*Proceedings of the 10th World Congress on Mechanical, Chemical, and Material Engineering (MCM'24)*
*Barcelona, Spain - August 22-24, 2024*
*Paper No. ICMIE 110*
*DOI: 10.11159/icmie24.110*

# Demand Forecasting Model To Reduce The Mean Absolute Percentage Error By Applying Seasonal Breakdown Tools In A Sme In The Tourism Sector

**Sayuri Arleth Renatta Ludeña-Roman[1], Sebastian Zelada-Collazos[2], Jorge Antonio Corzo-Chavez[3]**
Facultad de Ingeniería, Universidad de Lima, Perú
Av. Javier Prado Este 4600, Santiago de Surco, Lima, Perú
[1]20191164@aloe.ulima.edu.pe; [2]20192212@aloe.ulima.edu.pe; [3]jacorzo@ulima.edu.pe

***Abstract -*** The research work is based on the analysis of demand in a tourism company using mathematical models. The methodology design presents a correlational and descriptive scope where the company's sales are collected to calculate the mean absolute percentage error in demand. With the help of machine learning tools, a predictive analysis will be carried out to estimate the sales for the following year, seeking to reduce the error using one of the selected mathematical models, calculate the necessary sales force, and thereby reduce the economic impact equivalent to $16 789,02. The MAPE (Mean Absolute Percentage Error) in the tourism sector is 12,03%. Through calculations using Python and RISK, a value of 15,36% was obtained, reducing the MAPE by 4,24% compared to the year 2022. The Systematic Review of the Literature allows us to showcase the tools that can be developed in similar or atypical scenarios. The choice will depend on the behaviour pattern or trend.

***Keywords:*** Demand estimation, moving average, seasonal breakdown, smoothing exponential, simple average, mean absolute percentage error, Google Collaboratory, predictive analytics.

## 1. Introduction

The company to be evaluated is a SME (Small and Medium Enterprise) in the tourism sector, focused on providing high-value vacation programs or tourist packages, offering a dynamic, personalized, and innovative quality service for its client portfolio. Likewise, it began operations in 2016 and to date has firmly maintained its consolidation in the travel and tourism sector; reaching important commercial agreements with companies such as Assist Card, PlusUltra, CopaAirlines, Latam, Avianca, among others. Below is a summary of the historical data of annual sales from 2019 to 2022 as shown in table 1.

Table 1: Annual turnover.

| YEAR | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|
| BILLING ($) | 915 593,95 | 294 545,72 | 618 829,42 | 1 403 500,45 |

The present work of the Industrial Engineering degree addresses "Operation Research & Analysis" as a line of research for the application of mathematical modeling that allows the improvement of the sales forecasting process as support for the company's decision making.

The main objectives are to identify and determine the mathematical modeling tools to use, perform the sales forecast with the selected tools, design possible scenarios with medium-term forecasting methods, determine the size of demand, the economic impact on the company and decrease the mean absolute percentage error between sales forecasts and actual sales using the highest precision tool.

According to the research topic, the industrial engineering tools used to reduce the mean absolute percentage error between the sales forecast and actual sales are the following: Simple average, moving average and exponential smoothing. Additionally, to define the size of the sales force, the seasonal breakdown was used. With this tool, we will determine the number of average salespeople, assuming a constant level of productivity in the sales units.

## 2. State of the art

Regarding the technical gap, the mean absolute percentage error of companies in the tourism sector was identified as equivalent to 12,03% [1]. For its part, the company exceeds this value with 19,60% in 2022. Thus, creating a margin of opportunity to improve of 7,57%. Likewise, the quantitative analysis allowed us to calculate the economic impact caused by the absence of methods, lack of structure of the sales effort and the imbalance of data, resulting in a loss of $16 789,02.

With the help of the systematic review of the literature, it was possible to identify, evaluate and describe research articles classified according to the criteria according to their typology, as well as descriptive articles, case studies and experimental cases. Which contributed to the analysis of the tools to be used.

In accordance with what was written in the previous point and based on the objectives of the research, a comparative matrix will be established to define the proposal of each one. Two scientific articles have been selected for each tool to be evaluated. Regarding the objectives, those that respond to the main reasons for the problems are prioritized. Table 2 shows a comparative matrix.

Table 2: Comparative matrix and systematic literature review.

| **Root Causes** | **Engineering Tools** | **Scientific Article** |
|---|---|---|
| Unstructed database | Moving average | Carolina del Pilar (2017) [2] |
| Inaccurancy of sales records | Simple average | Ackerman, A. & Sellito, M. (2022) [3] |
| Fluctuations not considered | Smoothing exponential | Marroquin, R (2018) [4] |
| Inadequate sales force calculation | Seasonal breakdown | Cisneros-Martinez, JD & Fernandez-Morales, A (2016) [5] |
| Absence of sales indicators | Seasonal breakdown | |

## 3. Methodology

The methodology is oriented to the use of machine learning using demand forecast models using the historical data of a company in the tourism sector to calculate the mean absolute percentage error of the company. In addition, the aim is to project future sales and thereby calculate the number of salespeople needed until the end of the year. Therefore, clean and filtered data is required through exploratory data analysis to use the information in the models.

The tools used are seasonal breakdown for sales force calculation while smoothing exponential, simple and moving average for demand forecasting. On the other hand, Python codes will be used to estimate the company's sales in 2024 through Google Collaboratory. Likewise, exploratory data analysis will be necessary to obtain atypical data thanks to the Orange Data Mining software, which will also allow sales for the month of September 2023 to be calculated.

Likewise, data visualization must be used using tools such as Power Bi to analyse the results that will be delivered to the company. Below is figure 1 with the summary of the methodology.

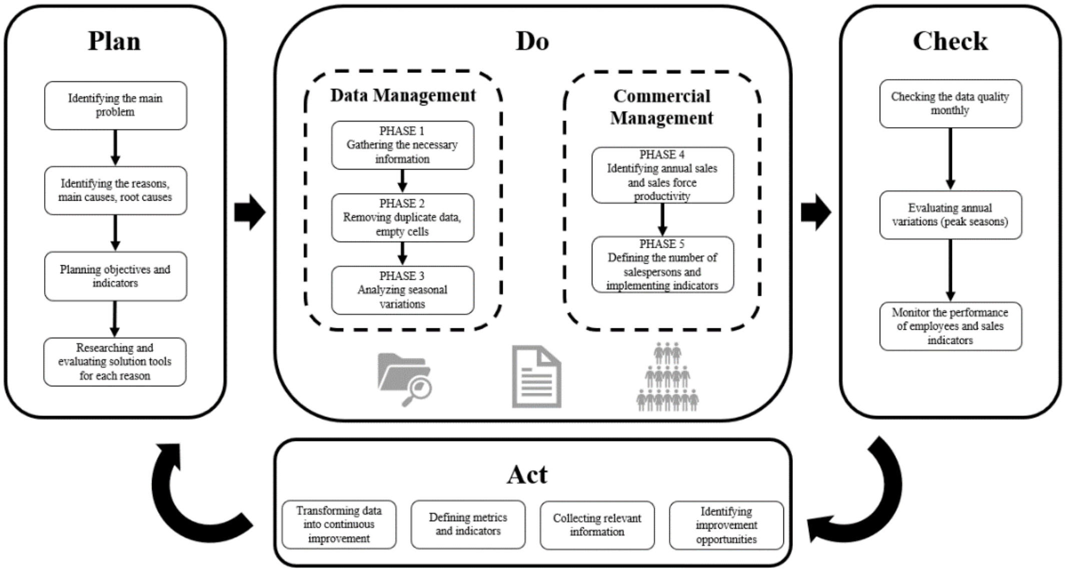

Fig. 1: Methodology.

### 3.1. Model components

Component 1 (Plan) To carry out the research, you must first know the company's problems and identify possible solutions. In this way, you can search for the models and tools necessary to make the improvement.

Component 2 (Do) Data management is important to ensure quality and analysis of seasonal variations in sales. The company database must be obtained that includes sales made per year with a monthly breakdown. The company began operations in 2019, therefore, the data was obtained from that year until the month of August 2023. The data was considered structured, since it was presented in Excel through a dynamic table with attributes such as date, seller, amount, and type of sale. Like all data, a cleaning process must be carried out to verify any problems that could lead to errors in the models such as atypical sales, duplicate data or the absence of data. In addition, the data must be consolidated with the necessary attributes, in the case of Orange and Google Collaboratory, it must include the date and annual amounts with a monthly breakdown while for the report in Power Bi, the name of the seller is also considered.

Component (Check + Act) The calculation of the forecast and the average absolute percentage error of all the years that have closed data (2019 to 2022) is carried out, each of the three selected models will be applied, which are simple, moving average and exponential smoothing. In this way, it is desired to compare and analyse the model that best fits the data, including the required seasonality. On the other hand, the data must be uploaded to Google Collaboratory so that the sales estimate for the remainder of the year 2023 and 2024 can be made. Regarding the information for the month of September, as it is a month in which the evaluation, and therefore, the data for the entire month is not available, the Orange Data Mining software must be used to predict the average value of the month. On the other hand, the data must be imported into Power Bi from the years with complete data without forecasting so that the company can observe the evolution of its sales, the scope of the sellers and the error they present to estimate their demand, which leads to the unnecessary hiring of sales force or lack thereof. In this component, the review of the previous ones is required, where, through the technical gap, the aim is to reduce the average absolute percentage error of the company. Currently, it has 19,60% while the sector average is 12,03%. For the improvement proposal, we will seek to achieve a percentage between 12 to 15 for the year 2024, so an analysis will also be carried out at the YTD (year-to-day) until the full month obtained from the Google Collaboratory estimate for the year 2024. In this way in this way, the company will look for strategies to achieve the forecasted sales. For the sales force, with what was obtained and the calculated productivity, the number of people that must be included in the sales force to achieve the objectives will be analysed.

## 4. Implementation and Results

### 4.1. EDA technique (exploratory data analysis)

All analysis is done with Orange Data Mining software to determine atypical sales, data duplication and thereby ensure the quality of our data set. For the exploratory data analysis, the data is first displayed in the "data table" module to verify that there are no missing data. In this case, the month of September is unknown data, so with the impute module and with using the average/most frequent technique, the data for the month of September is obtained for the year 2023. The technique used in the program will be shown in figure 2.

| | Date | Closed_Sales ($) |
|---|---|---|
| 1 | 2019-09-30 ... | 78893.6 |
| 2 | 2020-09-30 ... | 1039.48 |
| 3 | 2021-09-30 ... | 80696.9 |
| 4 | 2022-09-30 ... | 147252 |
| 5 | 2023-09-30 ... | ? |

| | Date | Closed_Sales ($) |
|---|---|---|
| 1 | 2019-09-30... | 78893.6 |
| 2 | 2020-09-30... | 1039.48 |
| 3 | 2021-09-30... | 80696.9 |
| 4 | 2022-09-30... | 147252 |
| 5 | 2023-09-30... | 76970.5 |

Fig. 2: EDA technique data table.

Subsequently, the table is displayed again with the data already verified in the "data table (1)". Finally, the "outliers" module is used to verify whether there are outliers, therefore, the "local outlier factor" method is used with a contamination parameter of 3% for the minimum amount of data. The modules used in the program are shown in figure 3.

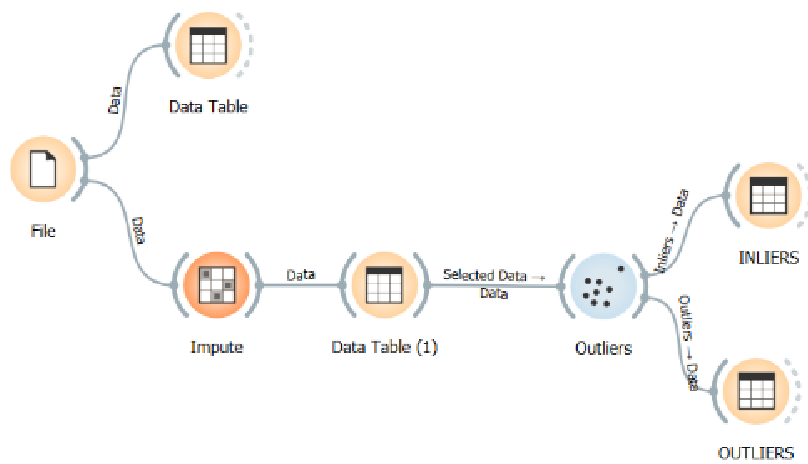


Fig. 3: EDA technique outliers.

### 4.2. Sales projections

Google Collaboratory is used to project sales from October 2023 to August 2024, using Python codes, within which the 3-month seasonality due to summer vacations is considered. In figure 4, the sales trend and the projection for the year 2024 are shown.

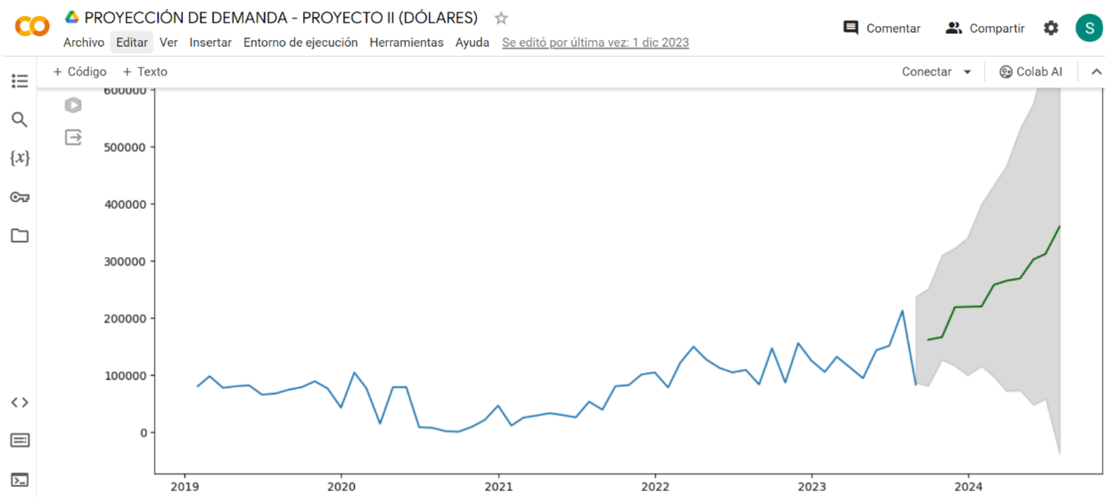


Fig. 4: Sales projections.

In addition, the numerical values are presented in the following table.

Table 3: Sales projections.

| Date | Forecast_Sales ($) |
|---|---|
| 31/10/2023 | 137 409,83 |
| 30/11/2023 | 172 038,80 |
| 29/12/2023 | 161 222,12 |
| 31/01/2024 | 151 410,08 |
| 29/02/2024 | 173 791,51 |
| 29/03/2024 | 166 726,55 |
| 30/04/2024 | 159 679,85 |
| 31/05/2024 | 176 327,17 |
| 28/06/2024 | 170 047,00 |
| 31/07/2024 | 201 097,24 |
| 30/08/2024 | 138 039,88 |

### 4.3. Comparison of mathematical models

With the complete and projected annual data, the three exposed models are developed, exponential smoothing, simple and moving average in such a way that the model that best suits the needs of the company is chosen, considering seasonality and the lowest average absolute percentage error. The following table shows the summary of the results obtained from each model.

Table 4: Comparison of mathematical models.

| Model (2023) | % |
|---|---|
| Smoothing Exponential | 31,0% |
| Simple Average | 23,4% |
| Moving Average | 15,19% |

As demonstrated, the moving average model will be used with a seasonality of 3 months, therefore, the calculation of the model with its respective sales and the forecast is presented.

Table 5: Comparison of mathematical models.

| Year 2023 | Closed Sales ($) | Forecast ($) | Error (Closed Sales - Forecast) | Absolute error |
|---|---|---|---|---|
| January | 105 689,51 | | | |
| February | 132 286,96 | | | |
| March | 113 301,53 | | | |
| April | 94 627,40 | 113 405,29 | -18 777,89 | 18 777,89 |
| May | 143 742,95 | 117 223,96 | 26 518,99 | 26 518,99 |
| June | 151 485,79 | 129 952,05 | 21 533,74 | 21 533,74 |
| July | 213 187,59 | 169 472,11 | 43 715,48 | 43 715,48 |
| August | 83 556,11 | 149 409,83 | -65 853,72 | 65 853,72 |
| September | 76 970,50 | 124 571,40 | -47 600,90 | 47 600,90 |
| October | 137 409,83 | 99 312,15 | 38 097,68 | 38 097,68 |
| November | 172 038,80 | 128 806,38 | 43 232,42 | 43 232,42 |
| December | 161 222,12 | 156 890,25 | 4 331,87 | 4 331,87 |
| | | **156 890,25** | | **23 823,80** |

On the other hand, the analysis of the complete years from 2019 to 2023 is presented in the following table, considering the year 2020 as an atypical year due to COVID-19.

Table 6: Analysis of the complete years from 2019 to 2023.

| | Forecast ($) | Mean Absolute error | % |
|---|---|---|---|
| Closed Sales 2019 | 69 676,73 | 11 293,96 | 16,21% |
| Closed Sales 2020 | 25 987,42 | 19 267,51 | 74,14% |
| Closed Sales 2021 | 96 178,05 | 17 659,53 | 18,36% |
| Closed Sales 2022 | 122 828,18 | 24 069,87 | 19,60% |

| | | | |
|---|---|---|---|
| Closed Sales 2023 | 156 890,25 | 23 823,80 | 15,19% |

Finally, a calculation is carried out at YTD (January-August) of all years using the same model, focusing on the year 2024 for the projection and determining the sales of the last months to equal the mean absolute percentage error to the year 2023 (year with the lowest MAPE since 2019). The results are shown in table 7.

Table 7: YTD calculated.

| | Forecast ($) | Mean Absolute Error |
|---|---|---|
| Closed Sales 2023 | 156 890,25 | 23 823,80 |
| Closed Sales 2024 | X | 19 222,93 |
| | X = 126 591,50 | |

The calculation of X represents the forecast that must be obtained for the months from September to December.

Table 8: Forecast from September to December.

| SET-DEC | Forecast ($) | Mean Asolute error | % |
|---|---|---|---|
| Closed Sales 2023 | 156 890,25 | 23 823,80 | 15,19% |
| Closed Sales 2024 | 126 591,50 | 19 222,93 | 15,19% |

### 4.4. Sales force calculation

Finally, applying the sales force formula, through the seasonal breakdown model, we have the number of workers necessary for the year 2023.

$$N = \frac{\text{Sales volume for 2023}}{\text{Productivity for each sales force 2023}} \tag{1}$$

$$N = \frac{1\ 585\ 519,07}{124\ 824,48} \tag{2}$$

$$N = 13 \text{ employees} \tag{3}$$

### 4.5. Data visualization

Power Bi software is used to view the general report that includes sales by year and month, the necessary sales force and the results and indicators that have been previously evaluated. The following figure shows the main page.

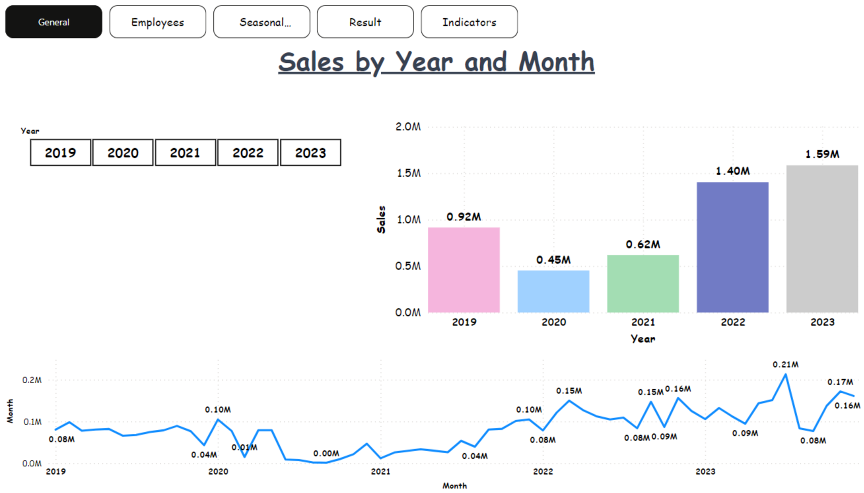

Fig. 5: Data visualization.

## 5. Discussion and Results

The design of the initial proposal based on the systematic literature review, it allowed us to identify and select the necessary industrial engineering tools to build the demand forecasting model. Likewise, to decrease the mean absolute error between the sales forecast and the actual sales. The choice of the appropriate tool will depend on the behavior pattern or trend of historical data in demand, also determined by their seasonality. Smoothing exponential is another tool that proposes an improvement in terms of forecasts, since, as time passes, each period decreases its weight. 7 different models were compared, among which the moving average, weighted moving average, exponential smoothing, holts, winters, and direction stand out; However, for the proposal of an inventory management system based on demand forecasts and the minimization of the MAPE error, the forecasting methodology with the lowest error is single smoothing exponential [4]. Based on historical sales data, a model is sought to help anticipate the demand that the company will receive. Demand estimation models for pellets for home use, industrial use and proposed process and information management for demand forecasting. To validate and check the quality of the data, the following metrics are considered: Mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE). Simple moving average with RMSE of 581,390, MAE of 422,716 and MAPE of 39% [3].

Google Collaboratory projections and the moving average tool were used to reduce the mean absolute percentage error for the year 2023, achieving a rate of 15,36%, whereas in the tourism sector, it stands at 12,03%. Nevertheless, there was a 4,24% reduction in MAPE compared to the year 2022. To compare the scenarios, a table will be created with the indicators of each scenario.

Table 9: Economic indicators of the improvement project.

| Criteria | Normal | Optimistic |
|---|---|---|
| MAPE | 15,36% | 12,03% |
| Economic NPV | $7 422,58 | $7 961,64 |
| Economic IRR | 26% | 27% |
| Benefit / Cost | 2,58 | 2,70 |
| Recovery period | 3 months y 5 days | 3 months y 12 days |

While it is true that in all scenarios the NPV (Net Present Value) is positive as well as the IRR (Internal Rate of Return) and Benefit/Cost, the difference between the scenarios lies in the Benefit, given that the project aims to increase income by reducing the mean absolute error.

For this reason, Risk will be used to define both input and output variables. First, the variables to be calculated must be defined. In this case, the main indicator is the MAPE, and the simulation will be based on this input indicator with its respective scenarios. Regarding the distribution, the triangular distribution will be selected due to three scenarios, as mentioned earlier, including a pessimistic one considering a 20%. This is shown in table 10.

Table 10: Criteria for Risk Simulator

| Criteria | Pessimistic | Normal | Optimistic |
|---|---|---|---|
| MAPE | 20% | 15,36% | 12,03% |

The average MAPE in the simulation was 15.7867%. However, in the project improvement, it was 15.36%. This means that the project improvement, along with the tools used, is feasible to achieve a lower MAPE, thereby increasing income by estimating an adequate demand. Finally, in figure 6, the simulation can be observed.

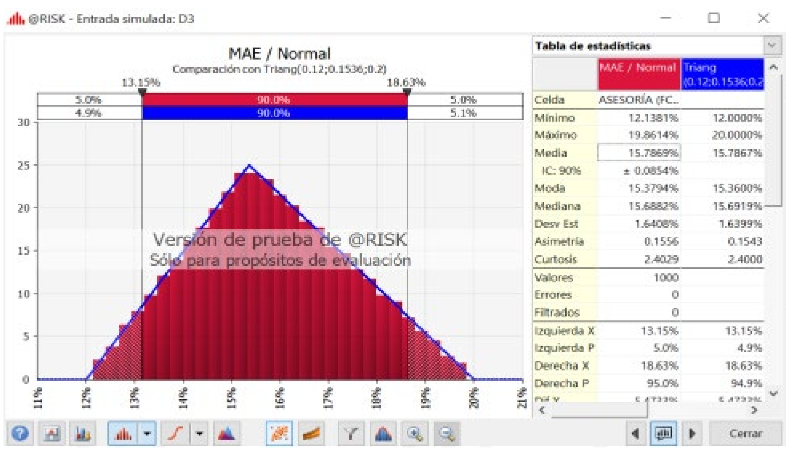


Fig. 6: Risk Simulator.

## 6. Conclusion

In conclusion, companies should adapt to technological advancements by implementing tools and strategies to remain competitive in the business world, thereby achieving growth. Within the improvement project, a tool was implemented that calculates the mean absolute percentage error of sales. This tool considers seasonality by annual rest periods and, to compare the data from the closed years, the missing information for 2023 was completed using Python codes in Google Collaboratory. This allowed annual sales to be forecast before the end of the year, enabling an optimal calculation of the mean absolute error for comparisons. Different techniques were used such as the exponential smoothing, moving average and simple, evaluating the annual error and highlighting that the moving average method was optimal, showing a 15,36% error. In the context of the tourism sector, an average MAPE of 12,03% is observed, and although this project reached 15,36%, it achieved a reduction of 4,24% compared to the previous year. In addition, a sales team of 20 workers was expected until 2023, although only 15 were reached until August. When analyzing the project with Risk Simulator, the effectiveness of the tool and Python to forecast is confirmed, generating an NPV of $7 422,58 with a mean absolute error of 15,36%. The literature review highlights the importance of choosing tools according to the behavior of historical data, influenced by the seasonality of monthly demand.

## 7. Reference

[1] Díaz, F. R., & Romaní, Y. L. (2022). Universidad del Zulia. https://doi.org/10.52080/rvgluz.27.98.13
[2] Sánchez Sepúlveda, Carolina del Pilar (2018). Medición en la precisión de los pronósticos de ventas y su efecto en los costos de la empresa. Universidad Andres Bello https://dspace.udla.edu.ec/handle/33000/9831
[3] Ackermann, A. & Sellito, M. (2022). Métodos de pronóstico de la demanda: una revisión de la literatura. Revista Innovar. https://doi.org/10.15446/innovar.v32n85.100979

[4] Marroquín, R. (2018). Propuesta de un sistema de gestión de inventarios a partir de pronósticos de la demanda dentro de una imprenta (Tesis de pregrado). Universidad de las Américas, Quito. https://dspace.udla.edu.ec/handle/33000/9831
[5] Cisneros-Martínez, J, & Fernández-Morales, A. (2016). Concentración estacional de la demanda hotelera en Argentina. Revista de Estudios Regionales.