

การประมวลผลข้อมูล รายละเอียดของเกมจาก ร้านค้า Steam



สมาชิก

กฤตเดช	เจียวยิระกุล	6104062630174
นครินทร์	สุขแพ	6204062610067
กิตติณาน	ศิริโสภณา	6204062620101



AWK

AWK

หาอัตราส่วนระหว่างคะแนนรีวิว
ด้านบวกกับด้านลบแล้วนำมาเพิ่ม
ลงฟิลด์ ratio_ratings ใหม่



Begin Processing

-

Body Processing

กำหนดฟิลด์ใหม่ \$19 ชื่อ ratio_ratings

End Processing

กำหนด if else โดย

ถ้าค่าคะแนนรีวิวด้านบวกเป็น 0 ให้แทนด้วย N/A

ถ้าค่ามากกว่า 0 จะคำนวณโดย $\$19 = \$13 / (\$13 + \$14)$

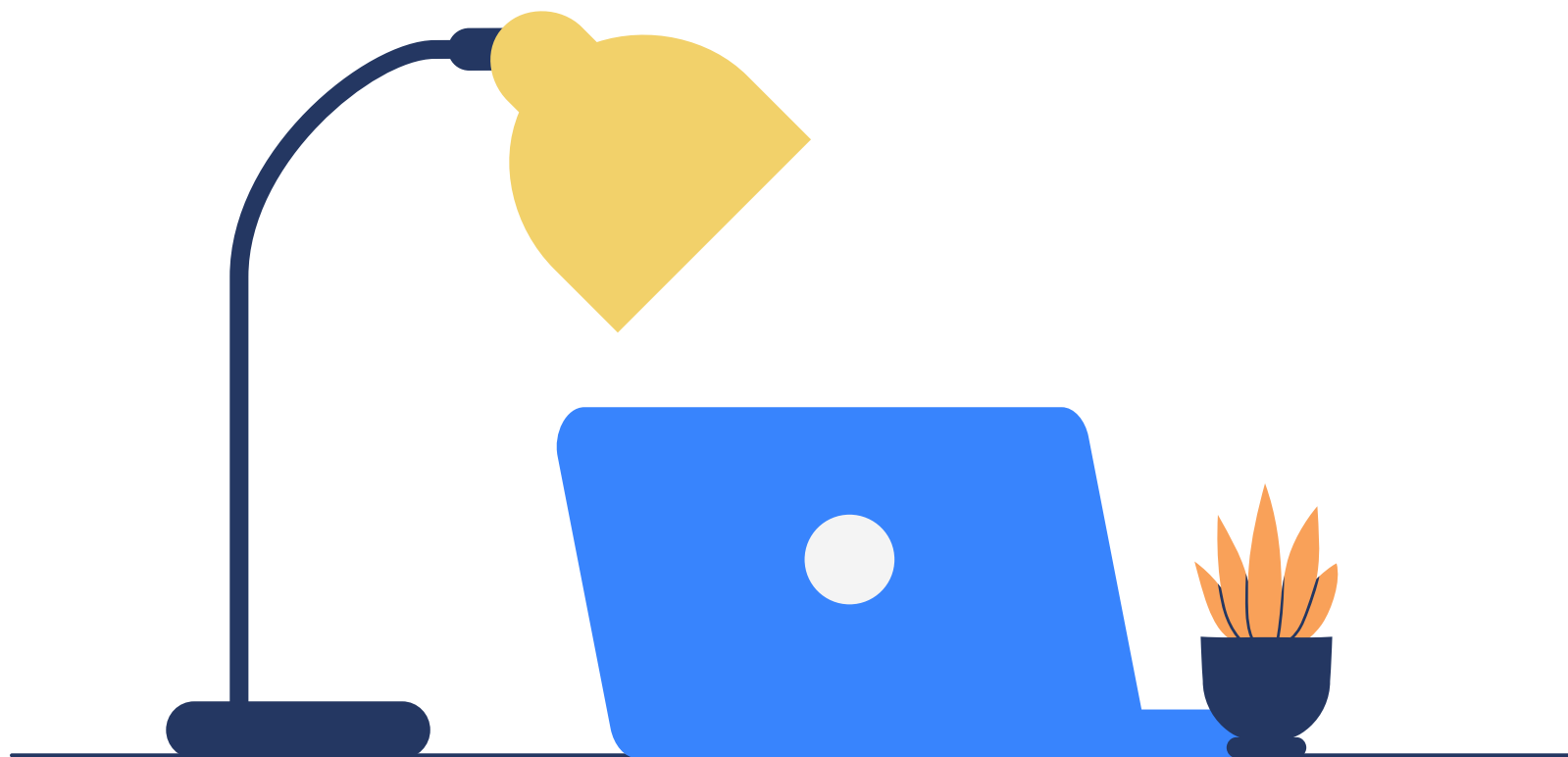
แล้วนำข้อมูลทั้งหมดมาแสดงผลใหม่

coding:

```
awk -F '|' 'NR==1{$19="ratio_ratings"}NR>1{if($13==0)
$19 = "N/A"; else $19 = $13/($13+$14)} {print $2 "|" $13
"|" $14 "|" $19> "steam3.csv"}' steam2.csv
```

AWK

แยกไฟล์ข้อมูลเกมที่มีคะแนน
รีวิวด้านบวกมากกว่าค่าเฉลี่ย
และต่ำกว่าค่าเฉลี่ย



Coding

```
awk -F '|' '{if($13>=1001) print $2 "|" $13 > "steam4.csv"}' steam2.csv  
awk -F '|' '{if($13<1001) print $2 "|" $13 > "steam5.csv"}' steam2.csv
```



Begin Processing

สร้างไฟล์เปล่า steam4.csv และ
steam5.csv



Body Processing

บรรทัดที่พบคะแนนรีวิวด้านบวกมากกว่าค่า
เฉลี่ยขึ้นไปให้บันทึกลงในไฟล์ steam4.csv

บรรทัดที่พบคะแนนรีวิวด้านบวกน้อยกว่าค่า
เฉลี่ยลงให้บันทึกลงในไฟล์ steam5.csv



End Processing

-



FILTER

FILTER

แสดงเฉพาะฟิลด์ name release_date developer publisher platforms
required_age genres price โดยแสดงเฉพาะเกมราคาสูงสุดของแต่ละประเภทเกม

ใช้ tail

นำเอาเฉพาะข้อมูลทั้งหมดยกเว้นบรรทัดแรก

ใช้ cut

ให้แสดงเฉพาะฟิลด์ของ name release_date developer publisher platforms required_age genres price

ใช้ sort

โดยเรียงลำดับเฉพาะฟิลด์ที่ 8 คือ price จากมากไปน้อย

ใช้ awk

คัดเอาเฉพาะเฉพาะฟิลด์ที่ 7 คือประเภทเกม เพื่อจัดเรียงให้เกมแต่ละประเภทไม่ซ้ำกัน

```
coding:cat steam.csv|tail +2|cut -d '|' -f2-3,5-8,10,18|sort -n -r -t '|' -k8|awk -F'|' '!a[$7]++'
```

แสดงฟิลด์ name average_playtime owners โดยเฉพาะฟิลด์ที่มีเวลาเล่นเฉลี่ยมากที่สุด 10 อันดับแรกของเกม Action

ใช้ tail

นำเอาเฉพาะข้อมูลทั้งหมดยกเว้นบรรทัดแรก

ใช้ sort

โดยเรียงลำดับเฉพาะฟิลด์ที่ 15 คือ Average_playtime จากมากไปน้อย

ใช้ cut

ให้แสดงเฉพาะฟิลด์ของ name average_playtime genres owners

ใช้ grep

คัดเอาเฉพาะเกมประเภท Action และตัดออกมา

ใช้ head

เอาเฉพาะเกมที่มีเวลาเล่นโดยเฉลี่ยมากที่สุด 10 อันดับแรก

code

```
cat steam2.csv|tail +2|sort -t '|' -  
    k15 -n -r|cut -d '|' -  
f2,15,11,17|grep Action|head  
-10 แสดงเฉพาะฟิลด์ name  
average_playtime genres  
owners
```





SED

A stylized illustration of a person with dark hair, wearing a white long-sleeved shirt and blue trousers, sitting on a yellow chair and working on a laptop at a white desk. The desk also holds a yellow mug and a white printer. A yellow desk lamp is positioned to the right of the laptop. On the wall behind the desk is a yellow calendar grid with several colorful sticky notes (orange, white, yellow) attached to it. To the right of the desk is a blue cylindrical pot containing a plant with orange and grey leaves. The entire scene is set against a plain white background with a thin dark blue horizontal line at the bottom.

```
sed '/^[0-9]\{1,\}[A-Za-z0-9[:space:]]\{1,\}[0-9]\{2\}-[0-9]\{2\}-[0-9]\{2\}|[0-1][A-Za-z0-9[:space:]]\{1,\}[A-Za-z0-9[:space:]]\&[-]\{1,\}[A-Za-z0-9[:space:]]\&[-]\{1,\}[0-9]\{1,\}[A-Za-z0-9[:space:]]\&[-]\{1,\}[A-Za-z0-9[:space:]]\&[-]\{1,\}[0-9]\{1,\}[0-9]\{1,\}[0-9]\{1,\}[0-9]\{1,\}[0-9]\{1,\}[0-9]\{1,\}/d' steam.csv > errordata
```

- 01 ใช้คำสั่ง substitution -> /pattern ที่ถูกต้อง/d
- 02 กำหนด pattern ที่ถูกต้อง จากนั้นตัดข้อมูลที่ถูกต้องตาม pattern ออก และนำข้อมูลที่เหลือที่ไม่ตรงรูปแบบบันทึกลงไฟล์ errordata
- ```
10 ([0-9]\{1,\})
CounterStrike ([A-Za-z0-9[:space:]]\{1,\})
01-11-0 ([0-9]*-[0-9]*-[0-9]\{1,\})
1 ([0-1])
Valve ([A-Za-z0-9[:space:]]\{1,\})
Valve ([A-Za-z0-9[:space:]]\{1,\})
windows;mac;linux ([A-Za-z,;]\{1,\})
0 ([0-9]\{1,\})
Multi-player;Online Multi-Player;Local Multi-Player;Valve Anti-Cheat
enabled ([A-Za-z0-9[:space:]]\/\&\;\-]\{1,\})
Action ([A-Za-z0-9[:space:]]\/\&\;\-]\{1,\})
Action;FPS;Multiplayer ([A-Za-z0-9[:space:]]\/\&\;\-]\{1,\})
0 ([0-9]\{1,\})
124534 ([0-9]\{1,\})
3339 ([0-9]\{1,\})
17612 ([0-9]\{1,\})
317 ([0-9]\{1,\})
100000000-200000000 ([0-9]\{1,\}\-[0-9]\{1,\})
7.19 ([0-9]\{1,\}\.[0-9]\{1,\})
```

# SED

ลบบรรทัดที่มีอักขระนอกเหนือ

จาก A-Z 0-9 | . ; - ออกจาก  
ไฟล์

01 ใช้คำสั่ง substitution -> /patternที่ต้องการ/d

02 กำหนด pattern ที่ถูกต้อง เพื่อใช้ในการค้นหาข้อมูลที่ต้อง  
ตัด จากนั้นตัดข้อมูลที่ต้องการตาม pattern และนำข้อมูลที่  
ตรงรูปแบบออก และเอาข้อมูลที่ไม่ตรงกับรูปแบบเก็บลง  
ไฟล์ raw2.csv

code:sed '/[^A-Za-z0-9\[:space:],\|,\|;\|.\|-\]/d' raw2.csv

