

# SOMANG (SO) HAN

sohan1@seas.upenn.edu | +1 (507) 321-3319 | Philadelphia, PA | <https://somang1418.github.io>

## EDUCATION

**University of Pennsylvania** | Master of Science in Engineering in Data Science

Philadelphia, PA | May 2022

**St. Olaf College** | Bachelor of Arts in Mathematics & Statistics, Magna Cum Laude

Northfield, MN | May 2018

## PROFESSIONAL EXPERIENCE

**Bluestem Brands | Decision Scientist, Marketing | Eden Prairie, MN | Nov 2018 – June 2020**

- Extracted millions of customer data across multiple platforms (Hadoop, SQL Server, SAS) and transformed large complex data to generate features for new customer acquisition and customer marketing models in Python
- Led a team of 2 to develop a customer experience tracking tool that categorized customer complaints by K-Means Clustering in R Shiny and increased customer satisfaction of fingerhut.com by 10% in four months
- Built ML models using Python on AWS to identify customer segments for campaigns and improved sales conversion rate by 15%
- Supported product managers with ad-hoc analysis and statistical analysis to enable data-driven decisions by business leaders

**University of Minnesota | Research Fellow | Minneapolis, MN | June 2018 – Sep 2019**

- Conducted research in Educational Psychology with 3 Ph.D. students to explore the association between intensity and breadth of out-of-school activity participation with academic outcomes by Hierarchical Linear Modeling in R
- Published research findings at the American Educational Research Association conference 2019

**Amplifon | Associate Data Analyst | Minneapolis, MN | Sep 2018 – Nov 2018**

- Consolidated annual budget reports for sales of Miracle-Ear products, validated metrics and KPIs by manipulation of complex Excel formulas and Pivot Tables, and improved the process efficiency by 5%

**ID Insight | Data Scientist Intern | Minneapolis, MN | June 2018 – Sep 2018**

- Identified different types of bank fraud by K-Means Clustering in R; generated data visualization in Tableau; presented findings to C-level leadership
- Rebuilt a bank fraud detection model using Logistic Regression in R and improved model performance by 20%

**St. Olaf College | Statistical Consultant | Northfield, MN | Aug 2017 – May 2018**

- Utilized statistical techniques in R to provide recommendations to 2 to 3 students per week for their projects
- Conducted NSF funded research in Topological Data Analysis with 3 students and presented research findings at the Underrepresented Students in Topology and Algebra Research Symposium 2018

## SELECTED PROJECT

**Reducing Plastic Pollution in Oceans**

(Summer Project 2020 with U.S. Census Bureau)

- Developed digital tools to increase the public awareness of plastic pollution in oceans and utilized several metrics to alert one's plastic usage and encourage to make informed choices to reduce the growing threat of plastic pollution

**TMDB Box Office Prediction**

(Kaggle Competition: Top 6%)

- Transformed 7,000 movie data to generate features and built ensemble models (LightGBM/Random Forest) to predict their overall worldwide box office revenue

**Elo Customer Loyalty Score Prediction**

(Kaggle Competition: Top 11%)

- Preprocessed 30,000 customer data and built ensemble models (XGBoost/LightGBM) to predict customers' loyalty score for Elo, one of the largest payment brands in Brazil

**Kaggle Days SF Hackathon - Product Defection Classification**

(Kaggle Competition: Top 32%)

- Teamed up with the CEO of Kaggle and a friend to build bi-classification model (XGBoost/LightGBM) to predict the likelihood of 400,000 products being defective

**Topological Data Analysis on Network of Game of Thrones**

(Summer Research 2017)

- Examined Game of Thrones network map data by a Topological Data Analysis tool and implemented link analysis to closely look characters' betweenness centrality and degree centrality to find the most important community in the network map
- Presented results of the research at the Joint Mathematics Meetings (JMM) 2018

## PUBLICATIONS

- Han, Somang, Taylor Okonek, Nikesh Yadav, Xiaojun Zheng 2020. "Distributions of Matching Distances in Topological Data Analysis." **SIAM Undergraduate Research Online**, vol. 13, 2020, doi:10.1137/18s017302
- Youngsoon Kang, Kyle Nickodem, Somang Han, Mireya Smith, Michael Rodriguez 2019. "Intensity and Breadth of Out-of-School Time Activity Participation: Connecting Latent Classes, Student Characteristics, and Educational Outcomes." **Paper accepted at American Educational Research Association Conference** at Vancouver, Canada

## ADDITIONAL INFORMATION

- **Skills:** Python, R, SQL, Excel, SAS, Tableau, Hadoop, AWS, HTML, CSS
- **Leadership:** Ambassador @ Women in Data Science (WiDS) 2020, Sergeant at Arms @ Toastmasters International, Guest Speaker @ MinneFRAMA (Business Analytics Conference), Co-chair @ St. Olaf Korean Cultural Association
- **Kaggle Expert:** <https://www.kaggle.com/somang1418>