

## ✓ Loading data


```
! gdown 1UCd-hnjyowsL1yJZAmrhuyh_wezAKA1Z
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
df = pd.read_csv('/content/1.csv')
df.head(8)
```

 [Показать скрытые выходные данные](#)

Далее: [Создать код с переменной df](#) [Посмотреть рекомендованные графики](#) [New interactive sheet](#)

## ✓ Cleaning

```
df['Income'].fillna(0.0, inplace=True)
df.head(8)
```

 /tmp/ipython-input-24-741927898.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col]

```
df['Income'].fillna(0.0, inplace=True)
```

	Age	Gender	Income	Favorite Transport
0	5	0	0.0	bicycle
1	8	1	0.0	scooter
2	10	0	0.0	bicycle
3	14	1	0.0	metro
4	16	1	0.0	metro
5	20	1	200.0	scooter
6	22	0	500.0	scooter
7	23	1	300.0	scooter



Далее: [Создать код с переменной df](#) [Посмотреть рекомендованные графики](#) [New interactive sheet](#)

## ✓ Encoding

```
label_encoder=LabelEncoder()
df['Gender']=label_encoder.fit_transform(df['Gender'])
df.head(8)
```

 [Показать скрытые выходные данные](#)

Далее: [Создать код с переменной df](#) [Посмотреть рекомендованные графики](#) [New interactive sheet](#)

## ✓ Model

```
X=df.drop(columns='Favorite Transport')
Y=df['Favorite Transport']
model=DecisionTreeClassifier()
model.fit(X,Y)
```

 [Показать скрытые выходные данные](#)

## ✓ Prediction

```
test_df=pd.DataFrame({
    'Age': [12],
    'Gender': [0],
    'Income': [0.0]
})
model.predict(test_df)

↵ array(['bicycle'], dtype=object)
```

- ✓ Если посмотреть на исходные данные, то есть похожая запись. Именно на основании этой записи модель предположила, что девочка 12 лет также любит кататься на велосипеде.

```
df.loc[2]
```

```
↵
```

	2
<b>Age</b>	10
<b>Gender</b>	0
<b>Income</b>	0.0
<b>Favorite Transport</b>	bicycle

**dtype:** object

## ✓ Exporting to the dot file

```
from sklearn import tree
tree.export_graphviz(model, out_file='decision_tree_model.dot', feature_names=['Age', 'Gender', 'Income'], filled=True, class_names=sorte
```

## ✓ Отображение диаграмм на основании данных для модели

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.histplot(x=df['Income'], hue=df['Favorite Transport'])
plt.show()
sns.countplot(x=df['Gender'], hue=df['Favorite Transport'])
plt.show()
```

