

OBJECTIVE

Passionate about leveraging AI and data science to solve real-world and high-impact business challenges.

Always exploring new ways to optimize business functions and push the boundaries of AI- and data-driven decision-making.

CONTACT

- danish.june87@gmail.com
- (479) 595-6093
- Dallas-Fort Worth, TX
- [LinkedIn Profile](#)

DEGREES

MIS, 2016 – University of Arkansas
MBA, 2011 – Jaypee University
B.Tech (CS Engg.), 2009 – Uttar Pradesh Technical University

SKILLS

- SQL
- Python
- Cloud Tools – AWS, GCP
- Generative AI, RAG

Danish Khan

DATA SCIENTIST

I am an award-winning Data Scientist with a strong background in machine learning, AI, and analytics-driven decision-making. I specialize in building and deploying predictive models, automation tools, and AI-powered solutions that drive efficiency and business impact.

EXPERTISE

Predictive Modeling & Forecasting, Optimization, Generative AI, RAG, Causal Inference, Reinforcement Learning, Risk Analytics, Business Intelligence

DOMAIN

Marketing/User Acquisition, Product, Finance, Community Management

INDUSTRY IMPACT

Designed AI-driven tools for campaign optimization, fraud detection, customer sentiment analysis, and more, with significant cost savings and revenue maximization

RECOGNITION

4 Industry Awards for AI and Data Science innovations

RESEARCH

Published Researcher in peer-reviewed journals.

MEMBERSHIP

Senior IEEE

Core Expertise

- **Predictive Modeling & Forecasting:** Building machine learning models to predict user behavior, lifetime value (LTV), churn, and revenue.
- **Data-Driven Marketing & Growth Strategies:** Utilizing ML-driven insights to refine customer segmentation, personalization, and ROI-driven campaign management.
- **Generative AI & Retrieval-Augmented Generation (RAG):** Implementing AI-powered solutions for natural language understanding, customer insights, and operational automation.
- **AI-Powered Automation & Optimization:** Developing in-house AI/ML tools to automate marketing optimization, fraud detection, and sentiment analysis, significantly improving efficiency and cost savings.
- **Business Intelligence & Visualization:** Creating dynamic, self-serve dashboards in Sigma, Looker, and Tableau for real-time decision-making.
- **Causal Inference & Experimental Design:** Leveraging causal models to identify high-value actions (HVAs), understand factors affecting key metrics such as retention, and drive growth strategies.
- **Reinforcement Learning for Decision Optimization:** Applying RL techniques for ad spend optimization, budget allocation, and campaign strategy improvements.
- **Fraud Detection & Risk Analytics:** Designing heuristic-based fraud detection tools that outperform industry solutions in identifying anomalies in incentivized campaigns.
- **Data Engineering & Scalable ML Pipelines:** Processing and integrating high-volume datasets from BigQuery, Snowflake, and other sources for robust analytics and model deployment.
- **Deep Learning & Neural Networks:** Experience with CNNs, boosting techniques, and dimensionality reduction for advanced modeling.

Market Intelligence Platform – Discord Sentiment Analysis

Dates: 03/2024 – 04/2024

Role: Data Scientist

Overview:

Developed an AI-driven platform to analyze player sentiments and feedback from Discord channels using natural language processing (NLP). The tool leverages LLMs and OpenAI's Assistants API to automate sentiment and thematic analysis, providing actionable insights for game developers and community managers.

Key Features:

- **Automated Data Retrieval:** The platform connects directly to Discord servers, allowing users to fetch messages from specific channels or threads within defined date ranges and query raw chat data in natural language. The tool drives specific insights from thousands of raw messages which otherwise would be practically inefficient or costly to perform
- **AI-Powered Sentiment and Thematic Analysis:** Leveraging latest AI tech including LLM and Assistants API, the tool goes beyond basic sentiment tagging by identifying recurring themes and nuanced feedback. It categorizes discussions into topics like game mechanics, player frustrations, and suggestions for improvement, offering a comprehensive view of player sentiments.
- **User-Friendly Interface:** Built with Streamlit and integrated with OpenAI APIs, the platform provides an intuitive interface for community managers. This design enables efficient processing of large data volumes and delivers real-time feedback to developers

Impact:

- Reduced manual analysis time by 80%, enabling faster responses to player concerns.
- Improved player satisfaction and retention by providing data-driven insights for game improvements.
- Integrated with 5 games, showcasing scalability and adaptability.

Achievements:

- Recognized with 3 industry awards in the field of Artificial Intelligence for developing this tool
 - Noble Award
 - Titan Business Award
 - Titan Innovation Award
- Received top management appreciation for the tool's scalability and engineering excellence.
- Independently designed and developed the platform, addressing a critical pain point for community and product managers.

Online Article: [Building an AI-Powered Tool to Uncover Key Player Insights from Discord](#)

Unified Campaign Management Tool

Dates: 04/2024 – 06/2024

Role: Data Scientist

Challenge:

Campaign management and optimization was a challenging and inefficient process, requiring time-consuming data retrieval and consolidation from multiple sources, customized calculations for different campaign types, and manual decision-making for each ad network. This labor-intensive workflow left the marketing team burned out, with no time to focus on management or strategic initiatives. Additionally, the performance of campaigns was subpar due to delayed optimizations and late marketing decisions, resulting in missed opportunities and higher costs.

Product Overview:

The Unified Campaign Optimization Platform is a Python-based solution designed to streamline campaign management and optimization for marketing teams. By integrating backend systems, data pipelines, and advanced predictive modeling, the platform enables faster and more strategic decision-making, reducing dependency on manual workflows.

Key Features:

- **Data Integration and Automation:** The tool retrieves raw data from backend and other integrated sources, processes costs, and applies adjustments to address inconsistencies in campaign data.
- **Predictive Modeling:** The platform is integrated with a machine learning model trained on granular datasets to predict LTV (player's lifetime value) by Day 3 (D3).
- **Optimization Framework:** The tool includes custom functions to accurately measure the performance of various ad campaigns at granular levels and provides optimized bid suggestions to maximize campaign efficiency and overall profitability.
- **Automated Reporting and Optimization:** The tool automates the generation of performance and optimization reports, which are used for both internal manual decision making and direct communication with ad networks for campaign adjustments. This automation significantly reduces time spent on report creation and enables rapid, data-driven decision-making.

Impact:

- **Significant Time Savings:** The tool automated key aspects of campaign management and optimization, reducing manual workload by over 80%. This allowed me to focus on strategic tasks, enhancing overall efficiency and productivity.
- **Significant Cost Savings:** Similar tools in the market come with hefty subscription costs, and the company was considering purchasing third-party solutions for these tasks. By independently designing and developing this tool in-house, I not only replicated the functionality but also saved the company a substantial amount in subscription fees, showcasing the value of innovative in-house solutions.
- **Faster and Accurate Decision-Making:** By integrating LTV models and having the predictions available on Day 3 (D3), the tool enabled decisions 4 days earlier compared to traditional methods. This early optimization reduced lag time and saved substantial costs by minimizing inefficiencies in campaign performance early on and without having to wait for over 8 days.
- **Streamlined Reporting and Collaboration:** Automated generation of performance reports provided timely updates to ad networks, enabling prompt actions and faster campaign optimization. This proactive approach significantly reduced costs by minimizing delays and avoiding high user acquisition expenses from late or unoptimized decisions.
- **Improved Campaign Outcomes:** The tool's data-driven insights and early optimizations directly contributed to higher ROI, improved campaign efficiency, and overall better marketing performance.

Achievements:

- Recognized with Noble awards in the field of 'Data & Analytics' for conceptualizing and developing this tool
- Independently conceptualized and developed the platform, transforming manual workflows into automated, data-driven processes.
- Managed the entire marketing team's workload single-handedly during a critical transition, ensuring business continuity.

Online Article: [From Chaos to Clarity: How a Unified Campaign Management Platform Transformed Marketing Operations](#)

Predictive LTV Modeling for Marketing Campaigns

Dates: 06/2022 – 08/2023

Role: Data Scientist

Overview:

The Predictive LTV Modeling project focused on designing and deploying machine learning models to forecast Customer Lifetime Value (LTV) by Day 3 (D3) for marketing campaigns across two of the company's most revenue-generating games. This initiative aimed to improve marketing efficiency by enabling accurate, early-stage insights into player value, overcoming the limitations of traditional methods.

Challenge:

Traditional LTV prediction models relied on user progression curves of historical data, providing late (8+ days) and less accurate forecasts due to their assumption of linearity. These delays hampered timely campaign optimizations, resulting in higher costs and suboptimal performance.

Key Features:

- **Data-Driven Insights:** The project utilized detailed historical player data to build accurate predictive models, ensuring robust handling of missing and inconsistent data. Advanced feature engineering techniques were applied to capture nuanced player behaviors, enhancing the quality and reliability of predictions.
- **Timely Predictions:** The models delivered LTV forecasts by Day 3 (D3), significantly reducing the traditional 8-day prediction timeline by over 50%. This expedited process allowed the marketing team to respond quickly to campaign performance metrics and optimize strategies in near real-time.
- **Integration and Scalability:** The predictive models were seamlessly integrated with backend systems, enabling automated updates and real-time access to predictions. Their scalable design ensured they could be refreshed and retrained easily, keeping up with evolving campaign requirements and player behaviors.
- **Optimization Focus:** The tool provided actionable insights that empowered the marketing team to make informed decisions about budget allocation and campaign adjustments. This strategic guidance improved the efficiency of marketing spend and drove better overall campaign outcomes.

Impact:

- Enabled **4-day faster decisions**, reducing campaign optimization delays and improving ROI.
- Improved accuracy of LTV predictions, leading to better-targeted marketing spend and cost savings.
- Ensured business continuity during team transitions, maintaining operational efficiency.

Achievements:

I independently built, trained, and deployed these predictive models, tailoring them for two high revenue games. This included data preprocessing, model selection and training, feature engineering, and ensuring backend compatibility for predictions. I addressed existing challenges by introducing scalable, non-linear models, significantly enhancing marketing efficiency and decision-making. This project highlights my expertise in predictive modeling, data-driven solutions, and my ability to independently deliver impactful innovations in marketing optimization.

Fraud Detection Tool for Marketing Campaigns

Dates: 06/2019 – 07/2019

Role: Data Scientist

Overview:

The Fraud Detection Tool is a heuristic-based solution specifically designed to address the limitations of existing third-party fraud detection systems. By analyzing historical campaign data, the tool established benchmarks for key metrics, defined allowable ranges, and implemented custom rules to identify fraudulent behavior. This tool successfully captured fraudulent activities across campaign types where Protect360 underperformed, offering a robust and reliable in-house alternative for fraud detection.

Challenge:

Detecting fraud in incentivized campaigns, such as cost-per-event and multi-reward campaigns, was a significant pain-point for the marketing team. These campaigns incentivize new players to perform specific in-game actions, like reaching certain levels or completing milestones, which fraudsters exploited to generate fake activity. Although our MMP's premium fraud detection product, Protect360 by one of the leading providers, effectively detected common fraud patterns, it failed to catch nuanced and advanced fraudulent activities. This gap resulted in financial losses and created inefficiencies, requiring a supplementary solution to fill the void and recover potential lost revenue.

Key Features:

- **Data Benchmarking:** Analyzed historical campaign data to establish benchmarks for detecting anomalies.
- **Custom Heuristics:** Designed rules to flag fraudulent activities, tailored to different campaign types.
- **Automation:** Fully automated the fraud detection process for scalability and efficiency.

Impact:

- **Saved \$250,000+** by detecting fraud missed by Protect360 and securing refunds from ad networks.
- Eliminated subscription costs for Protect360, making the tool cost-effective.
- Improved campaign ROI by redirecting budgets to legitimate activities.

Achievements:

- Independently conceptualized and implemented the tool, delivering significant cost savings and operational efficiency.
- Demonstrated the ability to identify gaps in existing systems and design innovative solutions.

Causal Inference Framework for Player Retention

Dates: 09/2024 - 10/2024

Role: Data Scientist

Overview:

The Causal Inference Framework for Player Retention is a structured, repeatable system I developed to uncover the true drivers of Day 7 (D7) player retention at Monumental LLC. Unlike traditional methods that only spot patterns, this framework uses advanced causal analysis to pinpoint what Day 0 (D0) actions—like upgrading cards or spending gems—actually cause players to stay or leave. It's a scientific approach that delivers clear, actionable insights, empowering game designers and data teams to boost engagement and retention effectively.

Challenge:

We were facing lower retention rates than our usual and expected baseline, which raised concerns about player engagement and long-term revenue. Standard analytics couldn't explain why; correlation-based methods showed what was happening but not what was causing it. We needed a reliable way to identify the key factors driving this drop so we could adjust game mechanics and marketing strategies to bring retention back up.

Key Features:

- **Comprehensive Data Preparation:** Processed raw JSON data for a large subset of players, using feature engineering and one-hot encoding to transform player actions into machine-readable formats.
- **Causal Analysis with Double Machine Learning (DML):** Applied DML to isolate cause-and-effect relationships. It predicts D7 retention with a baseline model, assesses D0 actions separately, and compares residuals to reveal true impacts—e.g., card upgrades increase retention by 5.8%.
- **Robust Validation:** Used 10-fold cross-validation to ensure results were consistent and reliable across different data splits, avoiding flukes.
- **Actionable Visualization:** Built bar plots and combo analyses (e.g., upgrades with pack purchases) to highlight key drivers, making insights easy to act on.
- **Scalable Design:** Created a flexible framework that adapts to growing datasets and new games, with automated steps for repeatability.

Impact:

- Reduced guesswork by identifying which player actions boost D7 retention and which actions lower them, along with the magnitude (%).
- Saved resources by focusing efforts on high-impact actions, lifting retention closer to baseline goals.
- Enabled faster, data-driven decisions, cutting retention strategy delays and improving player engagement.

Achievements:

- Delivered a first-of-its-kind causal inference system at Monumental LLC, recognized internally for its innovation and impact on retention strategies.

Deadhand Simulation for Game Optimization

Dates: 10/2024 – 10/2024

Role: Data Scientist

Overview:

The Deadhand Simulator is a Python-based simulation tool with an intuitive Streamlit interface that predicts the percentage of deadhands players are likely to encounter in their game lifetime. The tool incorporates game mechanics like card pools, combo detection, card locking, and final form cards, providing designers with actionable insights to balance the game's economy.

Challenge:

In our game, a "deadhand" situation occurs when two types of cards fail to fuse into a combo card due to an absent combination. While a small percentage of deadhands is necessary for maintaining a balanced game economy, excessive deadhands can frustrate players and negatively impact engagement. Accurately estimating the frequency of deadhands over a player's lifetime was previously impossible due to the randomness of card picking, granting, and drops, making manual estimations ineffective and unreliable.

Key Features:

- **Game Mechanics Simulation:** The simulator modeled critical game features such as combo detection, card locks, final form cards, and variations across IP pools. By replicating real player scenarios, it allowed the design team to test the mechanics under diverse conditions, ensuring the simulation closely aligned with the actual gameplay experience.
- **Extensive Case Testing:** The tool supported dynamic inputs, including player count, battle count, round count, IP deck selection, and customizable hand selection options. This flexibility enabled the team to test a wide variety of scenarios, ensuring the tool could adapt to different game mechanics and design iterations effectively.
- **User-Friendly Interface:** Leveraging the Streamlit framework, the tool featured a seamless and intuitive interface that simplified interactions for the design team. It enabled real-time analysis of simulation results, making the process efficient and accessible for users with varying technical expertise.
- **Automation and Logging:** Automated the entire simulation process, from calculating results to logging key outcomes, eliminating the need for manual interventions. The detailed logs provided a clear record of each simulation, enabling data-driven decisions for optimizing gameplay mechanics and ensuring reproducibility.

Impact:

- Enabled the design team to fine-tune deadhand percentages, improving game balance and player engagement.
- Saved significant time by automating deadhand estimations, allowing focus on strategic design elements.
- Provided actionable insights for pre-launch optimization, reducing post-release adjustments.

Achievements:

- Independently designed and developed the tool, showcasing expertise in Python and Streamlit.
- Collaborated with the design team to deliver a critical tool for game optimization.