The Beginner's Guide to Kaggle

July 13, 2017

Share Google Linkedin Tweet

Kaggle, a popular platform for data science competitions, can be intimidating for beginners to get into.

After all, some of the listed competitions have over $1,000,000 prize pools and hundreds of competitors.

Top teams boast decades of combined experience, tackling ambitious problems such as improving airport security or analyzing satellite data.

Recommended Reading

The Beginner's Guide to Kaggle

How to Handle Imbalanced Classes in Machine Learning

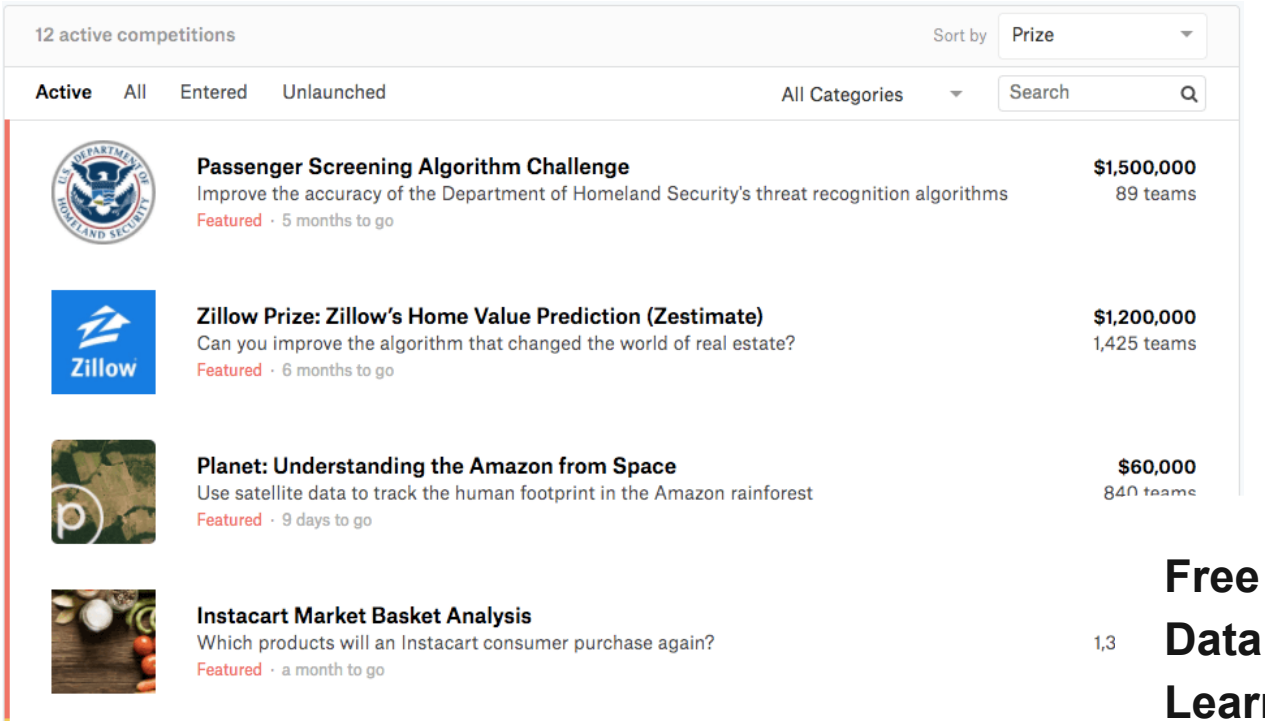
9 Mistakes to Avoid When Starting Your Career in Data Science

WTF is the Bias-Variance Tradeoff? (Infographic)

Free Data Science Resources for Beginners

Dimensionality Reduction Algorithms: Strengths and Weaknesses

Modern Machine Learning Algorithms: Strengths and Weaknesses



It's no surprise that some beginners hesitate to get started on Kaggle. They have reasonable concerns such as:

- How do I even start?
- Will I be up against teams of experienced Ph.D researchers?
- Is it worth competing if I don't have a realistic chance of winning?
- Is this what data science is all about? (If I don't do well on Kaggle, do I have data science?)
- How can I improve my rank in the future?

Well, if you've ever had any of those questions, you're in the right place.

In this guide, we'll break down everything you need to know about getting started, improving your skills, and enjoying your time on Kaggle.

Free 7-Day Crash Course: Data Science & Machine Learning!

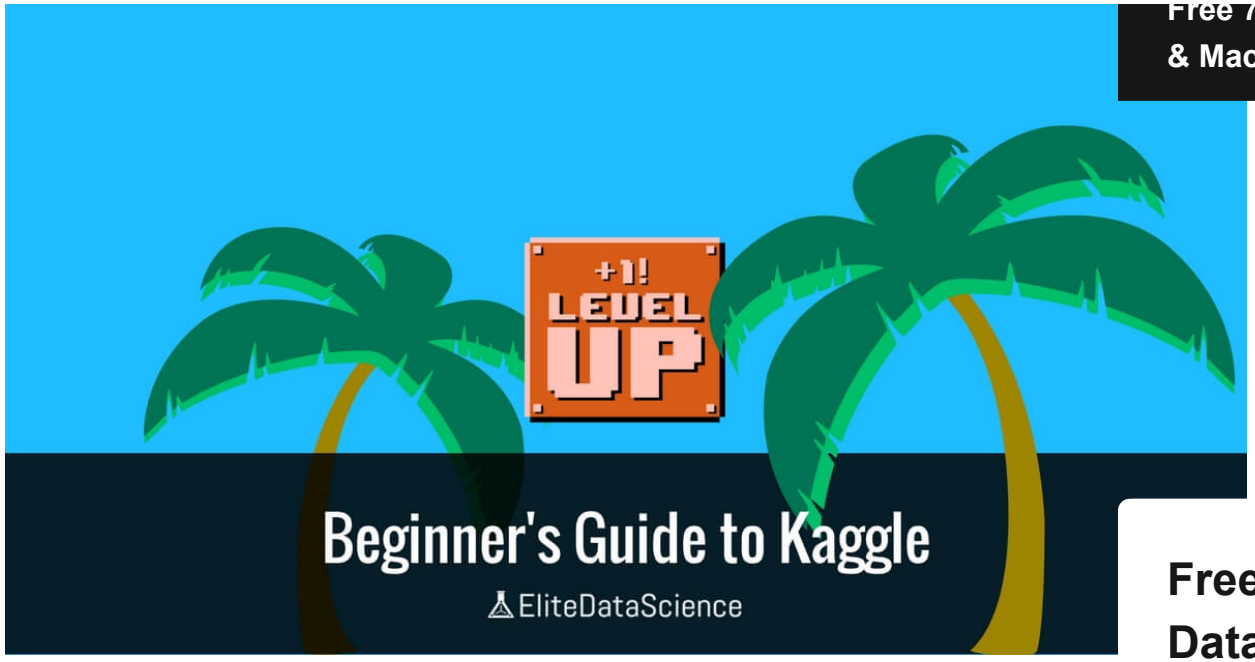
Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science & Machine Learning!



Kaggle vs. "Typical" Data Science

First, we need to make something very clear:

> *Kaggle competitions have important differences from "typical" data science, they still provide valuable experience if you approach them with the right mir*

Let us explain:

Kaggle competitions

By nature, competitions (with prize pools) must meet several criteria.

- **Problems must be difficult.** Competitions shouldn't be solvable in a single afternoon. To get the best return on investment, host companies will submit their

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science

Free 7-Day Crash Course: Data Science & Machine Learning!

biggest, hairiest problems.

- **Solutions must be new.** To win the latest competitions, you'll usually need t perform extended research, customize algorithms, train advanced models, etc.
- **Performance must be relative.** Competitions must crown a winner, so your solution will be scored against others'.

"Typical" data science

In contrast, day-to-day data science doesn't need to meet those same criteria.

- **Problems *can* be easy.** In fact, data scientists should try to identify low-hanging fruit: impactful projects that can be solved quickly.
- **Solutions *can* be mature.** Most common tasks (e.g. exploratory analysis, da cleaning, A/B testing, classic algorithms) already have proven frameworks. There's need to reinvent the wheel.
- **Performance *can* be absolute.** A solution can be very valuable even if it simply beats a previous benchmark.

Kaggle competitions encourage you to squeeze out every last drop of performar while typical data science encourages efficiency and maximizing business impac

So is Kaggle worth it?

Despite the differences between Kaggle and typical data science, Kaggle can sti great learning tool for beginners.

- **Each competition is self-contained.** You don't need to scope your own pro collect data, which frees you up to focus on other skills.
- **Practice is practice.** The best way to learn data science is to learn by doing. As long as you don't stress out about winning every competition, you can still practice interesting problems.

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

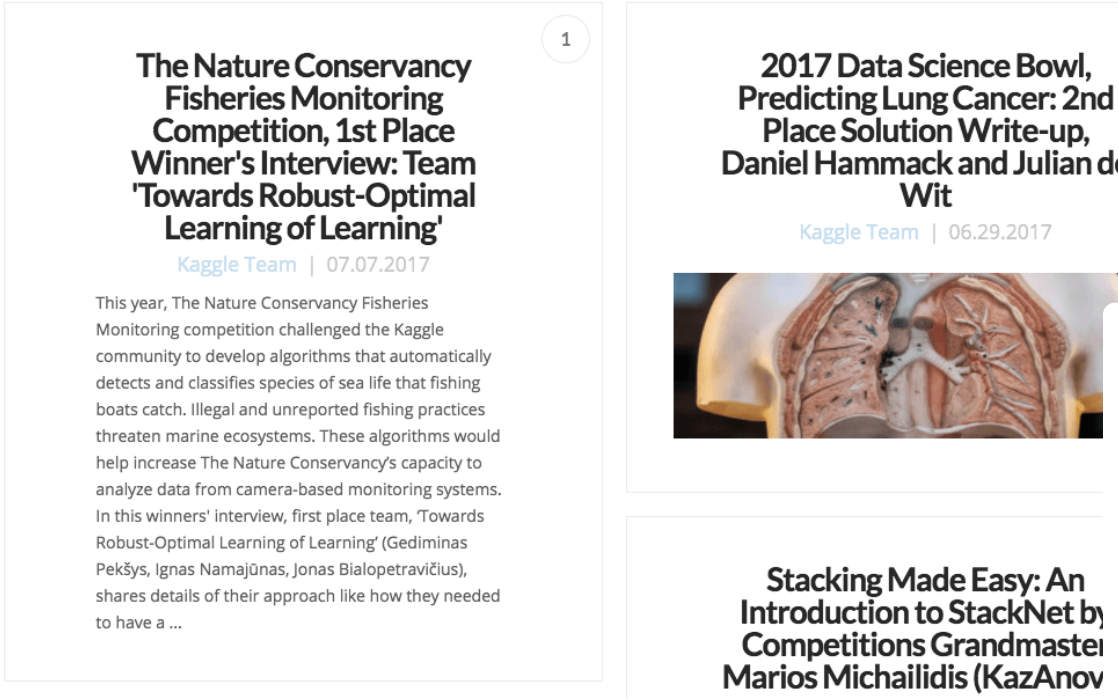
Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science

- **The discussions and winner interviews are enlightening.** Each competiti
its own discussion board and debriefs with the winners. You can peek into th
thought-processes of more experienced data scientists.

Free 7-Day Crash Course: Data Science & Machine Learning!



Kaggle winner interviews

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

How to Get Started on Kaggle

Next, we'll give you a step-by-step action plan for gently ramping up and compet
on Kaggle.

Step 1: Pick a programming language.

First, we recommend picking one programming language and sticking with it. Both Python and R are popular on Kaggle and in the broader data science community.

Free 7-Day Crash Course: Data Science

Free 7-Day Crash Course: Data Science & Machine Learning!

If you're starting with a blank slate, we recommend **Python** because it's a gener purpose programming language that you can use from end-to-end.

- R vs Python for Data Science
- How to Learn Python for Data Science

Step 2: Learn the basics of exploring data.

The ability to load, navigate, and plot your data (i.e. exploratory analysis) is the first step in data science because it informs the various decisions you'll make throughout model training.

If you go the route of Python, then we recommend the Seaborn library, which wa designed specifically for this purpose. It has high-level functions for plotting man most common and useful charts.

- Python Seaborn Tutorial

Step 3: Train your first machine learning model.

Before jumping into Kaggle, we recommend training a model on an easier, more manageable dataset. This will allow you to become familiar with machine learnin libraries and the lay of the land.

The key is to start developing good habits, such as splitting your dataset into se training and testing sets, cross-validating to avoid overfitting, and using proper performance metrics.

For Python, the best general-purpose machine learning library is Scikit-Learn.

- Python Scikit-Learn Tutorial
- 7-Day Applied Machine Learning Crash Course

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science & Machine Learning!

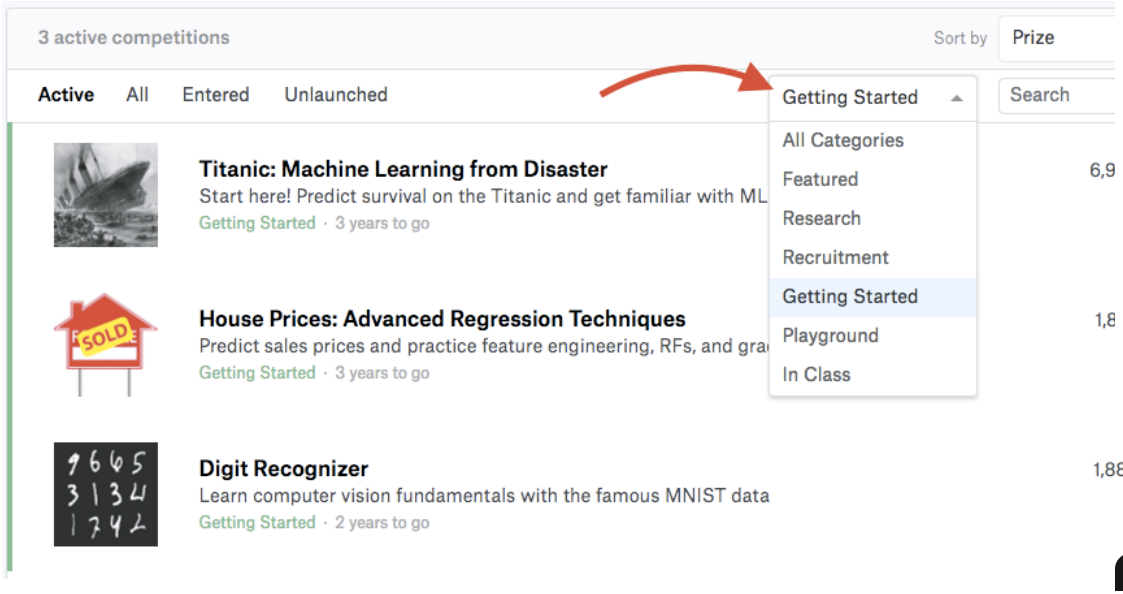
Free 7-Day Crash Course: Data Science & Machine Learning!

Step 4: Tackle the 'Getting Started' competitions.

Now we're ready to try Kaggle competitions, which fall into several categories. The most common ones are:

- **Featured** - These are usually sponsored by companies, organizations, or even governments. They have the largest prize pools.
- **Research** - These are research-oriented and have little to no prize money. They also have non-traditional submission processes.
- **Recruitment** - These are sponsored by companies who want to hire data scientists. These are still relatively uncommon.
- **Getting Started** - These are structured like featured competitions, but they h prize pools. They feature easier datasets, plenty of tutorials, and rolling subm windows so you can enter them at any time.

The 'Getting Started' competitions are great for beginners because they give you stakes environment to learn, and they are also supported by many community-created tutorials.



Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science & Machine Learning!

Step 5: Compete to maximize learnings, not earnings.

With that foundation laid, it's time to progress to 'Featured' competitions. In general, these will require much more time and effort to rank well.

For that reason, we recommend picking your battles wisely. Enter competitions that will expose you to techniques and technologies that align with your long-term goals.

While prize money is nice, the more valuable (and reliable) reward will be the skills you'll develop for your career.

Tips for Enjoying Kaggle

Finally, we'll cover our 7 favorite tips for making the most out of your time on Kag

Tip #1: Set incremental goals.

If you've ever played an addicting video game, you'll know the power of increme goals. That's how great games get you hooked. Each goal is big enough for a se accomplishment, yet realistic enough to be within reach.

Most Kaggle participants will never win a single competition, and that's completely fine. If you set that as your very first milestone, you may feel discouraged and lose motivation after a few tries.

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science & Machine Learning!

Incremental targets make the journey more enjoyable. For example:

- Make a submission that beats the benchmark solution.
- Score in the top 50% in one competition.
- Score in the top 25% in one competition.
- Score in the top 25% in three competitions.
- Score in the top 10% in one competition.
- Win a competition!

This strategy will allow you to measure your progress and improvement along the way.

Tip #2: Review most voted kernels.

Kaggle has a cool feature in which participants can submit "kernels," which are s scripts that explore a concept, showcase a technique, or even share a solution.

When you start a competition or when you hit a plateau, reviewing popular kerne spark more ideas.

Free 7-Day Crash Course: Data Science & Machine Learning!

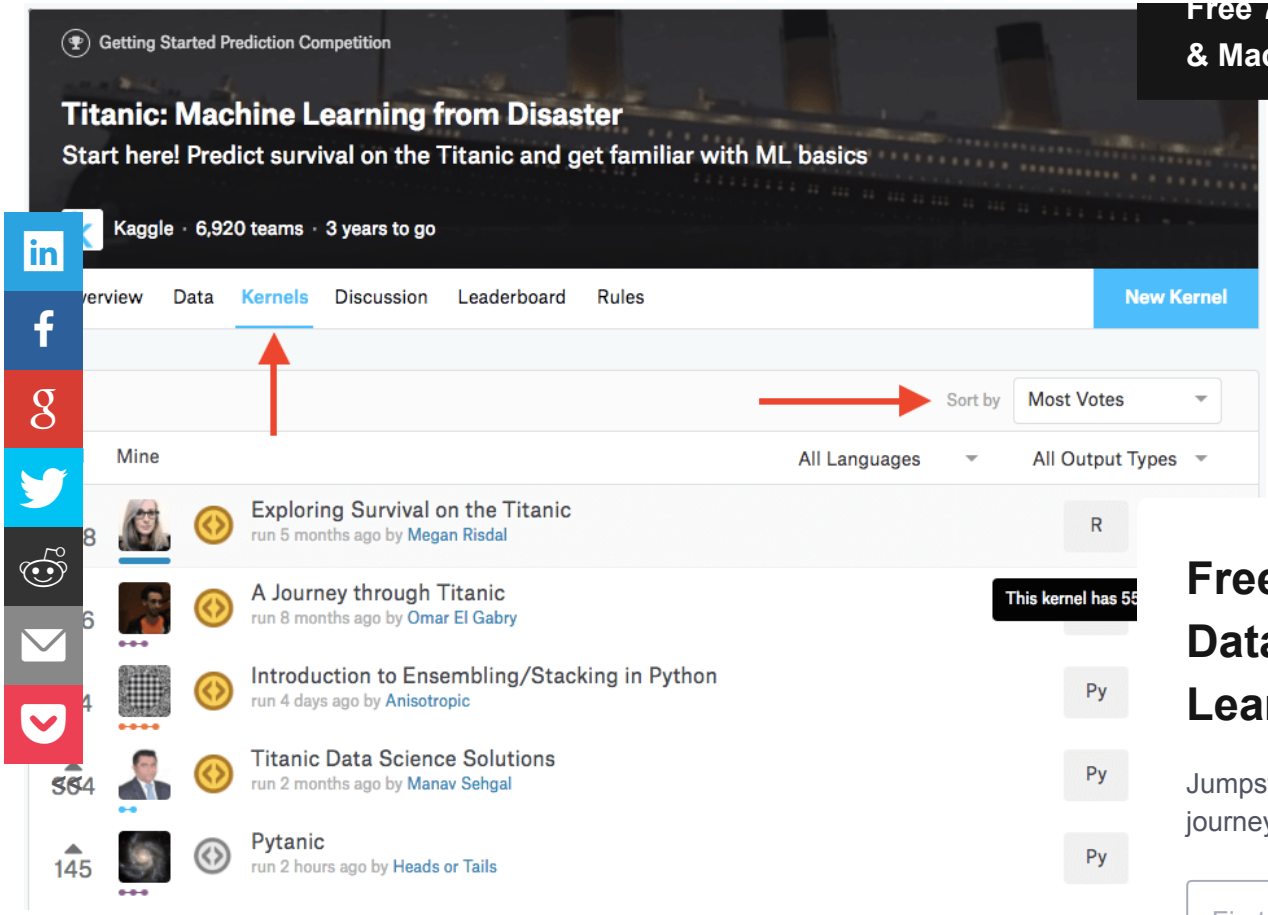
Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science



Free 7-Day Crash Course: Data Science & Machine Learning!

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Tip #3: Ask questions on the forums.

Don't be afraid to ask "stupid" questions.

After all, what's the worst thing that could happen? Maybe you get ignored... anc all.

On the other hand, you have plenty to gain, including advice and coaching from more experienced data scientists.

Tip #4: Work solo to develop core skills.

Free 7-Day Crash Course: Data Science

In the beginning, we recommend working alone. This will force you to tackle eve
of the applied machine learning process, including exploratory analysis, data cle
feature engineering, and model training.

If you start teaming up too early, you could miss opportunities to develop those cornerstone skills.

Tip #5: Team up to push your boundaries.

With that said, teaming up in future competitions can be a great way to push your boundaries and learn from others. Many past winners have been teams who joined forces to combine their knowledge.

In addition, once you master the technical skills of machine learning, you can co
with others who may have more domain knowledge than you do, further expandi
opportunities.

Tip #6: Remember that Kaggle can be a stepping stone.

Remember, you're not necessarily committing to be a long-term Kaggler. If you f
that you dislike the format, then it's no big deal.

In fact, many people use Kaggle as a stepping stone before moving onto their ov
projects or becoming full-time data scientists.

This is another reason to focus on learning as much as you can. For the long ru
better to target competitions that will give you relevant experience than to chase
biggest prize pools.

Tip #7: Don't worry about low ranks.

Some beginners never start because they're worried about low ranks showing up in their profile. Of course, competition anxiety is a real phenomenon, and it isn't limited t

Free 7-Day Crash Course: Data Science & Machine Learning!

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science

Free 7-Day Crash Course: Data Science & Machine Learning!

Kaggle.

However, low ranks are really not a big deal. No one else will judge you because they were all beginners at one point.

Even so, if you're still really worried about low rankings in your profile, you could create a separate practice account for learning the ropes. Once you feel comfort you can start using your "main account" to build your trophy case.

(Again, this is totally unnecessary!)

Conclusion

In this guide, we shared the 5 steps for getting started on Kaggle:

1. Pick a programming language.
2. Learn the basics of exploring data.
3. Train your first machine learning model.
4. Tackle the 'Getting Started' competitions.
5. Compete to maximize learnings, not earnings.

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science

Finally, we shared our 7 favorite tips for enjoying your time on the platform:

- Set incremental goals.
- Review most voted kernels.
- Asks questions on the forums.
- Work solo to develop core skills.
- Team up to push your boundaries.
- Remember that Kaggle can be a stepping stone.
- Don't worry about low ranks.

If you enjoyed this guide, then we invite you to join our community so we can giv heads up when we publish more guides. Plus, we'll send you a free 7-day email course on data science (with lessons not found on our blog).

Share Google Linkedin Tw

Leave A Response

*Name**

*Email**

Website

Free 7-Day Crash Course: Data Science & Machine Learning!

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science & Machine Learning!

Submit Comment

* Denotes Required Field

Copyright © 2017 · EliteDataScience.com · All Rights Reserved

Policy

Free 7-Day Crash Course: Data Science & Machine Learning!

Jumpstart your data science and machine learning journey with our practical, super intuitive course.

First Name

Email Address

Sign Up Now

Free 7-Day Crash Course: Data Science