KOBE SARAUSAD

Sammamish, WA, USA
Email: kobesarausad@gmail.com | Phone: 425-449-3681
LinkedIn | Portfolio

PROFESSIONAL SUMMARY

Data scientist and analytics professional with over 2 years of experience in sports analytics and revenue optimization. Expertise in developing predictive models for demand forecasting, customer segmentation, and anomaly detection using machine learning and statistical techniques. Proven track record of automating workflows and creating actionable business intelligence dashboards. Strong background in Python, R, and SQL with experience in cloud platforms including AWS, Azure, and Snowflake.

EDUCATION

University of Washington Seattle, WA
Bachelor of Science in Statistics: Data Science *Sep 2019 – Aug 2023*

- GPA: 3.53/4.00
- **Relevant Coursework:** Data Structures and Algorithms, Foundational Skills for Data Science, Statistical Computing, Data Visualization, Machine Learning, Applied Regression and Analysis of Variance, Nonparametric Statistics

PROFESSIONAL EXPERIENCE

Seattle Mariners Seattle, WA
Analyst, Revenue Insights *Oct 2022 – Present*

- Developed predictive models for demand forecasting, ticket sales projection, and customer lifetime value estimation using XGBoost, scikit-learn, TensorFlow/Keras neural networks, and Bayesian mixed-effects models with brms and rstanarm in R
- Built customer segmentation models using K-means, hierarchical clustering with scikit-learn, and neural network architectures to predict fan engagement tiers from behavioral data across 100K+ accounts
- Designed Bayesian attendance forecasting models incorporating promotional effects, opponent strength, temporal patterns, and weather variables; implemented using Stan and MCMC sampling methods
- Created promotion optimization system using linear programming with PuLP and lpSolve to maximize ROI across 81-game season while managing budget constraints and operational capacity
- Conducted queuing theory analysis using M/M/c models for stadium operations, simulating gate entry scenarios in R to optimize scanner capacity and reduce wait times by 15%
- Produced and maintained PowerBI dashboards with real-time KPIs, integrating data from SQL Server, Snowflake, and third-party APIs to inform executive decisions
- Automated reporting workflows using Python with pandas and numpy, and R with tidyverse and data.table, saving 10+ hours weekly and improving data pipeline reliability

University of Washington Seattle, WA
Data Science Intern *Sep 2022 – Aug 2023*

- Built D3.js interactive visualizations to translate deep learning model results for non-technical stakeholders, improving communication of complex analytical findings
- Fine-tuned deep learning models including LSTMs and transformers using TensorFlow and PyTorch on AWS SageMaker to generate forecasts of student academic performance using engagement data from learning management systems
- Experimented with large language models via AWS Bedrock to build a student chatbot for academic recommendations and advising support
- Developed scalable data preprocessing pipelines using AWS S3 for data storage and feature engineering workflows to support predictive modeling initiatives

Major League Baseball (MLB) New York, NY
Analytics Intern *Jun 2022 – Aug 2022*

- Created comprehensive data dictionaries and performed data mapping for relational databases in Snowflake cloud warehouse, improving data governance and cross-team accessibility
- Segmented MLB.tv customers using K-means and hierarchical clustering with scikit-learn to identify high-value customer segments and enhance targeted marketing campaign effectiveness
- Automated reporting workflows in R using R Markdown, knitr, and parameterized reports to streamline analytics tasks and reduce manual reporting time by 40%
- Collaborated with cross-functional teams including marketing, product, and engineering to translate business requirements into analytical solutions

RESEARCH & PROJECTS

NBA MVP Prediction (SPA DRP) | *R, NLP, Machine Learning*

- Predicted NBA MVP by combining season statistics with Reddit sentiment analysis using NLP techniques (tokenization, sentiment scoring) and the tidytext package in R
- Implemented ensemble machine learning models using randomForest and XGBoost to integrate structured statistical data with unstructured text features
- Achieved 85% predictive accuracy through feature engineering combining performance metrics (PER, win shares) and public sentiment indicators

American Homelessness Data Visualization (CSE 412) | *Vega-Lite, Interactive Visualization*

- Created interactive Vega-Lite visualizations with linked views and cross-filtering to present U.S. homelessness trends across states, demographics, and time periods
- Designed responsive interfaces with tooltips and drill-down capabilities, enabling stakeholders to explore multi-dimensional datasets and uncover regional patterns

Projecting Crime Frequency (Datathon 2023) | *D3.js, Deep Learning, Time Series*

- Compared deep learning architectures including LSTMs and temporal convolutional networks (TCN) using TensorFlow/Keras for spatial-temporal crime prediction
- Authored an interactive D3.js dashboard to visualize model performance metrics, prediction accuracy by crime type, and spatial distribution of forecasts

Uncovering Neglected Pro Bono Cases (DataFest 2023) | *R, Survival Analysis*

- Applied survival analysis using Cox proportional hazards models (survival package in R) to legal case data, modeling time-to-resolution as a function of case characteristics

- Identified high-risk cases requiring additional resources based on hazard ratios, case complexity scores, and historical resolution patterns

2020 Election Exit Poll Analysis (STAT 403) | *R, Bootstrap, Logistic Regression*

- Analyzed exit poll data using Bootstrap resampling methods (boot package) and logistic regression (glm) to predict state-level election outcomes based on demographic variables
- Implemented uncertainty quantification through 10,000 Bootstrap iterations to generate 95% confidence intervals for predicted vote shares

KPI Dashboard (Court1 Consulting) | *R Shiny, AWS, API Integration*

- Integrated Stripe, Toast, and business APIs with AWS Lambda functions (Python) and S3 for automated ETL pipelines, processing 50K+ daily transactions
- Created actionable R Shiny dashboards with real-time KPI tracking, custom metrics calculation, and automated email alerts for threshold breaches
- Implemented secure data storage and retrieval using AWS S3 buckets with proper IAM policies and encryption protocols

TECHNICAL SKILLS

Programming Languages: R, Python, SQL, JavaScript, HTML/CSS

Statistical Methods: Machine Learning, Bayesian Statistics, Time Series Analysis, Survival Analysis, Bootstrap Methods, Linear and Nonlinear Regression, Multivariate Analysis, Clustering, Classification, Nonparametric Statistics, Hypothesis Testing, MCMC Sampling

Machine Learning Frameworks: TensorFlow/Keras, PyTorch, scikit-learn, XGBoost, LightGBM, CatBoost, H2O.ai

Python Packages: pandas, numpy, scipy, statsmodels, matplotlib, seaborn, plotly, PuLP (optimization), NLTK, spaCy, transformers (Hugging Face)

R Packages: tidyverse (dplyr, ggplot2, tidyr), data.table, caret, brms, rstanarm, lme4, survival, glmnet, randomForest, lpSolve, shiny, R Markdown

Statistical Software: Stan, JAGS, BUGS, WinBUGS

Development Tools: Jupyter Notebook, RStudio, Git/GitHub, Docker, VS Code

Visualization: Tableau, PowerBI, D3.js, Vega-Lite, ggplot2, Plotly, matplotlib, seaborn

Cloud Platforms & Databases: AWS (Lambda, S3, EC2), Azure, Snowflake, SQL Server, PostgreSQL, MySQL

Business Tools: Microsoft Excel, PowerPoint, Word, Google Suite

Languages: English (Native), Japanese (Professional Working Proficiency)

ADDITIONAL INFORMATION

Professional Interests: Sports Analytics, Revenue Optimization, Predictive Modeling, Customer Analytics, Operations Research, Queuing Theory, Fantasy Sports Analytics

Certifications & Training: Machine Learning Specialization, Statistical Computing, Advanced Data Visualization

Professional Affiliations: Society for American Baseball Research (SABR), American Statistical Association (ASA)