

In [37]:

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

In [38]:

```
data= pd.read_csv("D:\\IPSITA\\Spaceship.csv")
```

In [39]:

```
df=pd.DataFrame(data)
```

In [40]:

```
categorical_features = df.select_dtypes(include=['object'])
for col in categorical_features:
    print(df[col].value_counts())
```

0001\_011
6136\_011
6141\_011
6139\_061
6139\_051
..
3126\_011
3124\_031
3124\_021
3124\_011
9280\_021
Name: PassengerId, Length: 8693, dtype: int64
Earth4602
Europa2131
Mars1759
Name: HomePlanet, dtype: int64
False5439
True3037
Name: CryoSleep, dtype: int64
G/734/S8
G/109/P7
B/201/P7
G/1368/P7
G/981/S7
..
G/556/P1
E/231/S1
G/545/S1
G/543/S1
F/947/P1
Name: Cabin, Length: 6560, dtype: int64
TRAPPIST-1e5915
55 Cancr i1800
PSO J318.5-22796
Name: Destination, dtype: int64
False8291
True199
Name: VIP, dtype: int64
Gollux Reedall2
Elaney Webstephrey2
Grake Porki2
Sus Coolez2
Apix Wala2
..
Jamelia Griffy1
Hardy Griffy1
Salley Mckinn1
Mall Frasp1
Propsh Hontichre1
Name: Name, Length: 8473, dtype: int64

In [41]:

```
df.info()
```

Out[42]:

```
PassengerId      0
HomePlanet       201
CryoSleep        217
Cabin            199
Destination       182
Age              179
VIP              203
RoomService      181
FoodCourt        183
ShoppingMall     208
Spa              183
VRDeck           188
Name             200
Transported       0
dtype: int64
```

In [43]:

```
categorical_cols = df.select_dtypes(include=['object']).columns
numeric_cols = df.select_dtypes(include=['int64', 'float64']).columns

for col in categorical_cols:
    df[col].fillna(df[col].mode()[0], inplace=True) # Filling categorical columns with mode

for col in numeric_cols:
    df[col].fillna(-1, inplace=True) # Filling numerical columns with
```

In [44]:

```
df.head(31)
```

Out[44]:

	PassengerId	HomePlanet	CryoSleep	Cabin	Destination	Age	VIP	RoomService	FoodCourt	ShoppingMall	Spa	VRDeck	Name	Transported
0	0001_01	Europa	False	B/0/P	TRAPPIST-1e	39.0	False	0.0	0.0	0.0	0.0	0.0	Maham Ofracculy	False
1	0002_01	Earth	False	F/0/S	TRAPPIST-1e	24.0	False	109.0	9.0	25.0	549.0	44.0	Juanna Vines	True
2	0003_01	Europa	False	A/0/S	TRAPPIST-1e	58.0	True	43.0	3576.0	0.0	6715.0	49.0	Altark Susent	False
3	0003_02	Europa	False	A/0/S	TRAPPIST-1e	33.0	False	0.0	1283.0	371.0	3329.0	193.0	Solam Susent	False
4	0004_01	Earth	False	F/1/S	TRAPPIST-1e	16.0	False	303.0	70.0	151.0	565.0	2.0	Willy Santantines	True
5	0005_01	Earth	False	F/0/P	PSO J318.5-22	44.0	False	0.0	483.0	0.0	291.0	0.0	Sandie Hineththews	True
6	0006_01	Earth	False	F/2/S	TRAPPIST-1e	26.0	False	42.0	1539.0	3.0	0.0	0.0	Billex Jacostaffey	True
7	0006_02	Earth	True	G/0/S	TRAPPIST-1e	28.0	False	0.0	0.0	0.0	0.0	-1.0	Candra Jacostaffey	True
8	0007_01	Earth	False	F/3/S	TRAPPIST-1e	35.0	False	0.0	785.0	17.0	216.0	0.0	Andona Beston	True
9	0008_01	Europa	True	B/1/P	55 Cancr i e	44.0	False	0.0	0.0	0.0	0.0	0.0	Erraiam Flatic	True
10	0008_02	Europa	True	B/1/P	TRAPPIST-1e	34.0	False	0.0	0.0	-1.0	0.0	0.0	Altadr Flatic	True
11	0008_03	Europa	False	B/1/P	55 Cancr i e	45.0	False	39.0	7295.0	589.0	110.0	124.0	Wezena Flatic	True
12	0009_01	Mars	False	F/1/P	TRAPPIST-1e	32.0	False	73.0	0.0	1123.0	0.0	113.0	Berers Barne	True
13	0010_01	Earth	False	G/1/S	TRAPPIST-1e	48.0	False	719.0	1.0	65.0	0.0	24.0	Reney Baketton	False
14	0011_01	Earth	False	F/2/P	TRAPPIST-1e	28.0	False	8.0	974.0	12.0	2.0	7.0	Elle Bertsontry	True
15	0012_01	Earth	False	G/734/S	TRAPPIST-1e	31.0	False	32.0	0.0	876.0	0.0	0.0	Justie Pooles	False
16	0014_01	Mars	False	F/3/P	55 Cancr i e	27.0	False	1286.0	122.0	-1.0	0.0	0.0	Flats Eccle	False
17	0015_01	Earth	False	F/4/P	55 Cancr i e	24.0	False	0.0	1.0	0.0	0.0	637.0	Carry Hughriend	False
18	0016_01	Mars	True	F/5/P	TRAPPIST-1e	45.0	False	0.0	0.0	0.0	0.0	0.0	Alus Upead	True
19	0017_01	Earth	False	G/0/P	TRAPPIST-1e	0.0	False	0.0	0.0	0.0	0.0	0.0	Lyde Brighttt	True
20	0017_02	Earth	False	F/6/P	55 Cancr i e	14.0	False	412.0	0.0	1.0	0.0	679.0	Philda Brighttt	False
21	0020_01	Earth	True	E/0/S	TRAPPIST-1e	1.0	False	0.0	0.0	0.0	0.0	0.0	Almary Brantuauez	False
22	0020_02	Earth	True	E/0/S	55 Cancr i e	49.0	False	0.0	0.0	0.0	0.0	0.0	Glendy Brantuauez	False
23	0020_03	Earth	True	E/0/S	55 Cancr i e	29.0	False	0.0	0.0	-1.0	0.0	0.0	Mollen Mcfaddenmon	False
24	0020_04	Earth	False	E/0/S	TRAPPIST-1e	10.0	False	0.0	0.0	0.0	0.0	0.0	Breney Jacostanley	True
25	0020_05	Earth	True	E/0/S	PSO J318.5-22	1.0	False	-1.0	0.0	0.0	0.0	0.0	Mael Brantuauez	False
26	0020_06	Earth	False	E/0/S	TRAPPIST-1e	7.0	False	0.0	0.0	0.0	0.0	0.0	Terta Mcfaddenmon	False
27	0022_01	Mars	False	D/0/P	TRAPPIST-1e	21.0	False	980.0	2.0	69.0	0.0	0.0	Alraium Disivering	False
28	0024_01	Europa	True	C/2/S	TRAPPIST-1e	62.0	False	0.0	0.0	-1.0	0.0	0.0	Penton Fulkided	True
29	0025_01	Earth	False	F/6/S	TRAPPIST-1e	15.0	False	0.0	225.0	0.0	998.0	0.0	Karard Brookenson	False
30	0026_01	Europa	False	C/0/P	55 Cancr i e	34.0	False	22.0	6073.0	0.0	1438.0	328.0	Anyoni Unconary	False

In [45]:

```
basic_statistics = df.describe()
```

In [46]:

```
basic_statistics
```

Out[46]:

	Age	RoomService	FoodCourt	ShoppingMall	Spa	VRDeck
count	8693.000000	8693.000000	8693.000000	8693.000000	8693.000000	8693.000000
mean	28.213735	219.988496	448.412976	169.548372	304.567813	298.240193
std	14.951705	660.526002	1595.796550	598.013969	1125.568266	1134.132115
min	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000	-1.000000
25%	19.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	27.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	37.000000	41.000000	61.000000	22.000000	53.000000	40.000000
max	79.000000	14327.000000	29813.000000	23492.000000	22408.000000	24133.000000

In [48]:

```
import warnings
warnings.filterwarnings("ignore")
sns.pairplot(df)
plt.show()
```

In [49]:

```
df.to_csv('cleaned_spaceship_data.csv', index=False)
```

In [ ]: