Kevin R. Pan

http://kevinrpan.com

Email : contact@kevinrpan.com
Mobile : +1-860-941-2128

Experience

- **The Brattle Group** — San Francisco, CA
  Research Analyst — *Jul 2016 - Present*
  - **Data Processing**: Sped up existing code run-time (by an average of 10x) without loss of functionality. Established firm-wide coding standards for R and SQL by giving formal trainings.
  - **Data Extraction**: Wrote and documented flexible functions for others to quickly pull new data productions. Created aggregate datasets from custom scraping hundreds of non-uniform PDFs, workbooks, metadata files, etc.
  - **Modeling**: Parallelized large model runs to finish a month's worth of model run-time in a week. Optimized performance of intensive scripts with automated memory flushing.
  - **Web Scraping**: Gathered bitcoin price data by scraping tables using Selenium browser automation and writing custom import functions using BeautifulSoup and rvest. Collected and processed data from various web APIs.
  - **Data Management**: Determined optimal process of data ingestion and management for data-heavy projects, balancing the trade-offs between database systems and in-memory manipulation.
  - **Data Presentation and Insights**: Discovered possible narratives based on data trends, supported by exhibits. Created interactive geospatial maps to visually examine patterns. Examined fund flows with D3 Sankey charts.
  - **Dimensionality Reduction**: Determined variable importance from hundreds of categorical variables using multiple correspondence analysis and t-SNE.
  - **Blockchain**: Calculated and created exhibits for implied interest rates from bitcoin spot and futures prices. Assembled blockchain research for Brattle whitepapers.
  - **Mapping**: Determined optimal zoning borders for various institutions by creating interactive geospatial maps.
- **Delta Analytics** — San Francisco, CA
  Data Fellow and Teaching Fellow (Nonprofit Volunteer) — *Jan 2017 - Present*
  - **Machine Learning Course**: Wrote IPython notebook modules on natural language processing, data visualization, Naive Bayes classification, and regression.
  - **BUILD**: Informed National Program Team about important features of program effectiveness. Provided markdown report with reproducible data cleaning, visualization, and regression tree analysis.
- **Bernoulli Health** — Philadelphia, PA
  Data Lead (Wharton Digital Health Club) — *Jan 2016 - Apr 2016*
  - **Bernoulli One**: Created model to assess health condition deterioration. Detected severe events 12 seconds ahead of the leading industry standard index (IPI). Presented analysis of false alerts and sensitivities to reduce nurse and doctor alarm fatigue.
- **Tudor Investment Corporation** — Greenwich, CT
  Quantitative Analytics (Intern) — *Jun 2015 - Aug 2015*
  - **Price Modeling**: Optimized bond price estimation for 30x speed, maintaining accuracy using SciPy and pandas. Unified modules for swap instrument pricing in Python.

Education

- **University of Pennsylvania** — Philadelphia, PA
  Bachelor of Arts in Mathematical Economics — Minors in Computer Science and Statistics — *Aug 2012 – May 2016*

Projects

- **'Handy' R Package**: Open source R package containing useful functions for common data handling tasks.
- **Shiny Web Application**: Data visualization built using R Shiny, using D3-style visualizations and live Google Sheet imports to analyze competition performance over time.
- **Kaggle NCAA**: Led team of 4 in modeling 2018 March Madness games with predictions ranking in top 30%.

Programming Skills

- **Languages**: R (package development, cleaning, visualization, modeling), Python (incl. scikit-learn), SQL
- **Technologies**: Excel, Selenium, Version Control (Git)