**18 Data Science Project Ideas for Beginners**

***Pick one or all of them — whatever looks like the most fun to you.***

Data science projects are a great way for beginners to get to grips with some of the very basic data science skills and languages that you’ll need to pursue data science as a hobby or a career. Tutorials, lessons, and videos are all great, but projects really act as a stepping stone to getting involved with data science and getting your hands dirty.

Data science projects for beginners are better for learning languages and skills because they’re stickier. I watched videos about learning Python about 10,000 times, but I only really started to understand Python when I took a project and did it myself. Data science projects are great because you’ve got much more personal vested interest than just watching an online tutorial. You’re motivated to see something through when you have a stake in the matter.

A good project can be anything from learning how to import a dataset all the way to creating your own website or something even more complex. Projects can be personal, just help you learn; they can serve as a portfolio to prove you know what you’re talking about.

This article will offer 18 data science project ideas for beginners. Pick one or all of them — whatever looks like the most fun to you. Let’s jump in.

**6** [**Data Science Project Tutorials for Beginners**](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#b0fb)  
· [Project 1: House Prices Regression](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#51cf)  
· [Project 2: Titanic Classification](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#87de)  
· [Project 3: Deep Learning Number Recognition](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#ac66)  
· [Project 4: YouTube comment sentiment analysis](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#959e)  
· [Project 5: Covid-19 Data Analysis Project](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#2a76)  
· [Project 6: YouTube APIs with Python](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#bc89)  
[**5 DIY Data Science Project Ideas for Beginners [Unlimited Data Science Project Ideas]**](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#db96)  
· [Project 7: Tidy Tuesdays](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#7b31)  
· [Project 8: The Pudding](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#4a34)  
· [Project 9: 538](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#f188)  
· [Project 10: NASA](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#d7e9)  
· [Project 11: The Tate museum](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#b74c)  
[**7 Skills-Based Data Science Projects**](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#fc18)  
· [Project 12: Collect data](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#9a82)  
· [Project 13: Clean data](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#9f81)  
· [Project 14: Explore data](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#5c9f)  
· [Project 15: Visualize data](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#5a8f)  
· [Project 16: Regression](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#ea46)  
· [Project 17: Statistics in general](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#5e74)  
· [Project 18: Machine learning](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#c990)  
[**Data science project ideas for beginners are unlimited**](https://towardsdatascience.com/19-data-science-project-ideas-for-beginners-3de3db9bf830#a514)

These seven data science projects are a mix of videos and articles. They cover various different languages depending on what you’re interested in learning. You’ll learn how to use APIs, how to run predictions, touch on deep learning, and look at regression.

These project tutorials for beginners are hands-on and specific, so they’re perfect if you want to get started but you don’t really know where. Pick one you like, see where you’re struggling, and use that to start building a list of other [data science skills](https://www.stratascratch.com/blog/most-in-demand-data-science-technical-skills/) you can learn.

**Project 1: House Prices Regression**

During the pandemic, I found myself spending a lot of time on Zillow. I loved looking at all the different houses because they were so rich in data. There are so many different aspects for me to investigate and lose myself in. That strange interest led me to [this tutorial](https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview) which allows you to predict the final price of homes in Ames, Iowa.

Sounds weird, but it’s fun.

You can use either R or Python to run through this project. Honestly, it’s an ambitious project especially if you’re brand-new to coding. But I’m starting with it because I think it speaks to a question a lot of people have — how much are houses worth? Humans are fundamentally curious, and the best data science projects exploit that curiosity to teach you skills.

What I love about this tutorial on Kaggle is that it has a ton of different options to complete it, and these different solutions are shared with the community. Anybody can upload their own code to this, so it’s a really good place to learn and copy from other people (which is really one of the best ways of learning how to code).

Get stuck in with predictions, a bit of machine learning, and some regression.

**Project 2: Titanic Classification**

One of the world’s best-known tragedies is the sinking of the Titanic. There weren’t enough lifeboats for everyone on board causing the death of over 1,500 people. If you look at the data though, it seems that some groups of people were more likely to survive than others.

The same website as in the project above, Kaggle, runs [this competition.](https://www.kaggle.com/c/titanic) They tried to figure out what factors were most likely to lead to success — socio-economic status, age, gender, and more. Similar to the house prices project, this project has access to the code of many other programmers that you can learn from. They also have their [very own tutorial](https://www.kaggle.com/alexisbcook/titanic-tutorial) they offer for total beginners. This is really useful for people who are new to Kaggle as well as coding.

At the end, you’ll have built a predictive model that answers that question. I recommend Python for this one.

Whether or not you actually join the competition, this is still one of the great data science projects for beginners to investigate.

**Project 3: Deep Learning Number Recognition**

Did you know computers can see? A lot of the latest interesting data science projects have to do with computer vision. This tutorial is great to teach you the basics of neural networks and classification methods. During the tutorial, your job is to correctly identify digits from a data set of tens of thousands of handwritten images.

This [competition/tutorial](https://www.kaggle.com/c/digit-recognizer/overview/tutorial) is also hosted by Kaggle — you can check out some of their own tutorials, or you can just get stuck in with user-submitted [code](https://www.kaggle.com/c/digit-recognizer/code).

In my opinion, this project isn’t as interesting as the Titanic or the house prices tutorial, but it’ll teach you some of the basics of a very complex subject. Plus, it’s pretty wild that you can teach a computer to see.

**Project 4: YouTube comment sentiment analysis**

Don’t read the comments! …Unless you’re doing a YouTube comment sentiment analysis data science project for beginners.

This tutorial of YouTube comment sentiment analysis is great because it’s truly for beginners. The creator of the [video tutorial](https://www.youtube.com/watch?v=kHOVWiZKpHM) is a beginner at natural language processing, which is the skill you’ll be learning in this tutorial. It’s a really cool video that’s about 14 minutes long, perfect for getting started with NLP. It’s also a great representation that shows how data science projects can run away with you, in a good way.

The video is really funny, and she links to the code in her [GitHub](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbW8yV1RWd2Ewdkxwd0NGdnRGM2FrNExpUG5vUXxBQ3Jtc0tsNFRYckhrZUVkUFhfU0xjeUQzRWw4ZHU2OVEwbVV0NG1nN3NzUUZnOFh2NV9EYV8yel93QndlSXd3d3lRcVBHdGxyampfZ0FxU25wWTZqLWRLSE1FWUpWYktKcTFYLURwME5ISFZFMldwVFJsYUhjcw&q=https%3A%2F%2Fgithub.com%2Fhellotinah%2Fyoutube_sentiment_analysis%2F). Feel free to get into it yourself!

**Project 5: Covid-19 Data Analysis Project**

During the pandemic, it felt like things were out of my control. It sounds silly, but one of the ways I grounded myself was just by keeping track of daily numbers. Sometimes it stressed me out, but I found myself looking to data as a way to understand the unimaginable.

The Python Programmer channel had a similar idea. In this tutorial, he teaches you to do Covid-19 data analysis using Python.

This [video tutorial](https://www.youtube.com/watch?v=MYU9W34dZh0) is a bit more serious than the previous one, and it goes a little bit more in-depth about how it’s done. He also covers the basics of some pretty key Python packages like pandas. It’s a really clear introduction to pandas and Python.

**Project 6: YouTube APIs with Python**

Speaking of APIs, working with APIs is a necessary skill set for all of the other scientists. When you’re choosing a project, make sure at least one of them teaches you to work with APIs to ensure you’ve covered this critical skill.

[This tutorial](https://www.youtube.com/watch?v=fklHBWow8vE) uses Python to walk you through making an API call to collect video statistics from a channel, and saving it as a pandas dataframe. It also offers you the Python notebook code and additional resources on GitHub.

**5 DIY Data Science Project Ideas for Beginners [Unlimited Data Science Project Ideas]**

There are practically millions of potential data science projects out there that I’ve been documented in tutorial and video form. But it’s also useful to know how to create your own project. Every other project tutorial out there we’ll talk about what other people want to do with — think about what you want to do.

Coming up with my own project was how I ended up getting into Python in the first place. I had a question, I needed an answer, the only way to get it was by analyzing my data with python. Rather than list more individual tutorials, I want to point you to some resources that can help you design your own data science projects from scratch.

**Project 7: Tidy Tuesdays**

This project relies on the [Tidy Tuesday GitHub repo](https://github.com/rfordatascience/tidytuesday). The great thing about this repo is that every Tuesday, brand-new untidy data is uploaded. The cohort analyzes it, visualizes it, and generally plays around with it. It’s a great place to learn from other people and experiment with it yourself.

This repo is best for people who want to learn R (though also good for some Python). It’s also best for basic data science skills, like reading files, doing introductory analysis, visualization, and reporting.

For example, this week’s Tidy Tuesday dataset was from the National Bureau of Economic Research. The way the dataset was structured meant that it was good to learn how to join tables. Maybe you’re interested in checking out the female representation of paper authors. Maybe you want to know about publishing frequency in summer versus winter. Either way, TidyTuesday can help you with some basic data science skills with new data every week. It goes back years, too, so you’ll be able to find something interesting no matter what kind of data you like, and you’ll never run out of data science project ideas.

**Project 8: The Pudding**

The Pudding does really jazzy visualizations and analysis, usually using JavaScript, Python, or R. TidyTuesday is great for sheer volume, but The Pudding offers some truly weird projects to take on.

Maybe you also are a huge Community fan like me, and you want to know [how many times](https://pudding.cool/2021/01/cool/) Abed says the word “Cool,” versus Jeff or Annie. Perhaps you love reading Agony Aunt letters and [this insight](https://pudding.cool/2018/11/dearabby/) into thirty years of American anxieties via Dear Abby letters intrigues you.

These projects offer a lot of cultural commentaries. They’re more challenging and niche than some others on this list, but they’re gripping and can teach you a lot about visualizations especially. The Pudding offers all their code on their [GitHub repo](https://github.com/the-pudding) which I encourage you to check out.

**Project 9: 538**

Sports and politics collide in the 538 blog, meeting in a glorious burst of statistics and math. Here, you can scroll through the articles, spot whatever interests you, and head to [the GitHub repo](https://github.com/fivethirtyeight/data/blob/master/index.csv) to see the code and analysis behind the findings. From there, you can dive into the data yourself.

One project I had a fun time digging into was Superbowl ads. The original article talked about how Americans love America, animals, and sex (as represented by their frequency in Superbowl ads). I was interested to know whether there were more sexual ads over the years. Find your own question and dive in!

**Project 10: NASA**

Who didn’t want to be an astronaut when they grew up? Now is (kind of) your opportunity to chase that dream.

[NASA’s data](https://data.nasa.gov/browse) isn’t as user-friendly as the three options I listed above. But the quantity (and general awesomeness) of the data on offer here makes it a must for any data science project list. Instead of trying to trawl through their dense literature and databases, I recommend you start with this “ [Space Science with Python](https://medium.com/space-science-in-a-nutshell/space-science-with-python-a-data-science-tutorial-series-57ad95660056)’’ tutorial series. For example, want to know how close the asteroid 1997BQ passed by Earth in May 2020? [Now’s your chance](https://github.com/ThomasAlbin/SpaceScienceTutorial/blob/master/part6/SpaceSciencePython_part6.ipynb) to find out.

**Project 11: The Tate museum**

Maybe you’re more of an Arts & Humanities buff. Luckily for you, there’s data available for you to create your own data science project too. Look no further than the [Tate museum’s data archive](https://github.com/tategallery/collection#usage-guidelines-for-open-data). Here, you can find the metadata for over 3,500 artists.

There’s a lot you can do for yourself with that data, but in case you’re already lost wondering where to begin, the Tate helpfully lists example data science projects others have done with access to this data. For example, Florian Kraeutli did some [gorgeous and introductory exploratory analysis](http://research.kraeutli.com/index.php/2013/11/the-tate-collection-on-github/) you can check out.

**7 Skills-Based Data Science Projects**

The first section of this blog post dealt with pretty specific tutorials. The second taught you where to look to create your own data science project ideas. This final one will point you in the right direction for skills-based data science project ideas. This is the most relevant for those who are putting together a resume or thinking about applying for a [data science job](https://www.stratascratch.com/blog/how-to-get-a-data-science-job-the-ultimate-guide/).

Each of these seven steps is worth being its own data science project for beginners, but once you’re ready, you can also use these seven to create a full project for more intermediate/advanced data scientists.

**Project 12: Collect data**

The very first step in any data science project is worth being a data science project itself: collect data.

Most times, data does not arrive perfectly formed into neat tables in your computer. You have to figure out how to get it from point A to point B in order to do everything else you want.

Turn it into a project and investigate how to collect data using some of the [most popular data science languages](https://www.stratascratch.com/blog/top-5-data-science-programming-languages/), like Python and SQL.

**Project 13: Clean data**

The data is in! But it’s messy. Learning how to clean data was one of the biggest letdowns in my Master’s when I was studying bird conservation. I thought I’d be able to get data in and start analyzing right away. Sadly, there were issues: duplicates, N/As, numbers stored as text, and just about every other issue you can think of.

Some folks say cleaning data is 80% of a data scientist’s job. It’s worth knowing how to do it.

I did my project using R, so if that’s you, I recommend [this tutorial](https://www.dataquest.io/blog/load-clean-data-r-tidyverse/) to learn how to load and clean data using R. If you’re a budding Pythonista, [this tutorial](https://realpython.com/python-data-cleaning-numpy-pandas/) helped me get to grips with cleaning data with Pandas and NumPy, both very common and useful Python packages.

**Project 14: Explore data**

Once your data is in and relatively tidy, it’s time for the exciting part: explore your data. This isn’t quite to the level of visualizing or analyzing it. Usually, there’s so much data you’re looking at that it helps to get a feel for what’s actually going on before you begin creating models. Think of this project like dipping your toe in the water to gauge the temperature.

[This 2.5 hour video tutorial](https://www.youtube.com/watch?v=kLDTbavcmd0) will teach you to build an exploratory data analysis project completely from scratch. It’s lengthy, and 100% comprehensive.

**Project 15: Visualize data**

There’s a lot you can do to visualize data, and a lot of data science skill is knowing which kind of visualization best represents the idea you’re trying to communicate. That’s why simply working on data viz is a great data science project idea for beginners.

[This Kaggle tutorial](https://www.kaggle.com/alexisbcook/bar-charts-and-heatmaps) is a bit boring but will teach you some of the basics of data visualization. With that knowledge, you can go on and create your own data science visualization project — this time using data that you care about.

**Project 16: Regression**

Regression is a super important predicting tool used in all avenues of data science. It’s what helps you statistically determine the relationship between X and Y. It’s the very basics of what will become machine learning.

You can create a project that focuses on regression with any dataset that has an X and Y variable. I did this myself with my bird data, predicting whether the size of the bird influenced the survival of the bird. Pick any dataset you like and use a method like Kaggle’s red wine quality data tutorial, linked [here](https://www.kaggle.com/madhurisivalenka/basic-machine-learning-with-red-wine-quality-data).

**Project 17: Statistics in general**

It’s easy to get caught up in the hype of NLP, ML, AI, DL, and every other data science acronym. But don’t forget data science of all kinds relies on statistics and math. To get the most out of any data science project idea you may have, ensure you have grasped the fundamentals of statistics underpinning the concepts of data science.

I’m cheating a little bit by grouping all these statistical fundamentals under a single subheader, but I recommend KDNuggets’s list of [eight basic statistics concepts](https://www.kdnuggets.com/2020/06/8-basic-statistics-concepts.html). From there, find a project that focuses on each of the eight. For instance, take the Tate dataset I linked above and learn about the “central tendency” by figuring out the median paint date of the artwork.

You can use any programming language for this project. I like Python since it’s great for beginners anyway, but R, SQL, JavaScript, or any of the other coding languages can accomplish the same goal.

**Project 18: Machine learning**

Let’s wrap up this list of data science project ideas for beginners with this one: machine learning. Any data scientist worth their salt knows about machine learning and can successfully use it to predict any number of things. Use what you learned from regression and apply it here.

To create a project that will teach you machine learning, nearly any dataset will do. For example, you can use [Uber’s pickup data](https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city) and ask questions like: does Uber make congestion worse? Alternatively, [this tutorial](https://towardsdatascience.com/movie-recommendation-system-based-on-movielens-ef0df580cd0e) which guides you through making movie recommendations could be a good project to tackle. I recommend using Python due to its TensorFlow package which is built for machine learning specifically.

**Data science project ideas for beginners are unlimited**

If you have an ounce of creativity and curiosity, you can trawl the web to find the data and tutorials you need to create your very own data science projects, no matter what your interest or skill levels. This article should serve as a signpost pointing you to potential options which you can peruse at your leisure.

If you enjoy reading stories like these and want to support me as a writer, consider [signing up to become a Medium member](https://zulie.medium.com/membership). It’s $5 a month, giving you unlimited access to stories on Medium. If you [sign up using my link](https://zulie.medium.com/membership), I’ll earn a small commission.