

Simpson's Paradox

Use `admission_data.csv` for this exercise.

```
In [2]: # Load and view first few lines of dataset
import pandas as pd
admits = pd.read_csv('admission_data.csv')
admits.head()
```

Out[2]:

	student_id	gender	major	admitted
0	35377	female	Chemistry	False
1	56105	male	Physics	True
2	31441	female	Chemistry	False
3	51765	male	Physics	True
4	53714	female	Physics	True

Proportion and admission rate for each gender

```
In [12]: print (len(admits[admits['gender']=='female']))
print (admits.shape[0])
```

257
500

```
In [13]: # Proportion of students that are female
(len(admits[admits['gender']=='female']))/admits.shape[0]
```

Out[13]: 0.514

```
In [14]: # Proportion of students that are male
(len(admits[admits['gender']=='male']))/admits.shape[0]
```

Out[14]: 0.486

```
In [15]: # Admission rate for females
len(admits[(admits['gender']=='female') & (admits['admitted'])])/(len
(admits[admits['gender']=='female']))
```

Out[15]: 0.28793774319066145