ERIC BOXER

ecboxer.github.io

Email : ecb2198@columbia.edu
Mobile : (347) 525-3516

Education

- **Columbia University** — New York, NY
  MS in Data Science; GPA: 3.8 / 4.0 — *Aug 2018 – Dec 2019*
  Applied Machine Learning, Databases, Machine Learning Pipelines, Algorithms, Network Analysis, Causal Inference, Statistical Inference and Modeling, Data Visualization

- **SUNY Stony Brook** — Stony Brook, NY
  BS in Mathematics and Economics, Minor in Philosophy; GPA: 3.9 / 4.0; Phi Beta Kappa — *Aug 2013 – Dec 2016*
  Symbolic Logic, Differential Equations, Topology and Geometry, Game Theory

Professional Experience

- **Swiss Re** *Associate Data Scientist* — Jan 2020 – Present

  Early Warning Signals: Developed a Python package to extract "signals" from an ElasticSearch database of news articles. Viewing the corpus as a network (with keywords as nodes and co-occurrence of keywords in an article as edges), we extract communities from the network and surface important articles for internal users in a PowerBI dashboard.

  Developed aspect-based sentiment analysis models for topics such as COVID-19 medical developments and client news. Model output was presented in a dashboard to medical experts and client managers to curate their news consumption.

  Accelerated Underwriting Trained an interpretable model to reduce referrals by 30%, for mid-sized US life insurer Sagicor, thereby reducing costs. Delivered applicant cohorts (learned in an unsupervised manner from self-organizing maps) and insights about the client's underwriting process that will be incorporated into their rules going forward.

  Plan Document Highlighting Created a tool for underwriters to more navigate plan documents (the "fine print" of an insurance contract). The tool OCRed scanned PDFs, split documents into meaningful chunks (via a custom spaCy sentence segmenter and tagger), selected relevant chunks (via regex matching from a pre-defined list of key terms), and presented output in a web app and annotated PDF file.

  Organized monthly meetups on projects related to data science for a global audience of Swiss Re employees.

- **Swiss Re** *Data Science Intern* — May 2019 – Dec 2019

  Early Warning Signals: Developed a novel outlier detection method for news articles, to identify potentially critical information for decision makers. Viewing the corpus as a network, we are able to identify important articles as those which form links between otherwise distant terms. Also, extended the node2vec method of computing node embeddings to include domain knowledge (by warm starting node embeddings) and early stopping (by computing link prediction scores on a hold-out validation network).

  Property & Casualty Solutions Performed exploratory data analysis, predictive modeling and risk scoring for marine insurer CV Starr. With financial and geospatial data from the client's portfolio, we explored drivers of attritional and named-event loss. Then, we incorporated in-house natural catastrophe models and publicly available fire incident data to create and validate policy- and broker-level risk scores. Recommendations were presented to the client, who expressed interest in continued engagement with the analytics team and reinsurance renewal with Swiss Re.

- **Frac.tl** *Data Visualization Developer* — March 2019 – June 2019

  Created visualizations and computed statistics for stories on medical student demographics and MLB attendance. Pulled data from physical sources about medical colleges and scraped websites for sports team attendance and wage statistics. Created interactive and static data visualizations in RShiny and ggplot2. For example, an interactive map of MLB ticket prices over the period 1950 to 2017, with icon size mapped to ticket prices and icon placement according to team location.

Data Science Projects

- **Pandas Contributor** *Columbia University* — Sep 2018 – Present

  Built documentation locally and pushed commits to the pandas-dev Github repo to be discussed with core contributors and merged into the master branch.

  Example pull request https://bit.ly/2DSbqjg

Programming Skills

- **Python**: Numpy, Pandas, PyTorch, Keras, Scikit-Learn, nltk, spaCy, Matplotlib, Scrapy
- **R**: Tidyverse, mlr3, base, stats, plsRglm, Rmarkdown
- SQL, Git, Elasticsearch, Linux, Javascript, D3, AWS, Azure, Jupyter, LaTeX, Emacs