**RoadMap to Master Data Scientist Skills**

**in 4 Months**

**Created by AI Coach John, CEO of ProITbridge.**

**Welcome to the RoadMap to Master Data Scientists Skills in 2025 in 4 Months – your personalized guide to clarity and direction.**

This isn't just another roadmap; This guide is for fresh graduates of 2025 who are serious about getting their first IT job in the next 6 to 9 months. If you're confused about coding, career options, or how to get started - this is your GPS.  
  
This roadmap includes weekly/monthly checkpoints, skill milestones, and alternate tracks (for both coding and non-coding paths).  
  
Remember: It’s not about aiming for MAANG on Day 1. It’s about building your real talent, getting into a solid job, and growing from there.  
  
Let’s begin.

**Targeted Job Positions**

* Data Scientist
* ML Engineer
* CV Engineer
* NLP Engineer
* LLM Engineer
* Data Analyst

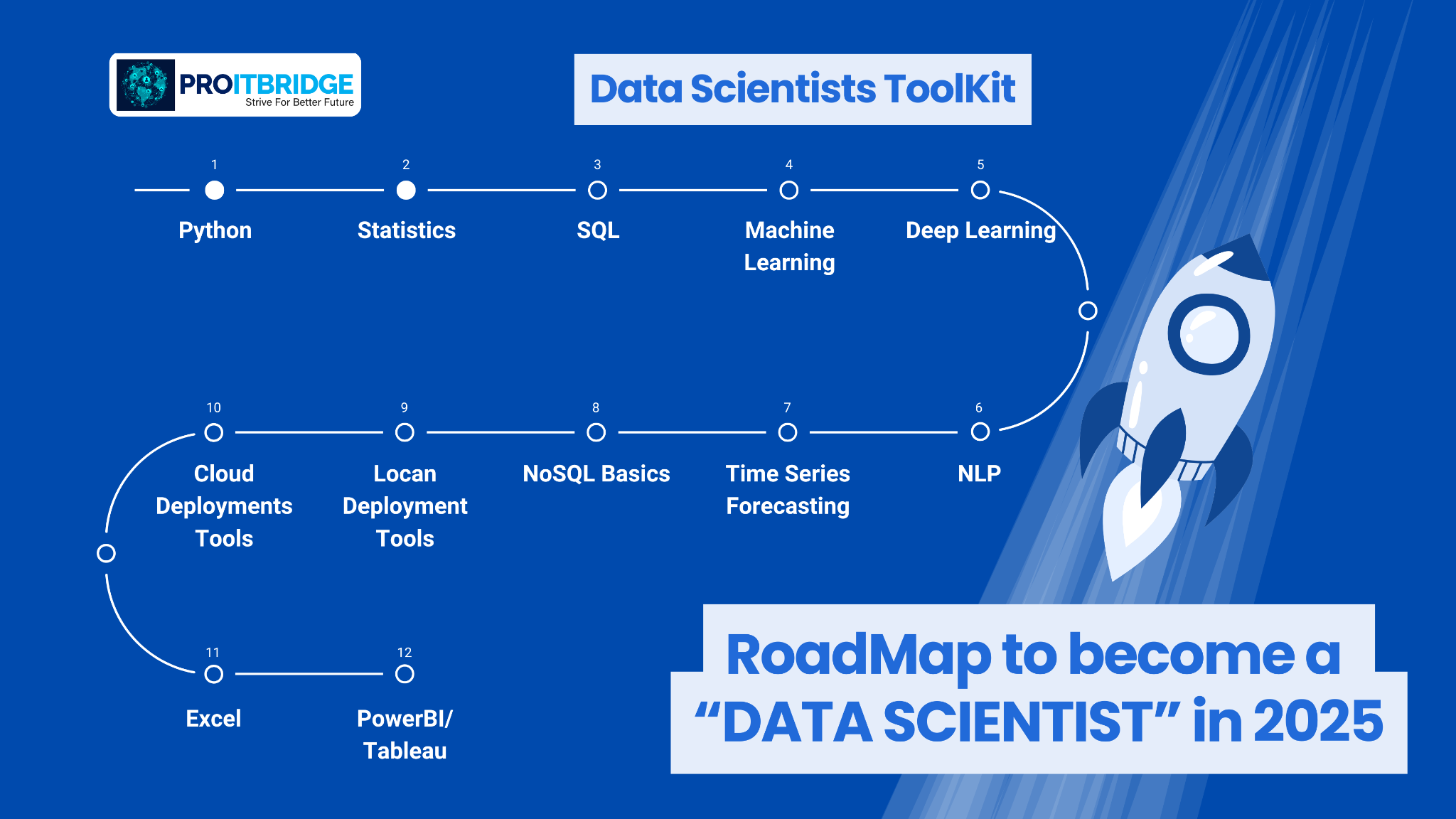
**Salary Range - 4 to 8 LPA**

**Step 1: IT Job Market Survey**

***99% of the Freshers will exclude this step and that’s why they suffer***

| **Step 1** | Do research in IT Job Market(Recruitment Sites) and find if we have opportunities, skills expected, Salary offered, etc., |
| --- | --- |
| **Step 2** | Get IT Job Market Awareness |
| **Step 3** | You can take informed decisions here |
| **Step 4** | Start with the Commitment and Own the Consequences   * Both positive and negative during the process |
| **Step 5** | You can win with Consistency in the Plan |

**Phase 1 - UPSKILL [4 Months]**



**Month 1 – Python Foundations, Data Handling, SQL, Statistics, Regression & Git**

**Week 1: Python Setup, Core Programming & Pandas Basics**

**Topics:**

* Python Installation & Environment Setup
* Explore IDEs: VS Code, Jupyter, Colab
* Jupyter Notebook Basics
* Python Basics: Syntax, Variables, Strings, Input/Output
* Loops, Functions, Conditional Statements
* Python Libraries Exploration: Pandas, NumPy, Seaborn, Plotly
* Python Data Structures: Lists, Tuples, Dictionaries, Sets

**Assignments:**

1. Write a Python script to calculate the factorial of a number.
2. Create a simple calculator using Python functions.
3. Solve 10 Python problems on HackerRank.
4. Analyze a small dataset using Pandas.
5. Create a contact book application using dictionaries.
6. Develop a file reader that processes structured text.

**YouTube Resources:**

* [Python for Beginners – Learn Coding with Python in 1 Hour](https://www.youtube.com/watch?v=rfscVS0vtbw)
* [Complete Python Pandas Data Science Tutorial – Codebasics](https://www.youtube.com/watch?v=vmEHCJofslg)
* [Python Data Structures – Lists, Tuples, and Dictionaries – Edureka](https://www.youtube.com/watch?v=WGJJIrtnfpk)

**Week 2: Pandas Deep Dive, Data Cleaning & Visualization**

**Topics:**

* DataFrames, Series, Indexing, Selection
* Data Input/Output
* Data Cleaning Techniques
* Handling Missing Data
* Data Visualization with Matplotlib, Seaborn, Plotly
* NumPy Arrays, Broadcasting, Indexing

**Assignments:**

1. Load and explore a dataset from Kaggle.
2. Perform basic data cleaning operations.
3. Clean a messy dataset by handling missing values and duplicates.
4. Create various plots (bar, line, scatter) using Matplotlib and Seaborn.
5. Perform numerical computations using NumPy.

**YouTube Resources:**

* [Pandas Tutorials for Data Analysis – Codebasics](https://www.youtube.com/watch?v=vmEHCJofslg)
* [Matplotlib Tutorial – Full Course – Programming with Mosh](https://www.youtube.com/watch?v=UO98lJQ3QGI)
* [Seaborn Tutorial – Full Course – DataCamp](https://www.youtube.com/watch?v=GcXcSZ0gQps)
* [Plotly Tutorial for Beginners](https://www.youtube.com/watch?v=9GYmFXBitBw&list=PLBSCvBlTOLa8rf2kGkP_Bx5xXqT-er4Yq&ab_channel=StatsWire)

**Week 3: Object-Oriented Programming (OOP), Statistics & Probability**

**Topics:**

* Object-Oriented Programming: Classes, Inheritance, Polymorphism
* Statistics Concepts: Mean, Median, Mode, Variance, Skewness, Kurtosis
* Random Variables, Probabilities
* Confidence Intervals, Hypothesis Testing
* Statistical Tests: T-Test, Z Test, ANOVA, Chi-square Test

**Assignments:**

1. Develop a class-based Python application for simple analytics.
2. Calculate statistical measures for a given dataset.
3. Conduct hypothesis testing on survey data.
4. Interpret the results and document findings.

**YouTube Resources:**

* [Object-Oriented Programming (OOP) in Python – Edureka](https://www.youtube.com/watch?v=JeznW_7DlB0)
* [Complete Statistics For Data Science In 6 hours By Krish Naik](https://youtu.be/LZzq1zSL1bs?si=v8IA5QKaFCQZLG5n)

**Week 4: SQL & Git**

**Topics:**

* SQL Beginners to Advanced
* Version Control with Git & GitHub – Setup, Repo, Branch, Merge

**Assignments:**

* Set up a Git repository, create branches, make commits, and merge branches on GitHub.

**YouTube Resources:**

* [Learn SQL Beginner to Advanced in Under 4 Hours](https://youtu.be/OT1RErkfLNQ?si=sqI0RXNj-NeTtJ2X)
* [Git Branching and Merging - Detailed Tutorial](https://www.youtube.com/watch?v=Q1kHG842HoI)

**Month 2 – Advanced Machine Learning, Model Tuning & Deployment**

**Week 1: Supervised Learning & Ensemble Methods**

**Topics:**

* Simple Linear Regression Multiple Linear Regression (till Pickle file)
* Logistic Regression
* Evaluation Metrics (Confusion Matrix, ROC, AUC)
* Regularization: Lasso, Ridge, ElasticNet
* Statsmodels for Data Analysis
* Decision Trees + GridSearchCV
* Ensemble Methods: Random Forest, AdaBoost, Gradient Boosting, XGBoost, LightGBM
* Model Validation Techniques (Cross-Validation, K-Fold)

**Assignments:**

* Implement Simple and Multiple Linear Regression models; save the Multiple Linear Regression model using Pickle.
* Build a Logistic Regression model; evaluate it using Confusion Matrix, ROC, and AUC.
* Apply Lasso, Ridge, and ElasticNet regularization techniques on regression models.
* Use Statsmodels to perform data analysis and interpret the results.
* Build a Decision Tree model and tune hyperparameters using GridSearchCV.
* Implement Random Forest, AdaBoost, Gradient Boosting, XGBoost, and LightGBM models; compare their performances.
* Apply Cross-Validation and K-Fold techniques to evaluate model performance.

**YouTube Resources:**

* [Simple Linear Regression, Clearly Explained!!!](https://www.youtube.com/watch?v=nk2CQITm_eo)
* [Statistics 101: Multiple Linear Regression, The Very Basics](https://m.youtube.com/watch?t=645s&v=dQNpSa-bq4M)
* [StatQuest: Logistic Regression](https://www.youtube.com/watch?v=yIYKR4sgzI8)
* [145 - Confusion matrix, ROC and AUC in machine learning](https://www.youtube.com/watch?v=Joh3LOaG8Q0)
* [Regularization Part 3: Elastic Net Regression](https://www.youtube.com/watch?v=1dKRdX9bfIo)
* [Introduction to statsmodels](https://www.youtube.com/watch?v=2BdfjqyWj3c)
* [Decision Tree Classifier and GridSearchCV In Python - Scikit Learn](https://www.youtube.com/watch?v=km71sruT9jE)
* [Part 35-Boosting algorithms in machine learning (AdaBoost, GBM ...](https://www.youtube.com/watch?v=D9A8zIiJSAo)
* [K Fold Cross Validation| Complete Explanation in 10 minutes](https://www.youtube.com/watch?v=Qep7dXjP1a8)

**Week 2: Other Supervised ML Algorithms & Feature Engineering**

**Topics:**

* KNN and SVM – Theory, Script
* Feature Engineering Techniques
* Handling Imbalanced Datasets
* Feature Importance Visualization

**Assignments:**

* Implement KNN and SVM algorithms; evaluate their performances.
* Perform feature engineering techniques such as encoding, scaling, and feature selection.
* Handle imbalanced datasets using appropriate techniques.
* Visualize feature importance using suitable methods.

**YouTube Resources:**

* [Master Naive Bayes, KNN & SVM: Complete Math Explained!](https://www.youtube.com/watch?v=RUFfqxcXAmc)
* [Feature Engineering Tutorial Python # 1](https://www.youtube.com/watch?v=pYVScuY-GPk)
* [Handling imbalanced dataset in machine learning](https://m.youtube.com/watch?v=JnlM4yLFNuo)
* [Interpreting and Visualizing Features Importance In Decision Tree](https://www.youtube.com/watch?v=0QgtrqNWkRQ)

**Week 3: UnSupervised Algorithms & Recommendations**

**Topics:**

* K-Means Clustering, DBSCAN, Hierarchical Clustering
* Cluster Validation Techniques
* Association Rules: Apriori – Theory & Script
* Recommendation Engine – Basic Script & Workflow
* Model Optimization Techniques – Part 2 (depth, estimators, learning rate)

**Assignments:**

* Apply K-Means, DBSCAN, and Hierarchical Clustering algorithms on datasets.
* Validate clustering results using appropriate techniques.
* Apply Apriori algorithm for association rule mining.
* Build a basic recommendation engine.
* Optimize models by tuning depth, number of estimators, and learning rate.

**YouTube Resources:**

* [K-means, Hierarchical Clustering, DBSCAN](https://www.youtube.com/watch?v=abGTypGb4MQ)
* [Association Rules: Apriori – Theory & Script](https://www.youtube.com/watch?v=9yl6-HEY7_s)
* [Recommendation Engine – Basic Script & Workflow](https://www.youtube.com/watch?v=ZspR5PZemcs)
* [Model Optimization Techniques – Part 2](https://www.youtube.com/watch?v=J5bXOOmkopc)

**Week 4: Dimensionality Reduction & Model Deployment**

**Topics:**

* Dimensionality Reduction: PCA, t-SNE
* Streamlit App Development & Deployment
* Introduction to MLflow for ML Lifecycle Tracking

**Assignments:**

* Apply PCA and t-SNE for dimensionality reduction.
* Develop and deploy a machine learning model using Streamlit.
* Track ML experiments using MLflow.

**YouTube Resources:**

* [Streamlit App Development & Deployment](https://www.youtube.com/watch?v=JwSS70SZdyM)
* [PCA : the math - step-by-step with a simple example](https://www.youtube.com/watch?v=S51bTyIwxFs&ab_channel=TileStats)
* [Visualizing Complex Data: PCA vs t-SNE Techniques](https://youtu.be/uYy9zJSCP2Q?si=kwAQzE8lOQIHWsr1)
* [MLFlow Tutorial | ML Ops Tutorial](https://www.youtube.com/watch?v=6ngxBkx05Fs&ab_channel=codebasics)

**Month 3 – Deep Learning, NLP & TSM Capstone Projects**

**Week 1: Deep Learning Foundations**

**Topics:**

* ANN: Perceptron, Activation Functions, Backpropagation
* Loss Functions: MSE, Cross-Entropy
* Build MNIST Digit Classifier from scratch
* Introduction to Deep Learning Frameworks: TensorFlow, Keras, PyTorch, Fast.ai

**Assignments:**

* Build a simple Artificial Neural Network using Keras
* Train a classifier on the MNIST dataset
* Implement Backpropagation from scratch in NumPy
* Compare TensorFlow vs PyTorch for a basic ANN

**YouTube Resources:**

* [Deep Learning with Python – Sentdex](https://www.youtube.com/playlist?list=PLQVvvaa0QuDfSfqQuee6K8opKtZsh7sA9)
* [Build Neural Network from Scratch – 3Blue1Brown](https://www.youtube.com/watch?v=aircAruvnKk)
* [TensorFlow vs PyTorch – Krish Naik](https://www.youtube.com/watch?v=GIsg-ZUy0MY)
* [Neural Network For Handwritten Digits Classification | Deep Learning Tutorial 7 (Tensorflow2.0)](https://youtu.be/iqQgED9vV7k?si=LGpnVpPfujMYWJcM)

**Week 2: CNNs & Architectures**

**Topics:**

* CNN Concepts: Filters, Kernels, Pooling, Padding
* CNN Architectures: LeNet, AlexNet, VGGNet, ResNet
* Transfer Learning: Feature Extraction, Fine-Tuning
* Image Classification Tasks using Pretrained Models

**Assignments:**

* Build a CNN from scratch using Keras/TensorFlow
* Fine-tune a pre-trained model (e.g., VGG or ResNet) on a custom dataset
* Compare CNN architectures using accuracy and speed
* Visualize feature maps from different convolution layers

**YouTube Resources:**

* [CNN Explained – Codebasics](https://www.youtube.com/watch?v=YRhxdVk_sIs)
* [VGG and ResNet Architecture – Stanford CS231n](https://www.youtube.com/watch?v=1hPgQWbWmEk)
* [What is Transfer Learning? Transfer Learning in Keras | Fine Tuning Vs Feature Extraction](https://youtu.be/WWcgHjuKVqA?si=mBbdeQ-Zi_JbdB1t)
* [Image classification vs Object detection vs Image Segmentation | Deep Learning Tutorial 28](https://youtu.be/taC5pMCm70U?si=AbaXcbIjKo4Gnymk)

**Week 3: Computer Vision + AutoEncoders + TSM**

**Topics:**

* OpenCV Basics: Image Loading, Processing, Drawing
* Object Detection: YOLO Overview & Basics
* AutoEncoders: Encoder-Decoder, Reconstruction Loss
* Use AutoEncoders for Image Denoising

**Assignments:**

* Build a basic object detection pipeline using OpenCV
* Train an AutoEncoder to denoise noisy images
* Explore pre-trained YOLO with OpenCV or Ultralytics
* Project: Real-time object detection using OpenCV webcam feed

**YouTube Resources:**

* [OpenCV Python Tutorial – Murtaza's Workshop](https://www.youtube.com/watch?v=WQeoO7MI0Bs)
* [YOLO Object Detection – Ultralytics](https://www.youtube.com/watch?v=Grir6TZbc1M)
* [AutoEncoders Explained – Krish Naik](https://www.youtube.com/watch?v=9zKuYvjFFS8)

**Week 4: NLP + Transformers + LLMs**

**Topics:**

* Text Preprocessing: Tokenization, One-Hot, BoW, TF-IDF
* Word Embeddings: Word2Vec, GloVe
* Sequence Models: RNN, LSTM, GRU
* Transformers: BERT, GPT (1–4)
* Hugging Face Transformers: Pipelines, Fine-tuning
* Generative AI: OpenAI API, Stable Diffusion
* Time Series Forecasting: ARIMA, Prophet, LSTM
* LangChain: Building LLM Workflows
* Time Series Forecasting

**Assignments:**

* Preprocess a dataset with BoW, TF-IDF, Word2Vec
* Build a sentiment classifier using LSTM or BERT
* Fine-tune a transformer model using Hugging Face
* Use OpenAI API to generate responses or summaries
* Forecast future sales or trends using Prophet

**YouTube Resources:**

* [NLP Playlist – Krish Naik](https://www.youtube.com/watch?v=05ONoGfmKvA)
* [RNN/LSTM Tutorial – Codebasics](https://www.youtube.com/watch?v=LHXXI4-IEns)
* [Transformers Demystified – Jay Alammar](https://www.youtube.com/watch?v=U0s0f995w14)
* [Complete Time Series Analysis for Data Science | Data Analysis | Full Crash Course | Statistics](https://youtu.be/A3fowDMo8mM?si=YSAT3EiXmaj0_KuG)
* [Finetune LLMs to teach them ANYTHING with Huggingface and Pytorch | Step-by-step tutorial](https://youtu.be/bZcKYiwtw1I?si=ACRscSPxIFIDYo4q)
* [How AI Image Generators Work (Stable Diffusion / Dall-E) - Computerphile](https://youtu.be/1CIpzeNxIhU?si=hww3I2WjpaW_-9Ci)
* [Building Effective Agents with LangGraph](https://youtu.be/aHCDrAbH_go?si=LopqifdBQ1WBLif_)

**Month 4 – NoSQL, Production, DevOps, Cloud, APIs & Professional Skills**

**Week 1: NoSQL, APIs, Web Frameworks & Front-End (Optional)**

**Topics:**

* NoSQL Basics
* FastAPI, Flask, Django Basics
* Build & Deploy ML models as APIs
* Create API Endpoints & Test with Postman
* Intro to Front-End (Optional): React.js, Angular
* Connect ML API to front-end
* Demo Project: Sentiment Analysis App with FastAPI + React

**Assignments:**

* Create a FastAPI ML API to classify text
* Test endpoints using Postman
* Deploy a Flask app on Render/Heroku
* Optional: Build a React front-end to interact with ML API

**YouTube Resources:**

* [NoSQL Database Tutorial – Full Course for Beginners](https://youtu.be/xh4gy1lbL2k?si=M7dktIHN1izPhJyK)
* [FastAPI Crash Course – Traversy Media](https://www.youtube.com/watch?v=0sOvCWFmrtA)
* [Deploy ML model as API](https://www.youtube.com/playlist?list=PLfFghEzKVmjuCMQqnONUX6N72RoJq4MSY)
* [Postman API Testing Tutorial | Postman Tutorial For Beginners 2025](https://youtu.be/CLG0ha_a0q8?si=By5RqdjJiKAJHVf3)
* [Frontend, API, Backend and Database explained](https://youtu.be/NzEYYemQ3_8?si=n-lS9s2VCLPZqAfP)

**Week 2: Cloud Platforms & Cloud Pipelines**

**Topics:**

* AWS, Azure, GCP for AI/ML
* Using S3, GCS, Azure Blob for data storage
* Basic ETL Concepts
* Automating Data Pipelines using AWS Glue, GCP Dataflow

**Assignments:**

* Upload a dataset to S3/GCS and read it in a Colab/Notebook
* Build a basic ETL job to clean and transform data
* Design a Cloud Data Pipeline for batch data ingestion
* Simulate a real-world ML pipeline with cloud storage

**YouTube Resources:**

* [AWS vs Azure vs GCP | Which one should you learn?](https://youtu.be/ECnqartfjIM?si=uPb5DlhRg6Dj5l1r)
* [Migrate AWS S3 Buckets to Azure Blob Storage using AzCopy](https://youtu.be/suyV-z5aoSY?si=tx2GY9S_eErn4rvo)
* [What is ETL | What is Data Warehouse | OLTP vs OLAP](https://youtu.be/oF_2uDb7DvQ?si=r7lxuXJJHixzCMyv)
* [How to Automate Event-based End-to-End ETL Pipeline using AWS Glue & AWS Lambda | Data Engineering](https://youtu.be/-ySaDk0Sgck?si=nBBBSGU4GKVCNgB_)

**Week 3: MLOps, CI/CD, Docker, DevOps & Spark**

**Topics:**

* MLOps Basics, CI/CD Workflows
* GitHub Actions & Jenkins Overview
* MLflow for Experiment Tracking
* Docker Basics for ML
* Intro to Kubernetes (high-level)
* DevOps for ML Workflow
* PySpark & Big Data Processing

**Assignments:**

* Set up a CI/CD pipeline with GitHub Actions
* Track experiments using MLflow
* Dockerize a Machine Learning Model
* Process large CSV data using PySpark
* Optional: Explore a basic K8s deployment conceptually

**YouTube Resources:**

* [MLOps Tutorial – Krish Naik](https://www.youtube.com/watch?v=_BE_XP5qD4U)
* [MLflow Tutorial – FreeCodeCamp](https://www.youtube.com/watch?v=ZoC8yT4g6Dk)
* [Docker for ML – Tech with Tim](https://www.youtube.com/watch?v=Gjnup-PuquQ)
* [Spark with Python – Data Arena](https://www.youtube.com/watch?v=_C8kWso4ne4)
* [CI/CD with GitHub Actions – Fireship](https://www.youtube.com/watch?v=R8_veQiYBjI)
* [Kubernetes Explained in 6 Minutes | k8s Architecture](https://youtu.be/TlHvYWVUZyc?si=JypJJoYeT5bJlCMN)
* [CI/CD Explained | How DevOps Use Pipelines for Automation](https://youtu.be/M4CXOocovZ4?si=bDO1eLONDGZljHrc)

**Week 4: Professional Skills + Resume Booster Projects**

**Topics:**

* Working with Business/Product Teams
* Aligning ML Models to Business Needs
* Communicating Insights to Non-Tech Stakeholders
* Resume Writing: Impact-focused
* Interview Preparation
* Finalize GitHub & Project Portfolio

**Assignments:**

* Present a completed ML project in a business context
* Finalize 2 Capstone Projects (NLP, CV, or LLM API-based)
* Refactor GitHub repos with proper README & documentation
* Practice mock interviews (storytelling + Q&A)
* Resume writing: Emphasize metrics and outcomes

**YouTube Resources:**

* [STAR Method for Interviews – CareerVidz](https://www.youtube.com/watch?v=K-CrEi0ymMg)

**Phase 2 - APPLY [2 Months]**

**Month 5: Real-World Projects & Application**

**Project 1: Machine Learning Project[2 Weeks]**

Title: Customer Churn Prediction  
Skills Used: Scikit-learn, Pandas, Matplotlib, SQL  
Extras: Flask API, Model Evaluation, GitHub Deployment

**Project 2: Deep Learning Project[2 Weeks]**

Title: Image Classification using CNNs  
Skills Used: TensorFlow or PyTorch, CNN, AWS/GCP (optional)  
Extras: Deploy on Streamlit or Flask + Docker

**Month 6: Real-World Projects & Application**

**Project 3: NLP Project[2 Weeks]**

Title: Resume Matcher using BERT  
Skills Used: Hugging Face Transformers, NLTK, Pandas, FastAPI  
Extras: LangChain intro, prompt tuning

**Project 4: Generative AI Project[2 Weeks]**

Title: Chatbot with OpenAI & LangChain  
Skills Used: OpenAI API, LangChain, FastAPI  
Extras: Front-end integration (optional), Docker

**Phase 3 - Placement Related [2 Months]**

**Month 6–7: Job Applications & Interview Preparation**

**Week 1:**

* Build your Data Science Resume & Portfolio (GitHub, LinkedIn, Kaggle)
* Record short project walkthrough videos (YouTube or Loom)
* Identify top job portals (Naukri, LinkedIn, Instahyre, Hirect)

**Week 2:**

* Mock Interviews (Technical + HR)
* Interview prep for:
  + Python, SQL Queries
  + ML Algorithms
  + Statistics & Probability
  + Scenario-based Business Questions

**Week 3:**

* Apply to jobs daily (target 10–15 quality applications/day)
* Attend webinars, workshops, AI bootcamps
* Network on LinkedIn, reach out to recruiters

**Week 4 (and beyond):**

* Track interview feedback and revise weak areas
* Learn from rejections & keep refining
* Optionally add one more small project (like a Time Series Forecasting or MLOps Pipeline)

**Topics to Upskill**

* Excel
* PowerBI
* Tableau

**Want 1:1 Guidance to Master these skills in 4 Months?**

<https://docs.google.com/forms/d/e/1FAIpQLSd6tvfZYeDnxbC33TtBau-FOqGaO1ksu_l3HxO-XnSYn4_sWg/viewform>

Regards,

**John-The AI Coach.**

**CEO & Founder of ProITbridge.**

Youtube: <https://www.youtube.com/@aicoachjohn>

LinkedIN:<https://www.linkedin.com/in/aicoachjohn/>

Instagram: <https://www.instagram.com/john_the_ai_coach/>

Website: <https://www.proitbridge.com/>

*Your journey starts today. Not tomorrow💪🔥*