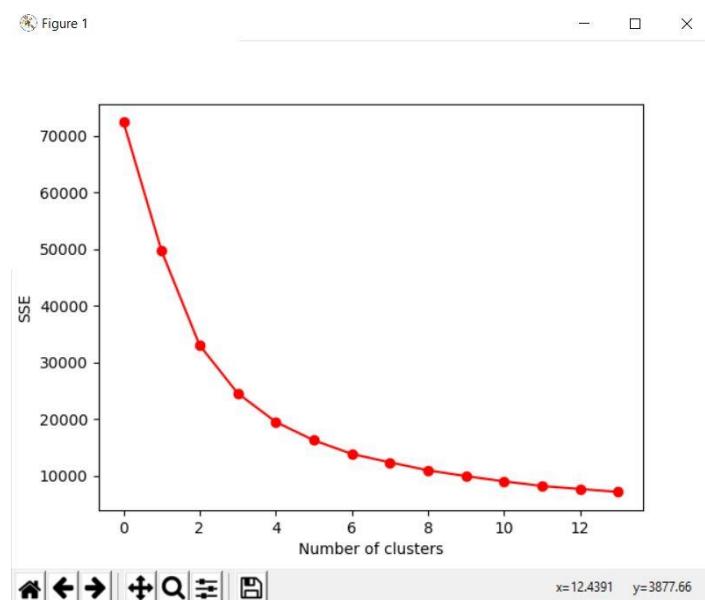


Analiza ponašanja korisnika kreditnih kartica

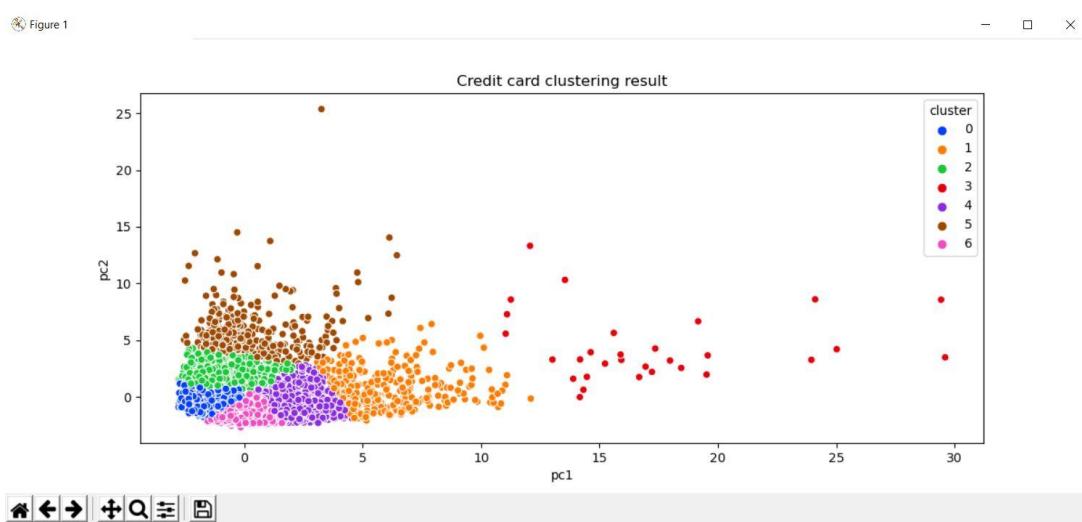
Klasterovanje

Na osnovu skupa podataka koji sumira ponašanje korisnika kreditnih kartica, grupisale smo korisnike koji imaju slično ponašanje korišćenjem *K-means* algoritma za klasterovanje. Optimalan broj klastera smo odredile pomoću *elbow* metode, čiji rezultat prikazuje Slika 1. Sa slike se vidi da se za broj klastera može uzeti broj u opsegu od 4 do 8. Zbog velikog broja podataka odabrale smo da broj klastera bude 7.



Slika 1 Grafički prikaz elbow metode

Prikaz vizualizacije klastera u 2D prostoru pokazuje Slika 2.



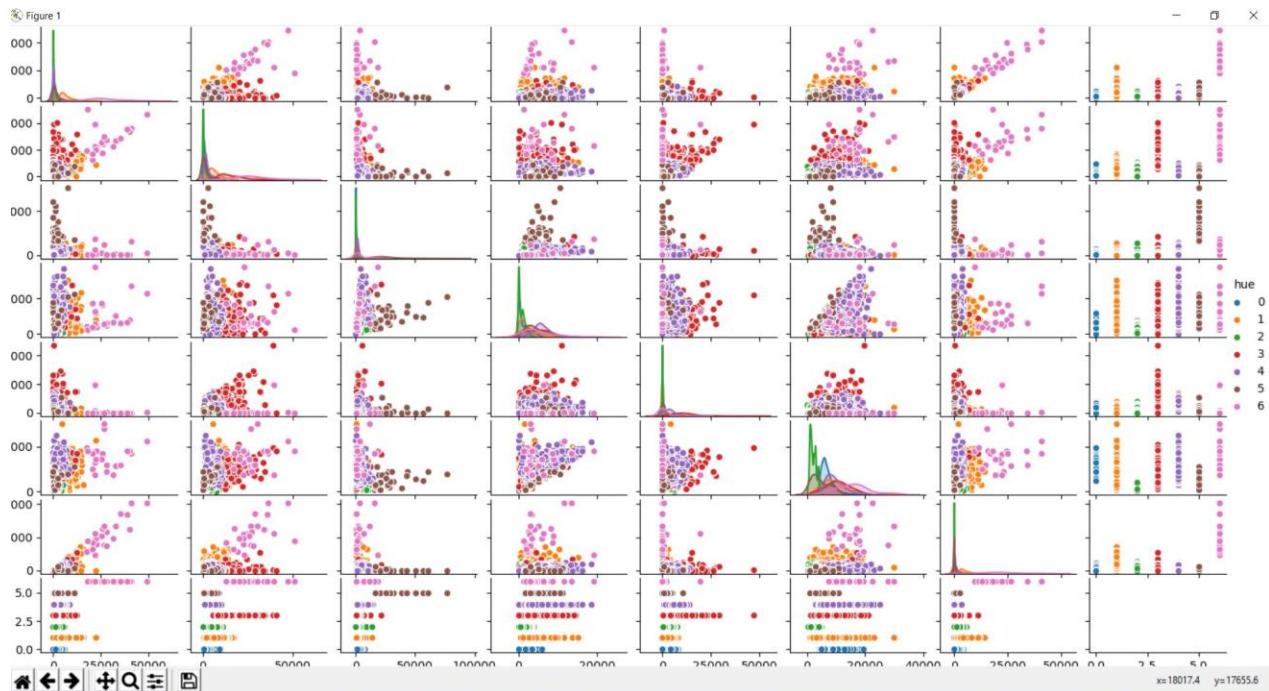
Slika 2 Vizualizacija klastera u 2D prostoru

Sumarizacija vrednosti atributa kojima su opisani klasteri

Svaki korisnik kreditne kartice iz skupa podataka je opisan sa 18 atributa. Međutim, radi sumarizacije izdvojile smo samo atribute koji jasnije odvajaju podatke. To su:

- PURCHASES - ukupan iznos potrošen na kupovinu
- PAYMENTS - ukupan iznos uplaćen na karticu
- MINIMUM_PAYMENTS - minimalan iznos koji je korisnik uplatio na karticu
- BALANCE - stanje na računu dostupno za kupovinu
- CASH_ADVANCE - iznos koji je korisnik uplatio unapred
- CREDIT_LIMIT - limit na kreditnoj kartici
- ONEOFF_PURCHASES - iznos potrošen na kupovinu jednokratno

Na osnovu *pairplot-a* (Slika 3) koji pravi poseban grafik za svaka dva atributa, može se videti prisutnost vrednosti atributa u klasterima. Svaki red i svaka kolona predstavljaju atrbute respektivno u odnosu na gorenavedenu listu.



Slika 3 Pairplot na osnovu kog se može raditi sumarizacija

U nastavku je opis svakog od klastera:

1. Klaster 1 (Plavi) – ova grupa ima najmanje minimalnih uplata (*minimum payments*), srednji limit na kartici (*credit limit*) i ukupan iznos potrošen na kupovinu je veoma mali (*purchases*)
2. Klaster 2 (Narandžasti) – ovu grupu čine korisnici sa opsegom od srednjeg do visokog stanja na računu (*balance*) i limita (*credit limit*), i imaju drugi najviši iznos potrošen na kupovinu (*purchases*)

3. Klaster 3 (Zeleni) – ovu grupu odlikuju korisnici sa najmanjim stanjem na računu (*balance*) i najmanjim limitom na kartici (*credit limit*), kao i najmanjom potrošnjom (*purchases*)
4. Klaster 4 (Crveni) – ovu grupu ima najviši iznos uplaćen unapred (*cash advance*) i visok ukupan iznos uplaćen na kartici (*payments*)
5. Klaster 5 (Ljubičasti) – ovu grupu čine korisnici koji imaju visoko stanje na računu (*balance*), visok limit na kartici (*credit limit*) i slabo zastupljenu jednokratnu kupovinu (*oneoff purchases*)
6. Klaster 6 (Braon) – grupa sa najviše minimalnih uplata (*minimum payments*), najmanjim ukupnim iznosom uplaćenim na kartici (*payments*) i najmanjim uplatama unapred (*cash advance*)
7. Klaster 7 (Roze) – ova grupa ima najveći protok novca na kartici, najviše troši na kupovinu (*purchases*) i najviše uplaćuje na karticu (*payments*)

Osim interpretiranja rezultata gledajući grafik, moguće je opisati klastere koristeći pravila koja su generisana pomoću stabla odlučivanja. Stablo radi tako što za svaki klaster, na osnovu podataka koji mu pripadaju, određuje opsege atributa koji odgovaraju tom klasteru. Radi poređenja sa vizualizacijom klastera, izdvojile smo neke od rezultata opisivanja klastera upotrebotom stabla odlučivanja. U zagradi je naveden broj podataka koji pripadaju klasteru.

1. Plavi klaster (1210) – CREDIT_LIMIT >1750.0 and CREDIT_LIMIT <=5475.0
and PURCHASES <= 1410.774
2. Narandžasti klaster (1447) – BALANCE <=5575.63 and BALANCE >670.002
and CREDIT_LIMIT >1750.0
and PURCHASES > 1410.774
3. Zeleni klaster (2740) – BALANCE <= 1965.446 and PURCHASES <= 1410.774
and CREDIT_LIMIT <= 1750.0
4. Crveni klaster (288) – PAYMENTS <= 11673.01 and CASH_ADVANCE > 5310.122
5. Ljubičasti klaster (402) – BALANCE > 3436.744 and CREDIT_LIMIT > 1750.0
6. Braon klaster (2833) – CASH_ADVANCE <= 2067.34 and PAYMENTS <= 909.53
7. Roze klaster (30) – PURCHASES > 1410.774 and PAYMENTS > 11673.01

Ako uporedimo rezultate, vidimo da se vrednosti atributa u većini poklapaju sa vizuelnom predstavom klastera. Možemo koristiti oba načina u svrhu interpretacije rezultata klasterovanja, međutim drugim načinom dobijamo preciznije opsege vrednosti.