

YASHPAL SAINI

New York, NY | (646) 289-1606 | yashsaini30668@gmail.com | [LinkedIn](#) | [GitHub Portfolio](#)

SUMMARY

Data Analyst with 2+ years of experience building SQL and Python analytics pipelines, experimentation frameworks, and forecasting models across enterprise-scale environments. Delivered measurable outcomes: 55% faster reporting cycles, 37% improvement in data accuracy, and 18% lift in campaign engagement.

EXPERIENCE

Research Foundation of CUNY | New York, NY

Aug 2025 – Present

Data Specialist Intern

- Develop and maintain SQL and BigQuery pipelines that have cut reporting time by 55% across 26 institutions, replacing manual spreadsheet exports with real-time Power BI dashboards that operations teams can use without analyst involvement.
- Improved data accuracy by 37% by designing validation and reconciliation frameworks across workforce and leave datasets, giving teams reliable numbers for operational planning and compliance reporting.
- Automated ETL pipelines and dashboard refresh cycles with Python and BigQuery, cutting weekly manual reporting work by roughly 50% and freeing time for higher-value analysis.
- Consolidated 7+ fragmented data sources into a unified GCP reporting layer, improving metric consistency across reports for 26 institutions and resolving long-standing reporting discrepancies.
- Identified a data integrity issue that had gone undetected for 2+ reporting cycles across 3 institutions by running Python-based EDA across the unified layer before a major reporting deadline.

I3 Infosoft Pvt. Ltd. | Noida, India

Sep 2022 – Jul 2024

Data Analyst

- Ran A/B experiments and built a Python customer segmentation model across 1M+ CRM records that lifted campaign engagement by 18%; findings were adopted by marketing to shape the following quarter's channel strategy.
- Replaced manual CRM reporting with automated SQL pipelines, reducing weekly report production time by 6 hours and giving the team capacity to support two additional client workstreams.
- Designed a Python time-series forecasting model that improved inventory planning accuracy by 20% and reduced stockout incidents by projecting monthly demand across product lines for the operations team.
- Implemented Power BI and Tableau dashboards tracking 15+ KPIs across SLA compliance, fulfillment rates, and cost metrics, taking ops response time from weekly reports to same-day visibility so teams could act on issues while they were still fixable.
- Migrated 7+ fragmented datasets into normalized relational SQL structures, cutting recurring query runtimes and reducing cross-team reporting turnaround by 25%.

PROJECTS

Customer Churn Prediction | [GitHub](#) | Python, Scikit-learn, SQL, Power BI

Jan 2026

- Achieved 84% classification accuracy and identified high-risk churn segments by contract type, tenure, and billing behavior using a Random Forest classifier developed with Python feature engineering and SQL-based EDA in SSMS.
- Enabled stakeholders to prioritize retention strategies without analyst support by delivering a Power BI dashboard surfacing churn rate by segment, customer lifetime value, and key retention KPIs derived from model outputs.

NYC Housing Equity & Policy Analytics | [GitHub](#) | R, Quarto, Random Forest, Spatial Analysis

Dec 2025

- Built an automated R pipeline integrating 2023 NYC housing records from 311, HPD, DOB, and ACS APIs, using spatial joins to aggregate complaint rates and resolution times across census tracts.
- Trained a Random Forest model ($R^2=0.54$), outperforming OLS and Elastic Net baselines, and used feature importance analysis to identify neighborhood income levels as a key predictor of longer complaint resolution times.

EDUCATION

Baruch College, Zicklin School of Business

Expected May 2026

M.S. in Business Analytics, GPA: 3.6/4.0 | Coursework: Data Mining, Database Management, Applied NLP, R Programming

TECHNICAL SKILLS

Languages & Query: SQL (Advanced), Python (Pandas, NumPy, Scikit-learn), R

Analytics: A/B Testing, Experimentation Analysis, Hypothesis Testing, Cohort Analysis, Forecasting, Supervised Learning, Customer Segmentation, Pricing Analytics, ROI Modeling

Cloud & Pipelines: Google Cloud Platform (GCP), BigQuery, GCP Data Lakes, Azure Data Studio, ETL Pipeline Development, Data Warehousing

BI & Visualization: Power BI, Tableau, Looker, Advanced Excel (PivotTables, VLOOKUP)

Data Quality: Data Validation, Data Quality Monitoring, Data Lineage Documentation, Data Governance

Certifications: Google Data Analytics Professional Certificate (Coursera)