

**RECENT PROFESSIONAL EXPERIENCE**

---

**Maritz Motivation, St. Louis, MO**

*Maritz Motivation pioneered the incentive and rewards industry over 100 years ago and continues to innovate by combining behavioral and data science to design and deploy strategies to optimize customer and employee behavior.*

**Director & Lead Data Scientist, Decision Sciences, 1/2019-Present**

Provide leadership on matters related to data pipeline and workflow, insight and model generation for clients and internal research and development. Responsible for establishing and applying data science and methodological standards across the Decision Sciences group and for developing the first ever data products leveraging cross-client, cross-industry data across the organization.

- Spearheaded the development of Maritz Motivation's first cross-client data product. Historically, models were built custom for individual clients without the benefit of cross-industry, cross-client data. This proved more costly and time consuming, as well as led to less accurate models for business outcomes such as retention and sales. User behavior across multiple clients from Maritz's employee recognition platform was collected into a single location (BigQuery). Data was normalized to account for how companies used recognition and the platform differently (e.g. higher volume of recognition, quarterly versus monthly point dispersions, etc.). Various predictors were developed and tested, with those summing behavior over a previous 6-month period having the best predictive performance. Generalized additive models with random effects were used to predict the likelihood of attrition at 1-, 3-, and 6-months into the future. The models were able to handle the different underlying relationships between tenure and attrition at each company, as well as each predictor's relation to attrition. Workgroup-level aggregate risk scores were developed to ensure anonymity, with high risk groups exhibiting 5x the attrition of low risk groups. Data and predictions delivered to clients through a Shiny application that included an interactive collapsible tree diagram to visualize and explore a company's organizational structure and risk scores.
- Developed a valuation methodology that serves as the standard through which Maritz communicates to clients the ROI of their investment in their employee recognition program. Individual recognition behaviors (i.e. receiving point-based recognition from a manager compared to peers, logging into the platform, giving a recognition, etc.) were each assigned a value based on how one unit of each behavior impacted an employee's likelihood to be retained while controlling for other behaviors. The value was derived from a generalized additive model that predicted retention as well as standard replacement costs of employees (see above). Client recognition strategies were not only valued more accurately, but also adjusted based on previously unknown insights around which behaviors provide the greatest ROI. The methodology is used as the basis for valuing all other vertical behaviors and programs (e.g. Sales Performance).
- Revamped a major automobile manufacturer's survey fraud detection process by creating a machine learning-based model to better identify potential cases of salesperson tampering, leading to financial loss for the OEM. Previous process identified suspicious customer experience surveys for investigation through simple rules and heuristics. Historical survey data and fraud investigation records were combined, and previously unused features created, such as whether IP addresses or emails were reused on the same day, hour of survey completion, rates of completion for non-required items on the survey, etc. A random forest was fit on balanced, synthetic data to guard against majority-case bias and make the model more robust to unseen cases. The model performed very well on test and validation data sets, recovering actual tampering cases over 50% of the time compared to 15% for the previous model.
- Devised and built a novel customer segmentation approach for a leading e-commerce company. Previously, client relied on blast communications that lacked sophisticated targeting to generate recurring revenue from its large customer base. Built initial clusters to identify those customers that were highly profitable, high revenue, and loyal over a 5-year period. Clusters were built using novel features generated from product purchase patterns as well as cost, margin, and discounts used and a combination of UMAP dimension reduction and k-means clustering. A classification model was built using random forests to identify which customers had the highest likelihood to evolve into customers similar to those in the two highest value clusters based on their first 12-months of activity. The model accurately identified these customers with a recall of 75%, and while precision was low (32%), the model was actionable in identifying a narrower group of high potential customers to target with customized offers to boost loyalty.
- Built a workflow for the Auto division, whereby a previously manual data collection and aggregation effort through Excel worksheets was automated and expanded using Google Drive, Cloud Storage, BigQuery, Data Studio, and R. Individual consultants filled out forms for thousands of dealerships in Excel and uploaded the forms to Google Drive. The daily R script downloaded, transformed, and combined the files and uploaded the combined response file to Google Cloud Storage via APIs. The script then loaded the new file daily into a BigQuery table, which was then used as the basis for a reporting solution designed in Google Data Studio. The workflow included extensive

error handling and logging, as well as the sending of automated reports to key internal stakeholders. Process made the expansion of the project feasible for over 2000 dealerships and saved the equivalent of \$1.4M in labor costs.

- Lead various learning sessions for colleagues that covered such topics as using hierarchical linear models, generalized additive models, developing applications with Shiny, calculating marginal effects for non-linear and non-parametric models, and web-scraping. Sessions provided practical introduction to each topic, as well as sample code and use cases specific to our internal and external clients.
- Create and maintain various internal R packages to standardize repeated operations within the Decision Sciences group, such as ETL of client and internal data, data transformation and model application preparation, and data visualization.

#### **The Gallup Organization, Washington, DC**

*Gallup has studied human nature and behavior for more than 75 years, delivering unrivaled insights into political, socioeconomic, consumer, and employee opinions and behaviors.*

#### **Global Data Strategist & Subject Matter Expert: Predictive Analytics & Data Science, 6/2012-12/2018**

Led global strategy development and implementation for Gallup in the areas of advanced analytics, modeling, and automation. Responsible for direct day-to-day management and development of five data scientists and statisticians, as well as recruiting, hiring, training, and development strategy for all advanced analytical consultants globally.

- Supported pursuits globally involving private businesses, government agencies, and research institutions where automation and predictive modeling played a significant role. Additionally, led the design and oversaw execution of custom studies and predictive model builds for clients.
- Developed global standards for conducting various analytical tasks, including (but not limited to) derivation of key drivers using dominance analysis and hierarchical regression as well as end-to-end predictive modeling protocols (e.g. standards for transforming and splitting data, which algorithms to use and when, acceptable metrics of evaluation for models, visualization of results).
- Designed and led the execution of a forecast improvement project for a major global supplier of key components for mobile electronic devices. The ensemble model produced quarterly forecasts at 3-, 6-, 9-, 12-, and 18-month windows for aggregate consumer demand and was based on regression-based weighting of multiple modeling techniques—e.g. ARIMA, Support Vector Machines, and k-Nearest Neighbors. Model leveraged diverse data sets from multiple sources (government, 3<sup>rd</sup> party, web-generated, and survey generated). The out-of-sample, quarterly mean absolute percentage error of the ensemble model reduced the error of the existing model by 77%. Additionally, provided framework and guidance on general model evaluation criteria for client to revamp their on-going forecasting processes across the organization, including staffing and organizational structure.
- Served as subject matter expert and lead architect of a decision science tool for a leading global non-profit. Goal was to maximize accuracy and provide transparency into major drivers of donation forecasts so that leadership could decide whether to use their limited resources in a given time period to boost donations, or whether external or environmental forces (e.g. economic confidence, disposable income, other external events, etc.) in a given period were likely to be the most dominant forces. Forecasting model was built using multivariate adaptive regression splines (MARS). Predictions generated at 1-, 2-, 3-, 4-, 5-, and 6-month windows into the future. Simulation tool provided the identification of the largest drivers of donations and allowed key executives to conduct “what-if” analyses by adjusting values of key features for a given time period.
- Developed a turnover model for a national educational firm focused on predicting instructor churn at the location level on a monthly basis, as well as uncovering insights into ways to reduce current turnover rates. Modeling leveraged data from disparate client systems (i.e. HRIS, employee survey and selection instruments, operational and accounting systems, facilities) and third-party data (i.e. local unemployment and wage data, competitor data, census data). Logistic regression, generalized additive models, and hierarchical regression were used to predict individual turnover within a 30-day window. Final model achieved strong precision and recall, and uncovered previously unknown importance of different recruiting channels, relative wages and unemployment rates, and critical moments in the on-boarding process. Application led to an initial 33% decrease in turnover and helped decrease time-to-fill by 27% by using predicted turnover as the trigger for the recruiting process.

#### **Senior Consultant, 6/2010-6/2012**

- Lead consultant, managing teams of 10-20 consultants executing complex engagements and responsible for managing relationships with strategic clients in the consumer goods, health care, hospitality, and pharmaceutical industries. Primary responsibility for the methodological design and execution of client research, as well as the P&L of engagements, revenue growth, and resource utilization for a \$3M+ portfolio.
- Designed and executed customized, mixed-methodology research for a world-leading luxury hotel management company to determine both the revenue potential of group guests as well as the causes of their disengagement with the brand. Analysis proved the significant revenue potential for this previously ignored guest segment as well as the key drivers for their loyalty and share of wallet. Identified various areas of improvement for the client,

particularly service and defect handling. Analysis and recommendations led to a \$30M+ increase in annual revenue.

- Created a decision framework for Area Directors and Branch Managers of a top-three global financial services firm in the United States, designed to align their management practices with the drivers of affluent customer loyalty and spend. The framework incorporated decision tree modeling of future branch performance using operational data, custom metrics built through logistic and dominance modeling of customer experience scores, and ethnographic research within the branch network. Use of the framework led to a doubling of customer experience scores in a three-month period, as well as a 200% increase in premium product sales and a 115% increase in cross sales compared to control branches. Client estimated an additional \$60M in annual profit among affluent customers based on a system-wide rollout.

## ADDITIONAL ACTIVITIES

---

### **Advisor, Data Science, Sport Stack, 2020-Present**

- Advisor to an early stage start up on issues related to data, analytics, and modeling of sensor and tracking generated data for college and professional baseball players and teams
- Co-led the creation of company's first product for use with TrackMan and Rapsodo data feeds, allowing for coaches at Division 1 and 2 schools to easily track player performance and identify key areas of focus, both process and outcome. Applications significantly reduced time of data compilation (30-40 hours per week) and optimized coaching focus for areas of improvement.

### **Multiple Major League Baseball Teams, Department of Baseball Operations, 2013-2017**

- Work covered by NDA.
- Provide advanced analytical consulting for several major league baseball front offices.
- Analysis used to inform current and future roster decisions (e.g. free-agents, trade targets, extension candidates, etc.) as well as in-game strategy (e.g. how specific pitchers should approach specific batters, and vice versa, game plan based on individual umpires and their ball-strike tendencies, etc.) and park renovations.

### **Public Projects**

- **Crash Course in R:** Created an introductory tutorial for those trying to get started with R from scratch. Formed the basis of a live courses delivered to dozens of colleagues to jump start the adoption of R at Gallup (<https://bit.ly/2ZTWg9G>)
- **The baseballr package for R:** Developed a package for the R programming language that includes various functions for acquiring and analyzing baseball data. Sample functions include the ability to scrape MLB and NCAA player data over custom time frames, calculate pitch tendencies to different parts of the strike zone, and calculate advanced metrics on user supplied data sets. (<https://goo.gl/ZnyFYL>)
- Additional projects can be viewed at [billpetti.github.io](http://billpetti.github.io)

## EDUCATION

---

### **University of Pennsylvania, Philadelphia, PA**

- M.A. (2006), pursued Ph.D. (ABD, 2013) Political Science
- Awarded Distinction, Comprehensive Exams
- Awarded the William Penn Fellowship (2003-2008)

### **Temple University, Philadelphia, PA**

- M.A. (2003), Political Science
- Awarded Distinction, Comprehensive Exams
- Awarded, University Graduate Student Fellowship (2001-2003)

### **The College of New Jersey, Ewing, NJ**

- B.A. (2000), Political Science (Honors Program)

## SKILLS

---

- **Statistics/ML:** Variety, including linear regression, logistic regression, hierarchical regression, generalized additive models, decision trees, random forests, stochastic gradient boosting, k-nearest neighbors, Gaussian mixture models and associated clustering methods (k-means, k-modes, etc.), UMAP dimensionality reduction.
- **Technical Stack:** R, MySQL, PostgreSQL, Git, Tableau, h2o, DigitalOcean, Google Cloud Platform (BigQuery, Storage, Data Studio), Bash, Shell.
- **Qualitative Skills:** Include small-n research design, focus groups, as well as ethnographic/observational research.