DATA VISUALIZATION

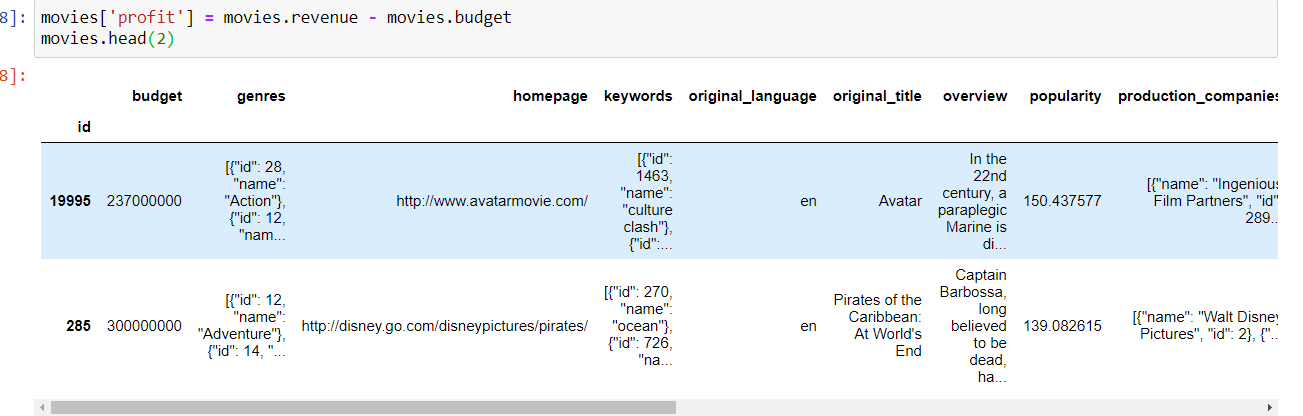
Project

He NI

Bader GUETARI

BI2

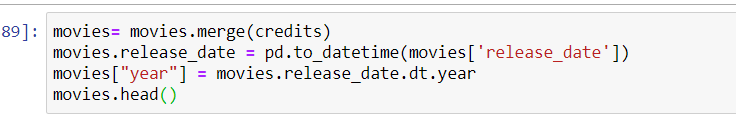
1. **The dataset link is:** [**https://www.kaggle.com/tmdb/tmdb-movie-metadata**](https://www.kaggle.com/tmdb/tmdb-movie-metadata)
2. **We chose this dataset to find out the trends in the movie market, what types of movies are more popular and rated higher, and our audience is the movie companies, with the aim of providing data analysis to the companies.**

**3. First we get the dataset and find two csv files, movies.csv and credits.csv**

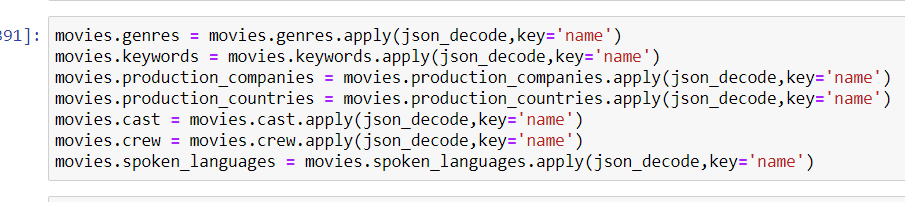


We need to merge and clean up these two data.

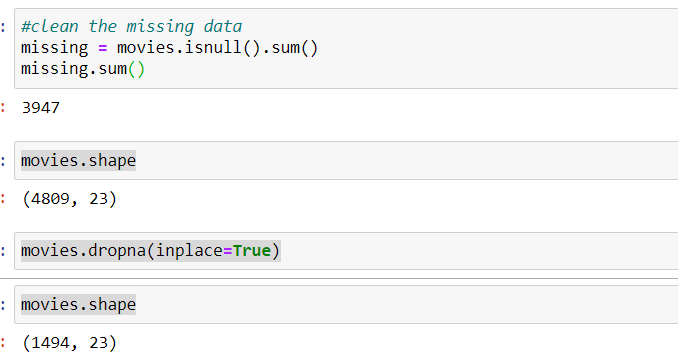
--Convert the year to pd.datetime and merge at the end of the table



--Perform data normalisation, i.e. data pre-processing, to split the features: crew, cast, production\_countries .... Perform splitting.



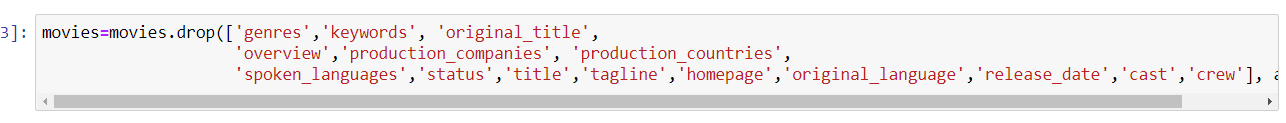
--then delete the empty value



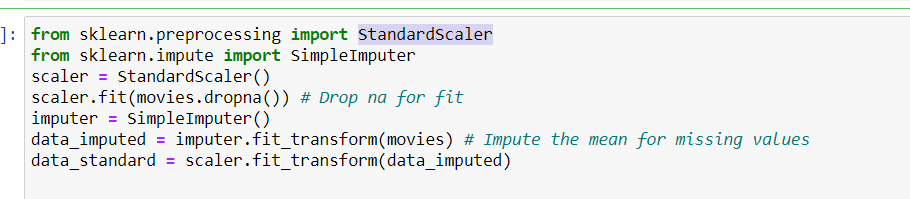
1. **Because there are many non-character types in the data, such as genres, we encode changes to them**



1. **drop those that cannot be encoded (too many columns)**

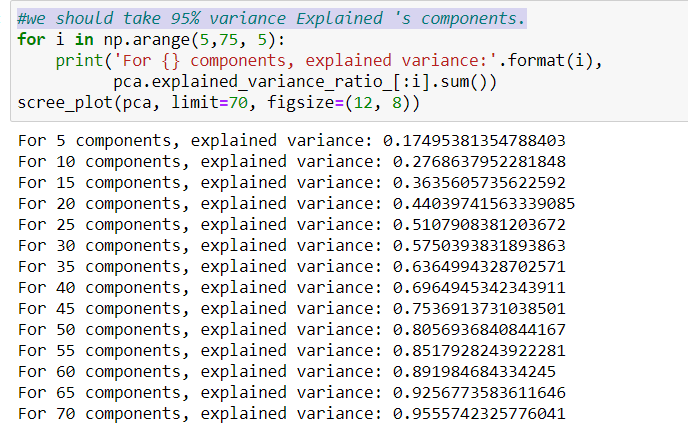


1. **we continue with the data pre-processing. Before we apply dimensionality reduction techniques to the data, we need to perform feature scaling so that the principal component vectors are not influenced by the natural differences in scale for features. Here we have chosen the Standard Scaler method.**

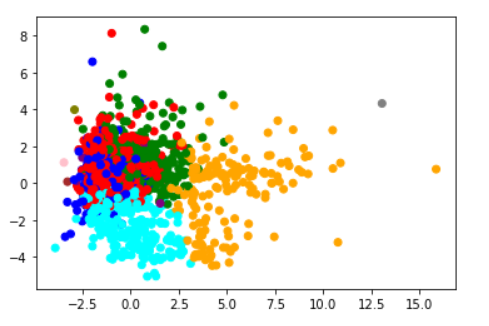


**Then using the PCA method for dimensionality reduction, because we have so many features, then we need to reduce the complexity of the data and identify the most important multiple features. So in comparison PCA is more appropriate and more efficient.**

1. **we should take 95% variance Explained 's components. So component should be 70.**



1. **Finally, the image is drawn using matplotlib.**



1. **Tableau dashboard**

[**https://public.tableau.com/profile/ni.he#!/vizhome/IMDB5000MovieDataset\_16089032498620/DashboardGenres?publish=yes**](https://public.tableau.com/profile/ni.he%23!/vizhome/IMDB5000MovieDataset_16089032498620/DashboardGenres?publish=yes)