Elmeri Jokinen, Fuzel Shaik, Anssi Savallampi, Juho Syrjä, Lauri Uuniemi

Store Sales Analysis & Forecasting in Ecuador

Business Intelligence: Applications and Projects

ITEE

April 2023

Sisällys

[1 Introduction (Lauri) 3](#_Toc134209677)

[2 Business Understanding (Lauri) 4](#_Toc134209678)

[3 Data Understanding (Fuzel) 6](#_Toc134209679)

[4 Data preparation 9](#_Toc134209680)

[5 Modelling (Elmeri) 10](#_Toc134209681)

[6 Evaluation (Anssi) 11](#_Toc134209682)

[7 Deployment (Juho) 13](#_Toc134209683)

[8 Analysis 14](#_Toc134209684)

[8.1 Regional 14](#_Toc134209685)

[8.2 Forecast by category 14](#_Toc134209686)

[8.3 Promotional effect on Sales 17](#_Toc134209687)

[8.4 Effects of holidays on Sales (Juho) 24](#_Toc134209688)

[8.5 Effects of crude oil price on Sales 26](#_Toc134209689)

[9 Personal Contributions 27](#_Toc134209690)

[10 References: 28](#_Toc134209691)

# Introduction

In this project, we are helping an Ecuadorian Retail Store Chain optimize their sales through a systematic analysis of regional, seasonal, and categorical sales. This Business intelligence solution can improve the overall profitability of the stores by optimizing inventory sizes for specific product categories, predicting changes in the demand of products, and analyzing the effectiveness of product promotion campaigns.

In this Project, we will be following the Cross-industry standard process for Data mining, CRISP-DM for short. This process framework entails a six-step cyclical process that leads the process from data to information and leads to knowledge and further iteration. This report will take the basic CRISP-DM process as its general structure to make informed decisions about the data being processed and to make relevant business insights that could be used to make valuable management decisions in the Favorita retail store chain.

The CRISP-DM structure is as follows:

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation.
6. Deployment

Business understanding involves understanding the project objectives and requirements, identifying the data mining goals and defining the success criteria. Data understanding pertains to gathering and exploring the data, to better understand its quality, completeness, and relevance to the project goals. Preparation involves cleaning, integration, transformation, and selection of the data. In modeling, the appropriate technique for models is assessed and tested. After that, the model is evaluated in contrast to the project objectives, and the appropriate decisions for applications are made. As the final step, the business intelligence solution is deployed in the business setting to improve organizational performance.

# Business Understanding

Favorita is a well-known retail store chain in Ecuador, established in 1952, that offers a wide variety of products and services to its customers, including groceries, clothing, household items, electronics, and appliances. With more than 200 stores throughout the country, Favorita has become one of the largest retail chains in Ecuador and has played a significant role in the country's economy. The company's success can be attributed to its ability to provide high-quality products at affordable prices, along with its commitment to customer satisfaction. Favorita's mission is to provide its customers with a convenient and enjoyable shopping experience, while also contributing to the development of the communities where it operates.

Corporación Favorita operates in multiple different industries: industrial manufacture, Real estate, and retail. Our analysis is focused in the retail store side of the business. The retail business includes various types of stores, such as supermarkets (Supermaxi & Aki) and hypermarkets (Megamaxi & Gran Aki), as well as smaller corner stores (Aki Vecino). In our assignment, we are focusing on the overall performance of 54 distinct store locations within Ecuador.

For retail stores businesses, being able to optimize storage space is a critical part of the competitiveness of the business in the contemporary business environment. Unoptimized and large inventories lead to higher costs and capital tied-up in inventory. Especially in the case of perishable goods (food & groceries), high inventory can lead to waste. It is imperative to be able to keep inventories smaller for operational efficiency.

Good estimations of local and global demand at a given time is crucial for being able to stock specific product categories and making sure that the demand is met with appropriate supply. Therefore, forecasting methods for demand are a crucial part of modern supply chain management. (Juga, 2021; Min & Mentzer, 2004).

The types of forecasts range from short (weeks/months) to long (+3 years). For the former one the individual fluctuations of demand are more important, whereas long-term forecasts rely more on general trends. In the case of Favorita, we are estimating a few different variations of patterns, such as general trends and seasonal variations in the demand. We will make item group and sales unit level inspection of demand and roll these individual observations up the forecasting hierarchy to regional and ultimately total business level strategic suggestions to make the suggestions consistent throughout the organizational structure based on the principles of the Forecasting pyramid (Vollman, Berry, Whybark, & Jacobs, 2005).

Our data will tackle the problems of demand management specifically, meaning that we will analyze the sales data and the potential events that affect demand on a regional and national basis, such as holidays, and the price of oil, which can lead to fluctuations in the oil-dependent Ecuadorian economy.

# Data Understanding

The dataset provided for the Ecuadorian Retail Store Chain contains sales table, stores table, transaction table, and holiday tables.

The sales table contains the information regarding the number of sales, promotion details, date of the sale, product family, and store id. This information can be used to investigate the total sales for a year, which product category has the greater number of sales, and from which store.

The stores table contains the basic information of the store such as state and city, store unique id, type of the store, and the group number in which they belong.

The transactions table contains the transaction information along with store id and date of transaction. The stores table and transactions table are merged with sales table on store unique id in Tableau. Check the table below to understand the datasets.

|  |  |  |  |
| --- | --- | --- | --- |
| Table name | Column name | Description | Column type |
| Sales | Date | Date on the sale made. | Date (dd/mm/yyyy) |
| Sales | Family | Name of sold product category. | Categorical |
| Sales | Id | Unique ID for the sale. | Categorical |
| Sales | Store Nbr | Store number where the sale happens. | Categorical |
| Sales | Onpromotion | The discount or the promotional value for the sale. | Numeric |
| Sales | Sales | Cost of total sales in $ | Numeric |
| Stores | City | City where the store presents. | Categorical |
| Stores | Cluster | Group number of the store. | Categorical |
| Stores | State | State where the store belongs. | Categorical |
| Stores | Store Nbr | Unique ID for the store. | Categorical |
| Stores | Type | Type of the store (A, B, C, D, and E) | Categorical |
| Transactions | Date | Date of the transaction. | Date (dd/mm/yyyy) |
| Transactions | Store Nbr | Number of the store from where the transaction was made. | Categorical |
| Transactions | Transactions | Total number of transactions. | Numerical |
| Holiday Events | Date | Date of the holiday event. | Date (dd/mm/yyyy) |
| Holiday Events | Type | Type of holiday | Categorical |
| Holiday Events | Locale | Scope of the holiday (National, local, or regional) | Categorical |
| Holiday Events | Locale Name | Demographic name. | Categorical |
| Holiday Events | Description | Description of the holiday. | Categorical |
| Holiday Events | Transferred | Transfer status of the holiday. | Boolean |

# Data preparation

The dataset used in this project is related to an Ecuadorian retail store chain. The original data was in CSV format. The dataset included information, for example on sales, stores, transactions, and holiday events. The data contained sales from various stores, categorized by product family, promotion, and date.

Before the data could be used for analysis, several data cleaning and transformation steps were performed. Firstly, the data was imported into Tableau from the CSV files. Then, the data was checked for any missing or duplicated values. It was found that there were no duplicates or missing values in the dataset, so no rows were removed.

To understand the data and gain insights, several exploratory data analysis (EDA) techniques were utilized. The data was visualized with various types of charts and graphs. Visuals such as bar charts, line charts, and scatter plots were created, and these visualizations helped to better identify trends, and compare key information between variables like store location and product categories, in addition to being able to identify any outliers or anomalies in the data. Furthermore, basic statistical analysis techniques, such as mean, median, and standard deviation were used to understand the distribution of sales across the different categories and regions.

The data preparation stage involved cleaning and transforming the dataset to prepare it for analysis. EDA techniques were used to understand the data and gain insights, including data visualization and statistical analysis, and the necessary information from the dataset was also aggregated based on relevant variables to create a summary table. The prepared data was then ready for the modelling stage.

# Modelling

The main objective of this project is to increase overall profitability of the Ecuadorian retail store chain Favorita by identifying the key factors that drive sales, optimizing inventory sizes for specific product categories, predicting changes in the demand of products, and analyzing the effectiveness of product promotion campaigns.

Initially the project plan was to utilize different machine learning models, but this was quickly realized to be too ambitious. In the end, the modeling part of this project involved utilizing Tableau’s time series forecasting feature, which employs exponential smoothing in its algorithm, to predict changes in sales for different product categories and for example simple linear regression on analyzing the effectiveness of promotion campaigns. The goal was to optimize inventory sizes and understand the factors that drive sales in order to improve the overall performance of the retail store chain. To achieve this, various algorithms and time series forecasting models were initially planned to be applied to the data to identify patterns and make predictions, but this level of granularity was deemed not achievable with the data and tools of this project.

The utilized models helped to identify the most important factors that influence sales, such as location, time of year, and promotions. The results of the modeling stage can be used to optimize inventory sizes and make more informed decisions about promotions. By predicting changes in demand and identifying the factors that drive sales, the Ecuadorian Retail Store Chain is able to make more accurate forecasts and avoid under or overstocking inventory. Additionally, by analyzing the effectiveness of promotions, the chain is able to allocate resources more effectively and improve overall sales performance.

# Evaluation

After the modeling phase, the next step in the CRISP-DM process is evaluation. This is where we assess the performance of our model and determine whether it meets the success criteria defined in the business understanding phase.

In the case of the Favorita retail store chain optimization project, evaluation stage will play a critical role in assessing the effectiveness of the business intelligence solution. During this stage, we will test and evaluate the models that were created during the modeling stage. We want to investigate if the models can optimize the size of the store's inventory for specific product categories, predict changes in product demand, and evaluate the efficiency of product promotion efforts. We will assess the models based on their accuracy, validity, and usefulness in giving insights that can be used to improve store profitability. In other words, we aim to ensure that the models are accurate, dependable, and provide helpful information to improve the business.

Assuming that our model performs well in evaluation, we believe that it can help the Favorita retail store chain meet its business success criteria. Our models can help the chain optimize its inventory sizes for specific product categories, predict changes in demand for products, and analyze the effectiveness of product promotion campaigns.

And by using our model to optimize inventory sizes, the chain can reduce costs associated with carrying large inventories while ensuring that it has enough stock to meet customer demand. This can help the chain increase its profitability and remain competitive in the market.

Our model's analysis of product promotion campaigns can help the chain determine which campaigns are most effective and which ones need to be improved or abandoned. This can help the chain optimize its marketing strategies and increase its sales and profits.

**Evaluate results**: To determine if the models meet the business success criteria, we need to compare the model results to the predefined business success criteria. If the model results meet or exceed the predefined success criteria, then the models can be considered successful.

Once we have determined that the models meet the business success criteria, we need to decide which model(s) should be approved for the business. This decision will depend on the specific needs and goals of the business, as well as the performance of the different models.

**Review process:** In the Business Understanding phase, we identified the project objectives and requirements and defined the success criteria. We also gathered an understanding of the Ecuadorian retail store chain, Favorita, and its operations. The Data Understanding phase involved exploring and gathering data to better understand its quality, completeness, and relevance to the project goals. In the Data Preparation phase, we cleaned, integrated, transformed, and selected the data. The Modeling phase involved assessing and testing the appropriate modeling technique, and in the Evaluation phase, we evaluated the model's performance against the project objectives. Finally, in the Deployment phase, the business intelligence solution is deployed in the business setting to improve organizational performance.

**Determine next steps**: Based on the results of the previous tasks, determining the next steps depends on whether the models meet the business success criteria and which one(s) should be approved for deployment. If the models meet the success criteria, the next step would be to proceed with deployment. However, if they do not meet the criteria, further iteration is required to improve the models.

# Deployment

The deployment phase of CRISP-DM covers four main steps, which are Plan Deployment, Plan Monitoring and Maintenance, Produce Final Report, and Review Project. As the project group lacks access to the technical details of Favorita information systems, the deployment plan for the business intelligence solution will not be described with integration to existing systems in mind, but instead to support decision-making activities at the relevant levels in the organization.

The deployment plan for the business intelligence solution developed is to deploy it for high-level overview of sales and promotional effects for decision makers who can best utilize it. The ability of the dashboard to aid in logistical planning is somewhat limited by the dominance of the capital region Picincha, which dominates in sales and most active stores by volume in the capital city of Quito. However, as perishable category goods such as groceries are sold across the country in large volumes, it should be possible to use the information provided to optimize inventory sizes and deliveries to some extent.

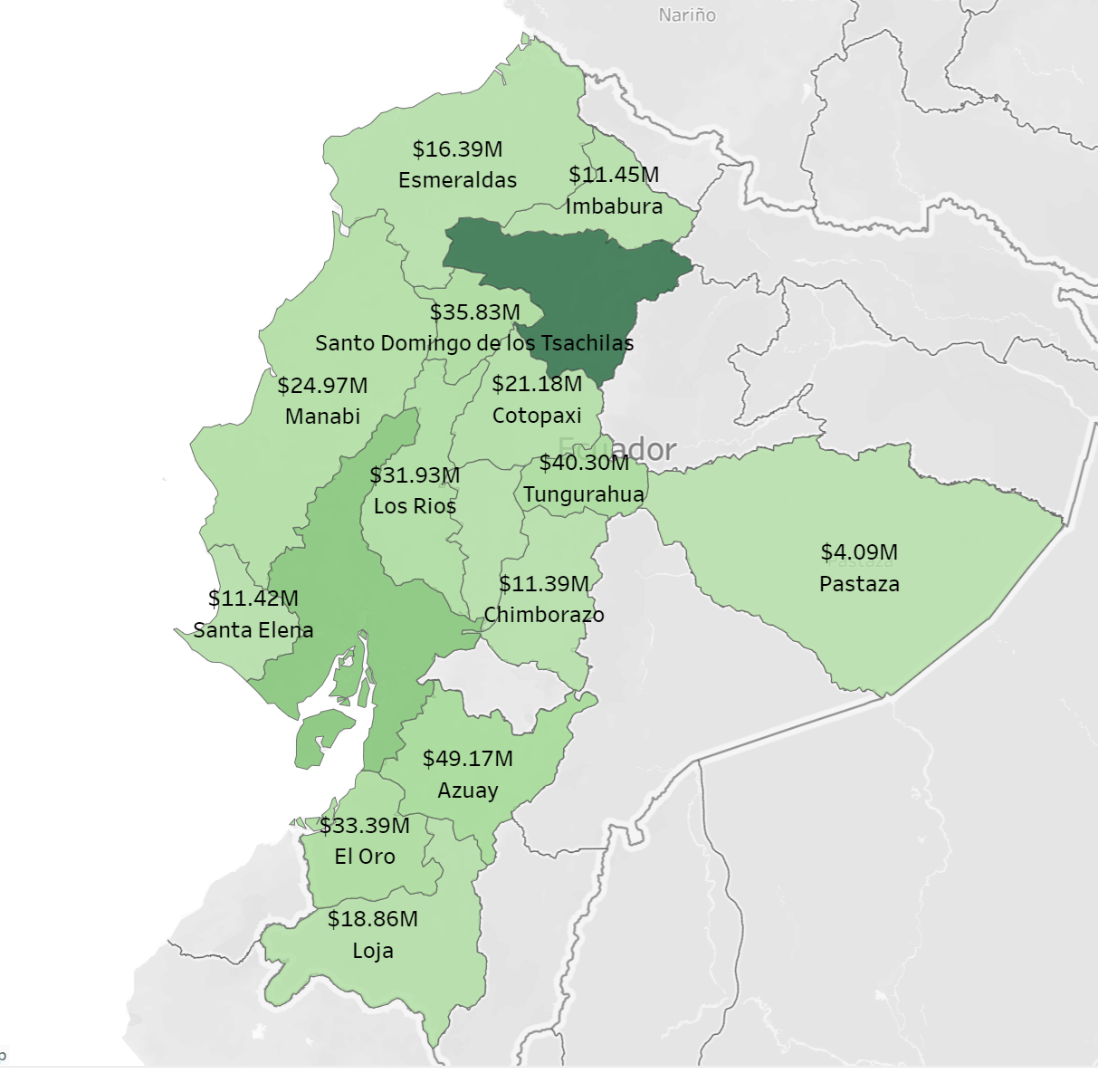
For monitoring and maintenance, observations will be covered in more detail in the following chapter (8. Analysis) are relevant. Particular attention should be given to the accuracy of the sales forecast as well as continuing to monitor the performance of promotional campaigns regarding product categories. For future development of more detailed analysis, we recommend developing machine learning models to evaluate the effect of local holidays on sales as this level of granularity is not achievable with the data or tools chosen for this project. As the data includes detailed sales figures for all stores, it is reasonable to assume that the company is capable of data warehousing and real-time data could be used with the dashboard. This would enable more detailed monitoring of the performance of the solution and infrastructure should be adapted to it.

The final report for this project is this report in combination with the presentation to be delivered on 5.5.2023. The report describes the methodology used in analyzing the business problems as well as data, and the decisions on which results of analysis will be included in the dashboard and the reasons for that. These deliverables will be presented to stakeholders in a meeting during week 19. The review process will be conducted after receiving feedback from stakeholders at the final project meeting.

# Analysis

## Regional

Regional analysis is carried out to find the regional sales distribution and identify potential markets in the country. The Geo map feature in tableau is used to perform this analysis. Where we use sum of the sales to create a heat map on the geo map to easily infer the information regarding sales across various states in Ecuador. The dataset provided has a total of 16 sates of Ecuador and their respective sum of sales from 2014 to 2017.

  
*Image 1: Geo map of the Ecuador region with total number of sales in each state.*

## Forecast by category

In this section, we present the results of our analysis on forecasting the sales based on product category. We utilized time series forecasting models to predict the sales for each category, with a focus on identifying the categories for which forecasting would be most useful.

Our analysis revealed that forecasting was particularly useful for the most sold product categories, with obviously bigger quantities sold, as well as bigger changes over time in the sales amounts. These categories included groceries, beverages, produce, cleaning, and dairy. These categories, especially groceries, beverages and produce, exhibited significant fluctuations in sales over time, and the forecasts can provide valuable insights for optimizing inventory sizes and identifying the factors that drive sales.

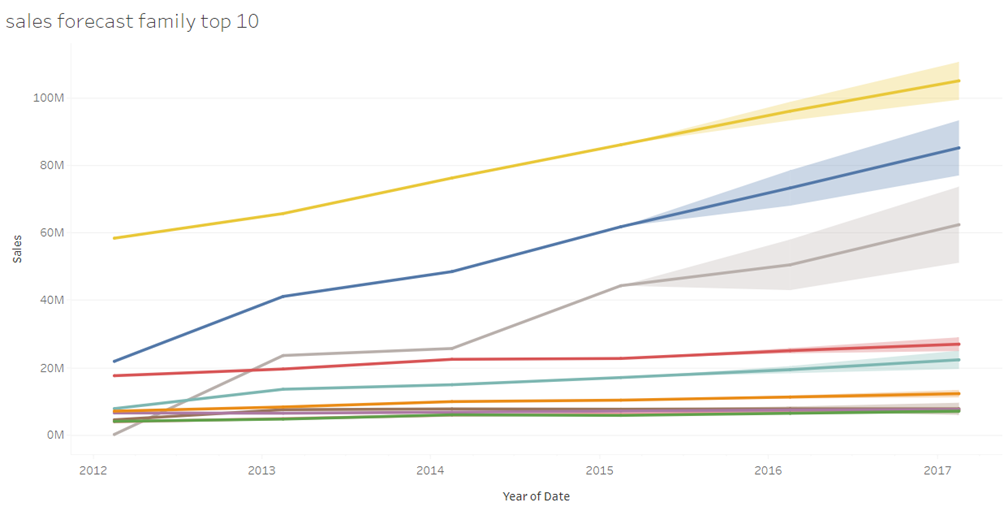
For the other categories, however, the sales were relatively consistent in previous years, and the forecasts therefore did not provide notable additional value. In these cases, historical sales data should be sufficient when making informed decisions about, for example inventory and sales strategy.

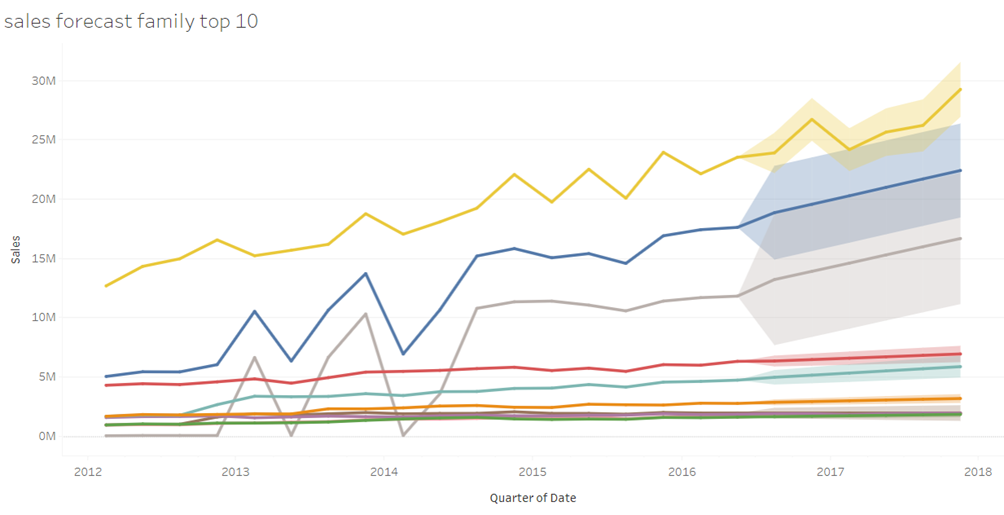
Furthermore, it was found that the choice of date range had a quite significant impact on the range of the forecasts. In the pictures below, differences in the different date range views can be seen, varying from yearly to weekly.

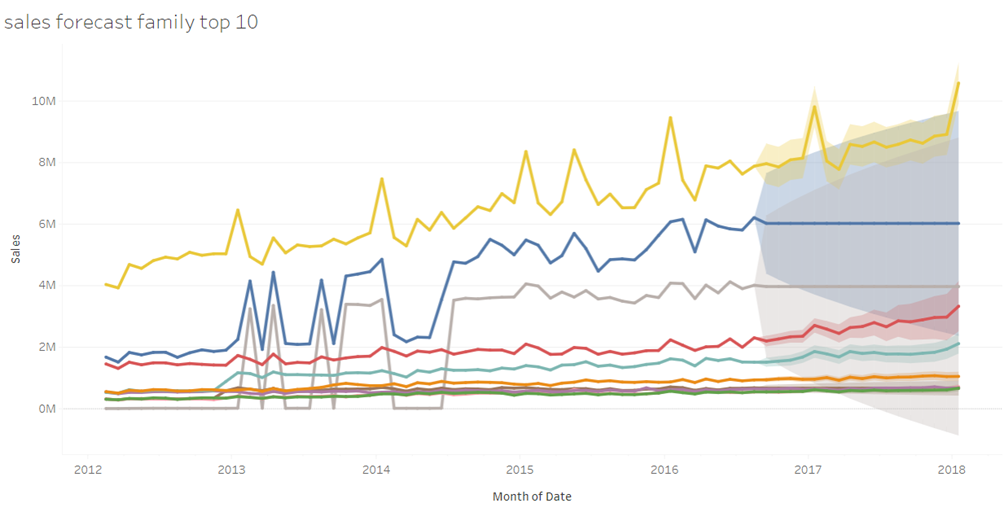
Depending on the date range used, the forecasts varied greatly, and it was necessary to choose a suitable range to obtain reliable forecasts. The monthly date range, while showing changes over time in more detail, did not generate a viable forecast especially for beverages and produce categories. The weekly date range on the other hand had even greater flaws, even forecasting large negative sales, and was thus omitted from the analysis. Had the monthly date range more accurately forecast beverages and produce categories, it would probably have been utilized.

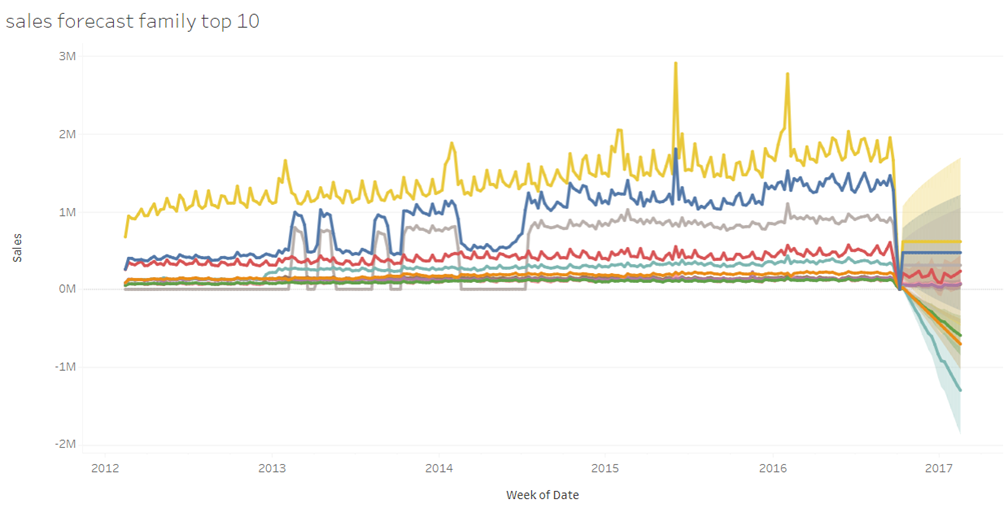
After considering the various options, we opted to use the quarterly date range for a somewhat reasonable forecast for the top categories. This also allowed for the capture of seasonal fluctuations in sales for the product categories, while making more accurate forecasts than in the yearly view.

Overall, our analysis of forecasting by category revealed that while forecasting was particularly useful for certain product categories, the choice of date range and the level of detail were critical factors in obtaining accurate and reliable forecasts. By utilizing time series forecasting models and carefully selecting the appropriate data range, we were able to generate valuable insights for future inventory optimization and overall performance improvements for the retail store chain.

*Image 2: Top 10 product families’ sales forecast on a yearly date range view.*

*Image 3: Top 10 product families’ sales forecast on a quarterly date range view.*

*Image 4: Top 10 product families’ sales forecast on a monthly date range view.*

*Image 5: Top 10 product families’ sales forecast on a weekly date range view.*

## Promotional effect on Sales

One part of our analysis dug deep into the effects of promotions on sales of the Favorita chain. We analyzed the regional effect of promotions over time, as well as the Pearson correlation of promotions and sales across product families and across time.

Regional promotion analysis

For the regional promotion analysis, we deemed a scatterplot with number of promotions on the horizontal axis and the number of sales on the vertical axis could give us a good insight into the progression of promotional efforts within the retail chain. For a more thorough analysis, we included a linear regression line for every year of promotions done so far (See Images 6 & 7).

Kuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisesti*Image 6: Effects of promotion on sales by store.*

Kuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisesti*Image 7: The Effects of promotion on sales by the Top 10 largest stores by sales*

As the retail chain started promotions of product families in 2014, the scatterplot for the year 2013 shows a straight line of observations and a flat zero slope trend line. With the year of 2014, the trend line seems to be very steep, and some stores have significantly increased their sales in contrast to previous year. This does seem to advocate that the promotions have a positive impact on sales (Image 6).

However, it does seem that most of the observations (stores) still lie within the original distribution of sales by year (1,5-6M USD) even though the number of promotions has increased drastically (From 0 to 40k-80k per year) (Image 6). This the effectiveness of promotions within most of the stores.

When we dig deeper by filtering the data by the top 10 stores by sales, we find out that these stores have significantly increased their sales over the timeframe and thus, their trend lines stay highly sloped over the period. Moreover, 8 out of these 10 stores are from the capital city of Ecuador, Quito. This seems to suggest that the more highly populated areas derive considerably more benefit from the promotion efforts in contrast to stores from more rural areas.

The linear regression line of 2017 for **all** stores seems to suggest that increasing the number of promotions by one would lead to an increase of $130 in sales (Image 6). In contrast, with the top 10 stores by sales, this same analysis would posit that the increase of one promotion would lead to an increase of $490 in sales (Image 7). This analysis should be taken with a grain of salt since the overall trend has been changing drastically over the period of analysis. However, this does suggest that the larger stores in the more populated areas benefit from the promotions and seem to have performed the promotions better over the period.

This analysis entails that the promotional efforts should focus on the high-population areas, and the promotional efforts could be toned down in the more rural areas. Although the promotional success could be a result of the network effects of the population, or a part of the marketing competence in these larger stores. The marketing teams should assess what the more successful stores are doing differently to make such success. If the promotional efforts are identical to those of other stores, then we can say for certain that the success is borne of the economies of scale in high-population (and high-sale) areas.

Our suggestion is that the marketing teams of the retail chain should further assess the content differences of the promotional efforts. Moreover, our analytics team could analyze the cost-efficiency of the promotions by contrasting the value derived with the promotion costs. To do this, our team would need data on the promotion budgets of different stores over time, but for this analysis, we don’t have the required data to do so.

Categorical Correlation Analysis

For the categorical analysis of promotional efficiency, we deployed two heatmaps on the Pearson correlation of sales and promotions. From these, the first one on the product family groups was deployed on the dashboard as a treemap (Image 8).

Kuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisesti*Image 8: The Pearson correlation treemap of sales and promotions by year and Product Family group*

The use of the Pearson correlation coefficient (coined r) is valid in the case of linear correlations. It is also useful in the sense that its results are easy to understand. r=1 means a perfect positive linear correlation, r=-1 means a perfect negative linear correlation, and r=0 means there is no correlation between the variables (Päkkilä, 2016). However, the difference with borderline values such as 0,5 or 0,65 is always subjective due to the differences between studied phenomena and industries (Cohen, 2013). In this case, we will use a rough categorization (Owens, 2022):

* r>=0,5 is considered strong association
* 0,5>r>=0,3 is considered moderate association
* 0,3>r>=0,1 is considered weak association
* 0,1>r>=0 is considered no association

The same rules apply in the negative side of the scale.

In the first heatmap (Image 9), all the product families are put into four distinct groups to track the overall efficiency of the promotional efforts of the retail chain. In the second heatmap (Image 10), we looked at the bottom 5 separate product families to identify the least effective promotional efforts.

Kuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisesti*Image 9: The Pearson Correlation of Sales and Promotion by Product Family Groups*

Kuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisesti

*Image 10: The Pearson Correlation of Sales and Promotion by Bottom 5 Product Families*

Overall, it seems that the promotional efforts are improving year by year, with most product family groups having a weak or moderate positive correlation in 2014 and ending in moderate to strong association in every product family group in 2017. This suggests that the promotions seem to be affecting the sales or at least growing along with the company. Our hypothesis is that we are observing some sort of organizational learning effect, where the promotional efforts are getting more consistent and effective on sales overall (image 9 & 11).

The correlation of promotions and sales in the dairy and meats product family group is quite close to zero in the beginning of the study period, meaning there is little to no correlation between these product families and promotions. This is the least correlating product family group in our dataset, which is understandable when looking at the stagnant sales progression of dairy and meat products over the study period. As a contrast, sales in home and specialty goods (e.g., home appliances, cleaning, electronics, etc.) have a high correlation with promotions every year (with a high of 0,73 in 2017).

The second heatmap (Image 10) reveals individual product families (not product family groups), where promotions have no association with the sales of these product families. All these product categories have stagnant or dropping sales over the period, with a slight increase in promotions. For these product families, we would consider revamping the promotions or downsizing them.

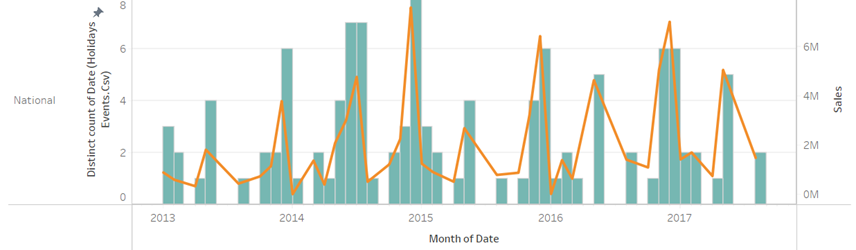
Kuva, joka sisältää kohteen kaavio

Kuvaus luotu automaattisesti

*Image 11: The number of sales per promotion (orange) versus the correlation of sales and promotions (blue). As the number of sales per promotion decreases, the correlation of sales and promotions increases.*

## Effects of holidays on Sales (Juho)

Ecuador has a number of different holidays, ranging from local to regional and national holidays spread throughout the year. When analyzing the effect of national holidays, two main holidays emerge as clear drivers of sales, namely Christmas and the Carnival, which is celebrated in the spring. Both holidays are followed by a corresponding drop in sales with post-Christmas being significantly larger.



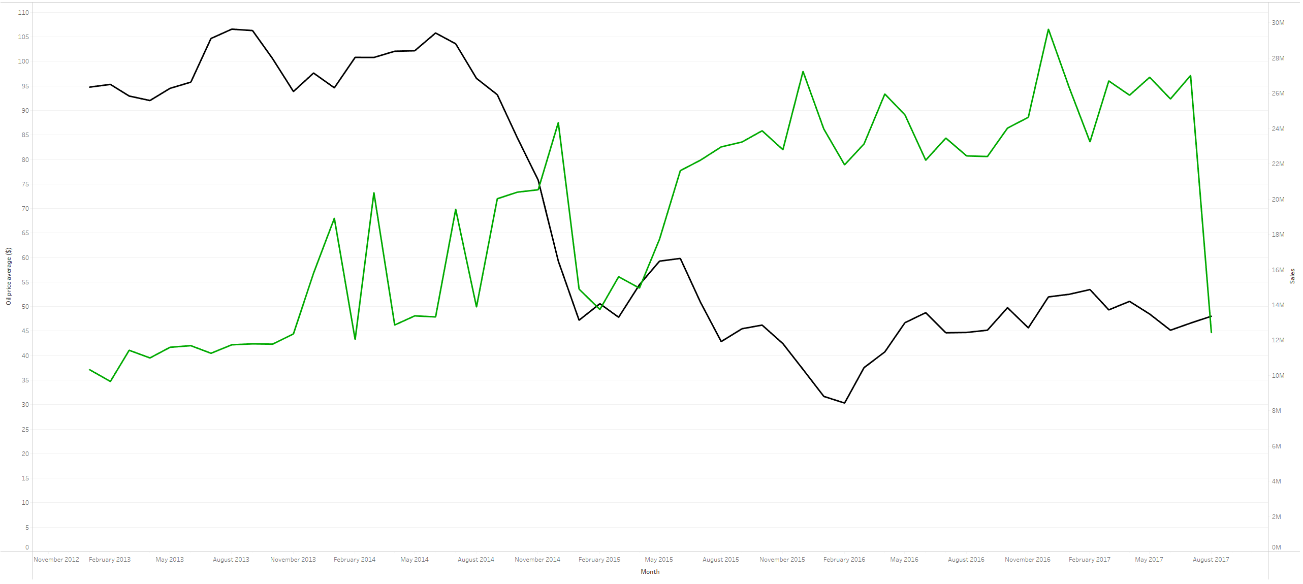
*Image 12: Christmas and spring Carnival sales peaks*

The holiday data includes two irregular events, the 2014 FIFA World Cup and the 2016 Ecuadorian earthquake which both were associated with increased sales (in case of the earthquake, with sales spiking immediately after the event) but which were filtered out of the data as extraordinary events for the analysis.

For local holidays, minor increases in sales are associated with local holiday events, but as expected the effect is much smaller when compared to the large national holidays. One possible use for local data would be to develop a more sophisticated machine learning model that could predict the local effects, but developing this through manual data analysis was determined to be too laborious for this project.

When looking at the overall sales, December was consistently the highest sales month for all the years in the dataset with sales increasing on average about 20 % compared to the preceding November. Observing the growth of sales for each holiday over the four-year period in the dataset it was noted that they mostly grew in proportion to the overall growth of sales of Favorita, with the interesting observation that two sales-oriented holidays (Black Friday and Cyber Monday) had comparatively low impacts and less overall growth of sales compared to other holidays.

## Effects of crude oil price on Sales

*Image 13: The average price of oil per month (black) versus the total sales of the month (green).*

This graph shows the average oil price per month (black) compared to Favorita's monthly sales (green). Ecuador is very oil-dependent economy and as you can see from the graph, when the price of oil decreases, the amount of sales seems to increase. This could be explained that the cost of transporting goods to supermarkets is affected by the price of oil since transportation companies require fuel to power their vehicles. Therefore, if the price of oil decreases, the cost of transporting goods to supermarkets may also decrease, which could affect the prices of products in the store and vice versa. Also, oil price decrease would affect raw materials used in the manufacturing of products sold in supermarkets, which could ultimately result in lower prices for the end products.

This would lead to consumers having more disposable income, as they are paying less for gasoline and other oil-based products and lead to an increase in overall consumer spending, which includes supermarkets. Lower prices can attract more customers, which can ultimately lead to higher sales. Lower oil prices can also benefit the economy as a whole, which can lead to increased consumer confidence and spending.

# Personal Contributions

Everyone’s contribution to this project is signified here:

Elmeri: Modeling, Data Preparation, Forecast by Category (text, analysis, and graphs).

Fuzel: Data understanding, regional analysis, Putting the dashboard together after everyone’s analyses, Dashboard demonstration.

Anssi: Evaluation, Effects of crude oil price on sales (text, analysis, and graphs)

Juho: Deployment, Effects of holidays on sales (text, analysis, and graphs)

Lauri: Introduction, Business understanding, Promotional effect on sales (text, analysis, and graphs excl. Image 8), PowerPoint slides and presentation(s).

# References:

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Routledge.

Juga, J. (2021): Jakelukanavat ja Kaupan ketjuliiketoiminta -Course material. Oulu Business School

Min, S., & Ph.D, J. (2004). Developing and Measuring Supply Chain Management Concepts. *Journal of Business Logistics*, *25*, 63–99. <https://doi.org/10.1002/j.2158-1592.2004.tb00170.x>

Owens, A (2022): How to use correlation analysis in R to improve marketing performance. Rpubs by Rstudio Internet-page. Referred 2.5.2023. Available at: https://rpubs.com/anitaowens/correlation-in-marketing

Päkkilä, J (2016): Tilastotieteen jatkokurssi (806119P) Luentomoniste. Oulu University course material. p25-27.

Vollman, T.E., Berry, L.W., Whybark, D.C.& Jacobs, F.R. (2005). Manufacturing Planning and Control for Supply Chain Management. McGraw-Hill, Boston.