Home | EDA

localhost:8888/notebooks/Desktop/newpython/EDA.ipynb

jupyter EDA Last Checkpoint: 58 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
train = pd.read_csv('C:/Users/sagar/Downloads/titanic/train.csv')
```

[3]:
```
train.head()
```

[3]:

	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

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
```

27°C Mostly clear

Search

ENG IN

21:41 07-06-2025

Home | EDA

localhost:8888/notebooks/Desktop/newpython/EDA.ipynb

jupyter EDA Last Checkpoint: 58 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code | JupyterLab | Python 3 (ipykernel)

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

[5]:
```
train.value_counts()
```

[5]:

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599	71.2833	C85	C	1
572	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0	2	0	11769	51.4792	C101	S	1
578	1	1	Silvey, Mrs. William Baird (Alice Munger)	female	39.0	1	0	13507	55.9000	E44	S	1
582	1	1	Thayer, Mrs. John Borland (Marian Longstreth Morris)	female	39.0	1	1	17421	110.8833	C68	C	1
584	0	1	Ross, Mr. John Hugo	male	36.0	0	0	13049	40.1250	A10	C	1
												..
328	1	2	Ball, Mrs. (Ada E Hall)	female	36.0	0	0	28551	13.0000	D	S	1
330	1	1	Hippach, Miss. Jean Gertrude	female	16.0	0	1	111361	57.9792	B18	C	1
332	0	1	Partner, Mr. Austen	male	45.5	0	0	113043	28.5000	C124	S	1
333	0	1	Graham, Mr. George Edward	male	38.0	0	1	PC 17582	153.4625	C91	S	1
890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C	1

Name: count, Length: 183, dtype: int64

27°C Mostly clear | Search | ENG IN | 21:42 07-06-2025

[6]:
```
train.tail()
```

[6]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

[7]:
```
train.describe()
```

[7]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

[10]:
```
train.nunique()
```

Home | EDA

localhost:8888/notebooks/Desktop/newpython/EDA.ipynb

jupyter EDA Last Checkpoint: 58 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[10]:
```
train.nunique()
```

[10]:
```
PassengerId    891
Survived         2
Pclass           3
Name           891
Sex              2
Age             88
SibSp            7
Parch            7
Ticket         681
Fare           248
Cabin          147
Embarked         3
dtype: int64
```

[11]:
```
train.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
```

[12]:
```
passenger=train.drop(['Embarked','Parch'],axis=1 )
```

[13]:
```
passenger.head()
```

27°C Mostly clear

Search

ENG IN

21:42 07-06-2025

Home | EDA

localhost:8888/notebooks/Desktop/newpython/EDA.ipynb

jupyter EDA Last Checkpoint: 58 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code | JupyterLab | Python 3 (ipykernel)

[12]:
```
passenger=train.drop(['Embarked','Parch'],axis=1 )
```

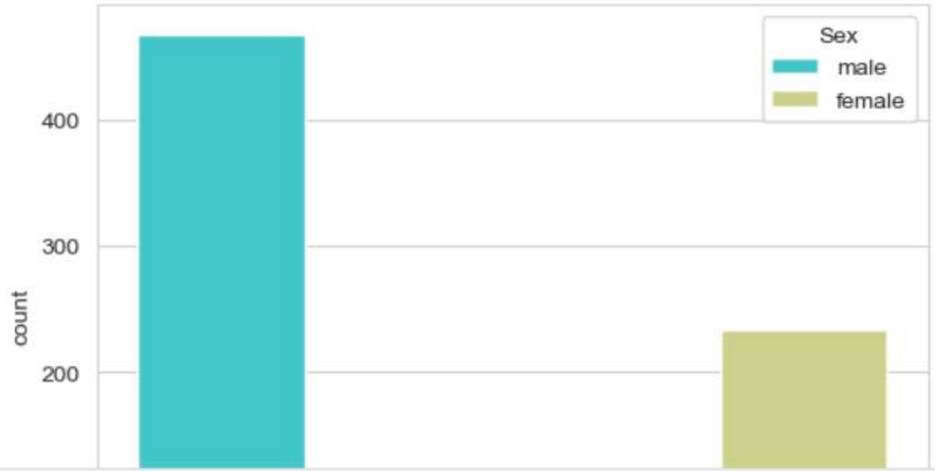
[13]:
```
passenger.head()
```

[13]:

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

[27]:
```
sns.set_style('whitegrid')
sns.countplot(x="Survived",hue='Sex',data=train,palette='rainbow')
```

[27]:
```
<Axes: xlabel='Survived', ylabel='count'>
```



27°C Mostly clear

Search

ENG IN

21:42 07-06-2025

Home | EDA

localhost:8888/notebooks/Desktop/newpython/EDA.ipynb

jupyter EDA Last Checkpoint: 58 minutes ago

File Edit View Run Kernel Settings Help

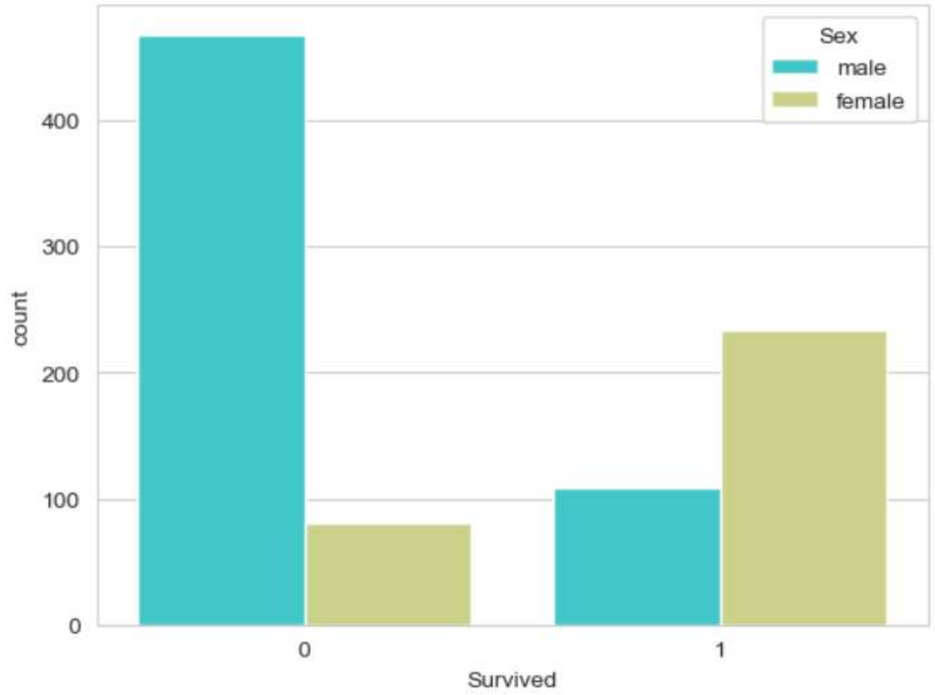
Trusted

Code

JupyterLab Python 3 (ipykernel)

[27]:
```
sns.set_style('whitegrid')
sns.countplot(x="Survived",hue='Sex',data=train,palette='rainbow')
```

[27]: <Axes: xlabel='Survived', ylabel='count'>



[28]:
```
sns.histplot(x='Age', data=train, kde=True, color='g')
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

27°C Mostly clear

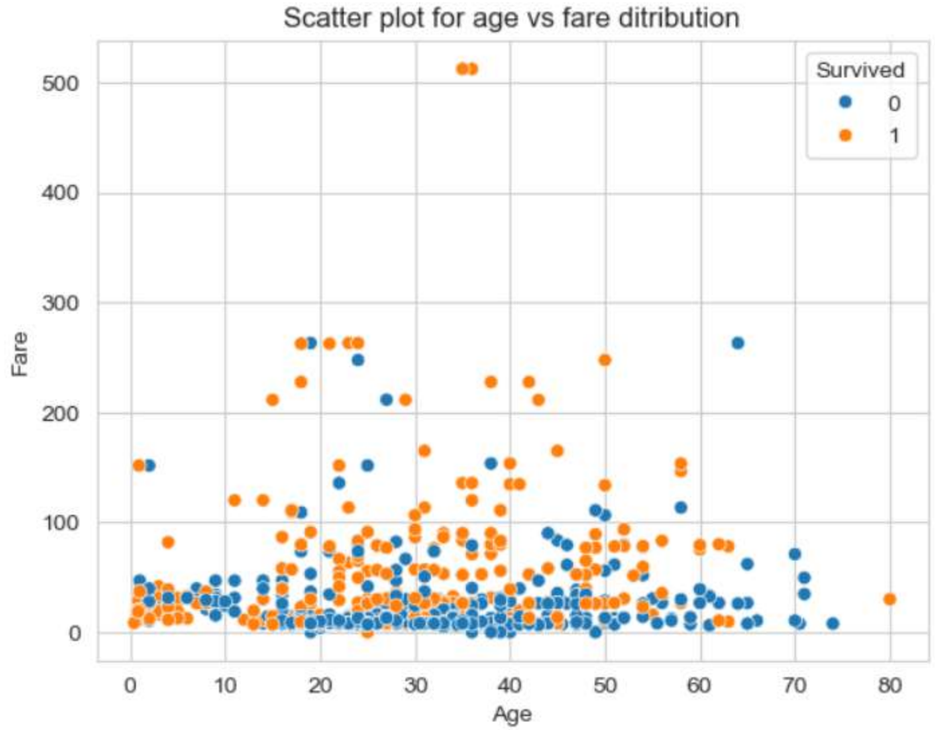
Search

ENG IN

21:42 07-06-2025

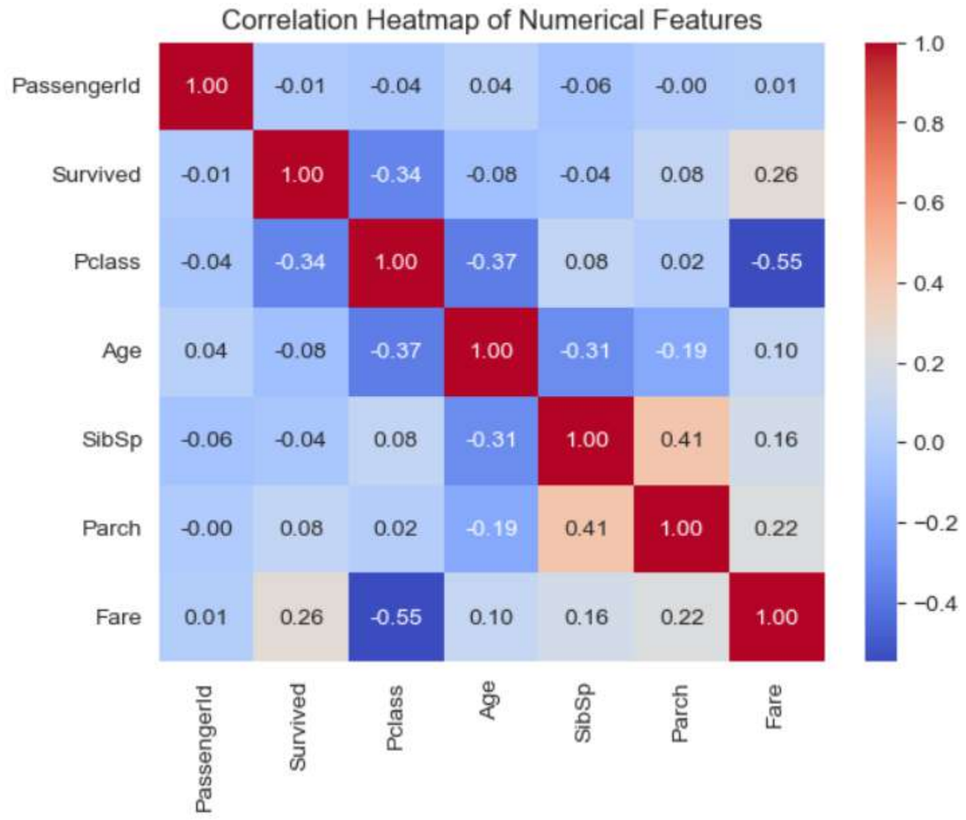
[28]:
```
sns.histplot(x='Age', data=train, kde=True, color='g')
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

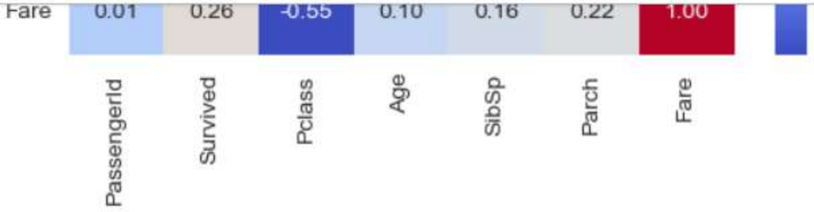
[29]:
```
sns.scatterplot(x='Age', y='Fare', data=train, hue='Survived')
plt.title('Scatter plot for age vs fare ditribution')
plt.xlabel('Age')
plt.ylabel('Fare')
plt.show()
```



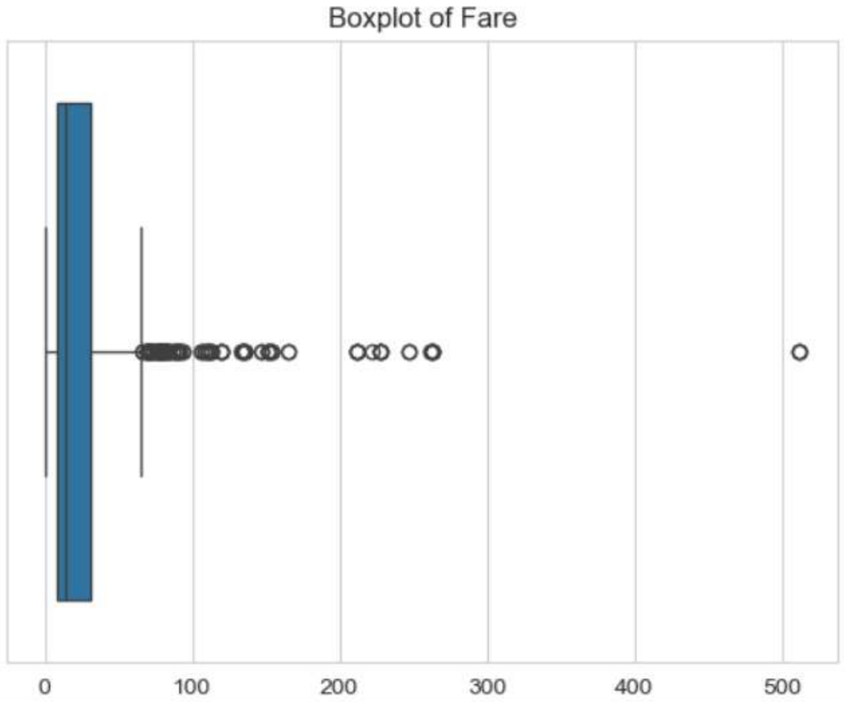
Age

[30]:
```
numerical_columns = train.select_dtypes(include=['int64', 'float64']).columns
sns.heatmap(train[numerical_columns].corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap of Numerical Features')
plt.show()
```

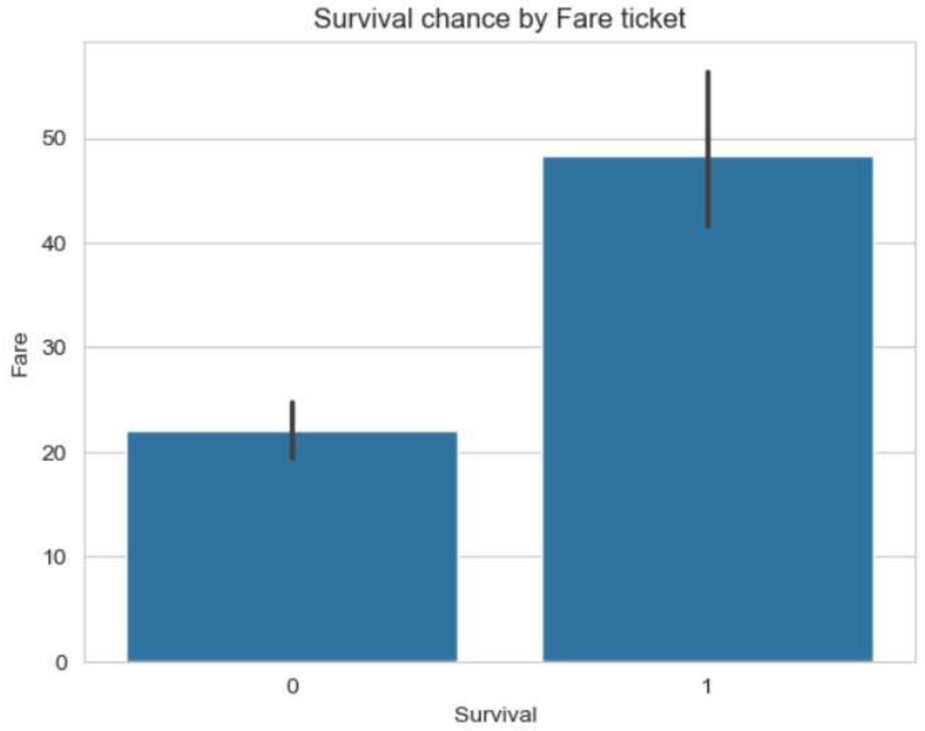




[31]:
```
sns.boxplot(x='Fare', data=train)
plt.title('Boxplot of Fare')
plt.show()
```



[32]:
```
sns.barplot(x='Survived', y='Fare', data=train)
plt.title('Survival chance by Fare ticket')
plt.xlabel('Survival ')
plt.ylabel('Fare')
plt.show()
```



[ ]: