

Arnav Sahai

New York, New York, United States | arnav.sahai@columbia.edu | (646) 316-6402 | github.com/Arnav504 | linkedin.com/in/arnav-sahai-370702116 | github.com/Arnav504

PROFESSIONAL SUMMARY

Data and modeling analyst with a Computer Science engineering background and a Columbia SIPA MPA focused on Data Science for Policy, Finance & Economic Policy, with specialized coursework in **macroeconomic modeling and forecasting**. Experienced in **data management** for analytical products—building and maintaining **data pipelines**, relational databases, and documentation to ensure **data quality assurance** and repeatability. Proficient in **database query languages** (SQL) and **macroeconomic frameworks** from technical undergraduate training. Bring applied quantitative modeling (Python, R, SQL) and clear client-facing communication to support macroeconomic analysis, forecasting-style time series work, and adoption of new analytical tools (including LLM-assisted workflows) across stakeholders.

EDUCATION

Columbia University

Master of Public Administration, *Data Science for Policy, Finance & Economic Policy*

Sep 2024 - May 2026

Guru Gobind Singh Indraprastha University

Bachelor of Technology, *Computer Science (Engineering)*

Aug 2015 - May 2019

SKILLS

Programming Languages: Python, R, SQL, Java

Frameworks & Libraries: Pandas, NumPy, scikit-learn, NLTK, Spring MVC, FastAPI, SQLAlchemy, React, TypeScript

Databases: PostgreSQL, MySQL, relational databases, SQL (database query language)

Tools & Software: ETL, data pipelines, data quality assurance, data management, data governance, data dictionaries, metadata documentation, version control (Git), Git, Docker, Jupyter, Excel (Advanced), Tableau, Power BI, Microsoft Office (Word, PowerPoint, Outlook), Splunk

Soft Skills: Structured problem-solving, market & desk research, insight synthesis, business analysis, client relationship management, project & program management with status updates, executive presentations, stakeholder alignment, mentoring, macroeconomic modeling and forecasting, macroeconomic frameworks

Spoken Languages: English (Fluent), Hindi (Fluent), Urdu (Conversational)

WORK EXPERIENCE

Sinclair Capital | Research Consultant | New York City, New York

Sep 2025 - Dec 2025

- Served as a client-facing consultant, developing an advanced statistical and financial model to investigate a dynamic business question and inform decision-making
- Structured an ambiguous problem into an executable analysis and communicated recommendations to a non-technical audience through written and oral presentation
- Produced well-documented model assumptions, inputs, and outputs to support reproducibility and stakeholder review in a fast-turnaround setting

International Finance Corporation (World Bank Group) | Development Impact Measurement Intern | New York City, New York.

Oct 2025 - Dec 2025

- Applied statistical and quantitative modeling and data mining across 15+ projects, designing custom analyses and benchmarking outcomes against sector metrics to support development/financial performance interpretation; applied macroeconomic forecasting-style approaches to model economic and financial outcomes
- Owned end-to-end **data management**: cleaned and validated complex datasets, built relational databases, and maintained documentation of methods/assumptions and **metadata documentation** to enable reuse, auditability, and stakeholder confidence
- Implemented **data quality assurance** checks and repeatable workflows (SQL/Python/R) to improve integrity of inputs feeding analytical models and reporting deliverables for senior stakeholders

Accenture | Application Development Associate (Pharmaceutical Client — Bristol Myers Squibb) | Bengaluru, India.

Oct 2019 - May 2021

- Supported production data systems for a global pharmaceutical company, building data-integration processes (Java, Spring MVC) and monitoring data across 1,500 interfaces (Splunk) to detect anomalies and quality issues
- Performed production support and defect triage with close attention to detail, contributing to stable data flows and timely resolution of data issues impacting downstream consumers
- Coordinated with cross-functional and global teams in an Agile environment, translating business requirements into technical updates and communicating technical detail to non-technical partners

The Smart Cube | Data Analyst Intern | Noida, India

Mar 2019 - May 2019

- Conducted procurement and supply-market research, benchmarking suppliers and analyzing cost trends and drivers to inform client decisions
- Designed custom analyses and Tableau/Power BI reporting and improved data accuracy by 70% through redesigned ETL and QA workflows, strengthening data integrity for decision-making

Fund for the City of New York | Community Fellow, Community Board 9 | New York City, New York.

Sep 2025 - May 2026

- Conducted research and stakeholder interviews; synthesized quantitative and qualitative insights into a policy framework and decision guide presented to leadership under tight time constraints
- Built an end-to-end **data pipeline** (citywide and block-level datasets with GIS) behind the analysis, including dataset structuring, transformations, and documentation to support accurate downstream analysis
- Established lightweight **data governance** practices (naming conventions, version control, and data dictionaries) to keep sources, assumptions, and transformations transparent across collaborators

- Researched technology and industry trends and translated complex analysis into clear, decision-oriented recommendations for stakeholders

- Guide and mentor participants in engagement workshops and coordinate cross-sector initiatives across residents, agencies, and nonprofits

- Support 500+ students and faculty, resolving issues under deadline with close attention to detail and accurate documentation

PROJECTS

Cyber Narrative Radar

- Designed and built a full-stack data platform end to end — relational schema, ETL pipeline, and containerized deployment (FastAPI, SQLAlchemy, PostgreSQL, Docker, React/TypeScript), with consistent documentation to support data lineage and maintainability

ASAN MACRO — AI-Powered Trade Intelligence Agent

- Built an agentic AI product processing real-time and historical maritime trade data to surface macroeconomic shifts in global goods flows, applying sector analysis to BRICS decoupling trends across minerals, textiles, and metals; used large language models for rapid prototyping and code generation within the build workflow

Russian Troll Twitter Activity Classifier

- Conducted text-as-data analysis on 200,000+ social media records using sentiment analysis and NLP (Python: NLTK, scikit-learn); built a Naive Bayes classifier distinguishing troll vs. authentic accounts with 71.9% accuracy

Egg Price Fluctuations & Bird Flu — Forecasting

- Modeled supply-side commodity price shocks via regression and time-series analysis in R, identifying a 2–3 month lag between H5N1 outbreaks and Grade A egg price surges, and evaluated equity responses (Cal-Maine +18%, Vital Farms +12%) to support forecasting-style scenario insights

Elon Musk Tweets & Tesla Stock — Predictive Modeling

- Built a multiple regression model quantifying relationships between tweet sentiment, frequency, and Tesla stock movements, demonstrating alternative-data signal extraction for financial analysis

SIPA Capstone, JPMorgan Chase — Digital Sovereignty

- Analyzed digital sovereignty regimes across the EU, India, and China (GDPR, CRA, EUCS; RBI, DPDP Act, SEBI; CSL, DSL, PIPL) and built a cross-jurisdictional conflict matrix translating regulatory divergence on data localization, cloud architecture, AI governance, and third-party risk into concrete risk exposure and a modular sovereignty architecture recommendation for a global bank

JPMC — Firmwide Cyber Risk Taxonomy

- Developed a board-ready cyber risk taxonomy using MITRE ATT&CK, ISO 27001, and NIST across five major risk categories; applied qualitative and quantitative risk assessment to evaluate likelihood, impact, and residual risk, integrating zero trust, multi-cloud resilience, AI-enabled defenses, and post-quantum planning

OHMAD — Governance & Compliance Framework

- Co-authored a sector-led deception and governance framework using high-fidelity IT/OT mirrors and agentic AI to attract pipeline ransomware actors, capture tooling, and generate early-warning intelligence; designed a governance and legal model aligning TSA, CISA, DOJ, and FBI information-sharing with GDPR/HIPAA compliance for critical infrastructure operators

CERTIFICATIONS & ACHIEVEMENTS

- McKinsey Forward - McKinsey
- DataCamp Data Scientist (Python; R), Intro to SQL - DataCamp
- IBM Data Science Foundations (Level 1) - IBM
- Simplilearn Statistics Essential for Data Science - Simplilearn
- Anthropic AI Fluency — Claude Code 101, Claude 101, Framework & Foundations, AI Capabilities & Limitations - Anthropic
- Columbia Prompt Engineering & Programming with OpenAI - Columbia University
- Bloomberg Finance Fundamentals - Bloomberg
- Bloomberg Spreadsheet Analysis (BQL) - Bloomberg
- Google Cybersecurity Professional Certificate - Google