

# ARNAV GOENKA

New York, NY | +1 (646) 249-3950 | [ag5235@columbia.edu](mailto:ag5235@columbia.edu) | [LinkedIn](#) | [GitHub](#) | [Portfolio Website](#)

## EDUCATION

### COLUMBIA UNIVERSITY

Master of Science, Applied Analytics (GPA: 3.92/4.30)

Dec 2026

New York, NY, United States

- **Focus:** Data Analytics, Machine Learning, Product & Business Analytics
- **Coursework:** Applied Machine Learning, Data Engineering, Cloud Computing, Storytelling with Data and AI, Blockchain & Crypto

### VELLORE INSTITUTE OF TECHNOLOGY

Bachelor of Technology, Computer Science Engineering (GPA: 3.96/4.00)

Aug 2025

Vellore, Tamil Nadu, India

- **Coursework:** Python, Object Oriented Programming (C, C++), Data Structures and Algorithms, Java, Probability and Statistics (R Studio), Software Engineering, Artificial Intelligence, DBMS (SQL), Linear Algebra, Calculus

## WORK EXPERIENCE

### Blinkit | Data Analyst Intern

Aug 2023 - Oct 2023

- Contributed to ~8% reduction in reverse stock pendency by developing & deploying a multi-level operational dashboard tracking Reverse Stock Transfer Orders (RSTO) across cities and stores, enabling early identification of bottlenecks
- Wrote SQL queries on Amazon Redshift to analyze store-, city-, and rider-level data and built dashboards used by operations managers to monitor last-mile performance and onboarding workflows, improving speed of operational decision-making.

### Invansys Technologies | Data Science Intern

Nov 2023 - Dec 2023

- Supported development of analytics and lightweight ML features within a client-facing SaaS platform by analyzing user interaction and feedback data using Python to identify usage patterns, engagement trends, and feature-level behavior insights.
- Built sentiment and user-behavior analysis and shared analytical outputs with the development team to help refine platform features and support more data-driven product decisions for the client.

### Finlatics | Business Analyst Intern

Nov 2024 - Jan 2025

- Evaluated margin improvement opportunities for a technology-focused company by analyzing revenue mix, regional margins, and contractor cost structures using Excel and Power BI, identifying drivers behind ~11% YoY margin growth versus ~26% industry benchmark and assessing margin uplift scenarios.
- Conducted feasibility analysis for a tourism project by projecting demand and revenues and analyzing fixed and variable costs across 3 potential locations, delivering structured investment recommendations.

### Finlatics | Data Science Intern

Mar 2025 - May 2025

- Analyzed a global media and technology dataset of hundreds of YouTube channels across multiple countries and 15+ performance variables using Python (Pandas, NumPy), performing data cleaning, missing value treatment, and feature structuring to prepare a reliable dataset for analysis.
- Used Matplotlib and Seaborn to evaluate subscriber growth patterns, geographic and category-level monetization trends, and revenue drivers, delivering a structured analytical report and presentation with actionable insights on content growth and platform performance strategy.

## PROJECTS

### CUStayWise - NYC Housing Livability Analytics Platform | Live App

- Built an end-to-end ETL pipeline by mapping NYC 311 complaints to Manhattan building, analyzing 90K+ complaints across 4k+ residential buildings in four ZIP codes near Columbia University using Python and PostgreSQL.
- Developed a building-level livability score and built an interactive Flask + Leaflet map of Manhattan to visualize per-category complaint counts at the building level, enabling comparative housing analysis and data-driven livability insights.

### Long Short-Term Network (LSTM) Based Prediction model for Nifty 50 Stocks

- Built a time-series based LSTM model predicting closing prices of stocks listed in Nifty 50( a benchmark index of Indian Stock Market) by feature engineering ~5 technical indicators (MACD, RSI, Bollinger Bands, Garman-Klass volatility, traded value) to capture market trends and applying data-smoothing (monthly aggregation, moving averages) to reduce noises in short term period.

### Patent(Approved, publication pending)-AI-driven dengue drug discovery using ChemBERT Modelling

- Generated 2K+ novel molecular structures from a set of 14 original dengue drugs using SMILES (Simplified Molecular Input Line Entry System) manipulation & RDKit(Chemoinformatic library in Python), then trained ChemBERT model (transformer-based model), used cosine similarity to flag & eliminate high-severity candidates, creating scalable, cost-effective AI-driven in silico dengue drug discovery pipeline.

## CORE SKILLS

- **Programming Languages:** Python, SQL, R, C, C++, Java
- **ML & Data Analytics:** ETL, NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PyTorch, Keras, OpenCV, Classification & Regression Models, Clustering, Bayesian Modeling, NLP Libraries (NLTK, VADER, Sentiment Analysis), Web Scraping (Beautiful Soup)
- **Data Visualization and Business Intelligence Tools:** Tableau, Power BI, MS Excel, Google Sheets
- **Statistical Analysis:** Descriptive Statistics, Regression Analysis, Time Series Analysis, Probability & Inferential Statistics

## CERTIFICATIONS

- AWS Certified Cloud Practitioner
- Salesforce-Data Visualization with Tableau
- Meta-Python Data Analytics
- Microsoft- Data Analysis and Visualization Power BI