**How to Start Your First Data Science Project When You’re a Beginner**

A 5-step guide to choosing your data science projects



Photo by [Bing Hui Yau](https://unsplash.com/@binghui?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) on [Unsplash](https://unsplash.com/s/photos/takeoff?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText" \t "_blank)

Where and how do you start Data Science Projects When You Are A Beginner?

As a data enthusiast, projects usually come in handy, whether you're starting in your data journey or want to reinvent yourself with the changing landscape of data science & analytics.

In any data-related job interview, at some point, you’ll find yourself talking about the projects you’ve worked on. But where do you even start?

There is only so much to learn and work on in data science. I started my data science journey four years back and I was a lost kid — I did not understand where to begin. Learning my way through working as a Senior Data Analyst, here is an approach to start small with data science, work your way through portfolio projects, and gather enough motivation to convert one of your projects into a research paper!

**Step 1: Start small, with the basics**

From 2018 to 2019, in a year I taught myself the basics of machine learning with understanding all over the place. As I got into graduate school, I started all over again, started small this time. I would try to work on projects in line with the model or topic I’d learn at school in classes and take real-life problem statements.

I spent a lot of time working on regression problems, data cleaning & feature engineering projects, in the beginning, to clear my basics & establish some best practices.

I’d recommend creating a process flow and target milestones that work the best for your pace, and knowledge level. For example —

1. Get comfortable with a programming language of your choice: Python, R, SQL for data querying (SQL is inevitable for data analytics)
2. Learn about data analysis, manipulation, and visualization: libraries like pandas are a treasure
3. Understand different machine learning models, use cases, and how to implement with code: scikit-learn can be a good package to start with
4. Continue to hone your knowledge of machine learning in more depth with advanced packages and libraries
5. Websites like DataCamp, Leetcode, and HackerRank can help with coding practice

**Step 2: Take an online certification for a defined approach**

Back in 2017 when I first read about Data Science, the use cases pushed me to learn about the concepts and machine learning being the hot keyword then (and even now). If you Googled Data Science in 2017, it was the same as today — tons of articles, websites, books, and videos to teach you about the tools and technology.

The internet can be overwhelming with plenty of information when trying to carve out a path on where to start in your Data Science journey. My mad scramble landed me at IBM Data Science Professional Certification where they teach you the basics of Python, data cleaning, data manipulation, and how to interpret your data modeling.

That certification opened me to my strengths, weaknesses, and areas of opportunities to expand. That is also the same course that encouraged my passion for data analytics and storytelling with data, of all.

In my opinion, when you are starting in Data Science, online courses and certifications create excellent navigation for you from the basics to a tangible outcome. And the certificate is always a bonus for your profile and resume!

**Step 3: Work through the Data Science lifecycle**

Any data science problem or a data professional works in and around the data science lifecycle to an extent. I love tying back my projects (even at work) to this lifecycle — the process is a direction on its own.

Data fuels the businesses of today. Companies use data to stay competitive and agile and as data professionals, it is quintessential you understand the business, and the value your work adds. You can consider each of the steps below as a project on its own with business understanding as to the hook.

**Business Understanding:** Ask relevant questions and define objectives for the problems to be solved

**Data Mining**: Gather and scrape the data necessary for the project

**Data Cleaning:** Fix the inconsistencies within the data, handle missing values, and treat data with principles of collinearity

**Data Exploration:** Form hypotheses about your defined problems by visually analyzing your data

**Feature Engineering:** Select important features and construct more meaningful ones using the raw data you have

**Predictive Modelling:** Train machine learning models, evaluate their performance, and use them to make predictions

**Data Visualization:** Communicate the findings with key stakeholders using plots and interactive visualizations aka dashboards

and back to business understanding….

**Step 4: Create a diverse portfolio of projects**

Now that you have an idea of the process flow, identify the categories of projects you want to work on —

**Clustering**: the type of projects where you group similar objects into sets or clusters. You showcase your ability to classify and categorize data relative to features and characteristics

**Regression:**projects that allow you to go back in time, away from messy, hard data to interpret towards a clearer and more meaningful model to predict the future

**Data cleaning:** A data professional can take up to 80% of their time cleaning the data than creating a model. Clean data is invaluable. You can get public data sets from the official websites of organizations, clean them and push them out on community channels for other enthusiasts to use. Share and grow, together!

**Web scraping:**data can never be enough. you can always create your data by parsing and extracting information available online. Creating your datasets adds to your skills of understanding the data and challenges associated with it

*(Warning: Web scraping is legal only if you scrape publicly available data on the internet. Be careful scraping personal data, intellectual property, or confidential data)*

**<< Pro-tip >>**

As you are working on these projects, make sure to push your code, create documentation and add a ReadMe file on your GitHub account (make one if you do not have it already) or if you’re ambitious, you can create a website and host all of your projects there!

**Step 5: Create visualizations & work on storytelling**

Storytelling with data is the strongest asset for a data professional

Everything in the world today can be visualized, given you have the right data to do so — from predicting occupancy at a hospital to facilities at a cricket stadium.

The only crucial part of an effective visualization is understanding your data, selecting a tool that provides capabilities for your data and problem statement, and choosing a correct chart type.

Some examples of data visualizations to consider —

1. Analyzing trends in hospitalization at a medical center
2. Impact of COVID-19 on modes of transportation
3. Capture market share of products across geography
4. Lyft/Uber rideshare trends in New York City
5. Inventory and supply chain management at retail stores

**The bottom line**

The careers of the future include robust data science, acquisition, and analytics across any and every industry — healthcare, finance, sports, retail & e-commerce, streaming, aviation, dating, marketing, weather, education, government, travel, and everything under the sun. If you find data and its landscape fascinating, now is the time to take up a project (and try not to leave it incomplete)

That’s it from my end for this blog. Thank you for reading! Let me know in the comments about your interesting projects, your journey in data, and what are looking for in 2022!