[27]:
```
import pandas as pd
import numpy as np

# Load Titanic dataset
df = pd.read_csv('Titanic-Dataset - Titanic-Dataset.csv')
df.head()
```

[27]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

[11]:
```
df.isnull().sum()
```

[11]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

[13]:
```
df['Age'] = df['Age'].fillna(df['Age'].mean())
```

[15]:
```
df.head()
```

[15]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

[17]:
```
df['Embarked'] = df['Embarked'].fillna(df['Embarked'].mode()[0])
```

[19]:
```
df.head()
```

[19]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

[21]:
```
df['Sex'] = df['Sex'].map({'female': 0, 'male': 1})
df['Embarked'] = df['Embarked'].map({'S': 0, 'C': 1, 'Q': 2})
df.head()
```

[21]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	NaN	0
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.0	1	0	PC 17599	71.2833	C85	1
2	3	1	3	Heikkinen, Miss. Laina	0	26.0	0	0	STON/O2. 3101282	7.9250	NaN	0
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	35.0	1	0	113803	53.1000	C123	0
4	5	0	3	Allen, Mr. William Henry	1	35.0	0	0	373450	8.0500	NaN	0

[25]:
```
# Create a new feature 'FamilySize'
df['FamilySize'] = df['SibSp'] + df['Parch'] + 1
df.head()
```

[25]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	FamilySize
0	1	0	3	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	NaN	0	2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	0	38.0	1	0	PC 17599	71.2833	C85	1	2
2	3	1	3	Heikkinen, Miss. Laina	0	26.0	0	0	STON/O2. 3101282	7.9250	NaN	0	1
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	35.0	1	0	113803	53.1000	C123	0	2
4	5	0	3	Allen, Mr. William Henry	1	35.0	0	0	373450	8.0500	NaN	0	1

[29]:
```
from sklearn.preprocessing import StandardScaler
# Standardize the 'Age' and 'Fare' columns
scaler = StandardScaler()
df[['Age', 'Fare']] = scaler.fit_transform(df[['Age', 'Fare']])
```

[31]:
```
df.head()
```

[31]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	-0.530377	1	0	A/5 21171	-0.502445	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	0.571831	1	0	PC 17599	0.786845	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	-0.254825	0	0	STON/O2. 3101282	-0.488854	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	0.365167	1	0	113803	0.420730	C123	S
4	5	0	3	Allen, Mr. William Henry	male	0.365167	0	0	373450	-0.486337	NaN	S

[37]:
```
corr_matrix = df.corr()
#shows the pairwise correlation between all numerical features in the df  with values between -1 and 1.
print("correlation matrix: ")
print(corr_matrix['Survived'].sort_values(ascending = False)) #Selects the correlation values between the target variable Survived and all other features
print('\n')

from sklearn.ensemble import RandomForestClassifier # model
model = RandomForestClassifier()
X = df.drop('Survived', axis = 1)
y = df['Survived']
model.fit(X, y)
feature_importances = pd.Series(model.feature_importances_, index=X.columns)
print('Feature Importances:')
print(feature_importances.sort_values(ascending=False))
```

```
correlation matrix:
Survived       1.000000
Fare           0.257307
Parch          0.081629
PassengerId   -0.005007
SibSp         -0.035322
Name          -0.057343
Age           -0.077221
Embarked      -0.163517
Ticket        -0.164549
Cabin         -0.254888
Pclass        -0.338481
Sex           -0.543351
Name: Survived, dtype: float64


Feature Importances:
Sex            0.229985
Ticket         0.133677
```