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 **on an average of 30%**[1].

By developing a prediction model that suggests timely purchases in order to avoid shortages, **we expect to lower the average cost of a shortage** 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.

[1] This 30% estimate was calculated using a first review of historical purchasing data and calculating the price difference between the price of medicines that have been sought with retailers and the prices that are usually obtained from the suppliers. The extra costs generated by the shortage are not registered, however it is estimated that they add 10% to that 30% calculation.

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

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.

Supplier profiles:

- Supplier (String): name of the supplier
- Telephone (Integer): telephone of contact
- Amount of days to deliver that are included in the inventory stock formula (float)

This dataset will be useful for matching the inventory stats with the corresponding Pharmacy suppliers, and then determining their reliability under different circumstances.

Methods

Descriptive statistics

The very first approach that we will take is to compute basic metrics on our data, i.e. the number of stockouts over the weeks, weekly number requisitions made by Dalinde's pharmacy, frequency of drugs requested, number of patients, etc.

Look for patterns in the data, i.e. seasonal behavior, are stockouts correlated with certain suppliers?, are specific families of drugs more propense to incur into a shortage?

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.

- **VAR:** Vector Autoregression (VAR) model, explains the process of a dependent time series variable with its own lagged values and the lagged values of other independent time series.
- **Sliding Window Method:** The sliding window method is a process that uses past time observations as input to predict the future observations by moving over a specific "window" of time. With this method we could analyze time blocks when the shortages occur.

Visualizations

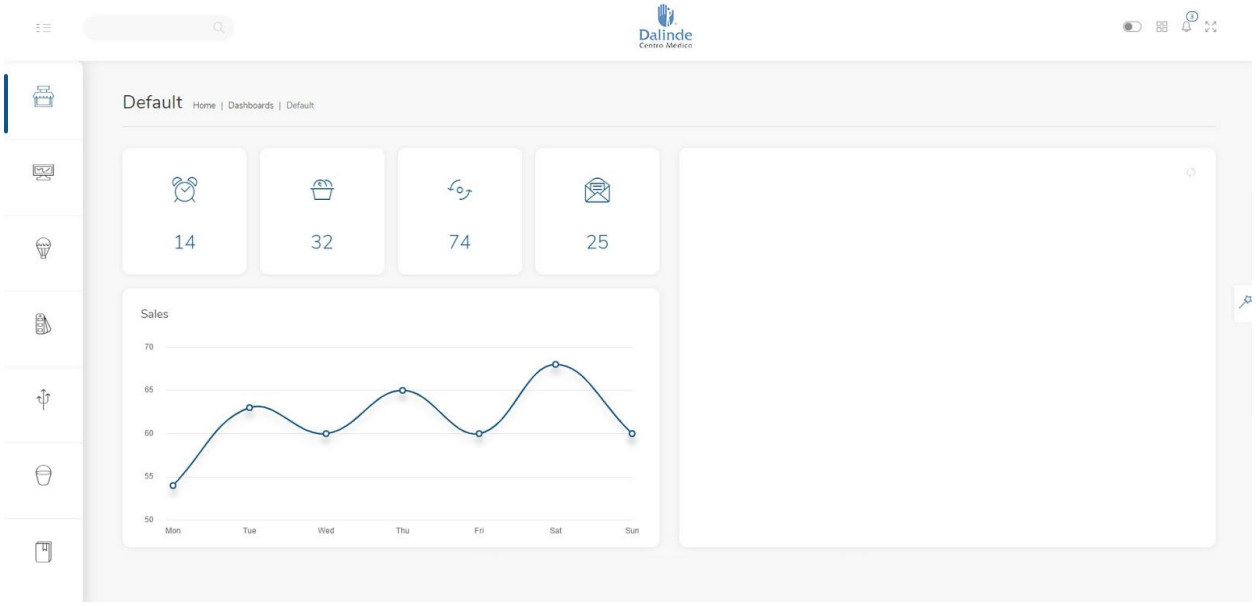
Because Dalinde's data is time series data, the next visuals will be useful to understand the problem better:

- Most and least requested drugs (in total and over time).
- Number of requisitions over time (daily, weekly, monthly).
- Number of stockouts over time (daily, weekly, monthly).
- Number patients (monthly).
- Correlation plots between requisitions, stockouts, and patients.

Interface

We designed the interface with the user experience in mind. We are also thinking of simulating queries to the database of their system and showing the results of the most recent status. The interface will not be public; instead, it is going to have and endpoint on a public IP on AWS.

It is necessary to establish a login interface because the information provided by the users will be sensitive and private, so it should not be accessed by third parties and it's not connected to the Hospital's system at all. We are going to simulate that the main user, in this case the manager of the hospital's pharmacy, will query his own system data with our data model. Then, our system will show him the data on different types of graphs and tables.



First approach for a clean UI, we think about it as v1.0

The user will be able to go through different options and become aware of the alerts from the suppliers that did not comply with the requests and other warnings provided by our predictive model. We believe the interface, as well as the system, should be clean and simple. This simple and clean architecture will be based on HTTP requests in order to give a fast and reliable experience to the user, with enough elements to satisfy his needs.

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	

*An ongoing activity will be to update this scoping document with EDA results.

Concerns

1) *The major concern about this project is that we don't get a good prediction model with the data that we are given.* There is a possible scenario in which the data we currently have access to has a low correlation with the problem we are trying to solve, which is the unexpected shortage in the Pharmacy stock. If this happens, we are counting on the hospital to provide us with additional data related to their patients and their anonymous profiles. We could introduce this data into our analysis as well, and build a prediction model with more independent variables. At the moment we don't have access to this data because patient information is more specific and it takes their system longer to retrieve it. The worst case scenario is that the hospital is not able to provide this data, and in this case we would need to explore other public sources that contain information about common illnesses in Mexican population over the last two years.

2) *Finding the best solution between performance and design*. It's possible that the model calculations will require most of the system resources, we have to find the best approach in order to give the user a good design and an exceptional functionality. We are thinking of a lightweight noSQL database for saving the data processed by the server side Python, and with PHP or Django/Flask solution to show the results to the user with JS Framework sharing some of the weight with the user's machine.

3) *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.