

Step by Step Guide

Data Scientist

Here's a general schedule for someone looking to become an advanced data scientist:

1. Beginner Level (0-6 months)

- Develop a solid foundation in mathematics (linear algebra, statistics, probability, etc.)
- Learn the basics of programming (Python, R, SQL)
- Get familiar with data manipulation and exploration using tools like Pandas and Numpy
- Start learning basic machine learning algorithms (Linear/Logistic Regression, Decision Trees, etc.)
- Practice solving simple data science problems and participate in online data science competitions

2. Intermediate Level (6-12 months)

- Get deeper into the mathematics behind machine learning algorithms
- Learn more advanced algorithms (Random Forest, Support Vector Machines, Neural Networks, etc.)
- Get familiar with deep learning frameworks such as TensorFlow and Keras
- Learn data visualization using tools such as Matplotlib, Seaborn, Plotly, etc.
- Start working on real-world projects and start building your portfolio

3. Advanced Level (12+ months)

- Start exploring specialized topics in data science such as Natural Language Processing, Computer Vision, Time-series Analysis, etc.
- Get into more advanced deep learning techniques such as Convolutional Neural Networks, Recurrent Neural Networks, etc.
- Start working on industry-level projects and build your network by participating in online communities and attending events
- Keep yourself updated with the latest trends and developments in the field of data science.

NOTE: It's important to note that the timeline may vary based on the individual's pace of learning, experience, and other factors. The above schedule is just a rough estimate and should be used as a guide rather than a strict rule.

Let's start with the **beginner level** of the schedule. Here's what you should aim to learn and cover during this period:

1. Mathematics:

- Linear Algebra: Understanding matrices and vectors, solving linear equations, eigenvalues, and eigenvectors
- Statistics: Descriptive statistics, probability distributions, hypothesis testing, estimation, and inference
- Probability: Conditional probability, Bayes Theorem, Random Variables, and Expectation

2. Programming:

- Choose either Python or R as your primary programming language for data science
- Learn the basics of the language syntax and data structures
- Familiarize yourself with libraries such as NumPy, Pandas, and Matplotlib for data manipulation and visualization

3. Data Exploration and Manipulation:

- Learn how to import, clean, and manipulate data using tools like Pandas and Numpy
- Get hands-on experience with data exploration and visualization techniques

4. Machine Learning:

- Learn the basics of machine learning algorithms such as Linear Regression, Logistic Regression, Decision Trees, and K-Nearest Neighbors
- Learn how to train and evaluate models using common metrics such as accuracy, precision, recall, and F1 score

5. Practice and Portfolio Building:

- Practice solving simple data science problems using online resources such as Kaggle and HackerRank
- Participate in online data science competitions to get hands-on experience and build your portfolio

NOTE: Remember, the goal of this beginner level is to lay a solid foundation in mathematics, programming, and machine learning. Try to spend around 6 months on this level, and aim to learn and practice as much as possible. This will prepare you well for the intermediate level, where you'll dive deeper into more advanced topics.

Let's start with the **Intermediate Level** of the schedule.
Here's what you should aim to learn and cover during this period:

1. Mathematics:

- Get deeper into the mathematics behind machine learning algorithms, such as gradient descent, regularization, and optimization techniques
- Study more advanced topics such as matrix factorization and singular value decomposition

1. Advanced Machine Learning:

- Learn more advanced machine learning algorithms such as Random Forest, Support Vector Machines (SVMs), and Neural Networks
- Study ensemble methods and how to combine multiple models to improve performance
- Learn how to perform feature engineering, feature selection, and dimensionality reduction techniques

2. Deep Learning:

- Get familiar with deep learning frameworks such as TensorFlow and Keras
- Learn how to build and train Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)
- Study transfer learning and fine-tuning pre-trained models for your specific use case

3. Data Visualization:

- Study more advanced data visualization techniques and tools such as Seaborn, Plotly, and Bokeh
- Learn how to create interactive visualizations and dashboards using frameworks such as Dash and Streamlit

4. Real-world projects and portfolio building:

- Start working on real-world projects and apply the concepts and techniques you've learned so far
- Build your portfolio and showcase your skills by participating in Kaggle competitions, open-source projects, or creating your own personal projects

NOTE: At this intermediate level, aim to spend around 6 to 12 months, and focus on hands-on learning and practical applications of the concepts and techniques you've learned. This will prepare you well for the advanced level, where you can start exploring specialized topics in data science.

Let's start with the **advanced level** of the schedule. Here's what you should aim to learn and cover during this period:

1. Specialized Topics:

- Study specialized topics in data science such as Natural Language Processing (NLP), Computer Vision, Time-series Analysis, and Graphical Models
- Learn how to use libraries and tools such as SpaCy, OpenCV, and Gensim for NLP and computer vision tasks

2. Advanced Deep Learning:

- Study advanced deep learning techniques such as Generative Adversarial Networks (GANs), Reinforcement Learning, and Transfer Learning
- Learn how to build and train models for specific use cases such as image classification, object detection, and generative models



3. Industry-level projects and network building:

- Start working on industry-level projects and apply the concepts and techniques you've learned so far
- Build your network by participating in online communities and attending events, such as data science conferences, meetups, and hackathons
- Staying updated with the latest trends and developments:
- Keep yourself updated with the latest trends and developments in the field of data science by reading research papers, following leaders in the industry, and participating in online courses and workshops

At this advanced level, aim to continue learning and practicing as much as possible, with a focus on hands-on projects and real-world applications. This level should continue throughout your career as a data scientist, as the field is constantly evolving and growing.

Note: that the timeline for this level is open-ended, as the goal is to continually improve and advance your skills as a data scientist. The focus should be on continuous learning and practical applications of the concepts and techniques you've learned so far.

Here are some **applications** that can be used at each level of the data science learning journey

1. Beginner level:


- Jupyter Notebook
- Google Colab
- R Studio
- Excel
- Python libraries such as NumPy, Pandas, and Matplotlib for data manipulation, analysis, and visualization

2. Intermediate level:

- Advanced Python libraries such as Seaborn, Plotly, and Bokeh for data visualization
- Machine learning libraries such as scikit-learn, XGBoost, and LightGBM for building and training models
- Deep learning frameworks such as TensorFlow and Keras for building and training neural networks

3. Advanced level:

- Specialized libraries and tools for specific tasks such as NLP (SpaCy, NLTK, Gensim), computer vision (OpenCV), and time-series analysis (Prophet, ARIMA)
- Advanced deep learning frameworks such as PyTorch and TensorFlow 2.0 for building and training complex models
- Tools for deployment and production such as Flask, Django, and TensorFlow Serving for deploying models and building production-level applications.



Note: that this list is not exhaustive, and there are many other tools and applications that can be used at each level, depending on your specific goals and use cases. The important thing is to continue learning and exploring new tools and technologies to expand your skills as a data scientist.

Here are some suitable **project** ideas that can be worked on at each level of the data science learning journey

1. Beginner level:

- Exploratory data analysis (EDA) on a publicly available dataset
- Building basic predictive models such as linear regression, decision trees, and k-nearest neighbors
- Data visualization projects, such as creating bar plots, histograms, and scatter plots to visualize relationships in the data
- Sentiment analysis on a text dataset, such as movie reviews or social media posts
- Anomaly detection in a time-series dataset

2. Intermediate level:

- Building more advanced machine learning models such as Random Forest, Support Vector Machines (SVMs), and Neural Networks
- Performing feature engineering, feature selection, and dimensionality reduction techniques
- Building deep learning models for image classification or object detection tasks
- Building a recommendation system for a product or service

- Time-series forecasting on a financial or environmental dataset

3. Advanced level:

- Building advanced deep learning models such as Generative Adversarial Networks (GANs) and Reinforcement Learning models
- Building NLP models for tasks such as text classification, entity recognition, and machine translation
- Building computer vision models for tasks such as image segmentation, object tracking, and face recognition
- Building models for specialized use cases such as personalized medicine, financial market prediction, and autonomous driving

Note: that this list is not exhaustive and there are many other project ideas that can be pursued at each level, depending on your specific goals and interests. The important thing is to continue learning and working on projects that challenge and stretch your skillset as a data scientist.

Here are some **job openings** in the data science field that can be pursued at each level of experience

1. Beginner level:

- Data Analyst
- Business Intelligence Analyst
- Junior Data Scientist
- Data Engineer

2. Intermediate level:

- Data Scientist

-
- Machine Learning Engineer
 - Data Engineer (with a focus on ML pipelines)
 - Business Intelligence Manager

3. Advanced level:

- Senior Data Scientist
- Machine Learning Manager
- Director of Data Science
- Principal Data Scientist

Note: that this list is not exhaustive, and there are many other job roles in the field of data science, depending on your specific skill set, interests, and goals. The important thing is to continue learning and acquiring new skills and expertise to advance your career in the field of data science. Also, keep in mind that the data science field is rapidly evolving, and new job roles and titles are constantly being created as the field grows and expands.

Here are some **companies** that have job openings in the data science field at each level of experience:

1. Beginner level:

- Accenture
- IBM
- Wipro
- TATA Consultancy Services (TCS)
- Infosys



2. Intermediate level:

- Amazon
- Google
- Microsoft
- Facebook
- Apple
- Airbnb
- Uber

3. Advanced level:

- OpenAI
- DeepMind
- NVIDIA
- Tesla
- Amazon Web Services (AWS)
- IBM Watson
- Google Cloud

Note: that this list is not exhaustive, and there are many other companies that have job openings in the data science field, depending on your specific skill set, interests, and goals. It is important to research companies in the industry, understand their needs and requirements, and tailor your job applications to best match their needs.

Here is a general idea of **average salaries** for each level:

1. Beginner level:

- Data Analyst: \$60,000 to \$80,000
- Business Intelligence Analyst: \$65,000 to \$85,000
- Junior Data Scientist: \$70,000 to \$90,000
- Data Engineer: \$75,000 to \$95,000

2. Intermediate level:

- Data Scientist: \$90,000 to \$140,000
- Machine Learning Engineer: \$100,000 to \$150,000
- Data Engineer (with a focus on ML pipelines): \$110,000 to \$165,000
- Business Intelligence Manager: \$120,000 to \$180,000

3. Advanced level:

- Senior Data Scientist: \$140,000 to \$220,000
- Machine Learning Manager: \$160,000 to \$250,000
- Director of Data Science: \$200,000 to \$300,000
- Principal Data Scientist: \$220,000 to \$350,000

NOTE: Again, it's important to note that these salaries are rough estimates and may vary widely based on location, company size, and other factors.

Here are some popular **online courses** that cover the material for each level of data science experience:

1. Beginner level:

- Coursera: Introduction to Data Science
- Udemy: Data Science Fundamentals
- edX: Data Science Essentials
- YouTube channel: Data School

2. Intermediate level:

- Coursera: Applied Data Science
- Udemy: Machine Learning A-Z
- edX: Artificial Intelligence (AI)
- YouTube channel: Sentdex

3. Advanced level:

- Coursera: Advanced Data Science
- Udemy: Deep Learning A-Z
- edX: Advanced Artificial Intelligence (AI)
- YouTube channel: Two Minute Papers

NOTE: These are some of the most well-regarded online courses and YouTube channels for data science, and they are a great place to start for anyone looking to learn more about the field. However, there are many other courses and channels available, and the best one for you may depend on your specific interests, needs, and learning style.

here are some of the best online platforms for preparing for data science **interviews** at each level

1. Beginner level:

- LeetCode (leetcode.com)
- HackerRank (hackerrank.com)
- DataCamp (datacamp.com)
- GeeksforGeeks (geeksforgeeks.org)

2. Intermediate level:

- InterviewBit (interviewbit.com)
- Pramp (pramp.com)
- AlgoExpert (algoexpert.io)
- Codeforces (codeforces.com)

3. Advanced level:

- Project Euler (projecteuler.net)
- Cracking the Coding Interview ([book by Gayle McDowell](#))
- CodeSignal (codesignal.com)
- Kaggle (kaggle.com)

ALL THE BEST