

# 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
  - a. Opportunities must be spread across three different functional areas
  - b. 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.
  - a. 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
  - a. Data and Data Architecture Strategy
  - b. 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:** Foodics

**Brief Company Description:** Foodics provides a cloud-based point-of-sale SaaS ecosystem with tools that help F&B, and retail businesses start, track and grow.

Customers use Foodics to accept payments, track inventory, monitor sales, process orders, digitize menus, manage employees, create analytics and smart reports, provide secure cloud storage and enable the integration of third-party apps.

## 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:** Potential restaurant and customer Identification [marketing]

**Project 2:** Digital assistant [product]

**Project 3:** Growth Prediction for the company [finance]

**Project 4:** Restaurant industry insights and statistics [product]

**Project 5:** Supply chain optimization [supply chain]

**Project 6:** Eligibility score for financial programs [finance]

**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:** Potential restaurant and customer Identification

**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: Automatic generation of marketing leads
- Role of data science in addressing the business problem: Automation
- Targeted business objective(s): revenue, customer acquisition

**2. Data Science Classification**

- Approach: unsupervised
- Type of Model: clustering algorithm (K-means, Fuzzy K-means, Hierarchical clustering, Mixture of Gaussians)

Justification: Adequate performance based on required work to implement.

Reference:

[https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUK Ewiw-9CY6IL9AhV9XaQEHTo8AtAQFnoECAsQAQ&url=http%3A%2F%2Fwww.diva-portal.org%2Fsmash%2Fget%2Fdiva2%3A1118107%2FFULLTEXT01.pdf&usg=AOvVaw1cF6\\_PENqiD4XoHDsecErp](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUK Ewiw-9CY6IL9AhV9XaQEHTo8AtAQFnoECAsQAQ&url=http%3A%2F%2Fwww.diva-portal.org%2Fsmash%2Fget%2Fdiva2%3A1118107%2FFULLTEXT01.pdf&usg=AOvVaw1cF6_PENqiD4XoHDsecErp)

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

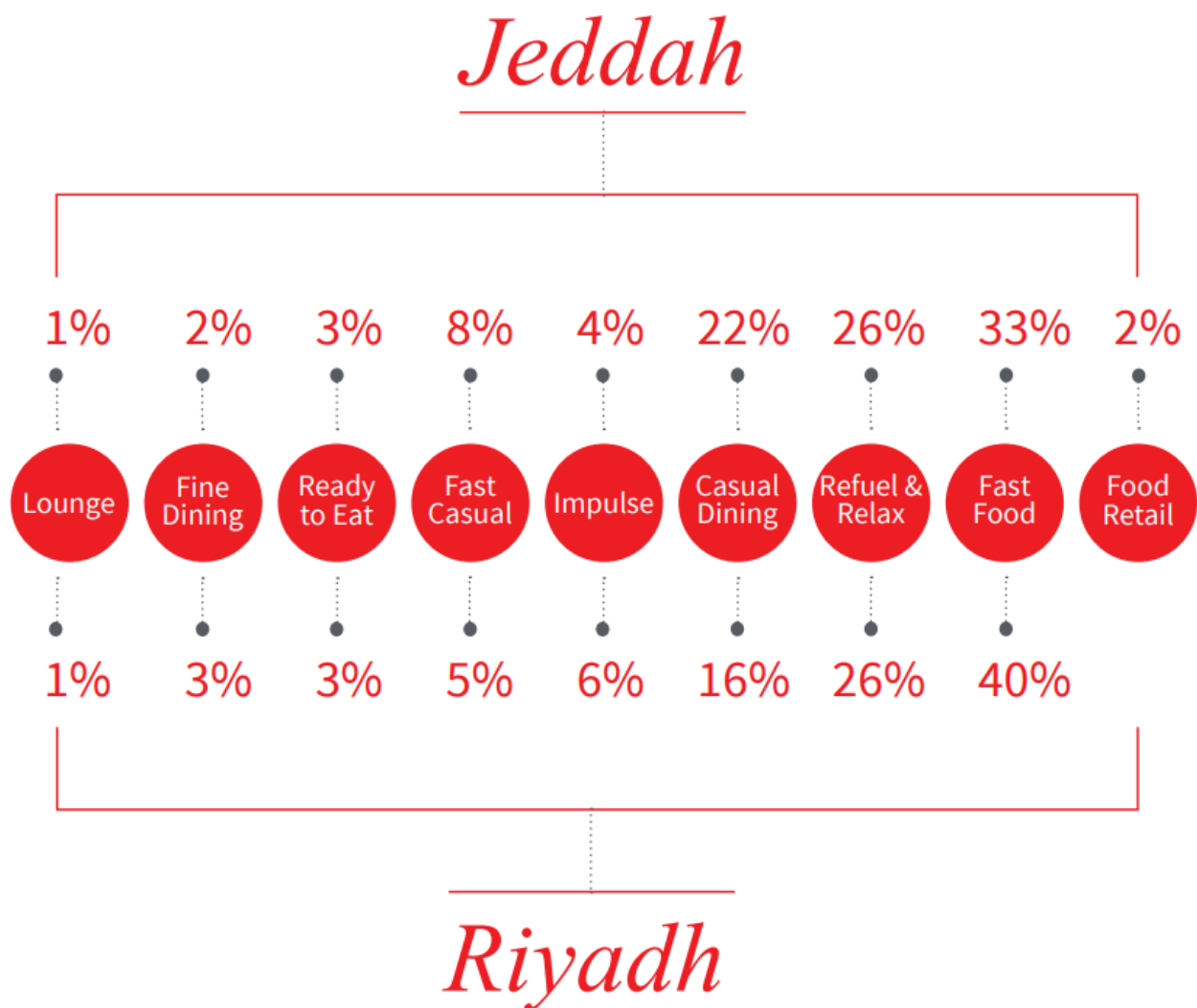
- in-house data
- Scraping of online data
- Purchase of data from restaurant delivery apps

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

High,

- 'The Chefz' delivery app was acquired by its compotator 'Jahez' for \$172 million. [ref](#)
- The food service market size was \$12.57 billion in 2021. [ref](#)
- Break down of the Saudi market categories:

Infographic 1 | Comparison % of units per category



[ref](#)

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

The cost can become considerable, but there are many online / open-source datasets that can be scraped and cleaned to offset costs of acquiring market data. As for project complexity from a data science prospective it is very simple and straight forwards given that sufficient data is present.

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

High, because a lead generation system will focus resources and attention to people and organizations that have a higher likelihood of signing up to the services and products provided.

## **7. Key Business Stakeholders**

- Executives.
- Finance department head.
- CFO
- Marketing department head.
- CMO
- 3<sup>rd</sup> Party applications.
- Data providers.

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

**Project 2 Name:** Digital assistant

**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: Creation of new future growth opportunities for the company through new products into different sectors.

- Role of data science in addressing the business problem: Creation of the new product

- Targeted business objective(s): revenue

**2. Data Science Classification**

- Approach: Deep learning

- Type of Model: NLP

Justification: Deep learning is the only way to create a sufficient digital assistance based on the complexity of the problem that we are wanting to solve.

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

Data needed: Text based human conversations.

- Online data
- Open-source datasets

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

- Medium, "The global intelligent virtual assistant market size to grow from USD 7.01 billion in 2021 to USD 47.57 billion, by 2028 with an estimated CAGR of 31.9% from 2022 to 2028" [ref.](#)

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

This project is of high complexity, and of high cost due to the amount of data computing power, computation time, and work hours required to complete such a project.

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

Medium, because the core of the project can be used as a start point for many other similar projects and the focus would be laser focused on Arabic language support at the beginning.

## **7. Key Business Stakeholders**

- Executives
- Data providers

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

**Project 3 Name:** Growth Prediction for the company

**Business Functional Area:** Finance

**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: What is the company's growth based on current trajectories.

- Role of data science in addressing the business problem: creation of the growth forecasting model

- Targeted business objective(s): revenue, cost reduction

**2. Data Science Classification**

- Approach: LSTMs (RNN based model)

- Type of Model: Deep Learning

Justification: These models have proved to be capable of recognizing historical patterns and change the estimation to reflect previous historical events.

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

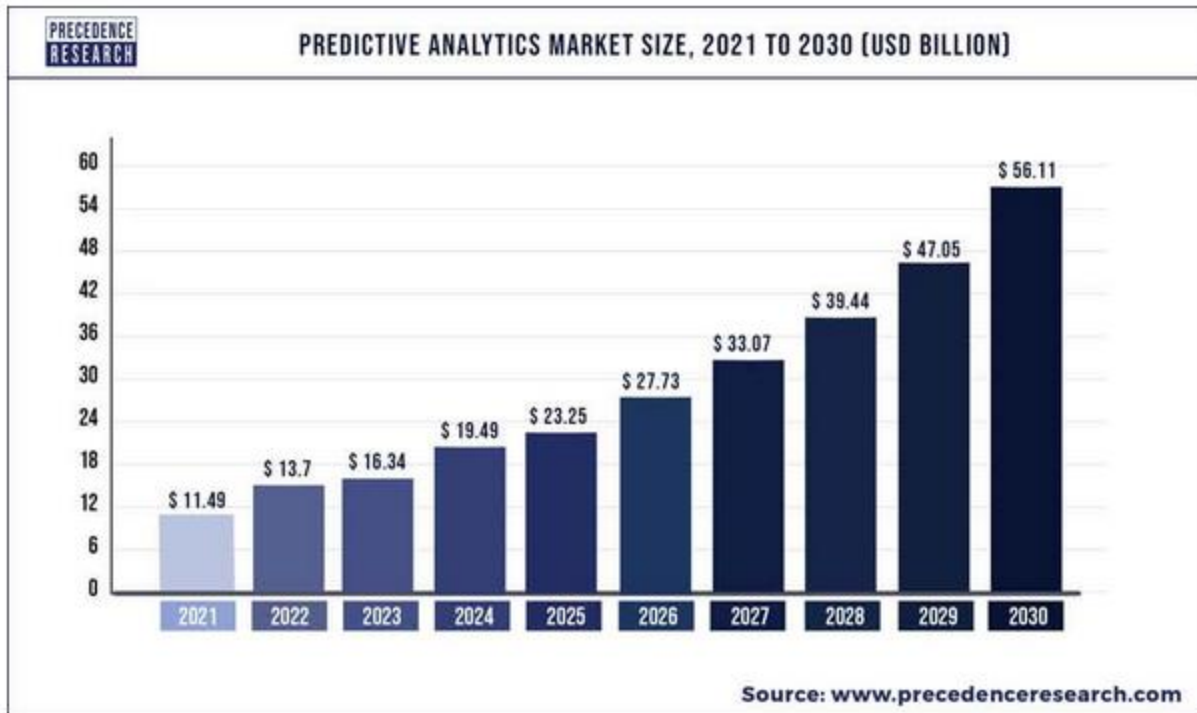
- Open-source datasets.

- Internal data.

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

- Medium, "The global predictive analytics market size was estimated at USD 11.49 billion in 2021 and it is expected to hit at around USD 56.11 billion by 2030 with a registered compound annual growth rate (CAGR) of 19.27% during the forecast period 2022 to 2030." [Ref.](#)





## 5. Cost and complexity of development and implementation

Given that the data is highly available, the cost would be lower and complexity would be average given that the data points considered by the model are limited.

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

Medium

## 7. Key Business Stakeholders

- Executives
- Finance department head.
- CFO

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

**Project 4 Name:** Restaurant industry insights and statistics

**Business Functional Area:** Products

**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: What is the market like now and what is it likely to look like in the future?

- Role of data science in addressing the business problem: Automatic generation of insights based on market data.

- Targeted business objective(s): revenue

**2. Data Science Classification**

- Approach: Deep learning

- Type of Model: NLG

Justification: The type of product that we want to create required the generation of text based on the input data and statistics and that task is best done by the NLG deep learning models.

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

- Open-source datasets

- Internal data.

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

- High, "The global revenue of the market research industry exceeded 76.4 billion U.S. dollars in 2021" [ref.](#)

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

Medium cost, and high complexity.

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

High, Because the company has a unique position in the market whereby very few have the data and costumer base that the company has.

## **7. Key Business Stakeholders**

- Executives
- Business department head.
- CIO
- CTO

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

**Project 5 Name:** Supply chain optimization

**Business Functional Area:** supply chain

**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: What is our forecasted inventory needs based on supply and demand?

- Role of data science in addressing the business problem: Creation of model to forecast inventory needs based on supply and demand.

- Targeted business objective(s): cost reduction

**2. Data Science Classification**

- Approach: Deep learning

- Type of Model: CNN

Justification: CNNs have proved to be amazing forecasting models.

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

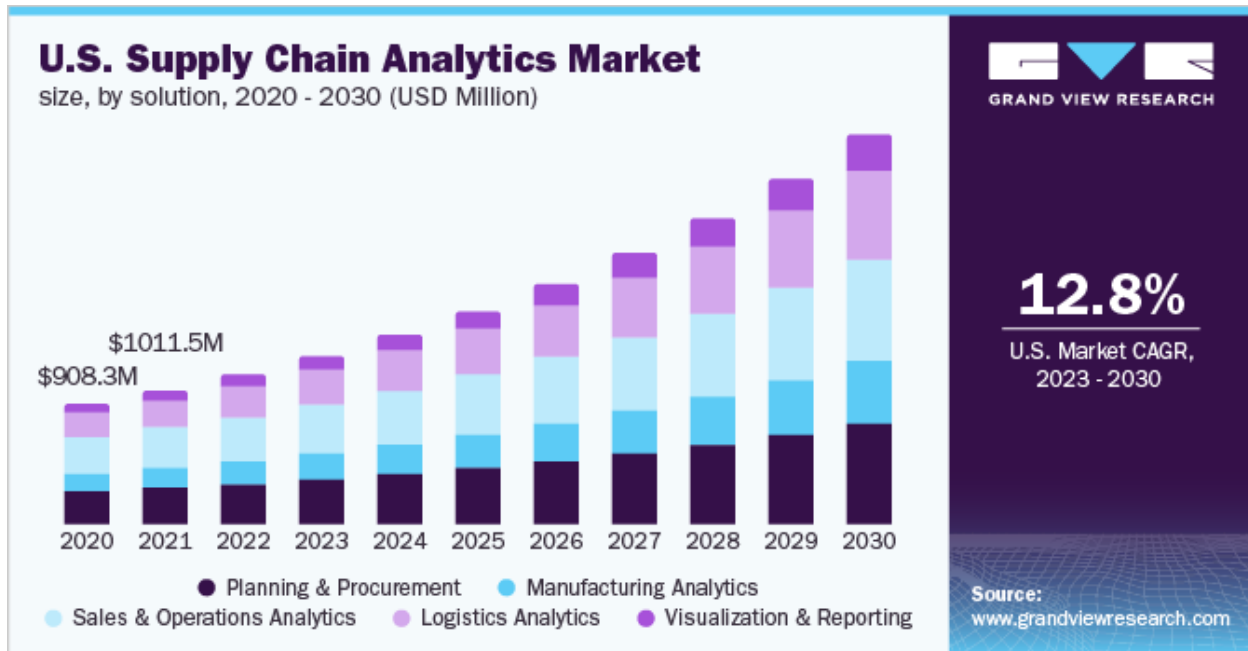
- Open-source datasets.

- Internal data.

- Data providers.

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

- Medium, "The global supply chain analytics market was valued at USD 6.12 billion in 2022 and is anticipated to expand at a CAGR of 17.8% from 2023 to 2030." [ref](#)



## 5. Cost and complexity of development and implementation

High complexity, high cost.

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

High, Efficient usage of resources at the right time will allow for more liquidity and better ability to respond to sudden shocks and changes to the supply chain.

## 7. Key Business Stakeholders

- Executives
- Warehousing department head
- CTO

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

**Project 6 Name:** Eligibility score for financial programs

**Business Functional Area:** Financial

**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: What is the eligibility of a costumer to apply for a financial program?

- Role of data science in addressing the business problem: Building the model for evaluating the costumer's eligibility score.

- Targeted business objective(s): revenue, customer acquisition.

**2. Data Science Classification**

- Approach: Supervised.

- Type of Model: Logistic L2 classification model.

Justification: A simple model with adequate performance with classification work loads.

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

Credit score datasets are required.

- Open-source datasets.

- Internal data.

- Data providers.

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

- High, "The global credit rating software market size was valued at \$0.42 billion in 2020, and is projected to reach \$1.92 billion by 2030, growing at a CAGR of 16.5% from 2021 to 2030." [Ref.](#)

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

High cost, low complexity.

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

High, because we already have a customer base and we already have the cloud computing and data processing capabilities even if we don't have the ability or will to provide the financial programs we can partner with our financial partners to create the financial products.

## **7. Key Business Stakeholders**

- Executives.
- Finance department head.
- CFO
- Financial partners.
- Data providers.

## 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
<b>Project 1:</b> Potential restaurant and customer Identification	5	3	3	4	4
<b>Project 2:</b> Digital assistant	3	1	1	2	3
<b>Project 3:</b> Growth Prediction	5	3	3	2	3



for the company					
<b>Project 4:</b> Restaurant industry insights and statistics	5	4	4	5	5
<b>Project 5:</b> Supply chain optimization	5	2	3	3	3
<b>Project 6:</b> Eligibility score for financial programs	5	2	4	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?)**

<b>Project Order</b>	<b>Project Title</b>	<b>Order Justification</b>
1	Restaurant industry insights	This will be an easy first project that will provide an immediate revenue stream that will allow us to carry over our momentum to the next project.

	and statistics	
2	Potential restaurant and customer Identification	This project requires more resources but will allow us to establish a data-based culture within the company after the success of the project.
3	Eligibility score for financial programs	This project be placed third since it is directly associated with finances, which has a high cost on failure.
4	Growth Prediction for the company	This project will be great after the trust of top management has been established and confidence levels are high.

**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:** Federated.

**Justification:**

We will start with a Federated model then slowly move to a Centralized Federated model, because the company is already has integrations with cloud providers; thus, governance and access to data for data science projects can be best maintained with these models.

## 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:

	<u>Position</u>	<u>Team</u>
1	Data Scientist	1
2	Data Scientist	1
3	Data Scientist	2
4	Data Scientist	2

Identify your roles and teams below:

	<u>Position</u>	<u>Team</u>
1	Data Scientist	1
2	Data Scientist	1
3	Data Engineer	1 / 2
4	Data Visualization Engineer	1 / 2
5	Data Scientist	2
6	Data Scientist	2
7	Machine Learning Engineer	2
8	Machine Learning Engineer	1
9	Full-Stack Developer	3
10	Back-End Developer	3

**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 Scientist	I'll start with an all-rounder data scientist that can work everything
2	Data Scientist	I'll offset the workload of the first hire with a data scientist that has more in-depth expertise
3	Data Engineer	I'll now hire a data engineer to ensure that data is available and ready for the data scientists before a project begins
4	Data Visualization Engineer	I'll now hire a data visualization engineer with great communication and presentation skills to communicate the results of the department's work

**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	Ensure that top management is on-board for a data-driven culture
2	Ensuring that everyone has basic data-access
3	Starting the habit of explaining analytical choices
4	Quantifying uncertainty
5	Choosing metrics with care
6	Encourage data scientists to innovate and start initiatives.

**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?	<ul style="list-style-type: none"> <li>- Single data warehouse</li> <li>- ETL from cloud-based data stores</li> <li>- Relational data store</li> </ul>
Data Governance	How will we promote data availability? (provide at least two ideas)	<ul style="list-style-type: none"> <li>- Data available to all employees</li> <li>- Announcement of new dataset integrations</li> <li>- Allow dataset requests for unavailable datasets</li> </ul>
	How will we promote usability? (provide at least two ideas)	<ul style="list-style-type: none"> <li>- A dictionary of the available data</li> <li>- Tags for easier searchability</li> <li>- A meta store alongside the data store</li> </ul>
	How will we guarantee integrity? (provide at least two ideas)	<ul style="list-style-type: none"> <li>- Regular check-ups on datasets</li> <li>- Reporting of issues by consumers of data</li> <li>- Data quality scores</li> <li>- Tracing of data</li> <li>- Regular scheduled back-ups of data</li> </ul>
	How will we guarantee security? (provide at least two ideas)	<ul style="list-style-type: none"> <li>- Training of employees</li> <li>- Firewalls</li> <li>- Encryption of data while moving</li> <li>- Masking of user information</li> </ul>
Technology	Identify the components of your Data Architecture	<ul style="list-style-type: none"> <li>- A data store that is the single source of truth</li> <li>- Access through SQL queries or BI applications</li> <li>- Access through APIs that expose information of the data store</li> </ul>
Skills and Capacity	How will we promote development of data literacy skills and capacity throughout the organization (provide at least three ideas)	<ul style="list-style-type: none"> <li>- Just in time training</li> <li>- Publishing data updates to the whole company</li> <li>- Rewarding successful data-driven projects</li> <li>- Promote data exploration and experiments that use the available data</li> </ul>
Support for Machine Learning	Give a brief description of the machine learning architecture and how it will interface with the data architecture	<ul style="list-style-type: none"> <li>- POC on local</li> <li>- Transition to cloud</li> <li>- API endpoints for access to ML models</li> </ul>



**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.