



**Mu Sigma**

**DO THE MATH**

Market mix policy simulation through Real Games - Proposal

*Mu Sigma Inc.*

*29 October 2018*

# Contents

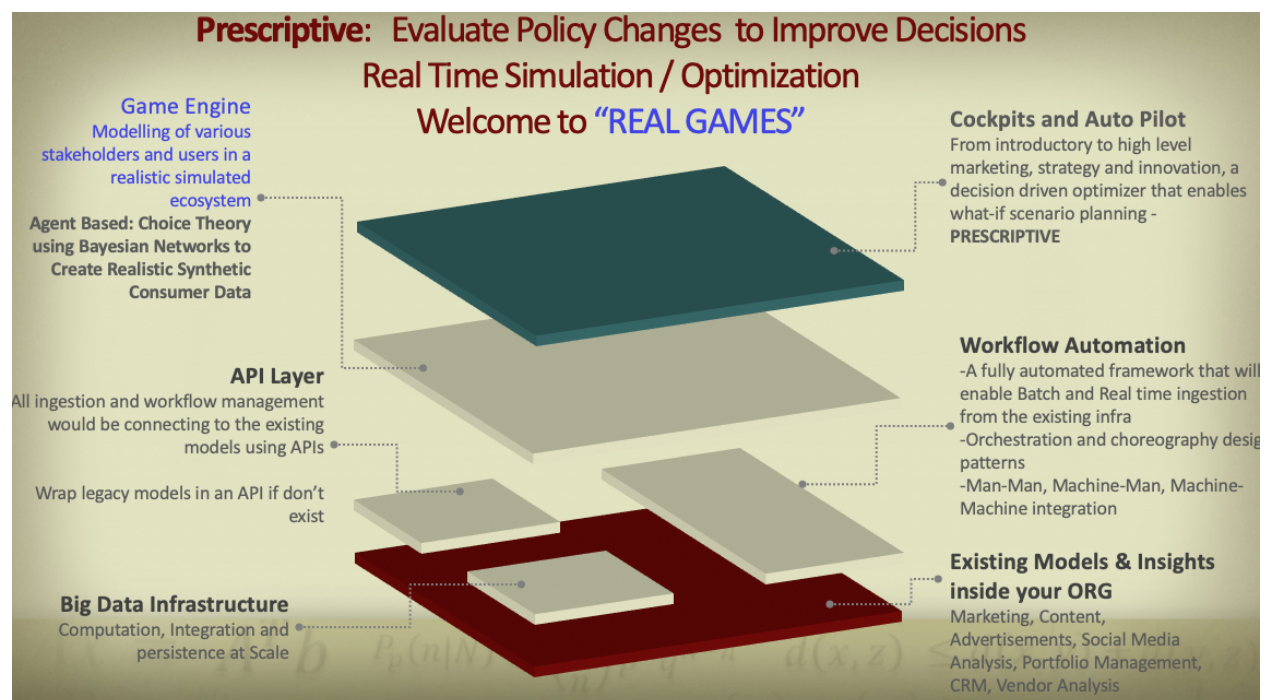
<b>Problem overview</b>	<b>3</b>
<b>Proposed solution - Real Games</b>	<b>3</b>
<b>Proposed engagement and timelines</b>	<b>4</b>
Stage 1 - Proof of Concept and assessment of existing infrastructure - [3 months]	4
Stage 2 - Scale to production - [3 months]	4
<b>Stage 1 - Details</b>	<b>4</b>
Stage 1.a. Causal model & simulation POC	4
Problem scope	4
Simulation objectives	4
Agent model	5
Process for developing and updating agent model	5
Possible future insights and decision support	7
Required data elements for observables and corresponding data sources	7
Final Deliverable	8
Timeline breakdown	8
Stage 1.b. Process improvement and automation - Audit & recommendation	8
Problem scope	8
Final deliverable	9
Timeline breakdown	9
<b>Kick-off checklist</b>	<b>9</b>
Business & Problem Understanding	9
Data understanding, engineering and computation infrastructure	10
Data Science	10
Decision Science	10
<b>Costing</b>	<b>11</b>
<b>Governance model</b>	<b>11</b>
<b>Contact details</b>	<b>13</b>

## Problem overview

Diageo has successfully rolled out a global marketing effectiveness platform, named *Catalyst*, based on econometric modeling to track ROI across channels, and optimize spend allocation. Diageo is looking to take marketing effectiveness to the next level by:

- Improving current processes and effectively automating the current platform where possible to reduce cycle time
- Moving from a correlation based econometric model to a causal model

## Proposed solution - Real Games



Mu Sigma proposes the **Real Games** approach, in order to simulate market behavior from the ground-up by replicating every consumer as an *agent* and defining their *interactions* with the product and market in different situations. This simulation can be used to evaluate various policy decisions in a cost-effective manner, once validated.

In order to realize the concept of **Real Games**, Mu Sigma recommends that the following components have to be put in place [either already existing (or) newly added], overlaid on top of the *existing modeling and Business Intelligence (BI)* setup.

- *Big Data infrastructure* - Ingesting, cataloguing and fusing data from various sources with differing levels of granularity
- *API and automation layer* - Serving of existing models through APIs for consistent re-use, as well as a workflow automation piece, designed to connect the different pieces of the puzzle and **reduce cycle time** - data pull and validation, monitoring and reporting on models, etc.
- *Game engine & cockpit* - Market simulation from the ground up i.e. simulating consumer behavior by modeling the decision-making mechanism of consumers, and building up to a macro picture of the market, as opposed to traditional macro-level analysis. In addition, a cockpit for simulating and trying policy decisions before rollout

## Proposed engagement and timelines

Mu Sigma proposes that the project be split into the following stages to ensure incremental impact is achieved, as well as outputs are validated and improved continuously.

**Note** - Stated timelines assume that required access to data and systems have been provided before the start of the project

### Stage 1 - Proof of Concept and assessment of existing infrastructure - [3 months]

- a. *Causal model and simulation - Proof of Concept*: Construct consumer-level agent causal model and validate simulation against macro-level indicators (from existing accepted models such as econometric MMX models) for the UK market
- b. *Process improvement and automation - Audit & recommendation*: Assess existing infrastructure and processes in terms of model development, monitoring and deployment of the current econometric MMX models and suggest automation improvements to reduce cycle time

### Stage 2 - Scale to production - [3 months]

- a. *Causal model and simulation*: Integrate causal model into the existing system as a challenger to econometric correlation-based models
- b. *Improved process automation*: Implement process improvements and automation to reduce cycle time

## Stage 1 - Details

Note - Items 1.a. and 1.b. will be executed in parallel over a 3 month period.

### Stage 1.a. Causal model & simulation POC

#### Problem scope

*Smirnoff* is the world's largest spirit brand, and has a huge presence in the UK market. There is a large marketing effort associated with the brand which is aimed at reaching out to consumers through a variety of channels such as TV, in-store promotions, brand and event partnerships and so on.

#### Simulation objectives

- Understand consumer level behavior across different segments and simulate micro behavior in order to build a macro-level picture based on market realities
- Fusion of different sources of data along with learnings/ findings from consumer behavior and market research studies, and integrate that information into the simulation
- Simulate policy decisions related to marketing mix and spend allocation, in order to identify optimal strategies
- Understand true cause-effect relationships between marketing activity and consumer response

## Agent model

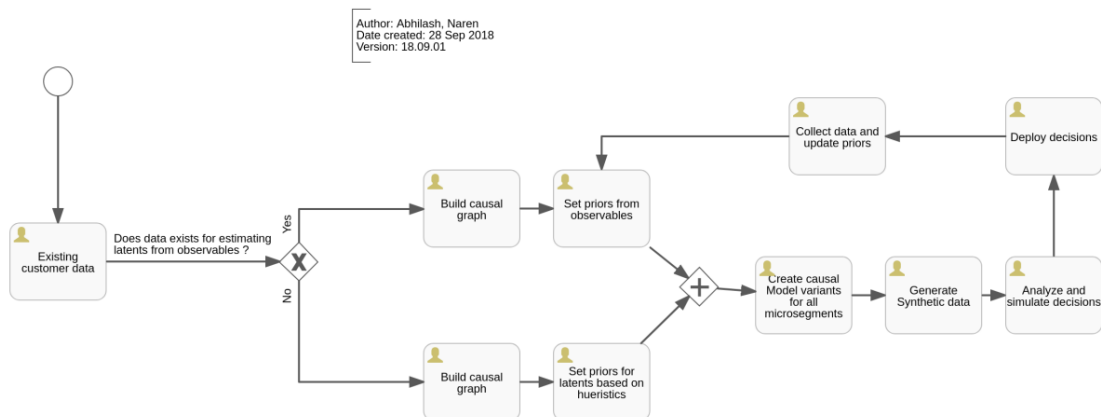
The cornerstone of the solution is the agent-level model, which essentially is a *digital twin* of the consumer emulating the decisions he/she would make given different situations.

The *brain* of the agent, which is its decision-making mechanism is represented, estimated, and implemented in the form of a **Bayesian network**. The Bayesian paradigm, specifically Bayesian networks provide the following advantages which make them the most appropriate tool for this exercise:

- Elegant and rigorous way to fuse data at differing levels of granularity, along with heuristic beliefs
- Structure and parameters of the network can be learnt from data, where available
- Where data is unavailable, prior beliefs from market research studies or opinion from a panel of experts can be incorporated in a consistent framework
- The network can be constantly validated and updated, allowing for continuous calibration based on real observed evidence - this includes validation of structure of the model & prior beliefs, as well incorporation of new evidence
- Provides a framework to identify and quantify **causal effects** from observational data, through the theory of causal calculus

This paradigm allows for businesses to estimate **true cause-effect** relationships, and hence make more informed policy decisions, *without* the need for Randomized Controlled Trials (RCTs) in all cases. The amount of field experimentation can be reduced substantially, therefore leading to generating more cost-effective signals for decision-making.

## Process for developing and updating agent model



## Initial consumer-level agent model

In order to be able to simulate the market based on consumer behavior, the following initial model of consumer decision-making in their interactions with the *Smirnoff* brand and product has been constructed. This conceptual model will continuously be iterated upon, based on **answers to questions** detailed below as well as **further insights**, as and when they come up.

The model includes decisions such as:

- Product purchase
- Product consumption

Competitors will be identified and up to 3 competitors will be considered in the simulation.

The causal model that defines the mechanism that consumers use to make decisions is represented as a Bayesian network consisting of:

- Observables - Data elements that can be measured
- Latents - Variables which are functions of the observables and represent more qualitative characteristics which cannot be measured
- Targets - Consumer decisions

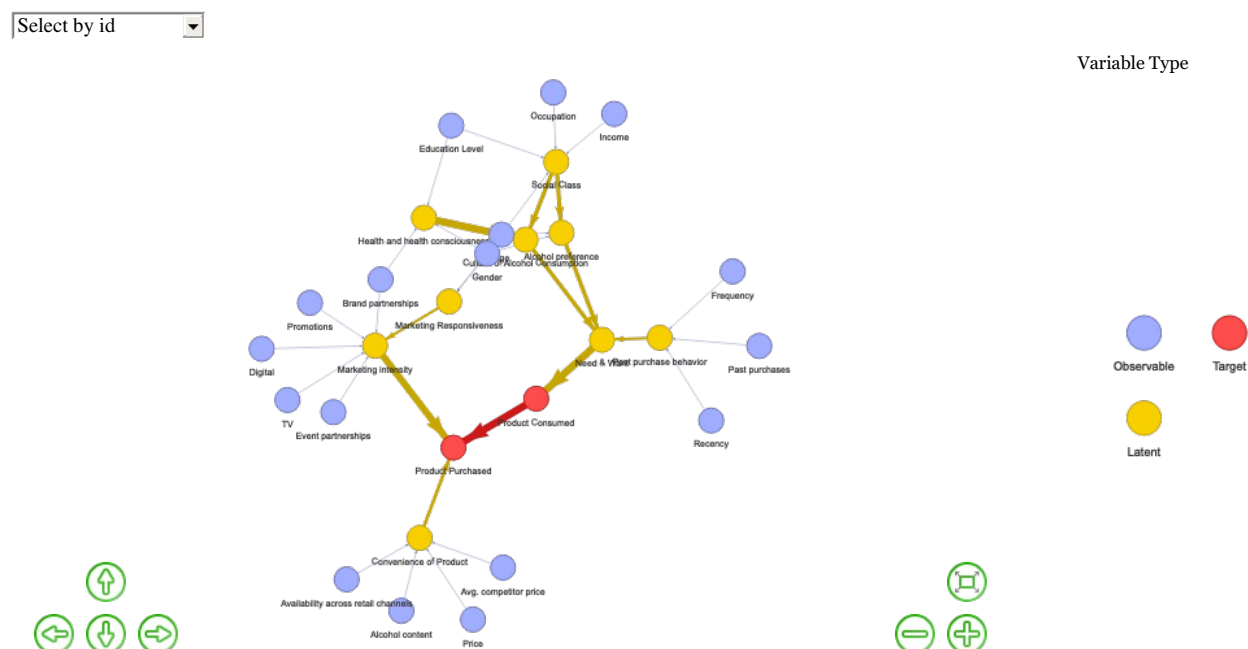
### Initial causal graph of determinants of alcohol consumption and purchase

#### List of nodes

```
kable(node_df %>% dplyr::arrange(group), format = "latex", longtable = TRUE) %>% column_spec(c(2,3), wi
```

id	label	group
1	Health and health consciousness	Latent
2	Social Class	Latent
3	Alcohol preference	Latent
4	Marketing Responsiveness	Latent
5	Culture of Alcohol Consumption	Latent
7	Need & Want	Latent
28	Convenience of Product	Latent
29	Marketing intensity	Latent
30	Past purchase behavior	Latent
8	Education Level	Observable
9	Income	Observable
10	Age	Observable
12	Occupation	Observable
13	Gender	Observable
14	Past purchases	Observable
15	Recency	Observable
16	Frequency	Observable
17	Promotions	Observable
18	Brand partnerships	Observable
19	Event partnerships	Observable
20	TV	Observable
21	Digital	Observable
22	Alcohol content	Observable
23	Availability across retail channels	Observable
24	Price	Observable
25	Avg. competitor price	Observable
26	Product Purchased	Target
27	Product Consumed	Target

#### Initial causal graph



### Possible future insights and decision support

- Formulate optimal marketing mix in a given market, by simulating various spend allocations realistically through a rigorous understanding of true cause-effect

### Required data elements for observables and corresponding data sources

In order to start with the development of the simulation, a certain set of initial data is required. Based on the initial conceptual model described above, the following data elements would be required, preferably at the level indicated. If the required level of granularity is not available, the most granular possible would suffice.

Further for macro validation, actual region level sales, active consumers, market share and units sold would need to be available in order to compare simulation results to actuals.

The following data elements are required for the UK market. Please indicate the following for each data element:

- Is data element available?
- If yes, what is the data source?
- If yes, what is the level at which the data element is available?

Variable	Type	Category	Preferred level of data
Education Level	Observable	Consumer demographics	Household
Income	Observable	Consumer demographics	Household
Age	Observable	Consumer demographics	Household
Occupation	Observable	Consumer demographics	Household

Gender	Observable	Consumer demographics	Household
Past purchases	Observable	Consumer behavior	Household
Recency	Observable	Consumer behavior	Household
Frequency	Observable	Consumer behavior	Household
Promotions	Observable	Marketing activity	Zip code
Brand partnerships	Observable	Marketing activity	Zip code
Event partnerships	Observable	Marketing activity	Zip code
TV	Observable	Marketing activity	Zip code
Digital	Observable	Marketing activity	Zip code
Alcohol content	Observable	Product characteristics	Overall
Availability across retail channels	Observable	Product characteristics	Zip code
Price	Observable	Product characteristics	Zip code
Avg. competitor price	Observable	Product characteristics	Zip code
Product Purchased	Target	Sales	Zip code
Product Consumed	Target	Consumption	Zip code

## Final Deliverable

- Construct consumer-level agent causal model and validate simulation against macro-level indicators (from existing accepted models such as econometric MMX models) for the UK market

## Timeline breakdown

- Month 1
  - Conduct a data audit and secure data. Confirm business questions
  - Estimate the initial causal models based on available data sources
  - Define heuristic priors from existing market research as well as expert opinions
- Month 2
  - Run simulations and conduct macro/micro validation
  - Update causal models
  - 2 iterations and review
- Month 3
  - Package final deliverable answering the question “Were we able to simulate consumer behavior accurately and identify causal effects in the consumer decision-making mechanism?”
  - Share experiment results

## Stage 1.b. Process improvement and automation - Audit & recommendation

### Problem scope

This part of the Proof of Concept deals with auditing the current set-up for marketing decision support to identify possibilities for:

- Process improvements
- Process automation
- Integration of required technology and infrastructure

The key objective of this exercise, is to recommend improvements in order to meet the business outcome of *reducing cycle time*



This includes understanding Diageo's current capabilities corresponding to each layer of the Real Games architecture, namely:

1. Existing processes with respect to model development, validation, deployment and monitoring for the econometric MMX models
2. Big Data infrastructure
3. Workflow automation capabilities
4. API capabilities - specifically for current models
5. Cockpits and consumption of insights for decision support

### **Final deliverable**

- Detailed recommendations on implementing the Real Games architecture, based on the audit of existing infrastructure and platforms in terms of data engineering, existing models, workflow automation, and consumption objectives

### **Timeline breakdown**

1. Month 1
  - Audit computational and data engineering infrastructure
  - Recommend additions/ changes for efficient data integration and fusion
2. Month 2
  - Audit model development & operationalization setup
  - Recommend additions/ changes for scaling and serving models with high availability through APIs
3. Month 3
  - Audit workflow automation capabilities and integrations with existing consumption platform
  - Recommend additions/ changes to enable orchestration of data transformation and validation through workflow automation, along with integration of model outputs into the existing marketing effectiveness platform

## **Kick-off checklist**

The following are a set of questions which provide Mu Sigma a more detailed understanding of the problem, and enable us to set up the simulation in the appropriate way, tied to required business outcomes.

### **Business & Problem Understanding**

1. Internal segmentation scheme
  1. Is there a current available internal segmentation scheme of consumers? If yes, please provide details of the same
  2. Are there multiple levels/ types of segmentation profiles based on purpose? If yes, please provide details of the same
2. Marketing Mix
  1. What are the various channels of promotion/marketing?
  2. What is the typical split of the budget across these channels?
  3. What are the most/least profitable channels? How strongly do they correlate to sales volume?

## Data understanding, engineering and computation infrastructure

1. Is Diageo's data engineering and processing infrastructure deployed on-premise, on-cloud, or a hybrid system? Please provide a brief overview
2. Data sources
  1. Please describe/list Diageo's existing data assets (internal, external, paid), along with details on what data elements are captured in each of them and their level of granularity. Also, please indicate if these are streaming or batch sources
  2. Please describe the data sources, their level of granularity, and confidence with which they are used, in the existing MMX modeling set-up
3. Ingestion
  1. Please indicate tools set up for ingestion of batch and streaming data
4. Processing and Storage
  1. What is the set-up used for storage of Big Data? Please provide details of data warehouses, data lakes, data marts, and so on
  2. What are available processing engines available, such as Spark cluster infrastructures, R Servers, and so on?
5. BI and Advanced Analytics
  1. Please list Business Intelligence tools used for marketing decision support, if any
  2. Please list platforms/ environments used for model development (such as R, Python, Spark)
  3. Please list any model operationalization platforms already in place, for model deployment and performance monitoring
  4. Please briefly describe Diageo's approach in terms of consumption channels of insights generated. (For eg. details such as whether Diageo prefers to be agnostic to BI tools, is tied to a specific BI tool, has purpose-built web applications and so on)

## Data Science

1. Marketing Mix Model
  1. Please provide a brief description of the MMX model set-up
  2. Please describe the training data schema for the MMX model
  3. Please list the current parameters of the champion model, along with the adopted approach for modeling
  4. Please describe how the existing MMX models are validated, and please indicate Diageo's confidence in the results
  5. Please describe the cadence with which the existing models are calibrated. Also, please indicate the key activities in the calibration process
2. Existing consumer behavior research
  1. Has Diageo conducted any market research studies on consumer behavior? If so, please provide details, models and findings from such studies
  2. Has Diageo conducted any market research studies to understand market share in the UK market? If so, please provide details and findings from such studies
  3. Does Diageo have (or can nominate) a set of experts, who can help set prior distributions of impact of various variables on outcome metrics, in the case that data for said variables is not available?

## Decision Science

1. Please describe your current understanding of the causal determinants of consumption and purchase of Diageo products amongst its customer base, if any
2. Please indicate, in your opinion, deficiencies in the current econometric implementation of the MMX model from a consumption perspective
3. Please indicate the key stakeholders involved in the development, monitoring and consumption of the MMX model and their roles

4. What is the level of automation in the MMX model development and consumption process?
5. Please indicate key outcomes of the Proof of Concept, from Diageo's perspective.
6. Please designate a Single Point of Contact from Diageo for this project, who can review progress and guide towards the final deliverable

## Costing

Mu Sigma proposes a three month timeline for Stage 1 costing USD 270,000

Cost for Prototype – Stage 1, Total 3 months				
Resource	Resource Estimate	Monthly Rate per resource (in USD)	Duration	Total Fee (in USD)
Offshore analyst	5	\$ 15,000	3 months	\$ 225,000
Onsite Associate	0.5	\$ 30,000	3 months	\$ 45,000
Grand Total for 3 months				\$ 270,000

### **Value-add from Mu Sigma:**


1. Access to Mu Sigma Art of Problem Solving Base Tool Kit (full suite of Mu Sigma's software, framework, tools, algorithms)
2. Access to Mu Sigma University, Mu Sigma LABS, its trainings and workshops

### **Cost Exclusions and notes:**

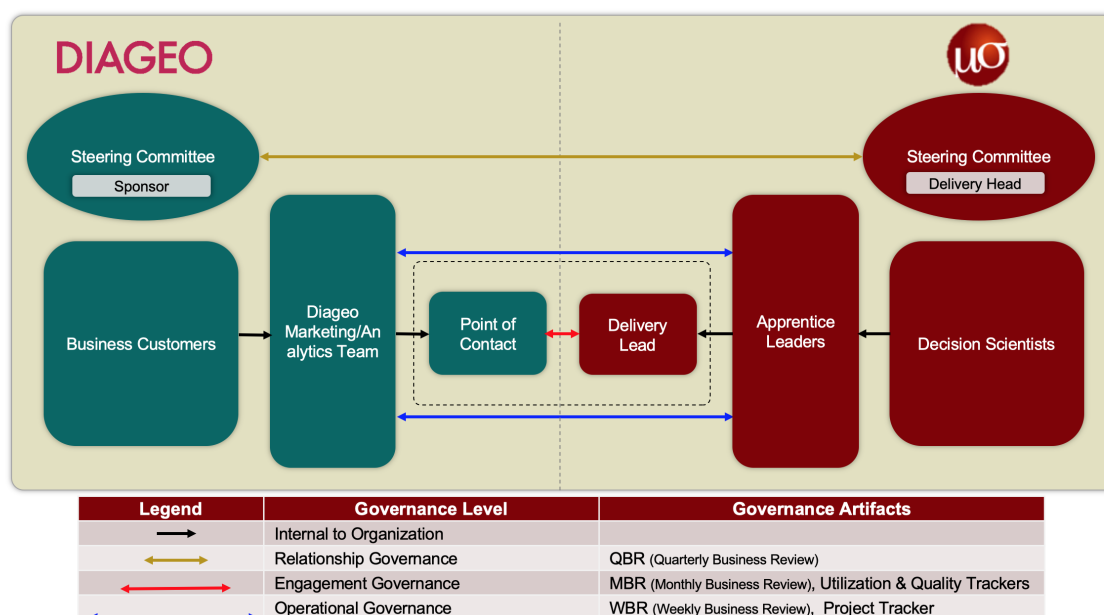
1. Client will have an option to pause the engagement at the check-in point at the end of the prototype phase
2. Mu Sigma will follow customer's travel policy for any travel associated with this work
3. Non-standard software/ hardware costs will be billed at actuals
4. Fees are exclusive of any present and future sales, use, service or other applicable taxes and other statutory levies chargeable in respect of Services rendered by Mu Sigma and shall be charged extra to Customer

## Governance model

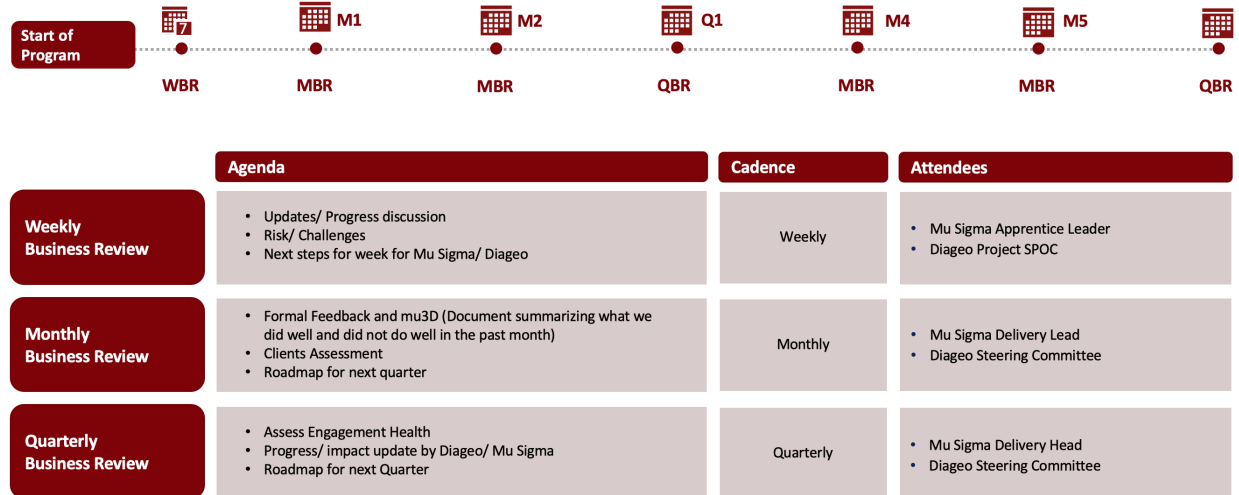
- The program will be steered effectively by a tight-knit governance mechanism enabled by leadership at all levels

DIAGEO			
Role	Responsibilities	Role	Responsibilities
Steering Committee	<ul style="list-style-type: none"> <li>This committee shall consist of the Sponsor, Point of Contact, Analytics Leads and Business Stakeholders</li> <li>This committee shall be responsible for setting goals and priorities for the Global Delivery team, and monitoring progress</li> </ul>	Steering Committee	<ul style="list-style-type: none"> <li>This committee shall consist of the Mu Sigma Delivery Head and Offshore Delivery and Apprentice Leads from Mu Sigma</li> <li>This committee shall be responsible for partnering with the Diageo Steering Committee and monitoring progress of the engagement</li> </ul>
Sponsor	<ul style="list-style-type: none"> <li>This is the highest point of escalation at Diageo for this relationship, and will be responsible for handling critical issues when required</li> </ul>	Mu Sigma Delivery Head	<ul style="list-style-type: none"> <li>The Mu Sigma Delivery Head is the highest point of escalation at Mu Sigma for this relationship, and will be responsible for handling critical issues when required</li> </ul>
Point of Contact	<ul style="list-style-type: none"> <li>This is the primary point of contact at Diageo for the complete Global Delivery Initiative. This person shall work closely with the Mu Sigma Engagement manager to execute and monitor projects and all activities as part of the partnership</li> </ul>	Delivery Lead	<ul style="list-style-type: none"> <li>The Delivery Lead is the primary point of contact at Mu Sigma for the complete Global Delivery Initiative. The DL shall work closely with Diageo Point of Contact to execute and monitor projects and all activities as part of the partnership</li> </ul>
Stakeholders	<ul style="list-style-type: none"> <li>These are Analytics Leads and Business Stakeholders who will help identify projects and tasks for Mu Sigma in partnership with the Point of Contact</li> </ul>	Apprentice Leaders	<ul style="list-style-type: none"> <li>These are Sr. Managers, Managers and Associate Managers from Mu Sigma who will be responsible for managing project delivery, appropriate communication and adherence to timelines</li> </ul>
Project Leads and Project Team	<ul style="list-style-type: none"> <li>These are the team members from Diageo with whom Mu Sigma will interact closely to understand operational details of the projects</li> </ul>	Project Lead and Decision Scientists	<ul style="list-style-type: none"> <li>These are Mu Sigma analysts, dedicated to working and delivering analytic solutions for Diageo</li> </ul>

- A robust governance model will be put in place for streamlined and effective management of the engagement



- Regular cadence of review ensures alignment between stakeholders and enables constant course correction to meet goals



## Contact details

In case of any clarifications, further details, or follow-ups please reach out to any of the following persons:

- Devyajyoti Ghosh - Devyajyoti.Ghosh@mu-sigma.com
- Julie Woods-Moss - julie@woodsmoss.com
- Hashir Mohamed - Hashir.Sitheekh@mu-sigma.com