

```
# STEP 1: Import required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# STEP 2: Load the dataset
df = pd.read_excel("Student Academic Performance.xlsx.xlsx")
df.head()
```

	Name	Gender	Phone Number	State of Birth	Email	Student ID	Qualification	Course	Semester
0	Lum Meu Mel	F	010-981 7433	Penang	meumel32@yahoo.com	21PMD30001	Degree	Hotel Management	S
1	Eric Hew Law Toh	M	019-752 8050	Kedah	law03toh@hotmail.com	21PMD30002	Diploma	Accounting	F
2	Hao Lin Min	F	019-455 0808	Perak	hoa32@crooks.org	21PMD30003	Degree	Information Technology	S
3	Cheng Zheng Xun	M	016-375 8448	Penang	zheng1xun@gmail.com	21PMD30004	Diploma	Computer Science	F
4	Eunice Cheong En Hui	F	014-550 5401	Kedah	enhui12@gmail.com	21PMD30005	Degree	Business	F

Next steps: [Generate code with df](#) [New interactive sheet](#)

```
# STEP 3: Basic data inspection
```

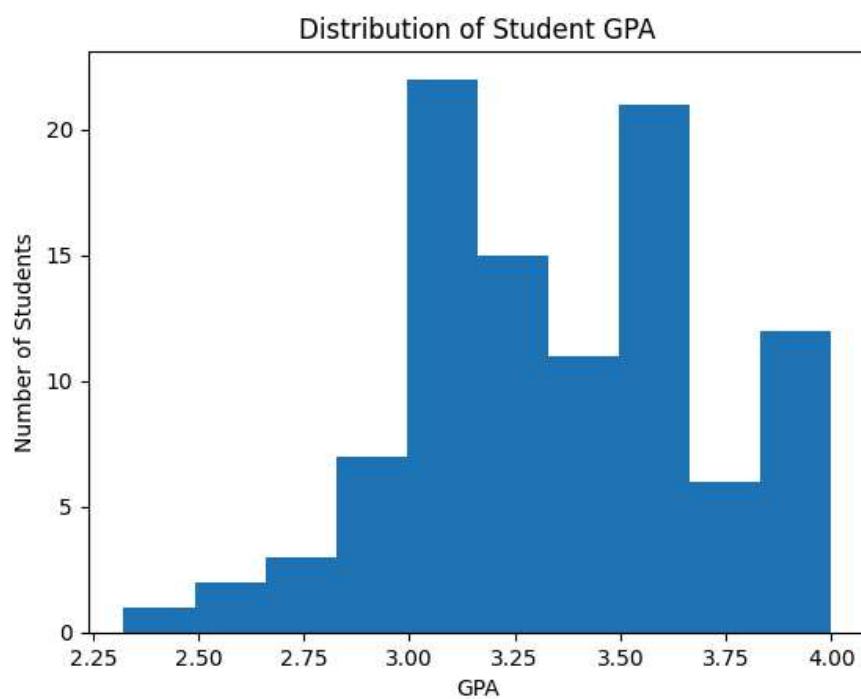
```
df.shape
df.columns
df.isnull().sum()
```

	0
<b>Name</b>	0
<b>Gender</b>	0
<b>Phone Number</b>	0
<b>State Of Birth</b>	0
<b>Email</b>	0

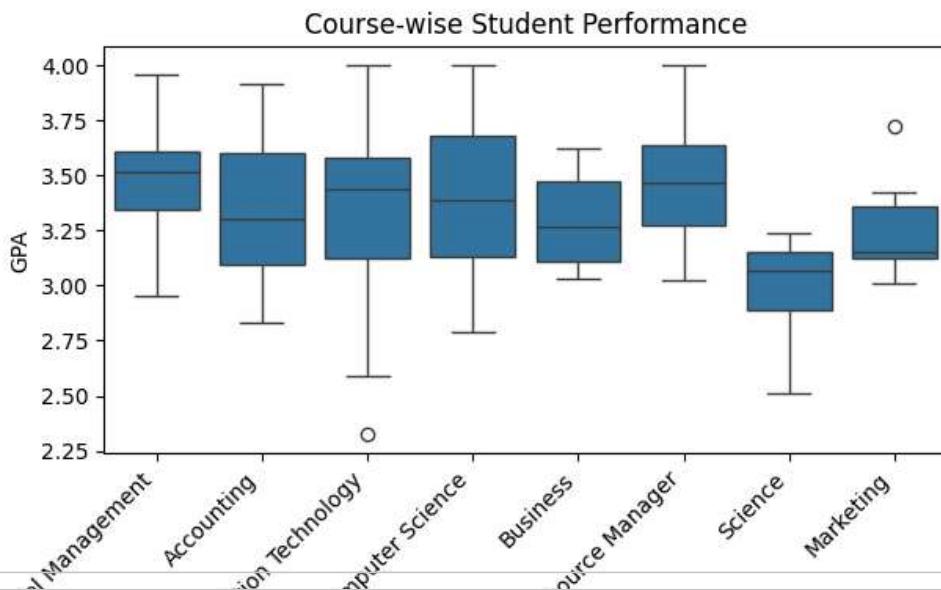
```
# STEP 4: Handle missing values  
df.fillna(0, inplace=True)
```

```
# VISUALIZATION SECTION  
# 1. Histogram - Distribution of Marks
```

```
plt.figure()  
plt.hist(df['GPA'], bins=10)  
plt.xlabel("GPA")  
plt.ylabel("Number of Students")  
plt.title("Distribution of Student GPA")  
plt.show()
```

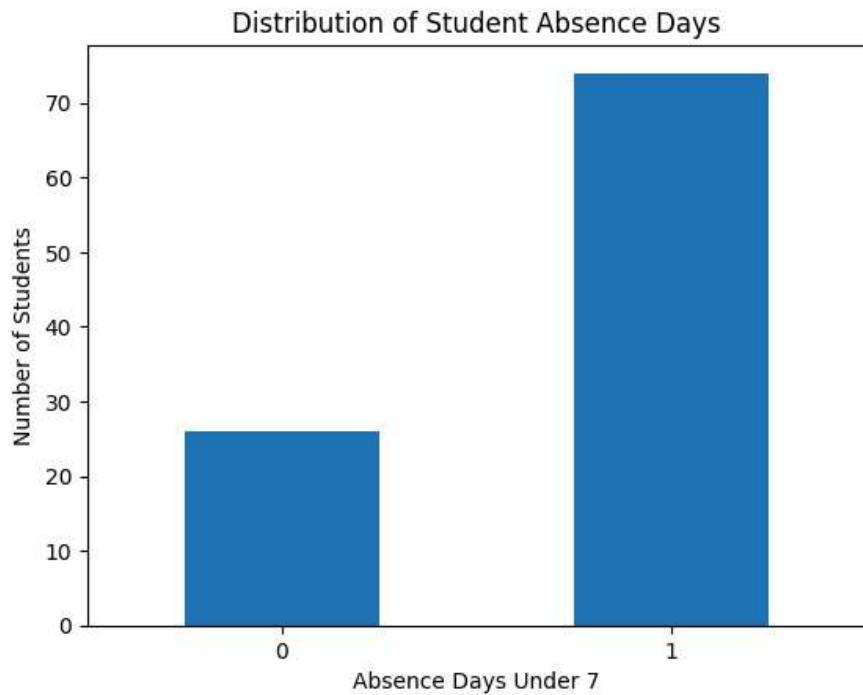


```
# 2. Box Plot - Subject-wise Performance  
plt.figure()  
sns.boxplot(x='Course', y='GPA', data=df)  
plt.xlabel("Course")  
plt.ylabel("GPA")  
plt.title("Course-wise Student Performance")  
plt.xticks(rotation=45, ha='right')  
plt.tight_layout()  
plt.show()
```



```
# 3. Bar Plot - Attendance Distribution
attendance_count = df['Absence Days Under-7'].value_counts().sort_index()

plt.figure()
attendance_count.plot(kind='bar')
plt.xlabel("Absence Days Under 7")
plt.ylabel("Number of Students")
plt.title("Distribution of Student Absence Days")
plt.xticks(rotation=0)
plt.show()
```



```
# 4. Attendance vs Performance
plt.figure()
sns.barplot(x='Absence Days Under-7', y='GPA', data=df)
plt.xlabel("Absence Days Under 7")
plt.ylabel("Average GPA")
plt.title("Attendance vs Student Performance")
plt.xticks(rotation=0)
```

```
plt.tight_layout()  
plt.show()
```

