

**Florida Gulf Coast  
University**



# **DA2I: Data Analysis with AI**

Leandro Nunes de Castro, Ph.D.

[Idecastrosilva@fgcu.edu](mailto:Idecastrosilva@fgcu.edu)



# **Data Analysis with AI OR AI-Enhanced Data Analysis**

## **Summary**

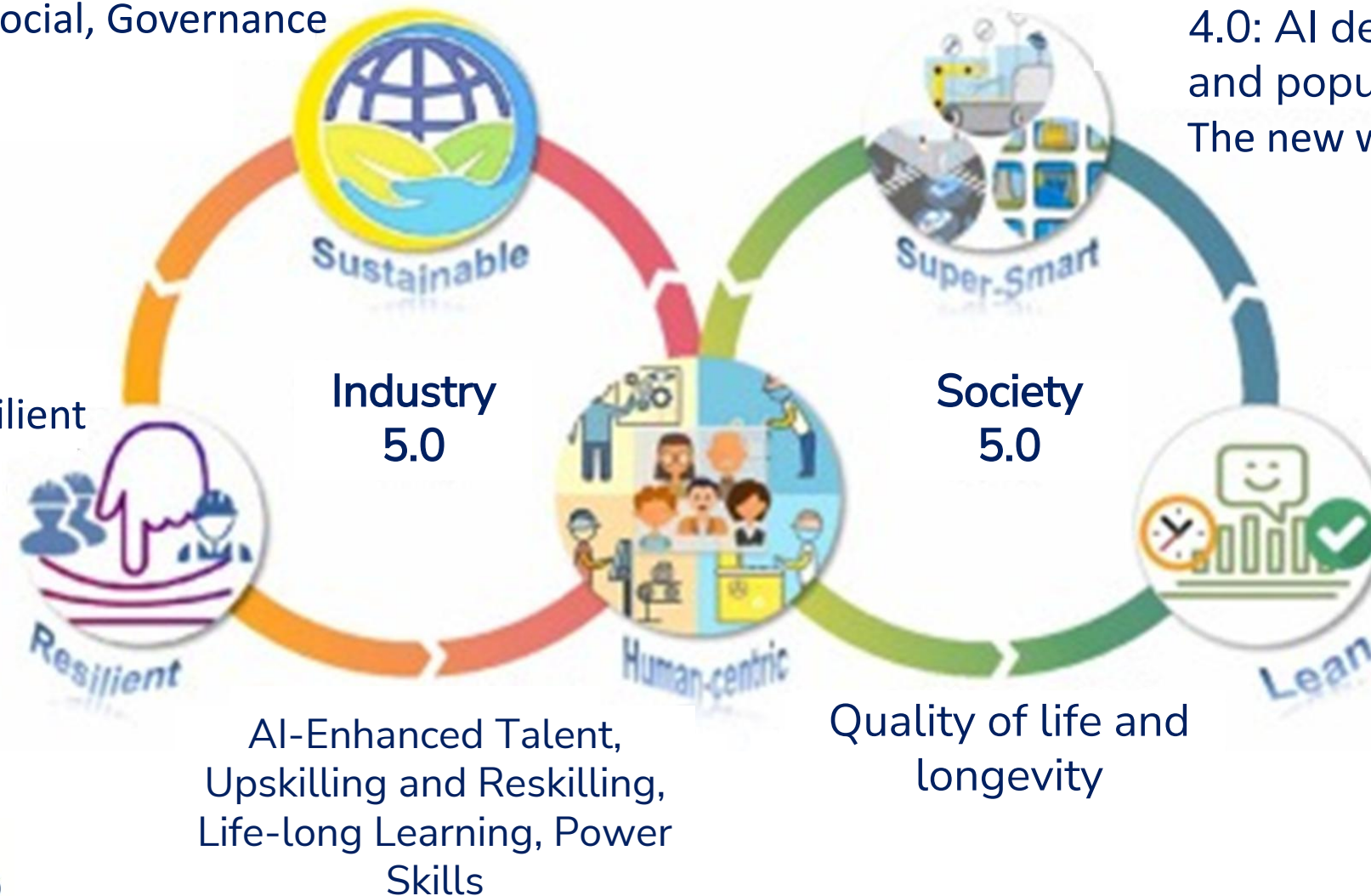
- AI and AI-Enhanced Data Analysis
- Main AI Tools to Be Explored
- Introduction to Data Analysis
- Main Careers in Data Science
- Terminologies and Courses
- From Bad to Good Practices
- Principles of Prompt Engineering

# Context





ESG:  
Environment, Social, Governance

Consolidation of industry  
4.0: AI democratization  
and popularization.  
The new workforce!

FARAs  
Technologies:  
Flexible, Agile, Resilient  
and Adaptable



## AI-Enhanced Data Analysis Classes

Class	Topics	AI Integration
 <b>Class 1</b>	Introduction to Data Analysis, AI Tools, Careers, Practices, Prompt Engineering	AI and AI-Enhanced Data Analysis
 <b>Class 2</b>	Data Analysis Workflow, Data Description, Data Preparation, Descriptive Data Analysis (DDA)	AI-Enhanced Data Preparation, AI-Enhanced DDA
 <b>Class 3</b>	Distribution, Associations, Amounts, Proportions, Evolution, Geospatial, Design Principles	AI-Enhanced Dataviz
 <b>Class 4</b>	Data Storytelling, Dashboard Design, Case Studies	AI-Enhanced Data Storytelling, AI-Enhanced Dashboard Design
 <b>Class 5</b>	Full Data Analysis Project, Requirements, Learning Objectives	Full Data Analysis Project with AI



# Learning Objectives

- Understand and explain the core principles of data analysis
- Understand the potential of AI in data analysis
- Use AI tools to assist in summarizing, visualizing, and interpreting data
- Use AI tools to develop compelling data narratives and dashboards
- Discuss the limitations, ethics, and accuracy in AI-generated outputs
- Apply the AI-augmented analysis workflow in a capstone project



# Dr. Leandro de Castro

- ✓ Converted Christian
- ✓ Husband and father of two



## ✓ Degrees

Ph.D. in El. Eng.



UFG



UNICAMP

## ✓ Leadership Positions

## ✓ Academic Career and Positions

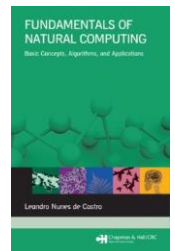
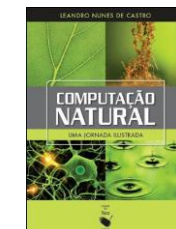
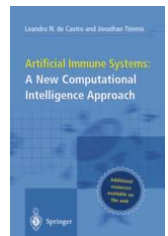
University of  
**Kent**



VNIVERSIDAD  
D SALAMANCA

## ✓ Researcher KPIs







## ✓ Entrepreneurial Career



# AI and AI-Enhanced Data Analysis

---

## Class 1 Overview

Topic	Description
 <b>AI</b>	AI and AI-Enhanced Data Analysis
 <b>AI Tools</b>	AI Tools to be Explored
 <b>Data Analysis</b>	Introduction to Data Analysis
 <b>Data Science Careers</b>	Careers in Data Science
 <b>Best Practices</b>	From Bad to Good Practices
 <b>Prompt Engineering</b>	Introduction to Prompt Engineering



# What is AI?

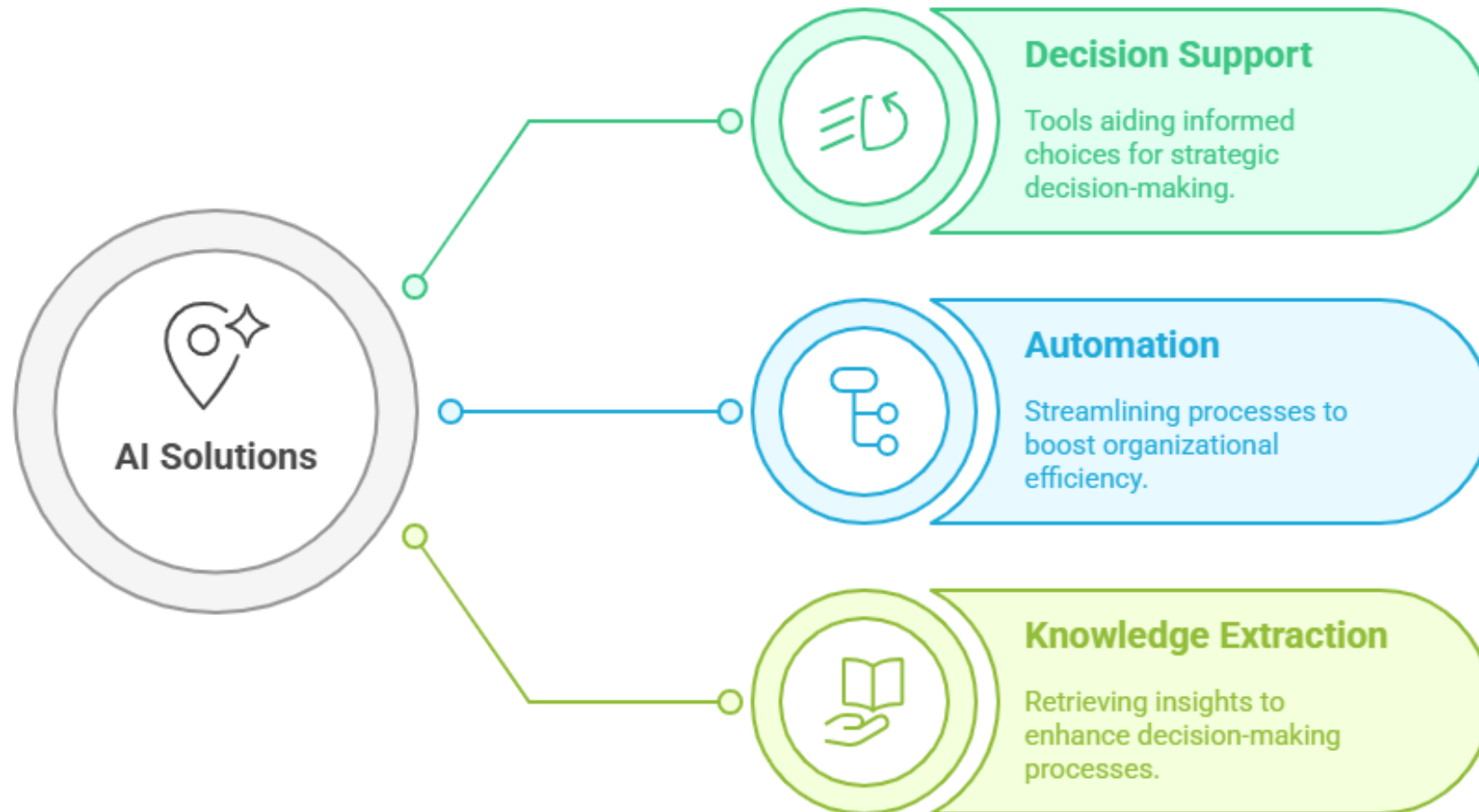
AI refers to computer systems designed to mimic human cognitive functions such as learning, problem-solving, perception, and decision-making. In the context of data analysis, AI encompasses technologies that can:

1. Interpret data
2. Learn from examples and experience
3. Make decisions with minimal human intervention
4. Adapt to new circumstances and data

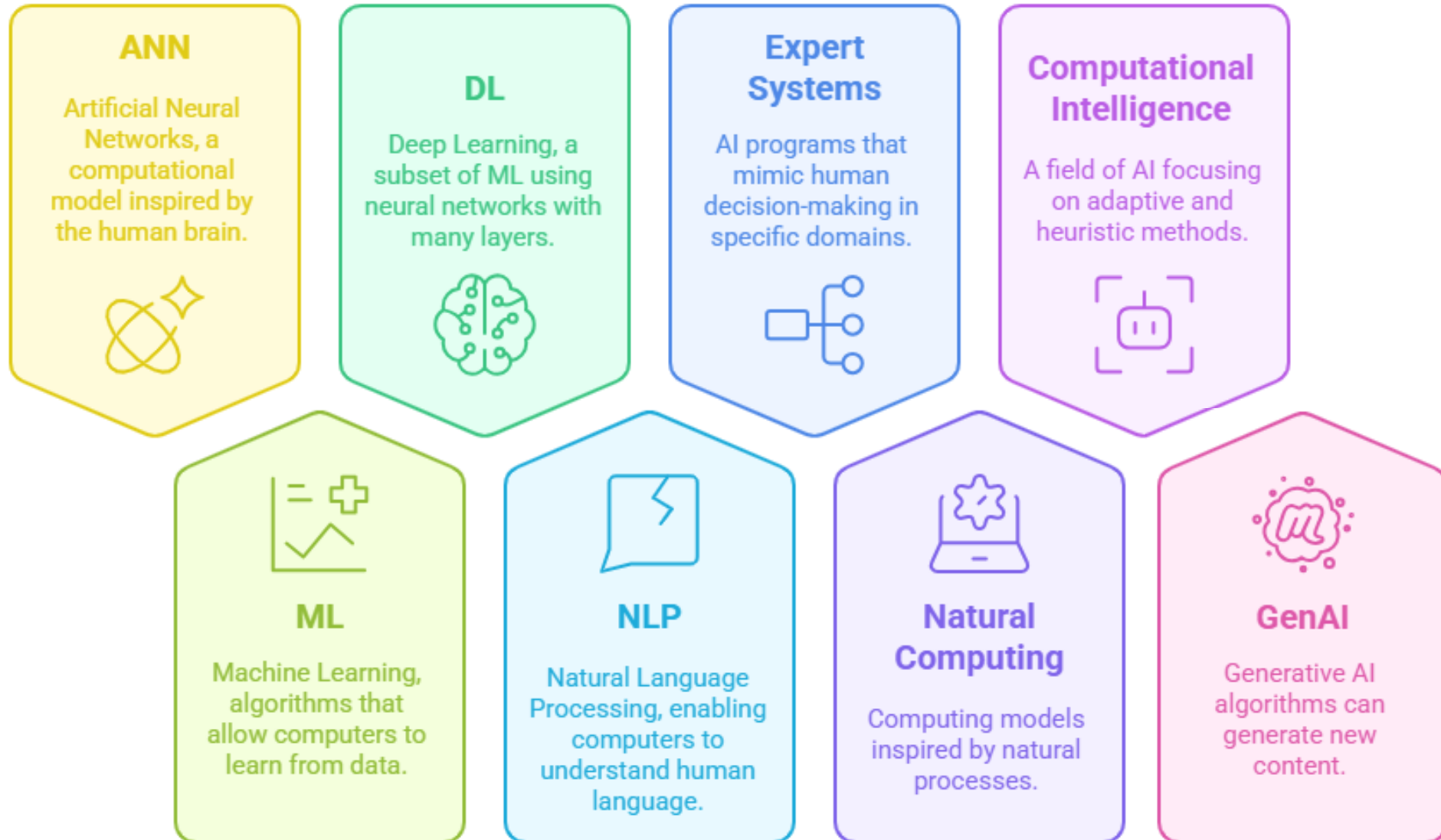
# Standard x AI-Based Software

Feature / Dimension	Standard Software	AI-Based Software
<b>Development Approach</b>	Rule-based, algorithmic; pre-defined instructions	Learn from the environment (e.g., data, interactions)
<b>Behavior Definition</b>	Deterministic - behavior is defined by code logic	Probabilistic - behavior emerges from learned patterns
<b>Input/Output</b>	Input → Logic → Output	Input → Model → Probabilistic Output
<b>Testing and Debugging</b>	Test against expected outputs	Test against metrics (e.g., accuracy, precision, recall)
<b>Update Process</b>	Modify the code manually	Retrain or fine-tune the model with new data
<b>Predictability</b>	High - same input always yields same output	Lower - small changes in input can affect output
<b>Failure Modes</b>	Logic bugs, edge cases	Data bias, concept drift, hallucinations
<b>Maintenance</b>	Code refactoring, dependency updates	Model retraining, dataset updates, drift monitoring
<b>Skillset Needed</b>	Programming, software engineering	Data science, ML/AI engineering, data engineering
<b>Example</b>	A tax calculator, CRM system	A recommendation engine, image classifier, chatbot

## Enhancing Organizational Strategies through AI



## Areas of Artificial Intelligence





## Generative AI

Creates original content  
and understands complex  
patterns



## Large Language Models

Enables natural language  
understanding and  
generation

Made with  Napkin

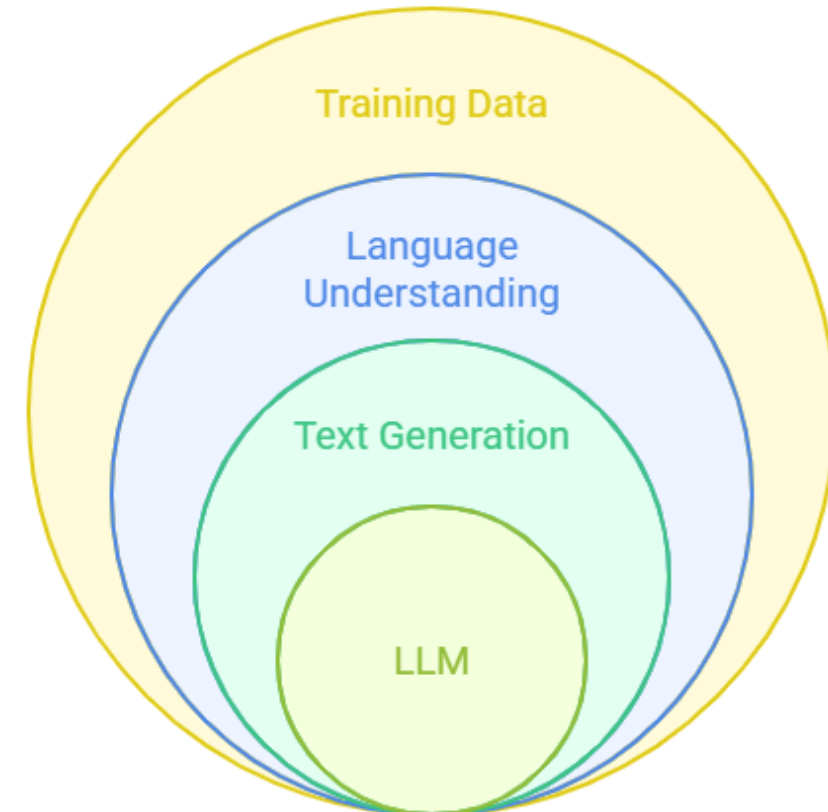
## Large Language Models Structure

Massive text datasets  
used for training

Comprehension of  
language patterns

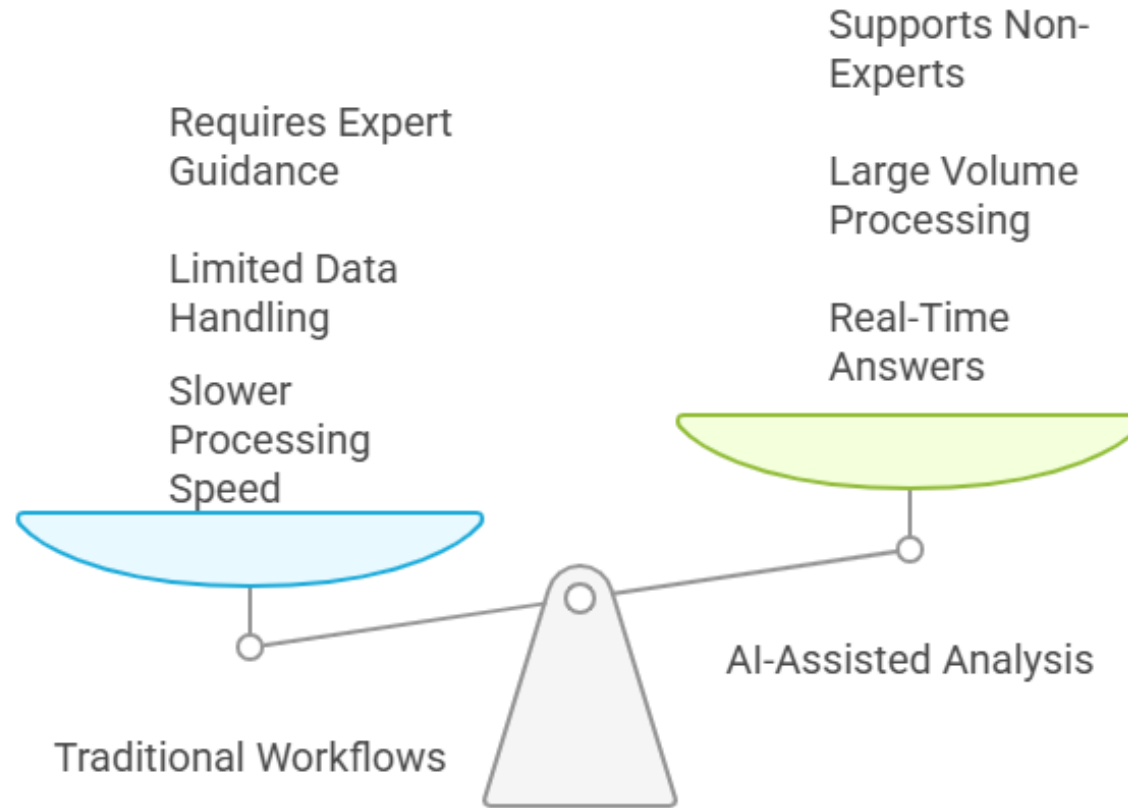
Ability to produce  
human-like text

Core function of text  
processing



Made with  Napkin

# Why AI in Data Analysis



AI enhances data analysis speed, scale, and support.

Made with  Napkin

**Main AI Tools to Be Explored  
in this Course:**

**ChatGPT, Claude.ai, Perplexity**

---



Feature	ChatGPT	Claude.ai	Perplexity
<b>Data Upload &amp; Analysis</b>	Yes, supports various formats (.csv, .xlsx, .pdf, .json up to ~50MB/file)	Yes, can process data with its "Analysis Tool" using JavaScript code.	Yes, allows file uploads for analysis.
<b>Code Interpretation</b>	Yes, can write and execute Python code for analysis.	Yes, can write and run JavaScript code for data processing.	Yes, utilizes coding capabilities for analysis within its Deep Research mode.
<b>Statistical Analysis</b>	Capable of various statistical analyses (comparative, correlation, etc.).	Can perform complex calculations and data manipulation.	Can identify patterns, trends, and anomalies.
<b>Data Visualization</b>	Can generate various charts (bar, pie, scatter, line, histograms, etc.).	Can create visualizations through its "Analysis Tool" and Artifacts.	Offers data visualization features to transform data into charts and graphs.
<b>Source Integration</b>	Can analyze data from Google Drive and OneDrive.	Focuses on uploaded data within the context window.	Integrates web search for broader context in analysis.
<b>Focus</b>	Broad AI capabilities including data analysis.	Strong natural language processing and expanding data analysis tools.	Research-focused with emphasis on data-driven insights.
<b>Real-time Data</b>	Requires web browsing for up-to-date external data.	Primarily analyzes uploaded data.	Strong real-time web search integration for current data.

# Icebreaker with AI Insights

- For the mammographic\_masses\_nominal dataset, prompt AI as follows:

*“Here is a dataset of mammographies. Can you tell me 3 surprising insights?”*

- Which insights came from AI? Which are real? What’s your overall assessment of the result?

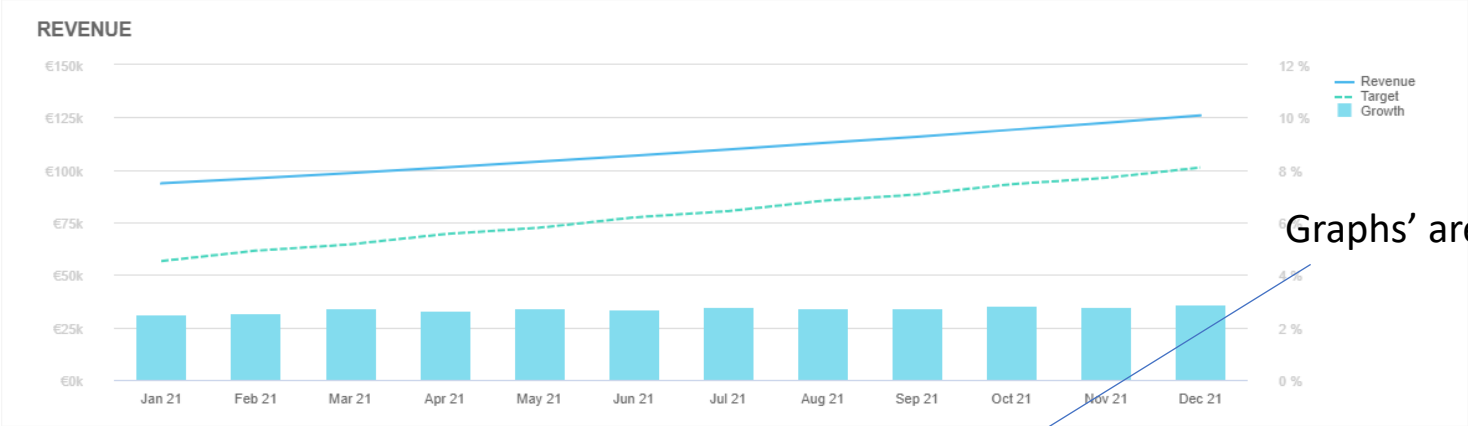
# Introduction to Data Analysis

---

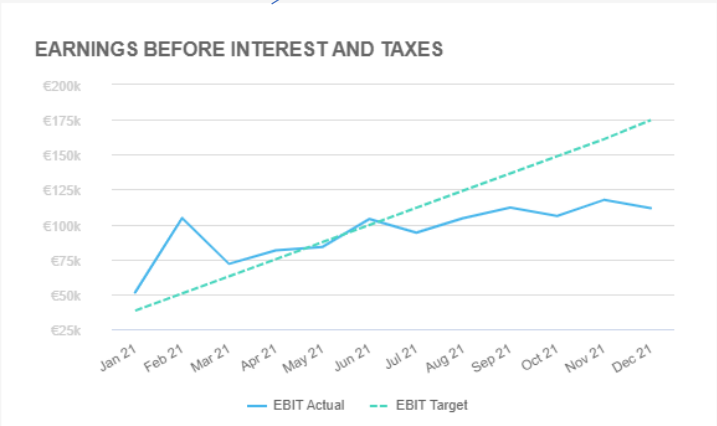
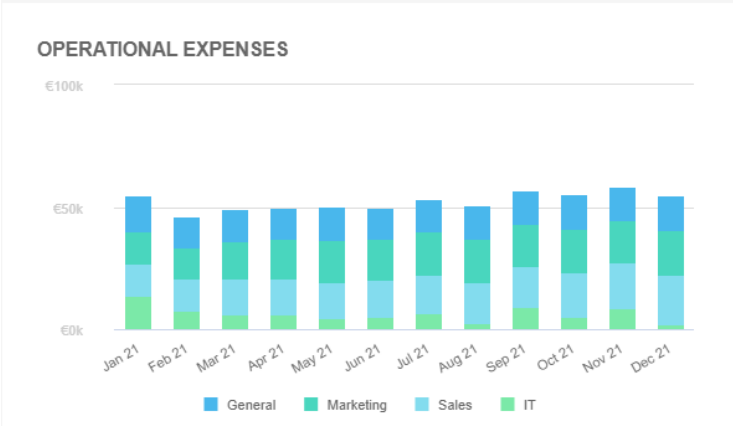
FINANCIAL DASHBOARD EXAMPLE

Key Performance Indicators (KPIs)

Table



Graphs' area



INCOME STATEMENT

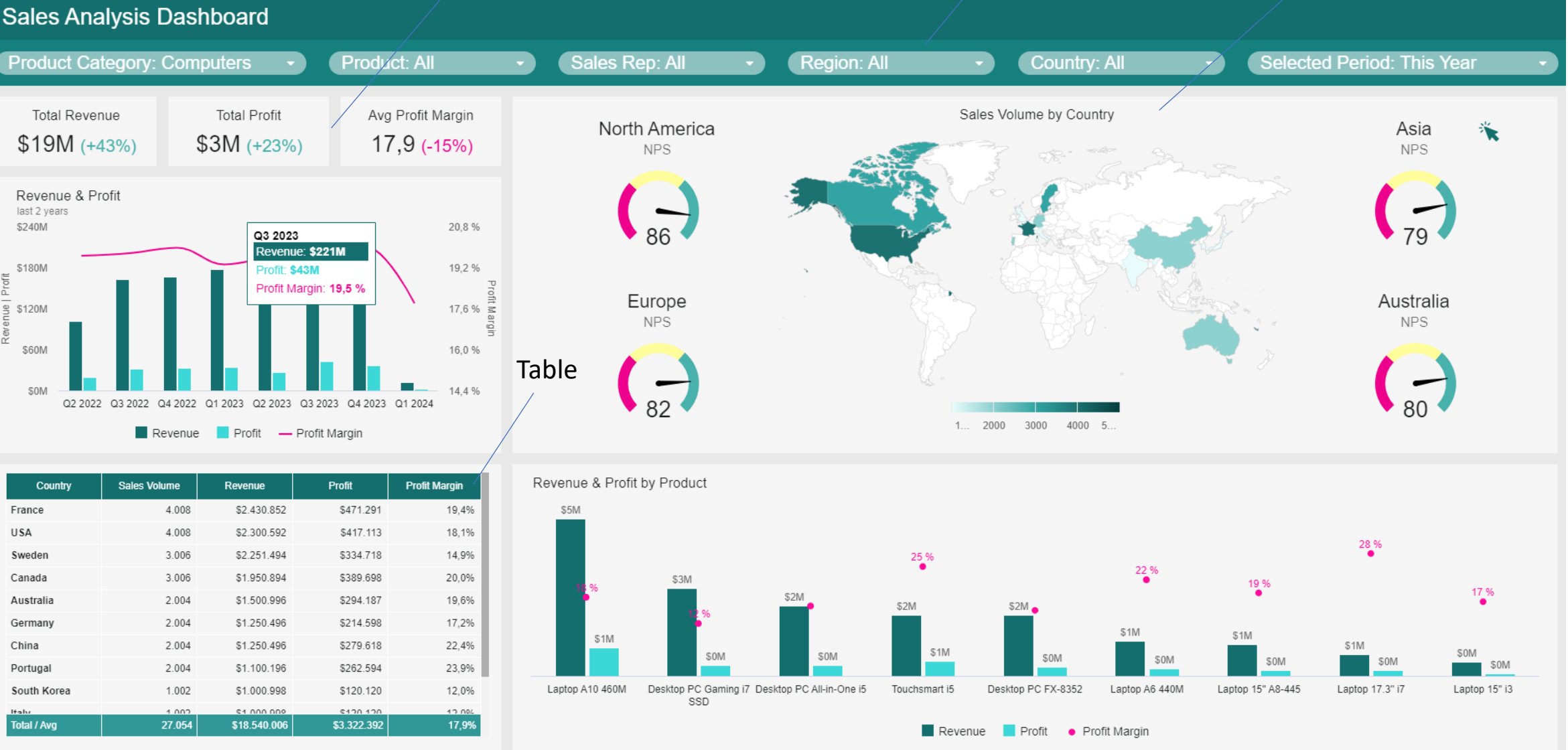
Revenue	1,305,507 €
COGS	208,453 €
GROSS PROFIT	1,097,054 €
OPEX	815,306 €
Sales	279,886 €
Marketing	192,710 €
IT	192,656 €
General & Admin	150,054 €
OTHER INCOME	2,130 €
OTHER EXPENSES	51,195 €
EBIT	232,684 €
Interest and Tax	38,244 €
NET PROFIT	194,440 €

# SALES ANALYSIS DASHBOARD EXAMPLE

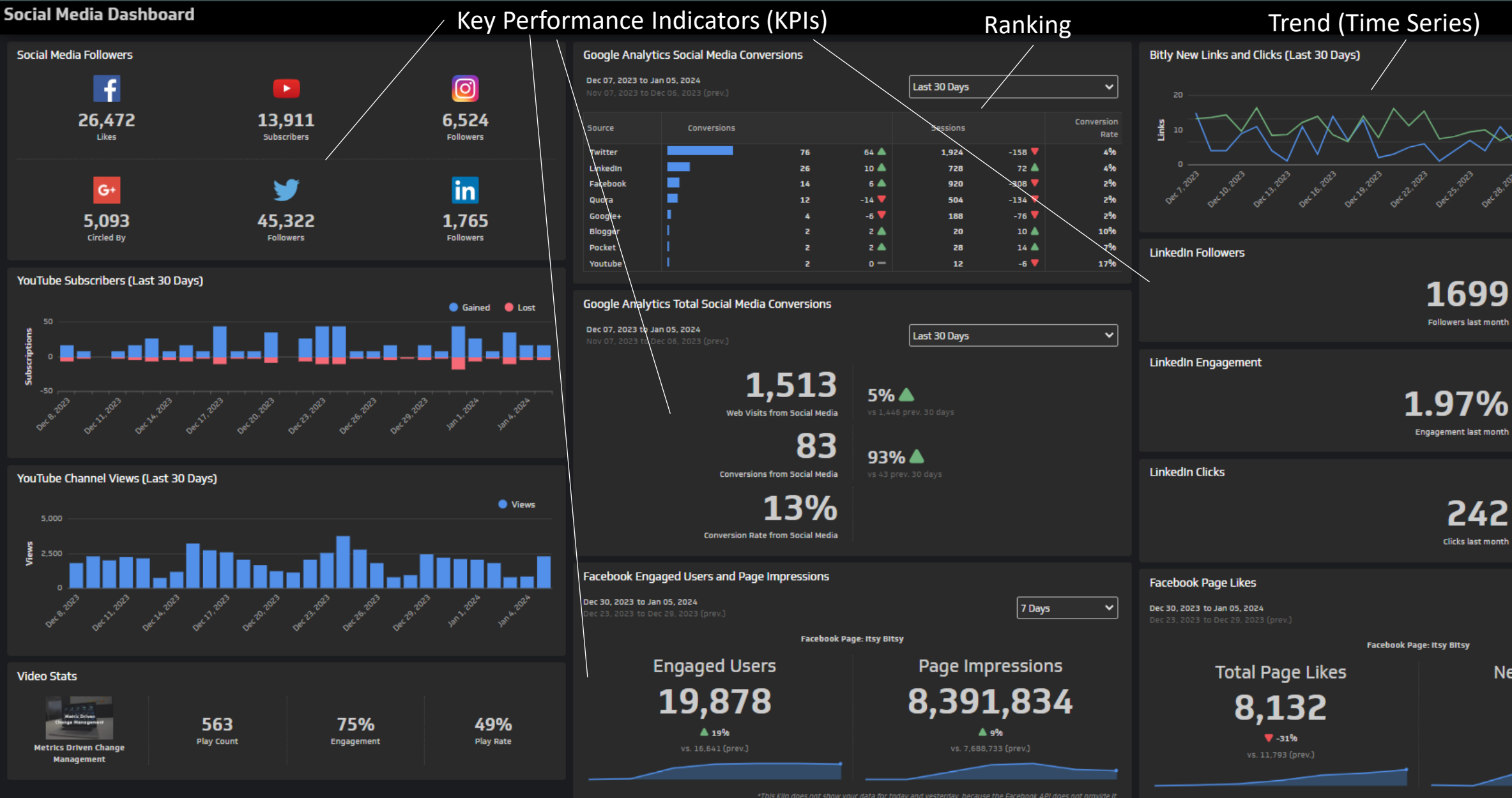
Key Performance Indicators (KPIs)

Filters

Geomap



# SOCIAL MEDIA DASHBOARD EXAMPLE



# What this course is and is not about

## It is about:

- Summarizing data by means of mathematics and statistics
- Visualizing different types of data
- Data storytelling and dashboard design
- Training Data Analysts
- Doing all that With and Without AI

## It is not about:

- Programming
- Web design
- Data science\*\*



# Who is a Data Analyst?

- Professional who collects, prepares and analyzes data, providing statistics, visualizations, interpretations, and **business-focused insights** about the data. Their **skills** include the use of manipulation and visualization software, like Excel, SQL, Tableau, Power BI, etc.
- **Common degrees:** economics, business administration, engineering, computing, statistics.



# What is Data Analysis?

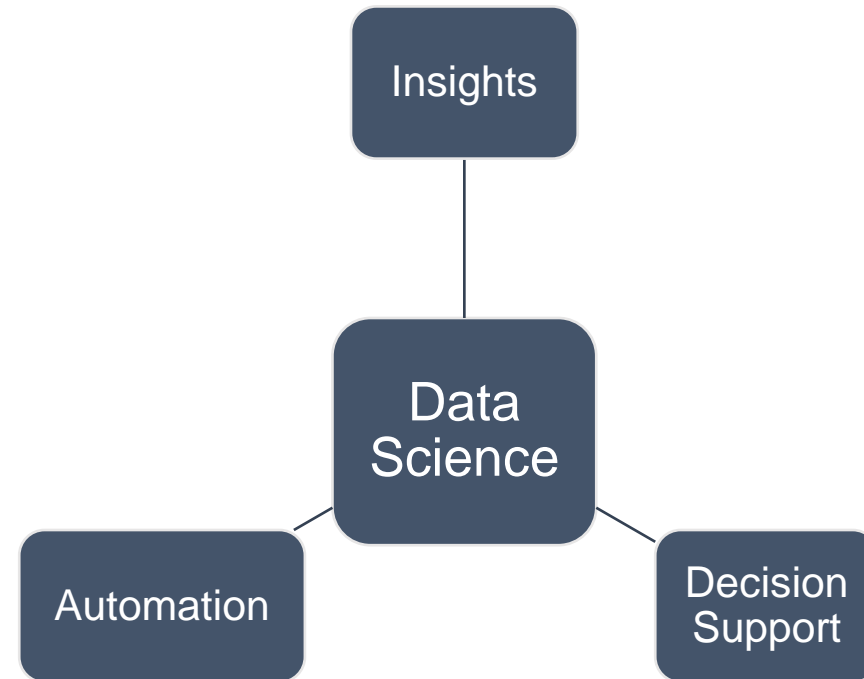
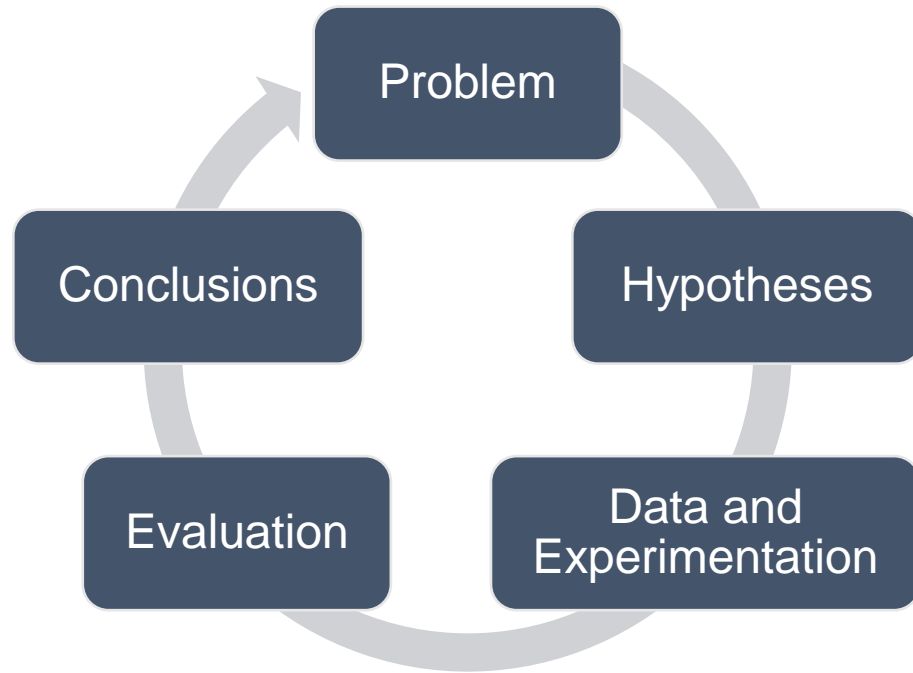
*Data Analysis is an important part in the journey of using data to build a data-driven culture focused on business-oriented insights, that is, a company in which data is intensively used to drive operations, decisions, and results.*

*DA2I focuses on the understanding, characterization, summarization, and visualization of data for business-oriented insights using standard and AI techniques.*

# Goals of DA2I

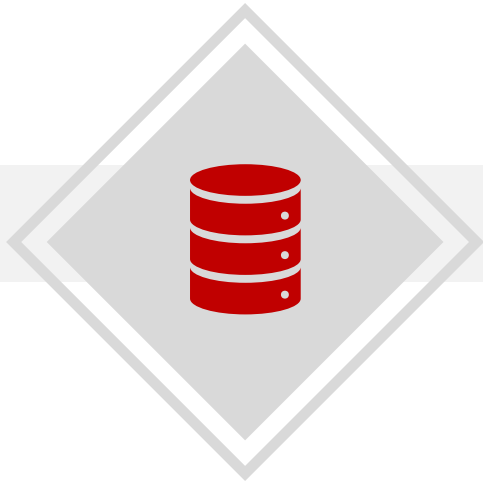
- To understand the distribution and structure of data.
- To summarize data characteristics.
- To extract business-oriented insights and indicators from data.
- To identify relevance and/or select variables.
- To visualize potential relationships between variables.
- To identify anomalies.
- To allow the application and/or selection of learning-based methods.

# Data Science: Data + Science



# Big Data

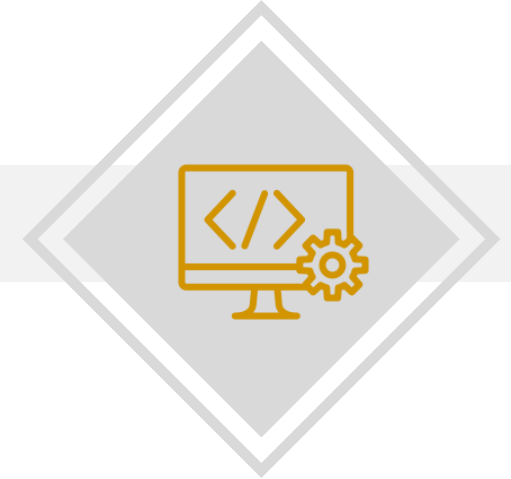
- Area that deals with the storage, processing and analysis of large datasets with specific characteristics: velocity, variety, and volume (3 Vs of big data). More Vs: veracity, value, variability, etc.



Data (3 Vs)



Hardware Infrastructure



Software Infrastructure



## Careers in Data Science

- Knowledge and Skills
- Business Analyst
- Data Scientist
- ML Engineer
- BI and Data Analyst
- Data Engineer
- Chief Data Officer

# Knowledge and Skills

## Mathematics and Statistics

- Machine learning
- Statistical modeling
- Probability
- Mathematical programming
- Neural networks

## Computation

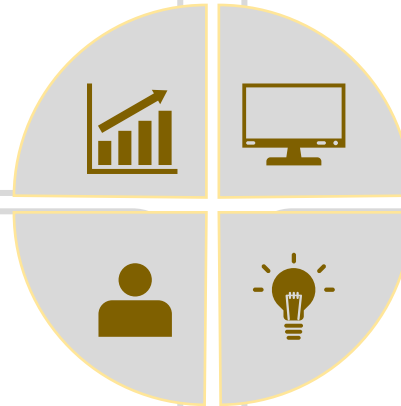
- Programming languages
- Databases
- High Performance Computing
- Software Engineering
- Parallel and distributed processing

## *Soft Skills*

- Communication
- Leadership
- Teamwork
- Troubleshooting
- Creativity

## Business

- Management
- Strategy
- Finance
- Processes





# Main Technical Skills and Knowledge

## Predictive Statistics

Linear algebra, calculus, probability, statistical inference, regression, Bayesian statistics, search and optimization, graph theory, etc., Learning paradigms, ML/DL/AI algorithms, experimental design and analysis, NLP, etc.

## Descriptive Statistics

Summary measures (central tendency, dispersion, form), measures of association, linear regression, the normal distribution, etc.

## Data Visualization

PowerBI, Tableau, Looker Studio, Metabase, QlickView, Excel, etc.

## Storytelling

Data visualization, interpretation, storyboarding and narrative, domain knowledge, story structuring, audience understanding, communication, data ethics, etc.

# Main Technical Skills and Knowledge

## Business Analysis

Exploratory data analysis (descriptive statistics, data exploration and visualization), data querying, experimental design and testing, data storytelling, data governance and ethics, critical thinking and problem-solving, data analysis tools, intro to ML/DL, business knowledge

## Software Engineering

Programming, software lifecycle, version control, documentation and testing, design and architecture, databases, web and API dev, deployment, DevOps, security, etc.

## Databases

ER modeling, relational data modeling, database design and normalization, indexing and querying, security, SQL, NoSQL, MongoDB, Cassandra, Neo4J, etc.

## Programming

Control flow and conditionals, data types and variables, functions and modular programming, data structures, OOP, file and error handling, debugging, best practices, R, Python, Matlab, C, Java, Matematica

# Main Careers

- Business Analyst
- Data Scientist
- BI and Data Analyst
- Data Engineer
- Data Architect
- Machine Learning Engineer
- Chief Data Officer
- Chief Scientist\*



# Main Knowledge/Skills

- Programming
- Software Engineering
- Databases
- Predictive Statistics
- Descriptive Statistics
- Business Analysis
- Storytelling
- Visualization

# Business Analyst

## Focus:

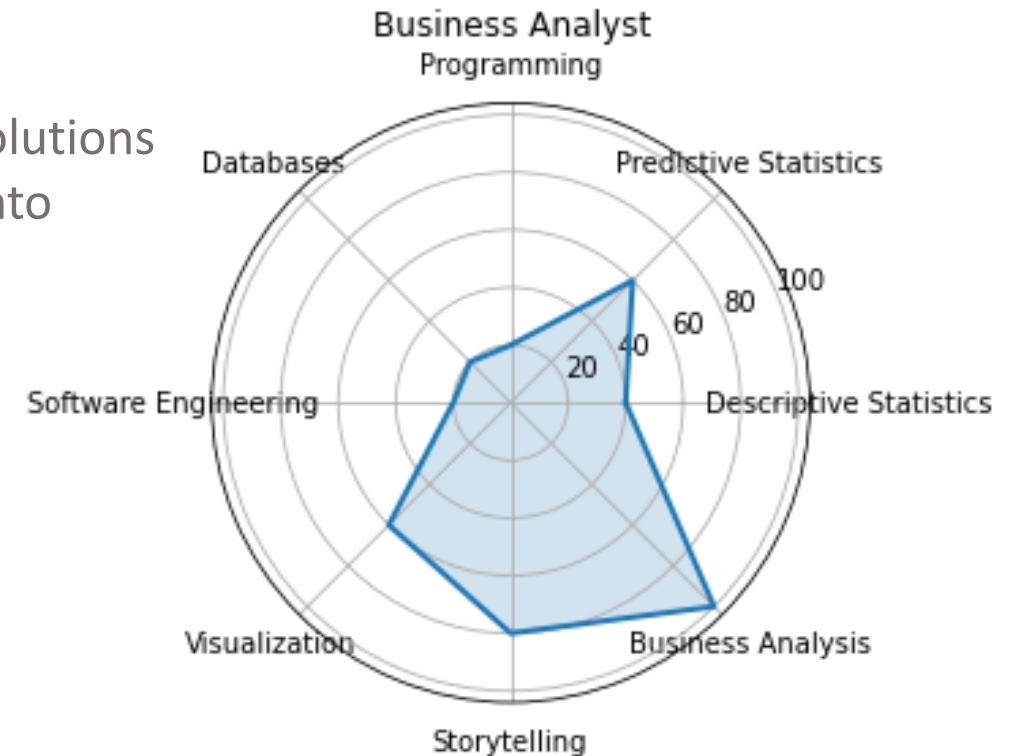
- Understand the customer's problem from the business perspective
- Gather requirements, develop use cases and structure solutions
- Identify the analytical potential of data and translate it into business

## Skills:

- Business and market knowledge
- Knowledge about the process and analytical tools
- Communication

## Degrees:

- Economics, Business, Engineering, Computing



# Data Scientist

## Focus:

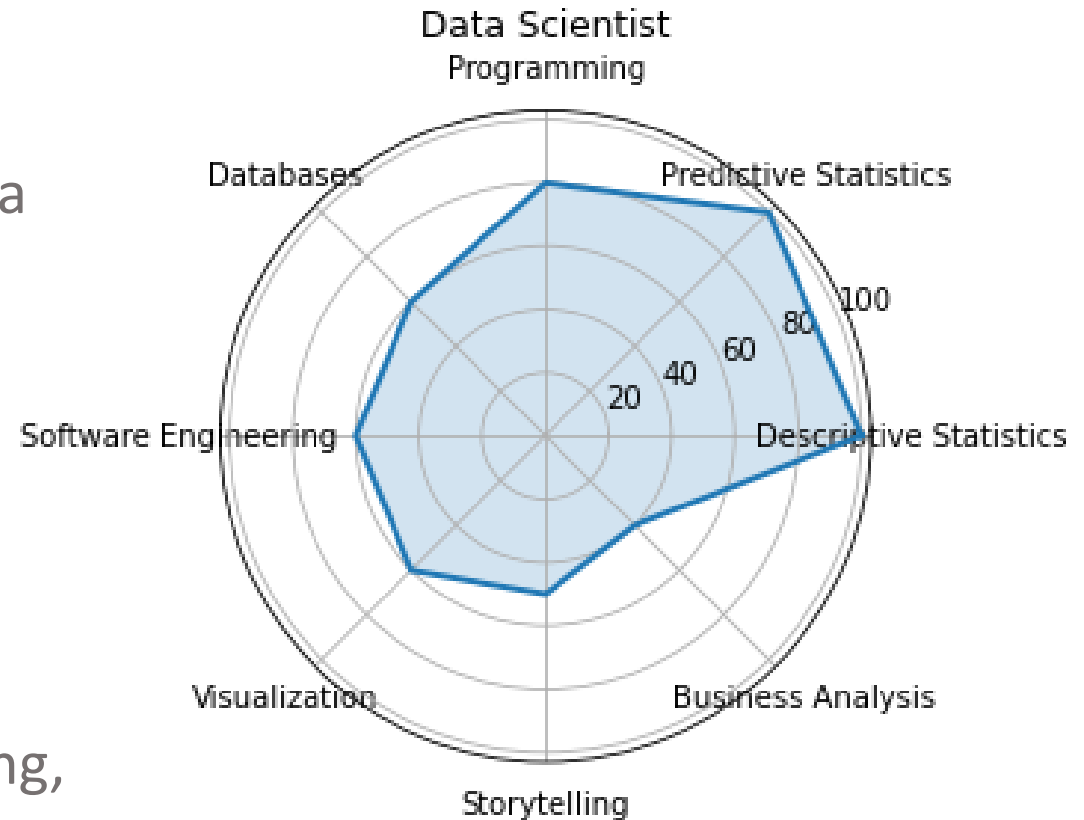
- Parametrize, tune and apply algorithms
- Translate business into data analytics
- Bridge between the dev and the business area

## Skills:

- Algorithms and programming
- Machine learning, statistics, neural networks
- Problem modeling
- Communication

## Degrees:

- Mathematics, Economics, Statistics, Computing, Engineering, Physics



# ML Engineer

## Focus:

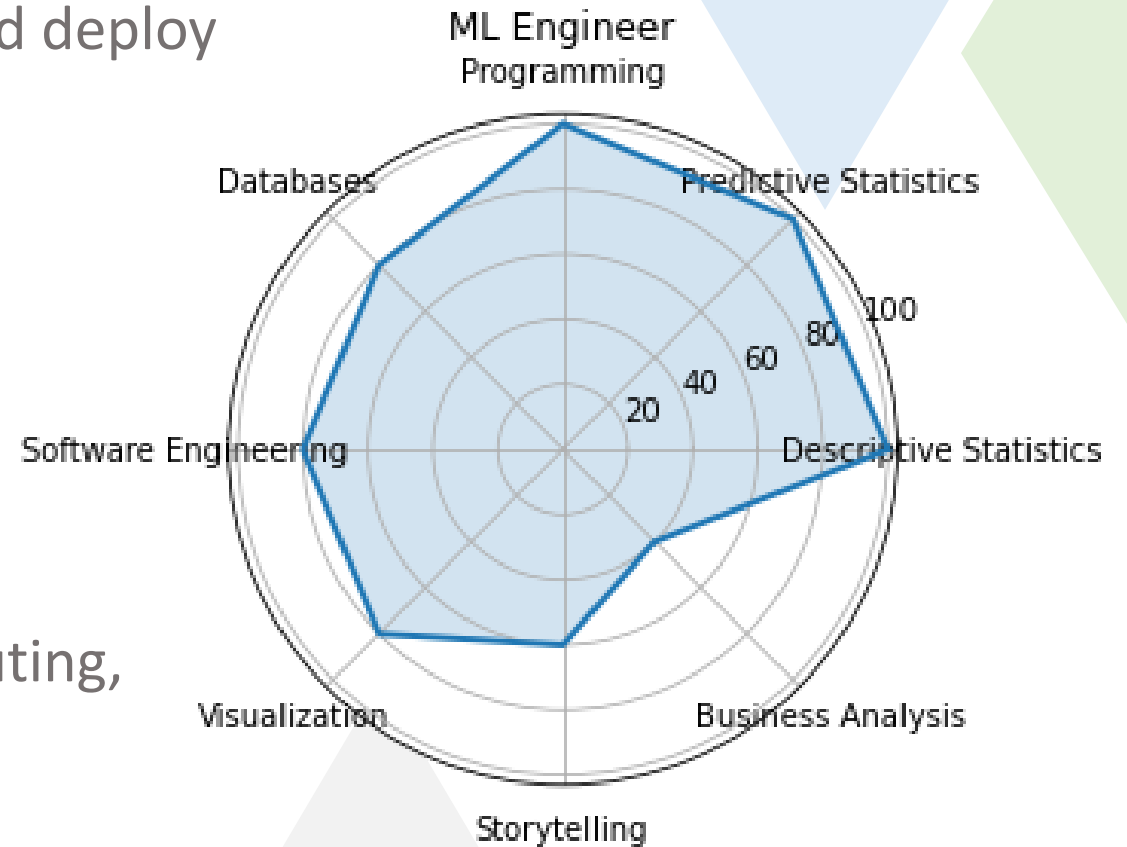
- Design and develop ML/DL algorithms
- Optimize (e.g., scalability, performance) and deploy ML/DL algorithms

## Skills:

- Algorithms and programming
- Software engineering
- Machine/Deep learning, statistics, ANN
- Problem modeling

## Degrees:

- Mathematics, Economics, Statistics, Computing, Engineering, Physics



# Data Analyst

## Focus:

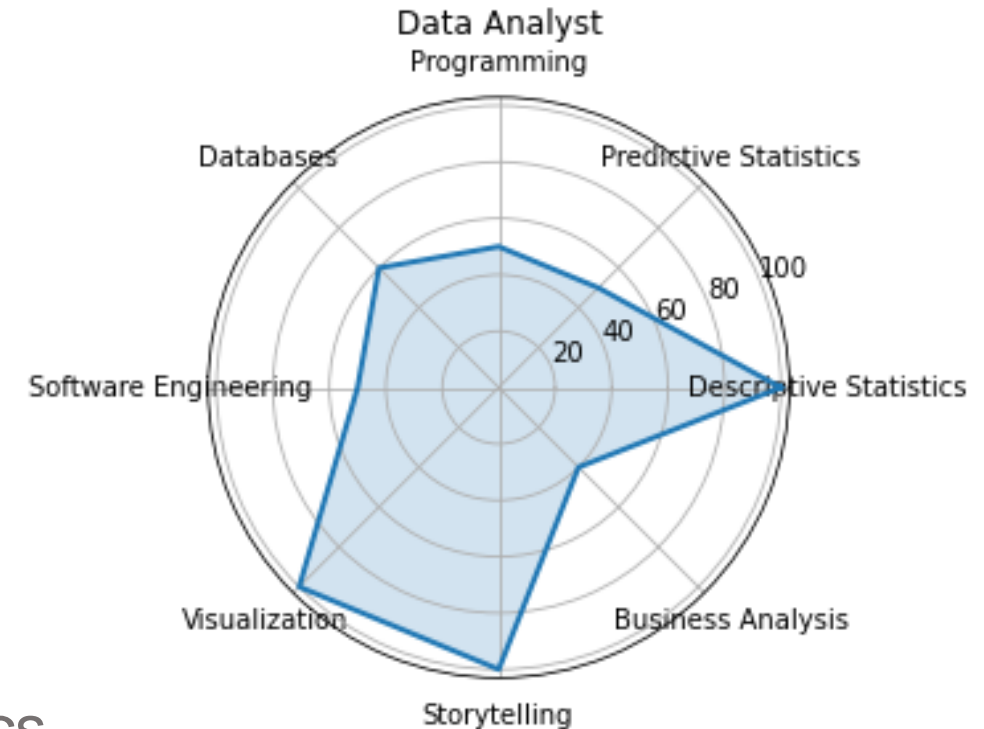
- Analyze the data
- Provide reports, visualizations, and interpretations
- Provide business-oriented insights from data

## Skills:

- Manipulation and visualization software (eg Excel, Tableau, Power BI, Metabase)
- Communication

## Degrees:

- Economics, Engineering, Computing, Statistics, Business





# Data Engineer

## Focus:

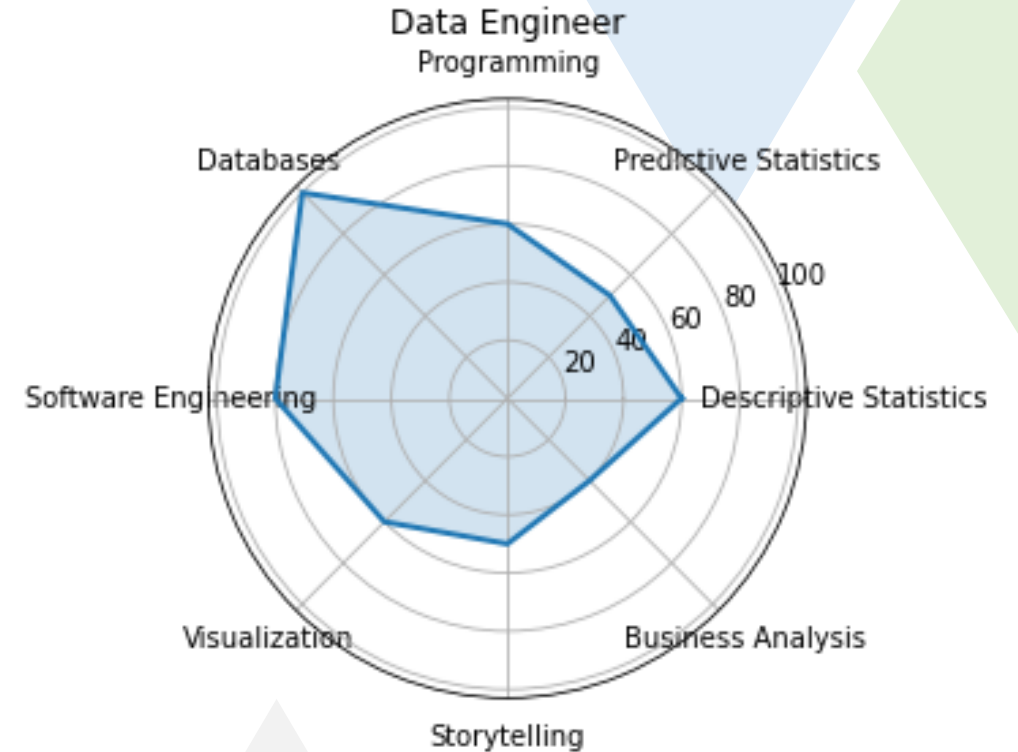
- Handle large masses of data
- Manage databases
- Prepare data
- Code models

## Skills:

- Software Engineering
- Programming languages and frameworks
- Databases
- Communication

## Degrees:

- Computing, Engineering



# Chief Data Officer

## Focus:

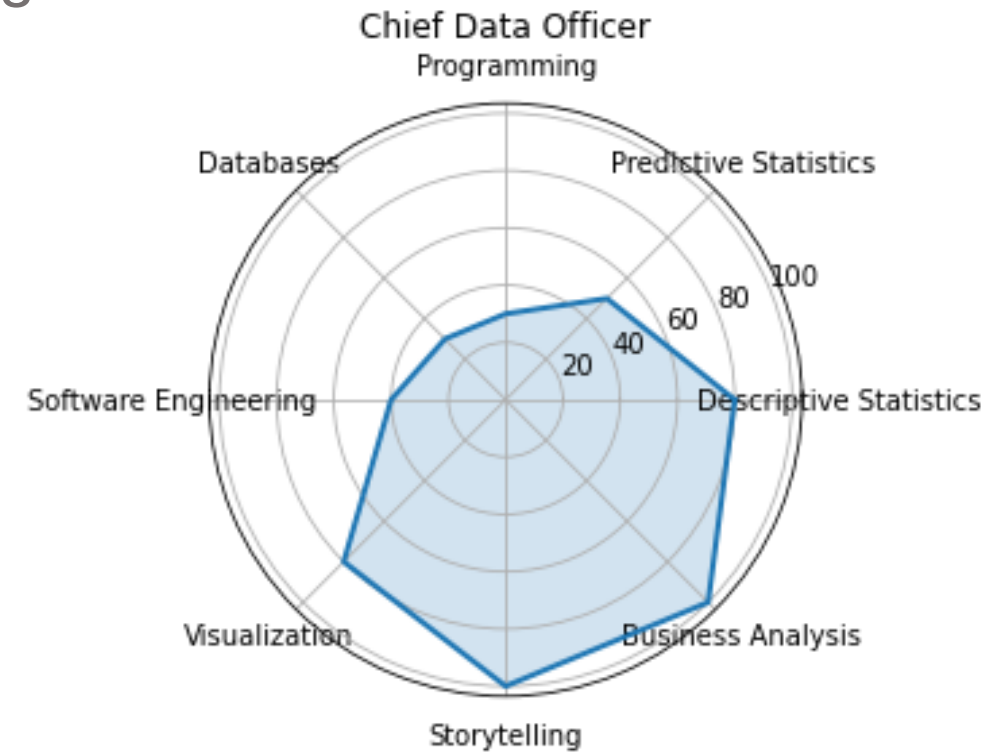
- Data governance for new products, services and strategic decision making
- See data as a company asset
- Directs the DS teams

## Skills:

- Computing
- Engineering
- Business
- Leadership

## Degrees:

- Computing, Engineering, Economics, Business

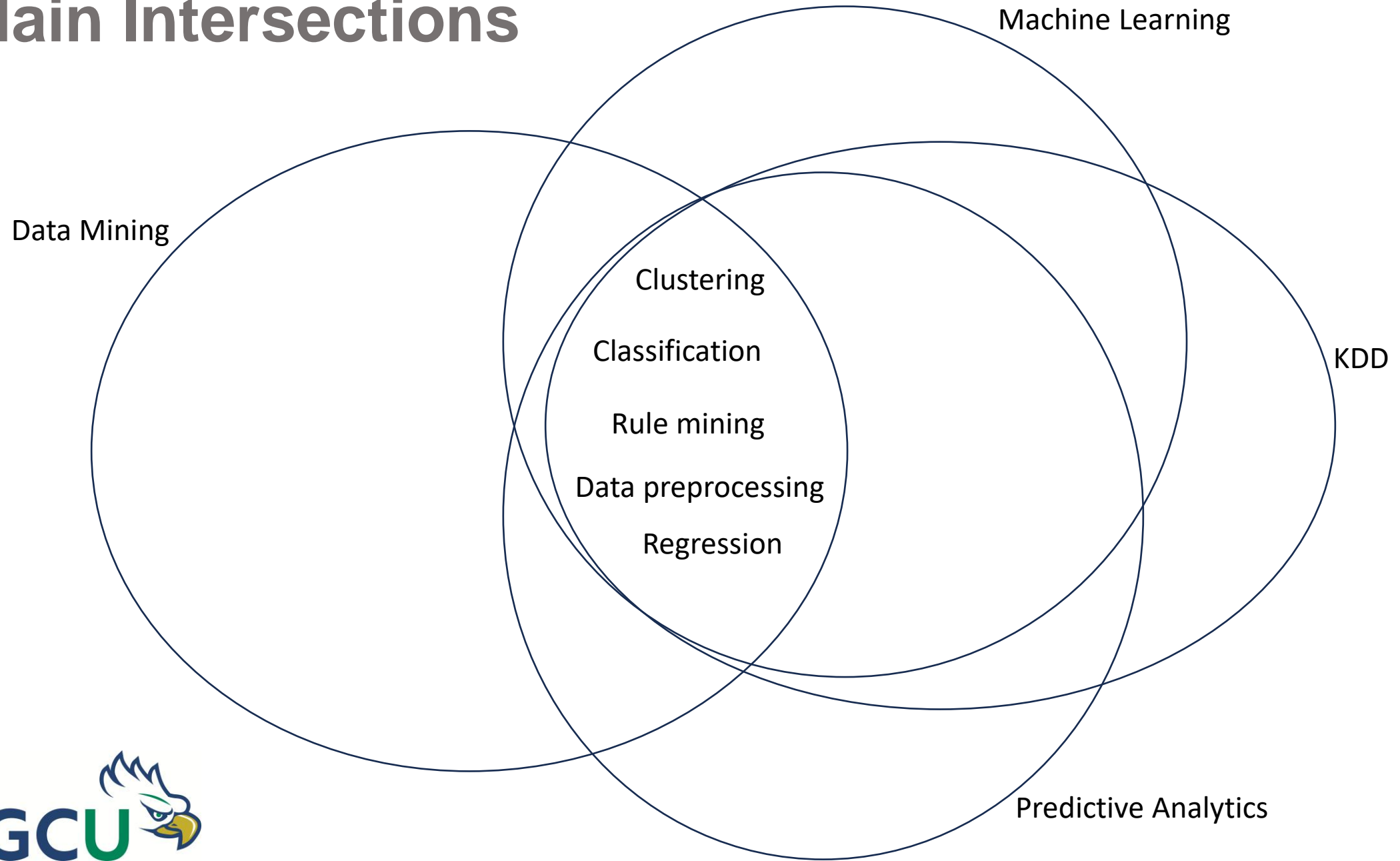


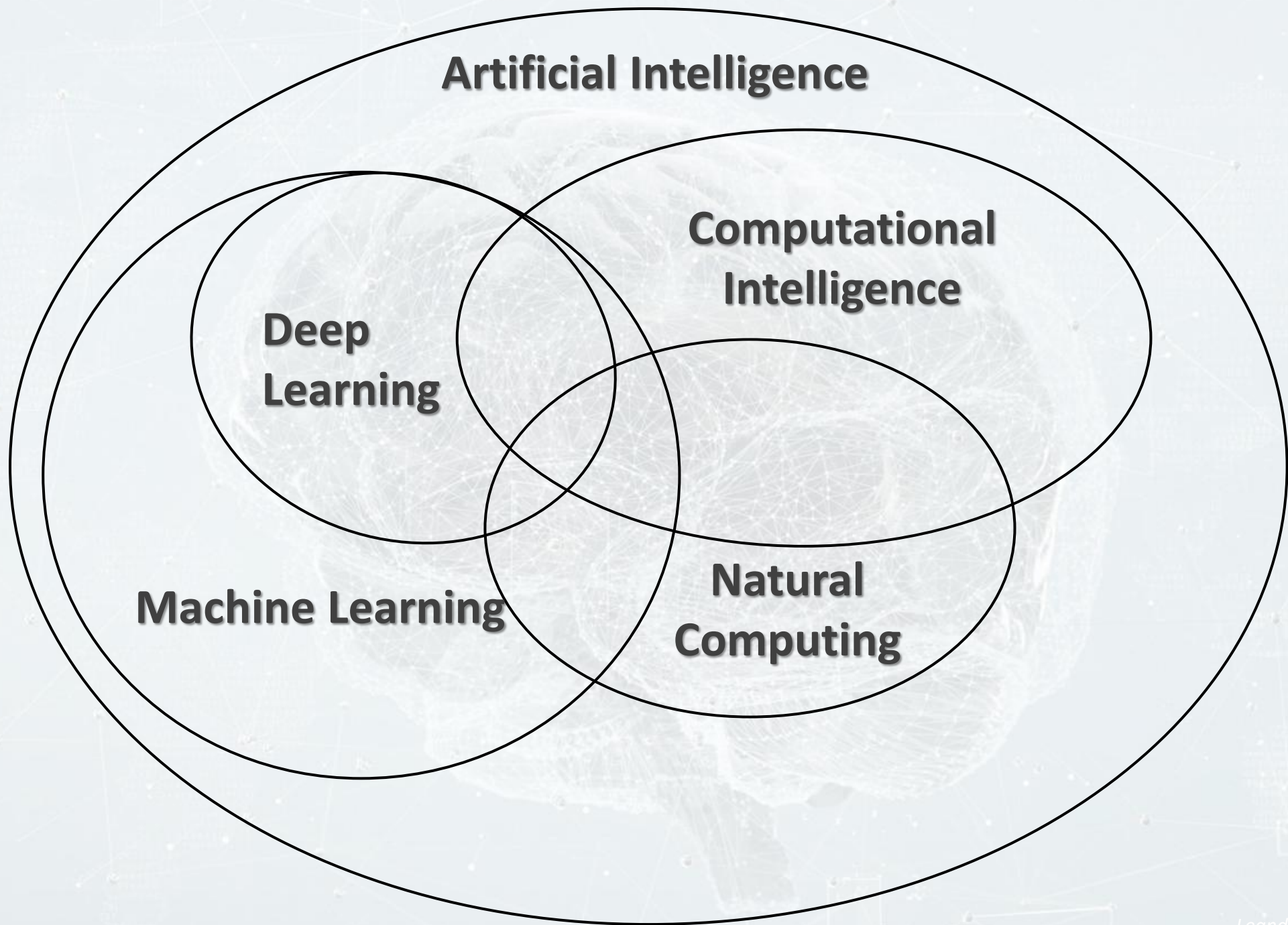
# Some Terminologies and Courses\*\*

---

\*\* There are variations

# Main Intersections





# Leandro Nunes de Castro

**ldecastrosilva@fgcu.edu**

<https://www.linkedin.com/in/Indecastro/>

<https://www.youtube.com/@Indecastro>

**Florida Gulf Coast  
University**



# DA2I: Basic Knowledge

Leandro Nunes de Castro, Ph.D.

[Idecastrosilva@fgcu.edu](mailto:Idecastrosilva@fgcu.edu)

# **From Bad to Good Practices**

## **(Group Activity)**

---



# Ticket Volume per Month: What's Wrong?

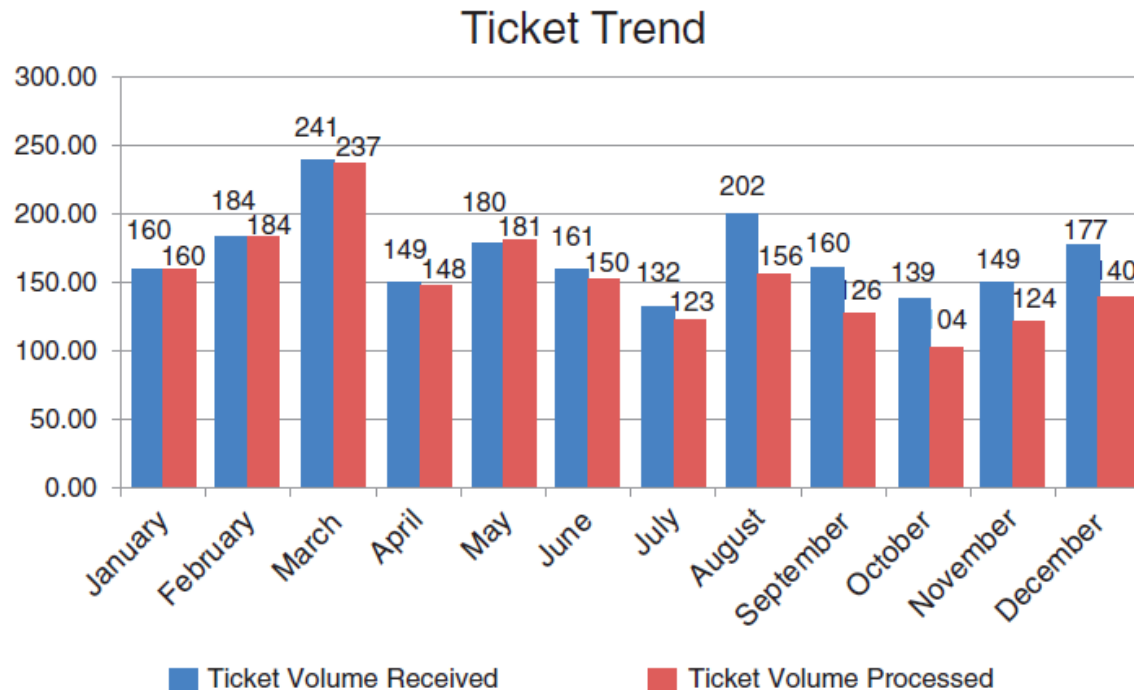


FIGURE 0.2 Example 1 (before): showing data

# Ticket Volume per Month: A Better Presentation

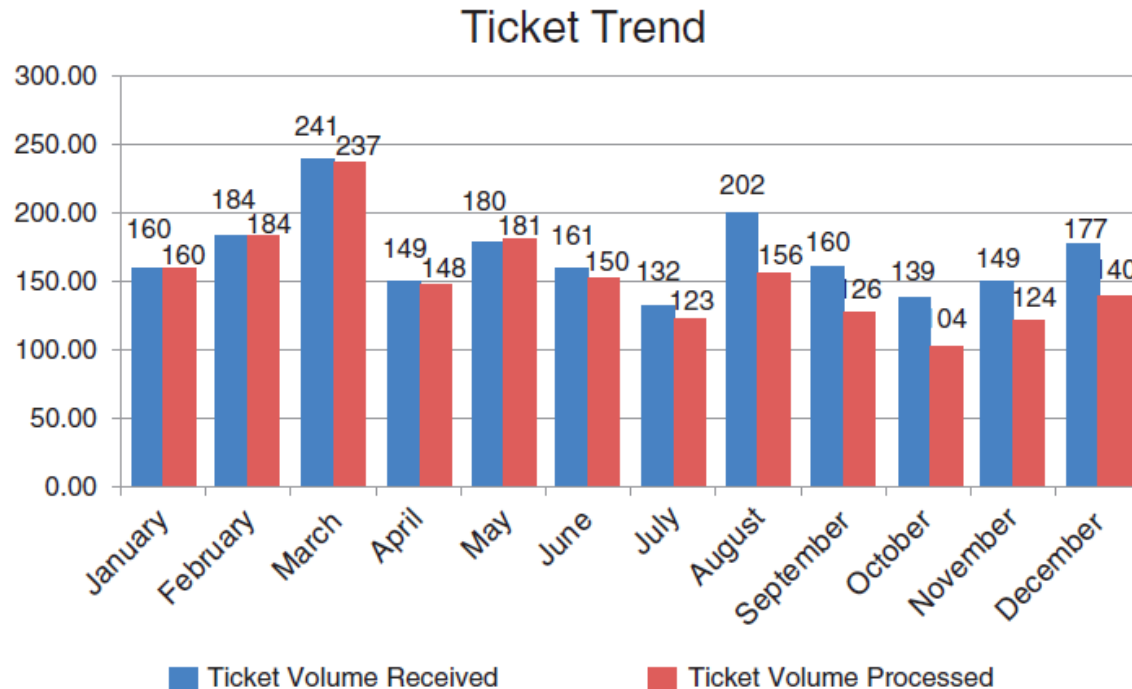


FIGURE 0.2 Example 1 (before): showing data



Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

FIGURE 0.3 Example 1 (after): storytelling with data

# Survey Results: What's Wrong?

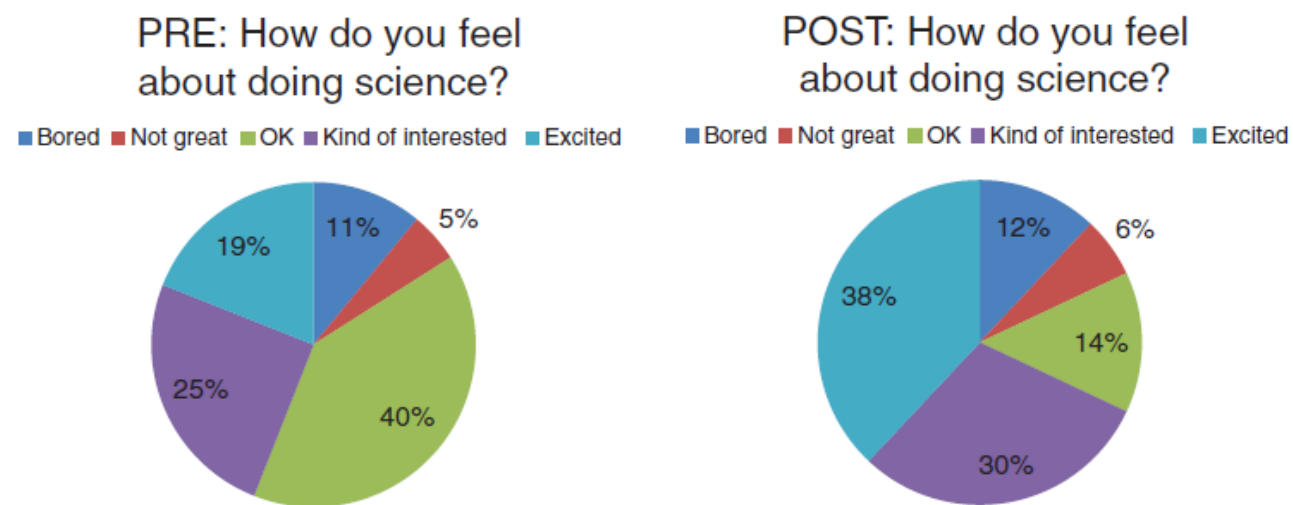


FIGURE 0.4 Example 2 (before): showing data



# Survey Results: A Better Presentation

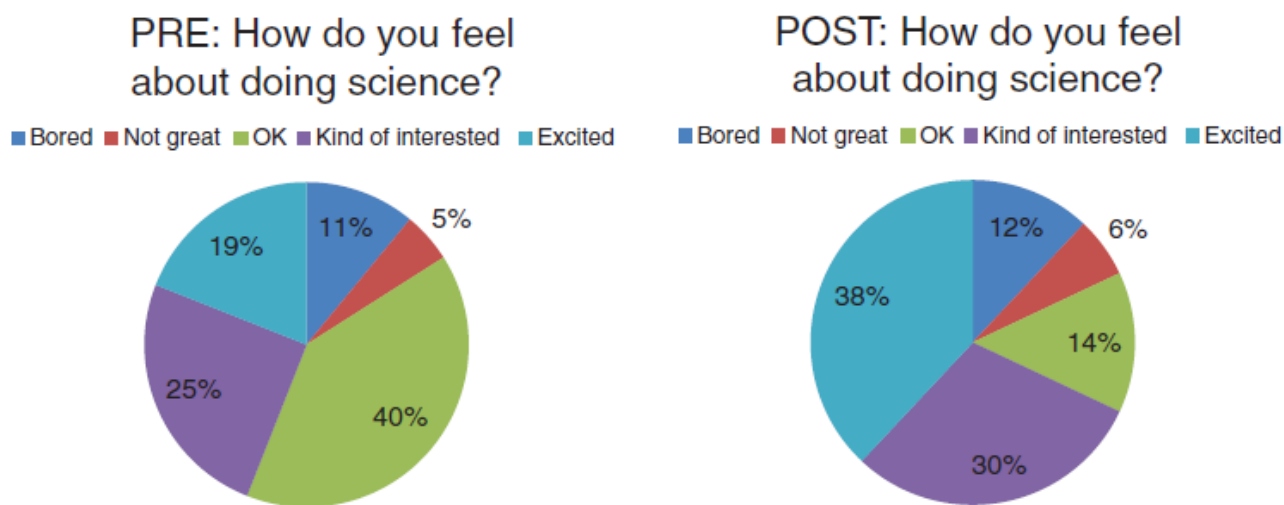
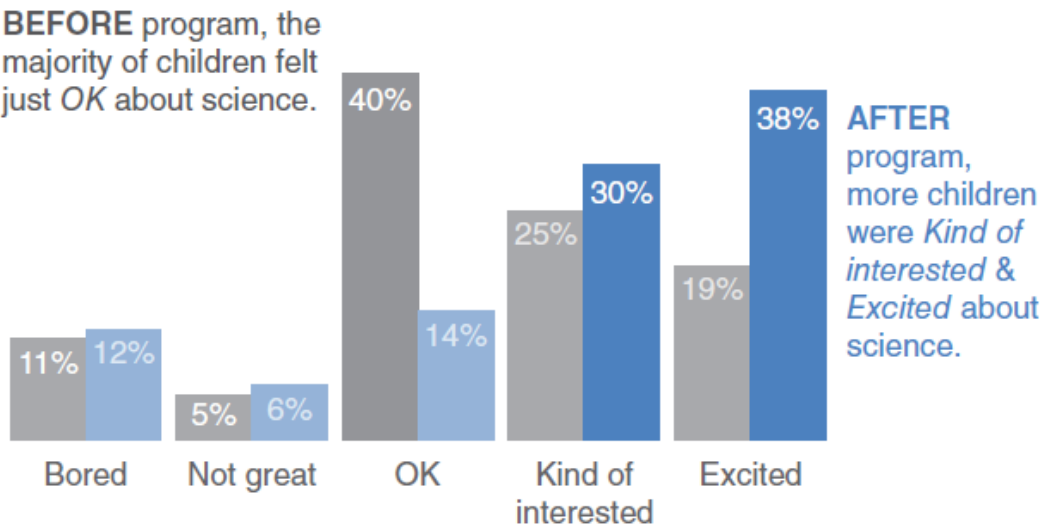


FIGURE 0.4 Example 2 (before): showing data

How do you feel about science?



Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 0.5 Example 2 (after): storytelling with data

# Avg Product Price per Year: What's Wrong?

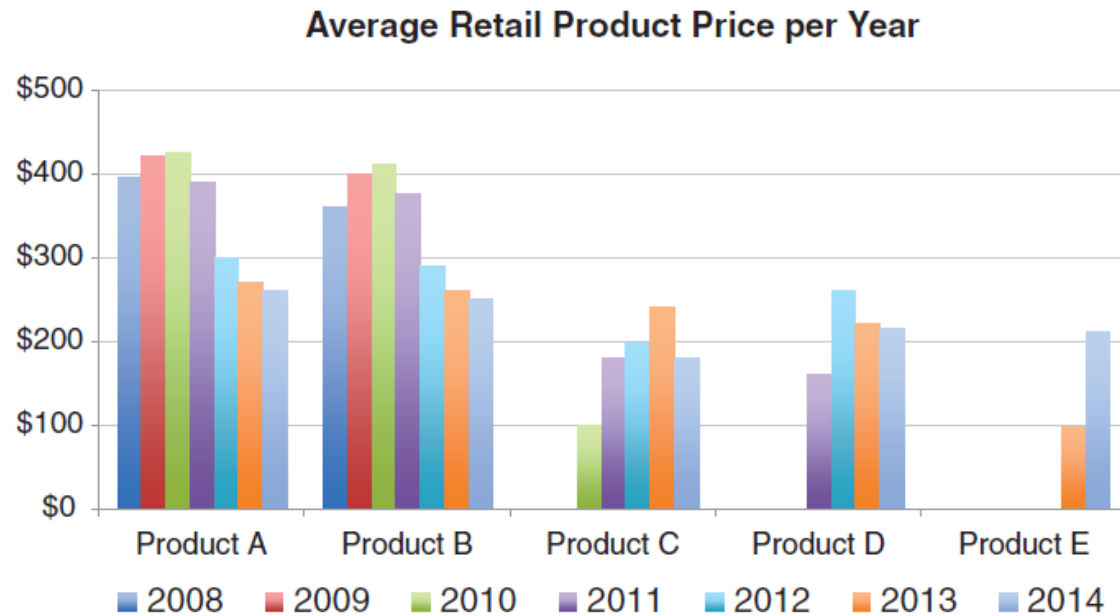


FIGURE 0.6 Example 3 (before): showing data

# Avg Product Price per Year: A Better Presentation

To be competitive, we recommend introducing our product *below the \$223 average price point in the \$150–\$200 range*

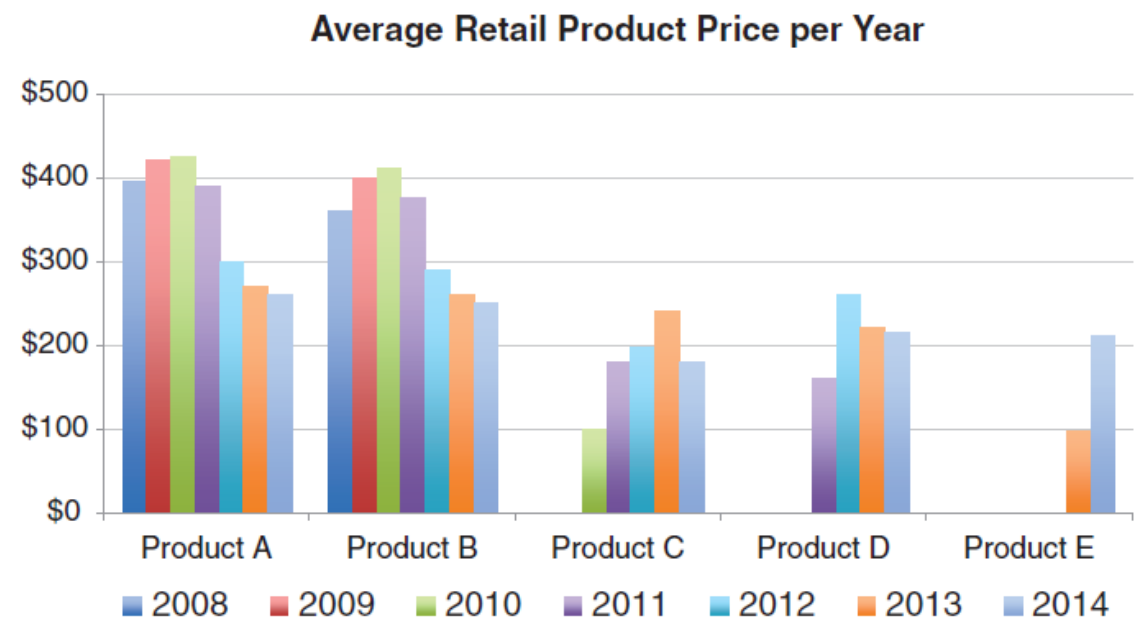


FIGURE 0.6 Example 3 (before): showing data

Retail price over time by product

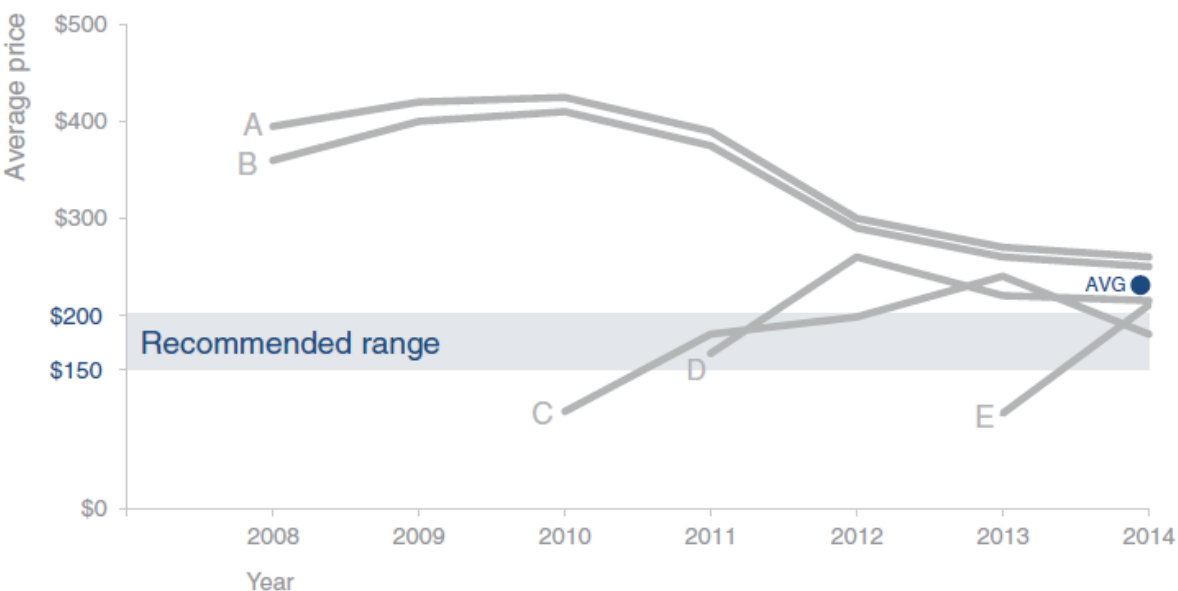


FIGURE 0.7 Example 3 (after): storytelling with data

# Analyzing Visuals with AI

- For the image file Class01\_FromGoodtoBadPractices\_Pictures.png use the AI tools to analyze the quality and choice of the visuals presented in the file. You may want to try Grok.com as well!

*“I want you to act as a data analyst and analyze the attached visuals in the search for improvements or better ways of visualizing the data.”*

- How do you compare the AI analysis with the one performed by your group?

# Data Visualization Showcase

---



# Data Visualization Showcase

- Ask students to bring in examples of effective or interesting data visualizations they have come across. (*10' search*)
- Have a class discussion on what makes these visualizations impactful, discussing design principles, clarity, and the story they convey. (*15' discussion*)

# Data Storytelling Case Studies

---

# Case Study 1: E-commerce Sales Analysis

## Objective:

Analyze and present insights from a dataset containing e-commerce sales data over the past year.

## Dataset Overview:

Contains information on product categories, sales revenue, customer demographics, and purchase patterns.

## Tasks:

1. Identify key elements in the dataset, such as top-selling products, peak sales periods, and customer preferences.
2. Propose a compelling data story that highlights trends, challenges, and opportunities within the e-commerce platform.
3. Design a layout that effectively communicates the story, including visualizations like tables and graphs.

# Case Study 2: Employee Satisfaction Survey

## Objective:

Interpret and communicate insights from an employee satisfaction survey conducted within a large organization.

## Dataset Overview:

Responses from employees on various aspects like work-life balance, job satisfaction, team collaboration, and feedback on company policies.

## Tasks:

1. Identify key themes in the employee satisfaction survey data, such as common concerns, positive trends, and areas for improvement.
2. Craft a data story that addresses the overall sentiment among employees, focusing on changes over time and potential correlations.
3. Develop a layout that effectively communicates the data story, utilizing visualizations and narratives to engage the audience.

# Case Study 3: Social Media Engagement

## Objective:

Analyze and present social media engagement metrics to improve business performance.

## Dataset Overview:

Includes data on social media platforms (likes, shares, comments), website traffic generated from social media, and business data related to social media campaigns.

## Tasks:

1. Identify key performance indicators (KPIs) for social media engagement.
2. Develop a data story that highlights the impact of social media efforts on the organization's goals, emphasizing success stories and areas for growth.
3. Propose a layout that effectively conveys the narrative, incorporating social media analytics, success metrics, and testimonials.

# Case Study 4: Public Health Analysis

## Objective:

Analyze public health data to identify patterns, correlations, and potential interventions for a specific health issue.

## Dataset Overview:

Contains data on disease prevalence, demographic factors, environmental variables, and healthcare access.

## Tasks:

1. Identify key patterns and correlations within the public health data, such as geographic hotspots, demographic vulnerabilities, and potential causes.
2. Construct a data story that communicates the urgency of the health issue, proposed interventions, and potential impact on affected communities.
3. Design a layout that effectively conveys the data story, utilizing maps, charts, and narratives to engage stakeholders and advocate for change.

# Leandro Nunes de Castro

**ldecastrosilva@fgcu.edu**

<https://www.linkedin.com/in/Indecastro/>