



---

# Data Science Xcellerator Program

## 21-Day Data Analytics, Data Science & AI Tools Bootcamp

**University-Style Syllabus & Program Handbook**

**Project-Based | Milestone-Oriented | AI-Enabled | Industry-Ready**

*Transform Data Into Decisions*

---

**Organized By:** SIN School of AI

**Duration:** 21 Days Intensive

**Program Type:** Data Analytics, Data Science & AI Tools

**Program Brand:** Data Science Xcellerator Program

**Learning Model:** 65% Practical | 20% Theory | 15% AI Workflow & Review

**Outcomes:** Portfolio, Dashboards, Models, GitHub, and Capstone Defense

**Document Purpose:** This handbook presents the complete upgraded syllabus, AI tool stack, challenge framework, scoring model, and capstone structure for the Data Science Xcellerator Program in a formal, attractive university-style format.

## Program Positioning

The Data Science Xcellerator Program is redesigned as a 21-day intensive, project-based Data Analytics and Data Science program with a dedicated AI tools layer. The objective is not only to teach Python, SQL, statistics, dashboards, and machine learning, but to make every student build visible proof of competence through GitHub repositories, AI-assisted analytics workflows, SQL case studies, BI dashboards, EDA reports, predictive models, executive decks, and a final capstone defense.

**Important Rule:** AI tools are allowed for productivity, but students must verify every AI-generated output. Blind copy-paste from AI tools is treated as a workflow failure, not a skill.

### Learning Model:

- **65% Practical Implementation:** live labs, datasets, dashboards, models, and repositories.
- **20% Conceptual Theory:** statistics, SQL logic, BI architecture, ML workflow, and metric governance.
- **15% AI Workflow, Business Communication & Review:** AI-assisted debugging, insight generation, documentation, portfolio, and viva.

## Final Student Output

By the end of the bootcamp, every student should have:

- 1 Python data processing project
- 1 SQL business analytics case study
- 1 ETL and data quality pipeline
- 1 AI-assisted EDA notebook
- 1 statistical A/B testing report
- 1 predictive modeling notebook
- 1 Power BI or Tableau dashboard
- 1 KPI dictionary and metric governance sheet
- 1 GitHub portfolio repository
- 1 model card with limitations
- 1 AI usage audit sheet
- 1 final capstone project
- 1 executive presentation deck
- 1 LinkedIn/GitHub proof-of-work post draft

## AI-Based Data Science Tool Stack

Tool Category	Tools / Platforms	How Students Use Them
AI Coding Assistants	GitHub Copilot, Cursor AI, Codeium / Windsurf, ChatGPT, Gemini, Claude	Generate draft code, debug Python errors, explain SQL logic, refactor notebooks, and write README files. All code must be tested manually.

AI for Data Analysis	ChatGPT Advanced Data Analysis, Gemini, Claude, Julius AI, NotebookLM	Explore datasets, draft analysis plans, summarize findings, generate hypotheses, and create insight narratives.
BI AI Assistants	Microsoft Power BI Copilot, Tableau Pulse / Tableau AI, Microsoft Fabric Copilot	Ask questions from semantic models, summarize dashboards, draft report pages, and explain KPI movement.
AutoML & Low-Code ML	PyCaret, AutoGluon, H2O AutoML, DataRobot, Azure AutoML, Vertex AI AutoML	Build baseline models quickly, compare algorithms, evaluate metrics, and learn model selection discipline.
ML Lifecycle & Experiment Tracking	MLflow, Weights & Biases, DVC	Track experiments, model versions, parameters, metrics, model cards, and reproducibility.
Data Quality & Monitoring	Great Expectations, Pandera, Evidently AI	Validate datasets, check schema drift, monitor model/data quality, and produce data quality reports.
LLM & NLP Tools	Hugging Face Transformers, LangChain, LlamaIndex	Build small text analytics, sentiment analysis, document Q&A, and retrieval-based analytics prototypes.
Presentation & Documentation AI	Canva AI, Gamma, Tome, ChatGPT, Notion AI	Create executive summaries, presentation outlines, project reports, and stakeholder-ready narratives.

### AI Ethics and Verification Policy:

- Every AI-generated SQL query must be executed and validated.
- Every AI-generated Python function must be tested on sample data.
- Every AI-generated chart recommendation must be checked against the business question.
- Every AI-written insight must be backed by actual numbers from the dataset.
- Every final project must include an **AI Usage Audit Sheet**: tool used, prompt used, output received, manual verification performed, and final correction made.

## 21-Day Milestone Roadmap

Phase	Days	Milestone	Final Output
Phase 1	Day 1-3	Python, NumPy, Pandas, GitHub, and AI-Assisted Coding Foundation	Cleaned dataset, Python scripts, GitHub repo, AI debugging log

Phase 2	Day 4-5	SQL and AI-Assisted Business Querying	SQL case study with joins, CTEs, window functions, and AI query validation notes
Phase 3	Day 6-7	Data Cleaning, ETL, and Quality Validation	Automated cleaning pipeline, Great Expectations/Pandera-style checks, data quality report
Phase 4	Day 8-10	Visualization, Dashboarding, and AI-Powered Visual Storytelling	Static chart pack, Plotly dashboard, Power BI/Tableau dashboard prototype
Phase 5	Day 11-14	Statistics, EDA, Hypothesis Testing, AutoML, and Predictive Modeling	AI-assisted EDA notebook, A/B testing report, ML model, model card
Phase 6	Day 15-18	BI, Semantic Metrics, Cloud Analytics, and AI BI Assistants	Power BI/Tableau dashboard, KPI dictionary, metric governance sheet, cloud architecture diagram
Phase 7	Day 19-20	GitHub, Kaggle, MLflow, Portfolio, and Grand Finale Build	Public project repo, README, reproducible code, experiment log, final challenge submission
Phase 8	Day 21	Capstone Defense, AI Audit, and Career Readiness	Live demo, executive deck, technical viva, AI usage audit, final grading

## Detailed 21-Day Curriculum

Day	Focus Area	Practical Work	Deliverable	AI Tool Layer
1	Orientation, Analytics Thinking, GitHub Setup	Program briefing, tool setup, GitHub setup, project folder structure, first notebook, role mapping	Workspace and GitHub repo created	ChatGPT / Gemini for learning plan and setup checklist

2	Python Core for Data Work	Variables, loops, functions, list comprehensions, file handling, reusable utility functions	Python utility script	GitHub Copilot / Cursor for code drafting and debugging
3	NumPy and Pandas Data Engineering	Arrays, vectorization, DataFrames, filtering, grouping, merging, missing values, type conversion	Cleaned Pandas notebook	AI-assisted code explanation and error debugging
4	SQL Fundamentals for Business Analytics	SELECT, WHERE, GROUP BY, HAVING, joins, subqueries, CASE WHEN, KPI extraction	SQL file with 10 business questions solved	AI-generated SQL reviewed manually
5	Advanced SQL and Window Functions	CTEs, DENSE_RANK, LAG, LEAD, rolling averages, cohort analysis, retention metrics	Funnel and cohort SQL case study	AI query optimization suggestions
6	Challenge 1: Raw Data Cleaning Blitz	Clean a corrupted dataset using automated scripts only	Solo challenge submission	AI allowed only for debugging, not final logic copying
7	ETL Pipeline and Data Quality Testing	Build ETL workflow, null checks, duplicate checks, range checks, freshness checks	ETL pipeline and data quality report	Great Expectations / Pandera-style validation
8	Static Data Visualization	Matplotlib and Seaborn charts, chart selection, data-to-ink ratio, business formatting	Executive chart pack	AI-assisted chart recommendation, manually verified
9	Interactive Visualization with Plotly	Interactive line charts, filters, hover data, drill-down charts, layout planning	Interactive HTML dashboard prototype	AI for dashboard layout critique
10	Challenge 2: Visual Storytelling Dashboard Battle	Teams build dashboard from a complex dataset and present insights	Pair dashboard submission	Power BI Copilot / Tableau AI / ChatGPT for narrative drafting

11	Statistics and EDA Foundation	Descriptive statistics, distributions, skewness, kurtosis, correlation, heatmaps, outlier interpretation	EDA notebook	ChatGPT / Claude for hypothesis generation
12	Hypothesis Testing and A/B Testing	Null and alternate hypotheses, p-values, t-tests, ANOVA, confidence intervals	A/B testing business report	AI-assisted test selection checklist
13	Predictive Modeling Foundation	Train-test split, linear regression, logistic regression, classification and regression metrics	Predictive model notebook	PyCaret / AutoGluon for baseline AutoML comparison
14	Model Validation and ML Lifecycle Basics	Residual analysis, overfitting, feature importance, model cards, limitations, reproducibility	Model card and validated ML notebook	MLflow / W&B / Evidently AI awareness
15	Challenge 3: Algorithmic Business Analytics Hackathon	Teams solve revenue drop or churn case using SQL, statistics, and strategy	SQL + notebook + 5-slide deck	AI allowed for deck polishing, not insight invention
16	Power BI Enterprise Dashboarding	Power Query, data modeling, star schema, relationships, DAX, KPI cards, slicers	Power BI dashboard v1	Power BI Copilot for report summaries
17	Tableau and Executive Storytelling	Dimensions/measures, parameters, dual-axis visuals, story points, dashboard actions	Tableau workbook or BI story	Tableau Pulse / Tableau AI for insight prompts
18	Industry Update Day: Cloud Analytics, dbt, Semantic Layer	Cloud analytics architecture, dbt-style modular SQL, metric dictionary, KPI governance	Architecture diagram and KPI dictionary	Microsoft Fabric Copilot / dbt concepts / semantic model prompts
19	GitHub, Kaggle, README, and Portfolio Build	Clean repo structure, README, screenshots, summary, Kaggle notebook publishing, LinkedIn draft	Portfolio-ready GitHub repository	AI-assisted README and LinkedIn draft

20	Challenge 4: Xcellerator Data Wars Grand Finale	Teams build predictive pipelines, submit outputs to leaderboard, document reproducible workflow	Model, repo, predictions, defense notes	AutoML baseline, MLflow experiment log, AI audit
21	Final Capstone Defense and Career Review	Live demo, technical viva, dashboard walkthrough, business recommendation, portfolio review	Final capstone and certificate eligibility	AI Usage Audit Sheet submitted

## Challenge Structure: Solo vs Team

**Blunt Rule:** Not every challenge should be team-based. The first technical challenge must stay solo. If everything is team-based, weak students can hide and strong students carry the output. Team challenges should begin only after individual fundamentals are tested.

Challenge	Mode	Total Points	Reason for Mode
Challenge 1: Raw Data Cleaning Blitz	Solo	100	Tests individual Python, Pandas, and cleaning logic. No hiding inside teams.
Challenge 2: Visual Storytelling Dashboard Battle	Team of 2	100	Dashboard work benefits from design and analytical collaboration without creating free riders.
Challenge 3: Algorithmic Business Analytics Hackathon	Team of 3-4	150	Business analytics requires SQL, statistics, strategy, and presentation. Squad format reflects real consulting work.
Challenge 4: Xcellerator Data Wars Grand Finale	Team of 3	200	Predictive modeling requires feature engineering, validation, documentation, and technical defense.
Final Capstone Defense	Hybrid: Team Product + Individual Viva	300	Students may build as a team, but each student must individually defend their work, code, dashboard, AI usage, and recommendation.

## Challenge 1: Raw Data Cleaning Blitz

**Mode:** Solo Individual Sprint

**Objective:** Program an automated cleaning engine that repairs a corrupted enterprise dataset under strict time constraints.

**AI Rule:** AI can be used for syntax debugging only. Final cleaning logic must be explained by the student.

**Deliverables:**

- Python cleaning script
- Cleaned dataset output
- Short data quality report
- Before-after comparison table
- AI usage log, if any AI tool was used

**Scoring: 100 Points**

<b>Evaluation Area</b>	<b>Points</b>
Missing value handling accuracy	18
Duplicate detection and removal	12
String cleaning and standardization	12
Outlier handling logic	12
Automation and reusable code structure	14
Data quality validation checks	12
Execution speed and script reliability	10
Documentation and final explanation	10

**Challenge 2: Visual Storytelling Dashboard Battle****Mode: Team of 2**

**Objective:** Convert a complex dataset into a clear, executive-level dashboard that explains the business problem and recommends action.

**AI Rule:** AI can help draft the dashboard story, but every insight must be backed by actual dashboard numbers.

**Deliverables:**

- Power BI, Tableau, or Plotly dashboard
- 1-page insight summary
- Dashboard screenshots
- 3-minute presentation
- AI-assisted insight verification sheet

**Scoring: 100 Points**

<b>Evaluation Area</b>	<b>Points</b>
Dashboard layout and visual clarity	18
Correct KPI selection	14
Depth of insights	18
Interactivity and filter design	14
Data storytelling and business narrative	14
AI-assisted summary accuracy and verification	10
Presentation quality	8
Team coordination	4

## Challenge 3: Algorithmic Business Analytics Hackathon

### Mode: Team of 3-4

**Objective:** Solve a simulated corporate revenue drop or churn problem using SQL, statistics, business reasoning, and executive communication.

**AI Rule:** AI can be used to polish deck language and generate analysis checklists. It cannot invent conclusions that are not proven by data.

### Deliverables:

- Optimized SQL query file
- Statistical analysis notebook
- 5-slide executive recommendation deck
- Business impact estimate
- AI usage and verification log

### Scoring: 150 Points

Evaluation Area	Points
SQL query correctness and optimization	22
Use of CTEs, joins, and window functions	18
Statistical reasoning and hypothesis testing	22
Business problem diagnosis	22
Quality of recommendations	22
Executive deck clarity	14
AI-assisted workflow transparency	10
Team defense and Q&A	20

## Challenge 4: Xcellerator Data Wars Grand Finale

### Mode: Team of 3

**Objective:** Build a predictive model from scratch, submit predictions, document the workflow, track experiments, and defend the solution before reviewers.

**AI Rule:** AutoML can be used only as a baseline. The team must explain manual feature engineering, model selection, validation, and limitations.

### Deliverables:

- Clean notebook or Python pipeline
- Prediction file
- GitHub repository
- MLflow or experiment tracking log
- Model card
- Technical defense notes
- AI usage audit sheet

### Scoring: 200 Points

Evaluation Area	Points
-----------------	--------

Model accuracy and leaderboard performance	35
Feature engineering quality	25
Validation strategy and metric interpretation	25
Code reproducibility	25
Experiment tracking and model versioning	20
Documentation and README quality	20
Model card and limitation analysis	20
Final technical defense	30

## Final Capstone Evaluation

### Mode: Hybrid - Team Product + Individual Viva

The final capstone is built as a team data product, but each student is evaluated individually through viva, code explanation, dashboard explanation, AI usage explanation, and contribution defense.

### Scoring: 300 Points

Evaluation Area	Points
Data ingestion and pipeline structure	30
Data cleaning and quality validation	30
EDA and statistical insight quality	35
Predictive model and metric interpretation	40
BI dashboard and KPI design	40
Business recommendation and impact estimate	35
GitHub repository, README, and reproducibility	25
AI usage audit, responsible AI workflow, and verification	25
Final presentation and technical viva	40

## Total Scoring System

Component	Points
Challenge 1: Raw Data Cleaning Blitz	100
Challenge 2: Visual Storytelling Dashboard Battle	100
Challenge 3: Algorithmic Business Analytics Hackathon	150
Challenge 4: Xcellerator Data Wars Grand Finale	200
Final Capstone Defense	300
<b>Total</b>	<b>850 Points</b>

## Top 12 Industry-Standard Updates Added

1. **AI-Assisted Analytics Workflow:** Students learn responsible AI use for SQL drafting, Python debugging, dashboard explanation, and insight summarization.
2. **AI Usage Audit Sheet:** Every final project documents prompts, outputs, verification steps, and corrections.
3. **Data Quality Testing:** Students create validation checks for nulls, duplicates, ranges, schema changes, and freshness.
4. **dbt-Style Analytics Engineering Awareness:** Students learn modular SQL thinking, documentation, lineage, and testing concepts.
5. **Cloud Analytics Architecture:** Students understand source-to-dashboard architecture using ingestion, transformation, warehouse/lakehouse, and BI layers.
6. **Semantic Layer and KPI Governance:** Students define business metrics clearly using a KPI dictionary.
7. **Executive Dashboard UX:** Students learn dashboard layout, filter logic, KPI hierarchy, and visual storytelling.
8. **AutoML Baseline Modeling:** Students compare manual models against PyCaret, AutoGluon, H2O AutoML, or similar tools.
9. **ML Lifecycle Basics:** Students create model cards, document limitations, and track experiments through MLflow-style workflows.
10. **LLM and NLP Exposure:** Students get basic exposure to Hugging Face, LangChain, and document/text analytics workflows.
11. **GitHub Reproducibility Standards:** Every project includes README, requirements, folder structure, screenshots, and run instructions.
12. **Capstone Defense and Viva:** Students must defend their own contribution instead of only submitting group work.

## Certification Eligibility

Students become eligible for the official program completion certificate only if they meet the following requirements:

- Minimum 90% attendance.
- Completion of all core assignments.
- Submission of all challenge deliverables.
- Public GitHub repository submission.
- AI usage audit sheet submission.
- Final capstone presentation.
- Individual viva clearance.

## Final Program Outcome

By the end of 21 days, students will not simply claim that they learned Data Analytics and Data Science. They will have a complete proof-of-work portfolio containing scripts, notebooks,

dashboards, business reports, machine learning models, AI usage records, GitHub documentation, and a final capstone project that can be shown to recruiters, institutions, and industry mentors.

## Reference Notes for AI Tool Layer

- Microsoft Power BI Copilot documentation: <https://learn.microsoft.com/en-us/power-bi/create-reports/copilot-introduction>
- GitHub Copilot documentation: <https://docs.github.com/copilot>
- MLflow documentation: <https://mlflow.org/docs/latest/>
- Hugging Face Transformers documentation: <https://huggingface.co/docs/transformers/en/index>