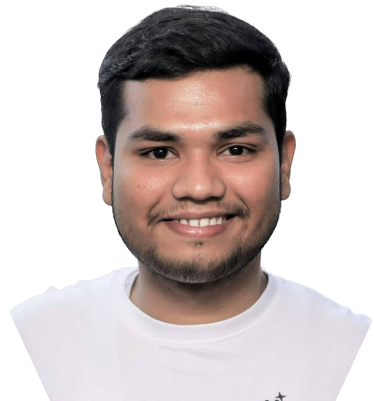


# Deciphering Competitive Data Science & Internships

**Date : 26th March 2023 | Speaker : Ayon Roy |**  
**Event : Data Science Meetup by GDG Noida & neo4j**

**Visit - [AYONROY.ML](https://AYONROY.ML)**



# Hello Buddy!

I am **Ayon Roy**

**Executive Data Scientist @ NielsenIQ**

**Z by HP Global Data Science Ambassador**

Mentored/Judged **100+** Hackathons

Delivered **60+** Technical Talks

Brought **Kaggle Days Meetup** Community in India for the 1st time

**If you haven't heard about me yet, you might have been living under the rocks. Wake up !!**

# Agenda

- What is Competitive Data Science ( CDS ) ?
- Why should you try CDS at least once?
- How should you start CDS? ( **Approach & Start** )
- Where to & How to get involved with core level CDS? ( **Launch** )
- Is CDS everything?
- The **What,Why,How** of Data Science Internships ?

# What is Competitive Data Science ?

A great opportunity to

- **Sharpen your programming & analytical skills**
- **Enhance domain knowledge**
- **Learn more about practical applications of data science & machine learning algorithms**

by participating in some real world Data Science Competitions hosted by organizations on various platforms.

**But why**  
**Competitive Data Science**  
**is gaining traction in 2023?**

It's possibly due to the



Organizations are having hard time to solve so many data science problems while their data science team is busy with other projects. So hosting a data science competition on certain platform can help & is helping.

**Data science competitions help organizations solve complex business problems while enabling data scientists to learn from the experience and win awards.**

Organizations need to define the problem, provide data and put a prize on the challenge. Competing data scientists build and present different algorithms to be the winner.



Why should you try  
**Competitive Data Science**  
at least once?

# To avoid situations like

when you have your first real-world  
adult experience after graduating



# And to

- Understand how to solve predictive modeling competitions efficiently
- Learn how to preprocess the data and generate new features
- Be taught advanced feature engineering techniques
- Be able to form reliable cross validation methodologies
- Gain experience in analyzing and interpreting the data
- Master the art of combining different machine learning models
- Get exposed to past (winning) solutions

How should you start your  
**Competitive Data Science**  
journey?

# The only thing you need to know **Before Starting** your CDS journey

“For participating in data science competitions, you only need an urge to constantly learn and improve. Getting a good ranking will follow.”

# Initial steps to start your CDS Journey

- Make sure your basics about Python & Mathematical concepts are clear enough.
- Focus on understanding core Data Science & Machine Learning algorithms
- Try to make self projects with the concepts you learned

# The next steps

- Try participating in Kudos/Knowledge Competitions ( Like Titanic etc.)
- Then try to learn about the approaches from other's notebooks
- Try to apply your learnings from those approaches in Featured/Prized Competitions
- Try exploring variety of techniques you can use to get better results

# How to approach a Competitive Data Science Problem?



1. **Start with a very simple baseline model**
2. **Understand the problem and data to create a good validation set**
3. **Try Feature engineering**
4. **Try building a variety of models**
5. **Try stacking or blending of these results using Ensembling**



Time is a very crucial factor in any data science competition.

**You should not waste your time writing the same snippets from scratch again and again in multiple competitions.** Instead, focus your valuable time on doing something new and better

# Where to get involved with Competitive Data Science ?

# My personal suggestions

- <https://www.kaggle.com/>
- <https://www.crowdanalytix.com/community>
- <https://zindi.africa/about>
- <https://datahack.analyticsvidhya.com/>
- <https://www.crowdai.org/challenges>
- <https://tianchi.aliyun.com/competition/gameList/activeList>
- <https://www.datasciencechallenge.org/>
- <https://www.drivendata.org/>

Know a few more platforms to kick start your CDS journey [here](#)

## Be a part of Communities like

1. Kaggle Days
2. Women in Machine Learning & Data Science
3. GDG, WTM
4. ODSC & a lot more....

## Be a part of as many hackathons as you can

who wants to miss networking, free food & swags alongside unlimited learning

# How to get involved more with Competitive Data Science ?

1. **Do such courses where the skills learnt in them can be used in Competitions.**
2. **Publish your competition research**, approaches on the forum & do write about the things that you want to share with others via blog etc.
3. **Participate in Discussion forums**, share your knowledge through answering questions & asking genuine questions.
4. **Make notebooks & share them along with great EDA, feature engineering etc** so that others can learn from it.
5. **Try to reproduce interesting kernels.**
6. **Be consistent** in whatever you are trying to share with the CDS community.

Is  
**Competitive Data Science**  
everything what the industry  
requires?



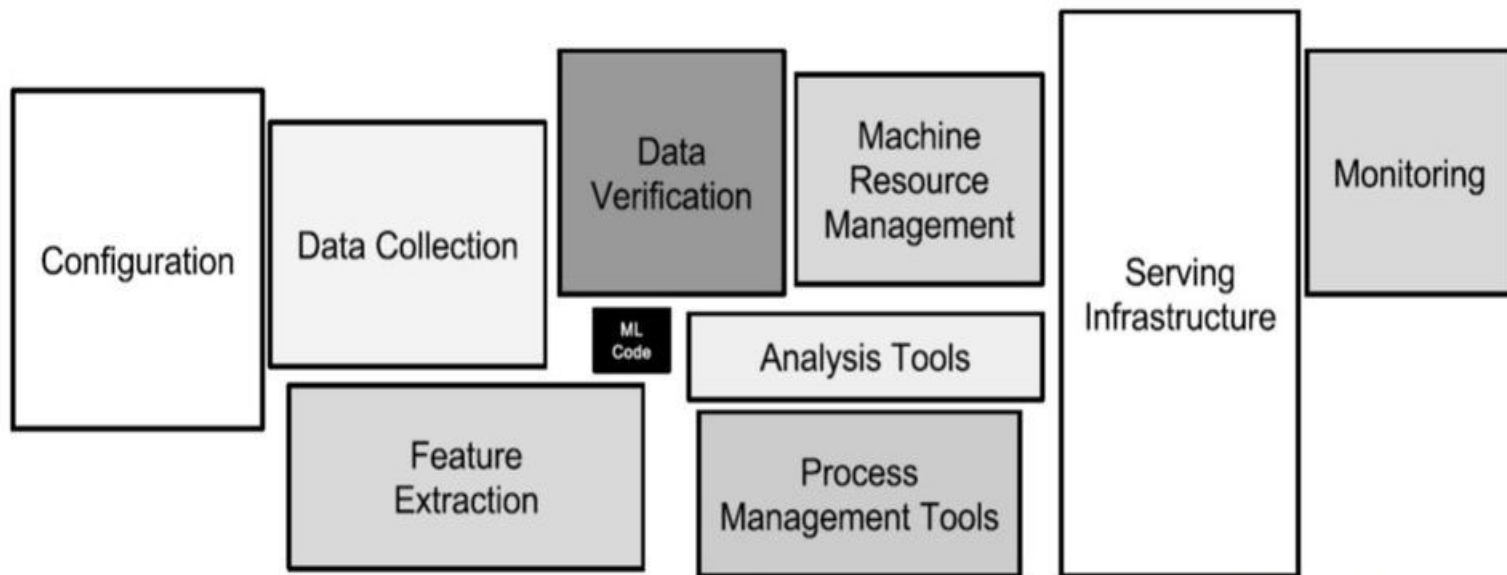


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

View the Google's Research Paper [here](#)

# A few important pointers to keep in mind

1. **Focus on understanding what business use case you are trying to solve** before applying Data Science, Machine Learning.
2. **Focus on Communication Skills to convey the result** of your Data Science concepts to the business stakeholders.
3. **Focus on DevOps** to make your models production ready.
4. **Focus on networking & showcasing your work** to the community.

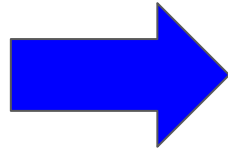
**& be prepared to rock the industry**

Take your baby steps  
with  
**Internships**

# The What of Data Science Internships

# What are the requirements of a decent Data Science Internship Opportunity ?

Let's divide the variety of Data Science Internships out there in the industry



## Requirements for **Entry Level Internships**

- Ideal for Second-Third Year Students of a 4 year Undergraduate course
- Basic Level of Python, OOPs, File Systems
- Good knowledge of Scraping, Numpy, Pandas & Data Visualization libraries
- High level overview of Machine Learning algorithms
- Few basic Data Pre-Processing & Exploratory Data Analysis Projects

## Requirements for **Middle Level Internships**

- Ideal for Third-Fourth Year Students of a 4 year Undergraduate course
- Sound knowledge of Data Science concepts & other ML algorithms with good grasp on statistics & concept of maths, SQL
- Good knowledge of Deep Learning concepts, DBMS, API development
- Self projects using ML algorithms

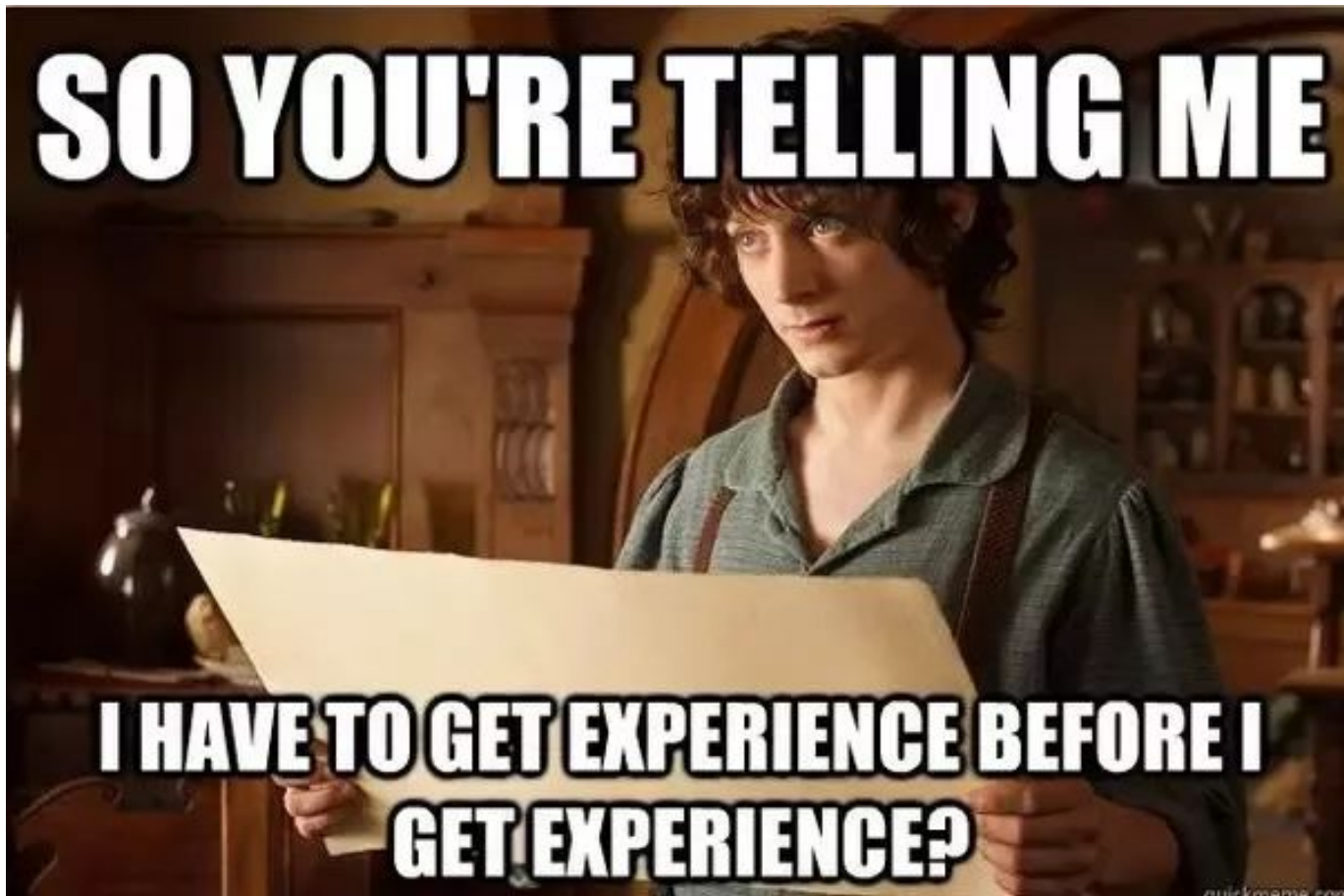
## Requirements for **Advanced Level Internships**

- Ideal for Final Year Students of a 4 year Undergraduate course
- Specialised domain of expertise like Computer Vision, Natural Language Processing etc.
- Proficient with whiteboarding of ML algos along with explanation of the basics
- Basic knowledge of Docker, Cloud
- 4-5 very good projects using the complex Deep Learning concepts like LSTMs, Transformers etc.

**Best way to get a Pre-Placement Offer before campus hiring starts**

# The Why of Data Science Internships





Why should students do **Data Science Internships** ?

- **To learn dealing with Messy, unstructured, incomplete data** ( This is real industry )
- **To experiment & learn new things** as an Intern so that you can save time as a Full Timer excluding the mistakes you did as an intern.
- **To understand how end to end real world Data Science applications works.**
- **To network with people** who look like Future of You.
- **To work under pressure & learn how to deliver** in tough deadlines too.

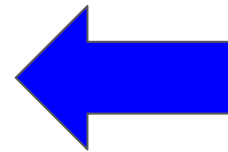
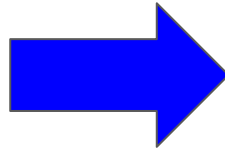
**To make like minded friends & throw a party at Starbucks**

# The How of Data Science Internships

# How to apply for Internships ?

- LinkedIn Jobs [ Worked for Me ]
- Angel.co [ Worked for Me ]
- Internshala [ Worked for Me ]
- Through Organization's Careers Page [ Worked for Me ]
- Commenting Interested on someone's LinkedIn Posts [ Not Worked for Me ]
- Career/ Internship Fairs [ Never Tried ]
- Via Winning Hackathons [ Worked for Me ]
- Asking for Referrals [ Worked for Me ]
- Community Events [ Have seen it work ]

My First  
Approach



My Second  
Approach

Visit - [AYONROY.ML](http://AYONROY.ML)

**WARNING**  
for  
of Data Science Internships

# Pay for Trainings, not Internships







# A few useful resources

1. <https://towardsdatascience.com/use-kaggle-to-start-and-guide-your-ml-data-science-journey-f09154baba35>
2. <https://www.coursera.org/learn/competitive-data-science#syllabus>
3. <https://towardsdatascience.com/how-to-successfully-manage-a-data-science-delivery-pipeline-33bdec1a9a27>
4. <http://kaggle.com/learn>

**Let me answer your Questions now.**

**Finally, it's your time to speak.**



# Danke Schoen

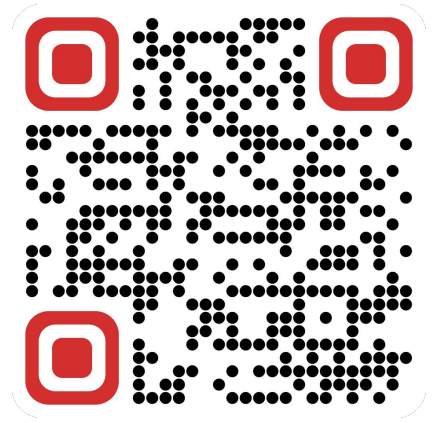
Questions ? Any Feedbacks ? Did you like the talk?  
Tell me about it.

If you think I can help you,  
connect with me via

Email : [ayon-roy@outlook.com](mailto:ayon-roy@outlook.com)

LinkedIn : <https://www.linkedin.com/in/ayon-roy>

Website : <https://AYONROY.ML/>



Download the slides