

gender_analiza

August 1, 2025

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
```

```
[2]: df_exercise = pd.read_csv("../datasets/gender_vjezba.csv")
```

```
[3]: df_exercise
```

```
[3]:      Unnamed: 0  Gender  ExerciseFrequency
0             0    Male           6.745071
1             1    Male           5.792604
2             2    Male           6.971533
3             3    Male           8.284545
4             4    Male           5.648770
..          ...    ...
195          195  Female           6.077976
196          196  Female           4.174214
197          197  Female           5.730588
198          198  Female           5.587313
199          199  Female           3.785545
```

[200 rows x 3 columns]

```
[4]: df_exercise.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 3 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            200 non-null   int64
1   Gender                200 non-null   object
2   ExerciseFrequency     200 non-null   float64
dtypes: float64(1), int64(1), object(1)
memory usage: 4.8+ KB
```

```
[5]: df_exercise.Gender.value_counts()
```

```
[5]: Gender
      Male      100
      Female    100
      Name: count, dtype: int64
```

```
[6]: genders = df_exercise.groupby("Gender")
```

```
[7]: genders["ExerciseFrequency"].describe()
```

```
[7]:
```

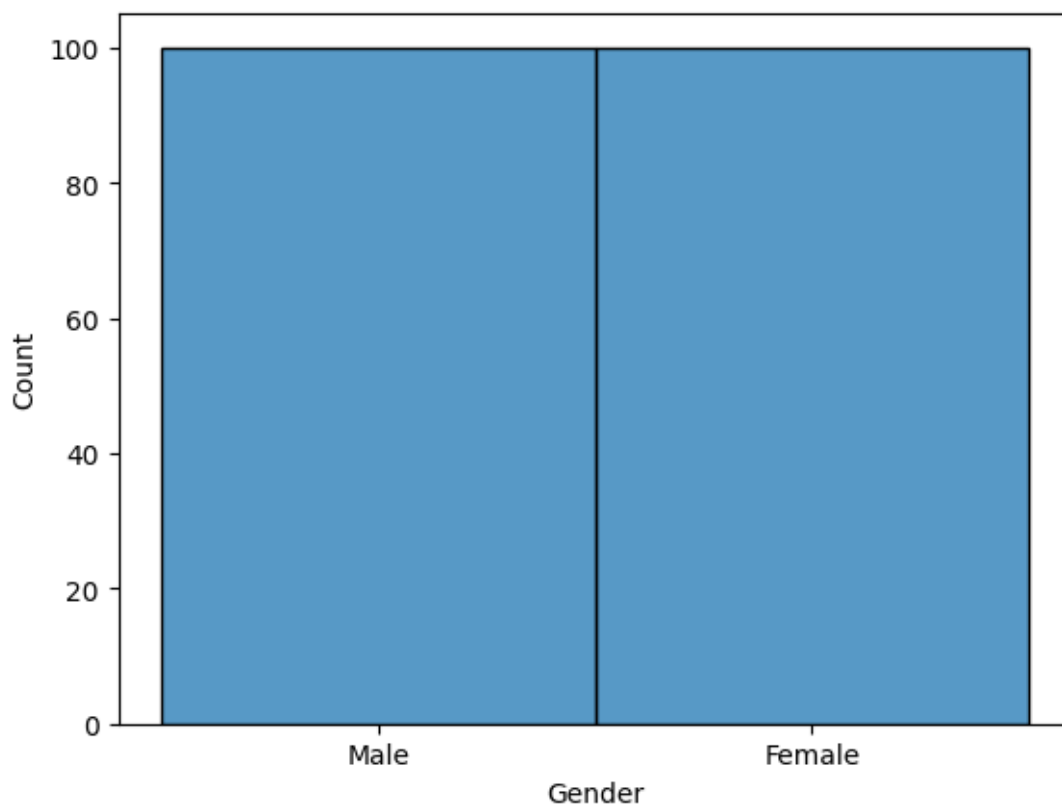
	count	mean	std	min	25%	50%	75%	\
Gender								
Female	100.0	5.533457	1.430503	2.621843	4.291509	5.626161	6.307256	
Male	100.0	5.844230	1.362253	2.070382	5.098641	5.809566	6.608928	


```
max
```

Gender	
Female	9.580254
Male	8.778417

```
[8]: sns.histplot(df_exercise.Gender)
```

```
[8]: <Axes: xlabel='Gender', ylabel='Count'>
```



```
[9]: df_exercise.corr(numeric_only=True)
```

```
[9]:
```

	Unnamed: 0	ExerciseFrequency
Unnamed: 0	1.000000	-0.083867
ExerciseFrequency	-0.083867	1.000000

```
[10]: male = df_exercise[df_exercise['Gender'] == 'Male']['ExerciseFrequency']  
female = df_exercise[df_exercise['Gender'] == 'Female']['ExerciseFrequency']
```

```
[11]: F_statistic, p_value = stats.f_oneway(male, female)
```

```
[12]: F_statistic
```

```
[12]: np.float64(2.4750968599505234)
```

```
[13]: p_value
```

```
[13]: np.float64(0.11725885247920977)
```