From Hesitation to Integration:

Navigating AI in Graduate and Undergraduate Computing and Data Science Programs

TIM KRAUSE, UNIVERSITY OF WISCONSIN – STEVENS POINT

MARIE-CLAIRE KOISSI-KOUASSI, UNIVERSITY OF WISCONSIN – EAU CLAIRE

An Undergraduate Example

APC – AI Learning Outcomes

Integrate Artificial Intelligence (AI) tools and principles while assessing their capabilities, limitations, and ethical implications

APC – AI Learning Outcomes

You will be able to:

- Evaluate the capabilities, limitations, and appropriate use cases of AI systems to inform technology decisions and solution design across IT contexts
- **Apply AI tools and technologies to solve business problems, including the use of generative AI for documentation, system analysis and design, project management, and for making decisions**
- Analyze machine learning approaches, including supervised, unsupervised, and reinforcement learning to determine their suitability for specific business and technical requirements
- Assess ethical implications and risks of AI systems, including bias, privacy, security vulnerabilities, transparency concerns, and recommend responsible AI practices

APC 330 – Professional IT Communication

- Lead the course and assignments with AI
  - AI as a communication productivity tool
  - Guidance for AI on individual assignments
  - Rubrics that account for AI usage and reflection

AI Guidance

- Empathy: understand and share the feelings of others
- Creativity: New ideas, concepts and approaches
- Intuition: Experienced-based judgments and decisions

- Human validation

AI Guidance – Tailored to Context

AI Guidance for This Assignment

This is a reflection on your own experience, so the writing needs to be your own. You might still choose to use AI (e.g., ChatGPT, the Editor tool in MS Word) to assist you with editing, revising, and any grammatical concerns.

AI Guidance for This Assignment

This first project step is a brainstorming activity. A core principle of systems thinking is shifting perspective and becoming aware of our own filters and lenses for viewing solutions. Begin by bringing your own ideas to the table before looking outside your own expertise or the organization.

If you use general web search strategies, AI-generated search summaries may be unavoidable and are acceptable. For this step, this is the limit on generative AI usage. Do not use generative AI to brainstorm or write your recommendations, idea log, talking points, or recording.

AI-Proofing

- Authentic
- Complex
- Requires Judgement and Human Element
- Tested and Revised with Generative AI Models (claude.ai)

AI – Proofing – First Iteration



AI Markers

- Delivering something else (that's close)
- Over-delivering
  - Design
  - Substance
- Easy markers:
  - Extra em-dashes and correct but unexpected punctuation
  - Tone (audience)

AI-Proofing Reflected in Rubrics

How did you use AI?	**5 pts** **Full Marks** Includes an example of both effective and less effective strategies. Includes considerations for future usage.	**3 pts** **Meets** Descriptive of current work, but may not fully anticipate implications going forward.	**0 pts** **No Marks** Missing or significantly off the mark.	5 pts
				Total Points: 25

From Hesitation to Integration:

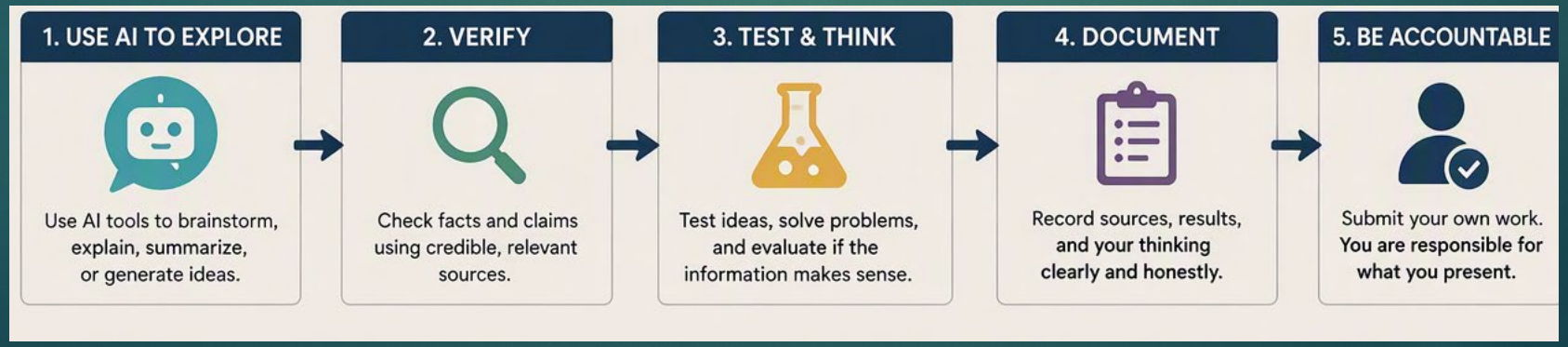
Navigating AI in Graduate and Undergraduate Computing and Data Science Programs

TIM KRAUSE, UNIVERSITY OF WISCONSIN – STEVENS POINT

MARIE-CLAIRE KOISSI-KOUASSI, UNIVERSITY OF WISCONSIN – EAU CLAIRE

Data Science
Responsible AI-assisted coding

Teach students to use AI tools while requiring them to verify, test, document, and remain accountable for their work



Data Science: AI-integration Learning Outcomes

- Statistical reasoning, methodological validation, and inference
- **Programming without AI assistance**
- **AI-assisted Programming and Coding (Pilot AI tools)**
- **Critical evaluation of AI-generated output**
- Documentation of AI-assisted workflows
- Ethical reasoning, privacy, and professional responsibility
- Communication, Interpretation, and decision support

Data Science:
AI-integration Learning Outcomes

1 Statistical Reasoning, Methodological Validation, and Inference



Students learn to:

- Formulate questions
- Choose appropriate methods
- Validate assumptions and methods
- Draw and communicate evidence-based inferences

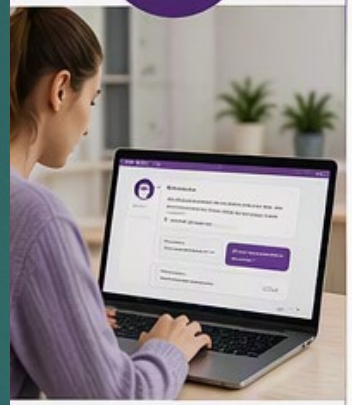
2 Programming without AI Assistance



Students learn to:

- Write clean, efficient, and well-structured code
- Debug and test independently
- Understand algorithms and data structures
- Build confidence in core programming skills

3 AI-Assisted Programming and Coding



Students learn to:

- Use AI tools to prototype, explain, and accelerate development
- Write better prompts and iterate effectively
- Integrate AI suggestions into their work responsibly

4 Critical Evaluation of AI-Generated Output



Students learn to:

- Assess accuracy, logic, and completeness
- Test with examples and edge cases
- Cross-check with credible sources
- Identify bias, limitations, and hallucinations

5 Documentation of AI-Assisted Workflows



Students learn to:

- Keep clear records of tools, prompts, and outputs
- Document validation steps and decisions
- Ensure transparency, reproducibility, and academic integrity

6 Ethical Reasoning, Privacy, and Professional Responsibility



Students learn to:

- Protect privacy and data
- Address bias and fairness
- Understand limits and risks of AI
- Act with integrity and professional accountability

7 Communication and Decision Support



Students learn to:

- Communicate methods, results, and limitations clearly
- Translate insights into actionable recommendation
- Support decisions with credible evidence

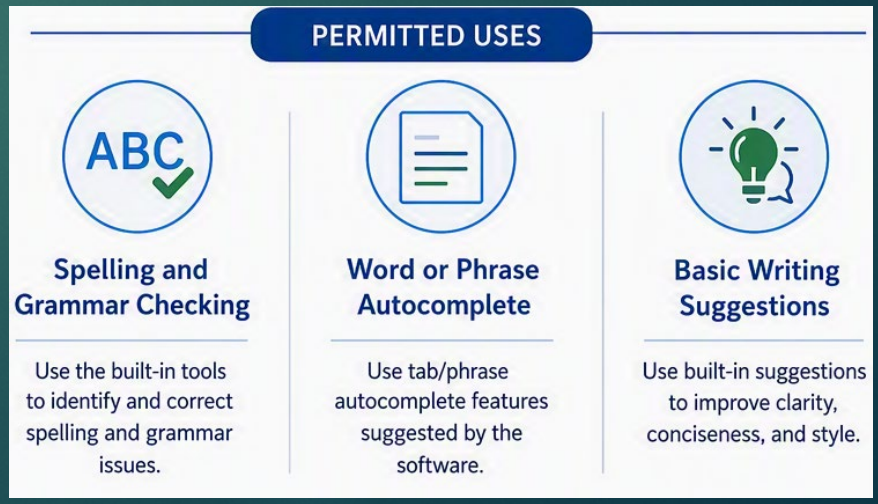
Data Science: Responsible AI-assisted coding Framework

The Baselines

Baseline A:
AI may be used to aid in understanding **instructional material** (such as notes, videos, readings, course discussion, and student help hours). AI may not be used for assignments, unless explicitly stated.

Baseline B: Assignment requirements (instructions/guidelines) should be taken as top priority, as they govern or constrain how AI should (or should not) be used in that assignment and/or how it should be verified or disclosed. Students must adhere to assignment requirements first and foremost.

Baseline C: Incorporated AI tools are allowed
Students are allowed to use standard word processing software, such as Microsoft Word, that incorporates an LLM (Large Language Model) in its spelling/grammar checker/tab autocompletion of phrases.



Data Science: Responsible AI-assisted coding Framework

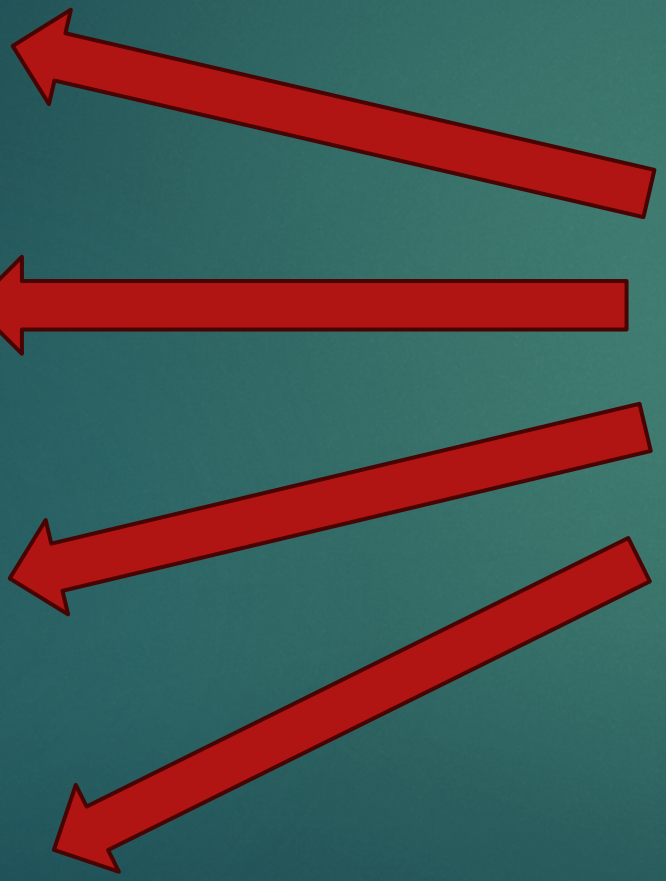
The Levels

AI-No

AI-Disclose

AI-Refine

AI-Yes



Baseline B:

Assignment requirements (instructions/guidelines) should be taken as top priority,

Data Science: Responsible AI-assisted coding Framework

The Levels (Note: Summarize main differences here)

- **AI-No**: AI Disallowed beyond what is specified in the baselines.
  - AI should not be used to produce submitted work; work should be completed independently to isolate learning of a foundational skill
  - Disclosure is not required because AI should not have been used in any part of completing the work.

- **AI-Disclose**: Collaborate and Adjudicate
  - AI may be used to assist completing the work
  - You must test and verify the results; though AI assisted, you are responsible for the results.
  - For your submission, in addition to the results, you must disclose how AI was used and how you verified the results.

[Student must attach first draft. Suggestion: prompts requested.]

- **AI-Yes:** AI to Assist
  - AI may be used to assist completing the work
  - You must test and verify the results; though AI assisted, you are responsible for the results.
  - For your submission, disclosure and verification explanations are not required. But students should cite which AI (name and version number) they used, if any.

Data Science: Responsible AI-assisted coding Framework

The Levels (Note: Briefly summarize main differences here)

- **AI-Disclose**: Collaborate and Adjudicate
  - AI may be used to assist completing the work
  - You must test and verify the results; though AI assisted, you are responsible for the results.
  - **For your submission, in addition to the results, you must disclose how AI was used and how you verified the results.[Student must attach first draft. Suggestion: prompts requested.]**

- **AI-Yes:** AI to Assist
  - AI may be used to assist completing the work
  - You must test and verify the results; though AI assisted, you are responsible for the results.
  - **For your submission, disclosure and verification explanations are not required. But students should cite which AI (name and version number) they used, if any**.

- **AI-Refine**: AI Disallowed to Produce but Allowed to Refine
  - AI should *not* be used to *produce* submitted work; work should be completed independently to isolate learning of a foundational skill
  - AI *may* be used to assist in *refining* work, e.g., to help find and explain errors in your work.
  - For your submission, you must disclose how AI was used to refine your work

[This level may be needed for the **capstone course,** which involves lot of re-writings]

Data Science: Responsible AI-assisted coding
Instructors Work: Modifying / Creating Assignments

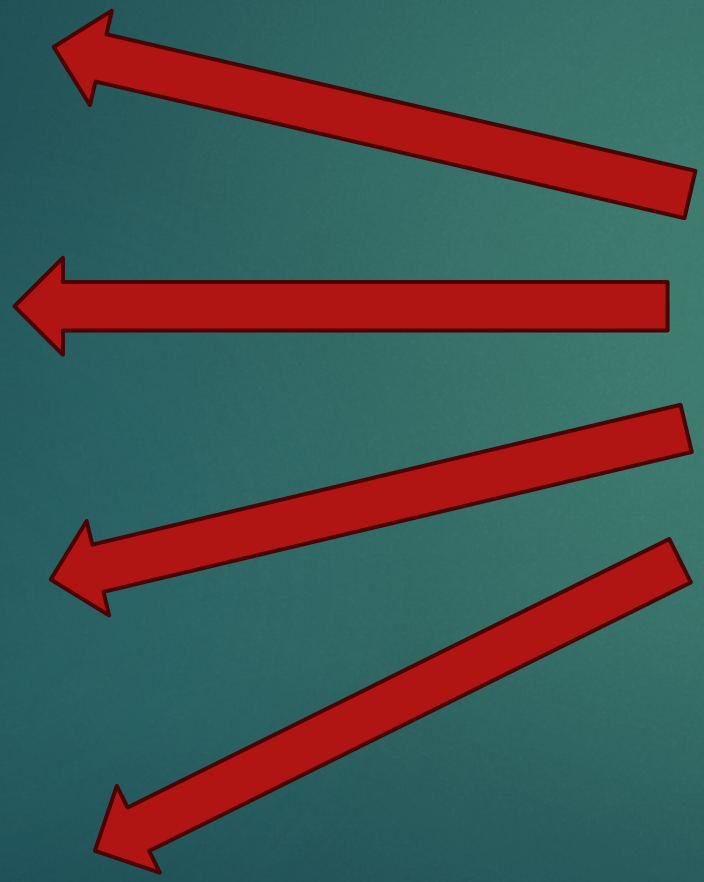
The Levels

AI-No

AI-Disclose

AI-Refine

AI-Yes



Assignments ?

[Modifying old / Creating new]

Data Science: Responsible AI-assisted coding
Instructors Work: Modifying / Creating Assignments

The Levels

AI-No

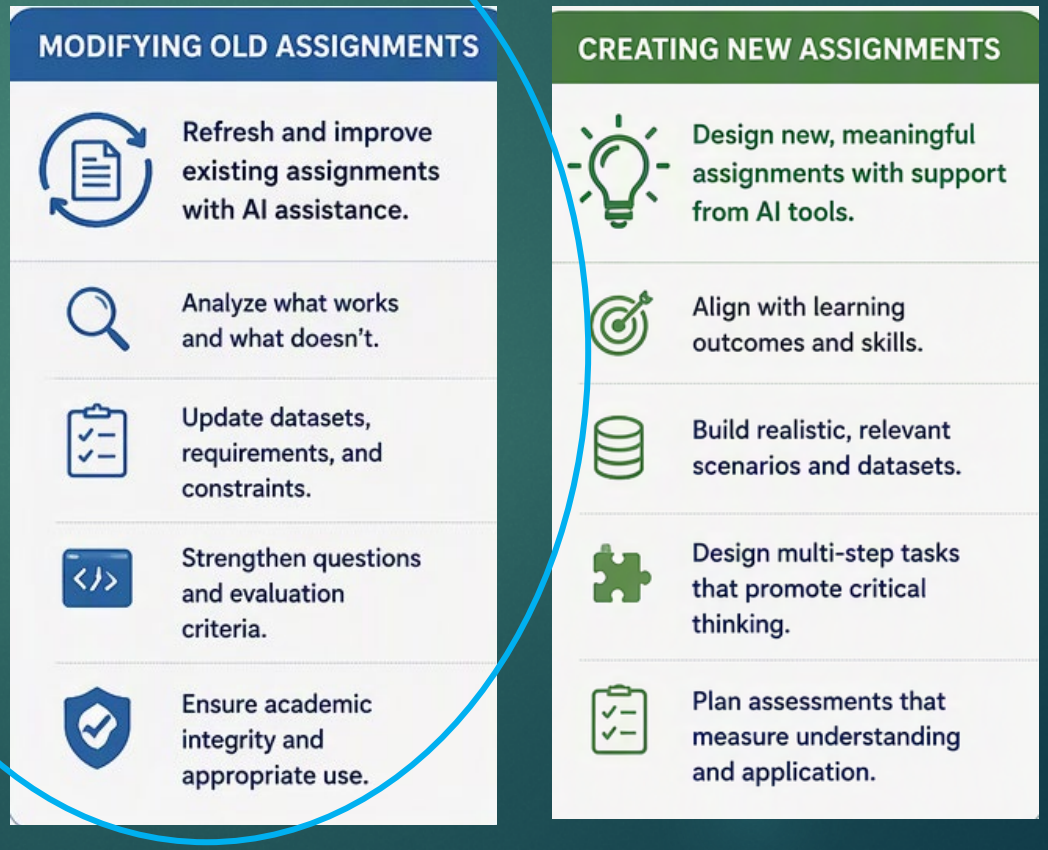
AI-Disclose

AI-Refine

AI-Yes

Assignments ?

[Modifying old / Creating new]



Data Science: Responsible AI-assisted coding Instructors Work: Modifying / Creating Assignments

Choosing an AI tool

- **Copilot** (Will use through school –UWEC- paid license)
- **ChatGPT**

[**Copilot** is built from Open AI's paid version ChatGPT so they function similarly when getting started with trying AI prompts.**]**

- **Google's Gemini AI**
- **Anthropic's Claude**
- **ASU's AI Labs tools** free tool from Arizona State University for image accessibility, question generator, rubric generator, and learning objective creator.

Example 1: DS 710 [Programming for Data Science (using Python)]

- Before 1

Assignment Overview

Purpose

The goal for this assignment is to use python plotting libraries to plot data in a meaningful way. The particular libraries you will use are likely `matplotlib`, which we explored in Lesson 5, plus built-ins via pandas, `seaborn`, or `plotly`. You are free to use any plotting tool you like to solve these tasks.

- After 1

Assignment Overview

Purpose

The goal for this assignment is to use python plotting libraries to plot data in a meaningful way. The particular libraries you will use are likely `matplotlib`, which we explored in Lesson 5, plus built-ins via `pandas`, `seaborn`, or `plotly`. You are free to use any plotting tool you like to solve these tasks. In this assignment, we will also begin to incorporate AI tools as coding collaborators. Watch out for the 🤖 emoji, which indicates that we expect you to use an AI tool on that task.

Example 1: DS 710 [Programming for Data Science (using Python)]

- Before 2

I wonder if the pay basis (hourly, annual, contract) impacts the scatterplot? Let's add that information.

> 🎯 Make a scatterplot of salary (y) vs years in job (x), for the entire data set.
>
> Use different markers for the different pay bases. For example, hourly could be dots, annual could be x's, and contractual basis could be triangles. Do whatever.

🎯 Make your code save the figure as `LAST_FIRST_plot_assign6b_task1-3.png`.

- After 2

How does the pay basis (hourly, annual, contract) impact the scatterplot? Let's add that information.

🎯 Make a scatterplot of salary (y) vs years in job (x), for the entire dataset.

Use different markers for the different pay bases. For example, hourly could be dots, annual could be x's, and contractual basis could be triangles. You should also use different colors; you might think about the accessibility of your colors; which colors are most accessible to viewers who are colorblind?

🤖 Here's a spot where you can ask an AI tool to help you choose a palette. Something to consider: how do AI tools differ in their responses here? ChatGPT, Google Gemini, and Claude are reasonable choices because they share broad capability but differ in how they handle code style, library preferences, and verbosity of explanation — so you may notice real differences even on a simple task.

🤖 Once you generate your scatterplot for the first time, you may notice that it's a bit difficult to read. Ask an AI tool to help you adjust the transparency of your markers; which should be most transparent? If you like the transparency option, go back to your first plot and see if adding transparency helps with readability in the lower left corner.

🎯 Have your code save the figure as `LAST_FIRST_assign6b_task1-3.png`.

Example 1: DS 710 [Programming for Data Science (using Python)]

- Level ? (ask the audience, if possible)

AI-No ? **AI-Disclose ?** **AI-Yes ?**

- After 2

How does the pay basis (hourly, annual, contract) impact the scatterplot? Let's add that information.

🎯 Make a scatterplot of salary (y) vs years in job (x), for the entire dataset.

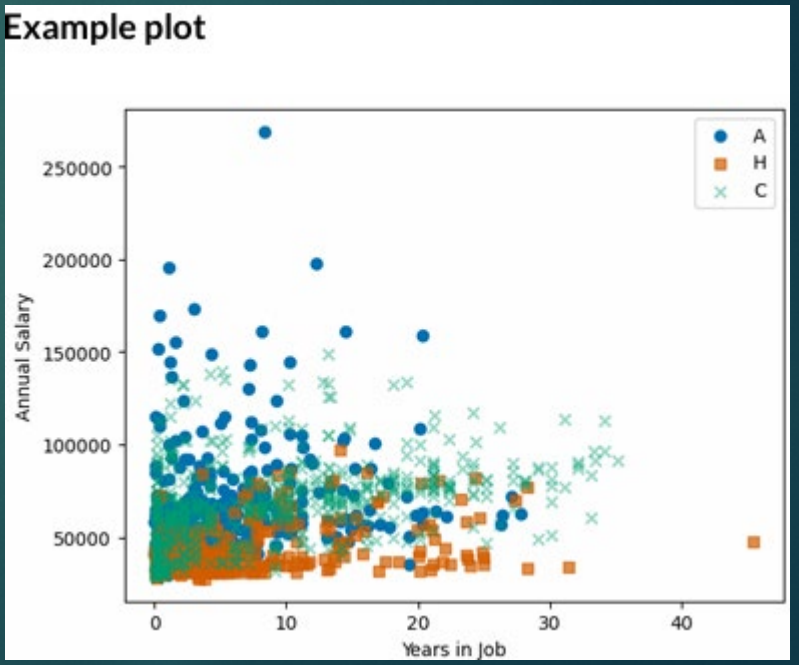
Use different markers for the different pay bases. For example, hourly could be dots, annual could be x's, and contractual basis could be triangles. You should also use different colors; you might think about the accessibility of your colors; which colors are most accessible to viewers who are colorblind?

🤖 Here's a spot where you can ask an AI tool to help you choose a palette. Something to consider: how do AI tools differ in their responses here? ChatGPT, Google Gemini, and Claude are reasonable choices because they share broad capability but differ in how they handle code style, library preferences, and verbosity of explanation — so you may notice real differences even on a simple task.

🤖 Once you generate your scatterplot for the first time, you may notice that it's a bit difficult to read. Ask an AI tool to help you adjust the transparency of your markers; which should be most transparent? If you like the transparency option, go back to your first plot and see if adding transparency helps with readability in the lower left corner.

🎯 Have your code save the figure as `LAST_FIRST_assign6b_task1-3.png`.

Example plot



Example 1: DS 710 [Programming for Data Science (using Python)]

- Before 3: visualization

Sometimes cramming all the data into one axes is too much, and we need multiple axes in the same figure. Fortunately, pyplot, pandas, and seaborn make this easy.

In this subtask, we'll compare how different (sub)departments give salaries based on years in position.

🎯 Make a figure with two subplots / sets of axes, side-by-side.

- In the left subplot: A scatter of salary (y) vs years in job (x) for the "facilities custodial services" department.
- In the right subplot: A scatter of salary (y) vs years in job (x) for the "mathematics" department.
- Make sure that both plots use the same axes bounds, otherwise the visual comparison is nearly meaningless.

- After 3: visualization

~~AI-No ?~~ **AI-Disclose ?** **AI-Yes ?**

🎯 Think about something else you would like to know about the dataset. For example, maybe you would like to directly compare two particular Sub Departments, or several different Job Titles. Up to you, surprise us!

🤖 Then, feed the raw dataset into two of the LLMs listed below, and ask each what type of visualization it would suggest to help answer your question. Keep track of your prompts, and use the *same* prompts in your chosen LLMs (in your reflection you will be asked for the exact prompt you used, copy-pasted verbatim). What are the similarities in responses? Differences?

Options for LLMs to try: ChatGPT, Google Gemini, and Claude.

Finally, implement the visualization in your code.

Example 2: DS 740 [Data mining and Machine Learning]

- Before

For each of the following situations, decide which type of data mining approach is most appropriate.

a. We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

- **A.** Regression
- **B.** Classification
- **C.** Unsupervised learning

- After ?

Example 2

- Before

For each of the following situations, decide which type of data mining approach is most appropriate.

a. We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

- **A.** Regression
- **B.** Classification
- **C.** Unsupervised learning

Prompt

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. Decide which type of data mining approach is most appropriate

- Anthropic Claude

Output

Supervised learning, specifically regression, with a focus on inference rather than prediction.

Why supervised? Each firm has a labeled response variable, CEO compensation, and a set of predictors: profit, number of employees, and industry. You want to relate the response to the predictors. Unsupervised methods like clustering or PCA have no response variable and wouldn't answer "which factors are associated with compensation."

Why regression? CEO compensation is a quantitative (continuous) outcome, so this is a regression problem rather than classification. Classification would only apply if you had turned compensation into categories, such as high versus low, which would throw away information.

Why inference? Your goal is to understand which factors are associated with compensation and how, not to forecast the pay of a new CEO. That favors interpretable models where you can examine coefficients, effect sizes, and significance:

- Multiple linear regression is the natural starting point. Industry enters as a categorical predictor via dummy variables.
- Because compensation is typically right-skewed, consider modeling log(compensation), and possibly log(employees) and log(profit) as well.
- Extensions such as interaction terms (e.g., profit × industry) or regularization (lasso) can help if there are many industry levels.

Flexible black-box methods like boosting or neural networks would be better suited if prediction accuracy were the main goal, but they give up the interpretability you need here.

Example 2 - Cont'd

- Before

For each of the following situations, decide which type of data mining approach is most appropriate.

a. We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

- **A.** Regression
- **B.** Classification
- **C.** Unsupervised learning

Prompt

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. Decide which type of data mining approach is most appropriate

- Copilot

The most appropriate **data mining approach is a supervised learning regression model**. CEO compensation is a **continuous numeric variable**, and you want to understand how it is related to other variables such as profit, number of employees, and industry. Regression is specifically designed for modeling and predicting continuous outcomes. [unstop.com], [docs.oracle.com]

Why regression?

- **Target (dependent) variable:** CEO compensation (numeric/continuous)
- **Predictor (independent) variables:**
  - Profit (numeric)
  - Number of employees (numeric)
  - Industry (categorical)

Possible models

- **Multiple Linear Regression** is the natural starting point because it:
  - Quantifies the relationship between compensation and each predictor.
  - Provides interpretable coefficients.
  - Allows testing which factors are significantly associated with CEO pay. [numberanalytics.com], [numberanalytics.com]

Example model

$$\text{CEO Compensation} = \beta_0 + \beta_1(\text{Profit}) + \beta_2(\text{Employees}) + \beta_3(\text{Industry}) + \epsilon$$

Alternative if prediction is the primary goal

If the objective is accurate prediction rather than interpretability, you could also consider:

- Regression Trees
- Random Forest Regression
- Gradient Boosting Regression

However, because the question asks which factors are **associated with** CEO compensation, **multiple regression (an explanatory/predictive regression approach)** is usually the most appropriate choice.

Answer: Use a **regression data mining approach**, specifically **multiple linear regression** as the initial method.

Example 2 - Cont'd

- Before

For each of the following situations, decide which type of data mining approach is most appropriate.

a. We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

- **A.** Regression
- **B.** Classification
- **C.** Unsupervised learning

Prompt

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. Decide which type of data mining approach is most appropriate

- Chat GPT

The most appropriate approach is **regression analysis**, specifically **multiple linear regression**.
The goal is to understand how CEO compensation (the **dependent variable**) is associated with several factors:

- **Profit** — quantitative predictor
- **Number of employees** — quantitative predictor
- **Industry** — categorical predictor, which can be represented using dummy variables

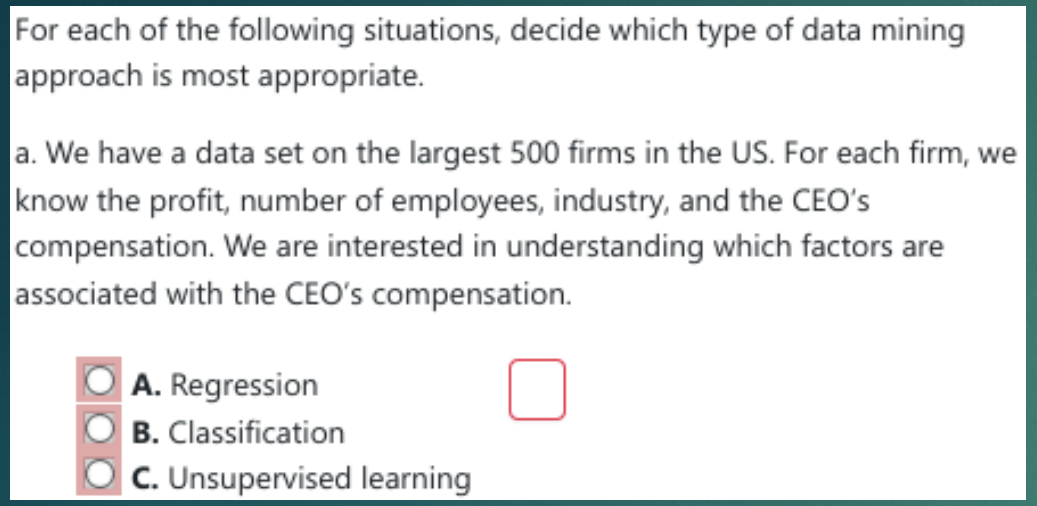
A multiple regression model could look like:

$$= \beta_0 + \beta_1() + \beta_2() + \beta_3() + \epsilon$$

This approach is appropriate because you are trying to **explain or predict a continuous outcome (CEO compensation) using multiple explanatory variables**, rather than simply grouping firms into categories.
Answer: Multiple regression (a supervised data mining approach).

Example 2

- Before



Prompt

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. Decide which type of data mining approach is most appropriate

- After → Level: AI – No

Reformulation ?

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. ~~Decide which type of data mining approach is most appropriate~~

Example 2

- Before

For each of the following situations, decide which type of data mining approach is most appropriate.

a. We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

- **A.** Regression
- **B.** Classification
- **C.** Unsupervised learning

Prompt

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. Decide which type of data mining approach is most appropriate

- After → Level: AI – No

Possible Reformulation

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

Which additional predictor other than revenue could be useful? Explain why.

Note: answer from AI (after revenue)
AI 1: CEO tenure
AI 2: Total Shareholder Return

Example 1

- Before

For each of the following situations, decide which type of data mining approach is most appropriate.

a. We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

- **A.** Regression
- **B.** Classification
- **C.** Unsupervised learning

Prompt

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. Decide which type of data mining approach is most appropriate

- After → Level: AI – Disclose

Possible Reformulation

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

Use AI to find why the logistic regression is not appropriate for this problem.

Disclose the AI tool used and your prompts (number of prompts).

Example 2

- Before

For each of the following situations, decide which type of data mining approach is most appropriate.

a. We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

- **A.** Regression
- **B.** Classification
- **C.** Unsupervised learning

Prompt

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation. Decide which type of data mining approach is most appropriate

- After → Level: AI – Yes

Possible Reformulation ?

We have a data set on the largest 500 firms in the US. For each firm, we know the profit, number of employees, industry, and the CEO's compensation. We are interested in understanding which factors are associated with the CEO's compensation.

You can Use AI for this problem. Indicate which tool was used.

A website states that the logistic regression is not appropriate because all these reason combined

(1) The response variable, CEO compensation, is a quantitative, continuous variable rather than a categorical outcome.
(2) The logistic regression is a supervised method.
(3) The CEO compensation is not in whole dollars, so we should use an unsupervised method.

Comment on these statements. Indicate which section of the lesson note s related to your answer.

Data Science: Responsible AI-assisted coding Instructors Work: Modifying / Creating Assignments

Choosing an AI tool

- **Copilot**
- **ChatGPT**

[**Copilot** is built from Open AI's paid version ChatGPT so they function similarly when getting started with trying AI prompts.**]**

- **Google's Gemini AI**
- **Anthropic's Claude**
- **ASU's AI Labs tools** free tool from Arizona State University for image accessibility, question generator, rubric generator, and learning objective creator.

Thank you