Prvi domaći zadatak iz Uvoda u nauku o podacima

Prvi domaći zadataka predstavlja projekat iz oblasti Eksplorativne analize podataka (Exploratory Data Analysis - EDA). Vaš skup podataka (u prilogu) predstavlja informacije o polovnim automobilima postavljenim na online oglasima. Vaš zadatak je da analizirate podatke na odgovarajući način tako da dobijete **što više** informacija.

Skup podatak sadrži sledeće varijable:

**Name** – ime proizvođača i modela vozila (poželjno je ovu kolonu podeliti u dve kolone)

**Year** – godina proizvodnje

**Selling\_price –** prodajna cena

**Km\_driven** – kilometraža

**Fuel** – koji tip goriva se auto koristi

**Seller\_type** – da li je prodavac fizičko lice ili auto plac

**Transmission** – tip menjača

**Owner** – koliko je vlasnika imao automobil

**Engine** – zapremina motora

**Max\_power** – kočena konjska snaga (snaga motora na zamajcu)

**Torque** – kolona koja pokazuje obrtni moment (Nm), pri određenom broju obrtaja (RPM). Poželjno je ovu kolonu podeliti u dve kolone.

**Seats** – Broj sedišta

Neke od korisnih informacija bi bile: broj postavljenih oglasa po proizvođaču (kao što možete videti postoji kolona koja ujedno sadrži informacije o proizvođaču i modelu proizvoda), kako broj pređenih kilometara utiče na cenu, da li ima više automobila sa pogonom na benzin ili dizel, koji automobili su najmanji potrošači, odnos između obrtnog momenta i obrtaja u minutu, koji su automobili najskuplji itd. **Ovo su samo neke od informacija, na Vama je da vidite koje korisne informacije možete izvući, te je radoznalost i kreativnost poželjna.**

Naravno, određene informacije je korisno predstaviti u grafičkom obliku.

Za varijable je potrebno uraditi deskriptivnu statistiku.

Neke od kolona koje sadrže brojeve nisu numeričke, već sadrže i jedinice mere. Takođe, da li u skupu postoje duplikati? Da li ima nedostajućih vrednosti?