titanic

August 12, 2023

1 Estudio Titanic

[ ]:
```
import seaborn as sbs
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[ ]:
```
titanic = sbs.load_dataset('titanic')
titanic.head()
```

[ ]:
```
   survived  pclass     sex   age  sibsp  parch     fare embarked  class  \
0         0       3    male  22.0      1      0   7.2500        S  Third
1         1       1  female  38.0      1      0  71.2833        C  First
2         1       3  female  26.0      0      0   7.9250        S  Third
3         1       1  female  35.0      1      0  53.1000        S  First
4         0       3    male  35.0      0      0   8.0500        S  Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True
3  woman       False    C  Southampton   yes  False
4    man        True  NaN  Southampton    no   True
```

[ ]:
```
# Declaramos las variables categoricas
titanic['survived'] = pd.Categorical(titanic['survived'])
titanic['pclass'] = pd.Categorical(titanic['pclass'])
titanic['sex'] = pd.Categorical(titanic['sex'])
titanic['age'] = pd.Categorical(titanic['age'])
```

[ ]:
```
survived_sex = pd.crosstab(index=titanic['survived'], columns=titanic['sex'])
survived_sex
```

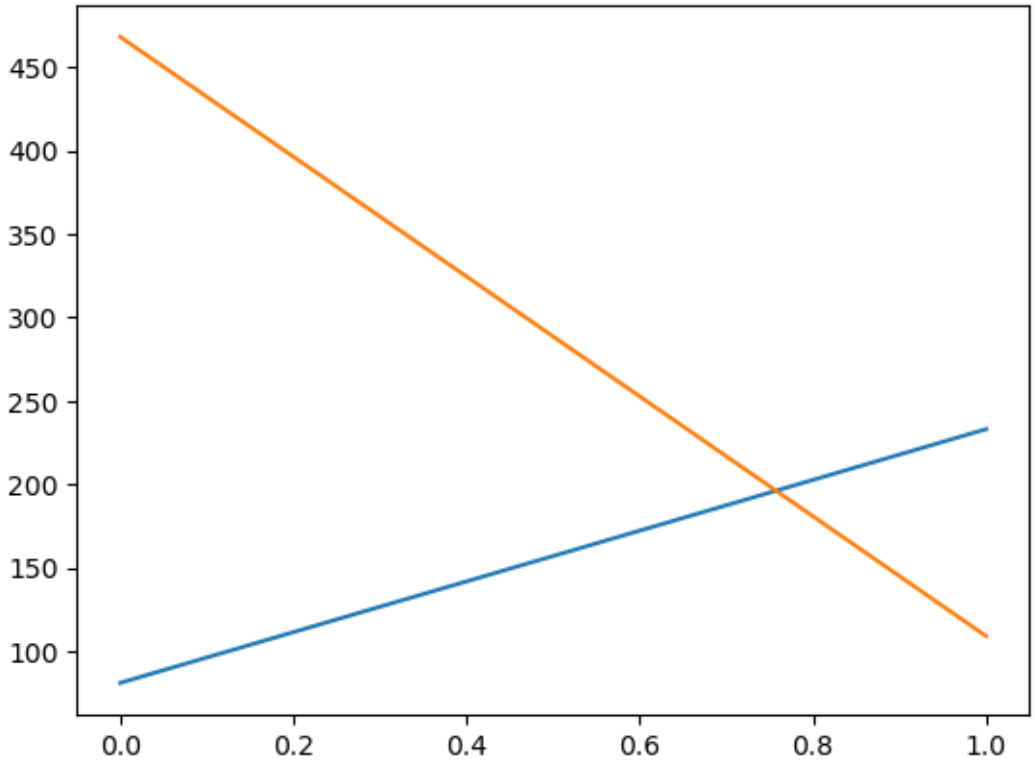
[ ]:
```
sex       female  male
survived
0             81   468
1            233   109
```

[ ]:
```
survived_sex / survived_sex.sum()
```

[ ]:
```
sex         female      male
survived
0         0.257962  0.811092
1         0.742038  0.188908
```

[ ]:
```
print(plt.plot(survived_sex))
```

```
[<matplotlib.lines.Line2D object at 0x7f44985f3690>, <matplotlib.lines.Line2D
object at 0x7f4498623b90>]
```



[ ]:
```
count_per_sex = pd.crosstab(index=titanic['sex'], columns='count')
count_per_sex
```

[ ]:
```
col_0   count
sex
female    314
male      577
```

[ ]:
```
percent_per_sex = count_per_sex / count_per_sex.sum()
percent_per_sex
```

[ ]:
```
col_0      count
sex
female  0.352413
male    0.647587
```

1.1 Tablas multi dimensionales

[ ]:
```
surv_sex_class = pd.crosstab(index=titanic['survived'],␣
↪columns=[titanic['sex'], titanic['pclass']], margins=True)
surv_sex_class
```

[ ]:
```
sex      female          male            All
pclass        1   2    3    1    2    3
survived
0             3   6   72   77   91  300  549
1            91  70   72   45   17   47  342
All          94  76  144  122  108  347  891
```

[ ]:
```
surv_sex_class / surv_sex_class.sum()
```

[ ]:
```
sex         female                     male                          All
pclass           1         2     3         1         2         3
survived
0         0.015957  0.039474  0.25  0.315574  0.421296  0.432277  0.308081
1         0.484043  0.460526  0.25  0.184426  0.078704  0.067723  0.191919
All       0.500000  0.500000  0.50  0.500000  0.500000  0.500000  0.500000
```

[ ]:
```
surv_sex_class['female'][2]
```

[ ]:
```
survived
0       6
1      70
All    76
Name: 2, dtype: int64
```

[ ]:
```
surv_sex_class['male'][1]
```

[ ]:
```
survived
0       77
1       45
All    122
Name: 1, dtype: int64
```

[ ]:
```

```