

## Sakib's 1-2 Year Data Science & AI/ML Roadmap

Embarking on this journey, remember that **you can do it step by step**. Data science blends coding with math and creativity, so take it one concept or project at a time <sup>1</sup>. As a future *Data Scientist*, you'll bridge the gap between raw data and real insights – a different role from data analysts (who focus on dashboards) or data engineers (who build data pipelines) [58†]. The phases below lay out a clear plan: build solid foundations, then tackle core ML skills and real projects, then specialize. Each phase lasts a few months so you steadily grow your portfolio and confidence.

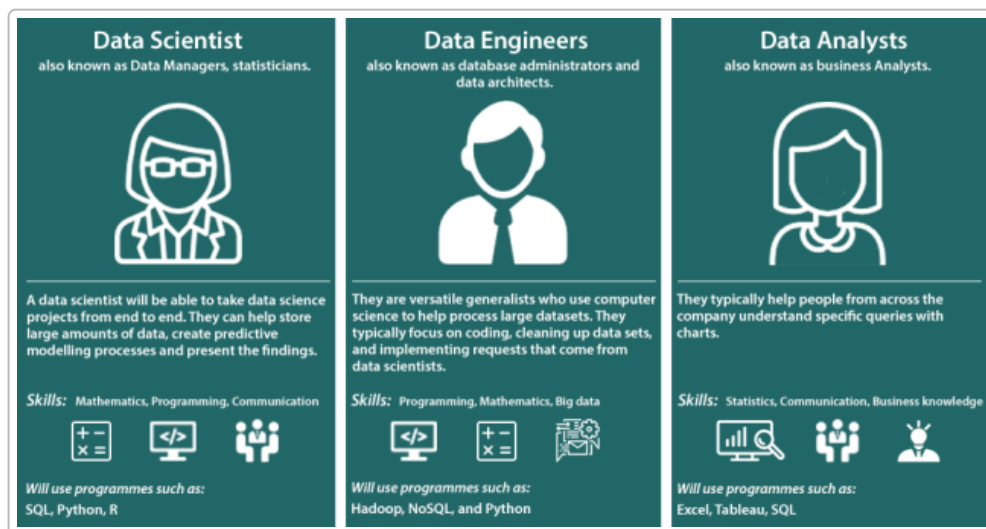


Figure: Data-related roles. A Data Scientist (left) handles end-to-end projects (cleaning, modeling, insights), unlike Data Analysts or Data Engineers [58†].

### Phase 1: Build Your Foundation (Months 1-3)

- **Strengthen Python & Tools:** You already know Python basics, Git, CLI and Pandas. Deepen this by solving problems daily. Use Kaggle's free *Python* and *Pandas* micro-courses <sup>2</sup> or YouTube tutorials (e.g. Corey Schafer) to reinforce syntax, data manipulation and version control. Practice on a Jupyter Notebook or Google Colab, and learn SQL basics if you can (managing data tables is essential later).
- **Learn Core Math & Stats:** Start *descriptive statistics* and *basic probability* in an applied way. For example, use Khan Academy videos or a Stanford *Intro to Statistics* Coursera course <sup>3</sup> to learn about means, medians, distributions and sampling. Immediately apply these: compute summary stats on a dataset. (You can also watch 3Blue1Brown's "Essence of Linear Algebra" videos for intuition when you reach vectors and matrices <sup>4</sup>.) Keep it practical: e.g. when you learn about means and standard deviation, calculate them for your data to see what they tell you.
- **Intro Data Projects:** Pick a simple real dataset (Kaggle's [Titanic](#) or a U.S. city dataset) and do basic analysis. Example project: scrape some data (since you know web scraping) such as weather or sports scores, then load it into Pandas and explore: clean missing values, plot distributions of key

columns, compute averages. Document your code and findings. Always **push this code to GitHub** and write a clear README – this begins building your portfolio <sup>5</sup>.

- **Visualization & Communication:** Learn one plotting library (Matplotlib/Seaborn) or even basic Excel graphs to visualize data. Create at least one report or slide deck summarizing a project's findings. Good visual communication is key: businesses prize clear charts and narratives.

**Quick Win:** Complete Kaggle's free "Intro to Machine Learning" micro-course or *Python* course <sup>2</sup>. These courses are short (few hours) and cover essentials. By the end of Phase 1, you should have 1–2 small projects (e.g. Kaggle Titanic, a scraped dataset analysis) on GitHub and feel comfortable with basic stats and data handling.

## Phase 2: Core Data Science Skills (Months 4–6)

- **Supervised Learning (ML Basics):** Learn common ML algorithms: linear regression, logistic regression, k-nearest neighbors, decision trees, and simple clustering (k-means). A hands-on way is scikit-learn tutorials or Kaggle's *Machine Learning* micro-course <sup>2</sup>. Focus on *understanding what each model does* rather than math details. For example, build a linear regression model on a housing dataset and see how coefficients relate to feature importance. This solidifies why linear algebra (vectors, dot products) is useful in ML <sup>4</sup>.
- **More Statistics:** Cover *hypothesis testing* and probability distributions (normal, binomial). Use real examples: e.g. test if two groups have different means (t-test) for some feature in your data. This practical practice reinforces the theory. The KDnuggets list mentions a free Stanford course covering exactly these topics <sup>3</sup>.
- **Feature Engineering:** Learn to clean and transform data for better models. Practice encoding categories (one-hot encoding), imputing missing values, and scaling features (normalization). These skills **improve model accuracy** in real projects.
- **Project: Intermediate Kaggle Challenge:** Apply these skills on a full Kaggle dataset like *House Prices – Advanced Regression Techniques* or *Titanic with more features*. Do the full cycle: EDA, feature engineering, model training, and submit predictions. Even if you don't top the leaderboard, use it as a portfolio piece. Push the completed notebook/code to GitHub. Writing a blog post or short report on your approach (see tips below) will make this even stronger <sup>5</sup> <sup>6</sup>.
- **Kaggle and Community:** Start a Kaggle profile if you haven't. Even entering beginner competitions is valuable practice. Kaggle encourages running code and learning; their forums and "kernels" (shared code notebooks) can inspire project ideas. Participating in Kaggle adds to your portfolio – it shows initiative and familiarity with real-world data challenges <sup>7</sup>.
- **Portfolio Tip:** As you complete each project, **write it up**. Commit code to GitHub and consider a short blog post or Medium article. According to data science experts, sharing your projects and tutorials online dramatically boosts visibility <sup>5</sup>. A portfolio with code *and* explanations demonstrates your skills clearly to scholarship committees and future employers <sup>6</sup>.

## Phase 3: Advanced Machine Learning & AI (Months 7–12)

- **Deepen ML & Intro to AI:** Study more advanced models (ensemble methods like random forests/gradient boosting) and begin *neural networks*. A practical start: build a simple neural net with Keras/TensorFlow or PyTorch on a dataset like MNIST (handwritten digits) or fashion images. This illustrates how **calculus (gradients) and linear algebra (matrix multiplications)** power learning in neural nets <sup>4</sup>. Keep the focus on making models work end-to-end.

- **Unsupervised & Special Topics:** Learn unsupervised techniques (PCA for dimensionality reduction, clustering analysis on unlabeled data) with hands-on projects (e.g. segment customers by behavior). Try at least one domain-specific mini-project: for example, do basic NLP (text classification of tweets or reviews) or computer vision (image classification) using pre-built libraries. This broad exposure will help you later choose a niche.
- **Data Pipeline & Tools:** Start understanding the *end-to-end pipeline*: how to collect data, version it, and deploy models. Learn a bit of SQL for databases if possible. Try deploying a model via a simple Flask or Streamlit app (this can be a side project). Even doing a “toy” deployment (model predicts on new input) shows maturity. (You don’t have to master MLOps now, but awareness of productionizing models sets you apart.)
- **Major Project:** Undertake a larger real-world project. Ideas: analyze a public dataset to solve a real problem (predicting air quality, forecasting sales, or building a recommendation system). Alternatively, join an open-source data science project or Kaggle competition with a team. Document it fully. For example, you might scrape tweets about a topic, do sentiment analysis, and present the findings in a dashboard.
- **Portfolio & Writing:** At this stage, refine your top 3 projects. Write detailed READMEs explaining your process and results. Consider writing a tutorial or making a short YouTube video about one project – teaching a topic is a known way to demonstrate expertise <sup>8</sup>. Keep updating your GitHub and Kaggle profiles with these accomplishments.

**Freelancing/Side Jobs Start (Month ~9+):** Once you have solid projects, create profiles on platforms like Upwork or Toptal. Highlight exactly the projects you completed (with links). As 365 Data Science notes, building a portfolio is *key* – push your code to GitHub, share work widely, and clients will take notice <sup>5</sup>. On your profile, emphasize concrete outcomes (“I built X model that achieved Y accuracy” with KPIs) <sup>9</sup>. Start small: bid on data cleaning or simple analysis jobs to get initial ratings. Focus on one platform (Upwork for volume; Toptal for high-end clients <sup>10</sup>) and do excellent work. Accumulate reviews and project hours gradually <sup>9</sup> – these will unlock better gigs.

## Phase 4: Specialization & Professional Growth (Months 13–24)

- **Choose a Niche:** By now you’ll notice fields you enjoy (finance data, healthcare, image analysis, NLP, etc.). Delve deeper: if, say, NLP excites you, take a specialized course or work on more language projects. Building expertise in one domain can make you stand out later.
- **Advanced Tools & Math:** Learn tools like PyTorch for deep learning, or Spark/BigQuery for big data, as needed. Study advanced math in context: for example, learn multivariable calculus concepts when fine-tuning neural nets, or linear algebra for optimization. Khan Academy or 3Blue1Brown videos can help with intuition. You don’t need full proofs, but know *why* algorithms work.
- **Capstone Project:** Plan an ambitious end-to-end project. Ideas: deploy an AI model as a web app, build a data pipeline from data ingestion to prediction, or collaborate on research. The goal is a project big enough to showcase (for example, a multi-model solution to a complex problem, documented in a portfolio case study). Aim to cover the full pipeline (data gathering, processing, model, and communication of results).
- **Networking & Contribution:** Join online communities (Kaggle forums, GitHub, local DS meetups or online Slack groups). Contribute to open-source (even minor fixes) or volunteer your skills on a small non-profit data project. Scholarship committees and employers love evidence of teamwork and service, as it “demonstrates commitment to the field” <sup>6</sup>.

- **Refine Portfolio for Scholarships:** By the end of two years, have **4–6 polished projects** on GitHub (with clear explanations) and an active Kaggle profile. Include variety (e.g. one visualization/analysis project, one ML project, one simple DL project). Use dashboards or slide decks to present your work visually. According to scholarship guides, a strong project/internship portfolio helps you **stand out to selection committees** <sup>6</sup> . Make sure your GitHub is tidy and your LinkedIn/Resume link to it.
- **Freelancing & Career Launch:** Now you're ready for serious freelancing or job search. With your portfolio, clients or recruiters will be impressed. On freelancing platforms, you can increase your rates. Target roles that match your niche. Always be learning: keep tabs on new DS/AI trends and incorporate them into projects. Eventually you might land a remote internship or job – treat that like a learning opportunity and resume builder.

## Continuous Learning Tips

- **Daily Routine:** Commit ~1.5 hours per day as you planned. One focused topic or small project each day will compound. As one data scientist advises, “focus on just one thing each day...don't overwhelm yourself” <sup>1</sup> . Over time this steady progress adds up.
- **Use Free Resources:** Make the most of free/low-cost platforms: Coursera (audit courses like Andrew Ng's ML course), edX, Khan Academy, FreeCodeCamp YouTube tutorials, and Kaggle Learn. For example, Stanford's Intro to Stats (free on Coursera) or Khan Academy's Linear Algebra. DataCamp offers beginner courses, sometimes free via student promotions <sup>11</sup> . Also, follow YouTube channels like *StatQuest* (statistics) and *freeCodeCamp.org* (Python, ML) for clear walkthroughs.
- **Stay Practical:** Always tie learning to a project. Learning a concept like regression feels much more real when you use it to predict prices or trends. Hackathons and Kaggle competitions are great opportunities to apply skills under some time pressure, as are guided capstone courses (many MOOCs have these).
- **Portfolio Maintenance:** Periodically update your portfolio. Remove any truly basic practice projects once you have stronger work, and keep the best ones pinned or highlighted. Add screenshots or visual summaries to README files. Link to any blogs or talks you've given. For scholarships, a clean and impactful portfolio page (GitHub or personal site) can make a lasting impression <sup>5</sup> <sup>6</sup> .

You now have a personalized, multi-phase plan. Stick with it patiently, build projects continuously, and share your work widely. Every small project, GitHub commit, and Kaggle submission moves you closer to mastery. Good luck, Sakib – your data science journey starts today!

**Sources:** This roadmap draws on expert guidance and resources (free courses, Kaggle tutorials) and emphasizes building a strong project portfolio <sup>5</sup> <sup>6</sup> <sup>2</sup> <sup>4</sup> . It integrates career advice on freelancing and scholarship-ready portfolios from industry veterans <sup>10</sup> <sup>6</sup> .

---

<sup>1</sup> DataScience RoadMap for 2025. DataScience RoadMap for 2024 | by Parvez Shah Shaik | Bootcamp | Medium

<https://medium.com/design-bootcamp/datascience-roadmap-for-2024-96b916a200e9>

<sup>2</sup> 7 Free Kaggle Micro-Courses for Data Science Beginners - KDnuggets

<https://www.kdnuggets.com/7-free-kaggle-micro-courses-for-data-science-beginners>

<sup>3</sup> 5 Free Courses to Master Statistics for Data Science - KDnuggets

<https://www.kdnuggets.com/5-free-courses-to-master-statistics-for-data-science>

4 Linear Algebra Concepts Every Data Scientist Should Know | by Benedict Neo | bitgrit Data Science Publication | Medium

<https://medium.com/bitgrit-data-science-publication/linear-algebra-concepts-every-data-scientist-should-know-18b00bd453dd>

5 7 8 9 10 How to Become a Freelance Data Scientist: 2025 Full Guide – 365 Data Science

<https://365datascience.com/career-advice/career-guides/freelance-data-scientist/>

6 Top 40 Computer Science Scholarships in 2025

<https://scholaroo.com/computer-science-scholarships/>

11 GitHub - Moataz-Elmesmary/Data-Science-Roadmap: Data Science Roadmap from A to Z

<https://github.com/Moataz-Elmesmary/Data-Science-Roadmap>