Ridhika Agrawal

ridhika123@gmail.com (224)241-6120 ridhika123 ridhikaagrawal

Education

- **New York University** — New York City, NY
  M.S. in Data Science — *Expected May 2024*
- **Grinnell College** — Grinnell, IA
  B. A. in Mathematics and Economics, with honors — *May 2020*
  - **GPA**: 3.8/4.0 | **Dean's List**: Spring 2018, Fall 2019 | **Grace Hopper Scholar**: excellence in STEM
  - Treasurer, International Student Organization | Math Student Educational Policy Committee | Statistics TA

Skills

- **Technologies:**
  - Python, SQL, R, C, Stata, SAS, Excel, Tableau, ArcGIS, VBA, Access, Minitab, Eviews
- **Tools:**
  - Data Analysis & Visualization: Matplotlib, Seaborn, NumPy, Pandas, dplyr, ggplot2
  - Machine Learning & Modeling: Scikit-learn, NLTK, GenSim, fastText, stats

Work Experience

- **Associate Analyst** — New York City, NY
  Securities and Finance, NERA Economic Consulting, Oliver Wyman Group — *Sep 2020 - Jun 2022*
  - Modeled 50K employee's unpaid overtime wages across 10 years using SAS and PROC SQL
  - Refuted damage estimation by identifying spuriousness and specifying 2SLS model with fixed effects in Stata
  - Projected $3.4 billion liability for illegal lending activity by wrangling 8 million+ row dataset in Python and SAS
  - Developed concise, easy-to-understand visualizations and tables in Excel to be included in legal expert reports
- **Behavioral Data Science Intern** — Los Angeles, CA
  mPulse Mobile [published paper link] — *May 2019 - Dec 2019*
  - Built neural network model with logistic activation function in Python to predict engagement with 78% accuracy
  - Conducted feature engineering using chi-square, t-test, correlation coefficients, normalizing and one-hot encoding
  - Built NLP pipeline and developed models using SVM to preprocess and label English and Spanish text messages
  - Composited indicator of health factors with hot deck imputation, equal weighting and k-means clustering
- **Data Science Consultant** — Grinnell, IA
  Data Analysis and Social Inquiry Laboratory — *Aug 2019 - May 2020*
  - Designed gradient tree boosted model using TDBoost in R to improve outreach by predicting alumni gift sizes
  - Assisted 30+ researchers with data cleaning, statistical modeling and coding constraints in a one-on-one capacity
- **Statistics Research Fellow** — Grinnell, IA
  Advisor: Dr. Jeffrey Jonkman [working paper link], Mathematics Department, Grinnell College — *May 2018 - Aug 2018*
  - Conducted meta-analysis using R to test efficacy of smoking cessation drug as a cure for alcohol dependence
  - Built mixed effects and random effects binomial models to measure drinking days, craving and adverse events
  - Used empirical Bayes and maximum likelihood methods to estimate heterogeneity and standardized means

Selected Projects

- **Recipe Recommendation System** [link]
  - Scrubbed 13K+ recipes using Beautiful Soup and Selenium, preprocessed data for feature extraction via BoW
  - Created representative vector using simple and TF-IDF weighted averaging on word embeddings by Word2Vec
  - Deployed streamlit app, which uses content-based filtering to recommend top recipes based on cosine similarity
- **Socio-Economic Determinants of Sanitation** [link]
  - Created panel data of 112 countries across ten years to explore factors that drive countries' access to sanitation
  - Performed linear regression analysis with logarithmic transformation and used a modified version of expectation-maximization algorithm to fit two-component beta mixture models in R
- **Influence of Ambient Light on Crime: Causal Inference** [link]
  - Designed sharp regression discontinuity model in Stata with daylight savings as exogenous shift in light
  - Wrote Python script to create 100 million+ row data of sunrise-sunset times for each latitude-longitude in the US
- **Suicide Rates Across Countries: An Exploratory Data Analysis** [link]
  - Conducted EDA on collected data using ANOVA, Tukey's test, and scatterplot of regressors faceted by income
  - Built multivariate regression models in R, evaluated accuracy using R-squared and normal probability curves