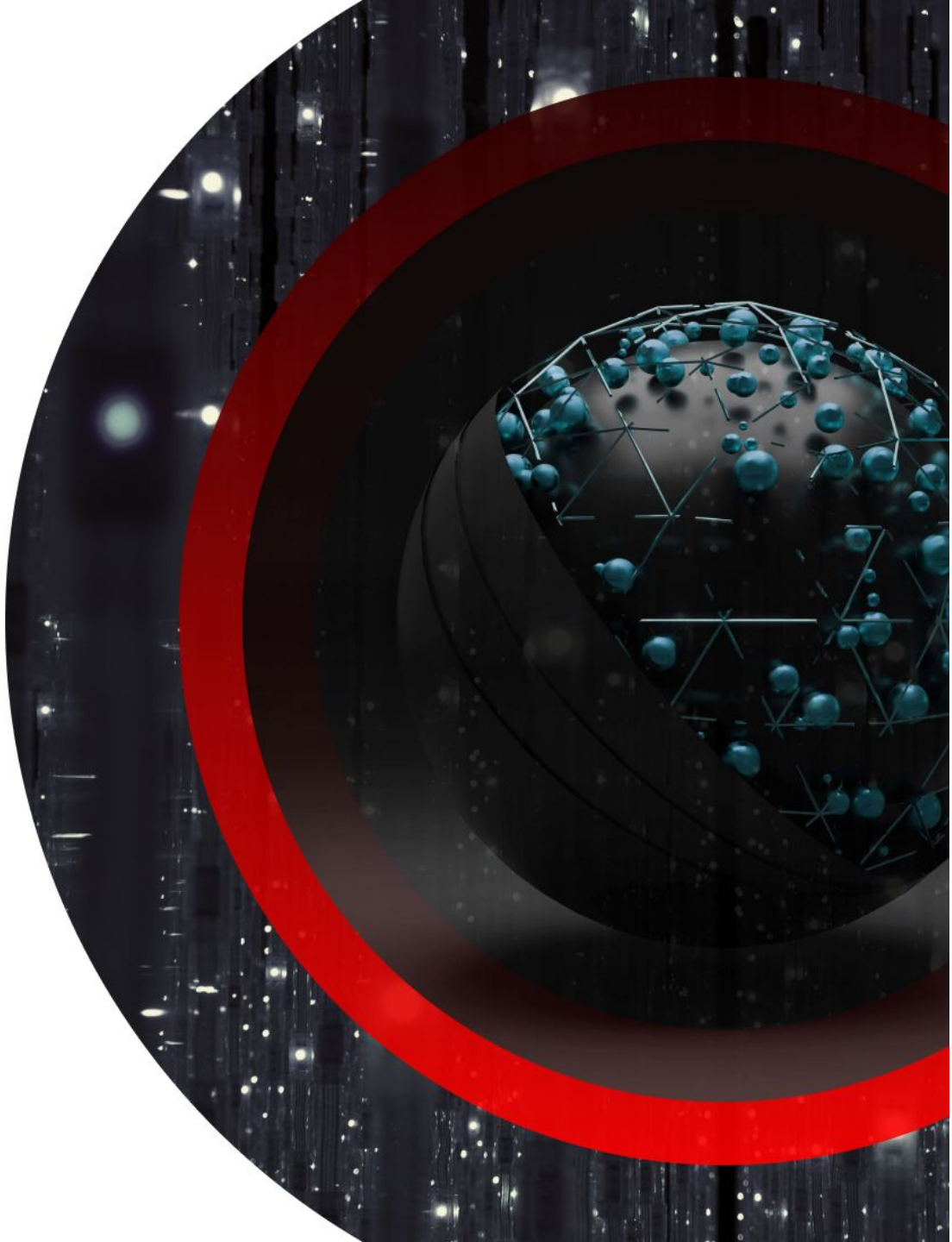
DATA SCIENCE AND ANALYTICS

Introductory Course



CLASS NORMS

COURTESY IN CLASS

Remaining on mute unless called on, exercising courtesy during breakout rooms, using the chat box for questions only

ATTENDANCE

100% attendance is expected and contributes to success in passing the course and the program

PARTICIPATION

Keeping an open mind in discussions and sharing experiences, making contributions during team assignments, submitting assignments in Canvas, and participating in discussion boards

USE OF CLASS RESOURCES

Follow along during the lecture with the lesson companion and download any in-class documents prior to class.

INSTRUCTIONAL TEAM

INSTRUCTIONAL TEAM

Name	Contact information
Name	Contact information
Name	Contact information

INTRODUCTORY COURSE OVERVIEW

- This course provides an accurate representation of the skills you will acquire and the projects you can hope to complete should you choose to pursue the full Data Science and Analytics Bootcamp.

- By the end of this course, you will grasp an understanding of the data-driven business world as you gain exposure to different career paths and a variety of analytic tools.

PROGRAM PATH

1 Introductory Course

2 SQL and Databases

3 Statistics and Probability

4 Data Storytelling

Milestone 1: Building and Presenting Data Stories

5 Python Programming

6 Data Wrangling

7 Visual Communications

8 Advanced SQL Programming

Milestone 2: Data Integration, Preparation, Reporting, and Presentation

9 Business Intelligence

10 Big Data

11 Machine Learning

12 Applied AI

Milestone 3: Capstone Project: Delivering Insights and Presentations

INTRODUCTORY COURSE PATH

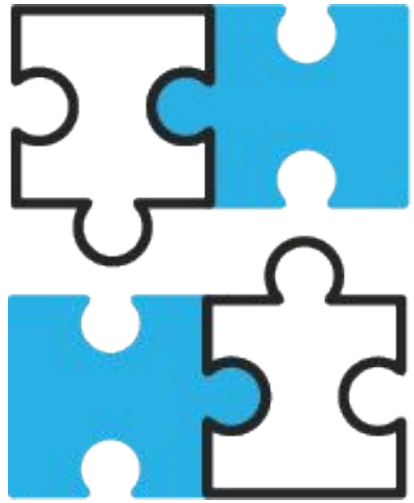
1 **INTRODUCTION TO DATA SCIENCE AND ANALYTICS**			
2 COMPUTING PRIMER	**3** PROGRAMMING CONCEPTS	**4** DISCOVERING AND CURATING DATA	**5** STRUCTURING AND ANALYZING DATA
6 CLEANING AND ENRICHING DATA	**7** VALIDATING AND PRESENTING DATA	**8** INTRODUCTION TO DATA SCIENCE PROJECTS	**9** ASSESSMENT NIGHT

MODULE OUTLINE

Lesson 1: Program Overview

Lesson 2: Getting Started with Data Science and Analytics

Lesson 3: Tools and Software Used by Data Professionals



MODULE GOALS

- Become familiar with the program, learning outcomes, hands-on projects, and program resources.
- Demonstrate understanding of what data science is and what data scientists do.
- Demonstrate an understanding of the tools and technology behind data science.
- Build your data analytics portfolio on GitHub.

LESSON 1: PROGRAM OVERVIEW

DISCUSSION

What attracted you to data science and analytics and made you decide to pursue a career in this field?

LESSON GOALS

WHAT ARE THE GOALS?

- Become familiar with the program, learning outcomes, hands-on projects, and program resources.

WHY ARE THEY IMPORTANT?

- Understand how you play an active role to complete this learning experience successfully.
- Utilize program resources.

PROGRAM OVERVIEW

- This market-driven program equips you with the skills and competencies needed to succeed as a data professional.
- It addresses the skills gap in the workforce caused by technological advancements and the constant evolving of the data field.
- You will experience scenarios and assignments that closely align with data challenges currently faced by businesses.

PROGRAM PATH

1 Introductory Course

2 SQL and Databases

3 Statistics and Probability

4 Data Storytelling

Milestone 1: Building and Presenting Data Stories

5 Python Programming

6 Data Wrangling

7 Visual Communications

8 Advanced SQL Programming

Milestone 2: Data Integration, Preparation, Reporting, and Presentation

9 Business Intelligence

10 Big Data

11 Machine Learning

12 Applied AI

Milestone 3: Capstone Project: Delivering Insights and Presentations

PROGRAM GOALS

- Organize business requirements into data requirements.
- Work responsibly with data.
- Leverage the power of artificial intelligence.
- Build visualizations that are accessible to everyone to help the business achieve its goals.
- Work on a variety of data ingestion, cleaning, and analysis assignments.

YOUR TAKE

What do you wish to accomplish by attending the Data Science and Analytics Bootcamp?

BREAK TIME

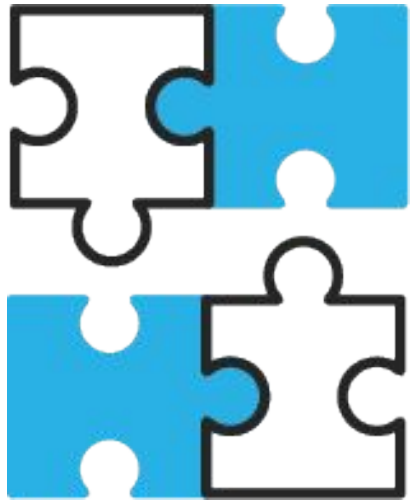
LESSON 2: GETTING STARTED WITH DATA SCIENCE AND ANALYTICS

LESSON OUTLINE

What is data science and analytics?

Careers in data science and analytics

Characteristics of data science and analytics careers



LESSON GOALS

WHAT ARE THE GOALS?

- Demonstrate an understanding of what data science and analytics is.
- Describe business solutions in different industries offered by data science and analytics.
- Recognize the difference between various careers in the data profession.
- Describe the daily job tasks of a data analyst.

WHY ARE THEY IMPORTANT?

- Understand and appreciate the business solutions provided by data science and analytics.
- Gain an understanding and appreciation of the wide variety of possibilities and career choices in the data science and analytics field.
- Reflect on your skills and character and determine which role(s) you would like to play in the data science and analytics world.

WHAT IS DATA?

Data is information in digital form that is stored on a computer or server and used as a basis for analysis reasoning, discussion, or calculation.

Information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful

Source: Stephanie Shen | Towards Data Science

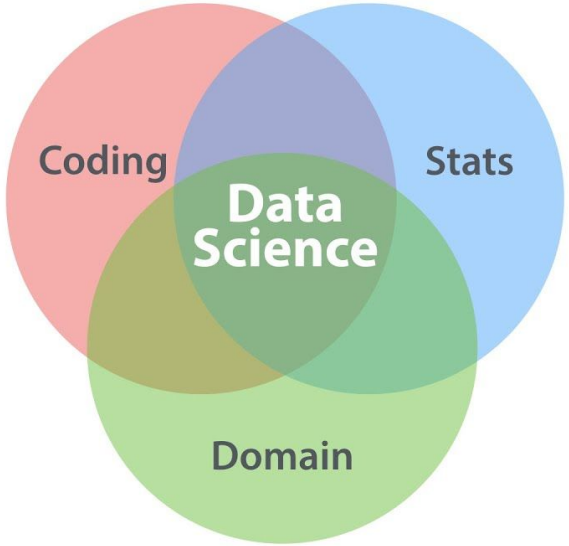
WHAT IS DATA SCIENCE?

Data science combines the scientific method, math and statistics, specialized programming, advanced analytics, AI, and even storytelling to uncover and explain the business insights buried in data.

DATA SCIENCE CORE SKILLS

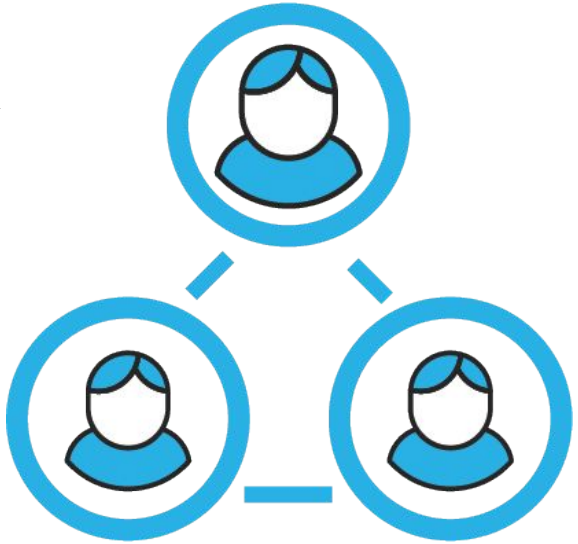
Cross-disciplinary set of skills

- **Statistics** to model and summarize data sets
- **Computer science** to design and use algorithms to store, process, and visualize data
- **Domain expertise** necessary to formulate the right questions and to put the answers in context



DATA SCIENCE IN OUR DAY-TO-DAY LIVES

- New friend suggestions by Facebook
- Google's help to complete a search phrase
- Television shows predicted by Netflix
- Music suggestions on Spotify



FROM DATA TO INSIGHT

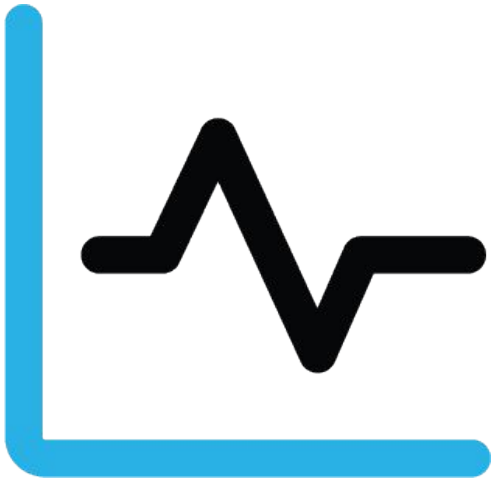
Data insights refer to the strategy businesses use to collect, analyze, and act on data related to the business and its clients and then make better decisions.

- Predictive analytics
- Artificial intelligence (AI)
- Machine learning
- Deep learning models

PREDICTIVE ANALYTICS

“Predictive analytics is a branch of advanced analytics that makes predictions about future outcomes using historical data combined with statistical modeling, data mining techniques, and machine learning. Companies employ predictive analytics to find patterns in this data to identify risks and opportunities.”

Source: IBM

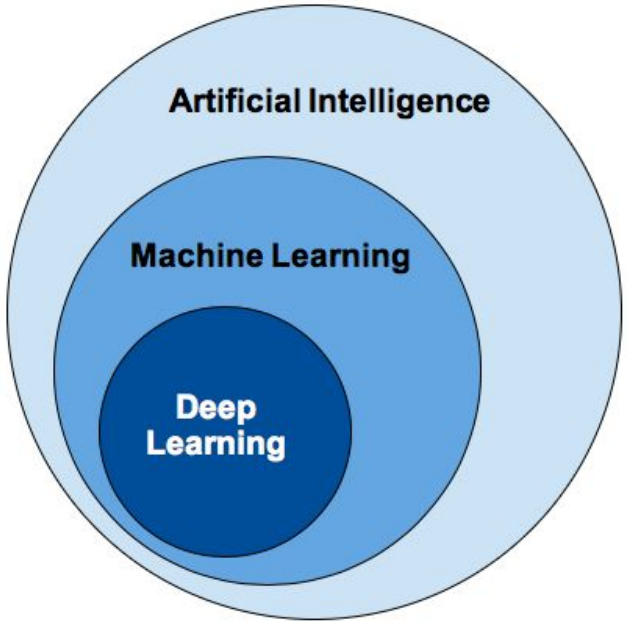


ARTIFICIAL INTELLIGENCE

In computer science, the term artificial intelligence (AI) refers to any human-like intelligence exhibited by a computer, robot, or other machine.

In popular usage, artificial intelligence refers to the ability of a computer or machine to mimic the capabilities of the human mind—learning from examples and experience, recognizing objects, understanding and responding to language, making decisions, solving problems—and combining these and other capabilities to perform functions a human might perform, such as greeting a hotel guest or driving a car.

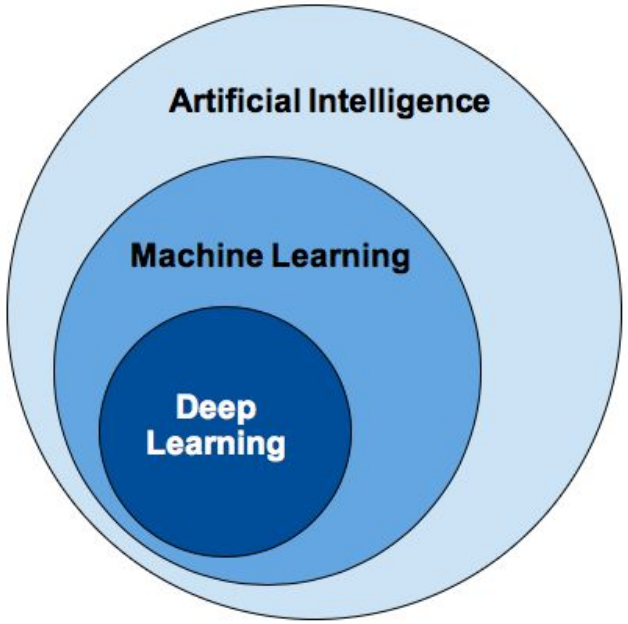
Source: WorldNoor



MACHINE LEARNING

"Machine learning is a form of AI that enables a system to learn from data rather than through explicit programming. However, machine learning is not a simple process. As the algorithms ingest training data, it is then possible to produce more precise models based on that data."

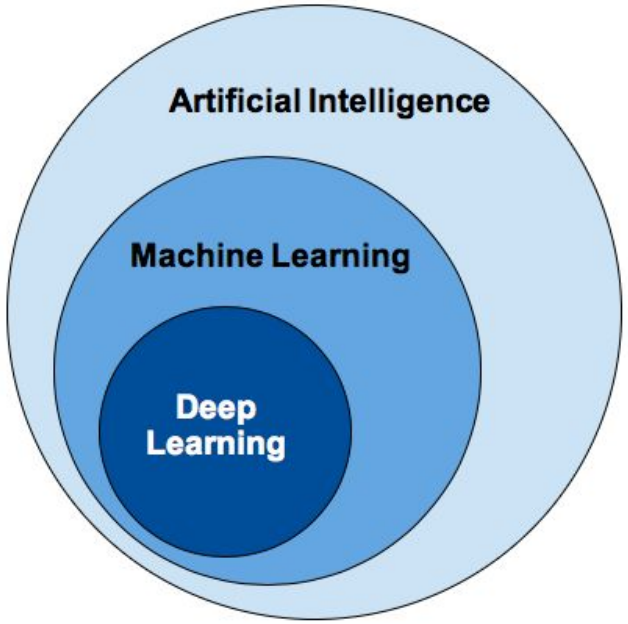
Source: IBM



DEEP LEARNING

“Deep learning is a subset of machine learning in which multi-layered neural networks—modeled to work like the human brain—'learn' from large amounts of data. Within each layer of the neural network, deep learning algorithms perform calculations and make predictions repeatedly, progressively 'learning' and gradually improving the accuracy of the outcome over time.”

Source: Tech Grid



DATA SCIENCE IN TODAY'S BUSINESSES

Many businesses have been able to transform their data into insights for better decision-making, better customer experience, and increased efficiency.

- People Analytics (Human Resources)
- Marketing Analytics

- Operations Analytics
- Production and Manufacturing

- Supply Chain Analytics
- Customer Service Analytics

DATA SOLUTIONS USING PREDICTIVE ANALYTICS

Insurance and banking companies

- Fraud detection
- Risk analysis
- Claims and processing analysis
- Process optimization

From: QuantZig

DATA SOLUTIONS USING PREDICTIVE ANALYTICS (CONT.)

Health care

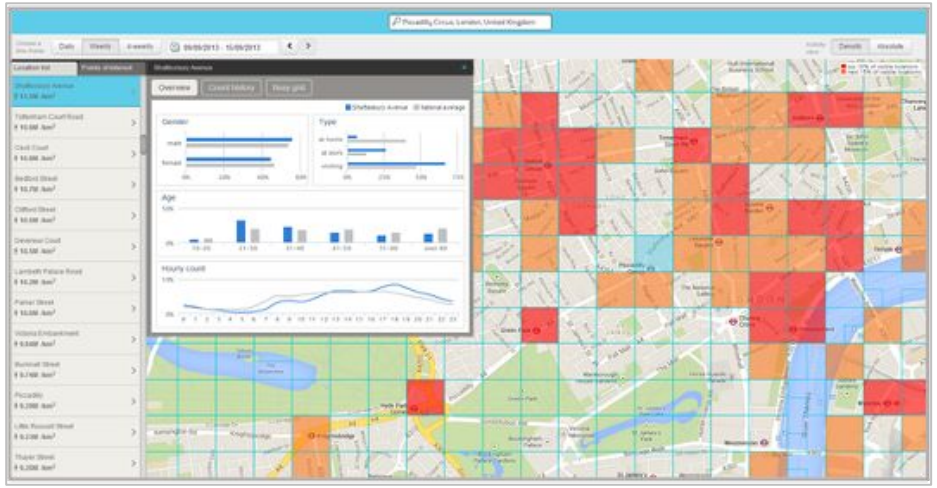
- Analyzing patient cases
- Improving patient health and well-being
- Identifying and predicting disease
- Identifying at-risk patients for developing disease (i.e., diabetes, heart disease, etc.)

From: Srikanth at Techiexpert.com

DATA SOLUTIONS USING PREDICTIVE ANALYTICS (CONT.)

Law enforcement

Law enforcement uses graphical solutions such as heatmaps and predictive analytics to predict and fight crime.



From: Mark Gibbs

DATA SOLUTIONS USING AI

DataRobot is an artificial intelligence (AI) platform that helps automate machine learning:

- To analyze, model, and present data
- To build custom reports using Python
- To democratize data

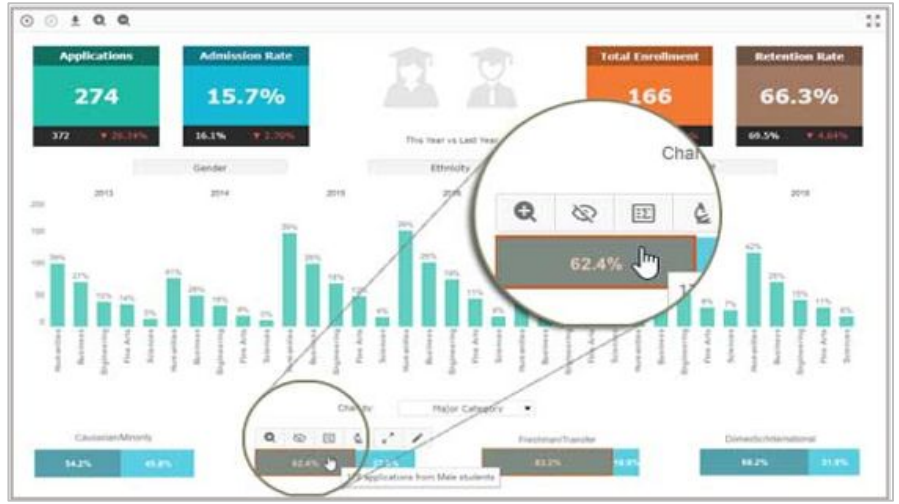
From: DataRobot

DATA SOLUTIONS USING AI (CONT.)

Education

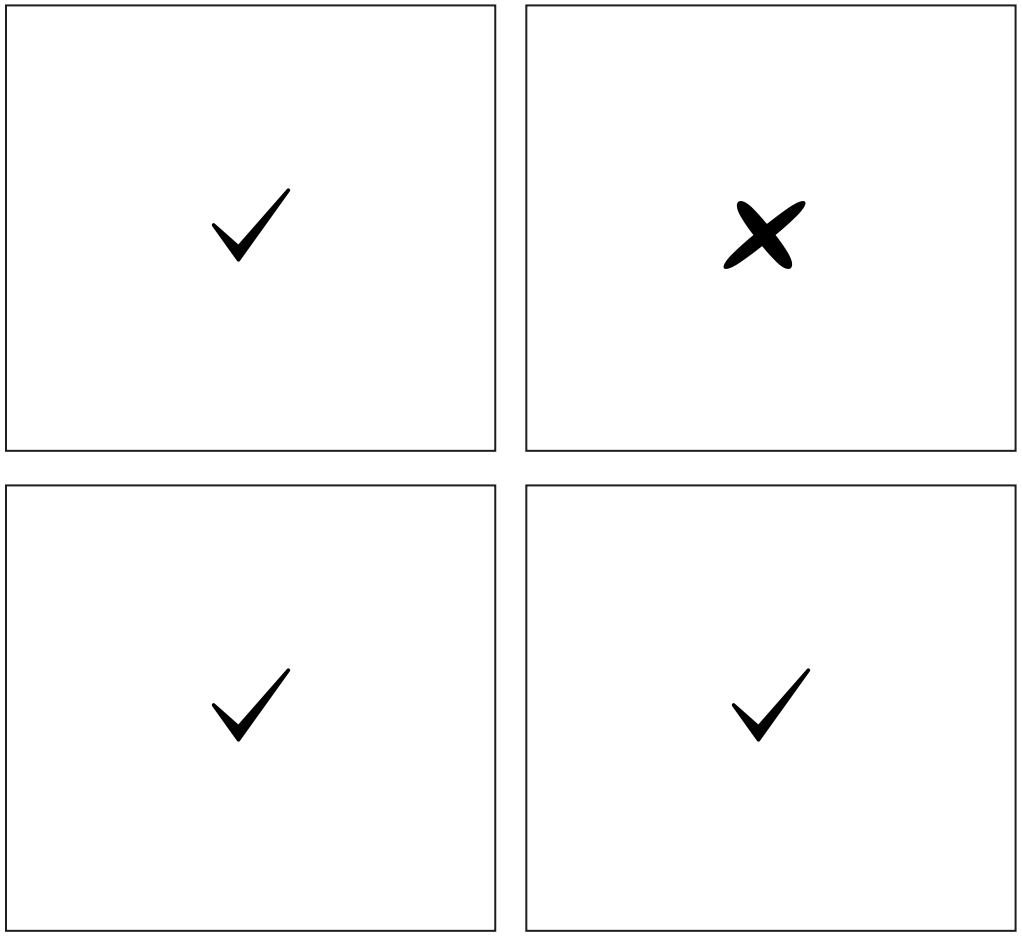
Adaptive learning platforms use machine learning and AI techniques to deliver content tailored to each individual student based on their needs and achievements.

- Analyzing student progress
- Identifying at-risk students
- Assessing student achievement
- Improving decision-making about students



From: inetSoft

POLL



Which of the following are types of solutions that can be found in banking and insurance?

A. Identifying risks

B. Identifying fraudulent activity

C. Processing claims

D. All the above

POLL FEEDBACK

Which of the following are types of solutions that can be found in banking and insurance?

A. Identifying risks

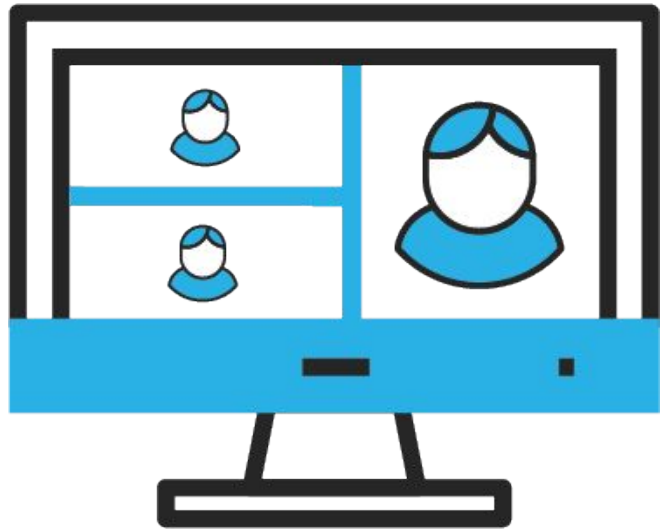
B. Identifying fraudulent activity

C. Processing claims

D. All the above

BREAKOUT ROOMS

- Pick an industry you are interested in and conduct a search for analytics solutions.
- Share your findings with your partner and the class.



BREAKOUT ROOMS

BREAKOUT REGROUP

DSA CAREER OPTIONS

- Data Engineer
- Data Analyst
- Data Scientist
- Business Intelligence (BI) Developer
- Reporting Analyst
- Machine Learning (ML) Engineer
- Data Architect
- Database Developer
- Technical Project Manager (TPM)
- Product Manager
- Scrum Master

ROLE COMPARISON

BI Professional	Data Analyst	Data Scientist
Often focused on data presentation layer	Scrutinizes data using data analysis tools	Develops and enhances data models and predictive models
Works with data visualization software	Cleans data for downstream analyses	Identifies and incorporates the right data into an analysis
May work with reporting software	Prepares reports	Develops and tests hypotheses using statistical techniques
May access data in various locations (database, CSV files, APIs, etc.)	Identifies various trends, summarizes or aggregates data	

DATA SCIENTIST VS. ANALYST

Data Scientist Role

- Prepare data for analysis and processing
- Perform advanced data analysis
- Present the results to reveal patterns and enable stakeholders to draw informed conclusions

Data Analyst Role

- Analyze company or industry data to discover new insights
- Build reports to share with decision-makers
- Work with data from multiple sources

BREAKOUT ROOMS

Pair up with a classmate and discuss your interests in data science.

- What types of data-related roles interest you?
- What characteristics do you share with a data scientist or analyst?
- Conduct a web search of roles that interest you and find a job description.

BREAKOUT ROOMS

BREAKOUT REGROUP

ACTIVITY: CAREER SERVICES PREPARATION

Complete the following:

- Career Services Prep Module
- Career Services Survey
- Job Exploration Worksheet
- Submit your resume and LinkedIn URL.

REVIEW AND WRAP-UP

Today you learned about:

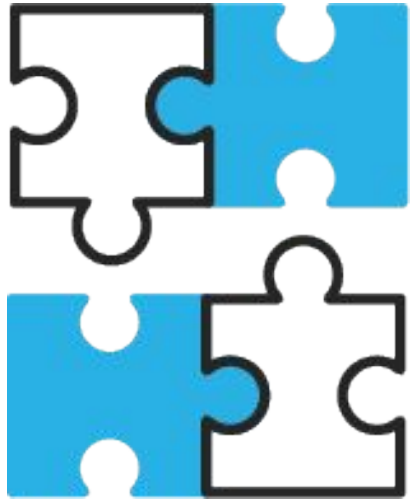
- What data science and analytics is
- Careers in data science and analytics
- Characteristics of data science and analytics careers

BREAK TIME

LESSON 3:
TOOLS AND SOFTWARE

LESSON OUTLINE

- Data software
- Programming languages
- Data visualization software
- Building a data analytics portfolio in GitHub



LESSON GOALS

WHAT ARE THE GOALS?

- Identify the programming languages used in data science.
- Identify the different types of software that data professionals use.
- Define GitHub and explain how it's used by data professionals.
- Build a data analytics portfolio in GitHub.

WHY ARE THEY IMPORTANT?

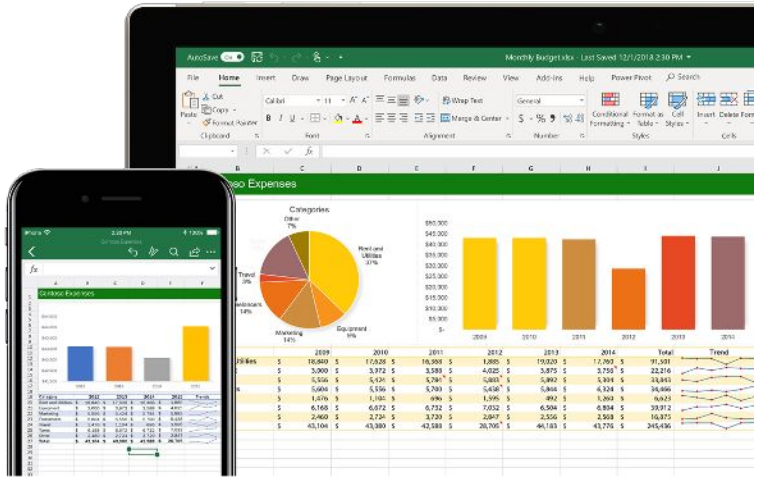
As a data professional, you will use a variety of tools to perform analysis on large data sets, as well as visualize and present the insights that solve these interesting problems that businesses face.

DATA SOFTWARE

- Data visualization
- Data quality
- Data preparation and data wrangling
- Data warehouse and data catalog
- Database

MICROSOFT EXCEL

Spreadsheets like Microsoft Excel and Google Sheets help data professionals analyze, clean, and visualize data.



PROGRAMMING LANGUAGES

- SQL
- Python

STRUCTURED QUERY LANGUAGE (SQL)

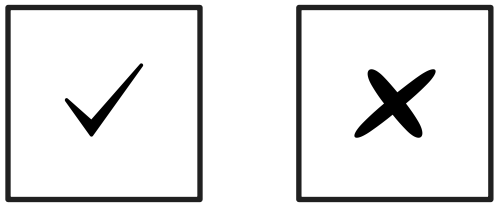
Structured Query Language (SQL) is a standardized language for defining and manipulating data in a relational database.

From: GeeksforGeeks

PYTHON

Python is an open-source programming language that is high level and works as a general-purpose language; it is most often compared to Ruby, JavaScript, and Scheme.

QUIZ



What is the most popular language used to query databases?

A. Python
B. Scala
C. R
D. SQL

QUIZ FEEDBACK

What is the most popular language used to query databases?

A. Python

B. Scala

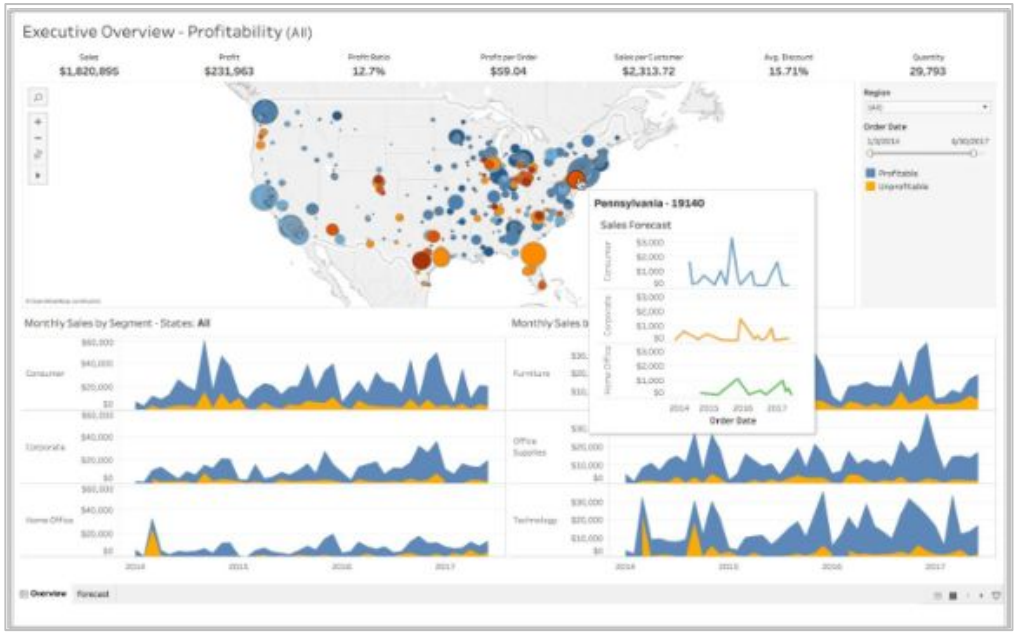
C. R

D. SQL

DATA VISUALIZATION

Tableau

- Takes raw data and transforms it into visual form
- Aggregates and summarizes large amounts of data



From: Tableau

GITHUB

GitHub

A place to store and showcase all your data projects as a portfolio for potential employers

“GitHub is a website and cloud-based service that helps developers store and manage their code, as well as track and control changes to their code.”

HOW DOES IT WORK?

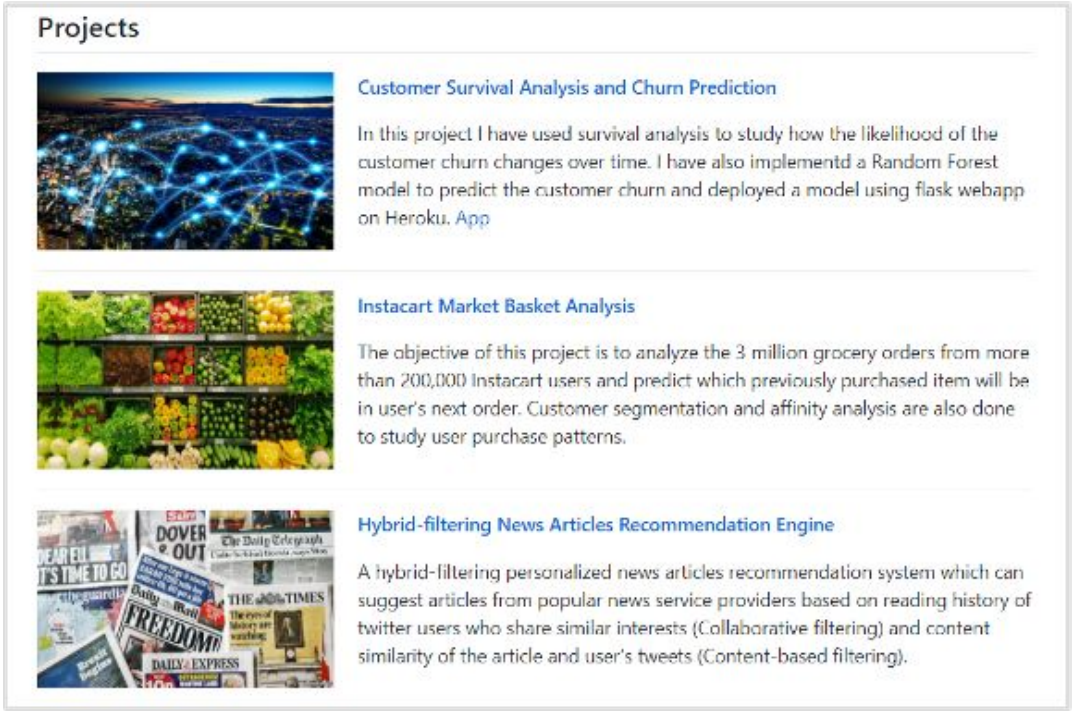
- Send and retrieve files.
- Collaborate with others.
- Store files in the cloud.
- Free vs. paid version

REPOS

A place to store your projects

Examples:

- Data-Analyst-Portfolio by mkumar7
- data_science_portfolio by melvfnz
- Data Science Portfolio by Arch Desai



From: github.com/archd3sai/Portfolio

1.1.3 ACTIVITY: BUILD YOUR GITHUB REPO

Project

Build your data analyst portfolio on GitHub to store and showcase data projects you work on throughout this course for potential employers to see.

Download the Activity Guide
Activity Guide: A complete guide on how to complete this project

Steps

1. Create a GitHub account.
2. Create a new repository.
3. Customize the project summary page.
4. Populate your portfolio with data projects that you work on throughout the Introductory Course.

REVIEW AND WRAP-UP

Today, you learned about:

- What data science and analytics is
- Careers in data science and analytics
- Tools and software that data professionals use
- How to build a portfolio in GitHub

QUESTIONS?

NEXT STEPS

- [x] Assigned Activities
- [x] Data Analyst Portfolio: GitHub Tasks
- [x] Other Reminders