# Galib Khan

(609) 412-3654 | ghk@uchicago.edu
https://majestickhan.com

## SUMMARY

Data Scientist specializing in large scale data analysis to provide business directions and implement cost saving methods. Able to use various modern tech stacks to create solutions ranging from exploratory data analysis to modeling development to large scale data automation. Known for being innovative and solving challenging data problems.

## SKILLS

- Python
- SQL
- R
- Tableau, Rmarkdown, Seaborn
- Spark *(Scala),* Hadoop, Impala

- OpenAI Gym / Stablebaseline3
- GCP/Azure
- Git
- Linux
- SnowFlake

## WORK EXPERIENCE

Kraft Heinz                                                                                              Chicago, IL
*Senior Data Scientist*                                                          March 2022 to Present
- Led the development of Spark based data solutions, enabling the analysis of multibillion row datasets and establishing the team's capability to handle high-volume data under my direction
- Optimized stroke length of pumps for filling cream cheese using Linear Regression, saving over $250,000 and 85% reduction in overfill
- Built and maintained a growing Delta Lake database in Databricks, exceeding 3.4 billion rows of sensor data, with daily updates that provide the team with the latest data for analysis and modeling
- Developed recommendation system to place products in certain manufacturing lines with a cost savings of over $370,000

Northern Trust                                                                                        Chicago, IL
*Data Scientist*                                                          August 2018 to March 2022
- Team lead in handling/managing transaction data using Spark and Impala
- Constructed framework in Python and R that aggregates and evaluates various classification models (*Random Forests, Support Vector Machine, Logistic Regression, Gradient Boosting, Adaboost*)
- Scaled hyperparameter tuning for ARMA-GARCH models on Linux servers, reducing computing time from 2 days to 12 hours

## ACADEMIC

University of Chicago, Physical Sciences Division                                        April 2020 to Present
- Master of Science in Analytics (MScA)                                                          GPA: 4.0/4.0
- Teaching Assistant for Reinforcement Learning

Columbia University                                                                                     Fall of 2019
- Grader for Dr. Wajahat Gilani for Applied Analytics

Rutgers University, Rutgers Business School – New Brunswick, NJ                               May 2018
- B. S., Business Analytics and Information Technology (BAIT)                               GPA: 3.9/4.0
- Graduated with the highest GPA in BAIT
- Teaching Assistant for undergraduate and graduate R and Python courses

## PROJECTS/RESEARCH

- Experimenting on how to develop, train, and transfer models from simulated environments to physical environments using CoppeliaSim, OpenAI Gym, and Legos
- Replicated Federal Reserve Bank of New York's Nowcasting model in Python based on Matlab version
- Contributed to Tensorflow/Adanet repository by developing a try/catch solution to run on Window