CT Academy | Data Labs

We will work with the data described in the *Orientation and Introduction* slide deck. Work will be conducted in teams; there should be at least one participant with some experience with analytical tools in each team. Identifying the various members' roles early will prove beneficial. There is no right or wrong answer/approach to these labs. We provide suggestions for each one; participants are encouraged to stray outside the guidelines as often as they need/want to.

Starting Point: prior to working with the data, develop a list of questions of interest about the scenario.

Lab 1: Data Awareness

Lab Description: this lab is designed to help participants explore how different datasets relate to one another and to encourage them to think about the trustworthiness of the data they use on a daily basis.

Lab Activities

- Create a "data dictionary" to explain the different fields and variables in the provided datasets.
- Identify some of the variables and measurements – are they all equally trustworthy?
- Discuss the criteria used to determine the trustworthiness of data.
- Are the datasets are structured or unstructured. Justify the reasoning behind your conclusions.

Learning Outcomes

- **Categorization Skills:** participants will develop an understanding of the various categories of data, enhancing their ability to navigate and interpret datasets effectively.
- **Critical Evaluation:** learners will cultivate the ability to critically assess the reliability of data sources, fostering a quantitative approach to evaluating data trustworthiness.
- **Structural Insights:** gain proficiency in identifying and describing the structure of datasets, including the importance and implications of structured versus unstructured data for analysis.

Lab 2: Data Ethics

Lab Description: this lab illustrates ethical considerations when using, analyzing, and presenting data.

Lab Activities

- Examine the datasets to uncover potential ethical issues in data usage, analysis, and presentation.
- Assess the potential impacts of identified ethical problems on stakeholders and society.
- Propose strategies and solutions to address and resolve these ethical dilemmas.

Learning Outcomes

- **Ethical Principles Application:** develop the ability to apply ethical principles systematically to scenarios involving data, ensuring responsible use and analysis.
- **Impact Awareness:** enhance understanding of the significant impacts that ethical lapses can have, both in terms of personal data privacy and societal implications.
- **Risk Mitigation Strategies:** acquire practical strategies for identifying, assessing, and mitigating ethical risks in data-driven projects.

Lab 3: Data Governance

Lab Description: this lab focuses on the foundational aspects of data governance, guiding participants through the identification and differentiation of reference data and master data within provided datasets.

Lab Activities

- Review the provided datasets to distinguish what you consider reference data versus master data.
- Reflect on these distinctions and generate examples of both reference data and master data from your own organization.
- Explore and identify potential strategies for managing both master and reference data within your organization.

Learning Outcomes

- **Data Typology Comprehension:** gain a solid understanding of the differences between reference data and master data, including their roles and importance in data management.
- **Organizational Application:** develop the ability to apply concepts of reference and master data to one's own organizational context, enhancing relevance and practical understanding.
- **Data Management Strategies:** acquire insights into effective data governance practices, focusing on the management of master and reference data to support organizational data integrity and quality.

Lab 4: Data Collection

Lab Description: this lab focuses on the methodologies behind data collection, challenging participants to analyze the provided data for clues on its collection process.

Lab Activities

- Analyze the provided datasets to infer the data collection methods used.
- Evaluate whether the collection method could influence data quality and the results of data analysis.
- Assess the potential impacts of the collection method on data integrity and analytical accuracy.
- Discuss the necessity of collecting new data to address your initial research questions or hypotheses.

Learning Outcomes

- **Collection Method Analysis:** develop the skill to infer data collection methods from existing datasets and understand their implications on data quality.
- **Impact Evaluation:** learn to critically assess how different collection methods might affect data integrity and the validity of analytical results.
- **Adaptive Data Strategy:** acquire the ability to determine when new data collection is necessary to address specific questions or improve data analysis outcomes, enhancing research design and execution.

Lab 5: Data Quality

Lab Description: this lab requires participants to critically assess the quality of provided datasets, encouraging a thorough examination of trustworthiness, identification of data anomalies, and exploration of underlying causes.

Lab Activities

- Assess the trustworthiness of the datasets and provide justification for your evaluation.
- Identify and flag instances of invalid entries, anomalous observations, missing values, or outliers within the data.
- Investigate and articulate potential root causes for the observed data quality issues.
- Propose short-term and long-term corrective actions to address and mitigate these issues.

Learning Outcomes

- **Quality Assessment Skills:** enhance the ability to evaluate the trustworthiness of data, supporting judgments with clear, evidence-based reasoning.
- **Issue Identification and Analysis:** master the skill of detecting and diagnosing data anomalies, including invalid data, outliers, and missing values, along with understanding their potential impacts.
- **Solution Formulation:** learn to develop actionable short- and long-term strategies to improve data quality, emphasizing preventive measures and corrective actions for identified issues.

Lab 6: Asking Questions

Lab Description: building on prior engagement with the data, this lab is designed to refine participants' inquiry process and project planning skills.

Lab Activities

- Develop a refined list of research questions based on your current understanding of the data and the scenario it represents.
- Compare your refined questions with your initial list to identify any overlap or shifts in focus.
- Evaluate the data for limitations that may impact the feasibility of answering certain questions.
- Outline what the deliverables for a project based on these questions might look like.
- Discuss possible ways to quantify or qualify success for this project.
- Identify the skillsets that would be necessary to successfully carry out this project.

Learning Outcomes

- **Strategic Question Refinement:** gain proficiency in refining research questions based on preliminary data analysis, ensuring they are aligned with data insights and project goals.
- **Data Limitation Assessment:** learn to critically assess the limitations of data in answering research questions, fostering realistic expectations and strategic adaptability.
- **Project Planning Skills:** develop the ability to conceptualize project deliverables and success metrics, enhancing project design and evaluation strategies.
- **Interdisciplinary Team Building:** understand the importance of diverse skill sets in project success, identifying the roles and expertise required to address complex research questions effectively.

Lab 7: Data Analytics

Lab Description: this lab provides participants with an in-depth exploration of the data analytics pipeline, from conceptual modeling to the practical aspects of software usage, data storage, and infrastructure requirements.

Lab Activities

- Outline the steps involved in a typical data analysis pipeline, from data collection to actionable insights.
- Identify and evaluate the software and analytical methods suitable for different types of data analysis.
- Develop a conceptual model to represent the situation or phenomena described by the data.
- Explore how data is stored, accessed, and managed in an analytics environment.
- Assess the potential infrastructure requirements necessary to support the analytics tools and methods selected.

Learning Outcomes

- **Analytics Pipeline Proficiency:** develop an understanding of the entire data analysis pipeline, including the sequential steps and processes involved in turning raw data into meaningful insights.
- **Software and Methods Selection:** gain the ability to select appropriate software and analytical methods based on the data characteristics and analysis objectives, enhancing analytical flexibility and effectiveness.
- **Conceptual Modeling Skills:** learn to construct conceptual models that accurately represent the data's underlying situations or phenomena, facilitating a deeper understanding of the analysis context.
- **Infrastructure Knowledge:** acquire insights into the infrastructure requirements for data analytics, including data storage and access considerations, to ensure efficient and scalable analysis projects.

Lab 8: Data Analysis

Lab Description: this lab advances participants' exploration and understanding of the dataset, pushing beyond initial assessments to conduct more sophisticated analyses.

Lab Activities

- Conduct an evaluation of the datasets to determine overall soundness, noting strengths and weaknesses.
- Design and run at least 2 sophisticated analyses that go beyond basic descriptive statistics or visualizations.
- Discuss methodologies for validating the results of these analyses to ensure their reliability and accuracy.
- Reflect on whether the results of these analyses address any of the initial questions or shed new light on the situation described by the data.

Learning Outcomes

- **Critical Dataset Evaluation:** enhance the ability to critically assess the quality of a dataset, identifying its strengths and areas for improvement to inform the analytical approach.
- **Advanced Analytical Techniques:** gain proficiency in designing and conducting sophisticated data analyses, leveraging advanced statistical methods or analytical techniques to extract deeper insights.
- **Results Validation:** learn strategies for validating analytical results, ensuring their reliability and relevance to the research questions or business scenarios.
- **Insightful Analysis Interpretation:** develop the skill to interpret the results of complex analyses in the context of the original research questions, evaluating how well the analysis addresses these questions or uncovers new perspectives.

Lab 9: Storytelling and Visualization

Lab Description: this lab is centered around data visualization, guiding participants through the creation of both exploratory and explanatory visualizations.

Lab Activities

- Create at least five univariate or bivariate visualizations to deepen your understanding of the dataset.
- Develop at least three "definitive" visualizations that effectively communicate key insights from the dataset, incorporating principles such as documentation, legends, annotations, and the Multiple I's. Aim for both content clarity and aesthetic presentation.
- Find a scenario where a dashboard would be particularly useful. Identify key questions the dashboard aims to answer or insights it seeks to provide. Determine the data sources and elements needed for the dashboard. Sketch a preliminary design of the dashboard, including mock charts, and evaluate its strengths, limitations, functionality, and design elegance.

Learning Outcomes

- **Advanced Visualization Techniques:** master the creation of both exploratory and explanatory visualizations, employing advanced design principles to enhance data interpretation and communication.
- **Principled Design Application:** apply comprehensive design principles to data visualization, ensuring that visual representations are not only informative but also engaging and accessible.
- **Dashboard Conceptualization and Design:** develop the ability to conceptualize and design functional and elegant dashboards tailored to specific analytical scenarios, incorporating storytelling elements to improve data presentation and decision-making.
- **Critical Design Evaluation:** cultivate the skill to critically assess the effectiveness of visualizations and dashboards, considering their functionality, user experience, and the ability to convey complex information succinctly.

Lab 10: Evidence-Informed Decision-Making

Lab Description: this lab is designed to help participants evaluate the adequacy of the data in supporting potential decisions.

Lab Activities

- Assess the provided datasets to determine if the data is sufficient to justify making certain decisions.
- Identify three potential decisions that could be made based on the current data and, for each, specify what additional information or evidence would be required to support those decisions confidently.

Learning Outcomes

- **Critical Data Assessment:** develop the ability to critically evaluate the sufficiency and relevance of data in supporting decision-making processes, enhancing analytical rigor.
- **Identification of Evidence Gaps:** gain skills in identifying gaps in the available data and articulating the specific types of additional evidence needed to fill those gaps, facilitating more informed decision-making.
- **Strategic Decision Planning:** learn to strategically plan decisions by integrating both current evidence and the identification of necessary additional data, ensuring decisions are well-grounded and evidence-informed.

Lab 11: Evaluating Outcomes

Lab Description: this lab introduces participants to the logic model framework, a structured method for planning and evaluating initiatives.

Lab Activities

- Construct a logic model for an initiative related to the datasets and scenarios we have been working on, mapping out the initiative's inputs, activities, outputs, outcomes, and impacts.
- Concentrate on detailing the outputs, outcomes, and impacts, articulating the distinctions and connections among these elements.
- Identify and list the types of data required to evaluate the specified outputs, outcomes, and impacts effectively.

Learning Outcomes

- **Logic Model Construction:** master the ability to develop a logic model for initiatives, providing a clear framework for planning and evaluation.
- **Outcome Focus:** learn to focus on and distinguish between outputs, outcomes, and impacts within the context of an initiative, enhancing strategic planning and evaluation skills.
- **Data Identification for Evaluation:** acquire the skills to identify specific data needed to evaluate the effectiveness of initiatives' outputs, outcomes, and impacts, promoting evidence-based assessment.