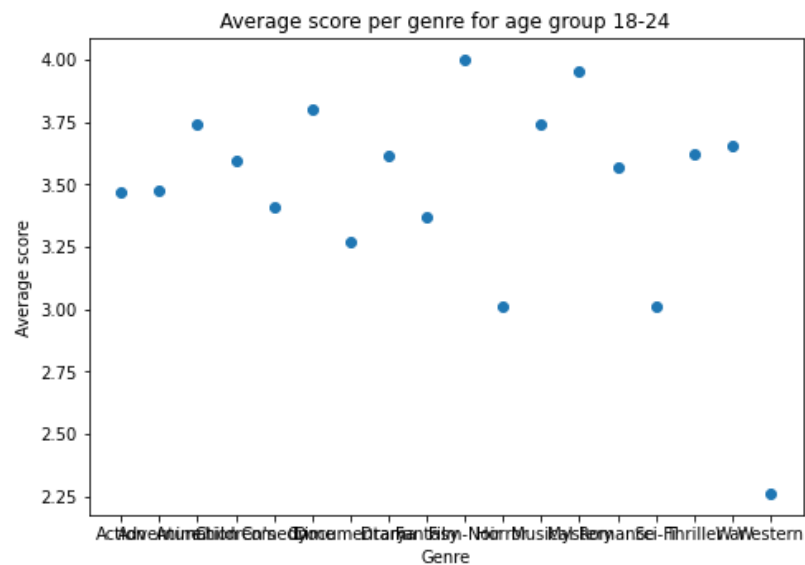


1. Jeg synes de viktigste parameterne er BrukerID (for å personalisere anbefalingssystemet til enhver bruker), Alder (for å bedre forutsi hvilke filmer forskjellige aldersgrupper liker best), Postkode (for å se hvilke filmer som er mer populære i enhver land), FilmID (for å lettere organisere filmer uten å bruke tittelen), Sjangere (brukes til å forutsi hvilke filmer enhver bruker liker best) og Rangering (viktig som input data til anbefalingssystemet, forteller oss om foretrekkede sjangere til enhver bruker).
2. Noen parametere som filene mangler og kan være viktige er: utgivelsesdato og spilletid til filmer. Utgivelsesdato kan, for eksempel, fortelle oss om hva forskjellige aldersgrupper synes om filmer fra forskjellige tiår. Spilletid kan fortelle oss om hvordan folk rangerer filmer av forskjellige lengder. Parameteren som jeg velger å bruke er: alder, sjangere og rangering. Disse tre er de viktigste parameterne og sammen kan de gi oss mange verdifulle resultater.
3. Jeg har visualisert en graf som viser snitt rangeringen til alle sjangere for hver aldersgruppe. For eksempel, ser du nedenfor denne grafen for aldersgruppen 18-24:



Koden er skrevet slik at samme grafen kan tegnes for alle aldersgrupper. Jeg har også jobbet på visualisering av en graf som viser den høyeste rangeringen til hver sjanger og en annen som viser den mest sette sjangeren til hver aldersgruppe, men ingen av de har en synlig output ennå.

4. Det er ikke lett å finne et mønster i grafene jeg har laget, så tenker jeg at training datasettet skal være ganske stort før den kan forutsi noen resultater. Dette betyr at det er også behov for et stort validation datasett. Sklearn i python kan brukes til å få dette til.