Crawl, Walk, Run: Defining a Strategic Roadmap from Messy Data to AI and Beyond

3/4/21 5:12 PM

Speaker: Tiffany Perkins-Nunn and Hamdan Azhar (BlackRock)
Date: 03/04/2021

Milestones on the road to increasing data science sophistication

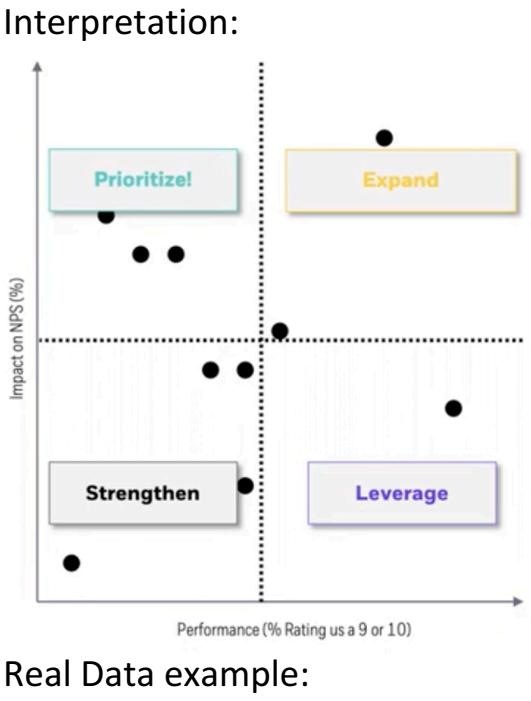
Level	Current state	Focus areas
Level 1	**Data, what's that??**	Educating stakeholders, evangelizing importance of data, and collecting data from disparate sources (databases, excel files, APIs, external platforms)
Level 2	**We have a lot of data, now what do we do with it??**	Cleaning data and stitching it together meaningfully to make it accessible to stakeholders; uncovering insights to understand clients; collecting more data through surveys (360 View, NPS)
Level 3	**We have some insights but we need to better understand drivers of behaviors and outcomes so we can inform busines decisions**	Statistical modelling (driver analysis, regressions, PCA, etc.), intuitive data visualizations (heat maps), dashboards
Level 4	**We have clean data and well-defined outcome variables, now we need to build models and productionize them**	Machine learning
Level 5	**We have the models, now we need a constantly running machine that is self-directing and auto-optimizing**	Artificial intelligence

Net Promoter Score

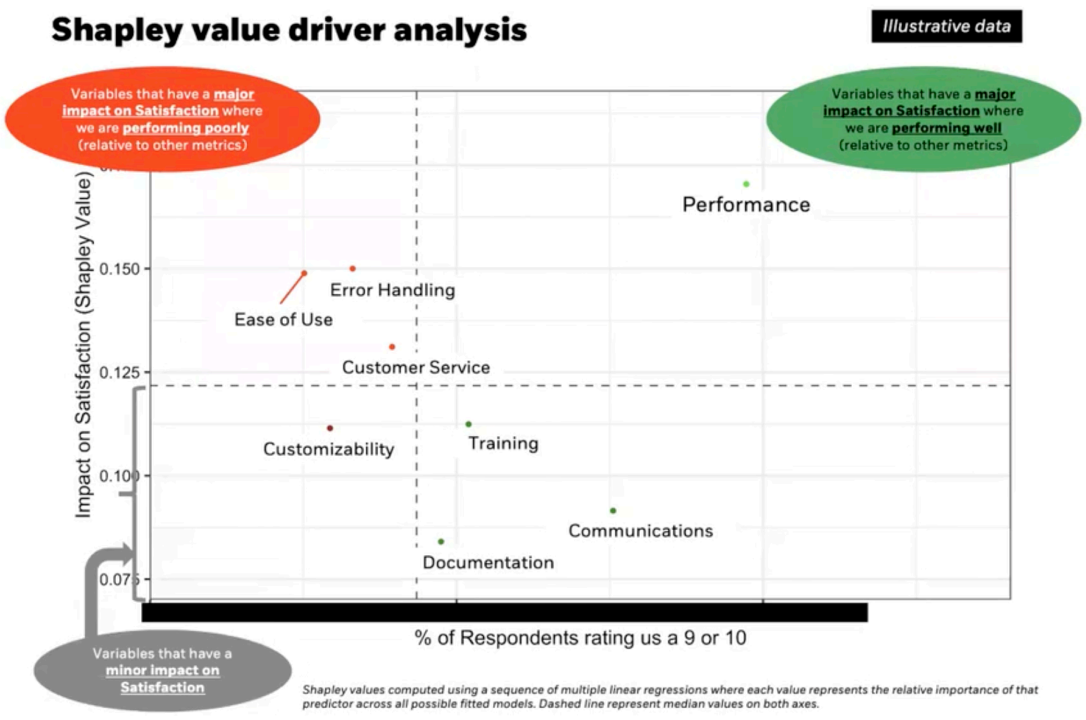
- 5 interaction channels:
  Emails, website, sales contact, trainings, discussions
- The more interactions there are, the more likely client are likely to recommend the company to a friend or colleague

Key Driver Analysis

- Explain driver of business outcome, tell us which components of the client experience are influencing satisfaction or NPS
- Methodology: compute "relative importance estimates" using Shapley value regression (estimating relative importance of multiple predictors in the presence of high multicollinearity)
- Interpretation:



- Real Data example:



-

Finding patterns in client feedback at scale

- Based on survey data, produce a heatmap
- Rows represent variables and columns are feedback break-down by job types, industries and region. Color is conditional formatted by value of each variable (what have been done well, what needs to be improved, etc.)