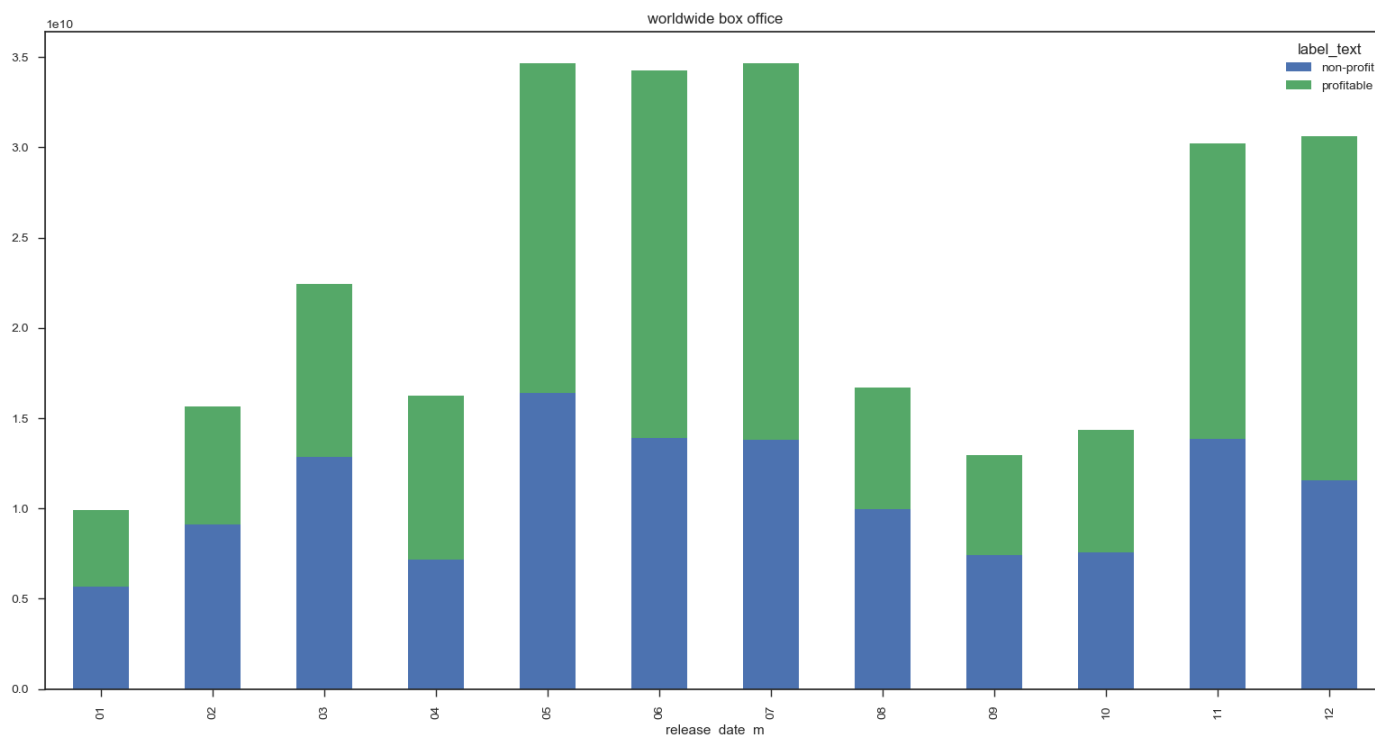
label_text
● non-profit
● profitable

全球票房

首周票房

Exploratory data analysis

各月份票房



暑假旺季 5,6,7
月

聖誕節旺季 11,12
月

Exploratory data analysis

依電影類型

有基本觀
眾群

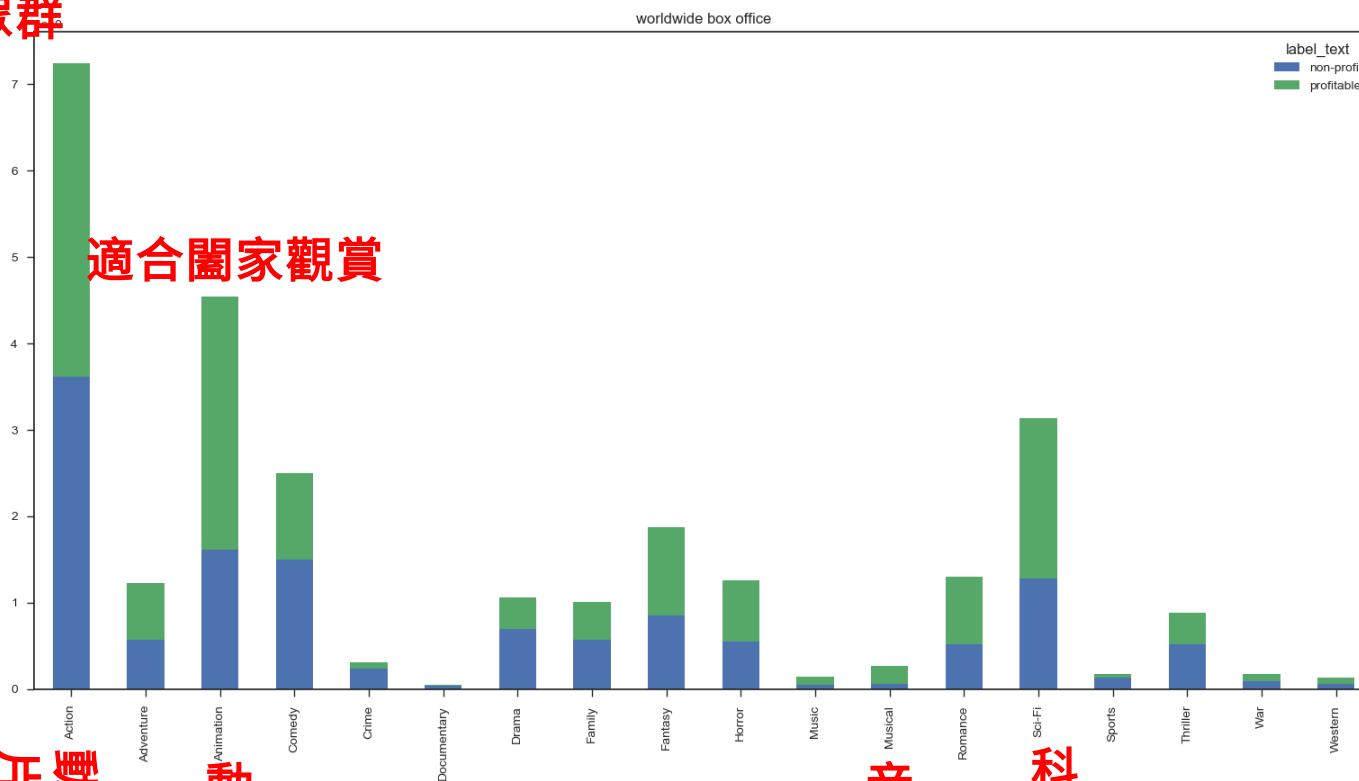
適合闖家觀賞

動作片

動畫片

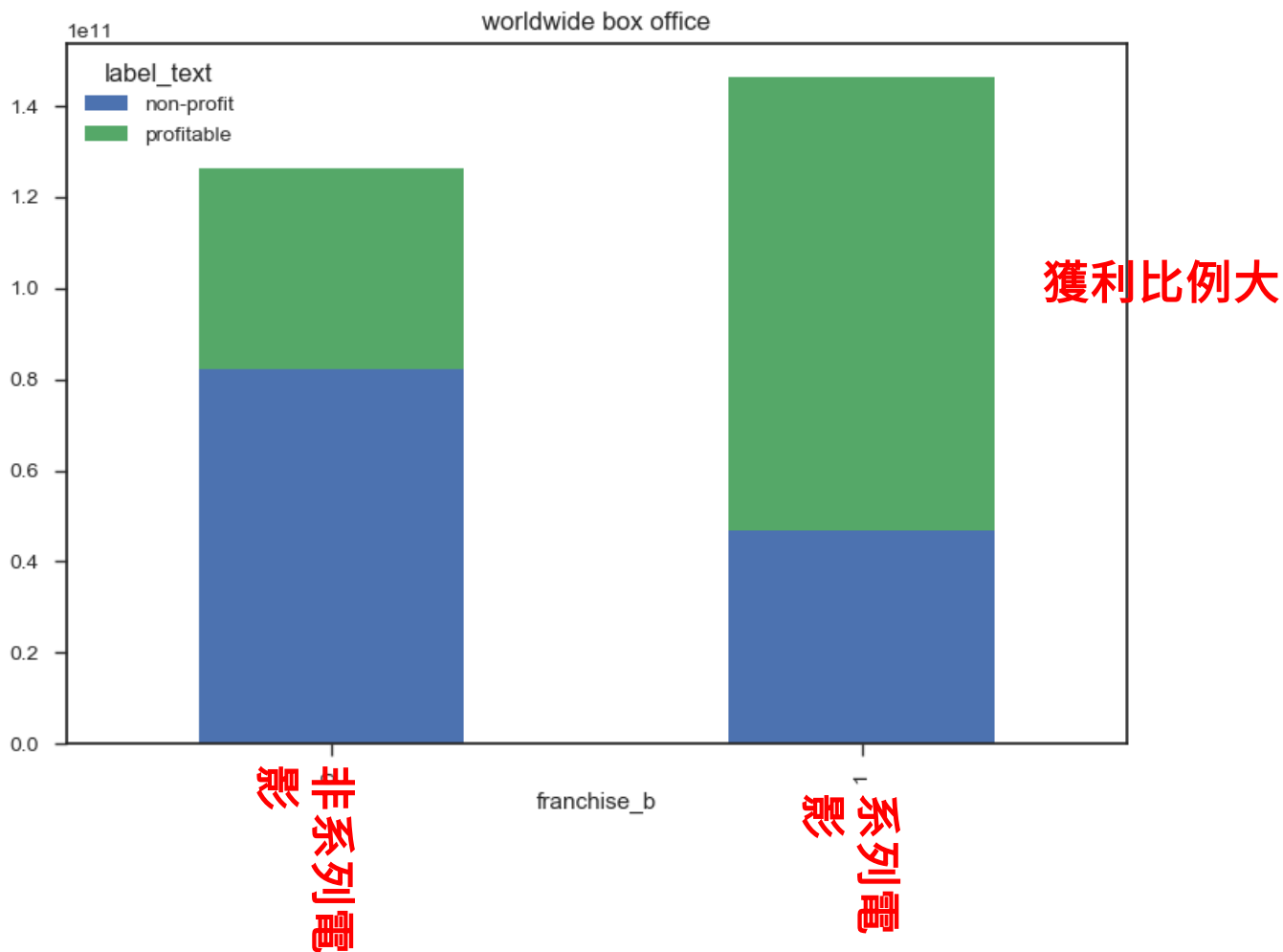
音樂劇

科幻片



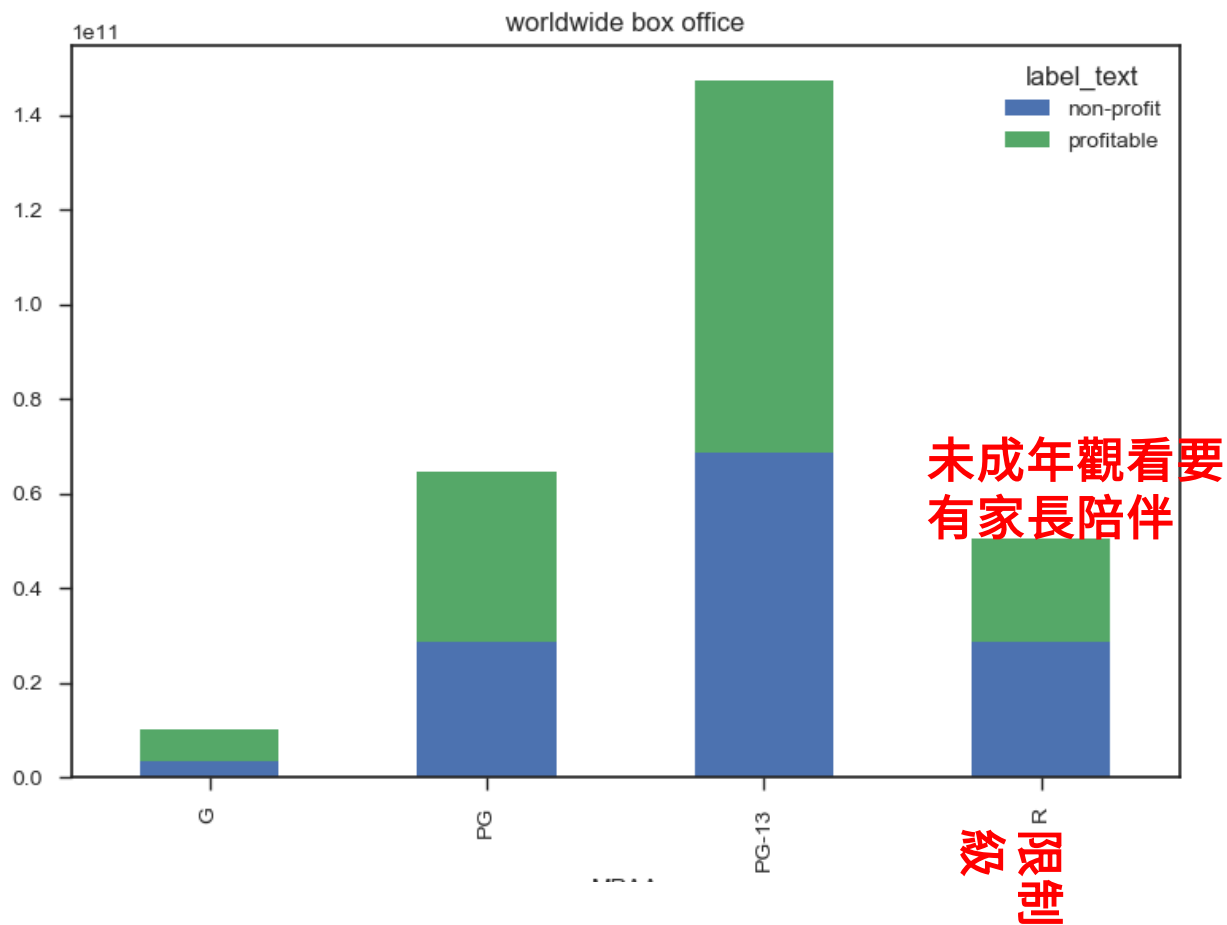
Exploratory data analysis

依系列電影分類



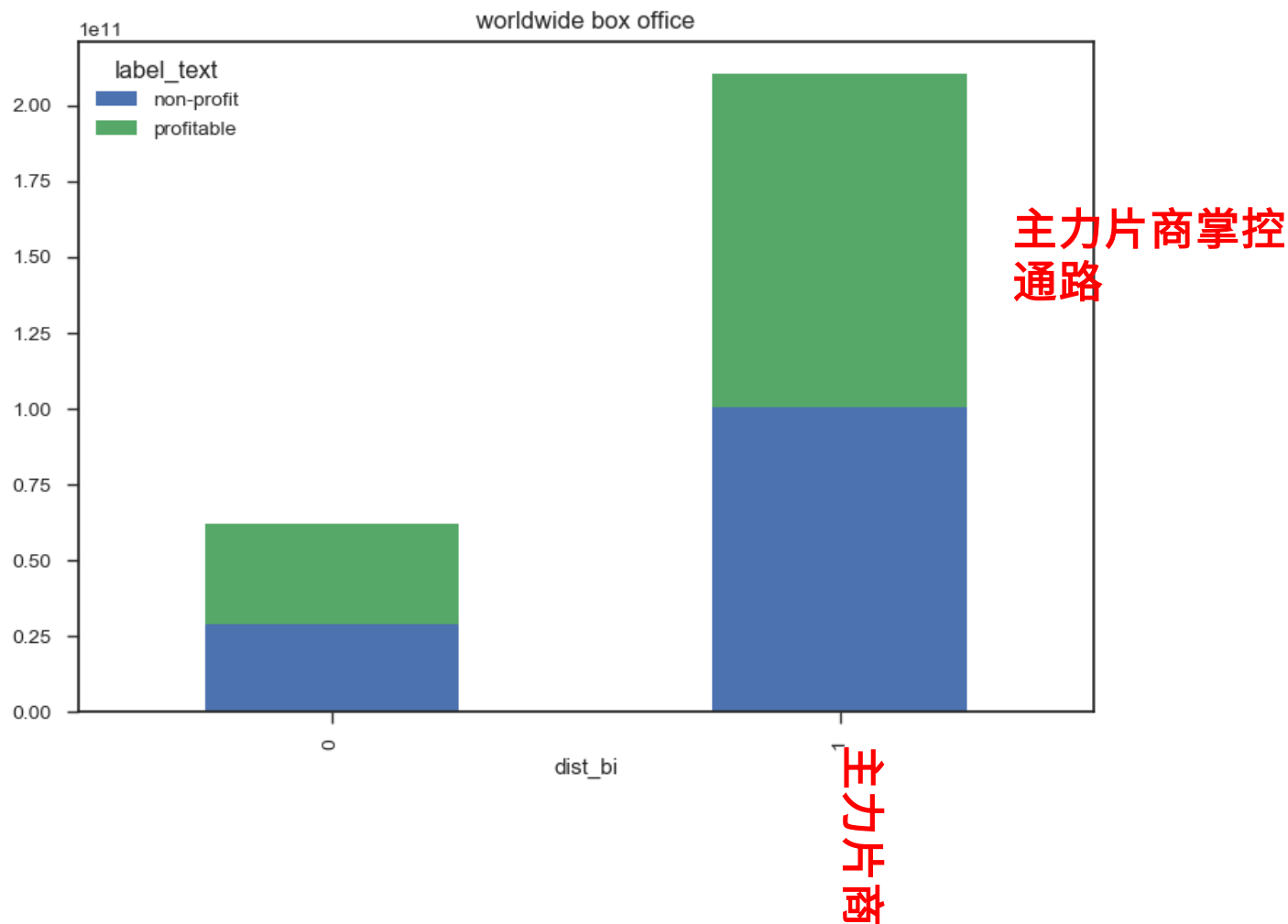
Exploratory data analysis

依電影分級分類



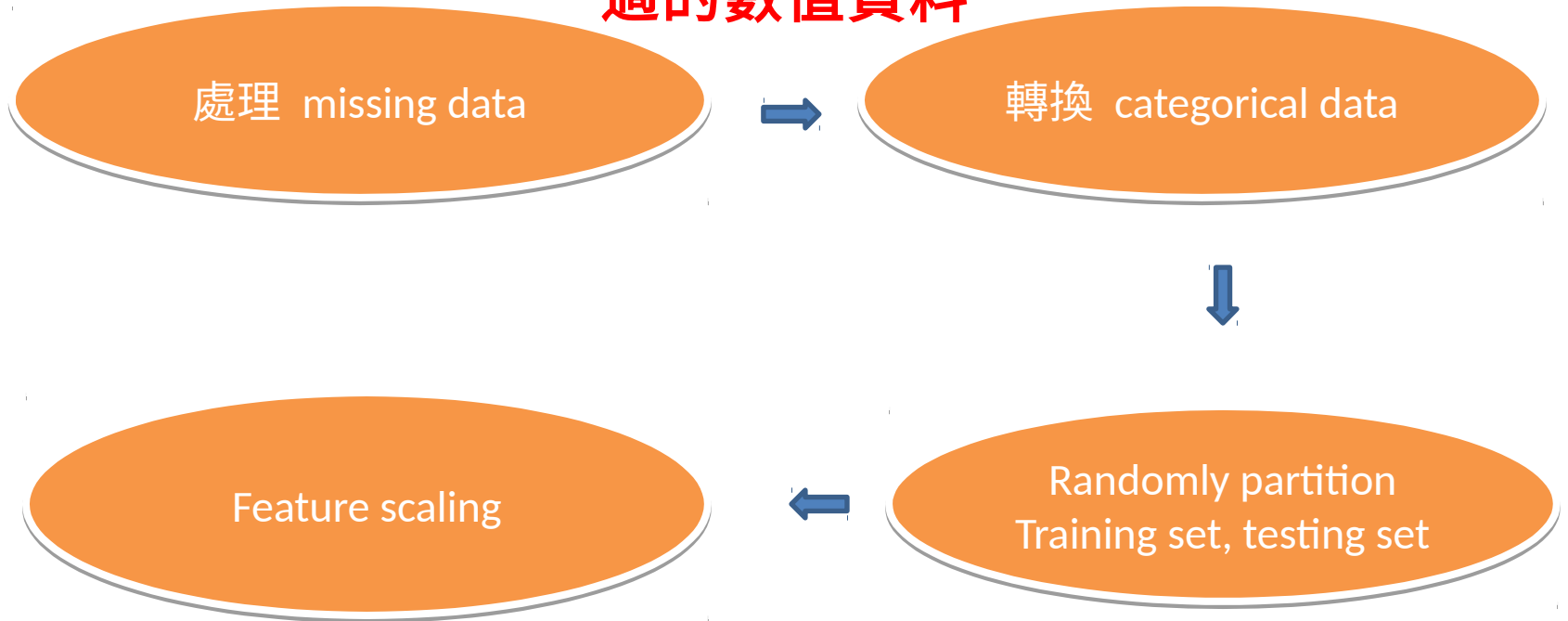
Exploratory data analysis

依片商分類



資料前處理

訓練模型需要合適的數值資料



Random Forest(隨機森林)

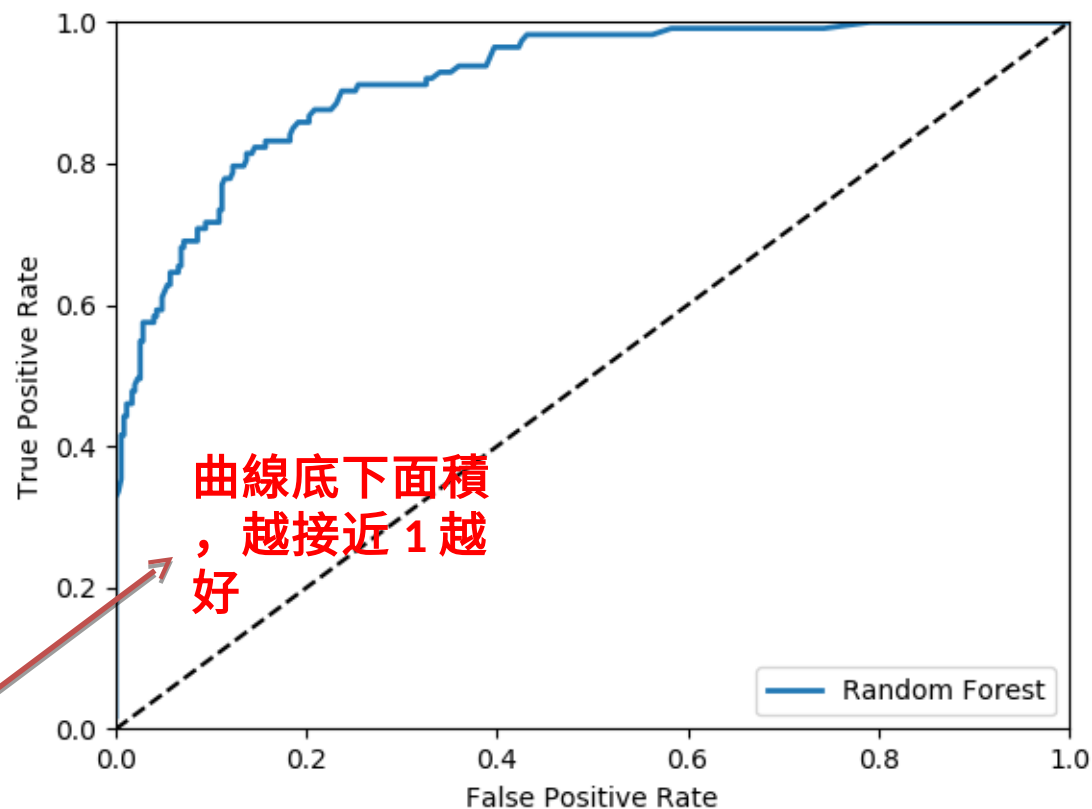
電影開拍前

評量指標	數值
Accuracy(train)	82.4%
Accuracy(test)	81.2%
Roc_auc	0.837

電影上映後

評量指標	數值
Accuracy(train)	89.5%
Accuracy(test)	87.0%
Roc_auc	0.917

Receiver Operating Characteristic (ROC)



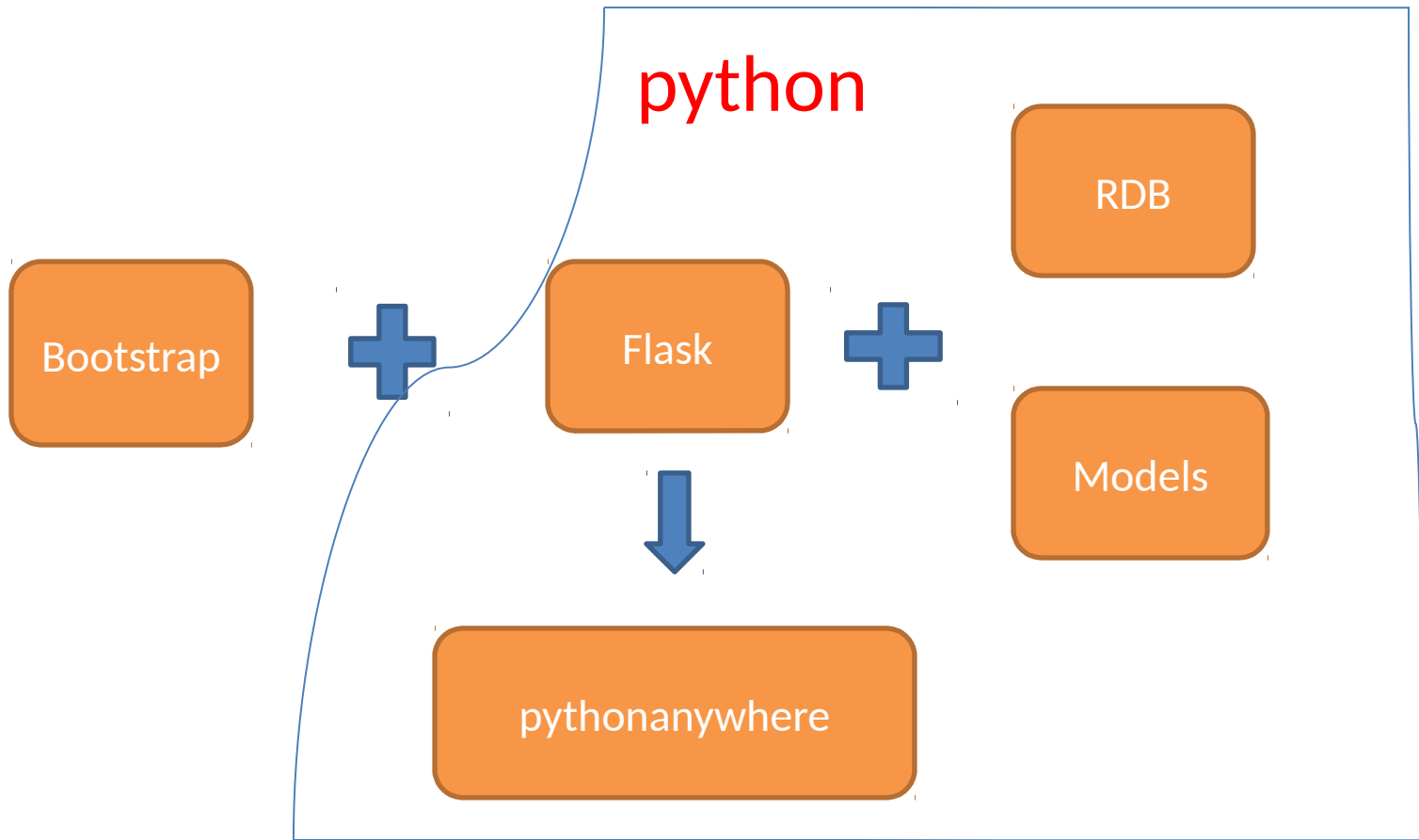
Random Forest

Feature importances

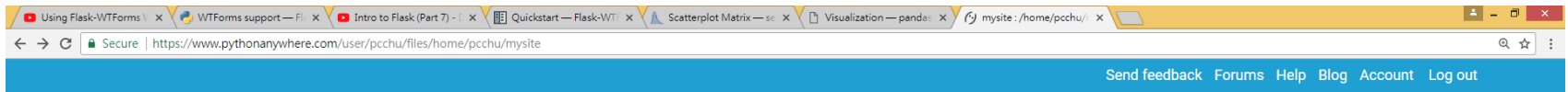
數值越高越重要

變數	數值
1) 首周票房	0.398510
2) 拍片預算	0.269303
3) 上映戲院數	0.084572
4) 電影片長	0.072779
5) 系列電影	0.016983
6) 電影分級 _PG	0.013339

網站建置



Pythonanywhere 雲端



我的
帳號

pythonanywhere

/home/pcchu/ mysite

Directories

Enter new directory name

New directory

__pycache__/
pkl_objects/
static/
templates/

模型放
在這裡

Files

Enter new file name, eg hello.py

New file

__init__.py	2018-04-02 10:49	25 bytes
flaskr.db	2018-04-02 10:49	3.0 KB
flaskr.py	2018-04-08 13:39	10.4 KB
fore_forms.py	2018-04-10 00:25	4.9 KB
reviews.sqlite	2018-04-09 10:21	16.0 KB
vectorizer.py	2018-04-02 10:50	821 bytes
vectorizer.pyc	2018-04-02 11:49	1.0 KB

Upload a file

Dashboard Consoles Files Web Tasks Databases

Open Bash console here

17% full - 88.1 MB of your 512.0 MB quota

目前使
用量

免費帳號的
硬碟容量

