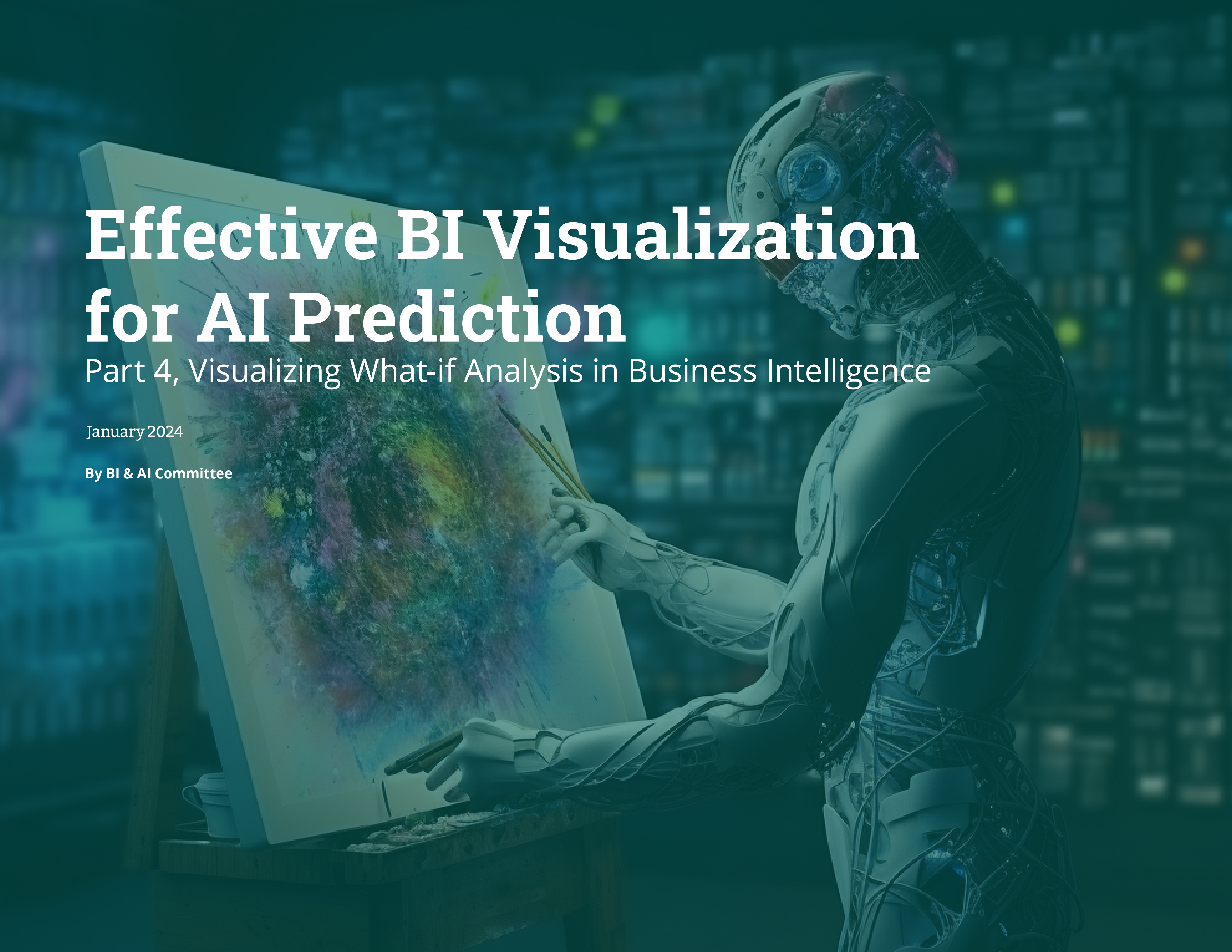


Effective BI Visualization for AI Prediction

Part 4, Visualizing What-if Analysis in Business Intelligence

January 2024

By BI & AI Committee

Introduction

It's no secret that by now Sally Sue has become a whiz with analytics, and one of her favorite techniques is the 'what if' analysis. This method allows her to play around with various scenarios, altering a few data variables to see how they affect the outcome.

For instance, Sally Sue can gauge the impact on sales, revenue, and profit if she decides to increase the product price by 10%. Or she might want to assess how a fresh marketing campaign could influence customer acquisition, retention, and loyalty. She can even evaluate whether recruiting more staff would positively affect productivity and quality.

So, how does Sally Sue go about it? She first establishes a baseline model that accurately represents the current situation or status quo. Then, she comes up with one or more alternative models, tweaking certain variables in the baseline model. Finally, she compares the outcomes of these alternative models against the baseline model to see the variance.

But the challenge that Sally Sue faces, like many others, is how to clearly and effectively present the results of a 'what if' analysis. It's not just about showcasing the final outcomes; it's equally important to highlight the drivers and factors influencing these outcomes. And it's crucial to demonstrate the level of confidence in the forecasts, as well as their sensitivity to changes in the variables.

This becomes all the more significant when Sally Sue uses AI-powered BI dashboards to carry out 'what if' analyses. AI models can generate predictions or scores based on data, but they also come with a degree of built-in uncertainty. This uncertainty means that any prediction or score can have a range of potential values, not just a single number. If Sally Sue ignores this uncertainty, she might reach inaccurate conclusions and make suboptimal decisions.

Visualizing Uncertainty in What-If Analysis

Sally Sue was well aware that uncertainty can sneak in from various sources, like the assumptions made or the predictive models utilized in 'what if' analysis. Sometimes, sufficient input data might not be available to make accurate assumptions, or complex algorithms employed could be prone to errors and biases. Hence, Sally Sue understood the importance of visualizations that accurately reflect the uncertainty in predictions. This helped her make informed decisions, mitigate risks linked with uncertainty, and comprehend her data better.

Uncertainty was a common hurdle in data visualization for Sally Sue, as it could skew the interpretation and credibility of the information presented. To tackle this, she employed data visualization techniques that used visual cues aligning with human perception and reasoning about uncertainty or variability. Techniques like blur to indicate a lack of precision, flicker to suggest instability, reduced saturation to imply vagueness, sketched outlines to convey approximation, and transparency to denote fragility.

For instance, Sally Sue used blur on a map showing the predicted path of a hurricane to highlight the areas most likely to be affected, and flicker to indicate less likely areas. When charting out a company's projected growth, she used reduced saturation to represent the lower and upper bounds of the estimate, and transparency to display the historical data. Similarly, when visualizing a molecule's structure, sketched outlines and transparency were used to depict uncertainty in the position and shape of atoms, and the bonds, respectively.

One of Sally Sue's go-to techniques to represent uncertainty was the boxplot, which presented a five-number summary and was great for showing a dataset's distribution, variability, outliers, and potential errors. She also used error bars, a simple graphical element that extends from a point estimate to display the range of possible values or the level of confidence associated with that estimate. Clearly labeled error bars were handy to represent measures of uncertainty like standard deviation, standard error, or confidence intervals.

Sally Sue also utilized graded confidence bands, a shaded area around a trend line, indicating confidence levels about the slope and direction of the line. This technique helped her compare different models or hypotheses and assess their fit with the data.

When performing "what if" analyses, Sally Sue often used Tornado plots, which showed how different variables affected the system's outcomes. For instance, when analyzing how various factors influenced a company's profit, these plots helped display the range of possible profits under different scenarios, like changes in sales, costs, taxes, etc.

Diverging bar charts was another technique Sally Sue used to understand the impact of key influencers on outcomes. These charts displayed positive and negative values of influencers on either side of a central axis, providing an easy comparison of performance.

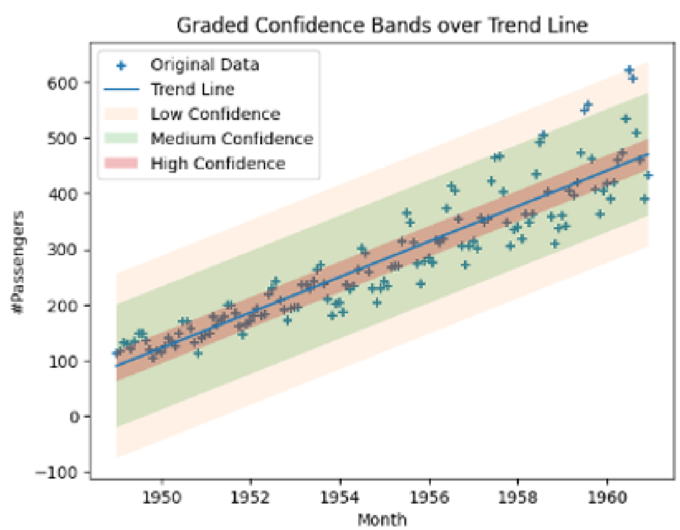
Lastly, Sally Sue used heatmaps, a powerful tool for visualizing the uncertainty arising from changing multiple variables simultaneously. These were particularly helpful while performing 'what if' analyses and exploring different scenarios and outcomes. Heatmaps helped Sally Sue quickly comprehend which input variables had the most significant impact on output variables, and which combinations had the highest or lowest uncertainty. They also aided in comparing different strategies and trade-offs.

How to Show How Good Your Forecast Is

One of the hurdles Sally Sue faced when conducting what-if analysis was demonstrating the robustness of a forecast. Given that not all outcomes hold the same likelihood or reliability, Sally

Sue knew it was crucial to express the strength of her forecasts in her analysis. Questions like "How confident are we in the accuracy and reliability of our predictions?" and "How sensitive are our results to changes in assumptions and parameters?" were always front and center in her mind.

One method Sally Sue used to express the robustness of a forecast was the confidence interval. This provided a range of values that covered a certain percentage of the possible outcomes. For example, when predicting the revenue of a new product, Sally Sue used a confidence interval to display the range of revenue values that could be expected with a 95% confidence level. This meant that 95% of the time, the true value of the model would fall within this range. Confidence intervals helped Sally Sue quantify the uncertainty and variability in her data and model, as well as demonstrate the level of confidence in her predictions. She often visualized the confidence interval using a line chart with shaded areas around the line, or a bar chart with error bars.



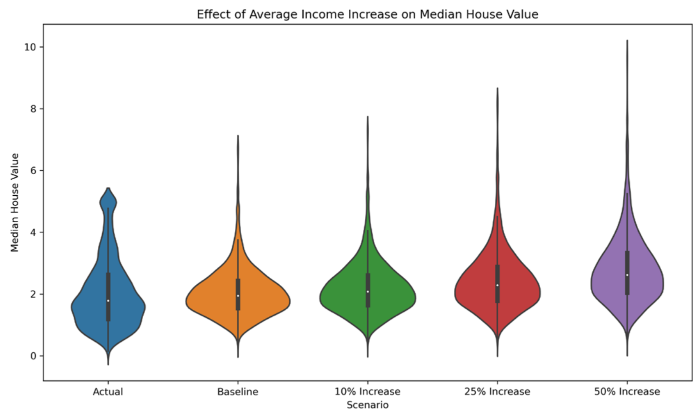
Another technique Sally Sue used to demonstrate the robustness of a forecast was scenarios or simulations. These were alternative outcomes based on different inputs or assumptions. For instance, when predicting the revenue of a new product, Sally Sue used scenarios to show how the revenue fluctuated under various market conditions, such as high demand, low demand, high competition, or low competition. Scenarios or simulations enabled her to test the robustness and sensitivity of her data and model, as well as demonstrate how different factors affected her predictions. She often visualized these scenarios or simulations using multiple line charts or bar charts.

The third method Sally Sue used to reflect the strength of a forecast was benchmarks or comparisons. These were reference values or standards that she used to evaluate her output variable. For example, when predicting the revenue of a new product, Sally Sue used benchmarks or comparisons to show how the projected revenue stacked up against competitors' revenue, industry average revenue, or historical revenue. Benchmarks or comparisons helped her assess the quality and validity of her data and model, as well as demonstrate how competitive or attractive her predictions were. She often visualized these benchmarks or comparisons using stacked bar charts or pie charts.

Quantifying and visualizing the variability of what if analysis

Another puzzle Sally Sue often had to solve during what-if analyses was determining whether an action or decision could be made based on her analysis. Questions like "How do we discern if we need to alter something or maintain the status quo?", "How do we confirm if we've met our target or objective?", or "How can we tell if our situation has improved or deteriorated?" were always on her mind.

SHAP values proved to be a potent tool for Sally Sue during 'what if' analyses. They helped her understand the influence each input variable had on a machine learning model's output and how this information could be harnessed to take actions or make decisions. For instance, let's assume Sally Sue had a machine learning model that forecasted customer satisfaction scores based on various factors like product quality, price, delivery time, and customer service. Sally Sue could use SHAP values to identify which factors were most crucial for the customer satisfaction score and how each factor affected it positively or negatively.

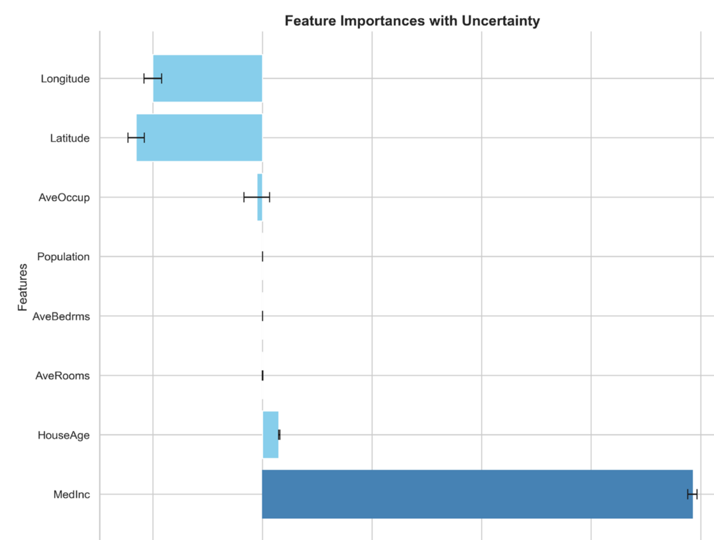


Furthermore, Sally Sue could use SHAP values to understand how altering one or more factors would impact the customer satisfaction score, and whether this change would ameliorate or exacerbate the situation. For example, she could determine how enhancing the product quality or reducing the price could increase the customer satisfaction score and by how much. She could also ascertain how decreasing the delivery time or improving customer service could boost the customer satisfaction score and by what margin. By employing SHAP values, Sally Sue could pinpoint the best actions or decisions to make based on her analysis, and whether she had achieved her goal or objective of maximizing the customer satisfaction score.

Visualizing ordering of variables by feature importance

One of the hurdles Sally Sue Somebody often faced during 'what-if' analyses was determining which variables to alter and to what extent. This is where the variable ordering came into play. Variable ordering involved ranking variables based on their significance or impact on the outcome. By doing this, Sally Sue could concentrate on the most influential variables and observe their effect on the results. She could also contrast different scenarios and pinpoint the best or worst cases.

One method Sally Sue used to order variables based on their importance was feature importance. This is a metric that ranks variables depending on their impact on a machine learning model's prediction. Sally Sue could order variables by feature importance using a bar chart or a dot plot, where the height or position of the bars or dots represented the feature importance score, and the order represented the rank of the variables.
For instance, if she had a model that predicted house prices based on variables like location, size, number of rooms, and age, she could use a bar chart to display which variable had the highest feature importance score. This helped her understand which variable was most pertinent for the prediction task.



In conclusion, Sally Sue's strategic use of visualization techniques effectively represented uncertainty in 'what if' analyses, aiding in informed decision-making and risk mitigation. She utilized various tools to demonstrate the robustness of her forecasts, quantify variability, and evaluate her predictions. Sally Sue also employed SHAP values to understand the influence of input variables and feature importance to prioritize variables based on their impact. Her approach exemplifies how data scientists can harness visualization techniques to comprehend data, make informed decisions, and communicate findings effectively.

Conclusion For This 4-Part *Effective BI Visualization for AI Prediction* Series

AI has changed our lives in the past few years, especially after the introduction of Generative AI. As Sally Sue continues to grow in her career, she should upgrade her skills to match the ever-changing technologies in data and analytics. Failure to do so may lead to an embarrassing moment in a presentation or even worse if the result leads to some wrong, disastrous conclusion. This whitepaper aims to demystify some AI misconceptions and leverage BI visualization to strengthen the understanding of AI-generated predictions.

When there is a way to quantify something, we unintentionally want to make it "prettier" to fulfill our prophecy. With those visualizations, you may be tempted to develop a model with a better-matched line described above. Stay away from this trap and let the visualization technique presented here be a thermometer revealing the truth instead of a thermostat to tune to what you want to see. While there's nothing wrong with pursuing perfection, watch out when you see something aligned way too well. Watch out for overfitting, which is a condition that indicates your model memorizes, rather than learns, the data set in which it's been trained.

Too good to be true? It probably is.

Authors

Cupid Chan

Cupid Chan, an industry veteran, holds the Senior Fellow position at CDHAI in Johns Hopkins and the University of Maryland. His career began in developing a top-tier BI platform and contributing to Linux Foundation ODPi. As Chair of LF AI & Data's BI & AI Committee, he drives the fusion of AI and BI, forging Cognitive Intelligence (CI), melding AI's speed with human-driven insights.

Deepak Karuppiah

Deepak Karuppiah, a Senior Architect in MicroStrategy's Augmented Analytics Group, merges AI and BI to craft technology stacks for AI-powered dashboards. Formerly, he specialized in secure data connectors for MicroStrategy. Proficient in machine vision and learning, he explores novel AI-BI intersections, evident through his role in the BI & AI committee. Beyond work, he volunteers technical skills to a local non-profit, showcasing dedication beyond his professional sphere.

Dalton Ruer

Dalton is a Senior Solution Architect at Qlik. He is a Data Scientist Storyteller, Analytics Evangelist and is an impassioned student of Generative AI. He is a seasoned author, speaker, blogger and YouTube video creator who is best known for dynamically sharing inconvenient truths and observations in a humorous manner. The passion which Dalton shares through all mediums, moves and motivates others to action.

Sachin Sinha

Sachin, Microsoft's Director of Technology Strategy, has honed his data engineering expertise post-graduation from the University of Maryland. Specializing in information management, he crafted systems aiding data-driven decision-making. His support extended to healthcare startups, fostering data-centric business models, and empowered public sector entities toward their missions via data-enabled decisions. Residing in Fairfax, VA, with his family, he passionately backs the Terps and Ravens while excelling in his professional pursuits.

Stu Sztukowski

Stu, a senior product manager at SAS, focuses on user-friendly AI through visual analytics. With a BS in Statistics from North Carolina State University and an MS in Advanced Analytics, he transitioned from data science, excelling in forecasting and BI. He champions accessible high-performance AI, aiming for universal comprehension. A mentor and versatile leader, he simplifies complex analytics for data scientists and business analysts, driven by a passion for public speaking and democratizing analytics.

twitter.com/LFAIDataFdn

linkedin.com/company/lfai

youtube.com/@lfaidatafoundation9555

github.com/lfai

LF AI & DATA

About the LF AI & Data Foundation

LF AI & Data is an umbrella foundation of the Linux Foundation that supports open source innovation in artificial intelligence (AI) and data. LF AI & Data was created to support open source AI and data, and to create a sustainable open source AI and data ecosystem that makes it easy to create AI and data products and services using open source technologies. We foster collaboration under a neutral environment with an open governance in support of the harmonization and acceleration of open source technical projects.Explore our current portfolio of projects and contact us to discuss hosting your open source AI or data project under our Foundation.

January 2024

To reference this work, please cite as follows:
https://wiki.lfaidata.foundation/pages/viewpage.action?pageId=35160417#BI&AI-PastPublication