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:** The Saudi British Bank

**Brief Company Description:** according to SABB website {SABB was established in 1978 as a Saudi Joint Stock Company and traces its origins in Saudi Arabia back almost 70 years, during which time it has been an active partner supporting the Kingdom’s economic growth and social development. As a partner of the HSBC Group, SABB is one of the leading corporate and institutional international banks in the Kingdom with a top retail banking and wealth management proposition. SABB is also a leader in Saudi Arabia and the region in trade finance, foreign exchange, equity and debt wholesale banking, digital service innovation, and ESG, paving the way for transformation and excellence.

SABB offers integrated financial and banking services, including corporate banking, investment, private banking, and treasury. SABB's paid-up capital is SAR 20.5 billion after the legal merger with Alawwal Bank, as the two banks became one entity and one of the largest banks in the Kingdom of Saudi Arabia.}

# 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:** [Enhance customers segmentation]

**Project 2:** [Personal finance & refinance prediction]

**Project 3:** [Credit Risk Scoring system]

**Project 4:** [Anti-money laundry and terrorism alert system]

**Project 5:** [customize the ways to attract mortgage customers]

**Project 6:** [Impact of e-learning on Staff performance]

**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:** [Enhance customers segmentation]

**Business Functional Area:** [Marketing & Salls]

**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: [digital world made all services and transactions done digitally without human interference. The AI will approach and attract customers. A good, enhanced customer segmentation result in a more customer focused business conduct and the products a tailored to meet every segment as the bank on the other hand should make every segment as a community and brand every segment according to targeted customer groups.]

- Role of data science in addressing the business problem:

[By collecting customers personal and financial information including employer and financial behavior]

- Targeted Business objective(s): [Customer On-boarding, Better Customer Experience and Increase Profit]

**2. Data Science Classification**

- Approach: [Descriptive]

- Type of Model: [Unsupervised learning , clustering]

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

* Customer income information
* Customers average balances
* Customers family members
* The profit generated from the relation

[The source of the data will be the bank data to be as a warehouse then a lake to gather new data and It does not has to be as dynamic]

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

[High magnitude. All banks are segmenting the customer with the focus on the business (usually three segments). to cluster customers in new ways that it will be easier to sell more tailored products with the customer focused approach with more sophisticated analysis]

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

[Medium cost with medium complexity]

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

[High, the new segments will help the bank to provide better customer experience as the segmentation will be more sophisticated. grouping the customers in smaller groups creating products and services in a more specialized way.]

**7. Key Business Stakeholders**

* Top MG
* Products
* Marketing
* IT
* Data warehouse

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

**Project 2 Name:** [Personal Finance & Refinance Prediction]

**Business Functional Area:** [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:

[It could be easier to identify existing customer who are eligible for refinance. The issue is that their eligibility changes with settlements. Moreover, we need to analyze customer spending. Approaching those customers manually will be difficult]

- Role of data science in addressing the business problem:

[AI to approach those customers]

- Targeted business objective(s): [Increase salles and revenue]

**2. Data Science Classification**

- Approach: [prescriptive]

- Type of Model: [unsupervised learning]

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

- Customer spending behavior

- Customer information

- Credit bureau history report

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

[High, personal finance is a core banking product and align with the bank’s targets. The project will increase salles and make profit]

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

[Medum to high]

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

[High, at a start the data needed are available in-house which will reduce the costs. On the other hand, training and experience data are clear to predict future purchases. Feedback and returning data will easy help us tailor offers in times of customers need.]

**7. Key Business Stakeholders**

* Top Management
* Business executives
* Marketing
* Sales
* IT
* Product

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

**Project 3 Name:** [Credit Risk Scoring System]

**Business Functional Area:** [Product, Sales and Risk]

**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:

[360-degree Credit risk analyses to the customer. Using preset KPIs by business and risk committee. Using the scores of the national (or international if possible) credit scoring bureaus like the Saudi credit scoring bureau (SIMAH). This will ease decision Making process to sell any financial product and reduce the risks of defaults and bad credit. It would also help on pricing based on the customer end score.]

- Role of data science in addressing the business problem:

[The system will allow end users to focus on business not the risk analyses of the customer. End user will rely on system decisions through reading, cross checking the customer and filtering them through the KPI’s to giving a final score to make the decision to finance the customer, give the customers a pricing based on their risk score (the higher the score the lower the finance cost)]

- Targeted business objective(s):

[Reducing cost through the reductions of bad credit and attracting good credit customers with special rates evidentially increase profit]

**2. Data Science Classification**

- Approach: [Prescriptive]

- Type of Model: [Semi-Supervised]

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

* Customer information
* Customer income information
* Customer SIMAH report
* Other credit scoring reports
* Peller’s used in the KPIs to give the decisions

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

[Very high, as it will reduce the costs of bad credits and allow us to approach new markets. Segmenting customers according to their data learned by the systems. High scores and good scores customers could be offered special prices to differentiate them from others.]

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

[High]

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

[High, the concept it-self is the bases of everyday retail finance. Since the biggening, banks suffered loses due wrong risk analysis and still suffering (one bad finance written-off will wipe the profit of many good finances). Currently, with all the available financial and non-financial data ML & AI and FINTECH reviewing customer risk will be digitalized for sales agents to be relied on for risk assessment decisions and keeping sales focusing on business.]

**7. Key Business Stakeholders**

* Top management
* Business
* IT
* Product
* SIMAH
* Retail Risk

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

**Project 4 Name:** [Anti-Money Laundry and Anti- Terrorism Alert System]

**Business Functional Area:** [Compliance and Legal]

**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 comply with international and national regulations by capturing changes in customers behaviors even accumulated small transactions. Cross check it with customer KYC (know your customer). AML team cannot handle the volume of the transactions to analyze each customer rather a system that will report a more filtered list of customers and transactions TO Investigated.]

- Role of data science in addressing the business problem:

[AI will alert only valid suspicions]

- Targeted business objective(s):

[Comply with AML and AT regulations and protect the bank and the society]

**2. Data Science Classification**

- Approach: [prescriptive]

- Type of Model: [supervised ]

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

* Customers statements
* Customers information
* Customer financial information
* Customers financial behavior

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

[High, as the data are all owned by the bank which will enable the system to detect more accurate variables and changes to suspected customers reducing work volume and increasing effectiveness of monitoring and prevention.]

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

[High]

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

[Medium, the whole world made a huge progress towards fighting money launderers and terrorists complying with Bazel 3 agreement. money launderers and terrorists find new ways inter the financial system without alerting AML monitoring systems]

**7. Key Business Stakeholders**

* Chief compliance officer
* Compliance
* AML
* Legal
* Operations
* IT

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

**Project 5 Name:** [Customize the Ways to Attract Mortgage Customers]

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

**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:

[Predicting potential mortgage customers attracting them with personalized marketing campaigns to increase sales and profits]

- Role of data science in addressing the business problem:

[Finding and predicting the customers reaction to the campaign through AI & ML thus increase the likely hood of the customer purchase

- Targeted business objective(s):

salles and revenue, market share, customer experience and customer special segmentation]

**2. Data Science Classification**

- Approach: [Prescriptive]

- Type of Model: [Unsupervised Learning]

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

* Customer information
* Customer financial information
* Number of family members
* Credit bureau
* 3rd party’s data about targeted customers
* Real estate market report

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

[Very high, as it is a core product of the bank retail finance and a strategical product, that is very profitable and easy to tailor. With the use of data science, the bank will not only sell their customers, but will reach new markets to on-board new customers increasing bank’s sales and market shares digitally]

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

[The cost will mostly be low to medium unless we gather 3rd party data for the customers to a more tailored and more personalized experience]

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

[High, from the experience data we have and 3rd parties’ data available attracting right potential customers at right timing with a personalized campaign to buy a mortgage, it will increase the possibility of purchase significantly. Conducting sophisticated analysis to past purchases related data, special segmentation, ML & AI (or deep learning) and the personalization in the way to approach the customer will increase sales]

**7. Key Business Stakeholders**

* Top Management
* Marketing
* Product
* IT
* Sales
* Business Executives

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

**Project 6 Name:** [The Impact of Soft Skills e-learning of Staff Performance]

**Business Functional Area:** [Human Resources L&D and Sales]

**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 bank and covid reduced the number of in-house training to the minimum and replaced it with e-learning. We need to analyze the impact of switching to e-learning when it comes to soft skills.]

- Role of data science in addressing the business problem:

[By identifying courses and changes in staff performance in general and if possible, in relation to that specific course]

- Targeted business objective(s):

[Increase sales through staff development at minimum costs at their own time]

**2. Data Science Classification**

- Approach: [Descriptive]

- Type of Model: [unsupervised]

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

* Training courses log history
* Staff performances
* Staff information and related data

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

[High to medium, as it will reduce the cost of staff development investments and improve business outcome]

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

[Low]

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

[Medium, staff performance could be measured in figures and ratios, a cross analysis between the trainings and the achievements will give us an indication that will also help suggest trainings to develop every staff]

**7. Key Business Stakeholders**

* Human resources
* Business Executives
* Learning and development

# 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**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Project No.** | **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:**  **[Enhance customers segmentation]** | 5 | 3 | 3 | 5 | 5 |
| **Project 2:**  **[Personal finance refinance prediction]** | 5 | 4 | 3 | 5 | 5 |
| **Project 3:**  **[Credit Risk Scoring system]** | 5 | 1 | 1 | 4 | 5 |
| **Project 4:**  **[Anti-Money laundry and anti- terrorism alert system]** | 4 | 3 | 2 | 4 | 3 |
| **Project 5:**  **[customize the ways to attract mortgage customers]** | 5 | 3 | 4 | 4 | 4 |
| **Project 6:**  **[The Impact of soft skills e-learning of staff performance]** | 4 | 5 | 5 | 3 | 4 |

**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 | [customize the ways to attract mortgage customers] | Mortgage is a strategical and highly profitable product, increasing its sales in alignment with the bank’s targets will be creating a high business value impact. The expectation to deliver is high in a short time with a medium 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 |
| 2 | [Personal finance refinance prediction] | With some similarities to the mortgage project and the high expectation to increase sales quickly. The profits generated from personal finance is less than mortgage but 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 |
| 3 | [The Impact of soft skills e-learning of staff performance] | The project has a medium business value impact, high feasibility, and medium likelihood of value capture. Yet, the results will support us to develop our e-learning systems that will increase moral, indirectly improve business outcomes, and could be used in the creation of data science culture |
| 4 | [Enhance customers segmentation] | High business value impact with a more advanced and sophisticated segmentation for a better understanding to our customers to create better products and enhance customer experience. Medium 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:** [Centralized Federated]

**Justification:**

[With very high sensitivity data, bank’s tech capabilities already in-house and the business functions (Retail, corporate and treasury) with different data science needs, a centralized federated model will be the most suitable option.]

**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:

|  |  |  |
| --- | --- | --- |
| **No.** | **Position** | **Team** |
| 1 | Data Scientist | 1 |
| 2 | Data Scientist | 1 |
| 3 | Data Scientist | 2 |
| 4 | Data Scientist | 2 |

Identify your roles and teams below:

|  |  |  |
| --- | --- | --- |
| **No.** | **Position** | **Team** |
| 1 | [Data Science Hacker] | 1 |
| 2 | [Data Science Hacker] | 2 |
| 3 | [Data Science] | 1 |
| 4 | [Data Science] | 2 |
| 5 | [Machine Learning Engineer] | 1 |
| 6 | [Machine Learning Engineer] | 2 |
| 7 | [Data Analyst] | 1 |
| 8 | [Data Analyst] | 2 |
| 9 | [Data Engineer] | Floating Team |
| 10 | [Data Visualization Engineer] | Floating Team |

**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 Science Hacker] | [With his/her general knowledge and bank capabilities it will enable us initiate first projects] |
| 2 | [Data Science Hacker] | [With his/her general knowledge and bank capabilities it will enable us initiate first projects] |
| 3 | [Machine Learning Engineer] | [Our first projects rely on machine learning] |
| 4 | [Data Visualization Engineer] | [As the first projects are business. Related data visualization for decision making to related to business mostly] |

**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 | [Align project results with business sales targets to show pure impact.] |
| 2 | [Data driven decision making as a culture throughout the bank from top management to front liners.] |
| 3 | [Focus on the usage of soft skills coaching through e-learning.] |
| 4 | [Special segmenting customers through AI will enable marketing to attract those customers with tailored marketing massage.] |
| 5 | [KPIs that based on data to major the performance for each staff within the bank to create a performance driven culture.] |
| 6 | [Create the need for our finance with the ease of digitalization.] |

**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? | * Single data warehouse * ETL * Relational data store |
| Data Governance | How will we promote data availability? (provide at least two ideas) | * Data available to all on need-to-know bases * Unify data driven decision making |
| How will we promote usability? (provide at least two ideas) | * Training as needed * Dictionaries, tags, etc * Aligning business needs with data science projects |
| How will we guarantee integrity? (provide at least two ideas) | * ETL checks and tracing * Misuse of data with lead to punishments, penalties and dismissal |
| How will we guarantee security? (provide at least two ideas) | * Firewall, encryption * Training to make data security as a culture * A digitally signed agreement of data privacy with all staff |
| Technology | Identify the components of your Data Architecture | * Data store represents single source of truth * Accessed via SQL * Spread sheets |
| Skills and Capacity | How will we promote development of data literacy skills and capacity throughout the organization (provide at least three ideas) | * Publish data updates organization wide * Promote and reward successful data driven projects * Promote a culture of experimentation and experimentation while leaving a space for mistakes |
| Support for Machine Learning | Give a brief description of the machine learning architecture and how it will interface with the data architecture | * As the data is very private and mostly stored in-house initial POC work to be done locally * Access ML through API endpoints to add a diverse look to the data |

**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.