Optimizing Dalinde Hospital's Pharmacy Stock

Mexico City | Team #8

Alejandro Córdoba, Roberto Corona, Emmanuel Hernández, Gerardo Ruiz, Esteban Marín, Victor Martínez

Business Problem

Located in the heart of Mexico City, in the Roma neighbourhood, Dalinde Medical center is a tertiary referral hospital with more than 70 years of experience supplying healthcare services.

One of the main problems identified with the Head of Pharmacy, Julio Pérez, is the inventory stock of patented medicines that consistently run out in their storage. Although Dalinde buys two drug families: patented and generic drugs, Mr. Pérez, points out that the hospital does not have a problem with the latter. This issue is very recurrent, and directly affects the treatment of patients, for example, it impacts scheduled surgeries and postpones them. When a medication is not in the pharmacy and it is urgently required, a resource from the Pharmacy department is appointed to its work post to find the medicine from the retail provider directly. This generates unforeseen expenses, and an excessive price of medicines. In addition to reducing the quality of services and creating a bad reputation for the hospital. The solution they seek is to be able to predict scarcity, find what factors it correlates with and anticipate it to avoid monetary losses.

Business Impact

With this analysis we pretend to lower the monetary cost of drug shortages. During 2019, **shortages increased the cost of a drug** that was not provided by the Hospital suppliers and had to be bought by the pharmacy independently.

By developing a prediction model that suggests timely purchases in order to avoid shortages, **<u>we expect to lower the average cost of a shortage</u>** below 30% . In addition to saving the cost generated by sending an employee to look for medicines, that is: salary, time, transportation costs, etc. Not counting the additional benefits of performing treatments or surgeries as scheduled.

Data

The provided datasets are reports retrieved from the IT system of the hospital. At the moment we have two datasets available containing the following fields:

Requested Orders vs Purchase Orders, from 2019 to 2020 (YTD):

- 56,000 records approximately (best estimate)
- Requisición (String): request number in-system
- Fecha (DateTime): request number in-system date timestamp
- Orden Compra (String): purchase order received at time of delivery to pharmacy storage
- Fecha Pedido (DateTime): request number in-system date timestamp
- Fecha Entrega (DateTime): delivery date needed for request
- Proveedor (String): supplier, registered legal name
- Artículo (String): medicine article code
- Descripción (String): chemical compound + amount + package presentation
- Salidas (Float): inventory stock level?
- Cantidad Requerida (Float): requested amount by Pharmacy
- Cantidad Pedida (Float): requested amount by Procurement

This data is specifically of the patent medicines handled by the hospital's Pharmacy. When the buying and delivering cells are empty, it means the Procurement department could not find a supplier and did not fulfill the requested order at the time. The presence of this empty data is a relevant indicator as it suggests a lack of drug supply. We consider that this dataset provides sufficient and precise data for us to develop a successful project.

RH System (orders outside procurement cycle due to out-of-stock), from 2019:

- 93 records
- Medicamento (String): description of the medicine
- Precio HR (Float): price HR
- Precio Pedido (Float): established price
- Diferencia (Float): price difference
- Fecha (DateTime): date of purchase

The RH system is a procedure the pharmacy staff use to cope with the lack of drug supply; that is, whenever the purchasing department is not able to acquire the requested drug from their reliable suppliers, the pharmacy staff then buys the drug from a retail supplier, thus increasing the cost of the purchase. This data is stored in this dataset.
This dataset has a relatively small amount of data since we only have access to the data from January to December 2019 at the moment. However, the complete data from the whole previous year will be provided soon.

Apart from the datasets described above, there is still a substantial amount of data not available to us at the moment to which we will have access in the following days. These datasets are:

Drug request history, from 2018 to 2019:

- Requisition (String): requisition number in-system
- Date of request (DateTime)
- Healthcare unit (String): unit or speciality where the patient is being treated
- Drug (String): name of the requested medicine and its dosage
- Quantity (Float): units of medicine requested
- Measuring unit (String): unit in which the medicine is measured.
- **Amount Consumed**: Amount of patent drugs consumed in a period of time
- **Amount Requested**: Amount of patent drugs requested by the Pharmacy apartment to the Purchase apartment.
- **Amount Purchased**: Amount of patent drugs requested to third providers (laboratories, distributors or pharmacies)
- **Amount Missing**: Amount of patent drugs not purchased because there was no availability in the market.

The dataset could serve to understand patient profile needs, their seasonality, current inventory levels, actual medicines dispatched, and shortage of medicines that Pharmacy requested to procure. However, this data has a monthly periodicity, which might reduce the time resolution of our data analysis when comparing the request history with the purchasing data. Despite this minor issue we are convinced that our analysis will still be precise enough.

Exploratory Data Analysis

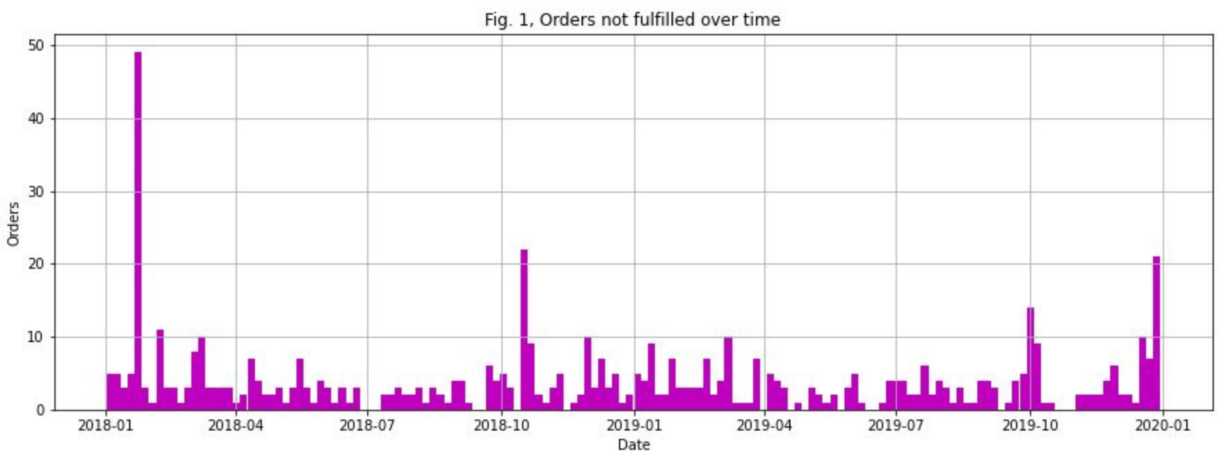
Databases

Our exploratory analysis begins with the medical and drug databases given by the Pharmacy Manager. The records are for a two year period from January, 2018 to December, 2019:

- Drug consumption by patients on a monthly frequency (also includes unitary cost sold).
- All Purchase orders by the Pharmacy department with the amounts requested, and amounts procured.
- Incomplete purchase orders, identified by the Pharmacy department.
- Patent drugs dictionary, engineered by our team, based on the Consumption and Purchase Orders to identify all patented drugs.

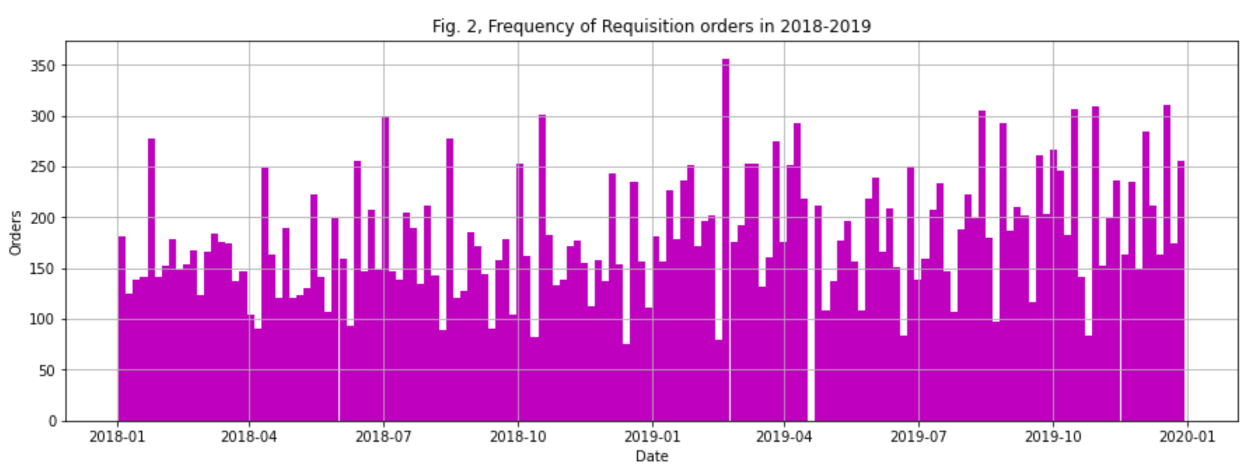
Analysis

Given the understanding of shortages by our stakeholder, we identified in the Purchase Order report, as a business rule, that the column "Amount Purchased" when equal to 0, it means that although there is a row for every requisition, this field is empty because the Procurement department could not fulfill the request and its required amount, so this translates as a shortage. In Figure 1, we look into how unfulfilled requests behave over time.



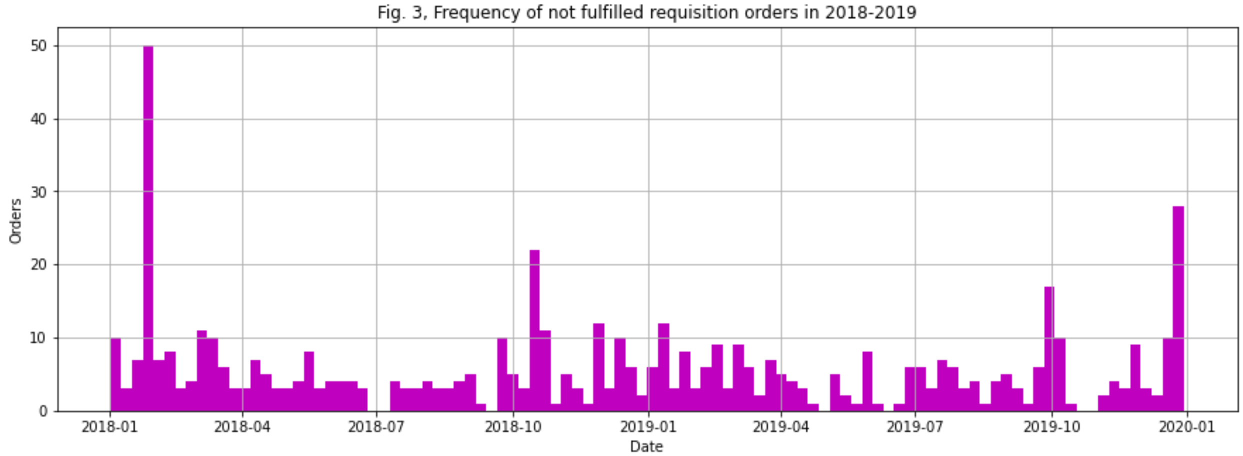
We used the Patent Dictionary to filter the Purchase Orders (PO) DataFrame to only keep those purchase orders that were for patented drugs given that the original PO records are for common medical supplies, and generic drugs.

Now that we only have purchase orders for patent drugs, we analyze through a histogram the frequency of requisitions through time to understand the needs of the Pharmacy department towards what the medical team (doctors and nurses) are requesting ultimately. In Figure 2, we see all requisitions for patent drugs, including fulfilled and not fulfilled.



In Figure 3, we used the filter 0 under Amount Required, again, previously identified as shortage business rule, to the purchase orders report for only patent drugs to understand the overall behavior of these drugs that are not fulfilled over time.

Fig. 3, Frequency of not fulfilled requisition orders in 2018-2019



Our focus on this analysis is to narrow down the most problematic medicines when being procured, so we want to understand the top 10 drugs that have not been fulfilled over time. In Figure 4, we identified the top 10 drugs that have had the most unfulfilled requests while in Figure 5, we also look at the volume of the top 10 patent drugs that were not fulfilled.

Fig.4, Top 10 Patent Drugs with Most Shortages on Requests over time

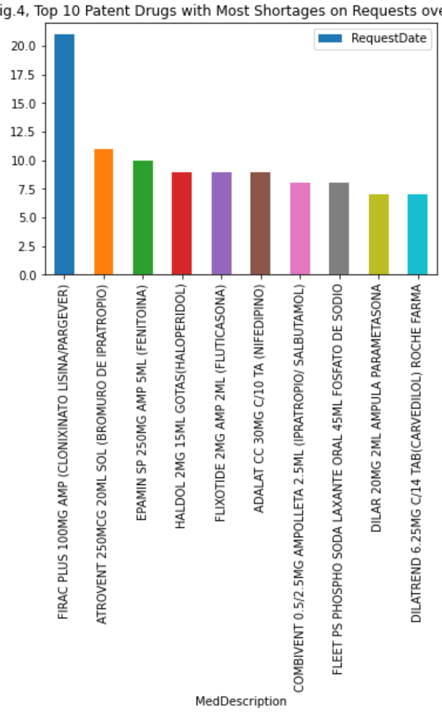
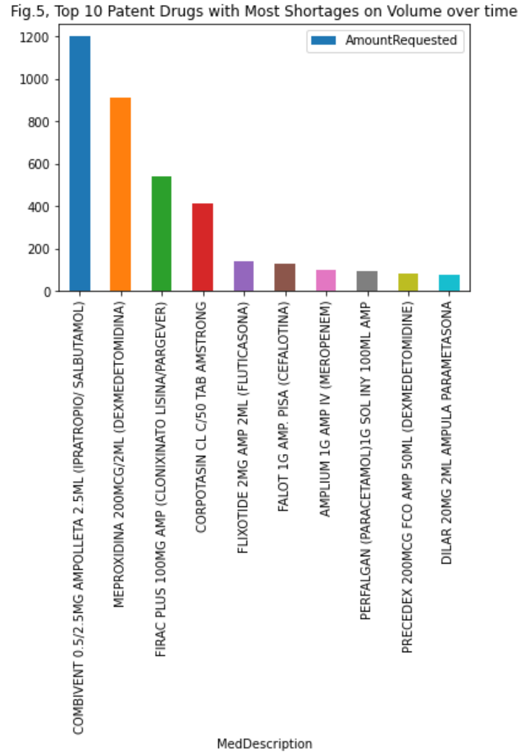
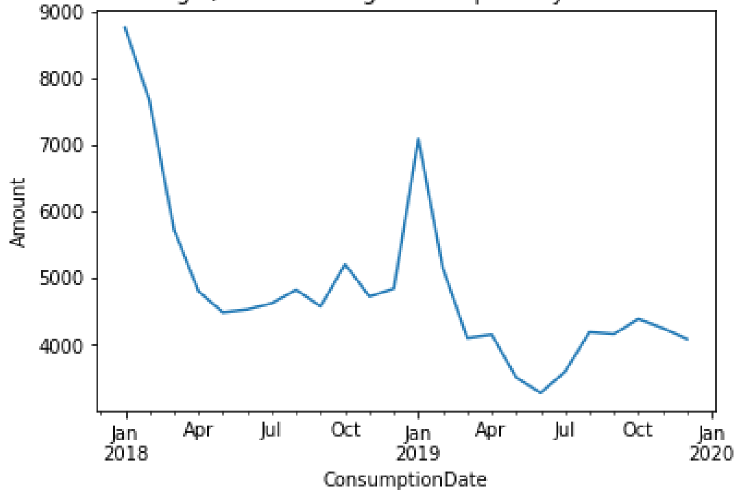


Fig.5, Top 10 Patent Drugs with Most Shortages on Volume over time



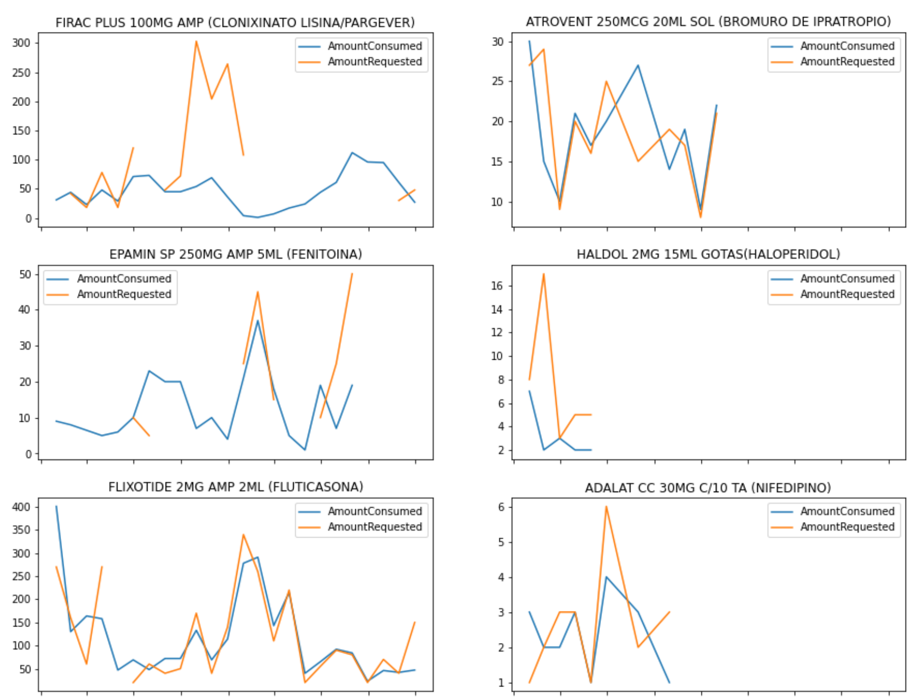
To understand the demand from the Medical area, we analyzed the patient consumption by volume over a two year period. Figure 6 shows the drug consumption behaviour in 2018 and 2019 where we see a specific periodicity for high demand starting in January, and decreasing in the first half of the year to later be picked up again in July, which holds true for both years.

Fig.6, Patients Drug Consumption by Volume



After studying the requisition, purchase and consumption data of these relevant drugs, we realized that some of these drugs are missing values in different months across the span of two years. Therefore, we proposed to filter the drugs by number of data points, and consider those that were consumed uninterruptedly throughout the two years.

Fig. 7 Top 10 Drugs with Most Shortages, Consumption vs Requested



In figure 7, we observe the 10 drugs that satisfied our previously mentioned criteria. After analyzing these top 10 drugs (using their brand and patent names) that had most shortages, we observed that there is not enough consumption and purchase order data, e.g.: Haldol was requested and consumed massively only in two months in 2018, and it stopped from being requested and consumed.. We also estimate that these inconsistent medicines probably do not cost the most to the Pharmacy department.

Given the latter, and on suggestion from Theresa Gebert, our project advisor and Teaching Assistant, we continued to narrow down our analysis and decided to shrink the universe of observations using the following criteria:

a. **Analyze by chemical compounds** and not by brand, patent names. By taking this way, we would look into the same compound regardless of different presentations. This is also an angle that Dalinde's management pointed during Meeting One could be of interest since they could look into bulk buying.
b. **Consider only those chemical compound medicines that have been constantly administered in the last 24 months.** In this case, we found over 90 chemical compounds.

Then, we analyzed the cost of those compounds which were more consumed by patients (using unitary cost sold to patients), in which we found **15 medicines with a total cost of $6'723,694 MXN ($353,878 USD)**. This represents 52% of the chemical compounds universe that were demanded uninterruptedly during the 2018-2019 period; and they also represent 19.44% of the total patent compounds from the same period.

Out of these top 15 compounds with the highest monetary value, we observed which had shortages at least once in the delivery of the suppliers, from those observations, we got 11 compounds that satisfied our criteria. The 11 compounds are listed in the Table 1:

Table 1: Compounds that have the highest monetary value and have reported at least one shortage

Compound Name	**Usage**
IPRATROPIO/ SALBUTAMOL	Respiratory system
VANCOMICINA	Anti-infectives for Systemic Use
BROMURO DE IPRATROPIO	Bronchodilator
NORFLEX	Musculoskeletal system
CO DIOVAN	Antihypertensive
KETOPROFENO	Musculoskeletal system
EMEND	Alimentary tract and metabolism
ALBUMINA	Blood and hematopoietic organs
FLUCONAZOL	Anti-infectives for Systemic Use
CIPROFLOXACINO	Anti-infectives for Systemic Use
CLARITROMICINA	Anti-infectives for Systemic Use

The consumption of the compounds listed in Table 1 adds up to $ 5’240,019 MXN ($275,790 USD) , which represents 41.25% of the 90 compounds that were demanded during the 24 months constantly; and represent the 15.15% of the total patent compounds in the period analyzed.

In order to give a monetary value to the shortages that the 11 compounds reported, we calculated the amount of medicines that were requested and not fulfilled, we named this metric “**Amount Requested and not fulfilled**”. Later on we multiplied the **Amount Requested and not fulfilled** times the average unitary cost reported for each of those 11 compounds, and got the **Shortage Value.** We can see these results on Table 2.

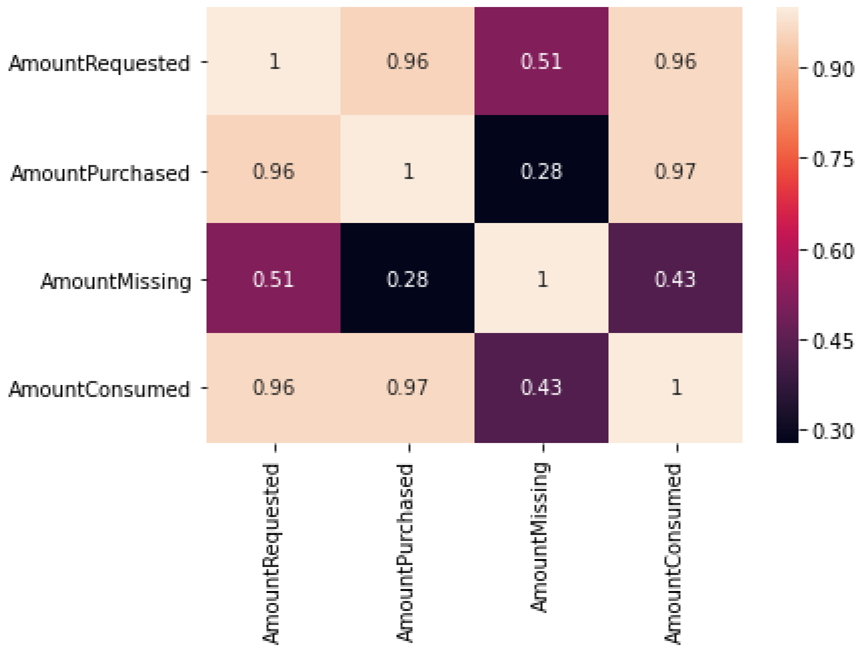
Table 2: Shortage Value of the 11 Most Consumed Compounds

Compound	Amount Requested and not fulfilled	Unitary Cost	Shortage Value
IPRATROPIO/ SALBUTAMOL	1,205	$35.80	$43,142.13
VANCOMICINA	55	$457.75	$25,176.37
BROMURO DE IPRATROPIO	39	$502.68	$19,604.49
NORFLEX	25	$317.81	$7,945.35
CODIOVAN	4	$1,217.91	$4,871.63
KETOPROFENO	60	$75.36	$4,521.48
EMEND	2	$2,215.05	$4,430.11
ALBUMINA	4	$578.67	$2,314.69
FLUCONAZOL	4	$568.43	$2,273.73
CIPROFLOXACINO	2	$311.83	$623.66
CLARITROMICINA	1	$436.54	$436.54

As we can observe from Table 2, the total cost of these shortages add up to $115,340 MXN ($ 6,070 USD).

In order to find relations among our time series we computed the existing correlation between our 4 time series, the results of this computation are shown in the following Correlation Matrix:

Fig. 8 Correlation matrix between the numerical values of the top 11 compounds

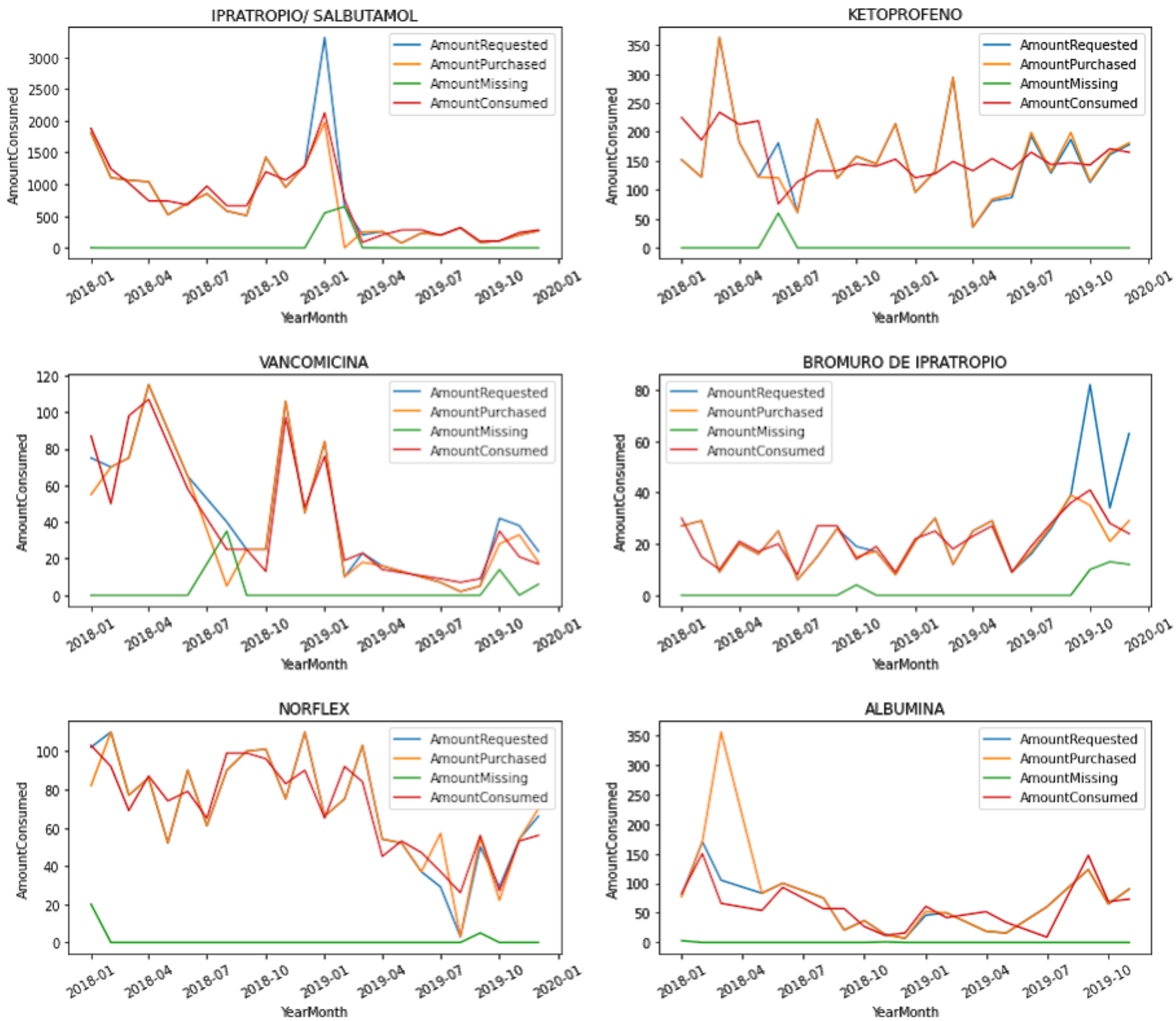


From the Correlation Matrix we can observe the following:

- There is a strong correlation between the amount of patent drugs consumed and the amount of patent drugs requested 97% .
- There is a slight correlation between the amount of patent drugs consumed and Missing Drugs, 43%
- There is a strong correlation between the amount of patent drugs consumed and the amount of patent drugs purchased 96% .

In order to visualize the latter Correlation Matrix, we plotted the 4 time series. The following graph (Fig 9) shows the behavior of the top 6 compounds from the 11:

Fig. 9 Line plot of numerical values of the top 11 compounds over time



As we can see, the most correlated time series for these compounds are amount requested and amount consumed, which strengthens our conclusion from the Correlation Matrix.

Next steps

1) Calculate the value of over-supply. At the present results we saw that this is a very important quantity and a possible loss of money, because drugs occasionally are supplied when it is not necessary.
2) Begin with the ARIMA modeling and Naive Bayes. At this point we have enough knowledge and information to begin the modeling.

3) Begin with the research on the PID models and how they are applied to inventory. This is a new idea and the expertise is only applied to electronic control, by this we have to learn about it and test it over our dataset to see if it can give us the results we are looking for.
4) Show the recent results to Dalinde hospital to get feedback. At this point we think we have enough information that could be helpful to Dalinde Hospital.

Models

- **ARIMA (AR)**: The Auto Regressive Integrated Moving Average (ARIMA) model analyzes a time series process through its own past observations (lags), so it can be used to predict future values in time. It has three parameters: *p, q*, and *d*. The parameter *p* is the order of the AutoRegressive (AR) term; the parameter *q* is the order of the MovingAverage process, or term; the parameter *d* states for the times the series has to be differentiated in order to make it stationary. This model can be very useful to check whether the lags of the requisitions for the drugs can predict the next shortage.

Interface

Our first approach was too complicated for the course time, we were thinking about communicating PHP and Python with a SQL DB, we decided based on the recommendations of our T.A. to simplify this, and use it directly from Python querying a PostgreSQL DB, the engine for displaying our graphs and tables is Django, a Python a high-level Python Web framework that encourages rapid development and clean, pragmatic design.

The login display it's going to be the same that Django uses on their core functions, this to ensure the privacy of hospital information.

The graphic engine it's going to be those used in the jupyter notebook, matplotlib and seaborn, with that in mind, we prepared this example:

Second approach for a clean UI, we think about it as v1.1

The administrator is going to pull all the views we prepared for him with the selector on the left, and then a different view of the same topic on the same topic on the submenu below the title bar.

With this approach we use the built-in power of Python to manipulate data, and with a lightweight UI that is better for the purposes of this project.

Milestones

Our project consists of two versions and will depend directly on the availability of the data complement and the time we have for its solution.

Version 1. A prediction model that determines what requisition is optimal according to the shortages that Dalinde Hospital has had over the last two years. This model includes data analysis and visualization of the requirements and needs of Dalinde's pharmacy. The probability to complete this version is 100%.

Version 2. A prediction model that includes external factors that influence pharmaceutical area operation, such as diseases, seasonality, number of patients, types of patients. The probability to accomplish this version is around 50%.

Timeline

Date	Deliverable	Details
Week 1	Team formation Environment setup	
Week 2	One-page summary Workflow setup Trello tasks	
Week 3	Scoping document Data access	
Week 4	Github training for the team Data cleaning Initial data exploration Q&A with stakeholder and NDAs	Victor, Ale: Github setup and review Esteban, Ale: Q&A, and NDAs ALL: Data cleaning and exploration.
Week 5	Continue data exploration	Review with Theresa.
Week 6	Advanced data exploration Initial modeling UI Wireframe	Victor, Ale, Roberto: UI development Emmanuel, Esteban, Gerardo: Modeling
Week 7	Continue modeling Application on cloud and testing environment	Emmanuel, Esteban, Gerardo: Modeling Victor, Ale, Roberto: Cloud application and testing environment
Week 8	Front-end complete Advanced modeling Update Final Report	Victor, Ale, Roberto: Front-End Emmanuel, Esteban, Gerardo: Modeling
Week 9	Fine-tune modeling Fine-tune application Write Conclusions section of Final Report.	Victor, Ale, Roberto: Front-End Emmanuel, Esteban, Gerardo: Modeling
Week 10	Finalize presentation Finalize report Finalize application	

Concerns

1) *Concern regarding modelling:* After seeing the results obtained with the data plotted according to the required and consumed medicines, we see that it is important to have a model that helps us minimize money losses at two points:
    a) Shortage
    b) Over-supply

   We believe that these two points can be solved using a predictive technique such as ARIMA that is based on using information from the past to predict the future. And we also find it interesting to use a control technique widely used in engineering and that work has currently been carried out to apply it in inventory control, this technique is Proportional, Integral and Differential (PID) control. Which works at a high precision level for electronic components. Applying this technique we hope to have enough precision that minimizes the losses of money.
   In these modeling scenarios, our concern is that neither of the two previous models work to achieve our objective. This is because ARIMA requires highly correlated information and this is not the case for all the data we have. And PID is a technique that requires finding the appropriate adjustment parameters.
   We are going to handle the information as simple as possible by first fitting simpler models, possibly using Naive Bayes in a binary way, to start with a minimal prediction model, this by recommendation of our TA.
   Before trying out more complex models we are going to inspect the data to see if we can find the most easy way to get the desired result, but due our inexperience it is possible to lose some important scenarios.

2) *A medium risk concern is the implementation of the model in a live environment*. Not the technical delivery from us, but how IT is going to implement it. A solution could be to implement the mathematical equation, as a daily prediction update, and as another column in the existing system to see which stock prediction amount recommendation is better, and match results, ultimately, with the actual amount requested by the Pharmacy.