**Portfolioopgave 1**

I dette dokument vil jeg kort beskrive min tankeproces under arbejdet med portfolioopgave 1. Samtidig vil jeg kort gøre redde for hvad datasættet opgaven omhandlede indeholdte.

Datasættet ”titanic.csv” indeholder en masse forskellige informationer omkring 887 af de 2224 personer ombord på RMS Titanic. Det er både informationer såsom deres navne, alder, passagerklasse, køn osv., samt om de overlevede rejsen. Disse er gemt i form af en CSV-fil med kommaer som separators.

Jeg har grebet opgaverne an på meget gradvis manér. Jeg er gået de 5 skridt igennem et ad gangen, og sørget for at ikke bare implementere den rigtige kode, men også forstå den. Jeg har både gjort brug af kilderne fra opgavebeskrivelsen, men også fundet flere / andre varianter via google og specielt reddit/r/learnpython. Jeg har prøvet, så vidt som muligt, at finde og bruge de værktøjer som pandas stiller tilgængelig, for at kunne løses opgaven bedst muligt. Jeg synes af og til at pandas har haft en lidt stejl læringskurve, men når det er sagt, så bliver det kun mere og mere tydeligt, for hver ny ting man lærer, at det er et utroligt brugbart værktøj. Til tider var jeg fristet til at omdanne visse kolonner i pandas dataframe til lister, og håndtere regneprocessen selv (specielt tidligt i arbejdet med opgave 3), indtil jeg fandt ud af at pandas funktioner kunne gøre præcis det samme, og mere til, uden at bruge nær så mange linjer af kode.

Den opgave der voldte mit flest problemer var uden tvivl opgave 4, som lød på at finde ud af om der var flere passagerer ombord med samme efternavn. Efter forgæves at have ledt efter en måde at ordne det på, kun ved hjælp af pandas, besluttede jeg mig for i stedet at lave navnekolonnen i dataframe’en om til en liste, fange alle efternavne ved hjælp at et for-loop og derefter konvertere listen om til en dataframe igen.

Jeg fik også gjort lidt brug af numpy ved opgave 5, hvilket lader til at spille rigtig godt sammen med pandas.