Student Guide for Sync Session

Week 1: Maths-Stats for Data Science

This guide is your roadmap to making the most of our online session. Packed with essential tips and strategies, it's designed to keep you engaged, prepared and ready to dive into a smooth and productive learning journey. Get ready to participate, learn and thrive!

Session Overview

Session title	Probability Distributions & Case Study on House Price Data Analysis
Session duration	3 hours
Session type	• **Lectures**: Conceptual understanding of probability distributions, types and applications. • **Case Studies**: Analysing house price data using R.
Scope	This session introduces fundamental probability distributions and their applications in data science: • Understanding discrete and continuous probability distributions. • Exploring binomial, Poisson and normal distributions. • Hands-on case study with R for data analysis.
Learning Objectives	

Objective	Core Capability
Understand the structure and working of probability distributions	Analytical thinking in probability modelling
Apply binomial and Poisson distributions to real-world scenarios	Ability to evaluate statistical models
Implement probability distribution functions using R	Practical coding and data analysis skills
Analyse and interpret probability distributions in datasets	Decision-making based on statistical models

Software/tools	• R Programming (Libraries: dplyr, ggplot2, tidyr, knitr) • Dataset: House Price Data (houseprice.csv) • Presentation Tool: PowerPoint

Pro Tips for Success

- **Ask Bold Questions**: No question is too small—curiosity is the key to learning!
- **Be Hands-On**: Practice coding to reinforce your understanding.
- **Collaborate**: Share insights and learn from your peers.

Session Details

Topic	A glimpse	Insight / Actionable
Introduction to Probability Distributions	Gain a quick understanding of probability distributions and their role in data analysis.	Reflect on how probability is used in everyday decisions like predicting weather.
Discrete vs. Continuous Distributions	Learn the difference between discrete and continuous probability distributions.	Example: Number of students in a class (discrete) vs. temperature readings (continuous).
Binomial Distribution	Explore Bernoulli trials and binomial probability formulas.	Example: Predicting email open rates based on user behaviour.
Poisson Distribution	Understand how Poisson distribution models rare events over time or space.	Example: Counting customers arriving at an ATM per minute.
Normal Distribution	Learn key characteristics and the empirical rule (68-95-99.7%).	Example: Heights of people, IQ scores and stock price fluctuations.
Case Study: House Price Data Analysis	Apply statistical techniques to real-world data in R.	Load, clean, visualise and interpret data insights.
Data Cleaning in R	Learn methods for handling missing values.	Example: Replacing missing values in `LotFrontage` using the mean.
Visualisation with R	Create histograms and boxplots for exploratory data analysis.	Example: Boxplot for house prices by quality rating.
Correlation Analysis	Use correlation matrices and heatmaps to identify relationships.	Example: Understanding how house price correlates with lot size.
Practical Considerations	Discuss trade-offs in statistical modeling.	Example: Choosing the right probability distribution for given data.

Post-Session Activities

Reflection challenge	What did you find exciting about probability distributions?
Explore more	• **Read**: R documentation on probability distributions and visualisation. • **Watch**: Tutorials on probability theory and data visualisation in R. • **Play Around**: Use different datasets to apply probability distributions in real-world scenarios.
Get inspired	Did you know probability distributions are used in risk assessment, finance,

	healthcare and AI? Mastering them will unlock numerous opportunities!
The journey ahead	Explore advanced statistical techniques like hypothesis testing and regression analysis to further strengthen your data science skills!