

Individuellt Projekt

Databehandling 25p

Syfte & Målbild

Detta projekt ämnar att ge dig som student möjligheten att sätta dina färdigheter i databehandling på sin yttersta spets genom att djupgräva in i, och dra insikter ur, ett verkligt realistiskt dataset mha Pandas & andra relevanta bibliotek i Python.

Du ska även få möjligheten att uppvisa förmåga att ta informerade & genomtänkta beslut när det kommer till datahantering och transformering. Utöver det ställer projektet även krav på dig att självständigt uppsöka information du kan behöva - främst genom relevanta dokumentationer.

Betygskriterier

Projektet betygsätts i skalan IG – G – VG.

För godkänt (G) krävs demonstrering av att du:

- behärskar de vanligaste metoderna i Pandas och mha dessa kunna extrahera relevanta grundläggande insikter ur ett dataset.
- klarar av att framställa enklare visualiseringar för att på ett i huvudsak tillfredställande vis uppvisa trender i data.
- Visar prov på självständig förmåga att hämta in grundläggande information från relevanta dokumentationer.

För väl godänt (VG) krävs demonstrering av att du:

- behärskar avancerade metoder i Pandas samt på ett systematiskt och korrekt vis kan extrahera avancerade insikter ur ett eller flera dataset samtidigt.
- klarar av att framställa avancerade visualiseringar som på ett mycket tillfredställande vis presenterar trender i data objektivt och otveksamt.
- Visar mycket god förmåga att självständigt hämta in avancerad information från relevanta dokumentationer.

Instruktioner

Ladda ner projektets notebook *chicago.ipynb* och följ sedan vidare instruktioner där.

Varje deluppgift ska besvaras i cellerna omedelbart nedanför. Dina svar ska bestå av kod (främst Pandas indexering/querying) där du

1. i kod framställer det begärda resultatet
2. en markdown cell där du beskriver resultatet du tagit fram.

Undvik så gott du kan att manuellt kopiera och klistra in resultat från större tabeller! Gör istället ditt bästa för att indexera fram endast det begärda svaret istället.

Du får även, där du känner att det är relevant och behjälpligt framställa visualiseringar med antingen Matplotlib, Seaborn eller Plotly / Plotly Express.

När du känner dig färdig ska du döpa om notebooken till följande format:

namn_efternamn_databehandling.ipynb

Därefter lämnar du in den på studentportalen.

Caveat

- Du behöver inte nödvändigtvis svara på alla deluppgifter för att få godkänt. Istället kommer det göras en helhetsbedömning. Det viktiga är att du uppvisar förmåga och gör ett ärligt försök.
- Deluppgifter markerade med * ger dig oftast tolkningfrihet. På dessa frågor förväntas oftast även ett svar också kräver djupare analys.
- Alla variabler du använder bör vara konstruerade och namngivna på ett sätt så att jag enkelt bara kan klicka mig igenom er notebook. Jag vill inte ha massa konflikter i er namngivning!
- Om möjligt, försök i första hand att inte använda interaktiva plottar. Anledningen är att för många sådana kan fylla upp RAM-minnet och således slöa ner hela notebooken.

Förstå detta rätt, ni får självklart använda interaktiva plottar – men försök vara lite sparsamma.