

# Arnav Malhotra

703-973-8096 | arnavmalhotra@gmail.com | Brooklyn, NY

Data expert with 5+ years of experience leveraging advanced analytics and machine learning using Python, SQL, and data visualization tools across a diverse set of industries including real estate, healthcare, and health insurance. Proven track record of solving key business challenges by developing new metrics and data-driven features, leading to actionable insights

## EDUCATION

### JOHNS HOPKINS

**BS IN BIOMEDICAL ENGINEERING**  
May 2018 | Baltimore, MD  
Conc. in Computational Biology  
College of Engineering &  
School of Medicine

### INDUS INTL. SCHOOL

May 2014 | Bangalore, India  
International Baccalaureate

## SKILLS

### LANGUAGES

• Python • SQL • SAS • R •  $\text{\LaTeX}$

### LIBRARIES

• pandas • shapely • GeoPandas  
• scikit-learn • fastai  
• matplotlib • folium • ggplot

### SERVICES

• AWS RedShift & Sagemaker  
• Snowflake • Redash • JIRA  
• Looker • Tableau • PowerBI

### CERTIFICATIONS

• SAS Certified Specialist: Base Programming

## PROJECTS

### KAGGLE COMPETITIONS

*Rainforest Species Audio, Bundesliga Highlights, iMet Collection*

Built multi-classification deep learning models, in Python using fastai

### DONO

*Co-Founder*

Started as a small-batch meal prep service specializing in curries from around Asia.

Now popping up at wine bars across New York City, serving modern, Indian-inspired cuisine.

8it's Best New Pop-Up NYC 2024 (@dono.nyc)

## LINKS

Github:// [amalhot5](#)

LinkedIn:// [arnavmalhotra](#)

## EXPERIENCE

### SENIOR DATA ANALYST | ANTIDOTE HEALTH

May 2025 - Sept 2025 | New York, NY

- Reported directly to Chief Product Officer
- Reduced time required for monthly CMS audits from 1 week to 2 days by automating data quality checks with Python and SQL
- Developed algorithm in Python to map over 60k providers' inputted specialty to CMS Specialties and Specialty Types
- Used ChatGPT and Python to infer the facility types for about 1,000 facilities from their listed services online
- Built dashboard in Looker to track metrics related to member enrollment, payments, and Social Determinants of Health
- Standardized tickets and how team tracks work using Agile methodologies

### DATA ANALYST | PERCHWELL

May 2023 - Jan 2025 | New York, NY

- Joined as a founding member of Data team
- Overhauled Street Search by utilizing GIS analysis in Python to find which street(s) that over 700k buildings in NYC are on and identified every intersection in NYC—allowing users to search for listings based their street
- Led development on Brokerage Similarity Score, which is an algorithm (built with scikit-learn) that finds the most similar brokerage based on many factors including locations, pricing, number, and performance of their listings
- Built prototypes in Python and defined logic for the more complex portions of the client and third party ingestion process—including finding over 10 million off-market sales from public deeds records, creating & updating about 15 million buildings & units, with their geometries, from public assessments and parcels data for four different markets (NYC, Cincinnati, Iowa, & California)
- Performed triage for high priority client complaints, finding bugs, and working with engineers to implement fixes—most notably implementing the RLS Discrepancy Detector and Cincinnati geography overwrite
- Built internal tools to automate manual processes—e.g. quarterly client reporting, data quality bulk fixes, and creating or updating neighborhood geographies (reducing time to onboard new counties from 5 days to 1 day)
- Worked with Engineering and Customer Service teams to create metrics and dashboards in order to reveal trends and drive strategy around data quality and customer satisfaction in Looker and Redash
- Gave presentations to executives, clients, and the whole company; demonstrating new features, prototypes, and market insights
- Mentored intern, conducted learning transfers & code reviews, interviewed prospective hires, & evaluated new BI tools

### HEALTHCARE DATA ANALYST | CAREJOURNEY

May 2021 - Apr 2023 | Arlington, VA

- Led development on the Provider Cost Score, which compares the actual cost of a medical episode to its expected cost—this was found by building a machine learning model in Python (with AWS Sagemaker) to take many risk factors related to the procedure, the provider, and their patients into account
- Developed 50+ medical appropriateness measures and provider-to-provider referral metrics using commercial health insurance claims data in SQL
- Worked with Data Science team to implement patented, multi-classification models to categorize doctors into sub-specialties

### INFORMATICIST 1 & 2 | HUMANA INSURANCE COMPANY

Feb 2019 - Jun 2020 & Jun 2020 - April 2021 | Arlington, VA

- Worked on the Business Intelligence Advanced Analytics Team to create metrics for other departments in order to define benchmarks, find trends, and show their impact on the company's OKRs; using Tableau & PowerBI