jupyter task5 Last Checkpoint: 2 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[2]:
```
import pandas as pd

df = pd.read_csv(r"train (1).csv")
df
```

[2]:

	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
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

jupyter task5 Last Checkpoint: 2 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[3]:
```
df.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
```

jupyter task5 Last Checkpoint: 2 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[4]:
```
df.describe()
```

[4]:

	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

•[7]:
```
df['Sex'].value_counts()  #If we want to see it in the single row its looks like these
```

[7]:
```
Sex
male      577
female    314
Name: count, dtype: int64
```

•[8]:
```
df.value_counts()   # If we want to see it on whole row
```

jupyter task5 Last Checkpoint: 2 hours ago

File Edit View Run Kernel Settings Help

Code JupyterLab Python 3 (ipykernel)

[8]:
```
df.value_counts()   # If we want to see it on whole row
```

[8]:
```
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
4            1         1       Futrelle, Mrs. Jacques Heath (Lily May Peel)        female  35.0  1      0      113803    53.1000  C123         S
1
7            0         1       McCarthy, Mr. Timothy J                             male    54.0  0      0      17463     51.8625  E46          S
1
11           1         3       Sandstrom, Miss. Marguerite Rut                     female  4.0   1      1      PP 9549   16.7000  G6           S
1
12           1         1       Bonnell, Miss. Elizabeth                            female  58.0  0      0      113783    26.5500  C103         S
1

..
872          1         1       Beckwith, Mrs. Richard Leonard (Sallie Monypeny)    female  47.0  1      1      11751     52.5542  D35          S
1
873          0         1       Carlsson, Mr. Frans Olof                            male    33.0  0      0      695       5.0000   B51 B53 B55  S
1
880          1         1       Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)       female  56.0  0      1      11767     83.1583  C50          C
1
888          1         1       Graham, Miss. Margaret Edith                        female  19.0  0      0      112053    30.0000  B42          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
```

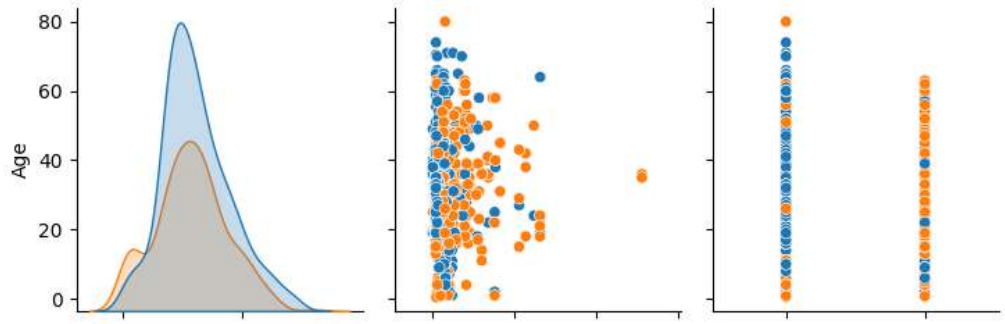
jupyter task5 Last Checkpoint: 2 hours ago

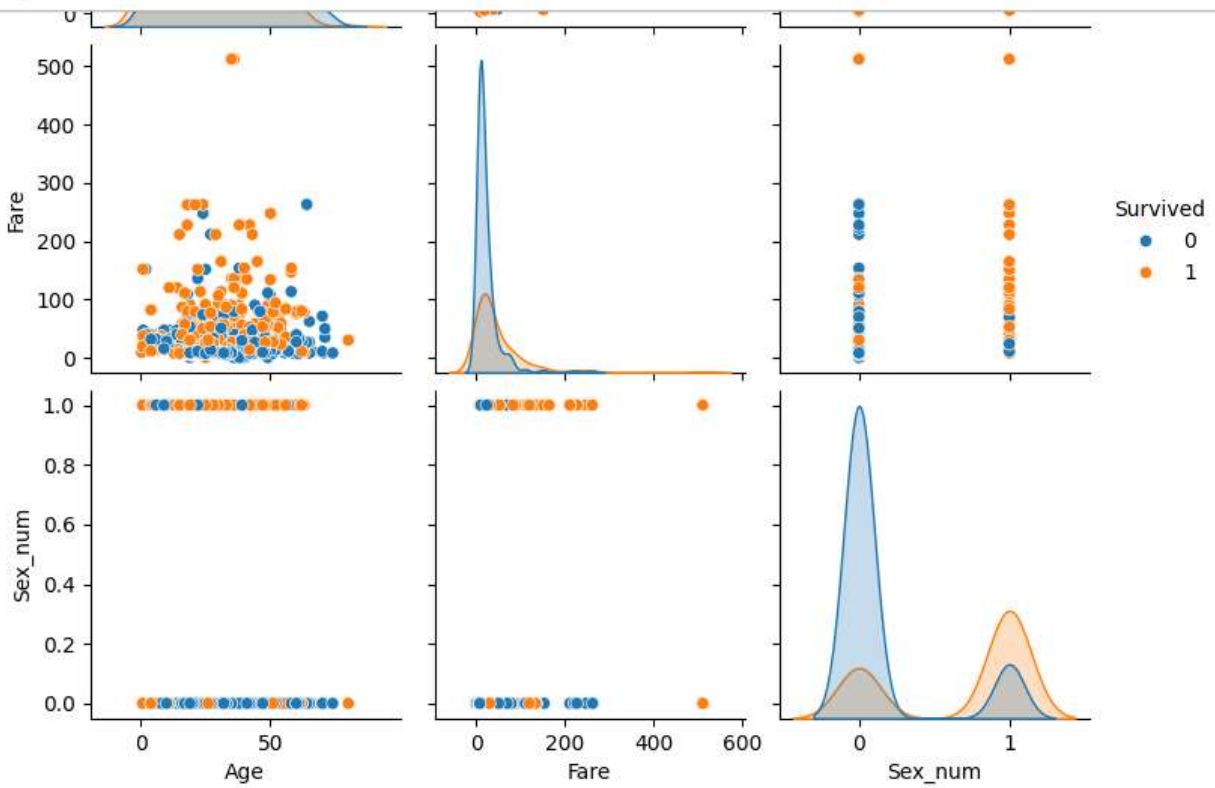
File Edit View Run Kernel Settings Help

Code JupyterLab Python 3 (ipykerne

•[27]:

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

df = pd.read_csv("train (1).csv")

df['Sex_num'] = df['Sex'].map({'male': 0, 'female': 1})

sns.pairplot(df[['Age', 'Fare', 'Sex_num', 'Survived']], hue='Survived')
plt.show()
```




Fare
Sex_num
Age
Fare
Sex_num
Survived
0
1

jupyter task5 Last Checkpoint: 2 hours ago

File Edit View Run Kernel Settings Help

Code JupyterLab Python 3 (ipykernel)

[20]:
```
df.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
```

[29]:

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

df = pd.read_csv("train (1).csv")


df[['Age', 'Fare', 'Survived']].hist(bins=20, figsize=(10, 6), color='skyblue', edgecolor='black')
plt.suptitle("Histograms of Age, Fare, and Survived", fontsize=14)
plt.tight_layout()
plt.show()


#  Histograms
# Plot: Histograms for Age, Fare, and Survived.

# Age: Distribution is right-skewed. Most passengers were between 20 and 40 years old. Some entries are missing.

# Fare: Highly skewed to the right. Most fares were below 100, but a few outliers paid extremely high fares.

# Survived: Binary distribution with slightly fewer survivors than non-survivors.
```

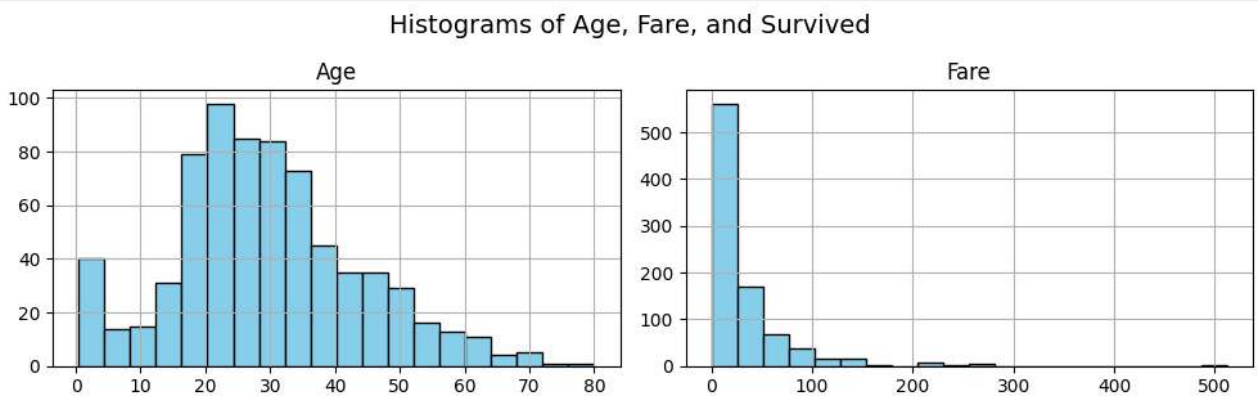
jupyter task5 Last Checkpoint: 2 hours ago

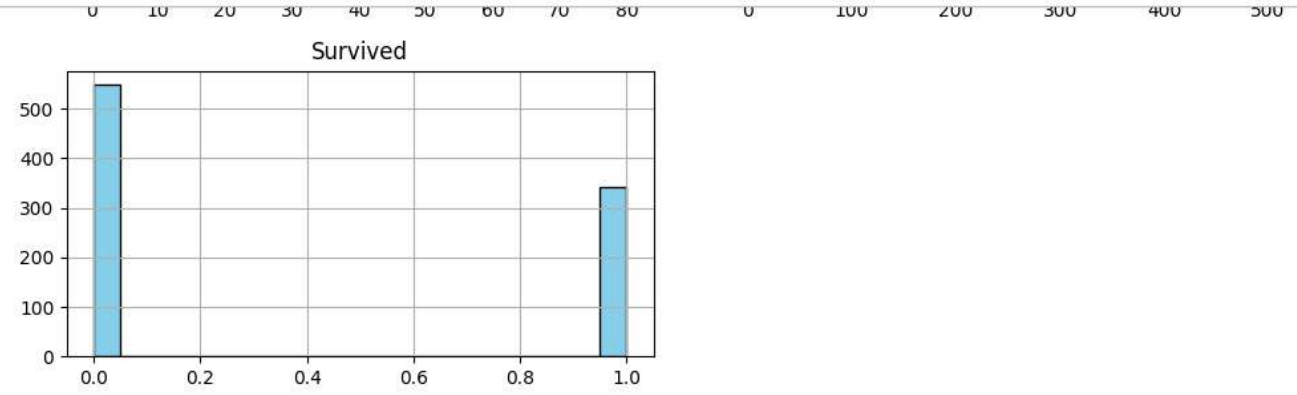
File Edit View Run Kernel Settings Help

Trusted

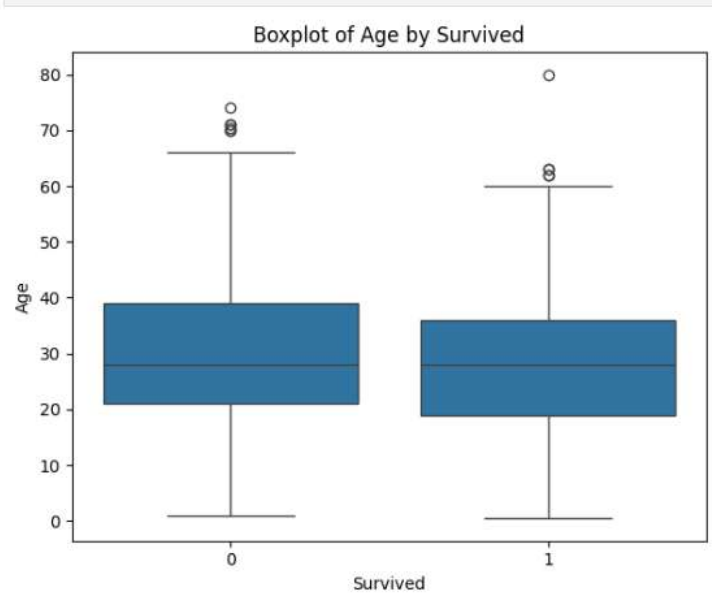
Code

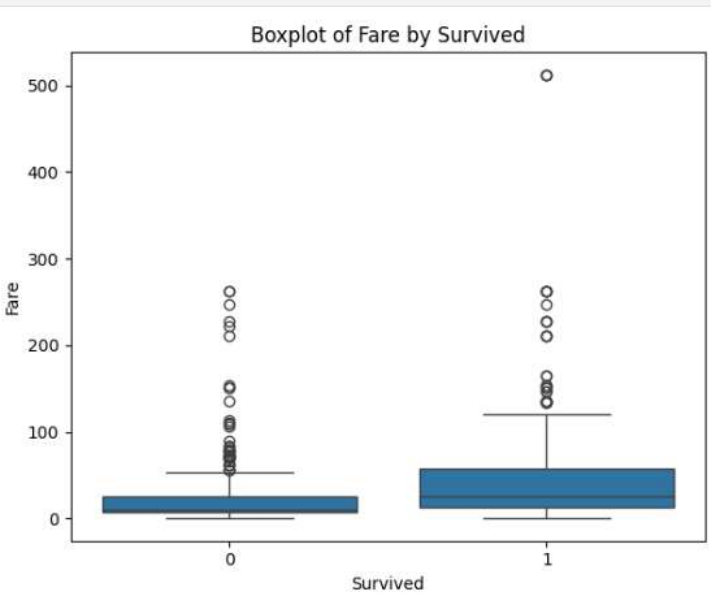
JupyterLab Python 3 (ipykernel)




0 10 20 30 40 50 60 70 80
0 100 200 300 400 500
Survived
500
400
300
200
100
0
0.0 0.2 0.4 0.6 0.8 1.0

[30]:
```
plt.figure(figsize=(12, 5))

plt.subplot(1, 2, 1)
sns.boxplot(x='Survived', y='Age', data=df)
plt.title("Boxplot of Age by Survived")

plt.subplot(1, 2, 2)
sns.boxplot(x='Survived', y='Fare', data=df)
plt.title("Boxplot of Fare by Survived")

plt.tight_layout()
plt.show()



#  Boxplots
# Plot: Boxplots for Age and Fare with respect to Survived.

# Age vs Survived: Median age is slightly lower for survivors, suggesting younger passengers had better survival rates.

# Fare vs Survived: Survivors had significantly higher fares, indicating that wealthier passengers (likely 1st class) were more likely to survive.

# Outliers are present in both plots, especially in the Fare column.
```


Boxplot of Age by Survived
Age
Survived
0
1
0
10
20
30
40
50
60
70
80

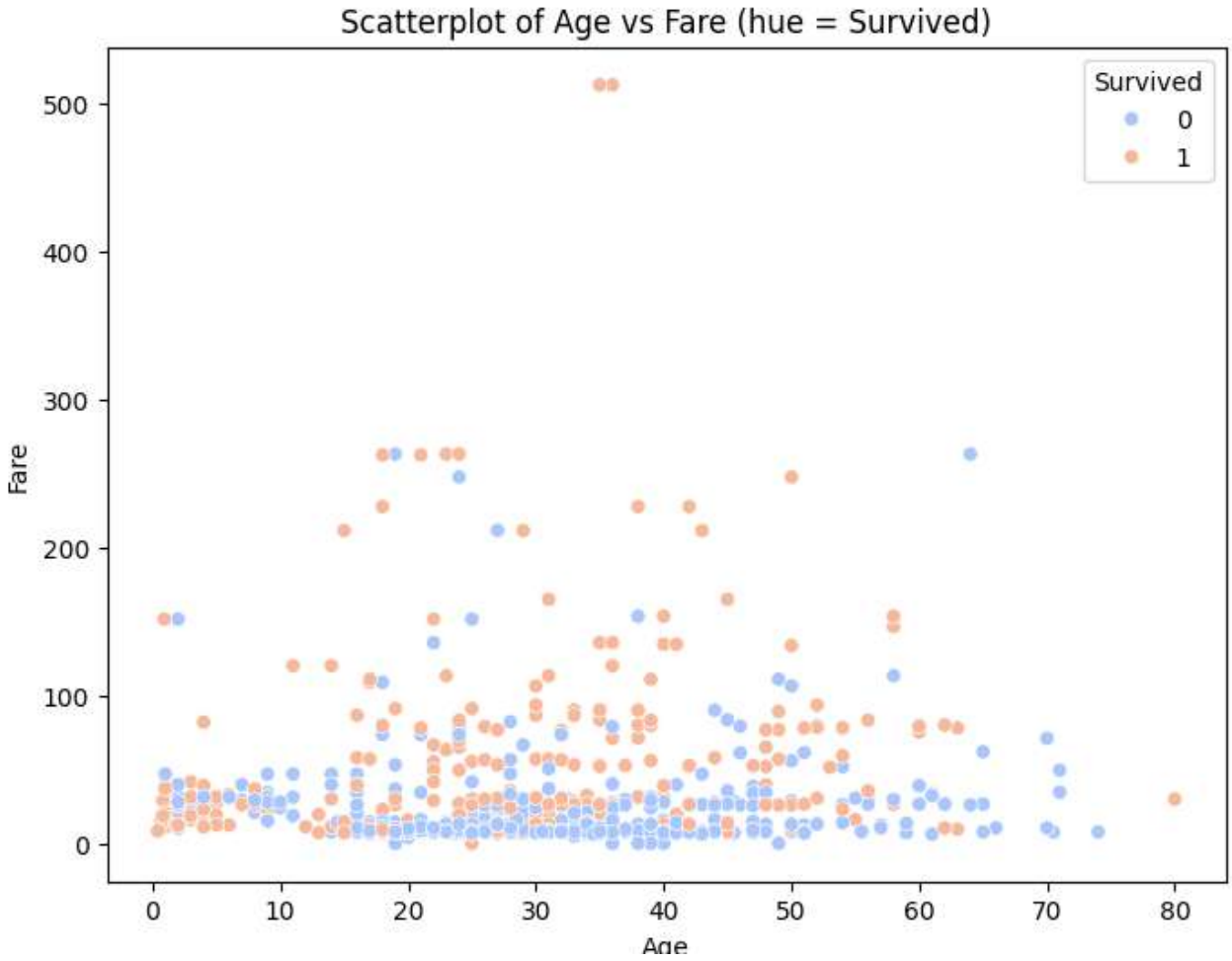

Boxplot of Fare by Survived
Fare
Survived
0
1
0
100
200
300
400
500

jupyter task5 Last Checkpoint: 2 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

*[31]:
```
df['Sex_num'] = df['Sex'].map({'male': 0, 'female': 1})

plt.figure(figsize=(8, 6))
sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df, palette='coolwarm')
plt.title("Scatterplot of Age vs Fare (hue = Survived)")
plt.show()


# Scatterplot
# Plot: Age vs Fare, hue = Survived.

# Most passengers cluster around low fare and age 20-40.

# Survivors are more common in the higher fare range, again pointing to class/wealth as a survival factor.

# No strong linear relationship, but visible clusters of survivors in specific regio
```


Scatterplot of Age vs Fare (hue = Survived)
Survived
0
1
Fare
Age
0
100
200
300
400
500
0
10
20
30
40
50
60
70
80