Portfolio 1 – Titanic CSV fil

Opgave 1: Indholdet af datasættet

a) Filen har forskellige datatyper som alder, køn og familienotater om passagerne. Denne data blev først åbnet in text editor, for at se hvad datasættet indeholdt af datatyper. Datasættet indeholder navne på passagerne, hvilken klasse de rejste med, om de overlevede, deres køn og alder. Dernæst indeholdt datasættet også prisen på deres billet, og om hvorvidt de havde søskende, partnere eller forældre ombord.

b) Af data der mangler, så er der informationer der ikke vides noget om. F.eks. er der et par personer, hvor der ikke vides om de havde familie ombord, eller hvad deres billet kostede.

Opgave 2a: Beskrivelse af datasættet med funktioner i Pandas

a) Efter at have indlæst datasættet i Spyder bruger jeg funktioner fra Pandas til at beskrive datasættet.
F.eks. finder jeg ud af hvor mange rækker der er ved hjælp af print(len(data)) = 887 rækker
Når jeg bruger kommandoen shape vises antal af alle rækker og kolonnerne = (887, 8)
Data.colums fortæller hvad navne kolonnerne har
= 'Survived', 'Pclass', 'Name, 'Sex', 'Siblings/Spouses Aboard', 'Parents/Children Aboard' og 'Fare'.
Data.dtypes fortæller hvilke typer af dataet er, f.eks. object, int eller floats

Opgave 3: Udtrækning og beregning af data i filen

Der bruges data.describe for at udregne bl.a. medianen og gennemsnittet.
Print(data['Survived']) viser dem der overlevede ud af de passagerer som er inkluderet i datasættet.
Sum bruges for at beskrive summen af de der overlevede = 342
Medianen af alder for at vise alderen på passagerne = 28.0

Opgave 4: Personer med samme efternavn?

Opgaven ønsker at finde ud af om personer har det samme efternavn. Her kan jeg lave en ny variabel last_names ud fra datasættets 'Name'

```
last_names = data['Name'].str.split(expand = True)
```

Str.split returnerer en liste af seperate ord i en string. Expand = true separerer string ind i kolonner, her 'Name'.

```
last_names[1].value_counts()
```

Value_counts tæller 'frequency counts' af elementer ved hjælp af Pandas

```
last_names = (data['Name'].str.split().str[-1])
```

-1 peger på det sidste tegn i string ved 'Name', og dermed efternavne nu

```
print(last_names.value_counts())
```

Her finder jeg efternavnene der går igen. F.eks. er der 9 passagere med navnet Andersson

```
Andersson    9
Sage         7
Skoog        6
Carter       6
Panula       6
             ..
Honkanen     1
Cann         1
Moraweck     1
hoef         1
Frolicher    1
Name: Name, Length: 664, dtype: int64
```

Opgave 5: Pivot tabel af de rejsendes klasser

Tabellen viser at flest mennesker på første klasse overlevede

Importerer numpy

```
import numpy as np
```

Med Pandas laver der en pivot-tabel hvor columns er klasse, og values er 'Age'

```
class_titanic = data.pivot_table(columns='Pclass', values='Age', aggfunc='count')
```

print(class_titanic)

```
In [15]: class_titanic = data.pivot_table(columns='Pclass', values='Age',
aggfunc='count')
    ...:
    ...: print(class_titanic)
Pclass    1    2    3
Age     216  184  487
```

Pivot-tabel over hvilke klasser der havde flest omkomne.

lost_lives_titanic = pd.pivot_table(data,index="Pclass", columns='Survived', values='Name', aggfunc='count')

Det var 3. klasse med 368 tabte liv.

print(lost_lives_titanic)

```
In [16]: lost_lives_titanic = pd.pivot_table(data,index="Pclass",
columns='Survived', values='Name', aggfunc='count')

In [17]: print(lost_lives_titanic)
Survived    0    1
Pclass
1          80  136
2          97   87
3         368  119
```

Litteraturliste:

Built-in Types¶. (n.d.). Retrieved from https://docs.python.org/3/library/stdtypes.html#str.split.

pandas.pivot_table¶. (n.d.). Retrieved from https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.pivot_table.html.

pandas.Series.str.split¶. (n.d.). Retrieved from https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.str.split.html.

pandas.Series.value_counts¶. (n.d.). Retrieved from https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.Series.value_counts.html.

Python Intro for Libraries. (n.d.). Retrieved from https://librarycarpentry.org/lc-python-intro/.