**MovieAPI**

Este proyecto pretende tomar dataset acerca de películas y sus plataformas, y mediante estos datos crear funciones con las cuales poder solicitarle a una API desplegada en Render la información necesaria, esto lo lograremos con las siguientes funciones:

* @app.get('/peliculas\_mes/{mes}'): Se ingresa el mes y la función retorna la cantidad de películas que se estrenaron ese mes históricamente.
* @app.get('/peliculas\_dis/{dis}'): Se ingresa el día y la función retorna la cantidad de películas que se estrenaron ese día históricamente.
* @app.get('/franquicia/{franquicia}'): Se ingresa la franquicia, retornando la cantidad de películas, ganancia total y promedio.
* @app.get('/peliculas\_pais/{pais}'): Ingresas el país, retornando la cantidad de películas producidas en el mismo.
* @app.get('/productoras/{productora}'): Ingresas la productora, retornando la ganancia total y la cantidad de películas que produjeron.
* @app.get('/retorno/{pelicula}'): Ingresas la película, retornando la inversión, la ganancia, el retorno y el año en el que se lanzó.
* @app.get('/recomendacion/{titulo}'): Ingresas un nombre de película y te recomienda las similares en una lista.

**ETL**

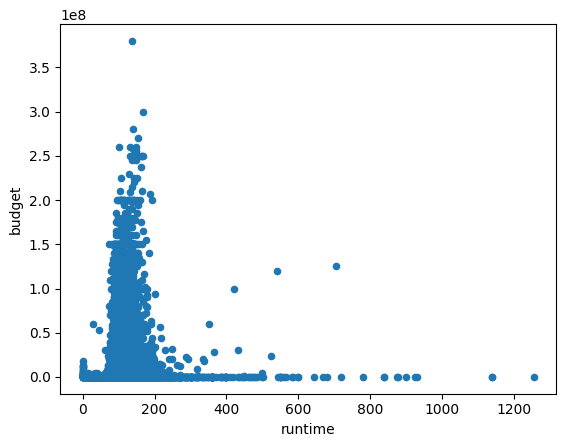
Antes de hacer las funciones con las que esta API va a trabajar, fue necesario hacer un trabajo previo de ETL donde será más fácil trabajar. Algunos de los procesos con los cuales se trabajó este ETL fueron:

* Algunos campos, como belongs\_to\_collection, production\_companies y otros (ver diccionario de datos) están anidados, esto es o bien tienen un diccionario o una lista como valores en cada fila, ¡deberán desanidarlos para poder y unirlos al dataset de nuevo hacer alguna de las consultas de la API! O bien buscar la manera de acceder a esos datos sin desanidarlos.
* Los valores nulos de los campos revenue, budget deben ser rellenados por el número 0.
* Los valores nulos del campo release date deben eliminarse.
* De haber fechas, deberán tener el formato AAAA-mm-dd, además deberán crear la columna release\_year donde extraerán el año de la fecha de estreno.
* Crear la columna con el retorno de inversión, llamada return con los campos revenue y budget, dividiendo estas dos últimas revenue / budget, cuando no hay datos disponibles para calcularlo, deberá tomar el valor 0.
* Eliminar las columnas que no serán utilizadas: video, imdb\_id, adult, original\_title, vote\_count, poster\_path y homepage.

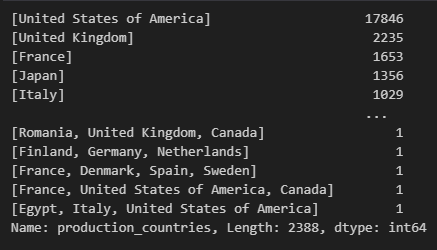
**EDA**

En el análisis exploratorio de datos podemos buscar patrones en los datos, correlaciones y tendencias.

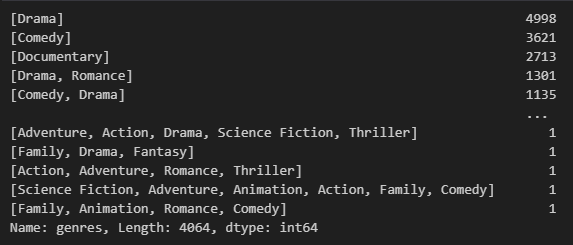
Por ejemplo, el siguiente grafico nos muestra la relación que hay entre el presupuesto y el tiempo de ejecución de películas:



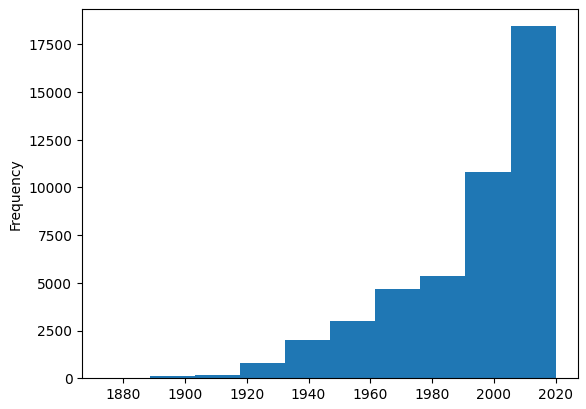
De este análisis también podemos encontrar los países que producen más películas.



Aquí están los géneros más producidos y las de menor producción.



También podemos ver la frecuencia por año de lanzamiento.



**Modelo de recomendación**

Una vez tenemos todo preparado, podemos trabajar un modelo de aprendizaje supervisado, en este caso fue el algoritmo support vector machine para ello se tomó como feature a las columnas Budget y reléase\_year así podremos visualizar y como target, utilizaremos la columna runtime para encontrar una recta que las relacione.