Universidad Técnica Particular de Loja

Integrantes: Tais Balcázar, José Luis Benítez y Ronin Carrión

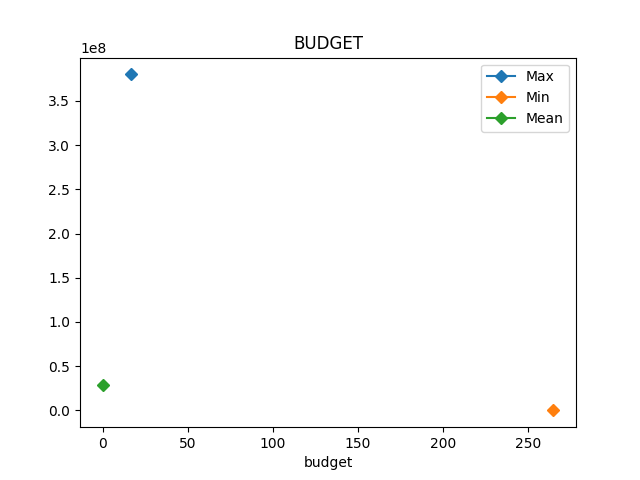
Fecha: 11/01/2023

Análisis Estadístico:

Para el análisis estadístico del dataset, hemos hecho uso de las librerías pandas y matplotlib, ideales para trabajar con datos y gráficos, respectivamente. Recuperamos los datos desde el CSV a través de pandas.read\_csv, enviándole como parámetro la dirección del archivo CSV, para que recolecte la información. Pandas reconoce directamente la data separada por comas, por lo que resulta intuitivo y muy fácil trabajar.

Luego de recolectar la data, hicimos uso de la librería matplotlib, para las gráficas, en donde le especificamos los valores para “X” y “Y”. Finalmente, para mostrar el gráfico usamos el método .show(). Cabe recalcar para que nos resulte más fácil el manejo de las librerías, a estas las hemos renombrado, para evitar redactar una sentencia larga de caracteres para hacer uso de los métodos de las librerías.

Budget:



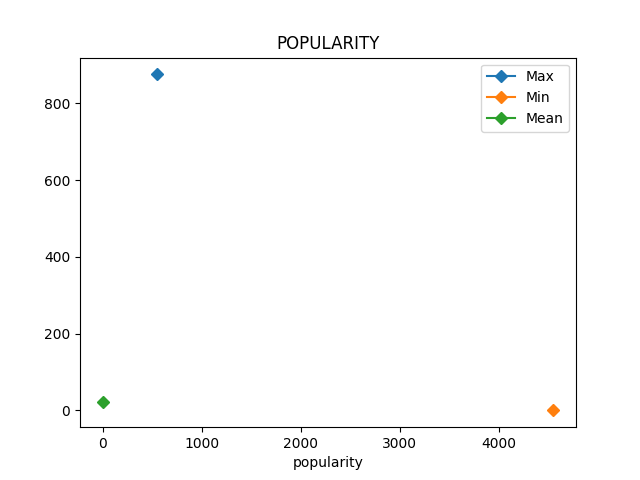
Tras el análisis descriptivo logramos identificar que entre los datos de “Budget” evidencia los siguientes resultados.

Maximo: $380000000 - Pirates of the Caribbean: On Stranger Tides.

Mínimo: $7000 – Primer.

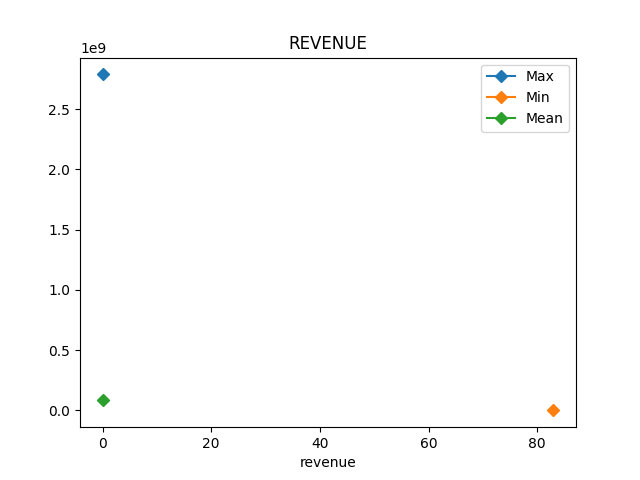
Media: $29045040.

Popularity:



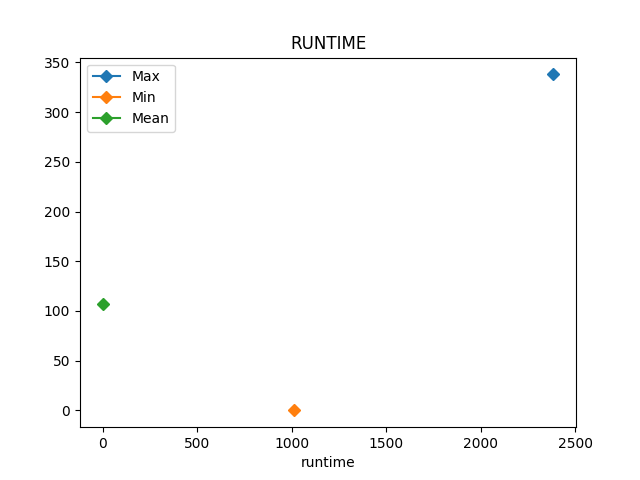
De acuerdo con la estadística descriptiva con los datos de "Popularity", podemos concluir que la película que tuvo más popularidad dentro de nuestra tabla fue "Minions", alcanzando una cifra de 875 millones, por otro lado, la película con menos popularidad fue la de "Penitenciary", la cual solamente llego a obtener 0.01117 millones, con esto también logramos analizar la media de nuestros datos, los cuales se situan alrededor de 21 millones.

Revenue:



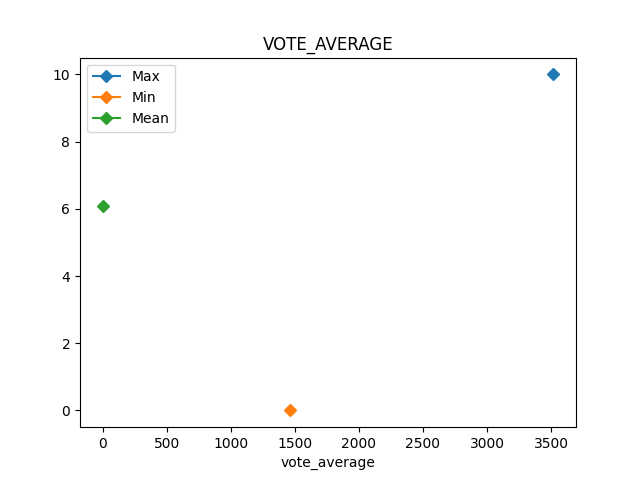
Analizando los datos de “Revenue” se ha podido determinar que los ingresos tienen como un máximo un 2,9. De la misma manera hemos encontrado que el promedio de ingresos de las películas en general se encuentra en 0.2. Finalmente, los ingresos más bajos que al menos, unas 80 películas han podido registrar es de 0,2.

Runtime:



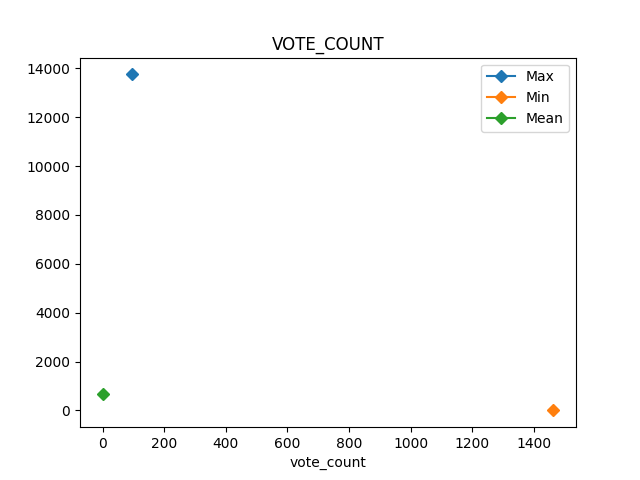
Tras el análisis estadístico de los datos de runtime, que quiere deci el tiempo de ejecución de la película, hemos encontrado que al menos unas 2400 películas han tenido un runtime de más de 300 horas. Mientras que, la media se encuentra por arriba de las 100 horas. Finalmente, encontramos que el mínimo tiempo de ejecución está en 0 horas, es decir, que no se han llevado a cabo su rodaje o su grabación.

Vote\_average:



Cuando hablamos sobre la aceptación que recibieron las películas por parte del público es decir el "Vote Averange", podemos sacar las siguientes estadísticas, la calificación más alta fue 10, y algunas de las películas que recibieron esta calificación fueron "Me and you five bucks" y "Little Big Top", mientras tanto la que tuvo la peor aceptación por parte del público fue "Dwengons" dándole una calificación de 0.5 ,y la media que obtuvieron las películas fue de 6.09.

Vote\_Count:



Tras el análisis descriptivo logramos identificar que entre los datos de “Vote\_Count” evidencia los siguientes resultados.

Maximo: 13752- Inception.

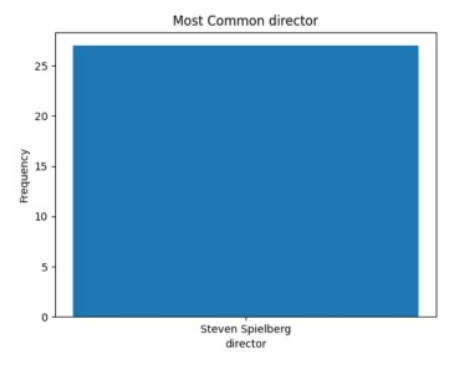
Mínimo: 530 – Four Rooms.

Media: 690.21.

**1.2. Obtención de conteo de atributos tipo cadena**

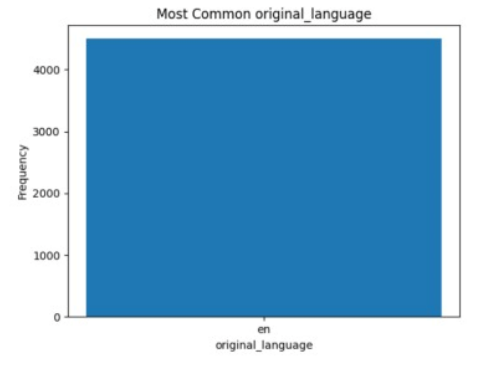
Trabajando con las columnas tipo cadena, genere el código de obtención de conteo y el (los) gráfico(s) correspondiente(s).

Director:



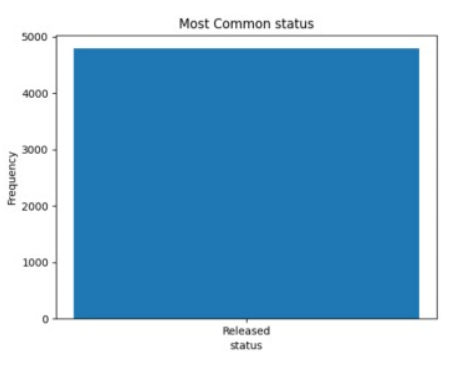
Tras el análisis descriptivo logramos identificar que entre los datos de “director” evidencia que la persona que más veces ha dirigido películas en esta base de datos es Steven Spielberg.

original\_language:



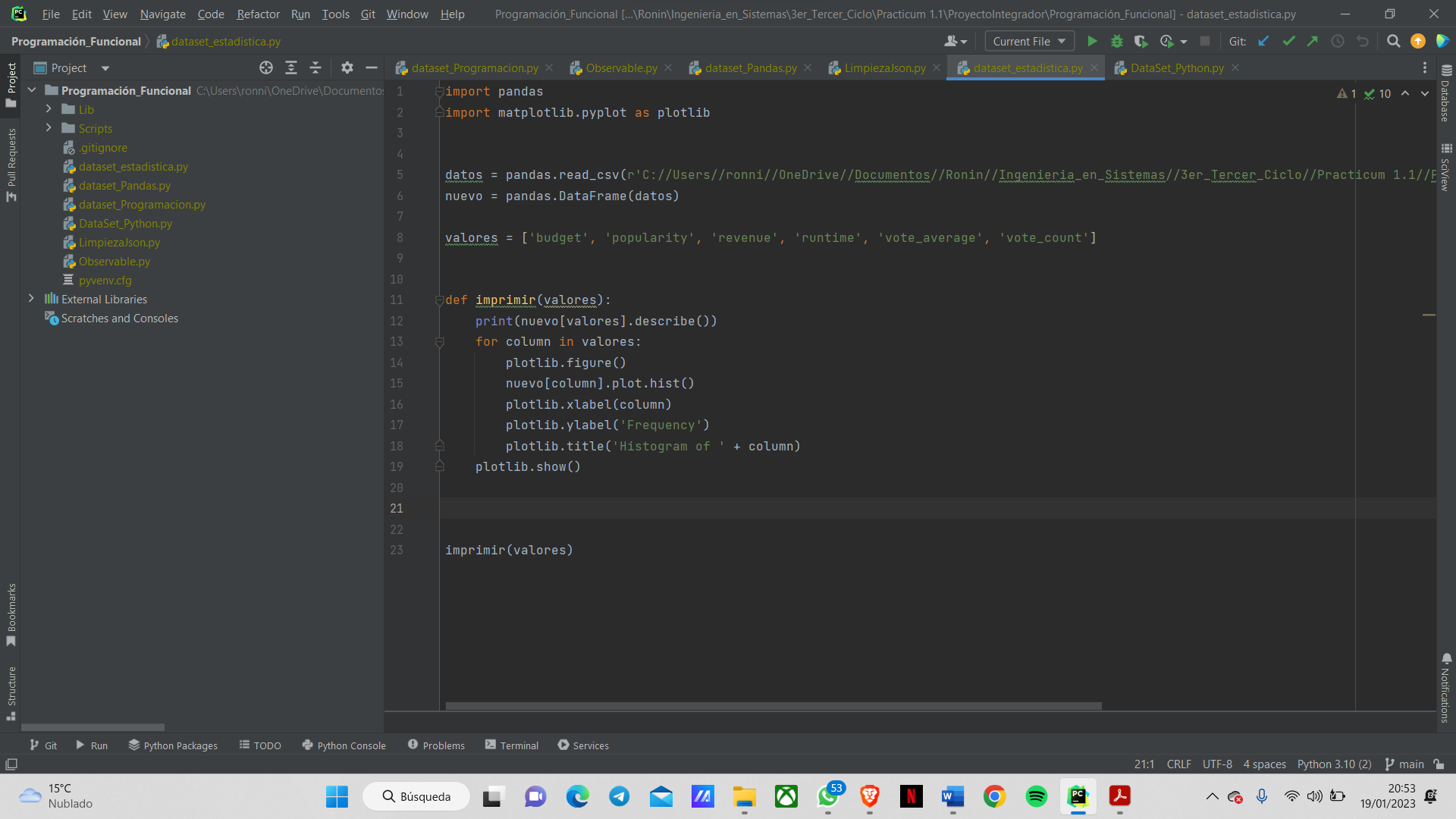
En el lenguaje original ("original\_language") de las películas logramos analizar que en más de 4000 películas el lenguaje que predomina, es el inglés (EN), obteniendo así más del 90% de todos los datos.

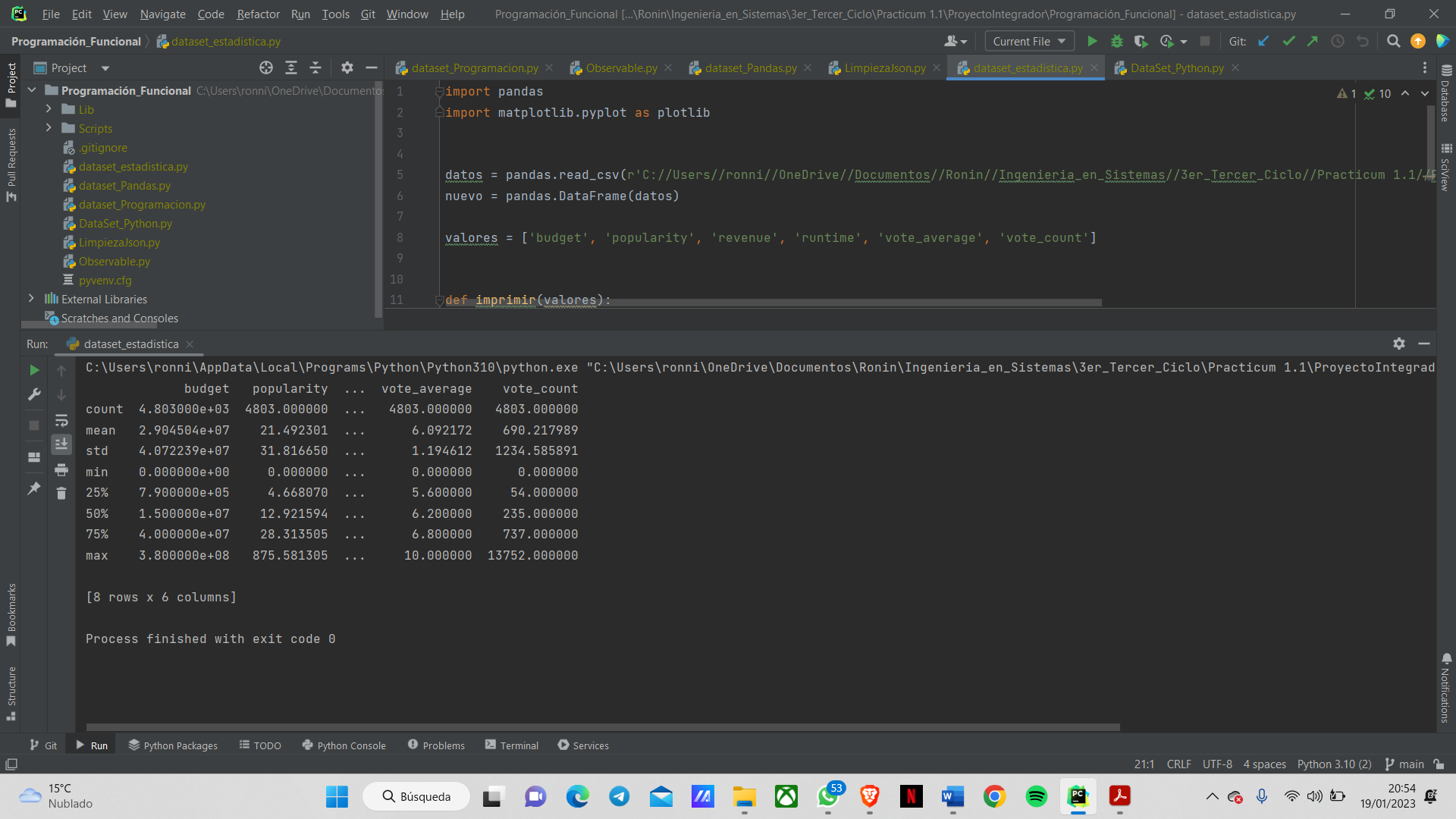
Status:



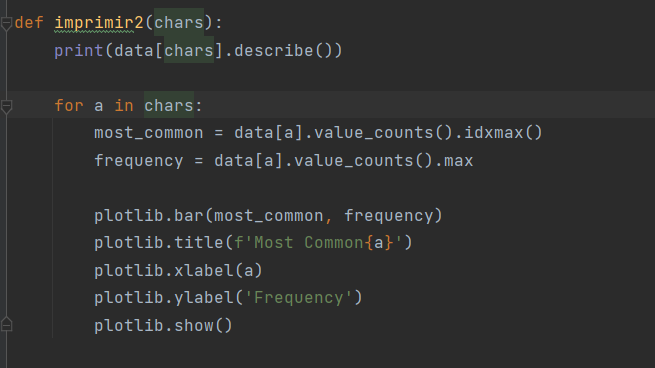
Según la gráfica podemos determinar que el estado más común es Publicado, dominando por arriba de las 4500 películas este estado, por lo que podemos afirmar que el estado más común y el dominante es “Publicado”.

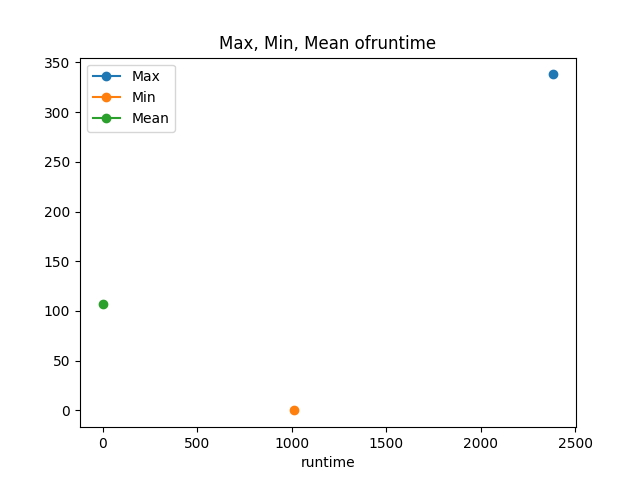
Código valores numéricos:





Código valores de cadena:





Gráfico

Descripción generada automáticamente