KUSH PATEL

Data Scientist | Applied ML · Forecasting · Experimentation · Causal Inference | Python · SQL · XGBoost · scikit-learn

Vancouver, BC +1 (778) 637-1602 dharma.patel552@gmail.com linkedin.com/in/kush-patel-48885719b github.com/KushPatel29
kushpatel29.github.io

Open to remote and relocation across Canada · Legally eligible to work in Canada · Available immediately

SUMMARY

- Applied-ML candidate with an MPS in Analytics (Applied Machine Intelligence, GPA 3.76) and 4+ years of paid operations/finance analytics, including SKU forecasting deployed as Python web apps.
- Runs a public wildfire forecast with prospective outcome grading. Public experiments and causal-optimization cases use labelled synthetic data; model selection compares simple baselines. Lifecycle evidence includes a local MLflow registry and scheduled public forecasts.

TECHNICAL SKILLS

Models: Python/pandas/NumPy/scikit-learn, XGBoost, logistic/Poisson models, isotonic calibration, anomaly detection, collaborative filtering and time-series forecasting.

Evaluation: Out-of-time/season splits, rolling-origin backtests, ROC/PR-AUC, top-decile capture, WAPE, recall@k, calibration and cluster bootstrap intervals.

Experiments & decisions: Randomized A/B tests, uplift, geo holdouts, difference-in-differences, power analysis and mixed-integer optimization (HiGHS/Gurobi), demonstrated in projects.

Data & lifecycle: SQL/T-SQL, PySpark, Databricks, DuckDB, Fabric, dbt, FastAPI, Streamlit, Docker, pytest and scheduled GitHub Actions; local MLflow registry with retrospective versions and prospective monitoring.

PROFESSIONAL EXPERIENCE

Two Rivers Specialty Meats

Dec 2023 – Jun 2026 · North Vancouver, BC

Data Analyst — Operations, Logistics & Enterprise Reporting

- Built SKU demand forecasts as Python web apps for procurement, alongside FEFO and lot-level traceability to support fewer stock-outs and less perishable waste.
- Delivered customer inactivity, lost-account and rep-coverage analysis for Sales.
- As sole analyst, integrated ERP, Square POS, workforce and REST feeds into Bronze/Silver/Gold datasets using Microsoft Fabric, Azure Data Factory/Synapse, SQL and PySpark.
- Built source-to-mart GL/P&L and inventory controls, cutting discrepancies an estimated ~45% and reporting errors ~30% (internal directional estimates, not audited).
- Automated schema, completeness, uniqueness and referential-integrity checks, with anomaly thresholds, exception logs and pipeline alerts.
- Owned the Power BI semantic model and DAX library; agreed KPI definitions and acceptance criteria with Sales, Finance and Supply Chain.

Shivam Investments

Dec 2020 – Jul 2022 · Vadodara, India (hybrid)

Financial Analyst — Data & Reporting

- Gathered requirements from portfolio and risk stakeholders; delivered Power BI dashboards with auditable measure definitions and reconciled GL/P&L figures across sources.
- Automated recurring preparation with Python and SQL views, replacing spreadsheet consolidation and cutting manual effort an estimated ~40% (internal directional estimate).

SELECTED PROJECTS — public code and CI-tested results; synthetic data is labelled

Canada Wildfire Risk — Python, XGBoost, scikit-learn, GitHub Actions · real public data
github.com/KushPatel29/wildfire-prediction-app · live: wildfire-prediction-app1.streamlit.app
Public forecasts, graded against outcomes: calibrated classifier and Poisson models over 753 cells and 4.0M cell-days of fire and station-weather records. Historical season-split ROC-AUC 0.881; the first 36 prospective days score 0.734 (95% CI 0.696–0.770), ahead of two baselines on ranking while the top-decile edge remains unresolved. Forecasts rebuild twice daily and are graded weekly against satellite detections using whole-day bootstrap intervals. Found and fixed an index-alignment bug that had put 97.7% of fires in the wrong cell.

Corridor Decision Intelligence — Python, DuckDB, PySpark on Databricks, HiGHS / Gurobi, FastAPI · synthetic customers
github.com/KushPatel29/decision-intelligence-platform · live: corridor-decision-intelligence.streamlit.app
Causal ML and decision optimization: uplift models trained on a ten-arm randomized offer trial feed one mixed-integer program over 112,563 customer-offer decisions, solved to proven optimality. The plan captures 75% of the perfect-knowledge value ceiling against 7% for propensity targeting, and the policy was itself randomized against business as usual in the simulation: +$1.40 per customer (95% CI $0.23 to $2.57). Features reproduced in PySpark on Databricks to within 2.2e-11.

Marketing Attribution & Incrementality — Python, SQL, causal inference · synthetic data with a known answer
github.com/KushPatel29/marketing-attribution-analytics · live: attribution-vs-truth.streamlit.app
Measurement that survives a known truth: six attribution models chase an answer built into the data and none finds it (last-touch gives “direct” a quarter of conversions; its true share is 1.9%). A powered geo holdout does, so the budget plan is calibrated on the experiment rather than on attribution.

Customer Recommendation Engine — scikit-learn, FastAPI, Streamlit · synthetic data
github.com/KushPatel29/Customer-Recommendation-Engine · live: cross-sell-rep-console.streamlit.app
Choosing the simpler model on evidence: a two-stage learned ranker scored 81.2% recall@10 on a warm-customer holdout and plain item-neighbourhood collaborative filtering scored 84.9% (popularity 75.4%), so the simple model ships. A proposed, pre-registered randomized pilot would measure commercial lift; no online lift has been established.

Ask Your Data — Python, DuckDB, ONNX embeddings, Claude API, Streamlit · synthetic data
github.com/KushPatel29/ask-your-data · live: ask-your-data-kp.streamlit.app
An LLM application that refuses rather than guesses: text-to-SQL over 71 tables in 11 domains, asked by voice or text. A keyless compiler binds questions to a profiled semantic layer; with a key, Claude writes the SQL through the same read-only guard, and its prose is checked against the returned rows. On a 58-question contract: 46 right, 0 wrong, 12 refused.

Additional anomaly detection, grounded text-to-SQL and fairness/survival cases at kushpatel29.github.io. MLflow records six historical wildfire versions retrospectively and ongoing prospective-monitoring runs in a local registry.

EDUCATION

Northeastern University — Master of Professional Studies in Analytics (Applied Machine Intelligence), GPA 3.76/4.0 · Vancouver, BC — **Sep 2022 – Jun 2024**
Coursework: machine learning, deep learning, NLP, time series, statistical inference.

Gujarat Technological University — Bachelor of Engineering, Computer Science · India — **Aug 2017 – Aug 2021**

AWARDS & CERTIFICATIONS

First prize, 2024 wildfire-prediction hackathon (now a live forecast) · IBM Python for Data Science, AI & Development · Google Data Analytics Professional Certificate · UC Davis SQL for Data Science · Databricks Fundamentals.