



Portfolio 1

Open Data Science

Eksaminand: Stephanie Rose Acampado Soelmark, PZG932

Gruppe:

Martine Ingemann Jørgensen, JPR328

Natacha Rylander Bech, TGZ940

Stephanie Rose Acampado Soelmark, PZG932

6 januar 2020

Denne opgave behandler Titanic datasættet, der indeholder data fra 887 passagerer. Herunder information om overlevelse, navn, alder med mere. Vi har analyseret denne data og i den forbindelse visualiseret resultaterne i form af tabeller. Denne rapport behandler således opgave 5 i opgavesættet.

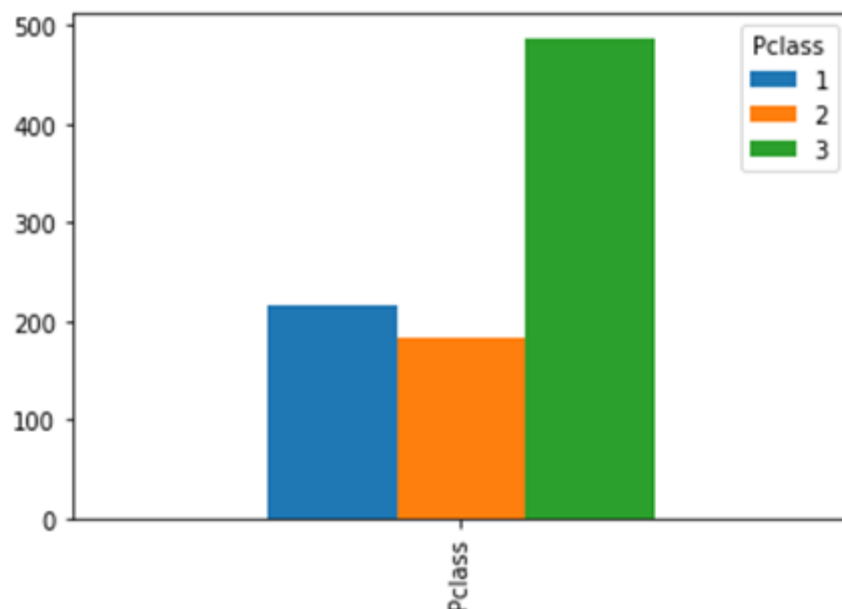
Opgave 5 – Pivot-tabeller

Vi har først genereret en pivot-tabel der viser antal rejsende inden for hver klasse (med brug af pandas funktioner). I dette tabel-format kan man angive hvilken kolonne, der skal vises, og hvilken funktion, der skal behandle den. Dette har vi valgt at drage fordel af i vores tabel. Dermed definerer vi columns til 'Pclass' for at få de diverse klasser og anvender aggfunc, som udgør den definition man ønsker at anvende, til at definere count, da vi ønsker en optælling af alle rejsende opdelt i klasser (pandas.pivot_table, u.å). Herefter anvender vi plot(kind='bar'), således vi også kan få et visuelt output.

```
class_tabel = df_titanic.pivot_table(columns='Pclass', aggfunc=({'Pclass': 'count'}))
```

```
print(class_tabel)
```

```
class_tabel.plot(kind='bar')
```



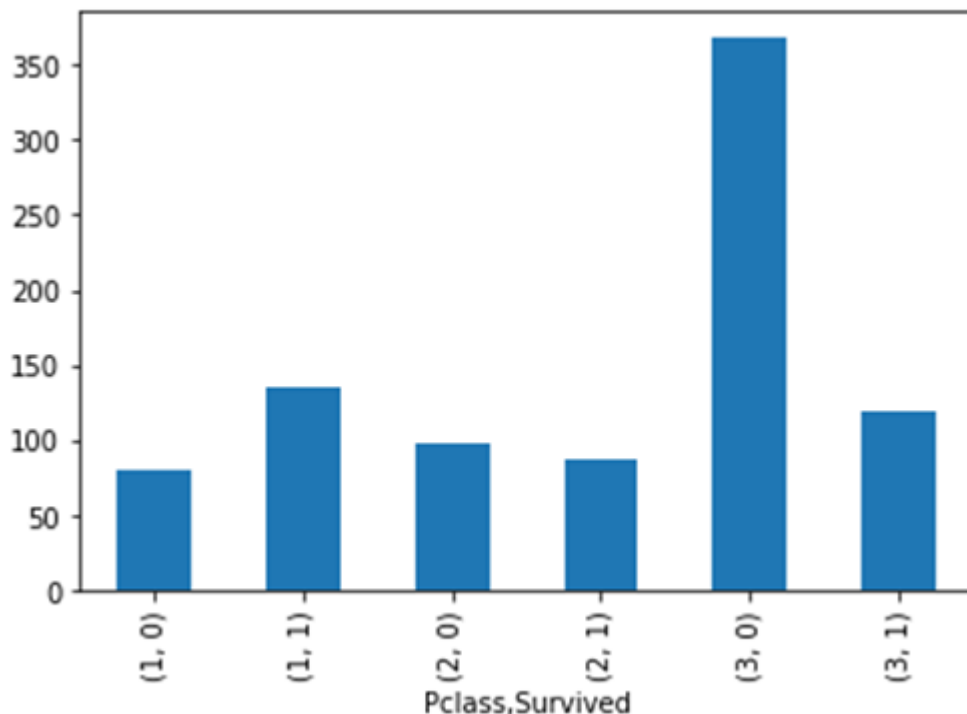
Således påvises det at der var 216 passagerer på første klasse, 184 på anden klasse og 487 på tredje klasse.

Herefter har vi lavet en pivot-tabel over antal overlevende og omkomne i hver klasse, hvor 0 er omkomne og 1 er overlevende. Funktionen optæller antallet for hver værdi i survived. Endnu engang har vi lavet en visuel fremstilling med `plot(kind='bar')` funktionen.

```
nyplot = df_titanic.groupby(['Pclass', 'Survived'])['Survived'].count()
```

```
print(nyplot)
```

```
nyplot.plot(kind='bar')
```



Således påviser søjlediagrammet at der på første klasse var 80 omkomne, på anden klasse var der 97 omkomne og på tredje klasse var der 368 omkomne. Dermed var 3 klasse den med flest omkomne

Slutteligt er endnu en pivot-tabel, der udelukkende viser fordelingen af overlevende i hver klasse. Først har vi lavet en ny dataframe, der består af kolonnerne, Survived og Class. Dernæst har vi anvendt `.groupby()` til at visualisere fordelingen af overlevende samt omkomne på de forskellige klasser.

For at generere pivot-tabellen, anvender vi den nye dataframe, værdien af de overlevende, som bliver opstillet i en kolonne sat i forhold til de tre klasser.

For at få antallet af overlevende anvender vi `sum()`. Funktionen finder således summen af alle værdierne, bestående af 0 og 1, og fremviser disse i tabellen ud fra hver klasse. Igen anvender vi `plot(kind='bar')` for at visualisere resultaterne.

```
ny_df = df_titanic[['Survived', 'Pclass']]
```

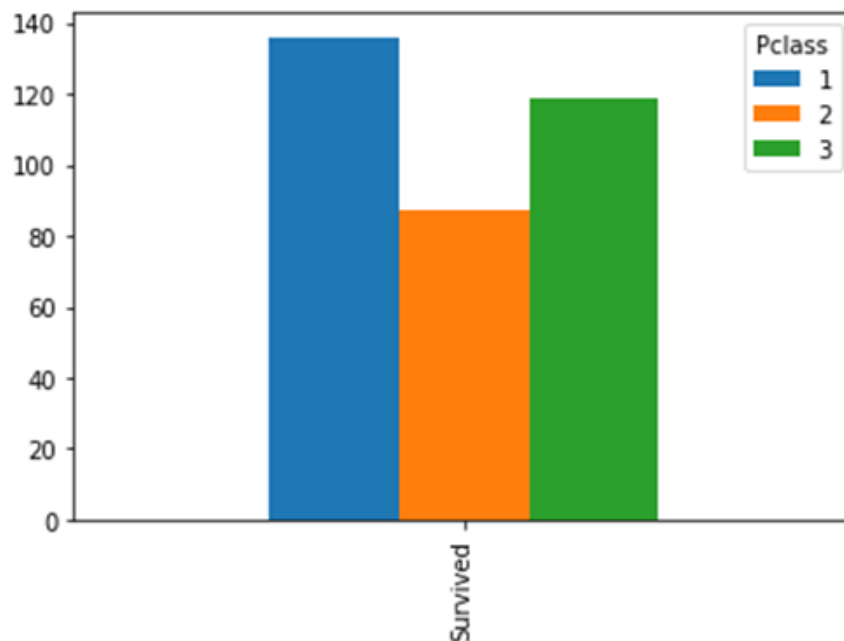
```
print(ny_df)
```

```
ny_df.shape
```

```
tabell = pd.pivot_table(ny_df, values= 'Survived', columns= 'Pclass', aggfunc= 'sum') #antal  
overlevende på de forskellige klasser
```

```
print(tabell)
```

```
vistabell = tabell.plot(kind='bar')
```



Overordnede søjlediagram påviser således antal overlevende per klasse. De fleste overlevende på henholdsvis første og tredje klasse (136 og 119). Mens de færrest overlevende var på anden klasse (87).

Litteraturliste

Pandas.pydata.org. (u.å). pandas.pivot_table. Lokaliseret d. 19. september 2019 på:
https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.pivot_table.html