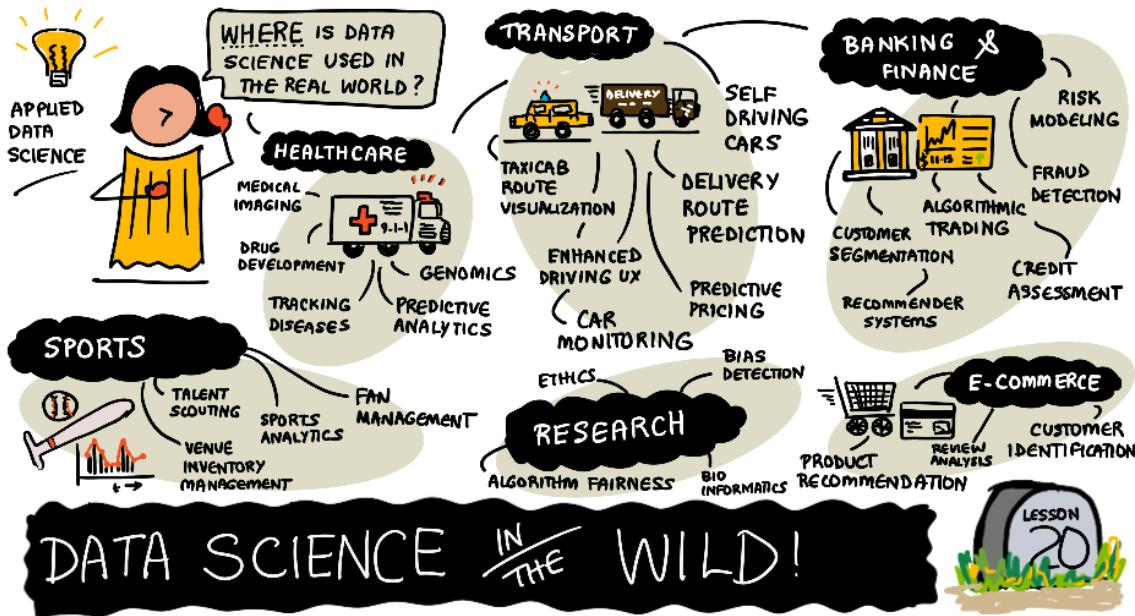


Data Science in the Real World



Data Science In The Real World - Sketchnote by [@nitya](#)

We're almost at the end of this learning journey!

We started with definitions of data science and ethics, explored various tools & techniques for data analysis and visualization, reviewed the data science lifecycle, and looked at scaling and automating data science workflows with cloud computing services. So, you're probably wondering: *"How exactly do I map all these learnings to real-world contexts?"*

In this lesson, we'll explore real-world applications of data science across industry and dive into specific examples in the research, digital humanities, and sustainability, contexts. We'll look at student project opportunities and conclude with useful resources to help you continue your learning journey!

Pre-Lecture Quiz

[Pre-lecture quiz](#)

Data Science + Industry

Thanks to the democratization of AI, developers are now finding it easier to design and integrate AI-driven decision-making and data-driven insights into user experiences and development workflows. Here are a few examples of how data science is "applied" to real-world applications across the industry:

- [Google Flu Trends](#) used data science to correlate search terms with flu trends. While the approach had flaws, it raised awareness of the possibilities (and challenges) of data-driven healthcare predictions.
- [UPS Routing Predictions](#) - explains how UPS uses data science and machine learning to predict optimal routes for delivery, taking into account weather conditions, traffic patterns, delivery deadlines and more.

- [NYC Taxicab Route Visualization](#) - data gathered using [Freedom Of Information Laws](#) helped visualize a day in the life of NYC cabs, helping us understand how they navigate the busy city, the money they make, and the duration of trips over each 24-hour period.
- [Uber Data Science Workbench](#) - uses data (on pickup & dropoff locations, trip duration, preferred routes etc.) gathered from millions of uber trips *daily* to build a data analytics tool to help with pricing, safety, fraud detection and navigation decisions.
- [Sports Analytics](#) - focuses on *predictive analytics* (team and player analysis - think [Moneyball](#) - and fan management) and *data visualization* (team & fan dashboards, games etc.) with applications like talent scouting, sports gambling and inventory/venue management.
- [Data Science in Banking](#) - highlights the value of data science in the finance industry with applications ranging from risk modeling and fraud detection, to customer segmentation, real-time prediction and recommender systems. Predictive analytics also drive critical measures like [credit scores](#).
- [Data Science in Healthcare](#) - highlights applications like medical imaging (e.g., MRI, X-Ray, CT-Scan), genomics (DNA sequencing), drug development (risk assessment, success prediction), predictive analytics (patient care & supply logistics), disease tracking & prevention etc.

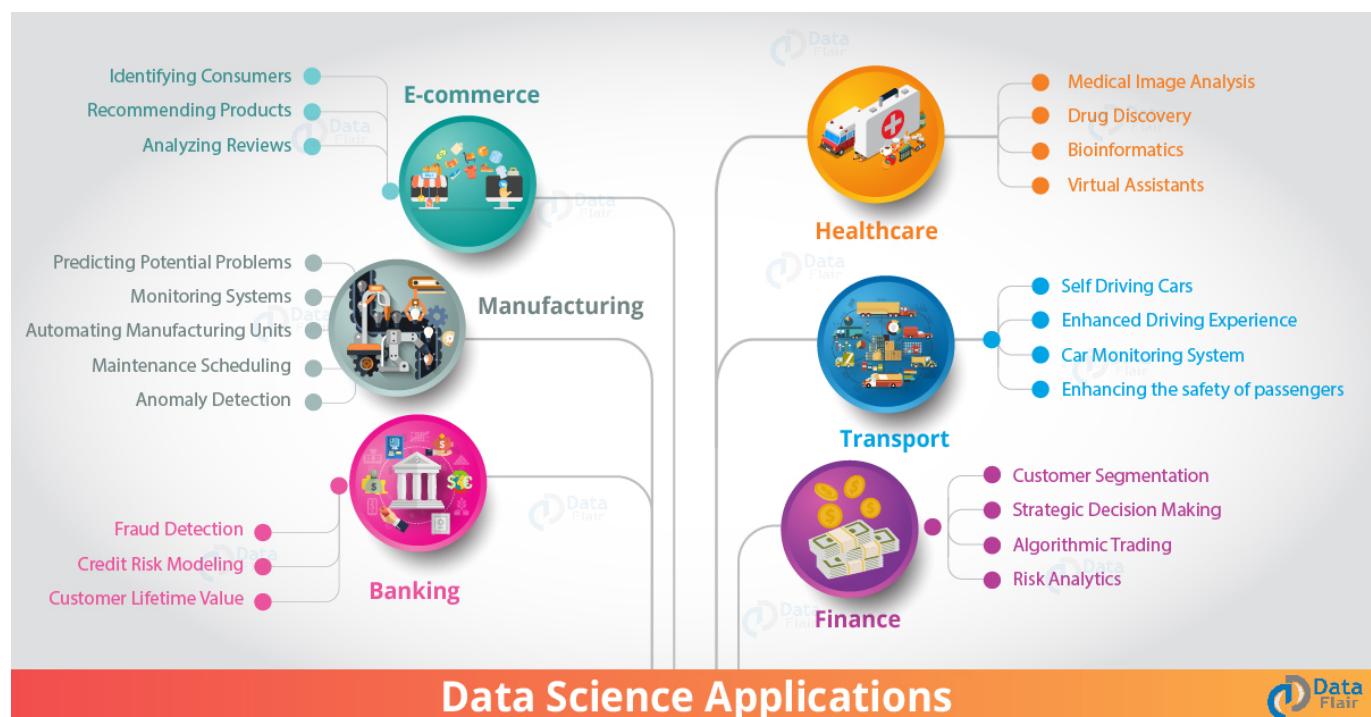
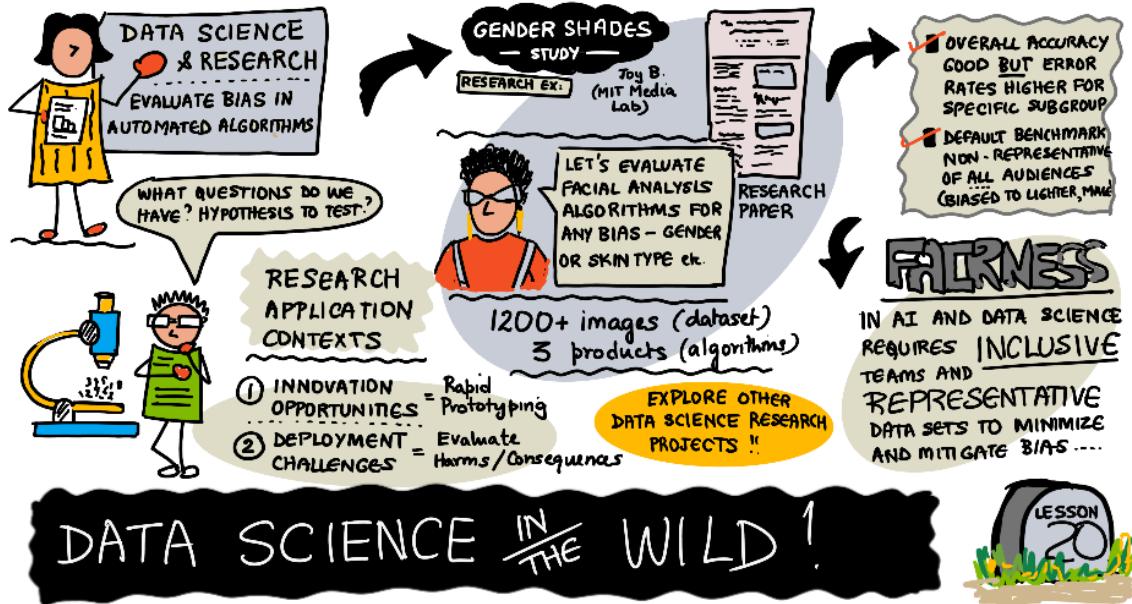


Image Credit: [Data Flair: 6 Amazing Data Science Applications](#)

The figure shows other domains and examples for applying data science techniques. Want to explore other applications? Check out the [Review & Self Study](#) section below.

Data Science + Research



Data Science & Research - Sketchnote by [@nitya](#)

While real-world applications often focus on industry use cases at scale, *research* applications and projects can be useful from two perspectives:

- *innovation opportunities* - explore rapid prototyping of advanced concepts and testing of user experiences for next-generation applications.
- *deployment challenges* - investigate potential harms or unintended consequences of data science technologies in real-world contexts.

For students, these research projects can provide both learning and collaboration opportunities that can improve your understanding of the topic, and broaden your awareness and engagement with relevant people or teams working in areas of interest. So what do research projects look like and how can they make an impact?

Let's look at one example - the [MIT Gender Shades Study](#) from Joy Buolamwini (MIT Media Labs) with a [signature research paper](#) co-authored with Timnit Gebru (then at Microsoft Research) that focused on

- **What:** The objective of the research project was to *evaluate bias present in automated facial analysis algorithms and datasets* based on gender and skin type.
- **Why:** Facial analysis is used in areas like law enforcement, airport security, hiring systems and more - contexts where inaccurate classifications (e.g., due to bias) can cause potential economic and social harms to affected individuals or groups. Understanding (and eliminating or mitigating) biases is key to fairness in usage.
- **How:** Researchers recognized that existing benchmarks used predominantly lighter-skinned subjects, and curated a new data set (1000+ images) that was *more balanced* by gender and skin type. The data set was used to evaluate the accuracy of three gender classification products (from Microsoft, IBM & Face++).

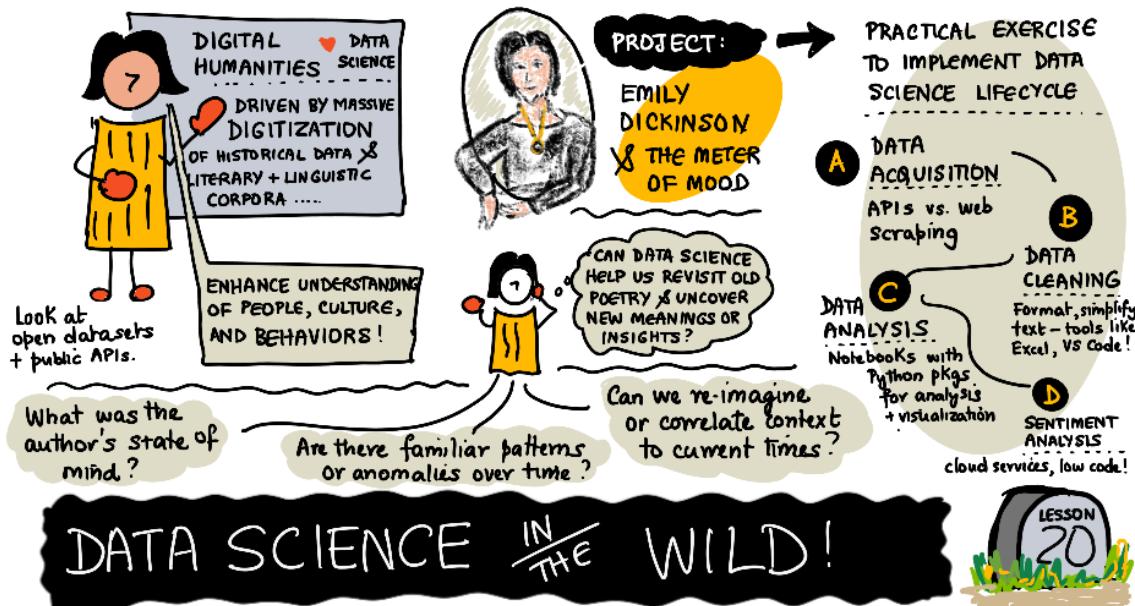
Results showed that though overall classification accuracy was good, there was a noticeable difference in error rates between various subgroups - with **misgendering** being higher for females or persons with darker skin types, indicative of bias.

Key Outcomes: Raised awareness that data science needs more *representative datasets* (balanced subgroups) and more *inclusive teams* (diverse backgrounds) to recognize and eliminate or mitigate such biases earlier in AI solutions. Research efforts like this are also instrumental in many organizations defining principles and practices for *responsible AI* to improve fairness across their AI products and processes.

Want to learn about relevant research efforts in Microsoft?

- Check out [Microsoft Research Projects](#) on Artificial Intelligence.
- Explore student projects from [Microsoft Research Data Science Summer School](#).
- Check out the [Fairlearn](#) project and [Responsible AI](#) initiatives.

Data Science + Humanities



Data Science & Digital Humanities - Sketchnote by [@nitya](#)

Digital Humanities [has been defined](#) as "a collection of practices and approaches combining computational methods with humanistic inquiry". [Stanford projects](#) like "*rebooting history*" and "*poetic thinking*" illustrate the linkage between [Digital Humanities](#) and [Data Science](#) - emphasizing techniques like network analysis, information visualization, spatial and text analysis that can help us revisit historical and literary data sets to derive new insights and perspective.

Want to explore and extend a project in this space?

Check out "[Emily Dickinson and the Meter of Mood](#)" - a great example from [Jen Looper](#) that asks how we can use data science to revisit familiar poetry and re-evaluate its meaning and the contributions of its author in new contexts. For instance, *can we predict the season in which a poem was authored by analyzing its tone or sentiment* - and what does this tell us about the author's state of mind over the relevant period?

To answer that question, we follow the steps of our data science lifecycle:

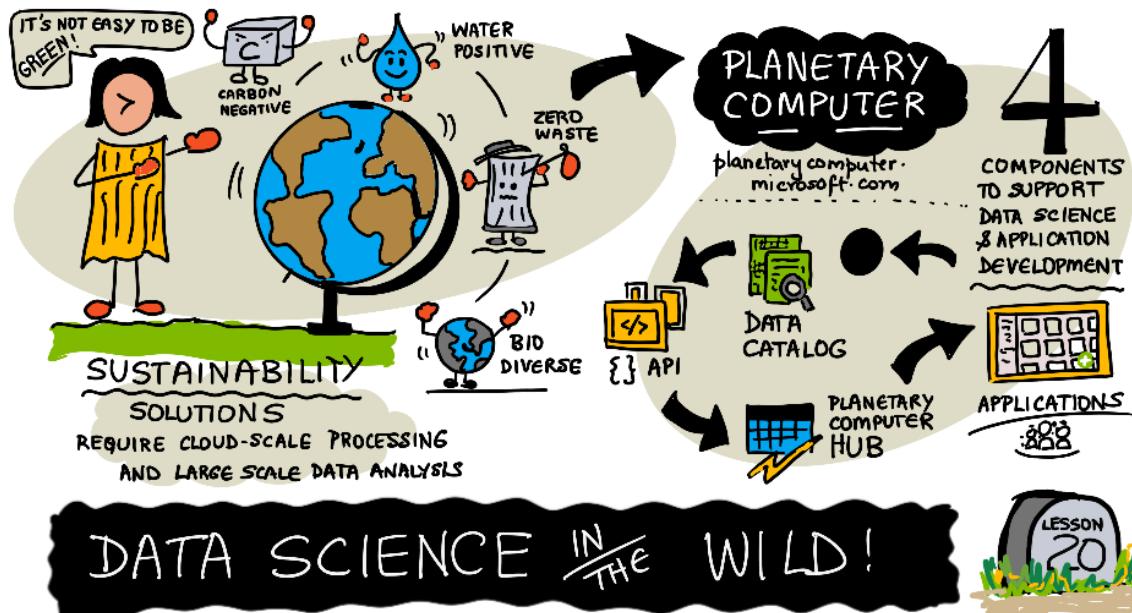
- **Data Acquisition** - to collect a relevant dataset for analysis. Options including using an API (e.g., [Poetry DB API](#)) or scraping web pages (e.g., [Project Gutenberg](#)) using tools like [Scrapy](#).

- **Data Cleaning** - explains how text can be formatted, sanitized and simplified using basic tools like Visual Studio Code and Microsoft Excel.
- **Data Analysis** - explains how we can now import the dataset into "Notebooks" for analysis using Python packages (like pandas, numpy and matplotlib) to organize and visualize the data.
- **Sentiment Analysis** - explains how we can integrate cloud services like Text Analytics, using low-code tools like **Power Automate** for automated data processing workflows.

Using this workflow, we can explore the seasonal impacts on the sentiment of the poems, and help us fashion our own perspectives on the author. Try it out yourself - then extend the notebook to ask other questions or visualize the data in new ways!

You can use some of the tools in the [Digital Humanities toolkit](#) to pursue these avenues of inquiry

Data Science + Sustainability



Data Science & Sustainability - Sketchnote by [@nitya](#)

The [2030 Agenda For Sustainable Development](#) - adopted by all United Nations members in 2015 - identifies 17 goals including ones that focus on **Protecting the Planet** from degradation and the impact of climate change. The [Microsoft Sustainability](#) initiative supports these goals by exploring ways in which technology solutions can support and build more sustainable futures with a [focus on 4 goals](#) - being carbon negative, water positive, zero waste, and bio-diverse by 2030.

Tackling these challenges in a scalable and timely manner requires cloud-scale thinking - and large scale data. The [Planetary Computer](#) initiative provides 4 components to help data scientists and developers in this effort:

- [Data Catalog](#) - with petabytes of Earth Systems data (free & Azure-hosted).
- [Planetary API](#) - to help users search for relevant data across space and time.
- [Hub](#) - managed environment for scientists to process massive geospatial datasets.
- [Applications](#) - showcase use cases & tools for sustainability insights.

The Planetary Computer Project is currently in preview (as of Sep 2021) - here's how you can get started contributing to sustainability solutions using data science.

- [Request access](#) to start exploration and connect with peers.
- [Explore documentation](#) to understand supported datasets and APIs.
- Explore applications like [Ecosystem Monitoring](#) for inspiration on application ideas.

Think about how you can use data visualization to expose or amplify relevant insights into areas like climate change and deforestation. Or think about how insights can be used to create new user experiences that motivate behavioral changes for more sustainable living.

Data Science + Students

We've talked about real-world applications in industry and research, and explored data science application examples in digital humanities and sustainability. So how can you build your skills and share your expertise as data science beginners?

Here are some examples of data science student projects to inspire you.

- [MSR Data Science Summer School](#) with GitHub projects exploring topics like:
 - [Racial Bias in Police Use of Force | Github](#)
 - [Reliability of NYC Subway System | Github](#)
- [Digitizing Material Culture: Exploring socio-economic distributions in Sirkap](#)- from [Ornella Altunyan](#) and team at Claremont, using [ArcGIS StoryMaps](#).

Challenge

Search for articles that recommend data science projects that are beginner friendly - like [these 50 topic areas](#) or [these 21 project ideas](#) or [these 16 projects with source code](#) that you can deconstruct and remix. And don't forget to blog about your learning journeys and share your insights with all of us.

Post-Lecture Quiz

[Post-lecture quiz](#)

Review & Self Study

Want to explore more use cases? Here are a few relevant articles:

- [17 Data Science Applications and Examples](#) - Jul 2021
- [11 Breathtaking Data Science Applications in Real World](#) - May 2021
- [Data Science In The Real World](#) - Article Collection
- Data Science In: [Education](#), [Agriculture](#), [Finance](#), [Movies](#) & more.

Assignment

[Explore A Planetary Computer Dataset](#)