

Fraud Analytics with Python & SQL

Table of Contents

A Production-Ready Workbook with Actuarial Depth

Part I — Core Technical Foundations

Chapter 1 — Python for Fraud Analysts

- Variables and data structures
- Conditional logic
- Functions
- Modular scoring rules
- Basic transaction simulation

Theory 1 — Fraud as a Statistical Decision System

- Observed fraud vs latent fraud intent
- Signal and noise
- Base rate constraints
- Why perfect detection is impossible

Chapter 2 — NumPy and Vectorized Computation

- Arrays vs lists
- Random number generation
- Batch scoring
- Efficient fraud simulation

Chapter 3 — Pandas for Transaction Data

- DataFrames
- Filtering and slicing
- Aggregations and group-by analysis
- Fraud rate calculations
- Date-time manipulation

Theory 2 — Rare Event Mathematics

- Bernoulli processes
- Class imbalance
- Precision as a function of base rate

Part II — SQL for Real-World Fraud Analytics

Chapter 4 — Relational Data in Banking Systems

- Transaction tables
- Customer and merchant tables
- Primary and foreign keys
- Star schema intuition

Chapter 5 — Core SQL for Fraud Analysts

- SELECT and WHERE

- GROUP BY and HAVING
- INNER and LEFT JOIN
- Subqueries
- Common Table Expressions (CTEs)

Chapter 6 — Window Functions (Critical)

- ROW_NUMBER()
- SUM() OVER()
- COUNT() OVER()
- Rolling aggregates
- Velocity feature construction

Chapter 7 — Building Fraud Features in SQL

- Rolling transaction counts
- Spend velocity
- Country-switch detection
- Tenure calculation
- Historical fraud rate features

Theory 3 — Information Extraction and Feature Validity

- Features as information summaries
- Predictive vs causal variables
- Data leakage
- Selection bias in monitoring systems

Chapter 8 — SQL and Python Integration

- Pulling SQL data into Pandas
- Feature parity validation
- Reproducible data pipelines

Part III — Supervised Fraud Modeling

Chapter 9 — Feature Engineering for Fraud

- Behavioral feature design
- Leakage avoidance
- Time-aware train/test splits
- Production consistency

Chapter 10 — Logistic Regression

- Odds and log-odds
- Maximum likelihood intuition
- Coefficient interpretation
- Marginal effects

Theory 4 — Bayes Decision Theory

- Likelihood ratios
- Optimal classification rules
- Thresholds as cost-based decisions

Chapter 11 — Model Evaluation Under Imbalance

- Confusion matrix
- Precision and recall
- ROC and AUC
- Threshold tuning
- Cost-sensitive evaluation

Chapter 12 — Scorecards and Interpretability

- Linear additive scoring
- Points-based systems
- Stability analysis
- Regulatory transparency

Theory 5 — Evaluation Geometry

- ROC invariance to base rate
 - Precision-recall tradeoffs
 - Why AUC does not equal profit
-

Part IV — Machine Learning Systems in Fraud

Chapter 13 — Decision Trees

- Recursive partitioning
- Impurity metrics
- Overfitting considerations

Chapter 14 — Random Forests

- Bagging
- Variance reduction
- Feature importance stability

Chapter 15 — Gradient Boosting

- Sequential learning
- Regularization
- Hyperparameter tuning
- Production constraints

Chapter 16 — Model Selection and Validation

- Cross-validation
- Temporal backtesting
- Stability over time

Theory 6 — Bias-Variance and Model Stability

- Generalization error
 - Variance in low base rate settings
 - Stability over time
-

Part V — Economic Decision Optimization

Chapter 17 — Fraud Economics

- Expected fraud loss
- Interchange revenue
- Customer friction cost
- False decline impact

Chapter 18 — Optimal Threshold Selection

- Profit curves
- Utility maximization
- Risk appetite calibration

Theory 7 — Frequency-Severity Decomposition & Credibility

- Fraud frequency modeling

- Fraud severity modeling
- Aggregate loss variance
- Tail exposure
- Segment-level shrinkage (credibility theory)
- Bühlmann intuition
- Application to merchant/customer segmentation

Chapter 19 — Portfolio-Level Fraud Risk

- Aggregate loss volatility
 - Unexpected loss
 - Stress testing
 - Capital intuition
-

Part VI — Adversarial and Dynamic Systems

Chapter 20 — Concept Drift

- Population drift
- Behavioral drift
- Fraudster adaptation

Chapter 21 — Monitoring and Governance

- Population Stability Index (PSI)
- Drift detection
- Backtesting frameworks
- Model risk management

Theory 8 — Fraud as an Adaptive Game

- Strategic fraudster behavior
- Feedback loops
- Equilibrium intuition

Chapter 22 — Strategy Evolution

- Champion/challenger testing
 - Rule overlays
 - Feedback loops
-

Part VII — Advanced Topics

Chapter 23 — Unsupervised Fraud Detection

- Anomaly detection methods
- Isolation forests
- Limitations of unsupervised systems

Chapter 24 — Graph-Based Fraud Detection

- Network modeling
- Entity resolution
- Connected components
- Graph-derived features

Chapter 25 — Real-Time Fraud Systems

- Low-latency scoring
- Feature stores
- API architecture
- Operational constraints

Part VIII — Capstone Projects

- Simulated end-to-end fraud environment
- SQL-based feature construction
- Model training and validation
- Profit-optimized threshold selection
- Adversarial drift simulation

Appendix — Statistical Foundations of Fraud Systems

A1 — Fraud as a Rare Event Problem

- Base rate mathematics
- Conditional probability
- False positive paradox
- Signal versus noise

A2 — Frequency Modeling

- Bernoulli processes
- Binomial modeling
- Poisson approximation
- Account-level vs transaction-level frequency

A3 — Severity Modeling

- Loss distributions
- Lognormal and Gamma intuition

- Heavy tails
- Outlier behavior

A4 — Frequency-Severity Decomposition

- Expected loss structure
- Compound loss models
- Aggregate fraud simulation
- Portfolio variance