Stand out in a crowd!

Understand what Data Science is and how it can boost your profile

Rutgers Libraries - NB Data Science Workshop Series

Pratiksha Sharma

Sept 29, 2022

Date Science

What's all the fuss about?

- You've all used one or many Data Science techniques at some point in your academic and/or professional journeys:
- Remember calculating a mean of a dataset in elementary school?
- Or perhaps collecting some information in a class and entering it on MS Excel?
- Granted, these are very trivial activities, but they are all little parts & techniques associated with Data Science!
- Over the past couple of decades, with the amount of data that is being generated, the ways we handle it have advanced too.

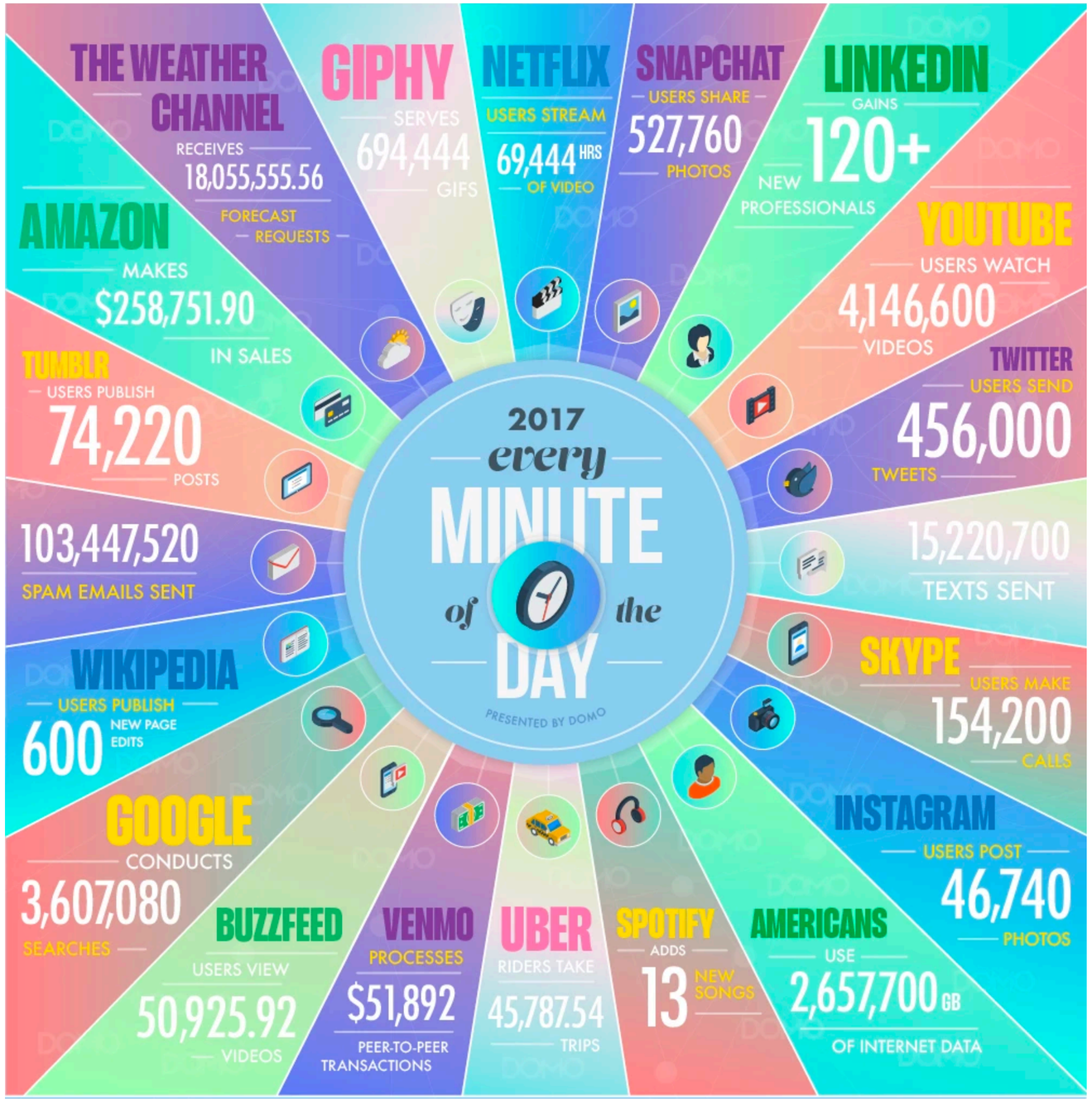
Source: UAtlantica.pt

Date Science

What's all the fuss about?

We are at a point where

DATA SCIENCE = NEED

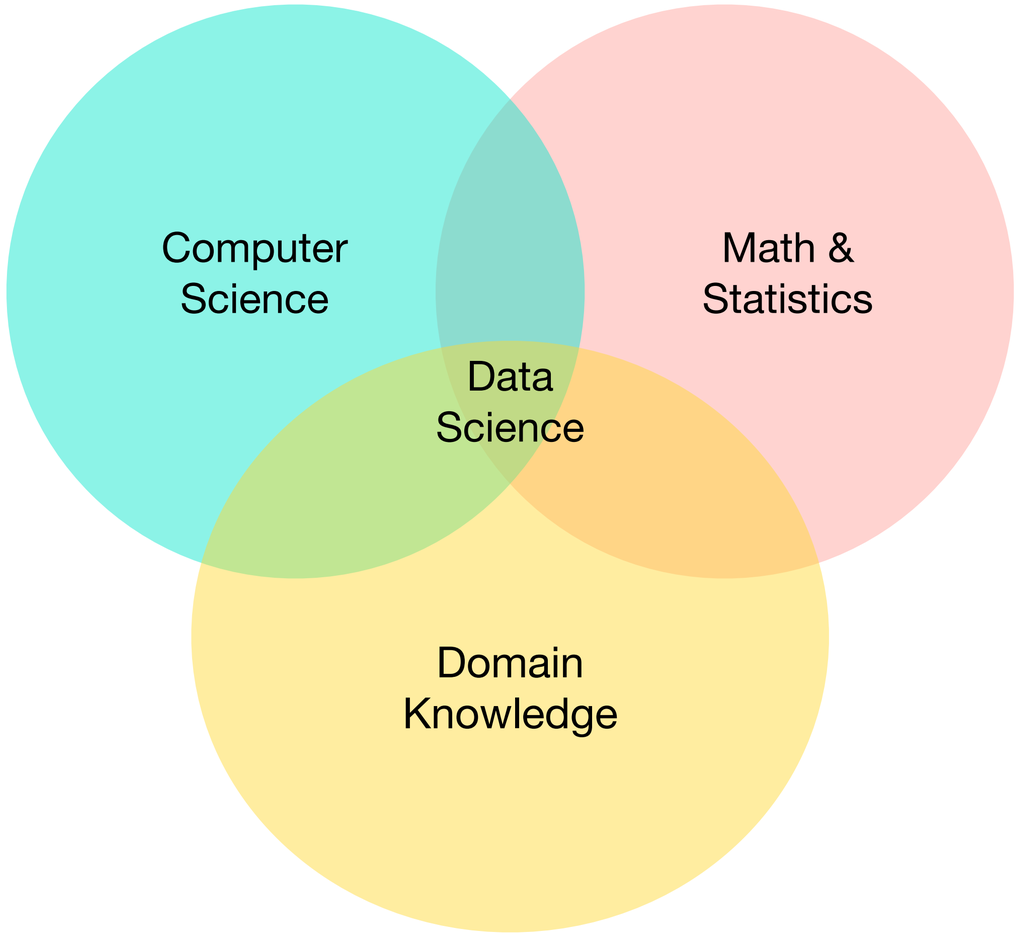


https://www.domo.com/learn/data-never-sleeps-5

Data Science

What's all the fuss about?

- A combination of Maths & Statistics, Computer Science and Domain Knowledge.
- How does it help me?



Data Science

Two Broad Approaches

- Start with a defined problem statement
- Collect relevant data to solve the defined problem
- Clean, pre-process, transform data
- Perform analysis/Model data
- Evaluate analysis/Test Models
- Extract meaningful insights/ Make predictions that help solve the problem

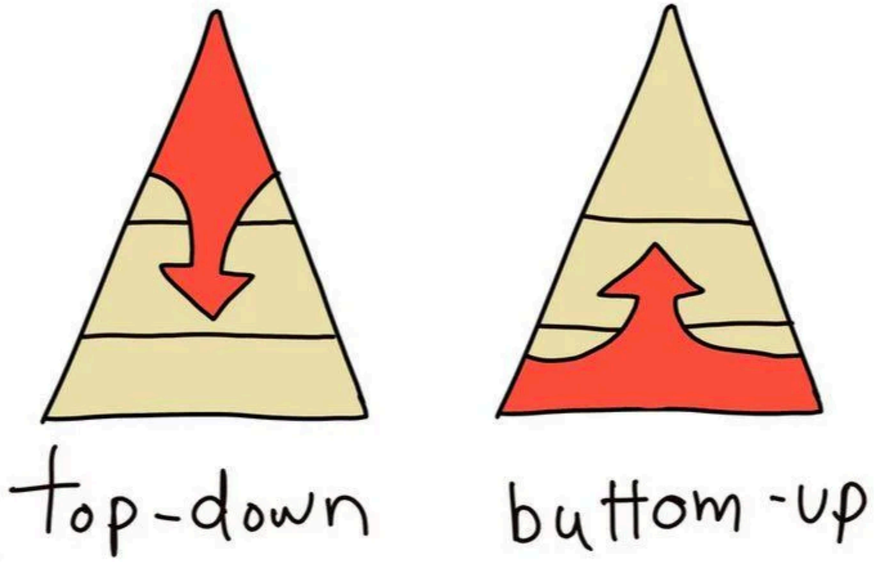


Image Source: FreshBooks.com

- Start with a bunch of data generated/collected
- Clean, pre-process, transform data
- Perform Exploratory Data Analysis (EDA), Visual Analytics etc
- Perform further analysis/Model data
- Evaluate analysis/Test Models
- Extract meaningful insights/ Make predictions that help improve processes

Data Science

Two Broad Approaches

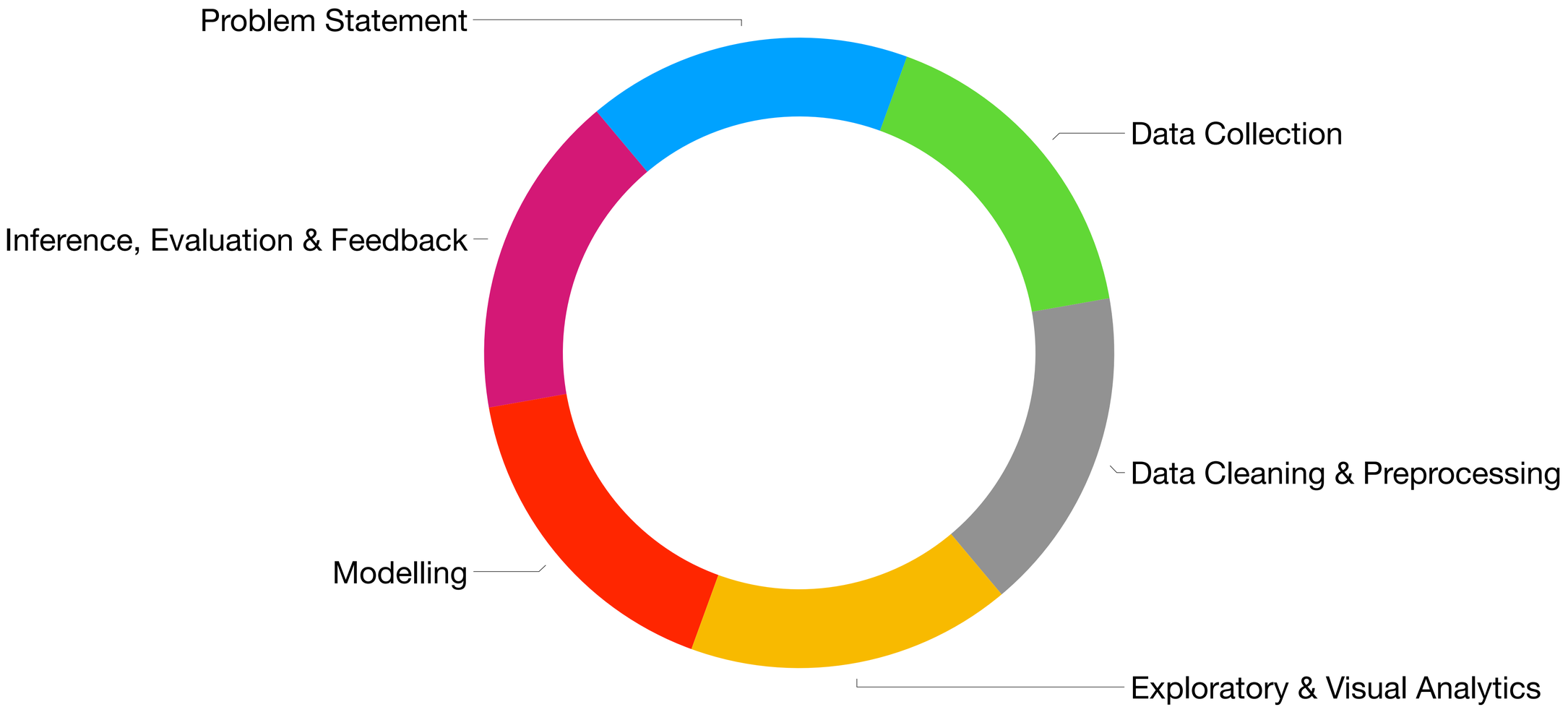
Top-Down

- Majority of applications today follow this approach
- Usually more expensive than a bottom-up approach
- Performed more on a need-basis & does not require proactiveness in anticipating the needs of tomorrow
- Loosely interpreted as a ‘cure-based’ approach

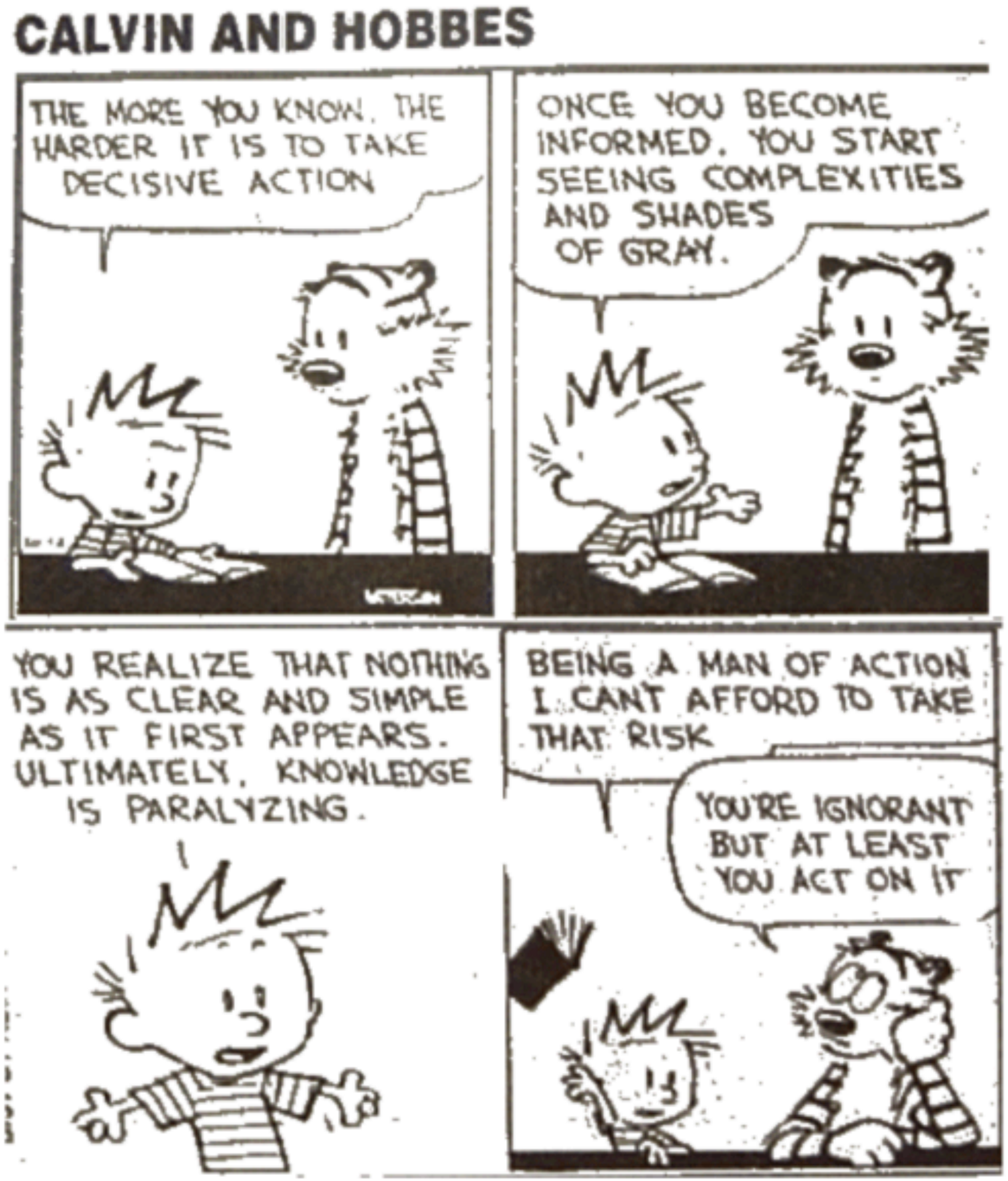
Bottom-Up

- Only a very small subset of applications (or individuals/ organizations) follow this approach
- Usually cheaper
- Requires constant investment & proactiveness to visualize the needs of the future
- Loosely interpreted as a ‘prevention-based’ approach

Data Science Cycle



Why is this a cycle?



Source: Bill Waterson | Universal Press Syndicate

Common Problems that are solved using Data Science

- Customer Acquisition, Marketing & Sales Analytics
- Spam Classification in your inboxes
- Recommendation Engines on your favorite websites: Netflix, Amazon, Walmart, Social Media
- Internet Searches & Chatbots
- Health Risk Assessment: Image processing, Simulation studies, Development of drugs
- A simple internet search!
- Organized data storage & retrieval: SQL & NoSQL Databases
- Much more..

Why is Data Science Useful?

- Needless to say, Data Science finds its applications in every domain, industry or discipline there exists
- It aims to solve issues bigger issues related to the 3V’s associated with the data being generated today - Volume, Velocity & Variety
- For automation of existing processes that involve risk of human error

Business

James

- James is a Business major & is looking forward to starting his very own clothing/apparel business.
- He & his team are looking for an ideal location(s) to open his very first set of stores.
- They take into consideration budgets, proximity of the store to vendors, store staff & locality.

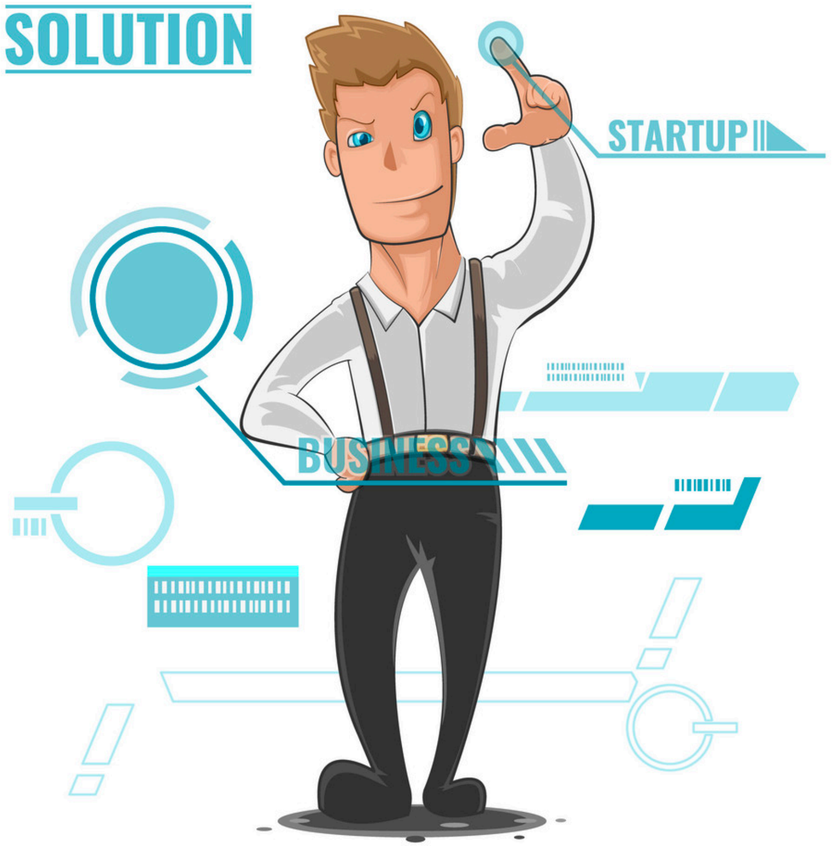


Image Source: vectorstock.com

Business Analytics

Particular Example

- James sends out consumer surveys with his product designs to locals
- James rewards customers who offer a feedback after a purchase of James' products
- James collects demographic information on people visiting his stores to aid targeting advertising
- James automates this task on the cloud!

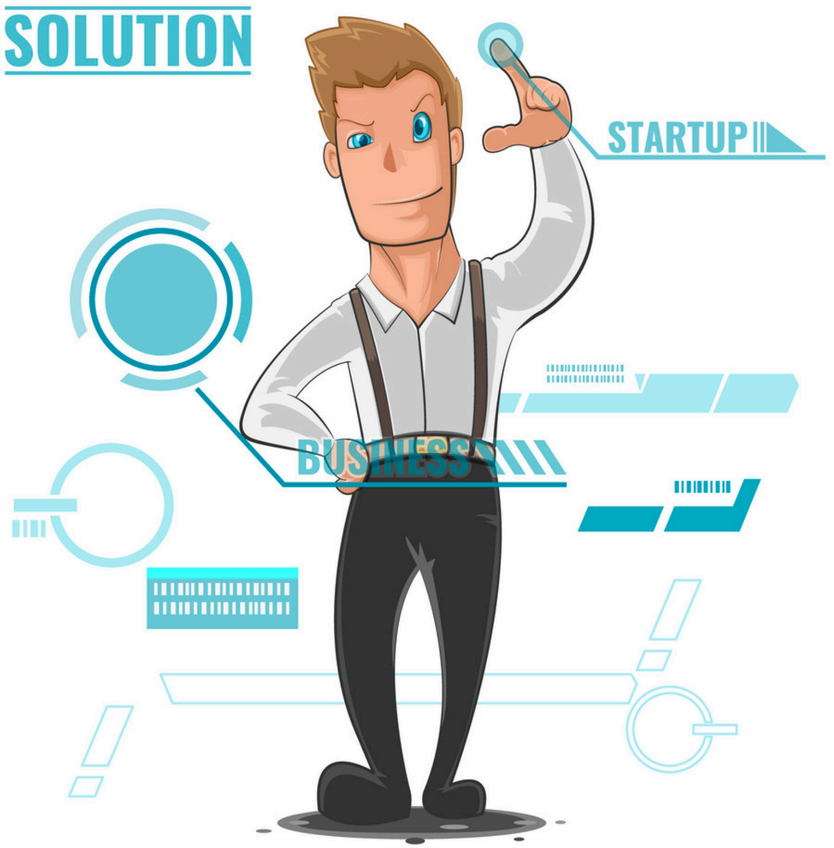


Image Source: vectorstock.com

Business

Use Cases

- Recommendation Engines
- Marketing & Sales Analytics
- Spatial-temporal Analysis
- Consumer Surveys & Product Improvements
- Quality Assurance/Quality Control

Finance

Michelle

- Michelle is a Finance professor who has also worked with Hedge-funds.
- She is developing cheaper customer alternatives to getting their portfolios hedged by financial advisors at a premium fee.
- She identifies the problem at hand, requirements from a solution and the risk involved.
- She has a team of students who are good at research to help her build a solution.

Image Source: vectorstock.com

Quantitative Finance

Particular Example

- Michelle filters her search of Research Studies by factors specific to her problem statement
- Michelle stores sample customer portfolio data in a SQL database
- Michelle's team uses Python libraries for building various statistical models from their data
- Michelle's team visualizes results from a particular strategy using Tableau!

Image Source: vectorstock.com

Finance

Use Cases

- Risk management, hedging of portfolios
- Tracking online transactions, data storage & locks
- User authentication systems & fraud detection
- Stock price modeling, Volatility modeling, Balance modeling
- Market-sector analysis: Efficient frontiers (or mean-variance optimization)
- Algorithmic Trading & performance backtesting

Scientific Disciplines

Thomas

- Thomas is a PhD fellow in Microbiology & is pursuing a project with the department of agricultural sciences in his University.
- He is conducting a field-study on the microorganisms present in soil based on a variety of factors.
- He wants to collect data periodically for a temporal research study.
- He has identified current and potential future problems in the logistics of this setup.

Image Source: vectorstock.com

Microbiology

Particular Example

- Thomas uses IoT sensors on-site to record & collect information such has temperature, humidity, pH levels etc.
- Thomas has connected his IoT sensors to the cloud by use cloud functions using Python Programming to transport data from on-site devices to online data stores
- Thomas pulls data from the cloud to his personal machine & uses MS Excel to calculate descriptive statistics & make plots

Image Source: vectorstock.com

Scientific Disciplines

Use Cases

- Generation of synthetic data for development of advanced models
- Heavy usage of relational databases for storage & retrieval of data
- Simulation Studies & Hypothesis Testing
- Setting up architectures for temporal studies & multi-group studies

Retail Sales

Hannah

- Hannah’s parents own a retail store that sells pet supplies & provides grooming and daycare services.
- She wants to increase their customer base by creating a substantial online presence & offering deliveries to customers from different states in the US.
- She understands the requirements of creating & maintaining an e-commerce website for the store & syncing inventories.
- She also understands optimization problems associated with out-of-state deliveries.

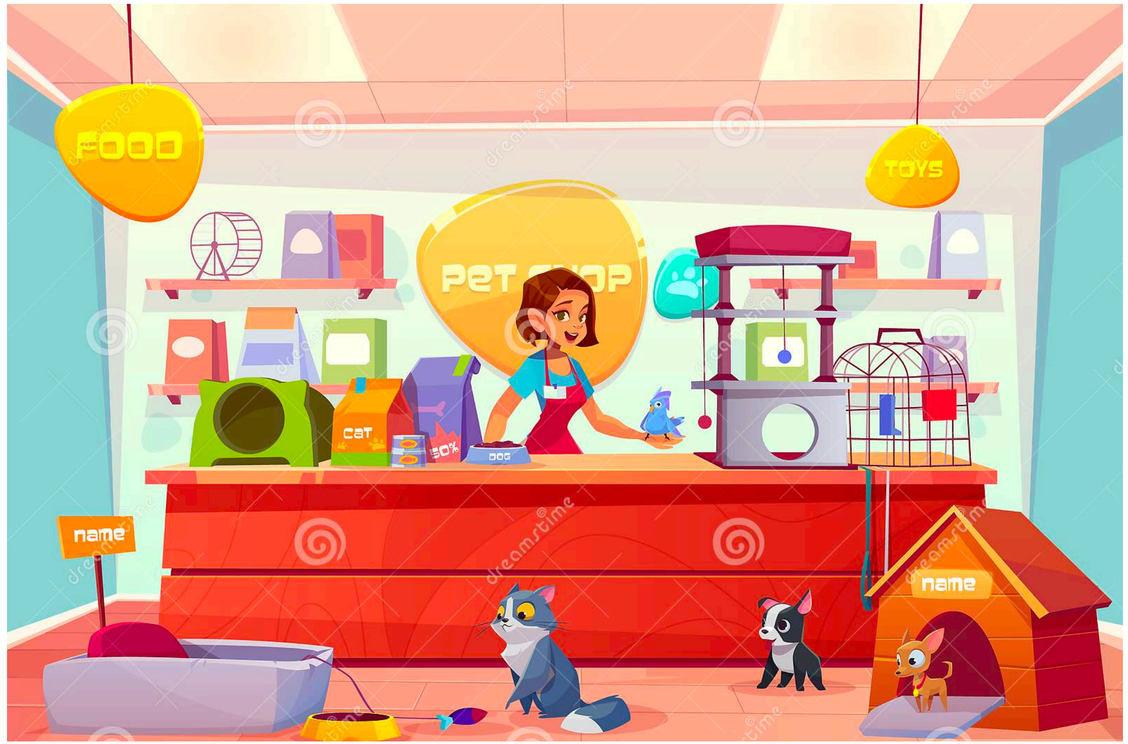


Image Source: dreamstime.com

Spatio-temporal Analysis

Particular Example

- Hannah maintains SQL databases that store and update inventory in each of the stores
- Hannah hosts her online e-commerce website & advertises it on social media where she constantly follows the response to her new strategies
- Hannah uses EDA on a history of out-of-state deliveries to determine if multiple products could be sent in lesser packages or lesser trips
- Hannah uses a temporal analysis in different geographic locations to predict the amount of inventory she should maintain seasonally

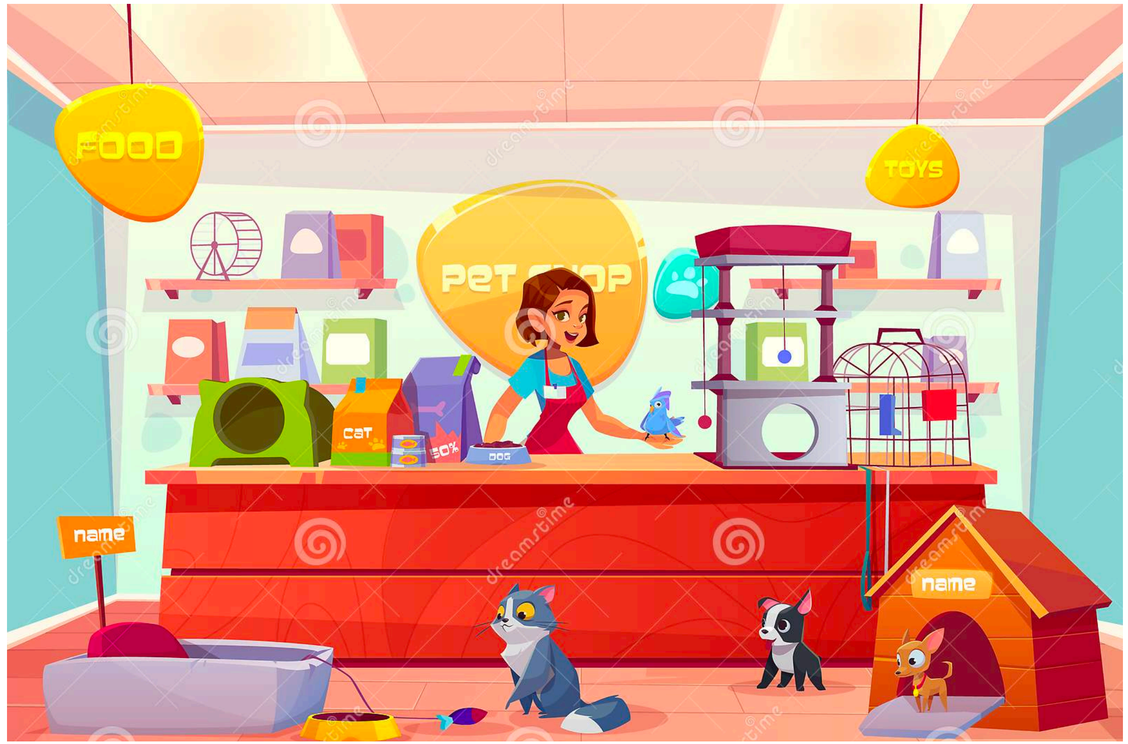


Image Source: dreamstime.com

Retail Sales

Use Cases

- Market-basket analysis to decide placement of products in a store
- Operational optimization by understanding patterns with historical data
- Social Media & Web Analytics to study how promotions or new strategies are received by the general public
- Sentiment Analysis to guide changes or improvements with products/ services
- Competitive pricing & demand prediction: dealing with Phantom Ordering & the Bullwhip effect

Technology

Mark

- Mark is a Comp Science major who loves gaming & has created an action-sequence game that he released for free online.
- He has data such as player statistics, count of clicks, character preferences, times the gaming application was opened/closed etc.
- He also has extra performance data that he has collected for each chapter/challenge in his game.
- He can additionally access data from discussion boards & session chats from the game.

Image Source: freepik.com

Bottom-Up Data Science

Particular Example

- Mark studies data about clicks to improve the user-interface of his gaming application
- Mark uses application open/close data to estimate if his application stopped responding on a host's computer device
- Mark uses player performance on a certain chapter of his game to determine the difficulty level of challenges/chapters released in the future
- Mark performs text-analysis on chat data to gain feedback on his application

Image Source: freepik.com

Technology

Use Cases

- Analysis of consumer activity & responses on Web pages for interface & content improvement
- Providing online customer service using chatbots, automatic email conversations
- Online Psychologist: Eliza!
- Development of new products for the market
- Enhancing searches and information retrieval

Healthcare & Medicine

Kim

- Kim is a Medical Resident.
- She is assisting on a clinical trial on Alzheimer's patients to monitor the effectiveness of a drug treatment.
- She wants to direct efforts towards round the clock monitoring of patient stress levels when they go home after taking their periodic medication & checking in personally at the hospital.
- She understands use of wearable technology & wants to find an association between patient stress levels & chances of detrimental risk.

Image Source: dreamstime.com

Clinical Trial

Particular Example

- Kim collects real-time data to determine patient stress levels & connects the feed to a dashboard that gives her all the required stats
- Kim provides a quick assisted chat option to her patients for problem resolutions based on past experiences & drug side-affects (if any)
- Kim collects geolocation data on her patients in case they are lost / out of reach
- Kim attempts to predict success/failure of the clinical trial based on continued monitoring of her patients' health

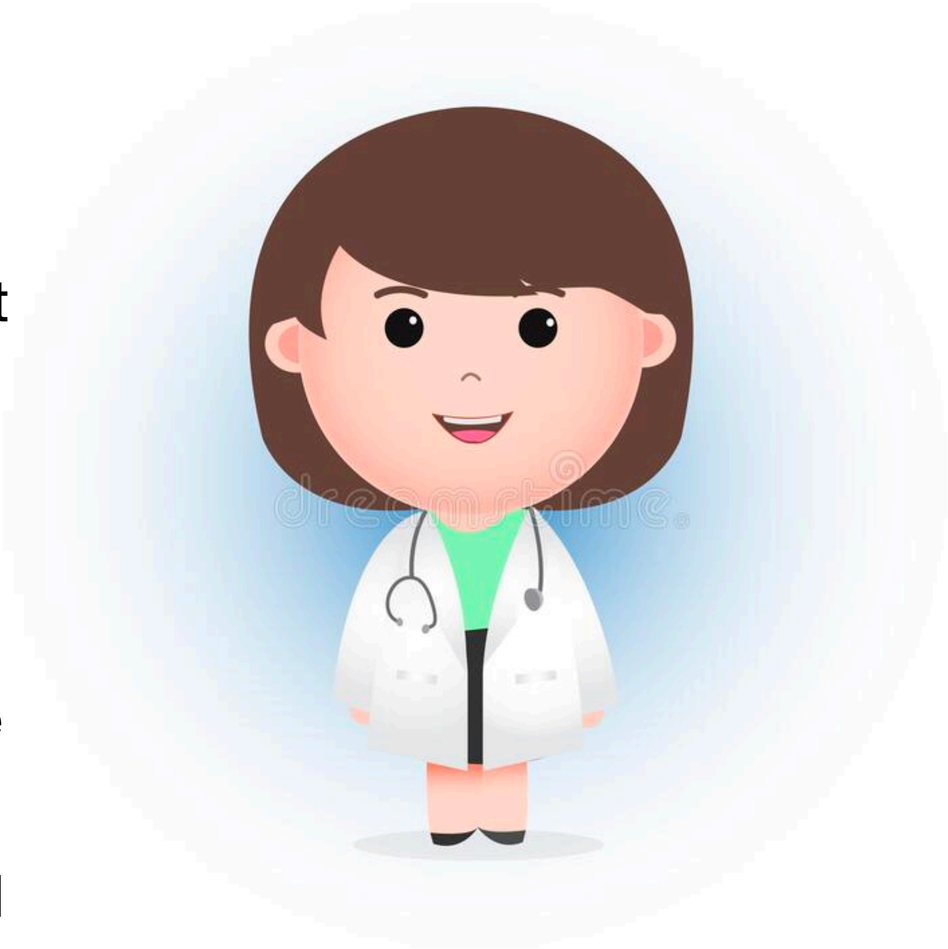


Image Source: dreamstime.com

Healthcare & Medicine

Use Cases

- Centralized access & real-time tracking of patient health
- Making treatments accessible to remote areas by developing virtual assistance standalone softwares
- Data transformation & sharing for medical research

Key Takeaways

- Every domain has numerous applications for Data Science, knowledge of which can boost your performance, efficiency & profile
- It can be daunting at first when you look at complex techniques, but there is always a way to start small
- You don’t have to be an expert at the three major disciplines to learn & use data science (you can though!)
- It’s no longer a want but very much a need
- There are tools available that make your journey with Data Science easier

But…

Source: Bill Waterson | Universal Press Syndicate

Upcoming Workshops

https://libcal.rutgers.edu/nblworkshops

- “Let the numbers talk!" Use Excel to learn how probability & statistics speak for your data : Oct 06
- “You will spend nearly 70% of your time doing this!" Organize & pre-process your Data : Oct 13
- The Power of Visual Storytelling: Learning Tableau Public : Oct 20
- “Make your computer work for you!" Learn how to use Python to program your tasks- Part 1 : Oct 27
- “Make your computer work for you!" Explore popular Data Science libraries in Python - Part 2 : Nov 03