

Data Science & Machine Learning

Duration	Live classes	Schedule	Batch size
12 weeks	120 hours	Mon to Fri, 2 hrs/day	10 students max

Tools you will use

 Python + SQL

 Pandas + NumPy

 Power BI (DAX + Power Query)

 Streamlit + Plotly

 Scikit-learn

10 students a batch. Exceptional mentors who guide you personally, and a **syllabus built from real job posts**, cuz we never skip the "hard parts" 😊

The class

10 per batch

Ten students. Never more.

A hard cap, not a target. Your mentor knows your name and your work.

The mentor

4–5 yrs in the job

Someone who does the job.

Working professionals, not full-time teachers. They read your work line by line.

The syllabus

1/3 is fundamentals

The hard parts, not just the basics.

Teaches everything the job needs and never skips the hard parts most courses skip, built from real job descriptions. Check the syllabus before you enroll.

Built into every course

Gate assessment

A short test before Day 1 finds your strengths and pace, then sets your plan.

A real, live project

Every course ends with a deployed project you can show employers.

Revision every 5th class

Every fifth class goes back over the week, so nothing piles up.

1:1 mock interview

A real interview practice with your mentor in the final week.

Certification exam

Pass the exam at the end of the course to earn your certificate.

CV and LinkedIn help

We help you rewrite your CV and LinkedIn before you apply.

Data Science & Machine Learning

Why This Course

Employers want analysts who can clean data, write SQL (the language for asking questions of a database), use statistics and explain results simply. This course teaches exactly that, in the order real analysts learn it.

- **You learn what employers actually want.** That is honest analysis, not just a list of tool names on your CV.
- **You learn the basics first.** Python comes before its data tools, SQL comes before models, and machine learning (teaching computers to predict from data) comes only at the end.
- **You spend a full week on Power BI.** It is the dashboard tool named most often in Kathmandu job posts.
- **You check everything AI writes.** Every query or formula drafted by AI is read and corrected by you.
- **You finish with work you can show and explain.** Your portfolio can be re-run by anyone, from the first step to the finished project.
- **We do not promise a salary.** Your pay depends on your skills, your portfolio and your local or remote clients.

Who Is This Course For

- **Complete beginners** who want a real path into data jobs.
- **People who work in Excel all day** in finance, operations or marketing and want to move to Python.
- **Fresh graduates** from any subject who want an analyst job.
- **Managers and decision makers** who want to check claims with data.
- **CS or IT students** who want practical industry work, not only theory.
- **Career changers** aiming for an analyst or junior data scientist job.

Course Snapshot

PARAMETER	DETAILS
Course Code	TR-02
Duration	12 Weeks (3 Months)

PARAMETER	DETAILS
Schedule	Monday to Friday, 2 Hours/Day
Total Hours	120 Hours
Batch Size	Maximum 10 Students
Course Fee	NPR 36,750
Level	Beginner-friendly, taught from the fundamentals (light pre-course prep listed below) → job-ready Data / BI Analyst
Outcome	Job-ready Data Analyst / Power BI Analyst (with the foundations to grow into a Junior Data Scientist role)

Prerequisites

- You do not need to know Python, statistics or maths. All three start from zero in Week 1.
- You need a laptop with at least 8GB of RAM.
- You should be comfortable using a computer, files and folders.
- You need basic English reading.
- You take the Saarathi Gate Assessment before Day 1. It is not a pass or fail test.

You do not need to install anything beforehand. We set up Python, VS Code and Jupyter together in the first class.

Optional warm-up. It is not required, but it makes the first weeks easier:

- Do the first few lessons of freeCodeCamp Scientific Computing with Python, free at [freecodecamp.org/learn](https://www.freecodecamp.org/learn).
- Do Khan Academy Statistics and Probability units 1-2, and Algebra Basics (variables and equations).
- Practise these 10 Google Sheets formulas: SUM, AVERAGE, IF, VLOOKUP, COUNTIF, SUMIF, MIN, MAX, SORT and FILTER.

Assessments at a glance

- **Gate Assessment.** You take it before Day 1. It is not pass or fail. It shows us how you learn and how fast.
- **Certification Exam.** It lasts 2 hours and has around 30 questions of different types, including code and short real-world cases.

- **Pass mark.** You need 70 percent. You must pass to get the Certificate of Completion.
- **Exam rules.** The exam runs in fullscreen, and we record if you switch tabs or apps or paste text. After three warnings, the exam ends.
- **Answer order.** You can answer in any order, but every question needs an answer before you submit.
- **Projects.** Your mentor gives you 2 projects during the course. They need to be approved before you get your certificate and before the Career Launchpad opens.
- **Recommendation letters.** We may give one to learners who pass, but it is never guaranteed.

Your Learning Week

DAY	ACTIVITY
Mon–Fri	2-hour live class with your mentor, where you practise on real projects
Mon–Fri	2 hours of self-study at home for projects, practice and fixing errors
Sunday	Open classroom for extra practice, or a day of rest
Day 5 (every week)	The last hour of class is for revision and clearing doubts

Class gives structure. Practice between classes gives depth.

Week-by-Week Curriculum

Phase 1 · Analyst Foundations & Data Handling (Weeks 1–4, 40 Hours)

The first four weeks (a third of the course) are pure base-building. Python alone gets a deliberate **1.5 weeks** (Week 1 + the first half of Week 2) before any data library, so the foundation actually holds.

WEEK	FOCUS AREA	TOPICS
Week 1	Python Fundamentals I: Setup & Core Language	– What data roles actually do (analyst, scientist, engineer) and where you fit – Setup from zero: Python 3.13, Jupyter (run code cell by cell), VS Code, Git, a virtual environment and a first GitHub commit – Variables, numbers, text and operators – Decisions (<code>if / elif / else</code>) and loops (<code>for</code> , <code>while</code>) – Functions: parameters, return values and scope – Containers: lists (ordered), dictionaries (labelled values), tuples (fixed), sets

WEEK	FOCUS AREA	TOPICS
		- Spreadsheet-to-Python bridge: SUM, IF, VLOOKUP, COUNTIF and their code equivalents - Notebook hygiene and AI intro together (one talk-segment block, room to settle in): markdown cells, clean cell order, Restart-and-Run-All, debugging habits, then ChatGPT and Copilot as a rubber duck, never run unread
Week 2	Python Fundamentals II + NumPy	<i>Python (first half, completing 1.5 weeks):</i> - List and dictionary comprehensions to fluency - Full File I/O with the standard library: <code>open()</code> , context managers, <code>pathlib</code> , <code>csv.DictReader</code> - Deeper error handling: <code>try / except / else / finally</code> , reading tracebacks - Functions deeper, modules and debugging together (extra hands-on time): default args, <code>*args / **kwargs</code> , reusable modules and imports, then print vs breakpoint and isolating a bug - Pure-Python CSV mini-analyzer mini-project (read, clean, summarise, no library) <i>NumPy (back half):</i> - NumPy arrays: vectorisation (fast column math), boolean masks, one broadcasting example - Why arrays beat Python lists for numbers (speed/memory), bridging into Pandas
Week 3	Pandas & AI-Assisted Data Cleaning	- Pandas Series and DataFrames (<code>.loc</code> vs <code>.iloc</code> , dtype awareness) - Reading CSV and Excel (JSON, Parquet awareness) - Analyst Excel: pivot tables, XLOOKUP/SUMIFS, basic charts, mirrored in <code>pivot_table / groupby</code> - Filtering, sorting, groupby, aggregation - Missing values, duplicates, type conversion - String and date cleaning (<code>.str</code> , <code>to_datetime</code> , timezone awareness) - Merging and reshaping together: <code>merge</code> , <code>concat</code> , picking join keys, then <code>melt</code> (wide to long) and <code>pivot</code> (long to wide) - End-to-end cleaning on a curated messy dataset, with AI as a partner: Claude to explore datasets, Copilot for Pandas code, always verified

WEEK	FOCUS AREA	TOPICS
Week 4	SQL, Window Functions & Python-SQL Integration	- Relational thinking, keys, normalisation and PostgreSQL setup - Single-table reads built up slowly in one block: SELECT, WHERE, ORDER BY, LIMIT, then summaries with GROUP BY, HAVING, aggregates, <code>date_trunc</code>, CASE bucketing - INNER JOIN and LEFT JOIN to fluency (RIGHT/FULL/self/anti awareness) - CTEs (WITH, named query stages) to fluency, a core SQL-screen skill beside joins and windows; correlated/nested subqueries awareness - Three core window functions to fluency: ROW_NUMBER, RANK, LAG (deeper fluency in the Week 12 drill) - Python-SQL integration: <code>read_sql</code>, SQLAlchemy, parameterised queries, avoiding SQL injection - Multi-source ingest: the same question from CSV, Excel, JSON API and Postgres, then reconciled

Phase 2 · Statistics, EDA & Decision Framing (Weeks 5–7, 30 Hours)

WEEK	FOCUS AREA	TOPICS
Week 5	Descriptive Statistics & Distribution Thinking	- A written data-check list run before any analysis starts: missing values, duplicates, wrong types, a row count that does not match the source - Summary numbers: mean, median, mode, variance, standard deviation, percentiles - Distribution shape: histograms, box plots, one KDE, skewness (violin/ECDF awareness) - Outlier detection (z-scores, IQR) and handling - Bivariate relationships and correlation-vs-causation together: scatterplots, relationship thinking, and why "moves together" never proves "causes" (confounders, reverse causation, selection bias) - Probability and sampling together: conditional/Bayes intuition, then sampling and bias - Central Limit Theorem intuition via simulation - Structured EDA summary habits: write down what you checked and found

WEEK	FOCUS AREA	TOPICS
Week 6	Probability, Sampling & Hypothesis Testing	- Confidence intervals: a defensible range, not false certainty (bootstrap awareness) - 2-sample t-test to fluency: real difference or chance? - Chi-square test to fluency: the same question for categories - p-values and Type I/II errors (multiple-comparison/Bonferroni awareness) - Effect size and practical significance - Mann-Whitney / ANOVA / Bayesian testing (awareness) - Checking AI-drafted statistics: re-run the numbers yourself, catch a seeded wrong join or wrong aggregation before it reaches the memo
Week 7	EDA, Visual Storytelling & Streamlit Dashboards	- Chart choice framework: comparison, distribution, relationship, composition - Seaborn to fluency (Matplotlib fine-control awareness) - One interactive Plotly chart (choropleth awareness) - North-Star KPI with leading/lagging indicators - A/B testing fundamentals and experiment framing - Streamlit: notebook to shareable dashboard with filters and caching (deploys in Week 11) - 10-minute filter-context primer (a slicer narrows what a number sums over), priming Week 8 DAX - Power BI vs Streamlit decision matrix - Analysis narrative and recommendations with honest uncertainty, drafted with AI for EDA: ChatGPT as a rubber duck, when to trust and when to push back

Phase 3 · Business Intelligence & Starter Machine Learning (Weeks 8–10, 30 Hours)

WEEK	FOCUS AREA	TOPICS
Week 8	Power BI: End-to-End Business Intelligence	- Power Query / M for ETL: profile, clean, fix types, pivot/unpivot, merge vs append, fact + dimension tables - Star-schema modelling: relationships, cross-filter direction, marked Date table, role-playing dimensions - DAX foundations: a precise plain-English metric definition first (what exactly counts as an "active user" or a "conversion"), then one slow context-transition build ("CALCULATE turns the current row into a filter"), measures vs calculated columns - DAX to fluency: CALCULATE with ALL/ALLEXCEPT, SUMX, VAR/RETURN, time intelligence (TOTALYTD,

WEEK	FOCUS AREA	TOPICS
		SAMEPERIODLASTYEAR) - DAX awareness: calculation groups, visual calculations, RANKX/AVERAGEX, REMOVEFILTERS/KEEPFILTERS, quick measures - Report design: slicers, drillthrough, bookmarks, tooltips, mobile, accessibility - Power BI Service: publish, scheduled refresh, gateway, static Row-Level Security; beginners ship Import mode (DirectQuery/Direct Lake, dynamic RLS awareness); Copilot in Power BI (GA DAX generation, report-grounded chat) for drafting measures, every one read and corrected, never run unread - Revision + one published end-to-end report, capstone mentor 1:1
Week 9	ML Workflow, Regression & Classification	- ML lifecycle: a precise written definition of the target (what exactly counts as "churned"), then split, prepare, train, evaluate; capstone topic signed off in the Week-8 Day-5 1:1 - Leakage prevention, scaling, encoding, scikit-learn Pipeline - Linear regression (predict a number) + RMSE, MAE, R² - Logistic regression (predict yes/no) + confusion matrix, precision, recall, F1 to fluency (ROC-AUC/PR-AUC awareness) - Cross-validation: k-fold to fluency (stratified/time-series awareness) - AI-drafted pipelines checked line by line: re-run every number, catch a seeded wrong join or aggregation
Week 10	Feature Engineering, Clustering & Model Explanation	- Tree models and class-imbalance together: decision trees and random forests + class-imbalance handling - Feature engineering: ratios, bins, date parts, interactions - Feature selection and leakage checks - K-Means clustering to fluency (elbow, silhouette, scaling); hierarchical/dendrograms awareness - Model explanation: SHAP global summary + one force plot + one PDP (permutation importance, gini pitfall awareness) - One-page manager memo, the headline answer first, never a chart dump, plus honest limitations

Phase 4 · Validation, Capstone & Career (Weeks 11–12, 20 Hours)

WEEK	FOCUS AREA	TOPICS
Week 11	Data Validation, Reproducibility & Capstone Sprint	– Day 1, lock: problem, dataset, success metric and a "do-nothing" baseline – Data validation and pytest together (one capstone-discipline block): completeness, consistency, freshness, uniqueness (<code>pandera</code> / <code>assert</code>), then fixtures and shape/dtype/null-rate assertions – Reproducible structure: <code>data/</code> , <code>src/</code> , <code>notebooks/</code> , pinned <code>requirements.txt</code> , fixed <code>random_state</code> , <code>.env</code> – Lightweight ETL thinking and delivery choice: notebook vs dashboard vs endpoint – Build days together: clean, EDA, model or experiment, then the dashboard or report; ship, peer review, demo rehearsal and final deliverables – MLOps awareness pointer plus a "what's next / lives in TR-10" note (time series, NLP, deep learning, Spark, all awareness)
Week 12	Portfolio, SQL Interviews & Career Launch	– Resume and LinkedIn for data roles – GitHub portfolio and README polish – SQL live-test drills: deep window-function fluency (LEAD, SUM OVER running totals, moving averages, partitioned ranking), timed, AI allowed, you explain every line out loud – Mock analyst + case interview (a real business question worked through to a recommendation) and a data scientist interview – Job search planning and application strategy – Demo day presentation

Skills You Will Gain

CATEGORY	TECHNOLOGIES
Languages	Python 3, SQL (PostgreSQL)
Data Handling	pandas, NumPy, data cleaning, Excel (pivot tables, XLOOKUP/SUMIFS, charts), ETL pipeline basics
Visualisation	Matplotlib, Seaborn, Plotly

CATEGORY	TECHNOLOGIES
Business Intelligence	Power BI: Power Query (M), star-schema data modelling, DAX, time intelligence, Row-Level Security, Power BI Service
Dashboards	Power BI, Streamlit
Machine Learning	scikit-learn (regression, classification, clustering, cross-validation)
Testing	pytest for data validation
Tools	Jupyter, Git, GitHub, VS Code
AI Assistants	GitHub Copilot, ChatGPT, Claude

Topic Depth and Awareness

Depth means you practise it until you can use it confidently. **Awareness** means we introduce it clearly so you can read about it and compare it later.

FOCUS AREA	LEARNING FOCUS
Python, Pandas, NumPy, SQL	Depth
Analyst Excel (pivot tables, XLOOKUP/SUMIFS, basic charts)	Depth
SQL INNER/LEFT joins, three core window functions (ROW_NUMBER, RANK, LAG), Python-SQL integration	Depth
Deep SQL window-function fluency (LEAD, SUM OVER, running totals, partitioned ranking)	Depth (Week 12 interview drill)
Descriptive stats, 2-sample t-test, chi-square, CLT intuition	Depth
Seaborn, one Plotly chart, local Streamlit	Depth
Matplotlib fine control, Plotly choropleth	Awareness
Power BI: Power Query, data modelling, DAX (context transition, CALCULATE, SUMX, time intelligence), static RLS, publishing in Import mode	Depth

FOCUS AREA	LEARNING FOCUS
Regression, linear/logistic models, confusion matrix + precision/recall/F1, k-fold CV, trees/forests, K-Means, feature engineering, SHAP	Depth
Pytest and reproducible project structure	Depth
Analysis narrative and stakeholder communication	Depth
Power BI storage-mode choice (Import vs DirectQuery vs Direct Lake; beginners ship Import), dynamic RLS, calculation groups, visual calculations, RANKX/AVERAGEX, paginated reports	Awareness
ROC-AUC/PR-AUC (read the chart), stratified/time-series CV, hierarchical clustering, permutation importance	Awareness
Tableau as a resume hedge	Awareness
TikTok Analytics, social platform APIs	Awareness
ETL pipeline tooling (Airflow, Prefect)	Awareness
Time-series forecasting (Prophet, ARIMA)	Awareness (covered in depth in TR-10)
Deep learning, TensorFlow, PyTorch	Awareness (covered in depth in TR-10)
Big Data tools (Spark)	Awareness
MLOps and model deployment	Awareness (covered further in TR-10)

Project Pool

Each student picks **one** final project (capstone) with their mentor at the end of Week 8. The list changes for each batch:

#	PROJECT	STACK FOCUS
1	Customer churn predictor (telecom / SaaS)	Python, SQL, scikit-learn
2	Banking / fintech Power BI dashboard	SQL, Power Query, data modelling, DAX, RLS
3	E-commerce funnel analysis	SQL, window functions, Power BI / visualisation
4	NGO impact tracker	Python, Pandas, Streamlit
5	Healthcare outcomes study	Python, statistics, SHAP
6	Trainer-scoped client brief	Whatever the brief requires

Career Paths & Trajectory

ROLE	TIMELINE	FOCUS
Data Analyst / Power BI Analyst	0–2 yrs	SQL, Power BI, cleaning, stakeholder explaining
BI Analyst / Analytics Associate	1–3 yrs	Power BI dashboards, data modelling, metric design, storytelling
Business Analyst	1–3 yrs	Requirements gathering, process mapping, data-driven decisions
Junior Data Scientist / Product Analyst	2–4 yrs	Problem framing, modelling, evaluation honesty
Senior Data Analyst / Data Scientist	4+ yrs	Judgment, reproducibility, high-trust analysis

FAQ

What does Data Science & Machine Learning cost and how do I pay?

NPR 36,750. Pay by cash at our Old Baneshwor campus, bank transfer, or Fonepay. Email info@saarathiacademy.com.np for the payment plan.

Do I need prior experience?

No. The Prerequisites section lists what you need. The Gate Assessment before Day 1 is not a pass or fail test. If you are not yet at that level, we suggest a short foundation course first.

How long is the course?

It runs for 12 weeks (3 months). You build your final project in Week 11. Practice interviews and a demo day happen in Week 12.

What tools will I use?

You use **Python, SQL (PostgreSQL), pandas, Power BI (Power Query, DAX, data modelling), Streamlit and scikit-learn**. Each part of the course ends with a public project online that you can show employers.

What roles does this prepare me for?

The Career Paths section lists them. We do not promise jobs. We help you build your portfolio, practise interviews, improve your CV and LinkedIn, and we refer you to companies in our hiring network.

How does the batch and mentorship work?

- Each batch has at most 10 students, and each student gets direct time with the mentor and project reviews.
- Classes are in person at Old Baneshwor. Recordings and online office hours are also available.
- The Sunday Open Classroom is free for enrolled students.

Enrollment

- **Next Batch:** see the live course page
 - **Location:** Old Baneshwor Chowk, Kathmandu, Nepal
 - **Mode:** Online + Offline
 - **Contact:** 9744442469, 9761095364
 - Limited to 10 seats per batch
-

Data Science & Machine Learning — full details, fees and next batch dates:
saarathiacademy.com.np/data-science-and-machine-learning-course-in-nepal