CSC8632: Executive Report

1 Executive Summary

Sage is currently facing concerns regarding ethical data management. As of now, data management decisions are made on a case-by-case basis, with no well-defined framework detailing the procedure to follow. This is negatively impacting our performance, as operations are inefficient, hard to keep track of, and risk suffering data leaks or giving undesired access to sensitive information. It is crucial to implement a solution to address these issues, as an increasing portion of our revenue comes from Cloud operations (which are very data-heavy) and more customers are trusting us with their valuable data. We must show to them that their trust is well-placed; we propose two solutions that will ensure this.

On one side, we must implement a Unique ID System, whereby data transactions are centralised and identified through a unique ID, meaning it is easy to back-track what data has been moved, where it has been moved from and to, and who were responsible for it. At the same time, we must carry out market research to try and find what exactly it is that people consider ethical. Of course, ethics are subjective and likely to vary between individuals in different countries and companies, so a wide variety of clients will have to be polled. This will allow us to create a new ethical rule-book which is to be followed when handling data to ensure we work consistently. To allow for further customisation, clients will be able to tweak their preferences to suit their needs.

While this may seem like an over-complicated issue, especially as it doesn't have a direct, measurable impact, we must remember that over a billion pounds a year in revenue is at risk if we fail to retain our customers' trust.

2 Context

The first step was to explore the problem statement. From our initial conversations with our problem owner, we used root-cause analysis to drill down to what we identified as the main issues behind our problem statement. The problem can be phrased as 'how does Sage, a company that offers cloud-based business solutions to small and medium sized businesses, establish a universal code of ethics for data science projects that could be applied business-wide to protect themselves from any potential damage caused by a miscalculation in what the public deem as acceptable data science practices?' Within this problem statement, we discovered that Sage must unify the ethical considerations of a wide range of stakeholders and actors, ranging from individual Sage employees to the wider public's ethics from multiple demographics in differing countries and business sizes. As well Sage has operations in multiple different countries which can complicate the ethical considerations further as ethics are rarely shared across borders. James (problem owner) highlighted to us how Sage's culture was 'risk adverse', meaning that they took a pragmatic approach to ethics, only doing something if they felt it was right; this means that conflicts can arise if different teams come to different conclusions. He gave the example of the wider cloud solutions industry selling customer data to third parties and how if Sage deemed this acceptable then they could become a disruptor in new markets.

3 Need

Initially the needs were identified based on general approaches often used for defining ethical boundaries, as due to the subjective nature of the problem it quickly became obvious that there is not right or wrong solution. As a group we needed to establish a middle ground among us regarding ethics that felt comfortable to operate in and provide the much-needed valuable solution for Sage. Based on the conversations during the meetings with the problem owner numerous needs were identified. First we must discover what would be deemed acceptable for a variety of age groups regarding data handling, and in general discovering the different ethical definitions among various demographics. Attempting to define all these different definitions is an exceptionally difficult task, so to overcome this problem we had discussions and came up with a framework that felt right for our age group.

Next, an important consideration emerges regarding what and how much of the customer's data should be shared both within the company and with external sources. Even if practices comply with GDPR and permission from the customer is gained, is it ethical for Sage to share sensitive information with third-party companies or vice versa? It is highly needed to develop an ethical/data-driven framework to set a 'rule book' that assists in highlighting the appropriate course of action. It is understandable that user

input will determine what action should be made for each case, although within the suggested framework the right course of action will be more easily distinguishable. Furthermore, this uncertainty of ethical decisions regarding projects and customer service is negatively impacting the internal communication and information sharing. These problems mainly emerge from not having a defined ethical boundary that employees share. At the same time misunderstanding what the customers deem acceptable. Nonetheless, difference in opinions is a key issue forcing employees to avoid taking risks due to the distorted "right" course of action, resulting to negatively affect projects. This situation is holding back the company and does not allow Sage to cater to new demographics expand the reach of their products while improving the company's reputation and safeguarding it from potential controversies. Our groups solution is based on these needs and considerations the initial customer journey map can be seen in Appendix A.

4 Vision

The next task ahead was to construct a plan and explain how the proposed solution can help to resolve the ultimate problem that Sage is facing. The first solution that was proposed has two sides. The first side was focused on creating an ethical boundary. For that we suggested that Sage should conduct market research in which they will gather the thoughts of the community regarding ethics and what they want from the companies that they deal with regarding ethics. Sage would subsequently be able to define a more generalised ethical boundary by considering the most common answers. Although ethics being a subjective topic means it is difficult to cover every aspect, this approach will allow Sage to deal with most scenarios.

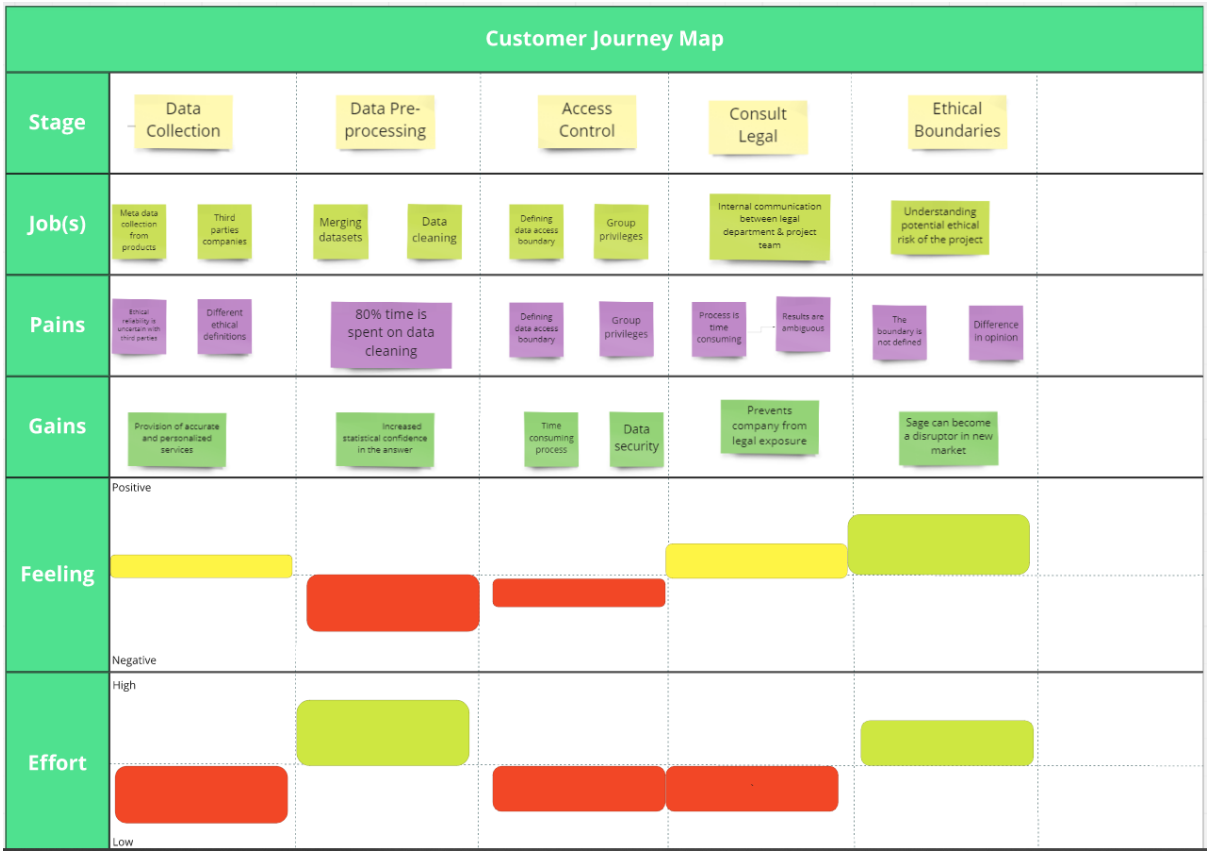
For the second side it was proposed that Sage should implement a preference control system in which users will be notified regarding the usage and sharing of their data and, if they are not comfortable with the information that is being shared or the company that it is being shared with, they can disable this action. The purpose of this is to give users control of their data in some aspects along with notifying them of practices applied to their data. Both sides of the solution are focusing on the customer's view of ethics. The next solution is for the organization itself; we propose that Sage should implement a unique ID system in which every customer will be assigned a unique identification or reference number. This number will be allocated before data pre-processing. Within the organization only authorized teams would be able to see all the details of the user while the rest of the members can only see the ID number along with the information they have rights to. For example: if the marketing team needs information regarding the customer's last purchase, they will have to request this information from the authorized team and they will share only the said data along with the user ID number. The main motive behind proposing this solution is to introduce data security which is the most essential ethical part. Appendix B contains the values that company would get for the proposed solutions and the business model. Appendix D contains a chart which shows the internal flow of the solution data-wise.

5 Outcomes

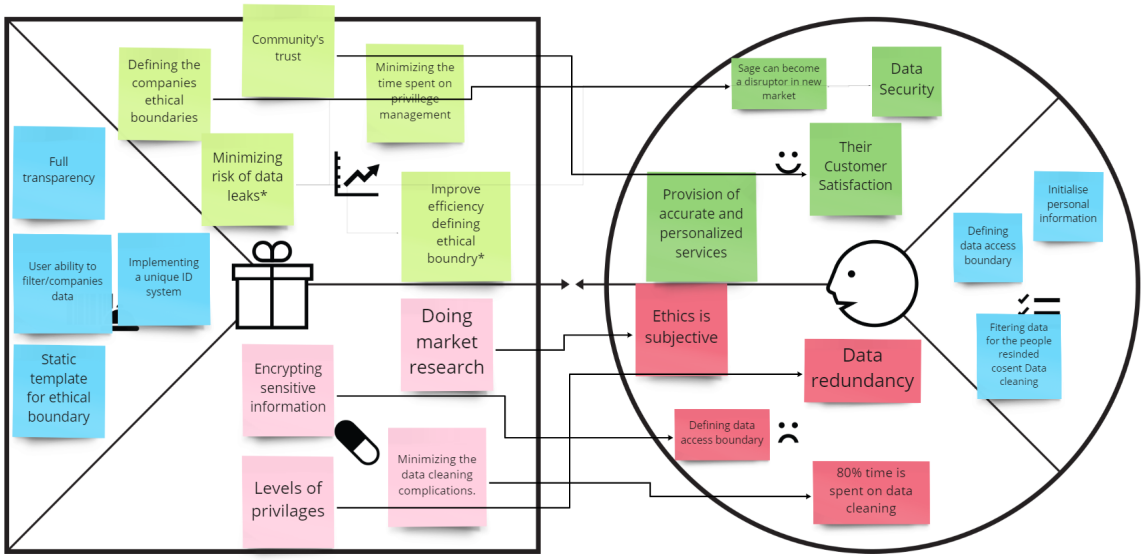
Before proceeding ahead, we have considered some key assumptions such as implementing ethical limitations which will assist Sage in becoming a fully trustworthy firm that wants its consumers to willingly share their data. These strongly outlined boundaries will reassure Sage's customers that their data is secure and being used in accordance with their wishes. The trust established during this process will be priceless and will contribute to Sage's image as an ethical company. To define the ethical boundaries, understanding customers is imperative. This issue can be tackled by doing market research on what the customers deem ethical and unethical, and will be undertaken by the marketing team within the institution in conjunction with external marketing agencies, and will take the form of polls, surveys and interviews. After the data is collected, data science techniques such as the Chi-squared test will be applied on it to compare results and look for commonalities. This will then help create a new consent form. The choice will be given to users to edit or completely remove their data and its uses. This will help provide a better, well-tailored service to customers. A brief overview of the expected time taken to develop a Minimum Viable Product is available in Appendix C.

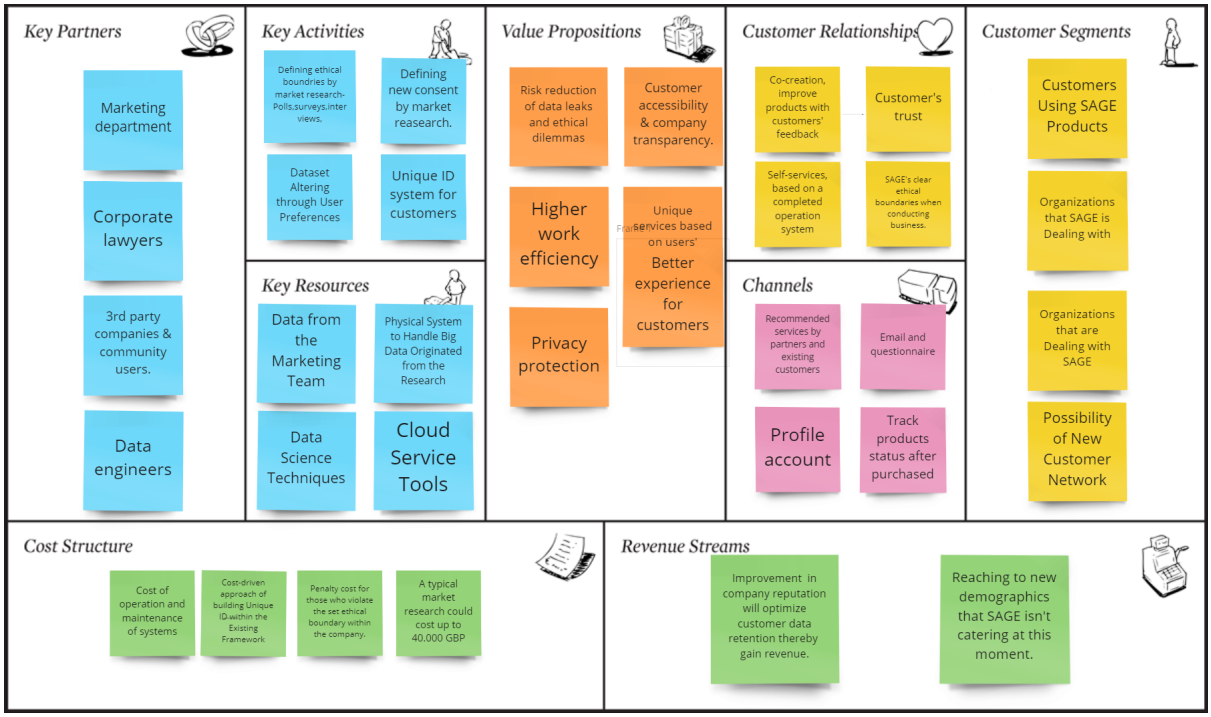
Appendix

A

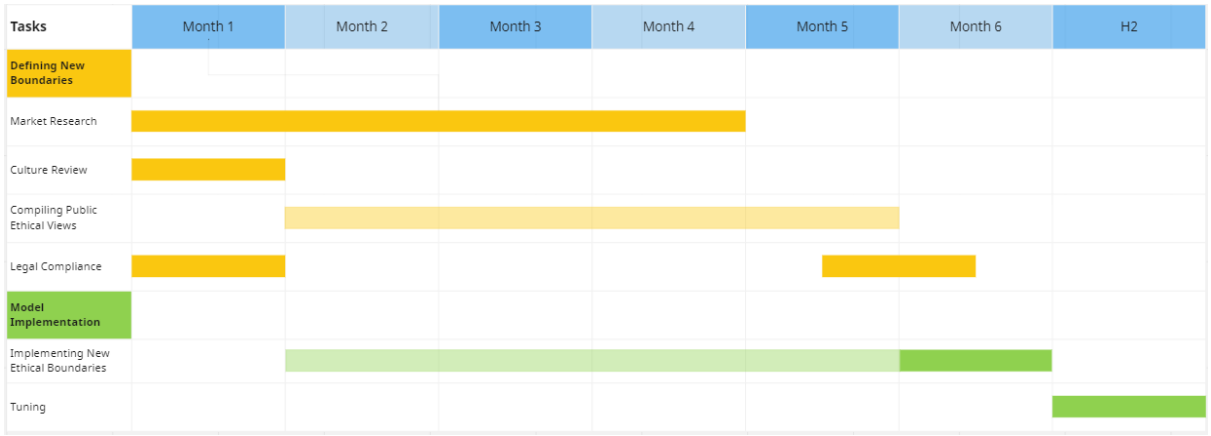

Customer Journey Map
Stage
Data Collection
Data Pre-processing
Access Control
Consult Legal
Ethical Boundaries
Job(s)
Meta data collection from products
Third parties companies
Merging datasets
Data cleaning
Defining data access boundary
Group privileges
Internal communication between legal department & project team
Understanding potential ethical risk of the project
Pains
Ethical reliability is uncertain with third parties
Different ethical definitions
80% time is spent on data cleaning
Defining data access boundary
Group privileges
Process is time consuming
Results are ambiguous
The boundary is not defined
Difference in opinion
Gains
Provision of accurate and personalized services
Increased statistical confidence in the answer
Time consuming process
Data security
Prevents company from legal exposure
Sage can become a disruptor in new market
Feeling
Positive
Negative
Effort
High
Low

B

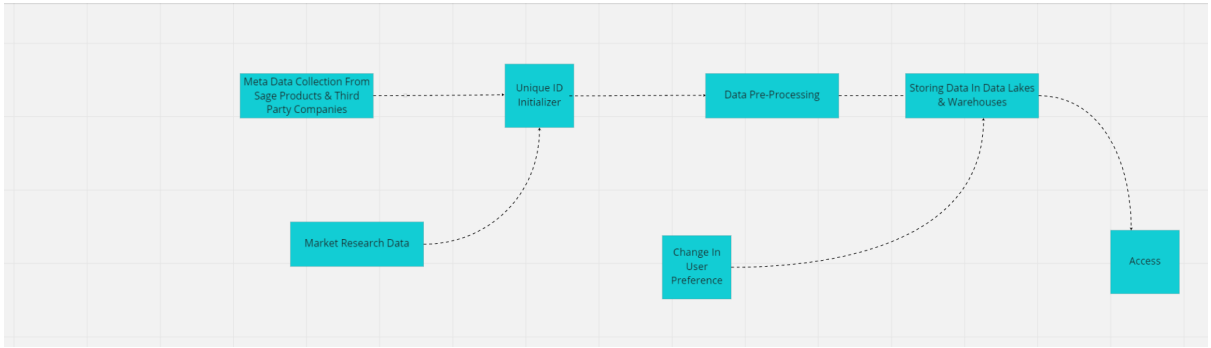

Community's trust
Defining the companies ethical boundaries
Minimizing the time spent on privillege management
Full transparency
Minimizing risk of data leaks*
Improve efficiency defining ethical boundry*
User ability to filter/companies data
Implementing a unique ID system
Static template for ethical boundary
Encrypting sensitive information
Doing market research
Levels of privilages
Minimizing the data cleaning complications.
Sage can become a disruptor in new market
Data Security
Their Customer Satisfaction
Provision of accurate and personalized services
Initialise personal information
Defining data access boundary
Ethics is subjective
Fitering data for the people resinded cosent Data cleaning
Data redundancy
Defining data access boundary
80% time is spent on data cleaning


Key Partners
Marketing department
Corporate lawyers
3rd party companies & community users.
Data engineers
Key Activities
Defining ethical boundries by market research- Polls,surveys,interviews.
Defining new consent by market reasearch.
Dataset Altering through User Preferences
Unique ID system for customers
Key Resources
Data from the Marketing Team
Physical System to Handle Big Data Originated from the Research
Data Science Techniques
Cloud Service Tools
Value Propositions
Risk reduction of data leaks and ethical dilemmas
Customer accessibility & company transparency.
Higher work efficiency
Unique services based on users'
Better experience for customers
Privacy protection
Customer Relationships
Co-creation, improve products with customers' feedback
Customer's trust
Self-services, based on a completed operation system
SAGE's clear ethical boundaries when conducting business.
Channels
Recommended services by partners and existing customers
Email and questionnaire
Profile account
Track products status after purchased
Customer Segments
Customers Using SAGE Products
Organizations that SAGE is Dealing with
Organizations that are Dealing with SAGE
Possibility of New Customer Network
Cost Structure
Cost of operation and maintenance of systems
Cost-driven approach of building Unique ID within the Existing Framework
Penalty cost for those who violate the set ethical boundary within the company.
A typical market research could cost up to 40.000 GBP
Revenue Streams
Improvement in company reputation will optimize customer data retention thereby gain revenue.
Reaching to new demographics that SAGE isn't catering at this moment.

C


Tasks
Month 1
Month 2
Month 3
Month 4
Month 5
Month 6
H2
Defining New Boundaries
Market Research
Culture Review
Compiling Public Ethical Views
Legal Compliance
Model Implementation
Implementing New Ethical Boundaries
Tuning

D


Meta Data Collection From Sage Products & Third Party Companies
Unique ID Initializer
Data Pre-Processing
Storing Data In Data Lakes & Warehouses
Market Research Data
Change In User Preference
Access