

AI4D-Lab Tanzania - The University of Dodoma, CIVE



Artificial Intelligence Short Course

Launching a Career as a Data Scientist

Zephania Reuben

June 28, 2022



Outline

1 Build your Career as a Data Scientist

- How do I land a career as a data scientist?
- Learning approaches
- Building portfolio of real world projects

2 Roles and Requirements

- How much math should I learn for DS / ML?
- What makes a good data scientist?
- Am I too old / too young to become a data scientist?

3 Future Proofing Your Career

- Libraries / pre-existing solutions, or code from scratch?
- How can I future-proof my skills and career?
- Making money from home / Are there remote opportunities?

4 General Advice for

- People/students with business/computer studies background
- Someone with no relevant work experience



Build your Career as a Data Scientist

What is Data Science?

Data Science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. **Wikipedia**



How do I land a Career as a Data Scientist?

All Data Scientists

- Data Analysis / Exploratory Analysis
- Data Preprocessing
- Applied Machine Learning



How do I Land a Career as a Data Scientist?

Business Data Scientists

- Insight knowledge
 - Marketing
 - Strategy
 - Operations
 - Inversiting
- Domain knowledge
- Communication and presentations



How do I land a Career as a Data Scientist?

Product Data Scientists

- Product knowledge
 - E-commerce
 - Entertainment
 - Banking
 - SaaS
- Software development basics
- Data pipelines



Building Portfolio of Real World Projects

Steps

- 1 Learn the core skills for data science and applied machine learning - **Roadmap**
- 2 Pick out a dataset to start with - **Datasets**
- 3 Start your project in Jupyter Notebook.
- 4 Explore the data and make sure you understand the features.
- 5 Define an interesting objective to pursue
- 6 Clean the data, engineer features.
- 7 Complete your analysis / train your models.
- 8 Write about your project directly inside your Jupyter notebook.
- 9 Upload your project to Github. **Repeat 1 - 9**



Roles and Requirements

How Much Mathematics Should I Learn for DS / ML?

- Probably very little Mathematics foundation.
- Depends on your goal and learning approach.



Roles and Requirements

What Makes a Good Data Scientist?

- Always drive toward business value.
- Develop elite communication skills.
- Know when when NOT to use ML.
- Understand tradeoffs / deadlines.
- Focus on feature engineering.
- Consider themselves lifelong students



Roles and Requirements

Am I too old / too young to become a data scientist?

Anyone can become a data scientist, regardless of age, educational background, or prior work experience.

- Develop real skills.
- Build a portfolio of projects that help to prove the real skills.



Future Proofing Your Career

Career Paths in Data Science

- Level Up
- Choose your own adventure



Future Proofing Your Career

Libraries / pre-existing solutions, or code from scratch?

For learning purposes, you can choose to code a few of your favorite algorithms from scratch.



Future Proofing Your Career

Coding From Scratch - Advantages

- Learning how the algorithms works.
- Customizable implementation.
- Potentially faster implementations.

Coding From Scratch - Disadvantages

- Higher math programming requirements.
- Takes a long time.
- Difficult to beat pre-existing libraries.



Future Proofing Your Career

Libraries and Pre-existing Solutions - Advantages

- Easier to learn.
- Much more commercial demand.
- Pre-optimized implementations allow you to focus on the application and building better models.

Libraries and Pre-existing Solutions - Disadvantages

- Cannot customize implementations.
- Cannot see each step of the algorithms.
- Limited in functionality by what's already there in the pre-existing library.



Future Proofing Your Career

How can I Future-proof my Skills and Career?

- Opportunity assessment
- Creativity
- Domain expertise
- Nuanced decision making
- Empathy



Future Proofing Your Career

Data Science Competitions

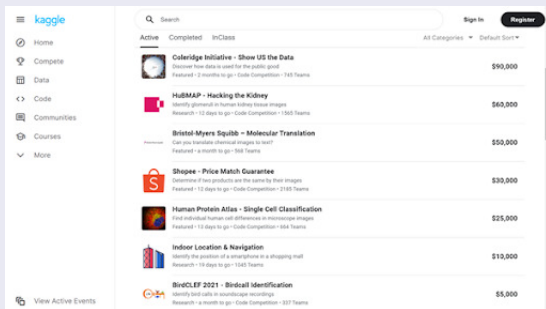
- **Data Science Competition** is the process of participating in data science challenge to solve complex business problems and share top solutions to the organizations that own the data.
- Data science competitions help organizations find solutions to their complex business problems while enabling data scientists to learn from the experience and win awards/prizes.



Future Proofing Your Career

Data Science Competition Platforms

■ Kaggle (Level: Beginner, Intermediate Advanced)



The screenshot displays the Kaggle homepage with a sidebar on the left containing navigation links: Home, Compete, Data, Code, Communities, Courses, and More. The main content area features a search bar and tabs for 'Active', 'Completed', and 'InClass' competitions. A list of active competitions is shown, each with a title, description, prize amount, and number of teams.

Competition	Prize	Teams
Coleridge Initiative - Show US the Data	\$90,000	145 Teams
HubMAP - Hacking the Kidney	\$60,000	1565 Teams
Bristol Myers Squibb - Molecular Translation	\$50,000	-
Shopee - Price Match Guarantee	\$30,000	2185 Teams
Human Protein Atlas - Single Cell Classification	\$25,000	884 Teams
Indoor Location & Navigation	\$10,000	1045 Teams
BirdCLEF 2021 - Birdcall Identification	\$5,000	337 Teams



Future Proofing Your Career

Data Science Competition Platforms

■ Zindi Africa (Level : Beginner, Intermediate Advanced))

The screenshot displays the Zindi Africa website interface. At the top, the ZINDI logo is followed by navigation links: Competitions, Hackathons, Data Scientists, Discussions, Jobs, Learning, About Us, and Sign in. Below the navigation bar is a horizontal menu with various categories: Safety, Health, Education, Transport, Disaster Services, Finance, Reinforcement Learning, Media, Quantitative, Biology, Communications, Customer Service, Weather, Logistics, Computer Vision, Genomics, Satellite, Arts & Culture, and others. Two competitions are featured on the page:

- CGIAR Crop Yield Prediction Challenge**: Can you predict maize yields on East African farms using satellite data? Prize: \$3,000 USD. 418 data scientists enrolled, 195 on the leaderboard, ended 17 days ago. Tags: Agriculture, Prediction, Computer Vision, Unstructured, Image, Satellite, Q&A.
- Indaba Grand Challenge: Curing Leishmaniasis by Deep Learning**: Indaba. The official competition of the Deep Learning Indaba 2020. Africa. 3000 Zindi Points. 234 data scientists enrolled, 38 on the leaderboard, 11 days to go. Tags: Health, Reinforcement Learning, Q&A.



Future Proofing Your Career

Data Science Competition Platforms

- Data Hack by Analytics Vidhya (Level: Beginner Intermediate))

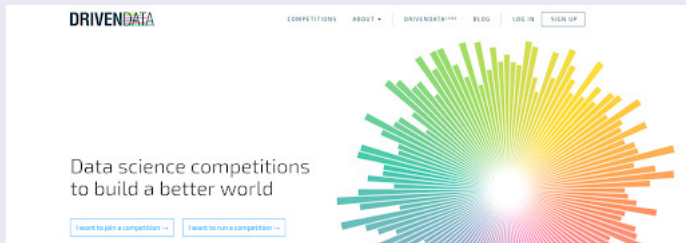
The screenshot displays the Analytics Vidhya website interface. At the top, the navigation bar includes links for Blog, Courses, Hackathons, BlackBelt, User Rankings, and Top Approaches, along with search and user profile icons. A banner below the navigation bar encourages users to "Choose from a wide range of competitions" and "Pick your competition to participate in from these categories. RSVP to events to meet like minded data scientists". Below the banner, there are filter buttons for All Contents, Hiring, Prize Money, Practice, Skill Tests, and Events. A status bar indicates the number of competitions in different states: Active (18), Late Submissions (40), and Closed (218). The main content area features four competition cards: "Data Science Blogathon 7", "Loan Prediction", "Big Mart Sales Prediction", and "Black Friday Sales Prediction". Each card shows a timer (4:10:26) and the type of problem (Assured Rewards or Practice Problem). A feedback button is visible on the right side of the page.



Future Proofing Your Career

Data Science Competition Platforms

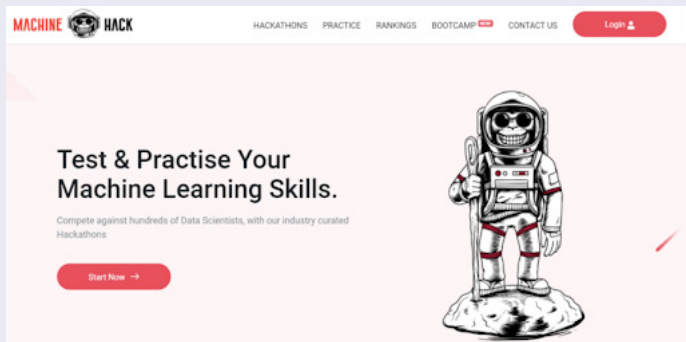
- DrivenData(Level: Beginner, Intermediate Advanced))



Future Proofing Your Career

Data Science Competition Platforms

- Machine Hack (Level: Beginner and Intermediate))



Future Proofing Your Career

Data Science Competition Platforms

■ HackerEarth(Level: Beginner and Intermediate)

The screenshot displays the HackerEarth homepage. At the top, there is a navigation bar with links for PRACTICE, COMPETE, COMPANIES, and LEADERBOARD, along with a search bar and LOGIN/SIGN UP buttons. A yellow banner below the navigation bar prompts users to refresh the page for their system timezone. The main content area is titled 'Hackathons, Programming Challenges, And Coding Competitions' and features tabs for LIVE, UPCOMING, and PREVIOUS events. Three featured challenges are shown: 'February Circuits '21' (RATED, 5330 participants, ends in 00:03:09), 'Sigma-Thon 2.0' (RACERATHON, 1846 participants, ends in 01:05:09), and 'University challenges are currently hidden' (ends in 01:06:08). A blue tooltip over the third challenge states: 'University challenges are currently hidden. Select this to view these challenges.' To the right, there is a section for 'CHALLENGE TYPE' with checkboxes for Competitive Challenges, Hackathons, Hiring Challenges (jobs), and University Challenges. Below this is a section titled 'HOST YOUR OWN HACKATHON' with a brief description.

h PRACTICE COMPETE COMPANIES LEADERBOARD Search developers, problems, etc. LOGIN SIGN UP

Please refresh the page to see the time in your system timezone

Hackathons, Programming Challenges, And Coding Competitions

LIVE UPCOMING PREVIOUS

February Circuits '21
RATED
5330
ENDS IN
00 : 03 : 09
START NOW

Sigma-Thon 2.0
RACERATHON
1846
ENDS IN
01 : 05 : 09
START NOW

University challenges are currently hidden.
ENDS IN
01 : 06 : 08
START NOW

CHALLENGE TYPE

- ☒ Competitive Challenges
- ☒ Hackathons
- ☒ Hiring Challenges (jobs)
- ☐ University Challenges

HOST YOUR OWN HACKATHON

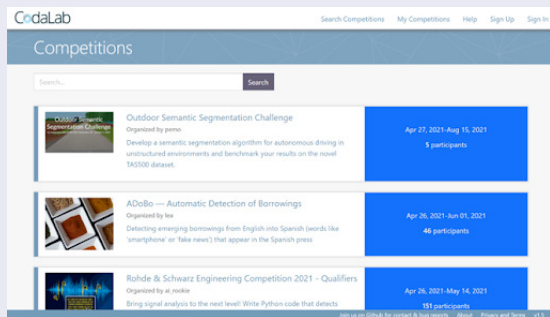
Source top talents and innovative ideas by conducting best-in-class hackathons.



Future Proofing Your Career

Data Science Competition Platforms

■ CodaLab(Level: Advanced))



The screenshot displays the CodaLab website's 'Competitions' section. At the top, there is a navigation bar with links for 'Search Competitions', 'My Competitions', 'Help', 'Sign Up', and 'Sign In'. Below this is a search bar with the placeholder text 'Search...'. The main content area lists three competitions, each with a thumbnail image, title, organizer, description, dates, and participant count.

Competition Title	Organized by	Description	Dates	Participants
Outdoor Semantic Segmentation Challenge	pemo	Develop a semantic segmentation algorithm for autonomous driving in unstructured environments and benchmark your results on the novel TAS500 dataset.	Apr 27, 2021-Aug 15, 2021	5 participants
ADoBo — Automatic Detection of Borrowings	lee	Detecting emerging borrowings from English into Spanish (words like 'smartphone' or 'Take news') that appear in the Spanish press	Apr 26, 2021-Jun 01, 2021	46 participants
Rohde & Schwarz Engineering Competition 2021 - Qualifiers	at, rookie	Bring signal analysis to the next level! Write Python code that detects	Apr 26, 2021-May 14, 2021	151 participants

At the bottom of the page, there is a footer with links for 'Join us on GitHub for contact & bug reports', 'About', 'Privacy and Terms', and a version number 'v1.5'.



Future Proofing Your Career

Benefits of Participating in Data Science Competitions

- Test your skills against Top Talent.
- Learn by doing and gain exposure.
- Earn income by doing what you enjoy and do best.
- Build your Profile and attract potential employers.
- Apply for Data Science job opportunities.
- Learn from others through collaboration and discussion.
- Networking with like-minded people.



Future Proofing Your Career

Types of Data Science Competitions

- Prize Competitions.
- Points or Medals Competitions.
- Knowledge Competitions. Recommended for beginners



Future Proofing Your Career

Making Money From Home / Are There Remote Opportunities?

- Freelancing: **UpWork**, **Toptal**
- Competitions and Hackathons e.g Zindi.
- Working with enterprise customers/ small businesses to solve really hard customer problems in the various domains. Doing consultancy work in data science.
- Work as part of a core ML/ Research team in an academic team at a major university.
- Writing technical articles, data science educational practice, and data blogging to teach, earn, and learn as well.
- Public speaking at paid large tech/ Machine Learning conferences, workshops, bootcamps, and meetups.



General Advice for

People/Students with IT/Business Studies Background

- 1 Data science is not only machine learning, analytical skills are crucial.
- 2 Skip most of the Math.
- 3 Be prepared for the mindset difference between software development and data science.
- 4 Domain knowledge can help you stand out big time
- 5 Practice your communication skills.



General Advice for

Someone with no Relevant Work Experience

- Develop the real skills capable of driving business value.
- Follow a top-down approach.



