STEPHEN BRACCO

New York, NY | 516-314-8346 | s.bracco2003@gmail.com | LinkedIn | GitHub

EDUCATION

BBA Statistics and Quantitative Modeling, Concentration in Data Science — **Expected May 2027**

CUNY Baruch College, Zicklin School of Business — New York, NY

Minor: Mathematics; GPA: 3.7/4.0

Relevant Coursework: Data Mining & Statistical Learning, Programming for Analytics, Statistical Computing

Clubs: Machine Learning and Data Science, Financial Quants and Engineers

SKILLS

Data Engineering: Python, SQL, ETL, GCP (BigQuery), Parquet, REST/Socrata API, Git, GitHub, AWS

Analysis and Visualization: pandas, GeoPandas, Matplotlib, Tableau, Seaborn, Streamlit, Excel (PowerQuery)

Machine Learning Methods: XGBoost, Logistic Regression, Random Forest, scikit-learn, Time-Series Validation, Probability Calibration, Risk Stratification

EXPERIENCE

Tech Talent Pipeline Residency (TTPR), Baruch College — **New York, NY**

Data Analyst Apprentice — June 2026 – Present

- Applied data cleaning, transformation, exploratory analysis, and statistical testing to real-world datasets.
- Constructed SQL queries including joins, filters, and aggregations to prepare data for analysis.
- Developed and fine-tuned classification, clustering, and time-series forecasting models, and evaluated performance.
- Applied Generative AI and LLMs via Python, APIs, and prompting to automate analysis & build AI-assisted workflows.
- Built Tableau and Excel dashboard that turn KPIs and model output into a clear narrative with recommendations.

Cha Physical Therapy, PC — **New York, NY**

Data Analyst — June 2025 – August 2025

- Automated financial reporting workflows, ingesting messy CRM data, saving 10+ hours per week and stopping $80 per customer visit in losses by identifying and resolving insurance plan discrepancies.
- Developed Python visualizations with pandas and matplotlib to analyze business trends, identifying 5+ key trends for strategic decision-making.
- Cleaned and structured 3,000+ financial and demographic records using Excel and Python to support operational decisions.
- A/B tested digital ad performance, improving ROI by 15% with optimized targeting based on CTR, CPC, and reach metrics.

PROJECTS

CabLink – MTA Bus Delay Prediction & Taxi Alert System – *Team of 4* — **July 2026 (Ongoing)**

- Built an ETL pipeline processing 15.5M GTFS-Realtime vehicle positions from AWS S3 into stop-level speeds across 372 route segments via per-vehicle trip chaining.
- Integrated five data sources, including 20M+ rows, into a unified hourly panel and resolved coordinate-system conflicts.
- Engineered a leak-free deviation index with trailing-median baselines to isolate unscheduled disruption from recurring rush-hour patterns; validated by detecting a 50% speed collapse on NYC's second-rainiest day of 2023.
- Benchmarked five model families (OLS, Ridge, LASSO, SARIMAX, XGBoost) under time-ordered validation, showing loss-function choice beat model complexity (5.3x vs 3.2x ranking lift); delivered an XGBoost classifier at ROC-AUC 0.802 ± 0.03 across eight walk-forward folds, with isotonic calibration cutting expected calibration error by 183x.
- Designed a rule-based alerting layer targeting taxi drivers to capitalize on stranded passenger demand from transit delays, emitting ~8 stop-level alerts/day at 15.6% precision against a 0.5% base rate.
- Quantified $120-146K/year revenue opportunity using a borough-benchmarked fair-share market analysis comparing local taxi capture rate (0.20%) against Brooklyn's own baseline (0.47%).

Hospital Readmission Risk Modeling — **April 2026**

- Built an end-to-end machine learning pipeline to predict 30-day hospital readmissions of diabetes patients using XGBoost, Logistic Regression, and Random Forest on imbalanced clinical data.
- Applied SHAP to identify key drivers of readmission risk, finding prior inpatient visitation and disease burden as dominant predictors.
- Evaluated with lift and cumulative gains – the top 20% of predicted-risk patients captured ~35% of readmissions, supporting a 15% probability cutoff for intervention targeting.
- Deployed the model as an interactive Streamlit application for real-time patient risk scoring and scenario analysis.