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:** Tom Tech Solutions

**Brief Company Description:** Tom Tech Solutions is a growing technology company that provides innovative software solutions for e-commerce businesses. Founded in 2015, the company has developed a suite of tools aimed at optimizing the supply chain, improving customer service, and driving sales growth for online retailers. With a team of 150 employees, Tom Tech Solutions operates primarily in the North American and European markets.

Despite its success, the company has yet to fully leverage the power of data science to improve its products and services. Data is collected from various touchpoints, including customer behavior, sales transactions, and supply chain operations, but it remains siloed and underutilized. As part of its strategy for scaling, Tom Tech Solutions is focusing on implementing data-driven decision-making across key functions, including marketing, customer experience, and operations.

# 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:** Customer Segmentation for Targeted Marketing [Marketing]

**Project 2:** Predictive Analytics for Sales Forecasting [Marketing]

**Project 3:** Supply Chain Optimization Using Predictive Maintenance [Supply Chain]

**Project 4:** Inventory Optimization with Demand Forecasting [Supply Chain]

**Project 5:** Credit Risk Modeling for Customer Financial Health [Finance]

**Project 6:** Fraud Detection System [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: Customer Segmentation for Targeted Marketing**

**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: Marketing teams often struggle to efficiently target the right customer segments for their campaigns, leading to wasted resources and ineffective outreach. Without customer segmentation, marketing efforts can be too generalized and not aligned with customer needs.

- Role of data science in addressing the business problem: Data science will enable the use of clustering algorithms (e.g., k-means) to group customers based on their behaviors, preferences, and demographics. This will help identify distinct customer segments, allowing for tailored marketing efforts. By analyzing purchasing patterns, browsing habits, and demographic information, we can predict and identify high-value customers.

- Targeted business objective(s):

* **Customer Acquisition:** By targeting the right segments, the company can attract more customers who are most likely to purchase.
* **Customer Retention:** Tailored campaigns will enhance the experience and loyalty of existing customers.
* **Revenue:** Personalized campaigns are likely to lead to increased conversion rates, boosting revenue.

**2. Data Science Classification**

- Approach: Supervised learning (classification) for identifying customer segments based on known behaviors, and unsupervised learning (clustering) for grouping customers based on similarity.

- Type of Model:

* Clustering models such as **K-means clustering** or **DBSCAN** for segmentation.
* Classification models like **logistic regression** or **decision trees** for predicting customer conversion based on segmentation.

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

**- Demographic data:** Age, gender, income level, education, etc.

**- Transactional data:** Purchase history, frequency, amount spent, time of purchase.

**- Behavioral data:** Browsing patterns, website interaction logs, time spent on site, product preferences.

**- Sources of Data:**

* Customer databases (CRM system, sales platform).
* Web analytics tools (Google Analytics, Adobe Analytics).
* Social media and other marketing platforms.

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

**Magnitude:** Medium

* **Justification:** This project can significantly improve marketing ROI by targeting the right customer segments, but the impact will vary depending on the market and how well the segments align with existing campaigns. The potential is high for expanding customer bases and increasing conversions, but it's not guaranteed without proper model tuning and data.

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

**- Cost:** Medium

* **Justification:** The cost of implementation includes data collection, model training, and testing. However, it can be managed by leveraging existing marketing data and using open-source machine learning tools.

- **Complexity:** Medium

* **Justification:** While the project is not highly complex, it will require a solid understanding of data science techniques for segmentation, data preprocessing, and the ability to integrate the model into marketing platforms.

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

**Likelihood of Value Capture:** High

* **Justification:** Given the importance of targeted marketing and the proven success of segmentation strategies, there is a high likelihood that personalized campaigns will lead to increased engagement and higher conversion rates. The business has existing data that can be leveraged, and the model can be fine-tuned for continuous improvement.

**7. Key Business Stakeholders**

**C-level Executives:**

* + **Chief Marketing Officer (CMO):**

Responsible for ensuring that the customer segmentation aligns with the company’s marketing strategy and goals. The CMO's involvement will help ensure that the data science initiative is closely aligned with the company's broader marketing objectives.

* + **Chief Data Officer (CDO):**

Ensures that the project leverages company-wide data resources, aligns with data governance standards, and utilizes the right data infrastructure to support the segmentation process.

* + **Chief Executive Officer (CEO):**

Provides overall strategic direction and ensures that the customer segmentation project aligns with the company's long-term business goals, including customer acquisition, retention, and revenue growth.

**Functional Leaders:**

* **Marketing Manager:**

Oversees the project and ensures it aligns with the marketing department's goals. The manager is responsible for driving the implementation of the customer segmentation insights into marketing campaigns.

* **Sales Team Leader:**

Ensures that the segmentation results are integrated into sales strategies, helping the sales team focus on high-value customer segments that are most likely to convert.

* **Data Science Team Lead:**

Leads the development, training, and implementation of the segmentation models, ensuring the accuracy and relevance of customer segments.

* **IT Department Lead:**

Provides support for integrating the model with existing marketing platforms and databases. The IT department ensures that technical systems are in place to capture and process the data required for the project.

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

**Project 2 Name: Predictive Analytics for Sales Forecasting**

**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:  
  The company faces challenges in predicting future sales with accuracy, leading to poor inventory planning, missed revenue opportunities, and inefficient marketing spend. The absence of reliable sales forecasting also limits the company's ability to strategize effectively for promotions and product launches.
* Role of data science in addressing the business problem:  
  Data science will enable the development of predictive models based on historical sales data, seasonality, market trends, and external factors such as promotions or events. By using machine learning techniques such as time series analysis or regression models, the system will predict future sales for different products, regions, and time periods, leading to improved planning and resource allocation.
* Targeted business objective(s):
* Revenue Growth: Accurate sales forecasts will allow for better alignment of resources and product availability, driving higher revenue.
* Cost Reduction: Efficient allocation of resources based on accurate forecasts will reduce overstocking and understocking costs, leading to better inventory management.
* Customer Acquisition: More effective marketing campaigns can be designed using insights from predicted sales, leading to increased customer acquisition.

1. **Data Science Classification**

* Approach: Supervised learning (time series forecasting, regression).
* Type of Model:
* Time series models like ARIMA or Prophet for forecasting.
* Regression models like Linear Regression or Random Forest to predict sales based on various factors.

1. **Data needed for the project and sources for that data**

* **Data Needed:**
* Historical Sales Data: Sales numbers, sales by region, product type, etc.
* External Data: Seasonal trends, market data, and economic indicators that can affect sales.
* Promotional Data: Information about past marketing campaigns and their effect on sales.
* Customer Demographics: Data on customer preferences and purchasing behavior.
* **Sources of Data:**
* Internal sales records (ERP system, CRM software).
* External market research data (industry reports, competitor performance).
* Marketing platforms (Google Ads, email campaigns data).

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

* **Magnitude**: High
* **Justification**: Accurate sales forecasting can significantly improve the business's bottom line by optimizing inventory management and marketing spend. Given the wide-reaching impact across various departments (marketing, sales, inventory management), the opportunity is large.

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

* Cost: Medium
* Justification: While the development of predictive models requires upfront investment in data collection, preparation, and model training, these costs can be mitigated by leveraging existing data and predictive modeling tools.
* Complexity: Medium
* Justification: Time series forecasting models and regression models are widely understood and used but still require expertise in data preprocessing, feature engineering, and model tuning to ensure accuracy.

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

* Likelihood of Value Capture: High
* Justification: Given the significant impact on revenue growth, cost reduction, and resource optimization, the likelihood of capturing value is high. The company can quickly translate accurate forecasts into actionable marketing and sales strategies, driving tangible improvements in operational efficiency.

1. **Key Business Stakeholders**

**C-level Executives:**

* **Chief Marketing Officer (CMO):**

Ensures that the predictive analytics project aligns with the company's marketing strategy, particularly for planning campaigns, product launches, and promotions based on accurate sales forecasts.

* **Chief Data Officer (CDO):**

Oversees the integration of data science models and ensures the data governance structure is in place to support accurate predictions and resource allocation.

* **Chief Executive Officer (CEO):**

Provides strategic guidance to ensure the project is aligned with the company’s long-term objectives, particularly for revenue growth and cost optimization through improved sales forecasting.

**Functional Leaders:**

* **Marketing Manager:**

Oversees how sales forecasts will be integrated into the planning of marketing campaigns and product promotions.

* **Sales Team Leader:**

Works closely with the marketing and inventory teams to align sales strategies based on forecasted demand.

* **Inventory/Operations Manager:**

Leverages the predictive model to adjust inventory levels and reduce costs associated with overstocking or stockouts.

* **Data Science Team Lead:**

Responsible for developing, training, and fine-tuning predictive models, ensuring their accuracy and relevance to sales forecasting.

* **CFO/Finance Leader:**

Ensures that the sales forecasting model aligns with financial projections and helps optimize budgeting and cost management across departments.

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

**Project 3 Name: Supply Chain Optimization Using Predictive Maintenance**

**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: The company faces unplanned downtimes and equipment failures that disrupt supply chain operations, leading to delays in production and distribution. These disruptions affect the overall supply chain efficiency, increase maintenance costs, and reduce product availability.

- Role of data science in addressing the business problem: Data science will leverage machine learning models to predict equipment failures before they occur. By analyzing historical maintenance data, sensor data, and environmental factors, predictive maintenance models can forecast when a piece of equipment is likely to fail. This allows for proactive maintenance, minimizing downtime and optimizing the supply chain.

- Targeted business objective(s):

* **Cost Reduction:** Predictive maintenance can reduce the costs associated with unplanned downtimes, expensive emergency repairs, and inefficient maintenance schedules.
* **Operational Efficiency:** By optimizing maintenance schedules, the company can ensure smooth supply chain operations, leading to faster production and distribution cycles.
* **Revenue Growth:** Increased operational uptime and reduced disruptions will result in higher throughput, directly contributing to revenue growth.

**2. Data Science Classification**

- Approach: Supervised learning (classification, regression).

- Type of Model:

* **Predictive models**: Using classification algorithms (e.g., Decision Trees, Random Forest) or regression models (e.g., Linear Regression, Support Vector Machines) to predict when equipment will fail based on historical data.
* **Time Series Forecasting**: To anticipate failure patterns based on temporal data from sensors.

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

* **Data Needed:**
* **Maintenance Logs:** Historical maintenance data, including dates, types of repairs, and failure reports.
* **Sensor Data:** Data from equipment sensors, including temperature, pressure, vibration, and other indicators of potential failure.
* **Operational Data:** Production and supply chain throughput data to understand how equipment failures affect overall operations.
* **Environmental Data:** Information on operating conditions such as temperature, humidity, and external factors that might influence equipment performance.
* **Sources of Data:**
* **Internal Data:** Data collected from maintenance logs and IoT sensors attached to the equipment.
* **External Data:** Environmental data can come from weather services, and operational data can come from ERP or manufacturing execution systems.

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

**Magnitude: High**

* **Justification:** Predictive maintenance can greatly reduce downtime, which directly impacts operational costs and throughput. The supply chain department will benefit from improved operational efficiency and reduced maintenance costs, thus justifying a high magnitude of opportunity.

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

* **Cost:** Medium
* **Justification:** While the project requires investment in IoT sensors, historical data analysis, and model development, the costs are mitigated by the potential long-term savings through optimized maintenance schedules and reduced downtime.
* **Complexity:** Medium to High
* **Justification:** Developing predictive models for maintenance involves working with complex sensor data, and the models must be integrated into existing maintenance processes and supply chain operations. This requires collaboration across multiple departments and expertise in data science, IoT, and supply chain management.

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

**Likelihood of Value Capture:** High

* **Justification:** By reducing equipment failures and downtime, predictive maintenance has a high likelihood of delivering value. The supply chain's operational efficiency will improve, leading to cost savings and increased throughput, making this project a high-value opportunity.

**7. Key Business Stakeholders**

**C-level Executives:**

* **Chief Operations Officer (COO):**

Ensures that the predictive maintenance project aligns with the company’s operational goals, particularly in improving supply chain efficiency, reducing downtime, and optimizing production.

* **Chief Financial Officer (CFO):**

Oversees the financial impact of the predictive maintenance project, tracking cost reductions from reduced downtime, lower emergency repair costs, and increased revenue from operational efficiency.

* **Chief Technology Officer (CTO):**

Supports the implementation of IoT sensors and integration of advanced machine learning models into the supply chain operations. Provides leadership on the technical aspects of the project.

**Functional Leaders:**

* **Supply Chain Manager:**

Ensures that predictive maintenance models align with supply chain operations, minimizing disruptions and improving throughput.

* **Operations Manager:**

Integrates predictive maintenance predictions into daily operations, ensuring equipment maintenance is carried out at optimal times to avoid disruption.

* **Maintenance Team Lead:**

Leverages predictive maintenance data to optimize maintenance scheduling, reducing downtime and improving equipment reliability.

* **Data Science Team Lead:**

Develops and fine-tunes predictive models and ensures their integration with existing operational systems and processes.

* **CFO/Finance Team Lead:**

Tracks the financial benefits of predictive maintenance, including cost savings from fewer unplanned repairs and downtime.

**Justification for C-Level Executive Involvement:**

* **Strategic Alignment:** C-level executives, particularly the COO and CFO, play a vital role in ensuring the project aligns with broader business goals—improving operational efficiency and profitability. Their buy-in is essential for cross-departmental cooperation and resource allocation.
* **Resource Allocation and Risk Management:** The CTO and COO help allocate the necessary resources for implementing IoT sensors, machine learning models, and integration into operational systems. Their involvement also helps manage the risks associated with technology adoption and implementation.
* **Organizational Change:** Predictive maintenance will involve process changes, including the optimization of maintenance schedules and production workflows. The support of C-level executives helps facilitate these changes across departments, ensuring smooth adoption and effective utilization of the models.

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

**Project 4 Name: Inventory Optimization with Demand Forecasting**

**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: Inefficient inventory management leads to either stockouts (loss of sales) or overstocking (increased storage costs). The company struggles with maintaining the right balance of inventory, especially for products with fluctuating demand. This results in missed sales opportunities and unnecessary storage costs.

- Role of data science in addressing the business problem: Data science will use predictive analytics and machine learning to forecast product demand at a granular level (e.g., per product, per store, per region). By analyzing historical sales data, seasonal trends, promotions, and external factors (e.g., holidays or weather), the models will provide more accurate demand forecasts. This will enable optimized inventory levels, reducing the risk of stockouts or overstocking.

- Targeted business objective(s):

* **Cost Reduction:** Minimizing excess inventory and storage costs while reducing the likelihood of stockouts, which can result in lost sales.
* **Revenue Growth:** Ensuring optimal stock levels improves product availability, increasing sales and customer satisfaction.
* **Operational Efficiency:** Streamlining supply chain processes to match inventory with actual demand, reducing operational inefficiencies.

**2. Data Science Classification**

- Approach: Supervised learning (regression, time series analysis).

- Type of Model:

* **Forecasting models:** Using time series analysis (e.g., ARIMA, Prophet) and machine learning models (e.g., Random Forest, XGBoost) to predict future demand based on historical data.
* **Optimization models:** Using mathematical programming or simulation techniques to determine the ideal stock levels based on forecasted demand.

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

* **Data Needed:**
* **Historical Sales Data:** Past sales data, including units sold, sales trends, and seasonality patterns.
* **Inventory Data:** Current and historical inventory levels, including stock turnover rates.
* **Product Data:** Information about product characteristics, pricing, and promotions.
* **External Data:** Factors like weather, holidays, and regional events that may impact product demand.
* **Sources of Data:**
* **Internal Data:** Sales and inventory management systems (e.g., ERP systems, point-of-sale data).
* **External Data:** Public sources for weather and event data, or data from third-party APIs for promotions and holidays.

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

**Magnitude:** High

* **Justification:** Efficient inventory management is crucial for a company’s bottom line. By optimizing inventory levels and improving demand forecasting accuracy, the company can significantly reduce costs associated with overstocking and stockouts while improving sales and customer satisfaction. The magnitude of this opportunity is high due to its direct impact on both operational costs and revenue.

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

* **Cost:** Medium
* **Justification:** The initial investment in technology and data infrastructure (e.g., upgrading forecasting software, collecting external data) can be moderate. However, the potential for cost savings and revenue increase justifies the investment.
* **Complexity:** Medium to High
* **Justification:** The complexity arises from integrating various data sources, cleaning and processing large datasets, and ensuring that the demand forecasting models account for a range of variables, including external factors. Additionally, the models must be integrated with inventory management systems to trigger automated inventory adjustments.

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

**Likelihood of Value Capture:** High

* **Justification:** Given that inventory management is directly tied to both revenue and cost, the likelihood of capturing value is high. Accurate demand forecasting will improve product availability and reduce operational inefficiencies, leading to cost savings and increased sales.

**7. Key Business Stakeholders**

**C-level Executives:**

* **Chief Supply Chain Officer (CSCO):**

Ensures that inventory management strategies align with the company’s supply chain goals, including reducing costs and improving product availability. The CSCO plays a key role in ensuring the integration of forecasting models with supply chain processes.

* **Chief Financial Officer (CFO):**

Oversees the financial impact of optimized inventory management, tracking cost savings from reduced overstocking and stockouts, as well as the increased revenue from improved product availability.

* **Chief Technology Officer (CTO):**

Supports the technical implementation of machine learning models and integration with existing ERP and inventory management systems. Ensures that the necessary infrastructure is in place for collecting and processing data.

**Functional Leaders:**

* **Supply Chain Manager:**

Leads the implementation of demand forecasting models in inventory management to reduce stockouts and overstocking, ensuring smoother operations.

* **Sales and Marketing Managers:**

Utilize demand forecasts to adjust sales strategies and promotions to match anticipated demand, improving customer satisfaction and increasing revenue.

* **Operations Manager:**

Coordinates the procurement and manufacturing schedules based on the demand forecasts, ensuring production is aligned with demand.

* **Finance Manager:**

Evaluates the financial outcomes of improved inventory levels, helping the business understand the savings from optimized stock levels and the growth from increased sales.

* **IT/Technology Manager:**

Works on the integration of forecasting models with the company’s existing systems, ensuring the necessary data flows smoothly and the models are updated as needed.

**Justification for C-level Executive Involvement:**

* **Strategic Alignment:** C-level executives, particularly the CSCO and CFO, are instrumental in aligning inventory optimization strategies with broader business goals. Their support ensures that demand forecasting is prioritized and integrated across the company.
* **Resource Allocation and Risk Management:** The CTO provides the necessary leadership in managing the technological resources, including machine learning infrastructure, data systems, and ensuring the seamless integration of forecasting models into the supply chain and inventory management systems.
* **Organizational Impact:** Optimizing inventory levels impacts multiple departments, including sales, operations, and finance. Involving C-level executives ensures alignment across these departments and facilitates the cross-functional collaboration necessary to fully realize the benefits of this project.

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

**Project 5 Name: Credit Risk Modeling for Customer Financial Health**

**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: The business faces challenges in accurately assessing the credit risk of potential customers. Without a robust credit risk assessment model, the company risks approving loans or credit to customers who may default, leading to significant financial losses. Additionally, the business may reject customers who are creditworthy, missing out on potential revenue.

- Role of data science in addressing the business problem: Data science will leverage historical financial data, customer behavior data, and external data sources to build predictive models that assess the likelihood of a customer defaulting on a loan. By using machine learning algorithms (e.g., logistic regression, decision trees, or ensemble methods), the company can better predict customer financial health and optimize loan approval decisions.

- Targeted business objective(s):

* **Revenue Growth:** By approving more qualified customers and rejecting higher-risk ones, the business can increase loan issuance and revenue.
* **Cost Reduction:** Reducing the risk of loan defaults lowers the potential for financial losses.
* **Customer Retention:** Ensuring that credit is extended to those who can manage it will improve customer satisfaction and long-term relationships.

**2. Data Science Classification**

- Approach: Supervised learning (classification).

- Type of Model:

* **Risk Prediction Model:** Using classification algorithms (e.g., logistic regression, decision trees, random forests) to predict the likelihood of default.
* **Feature Engineering:** Using financial history, demographic data, transaction behavior, and external credit scores as features in the model to improve accuracy.

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

* **Data Needed:**
* **Customer Financial Data:** Income, debt-to-income ratio, past loan history, payment patterns, and credit scores.
* **Behavioral Data:** Transaction history, spending habits, account balances, and payment timeliness.
* **External Data:** Credit scores from external credit bureaus (e.g., FICO, Equifax) and macroeconomic indicators (e.g., unemployment rates, interest rates).
* **Sources of Data:**
* **Internal Data:** Customer financial data from internal databases, such as CRM systems, transaction logs, and loan application forms.
* **External Data:** Data from third-party credit bureaus, economic reports, and public data sources.

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

**Magnitude:** High

* **Justification:** The ability to accurately assess and manage credit risk can significantly impact both revenue and costs. By improving the accuracy of loan approvals, the company can reduce the number of defaults and increase the approval rate for financially viable customers, leading to greater revenue. The potential to mitigate financial losses due to defaults makes this a high-impact project.

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

* **Cost:** Medium
* **Justification:** The initial costs involve gathering data, ensuring its quality, and building and testing predictive models. However, the project provides long-term cost-saving benefits by reducing defaults.
* **Complexity:** High
* **Justification:** The complexity comes from integrating various data sources, ensuring data accuracy, and developing a model that can effectively distinguish between low-risk and high-risk customers. The model will need to account for various factors, including economic conditions, customer behavior, and external data sources.

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

* **Likelihood of Value Capture:** High
  + **Justification:** Given the direct impact on loan profitability and customer management, the likelihood of capturing value is high. A well-built credit risk model can significantly reduce the risk of default and improve customer targeting, providing substantial financial returns.

**7. Key Business Stakeholders**

* **Finance Team:** To oversee the implementation of the credit risk model and integrate it with loan approval processes.
* **Risk Management Team:** To evaluate the model's accuracy and adjust risk thresholds based on real-time results.
* **IT/Technology Team:** To support the integration of the model into existing systems and ensure data availability.
* **Customer Service Team:** To assist in communication with customers regarding credit approvals or rejections.
* **Regulatory and Compliance Team:** To ensure that the credit risk model adheres to financial regulations and compliance standards.

**Key Executive Stakeholders:**

* **CEO (Chief Executive Officer)**: Responsible for ensuring that the credit risk modeling aligns with the company's overall business strategy and objectives.
* **CFO (Chief Financial Officer)**: Oversees the financial impact of the project, including cost-benefit analysis and resource allocation.
* **CRO (Chief Risk Officer)**: Leads the risk management efforts, ensuring that the model reduces the company’s exposure to default-related financial losses.
* **CIO (Chief Information Officer)**: Ensures the IT infrastructure supports the implementation and scalability of the credit risk model.
* **CMO (Chief Marketing Officer)**: Works to ensure that the model’s outputs align with customer acquisition and retention strategies.

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

**Project 6 Name: Fraud Detection System**

**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: Fraud is a significant threat to the business, leading to financial losses and damage to customer trust. The company needs an automated system to detect fraudulent transactions in real-time and reduce the risk of fraudulent activity that could impact both the company’s bottom line and reputation.

- Role of data science in addressing the business problem: Data science will leverage transaction data, customer behavior patterns, and historical fraud data to develop predictive models for fraud detection. By using machine learning algorithms (e.g., anomaly detection, supervised learning techniques), the system can flag suspicious activities and alert relevant teams. This will enable faster responses and prevent fraudulent transactions from being processed.

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

* **Cost Reduction:** Reducing financial losses from fraud.
* **Revenue Protection:** Ensuring fraudulent activities are prevented, which can otherwise erode profits.
* **Customer Retention:** Maintaining customer trust by providing a secure transaction environment.

**2. Data Science Classification**

- Approach: Supervised learning (classification) and unsupervised learning (anomaly detection).

- Type of Model:

* **Fraud Detection Model:** Classification algorithms like decision trees, logistic regression, or neural networks can classify transactions as legitimate or fraudulent.
* **Anomaly Detection Model:** Unsupervised algorithms (e.g., k-means, DBSCAN, or autoencoders) can identify patterns in transaction behavior that deviate from normal activity.

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

* **Data Needed:**
* **Transaction Data:** Date, time, amount, merchant information, payment methods, location, and transaction type.
* **Customer Data:** Account details, historical transaction patterns, and customer profile information.
* **External Data:** Blacklists, known fraudulent IP addresses, or data from financial authorities on known fraud schemes.
* **Sources of Data:**
* **Internal Data:** Transaction logs from payment systems, customer transaction histories, and behavioral data from customer accounts.
* **External Data:** Data from third-party fraud detection agencies, credit bureaus, and fraud databases.

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

**Magnitude:** High

* **Justification:** Fraudulent transactions can result in significant losses. By identifying and preventing fraud at an early stage, the company can save a considerable amount of money, protect its reputation, and retain customer trust. This project has a high potential to positively impact the company’s bottom line by significantly reducing fraud-related losses.

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

* **Cost:** Medium
* **Justification:** The costs mainly involve building the fraud detection model, integrating it into existing systems, and maintaining it. While the upfront investment is moderate, the long-term cost savings from reduced fraud will offset the initial development costs.
* **Complexity:** High
* **Justification:** The complexity lies in handling large volumes of transaction data, ensuring the model can detect fraud in real-time, and differentiating between legitimate anomalies and actual fraud. The model will need to be highly accurate to avoid false positives (legitimate transactions flagged as fraud) and false negatives (fraudulent transactions going undetected).

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

- **Likelihood of Value Capture:** High

* + **Justification:** The potential for value capture is high as the company can significantly reduce fraud-related financial losses. Moreover, the business will benefit from improved customer trust and satisfaction, which can translate into better customer retention and a stronger market position. The ability to detect fraud early also reduces the overall impact on operations.



**7. Key Business Stakeholders**

* **Finance Team:** To assess the financial impact of fraud and prioritize which transactions to monitor closely.
* **Fraud Prevention/Compliance Team:** To define what constitutes fraudulent behavior and ensure the model is aligned with regulatory requirements.
* **IT/Technology Team:** To support integration with payment systems and ensure the model runs efficiently in real-time.
* **Customer Service Team:** To manage customer inquiries regarding potential fraud alerts or blocked transactions.
* **Legal and Compliance Team:** To ensure that fraud detection practices comply with industry regulations (e.g., GDPR, PCI DSS).

**Key Executive Stakeholders:**

* **CEO (Chief Executive Officer)**: To ensure the fraud detection system aligns with the company's strategic goals and enhances overall business security.
* **CFO (Chief Financial Officer)**: Responsible for understanding the financial impact of fraud on the company's bottom line and justifying the investment in the system.
* **CRO (Chief Risk Officer)**: Oversees the company’s risk management efforts and ensures that the fraud detection system effectively mitigates financial risks.
* **CIO (Chief Information Officer)**: Ensures the infrastructure supports real-time data processing and the smooth integration of the system.
* **CMO (Chief Marketing Officer)**: Works to ensure that customer trust is maintained through effective fraud prevention and transparent communication.

# 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: Customer Segmentation for Targeted Marketing** | 5 | 3 | 3 | 4 | 4 |
| **Project 2: Predictive Analytics for Sales Forecasting** | 5 | 4 | 3 | 4 | 5 |
| **Project 3: Supply Chain Optimization Using Predictive Maintenance** | 4 | 2 | 4 | 4 | 5 |
| **Project 4: Inventory Optimization with Demand Forecasting** | 5 | 3 | 3 | 4 | 4 |
| **Project 5: Credit Risk Modeling for Customer Financial Health** | 5 | 3 | 4 | 5 | 5 |
| **Project 6: Fraud Detection System** | 5 | 2 | 5 | 5 | 5 |

**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 | [Predictive Analytics for Sales Forecasting] | This project aligns with immediate strategic goals of improving financial forecasting, resource allocation, and sales optimization. It’s relatively easy to implement and will provide quick wins that build momentum for further data science initiatives. |
| 2 | [Credit Risk Modeling for Customer Financial Health] | Following sales forecasting, this project will leverage financial data insights gained in the first project to enhance risk management and credit strategies. Its implementation complexity is higher, so it makes sense to tackle it after a foundation of predictive analytics. |
| 3 | [Inventory Optimization with Demand Forecasting] | After improving forecasting and credit risk models, optimizing inventory will leverage predictive insights to reduce operational costs. The experience gained from the first two projects can inform better data modeling for inventory management. |
| 4 | [Customer Segmentation for Targeted Marketing] | This project will use insights from the previous three, particularly sales forecasting and inventory optimization, to enhance marketing efforts. It builds on existing models and enhances customer engagement, which becomes more effective after having accurate predictive models in place. |

**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 Data Science Team

**Justification:** The **Centralized Data Science Team** is the most suitable organizational model for Tom Tech Solutions as it offers the following benefits aligned with the scope and needs of the projects in the roadmap:

1. **Alignment with Strategic Goals**: A centralized data science team can ensure that all data science efforts are aligned with the company’s strategic priorities and business goals. As the business is starting its data science journey, having a central hub allows for clear direction and consistent methodology across all projects.
2. **Expertise Consolidation**: The scope of the projects in the roadmap (such as predictive analytics, credit risk modeling, and customer segmentation) requires specialized expertise in machine learning, statistics, and business analysis. Centralizing the data science function ensures that the right talent is focused on these tasks and can collaborate effectively across projects.
3. **Resource Efficiency**: Given the relatively high complexity of the projects (e.g., fraud detection and supply chain optimization), a centralized model will ensure that resources (both human and computational) are efficiently utilized across all projects, without duplication of efforts or resources spread too thin across multiple business functions.
4. **Collaboration with Other Teams**: A centralized team can closely collaborate with marketing, finance, and supply chain departments to better understand their needs and deliver targeted solutions. With a central team, communication between departments is streamlined, ensuring that data science insights are effectively integrated into the decision-making process.
5. **Project Timing and Scalability**: For the first phase of the roadmap, which requires foundational work on predictive models and business insights, a centralized model will help keep the focus on delivering results quickly. As the company’s data science initiatives mature and the demand for data-driven insights grows, the team can evolve towards a **hybrid or decentralized model** to support specific departmental needs, particularly in marketing and finance.

**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 | Data Scientist | 1 |
| 2 | Data Scientist | 1 |
| 3 | Data Scientist | 2 |
| 4 | Data Scientist | 2 |
| 5 | Data Engineer | 1 |
| 6 | Machine Learning Engineer | 2 |
| 7 | Business Analyst | 1 |
| 8 | Data Architect | 1 |
| 9 | Project Manager | 1 |
| 10 | Data Analyst | 2 |

**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 (Team 1) | This role is the most urgent as it will directly contribute to the development and execution of foundational data science projects like customer segmentation and sales forecasting. The first data scientist is critical for initiating model development and analytics. |
| 2 | Machine Learning Engineer (Team 2) | Machine learning expertise is required to handle advanced predictive models, especially for projects like predictive maintenance and credit risk modeling. This hire will enable scalability and deployment of complex models into production. |
| 3 | Data Engineer (Team 1) | With the need for efficient data pipelines, data integration, and system scalability, this hire is necessary to ensure the data infrastructure can support model development. The data engineer will manage and preprocess large datasets for analysis. |
| 4 | Business Analyst (Team 1) | A business analyst will ensure that the data science efforts align with the company’s strategic goals and translate business needs into actionable insights for the data science team. This role will bridge the gap between data science and business stakeholders. |

**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 | Data Literacy Training for All Employees |
| 2 | Centralized Data Repository |
| 3 | Establish Data Science Champions in Every Department |
| 4 | Data-Driven Decision-Making Framework |
| 5 | Agile Data Science Projects |
| 6 | Leadership Data Engagement |

**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? | |  | | --- | | - **Customer Data**: Demographics, behavior, and preferences. |  |  |  |  | | --- | --- | --- | |  |  | - **Sales Data**: Transaction details, revenue, and product performance. |  |  |  |  | | --- | --- | --- | |  |  | - **Operational Data**: Inventory levels, supply chain data, production schedules. |  |  |  |  | | --- | --- | --- | |  |  | - **Marketing Data**: Campaign results, customer acquisition costs, and advertising spend. |  |  |  |  | | --- | --- | --- | |  |  | - **Financial Data**: Profit/loss, revenue, expenses, forecasts, and budgeting. |  |  |  |  | | --- | --- | --- | |  |  | - **Employee Data**: Employee performance, satisfaction, and HR metrics. | |
| Data Governance | How will we promote data availability? (provide at least two ideas) | |  | | --- | | - **Centralized Data Repositories**: Ensure data is stored in accessible, organized repositories. |  |  |  |  | | --- | --- | --- | |  |  | - **Cloud Storage**: Leverage scalable cloud solutions like AWS or Azure to ensure availability. | |
| How will we promote usability? (provide at least two ideas) | |  | | --- | | - **Data Standardization**: Use standardized formats and structures to enhance ease of use. |  |  |  |  | | --- | --- | --- | |  |  | - **Self-Service Analytics**: Provide tools like Tableau or Power BI to empower teams to access data independently. | |
| How will we guarantee integrity? (provide at least two ideas) | |  | | --- | | - **Data Validation**: Implement checks at the point of entry and during data processing. |  |  |  |  | | --- | --- | --- | |  |  | - **Regular Audits**: Conduct periodic data audits to ensure accuracy and consistency. | |
| How will we guarantee security? (provide at least two ideas) | |  | | --- | | - **Role-Based Access Control (RBAC)**: Assign data access based on roles to minimize exposure. |  |  |  |  | | --- | --- | --- | |  |  | - **Encryption**: Use encryption methods to protect sensitive data during storage and transfer. | |
| Technology | Identify the components of your Data Architecture | |  | | --- | | - **Cloud Data Warehouse**: Use AWS Redshift, Azure Synapse, or Google BigQuery for centralized storage. |  |  |  |  | | --- | --- | --- | |  |  | - **ETL Pipeline**: Implement tools like Apache NiFi or Talend to efficiently extract, transform, and load data. |  |  |  |  | | --- | --- | --- | |  |  | - **Business Intelligence (BI) Tools**: Integrate Tableau, Power BI, or Looker for reporting and visualization. |  |  |  |  | | --- | --- | --- | |  |  | - **Data Lake**: Store raw, unstructured data for future use (e.g., Amazon S3, Azure Data Lake). | |
| Skills and Capacity | How will we promote development of data literacy skills and capacity throughout the organization (provide at least three ideas) | |  | | --- | | - **Data Literacy Training**: Offer workshops and online courses for employees at all levels to improve their data skills. |  |  |  |  | | --- | --- | --- | |  |  | - **Cross-Functional Collaboration**: Foster collaboration between business and tech teams to improve data understanding. |  |  |  |  | | --- | --- | --- | |  |  | - **Mentorship Programs**: Develop mentorship opportunities where data experts coach others within the organization. | |
| Support for Machine Learning | Give a brief description of the machine learning architecture and how it will interface with the data architecture | |  | | --- | | - **ML Models**: Utilize cloud platforms like AWS Sagemaker, Azure ML, or Google AI for model development and training. |  |  |  |  | | --- | --- | --- | |  |  | - **Model Deployment**: Integrate models into production via API endpoints or batch processing. |  |  |  |  | | --- | --- | --- | |  |  | - **Real-time Feedback**: Link data architecture with machine learning models for dynamic decision-making. | |

**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.