



PyCon Tanzania

Build your Career as a Data Scientist

Arusha, Tanzania

Zephania Reuben

December 03, 2021



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 Requirement

How much math should I learn for DS / ML?

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

Roles and Requirement

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 Requirement

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

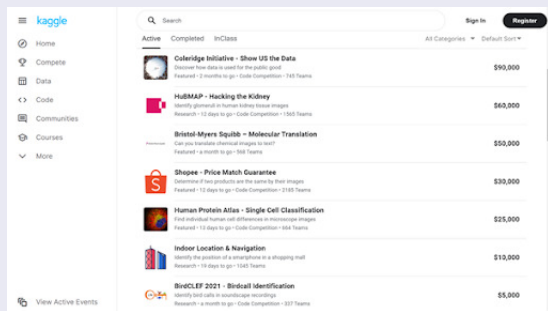
Data Science Competitions

- **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 website interface. On the left is a navigation sidebar with links for Home, Compete, Data, Code, Communities, Courses, and More. The main content area features a search bar and tabs for Active, Completed, and In-Class competitions. A list of active competitions is shown, each with a title, description, featured status, duration, number of teams, and prize amount.

Competition	Prize
Coleridge Initiative - Show US the Data Discover how data is used for the public good Featured • 2 months to go • Code Competition • 745 Teams	\$90,000
HubMAP - Hacking the Kidney Identify glomeruli in human kidney tissue images Research • 12 days to go • Code Competition • 1565 Teams	\$60,000
Bristol Myers Squibb - Molecular Translation Can you translate chemical images to text? Featured • a month to go • 588 Teams	\$50,000
Shopee - Price Match Guarantee Determine if two products are the same by their images Featured • 12 days to go • Code Competition • 2185 Teams	\$30,000
Human Protein Atlas - Single Cell Classification Find individual human cell differences in microscope images Featured • 13 days to go • Code Competition • 884 Teams	\$25,000
Indoor Location & Navigation Identify the position of a smartphone in a shopping mall Research • 19 days to go • 1245 Teams	\$10,000
BirdCLEF 2021 - Birdcall Identification Identify bird calls in soundscape recordings Research • a month to go • Code Competition • 337 Teams	\$5,000

Future Proofing Your Career

Data Science Competition Platforms

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

The screenshot displays the ZIND! website interface. At the top, the navigation menu includes 'Competitions', 'Hackathons', 'Data Scientists', 'Discussions', 'Jobs', 'Learning', 'About Us', and 'Sign in'. Below the navigation is a horizontal bar with various category tags such as 'Sales', 'Health', 'Education', 'Energy', 'Healthcare Services', 'Finance', 'Development/Learning', 'Media', 'Transportation', 'Agriculture', 'Customer service', 'Marketing', 'Logistics', 'Computer Science', 'Government', 'Science', 'Art & Culture', 'Retail', 'Security', 'Food for People', 'Location', 'Translation', 'Construction', 'Travel', 'Insurance', 'HR', and 'E-commerce'. There are also buttons for 'Active & Closed' and 'Any Rewards'.

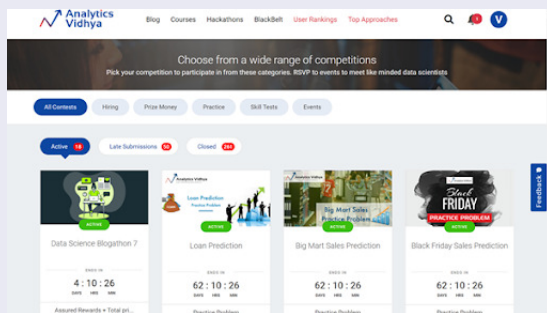
Two featured competitions are visible:

- CGIAR Crop Yield Prediction Challenge**: Can you predict maize yields on East African farms using satellite data? Prize: \$3,000 USD. 432 data scientists enrolled, 395 on the leaderboard, ended 17 days ago. Tags: Agriculture, Prediction, Computer Vision, Unstructured, Image, Satellite, GIS.
- Indaba Grand Challenge: Curing Leishmaniasis by Deep Learning**: Indaba. The official competition of the Deep Learning Indaba 2020. Africa. 234 data scientists enrolled, 38 on the leaderboard, 11 days to go. Tags: Health, Deep Learning Learning, GIS.

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. A search icon and a user profile icon are also present. Below the navigation is a banner with the text "Choose from a wide range of competitions" and "Pick your competition to participate in from these categories. RDVP to events to meet like minded data scientists".

Below the banner are several filter buttons: All Contents, Hiring, Prize Money, Practice, Skill Tests, and Events. Underneath these are status filters: Active (18), Late Submissions (0), and Closed (15).

The main content area features four competition cards:

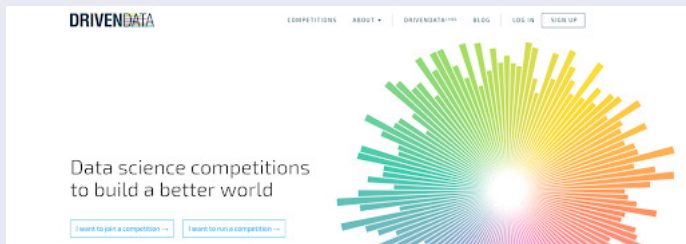
- Data Science Blogathon 7**: Status: Active. Duration: 4 : 10 : 26. Assured Rewards + Total pr...
- Loan Prediction**: Status: Active. Duration: 62 : 10 : 26. Practice Problem.
- Big Mart Sales Prediction**: Status: Active. Duration: 62 : 10 : 26. Practice Problem.
- Black Friday Sales Prediction**: Status: Active. Duration: 62 : 10 : 26. Practice Problem.

A vertical "Feedback" button is located on the right side of the competition grid.

Future Proofing Your Career

Data Science Competition Platforms

- DrivenData(Level: Beginner, Intermediate Advanced))

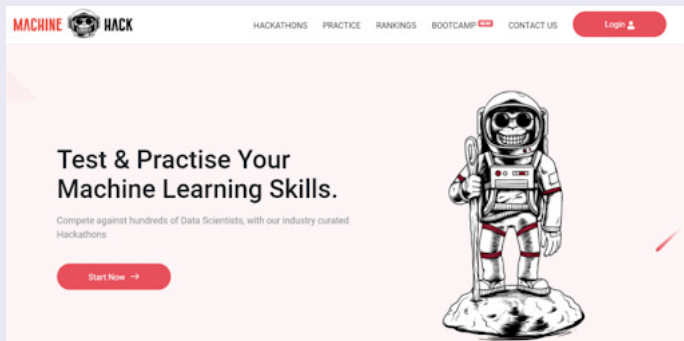


The screenshot shows the homepage of DrivenData. At the top left is the logo "DRIVEN DATA". To the right is a navigation menu with links for "COMPETITIONS", "ABOUT", "DRIVEN DATA", "BLOG", "LOG IN", and "SIGN UP". The main content area features the text "Data science competitions to build a better world" and two buttons: "I want to join a competition --" and "I want to run a competition --". On the right side of the page is a large, colorful graphic consisting of many thin, radiating lines in shades of green, yellow, orange, and red, resembling a stylized sunburst or data visualization.

Future Proofing Your Career

Data Science Competition Platforms

- Machine Hack (Level: Beginner and Intermediate))



The screenshot shows the homepage of the Machine Hack website. At the top, there is a navigation bar with the logo "MACHINE HACK" on the left, which includes a small robot head icon. To the right of the logo are links for "HACKATHONS", "PRACTICE", "RANKINGS", "BOOTCAMP" (with a red badge), and "CONTACT US". On the far right of the navigation bar is a red "Login" button with a user icon. The main content area has a light pink background. On the left, the text reads "Test & Practise Your Machine Learning Skills." followed by "Compete against hundreds of Data Scientists, with our industry curated Hackathons". Below this text is a red button labeled "Start Now" with a right-pointing arrow. On the right side of the main content area is a cartoon illustration of an astronaut in a white suit with red accents, holding a tool, standing on a small patch of grey ground. A red checkmark is visible to the right of the astronaut.

Future Proofing Your Career

Data Science Competition Platforms

■ HackerEarth (Level: Beginner and Intermediate)

The screenshot displays the HackerEarth website interface. At the top, there is a navigation bar with links for PRACTICE, COMPETE, COMPANIES, and LEADERBOARD. A search bar is located to the right of these links, and there are LOGIN and SIGN UP buttons. Below the navigation bar, a yellow banner contains a message: "Please refresh the page to see the time in your system timezone".

The main content area is titled "Hackathons, Programming Challenges, And Coding Competitions". It features three tabs: LIVE (selected), UPCOMING, and PREVIOUS. Below the tabs, there are three challenge cards:

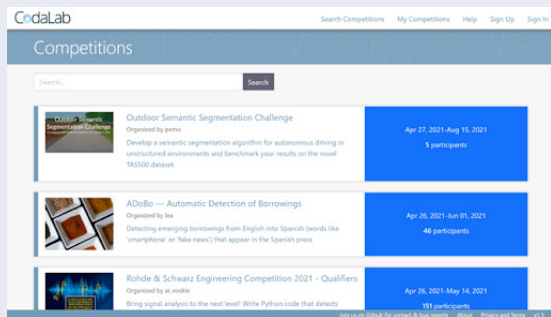
- February Circuits '21**: A RATED challenge with 5330 participants. It ends in 00:03:09. A green "START NOW" button is visible.
- Sigma-Thon 2.0**: A HACKATHON challenge with 1846 participants. It ends in 01:05:09. A green "START NOW" button is visible.
- University challenges**: A challenge with a blue tooltip that says "University challenges are currently hidden. Select this to view these challenges." and a "CLOSE" button. It ends in 01:06:08. A green "START NOW" button is visible.

On the right side of the page, there is a promotional banner for "Manipal Axis Bank job guaranteed program" with a "APPLY NOW" button. Below this, there is a "CHALLENGE TYPE" section with checkboxes for "Competitive Challenges", "Hackathons", "Hiring Challenges (jobs)", and "University Challenges". At the bottom right, there is a "HOST YOUR OWN HACKATHON" section with the text "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 shows the CodaLab website's 'Competitions' page. At the top, there is a navigation bar with links for 'Search Competitions', 'My Competitions', 'Help', 'Sign Up', and 'Sign In'. Below the navigation bar is a search bar with a 'Search' button. The main content area displays three competition cards:

Competition Title	Organized by	Start/End Dates	Participants
Outdoor Semantic Segmentation Challenge	pemo	Apr 27, 2021-Aug 15, 2021	5 participants
ADoBo — Automatic Detection of Borrowings	lee	Apr 26, 2021-Jun 01, 2021	46 participants
Rohde & Schwarz Engineering Competition 2021 - Qualifiers	at.rookie	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 '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.



ZIND!