100 Day Data Science Plan :

Building a Data Strategy

UDACITY

alteryx

Capstone Project for: Data Science for Business Leaders Nanodegree

Submitted by: Frederick Zoreta

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: [ALTERYX]

Brief Company Description: A US-based software company that specializes in analytics and data science. The company is known for having an analytics and data science platform that is highly capable of doing heavy analysis and making it available for the 'regular data worker' or 'citizen data scientist'. Their platform is considered a no-code / low- code platform.

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: [Spreadsheet Automations]
Project 2: [Advancing Analytics Dashboard]
Project 3: [Fraud Detection]
Project 4: [Sentiment Analysis]
Project 5: [Predicting Churn]
Project 6: [Quantitative Risk Management]

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: **[Spreadsheet Automations]**

Business Functional Area: **[Financial Management]**

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: Although this is mostly an automation-based project, to be more specific a Robotic Process Automation (RPA), Data Science plays a vital role here. RPA thrives with the use of ML and AI. Since 2019, 'Intelligent Automation' has been on the rise.

This would likely be a heavy utilization of UIPath with a lot of VBA , C# coding and also some minor ML engineering using Python libraries.

- Role of data science in addressing the business problem: 'Intelligent Automation' would not be possible without the use of data. Datascience would enable the automation process to be "faster, more precise and having more efficient results."

- Targeted business objective(s): [Place a

2. Data Science Classification
- Approach: **[PRESCRIPTIVE APPROACH]**
Since this specific project mainly focuses on automating the spreadsheet (both excel and google sheets) tasks, this approach would be ideal. This project has a 'specific task in mind'. Prescriptive approach is very ideal for having a 'bias towards action over monitoring'. This approach has also been proven to be utilized in automated marketing message testing.

- Type of Model: **[UNSUPERVISED LEARNING]**

This task is mainly for automating the spreadsheets. There is no specific data that needs to be wrangled, analyzed and visualized. The 'Intelligent Automation' part would come in at a much later period, which is outside the scope of this 100 day project timeline.

3. Data needed for project and sources for that data

There is no specific data set that is required for this specific project. More likely our team of RPA specialists together with a handful of data analysts; would look at past documentations of automation projects. We would also be analyzing past case studies form partner organizations.

4. Magnitude of opportunity (with justification)

This will have a huge impact on delivering projects on time-sensitive projects. A definite positive impact on: increase in productivity (almost all manual data entry and basic formulas would be eliminated), save on costs (at least 20 employees would be considered redundant. They were given 'upskilling' to eventually change to a different role), revenue would eventually increase since automation saves a lot of time & energy.

5. Cost and complexity of development and implementation

This would be a low cost & medium complexity type of project. Low cost because we have the necessary manpower to automate all our spreadsheets and its necessary functions. The specific infrastructure involved, which is UIPath Enterprise has already been purchased years ago. Our cloud based solutions would easily integrate with very minimal costs.

The complexity becomes medium level because just like all RPA related projects, it would involve several cycles and iterations of testing. We also have to make sure that the VBA code and the specific Power Query/M Language formulas are functioning well.

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

Because of the automation process, less hours and energy would be focused on redundant and rules-based tasks. As previously mentioned, this would save a lot of manpower that would eventually be upskilled and work in other departments where they could give more value.

7. Key Business Stakeholders

Key stakeholders here would be the following:
CTO, IT Dept, 1 Data Scientist, 2 Business Intelligence Specialists, 2 UX Researchers/Designers , 3 Agile Specialists , 1 RPA Consultant and 3 in house coders who specialize in VBA & C #.

The entire Finance dept is also a key stakeholder.

For each candidate project, please provide the following detail:

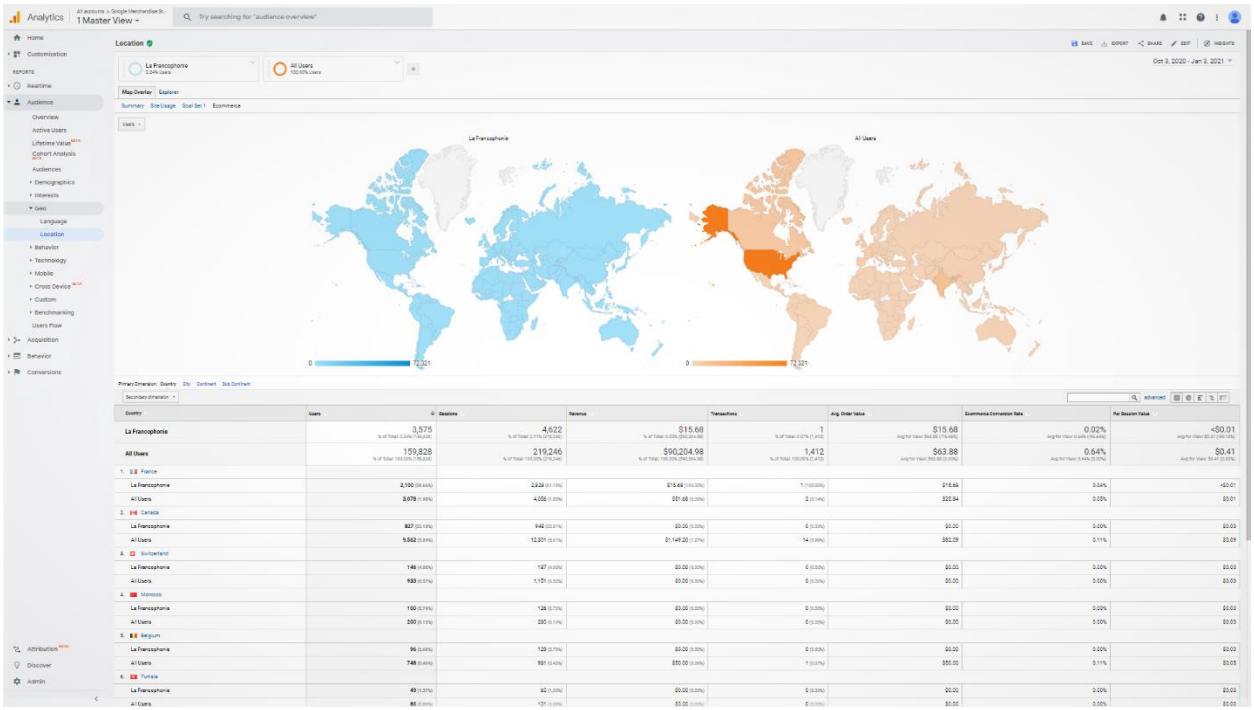
Project 2 Name: [Advancing the Analytics Dashboard]

Business Functional Area: [Digital Marketing]

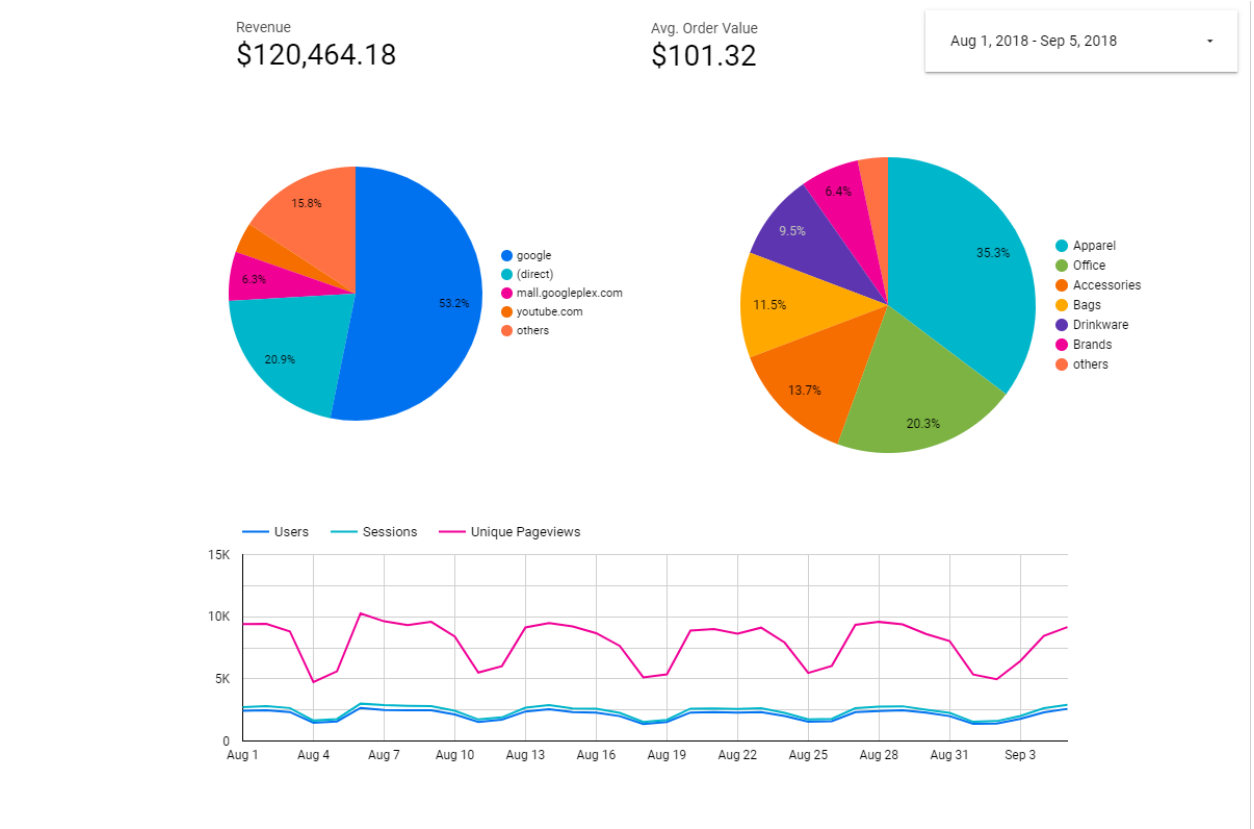
1. Description of the project (including business problem to be addressed, how data science will address that business problem, and the targeted business objective (revenue? customer acquisition? cost reduction?):

- Business Problem Addressed: The entire marketing department has been backlogged due to not having a centralized 'information radiator' or analytics dashboard. The entire team decided that there needs to be a truly dedicated analytics dashboard where several 'internal groups' within the marketing team could use. This way, they could track all their specific metrics and KPIs such as: Page Rank, Bounce Rate, Conversion Rates, Cost Per Acquisition (CPA), Costs Per Click (CPC), etc.

Below is a classical example of simple metrics using Google Analytics:



And a simple data viz using Google Data Studio:



- Role of data science in addressing the business problem: Although this may appear to simply involve the 'analytics and business intelligence' specialists, data science would surely have a huge impact here. Using machine learning algorithms would help us with a more accurate, real time analytics overview. Another main reason for having actual ML and (eventually) AI specialists working with our B.I. team, is that we have several datasets that have unstructured data such as: voice, videos and a lot of images. With this in mind, our team would eventually embed the ML specific tool kits into our BI powered analytics dashboards.

- Targeted business objective(s): Our organization wants to further solidify all our digital marketing efforts. Although both Google Analytics and Google Data Studio does a great job regarding tracking of KPIs and metrics, the CMO and CEO wants a deeper insights. They also want to eventually have a dashboard to monitor unstructured data such as: images, videos and possibly, sound files.

2. Data Science Classification

- Approach: **[DESCRIPTIVE APPROACH]**
- Type of Model: **[SEMI-SUPERVISED LEARNING]**

3. Data needed for project and sources for that data

A combination of both structured and unstructured data. 95% would be structured / relational data that would be derived from Google Analytics and Google Adwords. We already have a functional dashboard using both Tableau Server and Google Data Studio. Our main goal is to still keep these BI Tools functional, but C-suite wants to have a 'master level' analytics tools which eventually would be used for non structured data.

4. Magnitude of opportunity (with justification)

Tremendous opportunity, especially once we have established the non structured data. This would give us an edge over competitors and other market players that solely rely on data gathered from relational databases or simply Google Analytics.

5. Cost and complexity of development and implementation

This is medium complexity. The 'complex' part only comes when there has been a huge increase in unstructured data that needs to be further analyzed. Another possible issue of complexity is having the proper ML Algorithms that would match the specific insights and needs of the department.

In the future, this project may be "revisited" and have a version 3.0. With the growing popularity of APIs and 'social networks', we would be integrating specialists in GraphQL (gathering data from APIs) and graph database experts, specifically Neo4j.

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

This project has very high value once it is in full implementation. The entire similarity of this project would be that of having our 'personalized ElasticSearch stack'. This entire analytics platform would be able to fully give insights even with unstructured data.

7. Key Business Stakeholders

The key stakeholders here would be:
the marketing department, CTO, CMO, 2 Machine Learning Engineers, 3 data visualization specialists (Tableau Professional level and Tableau Server level). Also 2 UX/UI designers.

** In the future I may do a second version and add graph database consultants and possibly someone who is an expert level in the entire 'ElasticSearch' stack.

For each candidate project, please provide the following detail:

Project 3 Name: **[FRAUD DETECTION]**

Business Functional Area: **[FINANCE – ANTI MONEY LAUNDERING (AML) DIVISION]**

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: This project is of the highest priority. Our organization's finance department has had fraudulent documents and even transactions the past 8 months. Our prediction is that there may be more fraudulent transactions ahead if we do not act on it right away.

This would be addressing the issue of anti money laundering (AML). Reducing and eventually eliminating any form of fraudulent transactions means that we would not be losing revenue. This would also further solidify our organization as having a robust system of financial transactions.

- Role of data science in addressing the business problem:
Data science would enable the team to determine the patterns and anomalous transactions. This is were 'classical or old school machine learning' would come in.

- Targeted business objective(s): Eliminate the possibility of having fraudulent transactions. Anything that relates to fraud could automatically mean lost revenue. It would also tarnish the image of the company. This could lead to loss of potential and current customers (churn).

2. Data Science Classification
- Approach: **[PREDICTIVE APPROACH]**
- Type of Model: **[SUPERVISED LEARNING - CLASSIFICATION]**

3. Data needed for project and sources for that data

We already have a lot of datasets that were derived from both past transactions, past documentations and also some past case studies. This is a combination of both structured and unstructured data.

4. Magnitude of opportunity (with justification)

The revenue and the over-all image of the company is at stake here. Identifying and eventually avoiding any form of fraudulent transactions is of utmost priority. It is common knowledge that any form of fraudulent transaction, especially on a massive scale; would cause any organization a huge amount of money.

5. Cost and complexity of development and implementation

There would be a huge cost to this project. The over-all cost and complexity for this project is high. Detecting anomalies that leads to fraud usually takes time. This would also include a lot of testing with the 4 machine learning engineers that would be involved in this project. The implementation may actually last slightly longer than what is expected, but this could still vary.

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

Because of the correct predictions of anomalous transactions, we would then be able to eliminate anything that involves fraud. We won't risk having to deal with 'money laundering issues' within our organization. This would save a lot of time, money, energy and avoid embarrassment. The threat of lawsuits is also eliminated.

7. Key Business Stakeholders

3 Machine Learning Engineers, Chief Data Officer, Chief Finance Officer, 1 data visualization specialist, 2 data translators, 2 data architects

For each candidate project, please provide the following detail:

Project 4 Name: [SENTIMENT ANALYSIS]

Business Functional Area: [LEGAL DEPARTMENT]

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 legal team wants to have a much faster and efficient way of identifying patterns and behaviors within certain cases. This actually deals also with a lot of text-search patterns. Since the legal department deals with mainly qualitative data, sentiment analysis is the most appropriate method used for this issue.

- Role of data science in addressing the business problem: Sentiment analysis allows our data science team to further probe and look deeper into key specific insights regarding what the legal department wants/needs.

- Targeted business objective(s): [Place answer here]

2. Data Science Classification
- Approach: **[PREDICTIVE]**
- Type of Model: **[BOTH SUPERVISED (Classification) & UNSUPERVISED LEARNING (Clustering Algorithms)]**

3. Data needed for project and sources for that data
All the data and documents within the legal department. These would be a great majority of qualitative data. This would also be considered unstructured data and more likely stored in a document-oriented database like MongoDB. On the other hand, for future projects the legal department may also use a graph database (Neo4j) in identifying key players that have interconnections with certain documents/patterns or behaviors.

4. Magnitude of opportunity (with justification)

Although this is not of utmost priority, this project would eventually free our entire legal team with all the 'mundane tasks'. This is like the 'quantitative' equivalent of automating spreadsheets (Project # 1) and all the necessary basic calculations involved. Freeing our legal team from all these mundane tasks would eventually boost their productivity. This could also equate to having a great customer experience for our clients that needs to deal with legal issues.

5. Cost and complexity of development and implementation

This is of medium complexity. Implementing sentiment analysis is quite common especially in qualitative datasets. We also have qualified data scientists who can handle this task and deliver within a specific timeline. To add up, we already have access to the entire stack of documents which are electronically stored in a document oriented db; MongoDB.

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

As stated in the above specifics, we have access and full ownership of such documents and data. They are already stored in a non structured database. The main task here would be simply to create an algorithm and run the sentiment analysis, depending on what the needs and wants of a legal department team member is (this could be on a very specific task too).

7. Key Business Stakeholders

1 Machine learning engineer, 1 data architect, 1 data visualization specialist

For each candidate project, please provide the following detail:

Project 5 Name: [PREDICTING CHURN]

Business Functional Area: [DIGITAL 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: This specific project is somewhat connected to project # 2: Advancing Analytics Dashboard. The main purpose of that project was to give the entire marketing team a "bird's eye view" of the metrics, KPIs and performances.

The main goal of this project is to predict who would likely churn among our current customers. This eventually would involve further research, look at certain customer patterns and look for ways and methods to prevent customers from leaving our business.

- Role of data science in addressing the business problem: Our data science team would enable us to have much deeper insights on our customer's past, current and eventually on future journeys. This means that moving forward, our marketing team would have heavy collaborations with the data science and analytics team. In a highly competitive world, simple reliance on Google Analytics, Google Adwords and Facebook Banner Ads may simply be not enough.

- Targeted business objective(s): Minimize and possibly eliminate any customer churn. Every time that a customer/ set of customers leave our business, it is automatically lost revenue on our part. This is something that we should minimize.

2. Data Science Classification

- Approach: **[PREDICTIVE ANALYTICS]**
- Type of Model**: [SUPERVISED LEARNING – Regression & Classification AND Recommender Systems]**

3. Data needed for project and sources for that data

This would be a combination of both quantitative (all transactions, documentations, spreadsheets, RDBMS generated) and qualitative (those form survey results and studies done by the UX/CX research team).

The 'sentiment analysis' methods and algorithms used in project # 4 for our legal department could also be utilized here. This is in order for us to determine the sentiment of our users.

4. Magnitude of opportunity (with justification)

There is a very high magnitude of opportunity here. As a heavily for profit enterprise, we aim to make sure that our financial flow is continuous with very minimal or 'next to zero' damages. Predicting customer churn is a very valuable tool in such a way that would give the c0suite, marketing managers and even some of our finance specialist on how to maximize growth for the organization.

5. Cost and complexity of development and implementation

This is a highly complex project. Although not of the most urgent nature, predicting churn, on top of all the qualitative data that needs to be analyzed would take a lot of time and energy. This may eventually be a recurring project after the 100 day plan. Another issue that complicates this is how accurate all the UX research surveys were. This would be like using several projects on top of each other and eventually predicting how to prevent (or realistically; minimize) churning.

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

There is a likelihood of high value capture here. Predicting customer churn is almost an essential in any for profit organization these days. We are also confident that our team could definitely deliver on what is expected of them.

7. Key Business Stakeholders

Chief Data Officer, Chief Marketing Officer,
3 Data Scientists, 2 data visualization specialists, 2 machine learning engineers, 1 business analyst, 3 UX/UI engineers, 4 digital marketing analysts

For each candidate project, please provide the following detail:

Project 6 Name: [QUANTITATIVE RISK MANAGEMENT]

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: This project is all about building models in order to further understand the risks of financial portfolios. We also like to explore market-risk factors. This is a very vital task in our industry. We will be collecting data on the underlying risk factors that affects portfolio values & then analyze their behaviors.

- Role of data science in addressing the business problem: Will aid in identifying risk factors, real world returns and spotting volatile time-series. Machine Learning would be extremely effective in this particular task.

- Targeted business objective(s): Higher revenue and minimize all the possible financial risks involved.

2. Data Science Classification
- Approach: **[PREDICTIVE]**
- Type of Model: **[UNSUPERVISED LEARNING – Dimensionality Reduction]**

3. Data needed for project and sources for that data
All the specific financial data that pertains to our risk management portfolio. Majority of these would obviously be quantitative, but we will also analyze several qualitative data that is part of the finance department.

4. Magnitude of opportunity (with justification)

This would eventually have a huge impact on our revenue. This would enable us to fully understand risks in certain finance portfolios. With this, we could have a huge advantage over competitors. This would equate to higher revenue.

5. Cost and complexity of development and implementation

This is of high complexity since this would require a highly specialized data scientist and machine learning engineer to do. Also, there are a lot of factors and issues involved with regards to analyzing quantitative risk management portfolios. One also has to analyze to past and current market / trading trends, and also look at the past and current customer buying behaviors.

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

Since this would involve a high stakes financial value, plus having a presence of a highly competent data science, finance and analytics team, this project is considered a high value capture task.

7. Key Business Stakeholders

3 Machine Learning Engineers (2 of which specialize in financial ML),
1 Data Engineer, 2 Business Intelligence Analysts, 3 Financial Analysts , 1 Data Architect , 2 CPAs

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: [Spreadsheet Automation]	4	3	3	5	5
Project 2: [Analytics Dashboard]	5	4	4	4	4
Project 3: [Fraud Detection]	5	2	2	5	5
Project 4: [Sentiment Analysis]	3	5	5	3	3

Project 5: [Predicting Churn]	5	4	4	5	5
Project 6: [Quant Risk Mgmt]	3	3	2	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	Fraud Detection	**There has been an urgent need from the organization due to the rising occurrences of financial fraud.**
2	Spreadsheet Automations	**Since the main focus of the project is on an entire Automated Analytics Platform, automating most of the excel & google sheets is vital. There will be a lot of VBA, C# coding and a heavy reliance on an RPA platform (UIPath).**
3	Analytics Dashboard	**The marketing team is highly in need of a more robust BI tool. Although the current platform functions well, the CMO wants to have a real time insight into each KPIs.**
4	Quant Risk Management	**Building ML models to analyze several portfolios is a complex task. Although this would take up manpower and would fully utilize the entire specializations of the team, the urgency is not that crucial.**

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 – Functional Led]

Justification:
This is the most optimal since each of the 6 projects are totally different tasks. Although they are interconnected in some ways, having a federated model allows the functional teams to decide on their specific action and goals.

I would like to point out to the RPA-led project for spreadsheets. Although there would still be a Chief Data Officer that would be monitoring the over-all process, each of the functional teams would be reporting to a designated team lead. These team leads are highly skilled in those respective project tasks and are also experts in the field of Agile Methodology.

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	Machine Learning Engineer	Team 1
2	BI/BA Analysts	Team 1
3	UX/UI Engineer	Team 1
4	Deep Learning Expert	Team 1
5	Data Visualization Engineer	Team 2
6	Growth Product Manager	Team 2
7	Agile Consultant	Team 2
8	RPA (UI Path – specific) Expert	Team 3
9	Data Viz (Tableau- specific) Expert	Team 3
10	Elasticsearch Certified Professional	Team 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	3 Machine Learning Engineers	ML is extremely crucial in ALL our operations. These initial 3 ML experts would also be given advanced briefing regarding all 6 projects and the possibility of future tasks.
2	3 Agile Specialists	Agile Framework is extremely favored over Waterfall in all 6 projects. I aimed for 3 'Agilists'
3	2 RPA (UiPath) Experts	Having automation specialists is very vital although not as urgent as the first 2. Eventually having automated spreadsheets and other finance capabilities would aid a lot in the data science projects.
4	2 UX/UI Engineers	Usability is a huge factor, but since our projects are more on the data and automation side, the need for UX experts would only come at a later date. Their main task would be on the possibility of creating our own UI's for some of the projects.

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	Identifying more than specific data science methodology for each project, then deciphering if 1 or 2 methods are enough to arrive at a solution
2	Viewing each project not just as a stand alone task but also as highly inter-connected to other projects. After all, a poor performing project would eventually affect the performance of others.

3	Focus and drive all projects to specific proof of concepts. Look at them as a whole unit with the common goal of making our organization much better.
4	Maximizing the skills and talents of our employees and partner consultants, before looking externally for a potential consultant
5	Maximizing the current infrastructure we have; both on premise & cloud based.
6	Agile methodology is the main framework used all throughout the 6 projects. Since these projects would be unpredictable and involve certain iterations, Agile trumps waterfall in all 6 of them.

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?	This is a huge scope. This would include a lot of financial data and a lot of marketing (both digital & traditional). There will also be a certain amount of qualitative data that would be gathered from UX research.
Data Governance	How will we promote data availability? (provide at least two ideas)	All data that would be utilized is readily & openly available to the team members involved ONLY. We consider the data being used as HIGHLY CONFIDENTIAL, specifically for Project 3 (Fraud Detection) and Project 6 (Risk Management). All other data that deals with legal matters and marketing data should be kept with utmost confidentiality.
	How will we promote usability? (provide at least two ideas)	• Data usability would be on a case by case basis. As every project moves forward, certain groups and departments would be able to access the data that is being analyzed and presented. A simple scneario would be: a. All the data from the 'fraud detection' project could only be accessed by the team that works in the anti fraud / AML unit. B. The dashboard analytics that contains the marketing KPIs and near real-time monitoring would have certain 'viewing privileges' depending on one's current role and tasks. • C. The Risk Management data could not be viewed by anyone outside of their department.
	How will we guarantee integrity? (provide at least two ideas)	• Privacy and security workshops would be provided to all employees undergoing all these projects. • There will be a constant weekly audit to make sure that privacy & compliance are being observed.

	How will we guarantee security? (provide at least two ideas)	• Data access of given on a 'per role/ privilege' basis. • Data is behind a firewall. • All users would ALWAYS be required to enter the proper user name and password.
Technology	Identify the components of your Data Architecture	• Data would mostly come from our PostGreSQL databases that contains majority of our financial transactions. We would also utilize the MySQL database that contains all the metrics gathered form the digital marketing team. We also had a dedicated Neo4J graph DB that was started 2 years ago. This is mainly for the Fraud Detection project. Majority of our business intelligence capabilities would be utilizing the Tableau Server that we have been using for the past 6 years.
Skills and Capacity	How will we promote development of data literacy skills and capacity throughout the organization (provide at least three ideas)	• All of the employees involved these projects are already skilled in their respective roles. But this does not limit them from having additional on-line training in the areas of: machine learning, automation, A.I. fundamentals , IT security , risk and compliance. • All employees would have access to on line platforms & learn free. Such platforms are: Udacity, Coursera, DataCamp, TreeHouse, Udemy, EDX and Linux Academy • There will be constant BVIRs (Big Visual Information Radiators), whether they opt to work (1x a week due to COVID) at the office or a virtual version when working remotely (new norm). These BVIRs would have constant reminders embedded within regarding data literacy skills.
Support for Machine Learning	Give a brief description of the machine learning architecture and how it will interface with the data architecture	• We opted for building our data architecture in-house. We have a very robust and highly experienced team not just in data science and analytics, but also our automation specialists have the necessary skill sets.

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.