

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

```
df = pd.read_csv('/content/breast_cancer_survival_2.csv')
df.head()
```

	Age	Gender	Protein1	Protein2	Protein3	Protein4	Tumour_Stage	Histology	st
0	42	FEMALE	0.95256	2.15000	0.007972	-0.048340	II	Infiltrating Ductal Carcinoma	Pos
1	54	FEMALE	0.00000	1.38020	-0.498030	-0.507320	II	Infiltrating Ductal Carcinoma	Pos
2	63	FEMALE	-0.52303	1.76400	-0.370190	0.010815	II	Infiltrating Ductal Carcinoma	Pos
3	78	FEMALE	-0.87618	0.12943	-0.370380	0.132190	I	Infiltrating Ductal Carcinoma	Pos
4	42	FEMALE	0.22611	1.74910	-0.543970	-0.390210	II	Infiltrating Ductal Carcinoma	Pos

Next steps:

Generate code with df

 View recommended plots

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 334 entries, 0 to 333
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Age                    334 non-null    int64
1   Gender                 334 non-null    object
2   Protein1               334 non-null    float64
3   Protein2               334 non-null    float64
4   Protein3               334 non-null    float64
5   Protein4               334 non-null    float64
6   Tumour_Stage           334 non-null    object
7   Histology              334 non-null    object
8   ER status              334 non-null    object
9   PR status              334 non-null    object
10  HER2 status            334 non-null    object
11  Surgery_type           334 non-null    object
12  Date_of_Surgery        334 non-null    object
13  Date_of_Last_Visit     317 non-null    object
14  Patient_Status         321 non-null    object
dtypes: float64(4), int64(1), object(10)
memory usage: 39.3+ KB
```

```
df.isna().sum()

Age                0
Gender              0
Protein1           0
Protein2           0
Protein3           0
Protein4           0
Tumour_Stage       0
Histology           0
ER status          0
PR status          0
HER2 status        0
Surgery_type       0
Date_of_Surgery    0
Date_of_Last_Visit 17
Patient_Status     13
dtype: int64
```

```
df.fillna('no-null', inplace=True)
```

```
df.isna().sum()
```

```

Age                0
Gender             0
Protein1           0
Protein2           0
Protein3           0
Protein4           0
Tumour_Stage       0
Histology          0
ER status          0
PR status          0
HER2 status        0
Surgery_type       0
Date_of_Surgery    0
Date_of_Last_Visit 0
Patient_Status     0
dtype: int64

```

```

X = df.drop('Patient_Status', axis=1)
y = df['Patient_Status']
print(y)

```

```

0      Alive
1      Dead
2      Alive
3      Alive
4      Alive
...
329    Alive
330    Alive
331    Dead
332    Alive
333    Dead
Name: Patient_Status, Length: 334, dtype: object

```

```

from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score
label_encoders = {}
for column in X.select_dtypes(include='object').columns:
    label_encoders[column] = LabelEncoder()
    X[column] = label_encoders[column].fit_transform(X[column])
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
svm = SVC()
svm.fit(X_train, y_train)
y_pred = svm.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")

```

```
Accuracy: 0.7761194029850746
```

```

import matplotlib.pyplot as plt
labels = ['Accuracy']
values = [accuracy]
plt.figure(figsize=(5, 5))
plt.bar(labels, values, color='skyblue')
plt.ylabel('Accuracy')
plt.title('Model Accuracy')
plt.ylim(0,2)
plt.show()

```

