

Portfolio 1

1.)

Titanic datasættet består af følgende 8 kolonner og 887 rækker

```
>>> print(titanic.shape)
```

a.)

Survived - Pclass - Name - Sex - Age - Siblings/Spouses Aboard - Parents/Children Aboard - Fare

```
>>> print(titanic.columns)
```

b.)

ang. manglende data så kan jeg ikke se, at der er nogle "huller" i de rækker, som jeg har scrollet igennem, dog er der lidt forskellige "titler, lige fra Mr og Mrs, til Master, Lady og Major.

Desuden kan man ikke se nationalitet, i fald det skulle være relevant, men man kan se af efternavnene, at "passagererne er fra forskellig herkomst".

2.)

Datasættet undersøges ud fra de funktioner, som findes i pandas til beskrivelse af en dataframe - resultatet er oploadet som assignment1.py fil til min Github

3.)

- Hvor mange overlevede
 - >>>titanic['Survived'].sum() = 342
 -
- Gennemsnitsalderen på samtlige passagere
 - >>> titanic['Age'].mean() = 29.4
 -
- Medianen på personerne
 - >>> titanic['Age'].median() = 28.0
 -
- Hvor mange kvinder var med
 - >>>titanic[titanic['Sex'] == 'female'] = 314 rækker må være 314 kvinder (female som var ombord)
 -
- hvor mange af dem overlevede
 - 233 kvinder overlevede..
 -
- Personer med samme efternavn
 - ikke færdig, simpelthen for svær:)

4.)

pivot tabellen kan ses fint i Anaconda/Spyder ->

Pclass 1 2 3

Survived 136 87 119

hvilken rejseklasse havde flest overkomne, ... kommer i næste iteration:)