Koristile smo ‘Oscar nominations and filmographies since 1972’ dataset preuzet sa Kaggle sajta ([Oscar nominations and filmographies since 1972](https://www.kaggle.com/datasets/milanjanosov/oscar-nominations-and-filmographies-since-1972?resource=download&select=filmographies.csv)). Ovaj dataset se sastoji od dva csv fajla (filmographies.csv I nomination\_stats.csv).

Fajl nomination\_stats sadrži podatke o imenima nominovanih glumaca, kategorijama i godinama kada su nominovani, kao i ishodu nominacije (da li je dobitnik ili ne). U fajlu filmographies se nalaze podaci o nazivu, godini objavljivanja i tipu filma, kao i o imenu i poziciji (glumac, režiser, pisac…) osobe koja je učestvovala u kreiranju filma.

Nakon učitavanja fajlova u Python skriptu i analize varijabli koje se nalaze u njima, kreirale smo graf koji će se koristiti za prikazivanje glumaca i njihovih međusobnih saradnji. Iz dataseta filmographies izdvojile smo osobe čija je pozicija na filmu glumac/glumica. U čvorovima grafa se nalaze imena glumaca, broj filmova u kojima su glumili, broj nominacija i osvojenih nagrada.

Granama su povezani glumci koji su glumili u istom filmu, dok težine grana predstavljaju broj filmova u kojima su ti parovi zajedno glumili. Graf je neorijentisan, sadrži 546 čvorova i 19143 grana.

Nakon kreiranja grafa, izračunate su mere centralnosti kako bismo videle koji čvorovi, tačnije koji glumci su najvažniji u mreži. Pomoću **degree\_centrality**-a analizirale smo koliko veza ima svaki od čvorova tj. oni čvor koji ima puno veza sa drugima je popularan tj. taj glumac je glumio u mnogo filmova sa različitim partnerima. **Betweenness\_centrality** meri koliko puta neki čvor leži na najkraćem putu između drugih parova čvora. **Closeness\_centrality** smo koristile za računanje prosečne udaljensti jednog čvora od svih ostalih čvorova u mreži. **Eigenvector\_centrality** nam je poslužio da odredimo koliko su susedni čvorovi važni u odnosima između svih glumaca. Pored mera centralnosti, za svakog glumca smo izračunale broj filmova i broj nominacija.

Na osnovu izračunatih metrika za svakog glumca, kreirana je **korelaciona matrica** koja ima za cilj da proveri da li su različite centralnosti međusobno povezane i da li položaj glumca u mreži ima povezanosti sa njegovim karijernim uspehom tj. osvajanjem nagrade. Iz marice se može videti da postoji jaka korelacija između mera centralnosti gde je najveća korelacija između degree i eigenvector (0.99), što znači da glumci sa više veza u mreža imaju i veći uticaj prema eigenvector centralnosti. Korelacija sa num\_films je umerena (od 0.58 do 0.62), što znači da glumci koji su centralniji u mreži imaju bogatiju filmsku karijeru. Centralnost slabo utiče na broj nominacija (0.32-0.37), dok broj filmova još manje utiče na njih (0.058). Dakle, to koliko je glumac nagrađivan ne zavisi značajno ni od broja filmova ni od njegove centralnosti u mreži.

Kreirana je **matrica dostupnosti** u kojoj svaki red i svaka vrsta predstavlja ime glumca, dok svaka ćelija prikazuje broj filmova u kojima su ta dva glumca zajedno glumili. Na osnovu nje kreiran je dataframe **avg\_reach\_df** koji prikazuje prosečnu „dohvatljivost“ svakog glumca u mreži. Kreiran je još jedan dataframe pod **nazivom shortest\_paths\_df** koji prikazuje dužinu najkraćeg puta između svaka dva glumca.

Kako bismo bolje razumele strukturu mreže, koristile smo **diametar** radi merenja koliko je mreža kompaktna, tj. koliko su glumci daleko u mreži. Naš diametar iznosi 4, što znači da je potrebno najmanje 4 koraka da bi se stiglo od jednog do bilo kog glumca.

Kreirana su tri podgrafa sa po 50 glumaca sa najvećim brojem filmova, nominacija i osvojenih nagrada, kao i graf na kome su prikazani glumci koji su sarađivali međusobno najmanje 6 puta.

Klasterovanje 50 glumaca sa najvećim brojem filmova je odrađeno u tri kategorije: 140 ili manje filmova, 200 ili manje i više od 200.

Graf u kome se nalaze glumci sa najviše nominacija je podeljen takođe u tri klastera: glumci koji imaju 5 ili manje nominacija, oni koji imaju 10 ili manje, i ostali.

Za kraj, kreiran je graf koji klasteruje glumce po broju osvojenih nagrada. U prvu kategoriju spadaju glumci koji su osvojili jednu nagradu, u drugu oni koji su osvojili dve i u treću ostali.