**8-Week Applied AI/ML Internship Roadmap**

**Objective**  
Guide interns from database foundations to advanced AI techniques through one tightly-linked e-commerce case study. Each week adds skills and ends with a hands-on project that feeds the next stage, culminating in a production-ready, AI-powered analytics & chatbot solution.

**Week 1 – Relational Databases for Analytics**

**Concepts**

* ER-modelling, 1-NF → 3-NF
* SQL DDL: tables, constraints, indexes
* Views vs. materialized views
* Stored procedures, triggers

**Guided Practice**

Interns import the synthetic e-commerce bundle (11 CSVs) and run the pre-written schema script to build a complete OLTP store.

**Mini-Project**

“**E-Com Mini-Mart**”:

1. Load all 11 files.
2. Create two new problem-driven views (e.g., low-stock, VIP customers).
3. Write a stored procedure GetDailySales(yymmdd) that returns revenue by category.
4. Schedule a trigger that logs inventory whenever an order item is inserted.

Deliverables: SQL repo + short demo video + ER diagram.

Free resources

* *SQL Murder Mystery* interactive lab (free, Helix Academy)
* Oracle LiveSQL sandbox

**Week 2 – Data Cleaning & Statistical QA**

**Concepts**

* Pandas I/O, schema drift checks
* Missing-value strategies, outlier capping, data typing
* Aggregations & derived metrics (AOV, repeat-rate)
* Validation tests: variance, standard deviation, Shapiro–Wilk, Levene’s

**Guided Practice**

Clean the Week 1 tables in Jupyter, save “gold” Parquet files, and document tests with pytest+great\_expectations.

**Mini-Project**

“**Sales & Customer Gold Layer**”: build a reproducible ETL notebook that:

* Outputs tidy fact & dimension data frames
* Calculates RFM scores
* Runs an automated data-quality report (HTML) with statistical assertions

Free resources

* “Effective Pandas” course (Matt Harrison)
* *Python Data Cleaning Cookbook* (Packt – free to read on O’Reilly online library)

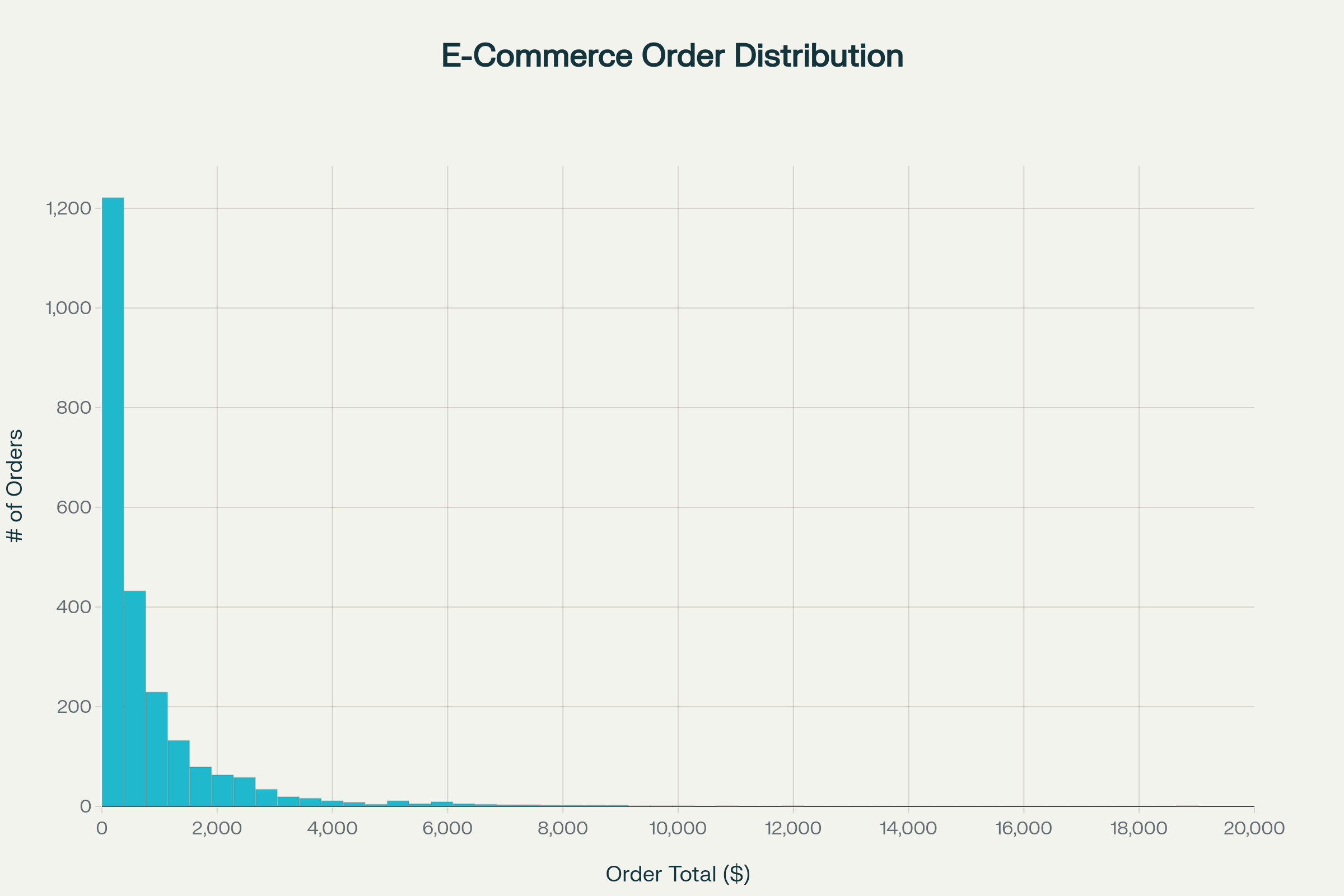
**Week 3 – Exploratory Data Analysis & Dashboards**

**Concepts**

* Story-driven EDA (hypothesis → chart)
* Matplotlib/Seaborn grammar
* Plotly Express & Dash basics
* KPI design for commerce

**Guided Practice**

Visualize sales seasonality, churn, product performance.



Order value distribution across 2,500 synthetic orders

**Mini-Project**

“**Executive KPI Dashboard**”: a Plotly Dash web-app that shows:

* Real-time revenue, AOV, conversion funnel
* Category heat-maps, top SKUs, promo lift
* Drill-down customer segmentation

Deploy to Streamlit/Dash Cloud and demo to mentors.

Free resources

* Plotly official 4-hour YouTube bootcamp
* TU Delft “Data Visualization” MOOC (edX)

**Week 4 – Feature Engineering & Classical ML**

**Concepts**

* Encoding, binning, interaction features
* Train/valid/test discipline
* Model zoo: logistic vs. tree vs. gradient boost vs. k-means
* Cross-validation, hyper-opt

**Guided Practice**

Create a labeled churn dataset and a regression target (CLV).

**Mini-Project**

“**Smart Targeting Models**”

1. Build at least three models (classification, regression, clustering) using engineered features.
2. Save pipelines with sklearn + joblib.
3. Compare with a model card (precision, recall, SHAP).

Free resources

* Kaggle “Feature Engineering” free course
* [Fast.ai](http://Fast.ai) tabular lessons (YouTube)

**Week 5 – Introduction to Deep, NLP, Time-Series & CV**

**Concepts**

* Feed-forward vs. CNN vs. RNN (high level)
* Transformers & embeddings primer
* Popular time-series models (ARIMA, Prophet)
* Text preprocessing, sentiment lexicons

**Guided Practice**

Use product reviews for sentiment and orders for time-series forecasting.

**Mini-Project**

“**Review Miner**”

* Web-scrape extra reviews with scrapy
* Fine-tune a small DistilBERT for sentiment
* Predict weekly demand with Prophet and overlay promo calendar

Free resources

* Hugging Face course chapters 1-3
* “Deep Learning for Time Series Forecasting” notebook series (Jason Brownlee – free)

**Week 6 – Computer Vision in Retail**

**Concepts**

* OpenCV image IO & preprocessing
* Pre-trained detection models (YOLOv8, SSD)
* Face detection vs. recognition ethics
* Evaluation metrics (mAP, IoU)

**Mini-Project**

“**Shelf-Scanner**” – simulate product-shelf images, then:

1. Fine-tune YOLOv8 to recognize five top SKUs.
2. Build a Streamlit app that flags out-of-stock facings via webcam.

Dataset hint: use Kaggle “Groceries” images + synthetic augmentation.

Free resources

* Ultralytics YOLOv8 free tutorials
* OpenCV-Python 3-hour crash course (freeCodeCamp)

**Week 7 – RL, LLMs & Retrieval-Augmented Generation**

**Concepts**

* Q-learning, policy vs. value
* ChatGPT / open-source LLM landscape
* Vector databases & RAG pipelines
* Streamlit deployment tips

**Guided Practice**

Create a simple gridworld in gymnasium and train Q-learning.

**Mini-Project**

“**E-Com Chat-Agent**”

* Embed product, order & FAQ texts into a FAISS index.
* Build an LLM (e.g., Llama-3–8B via API) with RAG to answer order-status and product questions based on the intern-built database.
* Front-end: Streamlit chatbot with session memory, deploy to Community Cloud.

Outcomes: demonstrate grounding vs. hallucination with test prompts.

Free resources

* OpenAI Cookbook (MIT-licensed)
* DeepMind “Reinforcement Learning Specialization” audit track (Coursera)

**Week 8 – Capstone & Evaluation**

**Capstone Brief**

“**Omni-Commerce AI Suite**” – interns integrate every artefact into a micro-service portfolio:

|  |  |  |
| --- | --- | --- |
| Layer | Component | Technology |
| Data | Postgres + gold-layer views | SQL, Airflow DAG |
| Analytics | Plotly KPI dashboard | Dash |
| ML API | Flask serving churn & CLV models | Uvicorn, FastAPI |
| CV Service | Shelf-scanner endpoint | YOLOv8, ONNX |
| Chatbot | Streamlit RAG UI | FAISS, Llama-3 |

**Demo Day**

* 15-min team presentation
* Live system walk-through
* GitHub repo & documentation hand-over

**Evaluation Rubric**

* Technical correctness (40%)
* Code quality & MLOps (25%)
* Business insight & storytelling (20%)
* Team collaboration (15%)

**Instructor Toolbox**

* **Starter data bundle** (11 CSVs + SQL schema)
* Detailed walk-through README and file inventory
* Teaching notebook templates per week (provided in repo)
* Slack channel & daily scrum check-ins
* Weekly quiz + code review checklist

**Closing Ceremony**

Interns receive:

* Digital certificate
* Code portfolio showcase link
* Recommended next-step learning path (MLOps, distributed training)

Happy learning and building!