Strategies & Best Practices for Teaching Data Science & AI Literacy

STEM Forward sySTEMnow Conference
November 1, 2023

Mark Zachar, Northwestern Mutual Data Science Institute

Ben LaDuke, UW-Milwaukee

Patrick Williams, UW-Milwaukee & AnYong Analytics LLC

Quentin Prince, UW-Milwaukee, Journey House & MKE Tech Hub Coalition

Session Agenda

1. 4 x 5min Lightening Introduction-Presentations

2. Best Practices Panel Discussion

3. Audience Q&A

Mark Zachar, Northwestern Mutual Data Science Institute

FOR 165+ YEARS,
NORTHWESTERN MUTUAL
HAS BEEN HELPING FAMILIES
AND BUSINESSES ACHIEVE
FINANCIAL SECURITY

Revenue $35 billion

#111 on FORTUNE 500

4.6+ Million Clients

9,300+ Financial Professionals

6,500+ Employees

Headquartered in
Milwaukee, Wisconsin

4

NORTHWESTERN MUTUAL
DATA SCIENCE INSTITUTE

Northwestern Mutual · MARQUETTE UNIVERSITY · UNIVERSITY of WISCONSIN UWMILWAUKEE

RESEARCH
Produce internationally recognized data research

TALENT
Build a strong pipeline of talent for the region

COMMUNITY
Strengthen our community through data science

Foundational Achievements 2018 - 2022

RESEARCH & INNOVATION	TALENT	COMMUNITY & PARTNERSHIPS
▪ **3 Sponsored research projects**	▪ **Internship talent pipeline**	▪ **Student capstone projects**
▪ **Research seminars**	▪ **Three new academic programs**	▪ **IMPACT Speaker Series**
▪ **NSF Engines grant app**	▪ **K-12 outreach**	▪ **Membership exploration**

Data Science Skilling: Lessons Learned

Data Science
- **Apply insights derived from data** to improve outcomes (business processes, etc.)

Data Analysis
- Analyze data and **generate insights from it**

Data Literacy
- Read, write, and discuss data **and the insights derived from it** in context
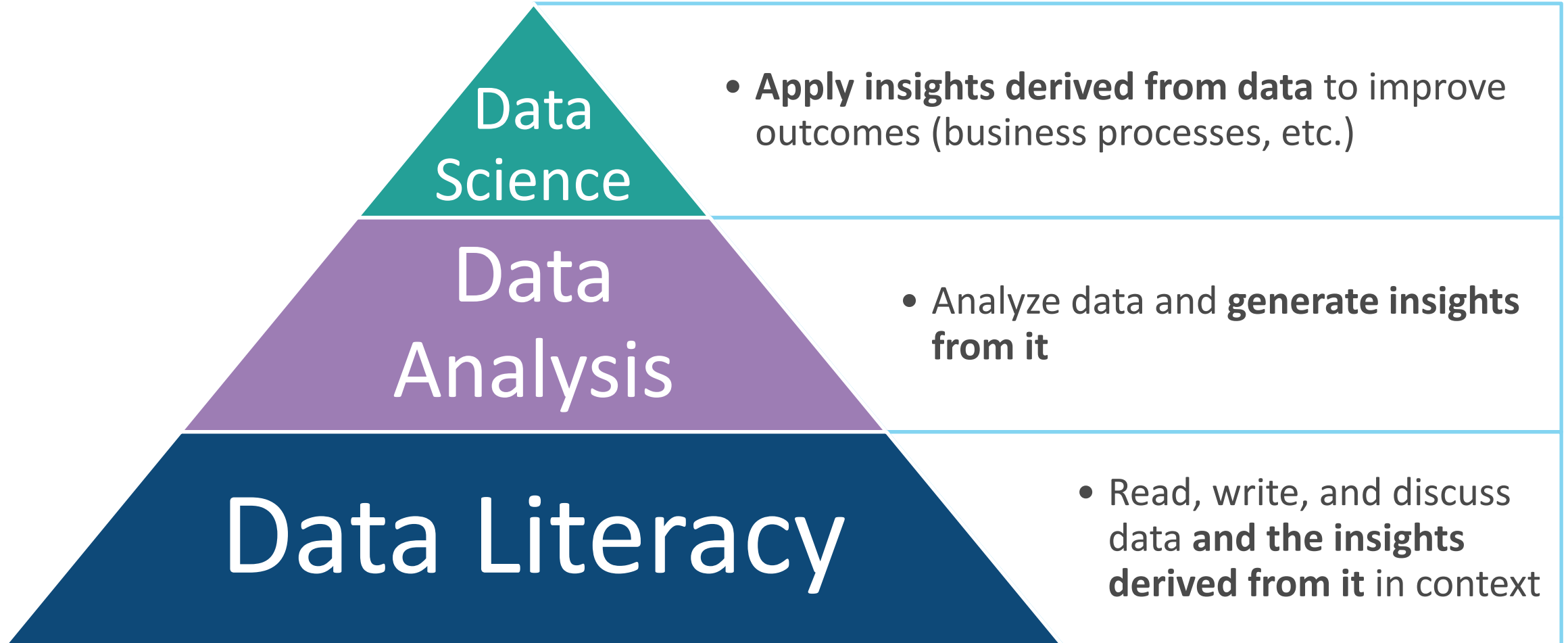
Foundational Knowledge vs Role-Specific Skills

Analytics Engineering

Data Analysis

Data Engineering

Data Governance

Data Science

Domain-specific knowledge

Add depth to foundational knowledge

Foundational Knowledge
SQL, Python*, Cloud, Bias, Statistics, Data Visualization

Core knowledge
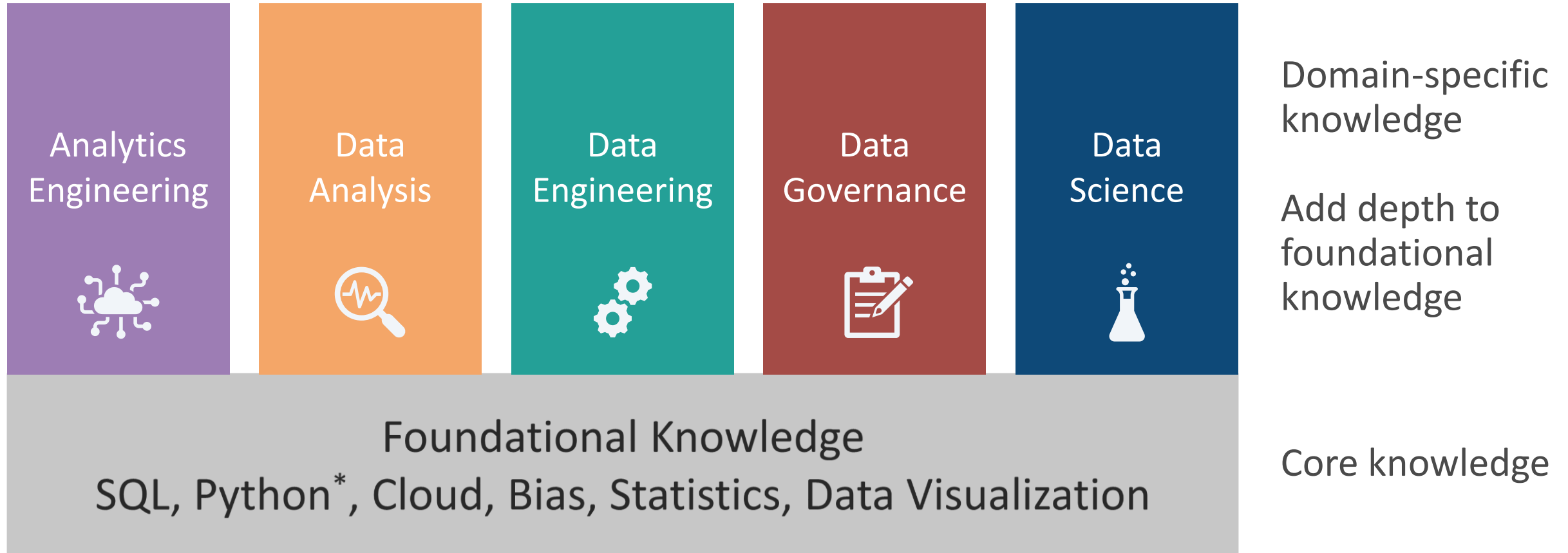
Ben LaDuke – Program Director, UWM College for Kids & Teens
UW-Milwaukee
School of Continuing Education

COLLEGE FOR KIDS & TEENS

The UWM College for Kids & Teens **Data Science Certificate Track** – In Partnership With Northwestern Mutual Data Science Institute

As part of a partnership with the Northwestern Mutual Data Science Institute, UWM's College for Kids & Teens is offering a certificate track that consists of a total of five courses related to the study of data science. In order to complete the certificate, the student must complete all three core classes offered every summer and two elective courses from a selection of 14 College for Kids & Teens offerings. Register for three core courses and two elective courses from a selection of offerings. All core courses are for grades 9-12 and can be completed within a three-year window.

Note: The elective course offerings are highlighted with the following symbol DATA *in the Schedule at a Glance. *Not all elective courses will be offered every summer. For more information on elective courses, course descriptions, course format and certificate specifics, please see our website.*

Core Courses

Introduction to Data Science (Hybrid)
$299
Pat Williams
8410-133, June 19-30, 9am-1:30pm
Gr. 9-12, Session I

Entrepreneurial Data Science (Hybrid)
$299
Quinten Prince
8410-209, July 10-21, 9am-1:30pm
Gr. 9-12, Session II

Statistical Story Telling: A Study of Probability & Statistics (Hybrid)
$299
Steve Augustine
8410-282, July 24-Aug 4, 9am-1:30pm
Gr. 9-12, Session III

Data Science Certificate Track - In Partnership With Northwestern Mutual Data Science Institute

As part of a partnership with the Northwestern Mutual Data Science Institute, UWM College for Kids & Teens is now offering a certificate in data science. This specialized learning track consists of a total of five courses related to the study of data science – introducing students to the exciting and dynamic world of data!

In order to earn the certificate, students must complete all three core classes offered every summer and two elective courses from a selection of 19 College for Kids & Teens offerings. All core courses are for grades 9-12 and can be completed within a three-year window. Elective courses may be offered for grades 6-12 and can be taken at any time previous or simultaneous to the finish of core classes.

Learning Outcomes

- Demonstrate an understanding of the basics of data science.

- Use data in meaningful ways and explore how it can be used in everyday life.

- Apply an introductory understanding of Python programming techniques.

- Examine the relationship between statistics and data.

- Utilize data and design thinking in an exploration of entrepreneurship.

Courses

Register for three core courses and two elective courses from a selection of offerings. All core courses are for grades 9-12 and can be completed within a three-year window.

Core Courses – 9-12th grade

➢ Introduction to Data Science

➢ Entrepreneurial Data Science

➢ Statistical Story Telling: A Study of Probability & Statistics

Elective Courses – 6-12th grade

➢ Advanced Geometry

➢ Adventures in Statistics

➢ Algebra Camp

➢ Algebra I

➢ Algebra II

➢ Build Your Own Business

➢ Discrete Mathematics and Combinatorics

➢ Introduction to Programming I &II

➢ Pre-Calculus

➢ Programming Primer

➢ Python Programming

➢ Roll the Dice

➢ Social Media Analytics

➢ Sportistics

➢ Statistical Literacy

➢ Survey of Calculus

➢ Trigonometry

➢ What's the Chance?

Certificate Data – Student Successes

In 2020, we began our certificate in partnership with Northwestern Mutual Data Science Institute (NMDSI). Due to Covid-19 we started the program as an entirely online program. In the four summers since, we have had more than our share of success stories. The classes are now offered each summer as in-person and hybrid to students in two-and one-week courses.

- 11 total students have completed the certificate - Including 3 this last summer 2023

- Since 2020, 161 students have completed at least 1 course towards this certificate, 34 students have completed at least 1 Core Course

- 7 students need to complete 1 Core Course to complete the certificate (5 of which also need to be complete elective courses)

- 2 students have completed all 3 Core Courses but need to finish elective classes

Our three instructors; Patrick Williams, Quentin Prince, and Stephen Augustine have developed insightful and open-ended learning environments for Pre-College students to start their learning journey in Data Science. With the help of our partners at the NMDSI and volunteers that have worked in the field we look forward to continuing to offer these unique learning opportunities for youth in the greater Milwaukee area.

Patrick Williams, UW-Milwaukee & AnYong Analytics LLC

Vitae

- President, AnYong Analytics® - Data Science Consulting (see AnYongAnalytics.com);
- Retired in 2018, Data Scientist at SAS® Advanced Analytics Lab;
- 36 Years of Experience in Predictive Modeling & Analytics Consulting and Education;
- MA and BA Economics, UW-Milwaukee, Graduate TA;
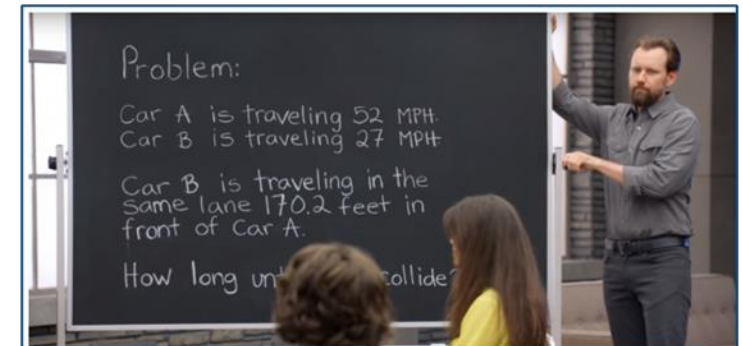- Instructor, UW-Milwaukee College for Kids & Teens Introduction to Data Science Course

Introduction to
Data Science

Course Approach

- **Hyper Hands-On**

  o Students install, configure and integrate Python®, R®, SQL and Jupyter Notebook®

  (Jupyter Notebook® is the Interactive Development Environment, or IDE)

  o on their own machines

  o using DOS (Windows®) or LINUX (Mac®) commands

  o to download, install and configure their data science environment

- **Hyper Immersive**

  o Students interact with increasingly complex code in IDE and batch

  o to create or extract, transform and load data (ETL)

  o analyze and visualize data

  o perform analyses and predictive modeling

  o on cross-sectional, time series and pooled data (e.g., COVID by location over time)

First Code Problem >

Problem:
Car A is traveling 52 MPH.
Car B is traveling 27 MPH.

Car B is traveling in the
same lane 170.2 feet in
front of Car A.

How long un... ...ollide?

```python
# And now, the solution, coded in Python:
# The distance differential is 170,2 feet. The speed differential is 25 MPH.
# Let's convert the speed differential to feet per second, then divide it into the
# distance differential (space between the cars), to solve for how many seconds it
# will take for Car A to collide with Car B.

distance_diff = 170.2                      # in feet
speed_diff = 25 / 1                         # in miles per hour
speed_diff = (5280 * 25) / (60 * 60)       # in feet per second
crash_time_diff = distance_diff / speed_diff
print('Collision time will be in ' + str(round(crash_time_diff, 2)) + ' seconds.')

# Note how the answer (output) for the code in this cell is printed below this cell.

Collision time will be in 4.64 seconds.
```
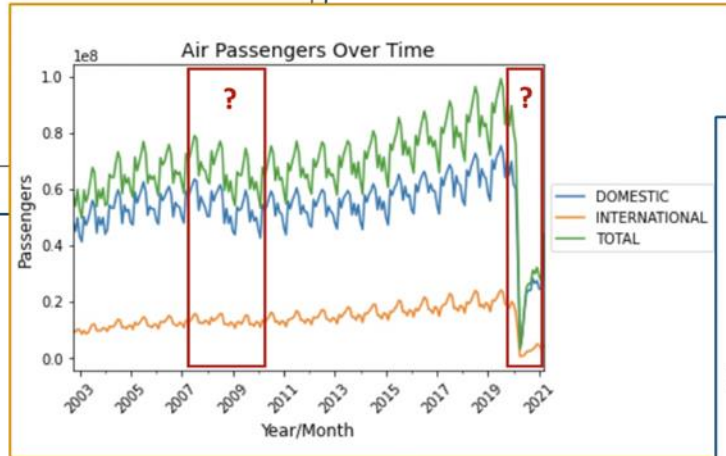
Course Approach
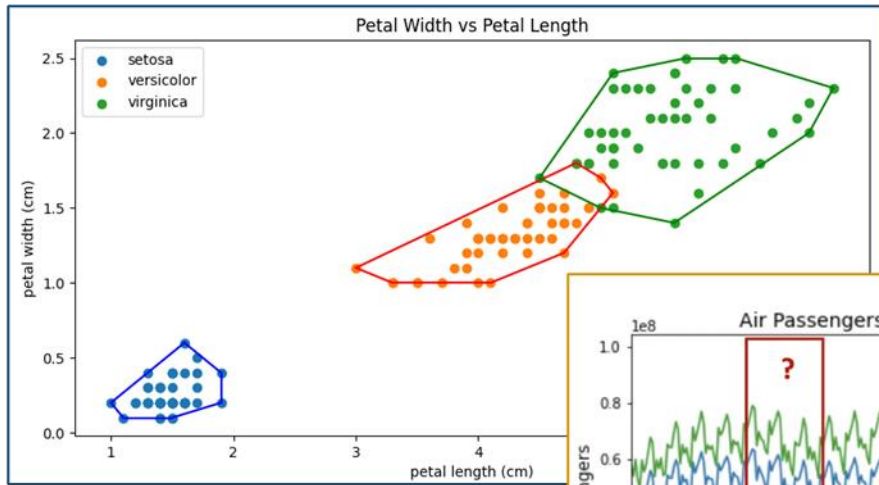
> Cluster Iris Data

Install & Config >

Explain Airline Passenger Data >

Classic Datasets and COVID Data

Model & Predict Salary Data >
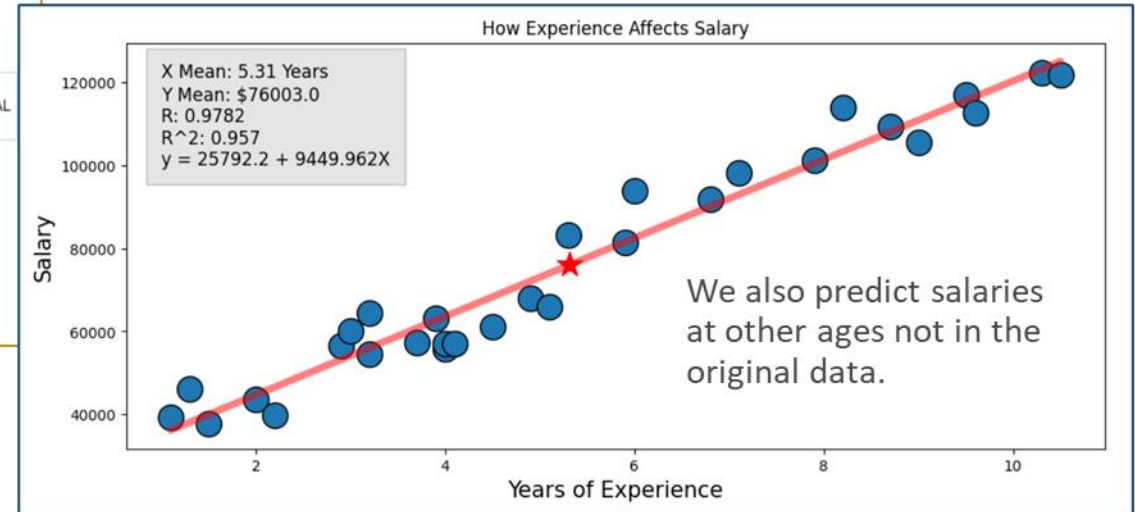
We also predict salaries at other ages not in the original data.

Introduction to
Data Science

Course Contents and 2024 Enhancements

Environment Preparation - 9:45 first Monday
1. Windows 10 Machines
2. Chromebook Machines
3. Mac Machines

I. Course Introduction
1. What is Data Science?
2. Will There be Any Math?
3. Will There be Any Tests or Homework?
4. What is "Big Data"?
5. What is "the Cloud"?
6. What are "in-memory" and "parallel" processing?
7. What are AI and M L?
8. What are Python, R, SAS and SQL?
9. Skills Requirements
10. Let's Look at Some Data!

1/ These items will be added in 2024.
2/ Currently only Tableau is explored. Power BI will be added in 2024.

II. Some Basics – "Pull up your socks!"
1. Spreadsheet Basics
2. Jupyter Notebook Basics
3. Code Execution, Python, R and SQL Basics

III. ETL & Data Sourcing
1. What is ETL?
2. Let's Get Some Data
3. Examining Data
   A. Cross-Sectional Data
   B. Time Series Data
   C. Pooled Cross-Sectional Time Series Data
4. Organizing Data in a Database - SQLite3 1/

IV. Data Exploration & Visualization – "Wudda we got"?
1. Exploring Data
   A. Descriptive Statistics
   B. Correlation of Data
2. Visualizing Data
   A. In Python and R
   B. In Tableau and Power BI 2/

V. Model Building & Implementation
1. What is a Model?
   A. Cross-Sectional Models
   B. Time Series Models
   C. Pooled Cross-Sectional Time Series Models
   D. "Show Me the Money!" - Evaluating Model Accuracy
2. Let's Build Some Models
   A. Linear Regression
      1. Cross-Sectional
      2. Time Series 1/
      3. Cross-Sectional Time Series (if time allows) 1/
3. Models that Group Data into Clusters
4. "What Now?" - Model Implementation 1/

Introduction to Data Science

2024 Course AI Enhancements: ChatGPT and Bard (TEST YOUR CODE!)

Default (GPT-3.5)

Bard Experiment

Settings

Help

✓ Bard was just updated. See update

P read a .csv file using Pandas

P Write a SQL program that summarizes all numeric columns to the state/month level, and then takes a one-period lag for each numeric variable by state

To read a CSV file using Pandas, you can utilize the `pandas.read_csv()` function. Here's an example of how to read a CSV file:

The following SQL program summarizes all numeric columns to the state/month level, and then takes a one-period lag for each numeric variable by state:
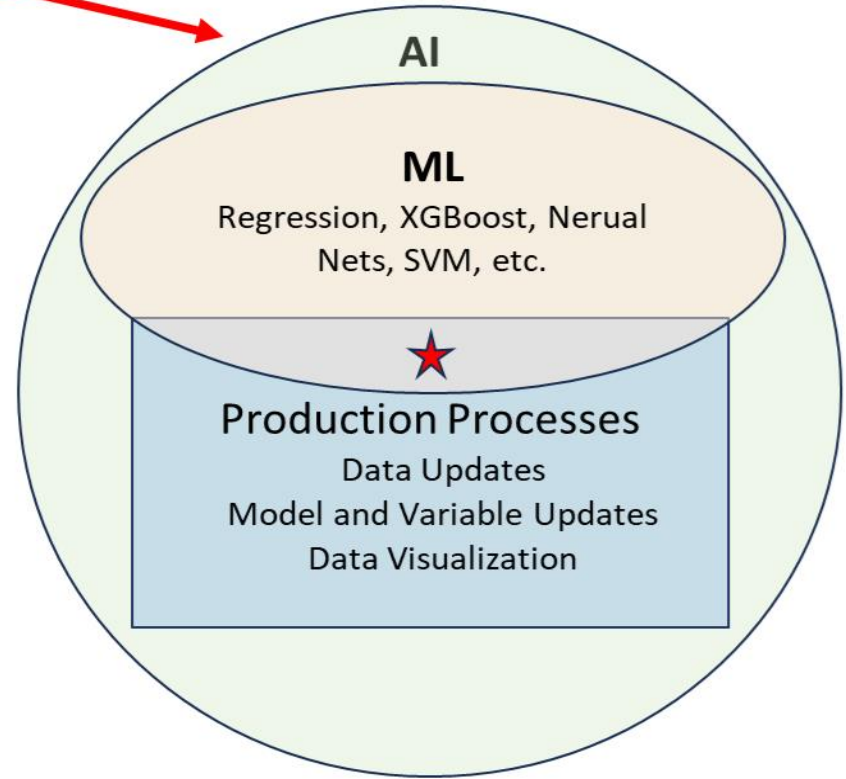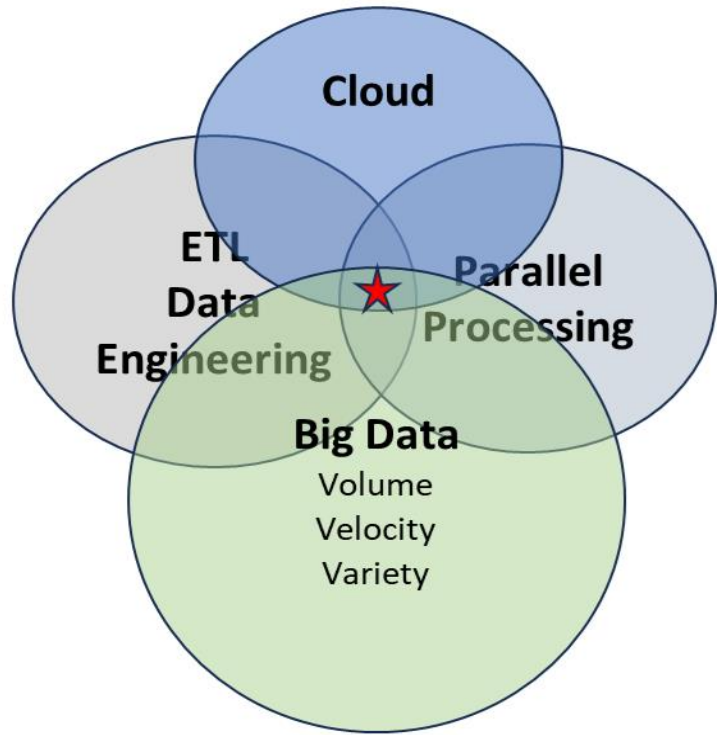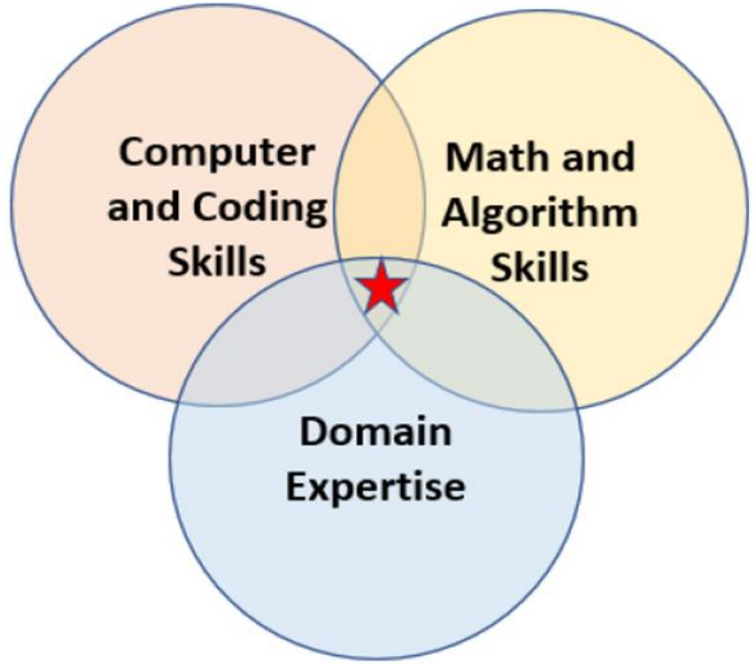
```python
import pandas as pd

# Specify the file path
file_path = "path/to/your/file.csv"

# Read the CSV file into a DataFrame
df = pd.read_csv(file_path)

# Print the DataFrame
print(df)
```

```sql
SELECT
  state,
  month,
  SUM(column_1) AS column_1_sum,
  AVG(column_2) AS column_2_avg,
  MAX(column_3) AS column_3_max,
  LAG(SUM(column_1), 1) OVER (PARTITION BY state ORDER BY month) AS column_1_lag,
  LAG(AVG(column_2), 1) OVER (PARTITION BY state ORDER BY month) AS column_2_lag,
  LAG(MAX(column_3), 1) OVER (PARTITION BY state ORDER BY month) AS column_3_lag
FROM table_name
GROUP BY state, month
ORDER BY state, month;
```

Copy code

Regenerate

Send a message

Enter a prompt here

Data Science - Observing/Analyzing Data

Introduction to Data Science

Skills → **System Resources** →

Computer and Coding Skills

Math and Algorithm Skills

Domain Expertise

Cloud

ETL Data Engineering

Parallel Processing

Big Data
Volume
Velocity
Variety

AI

ML
Regression, XGBoost, Nerual Nets, SVM, etc.

Production Processes
Data Updates
Model and Variable Updates
Data Visualization

Quentin Prince

UW-Milwaukee – Entreprenurial Data Science Instructor
Journey House – Director of Youth Leadership and Student Success
MKE Tech Hub Coalition – Program Manager

Course Description:

This Entrepreneurial Data Science class for high school students is designed to introduce students to the exciting world of data science and entrepreneurship. Over the course of two weeks, students will learn the basics of data science, explore the principles of entrepreneurship, and culminate in a pitch competition where they will apply their knowledge to develop and present their own data-driven business ideas.

Course Objectives:

- Understand the fundamentals of data science and its applications.

- Develop data collection and analysis skills.

- Learn data visualization techniques.

- Explore the principles of entrepreneurship and the traits of successful entrepreneurs.

- Generate innovative business ideas based on data-driven insights.

- Craft and deliver a persuasive business pitch.

- Apply data science and entrepreneurship principles in a real-world context.

Week 1: Understanding Data Science

Day 1 - Introduction to Data Science (Course Overview, Definition and importance of Data Science, Intro to real-world applications)

Day 2 - Data Collection and Analysis (Data collection methods, Data cleaning and preprocessing, Basic data analysis, Hands-on data analysis)

Day 3 - Data Visualization (Importance of data visualization, Tools for data visualization, Creating data visualizations, Group data visualization project)

Day 4 - Guest Speaker (Inviting a guest speaker from a local company or university to share their experiences with data science)

Day 5 - Data Science Project Kickoff (Introduction to the data science project, Group formation for the project, Initial project brainstorming)

Week 2: Exploring Entrepreneurship and Pitch Competition

Day 6 - Introduction to Entrepreneurship (What is entrepreneurship, Traits and skills of successful entrepreneurs, Real-world examples of teenage entrepreneurs)

Day 7 - Ideation and Opportunity (Generating innovative business ideas, Identifying market opportunities, Developing a unique selling proposition, Group discussion and idea sharing)

Day 8 - The Art of Pitching (Elements of a persuasive pitch, Storytelling in entrepreneurship, Crafting a compelling pitch, Pitch practice and feedback)

Day 9 - Preparing for the Pitch (Finalizing and rehearsing pitches, Preparing pitch materials, Pitch practice sessions in groups)

Day 10 - Pitch Competition (Student pitch competition, Each group presents their data-driven business idea, Panel of judges provides feedback and

An entrepreneurial data science program offers numerous benefits to high school students. First and foremost, it equips them with valuable skills that are increasingly in demand in today's data-driven world. By learning data analysis, statistics, and machine learning techniques, students gain a competitive edge in both higher education and the job market.

Moreover, this program nurtures problem-solving and critical thinking skills, enabling students to approach real-world challenges with data-driven solutions. It encourages innovation and creativity, fostering an entrepreneurial mindset that can be applied not only in business but also in various aspects of life. Additionally, the program offers exposure to cutting-edge technologies and industry trends, paving the way for exciting and rewarding career opportunities.

Overall, it empowers high school students with the tools and knowledge they need to excel in the digital age while nurturing their entrepreneurial spirit.
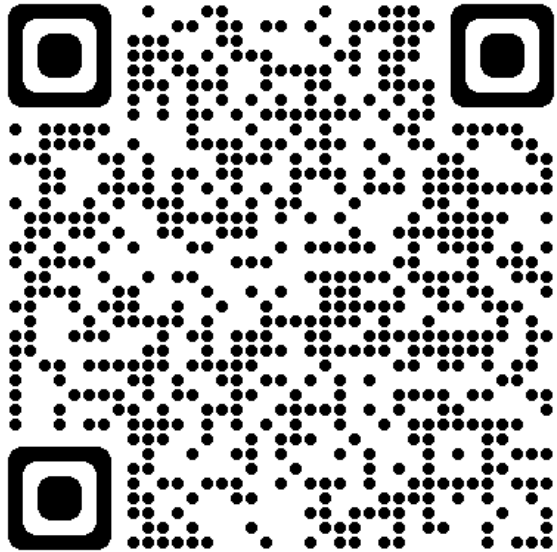
Session Agenda

1. 4 x 5min Lightening Introduction-Presentations

2. Best Practices Panel Discussion

3. Audience Q&A

Stay Connected With Us!

UWM College for Kids & Teens

Northwestern Mutual
Data Science Institute

TEACHER, ADMIN,
STUDENT RESOURCES