Import Required Libraries

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

```
df=pd.read_csv("/content/titanic_toy.csv")
```

```
df.drop(columns=['Family'],inplace=True)
```

```
plt.figure(figsize=(14,4))
plt.subplot(121)
sns.distplot(df['Fare'])
plt.show()
```

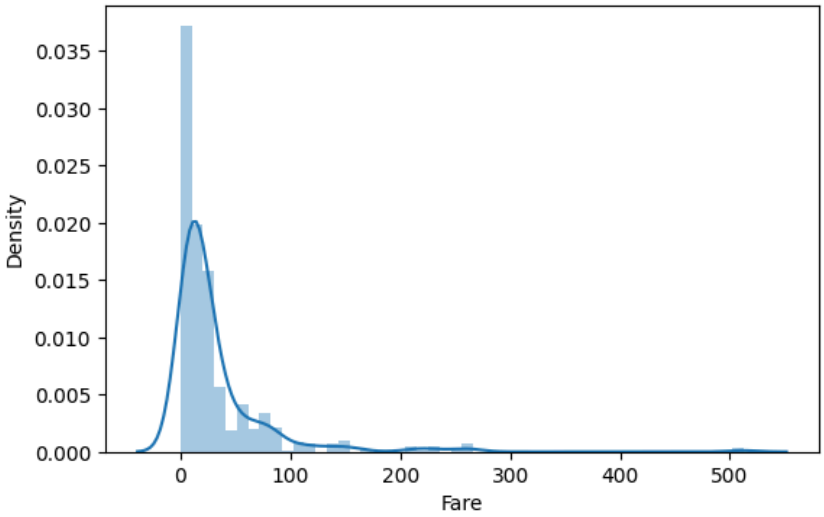
```
/tmp/ipython-input-3114248343.py:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['Fare'])
```



Check Skewness Of graph

```
df['Fare'].skew()
```

```
np.float64(4.819666600523025)
```

```
df['Fare'].describe()
```

	Fare
count	846.000000
mean	32.279338
std	50.305796
min	0.000000
25%	7.895800
50%	14.454200
75%	31.206250
max	512.329200

dtype: float64

📦 Outlier Detection (IQR Method)

```
percentile25=df['Fare'].quantile(0.25)
percentile75=df['Fare'].quantile(0.75)
```

```
iqr=percentile75-percentile25
iqr
```

```
np.float64(23.310449999999996)
```

```
upperlimit=percentile75+1.5*iqr
lowerlimit=percentile25-1.5*iqr
print("upperlimit",upperlimit)
print("lowerlimit",lowerlimit)
```

```
upperlimit 66.17192499999999
lowerlimit -27.06987499999999
```

trimming The outliers

```
df[(df['Fare']>upperlimit)|(df['Fare']<lowerlimit)]
```

	Age	Fare	Survived
1	38.0	71.2833	1
27	19.0	263.0000	0
31	NaN	146.5208	1
34	28.0	82.1708	0
52	49.0	76.7292	1
...	...	...	...
846	NaN	69.5500	0
849	NaN	89.1042	1
856	45.0	164.8667	1
863	NaN	69.5500	0
879	56.0	83.1583	1

110 rows × 3 columns

```
newdf=df[(df['Fare']<upperlimit)]
newdf
```

	Age	Fare	Survived
0	22.0	7.2500	0
2	26.0	7.9250	1
3	35.0	53.1000	1
4	35.0	8.0500	0
5	NaN	8.4583	0
...	...	...	...
885	39.0	29.1250	0
886	27.0	13.0000	0
887	19.0	30.0000	1
888	NaN	23.4500	0
890	32.0	7.7500	0

736 rows × 3 columns

Next steps: Generate code with newdf | New interactive sheet

📌 Outlier Treatment (IQR Capping)

```
new_df=df.copy()
new_df['Fare']=np.where(
    df['Fare']>upperlimit,
    upperlimit,
    np.where(
        df['Fare']<lowerlimit,
        lowerlimit,
        df['Fare']
    )
)
```

📊 Outlier Treatment Visualization (Before vs After IQR)

```
plt.figure(figsize=(14,6))
plt.subplot(121)
sns.boxplot(df['Fare'])
plt.title("Before IQR Outlier Applying")
plt.subplot(122)
sns.boxplot(new_df['Fare'])
plt.title("After IQR Outlier Applying")
plt.show()
```