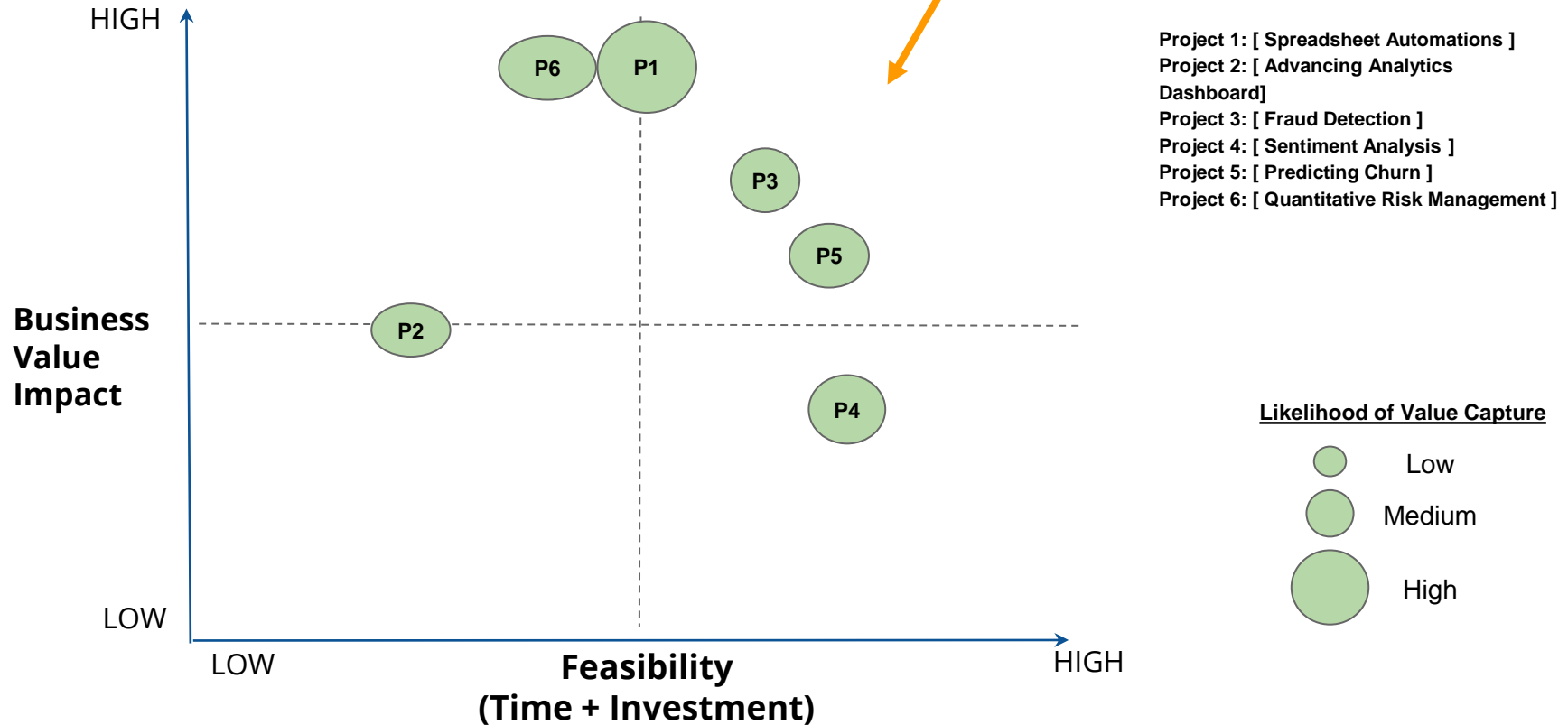


**100**

# **Day Data Science Plan: Building A Data Strategy**

**Step 2, Part 2:** Complete the “Data Science Opportunity Matrix” below by modeling each of the six projects in terms of feasibility (time & investment), business value impact, and likelihood of value capture



# Executive Summary

## Purpose of 100-day plan

- Establishing a data infrastructure that will pave the way for an enterprise 'Analytics Processing Automation' or APA. The over all analytics would utilize a cloud based
- autonomous database

## Approach

- Utilizing a an entire team with multi skill sets. There will be a heavy emphasis on cloud
- infrastructure where all the built in AI and ML engineering would take place. An RPA infrastructure
- platform would be utilized in order to handle the automation tasks

## Results

- A n established Analytics Processing Automation (APA) platform that would be utilized by the
- entire organization

# Scope of Work for First 100 Days

## **1. PURPOSE of 100-DAY PLAN**

- Establish a proper data infrastructure that can be utilized for various projects
- Assigning the proper team members for the specific roles and projects
- Ensure that all stakeholders have a clearer and solid picture of the 6 major projects

## **2. APPROACH**

- I have used a variety of approaches per project. This has always been a case by case basis
- All 3 – Descriptive, Predictive & Prescriptive approaches have been utilized
- I view each of the projects not as separate entities but also have inter-relations with each other. One project's success would eventually benefit the others.

# Candidate Data Science Projects

	Functional Area	Project Description
Project 1: Spreadsheet Automations	Financial Management	Majority of Excel & Google Sheets that involves basic to intermediate financial processing would eventually be automated
Project 2 Name: [Advancing the Analytics Dashboard]	Business Functional Area: [Digital Marketing]	Using advanced ML in order to have deeper key insights embedded into the dashboard for digital marketing campaigns & over-all marketing performance.

# Candidate Data Science Projects

<b>Project 3 Name: [FRAUD DETECTION]</b>	Business Functional Area: [FINANCE – ANTI MONEY LAUNDERING (AML) DIVISION]	<b>Heavy ML usage due to anomalies that occur within the transactions.</b>
<b>Project 4 Name: [SENTIMENT ANALYSIS]</b>	[LEGAL DEPARTMENT]	<b>Heavy analysis on a vast amount of documents. Using AI &amp; ML in order to see certain patterns and predict certain behaviours of clients.</b>

# Candidate Data Science Projects

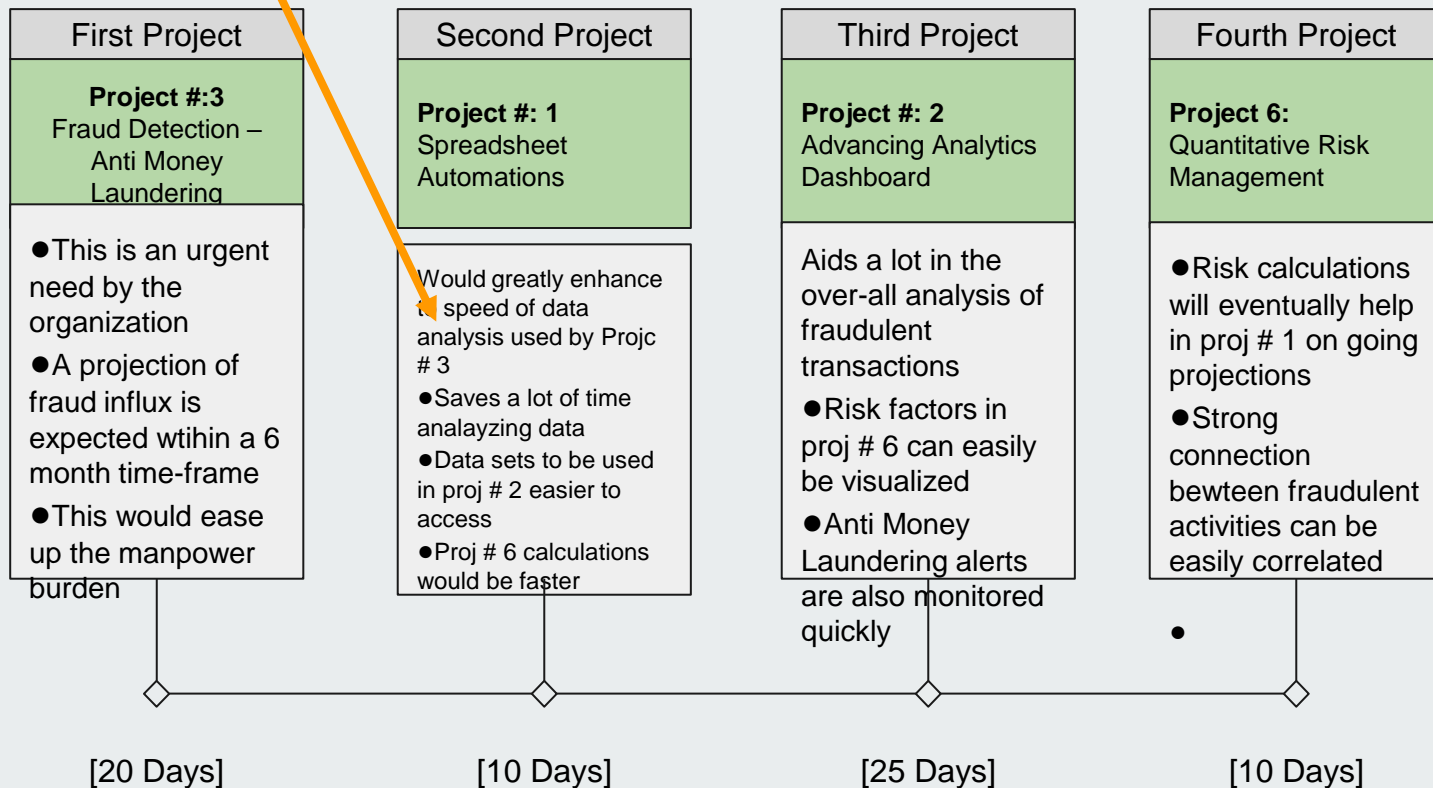
<b>Project 5 Name: [PREDICTING CHURN]</b>	Business Functional Area: [DIGITAL MARKETING]	<b>Reliance on Google Analytics &amp; Google Ads/ Bing Ads / FB Ads are simply not enough. We will use certain ML models in order to maximize the gains.</b>
<b>Project 6 Name: [QUANTITATIVE RISK MANAGEMENT]</b>	[FINANCE]	<b>Building models in order to understand the risks of financial portfolios.</b>

**Step 2, Part 3:** Complete the “Data Science Road Map” below with the first four data science projects chosen for implementation.

<u>Order</u>	<u>Project</u>	<u>Order Justification</u>
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.



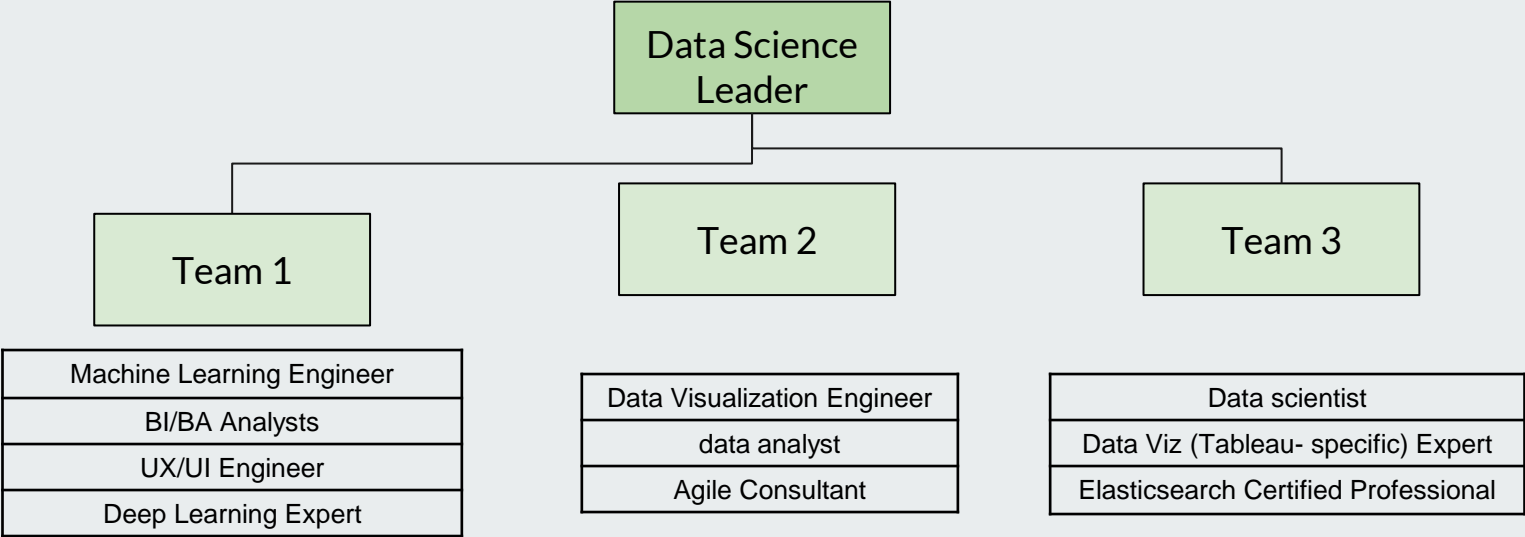
**Step 2, Part 3:** Complete the “Data Science Road Map” below with the first four data science projects chosen for implementation.



# Our Highest-Priority Data Science Projects

Order		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
First	Project 3: [Fraud Detection Anti Money Laundering]	4	3	3	5	5
Second	Project 1: [Spreadsheet Automations]	5	5	3	5	5

# Initial Structure of the Data Science Team



# I have identified six strategies for promoting a data-driven culture in our business

## Strategies for promoting a data-driven culture

	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.

# Technical Infrastructure Needed to Support the Data Science Organization

Data Requirements	What data should be included in the Data Strategy?	<ul style="list-style-type: none"> <li>This is a huge scope. This would include a lot of financial data and a lot of marketing (both digital &amp; traditional). There will also be a certain amount of qualitative data that would be gathered from UX research.</li> </ul>
Data Governance	Data Availability	<ul style="list-style-type: none"> <li>All data that would be utilized is readily &amp; 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.</li> </ul>
	Usability	<ul style="list-style-type: none"> <li>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 scenario 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.</li> <li>. the Risk Management data could not be viewed by anyone outside of their department.</li> </ul>
	Integrity	<ul style="list-style-type: none"> <li>Privacy and security workshops would be provided to all employees undergoing all these projects.</li> <li>There will be a constant weekly audit to make sure that privacy &amp; compliance are being observed.</li> </ul>
	Security	<ul style="list-style-type: none"> <li>Data access of given on a 'per role/ privilege' basis.</li> <li>Data is behind a firewall.</li> <li>All users would ALWAYS be required to enter the proper user name and password.</li> </ul>
Technology	Data Architecture Components	<ul style="list-style-type: none"> <li>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 from 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.</li> </ul>

# Technical Infrastructure Needed to Support the Data Science Organization



Skills and Capacity	Data literacy skills and organizational capacity	<ul style="list-style-type: none"><li>• 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.</li><li>• All employees would have access to on line platforms &amp; learn free. Such platforms are: Udacity, Coursera, DataCamp, TreeHouse, Udemy, EDX and Linux Academy</li><li>• 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.</li></ul>
Support for Machine Learning	Machine learning architecture	<ul style="list-style-type: none"><li>• 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.</li></ul>