Team members:
Álvaro Rivera
Mateo Rodríguez
Camilo Ulloa

INTERACTIVE SALES DASHBOARD

RESOURCES

Libraries/Packages/Technology used: *Python, Jupyter Notebook, Pandas, Numpy,Plotly, Plotly Express, HTML, CSS*

Resources:

- *database with coordinates from European countries* https://www.listendata.com/2020/11/zip-code-to-latitude-and-longitude.html
- *Online parapharmacy and pharmacy product distributor sales database*

WORK METHODOLOGY

The development of our interactive visualization panel consisted of 3 main phases:

1. Data preparation and cleaning
2. Graph elaboration
3. Visualization panel development

I. Introduction to the problem

We have obtained data from an online drug sales company called *Atida Mfarma* in the year 2017. This data is made up of 3 different tables: products, categories and sales (items_ordered_2years). Our goal in this project is to simplify the information for the user and express it in a visually attractive way so that conclusions about the company's performance can be reached.

II. Methodology

Data preparation and cleaning

In this phase, we have imported all the files into VSCode by means of Python for their own exploration. We have verified that all the information we plan to use is complete and without redundancy in the fields. In the same way, we have found large errors in the “items.csv” table where the “zipcode” and “city” variables had wrong values, mixed or even empty fields.

To fix this problem, we have eliminated the columns that were not going to be used for our analysis and we have created a dataframe where all the non-numeric postal codes were included. ie all fields with wrong values or UK postcodes (contains letters and numbers).

Once this table has been created, we have grouped the number of errors to simplify the correction process and we have exported them in csv format to be analyzed one by one using Excel. Once all the incorrectly entered fields have been corrected, the csv file has been re-imported through Python and then joined back to its original table through an *inner join* (merge). Regarding the other two tables provided, these did not have major errors, so their manipulation was minimal.

Elaboration of graphs

The methodology that was used to carry out the elaboration of the graphs that we created to visualize the data was the following. First of all we needed to define which graphics library would allow us to better execute and visualize the ideas we had in mind. Therefore, we decided to go ahead by using the Plotly graphics library, since it gave us a wide variety of graphics from which we could choose and customize to get the best result. However, the process was not as simple as choosing a graph and executing the code.

For the first graphs, we decided to make a timeline that will show the behavior of the sales provided by the database. For this graph, the first thing we did was find the sales using the "items" table, in which we created a new column called "sale" by multiplying the price of the product, the quantity purchased and 1 minus the discount applied. Having this, the next step was to obtain the date of each sale without including the hour and minute. We then group the total sales (the sum) for each day. With this done, we were able to create the appropriate graphic to express our ideas.

The process we follow to create the second graph is supplemented by some of the steps used for the first. For this new graph, the idea was to show the average sales during the different hours of the day. For this, the first step was to extract the time of each of the dates of the orders. Then, we got the average sales grouped by the results of the previous step. Having the required information, we create a bar chart and customize it to make it look its best.

With the third graph we decided to focus on visualizing sales by category. For this we use the “items” table and join it with the “products” table through the “product_id” field. Having this, the next step was to get the total sales obtained by each category (“analytic_category”). Following this, using the data already obtained to create

Finally, we have created a geographic heat map of total sales by zip code. For this, we have had to use the external database mentioned above, in order to obtain the coordinates of each postal code attached to each sale made. Then we've totaled all sales and grouped them by zip code. In order to have only the coordinates of those postal codes registered, we have joined the table of postal codes with the table of aggregate sales. The resulting data frame was graphed using Plotly Express using the total sales “Total__sales” as magnitude variable.

Dashboard

For the development of the dashboard, we decided to use the Plotly extension called Dash, which will allow us to create a dashboard on a web page. Once all the requirements have been installed and with all the graphs previously created with Plotly, we proceed to create

the Dash app. The design of the app is divided into two parts; The first was based on the creation of cards that were going to show us specific data of the pharmacy, such as the total sales or the sales by month for the year of analysis; and the second, where all the previously made graphs are included. Everything is included in a *container*, which is divided into rows and columns, to facilitate the orderly grouping of the cards, the graphics and text boxes.

First of all, we create three charts that contain the total product sales, the approximate sales per month and the three categories most sold by the pharmacy. This was done in order to make it easy for the reader to view the summarized information on Atmira's sales. Secondly, we add the four previously created graphs together with text boxes on their right side, which offer a short description with the most important aspects to be rescued from each one. Finally, we added a template offered by Dash that allowed us to design the display panel, creating a more friendly and formal appearance.

III. RESULTS OBTAINED

In the graph of sales over the days we can see three areas in which there is a peak in sales. Based on these three “peaks”, we could infer that there was possibly some marketing campaign implemented around that period since sales have had a very high increase for short periods of time. On the other hand, regarding the graph of the average sales per hour of the day, we can see that in the first hours of the day we see a growth in sales, which reaches a maximum at 6:00 and then drops a little and remains relatively constant until 8:00 p.m., when sales begin to decline. This change in sales may be related to the hours of operation of the pharmacy. In relation to sales, for each category we can clearly see that the category "cosmetics and beauty" is the one that generates the most income, with approximately 7.85 million euros, while the categories that generate the least income are veterinary and perfumery. This could be helpful for the company when it comes to knowing which categories it can focus on to maximize its sales, as well as identifying those categories and products that are not in demand and making decisions either to increase their demand or to stop selling them due to lack of demand. of demand.

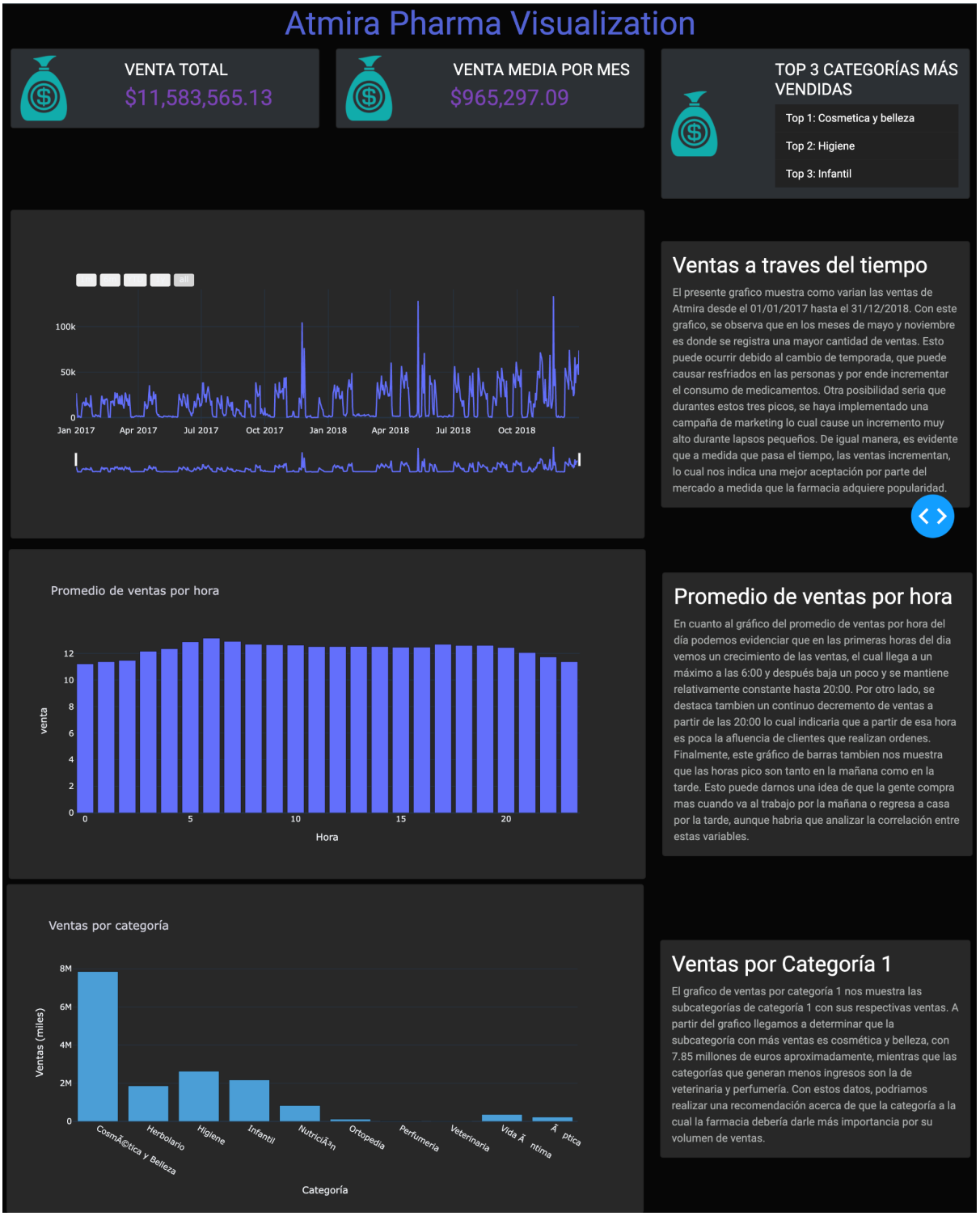
Through the heat map, we can reach more general conclusions that help us to get an idea of the geographical dispersion of customers. We can clearly see that there is a higher number of sales around the larger cities, specifically in Madrid, Bilbao, Valencia and Oviedo. We must bear in mind that they have also had sales in other countries such as Portugal and the United Kingdom, however their sales levels are very small in relation to the total sales within Spanish territory. We can clearly conclude that the company does have a very diversified market within Spain and that this can be seen as a strength given that its client portfolio is quite broad and this can help mitigate the negative effects of regional policies in Spain.

IMPROVEMENTS FOR THE FUTURE

The visualization panel developed helps the user to have a global vision of the company's sales status for a year. However, the lack of data from years before or after those provided generates a very big challenge when extracting useful information for process improvement.

As a team, we think that for the future, we could try to address this limitation by allowing the user to choose even more variables when interacting with charts. For example, allowing the user to choose the category and subcategory of the products that he wants to view and in the same way provide him with even more indices such as the elasticity of the products when they are subject to discounts. We would also like to include more external information such as weather information during the year to see the possible effects of the weather on the demand for products.

Screenshot of the first section of the dashboard:



Ventas por categoría

Heatmap de ventas

Por medio del mapa de calor, podemos llegar a conclusiones más generales y que nos ayudan a darnos una idea de la dispersión geográfica de los clientes. Podemos claramente ver que hay un mayor número de ventas en los alrededores de las ciudades más grandes en específico en Madrid, Bilbao, Valencia y Oviedo. Debemos de tener en consideración que también han tenido ventas en otros países como Portugal y el Reino Unido, sin embargo sus niveles de ventas son muy pequeños en relación al total de ventas dentro del territorio español. Podemos concluir a plena vista, que la empresa sí tiene un mercado muy diversificado dentro de España y que esto puede ser visto como una fortaleza dado que su cartera de clientes es bastante amplia y esto puede ayudar a
efectos negativos de políticas regionales en Espa