Week 1: Introduction to Data Science

- Define data science and explain its interdisciplinary nature.
- Identify key components of the data science workflow.
- Describe the CRISP-DM framework and its application in data-driven projects.
- Understand the importance of data-driven decision-making in various industries.
- Recognize the significance of data cleaning as a crucial step in data science.

Week 2: Types of Data & Visualization

Types of Data

- Define and differentiate between different types of data: nominal, ordinal, interval, and ratio.
- Determine which descriptive statistics are appropriate for each type of data.
- Explain why knowing the data type is essential for selecting proper analytical methods.
- Recognize the impact of data types on visualization and statistical analysis.

Data Visualization

- Define exploratory data analysis (EDA) and its role in understanding data.
- List the goals of data visualization, including identifying patterns and detecting anomalies.
- Interpret commonly used visualizations such as histograms, scatterplots, and box plots.
- Offer critiques of misleading or ineffective visualizations.
- Understand the relationship between visualization and statistical modeling.

Week 3: Probability Review & Bootstrapping

Probability Review

- Define probability and its role in statistical inference.
- Explain the concepts of random variables and probability distributions.
- Describe Bernoulli, Binomial, and Normal distributions and their applications in data science.
- Compute probabilities and understand the importance of probabilistic reasoning in data-driven decisions.

Bootstrapping

- Define the three steps of bootstrapping and explain its purpose.
- Define confidence intervals and their interpretation in statistical analysis.

- Specify a bootstrap procedure to estimate confidence intervals.

Week 4: Hypothesis Testing & Statistical Testing

Hypothesis Testing

- Identify the components of a hypothesis test, including the null and alternative hypotheses.
- Define p-values and explain their significance in hypothesis testing.
- Describe Type I and Type II errors and their implications.
- Work through examples of hypothesis testing in real-world contexts.

Statistical Testing

- Explain the logic behind statistical hypothesis testing.
- Determine appropriate hypothesis tests for different types of data.
- Compare parametric and non-parametric tests and understand when to use each.
- Implement statistical tests such as the t-test and chi-square test.

Week 5: Advanced Statistical Testing

- Review the importance of statistical tests in drawing conclusions from data.
- Compare and contrast one-tailed and two-tailed tests.
- Discuss the assumptions underlying different hypothesis tests.
- Explore non-parametric alternatives for scenarios where assumptions are violated.
- Interpret test results and their implications for scientific inquiry.

Week 6: Family-Wise Error and Multiple Hypothesis Testing

- Understand the problem of multiple hypothesis testing and its impact on statistical conclusions.
- Define family-wise error rate and why it matters in statistical inference.
- Implement multiple comparison correction methods (Bonferroni correction, α-spending, etc.).
- Recognize when family-wise error correction is necessary in hypothesis testing.
- Discuss trade-offs between different correction methods and their implications.

Week 7: Dataset Cards

- Define dataset cards and their role in responsible data documentation.
- Explain the importance of dataset transparency in machine learning and data science.
- Identify key components of a dataset card, including summary information and known biases.
- Discuss best practices for creating dataset documentation.
- Recognize the role of dataset cards in ensuring ethical data use and reproducibility.