**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 favorable 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 throughout 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

**Summaries 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

**Prioritizing Use Case**

Refer to the use case prioritization 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
   1. Create a system to extract reviews and ratings from the customers solely for the platform. For e.g. if the customer is downloading the app from AppStore, prompt the customer to rate the e-commerce app.
   2. Use BERTimbau - Portuguese BERT pre-trained model to perform sentiment analysis of the reviews.
   3. Use one Naive Bayes classifier using features like review score and review title and review comment and order ID etc. to classify the orders into good and bad reviews.   
        
      Segregating the app reviews and customer reviews will help in understanding the customer experience on the platform more correctly.
2. Non ML solution
   1. Customer service team calling customers to get feedback. This process will require a lot of resources incurring costs for the company

Selecting first ML solution because we already have reviews dataset

**Benefits of Proposed Solution**

**Process improvements -**

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

**Monetary benefits –**

The monetary benefits of sentiment analysis can be divided into two streams

1. 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.

1. Improve Sales Growth
   1. Current Number of customers daily - 272.44   
      (total number of orders in 1 year = 99442 / number of days in a year 365)
   2. Old Average order value - 100 Brazilian real (Assumed)
   3. New average order value - 105 BR (Assumed)
   4. Sales Growth = (272.44 \* 105) – (272.44 \* 100) = 1362.2 BR per day.

**Summarize the Solution**

Analyzing the sentiments of customers helps in extracting all the related information relevant to the concerns of the customers. We can extract information related to all the departments of the platform where the customer is facing the problem and take appropriate action.

It can also help in reducing churn, identifying fraud, and all other customer grievances

[However, there are certain pitfalls to sentiment analysis as well](https://www.toptal.com/deep-learning/4-sentiment-analysis-accuracy-traps).

**Limitation**

1. High complexity model - non linearity , thus non linear model might be needed - is the team capable or not.
2. Resources to build an efficient sentiment analysis model might not be available

**Define Appropriate Success Metrics**

1. Positive sentiments
2. Neutral Sentiments
3. Negative sentiments

**Prioritizing Use Case**

Refer to the use case prioritization framework

**Cite references**

1. How sentiment analysis can be used realtime:

<https://aws.amazon.com/blogs/media/how-to-capturing-and-amplifying-user-engagement-through-real-time-sentiment-analysis/>

1. Sentiment Analysis based on deep learning:

<https://arxiv.org/abs/2006.03541>

**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. 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.

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. This will provide insights into the factors driving customer churn, thus reinforcing its retention efforts.

**Proposed Solutions:**

1. ML Solution
   1. Create a ML model which can predict the ‘high risk’ customer and share the insights with CMO so that an appropriate retention strategy can be built.
   2. Identify driving factors for churn and share the insights with the relevant departments so that the respective departments can improve the services and enhance the experience for the customers
2. Non ML solution
   1. The marketing team can roll out small to medium size promotions and discounts to all the customers but this can inflate the customer acquisition cost

Selecting first ML solution because it is a more efficient data-based approach

**Benefits of Proposed Solution**

**Process improvements -**

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

**Monetary benefits –**

The monetary benefits of sentiment analysis can be divided into two streams

1. Reducing customer churn :  
   1. Reduction in the number of people who drop out - 20% (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 = 0.2 \* 272.44 \* 100 = 5448.8 per day.
2. Improve Sales Growth
   1. Current Number of customers daily - 272.44   
      (total number of orders in 1 year = 99442 / number of days in a year 365)
   2. Old Average order value - 100 Brazilian real (Assumed)
   3. New average order value - 105 BR (Assumed)
   4. Sales Growth = (272.44 \* 105) – (272.44 \* 100) = 1362.2 BR per day.

**Summarize the Solution**

Customer churn prediction is a classification problem. Predicting the high-risk customer and sharing the insights with CMO will create an opportunity to formulate a retention strategy targeted only to high risk customers will increase the revenue as well as reduce the cost of money spent on low risk customers.

Driving factors for the churn can be identified and the whole system can further be optimized

**Limitation**

1. Improper optimization will either not identify high risk customers and customer will be lost or low risk customers will exploit promotions and discounts

**Define Appropriate Success Metrics**

1. Churn Rate

**Prioritizing Use Case**

Refer to the use case prioritization framework

**Cite references**

1. How Gild Dropped Monthly Churn From 4% to <1%:

<https://predictablerevenue.com/blog/seeds-how-gild-dropped-monthly-churn>

1. A High-Performance Customer Churn Prediction System based on Self-Attention:

https://arxiv.org/abs/2206.01523

**Customer Acquisition Cost Optimization**

**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 optimization problem..

**Proposed Solutions:**

1. ML solutions -   
   1. Using the power of analytics, optimize the customer cost acquisition(CAC) and lifetime value (LTV) ratio. Your customers’ LTV needs to always be higher than your CAC in order for you to be making profits in the long run. A ratio of 1:1 means you’re spending too much, thus losing money for every acquisition you make. What you should be aiming for instead is a ratio of 3:1.
   2. Identify the lifetime habits of your top consumers by segmenting out different customer bases using clustering techniques and targeting them accordingly.
   3. Identity the top customers and using market basket analysis find a way to upsell and cross-sell
2. Non - ML solutions
   1. Improving the website and mobile app and making it more visually aesthetic for a better experience.
   2. Proper scaling to maintain low latency operations

Both ML and non ML solutions can be used hand in hand.

**Benefits of Proposed Solution**

**Process improvements -**

Customer Acquisition Cost Optimization more accurately will not impact any process in the e-commerce setup.

It will reduce the cost of acquisition.

**Monetary benefits -**

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

1. Reducing customer churn because of CAC optimization. Hence, increasing the revenue of the company.   
   1. Marketing Spent before CAC optimization per month: 65000 BR(Assumed)
   2. Marketing spent before CAC optimization per day: BR 2166 (65000/30)
   3. Marketing Spent after CAC optimization per month: 20000 (Assumed)
   4. Marketing spent after CAC optimization per day: 666.66 BR
   5. Cost saved: 2133-666: BR 1500

**Summarise the DS approach**

Calculate the LTV of customers based on the existing data. Target optimized marketing cost on customers based on LTV.

**Limitation**

1. Incorrect optimization will result in reducing the market budget too low causing customer churn
2. Creating a real-time optimization model can be challenging

**Define Appropriate Success Metrics**

1. CAC- Customer acquisition cost
2. LTV- Life time value
3. Cost savings after optimization

**Prioritising Use Case**

Refer to the use case prioritization framework

**Cite references**

1. Importance of Customer Acquisition and How AI Can Help You with it!

[https://medium.com/nerd-for-tech/importance-of-customer-acquisition-and-how-ai-can-help-you-with-it-d67f5975776c /](https://medium.com/nerd-for-tech/importance-of-customer-acquisition-and-how-ai-can-help-you-with-it-d67f5975776c%20/)

**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 organization against such actions. Any fraud that happens between independent sellers and buyers will harm the company’s image.E-commerce marketplaces are a platform that brings together sellers and buyers. The harm caused will have a direct impact on the revenue of the company.

**Proposed Solutions:**

1. ML solutions -   
   1. Building an unsupervised learning anomaly detection model to identify unusual patterns and activity.
   2. Using the historical data, identify the fraudulent transactions and label the dataset. (Fraud and not fraud). This will be a highly imbalanced dataset. Use sampling techniques like SMOTE to handle the imbalance and create a classification model to predict fraud transactions.
   3. Creating an ID-proof verification system using computer vision and OCR techniques to
2. Non - ML solutions
   1. Conduct regular site security audits.
   2. Make sure your store is PCI compliant.
   3. Bricks-and-mortar stores hire fraud prevention officers to catch shoplifters. You can protect your online store against fraudulent transactions by monitoring your store for suspicious activity. Monitor your accounts and transactions for red flags, such as inconsistent billing and shipping information, as well as the physical location of your customers. Use tools that track customer IP addresses and alert you to any addresses from countries known as a base for fraudsters.
   4. Use Hypertext Transfer Protocol Secure (HTTPS). HTTPS is the secure version of HTTP, which is the primary protocol used to send data between a customer’s web browser (like google) and your online store. HTTPS encrypts this data to protect sensitive information, such as customer names, addresses and credit card numbers. Using HTTPS prevents your online store from having its transactions broadcast in a way that’s easily viewed by hackers, cybercriminals, and fraudsters. You use HTTPS by buying an SSL certificate.

**Benefits of Proposed Solution**

**Process improvements -**

Preventing fraud will not only save costs but also save the company's image and reduce churn.

Customers may never trust the platform if they are victimized by a fraudster.

Fraud detection will also save customers

**Monetary benefits –**

Every $1 of direct fraud costs organizations between $2.48 to $2.82 in loss. This means that roughly 2.5x the amount lost to fraud is lost fighting fraud. [[LexisNexis](https://risk.lexisnexis.com/insights-resources/research/us-ca-true-cost-of-fraud-study)] And, merchants overall will spend 3%-5% over their overall revenue combating fraud operationally. [[BigCommerce](https://support.bigcommerce.com/articles/Learning/How-Fraud-Can-Impact-your-Business)]

Olist Revenue in 2018 – 25M (assumed)

Based on above report, assume 5% combating fraud operationally:

0.05\* 25,000,000= 1250000/365= 3424.65 BR per day

**Summaries the DS approach**

Using the historical data, create a model to classify normal and fraudulent transactions. Real-time fraud detection models can be built using unsupervised anomaly detection techniques which can study patterns in the data and detect abnormalities.  
  
[Does There Have to Be a Tradeoff Between Fraud Prevention and Customer Experience?](https://www.datavisor.com/blog/does-there-have-to-be-a-tradeoff-between-fraud-prevention-and-cx/)

**Limitation**

1. Building an efficient fraud detection system requires lot of domain expertise and highly skilled developers
2. The developed model can be highly complex.
3. Much more difficult is to create a real-time application with a fast response rate

**Define Appropriate Success Metrics**

1. Fraud score
2. Reduce False positives
3. Anomalies in case of anomaly detection

**Prioritizing Use Case**

Refer to the use case prioritization framework

**Cite references**

1. Deep learning for fraud detection in retail transactions:

https://medium.com/walmartglobaltech/deep-learning-for-fraud-detection-in-retail-transactions-564d31e5d1a3

1. Machine Learning for Fraud Detection in E-Commerce:https://arxiv.org/abs/2107.01979

**Price Optimization**

**Problem statement:**

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. Price optimization 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 optimization algorithm so as to maximize sales and revenue.

Similar to acquisition cost optimization, price optimization 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 -   
   1. Using various clustering techniques, perform customer segmentation. Based on location, the attitude of the customer, and competitor’s pricing, perform real-time price optimization system using a regression model in different segments.
   2. Try to include edge case scenarios like elections, whether to make the model more reliable
   3. Identify the items which have a high selling probability and a low selling probability. If the probability of selling is high, the product is priced low, and vice versa.
2. Non - ML solutions
   1. Maintaining prices of items based on certain margins.

Choosing ML based solution as it’s the key to maintain optimized price

Non ML based solutions won't consider all the factors which will lead to improper pricing and can increase the cost if the product is priced too high, churn rate will also increase.

**Benefits of Proposed Solution**

**Process improvements -**

Price optimization will increase revenue by increasing the customer base and reducing churn.

Companies using the efficient price optimization technique will always have an over the competitors.

**Monetary benefits -**

The monetary benefits are mentioned below:

1. Reducing customer churn because of optimization. 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.
2. An increase in customer base will increase the revenue:
   1. Current Number of customers daily - 272.44   
      (total number of orders in 1 year = 99442 / number of days in a year 365)
   2. Customers daily including the newly acquired customers: 300 (Assumed)
   3. Average order value - 100 Brazilian real (Assumed)
   4. Revenue increase = (300-272.44)\*100 = 27.56\*100= 2756

**Summaries the DS approach**

Using various clustering techniques, perform customer segmentation. Based on location, the attitude of the customer, and competitor’s pricing, perform a real-time price optimization system using a regression model in different segments. Try to include edge case scenarios like elections, whether to make the model more reliable

Identify the items which have a high selling probability and a low selling probability. If the probability of selling is high, the product is priced low, and vice versa.  
  
[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. Highly skilled in both machine learning and domain expertise needed
2. Data should be acquired for special cases like festivals, elections, competitors data etc.

**Define Appropriate Success Metrics**

1. Optimized Price

**Prioritizing Use Case**

Refer to the use case prioritization framework.

**Cite references**

1. Optimal pricing for maximum profit using Amazon Sage Maker: https://aws.amazon.com/blogs/machine-learning/optimal-pricing-for-maximum-profit-using-amazon-sagemaker/
2. Price Optimization with Practical Constraintshttps://arxiv.org/pdf/2104.09597.pdf