

**2023 IDEAS Conference panel discussion**  
**Chong Ho Alex Yu & Charlene Yang**

**1. Data science and business analysis are often seen as complementary disciplines. Can you explain the role of each in the context of driving business success, and how do they work together?**

This issue can be approached by highlighting the limitations of traditional business intelligence, and how DSML can remediate these shortcomings.

Traditional business intelligence has several limitations that can be addressed by modern data science and machine learning techniques:

Data Volume and Variety

Traditional BI relies on structured data like data warehouses using SQL for ETL. It struggles with unstructured data like text, images, video, etc.

Modern data science uses tools like NoSQL, techniques like NLP and computer vision to extract insights from unstructured data.

This allows a more complete view by incorporating more data sources.

Static Analytics

Traditional BI offers backward-looking reports and visualizations.

Machine learning adds predictive capabilities - forecasting, recommendations, etc.

This enables more proactive vs reactive decision making.

Lack of Adaptability

Traditional BI has predefined and rigid logic/models. It is more model-driven.

Machine learning algorithms automatically tune models and learn from new data. It is more data-driven.

This allows BI systems to evolve and adapt as patterns change.

Limited Automation

Traditional BI requires extensive manual data prep and coding/scripts.

Data science tools automate data processing pipelines and model building.

This increases efficiency and reduces errors.

Narrow Insights

Traditional BI only answers questions users think to ask.

ML techniques like anomaly detection and text mining unveil insights users may have missed.

## **2. Could you share examples of how data science and business analysis have been instrumental in solving critical business challenges or seizing opportunities?**

Examples in higher education:

- Predictive analytics to improve student recruitment and admissions. By analyzing applicant attributes and historical data, universities can better target marketing efforts, identify high propensity students, and shape incoming classes.
- Improving retention and graduation rates through early warning systems. Using machine learning on academic, engagement, and socioeconomic data, schools can identify at-risk students and provide targeted interventions to keep them on track.

Examples in business:

- Building dashboards in Power BI helps to visualize trends in lead generation.
- Helps to see associations between changes in home-buying prior to COVID, during, and after and how trends have changed.
- Building models to forecast budget, where funds should be allocated.

## **3. Data are often described as a valuable asset for businesses. What strategies can organizations employ to effectively collect, manage, and utilize data to gain a competitive advantage?**

- Build a data-driven culture - Foster an organization-wide understanding of the value of data. Incentivize data-based decision making at all levels. Discourage dichotomous decisions using the p value; rather, encourage pattern seeking.
- Invest in technology and talent - Deploy cutting-edge tools like cloud computing to enable advanced analytics. Hire data engineers, scientists, and analysts.
  - To add, investing in the right technology and knowing how to use it.
  - Example: Datorama is a tool that is particularly useful to marketing; however, it is a fairly sophisticated program that requires a data architect or engineer to create good models that can stand on their own so the ETL process can be smoother. Otherwise, constantly trying to fix breaks in the modules.
  - Hiring for a well-balanced team depending on the needs and tools of the business.
- Establish strong data governance - Institute data quality standards, access policies, metadata management, and governance to ensure trustworthy data.
- Focus on data quality - Dedicate resources to cleaning, validating, and maintaining reliable data over time. Avoid unnecessary and confusing coding (e.g., using 1 for yes and 2 for no). Garbage in leads to garbage out.
- Have and maintain good relationships with good third-party vendors or partners who help the company move forward.

#### **4. How does data-driven decision-making impact various aspects of a business, from marketing and customer engagement to operations and finance?**

- **Marketing Standpoint:** Very important. For example marketing for home-building, funds are allocated in different types of lead buckets and knowing where to give more budget to is very important in how it will drive business, how it drives lead generation. It's also important to know if the other buckets are not high-volume lead generators, why that is and if funds should be allocated to those buckets based on potentiality.
- **Customer Engagement:** Working in marketing for a homebuilder, it's very important to understand customer engagement. We are working with tagging management vendors who help us track that customer journey.
  - For example, on our website we have different articles that help the home-buying journey. Knowing what customers are interested in helps us to understand where they might be in their home-buying journey (e.g. reading articles about sustainability vs first-time homebuyer).
- Every company has a budget allocated specific to each department and if the budget is tight, it becomes even more important to make data-driven decisions so that efforts and finances are maximized.

#### **5. Data science involves advanced analytics and machine learning. How can businesses, regardless of their size, harness the power of these technologies to improve operations and customer experiences?**

- Start small, but start - Even small pilots of predictive models or basic automation can demonstrate value and allow for iterative expansion. Make sure the beta test results can inform us for the next course of actions.
- Focus models on high-impact areas - Target supply chain, customer churn, personalized recommendations - areas where even basic models can significantly improve KPIs. Don't waste time on trivial things.
- Lean on strategic partners - If lacking in-house skills, partner with expert agencies, system integrators, and consultants to build models (e.g., SAS Institute)
- Use Generative AI: ChatGPT, Claude, Google Bard...etc.

#### **6. Business analysts often serve as intermediaries between stakeholders and data scientists. What skills and qualities are essential for effective business analysis, and how can this role bridge the gap between business needs and data solutions?**

- **Communication** - Able to translate complex technical jargon into layperson language and vice versa. Act as a liaison between the two groups.
- **Critical thinking** - Ask probing questions to understand root causes and objectives. Identify assumptions and limitations of proposed solutions. In some situations finding a new question is more crucial than finding an old answer.

- **Business acumen** - Understand the organization's goals, processes, and functions. Provide use cases and context for data work.
- **Data literacy** - Comfortable working with data, metrics, databases, analytics, and data visualization to support decision-making. Know the context of the numbers.
- **Problem solving** - Unpack ambiguous problems into concrete requirements and steps for data teams. Break down a big problem into several sub-problems. Prioritize needs. Take one step at one time.
- **Collaboration** - Engage diverse teams, manage expectations, align interests, and foster partnerships, especially if you departments that have a more traditional mindset they may be resistant to change. So, fostering good collaboration and working relationships can help different parts of the organization to be more cohesive and successful.
- **Project management** - Drive projects from conception to implementation. Coordinate resources, requirements, budgets, timelines.
- **Synthesis skills** - Synthesize data findings into compelling insights tied to business impact and recommendations.

## 7. Ethics and responsible data usage are important considerations. What principles and practices should businesses follow to ensure they handle data ethically and responsibly?

- **Transparency** - Clearly communicate how data are collected and used. Seek consent when appropriate. Allow users control over their data.
- **Security** - Implement robust technical controls like encryption to protect sensitive data from breaches and unauthorized access.
- **Privacy** - Collect the minimum data needed. Anonymize or aggregate data to preserve privacy when possible. Allow user opt-outs.
- **Accuracy** - Establish processes to clean, validate and maintain accurate datasets. Quickly fix errors.
- **Accountability** - Enforce data policies through governance. Establish oversight measures and conduct audits.
- **Fairness** - Ensure algorithms and models are unbiased. Continuously assess for discrimination against protected classes.
- **Interpretability** - For high-risk AI models, ensure ability to explain how they arrive at decisions. Try to explain the “black box”.
- **Human oversight** - Keep humans in the loop for model reviews, audits and key decisions. Override the advice from AI, if necessary, because AI might not know what our ultimate goal is and the context of the problem (alignment problem).
- **Impact assessment** - Proactively evaluate data practices for possible negative externalities or unintended consequences.

At present, numerous individuals are advocating for increased government involvement in implementing regulations. Nevertheless, premature and excessive regulations could potentially stifle innovation. Reflecting on the 1990s, when the Internet first emerged, it operated with

minimal regulations. Over time, as society gained a deeper understanding of the Internet's significance, norms and standards were established. A similar approach can be adopted for Data Science and Machine Learning (DSML). We might also draw an analogy: just as early vehicle engineers didn't immediately impose laws requiring seat belts and airbags, sometimes, it's more prudent to allow standards to evolve organically.

**8. Data science projects often require collaboration across departments. How can organizations foster a data-centric culture and encourage interdisciplinary teamwork to maximize the value of data?**

- **Executive sponsorship** - Get buy-in from leadership. Set a data-driven vision. Tie analytics to business goals. Provide resources and oversight. This is a vertical top-down approach. It is easy to say but hard to do. Very often the leadership has no technical background to understand what is happening.
- **Training programs** - Offer education opportunities to improve data literacy and analytical skills across departments. It is a vertical bottom-up approach. Hopefully when there is a critical mass, some change can happen.
- **Break down silos** - Implement cross-functional teams, internal data sharing, and knowledge. This is a horizontal approach. I am a member of the AI Task Force and also a liaison of inter-department collaboration on DSML.

**9. Data-driven decision-making also involves visualization and communication of insights. What tools and techniques can businesses use to effectively communicate data findings to nontechnical stakeholders?**

- JMP Pro
- Tableau
- SAS Viya

Low-code or no-code solutions provide pre-built components, templates, and graphical user interfaces (GUI) that can accelerate development compared to programming. These tools allow users to focus more on the research question, the data, and the business logic rather than the syntax. However, no-code solutions inevitably involve some trade-offs in terms of flexibility, customization ability, scalability and performance compared to coding. Data science education should balance both sides.

**10. Looking ahead, what trends and innovations do you foresee in the field of data science and business analysis, and how might they impact the way organizations operate?**

Generative AI like ChatGPT, Bard, Claude, and other large language models may run even faster with quantum computing. IBM has been devoting efforts and resources to quantum computing, which is a rapidly-emerging technology that harnesses the laws of quantum

mechanics to solve problems too complex for binary-based computers. Specifically, currently the basic unit of information for existing computing devices must be binary (1 or 0), but for quantum computing, the basic element is the qubit, which has 3 possible states. If quantum computing becomes reality, AI will be much more powerful!

In the near future we can simply tell the AI system what we want e.g., “Computer. Based upon the last 10-year data, what are the top three predictors of customer loyalty? Run a model comparison. Select the best model using AIC as the criterion. Write a 10-page report with 5 figures and 2 video animations.”

Definitely more automation, not just on the business end but creating user-automation. Giving more choices to consumers that make it easier for them to make decisions faster.

- An illustrative case is Carvana. In the past, purchasing a car necessitated a trip to a dealership. However, Carvana has revolutionized the process, allowing buyers to acquire a vehicle without the inconvenience of interacting with salespeople and, in the process, saving the buyer several hours of valuable time.