**Delivery date prediction**

**Problem statement:**

The logistics team at Olist uses heuristics to provide an estimated delivery date for the orders placed. It is very conservative about the delivery dates. As a result, it is able to deliver the products much in advance. Although this is beneficial for the logistics team’s 'on time delivery' KPI, it is not favourable for the CMO. He found that on average, the estimated time to deliver products that are given to customers is twice that of the actual delivery time. Such a high expected delivery time is driving away Olist's customers. So, the CMO is looking to use ML to get a far more accurate expected delivery date.

**Proposed Solutions:**

1. ML solutions -   
   1. Use one regression model using features like destination location and source location, and historic delivery date, item size, proximity to distribution hubs, seller, etc to predict the estimated time of delivery. Add the delivery time to the order date and calculate the estimated delivery.
   2. Divide the entire delivery process into individual stages of transport.
      1. Estimate time for getting the goods from the vendor.
      2. Estimate the time that the item will stay in the warehouse.
      3. Estimate the time required to send the item from the warehouse to the delivery location.  
           
         Build individual models for estimating these times, add the time to get the overall delivery time and add it to the order time to get the estimated delivery date.   
         The advantage of building separate models is more fine control over the processes and better prediction.
2. Non - ML solutions
   1. A rule-based approach to predicting delivery times. There are set times for each step of the delivery process such as set time for getting goods from the warehouse to the transport hub, from the transport hub to the next transport hub etc.   
        
      Each of these times is mapped to the nearest shipping cutoffs wherever appropriate. Fixed additional processing times are added for weekends and holidays. Thus, the rule-based model doesn’t adapt based on recent performance changes and is designed based on heuristics.

Selecting First ML solution - because the data points for individual steps are not present.

**Benefits of Proposed Solution**

**Process improvements -**

Predicting delivery dates more accurately will not impact any process in the e-commerce setup.

Reduction in Inventory costs.

**Monetary benefits -**

The monetary benefits of accurate delivery can be divided into two streams

* Reducing customer churn because of accurate delivery date prediction. Hence, increasing the revenue of the company.   
  1. Reduction in the number of people who drop out - 10% (Assumed)
  2. Number of customers daily - 272.44   
     (total number of orders in 1 year = 99442 / number of days in a year 365)
  3. Average order value - 100 Brazilian real (Assumed)
  4. Revenue increased by = 0.1 \* 272.44 \* 100 = 2724.4 per day.
* Inventory management   
    
  Predicting the delivery date correctly will reduce the pressure on the supply chain. The through out rate of deliveries of the entire system increases but predicting the delivery date correctly. The warehouses will have to hold lesser inventory.
  1. Reduction of time items stay in warehouse = 40% (assumed)
  2. In the same time frame warehouse will be able to hold more items = 40% (assumed)
  3. Warehouse cost per item delivered will reduce = 25% (assumed)
  4. Initial warehouse cost per item delivered = 5 BR (assumed)
  5. Total savings per day = 272.44 \* 0.75 \* 5 BR = 1021.65 BR

**Summarise the DS approach**

The estimation of accurate delivery dates is a regression problem to be solved. You use various data to estimate the time needed for delivery, then add the time to order date to get the right delivery date.

Delivery date prediction is also a kind of balancing act between competitiveness and accuracy.   
You can always have long delivery dates and always be accurate, but might lose on sale to some competitor who can deliver quickly.

Or you can have extremely short delivery time promises and disturb the customer sentiment. Hence add a buffer  
  
[There is a constant trade-off between being accurate and being competitive and, of course, we would aim to optimize both.](https://towardsdatascience.com/delivery-date-estimation-5aff1a0ff8dc)

**Limitation**

1. RMSE might be high
2. High complexity model - non linearity , thus non linear model might be needed - is the team capable or not.
3. Think of edge cases - data for special cases is not available

**Define Appropriate Success Metrics**

1. Early delivery rates
2. Late delivery rates
3. Churn rate after estimated dates is shown

**Prioritising Use Case**

Refer to the use case prioritisation framework linked [here](https://docs.google.com/spreadsheets/d/1EGoVubzdetsv8YhrKeenSq_uaBSubtsgZm2WydL1U1I/edit#gid=704154787).

**Cite references**

1. Similar delivery date estimation application form example on amazon blogs. <https://aws.amazon.com/blogs/industries/how-to-predict-shipments-time-of-delivery-with-cloud-based-machine-learning-models/>
2. Paper suggesting various solutions for delivery date production <https://arxiv.org/pdf/2105.00315.pdf>

**Sentiment Analysis**

**Problem statement:**

The Chief Marketing Officer at Olist wanted to understand the experience of the customers based on the reviews received after the delivery of the orders. He also wanted to identify the areas of improvement based on these reviews. He had heard that NLP can be used for sentiment analysis and topic modeling, which will be useful in finding topics in customer reviews. However, he was also cognizant of the fact the customer reviews are in Portuguese, whereas the NLP algorithms are not so sophisticated in Portuguese.

**Proposed Solutions:**

1. ML Solution
2. Use pre-trained NLP models on Portuguese language data for sentiment analysis. This can be achieved using pre-trained language models likes BERT or fast-text for Portuguese language.
3. Apply topic modelling techniques to identify recurring themes or topics in negative reviews. This could provide insights into the causes of dissatisfaction among the customers.
4. Non ML solution

* Manual review and categorisation of feedback by a dedicated team proficient in Portuguese. This is a time-consuming process and inevitably prone to human bias and error. Given the large scale of the data, and the proficiency of NLP solutions in managing such tasks, the first ML solution is recommended.

**Benefits of Proposed Solution**

Process improvements:

Understanding the sentiment trend of customers will enable the business to identify and address issues affecting customer satisfaction. This could lead to various process improvements including, but not limited to, quality checks, packaging, relationships with suppliers and couriers.

Monetary benefits:

1. Reduction in customer churn rate due to better handling of complaints and dissatisfaction.
2. Increase in customer retention - Happy customers are likely to make repeat purchases
3. Positive word-of-mouth - Happy customers are potential brand ambassadors and this can generate free advertising for the business.

**Summarise the Solution**

The sentiment analysis requires Text processing, Feature generation, Training a model for sentiment prediction, and Once the sentiment is predicted, take relevant actions based on the sentiment.

Limitation

1. Contextual understanding of language - Sarcasm, irony or local idioms, expressions can be hard to detect for a model.
2. Language dialects/variations - Interpreting sentiment from different dialects might prove challenging for a model trained on standard Portuguese language data.
3. Model Interpretability - NLP models are black-box type which makes it difficult to interpret why they produce certain results. This can complicate the process of deriving actionable insights.

Define Appropriate Success Metrics

1. Reduction in churn rate
2. Increase in repeat purchase rate.

**Prioritising Use Case**

**Cite references**

Customer Churn

Problem statement:  
  
Customer churn is a critical metric for the CMO of any e-commerce company. OLIST wants to develop customer churn models to identify 'at-risk’ customers so that an appropriate retention strategy can be built. This will provide insights into the factors driving customer churn, thus reinforcing its retention efforts.  
  
Maintaining a large customer base is an important way of increasing revenue. However, as it happens in many businesses, customers tend to move between e-commerce companies. To prevent customers from constantly migrating, the company has built a churn model. The model is used to identify the customers who are likely to migrate. Now, the company wants to come up with a strategy to prevent churn.

Proposed Solutions:

1. ML solutions -
   1. Use classification algorithms such as Logistic Regression or Decision Trees to predict customer churn. These models would use features such as recency of purchase, frequency of purchase, value of purchase, customer service interactions, time since last engagement and more.
   2. Use association rule learning to understand the patterns in customer behaviour prior to churn, which can also indicate dissatisfied customers. This understanding can be used to develop targeted retention campaigns.
2. Non - ML solutions

Perform segmentation on the customer base for targeted marketing campaigns to engage customers.

Benefits of Proposed Solution

Process improvements:

Understanding the factors contributing to churn can lead to organizational improvements in areas like customer service, product pricing, or marketing.

Monetary benefits:

1. Increase in customer lifetime value - as improved retention leads to more repeat purchases
2. Improved understanding of customer needs leading to better targeted marketing campaigns, reducing cost of sales.

Summarise the DS approach

The customer churn predication is a classification problem to be solved using supervised learning models. We use various customer-related data to categorize the customers that are likely to leave.

Limitations

1. Concept and Data Drift
2. Data collection and processing - Depending on the size of the customer base, collecting and processing the data required for the churn model can be challenging.

Define Appropriate Success Metrics

1. Reduction in the churn rate
2. Increase in customer lifetime value

**Customer Acquisition Cost Optimisation**

Problem Statement:

The Marketing team at OLIST runs multiple promotional campaigns to acquire new customers. However, the CFO believes that the marketing team is burning significant cash by offering deep discounts on products and other benefits, which is inflating the customer acquisition cost. The CFO wants to initiate a new process to measure the effectiveness of the acquisition campaigns by comparing them against the lifetime value of customers.  
  
Another way of increasing revenue is to gain more customers. The money that a company spends on getting one customer is called the acquisition cost. For instance, suppose OLIST has to spend 30 BR to acquire one customer. In this case, 30 Brazilian Real (BR) is the acquisition cost of the customer. Obviously, it would be worth spending the 30 BR only if the customer generates more than 30 BR of lifetime revenue. So, the company wants to solve this optimisation problem.

Proposed Solutions:

1. ML solutions -
2. Use regression models to predict Customer Lifetime Value (CLV) using features like the user's demographics, geo-location, time of first purchase and average order value. This can help in determining which users are worth acquiring.
3. Conduct A/B testing on different marketing strategies to identify the most cost-effective methods.
4. Non - ML solutions

* Conduct market research to understand customer behaviour and tailor marketing strategies accordingly.

Given the complexity and need for accurate prediction, the first ML solution using regression models is recommended.

Benefits of Proposed Solution

Process improvements:

Optimising customer acquisition costs can lead to organisational improvements in areas like promotional planning, discount strategies, and campaign planning.

Monetary benefits:

1. Reduction in acquisition cost per customer
2. Increase in customer base due to optimised marketing campaigns
3. Improvement in profitability due to lowered marketing costs

Summarise the DS approach

Customer Acquisition Cost Optimisation involves using supervised learning models, like regression, to predict the lifetime value of newly-acquired customers. The cost of acquisition is then compared with this value to measure the effectiveness of marketing campaigns.

Limitations

1. Reliability of predictions - Customer behaviour is unpredictable and there may be significant variations in real-world scenarios.
2. Dynamic market conditions - Changes in competitions' marketing strategies, market trends etc can effect the acquisition costs.

Define Appropriate Success Metrics

1. Reduction in customer acquisition cost
2. Increase in the ratio of customer lifetime value to acquisition cost 3.Project ROI from marketing campaigns.

**Fraud Detection**

Problem Statement:

Fraud is one of the most challenging areas to deal with in the e-commerce industry, as it can result in huge financial losses. There can be fraud in the areas of merchant identity, advanced fee, wire transfer scams, chargeback transactions, etc. The CFO wants to use the power of analytics to identify fraudulent transactions so as to help guard the organisation against such actions.  
  
E-commerce marketplaces are a platform that brings together sellers and buyers. Any fraud that happens between independent sellers and buyers will harm the company’s image. The harm caused will have a direct impact on the revenue of the company.

Proposed Solutions:

1. ML solutions -
2. Use supervised learning algorithms such as Random Forest, Naive Bayes or SVM to classify transactions as fraudulent or non-fraudulent based on features like transaction amount, payment method, device used, location, time of transaction etc.
3. Use unsupervised learning algorithms like Isolation Forest or Autoencoders to detect anomalies in transactions which may indicate fraud.
4. Non - ML solutions

* Conduct regular audits and manually review transactions to identify any discrepancies, which can be time-consuming and inefficient.

Given the complexity and need for accurate detection, the first ML solution is recommended.

Benefits of Proposed Solution

Process improvements:

Increased fraud detection can lead to improved internal controls and risk management processes.

Monetary benefits:

1. Reduction in fraud losses
2. Retaining customer trust and hence loyalty, leading to improved sales

Summarise the DS approach

The fraud detection problem is a classification problem to be solved using either supervised (where historical fraud labels are available) or unsupervised learning algorithms (when labels are unavailable). Various transaction-related data points are used to classify transactions as fraudulent or non-fraudulent.

Limitations

1. Class imbalance - Fraud occurrences are usually rare compared to legitimate transactions, leading to imbalanced dataset.
2. Data and Concept Drifts
3. False positives - Any false warnings of fraud may inconvenience genuine customers.

Define Appropriate Success Metrics

1. Reduction in the amount of fraud losses
2. Decrease in false positives
3. Improvement in detection rate of fraudulent activities.

**Price Optimisation**

**Problem Statement:**  
  
Pricing is one of the most important aspects of business for an e-commerce organisation. It has a direct and profound impact on revenue, sales, profit and demand. Price optimisation is performed using a number of factors such as the location, the attitude of the customer, competitor’s pricing, etc. And, the data science algorithm predicts the customer’s segmentation to develop a response to the change in price. The OLIST sales team wants to build a price optimisation algorithm so as to maximise sales and revenue.  
  
Similar to acquisition cost optimisation, price optimisation is also a balancing act. There are multiple factors that go into deciding the price of a product such that a customer is most likely to buy it. If the product is priced high, then the probability of selling the product is low but the profit generated is high. On the other hand, if the price is low, then the probability of selling the product is high but the profit generated is low. Moreover, the probability of selling a product is dependent on multiple factors such as customer segments and special occasions.

Proposed Solutions:

1. ML solutions -
2. Implement a regression model to predict the optimal price points considering all variables that could influence the pricing decision like customer demand, competitive prices, time of year, product type, and more.
3. For price sensitivity, use classification models to segment customers into different groups, and adjust pricing strategies accordingly.
4. Non - ML solutions

* Use market research methods to understand customer behaviour, competitive pricing and tailor pricing strategies accordingly.

Considering the scale of operations and the complexities involved in price optimisation, the first ML solution is recommended.

Benefits of Proposed Solution

Process improvements:

Optimising prices can lead to improvements in areas like promotional planning, sales strategy, or supplier negotiation.

Monetary benefits:

1. More sales due to optimised prices
2. Increase in revenue due to better profitability
3. Improved customer retention due to competitive pricing

Summarise the DS approach

Price Optimisation is a prediction problem that requires using machine learning models, such as regression or classification, to predict the optimal price points or categorize customers based on their price sensitivities.

Limitations

1. Dependence on other variables - The optimal price can be influenced by a lot of external factors for which data might not be available.
2. Dynamic market conditions - Changes in competitors’ pricing strategies, market trends etc can affect the optimal price.

Define Appropriate Success Metrics

1. Increase in the sales volume
2. Profit margin growth.