


```
import pandas as pd
import numpy as np
```

```
train_data=pd.read_csv("/content/train.csv")
```

```
train_data.head()
```




	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2834

Next steps: [View recommended plots](#)

```
test_data=pd.read_csv("/content/test.csv")
```

```
test_data.head()
```



	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	

Next steps: [View recommended plots](#)

```
women = train_data.loc[train_data.Sex == 'female']["Survived"]
rate_women = sum(women)/len(women)
```

```
print("% of women who survived:", rate_women)
```

```
% of women who survived: 0.7420382165605095
```

```
men = train_data.loc[train_data.Sex == 'male']["Survived"]
rate_men = sum(men)/len(men)
```

```
print("% of men who survived:", rate_men)
```

```
% of men who survived: 0.18890814558058924
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
y = train_data["Survived"]
```


```
features = ["Pclass", "Sex", "SibSp", "Parch"]
X = pd.get_dummies(train_data[features])
X_test = pd.get_dummies(test_data[features])
```

```
model = RandomForestClassifier(n_estimators=100, max_depth=5, random_state=1)
model.fit(X, y)
```

```
model.predict(X_test)

predictions = model.predict(X_test)

output = pd.DataFrame({'PassengerId': test_data.PassengerId, 'Survived': predictions})
output.to_csv('submission.csv', index=False)
print("Your submission was successfully saved!")
```

 Your submission was successfully saved!