

Titanic Data Analysis - Insights & Summary

Brief Summary

This notebook analyzes the Titanic dataset to understand the factors influencing passenger survival. The work involves data cleaning, exploratory data analysis (EDA), feature engineering, model building using machine learning classifiers, and performance evaluation.

Code executed related to:

```
import kagglehub  
  
brendan45774_test_file_path = kagglehub.dataset_download('brendan45774/test-file')  
  
print('Data source import complete.')
```

Code executed related to:

```
import pandas as pd  
  
import numpy as np  
  
import matplotlib.pyplot as plt  
  
import seaborn as sns
```

Code executed related to:

```
df=pd.read_csv("tested.csv")
```

Code executed related to:

```
df1=df.copy()
```

Code executed related to:

```
df.head()
```

Code executed related to:

```
df.tail()
```

Code executed related to:

```
df.dtypes
```

Code executed related to:

```
df.info()
```

Titanic Data Analysis - Insights & Summary

Code executed related to:

```
(df.isna().sum()/len(df))*100
```

Code executed related to:

```
df.drop('PassengerId',axis=1,inplace=True)
```

Code executed related to:

```
df.duplicated().sum()
```

Code executed related to:

```
(df.isnull().sum()/len(df))*100
```

Code executed related to:

```
df['Age']=df['Age'].fillna(28.0)
```

Code executed related to:

```
df['Cabin']=df['Cabin'].fillna('unknown')
```

Code executed related to:

```
df.head()
```

Code executed related to:

```
df['Pclass'].value_counts()
```

Code executed related to:

```
df['Embarked'].unique()
```

Code executed related to:

```
df['Embarked']=df['Embarked'].fillna('unknown')
```

Code executed related to:

```
sns.countplot(x='Pclass',data=df)
```

```
plt.show()
```

Code executed related to:

```
sns.countplot(x='Sex',data=df)
```

```
plt.show()
```

Titanic Data Analysis - Insights & Summary

Code executed related to:

```
sns.countplot(x='Sex',data=df,hue='Survived')  
  
plt.show()
```

Code executed related to:

```
sns.countplot(x='Pclass',data=df,hue='Survived')  
  
plt.show()
```

Code executed related to:

```
df['Age']=df1['Age']
```

Code executed related to:

```
df["Age_group"] = pd.cut(x=df['Age'], bins=[0,10,20,30,40,50,60,70,80],  
labels=["0_10","10_20","20_30","30_40",'40_50','50_60','60_70','70_80'])
```

Code executed related to:

```
df['Age']
```

Code executed related to:

```
df['Age_group']
```

Code executed related to:

```
sns.countplot(x='Age_group',data=df,hue='Survived')  
  
plt.show()
```

Code executed related to:

```
df['Age'].isna().sum()
```

Code executed related to:

```
df['Age'].fillna(df['Age'].mean(),inplace=True)
```

Code executed related to:

```
df['Age'].isna().sum()
```

Code executed related to:

```
df['Age']=df['Age'].astype(int)
```

Titanic Data Analysis - Insights & Summary

Code executed related to:

```
df['Age'].max()
```

Code executed related to:

```
df['Age'].min()
```

Code executed related to:

```
df.dtypes
```

Code executed related to:

```
sns.countplot(x='Embarked',data=df,hue='Survived')  
  
plt.show()
```

Code executed related to:

```
sns.countplot(x='SibSp',data=df,hue='Sex')  
  
plt.show()
```

Code executed related to:

```
df[df['Age'].isnull()]
```

Code executed related to:

```
sns.boxplot(x='Pclass',y='Age',data=df1)  
  
plt.show()
```

Code executed related to:

```
def age_fillna(df):  
    for i in range(0,len(df)):  
        if(pd.isnull(df.Age[i])):  
            if(df['Pclass']==1):  
                df.Age[i]=38  
            elif (df['Pclass']==2):  
                df.Age[i]=28  
            elif (df['Pclass']==3):  
                df.Age[i]=24  
            else:
```

Titanic Data Analysis - Insights & Summary

...

Code executed related to:

```
age_fillna(df).head(10)
```

Code executed related to: