DSBL Capstone

# Step 0 - Introduction. 100-day Data Science Plan: Build a Data Science Strategy

Upon assuming a new leadership role within a company (whether from an internal move or joining the company anew), it is common for an executive to be asked to prepare a plan for their first 100 days in the job.

As part of this project, you will build/create the following:

1. Identification of six data science opportunities for the organization
   1. Opportunities must be spread across three different functional areas
   2. Detail the risks, challenges, and key factors for success for each of these opportunities
2. Prepare a roadmap for executing these six data science opportunities.
   1. Rack and stack evaluation of these opportunities
3. Prepare a Human Capital plan for your data science organization
4. Prepare a Technical plan for your data science organization
   1. Data and Data Architecture Strategy
   2. Machine Learning Architecture

The work product for this Capstone project will be a detailed presentation to the CEO, detailing your plan and the rationale behind your decisions.

This project asks you to prepare that 100-day data science plan for a company of your choosing; this could be your current company or some other existing company.

**Name of Company Chosen:** [Saudi Broadcasting Authority]

**Brief Company Description:** [The Saudi Broadcasting Authority (SBA) is a regulatory authority established back in 1962. It is a media organization headquartered in Riyadh, Saudi Arabia. It supervises the country's state-run television, radio, and digital media outlets. The goal of the SBA is to promote local cultural values as well as to support economic development initiatives in the Kingdom. In addition to its broadcasting activities, SBA also focuses on developing digital content and expanding its coverage to serve the many diverse communities within the Kingdom.]

# Step 1 - Identify Data Science Opportunities in the Business

Throughout the course, you have been exposed to multiple examples of data science projects implemented in a business setting. Now, based on your knowledge of your specific business context, you will generate six potential projects to be considered by the executive leadership team. These projects must span three unique functional areas of the business, with any one functional area representing no more than 3 projects:

Acceptable Project Mixes

\* 2 marketing + 2 supply chain + 2 finance

\* 2 marketing + 1 human resources + 1 procurement + 1 product + 1 manufacturing

\* 3 finance + 1 legal + 2 marketing

Unacceptable Project Mixes:

\* 3 marketing + 3 finance

\* 4 marketing + 1 product + 1 manufacturing

**Please identify your six projects here:**

**Project 1:** [Social Media Management]

**Project 2:** [Media Entertainment Services]

**Project 3:** [Strengthen the Digital Advantage Program]

**Project 4:** [Media Asset Management System]

**Project 5:** [Traffic and Scheduling System]

**Project 6:** [Preservation and Restoration Services]

**Note: You may choose to represent this information on slide 5 of the CEO Presentation Template**

**For each candidate project, please provide the following detail:**

**Project 1 Name:** [Social Media Management]

**Business Functional Area:** [Sales and Marketing]

**1. Description of the project (including business problem to be addressed, how data science will address that business problem, and the targeted business objective (revenue? customer acquisition? cost reduction?):**

- Business Problem Addressed: [The commission faces the challenge of delivering high quality audiovisual content to a population with shifting demands for content. As the population becomes more connected, the commission needs to be able to deliver more relevant and engaging audiovisual content to keep up with the trends and remain competitive.]

- Role of data science in addressing the business problem: [Help to identify target markets and potential customers for audiovisual investment, analyze customer data to discover potential trends or insights, optimize marketing strategies to maximize profits, and analyze financial performance to identify areas of improvement.]

- Targeted business objective(s): [Increase awareness of the company in the audiovisual space, increase market share gained from audiovisual investment]

**2. Data Science Classification**

- Approach: [Descriptive]

- Type of Model: [Unsupervised ML, clustering]

**3. Data needed for project and sources for that data**

* [Market research data]
* [Media data]
* [Audience Consumption data]

[The source of data will be the audiovisual content data stored in the commission and the market research data]

**4. Magnitude of opportunity (with justification)**

[The magnitude of opportunity is potentially high. The country is transitioning to a more digital economy and has seen an increased demand for audiovisual content due to the increased use of mobile phones and streaming services.]

**5. Cost and complexity of development and implementation**

[A medium cost with medium complexity is suitable for projects that require features such as content creation, scheduling, and analytics. These projects require a moderate level of technical expertise, a moderate budget for technology, and a moderate amount of time for development and implementation.]

**6. Likelihood of value capture (Low/Medium/High) with justification**

[High. Saudi Arabia is investing heavily in the audiovisual industry, with the aim of creating global distribution and attract inward investment. The Government has identified the sector as a priority for economic diversification and has committed to increasing spending on audiovisual content and production capacity. Therefore, there is likely to be strong value capture for all stakeholders involved in the Saudi audiovisual industry.]

**7. Key Business Stakeholders**

* [CMO]
* [Marketing]
* [Sales]
* [Customer service]
* [Product development]
* [IT teams]

**For each candidate project, please provide the following detail:**

**Project 2 Name:** [Media Entertainment Services]

**Business Functional Area:** [Marketing and production]

**1. Description of the project (including business problem to be addressed, how data science will address that business problem, and the targeted business objective (revenue? customer acquisition? cost reduction?):**

- Business Problem Addressed: [Providing a platform to make it easier and more efficient for customers to request rental of photographic, broadcasting and space equipment. This will make it simpler for customers to access the necessary equipment they need quickly and conveniently, with lower costs involved in the process.]

- Role of data science in addressing the business problem: [Data science can be invaluable in helping to address the business problem of providing a platform for requesting the rental of photographic, broadcasting and space equipment. By gathering data from user feedback and sales data, data science can be used to create predictive models and forecasts to better understand customer needs and preferences.]

- Targeted business objective(s): [Invest, increase sales and revenues]

**2. Data Science Classification**

- Approach: [Prescriptive]

- Type of Model: [Supervised learning]

**3. Data needed for project and sources for that data**

* [Customer needs and preferences.]
* [Rental agencies.]
* [Manufacturers.]

**4. Magnitude of opportunity (with justification)**

[Medium, having a platform for easily requesting rental of such equipment could help to streamline the process and make it easier for businesses to get the tools they need. This will help reduce costs associated with renting, as well as potentially increase economic opportunities for businesses in the space.]

**5. Cost and complexity of development and implementation**

[With medium cost and high complexity of this project are since it involves the rental of specialized equipment. This equipment typically requires a certain level of technical expertise, such as knowledge of camera, audio and video production, and space management. Additionally, this project will require the development of a secure platform to facilitate the rental process, as well as the integration of payment and shipping features.]

**6. Likelihood of value capture (Low/Medium/High) with justification**

[High, Renting out equipment is a lucrative business, especially within the broadcasting sector. The platform will provide customers a fast, efficient and secure way to access these services, prompting higher demand and value capture.]

**7. Key Business Stakeholders**

* [CFO]
* [Sales]
* [Finance]
* [Product managers]
* [Marketing]

**For each candidate project, please provide the following detail:**

**Project 3 Name:** [Strengthen the Digital Advantage Program]

**Business Functional Area:** [Human Resources]

**1. Description of the project (including business problem to be addressed, how data science will address that business problem, and the targeted business objective (revenue? customer acquisition? cost reduction?):**

- Business Problem Addressed: [To increase the efficiency of human resources departments by leveraging digital solutions and technologies. This includes optimizing the recruitment process, streamlining onboarding and training, leveraging modern analytics to help make data-driven decisions, and technology-enabling existing HR processes.]

- Role of data science in addressing the business problem: [Data science can be used to strengthen the digital advantage in Human Resources by empowering HR managers with analytics and predictive insights. The insights generated can help streamline operations, identify areas for improvement, visualize data trends, provide actionable insights for better decision making, and uncover hidden patterns and correlations in the data that may have not been previously discovered. Data science can also be used to drive better outcomes in recruitment, talent acquisition, training and development, employee engagement, and performance measurement.]

- Targeted business objective(s): [Develop and implement a comprehensive digitization strategy, improve the existing human resources digital process, procedures and operations, improve recruitment, onboarding, and talent management processes through automation and implement effective data and analytics tools for analyzing employee performance.]

**2. Data Science Classification**

- Approach: [Descriptive]

- Type of Model: [Unsupervised]

**3. Data needed for project and sources for that data**

* [Employee demographics.]
* [Training and development metrics.]
* [Talent retention data.]
* [Technology landscape.]

**4. Magnitude of opportunity (with justification)**

[High to medium, as the opportunity to strengthen the digital advantage through Human Resources is very large. By improving Human Resources processes, an organization can ensure that its talent pool is able to efficiently and effectively recruit and retain top-tier employees who have the skills and experience necessary to effectively and efficiently utilize digital technology in their roles. This potential can lead to increased productivity, cost savings, improved customer service and satisfaction, as well as enhanced market competitiveness.]

**5. Cost and complexity of development and implementation**

[The project can be implemented at a low cost and low complexity for several reasons. First, most of the work for the program can be done digitally, which eliminates the need for physical resources and allows for greater efficiency in implementation. Additionally, the program can be tailored to each individual user, which not only reduces complexity but also makes the program more effective at meeting its objectives. By utilizing digital tools and resources, the program can be implemented quickly and easily with minimal effort and cost.]

**6. Likelihood of value capture (Low/Medium/High) with justification**

[Medium, the potential is to capture significant value for the organization by improving their human resource capabilities. By strengthening their digital advantage, the Authority will be better equipped to efficiently and effectively recruit, manage, and reward its staff. This in turn could lead to improved employee engagement and productivity, helping them to remain competitive in the global environment.]

**7. Key Business Stakeholders**

* [CHRO]
* [Senior Management]
* [Human Resources Department]
* [IT Department]
* [Digital Content Department]
* [Training and Development Department]

**For each candidate project, please provide the following detail:**

**Project 4 Name:** [Media Asset Management System]

**Business Functional Area:** [Product]

**1. Description of the project (including business problem to be addressed, how data science will address that business problem, and the targeted business objective (revenue? customer acquisition? cost reduction?):**

- Business Problem Addressed: [The need to effectively manage and organize media assets such as videos, images, animations, and sound files. This project will help the organization to efficiently store, access, and utilize all of its media libraries, allowing for an improved utilization of resources and an improved delivery of content.]

- Role of data science in addressing the business problem: [Data science can help the organization to efficiently store, manage and use digital media assets such as images, audio and video materials. This can be done by categorizing digital media into large clusters or categories according to their features, using techniques such as clustering and classification algorithms. Such categorization can be used for automated retrieval of media from the organization’s vast library. Moreover, data science can also help in dealing with problems such as data compression for efficient storage and searchability of digital media records.]

- Targeted business objective(s): [To have a centralized repository of digital content and use the digital assets more effectively and efficiently.]

**2. Data Science Classification**

- Approach: [Prescriptive]

- Type of Model: [Supervised]

**3. Data needed for project and sources for that data**

[Customer service records]

[Organizational databases and servers]

[Media assets and associated metadata]

**4. Magnitude of opportunity (with justification)**

[High, the need for efficient and organized media asset management in the Saudi broadcasting authority due to the huge amount of content produced and broadcasted by them. A project that can help manage the content effectively, ensuring that only the authorized people have access to it and archiving it, can help the organization streamline its operations, allow for faster broadcasting, and content accessibility, as well as reduce costs associated with re-locating information. Additionally, with broadcasting standards constantly increasing, a system like this will enable the organization to meet and exceed these standards.]

**5. Cost and complexity of development and implementation**

[The cost and complexity of developing and implementing the project are high because it requires a comprehensive set of technological solutions to manage both physical and digital media. The system must be able to store, organize, and deliver data, such as images, video, audio, and text, as well as provide users with the ability to search, access, and share these assets. To successfully develop and implement, a team of experts must be assembled with expertise in various areas such as systems architecture, software engineering, media operations, database design and management and security protocols.]

**6. Likelihood of value capture (Low/Medium/High) with justification**

[High, Media Asset Management provides organizations with the ability to organize and store media-related assets quickly and conveniently. This allows businesses to receive higher returns from their investments by allowing them to efficiently access its stored media content, which can then be used to create products and increase visibility. Additionally, it helps in reducing costs associated with accessing, managing and delivering content, while improving collaboration across departments]

**7. Key Business Stakeholders**

* [CTO]
* [Production]
* [Product]
* [NOC]
* [IT team]
* [Operations]
* [Legal]
* [Advertising and marketing]

**For each candidate project, please provide the following detail:**

**Project 5 Name:** [Traffic and Scheduling System]

**Business Functional Area:** [Product]

**1. Description of the project (including business problem to be addressed, how data science will address that business problem, and the targeted business objective (revenue? customer acquisition? cost reduction?):**

- Business Problem Addressed: [The Traffic and Scheduling System project is to provide an effective solution to manage and monitor TV and radio broadcast schedules, allowing users to monitor, schedule, and control broadcast traffic efficiently and cost-effectively.]

- Role of data science in addressing the business problem: [By providing advanced analytics and insights on traffic patterns, scheduling strategies, and various other parameters. Data science techniques such as machine learning algorithms can be used to analyze large amounts of data to identify trends and correlations that can be leveraged in making decisions related to traffic and scheduling. In addition, data science techniques can be used to develop predictive models that can help forecast demand and optimize scheduling strategies. The use of data science thus provides a platform for the organization to better understand traffic patterns, enabling them to make informed decisions and improve the efficiency of their operations.]

- Targeted business objective(s): [To develop a traffic and scheduling system to enable efficient and effective monitoring and management of tv and radio programs. The system should allow efficient broadcasting of the programs, ensure timely screenings of ads, enable the automate import of program schedules, and provide accurate data regarding audience reach and program popularity.]

**2. Data Science Classification**

- Approach: [Prescriptive]

- Type of Model: [Semi-Supervised]

**3. Data needed for project and sources for that data**

[Transit network data]

[Traffic speed data]

[Volume of audiovisual collection]

[History reports of operations]

[End-users of the organization department]

**4. Magnitude of opportunity (with justification)**

[High, Traffic and Scheduling System is significant due to the importance of the organization in terms of providing content for the public. The ability to increase the efficiency of their operations and manage content effectively could have a large impact on their overall success in the market, making it a very valuable investment. Additionally, by providing an easy-to-use system with features like scheduling and traffic-analytics, the organization can increase their audience engagement and viewership. This can lead to improved revenue for both the broadcasting authority itself, as well as for those advertisers and companies who wish to buy commercial space.]

**5. Cost and complexity of development and implementation**

[The cost and complexity of the Traffic and Scheduling System Project are high due to the need to develop a comprehensive system that is capable of managing complex media scheduling operations. The complexity arises from the need to integrate various components such as media scheduling, broadcast traffic management, programming and content management, advertising, and back-office operations into one comprehensive package. Additionally, there is a need to develop software solutions that can be used across multiple platforms and are capable of handling large amounts of data. Moreover, the system must be able to integrate with different systems such as CRM, MAM, and financial systems.]

**6. Likelihood of value capture (Low/Medium/High) with justification**

[High. The Traffic and Scheduling System has the potential to capture value for the organization by optimizing traffic flows, increasing efficiency, and reducing costs associated with scheduling of staff and creation of content. Furthermore, it is also likely to enhance customer experiences through improved reliability, satisfaction, and safety for staff.]

**7. Key Business Stakeholders**

* [CTO]
* [NOC]
* [IT team]
* [Operations]
* [Software development team]
* [Finance]
* [Marketing]

**For each candidate project, please provide the following detail:**

**Project 6 Name:** [Preservation and Restoration Services]

**Business Functional Area:** [Product]

**1. Description of the project (including business problem to be addressed, how data science will address that business problem, and the targeted business objective (revenue? customer acquisition? cost reduction?):**

- Business Problem Addressed: [By the preservation and Restoration services project is the need to preserve and restore valuable audio-visual materials such as TV programs, films, and music recordings for the organization, so as to ensure their long-term viability and stability.]

- Role of data science in addressing the business problem: [To help preserve and restore films by extracting, analyzing, and archiving important data from them. Data analysis will allow for the extraction of a clear understanding of the film’s elements, including the film’s content, themes, music, structure, and more. This data can then be used to identify potential ongoing restoration and preservation needs for a given film. It can also help planners identify important films for preservation and recommend potential partners for their restoration. Additionally, data science can be used to classify films into genres and styles for further indexing and categorization.]

- Targeted business objective(s): [To provide high-quality preservation and restoration services to broadcast content producers in Saudi Arabia in order to ensure the preservation and longevity of key cultural content.]

**2. Data Science Classification**

- Approach: [Prescriptive]

- Type of Model: [Semi-Supervised]

**3. Data needed for project and sources for that data**

* [Historical images or records]
* [Technical specifications of the materials]
* [Environmental impact assessment data]
* [media databases]
* [Market research data]

**4. Magnitude of opportunity (with justification)**

[Very high, Saudi Arabia is rich with important historical sites that entice visitors from around the world. By engaging in preservation and restoration projects, the organization would be able to conserve the unique cultural heritage of the region more effectively, which thus making the area a more attractive place to visit. Additionally, engaging in such projects would also provide educational opportunities for people in the region and beyond, as it would allow future generations to learn about and appreciate the history of the region. Such projects could also provide an economic boost to the tourism industry in the region.]

**5. Cost and complexity of development and implementation**

[The cost and complexity of developing and implementing preservation and restoration services are very high due to several factors. First, the process of preserving and restoring materials requires specialized knowledge and expertise. Depending on the type of material being preserved or restored, this may require the use of specialized equipment, materials, and techniques. Additionally, many of these materials may be fragile or delicate in nature, making their preservation and restoration more challenging and potentially costly. Furthermore, the time it takes to complete a restoration project can be extensive, requiring both intensive labor and attention to detail, further increasing the cost.]

**6. Likelihood of value capture (Low/Medium/High) with justification**

[High, The value capture here is likely to be high as the service will help to preserve and maintain the traditional culture and heritage, which is of great importance to all people living in Saudi Arabia. With the right strategy and correct implementation, this service aims to give back a sense of pride and cultural respect to all Saudis. Additionally, this service also provides an opportunity to generate incomes for local businesses, as it can increase tourism within the country by allowing people access to restored historical areas.]

**7. Key Business Stakeholders**

* [COO]
* [Production]
* [Quality assurance]
* [NOC]
* [IT team]
* [Legal]

# Step 2 - Developing a Roadmap: Prioritizing Data Science Opportunities in the Business

A strategic approach to data science requires the business to consider the relative opportunities, costs, and risks of potential projects to identify the best order to carry out the projects. What should be tackled first? What is best pushed off until later? Completing the Data Science Roadmap requires stepping through key considerations to determine which project(s) should be considered ‘top priority’ and at what pace these and subsequent projects should be initiated.

**1. Complete this “Rack and Stack Exercise” worksheet to determine the relative strategic alignment, cost, complexity of implementation, certainty of value capture, and magnitude of benefit for each of the six projects**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Direct Alignment with Strategic Goals?** | **Cost** | **Complexity of Implementation** | **Certainty of Value Capture** | **Magnitude of Benefit** |
|  | 1=Low; 5=High | 1=High; 5=Low | 1=High; 5=Low | 1=Low; 5=High | 1=Small; 5=Large |
| **Project 1:**  **[Social Media Management]** | 5 | 3 | 3 | 5 | 5 |
| **Project 2:**  **[Media Entertainment Services]** | 4 | 3 | 5 | 5 | 3 |
| **Project 3:**  **[Strengthen the Digital Advantage Program]** | 4 | 4 | 5 | 3 | 4 |
| **Project 4:**  **[Media Asset Management System]** | 5 | 2 | 1 | 5 | 5 |
| **Project 5:**  **[Traffic and Scheduling System]** | 5 | 2 | 1 | 5 | 5 |
| **Project 6:**  **[Preservation and Restoration Services]** | 4 | 1 | 1 | 5 | 5 |

**Note: You may choose to represent this information on slide 8 of the CEO Presentation Template**

**Please complete Step 2, Part 2, the Data Science Opportunity Matrix, using slide 1 of the CEO Presentation Template (You may or may not decide to include this slide as part of your CEO presentation)**

**Step 2, Part 3: Complete the table below by referencing the first four data science projects chosen for implementation. Include your justification for each project's order of implementation (e.g., how will the third project benefit from being implemented after the completion of the first two projects?)**

|  |  |  |
| --- | --- | --- |
| **Project Order** | **Project Title** | **Order Justification** |
| 1 | Strengthen the Digital Advantage Program | The project has a medium business value impact, high feasibility, and medium likelihood of value capture. Yet, the results will support Improving digital infrastructure and capabilities is essential for businesses to remain competitive and successful. The Digital Advantage Program provides employees with the necessary tools and will also develop the momentum to build enthusiasm and support for data science from the top management, business units and other stakeholders and generate profits that could be used to invest in other data science projects. |
| 2 | Media Entertainment Services | The Media Entertainment Services is a strategical and highly profitable investment, increasing its sales in alignment with the organization’s targets will be creating a high business value impact. The expectation to deliver is high in a short time with low cost. This will develop the momentum to build enthusiasm and support for data science from the top management, business units and other stakeholders and generate profits that could be used to invest in other data science projects. |
| 3 | Social Media Management | It will help to improve customer engagement and loyalty, increase brand visibility, drive revenue growth and the high expectation to increase sales quickly. |
| 4 | Traffic and Scheduling System | High business value impact with a more advanced and sophisticated implementation for a better understanding to our customers to create better products and enhance customer experience. Low feasibility, and high likelihood of value capture with better understanding to more focused groups. |

**Note: You may choose to represent this information on slides 6 and 7 of the CEO Presentation Template**

# Step 3 - Establishing a Data Science Human Capital Strategy for your Data-driven Business

Now that we have established a roadmap for carrying out data science projects, our attention must turn to building and configuring the organization we will leverage to carry out this roadmap. The Data Science Human Capital Plan completed in this step will cover the organizational structure and talent configuration best suited to carry out the business’s roadmap, as well as the activities that the organization in particular -- and business more broadly -- must complete in order to promote a data-driven culture throughout the business.

**1. Identify the organizational model best suited for the data science organization that your business will need to deliver on the roadmap completed in Step 2. Provide justification for your selection based on the needs, scope, and timing of projects to be implemented in the Data Science Roadmap. If your organization should start with one model and evolve toward a different model, you may provide that detail and justification in your response.**

**Organizational Model:** [Agile, Centralized.]

**Justification:**

[As a centralized model allows for more efficient decision-making and ensures greater control over the data. For the organization, it may be beneficial to have a centralized organizational model, as this will allow them to more effectively manage the large amounts of data it produces and receives. This would enable them to more quickly and accurately assess data trends and customer preferences, as well as make better informed decisions when it comes to budget allocation and programming decisions.]

**2. Complete the “Human Capital Plan” Worksheet for your data science organization.**

**- Identify the first ten professional roles for which you would recruit. How would you organize these roles into teams within the organization?**

For example, if you had 4 data scientists split evenly into two teams, your response would look like this:

|  |  |  |
| --- | --- | --- |
|  | **Position** | **Team** |
| 1 | Data Scientist | 1 |
| 2 | Data Scientist | 1 |
| 3 | Data Scientist | 2 |
| 4 | Data Scientist | 2 |

Identify your roles and teams below:

|  |  |  |
| --- | --- | --- |
|  | **Position** | **Team** |
| 1 | [Data Science] | 2 |
| 2 | [Data Science] | 1 |
| 3 | [Data Science] | 1 |
| 4 | [Data Science] | 1 |
| 5 | [Machine Learning Engineer] | 2 |
| 6 | [Machine Learning Engineer] | 2 |
| 7 | [Data Analyst] | 1 |
| 8 | [Data Analyst] | 2 |
| 9 | [Data Engineer] | 3 |
| 10 | [Data Visualization Engineer] | 4 |

**Note: You may choose to represent this information on slide 9 of the CEO Presentation Template**

**Assume that leadership will allocate four new FTE’s for your data science organization during the current fiscal year. How would you prioritize your organizational buildout?**

|  |  |  |
| --- | --- | --- |
| **Order of Hire** | **Position** | **Justification** |
| 1 | Data Engineer | Ensure the availability and access of data to the organization and its customers or stakeholders. |
| 2 | Data Scientist | To use data science techniques such as predictive analytics, machine learning, natural language processing (NLP), and sentiment analysis to extract meaningful insights from data sets. |
| 3 | Machine Learning Engineer | Design, develop and maintain systems for analyzing and extracting training data for machine learning algorithms that can then be used to make better decisions. |
| 4 | Data Analyst | Analyze large datasets to uncover patterns and trends that can inform business decisions across various departments. |

**Craft a “Data-Driven Transformation Strategy” by identifying six specific initiatives that you would recommend the data science organization and/or the business undertake in order to promote a data-driven culture across the business.**

|  |  |
| --- | --- |
|  | **Strategy** |
| 1 | Establishing team-wide KPIs that identify how well the data-driven strategy is helping the business reach its goals will help monitor progress and measure success. |
| 2 | Automate the most frequently used tools, processes, and analytics across the data science organization and the business, to create an efficient operating environment that allows for more time to be spent on data-driven activities. |
| 3 | Develop a data-driven approach for all decision making, including establishing policies for data governance, access, sharing and protection. |
| 4 | Establish a centralized data hub to enable the aggregation, organization, and analysis of data across multiple departments and sources. |
| 5 | Establish a clear and precise vision for data-driven transformation among executive leadership and business stakeholders, in order to generate enthusiasm and support. |
| 6 | Invest in a comprehensive training program to ensure that all staff members are well-versed in data science concepts and techniques. |

**Note: You may choose to represent this information on slide 10 of the CEO Presentation Template**

# Step 4 - Establishing the Technical Infrastructure to Support the Data Science Organization

With a completed Data Science Roadmap and a Human Capital Plan for executing the data science strategy, we turn our attention to the technological capabilities that must be built to support the new Data Science organization.

Complete the table on the next page by entering strategic aspects your business might consider to meet its Data and Data Architecture needs.

**Data and Data Architecture Strategy for the business**

|  |  |  |
| --- | --- | --- |
| **Component** | | **Strategy** |
| Data Requirements | What data should be included in the Data Strategy? | * Data management guidelines, policies and procedures. * Data governance framework that defines roles and responsibilities. * Data quality assurance program. |
| Data Governance | How will we promote data availability? (provide at least two ideas) | * Educate both IT departments and business decision makers on the importance of data governance. * Establishing clear and simple guidelines and procedures outlining |
| How will we promote usability? (provide at least two ideas) | * Use the latest front-end technology. * Creating incentives for citizens to use the data governance platform. |
| How will we guarantee integrity? (provide at least two ideas) | * Developing data quality. * Establishing and enforcing strict access policies and data security measures * Assigning roles and responsibilities to personnel who manage the data. |
| How will we guarantee security? (provide at least two ideas) | * Implement technical and administrative controls. * Monitor data access activities. * Implement regular backups. |
| Technology | Identify the components of your Data Architecture | * Data acquisition. * Data analysis. * Data storage. * Data transformation. * Data visualization. |
| Skills and Capacity | How will we promote development of data literacy skills and capacity throughout the organization (provide at least three ideas) | * Create incentives for developing best practices. * Encourage collaboration across departments. * Set up a mentoring program. * Utilize skill-building technology. * Develop an organizational model. |
| Support for Machine Learning | Give a brief description of the machine learning architecture and how it will interface with the data architecture | To interface with the data architecture, the machine learning architecture should be connected to data sources such as databases or streaming services. The data should then be preprocessed and transformed into features that can be used as input for the models. |

**Note: You may choose to represent this information on slide 11 of the CEO Presentation Template**

# Step 5 (OPTIONAL) - Record a short video of you presenting your final slide deck to your CEO or Executive Committee (5 minutes)

You may wish to submit a short video of you presenting your final presentation to your CEO; while this is not a formal requirement for the Capstone project, it does provide an outstanding way to gain practice with communicating about data science in business contexts.