Data Collection & Presentation

1. Data Collection Methods

There are three main ways to collect data:

a. Survey: Asking people (e.g., online forms, interviews)

b. Observation: Watching and recording behavior (e.g., traffic patterns)

c. Experiment: Performing controlled tests (e.g., drug trials)

Example: You survey 10 students to ask their age and favorite subject.

---

2. Population vs Sample

Term	Meaning
Population	Entire group you want to study
Sample	Subset of population used for analysis

Example: If your college has 1000 students (population), and you survey 100 students (sample), you're using a sample to represent the population.

---

3. Sampling Techniques

Method	Description
Random Sampling	Every item has an equal chance
Stratified	Dividing into sub-groups (e.g., by gender)
Systematic	Every 5th or 10th item
Cluster	Dividing into groups and choosing whole group(s)

4. Frequency Distribution

A frequency distribution shows how many times each value occurs.

Age	Frequency
18	2
19	3
20	1

5. Data Visualization Techniques

These are used to present data clearly.

Chart Type	Use For
Bar Chart	Comparing categories
Pie Chart	Showing parts of a whole
Histogram	Frequency of numerical data
Line Graph	Trends over time
Box Plot	Summary of distribution (median, IQR)

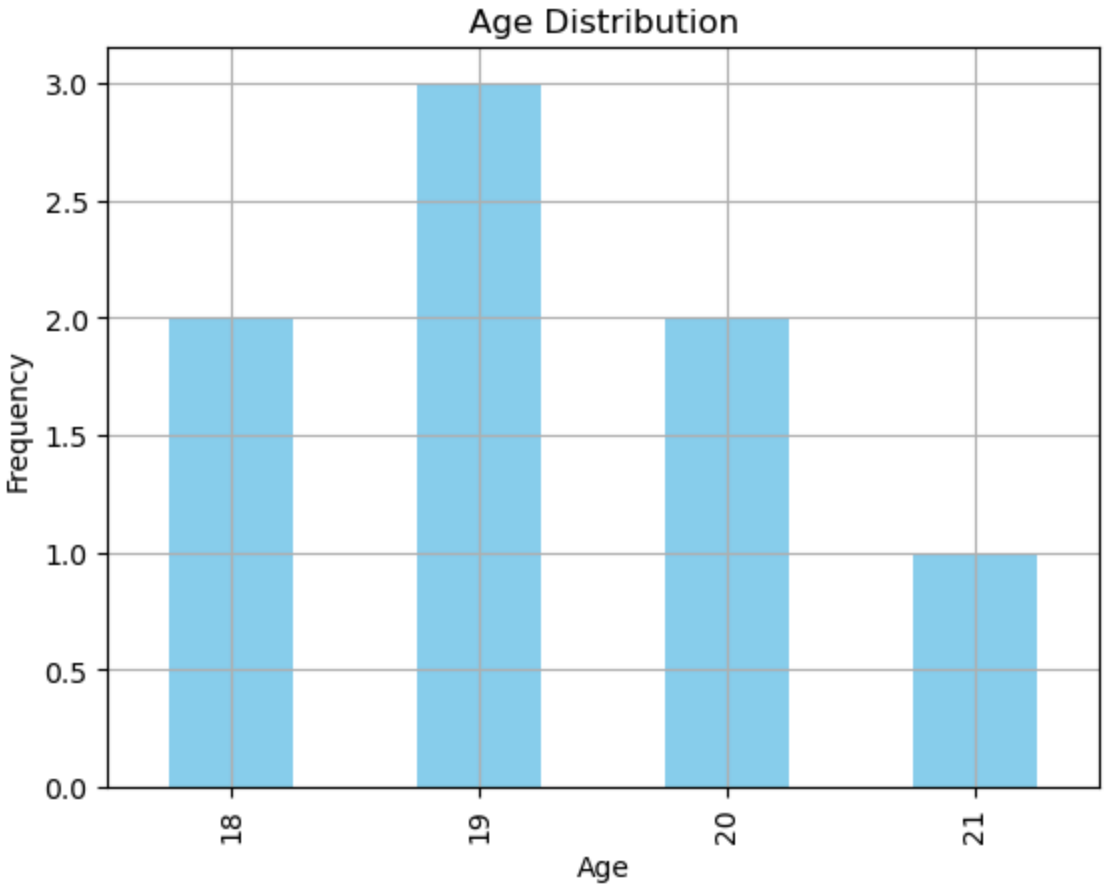
Python Example: Data Collection & Visualization

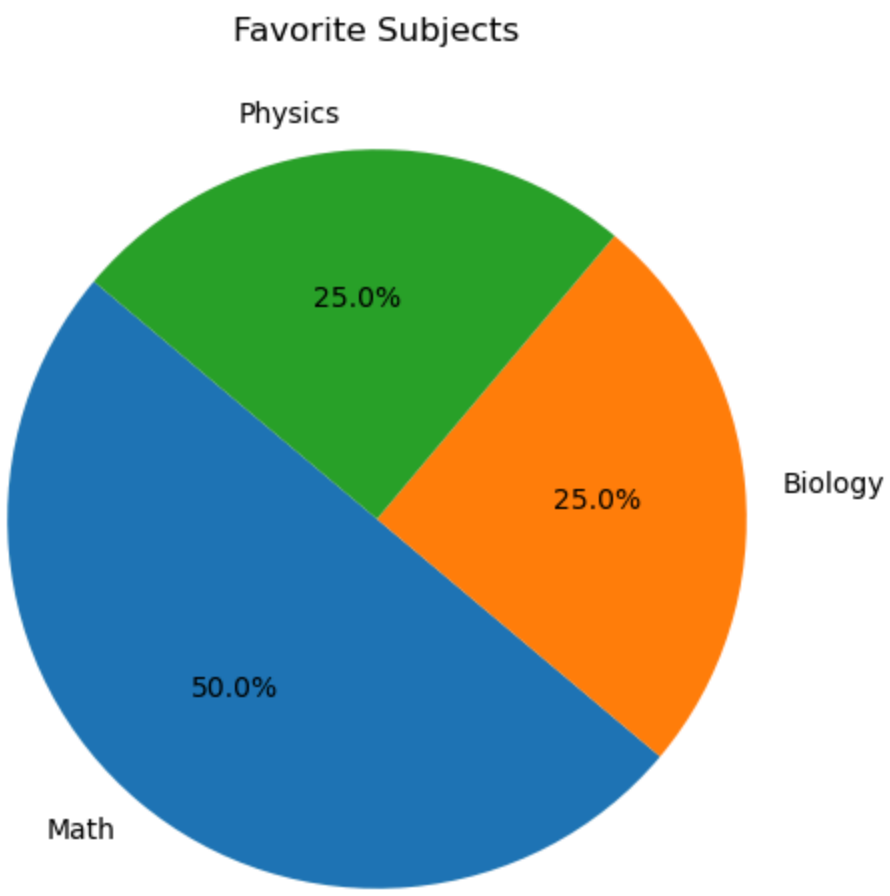
Let's create a simple dataset and present it using pandas and matplotlib.

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Step 1: Create a simple dataset (simulated survey data)
data = {
    'Name': ['Ali', 'Sara', 'John', 'Zara', 'Ahmed', 'Fatima', 'Leo', 'Ayesha'],
    'Age': [19, 20, 19, 18, 20, 18, 21, 19],
    'Favorite_Subject': ['Math', 'Biology', 'Math', 'Physics', 'Math', 'Biology', '
}

df = pd.DataFrame(data)

# Step 2: Frequency distribution of Ages
age_counts = df['Age'].value_counts().sort_index()
print("Age Frequency Distribution:")
print(age_counts)

print("::::::::::::::::::::::::::::::::::")
```

```
print("::::::::::::::::::::::::::::::::::")
print("::::::::::::::::::::::::::::::::::")

# Step 3: Bar chart for Age distribution
age_counts.plot(kind='bar', color='skyblue')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()

# Step 4: Pie chart for Favorite Subjects
subject_counts = df['Favorite_Subject'].value_counts()
subject_counts.plot(kind='pie', autopct='%1.1f%%', startangle=140, figsize=(6, 6))
plt.title('Favorite Subjects')
plt.ylabel('')
plt.show()
```

```
Age Frequency Distribution:
Age
18    2
19    3
20    2
21    1
Name: count, dtype: int64
::::::::::::::::::::::::::::::::::
::::::::::::::::::::::::::::::::::
::::::::::::::::::::::::::::::::::
```





✅ Output Explained:

Bar Chart: Shows how many students belong to each age.

Pie Chart: Shows percentage of students preferring each subject.

✅ Practice Task for You:

Try changing the dataset to include:

Gender

City

Test Scores

Then visualize:

Gender with a pie chart

Test scores with a histogram

In [ ]: