# Kevin Urban, PhD

kevin.ddu@gmail.com • (973) 464-6833

https://www.linkedin.com/in/drkrbn
https://github.com/krbnite
https://krbnite.github.io/

## Education

**Ph.D., Applied Physics**
New Jersey Institute of Technology & Rutgers University, Newark, NJ
Dissertation: The Hydromagnetic Structure of the Polar Cap and Its Interaction with the Solar Wind

**M.S., Applied Physics (Minor: Applied Mathematics)**
New Jersey Institute of Technology & Rutgers University, Newark, NJ
Thesis: Synoptic Variability of a CIR-Driven Open-Closed Boundary During Solar Minimum

**B.S., Applied Physics (Minor: Applied Mathematics)**
New Jersey Institute of Technology & Rutgers University, Newark, NJ

## Work and Professional Research Experience

**Early Signal (Cohen Veterans Bioscience)**　　　　　**New York, NY • July 2018 – Present**
*Associate Director of Data Science & Digital Health (June 2019 - Present)*
blasdfasb

- **Customer Behavioral Models**　　　　(Python, Sklearn, TensorFlow, Keras, AWS EC2/Redshift, SQL)

*Senior Data Scientist (July 2018 - Jun 2019)*
blasdfasb

- **Database Evaluation and Implementation for Wearables-to-Biological-Phenomenon Knowledge Graph Neo4j, MongoDB, MySQL, Postgres, etc**
  Reviewed the functionality and pros/cons of graph (Neo4j, AgensGraph, AWS Neptune), document (MongoDB), relational (MySQL, Postgres, TimeScaleDB), and key-value (Redis, DynamoDB) databases, as well as enhanced AWS cloud offerings, such as RDS, Athena, and Aurora * Identified graph databases as providing the best support for many-to-many/any-to-any relationships inherent in our concept mappings * Cypher from Neo4j
- **Literature Review: Suicide and Self-Harm Predictors and Risk Factors**
  Developed domain expertise in suicidal ideation and behaviors * Synthesized and reported on the state-of-the-art modeling efforts from several siloed communities of research, including efforts involving wearable and mobile phone sensors, electronic health records, brain imaging, paralinguistics, natural language processing, neurocognitive tests, clinical instruments (questionnaires) used for screening, diagnosing, and monitoring patients

**WWE (World Wrestling Entertainment)**　　　　　**Stamford, CT • October 2016 – July 2018**
*Data Scientist, Advanced Analytics Team*
Lead on numerous business-driven analytics research and predictive modeling projects for WWE Network – an over-the-top (OTT) content distribution network that can be described as "Netflix for Wrestling" and currently supports nearly 2 million daily active users, generating 1-3 million new rows of viewership/behavioral data in Redshift database every day

- **Customer Behavioral Models**　　　　(Python, Sklearn, TensorFlow, Keras, AWS EC2/Redshift, SQL)
  Development of data processing and machine learning pipelines (preprocessing, feature selection, dimensionality reduction, model selection, etc), resulting in predictive classification models (usually random forests and/or deep multilayer perceptrons) whose outputs could be regularly used by stakeholders for downstream decision support. Behavioral targets include customer segmentation, churn, winback, and lifetime value.
- **Wrestler Recognition Network**　　　　　　(Python, TensorFlow, Keras, AWS, EC2, GPU)
  Development of wrestler image recognition algorithm (transfer learning, convolutional neural networks) that could be used to tag video and image content.
- **Seasonal Subscriber Analysis**　　　　　　　　　　(R, Python, FFT)
  Spectral analyses of customer churn/winback behaviors to enhance customer segmentation efforts, help advise email campaigns, and improve churn/winback forecasting
- **Revenue Attribution Across Content**　　　　　　(R, tidyverse, SQL, Redshift)
  Development of ensemble approach to revenue attribution models for WWE Network content driven by viewership behaviour at the customer level (analysis of over a billion rows of data in Redshift)
- **Miscellaneous Projects**　　　　　　(R, tidyverse, SQL, Redshift, Tableau)
  - Fusion of fan panel survey response data with customer accounts on the WWE Network to better **quantify the relationship between subjective responses and objective viewership data**

- **Sentiment analysis** of customer cancellation surveys and social media content to uncover resolvable issues and inform on how to better serve our customer base
- Design of **interactive dashboards in Tableau** to help interested parties explore our data assets

### *Data Scientist, Content Analytics Team*
Lead on researching and developing multiple data engineering and automation efforts to establish consistent and efficient data collection across WWE's digital assets, including a variety of YouTube channels, 100's Facebook Pages and Twitter accounts, and live streaming across multiple online platforms.

- **YouTube Assets**   (Python, Google Client API, YouTube Data API, YouTube Reporting API, Redshift, SQL, Selenium, Hive, Presto, Tableau)
  - **Research into and documentation** of several YouTube APIs (Google Client, YouTube Data, YouTube Analytics, YouTube Reporting), establishing that 3rd party services were capturing only a fraction of the available content and viewership data from our YouTube assets.
  - **Persuasion for buy-in from leadership**: 3rd party consultants responsible for the collection/warehousing of our YouTube assets argued they were already collecting everything possible; thanks to my research and documentation, key decision makers supported my plan to revamp the YouTube data pipeline.
  - **Development and deployment** of an automated YouTube data collection and warehousing pipeline over a variety of WWE-owned channels, increasing the daily volume of valuable YouTube more than tenfold, resulting in significant cost savings by obviating need for 3rd party consultants
  - **Design and development** of a variety of live Tableau dashboards and automated emails to the appropriate stakeholders.
  - **Shorter Time to Insight**: YouTube takes its viewership metrics (e.g., number of view, likes, etc) very seriously and has many algorithms in place to detect and correct for fraudulent sources (robots, automated page refreshes, etc). In this spirit, there is a 3-day lag on any officially published data. However, our marketers, advertisers, and other leadership want insights and intuition on content performance immediately. By using Selenium for webscraping, I was able to extract "unofficial" YouTube viewership data with no lag – especially important on the day a video is live or published. This is something the data vendor was not getting for WWE.
  - **Hive/Presto vs Redshift**: Worked with data engineering team on a Hive/Presto solution on top of S3 to lessen our dependence on Redshift, which is more expensive, especially for the amount of data we were ingesting

- **Facebook/Instagram Assets**                    (Python, Facebook Graph API, Selenium, Redshift, SQL)
  - My success evaluating and re-engineering our YouTube data pipeline motivated us to take a look at other services provided by 3rd party vendors starting with Facebook (WWE owns and manages 100's of Facebook Pages, several dozen Instagram accounts, as well as several Facebook exclusive web series that air live weekly)
  - Researched and documented various Facebook APIs and technologies, including the various web-based data portals, GraphQL, and the Facebook Graph API.
  - By learning how to use the Graph API, I was able to query, pull, and storing all Page, Post, and Video Insights on a daily basis, which required flattening JSON files into a tabular form amenable to SQL queries in Hive or Redshift
  - By comparing my daily pull with our vendors, I identified that the vendors were only providing a limited subset of the Page , Post, and Video Insights in our daily Facebook load (and no Instagram data)
  - I also learned how track live videos at a high frequency, thus providing data for live videos on our various Facebook Pages, as well as live webisodes streaming on Facebook Watch – data the vendors were not able to provide

- **Multi-Platform Real-Time Dashboard for Live Events**      (Python, YouTube APIs, Facebook Graph API, Selenium, Redshift, SQL)
  - Lessons learned from YouTube and Facebook pipeline development motivated the development of a live Tableau dashboard that could be published within minutes of any live airing of a "kickoff show" (aired on YouTube, Facebook, Twitter, the WWE Website, and the WWE Network) before a main wrestling event (aired only the WWE Network)
  - This required an immediate, automated capture of viewership metrics across all platforms the event aired live on, replacing a highly manual process performed by several employees over multiple departments, resulting in a 12-24 hour latency
  - Required the automation of headless web browsers (using Selenium, Chrome, BeautifulSoup, and Cron) to scrape JavaScript-heavy Twitter and WWE websites

- **R Packages for Automated Reporting**                       (R, tidyverse, Redshift, SQL)
  Developed R packages that the team could use to automate various reporting responsibilities

**NJIT Center for Solar-Terrestrial Research**                 Newark, NJ • January 2012 – May 2016
*Research Scientist*

- Fusion and harmonization of data sets from a ground-based, globally-distributed network of instruments; high-altitude satellites, such as NOAA's DMSP fleet and NASA's Van Allen Probes; and interplanetary spacecraft (NASA's ACE)
- Provided new, empirical insights and interpretation of decades-old problem concerning the geomagnetic field's complex structure in Earth's polar regions
- Developed prediction scheme to forecast intensity and spatial-temporal distribution the magnetic weather in Earth's deep polar cap region using spacecraft upstream of Earth in the solar wind (NASA's ACE)
- Inverted the prediction scheme to leverage it as a remote-sensing technique such that data from ground-based magnetometers in Antarctica can be used to infer space weather parameters in near-Earth space and, if necessary, could essentially replace in-situ spacecraft for certain measurements
- Created innovative visualization and analysis techniques to better understand and transform our understanding of the Earth magnetic geography and its dynamic response to varying space weather conditions
- Supported myriad data analysis efforts and projects as a member of the RBSPICE instrument team for NASA's Van Allen Probes mission, which has been essential in understanding and forecasting hazardous conditions of Earth's radiation belt environment
- Developed software packages in Bash, R, MatLab, and IDL to enable efficient, streamlined analyses and reporting for a range of ongoing projects
- Published first-author publication, PhD dissertation, and currently have several papers in development

### NJIT Department of Mathematical Sciences                      Newark, NJ • 2008 − 2012
*Research Scientist, Mathematical Modeling*
- Developed dynamical system models for granular fluid systems
- Simulated and visualized the dynamical models in MatLab and R
- Investigated sources of chaos and instability in such systems by projecting them into a dimensionally-reduced parameter space and coding simulations of their Poincare maps in R
- Led effort in understanding and controlling the propagation of numerical errors, which otherwise prevent proper quantitative analysis and visualization of the underlying dynamical system
- Co-authored four peer-reviewed papers numerically modeling, visualizing, and analyzing granular fluid systems
- Led research on potential engineering applications of fractional calculus and co-authored peer-reviewed publication
- Explored real-world applications of advanced mathematical disciplines such as dynamical systems, topological analysis, manifold theory, abstract algebra, stochastic calculus

### NASA/CalTech Jet Propulsion Laboratory                      Pasadena, CA • Summer 2011
*Intern, Trajectory Optimization / Design*
- Collaborated with a team of engineers and scientists to develop a full-fledged spacecraft mission to the Trojan asteroids of Jupiter (from establishing and prioritizing science goals, to optimizing the science-engineering-financial parameter space, to the written proposal and presenting our mission design to the NASA review board)
- Gained an appreciation of rapid product development via concurrent engineering, and an understanding of the intricate interplay between the various engineering system designs, science goals, timeline requirements, and budget constraints.
- Published peer-reviewed paper documenting lessons learned

### Independent Research Consultant                      Clifton, NJ • August 2010 - January 2012
- Procured contracts from NJIT's Physics Department and Siena College's Department of Physics and Astronomy to continue the development of automated, real-time prediction of geomagnetic parameters using geospatially-distributed network of magnetometers
- Published peer-reviewed paper documenting proof-of-concept and lessons learned proof-of-concept [link]

### NJIT Department of Physics                      Newark, NJ • August 2008 - August 2010
*Research Assistant*
- Led project demonstrating the feasibility of a real-time surveillance, detection, classification, and prediction of local, regional, and global-scale geomagnetic events of interest
- Analyzed multi-channel instrument data from a spatially-distributed network of automated observatories in Antarctica.
- Developed data metrics and associated classification scheme for events of interest in the data sets.
- In-depth experience visualizing and analyzing data in R, MatLab, and IDL
- In-depth experience in various data analysis techniques, such as digital signal processing, time series analysis, spectral analysis, statistical methods, modeling, and regression.
- Hardware-software interfacing of instrumentation for data collection
- Published Master's thesis

### NASA Goddard Space Flight Center                      Greenbelt, MD • June − August 2007
*Intern, Observational Cosmology Laboratory*

- Developed software implementing a Stokes parameter analysis to run on simulation data for the proposed Absolute Spectrum Polarimeter [ASP] — an instrument designed to detect B-mode gravitational waves, which would provide evidence for Einstein's theory of general relativity and cosmological inflationary theory
- Worked with multiple programming languages, including MatLab, Bash (shell scripting), and IDL

**NASA Goddard Space Flight Center**          **Greenbelt, MD • June – August 2006**
*Intern, Heliophysics Division*
- Worked extensively in the UNIX programming environment, learning and developing Python, HTML, CSS, JavaScript, and PHP code
- Developed web content

## Selected Peer-Reviewed Publications

- Rethinking the polar cap: Eccentric-dipole structuring of ULF power at the highest corrected geomagnetic latitudes, Journal of Geophysical Research, Space Physics, Vol. 121, 2016.
- Analysis, simulation, and visualization of 1D tapping via reduced dynamical models, Physica D: Nonlinear Phenomena, Vol. 273, 2014.
- Mission to the Trojan Asteroids: Lessons learned during a JPL Planetary Science Summer School mission design exercise, Planetary and Space Science, Vol. 76, 2013.
- Dynamical systems model and discrete element simulations of a tapped granular column, Powders and Grains 2013: Proceedings of the 7th International Conference on Micromechanics of Granular Media. Vol. 1542. No. 1. AIP Publishing, 2013.
- Quiet-time observations of the open-closed boundary prior to the CIR-induced storm of August 9, 2008, Space Weather, Vol. 9, S11001, 2011.
- Tapping dynamics for a column of particles and beyond, Journal of Mechanics of Materials and Structures, Vol. 6, No. 1-4, 2011.
- Integrability analysis of regular and fractional Blackmore-Samulyak-Rosato fields, Condensed Matter Physics, Vol. 13, No. 4, 2010.

## Continuing & Extracurricular Education

**Google Developer Scholarship (Web Development) 2018**      Udacity (Currently Enrolled)
**DeepLearning.AI**      Coursera (Currently Enrolled)
**Deep Learning Nanodegree Foundation**      Udacity

## Computing Skills

**Languages & Tools**: Proficient in Python (Jupyter, TensorFlow-GPU, Keras, SciKit-Learn, Pandas, NumPy, MatPlotLib, Seaborn, Requests, BeautifulSoup, Selenium, etc), R (dplyr, tidyr, ggplot2, magrittr, RStudio, R Markdown, etc), SQL (Redshift, PostgreSQL, Hive / HiveQL, Presto, SQLite), Amazon Web Services (Redshift, S3, EC2, EBS), Tableau, UNIX (OS X) and Linux (Ubuntu) Shell Scripting (bash, awk, sed, grep, wget, ssh, sftp, crontab, etc), Git, Github, MatLab, IDL, LaTeX, Vim, HTML, CSS, JavaScript, MarkDown, Excel, Google Sheets; Limited Experience and/or Classroom exposure to Spark, Java, C/C++, Fortran

**Data Science & Analytics**: Structured & Unstructured Data (e.g., sensors, scientific instrumentation, spacecraft, images, text, time series, geospatial, simulation, marketing, viewership, commerce, HTML, JSON, XML, social media); Data Pre-Processing (e.g., table joins, up/down sampling, outlier detection, imputation, aggregation, feature selection / engineering / extraction, regex and rules-based scrubbing / transformations); Probability & Statistics (e.g., inference, hypothesis testing, resampling, conditioning, error analysis); Machine Learning (e.g., neural networks, deep learning, CNNs, RNNs, reinforcement learning, statistical models, regression, classification, clustering, prediction, PCA, LDA, decision trees, random forests, sentiment analysis, natural language processing); Time Series Analysis (e.g., parametric, nonparametric, spectral analysis, digital signal processing, digital filtering, forecasting, wavelets); Numerical Modeling (e.g., difference equations, numerical derivatives, numerical integration, linear/nonlinear models, dynamical systems, numerical stability, eigenvalue estimation)