3-1

實際操作圖



程式碼:

#3-1 於dp001平台的瀏覽影片時，操作行為名稱為「暫停」總共有幾次?

#以PseudoID 為71 的學生資料與影片操作的行為名稱為 暫停 資料內為 paused 進行過濾。

#印出過濾結果的長度。

import pandas as pd

df = pd.read\_csv('/content/edu\_bigdata\_imp1.csv', encoding='big5',low\_memory=False)

#跟據 PseudoID 為 71 過濾 DataFrame

df\_filtered = df[(df['PseudoID'] == 71)]

#跟據 dp001\_record\_plus\_view\_action 為 paused 過濾 DataFrame

df\_filtered = df\_filtered[(df\_filtered['dp001\_record\_plus\_view\_action'] == 'paused')]

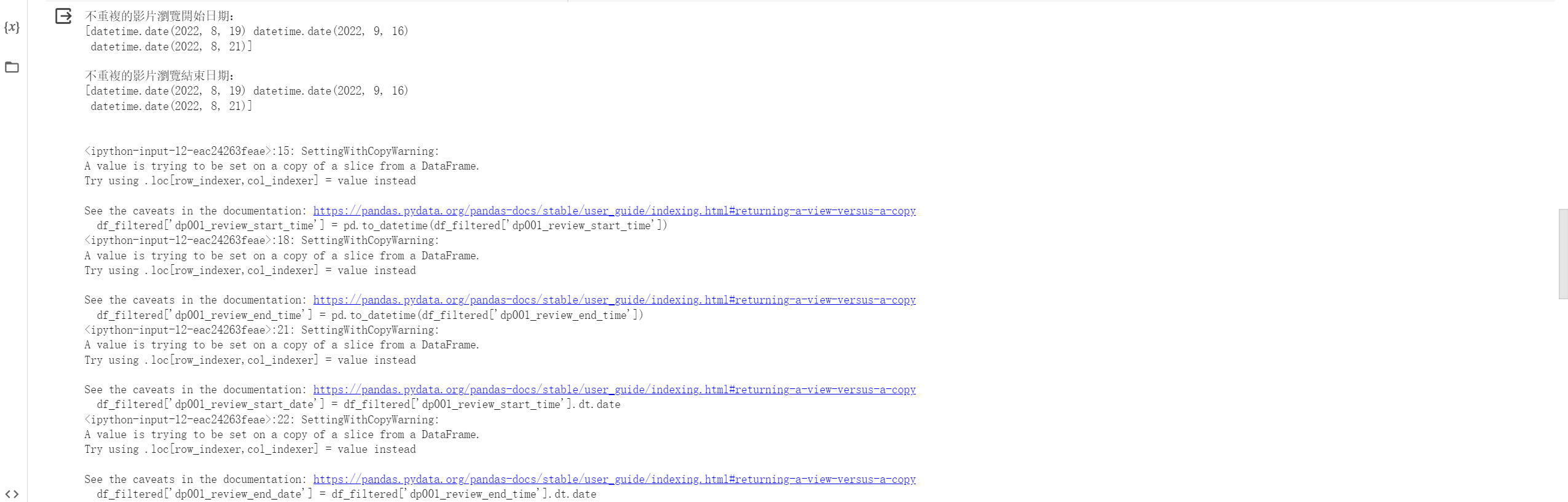
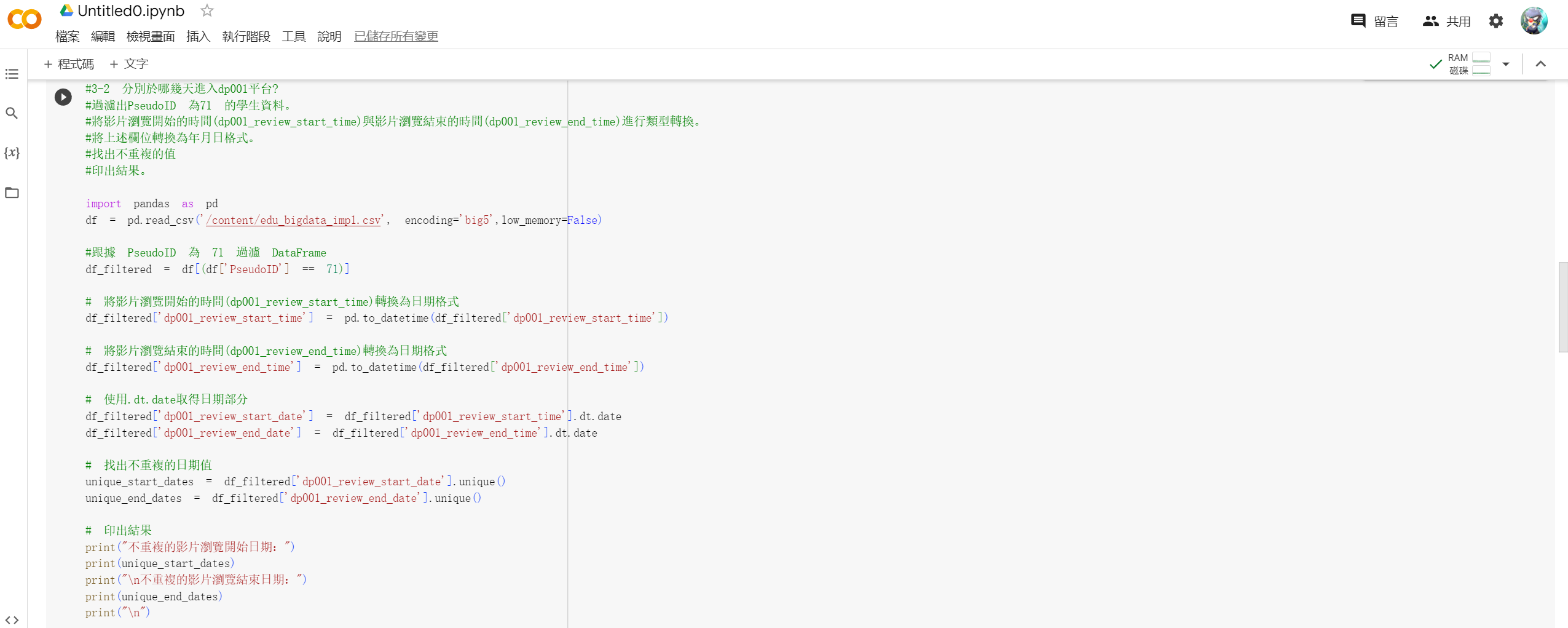
print(len(df\_filtered))

輸出:

578

3-2

實際操作圖



程式碼:

#3-2 分別於哪幾天進入dp001平台?

#過濾出PseudoID 為71 的學生資料。

#將影片瀏覽開始的時間(dp001\_review\_start\_time)與影片瀏覽結束的時間(dp001\_review\_end\_time)進行類型轉換。

#將上述欄位轉換為年月日格式。

#找出不重複的值

#印出結果。

import pandas as pd

df = pd.read\_csv('/content/edu\_bigdata\_imp1.csv', encoding='big5',low\_memory=False)

#跟據 PseudoID 為 71 過濾 DataFrame

df\_filtered = df[(df['PseudoID'] == 71)]

# 將影片瀏覽開始的時間(dp001\_review\_start\_time)轉換為日期格式

df\_filtered['dp001\_review\_start\_time'] = pd.to\_datetime(df\_filtered['dp001\_review\_start\_time'])

# 將影片瀏覽結束的時間(dp001\_review\_end\_time)轉換為日期格式

df\_filtered['dp001\_review\_end\_time'] = pd.to\_datetime(df\_filtered['dp001\_review\_end\_time'])

# 使用.dt.date取得日期部分

df\_filtered['dp001\_review\_start\_date'] = df\_filtered['dp001\_review\_start\_time'].dt.date

df\_filtered['dp001\_review\_end\_date'] = df\_filtered['dp001\_review\_end\_time'].dt.date

# 找出不重複的日期值

unique\_start\_dates = df\_filtered['dp001\_review\_start\_date'].unique()

unique\_end\_dates = df\_filtered['dp001\_review\_end\_date'].unique()

# 印出結果

print("不重複的影片瀏覽開始日期：")

print(unique\_start\_dates)

print("\n不重複的影片瀏覽結束日期：")

print(unique\_end\_dates)

print("\n")

輸出:

不重複的影片瀏覽開始日期：

[datetime.date(2022, 8, 19) datetime.date(2022, 9, 16)

datetime.date(2022, 8, 21)]

不重複的影片瀏覽結束日期：

[datetime.date(2022, 8, 19) datetime.date(2022, 9, 16)

datetime.date(2022, 8, 21)]

<ipython-input-12-eac24263feae>:15: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy>

df\_filtered['dp001\_review\_start\_time'] = pd.to\_datetime(df\_filtered['dp001\_review\_start\_time'])

<ipython-input-12-eac24263feae>:18: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy>

df\_filtered['dp001\_review\_end\_time'] = pd.to\_datetime(df\_filtered['dp001\_review\_end\_time'])

<ipython-input-12-eac24263feae>:21: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy>

df\_filtered['dp001\_review\_start\_date'] = df\_filtered['dp001\_review\_start\_time'].dt.date

<ipython-input-12-eac24263feae>:22: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy>

df\_filtered['dp001\_review\_end\_date'] = df\_filtered['dp001\_review\_end\_time'].dt.date

4-1

實際操作圖



程式碼:

#4-1 找出在dp001平台中，最多影片瀏覽行為的影片序號

#取得影片瀏覽序號欄位(dp001\_review\_sn),相同值的資料筆數。

#取得出现次數最多的值及出現次數。

#印出結果

import pandas as pd

df = pd.read\_csv('/content/edu\_bigdata\_imp1.csv', encoding='big5',low\_memory=False)

# 使用value\_counts()計算影片序號的出現次數

video\_sn\_counts = df['dp001\_review\_sn'].value\_counts()

# 找到出現次數最多的序號[.idxmax() 返回最大值的索引] 及 出現次數[.max()反回最大值]

most\_viewed\_video\_sn = video\_sn\_counts.idxmax()

max\_view\_count = video\_sn\_counts.max()

# 印出結果

print("最多影片瀏覽行為的影片序號是：", most\_viewed\_video\_sn)

print("出現次數：", max\_view\_count)

輸出:

最多影片瀏覽行為的影片序號是： 11043810.0

出現次數： 12410

4-2

實際操作圖



程式碼:

#4-2 請找出在dp002平台中，操作資源的知識架構分類中為「十二年國民基本教育類」總共有幾筆?

#過濾出操作資源的知識架構分類為十二年國民基本教育類的資料。

#印出結果長度即為答案。

#備註 : 請注意資料中操作資源的知識架構分類(dp002\_extensions\_alignment)的值,

#欄位中的值有標點符號存在,過濾時需注意。

import pandas as pd

df = pd.read\_csv('/content/edu\_bigdata\_imp1.csv', encoding='big5',low\_memory=False)

# 使用str.contains()方法過濾操作資源的知識架構分類為「十二年國民基本教育類」的資料

filtered\_data = df[df['dp002\_extensions\_alignment'].str.contains('十二年國民基本教育類', na=False)]

#備註 : str.contains()方法的主要部分，其中包含要搜索的目標字串，當na=False時，即不將包含NaN值的行視為包含指定字串。

# 印出過濾後的資料筆數即為答案

total\_records = len(filtered\_data)

print(total\_records)

輸出:

229

4-3

實際操作圖



程式碼:

#4-3 請找出在dp002平臺中，前3個最常發生的操作行為名稱

#從操作行為名稱欄位(dp002\_verb\_display\_zh\_TW) 過濾掉NA 或NAN 的資料。

#為欄位中的值進行計數。

#印出出现次數前三多的值及其出现次數

import pandas as pd

df = pd.read\_csv('/content/edu\_bigdata\_imp1.csv', encoding='big5',low\_memory=False)

# 過濾掉"dp002\_verb\_display\_zh\_TW"列中的NA或NAN值

filtered\_data = df.dropna(subset=['dp002\_verb\_display\_zh\_TW'])

# 使用value\_counts()方法計算操作行為名稱的出現次數

verb\_counts = filtered\_data['dp002\_verb\_display\_zh\_TW'].value\_counts()

# 找到出現次數前三多的值及出現次數

top\_3\_verbs = verb\_counts.head(3)

# 印出結果

print("前3個最常發生的操作行為名稱是：")

print(top\_3\_verbs)

輸出:

前3個最常發生的操作行為名稱是：

查看 126766

查詢參考 1564

下載 497

Name: dp002\_verb\_display\_zh\_TW, dtype: int64

4-4

實際操作圖



程式碼:

#4-4 請找出在dp002平台中，操作資源的知識架構分類中為「校園職業安全」總共有幾筆?

#過濾出操作資源的知識架構分類(dp002\_extensions\_alignment)「為校園職業安全」的資料。

#印出結果長度即為答案。

#備註 請注意資料中操作資源的知識架構分類的值，欄位中的值有標點符號存在，過濾時需注意。

import pandas as pd

df = pd.read\_csv('/content/edu\_bigdata\_imp1.csv', encoding='big5',low\_memory=False)

# 使用str.contains()方法過濾操作資源的知識架構分類為「校園職業安全」的資料

filtered\_data = df[df['dp002\_extensions\_alignment'].str.contains('校園職業安全', na=False)]

# 印出過濾後的資料筆數即為答案

total\_records = len(filtered\_data)

print(total\_records)

輸出:

5375

上述實際操作圖的程式碼網址 (怕網址掛掉附圖和程式碼在word檔):

https://colab.research.google.com/drive/1xFNr7vL6HFZRBVWTIPrzmgJ2zI6tBnId?hl=zh-tw#scrollTo=mkbe5Rs6lKfg