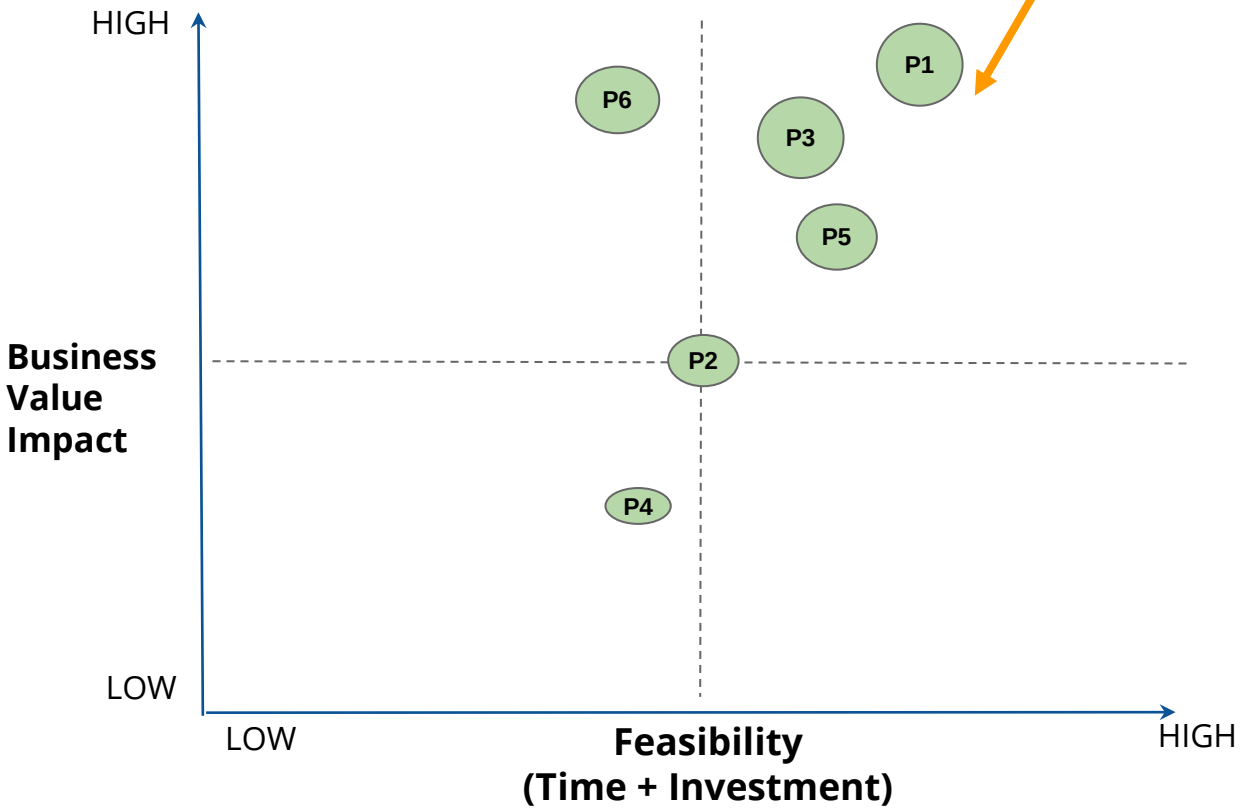

# 100-Day Data Science Plan: Building A Data Strategy

A Udacity Capstone Presentation by: **Frederick Zoreta**

**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




**Project 1:** [Spreadsheet Automations]

**Project 2:** [Advancing Analytics Dashboard]

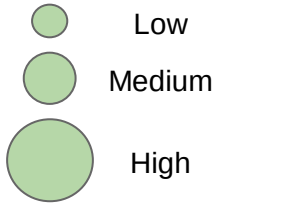
**Project 3:** [Fraud Detection ]

**Project 4:** [Sentiment Analysis ]

**Project 5:** [Predicting Customer Churn]

**Project 6:** [Quantitative Risk Management]

**Likelihood of Value Capture**



Alteryx Corporation

# Establising an Analytics Processing Automation: Merging AI/ML with RPA and a (Cloud based) Autonomous Database

Name : Frederick Zoreta
Position : AI & Data Product Manager , RPA Team Lead

Date: January 14, 2021

# 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

- An 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 infrasrtcuture that can be utilized for various projects
- Assiging the proper team members for the specific roles and projects
- Ensure that all stakeholders have a cleaer 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 bais
- 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.

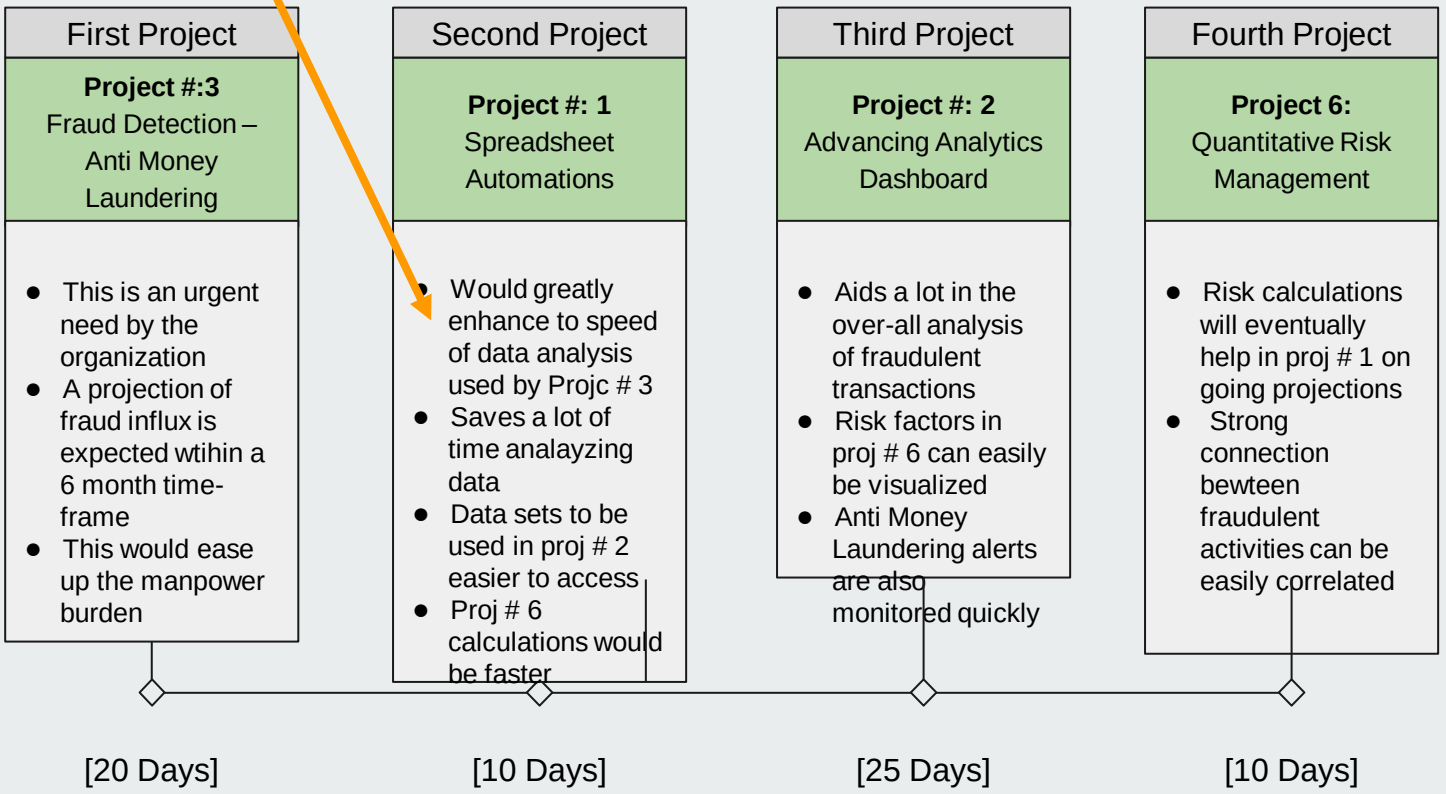
# Candidate Data Science Projects

| | 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: Advancing the Analytics Dashboard** | **Digital Marketing** | **Using advanced ML in order to have deeper key insights embedded into the dashboard for digital marketing campaigns & over-all marketing performance.** |
| **Project 3: Fraud Detection** | **Finance – Anti Money Laundering (AML)** | **Heavy ML usage due to anomalies that occur within the transactions.** |
| **Project 4: Sentiment Analysis** | **Legal Department** | **Heavy analysis on a vast amoung of documents. Using AI & ML in order to see certain patterns and predict certain behaviours of clients.** |
| **Project 5: Predicting Churn** | **Digital Marketing** | **Reliance on Google Analytics & Google Ads/ Bing Ads / FB Ads are simply not enough. We will use certain ML models in order to maximize the gains.** |
| **Project 6: Quantitative Risk Management** | **Finance** | **Building models in order to understand the risks of financial portfolios.** |

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

| Order | Project | Order Justification |
|---|---|---|
| 1 | **Project 3:** Fraud Detection – Anti Money Laundering | **There has been an unrgent need from the organization due to the rising occurences of financial fraud. The entire C-suite and even the IT Security team is concerned that a potential AML scheme may target our financial system.** |
| 2 | **Project 1:** 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 | **Project 2:** Advancing the 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. They also want the visualizations to be interactive with geospatial mapping.** |
| 4 | **Project 6:** Quantitative Risk Management | **Building ML models to analyze several portfolios is a complex task. Although this would take up manpower and would fully utilze the entire specializations of the team, the urgency is not that crucial.** |

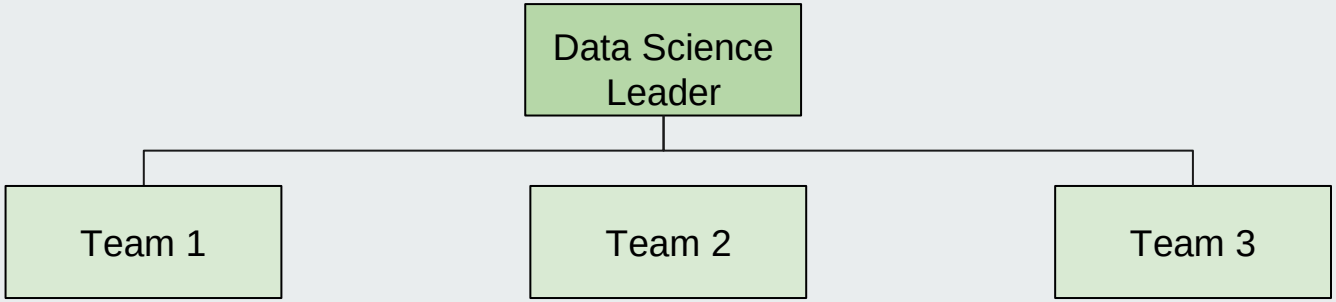
**Step 2, Part 3:** Complete the "Data Science Road Map" below with the first four data science projects chosen for implementation.



# 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




- 3 Machine Learning Engineers
- 1 Deep Learning Specialist
- 4 Data Scientists
- 2 Business Intelligence Analysts
- 1 Data Product Manager
- 2 UX/UI Engineers

- 1 Machine Learning Engineer
- 2 Data Visualization Specialists
- 3 Business Analysts
- 1 Growth Product Manager
- 2 Data Scientists
- 3 BI Specialist
- 3 Agile Specialist

- 2 Tableu Server Professionals
- 1 Elasticsearch Platform Specialist
- 2 UIPath Certified Professionals
- 1 Project Manager
- 1 UX Researcher

# 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 decipering if 1 or 2 methods are enough to arrive at a solution

Strategy 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.

Strategy 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 ogranization much better.

Strategy 4: Maximizing the skills and talents of our employees and partner consultants, before looking externally for a potential consultant

Strategy 5: Maximizing the current infrastructure we have; both on premise & cloud based.

Strategy 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? | • 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 | Data Availability | • 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. |
| | Usability | • 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.<br>A simple scneario would be:<br>a. All the data from the 'fraud detection' project could only be accessed by the team that works in the anti fraud / AML unit.<br>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.<br>• C. The Risk Management data could not be viewed by anyone outside of their department. |
| | Integrity | • Privacy and security workshops would be provided to all employees undergoing all these projects.<br>• There will be a constant weekly audit to make sure that privacy & compliance are being observed. |
| | Security | • Data access of given on a 'per role/ privilege' basis.<br>• Data is behind a firewall.<br>• All users would ALWAYS be required to enter the proepr user name and password. |
| Technology | Data Architecture Components | • 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.<br><br>Majority of our business intelligence capabilties would be utilizing the Tableau Server that we have been using for the past 6 years. |