

Python Data Science Projects For Boosting Your Portfolio

As Python is a critical data science skill, develop it through these five projects and build your Python data science project portfolio along the way.



The best way to learn about Data Science and machine learning is by doing practical projects. Data Science projects will give exposure to different dimensions of this field and help you hone your skills with hands-on experience in SQL, R, or Python. It will not only be helpful in upskilling yourself and building confidence in data science but also will help you make impressive resumes.

In this article, we will look at how python data science projects will be helpful for you to get a job and how you can build your own portfolio. We will also discuss some of the [data science project ideas for beginners](#) as well as for experienced professionals.

Python as a Must-Have Skill in Data Science

Data Science has had a boom over the past several years, and the drive in the field of artificial intelligence brought on by numerous advancements will only advance it to the next stage. As more industries start to recognize the possibilities of data science, the market brings out more job roles in this field.

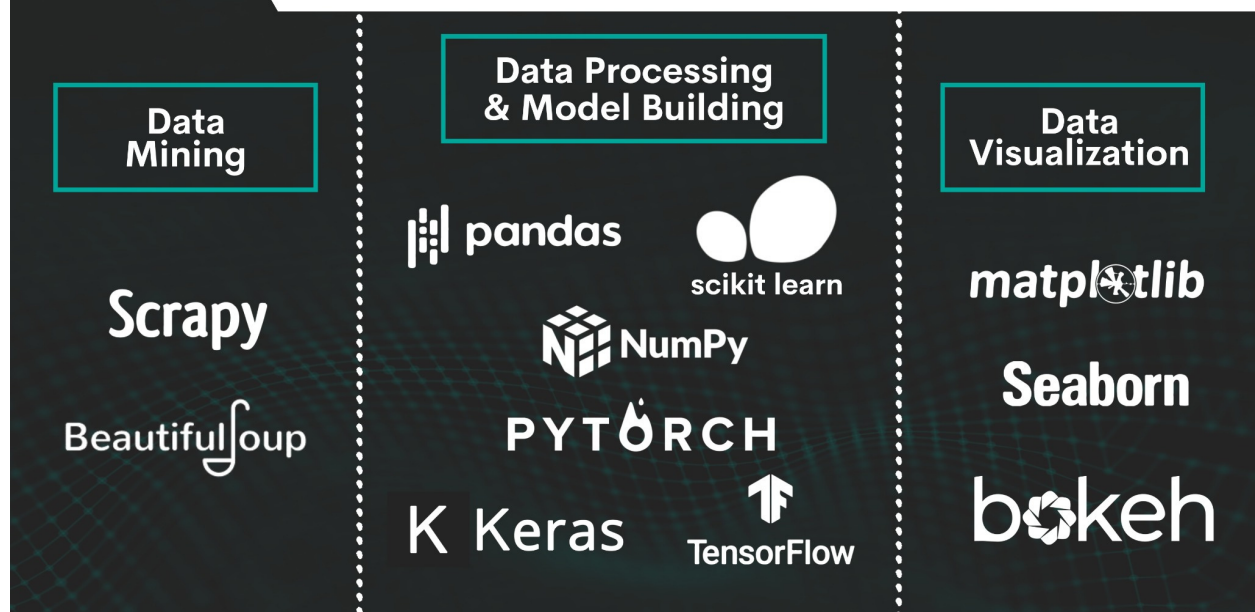
The [skills required for a Data Scientist](#) are quite wide. As a Data Scientist, you should be comfortable in working on the technical side as well as the management side. Data scientists need good technical and soft skills to succeed in this field. The Data Scientists technical skills can be categorized as below:

- Database and database design skills which uses SQL and Python
- Proficiency in working with Big Data which typically includes the use of SQL and Python
- Data Analysis and identifying patterns which includes the use of Python or R
 - Statistical analysis and hypothesis testing, which includes the use of Python or R
 - Machine Learning and model building which typically needs Python to implement and deploy models in production

As can be seen from the above list, Python is required in almost all the steps for a Data Scientist. Python is a very flexible programming language, and since it is open source, many developers have already built various packages/modules in python which are very useful for a Data Scientist.

Python is not only used by Data Scientists but also by software developers due to its ease of use and access to a lot of open source APIs. Python has open source libraries in all the areas of analytics that you can think of. It provides libraries for Data Mining, Data Processing, Model Building, and even for Data Visualization.

Python Libraries for Data Science



For Data mining, there are libraries such as [BeautifulSoup](#) and [Scrapy](#). Both of these libraries are useful if you want to scrape data from the web. They are very efficient in parsing through the HTML and XML files, and thus, the data extraction process becomes very simple. Developers have built many functions within these libraries, and thus, for Data Scientists, it becomes handy to use already existing functions and scrape the relevant data in a few lines of Code.

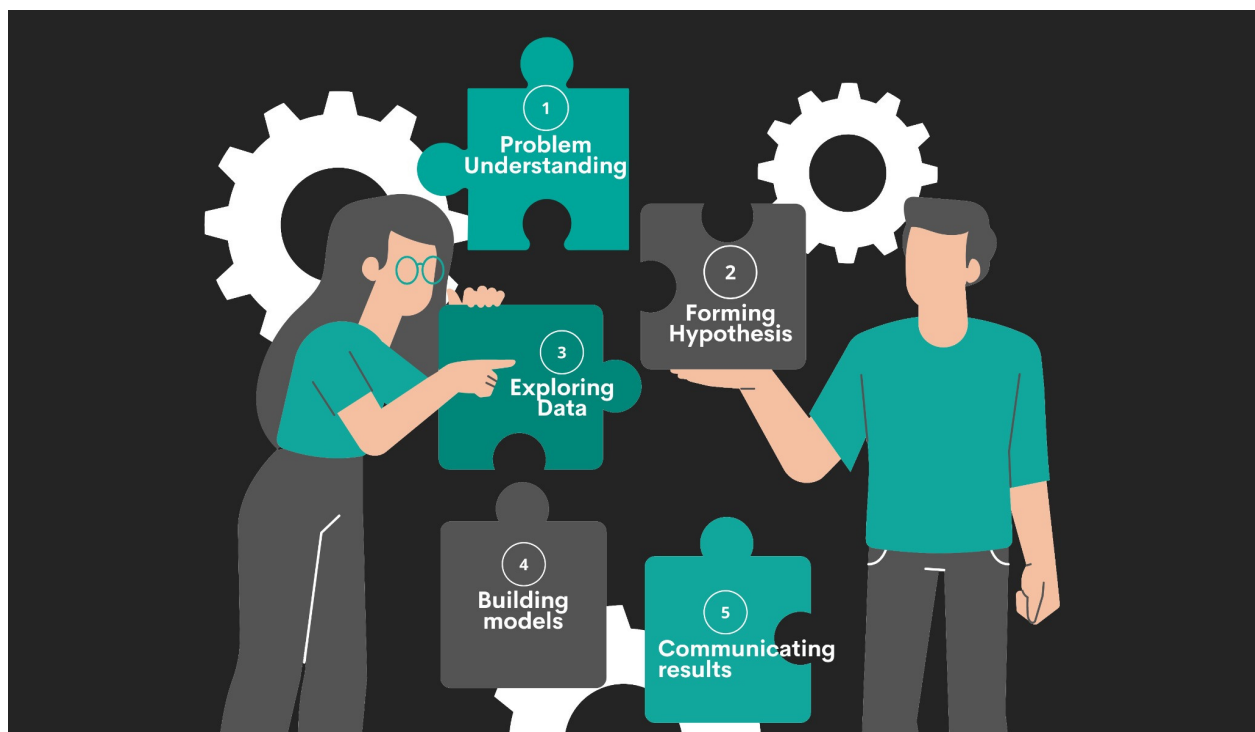
For Data processing and model building, there are many libraries available in Python. One of the most popular libraries is [pandas](#), and it provides many functions that are easy to use when analyzing data. This library has its own data types, which becomes very handy for the end users. There is another quite famous library in python for data processing which is [NumPy](#). Other libraries for the model building include [PyTorch](#), [Keras](#), [TensorFlow](#), [scikit-learn](#), etc.

For data visualization, there are many libraries available in python. These libraries help the Data Scientists to build basic visualizations, identify patterns and provide insights to the stakeholders. Libraries such as [matplotlib](#), [seaborn](#), and [bokeh](#) are the most popular ones for data visualization tasks.

Thus, Python has become one of the most popular programming languages due to its open-source nature and widely available and easy-to-use libraries.

How Python Data Science Projects Can Help Beginners Enhance Their Knowledge

The very fundamental data science skills and languages that you'll need to pursue data science as a hobby or a job can be learned through data science projects. While videos, lectures, and tutorials are all excellent resources, projects serve as a much better starting point for diving into data science and getting your hands dirty. By doing hands-on python data science projects, you will learn many skills:



- Understanding the problem and breaking it into smaller
- pieces Form hypothesis
- Explore the data
- Build models
- Communicate results

Understanding the Problem and Breaking Into Smaller Pieces

By doing the hands-on python data science projects, you will get experience by formulating the problem, understanding what is needed to be extracted from the data,

and then breaking the problem into small steps for analyzing it better. Breaking the problem down into small steps is critical, and it becomes easier to understand the next steps to carry out.

Form Hypothesis

Once you have formulated the problem statement, select one hypothesis that you think is relevant to the dataset. For example, if you want to understand why the sales are going down in the last quarter, form a hypothesis around the problem. Your hypothesis can be the sales dropped in the last quarters due to the seasonality and holiday season.

Explore the Data

By working on python projects, you will get experience in working with the dataset and exploring it from various angles. Exploratory data analysis is a very important part of any data science project. It involves a lot of techniques, but the main point is to understand your dataset better and fix any data issues you might encounter before applying machine learning.

It involves extracting the most important variables from the dataset and leaving the redundant ones out, identifying outliers, missing values, or any human errors in the data, understanding the relationship between different variables, etc. This step helps you understand the data, which is critical to solving problems.

Build Models

You'll eventually need to develop prediction models to back up your hypotheses. You'll need to write a program (code), for instance, to forecast revenue. You can investigate if and by how much an after-Christmas sale boosts profitability. Given the volume and overall profit, you can discover that some sales provide a higher profit than others.

Communicate Your Results

If you can't communicate the meaning of your analysis and technical findings to your stakeholders in a clear and compelling way, they will be of little use in the real world. You must master the crucial and undervalued art of data storytelling. You must provide a data visualization or presentation that communicates your findings to non-technical

people in order to complete your project. By working on a hands-on project, you would be able to create a data narrative from the analysis that you did in the above 4 steps.

How Python Projects Can Help Data Science Aspirants Land Their First Job

[Data analytics projects](#) will help you build a good portfolio that will be impressive for your next interview, and you'll also get practical skills that you will use at your job.

Data Science Projects in Python

There are many resources available online to get you started with Python data science projects. We will discuss some of the projects that you can pursue to build your portfolio and get a job in data science. These projects will help you to understand the applications of machine learning across different industries and give them the edge in getting hired.

Along with your education, if you have some good machine learning projects on your resume, you will have a higher chance of getting hired. Everyone interested in starting a career in data science then must have a hands-on project to show relevant experience in the interviews. Now let's look at some of the projects that you must have in your portfolio:

Project #1: DoorDash - Delivery Duration Prediction

Project: Delivery duration Prediction

Language: Python

Link: <https://platform.stratascratch.com/data-projects/delivery-duration-prediction>



There is a free project on StrataScratch about [DeliveryDuration Prediction](#). This data project has been used as a take-home assignment in the recruitment process for the data science position at DoorDash.

Description

When a consumer places an order on DoorDash, they show the expected time of delivery. It is very important for DoorDash to get this right, as it has a big impact on consumer experience. In this project, you will build a model to predict the estimated time taken for delivery. In this project, you will be given a dataset (csv file) containing a subset of deliveries received at DoorDash in early 2015 in a subset of cities. There are many features in this dataset that will be useful for predicting delivery times.

This project will be very helpful for beginners who have some knowledge of Python. This project covers all the steps we discussed in the above section regarding problem-solving. For example, you will be able to understand the problem and break it into smaller pieces. You will be able to perform exploratory data analysis, which will help you in understanding the data, identifying the patterns, identifying relevant features required to predict the delivery times.

This end-to-end project will help you gain/revise most of the skills that are required to land a job as a Data Scientist.

Project #2: AirBnb - Market Analysis Dublin

Project:Market Analysis - Dublin

Language:Python

Link: <https://platform.stratascratch.com/data-projects/market-analysis-dublin>



This data project has been used as a take-home assignment in the recruitment process for the data science positions at Airbnb.

Description

A new city manager for Airbnb has started in Dublin and wants to better understand:

- what guests are searching for in Dublin,
- which inquiries hosts tend to accept.

Based on the findings, the new city manager will try to boost the number and quality of hosts in Dublin to fit the demands from guests. The goal of this challenge is to analyze, understand, visualize, and communicate the demand/supply in the market. For example, you may want to look at the breakdown of the start date day of the week, or the number of nights or room type that is searched for, and how many hosts accepted the reservation. In particular, we are interested in:

- what are the gaps between guest demand and host supply that the new city manager could plug to increase the number of bookings in Dublin,
- what other data would be useful to have to deepen the analysis and understanding.

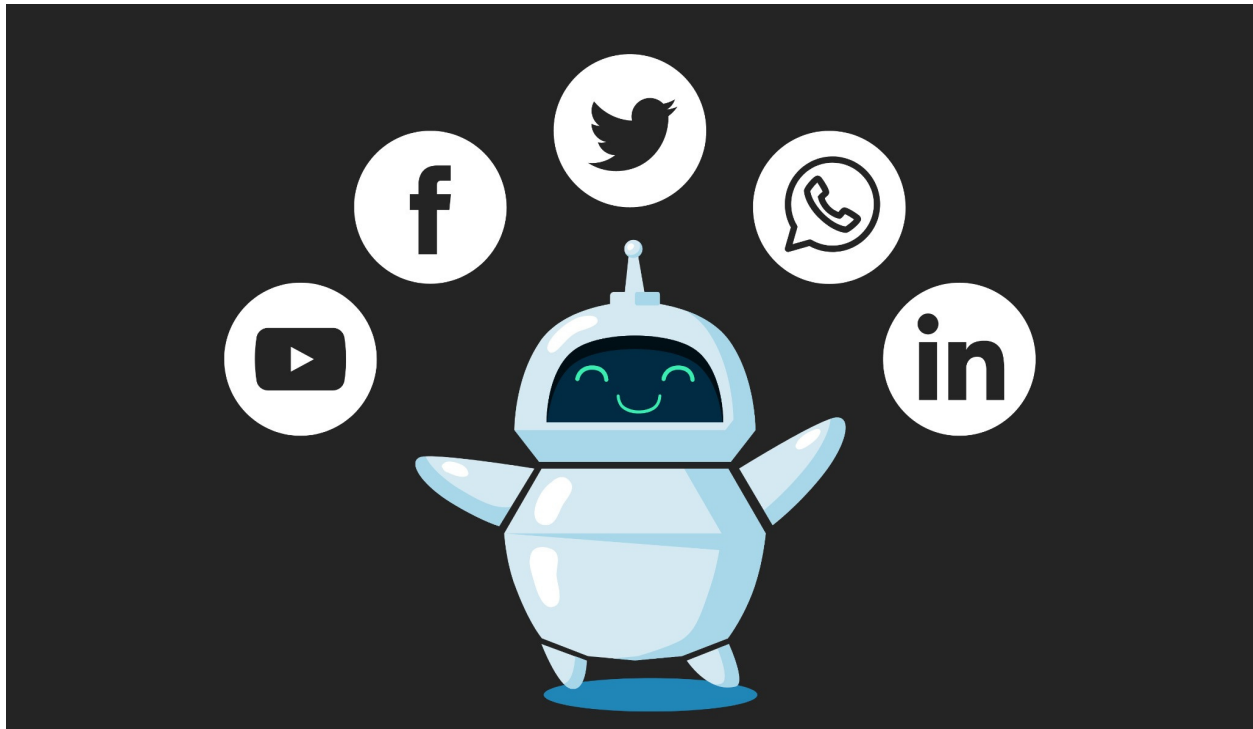
This project will also help you in gaining end-to-end experience of a data science project. In this, the end goal is not to predict anything but to find out useful insights that will be helpful for the new manager at Airbnb Dublin.

Project #3: Chatbots

Project:Build Chatbot

Language:Python

Link: <https://dzone.com/articles/python-chatbot-project-build-your-first-python-pro>



Description

Chatbots, also known as chatter-bots or conversational agents, are software programs that are usually used instead of living agents to solve customers' problems. Have you ever been to a customer support website and chatted with someone from customer service and realized that, in fact, you are chatting with a "robot"? Then you know what chatbots are!

Visitors can access chatbots mostly through web-based apps or a standalone app. The real-world application of chatbots is mostly in the customer service industry these days. Chatbots usually take over tasks that were previously handled by real people, like support agents or customer satisfaction representations.

Chatbots are an intelligent piece of software that reads the chat from the customer (text) and decides what would be the correct response. All these bots use Natural Language Processing (NLP) which is typically composed of two steps: natural language understanding which converts the text from the customer and breaks it down, applying machine learning models to understand and extract the meaning of that sentence. The second step is natural language generation which generates the reply to the customer's text based on the meaning generated in the first step. NLP, in general, is the core of building a chatbot.

In this project, you will explore a library called ChatterBot which is designed to deliver automated responses to the user inputs. It uses a combination of machine learning algorithms to identify the correct response for the given statement. In this project, you will learn how to install the dependencies required for the ChatterBot library and create/use a new python environment for the installation. You will learn how to train and test the model and also understand parameter tuning of the model to increase the accuracy.

This project will provide you with experience in NLP and model building which is a must-have in a Data Scientists portfolio.

Project #4: Twitter Sentiment Analysis

Project: Sentiment Analytics with Twitter Data

Language: Python

Link: <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>

sentiment analysis



Description

Sentiment analysis, also known as opinion mining, is a method to understand the overall sentiment of your customers by using text mining and natural language processing techniques. Sentiment analysis is used in a lot of fields, for example, in marketing, companies use sentiment analysis to understand what their customers or potential customers are talking about, what they like and dislike, and so on.

Sentiment analysis has become a crucial part of most businesses that are trying to understand their customers better.

In this project, you will perform sentiment analysis from the Tweets. You will extract the data using a Twitter API, clean the dataset and remove unwanted information from the tweets and build the model to identify whether a tweet is positive, negative, or neutral. This is a great project to have in your portfolio since you need some knowledge about APIs and how to extract data from the Twitter API to perform sentiment analysis.

Project #5: House Price Prediction in California

Project:House Price Prediction

Language:Python

Link: <https://www.kaggle.com/datasets/harrywang/housing>



This is one of the most famous projects in the data science field. There are many resources where you can find the sample datasets regarding the house prices, and you would need to build a machine learning model to predict the price of the house based on some features of the house; the number of bedrooms, location, area in square feet, etc.

One such dataset is available on [Kaggle](https://www.kaggle.com/datasets/berkeleypost/house-prices-in-berkeley). This is a great website to work on various datasets and be part of data science competitions.

In this project, you will be able to define the overall objective of the project, understand and make sense of the features available in the data, split the data into training and testing, find out the correlation between the independent variables and identify which variables are required to predict the house price.

Kaggle also provides projects that are done by other people, and it will be a great learning resource for you if you are just starting out in data science. You will learn many approaches that people follow to problem-solving.

Having this project in your portfolio is a must. This project will provide you with hands-on experience in understanding various features, identifying relationships between them, and using them to predict the value of a target variable.

Summary

From this article, you can see how important Python is for a Data Scientist. Python is a must-have skill for anyone who is interested in getting into the Data Science and Machine Learning field. Python, as a programming language, is a very flexible language as compared to C or C++, and due to this flexibility, a lot of programmers have built various packages in Python that can be used by data scientists. We discussed some packages like BeautifulSoup and Scrapy to extract the data from the web. Due to the availability of such open source packages in Python, it's the most sought-after skill in Data Science.

We also discussed some of the projects that you should have in your portfolio to get hands-on experience with Python and data analysis in general. On StrataScratch, you can get access to many projects in Data Science that were asked in the interviews as take-home assignments. All the top companies usually have either a coding round or they will give a take-home assignment, and it's important to practice to feel strong and confident in your interview.

I hope you enjoyed the article, and it gave you good clarity on how to build a data science project portfolio. Good luck with your next interview, and have fun practicing all the concepts on StrataScratch.