



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

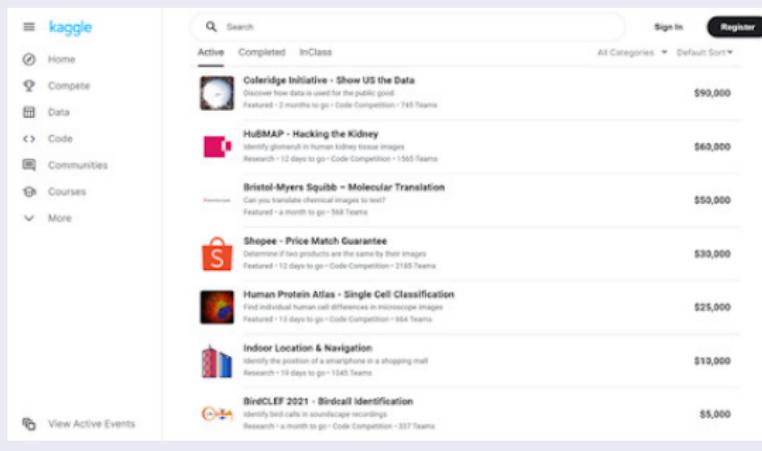
Data Science Competitions

- **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 shows the Kaggle homepage with a sidebar on the left and a main content area on the right.

Left Sidebar:

- Home
- Compete
- Data
- Code
- Communities
- Courses
- More

Right Content Area:

Search bar: Search

Filter buttons: Active, Completed, InClass

Sign In and Register buttons

Sort buttons: All Categories, Default Sort

List of competitions:

- Culveridge Initiative - Show US the Data**
Discover how data is used for the public good
Featured - 2 months to go - Code Competition - 748 Teams
\$90,000
- HuBMAP - Hacking the Kidney**
Identify diseases in human brain 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 - 384 Teams
\$50,000
- Shapee - 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 microscopic images
Featured - 13 days to go - Code Competition - 564 Teams
\$25,000
- Indoor Location & Navigation**
Identify the position of a smartphone in a shopping mall
Research - 19 days to go - 1045 Teams
\$10,000
- BirdCLEF 2021 - Birdcall Identification**
Identify bird calls in soundscape recordings
Research - a month to go - Code Competition - 337 Teams
\$5,000

View Active Events

Future Proofing Your Career

Data Science Competition Platforms

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

The screenshot shows the Zindi Africa website interface. At the top, there is a navigation bar with links for Competitions, Hackathons, Data Scientists, Discussions, Jobs, Learning, About Us, and Sign In. Below the navigation bar is a horizontal menu bar with various categories: Geospatial, Image, Text, NLP, Computer Vision, Deep Learning, Reinforcement Learning, Data, Machine Learning, Biology, Chemistry, Customer Service, Predictive, Signals, Computer Vision, Computer, Python, Data & Culture, Security, Fraud Detection, Location, Education, Construction, Health, Research, and Zindi. There are also buttons for 'Active & Closed' and 'Any Reward'.

CGIAR Crop Yield Prediction Challenge
Can you predict maize yields on East African farms using satellite data?
Kenya
Agriculture, Prediction, Computer Vision, Unstructured, Image, Geospatial, Zindi

Indaba Grand Challenge: Curing Leishmaniasis by Deep Learning
Indaba
The official competition of the Deep Learning Indaba 2020
Africa
Health, Reinforcement Learning, Zindi

Prizes
\$3,000 USD
353 data scientists awarded
109 on the leaderboard
ended 17 days ago

Points
3000 Zindi Points
234 data scientists awarded
38 on the leaderboard
11 days to go

Future Proofing Your Career

Data Science Competition Platforms

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

The screenshot shows the Analytics Vidhya Data Hack platform. At the top, there is a navigation bar with links to Blog, Courses, Hackathons, BlackBell, User Rankings, and Top Approaches. There are also search and social media icons. The main header reads "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 this, there are several competition cards. One card is highlighted with a blue speech bubble containing "Active". The cards include:

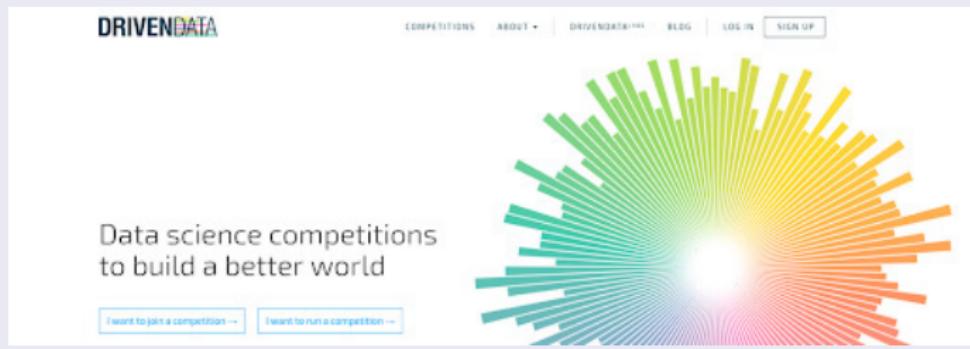
- Data Science Blogathon 7 (Active)
- Loan Prediction (Active)
- Big Mart Sales Prediction (Active)
- Black Friday Sales Prediction (Active)

Each card displays the competition name, status, and a "Participate" button. Below the cards, there is a "Feedback" button. At the bottom, there is a footer with text about assured rewards and total participants, and a "Practice Problem" section.

Future Proofing Your Career

Data Science Competition Platforms

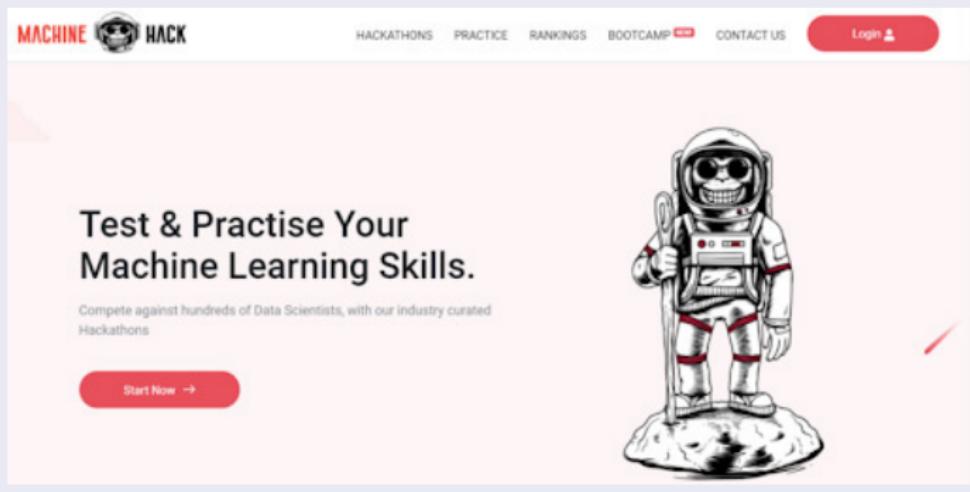
- DrivenData (Level: Beginner, Intermediate Advanced))



Future Proofing Your Career

Data Science Competition Platforms

- Machine Hack (Level: Beginner and Intermediate))

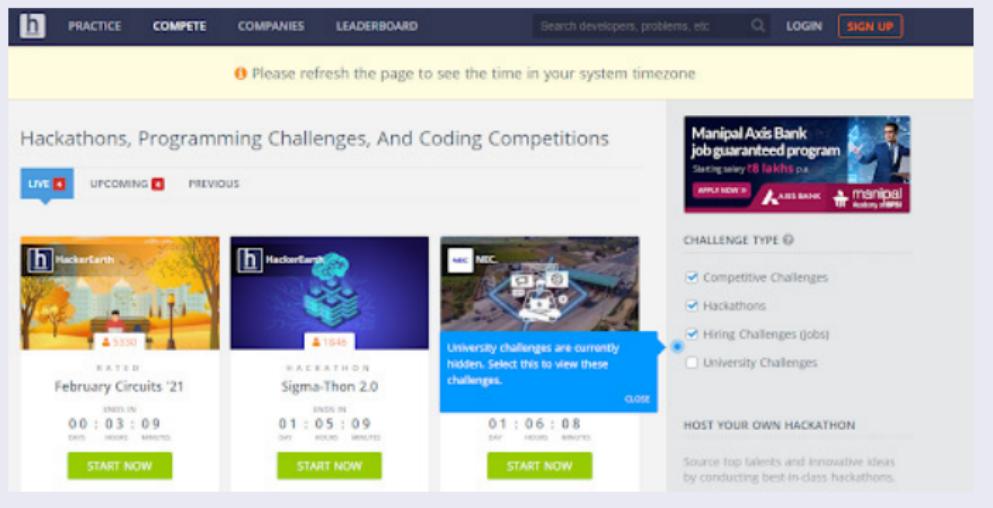


The screenshot shows the homepage of the Machine Hack website. At the top, there is a navigation bar with the logo 'MACHINE HACK' on the left and links for 'HACKATHONS', 'PRACTICE', 'RANKINGS', 'BOOTCAMP', 'CONTACT US', and 'Login' on the right. The main content area features a large, stylized illustration of a skeleton wearing a white space suit and goggles, standing on a small planet with a red rocket trail in the background. The text 'Test & Practise Your Machine Learning Skills.' is displayed prominently in the center. Below this, a subtext reads 'Compete against hundreds of Data Scientists, with our industry curated Hackathons'. At the bottom left is a red 'Start Now →' button.

Future Proofing Your Career

Data Science Competition Platforms

■ HackerEarth (Level: Beginner and Intermediate)



The screenshot shows the HackerEarth homepage with the following details:

- Header:** Practice, Compete, Companies, Leaderboard, Search bar, Login, Sign Up.
- Middle Section:** A message: "Please refresh the page to see the time in your system timezone".
- Section Title:** Hackathons, Programming Challenges, And Coding Competitions.
- Filters:** LIVE (selected), UPCOMING, PREVIOUS.
- Challenges:**
 - February Circuits '21:** RATED, ENDS IN 00 : 03 : 09, START NOW button.
 - Sigma-Then 2.0:** RATED, ENDS IN 01 : 05 : 09, START NOW button.
 - University challenges:** University challenges are currently hidden. Select this to view these challenges. ENDS IN 01 : 06 : 08, START NOW button.
- Advertisement:** Manipal Axis Bank job guaranteed program, Starting salary ₹8 lakhs p.a.
- Challenge Type Filter:** Competitive Challenges, Hackathons, Hiring Challenges (jobs) (selected), 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))

 CodaLab

Search Competitions My Competitions Help Sign Up Sign In

Competitions

Search...

 Outdoor Semantic Segmentation Challenge Organized by peneo Develop a semantic segmentation algorithm for autonomous driving in unstructured environments and benchmark your results on the novel TA5500 dataset.	Apr 27, 2021-Aug 15, 2021 5 participants
 ADoBo — Automatic Detection of Borrowings Organized by lee Detecting emerging borrowings from English into Spanish (words like 'smartphone' or 'fake news') that appear in the Spanish press	Apr 26, 2021-Jun 01, 2021 46 participants
 Rohde & Schwarz Engineering Competition 2021 - Qualifiers Organized by ai_rockie Bring signal analysis to the next level! Write Python code that detects	Apr 26, 2021-May 14, 2021 151 participants

Join us on GitHub for contact & bug reports [About](#) [Privacy and Terms](#) 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.



ZINDI