

Eda Palazoğlu

Instructor: Onur Varol

CS 210 Introduction to Data Science

Final Project

Introduction

In this data science project, I analyze my movie-watching habits based on dates using the extracted data from my Letterboxd account. Letterboxd is a platform that you allows you to log, rate and review movies into your own dairy in your profile. By importing the data into Python data frame file, the focus of is on the Dairy data set to its comprehensive information.

The general insight is to recognize my view patterns, trends. The analysis includes data types, missing handling value and visualization. In addition, the methods used can be seen as data cleaning, transformation and creation visuals to enhance the understanding.

Methods

Data import:

Importing the movie data from my Letterboxd account in the form of csv files.

Data analysis by examining the data frame's structure, data types and available columns to recognize key feature.

Data cleaning:

Dairy data set is chosen in terms of its richness in information. Missing values in the rewathced column were fixed by transforming NaN values into yes or no format. For this, a code implemented to convert numerical values (1 and 0) as yes or no.

Date columns were converted to the Pandas datetime format.

Visualization:

Line charts were used for visualization the trends over time.

The first chart shows tyears of the movie I have watched.

The second charts illustrates the years I have watched the movie in and the year of that movie itself.

The third chart displays number of movies I have watched month by month.

Conclusion

In conclusion, this data science project provided a detailed frame of my movie watching habits.

Through data import, cleaning and visualization of my data patterns and trends emerged over time.