Bank Tickets Cancellation Rate Analysis

Case Study

26/02/2019

Part 1: DNVB Business Model

Digitally Native Vertically-Integrated Brands

DNDTC - Digitally Native Direct To Consumer

- Quickly summarize the digitally native vertical brand (DNVB) business model and explain the main differentiating factors vs traditional legacy brands as well as multi-brand stores.
- Mention 3 examples of DNVB brands in the US; one not being in fashion.
- Explain which categories of branded consumer goods are most ideal for the DNDTC business model and why.

DNVB Business Model

Key Principles of a DNVB Business Model:

- ➔ Sell online
- ➔ Directly to consumers (no third parties),
- ➔ Control your product all the way from warehouse to customers (no middlemen)

Main competitive advantages:

- ➔ Unit economics
- ➔ Branding

Growing 3x faster than traditional brands!

What differentiates a DNVB from a traditional legacy brand ?

- ➔ Impeccable customer Experience
- ➔ Higher 'intimacy' with the customer ⇐ more customer data
- ➔ Digitally born - initially created for selling online and not differentiated into it
- ➔ Omnichannel - Guide Shops, partnerships with marketplaces and influencers

Sources:
Kevan Lee, 12/01/2019
Digital Commerce, 05/04/2018
Medium, Andy Dunn, 09/05/2016

DNVB brands examples: USA

Proper Cloth - custom dress shirts

Ledbury - clothes

Thirdlove - bras

Ministry of Supply - fashion

Jack Erwin - shoes

Harry's Inc - razors

Brosa - furniture

Bougs - flowers

Main source:
The DNVB Encyclopedia by Andy Dunn

Explain which verticals/categories of branded consumer goods are most ideal for the DNDTC business model and why

Consumer goods: items used daily by the average consumer → opportunities for unit economics, transaction costs reduction and logistics efficiency

However, need for frequent replacement → leads to low switching costs

→ Brands with emphasis on **brand loyalty**

Importance of **differentiation**, importance of customer **data** and **experience**

→ lengthening the product life cycle by improving quality and increasing customization

- ➔ fashion brands
- ➔ daily use products
- ➔ cosmetics
- ➔ self-care products

Part 2. Case Study

The main activity you must do is a data exploration to bring actionable insights on how to reduce cancelation rate. Please present the main findings and action plans.

Steps of analysis

Data cleaning and assigning new values *- working with outliers, excluding potential fraud activities, calculating KPIs and intermediary fields*

‘Look at the data’ *- graphs, cross-tabs, heat maps. Google Data Studio + Seaborn*

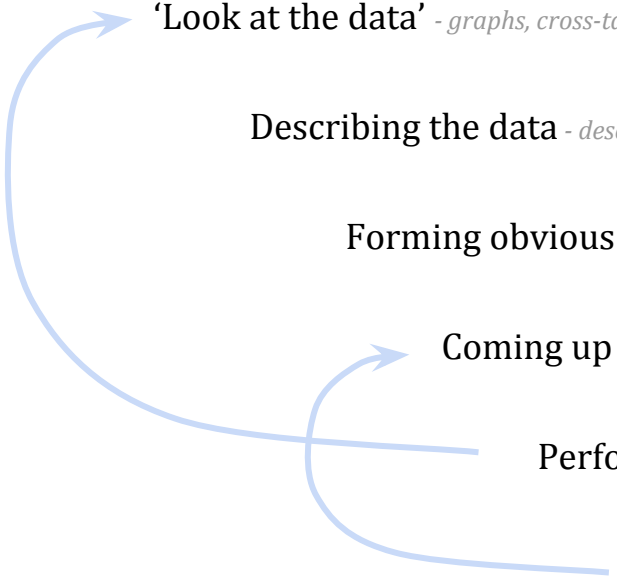
Describing the data *- description tables. Pandas*

Forming obvious hypotheses *- forming the simplest ones based on common sense and data descriptions above*

Coming up with ways to check hypotheses

Performing analysis and interpreting the results *- justifying the business sense and insights*

Coming up with new hypotheses *- based on the results of previous analysis*



Data Description

Dataset consists of 4 tables:

- ➔ Fact tables **Orders** & **Order Items**
- ➔ Dimension tables **Customers** & **Products**

6 months observation period: **January** through **June 2016**

61 499 orders with total of **136 774** order items, made by **30 302** unique customers

Data extracted on ~ 17/12/2019

Data Preparation and Assumptions
Orders table

Outliers and excluded data

- Orders with **order subtotal** of above 8000
- Orders with **order total** of above 2800
- Customer with **ID** 32725308 *(potential fraud)*

Main Dimensions

- **Payment methods**: 'Boleto' & 'BOLETO' grouped as same payment method
- **Order status**: All 'Cancelled' considered for cancellation rate, without taking into account combinations of flags **is_placed_order** and **is_shipped_order.**

New Fields

- Boolean **payment_method_b:** for comparing Bank tickets (True) vs. all other payment methods (False)
- **GMV**: order total where is placed order = True
- **Revenue**: order total when is_shipped_order = True
- **Order month**

Data Preparation and Assumptions
Order Items

New Fields

- **Profit margin:** (promotional price - product cost)/ promotional price

Cancellation Rate

Cancellation Rate

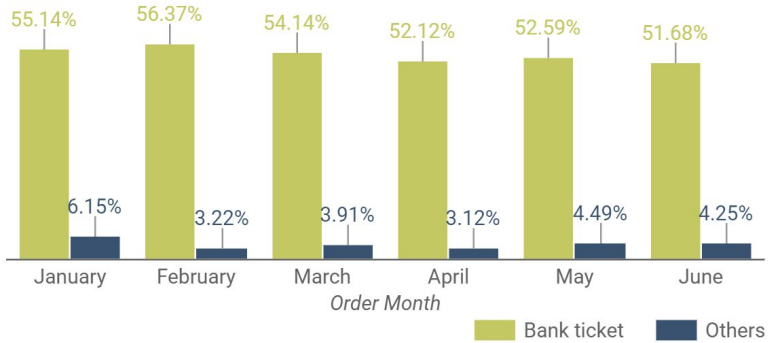
Order-based approach to calculating cancellation rate:

Cancellation Rate - cancelled orders in any given month divided by total number of orders in that month.

- ➔ More than half of all bank ticket orders get cancelled!
- ➔ Cancellation rate of all other payment methods is considerably lower than that of bank tickets.

Overall cancellation rate
in the observed period

Cancellation rate -
Bank tickets compared to other payment methods



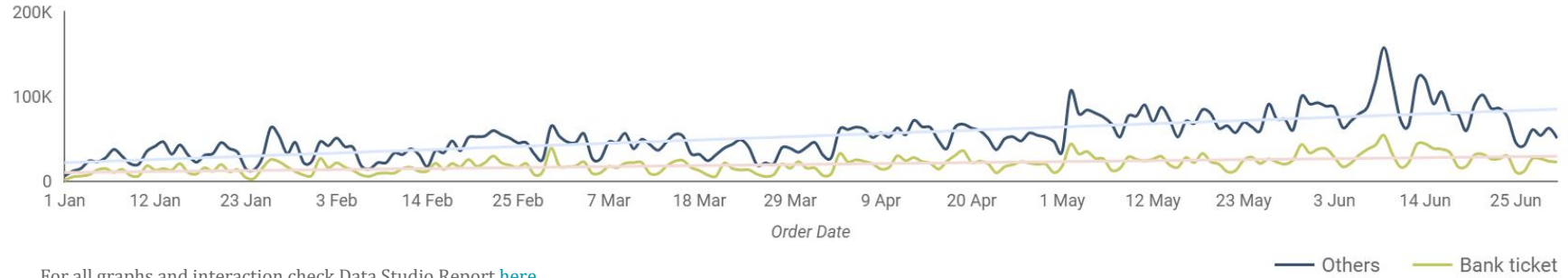
For all graphs and interaction check Data Studio Report here.

Gross Merchandise Volume

GMV - total sales value. Commonly used in online retailing

In AMARO terms:
GMV - total value of placed orders
Revenue - total value of shipped orders

Throughout the observed period, the GMV is quickly growing for all payment types, however, the growth rate of GMV from orders made with bank tickets is substantially lower (trendline below)

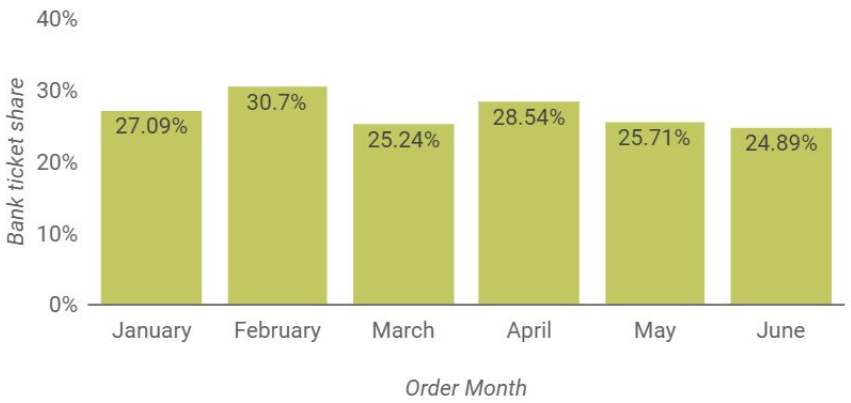


For all graphs and interaction check Data Studio Report here.

Bank tickets

Bank tickets represent a large share in all orders throughout the observed period, and account for a large share in GMV

Share of bank tickets in all orders



Share of bank tickets in GMV

For all graphs and interaction check Data Studio Report here.

Lost Revenue

due to cancellations

328 th. R$

monthly
estimated lost revenue
from bank ticket cancellations*

=18%

of total monthly revenue

- ➔ **Almost 90%** of lost revenue from cancellations is due to bank tickets
- ➔ Lost revenue grows marcably over the observed period, the monthly average almost doubles when comparing first and last months.

* under a rough assumption that each bank ticket cancellation results in lost revenue, without considering replaced orders

For all graphs and interaction check Data Studio Report here.

Hypotheses of analysis

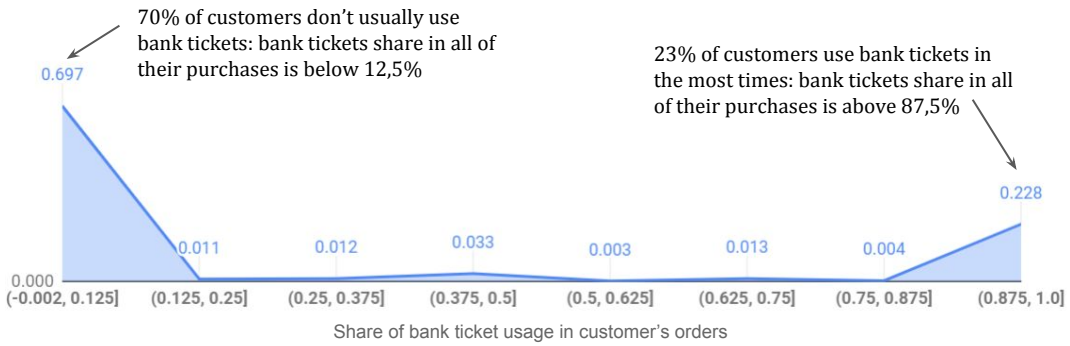
Hyp 1. Customers that prefer to pay with bank tickets are the same ones

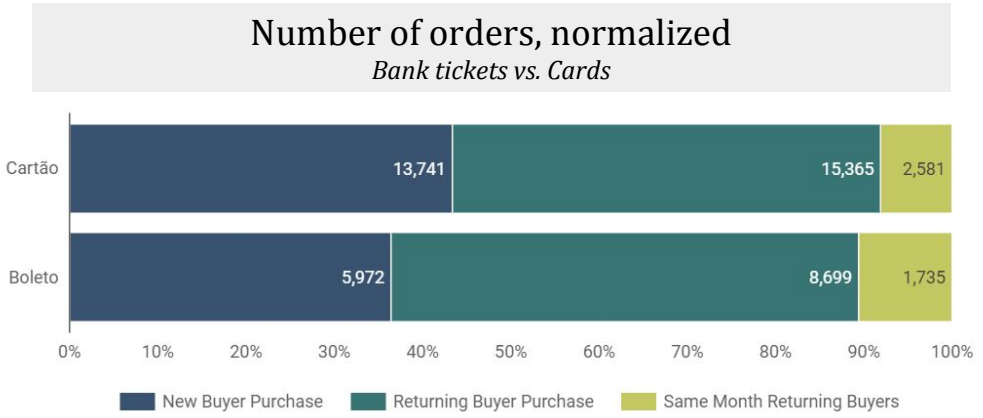
1. Analysis of preferred **payment method** (either bank ticket or others) has shown that **93%** of users don't normally switch between payment methods.

2. Another way to analyze whether bank ticket users are generally the same ones is by looking at customer categories *(is_purchase_of_new_buyer)*.

When compared to **Cartão**, bank ticket orders have a much lower share of **New Buyer Purchases**, indicating lower volatility of customers that choose to pay with bank tickets.

→ These insights allow to observe the 2 groups of orders - Bank Ticket vs. Others and their properties and compare them directly, without deepening the analysis to interdependence of bank ticket and non-bank ticket orders on the level of each customer.





For all graphs and interaction check Data Studio Report here.

Hyp 2. Cancelled orders placed with a bank ticket are followed by the next order sooner than orders made with other payment methods

Initial idea:

- Maybe some customers placed their order with a bank ticket mistakingly and therefore immediately replaced it with the next one.

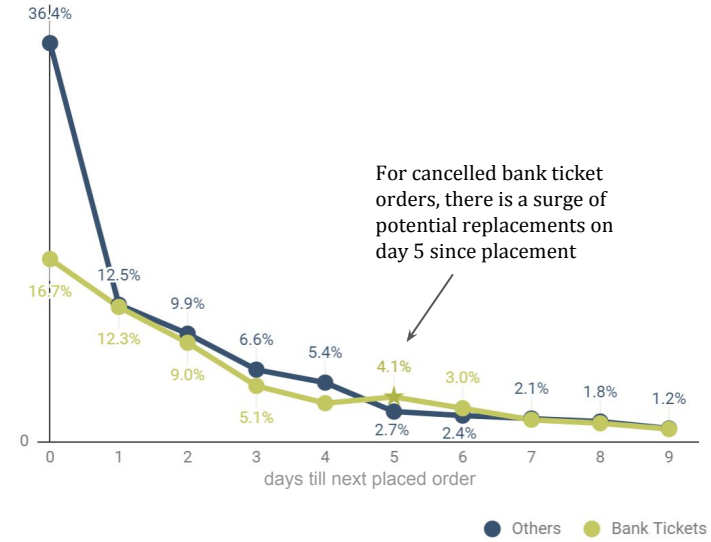
However, the graph contradicts this idea.

For cancelled orders with a payment method other than bank ticket, **36%** of following orders **were placed on the same day**, and with bank tickets - only **17%** were placed on the same day.
Probably this happens because a cancellation of a card payment is immediately visible, and the repeated payment is already counted as the next order → *needs further investigation of the website/app*

It's interesting to note though how the curve for bank tickets breaks on day 5, with a visible surge in potential replacements.

Most likely this is the day when the initial order is actually cancelled, incentivizing an **order replacement**.

Cancelled orders by day till next placed order
Bank tickets vs. Others



Hyp 4-6. Cancelled orders placed with a bank ticket are different from cancelled orders made with other payment method in terms of AOV (H4), discounts (H5), or profit margin (H6)

Payment method / AOV		
Order Status	Cartão	Boleto
DELIVERED	265.73	220.51
CANCELLED	419.93	224.62

Payment Method / Order Discount		
Order Status	Others	Bank ticket
DELIVERED	37.92	44.29
CANCELLED	45.32	13.68

Payment Method / Profit Margin		
Order Status	Others	Bank ticket
DELIVERED	59.48%	52.32%
CANCELLED	50.80%	53.19%

First of all, it's important to note that when it comes to cancelled orders, the AOV of bank ticket payments is actually significantly lower than that of the most common method of payment - **Cartão**. Therefore, we can't say that its the high AOV that pushes bank ticket customers to cancel their orders.

Next, it's also clear to see that there is no significant difference between the AOV of cancelled and delivered orders made with bank tickets.
This hypothesis can't be accepted.

For bank ticket orders, the total discount is lower in cancelled orders than it is for other payment methods, and vice-versa for Delivered orders: the discount is higher in bank ticket orders than it is in orders with other payment methods.

However, these differences don't appear to be significant, inviting further investigation.

The results are pretty similar to those that we saw in the AOV analysis: when it comes to cancelled orders, the orders paid with bank tickets appear to have had a slightly higher profit margin. However, when we compare successful orders - those with status **Delivered**, non-bank ticket orders seem to have a slightly higher average profit margin.

In whatever case, the differences don't appear to be very suspicious and might turn out to be due to variation in the sample.

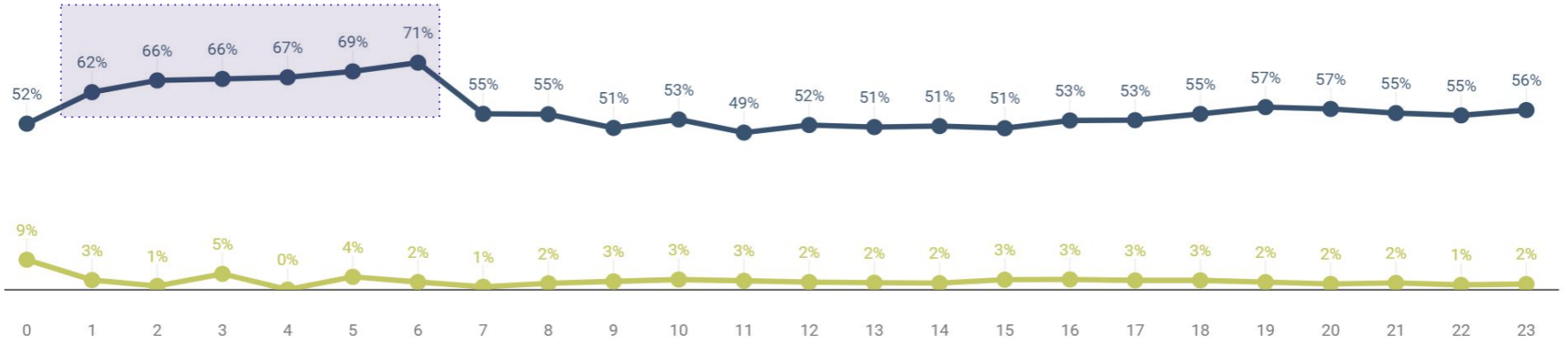
* For the analysis, bank ticket orders were compared only to purchases made with 'Cartão', as the AOV was very volatile with several other categories, especially due to paying in several installments

For all graphs and interaction check Data Studio Report here.

Hyp 8. Bank ticket cancellations are more common in certain times of the day

It appears that there is quite a large fluctuation in the cancellation rate during the night time for the orders made with bank tickets. With the cancellation rate hitting its absolute maximum at 6 am - 71%, and being on an increasingly high level during the hours 2-5 am.

All other methods of payment, on the other hand, don't follow this same pattern. The cancellation rate for them is slightly higher at the night time, but we can't observe such a visible peak like in orders paid by bank tickets.



Further directions of study: Data analysis

- ➔ Study the outlying customers to check whether we are dealing with fraudulent activities (outliers in terms of average order total)
- ➔ Explore differences in customer groups: compare the properties of customers who predominantly use bank tickets (>85% of customer's orders) with those who don't
- ➔ Coincidence in order items of subsequent orders made with different payment methods and with a small gap in between them (less than 2 days). Compare order items in each order, label orders as 'replaced orders' after a certain threshold of coincidence, explore the patterns of replacement.
- ➔ Observe dynamics of bank ticket purchases and cancellations on different week days
- ➔ Study properties of customers that place orders in the frame 01:00 - 06:00 am and don't pay them
- ➔ Applying ML to try to predict whether bank ticket will be paid or not (see next section)

Further directions of study: Experiments

- ➔ Run a test with some regular bank ticket customers: allocate two days instead of just one till bank ticket expiration date:
  - \+ If most customers 'forget' to pay the bank ticket the next day, increasing the period by one day might significantly decrease cancellations
  - \+ 1 day bank tickets are rare on the Brazilian market
  - \- Might be incompatible with AMARO's extreme agility and stock minimization, as items are reserved for more than one day
- ➔ Run a test with other payment conditions for orders made in the period 01:00 through 06:00 am
- ➔ Run a one-question survey with clients whose bank tickets were cancelled. In the cancellation email (day 6), instead of this question of the in-built survey

Ask:

What was the main reason you couldn't pay the bank ticket?

Other recommendations

UX and process recommendations based on a test purchase on the website

Beyond the insights and action plans from the data exploration, do you have any suggestion from the business point of view on how to improve this process?

1 Placing an order

- ✓ Easy to select bank ticket as a payment option
- ✓ Conditions well-explained, good management of customer expectations
- ✓ The default payment method is card

2 Order Confirmation

Should rather be *‘1 dia útil a partir da data de confirmação do pagamento’*

Should rather be highlighted. Too little attention drawn to the fact that the ticket still needs to be paid, check mark in the top also contributed to this confusion

Should probably say *‘Visualizar boleto’*

3 My orders

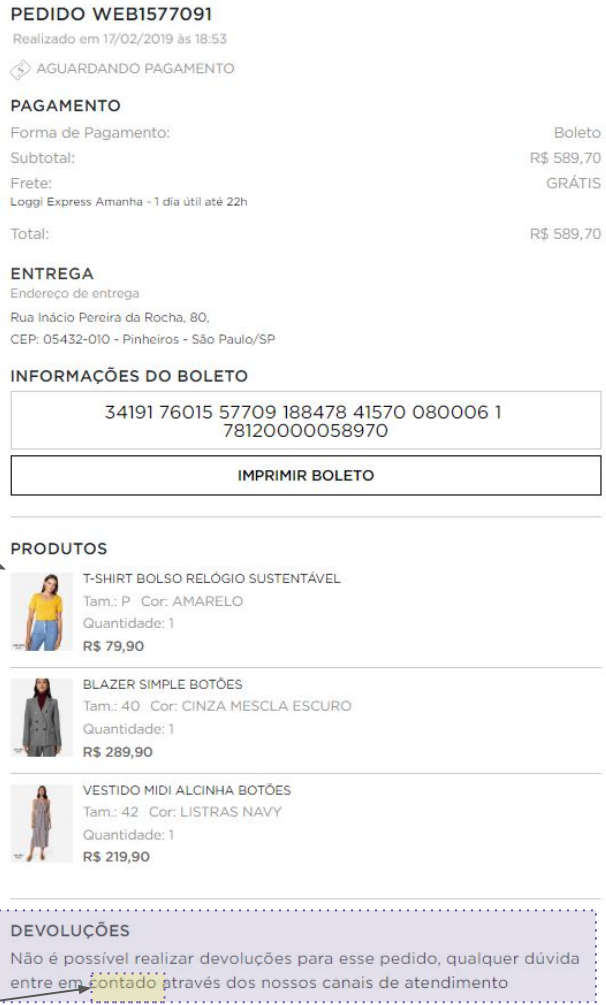
Day 0 - right after placement

✓ Names of the products are clickable

However, would be great if the whole ‘cell’ of the product info led to that product’s page

➔It’s not clear whether it’s not possible to return this purchase because it’s made with a bank ticket or for some other reason. If the former is true, should be stated on previous stages!

➔It should also be mentioned somewhere that in order to cancel this order it’s enough to just not pay the ticket



A little typo ;)

4 My orders

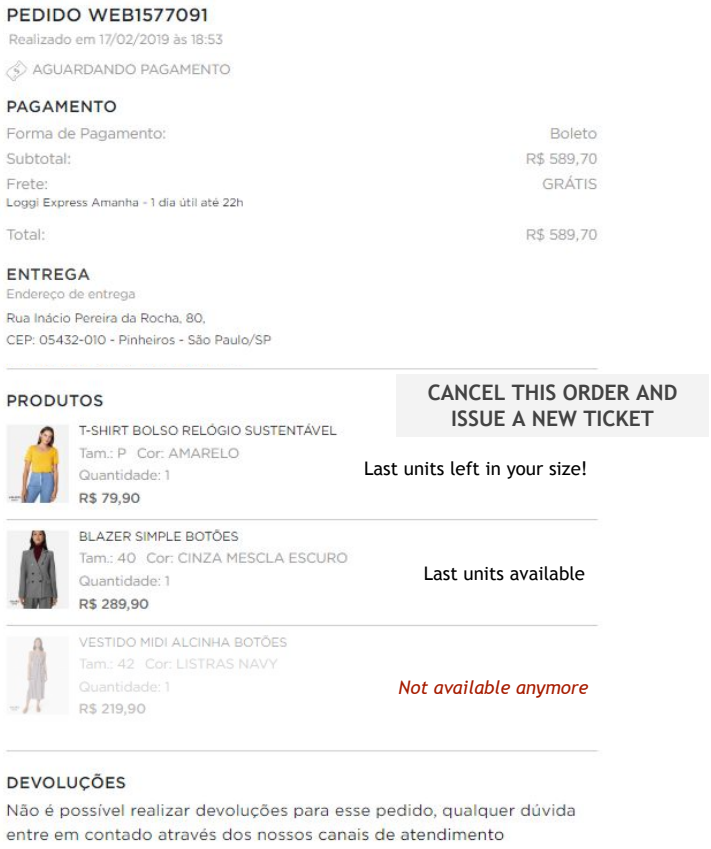
Days 1-5: Aguardando pagamento

Purchase was made on 17/02 - *day 0*

On days 18/02-22/02, the option to ‘Print bank ticket’ was still available, even though the bank ticket has expired and it can’t be paid once expired.

Probably should be available for visualization and not for printing.

Add button **CANCEL ORDER AND ISSUE A NEW BANK TICKET** to be available on any day starting from day 1: attempt to replicate the same shopping card, only excluding the items that aren’t available anymore.



5 My orders

Days 6 and on: Cancelled

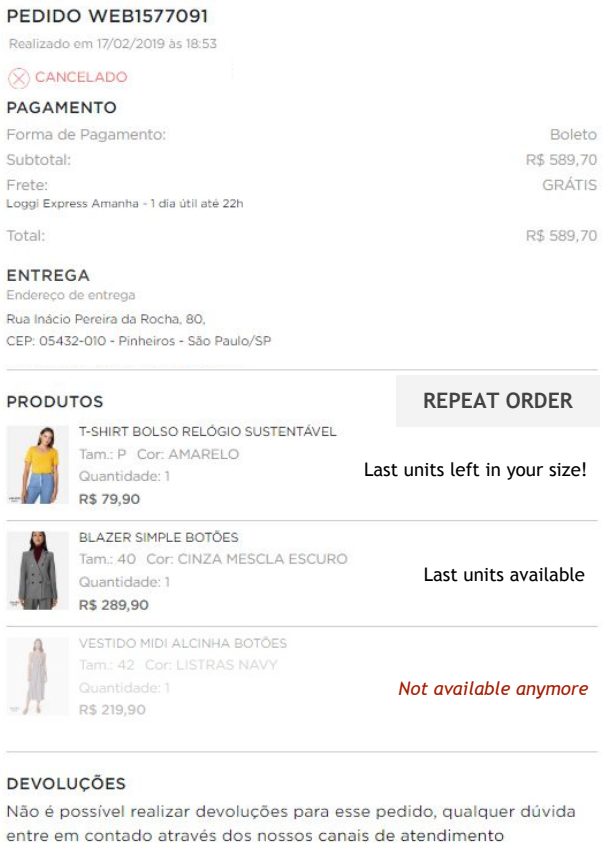
Purchase was made on 17/02 - *day 0*

On days 18/02-22/02, the option to ‘Print bank ticket’ was still available, even though the bank ticket has expired and it can’t be paid once expired.

Probably should be available for visualization and not for printing.

Add button **REPEAT ORDER** to be starting from day 6, once the order is officially cancelled - facilitate order replacement

Goes in line with findings discussed in Hyp.2: surge in subsequent orders on day 5 - Once the customer knows for sure that the order is cancelled, she replaces it ← *needs further investigation*



ML Application

Do you imagine that it would be possible to apply machine learning to solve any of the issues found in the data exploration?

ML Application for further investigation: two lines of analysis

Supervised learning can be applied to further explore characteristics of:

orders made with bank tickets	customers that use bank tickets*

by dividing them into following classes in the training set:

- delivered orders - cancelled orders	- customers that usually pay their bank tickets - customers whose bank ticket orders get cancelled

and then building a **K-nearest neighbors classifier** based on:

k parameters of orders	k parameters of customers

to find out how well the classifier will predict the proximity of each item in the validation set to this or that class. Repeat the stages for another set of k parameters to find the best classifier.

* For this line of analysis, we'll first need to select a threshold that defines that user can be considered a regular bank ticket user

Conclusion

My main educated guess based on results of analysis is that:

- ➔ The majority of customers forget or don’t have the time to pay their bank tickets on the next day, and the bank ticket ends up expiring.
- ➔ Several improvements in the UX of placing and replacing bank ticket orders (some of them being quick-wins and others relatively hard to implement) might help speed up order replacement for such orders