Probability Distributions (Normal, Binomial, Uniform)

Goal: Understand how probability distributions describe real-world data and how to visualize them using Python.

🧠 1. What is a Probability Distribution?

A **probability distribution** shows how the values of a variable are spread or how likely each possible outcome is.

In simple terms — it tells us **"how often we expect something to happen."**

🔹 Types of Distributions We'll Study

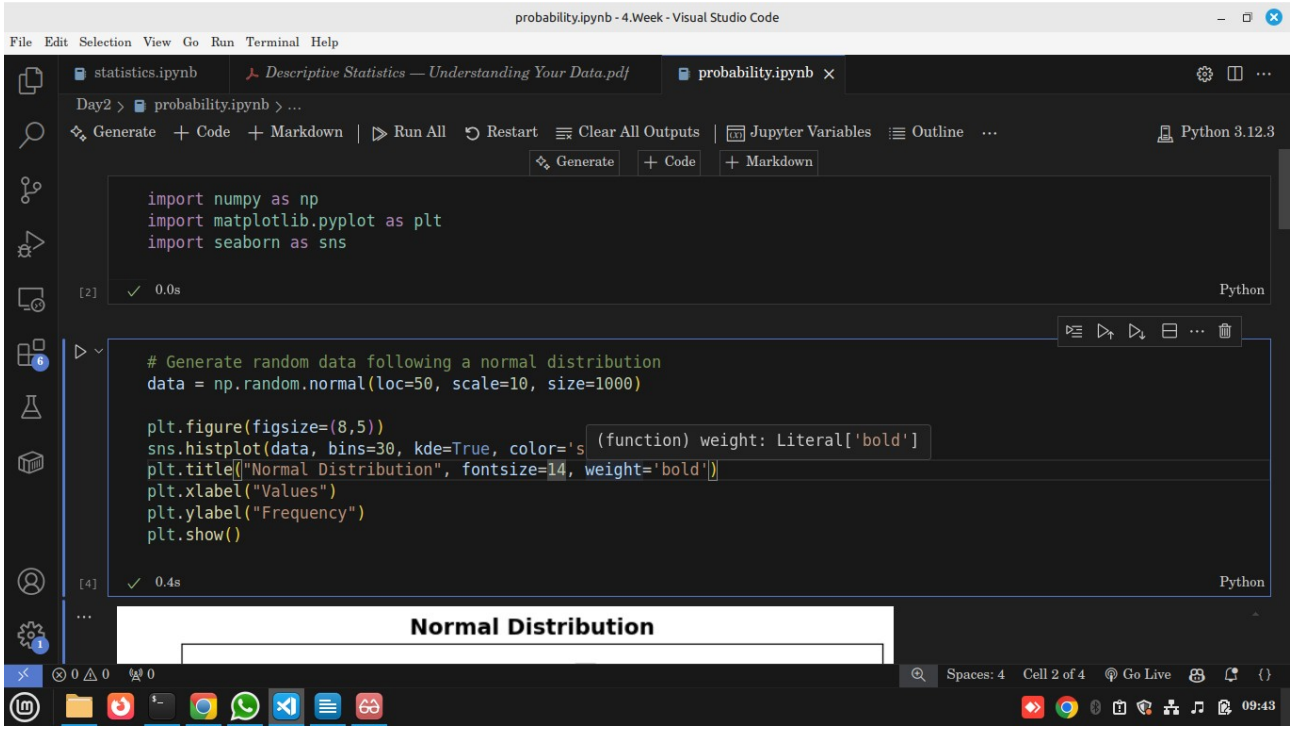
1. **Normal Distribution** – Bell-shaped curve, data around the mean.
2. **Binomial Distribution** – Used for outcomes like success/failure.
3. **Uniform Distribution** – All outcomes equally likely.

We'll use **NumPy**, **Matplotlib**, and **Seaborn** for demonstration.

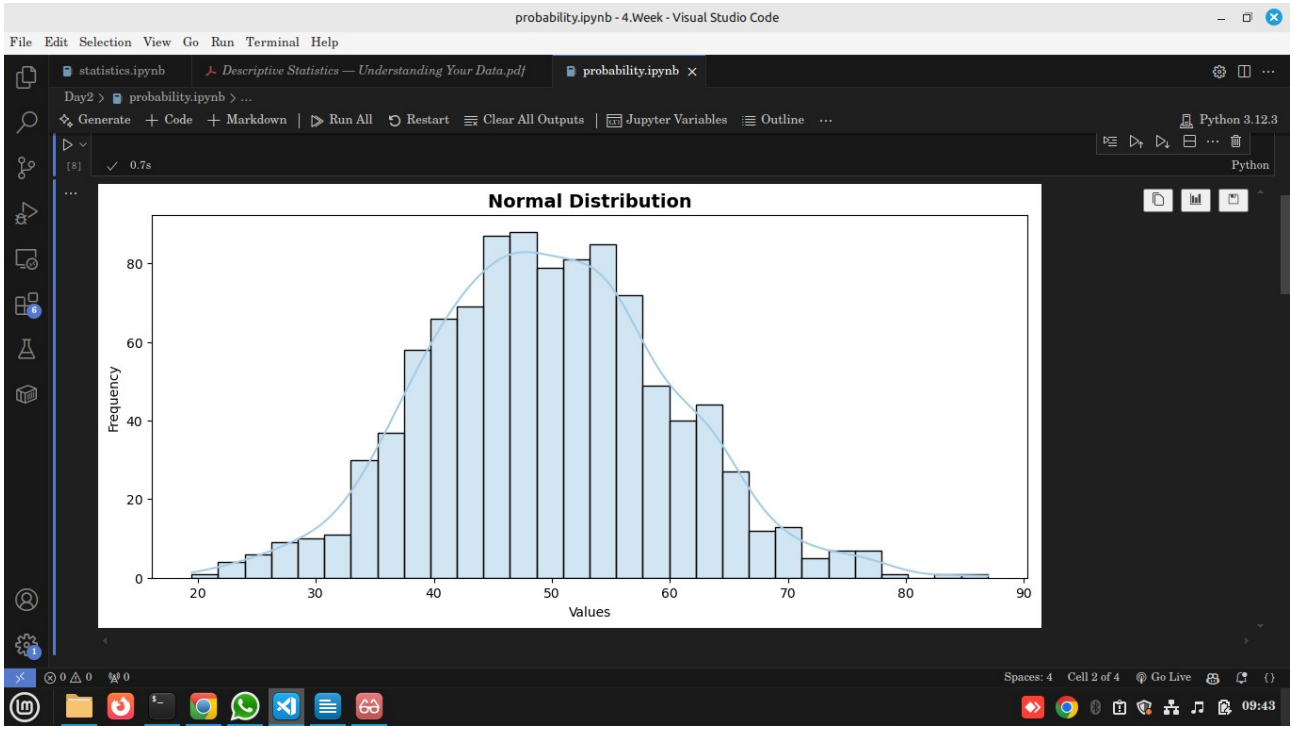
🔧 Import Required Libraries

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

🎯 2. Normal Distribution



Output



🧠 Explanation:

- **loc = 50** → Mean of the data (center of the curve).
- **scale = 10** → Standard deviation (spread).
- **kde=True** → Draws the smooth curve to show the density.

The shape is **bell-like** — most values cluster around the mean (50), and fewer appear as you move away from it.

📊 Insight:

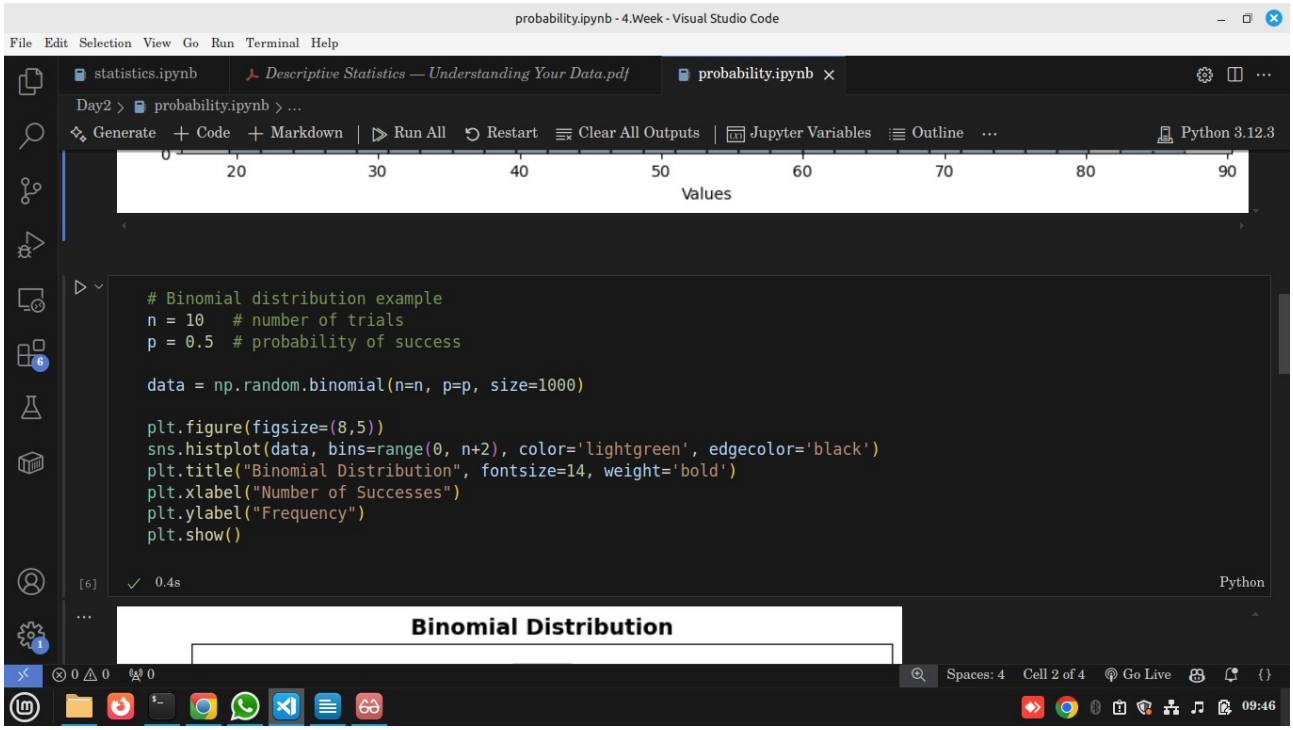
This is one of the most common distributions in nature and business.
Examples:

- Human height
- Exam scores
- Daily temperature

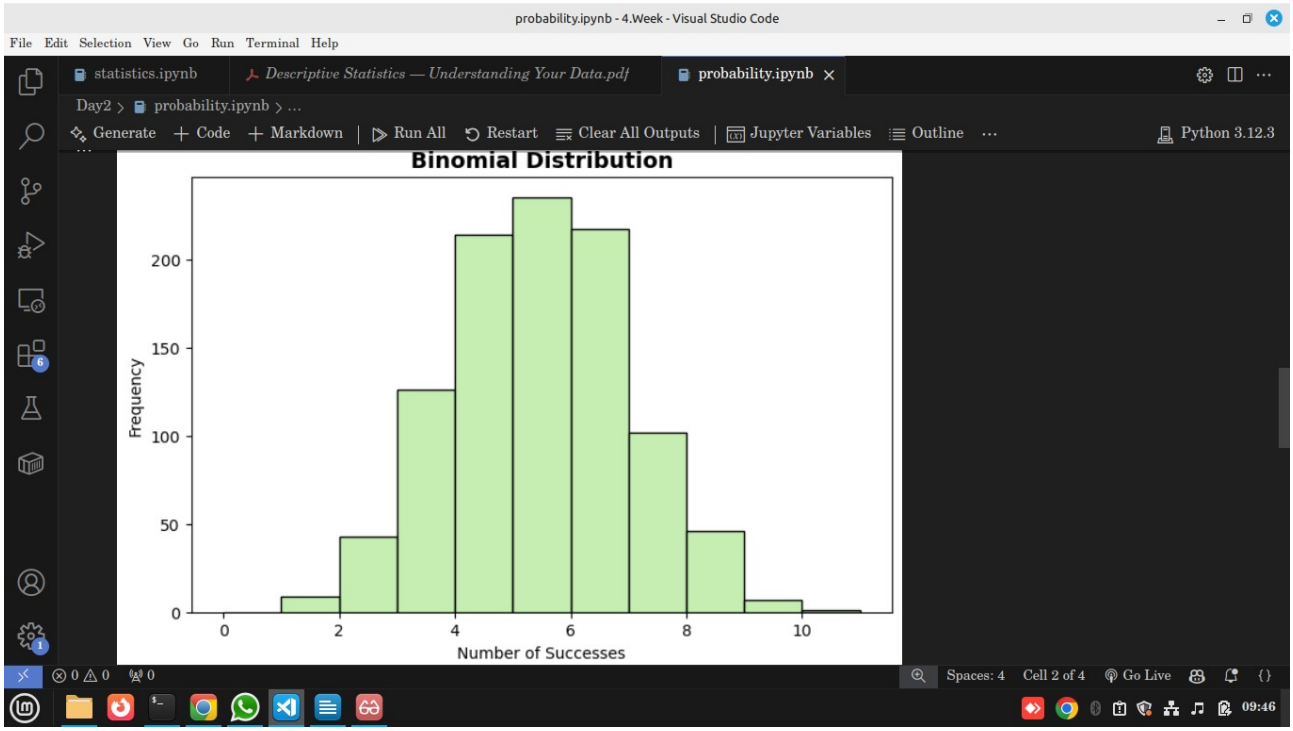
Why it matters:
The normal distribution helps analysts predict probabilities — for example, estimating how many students will score above average in an exam or how likely a machine part will last beyond a certain time.

🎯 3. Binomial Distribution

🧩 Code:



Output



🧠 Explanation:

- **n = 10** means each trial has 10 attempts (e.g., tossing a coin 10 times).
- **p = 0.5** means the chance of success (heads) is 50%.
- The most frequent result (the peak) occurs around **n × p = 5**.

This distribution is **discrete**, meaning it counts how many successes occur.

📊 Insight:

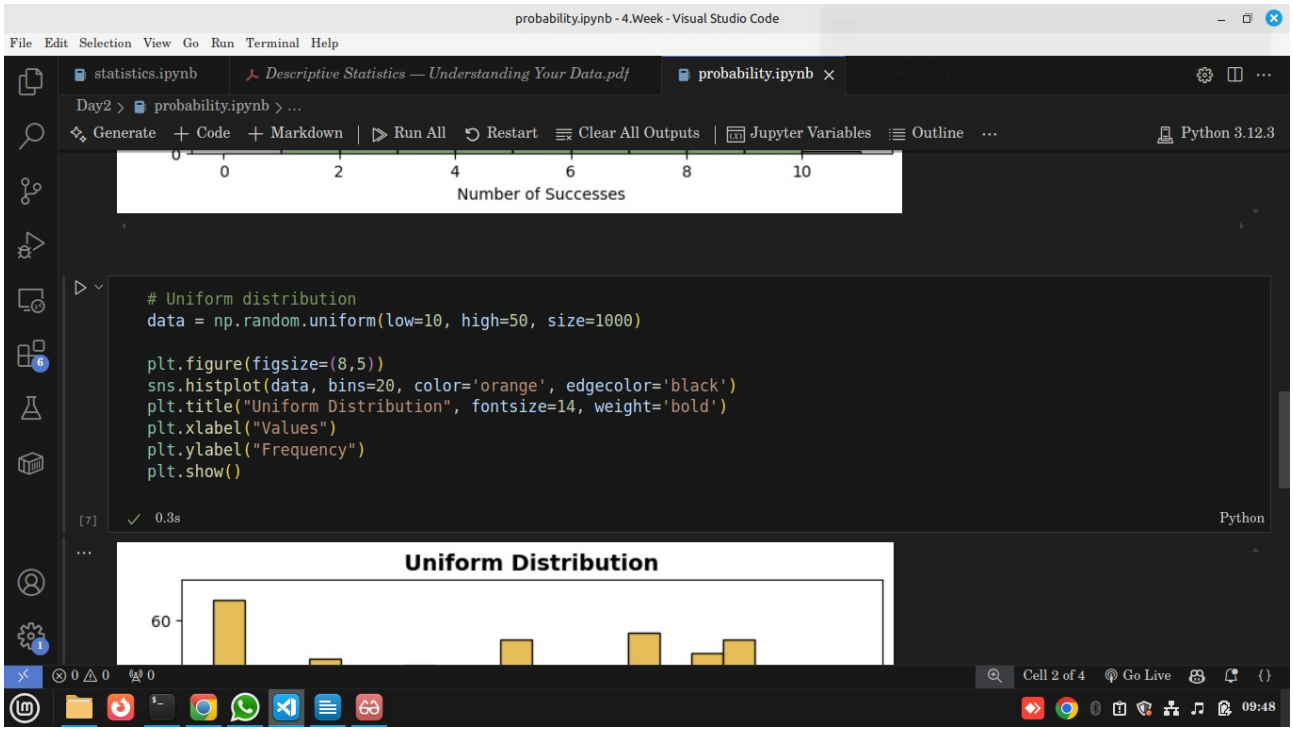
Used when outcomes are yes/no, success/failure, or click/no-click.
Examples:

- Number of customers who make a purchase out of 10 visitors
- Number of defective items in a batch
- Number of students who pass an exam

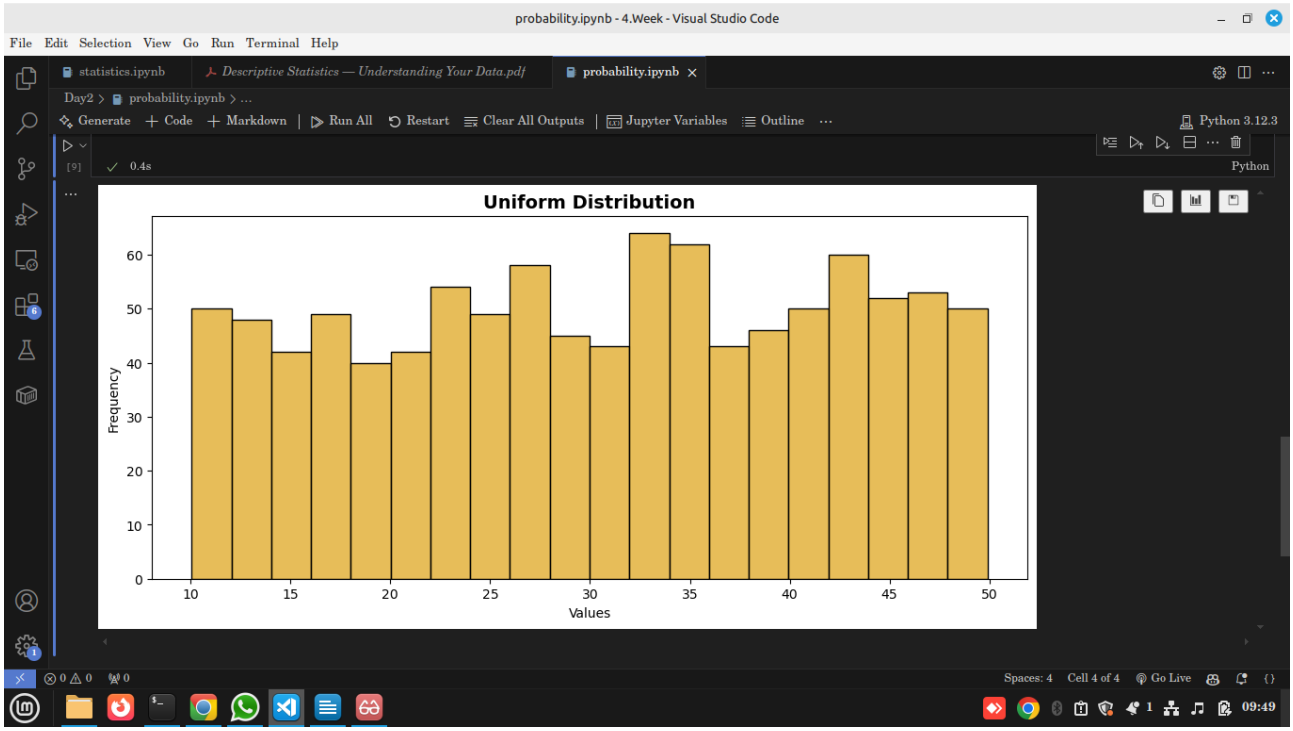
Why it matters:
Binomial distribution is critical for predicting probabilities in marketing, manufacturing, and medicine — any scenario with repeated binary outcomes.

🎯 4. Uniform Distribution

🧩 Code:



Output



🧠 Explanation:

- **low = 10, high = 50** means all numbers between 10 and 50 are equally likely.
- The histogram shows a **flat** pattern — each range has roughly the same frequency.

📊 Insight:

Used when randomness or fairness is required.
Examples:

- Random number generation
- Shuffling playlists
- Selecting random samples in research

Why it matters:
Uniform distribution ensures fairness in games, experiments, or simulations where no bias should exist.

📈 5. Comparison of the Three Distributions

Distribution	Type	Shape	Real-world Example	Key Use
Normal	Continuous	Bell curve	Heights, IQ, temperature	Natural and human behavior modeling
Binomial	Discrete	Stepped	Coin toss, pass/fail	Modeling success/failure rates
Uniform	Continuous	Flat	Random draws, fairness	Random sampling and simulations

---

💡 6. Real-world Importance

when you plot or analyze a normal distribution in a business setup, you're not just looking at the pretty bell curve — you're looking for patterns, thresholds, and probabilities that guide decisions.

Here are **practical, real-world insights** you'd look for 👇

💼 1. Understanding Customer Behavior

If your data shows **customer spending** follows a normal distribution:

- 📊 **Mean (Average spend):** tells you what a *typical* customer spends — e.g., KSh 1,500 per visit.
- 🎯 **Standard deviation:** shows how consistent spending is.
  - Low σ → customers spend almost the same every time (stable).
  - High σ → customers vary widely (some spend a lot, some spend very little).

- 💡 **Insight:**
  You can segment customers — e.g., target high spenders for VIP programs and low spenders for discounts.

📦 2. Product Quality Control (Manufacturing)

Let's say you measure **the weight of packaged goods** (e.g., 500g flour packets):

- If the weights form a normal curve centered around 500g:
  - ✅ Most products are close to the standard (good).
  - ⚠️ If the curve shifts left (mean < 500g): your machine is underfilling.
  - 🚨 If the curve widens (large σ): your process is inconsistent.

Insight:
The normal distribution helps detect **production drift** early — saving costs and preventing complaints.

🧍 3. Employee Performance Analysis

In HR or management:

- Suppose you collect **employee performance scores**.
- They often follow a normal distribution — most employees perform at an average level, few are very high or low.

Insight:

- The mean gives the "typical performance level."
- The tails identify **top performers** (for promotion) and **low performers** (for training).

💰 4. Sales Forecasting

If your daily or weekly **sales volume** follows a normal pattern:

- You can predict the **range of expected sales** (e.g., 68% of the time, sales will be within ±1σ).
- Helps plan **inventory** — avoid overstocking or running out.

Example:
If mean = 10,000 KSh, σ = 2,000
→ 68% of days sales will be between 8,000 and 12,000.

✅ This helps with staffing, supply ordering, and setting realistic targets.

📞 5. Customer Wait Times or Service Durations

In customer service or call centers:

- If call durations or waiting times are normally distributed,
    - The mean shows average service time.
    - Long right tail may indicate some customers wait *too long* → you can fix staffing or processes.

📈 6. Marketing Campaign Results

If you test new campaigns and measure **conversion rates**:

- Normal distribution can help you find **what's typical performance**.
- Outliers on the right (very high conversions) can reveal **strategies worth repeating**.

🧩 7. Assignment

Use Python and Seaborn to:

1. Plot a **Normal Distribution** with mean = 100 and std = 15.
2. Plot a **Binomial Distribution** with n = 20 and p = 0.3.
3. Plot a **Uniform Distribution** between 0 and 1.
4. For each:
    - Describe the shape.
    - Find the mean and variance using NumPy.
    - Give one real-life example.

🧭 Summary

- **Normal Distribution:** Data clusters around the mean.
- **Binomial Distribution:** Measures number of successes in fixed trials.
- **Uniform Distribution:** Equal likelihood for all outcomes.
- These form the **foundation of statistical modeling**, allowing data scientists to simulate real-world randomness, estimate probabilities, and make predictions.