Bias analyse of Adult dataset.nbconvert

January 19, 2020

1 Analyse bias dataset

This notebook main goal is to : - See if data is biased on protected attributes

1.1 Load packages

[1]:
```
import pandas as pd
from IPython.display import display, Markdown

# transparentai package : https://github.com/Nathanlauga/transparentai
from transparentai.datasets import ClassificationDataset
```

1.2 Load dataset

[2]:
```
dataset = pd.read_csv('../data/adult.csv', sep=',')
```

[3]:
```
target = 'income'
target = None if target not in dataset.columns else target
```

1.3 Create ClassificationDataset

[4]:
```
fair_dataset = ClassificationDataset(df=dataset,
                                     label_name=target,
                                     privileged_values={'marital-status':␣
↪['Married-civ-spouse', 'Married-AF-spouse'], 'race': ['White'], 'gender':␣
↪['Male']})
```

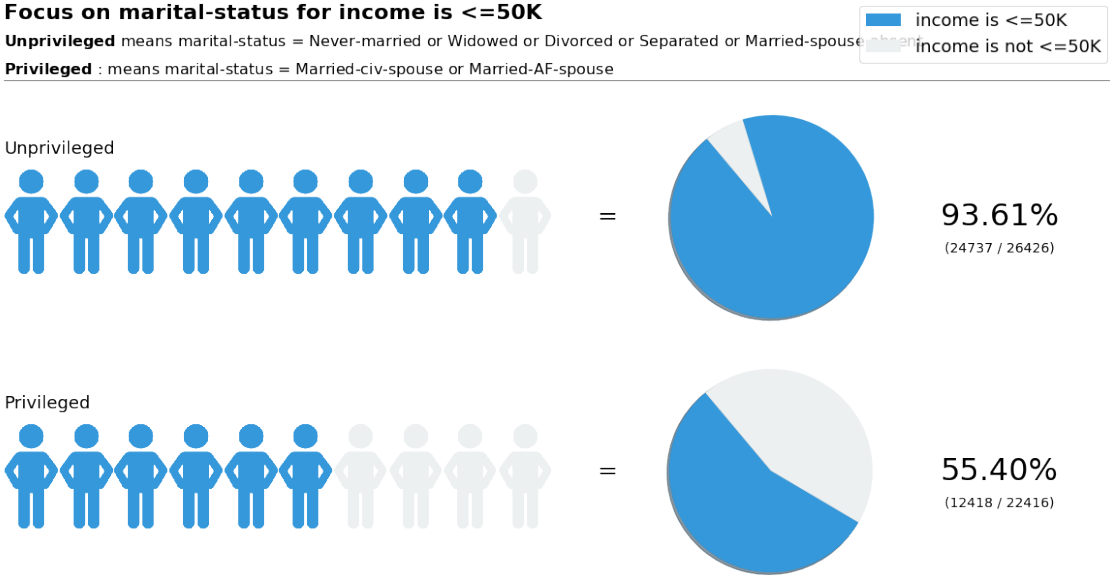
1.4 Show if dataset is biased on protected attributes

[5]:
```
fair_dataset.show_bias_metrics()
```

Focus on marital-status for income is <=50K

Unprivileged means marital-status = Never-married or Widowed or Divorced or Separated or Married-spouse-absent

Privileged : means marital-status = Married-civ-spouse or Married-AF-spouse



Disparate impact

$$\frac{0.94}{0.55} = 1.69$$

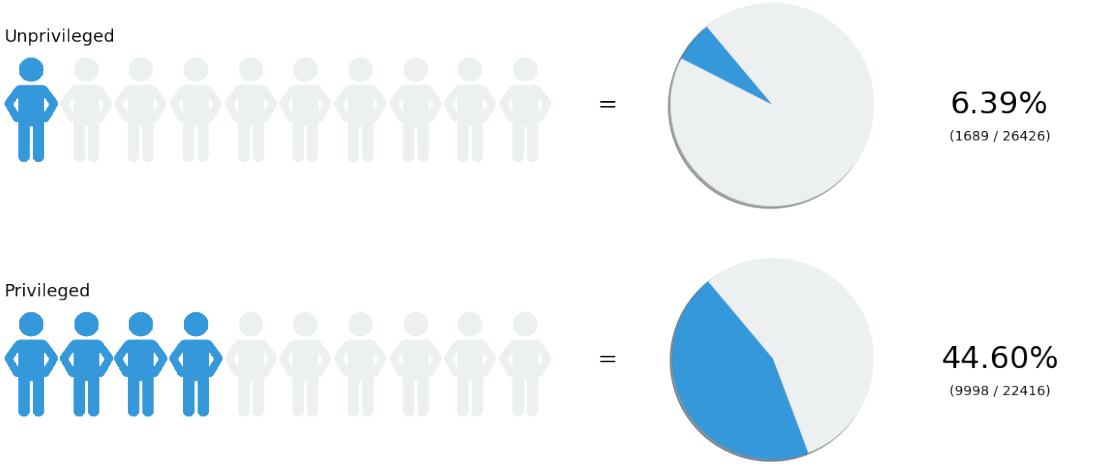
Statistical parity difference

$$93.61 - 55.40 = 38.21$$

Focus on marital-status for income is >50K

Unprivileged means marital-status = Never-married or Widowed or Divorced or Separated or Married-spouse-absent

Privileged : means marital-status = Married-civ-spouse or Married-AF-spouse



Disparate impact

$$\frac{0.06}{0.45} = 0.14$$

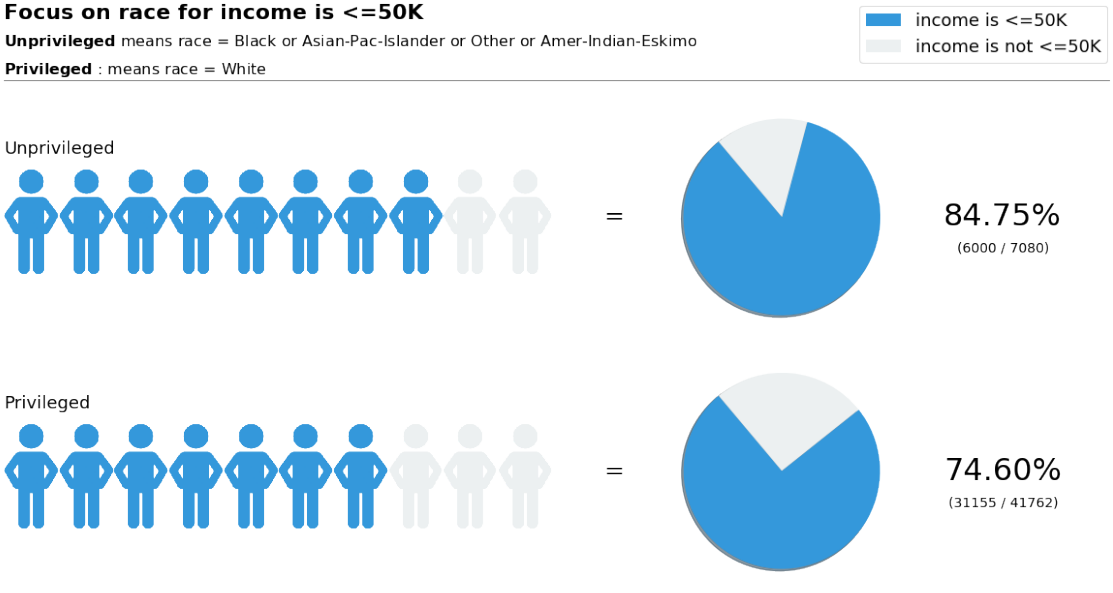
Statistical parity difference

$$6.39 - 44.60 = -38.21$$

Focus on race for income is <=50K

Unprivileged means race = Black or Asian-Pac-Islander or Other or Amer-Indian-Eskimo

Privileged : means race = White



Disparate impact

$$\frac{0.85}{0.75} = 1.14$$

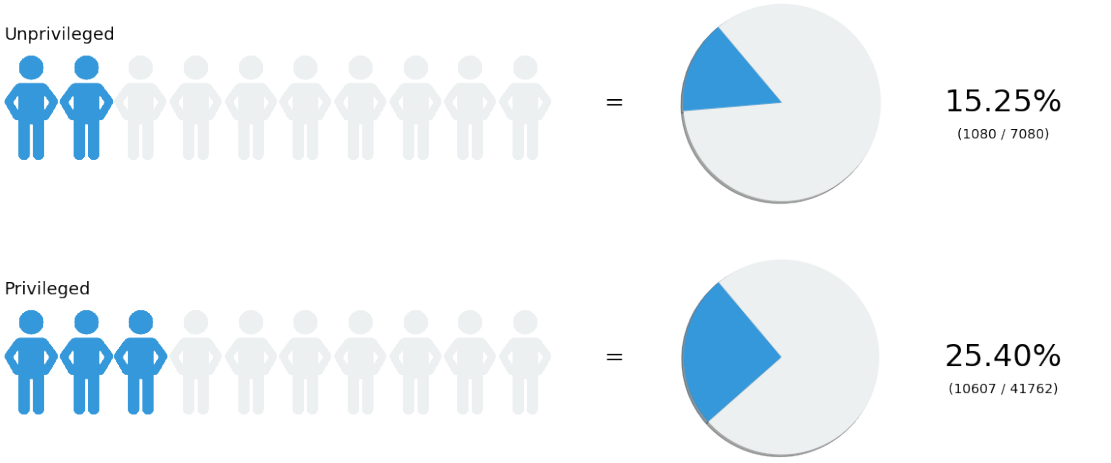
Statistical parity difference

$$84.75 - 74.60 = 10.14$$

Focus on race for income is >50K

Unprivileged means race = Black or Asian-Pac-Islander or Other or Amer-Indian-Eskimo

Privileged : means race = White

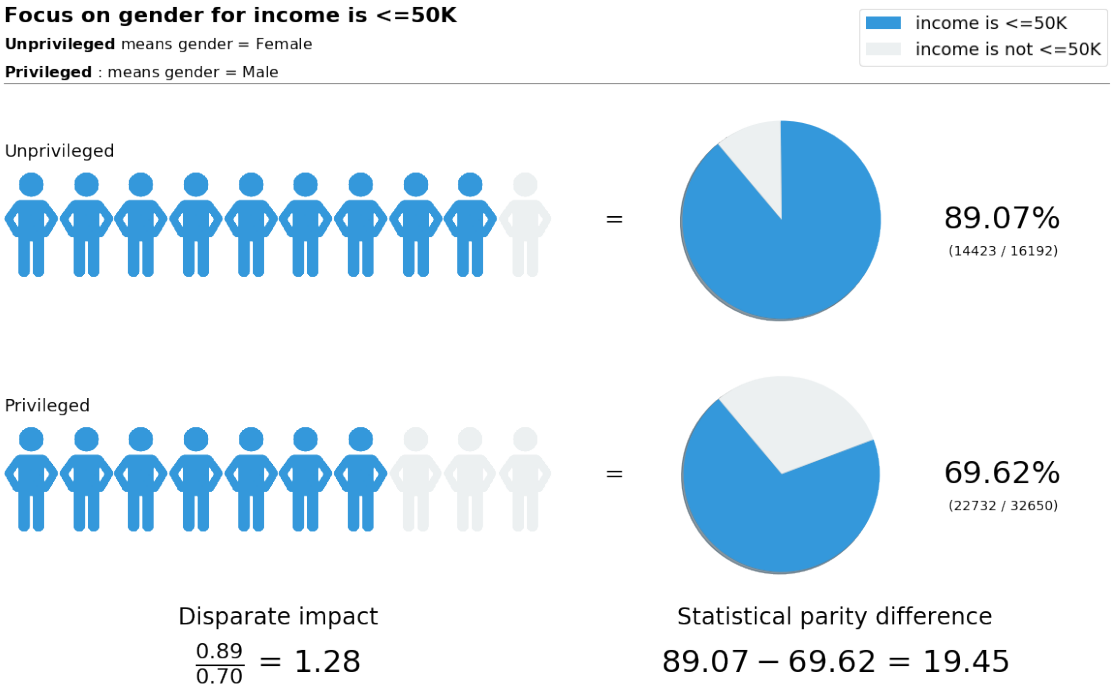


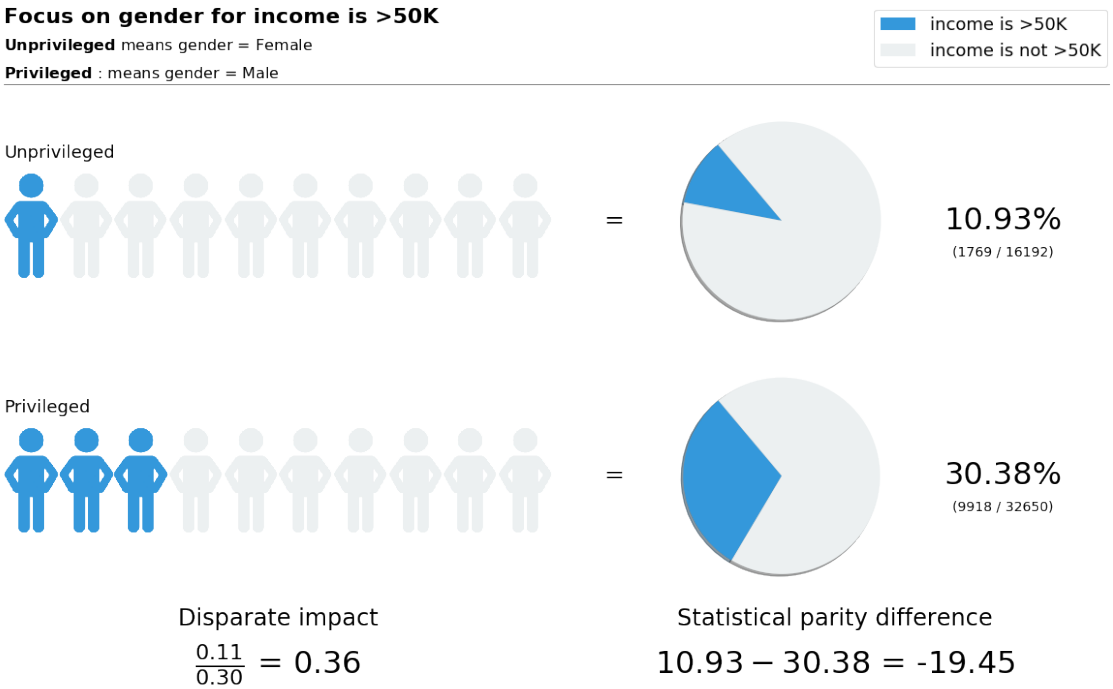
Disparate impact

$$\frac{0.15}{0.25} = 0.60$$

Statistical parity difference

$$15.25 - 25.40 = -10.14$$


Focus on gender for income is <=50K
Unprivileged means gender = Female
Privileged : means gender = Male
income is <=50K
income is not <=50K
Unprivileged
=
89.07%
(14423 / 16192)
Privileged
=
69.62%
(22732 / 32650)
Disparate impact
0.89 / 0.70 = 1.28
Statistical parity difference
89.07 − 69.62 = 19.45


Focus on gender for income is >50K
Unprivileged means gender = Female
Privileged : means gender = Male
income is >50K
income is not >50K
Unprivileged
=
10.93%
(1769 / 16192)
Privileged
=
30.38%
(9918 / 32650)
Disparate impact
0.11 / 0.30 = 0.36
Statistical parity difference
10.93 − 30.38 = -19.45

The end