

Data Science and Artificial
Intelligence in Marketing

BUSI 488 / COMP 488

Daniel M. Ringel

UNC Kenan-Flagler Business School
Spring 2023

January 10th, 2023

Class 01: Welcome and Course Introduction

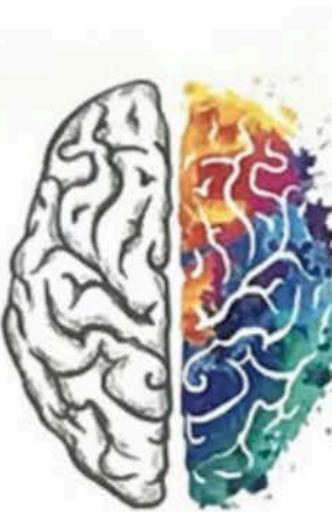
Sections 001 and 002



UNC
KENAN-FLAGLER
BUSINESS SCHOOL



Scientists of Discovery



Artists of Innovation

Daniel M. Ringel (He/Him)

Assistant Professor of Marketing for Data Science and AI

Chapel Hill, North Carolina, United States · [Contact info](#)

500+ connections

<https://www.linkedin.com/in/dmr76/>

Office Location: McColl 4418
Office Phone: 919-962-8746

Email: dmr@unc.edu
Office Hours: By appointment



**UNC Kenan-Flagler
Business School**



Goethe University

Today's Agenda

1 Welcome

2 Artificial Intelligence

3 Walking the Talk

4 Course Overview

5 What is Data Science?

6 Python and CoLab

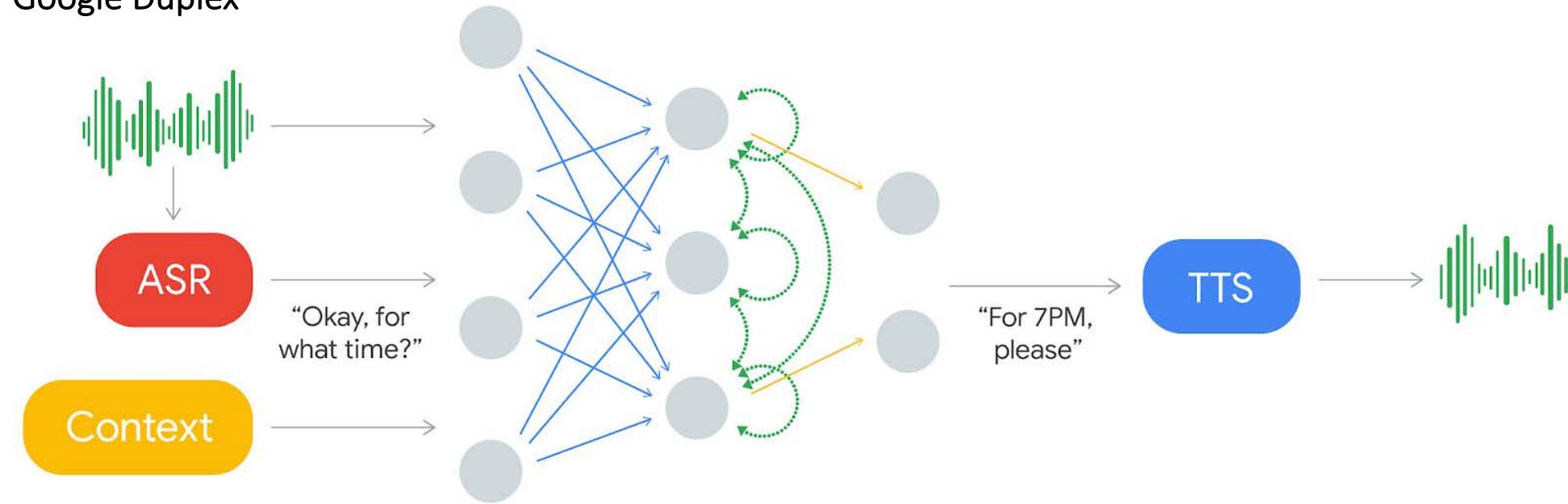
7 The Dark Side of Data Science and AI

Prep-Check:

- ✓ Visited Canvas
- ✓ Read Syllabus
- ✓ Google Account
- ✓ DataCamp Account
- ✓ CoLab Tutorial

AI for Interactive Assistants

Google Duplex



<https://ai.googleblog.com/2018/05/>



Audio: <https://www.gstatic.com/b-g/5717BWTLRKCB8JUDQ1QUA6HMC26WL238301173.mp3>

Audio: <https://www.gstatic.com/b-g/KOK4HAMTAPH5Z96154F6GKUM74A3Z1576269077.mp3>



AI explains my jokes

Explaining a joke

Prompt

Explain this joke:

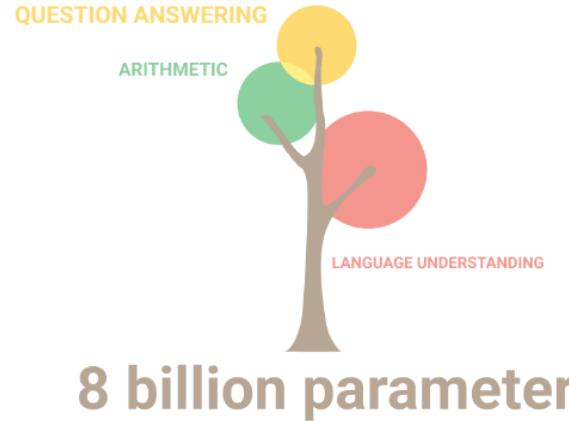
Joke: Did you see that Google just hired an eloquent whale for their TPU team? It showed them how to communicate between two different pods!

Model Response

Prediction: TPUs are a type of computer chip that Google uses for deep learning. A "pod" is a group of TPUs. A "pod" is also a group of whales. The joke is that the whale is able to communicate between two groups of whales, but the speaker is pretending that the whale is able to communicate between two groups of TPUs.



Powerful AI Models: Google Pathways Language Model (PaLM)



Crazy Expensive!

- Team of 70 Scientists
- Requires Terabytes of Text:
 - Websites, Books, Wikipedia, News, Chats, Social Media, etc.
 - In multiple Languages
- PaLM was trained on 6144 TPUs (Tensor Processing Units)
- Estimated Training Cost (per 1 training): 8 Million USD

Empathetic AI?

The screenshot shows the OpenAI Playground interface. On the left, there's a large text area labeled "PROMPT" containing a conversation about a lost suitcase. On the right, there are several configuration options: Mode (set to "Q&A"), Model (set to "text-davinci-002"), Temperature (0), Maximum length (100), Stop sequences (empty), Top P (1), Frequency penalty (0), Presence penalty (0), Best of (1), Inject start text (checkbox checked, value "+A:"), Inject restart text (checkbox checked, value "+R:Q"), and Show probabilities (Off). At the bottom, there are "Submit" and "Cancel" buttons, and a status bar showing "668".

PROMPT:

I am an empathetic question-answering bot. If you ask me a polite question, I will give you a friendly answer. If you ask me an angry question, I will confirm your feeling.

INTERACTIVE Q&A

Q: My whole life was in that bag. Now it is lost forever!

A: I am so sorry about your loss. We will do everything we can to find your bag.

Q: I am so angry right now I can hardly speak!

A: I can understand why you are so angry. Losing a suitcase can be a very stressful experience.

Q: Who the h* do you think you are?**

A: I am a customer service representative for this airline.



<https://beta.openai.com/examples>

- Pre-Trained GPT-3 Large Language Model from OpenAI
- Question-Answer Tuned
- Uses Prompts to understand context and expected behavior

Converse with AI?

ChatGPT



Examples

"Explain quantum computing in simple terms"



Capabilities

"Got any creative ideas for a 10 year old's birthday?"

Allows user to provide follow-up corrections

"How do I make an HTTP request in Javascript?"

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

<https://chat.openai.com/chat>



Creative AI?

transformer.huggingface.co/doc/gpt2-large

Write With Transformer gpt2 ⓘ

unicorn icon

Shuffle initial text Trigger autocomplete or tab Select suggestion ↑ ↓ and enter Cancel suggestion esc Save & Publish

Model & decoder settings ⓘ

Model size **gpt2/large**

Top-p **0.9**

Temperature **1**

Max time **1**

See how a modern neural network auto-completes your text 😊

This site, built by the [Hugging Face](#) team, lets you write a whole document directly from your browser, and you can trigger the Transformer anywhere using the Tab key. It's like having a smart machine that completes your thoughts 😊

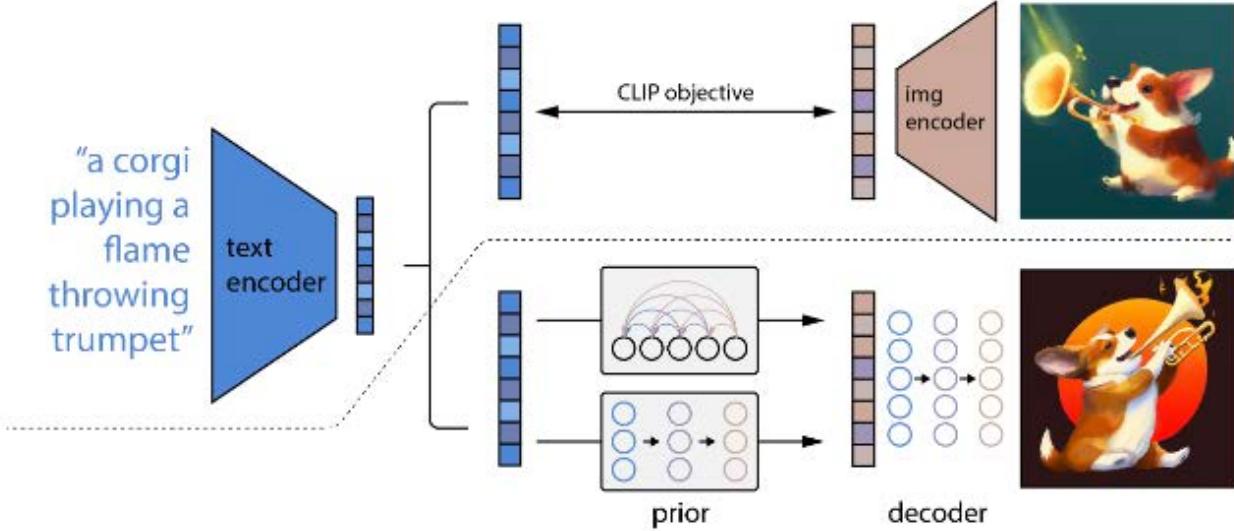
Get started by typing a custom snippet, [check out the repository](#), or [try one of the examples](#). Have fun!

<https://transformer.huggingface.co/doc/gpt2-large>



Artistic AI?

- DALL-E 2: Text to Images
- Generate Images from Text
- Published in April 2022 by OpenAI
- Text Analysis to Image
- Trained on 650 Million Images



<https://github.com/lucidrains/DALLE2-pytorch>
<https://arxiv.org/pdf/2204.06125.pdf>

<https://openai.com/dall-e-2/#demos>

Text Prompt:

„Teddy bears working on new AI research underwater with 1990s technology“



What is Artificial Intelligence (AI)?

Artificial intelligence

- Leverages computers and machines
 - to mimic the problem-solving and decision-making capabilities
 - of the human mind
- [IBM 2020]
-
- The science and engineering of making intelligent machines, especially intelligent computer programs
 - Related to the similar task of using computers to understand human intelligence
 - AI does not have to confine itself to methods that are biologically observable
- [John McCarthy 2004: [Paper](#)]

Can machines think?

Turing Test: Human interrogator tries to distinguish between a computer and human text response
[Alan Turing 1950: [Paper](#)]

Weak vs. Strong AI

Weak AI — also called Narrow AI or Artificial Narrow Intelligence (ANI)

- Trained and focused to perform specific tasks
- Drives most of the AI that surrounds us today
- Enables some very robust applications
- Examples: Apple's Siri, Amazon's Alexa, Autonomous vehicles

Strong AI — Artificial General Intelligence (AGI) + Artificial Super Intelligence (ASI)

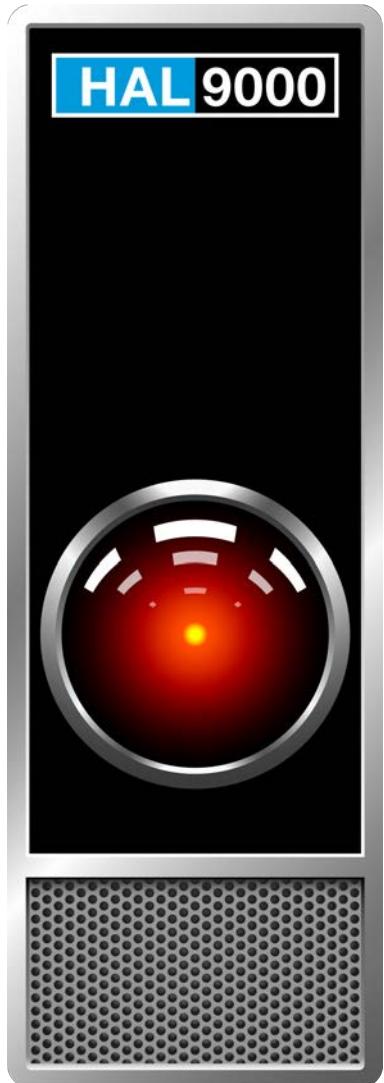
Artificial general intelligence (AGI): Theoretical form of AI

- Machine has an intelligence equaled to humans
- Has self-aware consciousness
- Ability to solve problems, learn, and plan for the future

Artificial Super Intelligence (ASI)

- Surpass the intelligence and ability of the human brain
- Still entirely theoretical with no practical examples in use today
- Best example of ASI might be from science fiction: HAL in *2001: A Space Odyssey*

[Source: IBM 2022]



What brought Life to AI

Gaming



Digitalization

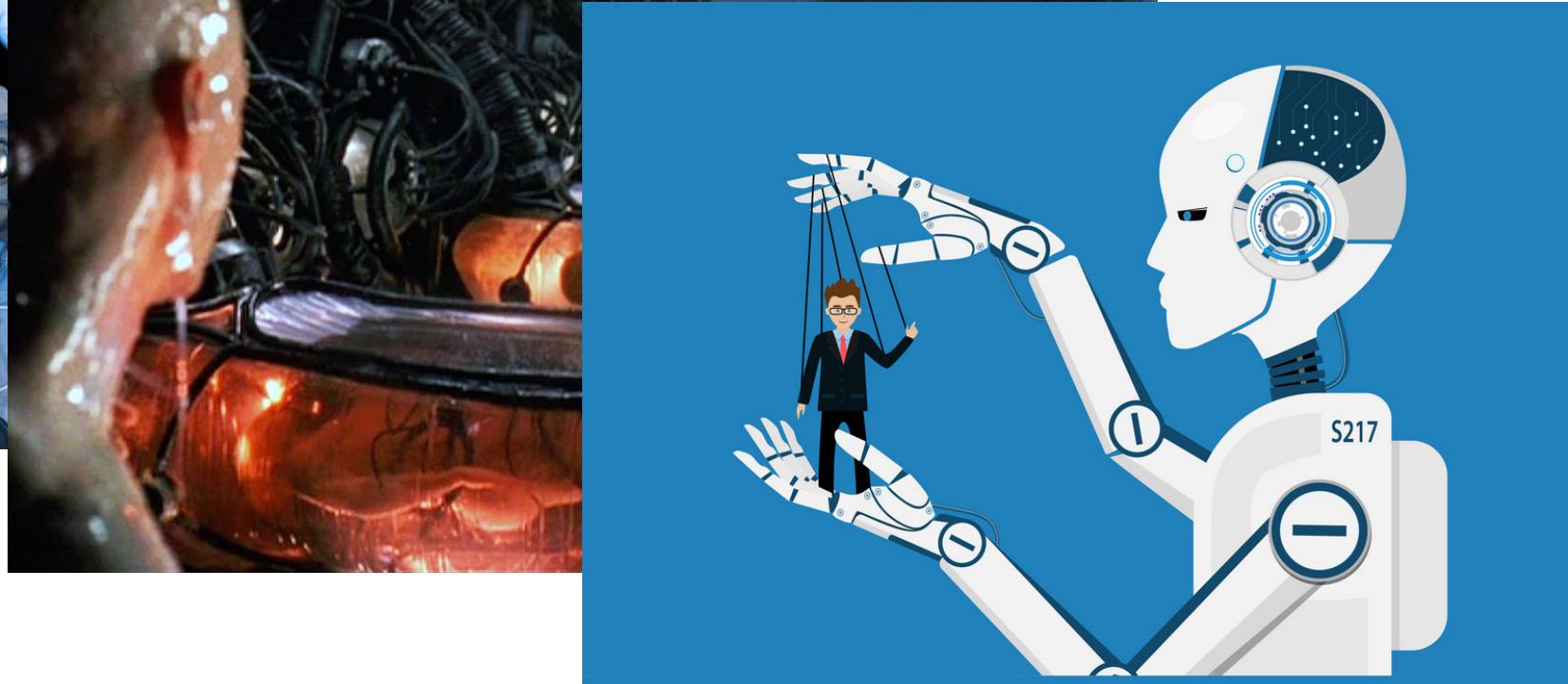
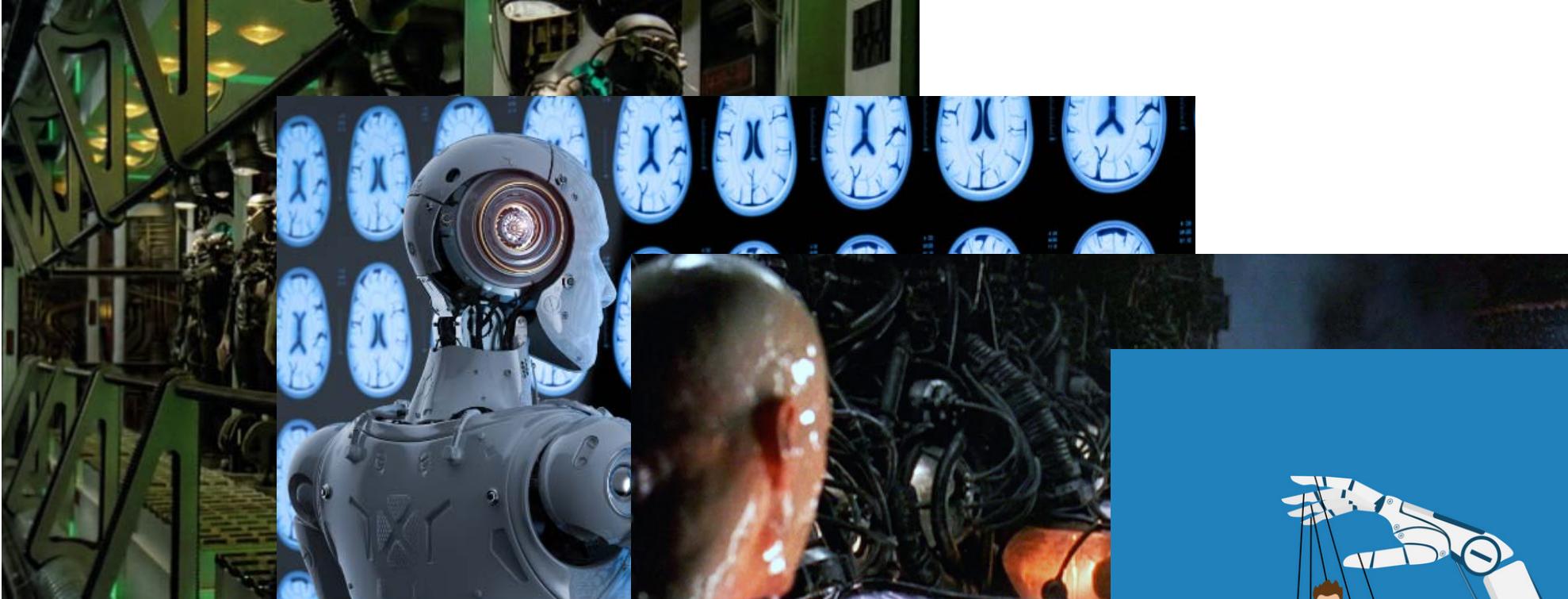


Serendipity

Novel Algorithms



Will AI Control Humans?



Who Controls Humans?



Everyday AI



Google

A screenshot of a Google search results page. The search query "learn about artif" is typed into the search bar. Below the search bar is a list of suggested search terms related to learning artificial intelligence. At the bottom of the search results are two buttons: "Google Search" and "I'm Feeling Lucky". There is also a link to "Report inappropriate predictions".

- learn about artificial intelligence
- learning about artifacts
- learn artificial intelligence from scratch
- learn artificial intelligence online
- learn artificial intelligence for beginners
- learn artificial intelligence with python
- learn artificial intelligence free
- learn artificial intelligence pdf
- learn artificial neural network
- learn artificial intelligence reddit

A screenshot of an Apple device displaying the "Hyper Detection Transformer" feature. The screen shows a woman's face with a circular overlay. A message at the bottom says "Move your head slowly to complete the circle." Below the device is a grid of five images labeled "ORIGINAL" and "SKY", "HAIR", "PERSON 1", and "PERSON 2". Each image has a small label indicating the detected object.



A screenshot of a Google Maps navigation interface. The starting point is "Your location" and the destination is "Top of the Hill Restaurant & Brew...". The map shows a route with three segments: a blue line for 9 min, an orange line for 17 min, and a green line for 31 min. Other icons represent walking, cycling, and public transit. The map includes labels for "FRANKLIN-ROSEMARY HISTORIC DISTRICT", "Chapel Hill", "The Carolina Inn - destination by Hyatt Top rated", "University of North Carolina at Chapel Hill", and "Dean E. Smith Center". At the bottom, there are buttons for "Steps", "Start", and "Pin".



Walk the Talk and Look under the Hood

Torture the data, and it will
confess to anything

Ronald Coase

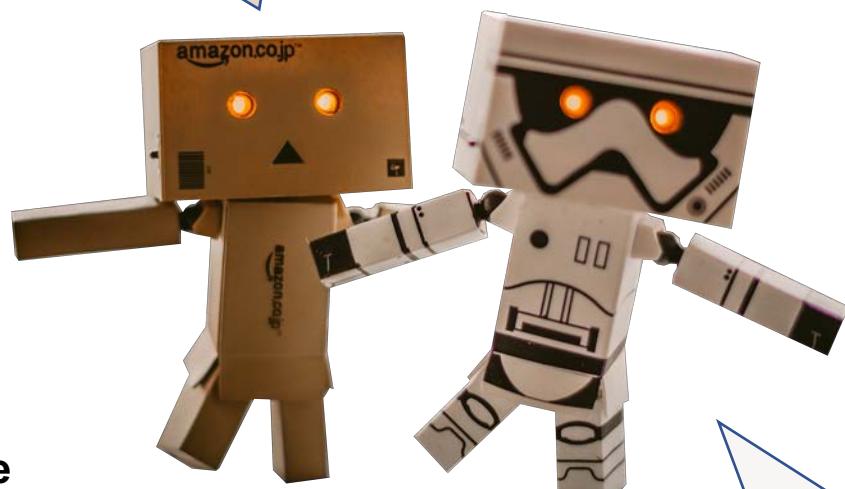
There are only **two industries** that
refer to their customers as “users”:
Illegal Drugs and Software

Edward Tufte

The rise of powerful **Artificial Intelligence**
will either be the best or the worst
thing to ever happen to **humanity**

Stephen Hawking

To **Understand** that



You gotta **Study** this

Data Science is an interdisciplinary field that uses methods, processes, algorithms and systems to extract **knowledge and insights** from structured and unstructured data

Machine learning is the study of algorithms that allow computer programs to **automatically improve** through experience

Artificial Intelligence is the science of making **computers behave** in ways that we thought required human intelligence

Must be a business or a computer science major/minor to take this course

- Interdisciplinary Teams
50% Business, 50% CS
- Real Business Problems

- Data Processing
- Machine Learning
- Deep Learning / NLP

- Underlying Concepts
- Hands-on Application
- Downstream Implications

- Learn Python Coding
- Configure and Train Models
- Tell your Story with Data

Hard-Skills
+ Soft-Skills
+ Critical Thinking
= **YOUR Toolbox!**



Source: WOLFRAM

Your Journey through the Data Science Pipeline

Syllabus

Only the (online) Syllabus on Canvas is valid and will be kept up to date

Course Objectives

Teach students how to apply contemporary data science and AI-enabling methods to real-world marketing problems:

- develop data-driven perspective on marketing problems, that
- informs managerial decision making, and
- present insights in such a way that nonexperts can understand them

Course Format

Pre-Class: DataCamp Homework

In-Class: Explain concepts, discuss applications and limitations, team-presentations

Post-Class: Review class materials → ***Experiential Learning!***

Take-Home Quizzes, Optional Evening Tutorials via ZOOM

Teamwork: Apply methods and algorithms to real data in Team Assignments and a Major Class Project to extract insights and derive managerial implications. Create a mini-workshop on Data and AI along the Consumer Journey with your team and run it in class.

Course Logistics

Semester Length : 3 Credits count once towards CS or BABS degree. Counts for Data Science Minor. Multiple sections: S001 / S002 TTH 2:00pm/3:30pm in person

Learning Success Directors are there to help you! They have ZOOM office hours. 18

Quizzes and Final Exam

3 Quizzes and a Final Exam

- Online, take-home, multi-day window.
- Closed book, closed notes, closed web search.
- No collaboration.

All course material is fair game!

Any material from class, discussions, supplemental readings, and exercises can appear on Quizzes and/or the Final Exam.

No coding questions on Quizzes and Final Exam.

Hierarchy of Information Sources for Quizzes and Final Exam (in case of contradicting statements and/or definitions):

*External Information < Other Courses < DataCamp
< Readings < Classroom < Instructor*

You must have prior approval to take a quiz or the final exam at a time other than the scheduled dates



CANVAS

BUSI488: Data Science in the Business World

2022 / 2023 BSBA Elective Sp...

Recent Announcements



Welcome to 488! PLEASE READ!

Dear Students, Welcome to...

Posted on:

Dec 21, 2022, 10:44 AM



BONUS Assignments for Quizzes: How to Boost your Score

Dear Students, life can som...

Posted on:

Dec 21, 2022, 10:44 AM



Link to Sign-Up to 6 months FREE access to DataCamp

Dear Students, the DataCa...

Posted on:

Dec 21, 2022, 10:43 AM

Search this course

View Course Calendar

View Course Notifications

To Do

[Homework 1: DataCamp](#)

BUSI488: Data Science in the Business World
4700 points |
Jan 17 at 11:59pm

[Team Assignment 1: Data and AI along the Consumer Journey](#)

BUSI488: Data Science in the Business World
8 points |
Jan 24 at 11:59am

[Homework 2: DataCamp](#)

BUSI488: Data Science in the Business World
7400 points |
Jan 24 at 11:59pm

[Quiz 1 - Data Science and Data](#)

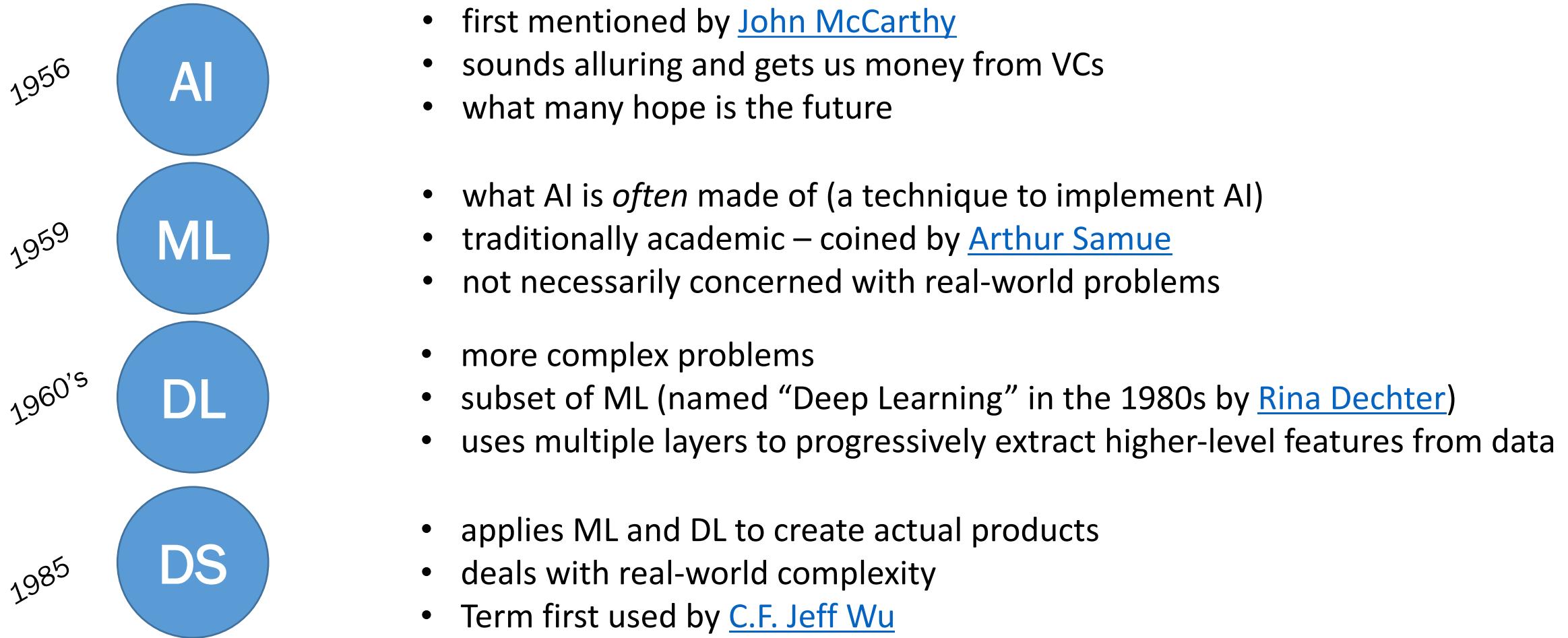
BUSI488: Data Science in the Business World

The course will be hosted on CANVAS

It is crucial that you regularly check your CANVAS for this course because:

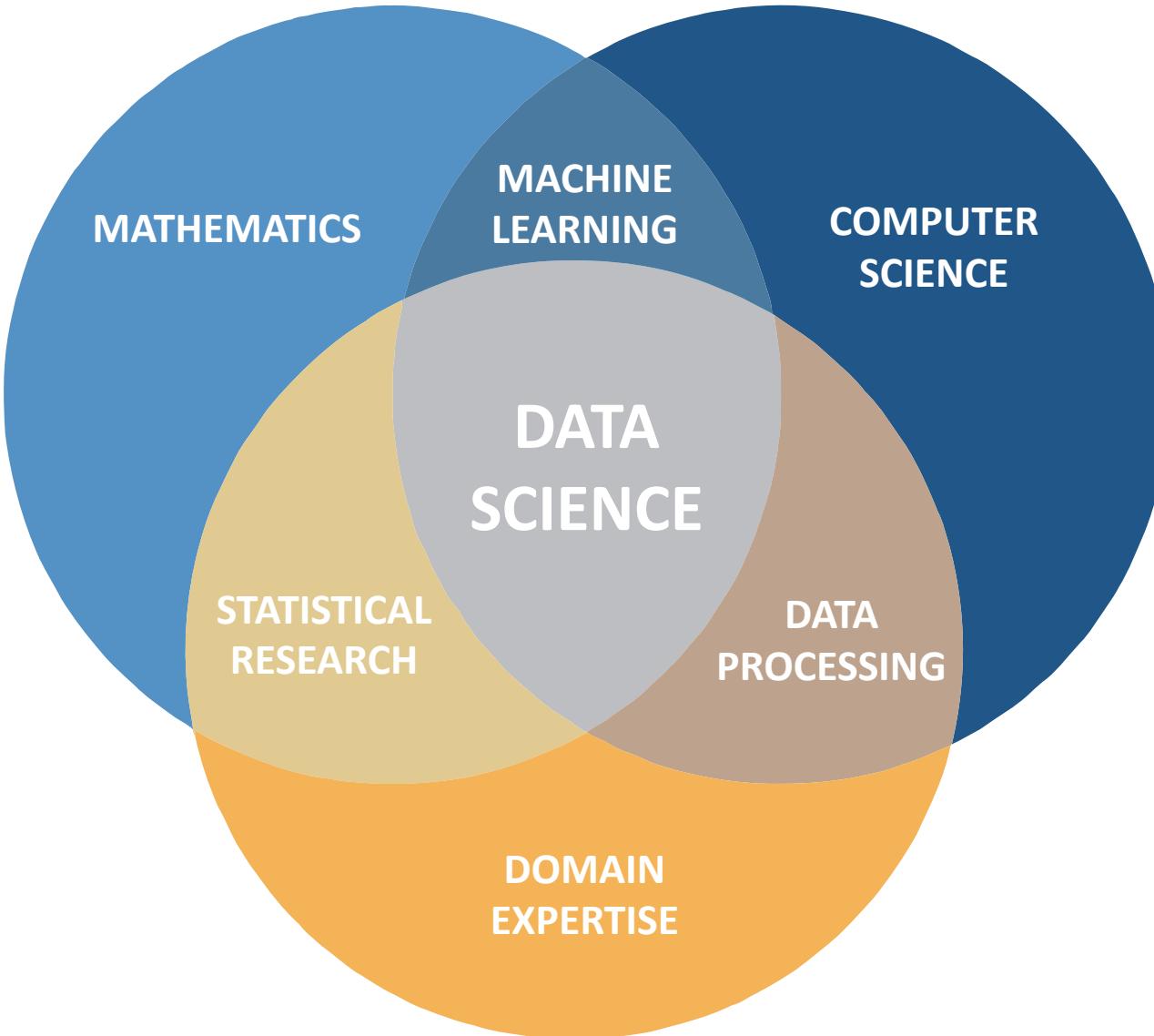
- It is our primary communication medium
- Course resources are available here
- You submit your assignments here
- You take your quizzes and final exam here
- Any changes or updates are communicated here (and may be repeated in class – but perhaps not as timely as on CANVAS)

Data Science, Machine Learning, Deep Learning, Artificial Intelligence



Adapted from *A simplified explanation of AI, Machine Learning, and Data Science* by Suraj Jena, June 10, 2018

What is Data Science?



Source: www.edureka.co

Data Science: More than just Coding

Key skills for data scientists ***are not just the abilities*** to build and use deep-learning infrastructures

Key skills in Data Science are:

1. to learn on the fly
2. to communicate well in order to answer business questions
3. and to explain complex results to nontechnical stakeholders

Aspiring data scientists should focus ***less*** on ***techniques*** and ***more*** on ***questions***

New techniques come and go, but ***critical thinking*** and quantitative, ***domain-specific*** skills will remain in demand

What Data Scientists Really Do, According to 35 Data Scientists

Hugo Bowne-Anderson, HBR podcast, August 15, 2018



The image shows a dark gray rectangular background with a faint, multi-colored watermark of XML code visible through it. The XML code is a snippet of SVG (Scalable Vector Graphics) definition, including declarations for namespaces, extensions, and various styling elements like width, height, and XML space.

```
<!--NS  
EXTENSIONS/  
.Y="0PX" WIDTH="/"  
.ORG/2/SVG" XMLNS="H  
'PXEWEREXTENSI X="0P  
'WWW.W6.ORG/2000/SV  
'P://WWW.0.ORG/19  
'HTTP://WWW." VIEWBO  
'P://NS.AD?0E.C9M/AD  
'NSIONS/3.0/"VIEWBO 0  
... 0PX" Y="0PX" WIDT "484.  
'//WWW.W HEIGHT="398.6PX" VIEW  
'XMLNS:A 398.6" STYLE="ENABLE  
'0E.COM/ADO 0 499.9 428.6;"  
'XTENSIONS/3.0 XML:SPACE="P  
'P://NS.A900E.C <STYLE TYPE="T  
'VGVIEWERE  
'0PX" Y="0PX"  
'="484.9PX"  
'HT="398.6PX"  
'IEWBOX="0 0 484  
'YLE="ENABLE-P  
'0 0 484.9  
' :SPAC  
'YF
```

Why Data Science + Business?

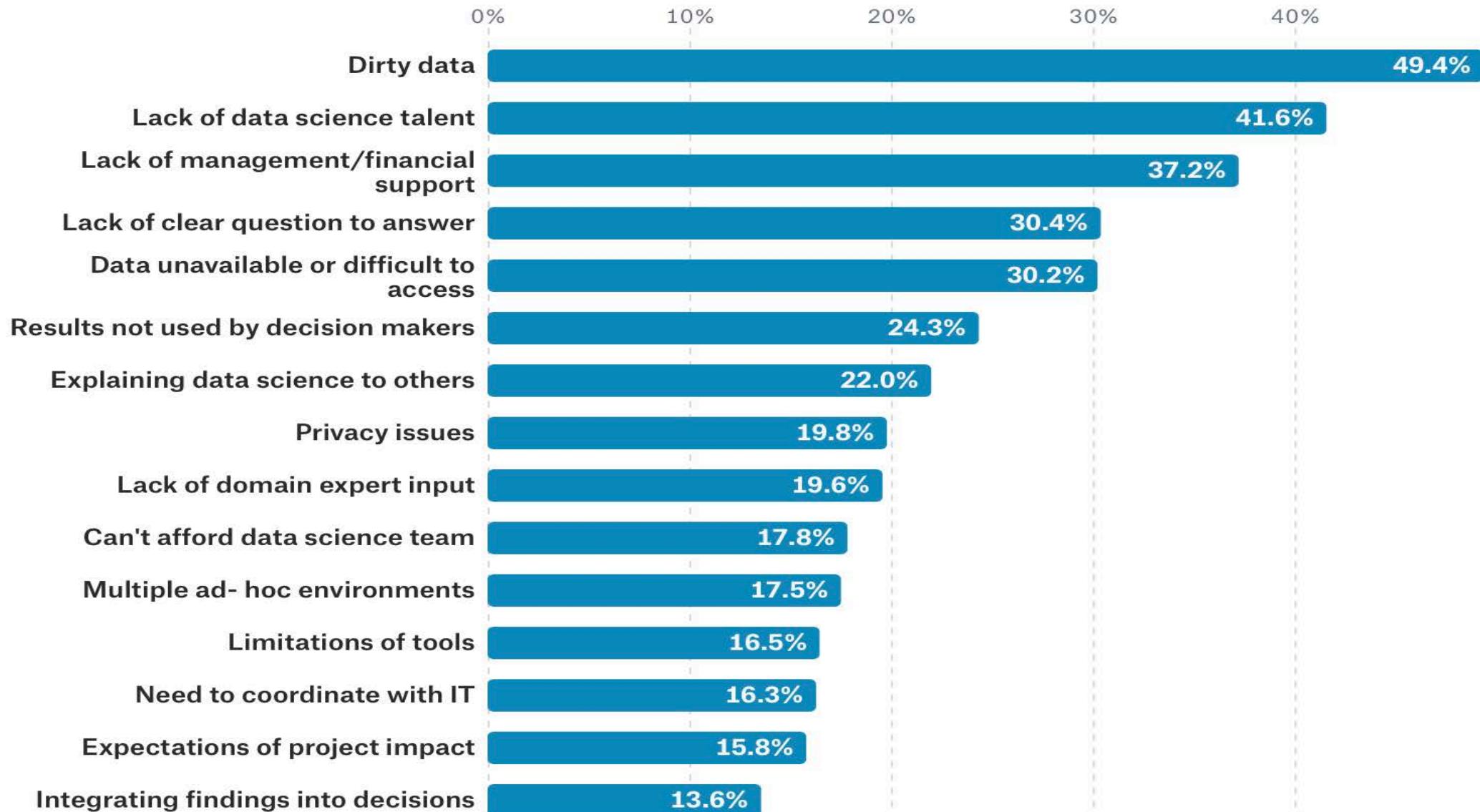
- Data science is growing up fast
- Find signals in the unfathomable volume of noise
- Changes our relationship to fields as varied as language translation, retail, health care, and basketball
- Despite the success stories, many companies aren't getting the value they could from data science
- Even well-run operations that generate strong analysis fail to capitalize on their insights
- Efforts fall short in the last mile, when it comes time to explain the stuff to decision makers
- Until companies can successfully traverse that last mile, data science teams will underdeliver

Data Science and the Art of Persuasion

Scott Berinato, Senior Editor HBR

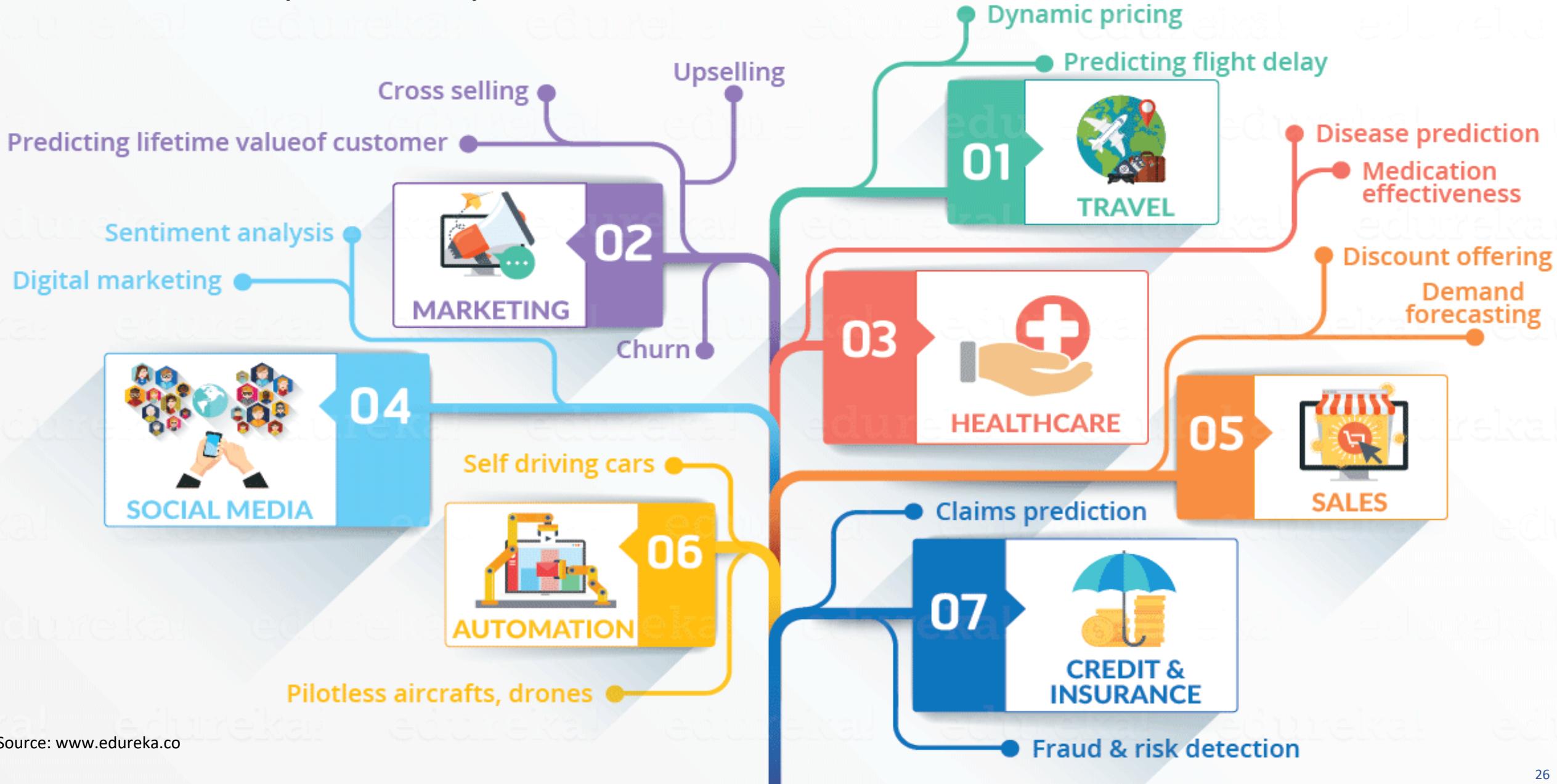


Barriers for Data Science in Firms



Source: Kaggle 2017 Data Science Survey

Domains Impacted by Data Science



Data Science vs. Business Analytics

Analyst

Business Administration

Data Exploration, Analysis and Insights

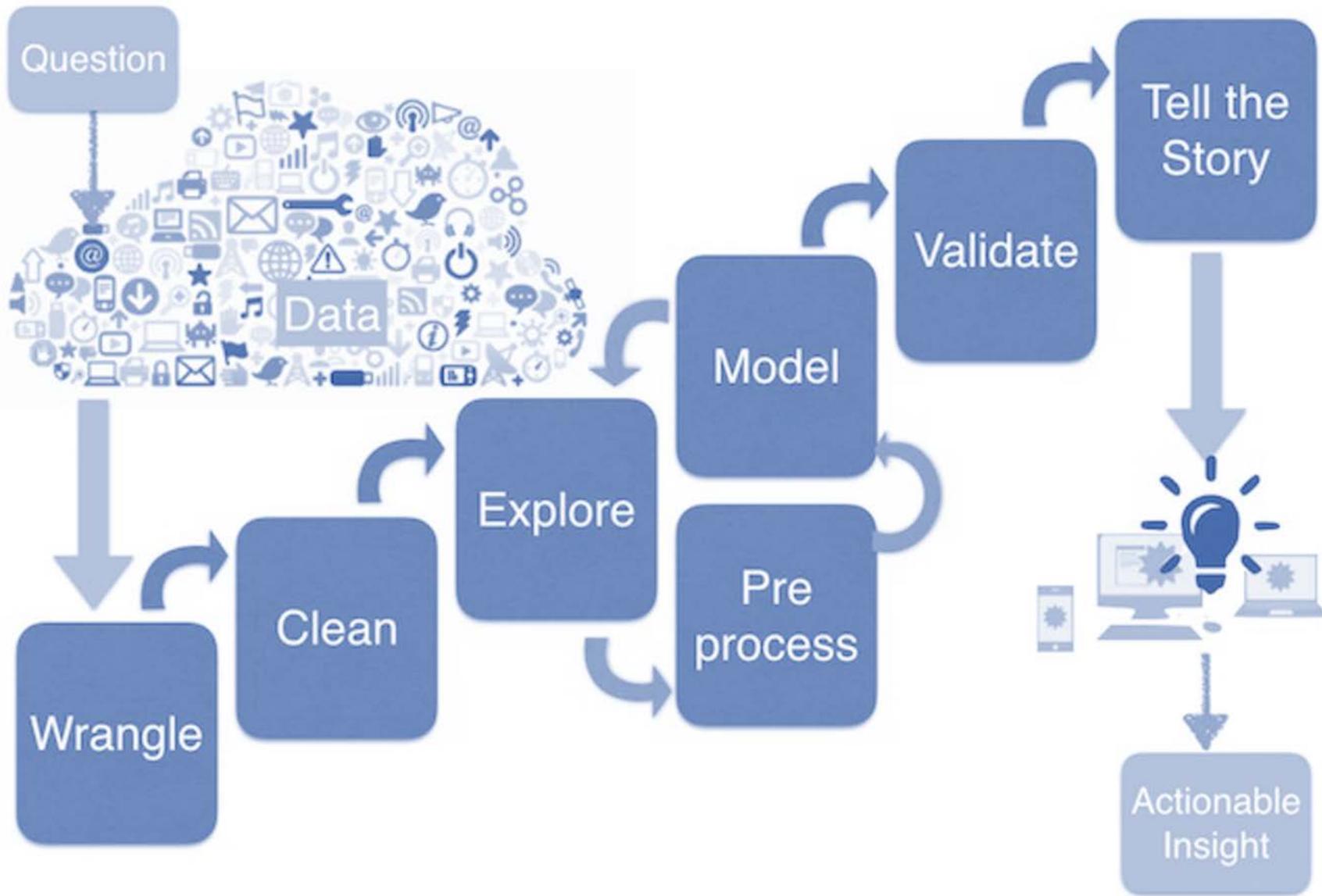
Advanced Algorithms, Machine Learning (AI)

Data Product Engineering

Data Scientist

This Course

The Data Science Pipeline



Source: WOLFRAM

PYTHON



Python Coding in this Course

- *Easy to learn and code*
- *Extensive libraries and frameworks that support a plethora of applications*
- *Incorporated the variants of Java and C like CPython, Jython, etc.*
- *GUI support*
- *Companies working on Python: Intel, Facebook, Spotify, Netflix, etc.*

Coding Exercises for You

Assignments / Everyone ▾

ACTIVE PAST DUE ARCHIVED

Active Assignments

Q Search assignments...

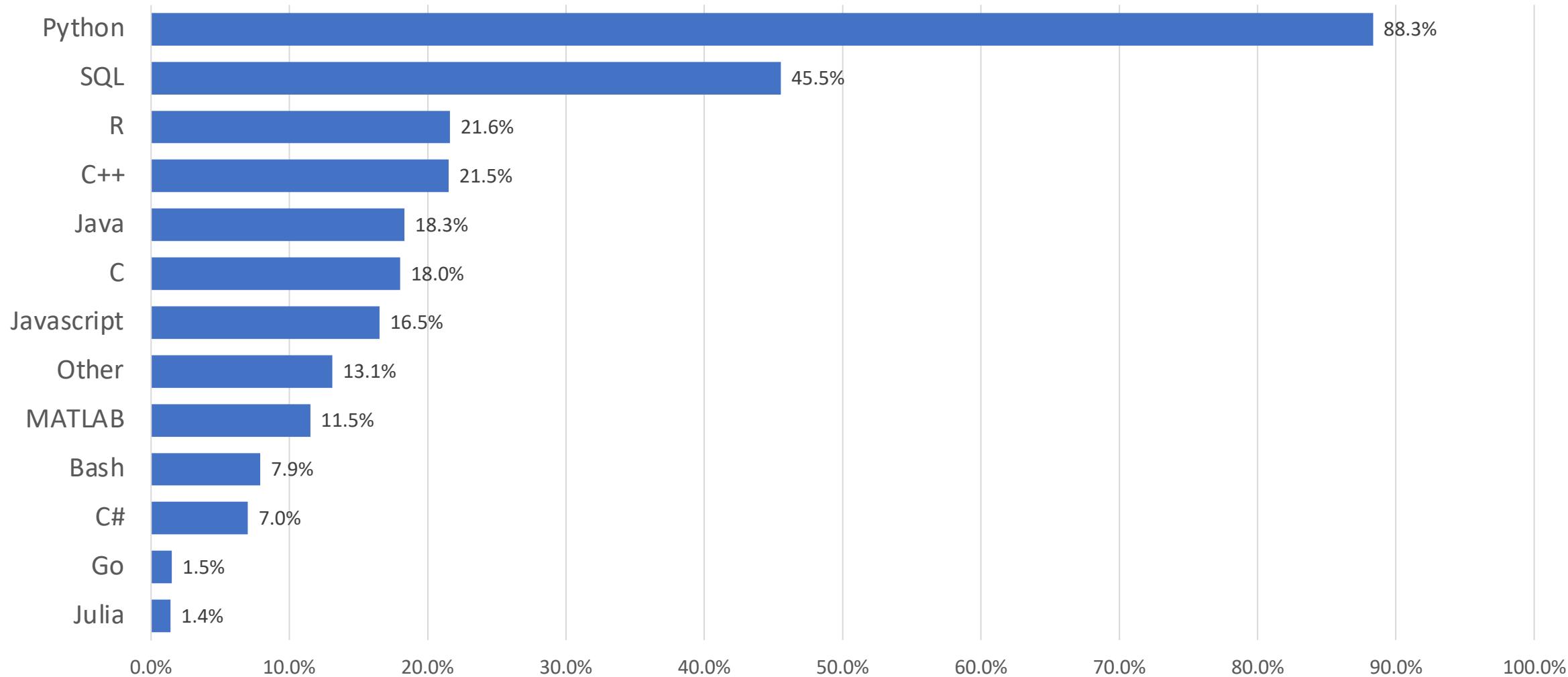
TITLE	ASSIGNEES	STATUS	DUE BY	C	A	CR	DETAILS
Introduction to Python Course	Organization	Active	Jan 18, 22:58 CST	13	31	41%	View
Python Programming Assessment	Organization	Active	Jan 18, 22:59 CST	4	31	12%	View
Intermediate Python Course	Organization	Active	Jan 24, 22:58 CST	6	31	19%	View
Data Visualization with Python Assessment	Organization	Active	Jan 24, 22:59 CST	1	31	3%	View
Data Manipulation with pandas Transforming DataFrames Chapter	Organization	Active	Jan 31, 22:55 CST	2	31	6%	View
Data Manipulation with pandas Aggregating DataFrames Chapter	Organization	Active	Jan 31, 22:56 CST	2	31	6%	View



- You get **free access** for the duration of this course to all of DataCamp's courses and chapters
- Assigned courses are compulsory Homework
- Each chapter of a course should take approximately 45min to 1 hour—even for students without any coding experience
- You can take these courses/chapters anytime and anywhere. And you can pause and continue any chapter.
- You can take as many courses and chapters as you like: There are over 360 courses taught by over 270 expert instructors!

Why Python?

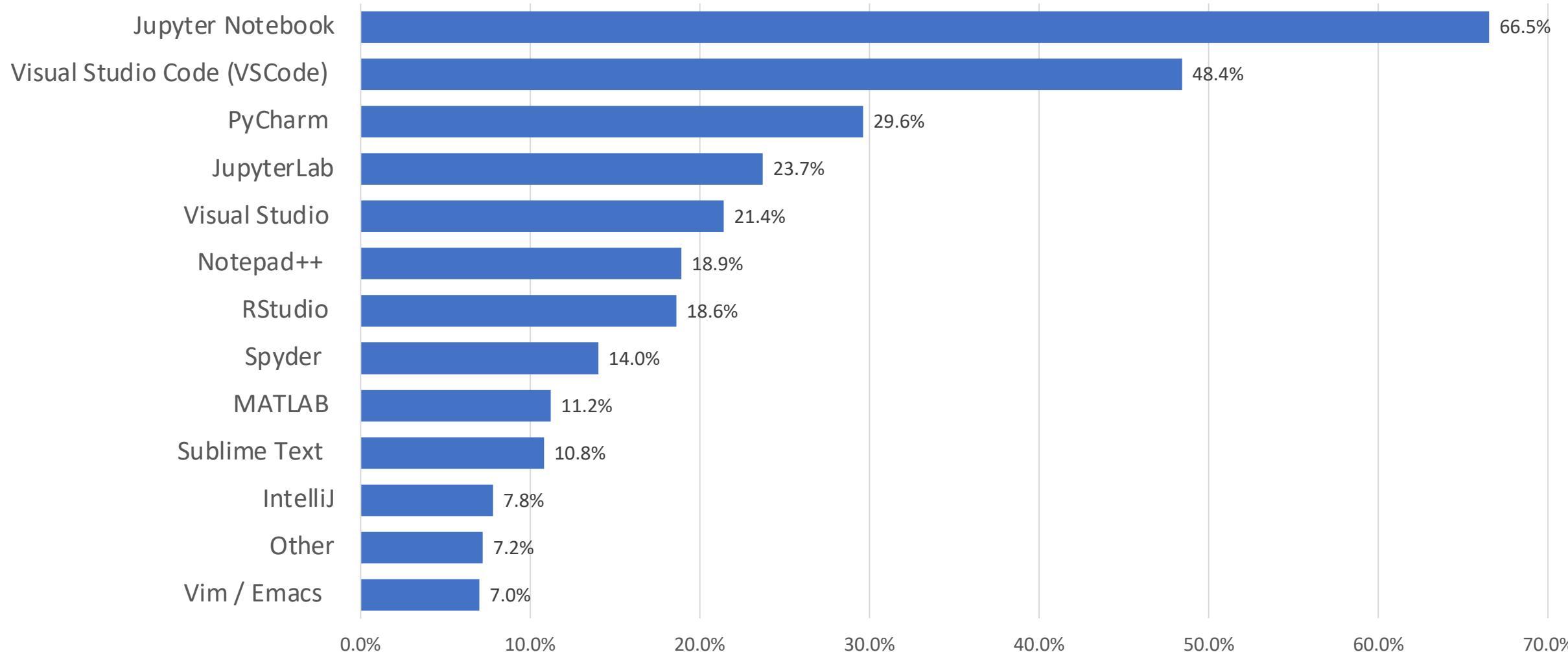
What programming languages do you use on a regular basis?



Source: 2022 Kaggle Machine Learning & Data Science Survey
<https://www.kaggle.com/competitions/kaggle-survey-2022>

Why Google CoLab?

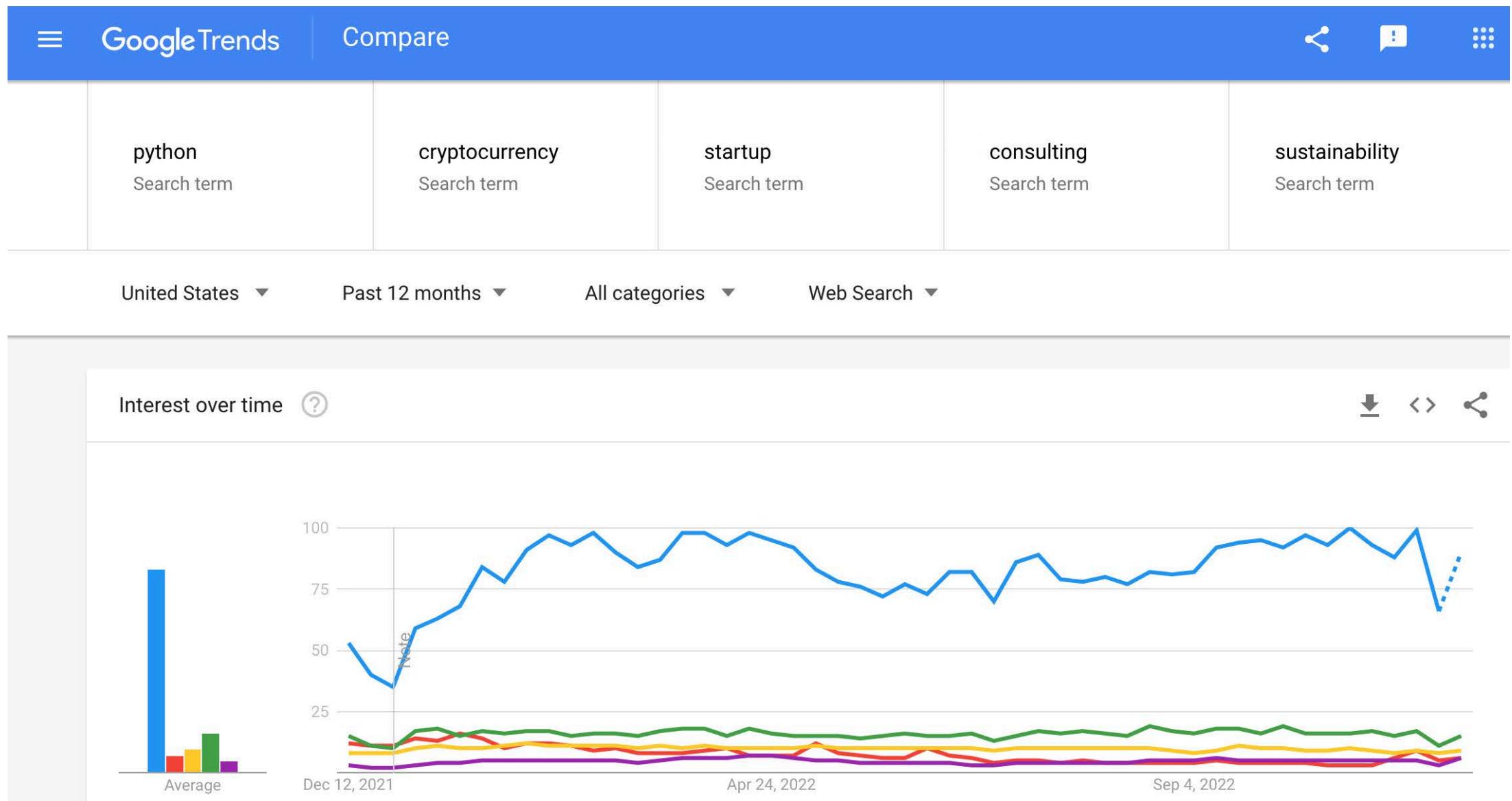
Which of the following integrated development environments (IDE's) do you use on a regular basis?



Source: 2022 Kaggle Machine Learning & Data Science Survey

<https://www.kaggle.com/competitions/kaggle-survey-2022>

Python Searches on Google in 2022



Python Notebooks

Python Notebooks are incredibly powerful tools to interactively develop and presenting data science projects

A notebook integrates code and the code's output into a single document that combines:

- Code itself
- Visualizations
- Narrative text
- Mathematical equations
- Other rich media (e.g., images)

Jupyter and **Google CoLab** are essentially advanced word processors that can also load data, process it, run complex models and save data.



Markdown is a lightweight, easy to learn markup language for formatting plain text.

A **Cell** is a container for:

- Text (or other media) to be displayed in the notebook
- Code to be executed by the notebook's kernel



A **Kernel** is a “computational engine” that executes the code contained in a cell of a notebook

Google Colaboratory (CoLab)

Colaboratory, or “**Colab**” for short, is a product from **Google Research**. **CoLab** allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

CO Create, open, edit, run, save python (jupyter) notebooks

CO Upload Files and Datasets to Google Drive

CO Export Data/Files to Google Drive

CO Runtimes – use GPUs or even TPUs

CO Clone Repositories from Github

Copy the clone link of the Github repository to CoLab

CO Terminal Commands

`!pip install library_name`

CO Collaborate: Share your notebook

This course requires YOU to use CoLab

- Follow instructions on Canvas: [Google CoLab](#)
- Complete CoLab [Tutorial](#) before Class 02 (CANVAS>Files>CoLab_Tutorial)

The screenshot shows a Google Colaboratory notebook titled "BUSI488_2022_CoLabTutorial.ipynb". The code cell contains the following Python script:

```
# ALTERNATIVE to loading the datafile into Session: Mount your Google Drive and open the file from it
from google.colab import drive
drive.mount('/content/gdrive')

# Let's navigate to the folder that contains our data file 'DrDsAmazinGroceryStore2.csv'
# - %cd is a command that means "change directory". In other words: go to the following folder
# - following %cd, you need to provide the "path" (i.e., hierachie of folders) to the folder where your data file
# - keep the first part /content/gdrive/MyDrive/ the same as below
# - change the part /Teaching/2022/BUSI488_2022/CoLab_Tutorial to your own folder structure (i.e., the folders yo
#   - for example, if you created a folder in your google drive called BUSI488,
#   - and within that folder (BUSI488) you created another folder called Class01,
#   - then you need to write the following: %cd /content/gdrive/MyDrive/BUSI488/Class01
# - make sure that the files you want to load are actually in that folder, that is, you uploaded them there (on y
# - to upload the files from Canvas to Google Drive, you first need to download them to your computer and then up
#   (unless you are automatically syncing your computer to the Google Cloud)

%cd /content/gdrive/MyDrive/Teaching/2022/BUSI488_2022/CoLab_Tutorial

!ls # special shell command to view the files in the home directory of the notebook environment
```

Below the code cell, a section titled "4. Read data using Pandas" is expanded. The text states: "We will use a dataset on a fictitious grocery store, "Dr.D's Amazin' Grocery Store", to demonstrate some basic python code and how to run it in CoLab. We will revisit these data later in the course to explore machine learning concepts."

The code cell under "4. Read data using Pandas" contains:

```
# Read the TSV file into a Pandas
data = pd.read_csv('DrDsAmazinGroceryStore2.csv')
```

The Dark Side of Data Science and AI



Where are the Data coming from?

- Census surveys
- IRS Records
- Medical records
- Insurance records
- Search logs
- Browse logs
- Shopping histories
- Photos
- Videos
- Smart phone Sensors
- Mobility trajectories
- Many more ...

!!! VERY SENSITIVE INFORMATION !!!

**YOUR
DATA
FOR SALE**

Everything about you
is being tracked—
get over it
BY JOEL STEIN

!!! VERY SENSITIVE INFORMATION !!!

Looking Ahead



Next Class: Thursday, January 12th, 2023

*Data Formats, Data Types, Data Quality,
and Outlier Detection*

DataCamp Homework 1:

- Introduction to Python
(approximately 4 hours)
- Due on January 17th by 11:59pm