PROJEKTNI ZADATAK ZA NoSQL Baze i big data analitika 1. dio

**Svaki student temeljem dobivenog seta podataka za uspješno dovršenje projekta za ispit mora u MongoDB okružju koristeći Java script odraditi naredne zadatke :**

1. Sve nedostajuće vrijednosti kontinuirane varijable zamijeniti sa -1, a kategoričke sa „empty“.
2. Za svaku kontinuiranu vrijednost izračunati srednju vrijednost, standardnu devijaciju i kreirati novi dokument oblika sa vrijednostima, dokument nazvati: statistika\_ {ime vašeg data seta}. U izračun se uzimaju samo *nomissing*  vrijednosti .

|  |  |  |  |
| --- | --- | --- | --- |
| Varijabla | Srednja vrijednost | Standardna devijacija | Broj nomissing elemenata |
| A |  |  |  |
| B |  |  |  |
| C |  |  |  |
| D |  |  |  |
| …. |  |  |  |

1. Za svaku kategoričku vrijednost izračunati frekvencije pojavnosti po obilježjima varijabli i kreirati novi dokument koristeći nizove, dokument nazvati: frekvencija\_ {ime vašeg data seta} . Frekvencije računati koristeći *$inc* modifikator.

|  |  |
| --- | --- |
| Varijabla | Pojavnost |
| A | {(a1:20), (a2:30)} |
| B |  |
| C |  |
| D |  |
| …. |  |
|  |  |

1. Iz osnovnog dokumenta kreirati dva nova dokumenta sa kontinuiranim vrijednostima u kojoj će u prvom dokumentu biti sadržani svi elementi <= srednje vrijednosti , a u drugom dokumentu biti sadržani svi elementi >srednje vrijednosti , dokument nazvati: statistika1\_ {ime vašeg data seta} i statistika2\_ {ime vašeg data seta} i
2. Osnovni dokument kopirati u novi te embedati vrijednosti iz tablice 3 za svaku kategoričku vrijednost, : emb\_ {ime vašeg data seta}
3. Osnovni dokument kopirati u novi te embedati vrijednosti iz tablice 2 za svaku kontinuiranu vrijednost kao niz : emb2\_ {ime vašeg data seta}
4. Iz tablice emb2 izvući sve one srednje vrijednosti iz nizova čija je standardna devijacija 10% > srednje vrijednosti koristeći $set modifikator
5. Kreirati složeni indeks na originalnoj tablici i osmisliti upit koji je kompatibilan sa indeksom

Na ispitu je potrebno pokazati rezultate analize. Kod za svaku od navedenih točaka. Kod mora biti pisan kao **.js** file /fileove poslane najmanje 2 dana prije ispita sa jasnom naznakom u kodu na koju točku se odnosi.

Na ispitu student će aktivirati kod i pokazati rezultate , te će morati objasniti pojedine tehničke detalje realizacije projekta.

Na ispitu student će također dobiti dva pitanja vezana uz Big data koncept i Lambda arhitekturu.