Introduction to the Year 8 Data Science Program

**Program Title:** Data Science: Unlocking Insights in the Real World

**Target Audience:** Year 8 Students (Ages 13-14) - Selective School Edition

**Program Overview:**

Welcome to the Year 8 Data Science program, an exciting and innovative course designed to equip students with essential skills for the 21st century. In today's data-rich world, the ability to understand, analyse, and interpret data is no longer a niche skill but a fundamental competency. This program provides a foundational understanding of data science principles and practices, empowering students to become data-literate citizens and future innovators.

**Alignment with NSW Stage 4 Science Syllabus and Australian Curriculum: Science (Version 8.4):**

This program is meticulously designed to align with the NSW Stage 4 Science Syllabus, specifically addressing the "Data Science 1" focus area (SC4-DA1) and the broader Working Scientifically skills (SC4-WS). It also strongly integrates with the Australian Curriculum: Science (Version 8.4), particularly the Science Inquiry Skills strand.

**Key Curriculum Links:**

* **NSW Stage 4 Science Syllabus - Data Science 1:**
  + **SC4-DA1-01:** Explains how data is used by scientists to model and predict scientific phenomena.
  + **SC4-WS-06:** Uses data to identify trends, patterns and relationships, and draw conclusions.
  + **SC4-WS-07:** Identifies problem-solving strategies and proposes solutions.
  + *Throughout the program, students will develop skills across all Working Scientifically outcomes (SC4-WS-01 to SC4-WS-08), including observing, questioning, planning investigations, processing and analysing data, problem-solving, communicating, and reflecting on scientific methods.*
* **Australian Curriculum: Science (Version 8.4) - Science Inquiry Skills (Year 7-8):**
  + **ACSIS124:** Formulating questions or hypotheses that can be investigated scientifically.
  + **ACSIS125:** Planning, selecting and using appropriate investigation types, including fair tests and controlled experiments, to collect reliable data; assessing risk and addressing ethical considerations.
  + **ACSIS126:** Processing, analysing and evaluating data; identifying patterns and relationships, and summarising data using a range of representations.
  + **ACSIS127:** Communicating ideas, findings and evidence-based conclusions using scientific language, and appropriate representations.
  + **ACSIS131:** Evaluating investigation methods and conclusions, including identifying limitations and suggesting improvements.

**Pedagogical Logic: Real-World Analysis with Modern Tooling**

This program is built upon a pedagogical approach that emphasizes:

* **Inquiry-Based Learning:** Students are active learners, encouraged to ask questions, investigate phenomena, and discover insights through data exploration and analysis. The program moves beyond rote memorization to foster genuine understanding and critical thinking.
* **Hands-On Activities:** Learning is grounded in practical experience. Students engage in hands-on data collection, manipulation, visualization, and analysis activities, reinforcing theoretical concepts through direct application.
* **Real-World Relevance:** The program connects data science concepts to real-world problems and applications across diverse scientific and societal domains. This approach enhances student engagement by demonstrating the practical value and impact of data science in addressing contemporary challenges and making discoveries.
* **Development of 21st-Century Skills:** The program explicitly cultivates essential 21st-century skills, including:
  + **Digital Literacy:** Students become proficient in using modern digital tools for data science, including Observable notebooks and Python programming.
  + **Computational Thinking:** Students develop computational thinking skills through data analysis, algorithmic thinking (implicitly through AI interaction and potentially simplified coding), and problem-solving using computational methods.
  + **Critical Thinking and Data-Driven Reasoning:** Students learn to evaluate data critically, identify biases, draw evidence-based conclusions, and understand the limitations of data and models.
  + **Communication and Collaboration:** Students develop communication skills through data visualization and presentation, and collaboration skills through group projects and peer feedback activities.
* **Gradual Skill Development:** The curriculum follows a structured progression, starting with foundational concepts of data and data science in Week 1 and gradually building towards more advanced topics like data wrangling, statistical analysis, and an introduction to AI in data analysis by Week 7. This scaffolded approach ensures students develop a solid understanding and skillset incrementally.
* **Integration of Modern Tools:** The program leverages cutting-edge, industry-standard tools to provide students with authentic data science experiences:
  + **Observable Framework:** Provides an interactive, web-based environment for data exploration, visualization, and computational notebooks. Its real-time interactivity and collaborative features enhance student engagement and learning.
  + **Python Programming:** Introduces Python as a powerful yet accessible language for data manipulation, analysis, and visualization. Students gain practical coding skills that are highly valuable in data science and STEM fields.
  + **AI Tutor Integration:** Incorporates an AI Tutor as a personalized learning assistant, providing students with on-demand support, feedback, and guidance. This AI integration aims to enhance personalized learning and cater to diverse student needs.

**Program Goal:**

The overarching goal of this Year 8 Data Science program is to empower students to become confident, data-literate individuals who can:

* Understand the role of data in scientific inquiry and real-world problem-solving.
* Apply data science principles and techniques to investigate scientific questions.
* Effectively communicate data-driven insights using visualizations and evidence-based reasoning.
* Critically evaluate data and models, understanding their strengths and limitations.
* Appreciate the ethical considerations of data collection and use in a digital society.

By the end of this program, students will be well-prepared to engage with data critically and creatively, fostering a lifelong curiosity for scientific inquiry and a strong foundation for future STEM studies and careers in a data-driven world.

**Overall Theme:** Data Science: Unlocking Insights in the Real World

**Week 1: Data Science Foundations & Digital Immersion**

* **Focus:** Rapid Introduction to Data Science, Digital Tool Immersion (Observable, Python, AI Tutor).
* **Content:**
  + What is Data Science? Applications in cutting-edge science.
  + Data Sources: Diverse scientific data sources (real-time, simulations, large datasets).
  + Observable Notebooks: Intensive introduction - navigation, markdown/code cells, Python basics, sharing.
  + AI Tutor: Introduction and practical exercises for Q&A and code assistance.

**Week 2: Data Collection Mastery & Digital Responsibility**

* **Focus:** Data Collection Techniques (rigorous methods), Digital Footprint & Ethics, Accuracy, Precision, Validity (data quality).
* **Content:**
  + Lesson 1: Digital Footprint & Data Ethics: Online safety, privacy, responsible data use, ethical considerations.
  + Lessons 2 & 3: Data Collection & Quality:
    - Primary and Secondary Data, data collection methods design.
    - Accuracy, Precision, Validity: Practical exercises, error analysis, experimental design for data quality.
    - Digital Data Capture in Observable.

**Week 3: Data Visualization & Interactive Storytelling**

* **Focus:** Data Visualization Techniques (graph types), Interactive Dashboards in Observable, Visual Clarity & Data Storytelling.
* **Content:**
  + Types of Graphs: Comprehensive exploration, strengths/weaknesses for data types.
  + Python Plotting Libraries (matplotlib, Plotly, Seaborn): Hands-on visualization creation in Observable.
  + Interactive Elements: Dashboards with sliders, dropdowns, tooltips in Observable.
  + Peer Review & Critique of data visualizations.

**Week 4: Descriptive Statistics & Data Interpretation**

* **Focus:** Descriptive Statistics Mastery (mean, median, mode, standard deviation, variance, percentiles), Data Distribution & Outliers, Statistical Interpretation.
* **Content:**
  + Central Tendency & Dispersion: In-depth exploration, formulas, appropriate use.
  + Data Distribution: Histograms, frequency distributions, normal distribution, skewness, kurtosis (conceptual).
  + Outlier Detection & Analysis: Methods for identification, causes, handling outliers.
  + Python for Statistical Calculation (NumPy, SciPy) in Observable.

**Week 5: Scientific Question Formulation & Experimental Design - Advanced**

* **Focus:** Formulating Complex Scientific Questions, Advanced Experimental Design, Hypothesis Testing & Prediction.
* **Content:**
  + Question Types: Descriptive, comparative, correlational, causal – design for each.
  + Experimental Design Principles: Control groups, variables, randomization, replication, blinding.
  + Virtual Experiments & Simulations in Observable for hypothesis testing.
  + AI Scenario Generator for complex experimental design challenges.

**Week 6: Advanced Data Wrangling & Real-World Datasets**

* **Focus:** Handling messy real-world datasets, advanced data cleaning (Python/Pandas), data transformation, feature engineering (conceptual).
* **Content:**
  + Working with real-world datasets (provided and student-sourced).
  + Data Cleaning Techniques in Python (Pandas): Handling missing values, inconsistent formats, data validation.
  + Data Transformation: Reshaping, merging, aggregating data.
  + Feature Engineering (Conceptual): Creating new variables from existing data for better analysis.

**Week 7: AI-Powered Data Analysis & Predictive Modelling (Introduction)**

* **Focus:** Introduction to AI in Data Analysis (pattern recognition, clustering, classification - conceptual), AI tools for data summarization and trend analysis, Basic Predictive Modelling (simplified introduction).
* **Content:**
  + AI Concepts in Data Analysis: Pattern recognition, clustering, classification (conceptual overview).
  + AI Tools for Data Analysis in Observable (if available/suitable - or explore external simplified AI tools).
  + Simplified Introduction to Predictive Modelling: Basic concepts, visual tools (if feasible), understanding model limitations.
  + Ethical considerations of AI in data analysis and prediction.

**Week 8: Group Project - Real-World Data Science Challenge**

* **Focus:** Collaboration, problem-solving, applying data science to real-world scientific/societal issues, scientific communication.
* **Content:**
  + Team-based projects: Students select a real-world problem and relevant dataset.
  + Project Stages: Question formulation, data acquisition, data analysis, visualization, model building (if applicable), presentation.
  + Emphasis on collaborative work, project management, and effective communication of findings.
  + Regular AI-driven feedback on project progress and data analysis.

**Week 9: Project Refinement & Advanced Review**

* **Focus:** Refining project analysis, visuals, presentations, Comprehensive knowledge checks, Exam preparation.
* **Content:**
  + Peer and AI Review of group projects: Iterative improvement based on feedback.
  + Advanced Practice Exam Questions: Challenging data analysis problems mimicking exam format.
  + Focus on consolidating data science skills and preparing for summative assessment.
  + Refinement of Observable notebooks and project presentations.

**Week 10: End-of-Semester Exam & Future Pathways**

* **Focus:** Formal assessment of data analysis, visualization, coding, scientific reasoning, Self-evaluation and future pathways.
* **Content:**
  + End-of-Semester Exam:
    - Section 1: Data Interpretation (graphs, statistics, short-answer questions).
    - Section 2: Coding and Notebook Tasks (Python in Observable).
    - Section 3: Real-World Problem Scenarios (evidence-based problem-solving).
  + Exam review and feedback.
  + Discussion of future pathways in data science and related STEM fields.
  + Student self-reflection on learning and skill development.

2. Complete Set of 30 Lesson

Week 1: Data Science Foundations & Digital Immersion

Lesson 1.1: Introduction to Data Science: What, Why, and Where?

* **Title:** Data Science: Unveiling the Power of Data in the 21st Century
* **Learning Outcomes:**
  + Students will be able to define data science and explain its interdisciplinary nature.
  + Students will be able to identify key applications of data science in various scientific fields and real-world scenarios.
  + Students will understand the increasing importance of data science in modern society and scientific discovery.
  + *NSW Curriculum Outcomes:* SC4-DA1-01 (explains how data is used), SC4-WS-07 (identifies problem-solving strategies)
* **Overview:** This introductory lesson provides a compelling overview of data science, its definition, interdisciplinary nature, and significant applications, emphasizing its relevance in today's world and scientific advancements.
* **Activities:**
  + **Warm-Up (10 mins):** "Data Science Headlines" - Show a series of news headlines or short summaries related to data science breakthroughs in different fields (e.g., AI in medicine, climate change modelling, astronomical discoveries, personalized learning). Ask students: "What do these seemingly different stories have in common?" Lead to the idea of data being central.
  + **Main Activity 1: "Defining Data Science - The Interdisciplinary Web" (25 mins):**
    - Introduce the definition of Data Science: "an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from data in various forms."
    - Visually represent Data Science as a hub with connections to different disciplines: Statistics, Computer Science, Domain Expertise (e.g., Biology, Physics, Social Sciences), Communication, Ethics.
    - Discuss each connection: How does Statistics contribute? Computer Science? Why is domain expertise crucial? Why is communication and ethics essential in data science?
  + **Main Activity 2: "Data Science in Action - Real-World Impact" (25 mins):**
    - Present a series of short case studies showcasing the impact of data science in diverse fields (tailored to be engaging for selective students):
      * **Precision Medicine:** How data science and AI are revolutionizing disease diagnosis, personalized treatment plans, and drug discovery (e.g., genomics, AI-powered diagnostics).
      * **Climate Change Modelling & Prediction:** How data science is used to analyse vast climate datasets, build complex models, and predict future climate scenarios (e.g., climate simulations, extreme weather forecasting).
      * **Astrophysics & Big Data Astronomy:** How data science enables astronomers to analyse massive datasets from telescopes, discover new celestial objects, and understand the universe (e.g., exoplanet discovery, galaxy evolution).
      * **Materials Science & Accelerated Discovery:** How data science and machine learning are speeding up the discovery and design of new materials with specific properties (e.g., new battery materials, high-performance alloys).
    - For each case study, guide students to identify:
      * What is the scientific problem being addressed?
      * What types of data are being used?
      * How is data science contributing to solving the problem or making new discoveries?
    - Brief group discussions on each case study, followed by class sharing.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - Why is data science considered a "game-changer" in science and many other fields?
    - What are some of the most exciting potential future applications of data science you can imagine?
    - What skills and qualities do you think are important to succeed in data science?
  + **Assessments:**
    - Formative Quiz (5 mins): Short multiple-choice quiz on the definition of data science, its interdisciplinary nature, and key application areas. AI-generated quiz questions.
    - Observation of student participation in discussions and group activities.
  + **Student Workbook:**
    - Definition of Data Science and its interdisciplinary components (diagram).
    - Summary table for case studies, identifying the problem, data types, and data science impact.
    - Reflection questions on the importance and future of data science.
  + **Extension Activities:**
    - Students research a specific data scientist working in a field that interests them and create a short presentation or profile.
    - Challenge: Debate: "Is data science just a new name for statistics, or is it a fundamentally different field?"
* **Lesson Materials:**
  + Slide Deck: Introduction to Data Science, definition, interdisciplinary web diagram, case study examples (with visuals).
  + Handout: Case study descriptions, summary table for case study analysis, definition of data science.
  + Formative Quiz (printable or digital).
  + (Optional) Short video clips showcasing data science applications.

Lesson 1.2: Digital Toolkit: Getting Started with Observable & AI Tutor

* **Title:** Your Data Science Lab: Navigating Observable and Meeting Your AI Assistant
* **Learning Outcomes:**
  + Students will be able to create and navigate an Observable notebook environment.
  + Students will be able to create and use markdown and code cells within Observable.
  + Students will be able to execute basic Python code in Observable and view output.
  + Students will be able to interact with the integrated AI Tutor for questions and assistance within Observable.
  + *NSW Curriculum Outcomes:* SC4-WS-03 (Use digital technologies), SC4-WS-06 (Process data and information)
* **Overview:** This hands-on lesson introduces students to their primary digital tools: the Observable notebook and the AI Tutor. Students will learn to navigate Observable, create notebooks, use markdown and code cells, and start interacting with the AI Tutor for support.
* **Activities:**
  + **Warm-Up (10 mins):** "Digital Tools We Use" - Quick brainstorm: "What digital tools do you already use for learning, creating, or exploring information?" (e.g., Google Docs, online simulations, coding platforms, online encyclopedias). Discuss: What makes these tools useful? What are the benefits of digital tools in learning? Lead to the idea of Observable as a powerful digital tool for data science.
  + **Main Activity 1: "Observable Interface Expedition" (25 mins):**
    - Guided, step-by-step walkthrough of the Observable interface:
      * Account login and homepage overview.
      * Navigating to "Notebooks" and creating a new notebook ("+ New").
      * Introduction to the notebook interface: Title, cells (markdown and code), toolbar.
      * Creating a Markdown Cell: Click "+", choose "Markdown". Type a title (e.g., "My First Observable Notebook"). Demonstrate markdown formatting basics (headings, bold, italics). Run the cell (Shift+Enter).
      * Creating a Code Cell: Click "+", choose "Code". Type simple Python code: print("Hello Data Science!"). Run the cell. Explain output display.
      * Adding more markdown and code cells. Rearranging cells (drag and drop). Saving notebooks (automatic, but emphasize). Sharing notebooks (briefly introduce "Share" button).
    - Encourage students to actively follow along and experiment within their own Observable notebooks as you demonstrate.
  + **Main Activity 2: "Meet Your AI Tutor - Ask Me Anything (Data Science Edition)" (20 mins):**
    - Introduce the integrated AI Tutor (specific implementation will depend on the chosen AI tool).
    - Demonstrate how to access and interact with the AI Tutor (voice or chat interface).
    - Guided practice using the AI Tutor:
      * Ask basic questions about Observable: "How do I create a code cell?", "What is markdown used for?", "How do I save my notebook?".
      * Ask basic data science concept questions: "What is data?", "What is an algorithm?", "What is statistics?".
      * Ask for help with simple Python code: "How do I print something in Python?", "How do I create a list in Python?".
    - Encourage students to ask their own questions and explore the AI Tutor's capabilities. Emphasize it's a tool for help and learning, not just getting answers.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - What are your first impressions of using Observable? What do you find interesting or useful?
    - How do you think the AI Tutor can support your learning in data science? What are its potential benefits?
    - What are you most looking forward to doing with Observable and data science in the coming weeks?
  + **Assessments:**
    - Formative Check (5 mins): Quickly review student Observable notebooks to ensure they have created a notebook, added markdown and code cells, and run basic code.
    - Observe student interaction with the AI Tutor.
  + **Student Workbook:**
    - Step-by-step guide to creating an Observable account (if needed) and creating a basic notebook.
    - Key features of the Observable interface (screenshot and labels).
    - Basic markdown formatting guide.
    - Instructions for accessing and using the AI Tutor with example questions.
    - Practice exercises for creating markdown and code cells and running Python code.
  + **Extension Activities:**
    - Explore Observable's example notebooks in the "Explore" section to see more advanced uses of the platform.
    - Challenge: Experiment with more advanced markdown formatting (links, images, tables). Try slightly more complex Python code in code cells (basic arithmetic, variable assignment).
* **Lesson Materials:**
  + Slide Deck: Introduction to Observable, interface tour, step-by-step guide for notebook creation, examples of markdown and code cells, introduction to AI Tutor.
  + Handout: Observable guide, AI Tutor usage guide, practice exercises for Observable.
  + Computers/Tablets with internet access for Observable.
  + (Optional) Pre-lesson instructions for Observable account setup if needed.

Lesson 1.3: Data Exploration: Types of Data in Science

* **Title:** Data Under the Microscope: Exploring the Variety of Data in Science
* **Learning Outcomes:**
  + Students will be able to define "data" in a scientific context.
  + Students will be able to differentiate between qualitative and quantitative data, and provide scientific examples of each.
  + Students will be able to further classify quantitative data as discrete or continuous, and provide scientific examples.
  + Students will be able to identify examples of different data types in real-world scientific datasets.
  + *NSW Curriculum Outcomes:* SC4-DA1-01 (examines a range of sources of data), SC4-WS-06 (analyses data and information)
* **Overview:** This lesson delves into the fundamental concept of data, focusing on the different types of data encountered in scientific investigations. Students will learn to classify data as qualitative or quantitative, and further refine quantitative data into discrete and continuous types, using scientific examples throughout.
* **Activities:**
  + **Warm-Up (10 mins):** "Data Around the Room" - Quick observation activity. Ask students to look around the classroom and identify things that could be considered "data" related to the classroom environment. Examples: colours of objects, number of desks, temperature reading, types of materials, student opinions (if surveyed). Briefly list these on the board.
  + **Main Activity 1: "Qualitative vs. Quantitative Data - The Great Divide" (25 mins):**
    - Introduce the fundamental distinction: Qualitative Data (descriptive, categorical, non-numerical) vs. Quantitative Data (numerical, measurable).
    - Use examples from the "Data Around the Room" brainstorm to classify each item as Qualitative or Quantitative.
    - Provide more scientific examples of each data type across different scientific disciplines:
      * **Qualitative:** Colour of chemical precipitates, descriptions of animal behaviour, types of rock formations, patient symptoms descriptions, survey responses with open-ended questions.
      * **Quantitative:** Temperature readings, height of plants, reaction rates, number of species, population size, experimental measurements (mass, volume, time).
    - Interactive sorting activity: Provide cards or a digital drag-and-drop activity with various scientific data examples. Students classify them as Qualitative or Quantitative. Discuss and clarify classifications as a class.
  + **Main Activity 2: "Discrete vs. Continuous Data - Refining the Numbers" (20 mins):**
    - Focus on Quantitative Data. Introduce the sub-classification: Discrete Data (countable, whole numbers) vs. Continuous Data (measurable, can take any value within a range, often includes decimals/fractions).
    - Use previous Quantitative Data examples to classify them as Discrete or Continuous.
    - Provide more examples and scenarios to solidify understanding:
      * **Discrete:** Number of leaves on a plant, number of bacteria colonies, number of trials in an experiment, number of students in a class.
      * **Continuous:** Height of a tree, temperature of a solution, time taken for a reaction, mass of a sample, pH value.
    - "Data Type Challenge": Present scientific scenarios or research questions. Students identify the type of quantitative data that would be collected (Discrete or Continuous) and justify their choice.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - Why is it important to distinguish between different types of data in science?
    - How does the type of data you collect influence how you can analyse it and what kinds of conclusions you can draw?
    - Can some data be considered both qualitative and quantitative in certain situations? (e.g., colour intensity can be described qualitatively or measured quantitatively with instruments).
  + **Assessments:**
    - Formative Quiz (5 mins): Short multiple-choice quiz on defining data, differentiating between qualitative and quantitative data, and classifying data as discrete or continuous (scientific examples). AI-generated quiz.
    - Observation of student participation in sorting activities and class discussions.
  + **Student Workbook:**
    - Definitions of data, qualitative data, quantitative data, discrete data, continuous data (with examples).
    - Sorting table for classifying scientific data examples (Qualitative/Quantitative, Discrete/Continuous).
    - "Data Type Challenge" exercises with scientific scenarios.
  + **Extension Activities:**
    - Students find real-world scientific datasets online and identify examples of qualitative, quantitative, discrete, and continuous data within those datasets.
    - Challenge: Research measurement scales (nominal, ordinal, interval, ratio) and try to classify the data types (qualitative/quantitative, discrete/continuous) according to these scales (optional – more advanced).
* **Lesson Materials:**
  + Slide Deck: Data definition, qualitative vs. quantitative data (definitions, examples), discrete vs. continuous data (definitions, examples), real-world data examples.
  + Handout: Data sorting cards or digital drag-and-drop activity, "Data Type Challenge" scenarios worksheet, definitions of data types.
  + Formative Quiz (printable or digital).

Week 2: Data Collection Mastery & Digital Responsibility

Lesson 2.1: Digital Footprint & Data Ethics (Consolidated Lesson)

* **Title:** Navigating the Digital World: Your Footprint and Ethical Data Use
* **Learning Outcomes:**
  + Students will be able to define and explain the concept of a digital footprint.
  + Students will be able to identify various online activities that contribute to their digital footprint.
  + Students will be able to analyse the potential implications of their digital footprint for privacy and security.
  + Students will understand ethical considerations related to data collection, use, and privacy in the digital age, particularly in scientific contexts.
  + *NSW Curriculum Outcomes:* (Implicitly linked to broader digital literacy, ethical understanding, and responsible data use within SC4-DA1-01 and SC4-WS-07 – focus on responsible digital citizenship).
* **Overview:** This consolidated lesson focuses on digital footprints and data ethics. Students will explore what constitutes a digital footprint, its implications, and the ethical responsibilities associated with data in the digital world, including scientific data. This lesson is designed to be more engaging and thought-provoking for selective school students, encouraging critical reflection on their online presence and ethical data practices.
* **Activities:**
  + **Warm-Up (10 mins):** "Digital Footprint Brainstorm & Scenario" -
    - Quick brainstorm: "What are some things you do online that leave a trace or record behind?" (e.g., social media posts, searches, online shopping, website visits, app usage). Write these on the board.
    - Present a short, thought-provoking scenario: "Imagine a future employer or university admissions officer is looking at your online activity. What kind of picture would your digital footprint paint of you? Are there things you would want to change or be more mindful of?" Brief class discussion.
  + **Main Activity 1: "Deconstructing the Digital Footprint" (25 mins):**
    - Introduce the definition of a digital footprint: passive and active footprints.
    - Categorize brainstormed examples from warm-up into passive and active footprints.
    - Discuss different types of data collected online: personal information, browsing history, location data, social media activity, etc.
    - Explore the potential implications of digital footprints:
      * Privacy concerns and data security risks.
      * Online reputation and long-term consequences.
      * Targeted advertising and filter bubbles.
      * Use of digital footprint data by companies, governments, and other entities.
  + **Main Activity 2: "Ethical Data Dilemmas in Science" (20 mins):**
    - Shift focus to data ethics, particularly in scientific contexts.
    - Present ethical dilemma scenarios related to data science (designed to be more complex and nuanced for selective students):
      * **Scenario 1: Anonymized Medical Data:** Researchers have a large dataset of anonymized patient medical records for studying disease patterns. Is it ethical to use this data without explicit consent, even if anonymized? What are the potential benefits and risks?
      * **Scenario 2: Environmental Monitoring & Surveillance:** A city wants to use AI-powered surveillance cameras to monitor pollution levels in real-time. This data could also be used for general public surveillance. What are the ethical trade-offs between environmental protection and privacy?
      * **Scenario 3: AI Bias in Data Analysis:** An AI algorithm used to predict student success in science is found to be biased against certain demographic groups due to biases in the training data. How should this be addressed? What are the ethical responsibilities of data scientists in mitigating bias?
    - Small group discussions on each scenario, followed by class sharing and debate. Encourage students to consider different perspectives and justify their ethical reasoning.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - What are the key takeaways about managing your digital footprint responsibly?
    - Why are ethical considerations particularly important in data science and scientific research?
    - How can we ensure data is used ethically and responsibly in science and technology?
  + **Assessments:**
    - Formative Task (5 mins): "Digital Footprint Reflection" - Students write a short paragraph reflecting on their own digital footprint and one action they will take to manage it more responsibly.
    - Observation of student participation in discussions and group work, focusing on the depth of their ethical reasoning.
  + **Student Workbook:**
    - Definitions of digital footprint (active/passive).
    - Table to categorize online activities and their footprint types.
    - Summary of ethical considerations in data science.
    - Space for reflection on personal digital footprint and action plan.
  + **Extension Activities:**
    - Research current events or news articles related to data privacy breaches, ethical AI dilemmas, or digital surveillance. Present findings to the class and lead a discussion.
    - Challenge: Design a "Digital Footprint Awareness Campaign" for the school, aimed at educating students about responsible online behaviour and data ethics.
* **Lesson Materials:**
  + Slide Deck: Digital footprint definition, types, implications, ethical data dilemmas scenarios, key ethical principles.
  + Handout: Digital Footprint Brainstorm worksheet, Ethical Dilemma scenarios descriptions, Digital Footprint Reflection prompt.
  + (Optional) Short video clip or news report related to digital privacy or data ethics to spark discussion.

Lesson 2.2: Data Collection Techniques: Primary vs. Secondary Data (Hands-On Design)

* **Title:** Becoming Data Collectors: Designing Scientific Investigations
* **Learning Outcomes:**
  + Students will be able to differentiate between primary and secondary data sources in scientific research.
  + Students will be able to describe various methods of primary data collection (experiments, surveys, observations, simulations).
  + Students will be able to design basic data collection plans for different scientific questions, specifying data types and collection methods.
  + *NSW Curriculum Outcomes:* SC4-WS-02 (Planning investigations), ACSIS125 (Conducting investigations safely), SC4-DA1-01 (Examining data sources).
* **Overview:** This lesson shifts to practical data collection. Students will learn about primary and secondary data and focus on designing their own primary data collection methods. The lesson emphasizes active learning and experimental design skills, tailored for a selective student group.
* **Activities:**
  + **Warm-Up (10 mins):** "Data Source Challenge" - Present a series of scientific questions (designed to be engaging and potentially complex for selective students):
    - "How does caffeine affect reaction time in teenagers?"
    - "Is there a correlation between air pollution levels and respiratory illness in our city?"
    - "What is the biodiversity of insect species in our schoolyard?"
    - "How do different types of exercise affect heart rate recovery time?"  
      For each question, ask students: "Where could scientists get data to answer this question? Would it be primary or secondary data? What are some potential sources?" Class discussion to elicit different data source ideas.
  + **Main Activity 1: "Primary vs. Secondary Data Deep Dive" (20 mins):**
    - Explicitly define Primary Data (collected firsthand for a specific purpose) and Secondary Data (collected by others, for other purposes).
    - Discuss the advantages and disadvantages of each data type in scientific research (e.g., primary - control, relevance, but time-consuming; secondary - readily available, large datasets, but potentially less relevant, quality concerns).
    - Examples of Primary Data Collection Methods: Experiments, Surveys, Observations, Simulations. Briefly describe each method and when it is most appropriate.
  + **Main Activity 2: "Design Your Data Collection Plan" (30 mins):**
    - Divide students into small groups.
    - Assign each group one or two scientific questions from the warm-up or provide new, more complex scientific questions suitable for Year 8 selective students (e.g., related to local environmental issues, human behaviour, simple physics phenomena).
    - Each group task: Design a *primary* data collection plan to investigate their assigned question. The plan must include:
      * **Scientific Question (clearly stated).**
      * **Type of Primary Data Collection Method(s) to be used (Experiment, Survey, Observation, Simulation – or combination). Justify your choice.**
      * **Specific Data to be Collected (what variables will you measure or observe? Be precise).**
      * **Brief outline of the Procedure (how will you collect the data? What steps will you follow? Consider safety and ethical aspects).**
      * **Data Types (will you collect qualitative, quantitative, or both?).**
    - Groups work on their plans. Teacher circulates to provide guidance and feedback, encouraging more rigorous design and consideration of potential challenges.
    - Groups briefly present their data collection plans to the class. Class feedback and discussion – peer review of plans.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - What are the key considerations when designing a good data collection plan for scientific research?
    - What are some potential challenges or limitations you anticipate in carrying out your data collection plans?
    - How does the choice of data collection method influence the type of data you get and the conclusions you can draw?
  + **Assessments:**
    - Formative Task (5 mins): "Data Collection Method Matching" - Quick matching exercise (worksheet or digital) where students match scientific questions to appropriate primary data collection methods.
    - Review of group data collection plans – assess clarity, completeness, and feasibility.
  + **Student Workbook:**
    - Definitions of primary and secondary data with advantages/disadvantages.
    - Descriptions of primary data collection methods (experiments, surveys, observations, simulations).
    - Template/worksheet for designing a data collection plan (with prompts for question, method, data, procedure, data types).
  + **Extension Activities:**
    - For a chosen scientific question, design *multiple* data collection plans using different primary data collection methods and compare the potential data and insights each method could provide.
    - Challenge: Research examples of real-world scientific studies that used specific primary data collection methods (e.g., a famous experiment, a large-scale survey). Analyze why that method was chosen and its strengths and limitations in that context.
* **Lesson Materials:**
  + Slide Deck: Primary vs. secondary data definitions, examples of primary data collection methods, data collection plan design process.
  + Handout: Data Source Challenge questions, Data Collection Plan template/worksheet, Data Collection Method Matching exercise.

Week 3: Data Visualization & Interactive Storytelling

Lesson 3.1: Data Display: Tables - Organising Information

* **Title:** Data in Order: Mastering Tables for Scientific Clarity
* **Learning Outcomes:**
  + Students will be able to construct tables to effectively organise and present scientific data.
  + Students will be able to identify the key components of a well-designed table (title, headings, data cells, units).
  + Students will be able to choose appropriate table formats for different types of scientific data.
  + Students will be able to interpret data presented in tables and draw basic conclusions.
  + *NSW Curriculum Outcomes:* SC4-WS-06 (Process data and information), SC4-WS-07 (Communicate information)
* **Overview:** This lesson focuses on tables as a fundamental method for organising and presenting scientific data. Students will learn the principles of table design, practice constructing tables, and develop skills in interpreting tabular data to extract information and draw conclusions.
* **Activities:**
  + **Warm-Up (10 mins):** "Information Overload Challenge" - Present students with a paragraph of raw scientific data (e.g., a list of plant heights, temperatures, reaction times in text format). Ask them to quickly find specific pieces of information (e.g., "What is the maximum temperature?", "What is the height of plant number 3?"). Discuss how difficult it is to extract information from unstructured text data. Lead to the need for organised data presentation methods like tables.
  + **Main Activity 1: "Anatomy of a Table - Deconstructing Good Design" (25 mins):**
    - Introduce the key components of a scientific table:
      * **Table Number and Title:** Clear, concise, informative title describing the table's content.
      * **Column Headings:** Descriptive headings for each column, indicating the variable being measured and units of measurement (if applicable).
      * **Row Headings (or first column):** Labels for each row, indicating categories or experimental conditions.
      * **Data Cells:** Cells containing the actual data values.
      * **Units of Measurement:** Clearly stated units (e.g., cm, °C, seconds) in column headings or below table if consistent for all values.
    - Show examples of well-designed and poorly designed scientific tables. For the poorly designed examples, ask students to identify what is wrong and how it could be improved. Focus on clarity, organization, and completeness of information.
    - "Table Critique": Provide students with poorly designed tables (pre-prepared or from scientific publications - simplified if necessary). In groups, students critique the tables, identifying weaknesses and suggesting specific improvements to enhance clarity and readability.
  + **Main Activity 2: "Table Construction Challenge - From Raw Data to Organised Table" (20 mins):**
    - Provide students with sets of raw scientific data (different datasets for different groups - e.g., plant growth data, reaction time data, survey results). Data can be provided in text lists or simple spreadsheets.
    - Each group task: Construct a clear and well-designed table to present their assigned dataset. They need to decide on appropriate column and row headings, units, and table title. Encourage them to consider the most effective way to organise the data for clarity and easy interpretation.
    - Groups share their constructed tables with the class. Class feedback and discussion – peer review of table design. Discuss different approaches and best practices for table construction.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - Why are tables such a fundamental and important tool for presenting scientific data?
    - What are the key principles of good table design for scientific communication?
    - When is it most appropriate to use a table to present data, compared to other visualization methods like graphs?
  + **Assessments:**
    - Formative Task (5 mins): "Table Component Identification" - Provide a well-designed table. Ask students to label and identify the key components (title, headings, data cells, units).
    - Review of group-constructed tables – assess clarity, completeness, and adherence to table design principles.
  + **Student Workbook:**
    - Diagram of a well-designed scientific table with labels for key components and descriptions.
    - Checklist of criteria for good table design.
    - Raw data sets for table construction exercises.
    - "Table Critique" worksheet with poorly designed tables and prompts for analysis and improvement suggestions.
  + **Extension Activities:**
    - Students find scientific tables in research papers or reports and analyse their design. Critique their effectiveness and suggest any improvements.
    - Challenge: Explore more advanced table formatting options (e.g., using borders, shading, conditional formatting – in software like spreadsheets or online table generators). Create tables with more complex data structures (e.g., hierarchical tables, tables with multiple levels of headings).
* **Lesson Materials:**
  + Slide Deck: Introduction to tables, components of a well-designed table, examples of good and bad tables, table design principles.
  + Handout: Raw data sets for table construction exercises, "Table Critique" worksheet with poorly designed tables, checklist for good table design.
  + Formative Task (printable or digital).

Lesson 3.2: Visualizing Relationships: Bar Graphs

* **Title:** Bar Graphs: Comparing Categories and Showing Differences
* **Learning Outcomes:**
  + Students will be able to construct bar graphs to visually represent categorical data and compare quantities across different categories.
  + Students will be able to identify the key components of a bar graph (axes, bars, labels, title, scale).
  + Students will be able to choose appropriate bar graph types (vertical, horizontal, grouped) for different datasets and analytical goals.
  + Students will be able to interpret bar graphs and extract meaningful comparisons and conclusions from them.
  + *NSW Curriculum Outcomes:* SC4-WS-06 (Process data and information), SC4-WS-07 (Communicate information), SC4-DA1-01 (Analyse a model to identify data and trends)
* **Overview:** This lesson introduces bar graphs as a powerful visualization tool for comparing categorical data. Students will learn the components of bar graphs, practice constructing them, and develop skills in interpreting bar graphs to identify comparisons and draw conclusions about categorical relationships in scientific data.
* **Activities:**
  + **Warm-Up (10 mins):** "Categorical Data Brainstorm" - Ask students to brainstorm examples of categorical data in science and everyday life (e.g., types of animals, colours of flowers, types of rocks, favourite subjects, survey responses with categories). List these on the board. Discuss: What kind of questions can we answer with categorical data? How can we visually compare categories?
  + **Main Activity 1: "Anatomy of a Bar Graph - Building Blocks of Comparison" (25 mins):**
    - Introduce the key components of a bar graph:
      * **Axes:** Horizontal (x-axis) for categories, Vertical (y-axis) for frequency or quantity. Label axes clearly (category name, quantity name, units if applicable).
      * **Bars:** Rectangular bars representing each category. Bar height or length corresponds to the frequency or quantity for that category. Bars should be visually distinct and have appropriate spacing.
      * **Scale:** Appropriate scale on the y-axis to accommodate the data range, with clear tick marks and labels.
      * **Title:** Clear, concise title describing what the bar graph represents.
      * **Labels:** Category labels along the x-axis, and y-axis labels with units.
    - Show examples of well-designed and poorly designed bar graphs. For poorly designed examples, ask students to identify weaknesses and suggest improvements. Focus on clarity, accurate representation of data, and effective visual comparison of categories.
    - Introduce different types of bar graphs:
      * **Vertical Bar Graph:** Bars extend vertically from the x-axis. Common for comparing categories along the x-axis.
      * **Horizontal Bar Graph:** Bars extend horizontally from the y-axis. Useful when category labels are long or when comparing many categories.
      * **Grouped Bar Graph (Clustered Bar Graph):** Used to compare multiple categories for different groups or conditions. Bars for each group are clustered together for each category.
    - "Bar Graph Type Match": Present scenarios or research questions with categorical data. Students identify which bar graph type (vertical, horizontal, grouped) would be most appropriate and justify their choice.
  + **Main Activity 2: "Bar Graph Construction Challenge - From Data to Visual Comparison" (20 mins):**
    - Provide students with datasets suitable for bar graphs (different datasets for different groups - e.g., species counts in different habitats, survey results on favourite science topics, experimental results comparing different treatments). Data can be in tables or lists.
    - Each group task: Construct a bar graph (using paper and pencil initially, or digital tools if available - spreadsheet software, Observable) to visualize their assigned dataset. They need to choose the appropriate bar graph type, label axes, set scale, and add a title. Encourage them to focus on creating a clear and visually effective bar graph for comparison.
    - Groups share their constructed bar graphs with the class. Class feedback and discussion – peer review of bar graph design and effectiveness in communicating comparisons.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - Why are bar graphs effective for visualizing and comparing categorical data?
    - What are the key elements of a well-designed bar graph for scientific communication?
    - When is it most appropriate to use a bar graph compared to tables or other graph types? What types of questions are best answered with bar graphs?
  + **Assessments:**
    - Formative Task (5 mins): "Bar Graph Component Identification" - Provide a bar graph. Ask students to label and identify key components (axes, bars, labels, title, scale).
    - Review of group-constructed bar graphs – assess clarity, accuracy, and appropriate choice of graph type and design.
  + **Student Workbook:**
    - Diagram of a well-designed bar graph with labels for key components and descriptions.
    - Checklist of criteria for good bar graph design.
    - Data sets for bar graph construction exercises.
    - "Bar Graph Type Match" worksheet with scenarios and prompts for choosing appropriate bar graph types.
  + **Extension Activities:**
    - Students find bar graphs in scientific publications or news articles and analyse their design and effectiveness in communicating data. Critique their strengths and weaknesses.
    - Challenge: Explore variations of bar graphs (e.g., stacked bar graphs, 100% stacked bar graphs) and investigate when these types are most useful. Create examples of these more complex bar graph types using digital tools.
* **Lesson Materials:**
  + Slide Deck: Introduction to bar graphs, components of a bar graph, types of bar graphs (vertical, horizontal, grouped), examples of good and bad bar graphs, bar graph design principles.
  + Handout: Data sets for bar graph construction exercises, "Bar Graph Type Match" worksheet with scenarios, checklist for good bar graph design.
  + Formative Task (printable or digital).
  + Graph paper, rulers, coloured pencils (for paper-based construction) OR access to spreadsheet software, Observable, or online graph makers (for digital construction).

Lesson 3.3: Showing Trends Over Time: Line Graphs

* **Title:** Line Graphs: Revealing Trends and Changes Over Time
* **Learning Outcomes:**
  + Students will be able to construct line graphs to visually represent continuous data and show trends or changes over time or a continuous variable.
  + Students will be able to identify the key components of a line graph (axes, data points, lines, labels, title, scale).
  + Students will be able to choose appropriate scales and line styles for effective line graph construction.
  + Students will be able to interpret line graphs to identify trends, patterns, and relationships in data over time or a continuous variable.
  + *NSW Curriculum Outcomes:* SC4-WS-06 (Process data and information), SC4-WS-07 (Communicate information), SC4-DA1-01 (Analyse a model to identify data and trends)
* **Overview:** This lesson introduces line graphs as the primary tool for visualizing trends and changes in continuous data, particularly over time. Students will learn the components of line graphs, practice constructing them, and develop skills in interpreting line graphs to identify trends, patterns, and relationships in scientific data.
* **Activities:**
  + **Warm-Up (10 mins):** "Trend Spotting Challenge" - Show students several simple line graphs depicting trends (e.g., temperature changes over a day, population growth over years, reaction rate changing with temperature). For each graph, ask: "What trend or pattern do you see in this graph? Is it increasing, decreasing, fluctuating, stable? What does this trend tell you about the data?" Focus on quickly identifying visual trends.
  + **Main Activity 1: "Anatomy of a Line Graph - Connecting the Dots of Change" (25 mins):**
    - Introduce the key components of a line graph:
      * **Axes:** Horizontal (x-axis) usually for time or a continuous independent variable, Vertical (y-axis) for the dependent variable being measured. Label axes clearly (variable names, units).
      * **Data Points:** Plotted points representing data values for each time point or x-value.
      * **Lines:** Lines connecting consecutive data points to show trends. Use straight lines between points. For multiple datasets on the same graph, use different line styles or colours to distinguish them.
      * **Scale:** Appropriate scales for both axes to accommodate data ranges, with clear tick marks and labels. Choose scales that spread out the data effectively and make trends visible.
      * **Title:** Clear, concise title describing what the line graph represents.
      * **Labels:** Axis labels with variable names and units, and potentially a legend if multiple lines are plotted.
    - Show examples of well-designed and poorly designed line graphs. For poorly designed examples, ask students to identify weaknesses and suggest improvements. Focus on clarity, accurate representation of trends, and effective visual communication of change over time.
    - Discuss choosing appropriate scales for line graphs: linear vs. logarithmic (briefly introduce log scale concept for future, focus on linear for Year 8). Discuss how scale choice can affect the visual appearance of trends.
    - Discuss line styles: solid lines, dashed lines, dotted lines for distinguishing multiple datasets.
  + **Main Activity 2: "Line Graph Construction Challenge - Revealing Time-Based Trends" (20 mins):**
    - Provide students with datasets suitable for line graphs (different datasets for different groups - e.g., temperature readings over time, plant growth measurements over weeks, reaction rate data at different temperatures, population data over years). Data can be in tables or lists.
    - Each group task: Construct a line graph (using paper and pencil initially, or digital tools - spreadsheet software, Observable) to visualize their assigned dataset. They need to choose appropriate axes, scales, plot data points, connect lines, label axes, and add a title. Encourage them to focus on creating a clear line graph that effectively shows trends and changes over time.
    - Groups share their constructed line graphs with the class. Class feedback and discussion – peer review of line graph design and effectiveness in communicating trends. Discuss different scale choices and their impact on visual interpretation.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - Why are line graphs the best choice for showing trends and changes over time or a continuous variable?
    - What are the key elements of a well-designed line graph for scientific communication?
    - When is it most appropriate to use a line graph compared to tables or bar graphs? What types of scientific questions are best answered with line graphs?
    - How can misleading scales or poor design distort the trends shown in a line graph?
  + **Assessments:**
    - Formative Task (5 mins): "Line Graph Component Identification" - Provide a line graph. Ask students to label and identify key components (axes, data points, lines, labels, title, scale).
    - Review of group-constructed line graphs – assess clarity, accuracy, appropriate scale choice, and effectiveness in showing trends.
  + **Student Workbook:**
    - Diagram of a well-designed line graph with labels for key components and descriptions.
    - Checklist of criteria for good line graph design.
    - Data sets for line graph construction exercises.
    - "Scale Choice Challenge" - scenarios where students choose appropriate scales and explain their reasoning.
  + **Extension Activities:**
    - Students find line graphs in scientific publications or reports and analyse their design and interpretation. Critique their effectiveness in communicating trends and relationships.
    - Challenge: Explore more advanced line graph techniques (e.g., multiple lines on one graph with legend, shaded areas under lines to represent uncertainty or ranges). Create examples of these more complex line graph types using digital tools. Investigate the concept of trendlines or best-fit lines (briefly introduced conceptually).
* **Lesson Materials:**
  + Slide Deck: Introduction to line graphs, components of a line graph, examples of good and bad line graphs, line graph design principles, scale choice considerations.
  + Handout: Data sets for line graph construction exercises, "Scale Choice Challenge" scenarios worksheet, checklist for good line graph design.
  + Formative Task (printable or digital).
  + Graph paper, rulers, coloured pencils (for paper-based construction) OR access to spreadsheet software, Observable, or online graph makers (for digital construction).

Week 4: Descriptive Statistics & Data Interpretation

Lesson 4.1: Central Tendency: Mean, Median, Mode - Finding the "Average"

* **Title:** Unlocking the Center: Mean, Median, and Mode – Finding the Typical Value
* **Learning Outcomes:**
  + Students will be able to define and calculate the mean, median, and mode for a given dataset.
  + Students will be able to explain the differences between mean, median, and mode and identify situations where each measure is most appropriate.
  + Students will be able to calculate mean, median, and mode using Python in Observable notebooks.
  + Students will be able to interpret the mean, median, and mode in the context of scientific data and draw basic conclusions about central tendency.
  + *NSW Curriculum Outcomes:* SC4-WS-06 (Process data and information), SC4-WS-07 (Communicate information), SC4-DA1-01 (Analyse a model to identify data and trends)
* **Overview:** This lesson introduces the concept of central tendency and the three key measures: mean, median, and mode. Students will learn to calculate each measure, understand their differences, and apply them to scientific data, using Python in Observable for calculations.
* **Activities:**
  + **Warm-Up (10 mins):** "Guess the Average" - Present a few sets of simple numbers (e.g., ages of students in a small group, number of petals on different flowers). Ask students to quickly estimate the "average" value for each set. Discuss: What does "average" mean in everyday language? Are there different ways to think about "average"?
  + **Main Activity 1: "Mean, Median, Mode Definitions and Calculations" (25 mins):**
    - Introduce and define Mean (arithmetic average), Median (middle value when data is ordered), and Mode (most frequent value).
    - Step-by-step calculation examples for each measure using small, simple datasets (done by hand initially).
      * **Mean:** Sum of values divided by the number of values.
      * **Median:** Ordering data and finding the middle value (or average of two middle values for even datasets).
      * **Mode:** Identifying the value that appears most frequently (datasets can have no mode, one mode, or multiple modes).
    - Worksheet practice: Provide students with several small datasets. Students calculate the mean, median, and mode for each dataset by hand. Check answers as a class.
  + **Main Activity 2: "Mean, Median, Mode in Python (Observable)" (20 mins):**
    - Introduce how to calculate mean, median, and mode using Python in Observable notebooks.
    - Guided code-along in Observable:
      * Create a code cell.
      * Create a Python list of numbers representing a scientific dataset (e.g., temperature readings, plant heights).
      * Use numpy.mean(), numpy.median(), and scipy.stats.mode() functions to calculate the measures.
      * Print the results with clear labels (e.g., "Mean Temperature:", mean\_temp).
    - Student practice: Provide students with new datasets in Observable. Students write Python code to calculate and display the mean, median, and mode.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - What are the similarities and differences between mean, median, and mode?
    - When might the mean be the most appropriate measure of central tendency? When might the median be better? When might the mode be useful?
    - How can understanding central tendency help us analyse and interpret scientific data? What kind of information does it give us?
    - Are there any limitations to using just central tendency to describe a dataset? (Lead to the idea of data dispersion in the next lesson).
  + **Assessments:**
    - Formative Quiz (5 mins): Short multiple-choice quiz on definitions of mean, median, mode and their basic calculation. AI-generated quiz.
    - Review of student Python code in Observable notebooks – check for correct calculations and code syntax.
  + **Student Workbook:**
    - Definitions of mean, median, and mode with formulas.
    - Step-by-step calculation examples for each measure.
    - Worksheet with datasets for manual calculation practice.
    - Observable notebook exercises with Python code templates for calculating mean, median, and mode.
    - "Choosing the Right Measure" scenario worksheet with prompts to decide when to use mean, median, or mode.
  + **Extension Activities:**
    - Students research real-world examples where mean, median, and mode are used in different scientific fields (e.g., average temperature, median income, modal species in an ecosystem).
    - Challenge: Investigate the concept of weighted mean and when it is used. Calculate a weighted mean for a given dataset.
* **Lesson Materials:**
  + Slide Deck: Introduction to central tendency, definitions of mean, median, mode, calculation examples, choosing the appropriate measure, Python code examples.
  + Handout: Worksheet with datasets for manual calculation practice, "Choosing the Right Measure" scenario worksheet, Python code templates for Observable exercises.
  + Formative Quiz (printable or digital).
  + Computers/Tablets with internet access for Observable.

Lesson 4.2: Data Dispersion: Range - Understanding Data Spread

* **Title:** Beyond the Average: Range – Measuring the Spread of Data
* **Learning Outcomes:**
  + Students will be able to define and calculate the range of a dataset.
  + Students will be able to explain how range describes the dispersion or spread of data.
  + Students will be able to calculate range using Python in Observable notebooks.
  + Students will be able to interpret the range in the context of scientific data and compare the spread of different datasets.
  + Students will understand the limitations of range as a measure of dispersion and recognize the need for more robust measures (to be introduced later).
  + *NSW Curriculum Outcomes:* SC4-WS-06 (Process data and information), SC4-WS-07 (Communicate information), SC4-DA1-01 (Analyse a model to identify data and trends)
* **Overview:** This lesson introduces the concept of data dispersion, focusing on the range as a simple measure of spread. Students will learn to calculate range, understand its meaning, use Python in Observable for calculations, and recognize its limitations, setting the stage for more advanced measures of dispersion.
* **Activities:**
  + **Warm-Up (10 mins):** "Data Spread Visualisation" - Show two sets of data visually represented (e.g., two line graphs with same mean but different spread, or two sets of bar graphs with different variability). Ask students: "Which dataset is more spread out? How can you describe or measure this 'spread'?" Lead to the idea of data dispersion and the need for measures beyond central tendency.
  + **Main Activity 1: "Range Definition and Calculation" (25 mins):**
    - Introduce and define Range: The difference between the maximum and minimum values in a dataset.
    - Step-by-step calculation examples for range using small datasets (by hand). Example: Dataset: [12, 15, 18, 20, 25]. Range = 25 - 12 = 13.
    - Explain how range describes the total spread of the data – how much the values vary from the lowest to the highest. A larger range indicates greater dispersion.
    - Worksheet practice: Provide students with several datasets. Students calculate the range for each dataset by hand. Check answers as a class.
  + **Main Activity 2: "Range in Python (Observable)" (20 mins):**
    - Introduce how to calculate range using Python in Observable. (Note: NumPy doesn't directly have a 'range' function in the statistical sense, so we'll use max() and min() and subtraction).
    - Guided code-along in Observable:
      * Create a code cell.
      * Create a Python list of numbers representing a scientific dataset.
      * Use numpy.max() to find the maximum value, numpy.min() to find the minimum value, and subtract to calculate the range.
      * Print the result with a clear label (e.g., "Data Range:", data\_range).
    - Student practice: Provide students with new datasets in Observable. Students write Python code to calculate and display the range.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - What does the range tell us about a dataset that central tendency measures (mean, median, mode) don't?
    - Compare datasets with the same mean but different ranges. What does the range reveal in this case?
    - What are the limitations of using range as the *only* measure of data dispersion? (e.g., sensitive to outliers, doesn't describe the distribution within the range).
    - When might range be a useful measure despite its limitations? (e.g., quick overview of spread, in situations where outliers are not a major concern).
  + **Assessments:**
    - Formative Task (5 mins): "Range Calculation Quiz" - Quick calculation of range for a very simple dataset. AI-generated quiz.
    - Review of student Python code in Observable notebooks – check for correct range calculations and code syntax.
  + **Student Workbook:**
    - Definition of range with formula.
    - Step-by-step calculation examples for range.
    - Worksheet with datasets for manual range calculation practice.
    - Observable notebook exercises with Python code templates for calculating range.
    - "Interpreting Range" scenario worksheet with prompts to compare datasets based on their range and central tendency.
  + **Extension Activities:**
    - Students compare the range of different real-world scientific datasets (e.g., temperature ranges in different cities, height ranges of different plant species).
    - Challenge: Investigate and compare range to other measures of dispersion like interquartile range (IQR) and standard deviation (brief conceptual introduction to IQR).
* **Lesson Materials:**  
  \* Slide Deck: Introduction to data dispersion, definition of range, calculation examples, interpreting range, Python code examples, limitations of range.  
  \* Handout: Worksheet with datasets for manual range calculation practice, "Interpreting Range" scenario worksheet, Python code templates for Observable exercises.  
  \* Formative Quiz (printable or digital).  
  \* Computers/Tablets with internet access for Observable.

Lesson 4.3: Outliers: Identifying and Interpreting Unusual Data Points

* **Title:** Data Detectives: Unmasking Outliers – Spotting the Unusual Suspects in Data
* **Learning Outcomes:**
  + Students will be able to define outliers as unusual or extreme values in a dataset.
  + Students will be able to identify outliers visually using graphs (e.g., box plots, scatter plots) and numerically using simple rules (e.g., beyond a certain number of standard deviations from the mean – conceptual).
  + Students will be able to interpret potential causes of outliers (errors, natural variation, interesting anomalies).
  + Students will be able to discuss appropriate strategies for handling outliers in data analysis (investigate, correct, remove with justification, or keep).
  + *NSW Curriculum Outcomes:* SC4-WS-06 (Process data and information), SC4-WS-07 (Communicate information), SC4-DA1-01 (Analyse a model to identify data and trends)
* **Overview:** This lesson introduces outliers as unusual data points and equips students with methods to identify and interpret them. Students will learn to spot outliers visually and numerically (conceptually), understand their potential causes, and discuss responsible strategies for handling outliers in scientific data analysis.
* **Activities:**
  + **Warm-Up (10 mins):** "Spot the Odd One Out" - Present a series of lists of numbers or objects where one item is clearly an outlier (e.g., ages of students in a class with one value of 90, heights of plants with one value much taller than others). Ask students: "In each list, which value seems unusual or doesn't fit with the others? Why do you think it's unusual?" Lead to the concept of outliers as data points that are different from the rest.
  + **Main Activity 1: "What are Outliers? Definition and Visual Identification" (25 mins):**
    - Define Outliers: Data points that are significantly different from other values in a dataset. They lie far away from the main cluster of data.
    - Discuss potential causes of outliers:
      * **Errors in data collection or measurement:** Mistakes in recording data, faulty equipment.
      * **Natural variation:** Genuine extreme values within a distribution (e.g., exceptionally tall person in a height dataset).
      * **Interesting anomalies or new discoveries:** Outliers might represent something significant or unexpected that needs further investigation (e.g., unexpected experimental result, indication of a new phenomenon).
    - Visual identification of outliers using graphs:
      * **Box Plots:** Show how outliers are represented as points beyond the whiskers of a box plot.
      * **Scatter Plots:** Show how outliers are points far away from the main cluster in a scatter plot.
      * Demonstrate creating simple box plots in Observable (using Plot library - basic example).
    - "Outlier Spotting Challenge (Visual)" - Present students with graphs (box plots, scatter plots) showing datasets with outliers. Students visually identify the outliers and describe their location on the graph.
  + **Main Activity 2: "Numerical Outlier Identification (Conceptual)" (20 mins):**
    - Introduce the concept of numerical outlier detection (without going into complex formulas at Year 8 level).
    - Explain the idea of using standard deviation (conceptually - "how spread out the data is") or IQR (Interquartile Range - if introduced in previous extension activity) to define outlier boundaries.
    - Simplified rule example: "Values that are much further than the typical spread from the mean or median might be outliers." (Avoid precise formulas at this stage).
    - "Outlier Spotting Challenge (Numerical - Simplified Rule)" - Provide datasets and a simplified rule (e.g., "values more than 2 times the range away from the median are potential outliers"). Students apply the rule to identify potential outliers. Discuss that these are just guidelines, not definitive rules.
  + **Reflection & Discussion (10 mins):** Class discussion:
    - Why is it important to identify outliers in scientific data?
    - What are some possible reasons for outliers to occur?
    - What should scientists do when they find outliers in their data? (Discuss different approaches - investigate, correct, remove, keep - and the importance of justification and transparency).
    - Is it always "bad" to have outliers? Can outliers sometimes be important or interesting?
  + **Assessments:**
    - Formative Task (5 mins): "Outlier Definition Check" - Short answer question defining outliers and listing possible causes. AI-generated question.
    - Observation of student participation in outlier spotting activities and discussions.
  + **Student Workbook:**
    - Definition of outliers and potential causes.
    - Examples of outliers in different types of graphs (box plots, scatter plots).
    - "Outlier Spotting Challenge (Visual)" worksheet with graphs.
    - Simplified rule for numerical outlier identification (conceptual).
    - "Outlier Spotting Challenge (Numerical - Simplified Rule)" worksheet.
    - "Handling Outliers" scenario worksheet with prompts to discuss appropriate actions for different outlier situations.
  + **Extension Activities:**
    - Students research real-world examples of significant scientific discoveries that were initially considered outliers or anomalies (e.g., discovery of penicillin, unexpected astronomical observations).
    - Challenge: Investigate more formal methods for numerical outlier detection (e.g., using z-scores, IQR rule – more detailed explanation of IQR). Implement a simple outlier detection function in Python in Observable using one of these methods (if appropriate for Year 8 level).
* **Lesson Materials:**  
  \* Slide Deck: Introduction to outliers, definition, causes, visual outlier identification (box plots, scatter plots), numerical outlier identification (conceptual), handling outliers, importance of outliers.  
  \* Handout: "Outlier Spotting Challenge (Visual)" worksheet with graphs, "Outlier Spotting Challenge (Numerical - Simplified Rule)" worksheet, "Handling Outliers" scenario worksheet, definitions of outliers and causes.  
  \* Formative Quiz (printable or digital).  
  \* Computers/Tablets with internet access for Observable.

Week 5: Formulating Scientific Questions & Designing Investigations - Advanced

* **Lesson 5.1:** Asking Testable Questions: From Observation to Inquiry
  + Focus: Transforming observations into testable scientific questions. Characteristics of good scientific questions (testable, specific, focused). Different types of scientific questions (descriptive, comparative, correlational, causal - conceptually).
  + Activities: Observation exercises, question formulation practice for different scenarios, classifying question types, refining broad questions into testable ones.
* **Lesson 5.2:** Experimental Design: Variables, Controls, and Groups
  + Focus: Key elements of experimental design: Independent and dependent variables, control variables, control groups, experimental groups. Importance of controls and variables.
  + Activities: Identifying variables in experimental scenarios, designing experiments to test specific questions, analysing flawed experimental designs and suggesting improvements to controls and variable manipulation.
* **Lesson 5.3:** Designing Virtual Experiments in Observable
  + Focus: Using Observable notebooks to simulate simple experiments. Setting up variables, simulating data collection, and analysing results in Observable.
  + Activities: Guided virtual experiment design in Observable (teacher-provided template or example), student-designed virtual experiments to test simple hypotheses, data analysis and interpretation of simulated data in Observable.

Week 6: Advanced Data Handling (Cleaning, Sorting, Filtering)

* **Lesson 6.1:** Real-World Data Challenges: "Messy" Datasets and Data Quality Issues
  + Focus: Introduction to the realities of real-world data: missing values, inconsistent formats, errors, biases. Importance of data cleaning and quality control.
  + Activities: Exploration of "messy" datasets (provided examples), identification of data quality issues, discussion of common data cleaning challenges and strategies.
* **Lesson 6.2:** Data Cleaning Techniques in Python (Pandas - Basic)
  + Focus: Hands-on introduction to basic data cleaning techniques in Python using Pandas (within Observable). Handling missing values (removing rows/columns, imputation - briefly conceptually), standardizing formats, basic data validation.
  + Activities: Code-along data cleaning exercises in Observable using Pandas on provided "messy" datasets, error analysis and debugging of data cleaning code.
* **Lesson 6.3:** Data Sorting and Filtering for Focused Analysis
  + Focus: Techniques for sorting and filtering data to focus analysis on specific subsets or patterns. Using Python/Pandas for sorting and filtering in Observable.
  + Activities: Data sorting and filtering exercises in Observable using Python/Pandas on provided datasets, applying sorting and filtering to answer specific research questions, data exploration through sorting and filtering.

Week 7: AI-Powered Data Analysis & Predictive Modelling (Introduction)

* **Lesson 7.1:** Introduction to AI in Data Analysis: Pattern Recognition and Insights
  + Focus: Conceptual introduction to AI and machine learning in data analysis. Pattern recognition, clustering, classification (conceptual overview, no complex algorithms). AI for data summarization and trend analysis.
  + Activities: Examples of AI applications in data analysis (case studies), exploring simplified visual AI tools for pattern detection (if available), discussion of the potential and limitations of AI in scientific data analysis.
* **Lesson 7.2:** Simplified Predictive Modelling (Conceptual Introduction - Visual Tools if feasible)
  + Focus: Basic concepts of predictive modelling: training data, features, target variable, model building (simplified explanation), prediction, model evaluation (conceptually). If possible, introduce a very simplified, visual predictive modelling tool (e.g., drag-and-drop interface).
  + Activities: Conceptual explanation of predictive modelling steps, using a simplified visual tool for basic predictive modelling tasks (if feasible), exploring example predictive models, discussion of the ethical implications of predictive models.
* **Lesson 7.3:** Ethical Considerations of AI in Data Science
  + Focus: Ethical implications of using AI in data science: Bias in AI, data privacy concerns, transparency and explainability of AI models, responsible AI development and deployment.
  + Activities: Ethical dilemma scenarios related to AI in data science (case studies), group discussions on ethical considerations, brainstorming strategies for mitigating bias and promoting ethical AI practices.

Week 8: Group Project - Real-World Data Science Challenge

* **Lesson 8.1:** Project Launch: Defining Problems and Choosing Datasets
  + Focus: Project introduction, team formation, brainstorming real-world scientific/societal problems suitable for data science investigation, dataset identification and selection strategies.
  + Activities: Project overview presentation, team formation activity, group brainstorming of project ideas, dataset research and selection activity (students begin to find or are provided with potential datasets), project proposal outline.
* **Lesson 8.2:** Data Exploration and Question Formulation for Group Projects
  + Focus: Group work session focused on exploring chosen datasets in Observable, initial data visualization and descriptive statistics, refining project questions based on dataset exploration, developing a specific project plan and timeline.
  + Activities: Group work time in class, teacher guidance and consultation, dataset exploration in Observable notebooks, preliminary data analysis, project question refinement, project plan development (deliverables, timeline, roles).
* **Lesson 8.3:** Data Analysis and Visualization - Project Work Session
  + Focus: Group work time dedicated to data analysis and visualization for projects. Students work on data cleaning, applying appropriate statistical methods, creating relevant visualizations in Observable to address their project questions.
  + Activities: Group work time in class, teacher guidance and support for data analysis and visualization techniques, peer consultation and collaboration, progress checkpoint and feedback.

Week 9: Project Refinement & Advanced Review

* **Lesson 9.1:** Project Presentations - Peer Feedback and Iteration
  + Focus: Group project presentations to the class (interim presentations or final presentations depending on project complexity). Peer feedback sessions to provide constructive criticism and suggestions for improvement.
  + Activities: Group project presentations (brief presentations to allow time for feedback), peer feedback activity (structured feedback forms or guidelines), group reflection on feedback and planning for project refinement.
* **Lesson 9.2:** AI-Driven Project Review and Refinement
  + Focus: Utilizing AI tools for automated feedback on project notebooks, code quality, data visualizations, and analysis (if feasible - depending on AI tool capabilities). Group work time to address AI and peer feedback and refine projects.
  + Activities: AI-driven feedback session (if tool available), group work time to review AI feedback and peer feedback, project notebook and presentation refinement, teacher consultation and guidance.
* **Lesson 9.3:** Exam Preparation and Advanced Data Analysis Practice
  + Focus: Exam review session covering key concepts and skills from the course. Practice with advanced data analysis problems and exam-style questions.
  + Activities: Exam review session (teacher-led or AI-assisted review materials), practice data analysis problems (more complex datasets and questions), exam-style question practice, Q&A session, review of key formulas and concepts.

Week 10: End-of-Semester Exam & Future Pathways

* **Lesson 10.1:** End-of-Semester Exam (Part 1 - Data Interpretation and Short Answer)
  + Focus: Section 1 of the end-of-semester exam: Data Interpretation (graphs, tables, statistics) and Short Answer questions.
  + Activities: Supervised exam session (Part 1).
* **Lesson 10.2:** End-of-Semester Exam (Part 2 - Coding and Notebook Tasks)
  + Focus: Section 2 of the end-of-semester exam: Coding and Notebook Tasks in Observable (basic Python operations, data analysis tasks).
  + Activities: Supervised exam session (Part 2).
* **Lesson 10.3:** Future Pathways in Data Science & Course Reflection
  + Focus: Post-exam discussion and reflection. Exploring future pathways in data science and related STEM fields. Course reflection and student self-evaluation.
  + Activities: Post-exam debrief and Q&A, discussion of career paths in data science, further study options, resources for continued learning, student self-reflection activity on course learning and skill development, course wrap-up and feedback.