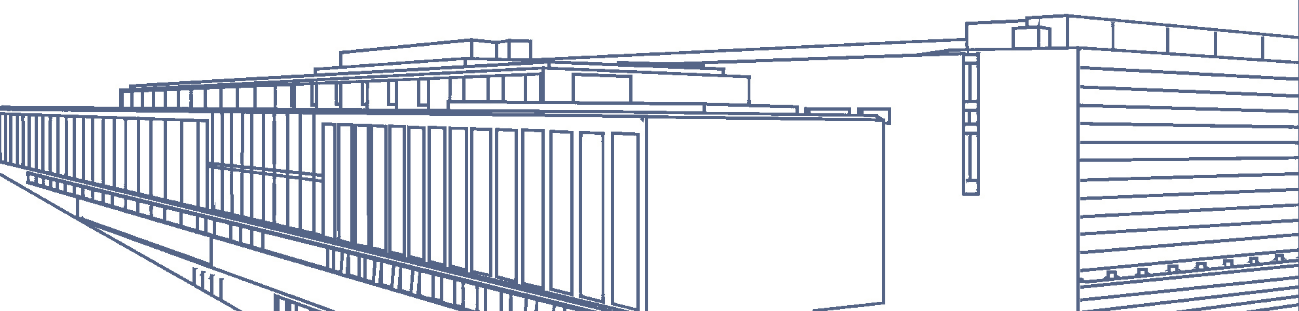
|  |
| --- |
| Sistemes d’Informació a les Organitzacions |
| **Pràctica 1 (PART 1):**  Imputació de valors desconeguts: A la recerca de les millors prediccions |
| 16/03/2020 |

|  |
| --- |
| Victor López Romero  Aleix Sancho Pujals |



Contingut

[1. Enunciat de la pràctica 3](#_Toc35113821)

[2. Treball realitzat 4](#_Toc35113822)

[2.1. Gràfiques 4](#_Toc35113823)

[2.1.1. Recompte de puntuacions en intervals 4](#_Toc35113824)

[2.1.2. Probabilitat de cada puntuació 5](#_Toc35113825)

[2.1.3. Recompte de les mitjanes de puntuacions en intervals 5](#_Toc35113826)

[2.1.4. Positivisme dels usuaris 6](#_Toc35113827)

[2.1.5. Positivisme dels restaurants 7](#_Toc35113828)

[2.1.6. Nombre de visites dels restaurants per la seva mitjana 8](#_Toc35113829)

[2.1.7. Probabilitat que una mateixa puntuació la faci diferents usuaris 8](#_Toc35113830)

[2.1.8. Comparació entre la puntuació mitja dels usuaris i la seva moda (I) 9](#_Toc35113831)

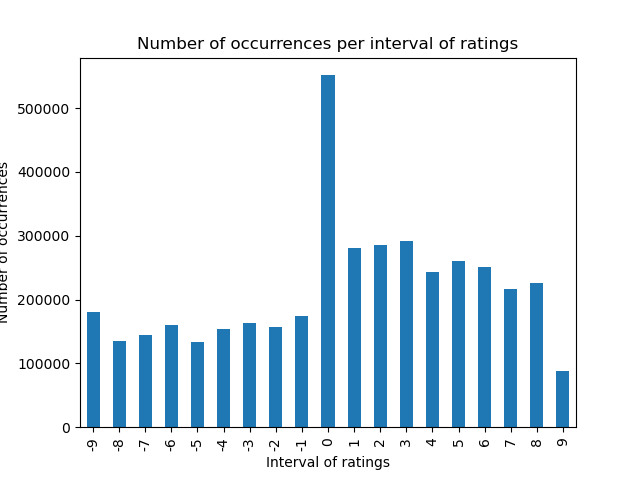
[2.1.9. Comparació entre la puntuació mitja dels usuaris i la seva moda (II) 10](#_Toc35113832)

# Enunciat de la pràctica

# 2. Treball realitzat

## 2.1. Gràfiques

### 2.1.1. Recompte de puntuacions en intervals

****

|  |  |
| --- | --- |
| **Puntuació** | **Vegades que apareix** |
| -10 | 96142 |
| -9 | 149580 |
| -8 | 139251 |
| -7 | 152440 |
| -6 | 145807 |
| -5 | 143189 |
| -4 | 165200 |
| -3 | 154172 |
| -2 | 167809 |
| -1 | 211028 |
| 0 | 299134 |
| 1 | 261463 |
| 2 | 286576 |
| 3 | 290189 |
| 4 | 273478 |
| 5 | 253191 |
| 6 | 259993 |
| 7 | 230491 |
| 8 | 212657 |
| 9 | 204570 |

Figura 1

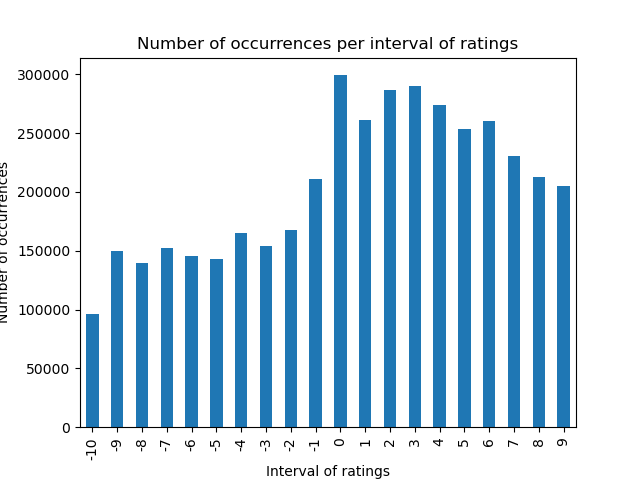


Figura 2

En aquesta gràfica es mostren les dades que hem extret a partir de les diferents puntuacions que han realitzat els usuaris, com tenim moltes dades diferents, hem decidit agrupar aquestes dades en diferents intervals enters ( [-x.99,x.99]) a les dades analitzades. Un cop agrupades en aquests intervals em fet un recompte del nombre de puntuacions que cauen en cada interval, en la Figura 1 es pot veure que la majoria de puntuacions cauen en l’interval [-0.99, 0.00] això es degut que s’ha utilitzat un sistema de truncament de les dades, ja que en la Figura 2 s’han arrodonit les dades i es mostra una major distribució d’aquests puntuacions. Cal dir que les puntuacions al voltant del 10 s’han agrupat al 9 perquè eren molt poques.

Com a conclusió final d’aquestes gràfiques podem concloure que la majora de puntuacions estan en l’interval [-0.99, 0.99], però que la majoria d’elles tendeixen a anar a l’alça. També que entre totes les puntuacions s’han realitzat més puntuacions positives que negatives.

### 2.1.2. Probabilitat de cada puntuació

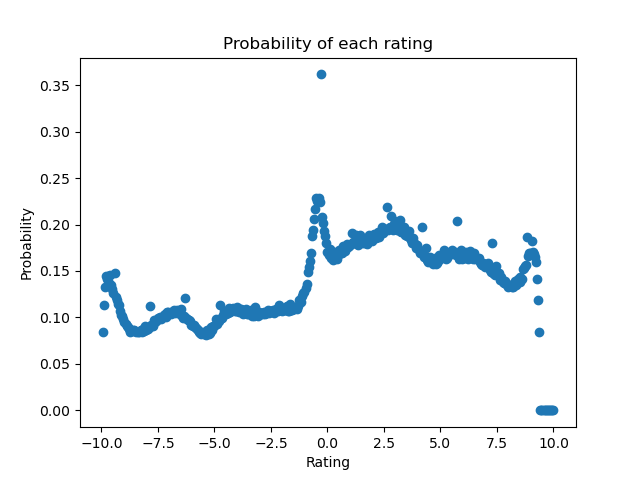
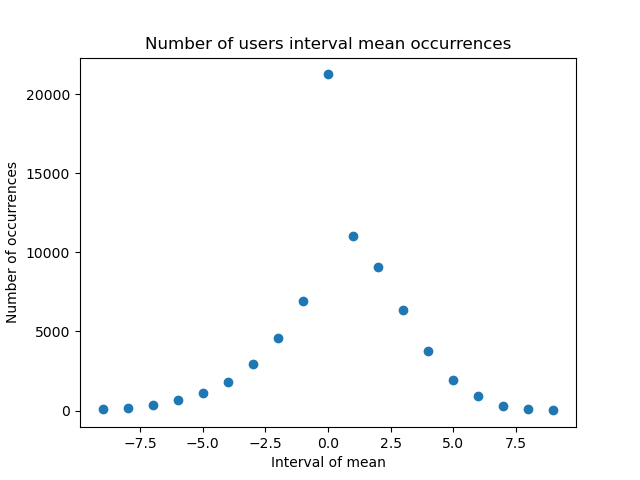
****

Figura 3

A la gràfica anterior es pot observar les probabilitats que te una puntuació de ser escollida. Com es pot observar les puntuacions menys extremes son les més probables de sortir. Aquesta gràfica també es complementa amb les anterior on ens diu que la puntuació que es més probable que es faci es una que estigui sobre el 0 i amb una mica menys de probabilitat qualsevol puntuació positiva que no sigui molt extrema. Tot i així no podem extreure molta informació d’aquesta gràfica, ja que les probabilitats són baixes.

### 2.1.3. Recompte de les mitjanes de puntuacions en intervals

Figura 4



|  |  |
| --- | --- |
| **mean** | **count** |
| -10 | 26 |
| -9 | 100 |
| -8 | 223 |
| -7 | 505 |
| -6 | 854 |
| -5 | 1453 |
| -4 | 2283 |
| -3 | 3663 |
| -2 | 5663 |
| -1 | 8399 |
| 0 | 10868 |
| 1 | 11618 |
| 2 | 10205 |
| 3 | 7781 |
| 4 | 4924 |
| 5 | 2705 |
| 6 | 1410 |
| 7 | 536 |
| 8 | 159 |
| 9 | 46 |

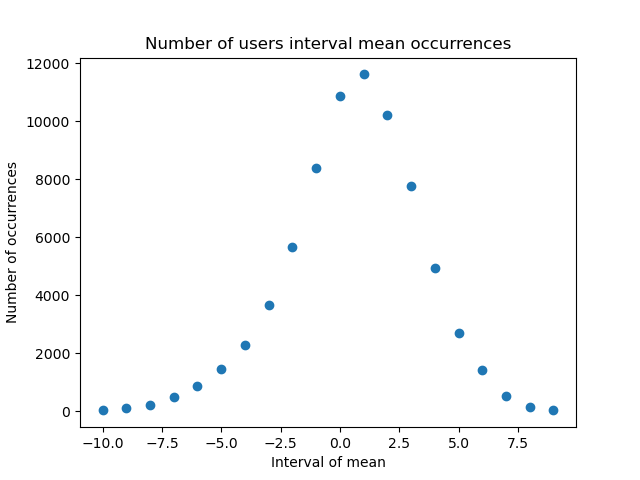
A la gràfica anterior es pot observar el nombre d’aparicions dels diferents intervals enters ( [-x.99,x.99]) de les mitjanes de puntuacions de cada usuari a les dades analitzades. Es pot observar fàcilment com l’interval [-0.99, 0.99] és el més trobat a les dades. Es una proba similar a l’anterior però treballant amb les mitjanes. De la mateixa manera que abans tenim la Figura 4 realitzada amb truncament de dades i la Figura 5 amb arrodoniment.

Figura 5

### UserPositivism.png2.1.4. Positivisme dels usuaris

Figura 6

A la gràfica anterior es pot observar el nombre d’aparicions de les mitjanes de les puntuacions dels usuaris dividits en dos parts, els valors positius y els negatius. Com es pot observar a la gràfica la majoria dels usuaris tenen mitjanes positives. El valor False fa referència a un usuari negatiu i el True a un positiu, ens hem basat en que buscar el nombre de vegades que un usuari puntua per sobre de 0 o per damunt d’aquesta puntuació.

### 2.1.5. Positivisme dels restaurants

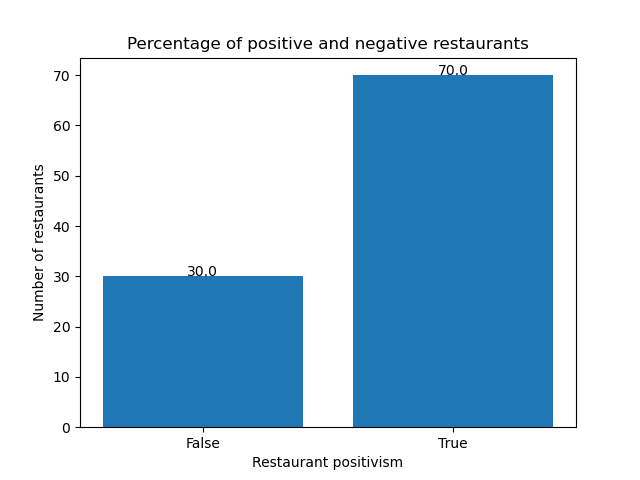


Figura 7

A la gràfica anterior es pot observar el nombre d’aparicions de les mitjanes de les puntuacions dels restaurants dividits en dos parts, els valors positius y els negatius. Com es pot observar a la gràfica la majoria dels usuaris tenen mitjanes positives.

El valor False fa referència a un restaurant negatiu i el True a un positiu, ens hem basat en que buscar el nombre de vegades que un usuari puntua per sobre de 0 o per damunt d’aquesta puntuació.

### 2.1.6. Nombre de visites dels restaurants per la seva mitjana

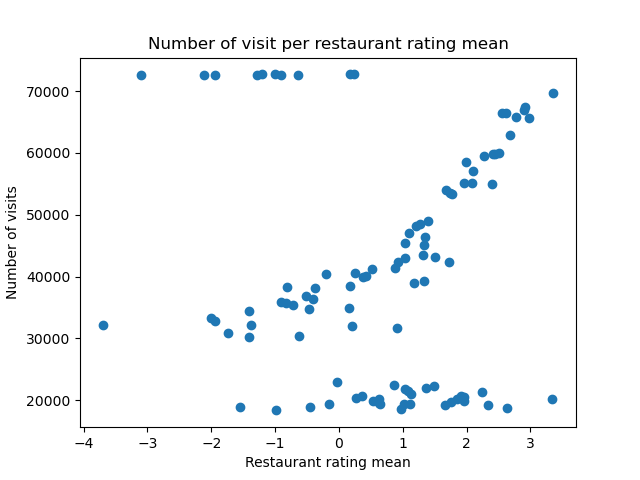


Figura 8

En aquesta gràfica volem analitzar quina afluència té cada restaurant respecte la seva puntuació mitja, és a dir si depèn aquesta afluència de la puntuació que ha obtingut com a mitjana. Podem veure en la Figura 8 que es diferencien 3 grans grups que es podrien analitzar de manera separada. El primer d’ells es tracta dels restaurants que tenen una puntuació mitjana entre -4 i 0 i tenen una gran quantitat de visites, en aquest grup hi pertanyen 10 restaurants. Un altre grup seria just el cas contrari, on els restaurants tenen una mitjana de puntuació més elevada, entre -1 i 4 però tenen una afluència de gent molt baixa, hi agrupem 30 restaurants. Per tant la resta de restaurants, 60, segueixen una tendència lineal, on podem concloure que com més afluència tenen, reben una puntuació mitja més positiva.

### 2.1.7. Probabilitat que una mateixa puntuació la faci diferents usuaris

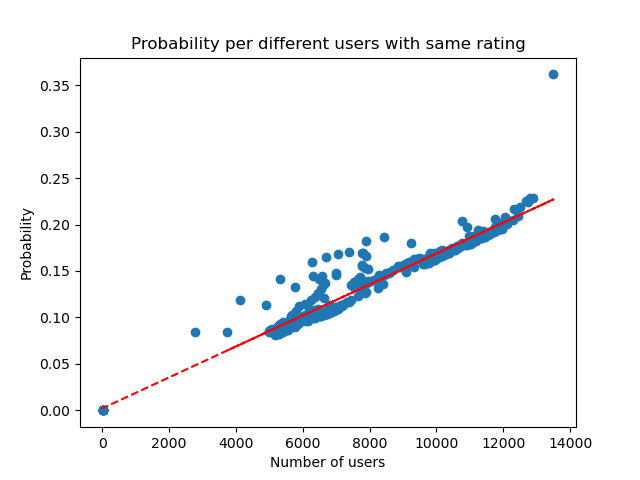


Figura 9

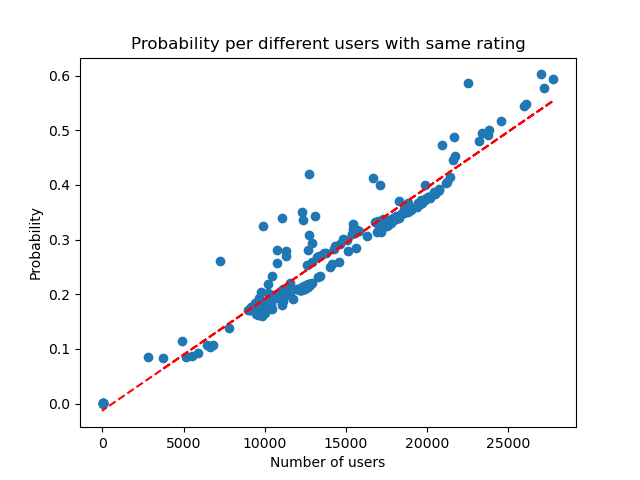


Figura 10

Per poder fer aquest anàlisi el primer que vam decidir fer va ser disminuir el nombre de puntuacions per poder agrupar els usuaris en grups més grans, d’aquesta manera les puntuacions s’han truncat/arrodonit en la Figura 9 i 10 respectivament a 1 decimal. Una vegada fet això, s’ha afegit la probabilitat que hi ha que es repeteixi una puntuació, calculada anteriorment i finalment hem contat el nombre d’usuaris diferents que han puntuat amb aquesta puntuació. Com s’ha esmentat en la gràfica mostrem la relació que segueix el nombre d’usuaris diferents que han puntuat amb certa nota i la probabilitat que té aquesta d’aparèixer. Els resultats obtinguts són els esperats i es que com més usuaris diferents puntuen amb certa nota, es més probable que aquesta es repeteixi, per tant ens assegurem que no hi ha cap usuari que faci servir moltes vegades una mateixa nota i això faci analitzar malament les gràfiques anteriors.

### 2.1.8. Comparació entre la puntuació mitja dels usuaris i la seva moda (I)

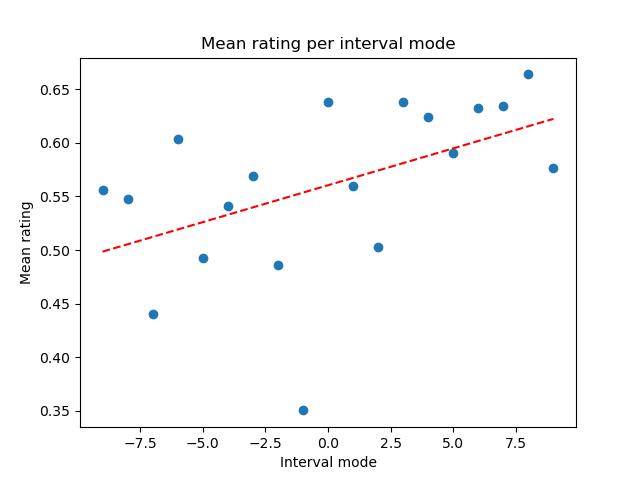


Figura 11

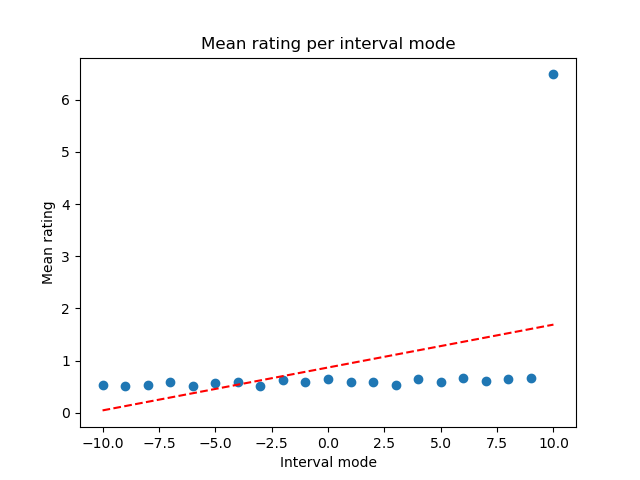


Figura 12

Per començar a analitzar la gràfica s’ha de comentar que per poder treballar amb la moda dels usuaris, aquesta s’ha truncat en la Figura 11 i s’ha arrodonit en la Figura 12 a un nombre enter. Per agrupar la moda amb la mitjana de cada usuari hem decidit fer la mitjana de la mitjana de cada usuari que comparteixin la mateixa moda, de tal manera que podem concloure que la moda es bastant semblant a la mitjana, és a dir que les puntuacions dels augmenten conforme també augmenta la moda.

### 2.1.9. Comparació entre la puntuació mitja dels usuaris i la seva moda (II)

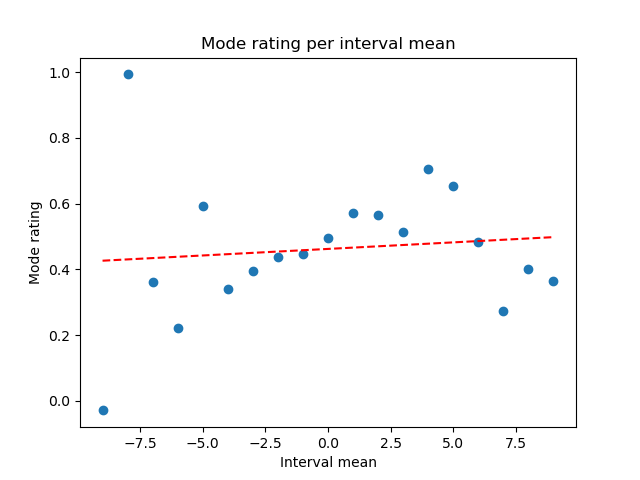


Figura 13

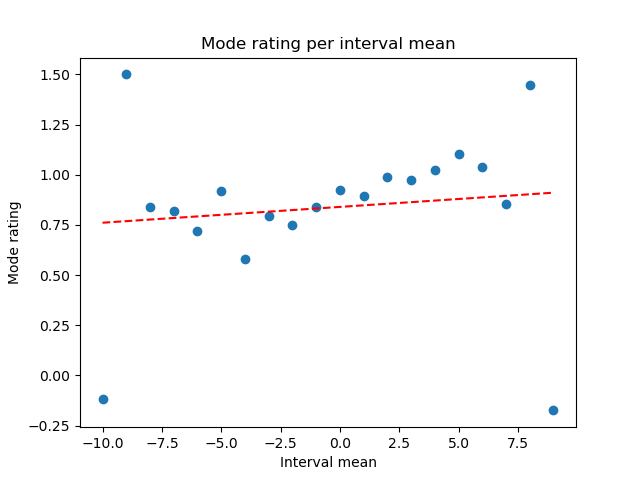


Figura 14

En aquest estudi fem un anàlisi similar a l’anterior, partim amb les mateixes dades, però les analitzem de manera una mica diferent. Cal dir que de la mateixa manera la primera figura ho fem amb truncament de dades i al segona amb arrodoniment d’aquestes. Per tal d’obtenir aquest gràfic el que farem es reduir el nombre de mitjanes de agrupant-les en grups d’enters, d’aquesta manera es redueix a 20 dades i posteriorment les podem comprar amb les modes. Per tant el que farem serà agrupar totes les mitjanes que tinguin el mateix valor i farem la mitjana de les modes corresponents. Per tant, extraiem que a mesura que la mitjana de puntuacions incrementa també ho la moda que pertanyia a aquestes puntuacions mitges.

### 2.1.10. Desviació estàndard respecte la mitjana dels usuaris

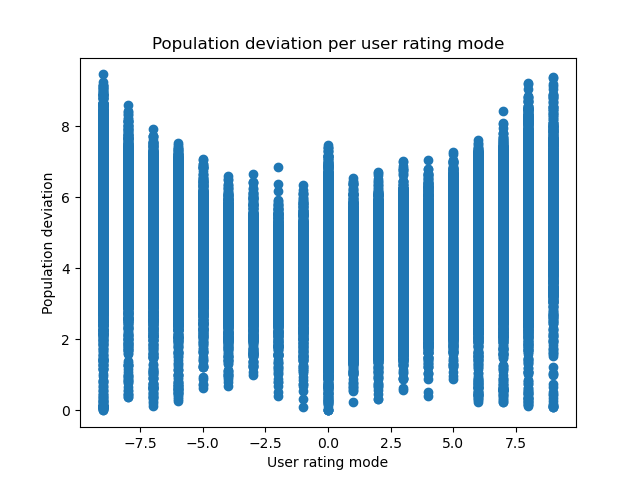


Figura 15

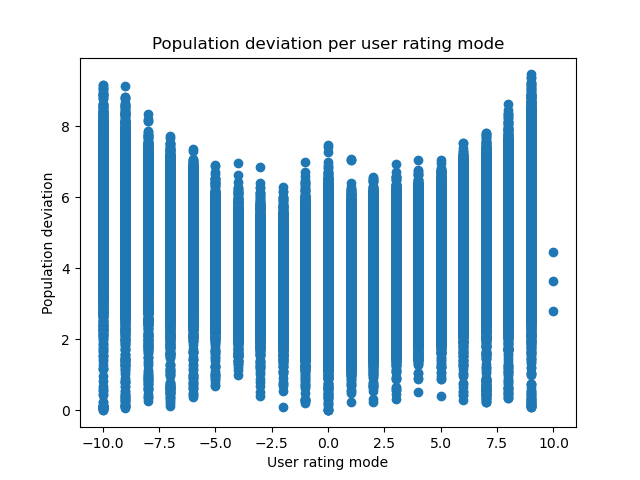


Figura 16