face-mask-dataset-statistics

April 3, 2021

[6]:
```
import pandas as pd
import numpy as np
import seaborn as sns #visualisation
import matplotlib.pyplot as plt #visualisation
%matplotlib inline
sns.set(color_codes=True)
```

[3]:
```
df = pd.read_csv('train_labels.csv')
```

[7]:
```
# To display the top 5 rows
df.head()
```

[7]:

	filename	label	label (0/1)	label (no/yes)
0	Image_1.jpg	without_mask	0	no
1	Image_2.jpg	without_mask	0	no
2	Image_3.jpg	without_mask	0	no
3	Image_4.jpg	without_mask	0	no
4	Image_5.jpg	without_mask	0	no

[8]:
```
# Checking the data type
df.dtypes
```

[8]:
```
filename          object
label             object
label (0/1)        int64
label (no/yes)    object
dtype: object
```

[9]:
```
# Total number of rows and columns
df.shape
```

[9]:
```
(11264, 4)
```

[10]:
```
# Finding the null values.
print(df.isnull().sum())
```

```
filename          0
label             0
```

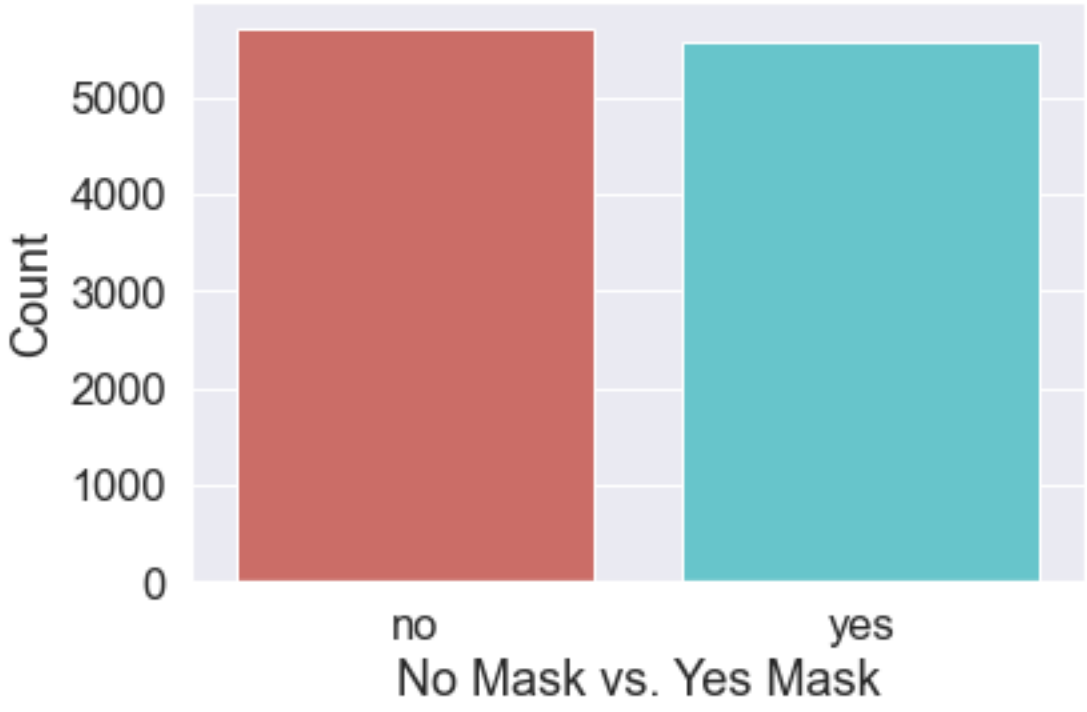
```
label (0/1)       0
label (no/yes)    0
dtype: int64
```

[13]:
```
# Describe statistics of the dataset
df.describe()
```

[13]:

	label (0/1)
count	11264.000000
mean	0.493874
std	0.499985
min	0.000000
25%	0.000000
50%	0.000000
75%	1.000000
max	1.000000

[30]:
```
# Check the dominant class.
countplt=sns.countplot(x='label (no/yes)', data=df, palette ='hls')
plt.xlabel("No Mask vs. Yes Mask")
plt.ylabel("Count")
plt.show()
```



[23]:
```
# Count the number of images with no mask
count_no_sub = len(df[df['label (no/yes)']=='no'])
print("The number of images with NO mask : {}".format(count_no_sub))

# Count the number of images with mask
count_yes_sub = len(df[df['label (no/yes)']=='yes'])
print("The number of images with mask    : {}".format(count_yes_sub))
```

```
The number of images with NO mask : 5701
The number of images with mask    : 5563
```

Given the results, the dataset is approximately evenly split between the two classes, i.e. images with no mask and images with mask.