Soomin Kim

New York, NY 10128 (929) 837-0434
soomin.kim@nyu.edu linkedin.com/in/soominkim231 github.com/soominkim231

EDUCATION

New York University **May 2023**
M.S. in Data Science. GPA: 3.76/4.0 *New York, NY*

- **Relevant Coursework***:* Intro to Data Science, Probability and Statistics, Linear Algebra, Programming for Data Science, Data Visualization, Big Data, Machine Learning, Natural Language Processing (ongoing), Ethics of Data Science (ongoing)
- **Leadership**: Center for Data Science Student Leadership (Executive), Women in Data Science (Executive)

Dartmouth College **Jun 2020**
B.A. in Quantitative Social Science and Music *Hanover, NH*

- **Honors**: College Honor List, Citations for Academic Excellence

Relevant Certification: IBM Data Science Professional Certificate

TECHNICAL SKILLS

Languages: Python (pandas, NumPy, SciPy, sklearn, Matplotlib, PyTorch, Tensorflow) | R | SQL | MATLAB | STATA
Tools: Spark | Hadoop | Git | Unix | BigQuery | Tableau | ArcGIS | Google Colab | Jupyter Notebook | Excel | AWS

EXPERIENCE

Attentive*, New York, NY* **Sep–Dec 2022**
Machine Learning Engineer Intern

- Deploy state-of-the-art **neural network** model for **product recommendations** that outperforms existing model's performance
- Query into **Snowflake relational data warehouse** of 40 million different products over 8,000 distinct brands using **SQL**
- Load data of 8 million+ records and perform **EDA** on **Jupyter Notebook** via **AWS SageMaker** Notebook Instance Environment

Sony Music Entertainment*, New York, NY* **Jun–Aug 2022**
Data Science Intern

- Collected and **preprocessed** data pulled from internal **cloud databases** and Chartmetric/Spotify **API**s using **Python** and **SQL**
- Built **multivariate linear regression** model to predict Spotify first day streams; fine-tuned model from R^2 0.51 to 0.92 whilst validating model assumptions, which provided actionable **metric recommendations** to use when targeting new artists to sign
- Identified songwriter groups by track popularity using **unsupervised clustering** algorithms with **TF-IDF** method on text data and applied **PCA** to reduce vector dimensionality for cluster **visualization**, delivering songwriter collaboration recommendations

The Ripolles Lab, New York University*, New York, NY* **Oct–Dec 2021**
Data Analyst

- Developed **Python** algorithm that detects and visualizes music-induced goosebumps from real-time video of skin captured by prototype wearable sensor of **Raspberry Pi** architecture; expanded prototype use to 300 people in NYC area
- Set up virtual environments via **Conda**/**Python** on **Raspberry Pi Zero** that ensured **Python** script runs smoothly
- Utilized **Git** and worked on separate branches to allow code collaboration and to prevent merge conflicts

Ernst & Young*, Seoul, South Korea* **Sep–Dec 2020**
Technology Solutions Consulting Intern

- Collaborated on client-facing project to benchmark 8 Engineering Procurement Construction (EPC) companies' global operations
- Interviewed 40+ experts (C-level executives) on 30-120 min calls, successfully narrowing team's focus to 3 EPC companies
- Pivoted and adapted quickly as client guidance evolved, whilst always keeping the ultimate project objective in mind
- Summarized research on organizational structure & strategy into **Excel**/**PowerPoint**, effectively reporting to senior management

PROJECTS

Clustering Analysis of Korean Restaurants in NYC using Machine Learning and Python **Apr–Jun 2021**
IBM Data Science Professional Certificate

- Pulled and **prepared geodata** of 10K+ venues and 306 NYC neighborhoods (via **Foursquare API**) using **Python**
- Performed comprehensive **EDA** and **visualized** relationship between neighborhoods and number of Korean restaurants
- Employed **Elbow Method** to identify optimal number of clusters and **K-Means Clustering** to cluster neighborhoods based on similar mean frequency of Korean restaurants, identifying optimal locations to start Korean restaurant business in NYC

Factors Determining Sentence Length and Severity for U.S. Inmates using R **Jan–Mar 2019**
Dartmouth College Government Department

- Worked closely with 3 other undergraduates to conduct **multivariate linear** and **logistic regression** using **R** to investigate the statistically significant predictors of sentence length and severity for U.S. inmates across State and Federal Prisons
- Visualized results showing expected values of receiving an extreme sentence as a function of various predictors