Arnav Malhotra

703-973-8096 | arnavmalhotr@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 innovative solutions for many stakeholders—creating new metrics, data-driven features, and insights leading to increased efficiencies and customer retention.

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 •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

Curry In A Hurry
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.
(@curryinahurrynyc)

LINKS

Github:// **amalhot5**
LinkedIn:// **arnavmalhotra**

EXPERIENCE

DATA ANALYST | Perchwell
May 2023 - Present | 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 on which streets they are on
- Led development on Brokerage Similarity Score, which is an algorithm (built with scikit-learn) that finds the most similar brokerage based on the 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 Discrepany Detector and Cincinnati geography overwrite
- Built internal tools to automate manual processes—for example quarterly reporting for brokerages, many data quality bulk fixes, and creating or updating neighborhood geographies (reducing time to bring on a new county from over 1 week to 1 day)
- Worked with Engineering and Customer Service teams to create metrics around data quality and customer satisfaction, as well as dashboards to track those metrics, reveal trends, and drive strategy in Looker and Redash
- Gave presentations to executives, clients, and the whole company; demonstrating new features, prototypes, and market insights
- Mentored intern, 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 and SAS along with algorithms to get around limitations of that data
- Supported the Platform team's new product launch by interpreting wireframes and providing the data and metrics
- Worked with Data Science team to implement patented, multi-classification models to categorize doctors into sub-specialties

INFORMATICIST II | Humana Insurance Company
Jun 2020 - April 2021 | Arlington, VA

INFORMATICIST I | Humana Insurance Company
Feb 2019 - Jun 2020 | Louisville, KY

- Worked on the Business Intelligence Advanced Analytics Team to create and monitor metrics for other departments in order to create benchmarks, aid in goal-setting, find trends, and show their impact on the company's OKRs; using Excel, Tableau, & PowerBI
- Built predictive models with Python and scikit-learn to discover what factors influenced our members' impression of Agent Knowledge
- Performed sentiment analysis and categorization of agent case comments, web chat transcripts, and free-text responses from surveys using Python