

Práctica 1 Tipología y Ciclo de Vida de los Datos

1. Contexto

El conjunto de datos generado como parte de esta asignatura consta de las puntuaciones y otra información sobre los álbumes de música que los usuarios de la página web, 'bestalbumsever', considera como los 500 mejores.

2. Definición del Dataset

Best_500_albums_information

3. Descripción del dataset

Como ya se ha comentado anteriormente, el dataset pretende recopilar la información relativa a los mejores 500 álbumes de la historia de la música, guardando la banda que lo compuso y lo grabó, el año de grabación y la puntuación que recibe el álbum por los usuarios de la página web en cuestión.

4. Representación gráfica

A continuación se presenta una imagen y dos gráficos que ayudaran a entender el dataset y el propósito de la recopilación de esta información.

Position Score			Album	Band	Release Year
0	1	92	OK Computer	Radiohead	1997
1	2	92	The Dark Side Of The Moon	Pink Floyd	1973
2	3	92	Abbey Road	The Beatles	1969
3	4	91	Revolver	The Beatles	1966
4	5	91	Sgt. Pepper's Lonely Hearts Club Band	The Beatles	1967
...
495	496	77	Out Of Time	R.E.M.	1991
496	497	79	Pretty Hate Machine	Nine Inch Nails	1989
497	498	78	Out Of The Blue	Electric Light Orchestra	1977
498	499	81	For Your Pleasure	Roxy Music	1973
499	500	83	Giant Steps	John Coltrane	1960

500 rows × 5 columns

Figure 1: Overview del dataset

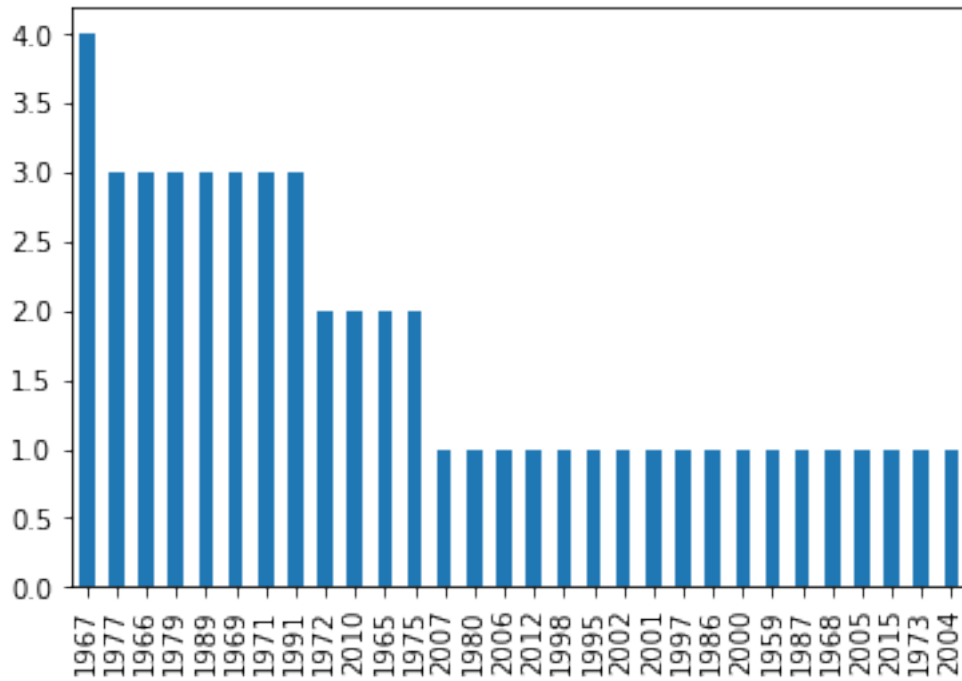


Figure 2: Número de álbumes en el top50 por años

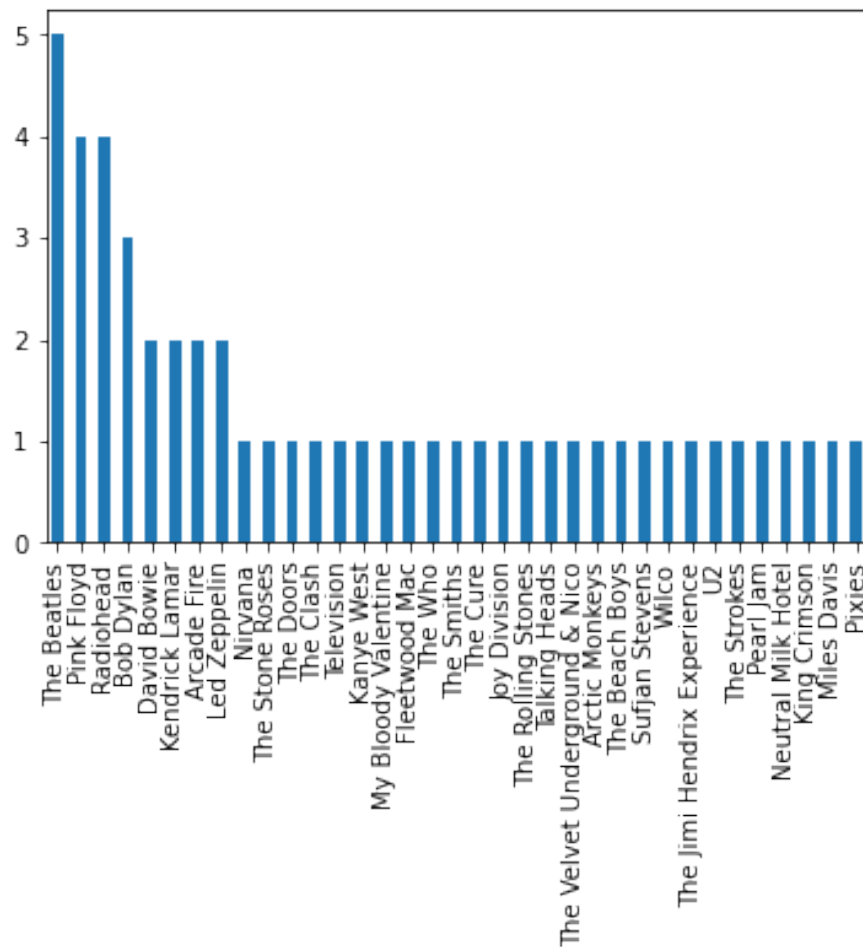


Figure 3: Número de álbumes en el top50 por banda

5. Contenido

El data set consta de los siguientes atributos:

- **Position:** Variable que guarda la posición del álbum en la lista.
- **Score:** Variable que guarda la puntuación sobre 100 del álbum.
- **Album:** Variable que guarda el nombre del álbum.
- **Band:** Variable que guarda el nombre de la banda que grabó el álbum.
- **Release Year:** Variable que guarda el año en el que el álbum se grabó.

Paralelamente, otro script nos guardara la imagen de la portada de cada álbum, siguiendo el orden de la posición en la lista, para el primer álbum de la lista, la imagen sera 'picture0.jpg'.

6. Agradecimientos

Agradecimientos a la página web por permitir el acceso a los datos y sobre todo los usuarios que con sus puntuaciones permiten que se puedan extraer estadísticas interesantes acerca de la historia de la música.

7. Inspiración

Me parece interesante el dataset por que permite sacar conclusiones acerca de los mejores grupos de la historia de la música y de los mejores años en la historia de la música, siendo la clasificación un computo de la opinión de miles de usuarios.

8. Licencia

Escojo la licencia CC BY-SA 4.0 License por que se debe proveer el nombre del creador del conjunto de datos indicando los cambios que se han realizado, de forma que las modificaciones o aportaciones que cualquiera ejecuten quedan reflejadas en el trabajo. Otro aspecto a tener en cuenta es que cualquier distribución que se haga de este mismo trabajo, tendrá que hacerse bajo la misma licencia, asegurándonos que las condiciones que queremos imponer con esta licencia se mantengan aun que terceros manipulen y distribuyan el trabajo.

9. Código.

El código se puede ver en la carpeta /source, hay dos archivos de extensión .py.

10. Dataset

Link en Zenodo:

- <https://zenodo.org/record/4266344#.X6qXDGHKhPZ>