

Övning: Psykisk ohälsa bland studerande - analys av en enkätundersökning

I den här övningen ska vi arbeta med ett större dataset än i tidigare övningar!

Datasetet kommer från en enkätundersökning kring om studerande på universitetsnivå i Indien anser sig vara deprimerade eller inte och hur detta korrelerar till deras studiebörd, arbetsbörd, antal timmar de arbetar/studerar per dag, hur mycket de sover, om de upplever att de äter hälsosamt eller inte etc.

Datasetet är hämtat härifrån, innehåller ca 27 900 enkätsvar och är i CSV-format (kommaseparerat).

Ladda hem datasetet

Även om vår mall har funktionalitet för att läsa in CSV-formaterade dataset, så ska vi i det här fallet importera datan till SQLite, då detta möjliggör att vi kan använda SQL-frågor för att arbeta med datan.

Du kan använda SQLiteStudio:s importfunktion för detta:

1. Skapa en ny databas.
2. Skapa en tabell med kolumner motsvarande de i datasetet. Tänk igenom val av datatyp och kolumnnamn. Det kan vara bra om du inte har mellanslag eller konstiga tecken i kolumnnamn och följer en namngivningskonvention - t.ex. camelcasing!
3. Importera data från CSV-filen genom att välja Tools -> Import och följa instruktionerna i den dialogruta/wizard som dyker upp!
4. Fundera även på om någon data behöver transformeras, t.ex. vore kanske Sleep Duration en bra kandidat. Denna kolumn innehåller just nu textbaserad data, men borde gå att omvandla till numeriska gränsvärden som går att räkna på. Finns det andra kolumner som behöver transformeras för att det ska bli lättare att räkna med dem statistiskt?
5. Utgå från vår mall i version 3 och se till att kopiera in databasen i den. Obs! I och med att det här blivit en inlämningsuppgift är det version 5 eller senare av mallen som gäller!

6. Se även till att skapa ett git-repo på GitHub med filerna från mallen + SQLite-databasen! (Använd gärna GitHub Desktop för att skapa repositoret och, viktigt, se till att du git-ignorerar node_modules-mappen!)
7. Börja undersöka datasetet - kan du hitta samband mellan olika kolumner/mätvärden? Detta kan du göra direkt med SQL-kommandon (och om du vill i SQLiteStudio) eller genom att använda vår mall för att ställa SQL-frågor, med fördelen att du får tillgång till att använda Simple Statistics på resultat (efter mappning till numeriska listor/arrayer), samt Google Charts.
8. Finns det en story att berätta här? Är det tydligt att vissa faktorer verka påverka de studerandes psykiska hälsa/ohälsa? Vilka är i så fall de som påverkar mest...
9. Kan du berätta den här historien på ett intressant sätt, genom att blanda en berättande text, tabeller med delresultat och diagram som pekar på viktiga tendenser?

Detta är en relativt stor övning så du behöver inte göra klart den under veckan, men arbeta gärna med den under fredag och måndag till dess att vi ses igen nästa tisdag! Då kan vi prata om hur långt du kommit, vad du har upptäckt och vad det finns för story att berätta kring datan.