OLIST

Data Science Adoption Strategy – Case Study

Name : Naveen Kumar J
Position : DS Consultant
Date : 24/01/2023

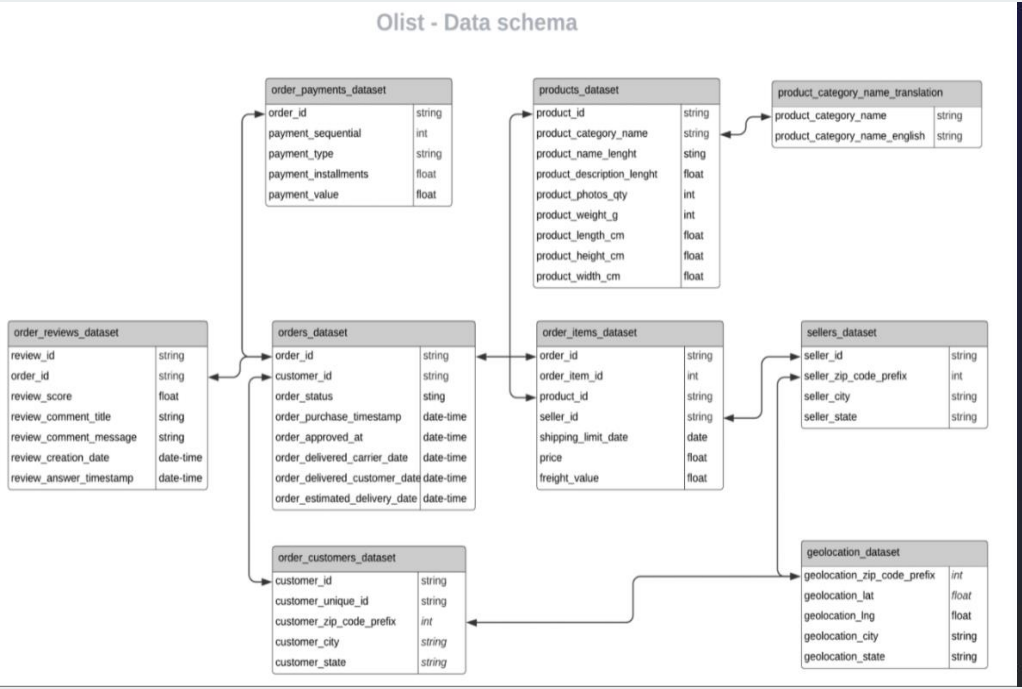
Why Data Science Adoption is needed?

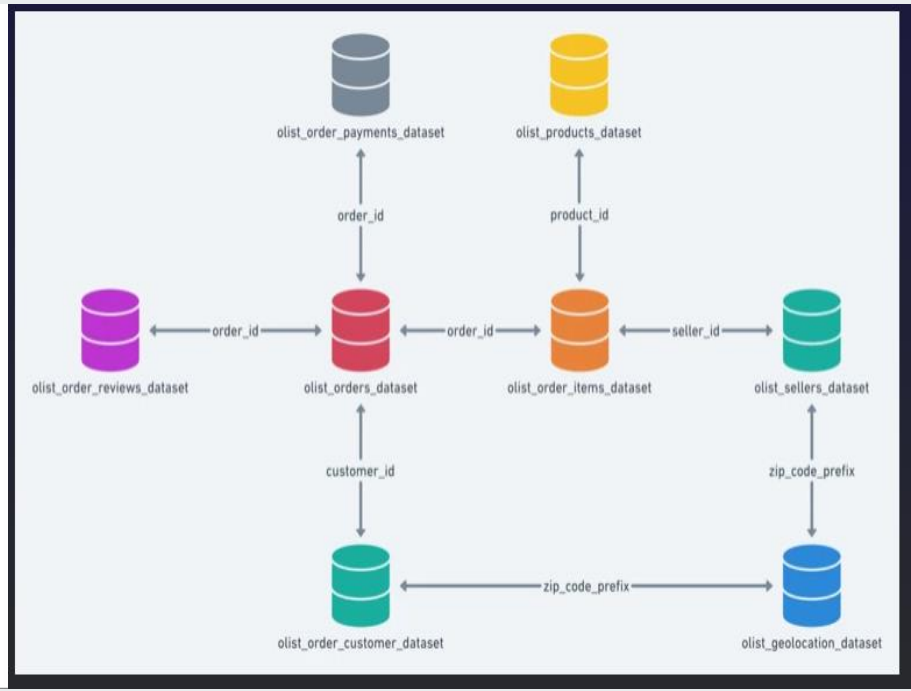
- Almost all companies are trying to adopt DS solutions and techniques into their business programs.
- The adoption of DS offers certain unique capabilities and opportunities that impact both sides of the balance sheet, increase revenue and reduce costs.
- For some businesses like fleet management and OTT platforms, DS adoption is easy.
- However, it is difficult for some other businesses like manufacturing and healthcare.
- Usually, the companies that are at an advanced stage in DS adoption have in-house data science teams, while other companies start by outsourcing projects to consultants

BUSINESS GOALS

- Increasing the number of active customers
- Increasing revenue
- Increasing the efficiency of services
- Improving customer experience

OLIST Data Schema

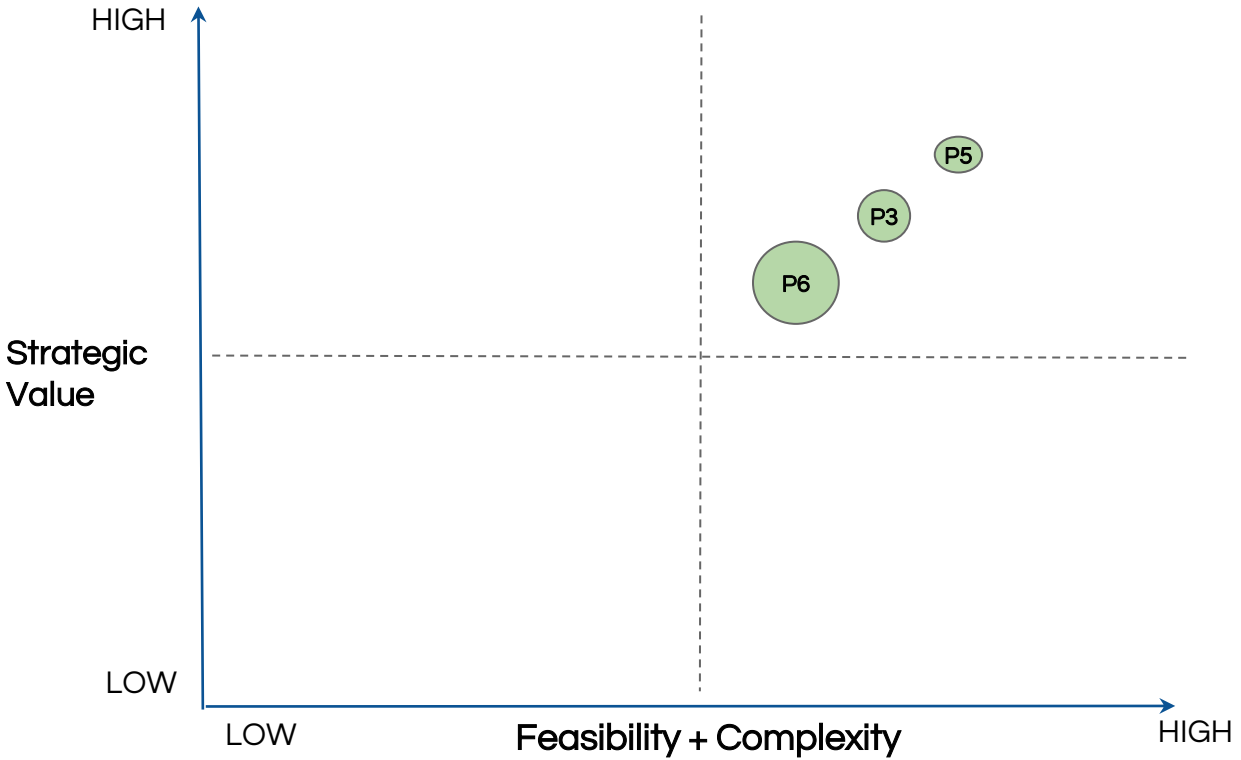




Candidate Data Science Projects

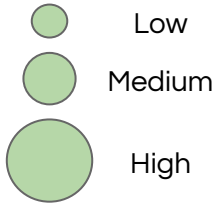
	Functional Area	Project Description
Project 1: Delivery Date Prediction	Supply Chain	Accurately Predicting delivery using Machine learning
Project 2: Sentiment Analysis	Marketing	Analyzing the sentiments of the customers using NLP and Machine learning
Project 3: Customer Churn	Marketing	Identifying ‘High-Risk ‘ customers using Classification model
Project 4: Customer Acquisition Cost Optimization	Finance	Optimizing effectiveness of the acquisition campaigns by comparing them against the lifetime value of customers.
Project 5: Fraud Detection	Security	Identifying fraudulent transactions using the power of analytics
Project 6: Price Optimization	Sales	Real time Price optimization using regression model

High Strategic Value

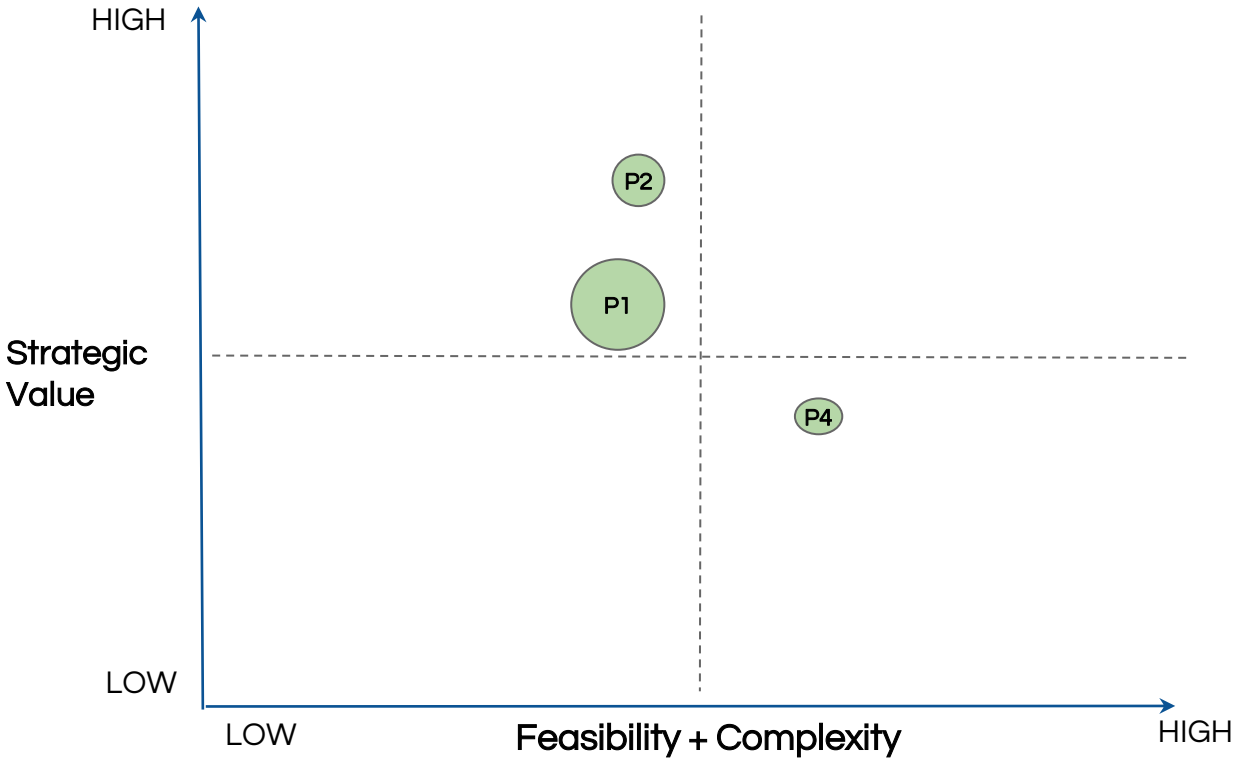


Project 1: [Delivery date prediction]
Project 2: [Sentiment Analysis]
Project 3: [Customer Churn]
Project 4: [Customer Acquisition Cost optimization]
Project 5: [Fraud Detection]
Project 6: [Price Optimization]

Business Value

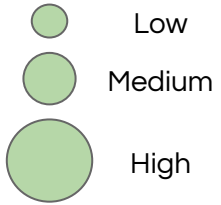


Low Strategic Value



Project 1: [Delivery date prediction]
Project 2: [Sentiment Analysis]
Project 3: [Customer Churn]
Project 4: [Customer Acquisition Cost optimization]
Project 5: [Fraud Detection]
Project 6: [Price Optimization]

Business Value



Highest-Priority Data Science Projects

Order	Project	Data feasibility	Infrastructure feasibility	Complexity	Strategic Value	Business Value
		1=Low; 5=High	1=High; 5=Low	1=High; 5=Low	1=Low; 5=High	1=Small; 5=Large
First	[Customer Churn]	4	4	4	5	5
Second	[Price Optimization]	4	4	4	4	4
Third	[Delivery Date Prediction]	4	4	3	4	3
Fourth	[Sentiment Analysis]	3	4	2	3	3
Fifth	[Fraud detection]	3	4	2	4	3
Sixth	[Customer Acquisition Cost Optimization]	4	4	2	2	2

Complete the “Data Science Road Map” below with the first four data science projects chosen for implementation.

Order	Project	Order Justification
1	**Project 3:** [Customer Churn]	Customer Churn prediction is the highest priority because maintaining large customer base is most important way to increase revenue and avoid high customer acquisition cost
2	**Project 6:** [Price Optimization]	Pricing is one of the most important aspects of business for an e-commerce organization. It has a direct and profound impact on revenue, sales, profit and demand
3	**Project 1:** [Delivery Date prediction]	Predicting correct delivery date will have an indirect impact on the revenue by reducing the churn and improving the shopping experience and delight.
4	**Project 5:** [Fraud Detection]	Protecting the customer and company from scams and fraudsters should be one of the highest priority as it will save both image and cost of customer and company.

Technical Infrastructure Needed to Support the Data Science Organization

Data Requirements	What data should be included in the Data Strategy?	• Data required is in the form of structures data and is available in respective SQL data bases. Unstructured data like images, videos and call recording should be included to create sentiment analysis and fraud detection models.
Data Governance	Data Availability	• All the relevant data is available in respective datasets. Data on special occasions like elections and festivals is not available and should be acquired. MSRP to create price optimization model should be included. Data related to the sentiments and ratings of ecommerce platform is missing. 'Fraud' and 'non-fraud' data is missing
	Usability	• Data is clearly structured, documented and labeled. Labels 'Churn' and 'Not Churn' , 'Fraud' and 'non-Fraud' are not created, and data preprocessing should be done to create those columns.
	Integrity	• data retains its essential qualities even as it is stored, converted, transferred and viewed across different platforms is ensured.
Skills and Capacity	Data literacy skills and organizational capacity	• Resource mix for AI team including Data Engineers, Data Analyst, Data Scientists are required to extract the data, analyze and create machine learning models. • Representatives from ross-functional team (CFT) of respective domains should be available to support machine learning engineers.
Support for Machine Learning	Machine learning	• The choice between on-premise and cloud • Substantial storage for all formats of data without limitations • Powerful core processors, GPUs etc. for high compute Algorithms • Tools and software necessary for AI execution • Remote access, virtual machines etc.

Conclusion

- Even though it is difficult to implement DS projects, but the benefits they offer outweigh the efforts one needs to put in.
- Manufacturing companies, workers are required to wear safety equipment.
- It will be extremely labor intensive for security personnel to look after all the CCTV footage at once to keep a watch on the workers.
- A DS solution can do this quite easily. It can use a computer vision model to consume all the CCTV footage, identify employees who are not wearing safety equipment and highlight these employees to security personnel.
- The security personnel can then decide the action that needs to be taken.