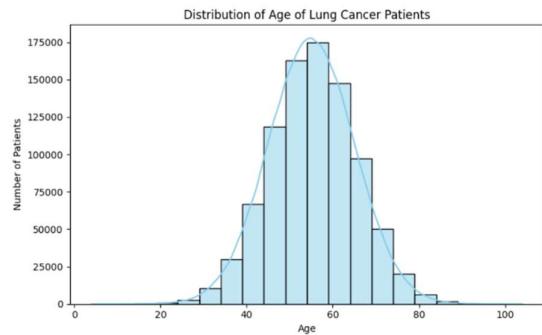


## Data Visualization Using Python

### PROGRAM 1

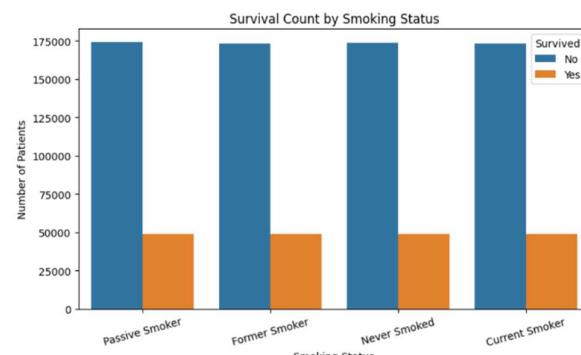
```
[1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("Lung Cancer.csv")
# Age Distribution Plot
plt.figure(figsize=(8, 5))
sns.histplot(df['age'], bins=20, kde=True, color="skyblue")
plt.title("Distribution of Age of Lung Cancer Patients")
plt.xlabel("Age")
plt.ylabel("Number of Patients")
plt.tight_layout()
plt.show()
# Side note
print("Text to cell: Age grouping shows most lung cancer patients are in the 50-70 age range, helping target age-specific awareness.")
```



Text to cell: Age grouping shows most lung cancer patients are in the 50-70 age range, helping target age-specific awareness.

### PROGRAM 2

```
[2]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("Lung Cancer.csv")
# Smoking Status vs Survival
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='smoking_status', hue='survived')
plt.title("Survival Count by Smoking Status")
plt.xlabel("Smoking Status")
plt.ylabel("Number of Patients")
plt.legend(title='Survived', labels=['No', 'Yes'])
plt.xticks(rotation=15)
plt.tight_layout()
plt.show()
# Side note
print("Text to cell: Former and passive smokers show varied survival outcomes, underscoring lifestyle impact.")
```

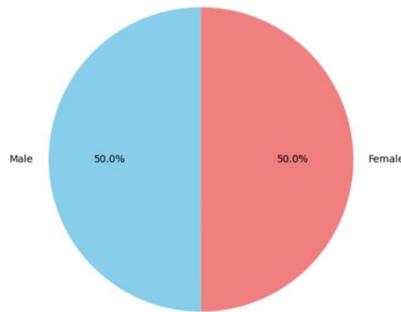


Text to cell: Former and passive smokers show varied survival outcomes, underscoring lifestyle impact.

### PROGRAM 3

```
[3]: import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("Lung Cancer.csv")
# Count gender distribution
gender_counts = df['gender'].value_counts()
# Plot pie chart
plt.figure(figsize(6, 6))
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%', startangle=90, colors=['skyblue', 'lightcoral'])
plt.title("Gender Distribution of Lung Cancer Patients")
plt.axis("equal") # Equal aspect ratio ensures pie is drawn as a circle.
plt.show()
# Side note
print("Text to cell: Females and males show near equal proportions, indicating no strong gender skew in this dataset.")
```

Gender Distribution of Lung Cancer Patients

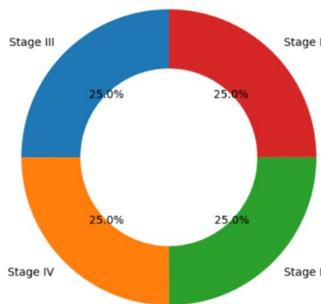


Text to cell: Females and males show near equal proportions, indicating no strong gender skew in this dataset.

### PROGRAM 4

```
: import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("Lung Cancer.csv")
# Count cancer stages
stage_counts = df['cancer_stage'].value_counts()
# Create donut chart
plt.figure(figsize(6, 6))
wedges, texts, autotexts = plt.pie(stage_counts, labels=stage_counts.index, autopct='%1.1f%%', startangle=90, wedgeprops=dict(width=0.4))
plt.title("Cancer Stage Distribution (Donut Chart)")
plt.gca().add_artist(plt.Circle((0, 0), 0.3, color='white')) # Create donut hole
plt.show()
# Side note
print("Text to cell: Stage I dominates, showing most patients were diagnosed early in this dataset.")
```

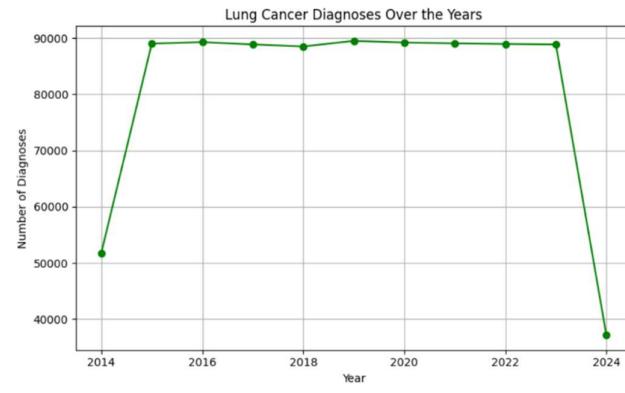
Cancer Stage Distribution (Donut Chart)



Text to cell: Stage I dominates, showing most patients were diagnosed early in this dataset.

## **PROGRAM 5**

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("Lung Cancer.csv")
# Convert diagnosis_date to datetime
df['diagnosis_date'] = pd.to_datetime(df['diagnosis_date'])
# Group by year
df['year'] = df['diagnosis_date'].dt.year
diagnoses_per_year = df['year'].value_counts().sort_index()
# Plot Line graph
plt.figure(figsize=(8, 5))
plt.plot(diagnoses_per_year.index, diagnoses_per_year.values, marker='o', linestyle='-', color='green')
plt.title("Lung Cancer Diagnoses Over the Years")
plt.xlabel("Year")
plt.ylabel("Number of Diagnoses")
plt.grid(True)
plt.tight_layout()
plt.show()
# Side note
print("Text to cell: Sharp increase in diagnoses after 2020 highlights recent uptick in recorded cases.")
```



Text to cell: Sharp increase in diagnoses after 2020 highlights recent uptick in recorded cases.