Assignment 3 – Probability in Excel

This assignment is based on the **Transport Delays Dataset**. The goal is to apply the skills learned during the third Excel lab session to calculate probabilities, apply probability rules, and simulate probability distributions.

Dataset Information

The dataset contains 5,000 transport records with the following variables:

- **Trip_ID**: Unique identifier
- **Transport_Type**: Categorical (Bus, Train, Flight)
- **Region**: Categorical (Africa, Asia, Europe, North America, South America, Oceania)
- **Delay_Minutes**: Numerical (0–180, skewed distribution – shorter delays more common)
- **On_Time**: Categorical (Yes/No)
- **Weather_Condition**: Categorical (Clear, Rain, Snow, Fog, Storm)
- **Cancelled**: Categorical (Yes/No)

Objectives

By completing this assignment, you will demonstrate your ability to:

- Apply probability rules (addition, multiplication)
- Calculate probabilities from categorical data
- Compute expected values for discrete distributions
- Simulate binomial and normal distributions in Excel
- Interpret real-world probability results

Tasks

1. Basic Probability

- What is the probability that a trip is a **Flight**?
- What is the probability that a trip was **Cancelled**?
- What is the probability that a trip occurred in **Asia**?Calculate Variance and Standard Deviation

(Hint: Use counts and relative frequency.)

2. Addition Rule

- What is the probability that a trip is either a **Train** or was **Cancelled**?
  (Show calculation using Addition Rule: $P(A) + P(B) – P(A \cap B)$).

3. Multiplication Rule

- What is the probability that a trip was a **Bus** AND was **On Time**?
  (Show calculation using Multiplication Rule: $P(A) \times P(B|A)$).

4. Discrete Probability Distribution

Using **Delay_Minutes** (grouped into categories):

- 0–30 mins = Short Delay
- 31–60 mins = Medium Delay
- 61–120 mins = Long Delay
- 121–180 mins = Severe Delay

- Create a frequency table and compute **relative probabilities** for each delay category.
- Compute the **expected delay value (mean)** using probability weights.

5. Binomial Probability

Suppose the probability that a trip is **On Time = 0.7**:

- Using **BINOM.DIST**, calculate the probability that exactly **4 out of 6 trips** are on time.
- Calculate the probability that **at least 5 out of 6 trips** are on time.

6. Normal Distribution Simulation

Using **Delay_Minutes**:

- Compute the **mean** and **standard deviation** of delays.
- Using NORM.DIST, calculate the probability that a delay is:
  - Less than 30 minutes
  - Between 30 and 60 minutes
  - Greater than 60 minutes

7. Interpretation

Write **3–4 short bullet-point insights**, for example:

- Which transport type is most likely to be delayed?
- What weather condition is most often associated with cancellations?
- Do delays follow a roughly normal distribution?

Submission Guidelines

- Submit a single Excel file with the following sheets:
- **Sheet 1**: Cleaned dataset
- **Sheet 2**: Basic Probability
- **Sheet 3**: Addition & Multiplication Rule
- **Sheet 4**: Discrete Probability Distribution
- **Sheet 5**: Binomial Probability
- **Sheet 6**: Normal Distribution
- **Sheet 7**: Insights (typed in cells)

Submission Deadline: Sunday 12th October 2025 11:59pm GMT.

NOTE: Please save your assignment with your full name (e.g., *John Doe Assignment 1*) and send it to the following email addresses:

ruchinbox@gmail.com, tijanijumoke410@gmail.com & charlesopondo1376@gmail.com