OLIST

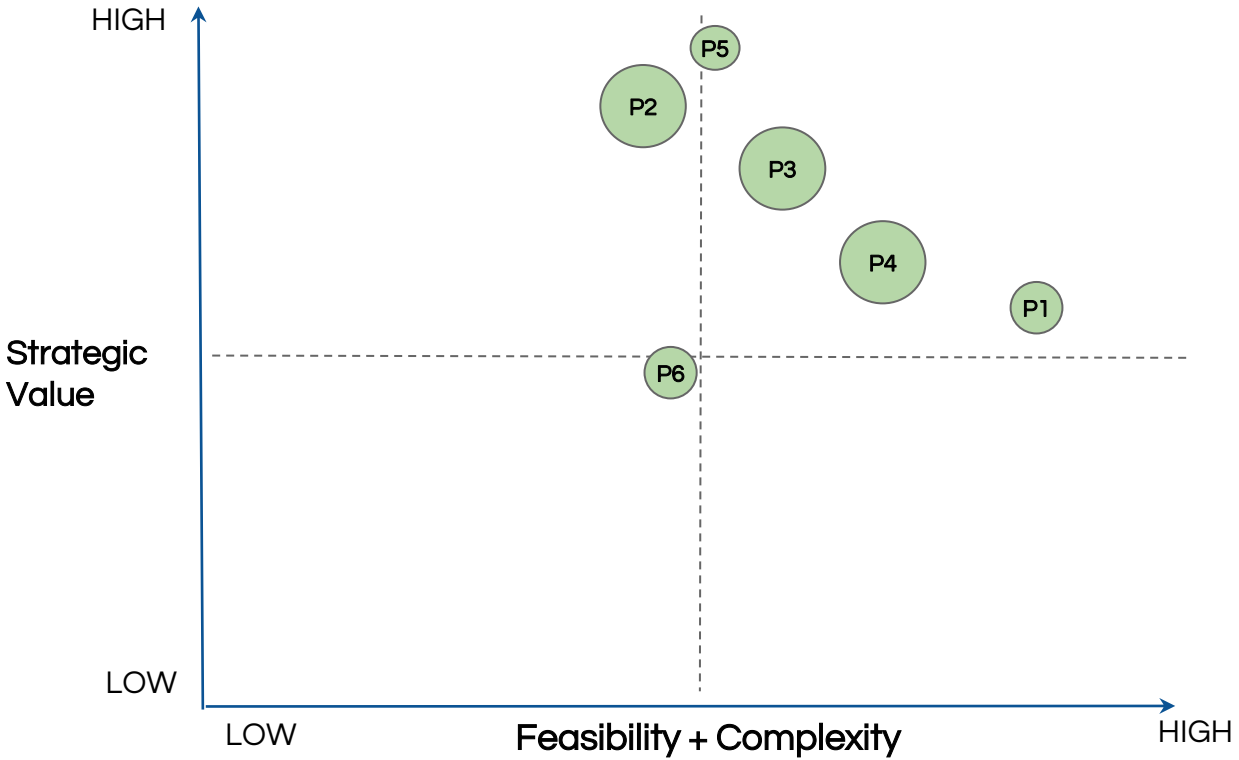
Data Science Adoption Strategy

Pieush Vyas
UOA-MS(DS)
13-06-2023

Candidate Data Science Projects

	Functional Area	Project Description
Project 1: Delivery Date Prediction	Operations	Delivering on time delights customers. Customer satisfaction and company competitiveness depend on data supply. Accurate date forecasts set the delivery team up for better inventory management and last-mile delivery.
Project 2: Sentiment Analysis	Customers	Customer sentiment is a business's intangible asset. Like any resource sentiment, it needs management. Sentiment analysis can be done in many ways, but this use case uses customer reviews to identify OLIST's happy and unhappy customers.
Project 3: Customer Churn	Sales	Customers switch e-commerce companies, as in many businesses. The company uses a churn model to keep customers from leaving. Customers likely to migrate are identified by the model. Now, the company wants to prevent churn.
Project 4: Customer Acquisition Cost Optimization	Sales	OLIST might spend 30 BR to acquire one customer. Customer acquisition costs 30 Brazilian Real (BR). Only if the customer generates over 30 BR in lifetime revenue is it worth spending 30 BR. Thus, the company seeks optimization.
Project 5: Fraud Detection	Operations	E-commerce marketplaces connect buyers and sellers. Independent seller-buyer fraud will damage the company's image. Damage will affect company revenue.
Project 6: Price Optimization	Sales	High prices reduce sales but increase profits. Conversely, a low price increases sales but decreases profit. Customer segments and special occasions also affect product sales.

Complete the Data Science Opportunity Matrix below by modeling each of the six projects in terms of feasibility, complexity, strategic and business value impact.



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
Medium
High

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
Frist	Sentiment Analysis	4.3	5	5	5	4
Second	Fraud Detection	4.6	5	4	4	2

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

Order	Project	Order Justification
1	**Sentiment Analysis**	Knowing what your customers are saying is important to any type of business as it is cheaper to maintain a current customer than trying to bring in new customers.
2	**Fraud Detection**	Fraud Detection is important for the continuation of business as if orders are not screened for fraudulent transactions chargebacks can occur leading to large forecasted expenses.
3	**Delivery Date Prediction**	Customers always expect for their parcel to arrive on-time, should this not be the case it is most likely they will not shop with us again.
4	**Price Optimization**	Price optimization is important to ensure all prices listed on the site are competitive when compared to similar sites.

Technical Infrastructure Needed to Support the Data Science Organization

Data Requirements	What data should be included in the Data Strategy?	• Data required includes but not limited to Customer Profiles, Purchase data, Orders, Product Data, Location Data of Customers (If Possible)
Data Governance	Data Availability	• Data should be updated every week or two weeks to help develop the model further with more usable data
	Usability	• Data should be relevant to its use case ie we should have access to order and delivery information when predicting delivery dates and access to payment info and customer info when using it for fraud analysis
	Integrity	• Data should not be missing any important fields that are considered required for any of the ear-marked projects otherwise rows that are missing important info will need to be removed from the data frame.
Skills and Capacity	Data literacy skills and organizational capacity	• Python Knowledge • Pandas • TensorFlow / Scikit-Learn Usage • Data Cleaning Techniques
Support for Machine Learning	Machine learning	• Google Colab Access or alternative Jupyter Notebook instances • Access to CI Tools such as Apache Airflow and MLkit