

**2021.AI**an algorithmic and technology company

**Global Hotel Alliance**

**Data Science Report**

**By 2021.AI**

April 15th 2017

VERSION 1.0

**Contents**

1. Executive Summary 4

1.1. Business context and project background 4

1.2. Key recommendations 4

2. Background and context 9

2.1. Project background and context 9

2.2. Project deliverables 9

2.3. Project structure and approach 10

2.4. Workshop agenda 10

2.5. Opportunity mapping 11

2.6. Report structure 12

3. Business Case – Data as an asset 13

3.1. Introduction 13

3.2. Market matrix 13

3.3. Industry, competitors and market 14

3.4. Case framework: Methodology 16

3.5. Analysis: Current State of Data 20

3.6. Asset Value of Data: Information Measures 20

3.7. Asset Value of Data: Economics Measures **Error! Bookmark not defined.**

3.8. Recommendations 23

4. Technology and data 25

4.1. Introduction 25

4.2. Technology and data As-is 25

4.3. Technology and data To-be environment 28

5. Use Cases and POC Candidates 32

5.1. Introduction 32

5.2. Use Case shortlist 32

5.3. POC candidate recommendation 33

6. Appendix 39

6.1. Use Case long list 39

6.2. Non-prioritized use cases 41

6.3. Overview of data processing during the POCs 41

6.4. Algorithms and models 44

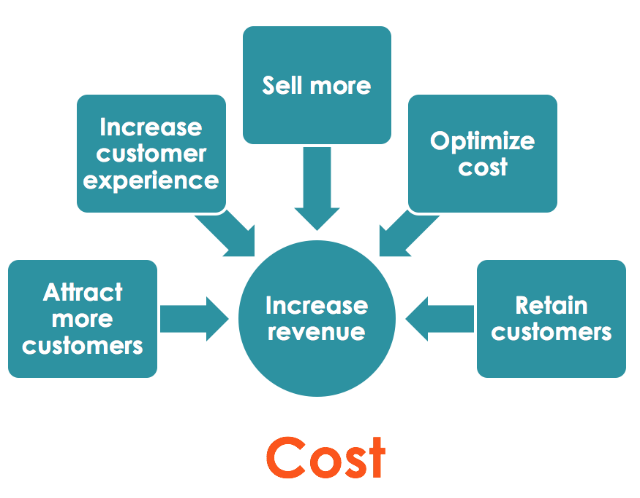
6.5. Overview of relevant articles 44

# Executive Summary

## Business context and project background

Within the context of this Data Sciences Report, GHA’s key objective has been defined to increase revenue, with the latest technology advanced, and data science in particular. In the making of this report, several revenue drivers for achieving this objective have been defined and detailed in the specific use cases. Five main revenue drivers have been illustrated in the figure below. Effective analytics and data sciences is a pre-requisite for implementing all of them. While data also is a fundamental component of these revenue drivers, if leveraged effectively, data becomes much more than a mere component: it becomes a revenue driver and an asset. With this in mind, we have pursued a new approach assessing the value the data assets, as part of the business case for this data science initiative itself.

However, leveraging the data asset demands a new way of working, which is constrained by the traditional BI and data analytical approach employed in most organizations, which is why GHA wants to leverage new advanced technology and data science in order to improve the capacity to increase revenue by harvesting more value from data.



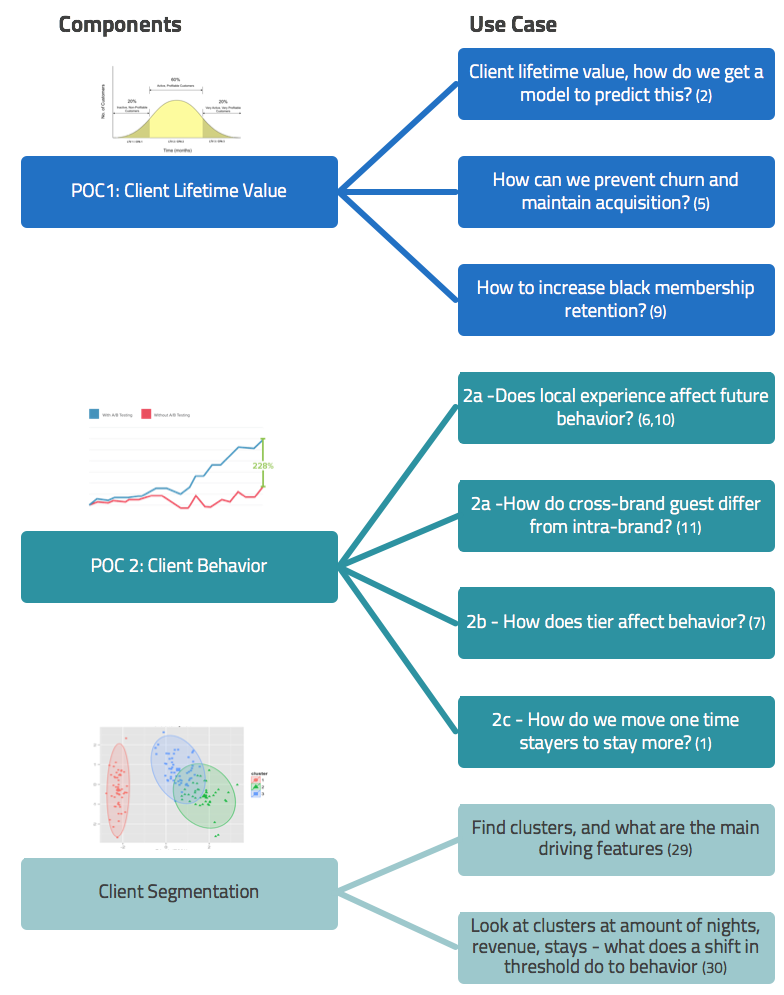
*Revenue and revenue drivers*

## Recommendations - Proof-of-Concept (POCs)

GHA commissioned 2021.AI to run the first 4-week phase of project to initiate this transformation by identifying where and how GHA can achieve this goal by mapping opportunities and making specific recommendations for implementing suitable opportunities rapidly. These recommendations take the form of Proof-of-Concept (POCs), and are developed through an agile process based on identifying use cases, which are suitable from a Data science standpoint and to assess the value of data, its underutilized potential, and prioritizing them into POC candidates. The business case approach is influenced by GHA’s desire to ideally quantify data as an asset and identify the underutilized potential of this data asset.

2021.AI employed an agile methodology to achieve these objectives and two-day London workshop were at the heart of activities. During this workshop, the final short list of thirteen “Priority 1” use cases where defined. From this list, four POC candidates were synthesized. The recommendation is to run these in two main POCs as illustrated hereunder:

* **POC1: Client Lifetime Value**
* **POC2: Client Behaviour (with three sub POC’s in the next phase of the project):**



These POCs are highly suitable because:

* They address the core business objective of driving an increase in revenue
* Data science and machine learning algorithms are available
* Data is available – even though it needs to be analysed further
* They can be implemented rapidly.

Our recommendation is to start with POC2 Client behavior.

## Business case “Data as an Asset”

Assessing the value of GHA’s data asset is not a trivial task, and we will need more information and time to reach a more precise conclusion. Several approaches, have been evaluated, and our recommendation is to continue the analysis during the next phase, based on the Gartner approach based on two main groups and six different evaluation methods:

**The Foundational measures:**

* 1. Intrinsic value
  2. Performance value
  3. Business value

**Financial measures:**

* 1. Cost value
  2. Market value
  3. Economic value

Ideally, we will present these six values in a dashboard for GHA in order to provide a “helicopter-view” of the Asset Value of Data, and the subsequent dynamic valuation of it, given that this changes over time.

In the long term, once the valuation model is robust and accepted, GHA can consider declaring data assets on its balance sheet if the opportunity to do so arises.

**1.3.1 Initial valuation of data**

Only five of eight use cases

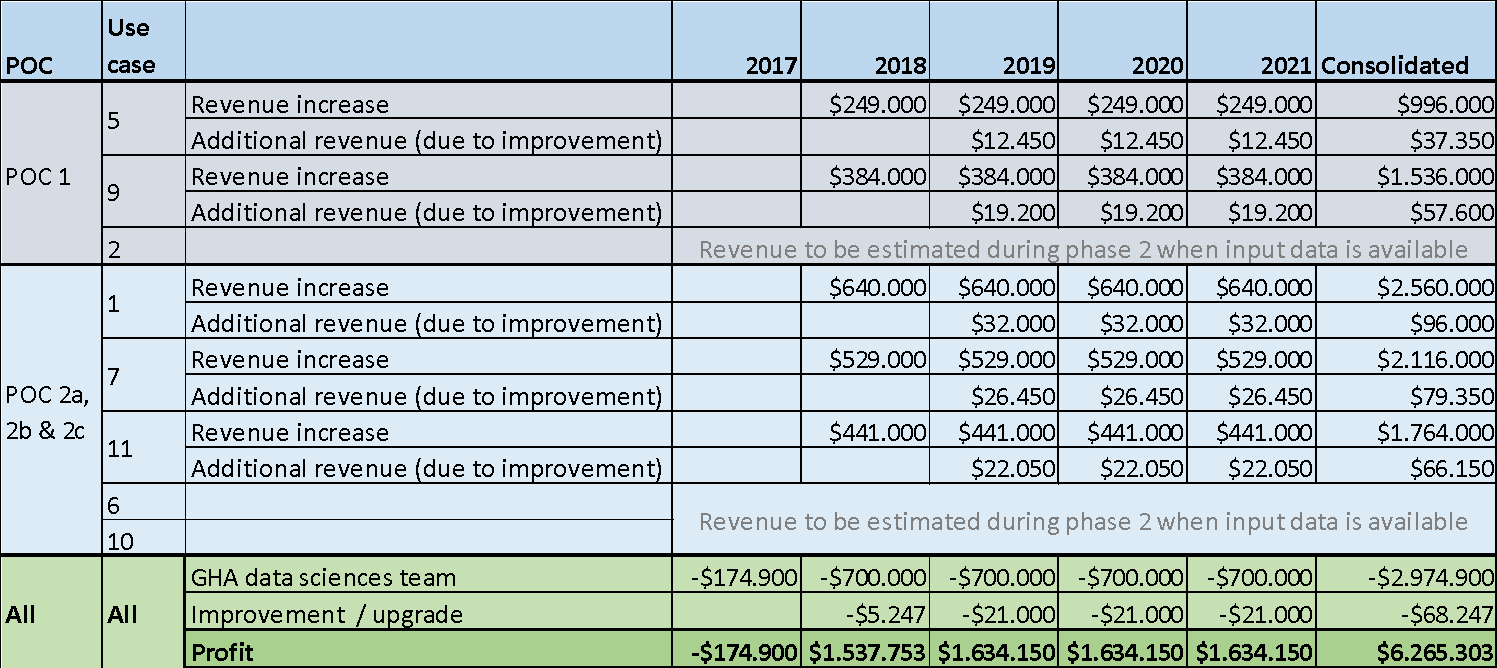
For the first business cases, running POC1 and POC2 as proposed in section 1.2, we have estimated the total revenue increase for five of the eight uses cases in the POCs (three lacked sufficient input data for the revenue calculation).

Only one of six valuation elements included

Our valuation is based on just one of these six valuation elements (the Economic Value of Information). In the next phase we will be able to add the additional five value elements to arrive at a total value - as more data becomes available.

In the table 1.0 below the **revenue increase** per year is combined an **additional revenue** component (yield pick-up by 5% per year, as we fine tuning models and add data) and the **actual cost of operating the GHA data sciences team** as proposed by 2021.AI.

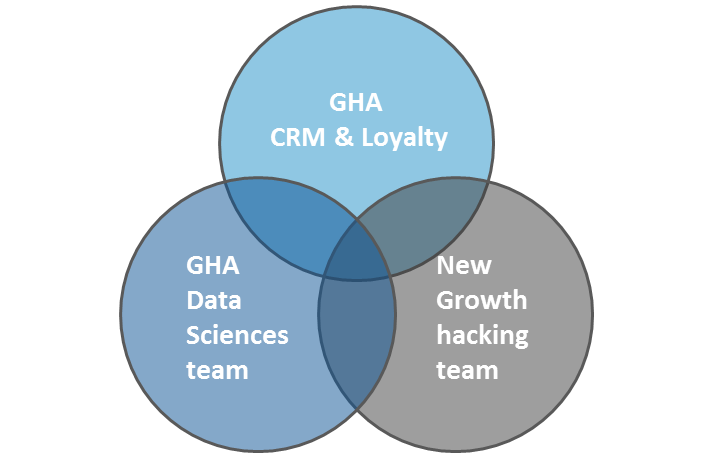
This table clearly illustrates that with a very limited scope and implementation, and based on a very conservative valuation, the yearly profit from investing in a dedicated GHA data sciences team, results in a ROI of more than 200%. In practice the actual returns will be significantly higher, as the data science team has the capacity to implement two to three times as many POCs, as those illustrated in table 1.0.



*Table 1: Business case overview*

The assumptions for the revenue increase calculations are detailed in section 3.6 and in the presentation “GHA Financials”. Please note that this is based on a conservative estimate of no revenue being generated in 2017, as this period in the calculation is used primarily for setting up the program.

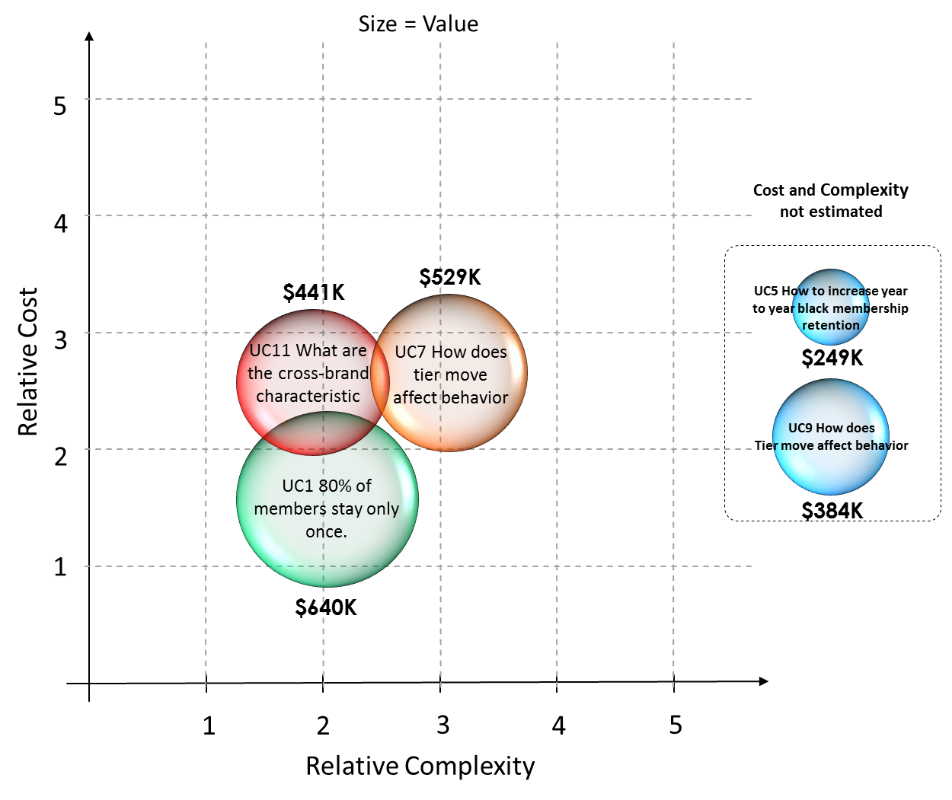
To ensure focus on optimizing the value from the insights generated from GHA’s data science team, 2021.AI proposes establishing a new small, 2-person GHA growth hacking team, whose only focus is to take action based on new insights, working close with the CRM or Loyalty team at GHA. The cost of operating this team is not included in table 1.0 as 2021.AI need more insight to evaluate how much of such team already exists within GHA, before deciding where and how to incorporate.



Though we not perfected the valuation method for estimating the total value of GHA’s data asset, and potential data asset, this first valuation suggests that there is significant value, that GHA will be able to harvest by applying advanced technology and data sciences effectively which will be a double digit X the current estimated value of more than $ 6 mill, which will take the value in the tribble digit $ mill. figure. This is well in line with the current assumptions across industries and analysis, summarized well in this quote:

“For most companies, their data is their single biggest asset. Many CEOs in the Fortune 500 don’t fully appreciate this fact.” – Andrew W. Lo, Director, MIT Laboratory for Financial Engineering

The illustration hereunder indicates relative cost versus the relative complexity for three of the five selected use cases, with the size of the bubble indicating the value.



*Use case opportunity mapping.*

## Technology Recommendations

GHA needs an environment and architecture that supports the implementation of data science and machine learning without infringing, or being constrained by, GHAs existing architectures and Discovery business intelligence platform. If selected, this can be implemented easily with the 2021.AI Platform, given that it meets these constraints and be integrated and used during the POC implementation.

## Data Recommendations

Given the central data mart the main recommendation for data is to enrich it with data from at 2 main sources:

1. Member hotels and their data
2. Social and other third party data

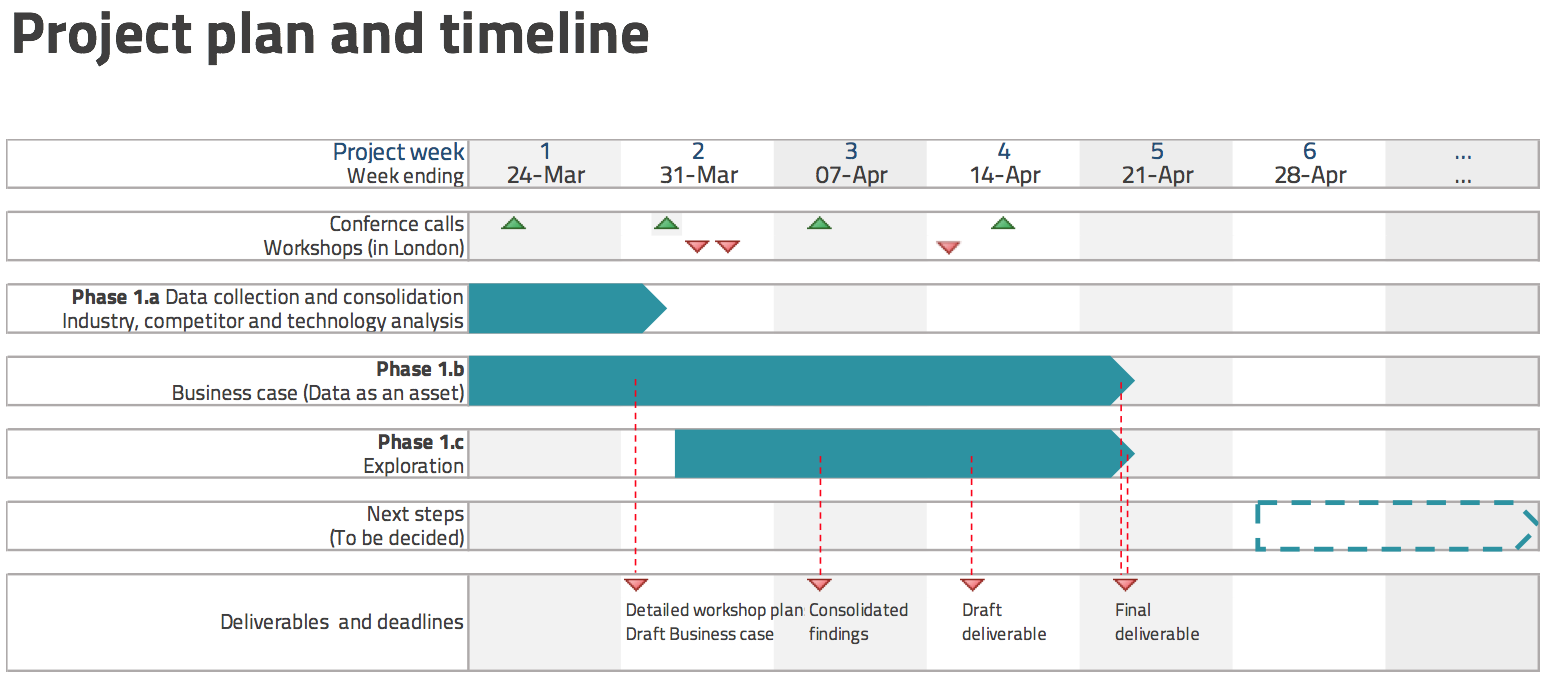
# Background and context

## Project background and context

Based on the first agreement between GHA and 2021.AI to produce a Data Sciences Opportunity Mapping for GHA, 2021.AI have produced this GHA Data Sciences Report. The focus and content of this report is fully aligned with the list of project deliverables (see figure in section 2.2)

Due to the project’s relatively short timeframe (see figure below) it has not been possible for 2021.AI to get access to GHA data. This has limited the analysis of current data and to some extent the analysis of the cost of getting good data. Once 2021.AI receives data from GHA these analyses can be completed. To compensate, other areas in the project and report have been extended to include more details than initially planned, particularly on use case shortlist and their implementation in phase 2. The technology assessment has also been extended, and a full solution to support the POC in phase 2 has already been validated.

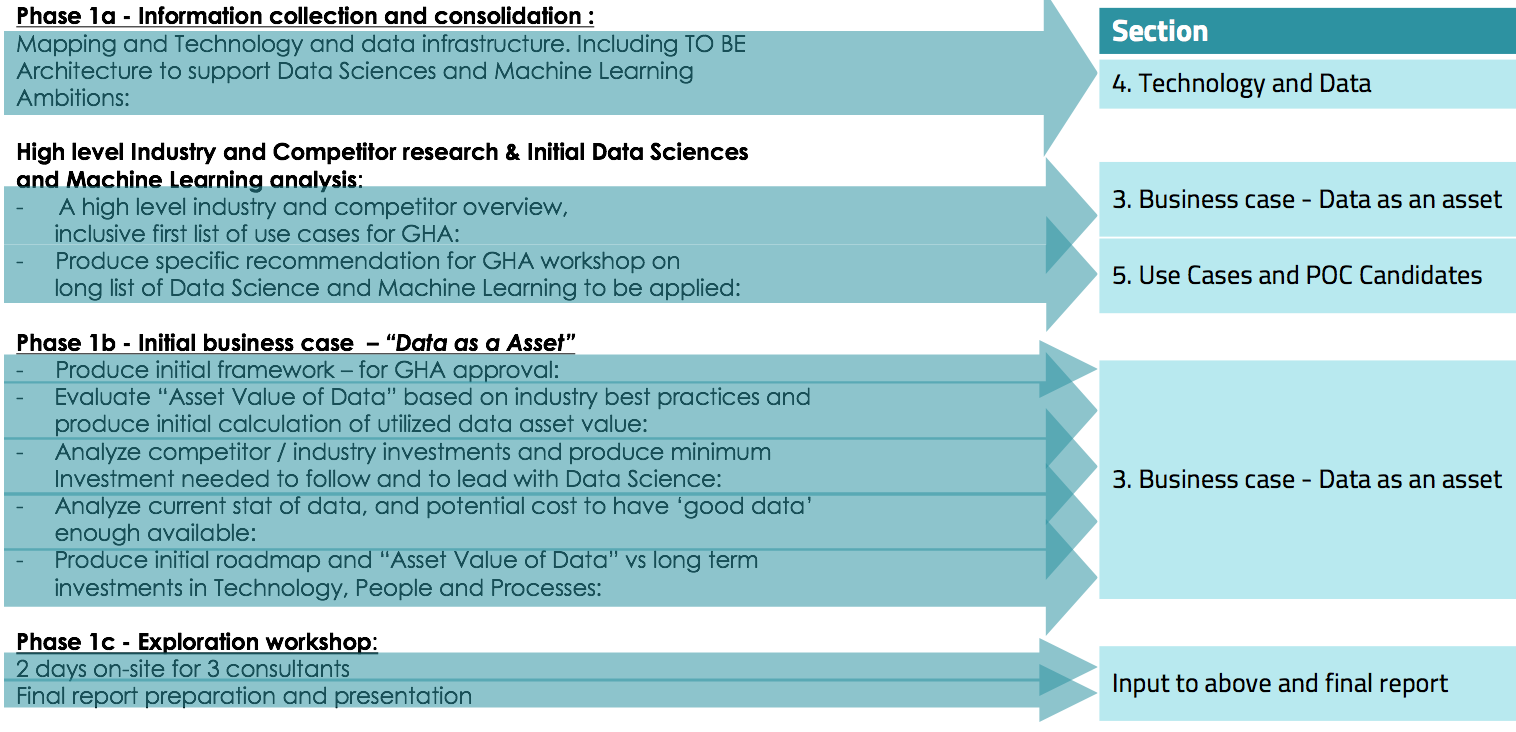
Within the *“Data as an Asset”* analysis the overall competitor and industry investment element has been included, but not to the extent of the original plan, based on guidance from GHA. The focus of this analysis is on evaluating and producing a framework, to enable GHA to conduct concrete assessments of the value of its data in detail in phase 2.



*Project timeline*

## Project deliverables

The list of agreed deliveries from the original agreement is included (in figure below) in a format where, we have mapped the deliveries to sections in this report.



*Project deliverables*

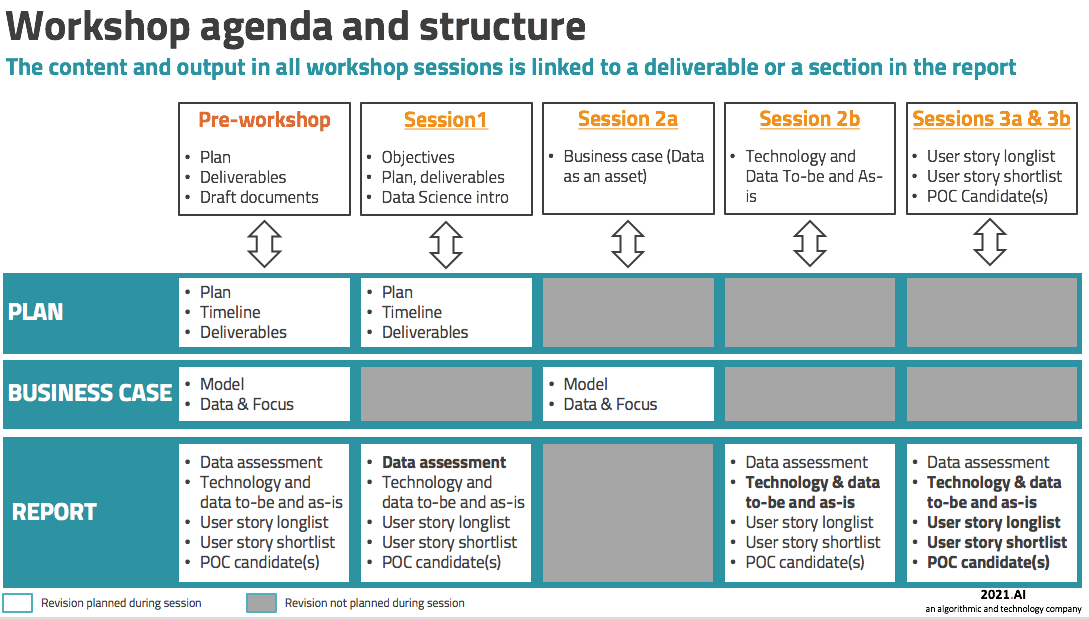
## Project structure and approach

Given the three key themes: the business case, the use cases and the underlying technology and data, project activities were structured within three main workstreams.

1. The **Business case ‘Data as an Asset’** **workstream** focused on assessing how business data could be assessed with the aim of developing a business case to fund initiatives related to this scope of work in the short and long term.
2. The **Technology and Data workstream** focused on understanding GHAs current technology and data environment, with the aim of understanding and addressing implementation constraints.
3. The **Use case workstream** focused on transforming business issues to use cases, and assessing and prioritizing them into POC’s with the aim of implementing them in the next phase.

## Workshop agenda

One of the key activities in the project was the 2-day workshop in London on the 29th and 30th of March. Each session in this workshop (see figure below) was linked to a key theme, deliverables and report section:



*Workshop structure*

## Opportunity mapping

Opportunities have been synthesized from prioritized use cases.

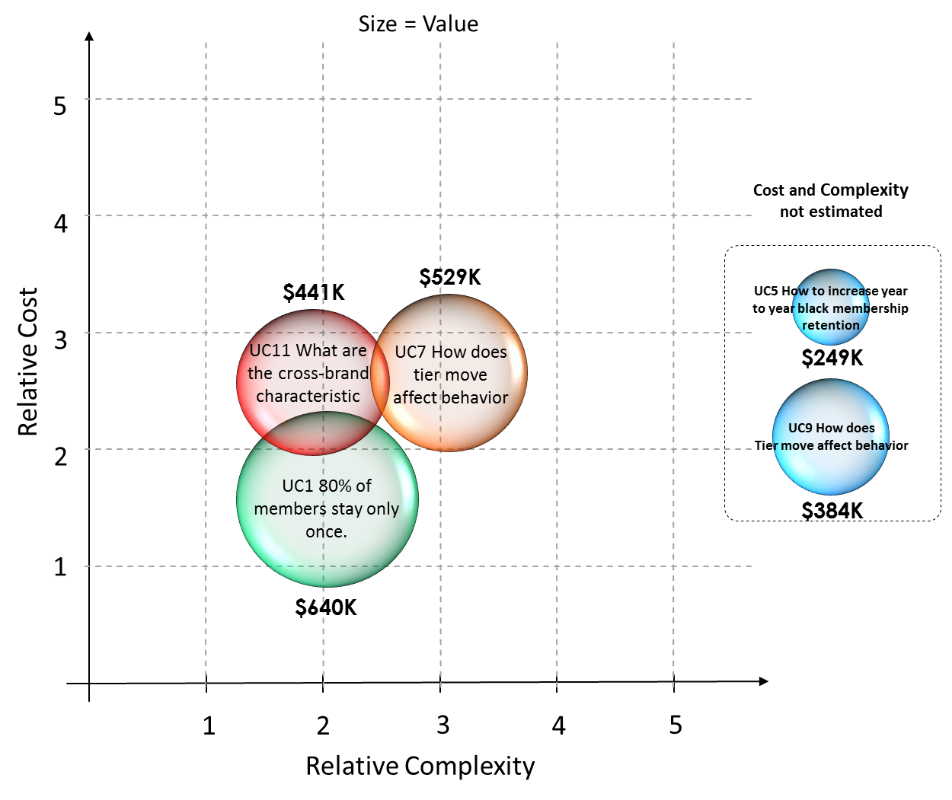
The outcome of one of the workshop sessions was a validated list of 33 use cases. These were prioritized by GHA and analysed by a 2021.AI. Initial findings indicate that there are several suitable use cases and at least four of them are highly suitable.

These use cases relate to two main issues:

1. Understanding and predicting Client Lifetime Value
2. Understanding Client Behaviours
   1. What affects cross brand stay?
   2. What affects a tier move?
   3. 80% of members stay only once, how can we change this?

The goals of these use cases are to drive business value, to enable better understanding of GHA Discovery members, as well as gaining the insight to launch and measure initiatives, and to direct members to become more engaged.

The opportunity map (in figure below) illustrates the relative cost, value and complexity of the proof of concept candidates.



*Use case opportunity map*

## Report structure

The structure of this report is aligned with the structure and approach of the project and the key sections are 3, 4 and 5. The **Business case ‘Data as an Asset’ theme** is addressed in Section 3, **Technology and Data** is addressed in Section 4 and the **Use case theme** is addressed in Section 5.

# Business Case – Data as an asset

## Introduction

Global Hotel Alliance has some characteristics that are not similar to normal hotels, chains and brands. Being a loyalty/partner program, the data GHA gathers has large volumes and applies across continents and the hospitality sector - giving them an informational insight, which none of the hotel brands would have been able to gather themselves.

Increasing, enriching and utilizing this data asset to its full potential is key to GHA achieving competitive advantage. Data science is a mean to leverage asset.

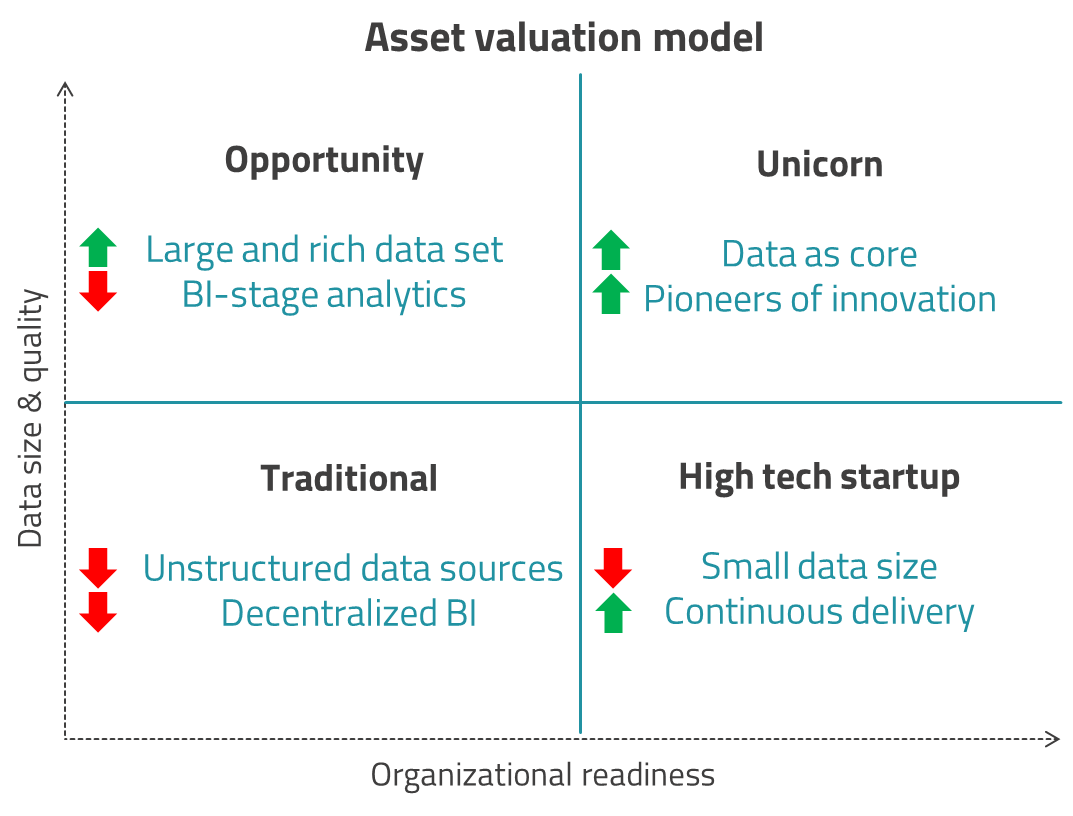
At the same time GHA have challenges that can and should be taken into consideration, because they need to be resolved in the short, medium and long term:

1. How can we clearly distinguish the revenue GHA generate for their hotel brands, from the revenue that the hotels generate themselves?
2. How can we improve stickiness for hotel brands to stay in the alliance?
3. How can we quantify and valuate GHA’s data asset?
4. How can we make better decisions on how and what to do in order to improve the data collection process?

Through solving these underlying question, the technology stack of data sciences can help utilizing the data asset to extract the maximum value from the advantage of having access to cross hotel brand information.

## Market matrix

Different organizations will have very dissimilar potential in harvesting value from their data assets, depending on the type of organization and current data available. This also means that not all organizations will have the same potential *“underutilized data asset”*. GHA has tremendous potential and is in the Opportunity quadrant in the figure below.



*Market matrix*

## Industry, competitors and market

In order to illustrate what potential data science has and how other companies have applied it to their business operations, we have collected a set of background material and references cases. These are a mix of key takeaways from panel debates, academic reports, actual use cases of how data sciences have been applied, and specific targets and measurements related to the return of investment for such projects.

### 3.3.1 Expedia

The online travel agency Expedia which owns more than 200 travel booking websites, including industry leaders such as Hotels.com, Trivago and Travelocity, decided to experiment with artificial intelligence in 2016. They hosted a competition in collaboration with an online network for data scientist, with a prize purse of $25.000.

Expedia laid out a detailed - but anonymized - data set, for their client bookings and gave the following question to the participants:

“Which type of hotel will an Expedia customer book?”

To give a sense of the stage that Expedia was at in terms of utilizing their data asset, they put out this statement together with the description of the competition: “Currently, Expedia uses search parameters to adjust their hotel recommendations, but there aren't enough customer specific data to personalize them for each user. In this competition, Expedia is challenging participants to contextualize customer data and predict the likelihood a user will stay at 100 different hotel groups.”

This is one kind approach to work with AI, and gives a taste of the potential that predictive analytics has on industries and companies with a deep and rich set of data.

### 3.3.2. TripAdvisor

In 2016 we visited TripAdvisor to learn more about the machine learning and how they utilized their data. They have over 350 million reviews, which constitutes an immense foundation for applying AI-models.

These are the key takeaways gathered from our visit:

1. Everything data-driven, not super sophisticated machine learning, but rather continuously running a/b tests optimizing booking and click-through rates.
2. Simultaneously between 20-30 different tests running on the site – driving click through rates and revenue.
3. Heavily optimized using data and tracking, in real-time, semi real-time.
4. Daily bidding on keywords to google to drive traffic towards TripAdvisor.
5. Very structured and industrialized way they used data and statistics to drive traffic to TripAdvisor in continuous manner with very good reporting, continuously tracked and optimized.
6. Engaged culture to develop and drive the business on data and tests.
7. Organizational support from senior management, product management and technology.

### 3.3.3. IRS case

The IRS in United States, made a predictive analytics project on flagging suspicious-looking tax returns. The previous model that the government had used had a success rate of 1% - meaning that just 1 out of 100 inspected cases turned out to be fraud. The new AI-model predicted fraud correctly in 1 out of 4 cases, de-facto decreasing workload by 25x for the IRS.

### 3.3.4. AIIM

The AIIM community held a leadership summit in the summer 2016 with the topic of:

*“How do you measure the value of information?”*

The summit had two dozens of key stakeholders within the industry of infonomics as they have coined the term for the morph of information and economics. They arrived at four key findings:

1. Assigning a value to information assets is not as easy as it sounds
2. How you value information assets ultimately depends on the type of asset and how it is used
3. Infonomics has potential as an umbrella term for this discipline, but is still largely not well understood in the user community
4. The accounting profession may force the question of the valuing of information assets, but likely not anytime soon

These key findings are in line with our research and demonstrates that we are pushing the boundaries for currently available best pratices.

### 3.3.5. IHG

IHG have not cracked the code of data sciences and predictive analytics yet, but are on the verge. They have explored and analysed their user journey, and aligned all marketing to the phase of which they interact with the users. This might be an obvious initiative to do, but it is one that many companies that operate in the digital client journey space have difficulty to master.

The four phases are *dream, plan, book, stay*. Each of the marketing efforts are centred not around the tool itself, but around the phase and in collaboration with other tools - so that the marketing effort feels personalized and user-centric.

Nevertheless, they are primarily using business intelligence techniques and historical data to adjust and cater for each phase of the user journey and using sub-optimizing techniques in each phase in isolation.

There is no sign of predictive analytics or data sciences in their data utilization, but looking at their current state, it is only a matter of time before they will explore AI, since they have closed their gaps on all the other aspects.

## Case framework: Methodology

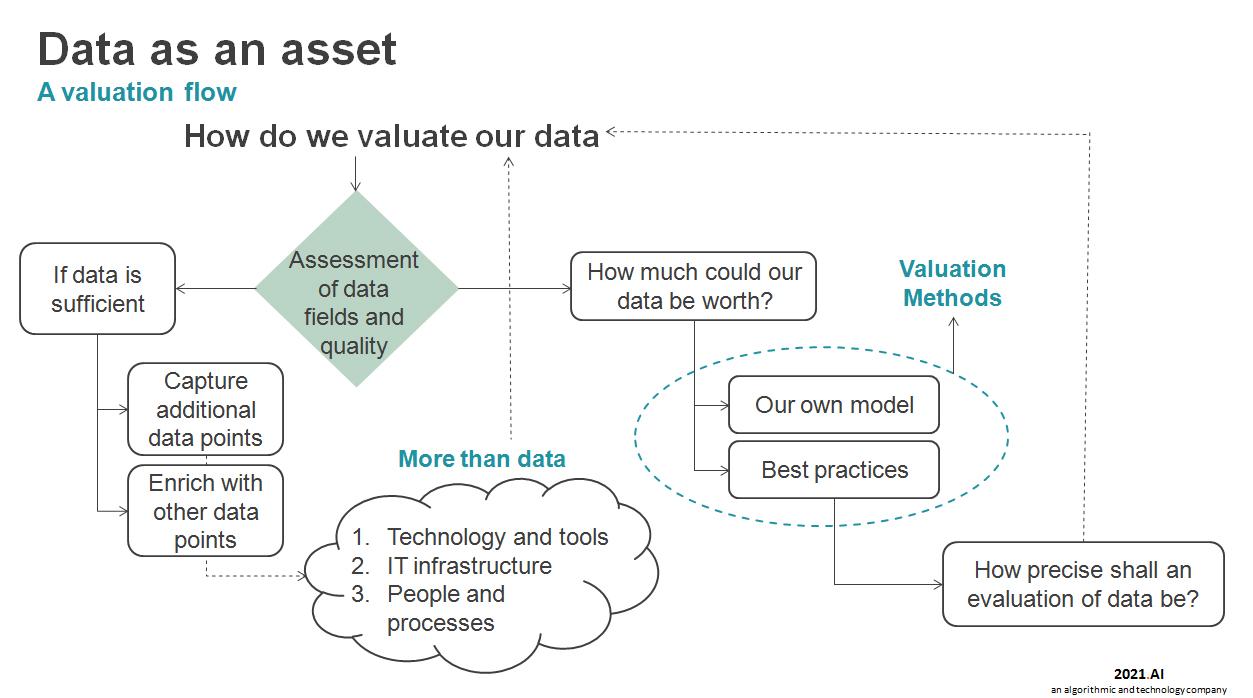
Assessing the value of data in a company is not an easy methodological task. In the AIIM summit they asked if the generally accepted accounting principles in the next 5 years will require information assets to be valued and put on the balance sheets. Just 30% of the user group (consisting of seniors and executives) expects this.

This has pros and cons. The downside is that companies may operate with different forms of valuation of their data asset, making them incomparable. For example, have Netflix valued their algorithm at $1 billion on a yearly basis. This is probably correct, but the valuation method they have been using, could be very different - and probably is - from the one they user at Uber or AirBnB.

On the other hand the upside is that a company with a large and rich data set can utilize and exploit it to a much greater extent and gain differentiation on competitive advantages in a level they have not seen before. When speaking of the huge gap between the book-to-market value of high-tech companies such as Facebook on the exchange or Microsoft’s purchase of Skype and LinkedIn - the AIIM underlines the fact the inability to properly measure the value leads to an undervaluation of the asset: “Simply stated - if you can’t measure it, you won’t value it.”

### 3.4.1. Assessment method: 2021.AI

There is one important factor to keep in mind when you start to assign large and complex data sets with dollar values; the data set is in constant change, which means the subsequent valuation is as well.



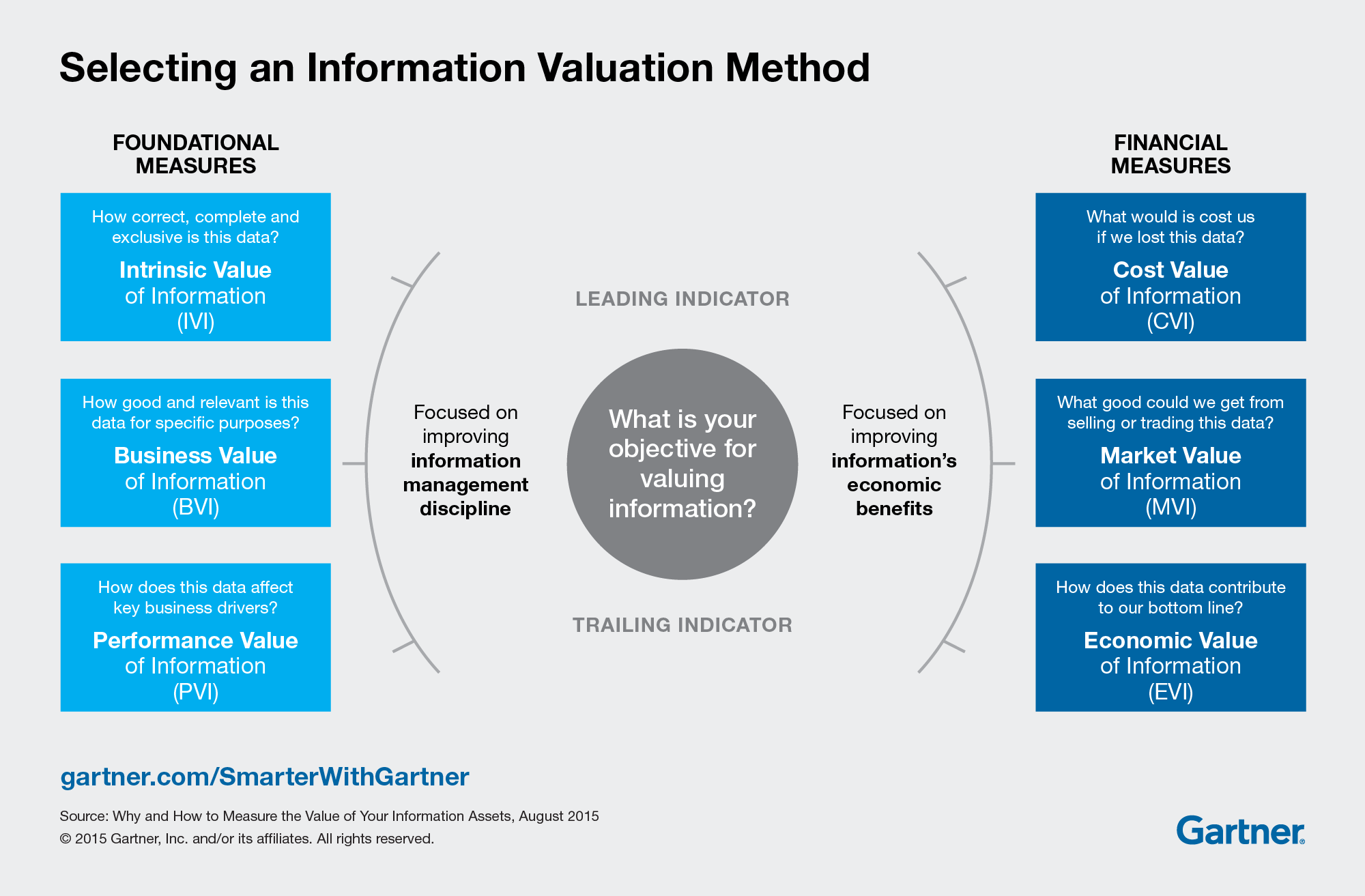
*Data workflow model*

This is something we take into account when assessing your data as an asset. Our model shows the continuous workflow in achieving an accurate view of the value of the data at hand. GHA’s current 10 million data records are of course increasing for every booking that happens, but likewise, the historical data becomes less and less important relative to the new data that flows in.

This lifecycle of data in combination with our workflow model (in figure below), gives the best real-time data assessment of the current data value.

### 3.4.2. Valuation method: Gartner

The 2015 report on valuation of data from Gartner, is a model that quantifies and qualifies data sets in six different measures, which can be applied across industries and on various sizes of the data asset. The model (in figure below) is the predominant in the field of infonomics:



*Gartner valuation method*

**Foundational Measures**

1. **Intrinsic value of information** – breaks data into characteristics such as accuracy, accessibility, completeness, then rates each characteristic and tallies for final score.
2. **Business value of information** – measures data characteristics in relation to business processes.
3. **Performance value of information** – measures data impact on key performance indicators over time.

**Financial Meassures:**

1. **Cost value of information** – measures the cost of “acquiring or replacing lost information” as well as lost revenue caused by loss of data
2. **Economic value of information** – measures how data contributes to revenue
3. **Market value of information** – measures the revenue generated by “selling, renting or bartering” corporate data.

The valuation methods raise the following questions:

1. How can GHA increase the data accuracy to create more intrinsic value?
2. Which ways can we gather more meta data on hotel guests to create better business value for GHA?
3. Are we good enough in utilizing data for GHA’s key business drivers to enhance performance value?
4. How much does GHA spend on data security to protect the cost value of information and is this amount justifiable?
5. Can we use the market value of data to create transparency around what GHA’s cross-brand anonymized data brings of knowledge, to each hotel brand in the alliance?
6. How can data bring more economic value to the hotel and the GHA bottom line?

The purpose of creating and using such an infonomics valuation, is to ask these questions and following up with actions to maximize these values. The ongoing data assessment method, will lead to an updated view of the current data value at hand, at any given time. Thus, any given effort in increasing data value can be measured and evaluated.

This model is applicable and beneficial for GHA’s data set of +10m data records.

## Analysis: Current State of Data

The baseline for working with the GHA data, is the more than ten million data records, each with 129 data attributes, that sum up the stays/bookings done by all members.

The initial assessment of the intrinsic value of the data, is the first step towards slicing and dicing the data to be able to give a holistic overview of the value at hand.

The intrinsic value analysis will be based on the following questions:

1. How can we group the 129 attributes into segments with different values? (i.e. time of booking is worth more than which browser you used when booking)
2. How many of the 129 attributes are available in each record - what are the means and averages?
3. How old is the data, and what is the time value of a data record? (i.e. a booking in 2011 is worth less than a booking in 2016)
4. How enriched is the data set with social media sources? Can we link social media tracking to the data set?
5. Can we link data records with website searches and email campaign records? If yes, how does that affect the mean and average of a given data record - and the number of attributes?

In assessing the valuation of the data asset, we will be looking at value indicators at a micro- and macro level.

1. Micro: Use cases in the POC (performance value)
2. Macro: Data asset as a whole (intrinsic, cost and market value)

The micro-perspective is analysed in the following section, whereas the macro-perspective will be analyzed through phase 2.

## Asset Value of Data

There is a bit of a journey from predictive analytics to a dollar-based valuation of the entire data set at a given company.

The use cases are nothing but ideas on where we think that data analysis could lead to better insights. If our underlying hypotheses in the use cases are correct, the data analysis will give us rough indications on which patterns that leads to different actions, thus giving us the ability to generate predicted user journeys.

The benefits of these insights are only reaped if we know how to adjust existing marketing effort to pull and push information to the guests based on these pattern analysis. These efforts must then lead to more engagement, more bookings and a longer life-cycle at GHA. Each metric is linked back to a KPI which has an underlying dollar-based value. A given change in that value can be pointed back to the effort done on the basis of a use case.

The sum of the use cases gives you the performance value of information, and is a continuous value assessment that must be maintained and improved, year for year.

A few assumptions are has been made:

1) The distinction between which stays are in the same hotel, intra-brand and cross-brand have not been made yet. This will be done in phase 2, after a mapping between stay codes and guests. For now, a standard 3% of room revenue is attributed GHA.

2) The number of active members in each category is calculated using the Q1 2017 numbers for stays and forecast these numbers for 2017. They are then multiplied it by the average stay per membership type (2016 numbers). They also reflect the 96%, 3%, 1% split as discussed during the workshop.

3) The local redemption cost has not been factored into the equations, but is of high importance in the long-term calculation, and requires some extracts on average reimbursement and amount of redemption per segment and so forth.

### 3.6.1. Use case #1: Second-night bookers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | Use case | Affected measurement | Performance indication  (1-3%) | Conservative target (1%) | Value per booking | Annual use case return |
|  | 80% of all guests only make one booking. Can data science help them on their journey to do the second booking? | The measurable factor is the value of the second booking. But the long-term benefit is the loyalty value. | What is the value if we can get 1-3% of current Gold members to book a second time? | 54,844 | $ 11.67 | $ 640,029 |

**Assumptions**: Only the value of the second booking has been factored in. The enhanced loyalty from members is an additional benefit and is not included. The potential value per booking is measured by taking the average GHA revenue (3%) per booking in the gold segment. The conservative target (1%) is derived from the lowest end of the estimate range.

### 3.6.2. Use case #11: Cross-brand / intra-brand

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 11 | Use case | Affected measurement | Performance indication  (1-3%) | Conservative target (1%) | Revenue per stay | Annual use case return |
|  | How do we distinguish and affect cross-brand guests? Cross-brand stays generate 3x the profit (7.5%) than recurring hotel and intra-brand (2.5%) | The GHA size of revenue from hotel brands. Probably only realistic to do so in Platinum and Black segment as long as use case 1 is at the 80% level. | Can we move 1-3% of the overall stays in the Platinum & Black segment into cross-brand hotels? | 21,654 | $ 20.40 | $ 441,733 |

**Assumptions:** The first assumption is that all Gold member stays cannot be affected, as the majority of these are one-off stays and thus have limited data to perform a cross-brand marketing effort. Therefor it is worth noticing that the value of this use case will increase, if you can solve use case #1 (getting Gold members to book more than once).

The revenue per stay is measured as the average spend per stay (ASPS) across segments, times the difference between a cross-branded stay and an intra-branded stay (2.5% versus 7.5%). The conservative target (%) is derived from the lowest end of the estimate range.

### 3.6.3. Use case #7: Tier upgrades / behaviour

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 7 | Use case | Affected measurement | Performance indication  (5-10%) | Conservative target (5%) | Value per guest | Annual use case return |
|  | Will a tier upgrade affect the behavior of a guest, and is there any way that we steer the client towards that behavior? | Membership type: A black member is on average more profitable than a Platinum member, by a factor 2.5-3x. | What is the value if we can get 5-10% of current Platinum members to become Black? | 9,851 | $ 53.71 | $ 529,109 |

**Assumptions:** Value per guest is the difference between what an average Black member generates to GHA per year compared to the average Platinum member. The conservative target (5%) is derived from the lowest end of the estimate range.

### 3.6.4. Use case #5: Minimize client churn

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 5 | Use case | Affected measurement | Performance indication  (10-15%) | Conservative target (10%) | Value per guest | Annual use case return |
|  | 46% of Platinum earners in 2015 did not return in 2016. Can we decrease that number through AI? | The churn rate is high. A 10% decrease would be significant (4.6% of earners) | What is the value if we decrease churn of Platinum members by 10-15%? | 9060 | $ 27.52 | $ 249,426 |

**Assumptions:** Value per guest is the difference between what an average Platinum member generates to GHA per year compared to the average Gold member. A decrease of 10% is in fact just 4.6% of the total Platinum member base. The conservative target (10%) is derived from the lowest end of the estimate range.

### 3.6.5. Use case #9: Black membership retention

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 9 | Use case | Affected measurement | Performance indication  (15+%) | Conservative target:  50% re-win | Value per guest | Annual use case return |
|  | Minimizing churn in the most profitable guest segment is important. How to increase year to year black membership retention (only 35% re-won black tier from 2015 to 2016) | The additional value from a Platinum to a Black member is the affected measurement | With such a low current re-winning rate of clients, improving this measure to 1 out of 2 clients should be possible. | 7166 | $ 53.71 | $ 384,873 |

**Assumptions:** The underlying data for a Black member is much larger than for the average Gold and Platinum member, thus the validity of the algorithmic models will be greater and the insights and predictions more accurate. The conservative target (15%) is the minimum we should be able to improve this KPI with.

### 3.6.6. Use case #10: Local experience / non-redeemers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 10 | Use case | Affected measurement | Performance indication | % in numbers | Value per redeem | Annual use case return |
|  | The underlying anticipation for use case 6. If use case 10 is generating patterns, then use case 6 is viable. The expected behavior is unknown. | Cluster analysis and other machine learning techniques will determine the affected measurement. | Activation of non-redeemers might be profitable in the long term. | Can only be estimated in phase 2 | Can only be estimated in phase 2 | Can only be estimated in phase 2 |

**Assumptions:** There are no assumptions in this use case, which is the reason we do not predict the return of this one. The exploratory methodology to solve this use case, will lead to given metrics - which will be the underlying metrics of overall KPI's.

### 3.6.7. Use case #6: Local experience / cross-brand

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 6 | Use case | Affected measurement | Performance indication | % in numbers | Value per redeem | Annual use case return |
|  | Cross-brand stays have different LE-reward opportunities. If redemption affects that behavior, how do we get more guests to redeem? | Clients that stay cross-brand are more profitable for GHA than intra-brand guests - also after considering the LE cost. | If our anticipation is correct, we should steer our attention towards getting more guests to redeem their LE. The hypothesis is a pre-requisite for calculation. | Can only be estimated in phase 2 | Can only be estimated in phase 2 | Can only be estimated in phase 2 |

**Assumptions:** The value can only be calculated as the result of the leading metric identified in use case 10.

## Recommendations

To further evaluate the Asset Value of Data and calculate the underutilized data asset value, our recommendation is to continue the analysis into phase 2 and evaluate according to the six different measurements identified under phase 1. The six measurements are group in a financial and a non-financial group:

1. **The informational measures (intrinsic, performance & business)**
2. **The economical measures (cost, market & economic)**

Ideally, we should evaluate how to make these six values presentable in a dashboard for GHA to have a helicopter-view of the Asset Value of Data, and the subsequent dynamic valuation of it, as the data sets changes over time.

We also recommend using such a dashboard to gather insights and keep track of the outcome of the Data science initiatives in place to increase the data pool with new data or enhance and enrich existing data. The value-add from such initiatives will directly inflict one or more of the six data value measurements.

In the long term, once the valuation model is robust and accepted, GHA can consider declaring data assets on its balance sheet if the opportunity to do so, exists at that point of time (instead of being limited to using "goodwill" or "intangible assets".

Finally, yet importantly, we recommend that 2021.AI on GHA’s behalf, continues to build up a knowledge base - together with GHA - of market research and competitor insights within the hospitality sector and the affiliated data utilization of such. This is to position GHA as an industry leader of advanced analytics to gain a competitive advantage through utilization of the latest developments within Data science.

# Technology and data

## Introduction

This section looks at the current GHA technology and data environment, identifying its strengths and weaknesses relative to the requirements for data science and the implementation of new technology. During the process it identifies issues and it concludes by making recommendations for resolving them..

## Technology and data As-is

GHA uses the Oracle CX Cloud suite, as well as very few bespoke components to enable GHA member brands without the Oracle infrastructure to upload stays and get vital information about guests that are members of the Discovery loyalty program.

### 4.2.1. GHA Technology Components

GHA Core Components

* Discovery (GHA Central Solution ORS/OCIS - OPERA Reservation System/Information System
* Oracle Business Intelligence
* Oracle Responsys (CRM)
* GHACRS (Reservation System)
* OWS OPERA Web System (interfacing)

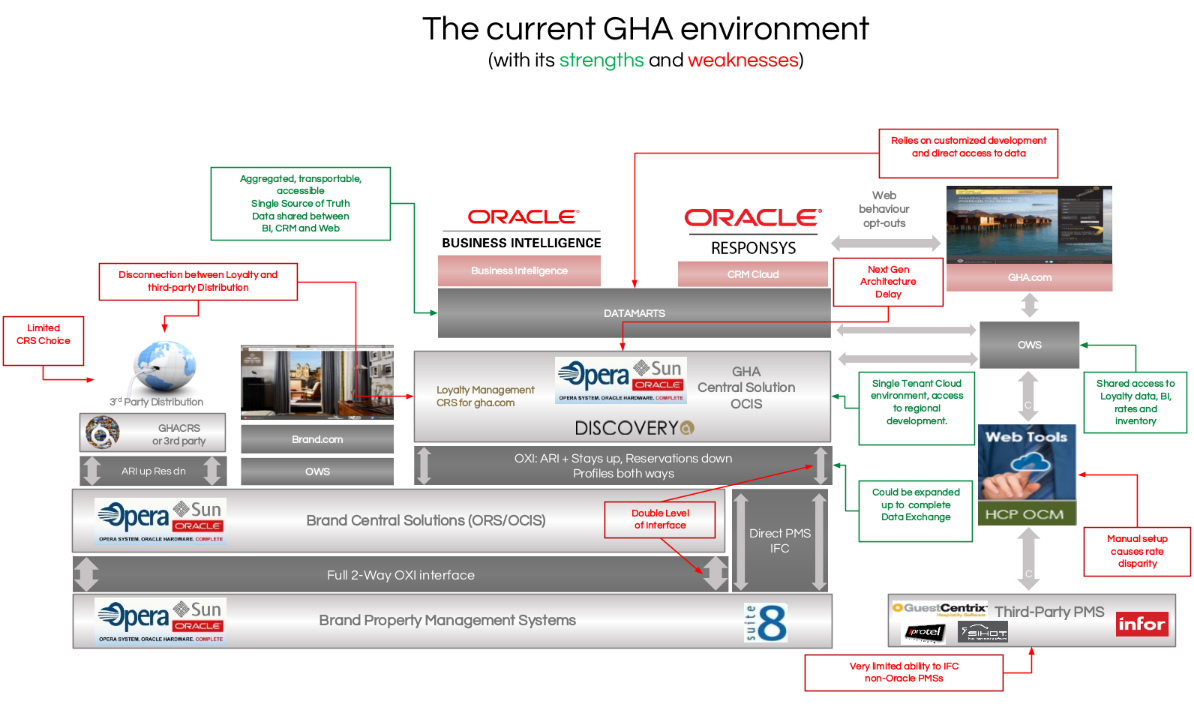
Brands Core Components

* Brands Oracle PMS
* Brands ORS/OCIS - OPERA Reservation System/Information System

Brands 3rd Party Components (non-Oracle)

* Third Party PMS
* Web Tools (manual entry)

These components can also be see in the figure below.



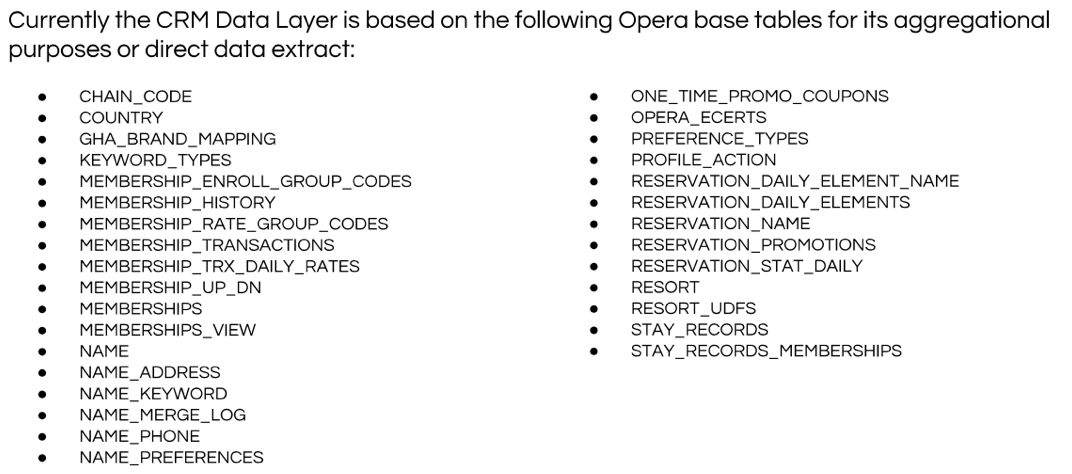
### 4.2.2. Data

Due to the streamlined infrastructure of GHA data environment, it is relatively easy to access data, as virtually all the data is available centrally. GHA does not struggle with the common issues of enterprises, namely with multiple systems, departments, and a mix of legacy systems and multiple sources of the same data.

This streamlined infrastructure is an enabler for quick implementation, as the whole technical and data infrastructure can be seen as normalized and as the “golden master” of all data records.

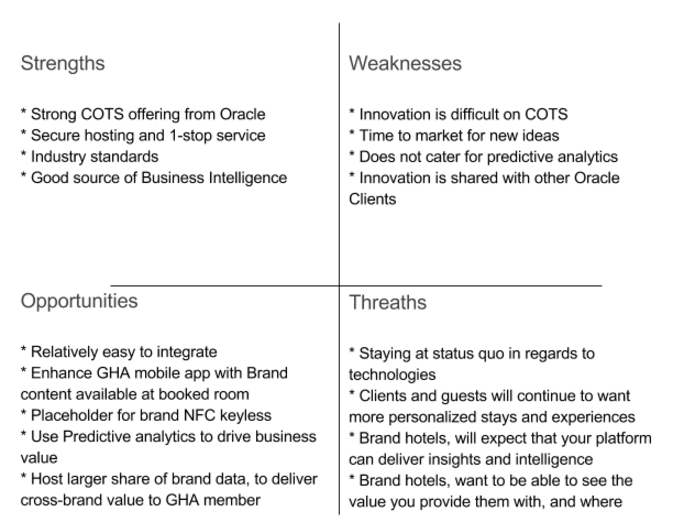
Aggregated guest data is stored in a table with 155 columns, giving GHA employees access to a wealth of business intelligence. This table is updated daily - and always available 24/7. Brand Hotels have access to their own guest data set. Once a guest has completed a stay at another brand in the GHA family, this brand also has access to this guests information from the time of his/her first stay at the brand.

This data is aggregated from the following core tables (see figure below):



### 4.2.3. Technology and Data - SWOT

Based on the current level of understanding 2021.AI have produces a bulleted assessment list of the areas related to technology and data in regards to the current GHA technology and data environment:



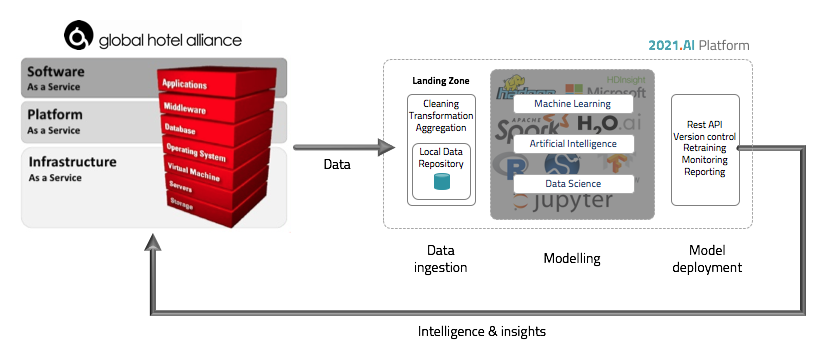
*SWOT analysis*

## Technology and data To-be environment

To be able to leverage the data assets that GHA have accumulated it is key for any future environment to support the use of algorithms and experimentation, as well as an analysis of the large datasets, GHA needs to have a strategy for implementing advanced technologies that enable the processing of large datasets to deliver analytical insights that can be channelled directly into the current business intelligence platform, namely Oracle.

Given the importance of Oracle, the To-be environment needs to be implemented with a phased, non-invasive approach.

1. POC - run fully hand held, and run using GHA Discovery Data, in a 2021.AI hosted self-sustained secure cloud environment - to prove business case(s) and that the use cases and POC hypothesis are met. (see figure below)
2. POC - results delivered to GHA Discovery Platform. In parallel to the 2 steps above, investigations are conducted between GHA, Oracle and 2021.AI to find best integration strategy. This investigation includes the mapping and identification of resources needed to use and maintain the platform, both from a GHA staffing point of view but also any other related costs.
3. Use findings from POC Setup of sandbox environment based on recommendations found in integration strategy.
4. Creating a “Go-Live” Plan for a GHA “To-be” infrastructure based on 1-3.



*Overview of 2021.AI Platform, its components and their integration*

### 4.3.1 Technology Recommendations

GHA needs an environment and architecture that supports the implementation of data science and machine learning without infringing, or being constrained by, GHAs existing architectures and Discovery business intelligence platform. If selected, this can be implemented easily with the 2021.AI Platform, given that it meets these constraints and be integrated and used during the POC implementation.

The 2021.AI Platform

The core of the 2021.AI platform consisting of a Linux node, with python and R as main programming kernels along with the Jupyter superstructure. The core platform caters for the main underlying frameworks for deploying and utilizing models and algorithms for R and Python libraries.

Depending on the amount of data and data sources, either .csv files or a local staging database will be deployed. If the data volumes require larger datasets one or more instances of Hadoop will spawned to cater for this. For the POCs the .csv approach should suffice.

Access to the 2021.AI environment is possible in several ways: a graphical user interface can be opened via a browser, or direct access (manual or programmatic) can be implemented using the industry standard for secure FTP.

Integration with Oracle and the existing environment will be managed with a series of steps, in close collaboration with GHA’s technical team:

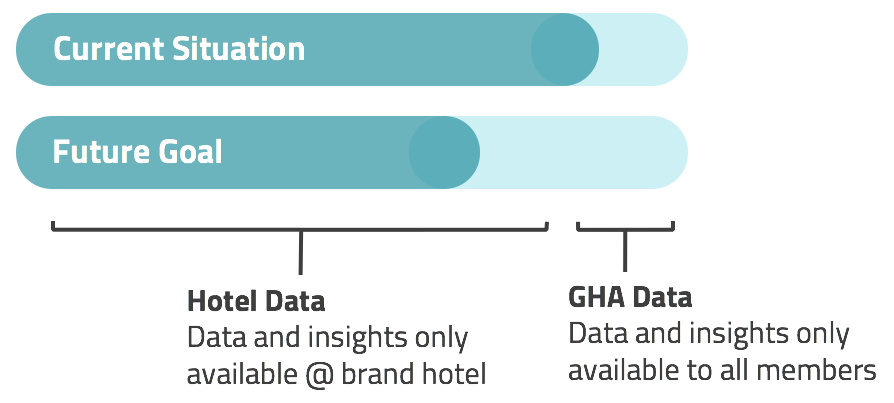
1. Design and implementation of Oracle integration points
   1. We are already working with technical consultants in Oracle Nordics, and don’t expect issues in this regard.
2. Design and implementation of Data inputs and outputs
   1. We have already created repository on our 2021.AI cloud so that GHA can securely (the connection is encrypted using RSA with 1280 bits and the RSA SHA256 algorithm) FTP their files to and from the 2021.AI platform.
   2. Extensions to the current database schema to accommodate for the measures calculated by the 2021.AI platform.
   3. A description of the data that is required to run the models and algorithms. This will be analysed.
   4. A description of the format that the results of the models and algorithms is delivered in. >> .csv’s
   5. Description and guide to the manual upload process for the POC.
3. Design and implementation of the processes for the POC candidates - from data ingestion to result delivery, please see Appendix 4 - Framework for conducting the POC candidates.
4. Detailed requirements and installation of hardware and infrastructure as well as recommendations for security.

### 4.3.3 Data Recommendations

Given the central data mart the main recommendation for data is to enrich it with data from at 2 main sources:

1. Member hotels and their data
2. Social and other third party data

The advantages with the first source are not limited to GHA. Not only does GHA stand to gain, but the member hotels and guests as well. The figure below shows that more insights and data that has business value currently only resides at the GHA brand members and is not centrally available for analytics and personalization.



*Illustration of valuable insights not in GHA data mart*

Given that data is already shared between these parties, the likelihood of regulatory constraints is significantly lower. While the second source is more difficult to access, its advantages cannot be discounted and where available, should be leveraged. Given the complexity of this issue, however, it will need to be addressed in further detail during a late stage.

Benefits of centralizing data

* Give more insights to member and cross-brand hotels
* Ensure Guests get what they want – by sharing data
* Better cross selling opportunities
* Makes exploratory analysis possible
* Enriching GHA data with brand insights
* Win-win - making intelligence available to GHA and members

# Use Cases and POC Candidates

## Introduction

This section is at the heart of the 2021.AI methodology, which focuses on implementing solutions quickly. To do so the approach employed is to define business requirements as use cases, with the aim of consolidating one or more of them and implemented them rapidly in POCs.

During this project, a long list of use cases was developed and eventually transformed into four POC candidates.

A good POC candidate is one where the required data is available, and the value of the insights are quantifiable, and where business experts can run experiments to see if it is possible to change the behaviours. This can be done using A/B testing, as well as benchmarking against historical data.

Our way-of-working with use cases and POC candidates is as follows:

* Inform about various stages of the opportunity mapping
* Data science and machine learning background
* Walkthrough various illustrative use cases
* Opportunity mapping using Co-Creation
* Ranking and prioritization of the use cases
* Analysis of high priority use cases by 2021.AI data scientist
* Revert with finding and grouping of the priority 1 use cases to GHA
* Select of 2-4 use cases as POC candidate(s)
* Make high-level plan for executing the POCs, including measurable success criteria - as well as the POC ranking parameters: Business Value, Complexity, Time to Market and Cost.

## Use Case shortlist

During the London workshop, 33 use cases (see appendix) were ranked on a scale of 1 to 3, resulting in an initial list (see table below) of 11 “Priority 1” candidates:

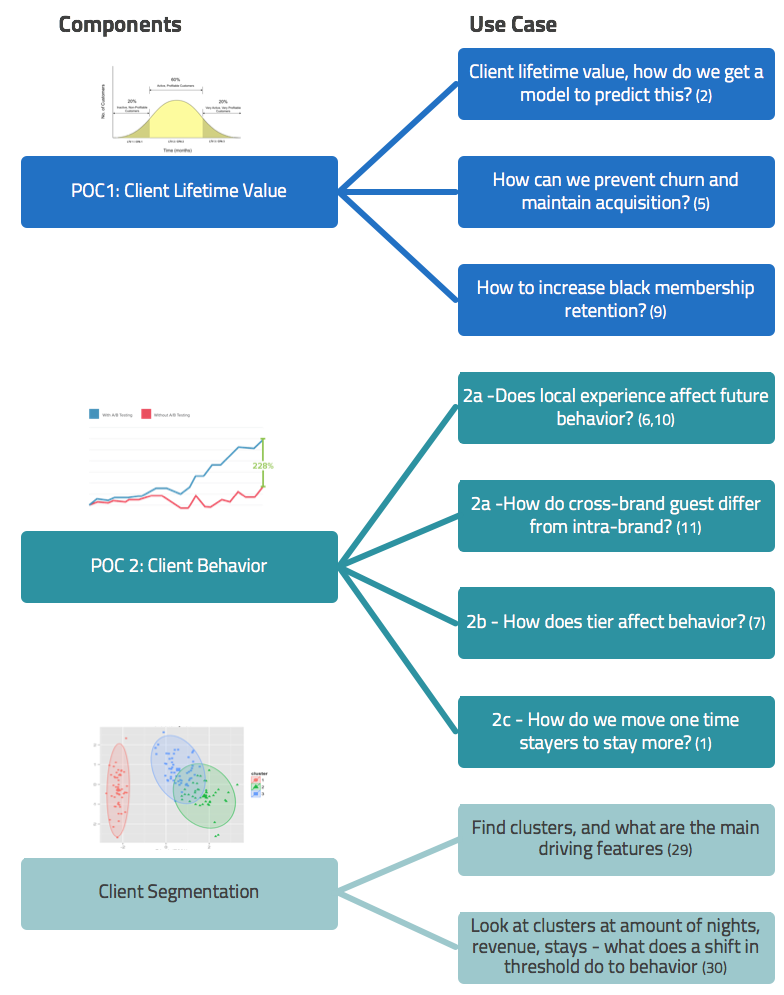
|  |  |  |
| --- | --- | --- |
| **Priority** | **UC#** | **Use cases** |
| 1 | 1 | 80% of members stay only one stay - how do we move these to stay another night? |
| 1 | 2 | Client lifetime value (CLV) how do we get a model to predict this? can we use historical data and historical outcome to estimate this - what is the accuracy of our model? |
| 1 | 3 | Codes for stay channel - need golden source OTA vs. non-OTA (how good are we at turning OTA guest into direct bookers |
| 1 | 4 | Fast movers: who is accelerating |
| 1 | 5 | GHA has high acquisition and high churn (how can we illustrate the hot-spots) and how can we prevent the churn, and maintain the acquisition |
| 1 | 6 | How does local experience redemption affect cross brand stays |
| 1 | 7 | How does tier affect behavior - Do your frequency change/avg. room rate change |
| 1 | 8 | How to get the rights segments for “Formula One” and "Ultratravel Collection" ... together with clusters of hotel categorization |
| 1 | 9 | How to increase year to year black membership retention (only 35% re-won black tier from 2015 to 2016) |
| 1 | 10 | Does Local Experiences redemption affect future behavior compared to non-redeemers |
| 1 | 11 | What are the cross-brand characteristic - how do cross brand guest differ from intra-brand |

*Initial list of 11 “Priority 1” candidates*

## POC candidate recommendation

The “Priority 1” use cases were linked to key components (see figure below) based on the similarity of underlying algorithms and methodologies in them as well as the relationship between these use cases and two key themes: Client behaviour and Client lifetime value:

1. POC 1 – relates to client lifetime value
2. POC 2 – relates to client behaviour



*Relationship between key components and Use cases*

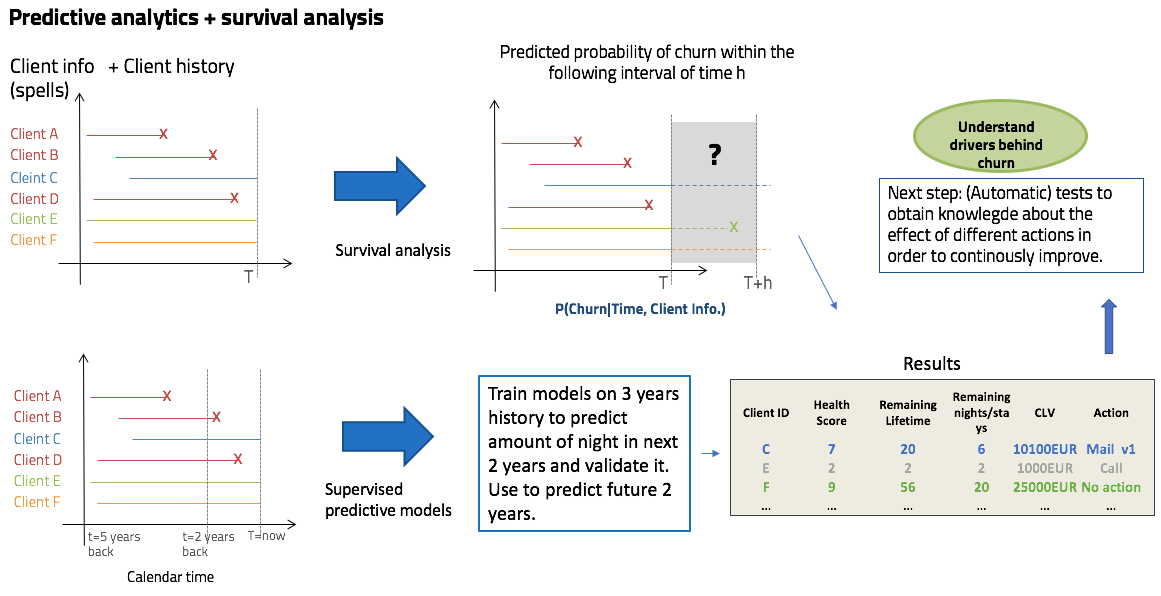
We recommend starting with POC2c – “How to get clients to stay more”, as this has the highest business priority, and our initial analysis shows that the correlation between behaviour or patterns that can easily be affected by direct targeting of a select group of Discovery Loyalty members, can drive an increase in revenue.

### 5.3.1 Candidate 1: Client lifetime value (UC#2+5+9)

An estimate of a client’s life time value requires good insight into customers stay history, especially with regards to hotel bookings, as there is often a long time between bookings. While it is difficult to say exactly how much historic client life time data should be collected, the rule of thumb is 5 years.

The proper way of estimating client’s life time would be through a prediction of the number of stays for a set future period of time. From a data science standpoint this would require survival analysis and a supervised learning approach to predict either predict the probabilities for a given customer to stay in the future, or train a model to predict the amount of future stays for a given period. This is then transformed to client lifetime value based on the client's previous spending and stays.

Besides putting a value on client spending, we will also benefit from predictions of the number of stays in the future for each client, to identify clients that are on the way to churn or downgrade.



*Illustrative overview of the data science algorithm employed for Candidate 1*

Data requirements: This task would require as data for the last 5 years including meta data on customers. Key data fields include Signup date, history of all stays with date of stay and spending and possibly other variables such as experience redemption or not.

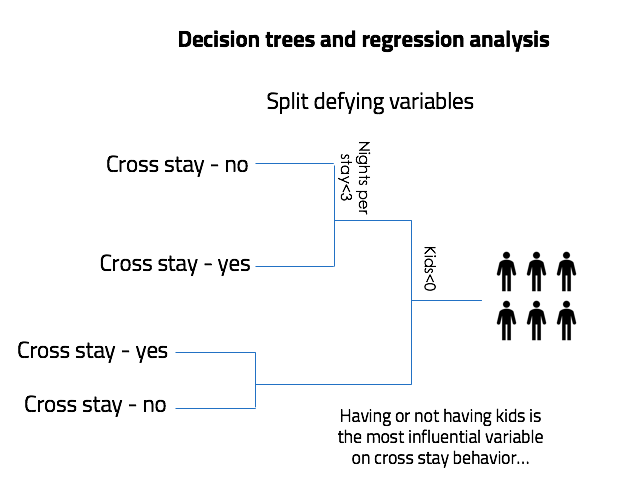
Implementation complexity: 3/5

Metrics: This exercise would be performed on all customers and provide insight on future client behaviour which should act as facilitator for actions.

Algorithms and models: Survival analysis + std. supervised learning (cox , elastic net, tree methods: random forest, decision tree, boosted trees, support vector machine, neural nets)

### 5.3.2 Candidate 2a: Client behaviour - What affects cross brand stay? (UC#11+10+6)

Using regression analysis or trees method, it is possible to identify what has an effect on the desired measure of outcome. Required input data includes what potentially can have an effect on the outcome, i.e. cross brand stay. The outcome would identify how big an impact (if any), an average number of stays per night for example, has on cross brand stay. Besides identifying drivers, it will also be possible to apply the model to predict probabilities for cross brand for all clients.



*Illustrative overview of the data science algorithm employed for Candidate 2a*

Data requirements: For this exercise, explanatory variables for each clients are very important. Historic data for clients, including those who did or did not have a cross brand stay. Required input data also includes available metadata on client stay behaviour (nights per stay, frequency, type of stay…), experience redemption etc.

Implementation complexity: 2/5

Metrics: tbd

Algorithms and models: Regression analysis, and here there are several options depending on the variables and trade-offs between complexity and understanding of behaviour. There are both linear and nonlinear approaches each containing different models that can be applied.

### 5.4.3 Candidate 2b: Client behaviour - What affects a tier move? (UC#7)

Tier move is directly related to amount of stays per year/period which is the outcome variable to be investigated. As in the earlier case, if the purpose is to find what has effect on this, a regression analysis should be performed to investigate the effect of the different explanatory variables. Another approach could be to construct a transition probability matrix for each customer given the state they are in.

These predictions could serve as an indicator for a potential move from one tier to another. For this exercise, explanatory variables for each customer is very important. Required input data includes historic data for clients who did or did not have a cross brand stay and any available metadata on clients, stay behaviour (nights per stay, frequency, type of stay…), experience redemption etc.

Data requirements: Same as candidate 2a.

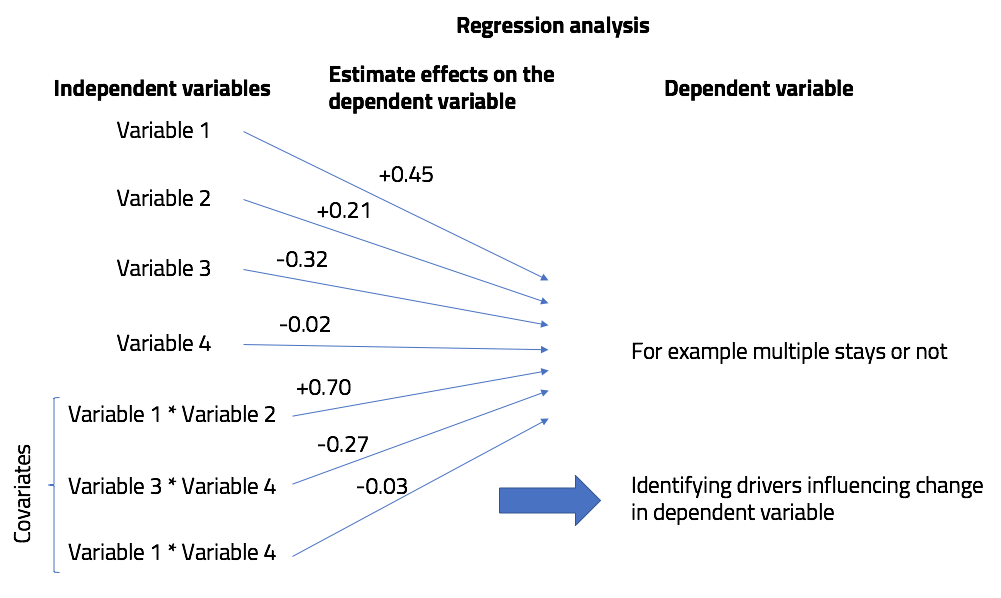
Implementation complexity: 3/5

Metrics: tbd

Algorithms and models: Same as candidate 2a

### 5.2.4 Candidate 2c: Client behaviour - 80% of members stay only once. How do we change this behaviour? (UC#1)

Besides asking customers directly through surveys on their different areas of interest to identify causes and reasons for behaviour, we can also perform regression analysis using variables of interest to identify influential characteristics, on who stays in GHA more than once. It is important to gather variables that might not be trivial in the beginning but turn up to explain the variation for example – difference in time between registration for Discovery and first stay (maybe those who sign up on arrival are less likely to stay again).



*Illustrative overview of the data science algorithm employed for Candidate 2c*

Data requirements: A key factor for gaining insight to this question, as well as those above, is to gather data that can potentially have an effect, besides the obvious metadata on customers. For this task, only data up to first stay is needed, as well as an indicator if a customer then stayed more than once or not. Other parameters could be if the guests have created a profile prior to stay, updated interests immediately after stay etc.

Implementation complexity: 2/5

Metrics: tbd

Algorithms and models: Same as candidate 2a

Beside finding desired drivers behind the investigated behaviour, such as cross-brand stays and tier move (number of stays per year), we should focus on building tools to predict future stays (which can lead to client lifetime value) as well as probabilities to switch tiers.

# Appendix

## Use Case long list

This list of use cases (in table below) was reviewed and prioritized during the 2-day London workshop.

|  |  |  |
| --- | --- | --- |
| **Priority** | **UC#** | **Use Case** |
| 1 | 1 | 80% of members stay only one stay - how do we move these to stay another night? |
| 1 | 2 | Client lifetime value (CLV) how do we get a model to predict this? Can we use historical data and historical outcome to estimate this - what is the accuracy of our model? |
| 1 | 3 | Codes for stay channel - need golden source OTA vs. non-OTA (how good are we at turning OTA guest into direct bookers |
| 1 | 4 | Fast movers: who is accelerating |
| 1 | 5 | GHA has high acquisition and high churn (how can we illustrate the hot-spots) and how can we prevent the churn, and maintain the acquisition |
| 1 | 6 | How does local experience redemption affect cross brand stays |
| 1 | 7 | How does tier affect behavior - Do your frequency change/avg. room rate change |
| 1 | 8 | How to get the rights segments for “Formula One” and "Ultra travel Collection" ... together clusters on hotel categorization |
| 1 | 9 | How to increase year to year black membership retention (only 35% re-won black tier from 2015 to 2016) |
| 1 | 10 | Does Local Experiences redemption affect future behavior compared to non-redeemers |
| 1 | 11 | What are the cross-brand characteristic - how do cross brand guest differ from intra-brand |
| 2 | 12 | “Surprise and Delight” (how does this affect guest performance) - this is a non-published perk given to some when they are on the verge of hitting a next tier |
| 2 | 13 | After upgrade motivation —> then drops off? |
| 2 | 14 | Are there any certain behaviors at downgrade? Do guests change their stay pattern? What patterns are typical for a tier-downgraded member, and what measures can we use to counter this? |
| 2 | 15 | Change in profile —> what does it mean? – That a member changes something on their profile should GHA contact them in some way |
| 2 | 16 | Close to next tier behavior - can we keep the momentum? |
| 2 | 17 | High Status “danger list” abrupt change is high status behavior |
| 2 | 18 | How effective is the “end of October” challenge? |
| 2 | 19 | how is a status match vs. earned status behavior |
| 2 | 20 | How to maximize cross brand sales…. does location play a role? Number of hotels in a given location |
| 2 | 21 | Intra-Brand - from single hotel to more within brand - if we can prove that is due to GHA activity |
| 2 | 22 | What is the “right” amount of partner hotels in a city (measure cross-brand stays) |
| 2 | 23 | What is the “right” amount of partner hotels in a city (measure cross-brand stays) - use the size of the city as parameter |
| 3 | 24 | “Weather Data” at location and at destination to trigger email offers based on 1) weather 2) preferences. |
| 3 | 25 | 3 day in advance of weekend good weather in “Paris” |
| 3 | 26 | How not to become dormant |
| ignore | 27 | What would membership tier levels look like if everything was rolling membership dates - compare this to generic tier stays - is it worthwhile changing this? |
| x | 28 | Average days between stays (and trend) - stable - up – down |
| x | 29 | Find cluster, without knowing the features… |
| x | 30 | Find clusters, and what are the main driving features |
| x | 31 | Look at clusters at amount of nights, revenue, stays - what does a shift in threshold do to behavior |
| x | 32 | What is response rate and value of emails and campaigns per region, country, language, brand - make measure and actions scientific and structured. |
| x (fact) | 33 | Non-GHA booked booking not known prior to completed stay. |
| x | 34 | Local Experiences redeems with/without stay |

## Non-prioritized use cases

The following priority 1 use cases have been down-prioritized

* Codes for stay channel (UC#3) - need golden source OTA vs. non-OTA (how good are we at turning OTA guest into direct bookers:
* If the information is provided it is of course possible to analyze how many second timers book directly instead of through OTA and what differentiates them.
* If GHA is actively trying to do something (like sending emails) to induce this behaviour, then the task is difficult if no control group have been left out. Without a control group it would be very difficult to correctly say whether any change is due to email or something else.
* How to increase year to year black membership retention (only 35% re-won black tier from 2015 to 2016) (UC#9)
* Most of the analysis needed for this use case is part of POC candidate 1: Client Lifetime Value and Churn which would indicate which clients are on the way out/downgrade. With this information, one can test different approach to customers, in order to keep them active and identify what works or not.
* A clustering/regression exercises could be tested on black tier members only to see if there is anything that can explain behaviour.
* Use Cases 4 & 8 have not been assessed as of now, as mentioned in the start of this section.

Through data science we can only provide tools, such as indications and explanations. If no actions are taken to impact customers and their behaviour, then there is not much use of the intelligence gathered.

2021.AI can introduce a scientific approach to testing the effect of actions taken (through A/B testing - for more information about A/B testing and how Netflix's uses this to provide a more personalized offering please see appendix - List of relevant articles) in order to gain insight into what “remedies” works better.

## Overview of data processing during the POCs

Working with data is a non-trivial exercise and involves a series of steps:

### 6.3.1. Get the data

Acquire data and load it to the integrated development environment (IDE) such as R Studio, Python Notebook or Jupyter Notebook that allows to use both R and Python as programming languages.

### 6.3.2. Explore the data

Do a simple exploration of data to make sure that there is nothing strange happening in the data. For example, if the data shows that at some particular date many more customers got registered than normally, then maybe there has been some kind of anomaly/error in how this data got recorded, which would need to be checked as it might in the end produce incorrect results.

### 6.2.3. Clean and prepare the data

Here one would investigate missing values and determine depending on variable types what to do, change to zero, remove that subject or assign an “unknown” category. One could also consider predicting what should have been there instead of missing values in order not to remove too many subjects.

Are there any negative values for measures that should only be positive?

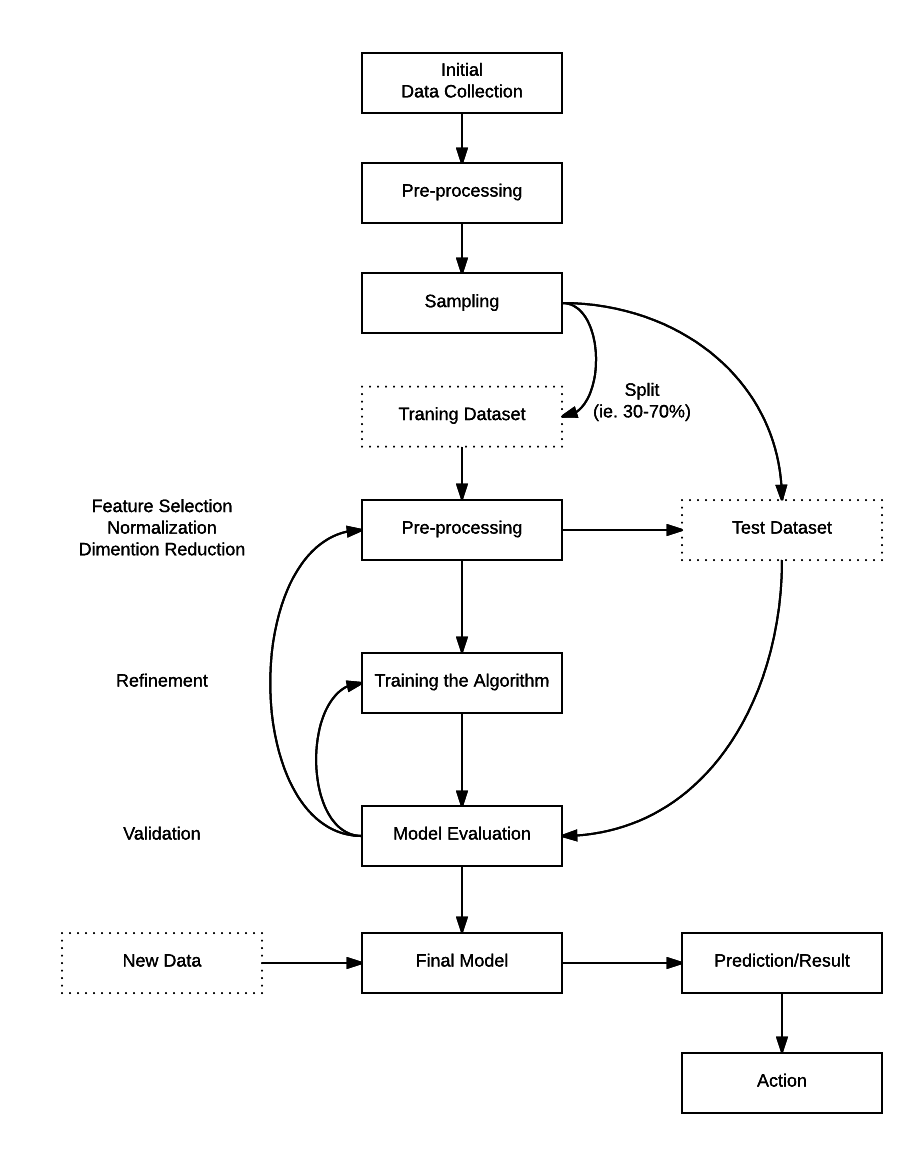
Are there outliers that could influence the results towards a less realistic?

It could also be that some variables contain strings with multiple input (like hotel stayed in “AT, GT, NY”) which would need to be separated to separate columns.

It might also be necessary to compute some variables based on others to potentially help explain the outcome variables (for example: average nights per stay, time between stays, time between booking and stay)

For some cases, it might be necessary to normalize some variables in order to make them comparable and avoid that one variable is more dominating the other just because it has a different range. This is done to minimize variance influence of variables of different scales in for example principal component analysis.

Transformation of variables could also be part of preparation. This is needed to fit data to a particular distribution, often to a normal distribution as many algorithms work under assumption that the variables follow a normal distribution.



*Overview of steps in data*

### 6.3.4. Structure data for the model

Before modelling one needs to make sure that the data is structured correctly for the type of model one will be using. For example, time series analysis would require timely data for each customer (multiple rows). While regression analysis or decision trees would just need one row for each with all the descriptive variables.

Now the modelling can begin.

### 6.3.5. Monitor and updating the model

There is not a single response on how models are kept up to date in order to perform best possible after implementation. First of all models are trained and tested before placed in production, which gives an idea of how well they perform. Secondly, after the model has been put in production, continuous model retraining over time ensures that new and more data is used for training which should result in model adoption if there are changes in customer behaviour.

Some models of the machine learning field can incorporate a response/feedback loop to provide a kind of check whether the recommendation/prediction was correct or wrong, and include this information in future estimations. This is similar to Netflix movie ratings, but instead of ratings, a feedback of opened email, clicked link or booked hotel can be used. Such a feedback loop, if not present yet can be introduced, to provide better model predictions in the future as more data is gathered.

It might also be needed to update models with new variables (if these were not available or existed before), or a better model if possible. Making changes to models and at what frequency should however be considered with care.

Some model predictions, such as client lifetime value are often used to measure performance over time. Any updates to such model will cause new predictions to be impossible to compare to previous ones, unless one also re-computes them with use of the updated model.

## Algorithms and models

Algorithms and models are data dependant and various models are tried and tested, to see which model gives the best predictions. Once a well working algorithm is found, the following steps are taken to increase precision of the predictions.

Please find below examples of algorithms we will use to address the selected user stories.

### 6.4.1. Client lifetime value

Survival analysis + std. supervised learning (cox, elastic net, tree methods: random forest, decision tree, boosted trees, support vector machine, neural nets).

### 6.4.2. Client behaviour

Regression analysis, this with several options depending on the variables and trade-off between complexity and understanding of behaviour. Both linear and nonlinear approaches will be used, each containing different models.

### 6.4.3. Client clustering

K-means, hierarchical clustering, random forest.

## Overview of relevant articles

### 6.5.1. The power of personalisation

The drive to personalise consumer offers is moving from targeted customers to include the wider family unit.

The Marriott approach illustrates that learning on the hoof is a must. This will make artificial intelligence (AI), with its learning capacity, important and brand consultancy Soul is looking at embedding AI into the customer experience of a loyalty program client.

It is ironic that where the race to automate everything seemed to obviate the human factor, technology is now being used to personalize and humanize service.

Source: [*https://www.raconteur.net/business/the-power-of-personalisation*](https://www.raconteur.net/business/the-power-of-personalisation)

### 6.4.2. Background info on IHG's analytics adventure

ROI discussion and IHG case described.

"Multi-channel analytics significantly reduces the guesswork from design, waste in process and time, and better enables our commercial teams to focus on the right things. We’ve learned, for example, through urgency message testing “Book now – just four rooms left at this price!” has little impact on people traveling tonight or tomorrow, but a major influence on those traveling a week from today. These insights allow us to understand our customers’ motivations like never before.

Generally speaking, we can see returns in the order of 600-800 percent for single tests in terms of ROI. And we can tell we’re doing things right when our annual results reports are full of digital success stories about the outcomes of our campaigns."

Source: *https://blog.ebiquity.com/2015/04/how-multi-channel-analytics-helps-ihg-understand-its-customers*

### 6.4.3. The promise of AI as a social and business growth tool

Everyday more companies are adopting or planning to adopt artificial intelligence as a growth tool in different industries. Counting on a wide range of applications, AI is set to be one the main business and social changing forces of the new digital age.

Source: *http://www.cio.com/article/3187084/it-industry/the-promise-of-artificial-intelligence-as-a-social-and-business-growth-tool.html?nsdr=true*

### 6.4.4. The Netflix Experimentation Platform: It’s All A/Bout Testing

Ever wonder how Netflix serves a great streaming experience with high-quality video and minimal playback interruptions? Thank the team of engineers and data scientists who constantly A/B test their innovations to our adaptive streaming and content delivery network algorithms. What about more obvious changes, such as the complete redesign of our UI layout or our new personalized homepage? Yes, all thoroughly A/B tested.

Source: *http://techblog.netflix.com/2016/04/its-all-about-testing-netflix.html*