# Integrating statistical analysis into everyday business operations using mobile technologies

Lyazat Naizabayeva[1,†] and Gulzat Turken[2,*,†]

[1] International Information Technology University, 34/1 Manas St., Almaty, Kazakhstan
[2] Al-Farabi Kazakh National University, Al-Farabi Avenue 71, 050040, Almaty, Kazakhstan

**Abstract**

This study focuses on analyzing the company's production and sales, and developing a mobile application for statistical data analysis. Using ARIMA methodology for sales modelling and Flutter, MySQL, Laravel PHP and Python technologies for application development, we conducted a comprehensive analysis of the company's business processes. The results show seasonality of sales with peaks in spring and December, with average daily sales of 118 units and revenue of KZT 532,930. The developed mobile application improves the efficiency of operations and supports the decision-making process. The study demonstrates the importance of adopting modern technology to improve competitiveness and optimize production. Recommendations include optimizing inventory, adapting marketing strategies to seasonal fluctuations and strengthening environmental product positioning to stimulate sales during periods of low demand.

**Keywords**

mobile application, product management, data analysis, MYSQL, LARAVEL, ARIMA model

## 1. Introduction

In the context of the accelerating digitalization of economic processes, the integration of statistical analysis into the operational activities of enterprises is becoming critical for maintaining their competitiveness. Despite significant progress in the field of analytical tools, there is a noticeable gap between the potential of modern statistical analysis methods and their practical application in everyday business operations. This study aims to explore how this gap can be bridged through the use of mobile technologies.

The methodological framework of the study relies on a synthesis of quantitative and qualitative methods. The central element is a mathematical model of sales forecasting developed by the authors, based on multiple linear regression, taking into account seasonal fluctuations and long-term trends.

The innovativeness of the approach lies in the integration of this model into a mobile application, which allows decentralizing the decision-making process and increasing the responsiveness to changes in market conditions. The empirical part of the research includes the analysis of the effectiveness of the implementation of the developed application in various functional divisions of the company.

Theoretical significance of the work consists in the development of conceptual foundations for the integration of statistical analysis into business processes using mobile technologies. The practical significance lies in the development of specific recommendations for the implementation of such systems in organizations of different profiles.

DTESI 2024: 9th International Conference on Digital Technologies in Education, Science and Industry, October 16–17, 2024, Almaty, Kazakhstan
* Corresponding author.
† These authors contributed equally.
naizabayeva@gmail.com (L. Naizabayeva); turken.gulzat@gmail.com (G. Turken)
0000-0002-4860-7376 (L. Naizabayeva); 0000-0003-4981-514X (G. Turken)

The study also considers potential limitations and challenges associated with the implementation of the proposed approach, including information security, staff training and integration with existing enterprise information systems. Strategies to minimize the identified risks are proposed.

The purpose of this paper is to develop and empirically validate a methodology for integrating complex statistical analysis into routine business processes through the lens of mobile applications. Elite Baby Production, a company specializing in baby products, is chosen as the object of the study, which allows for a detailed analysis in the context of a specific industry.

The results of this study contribute to the development of the theory and practice of statistical analysis in business process management and can serve as a basis for further research in the field of integration of analytical tools into the operational activities of enterprises.

## 2. Literature review

The literature review of all articles emphasizes the importance and development of the application of statistical analysis in various fields. Efficient optimization techniques play a vital role in sales forecasting [1]. The digital age has significantly impacted several industries such as retail sector, financial services, healthcare and education depending on the use of modern technology. The ability to store, process and analyze huge amounts of data is crucial to make informed decisions [2]. This research work [3] analyses d datasets using appropriate statistical techniques. As a result, we have obtained appropriate charts for our datasets, which is very useful for easy and quick communication of summary and conclusions to the viewers.

The following research paper [4] provides an overview of the key concepts of statistical data analysis. The main objective was to provide a quick reference guide to the most commonly used concepts in this field. The research paper [5] conducted an in-depth study on the design of data collection process, preparation of data for analysis, data analysis and communication of data analysis results. The research paper [6] used different technologies to compare algorithms for urban road network trajectory analysis conducted with different amounts of big data to determine the calculation time, accuracy and ways to improve the calculation time. This article [7] focuses on defining data analysis and the concept of data preparation. It provides an overview of data analysis methods, starting with a brief description of six main categories. Following that, the paper delves into the most widely used statistical methods, including descriptive, explanatory, and inferential analyses. The following article [8] is dedicated to providing a comprehensive understanding of data analysis and the foundational concept of data preparation. It begins by exploring the definition of data analysis, shedding light on the essential processes involved in preparing data for analytical purposes. The researchers in this article [9] addresses the common challenge faced by many researchers, particularly beginners, in choosing the appropriate statistical method for analyzing research data. It offers a clear and straightforward guide to assist in the selection of statistical techniques. The paper outlines key factors that should be considered when deciding on a statistical method for data analysis, and provides recommendations for suitable methods based on specific research problem scenarios.

## 3. Mathematical model and methods

The ARIMA model provides another method for time series forecasting. Exponential smoothing and ARIMA models are two of the most widely used time series forecasting methods, and based on the expansion of these two forecasting methods, many other forecasting methods have been created. Unlike the exponential smoothing model, which focuses on trends and seasonality in the data, the ARIMA model aims to depict autocorrelations in the data. When we combine the difference and autoregressive models with the moving average model, we can obtain a non-seasonal ARIMA model. ARIMA stands for Auto Regressive Integrated Moving Average.
The mathematical model can be represented in the following form:

$$y_t^{'}=c+\phi_1 y_{(t-1)}^{'}+\cdots+\phi_p y_{(t-p)}^{'}+\theta_1 \varepsilon_{(t-1)}+\cdots+\theta_q \varepsilon_{(t-q)}+\varepsilon_t \quad (1)$$

Here $y_t^{'}$ - the Stationarized Time Series; c – Constant; $\phi_1 y_{[t-1]}^{'} + \cdots + \phi_p y_{[t-p]}^{'}$ - Autoregressive (AR) Component; $\theta_1 \varepsilon_{[t-1]} + \cdots + \theta_q \varepsilon_{[t-q]}$ - Moving Average (MA) Component; $\varepsilon_t$- Current Error Term.

In the above equation, $y_t^{'}$ is the difference sequence. The “predictor variables” on the right hand side include the delayed value of $y_t$and the error of the delay. We call this model the ARIMA (p, d, q) model with parameters p - Autoregressive model order, d - Differential order and q - Moving average model order. The smoothness and reversibility conditions in the autoregressive and in the moving average models still apply in the ARIMA model.

Once we start combining different models to form complex models, the delay operator becomes extraordinarily easy. Equation (1) can be expressed as:

$$\underbrace{\left(1-\phi_1 B-\cdots-\phi_p B^p\right)}_{\text{AR(p)}} \underbrace{(1-B)^d}_{\text{d differences}} y_t = c + \underbrace{\left(1+\theta_1 B+\cdots+\theta_q B^q\right)}_{\text{MA(q)}} \varepsilon_t \tag{2}$$

The auto.arima () function is very useful, but relying entirely on automated programs is a dangerous behavior. So even if you have used a program to automatically select a model, it is still essential to understand the characteristics and behavior of the model. The value of p  is important when there are cycles in the data. In order to make periodic predictions, p ≥ 2p ≥ 2 is one of the necessary conditions, For the our conducted  model, the model is  characterized by periodicity when $\varnothing_1^2 + 4\theta_2 < ¿0$. In this case, the average length of the cycle is

$$\frac{2\pi}{\arccos\left(-\frac{\phi_1(1-\phi_2)}{4\phi_2}\right)} \tag{3}$$

Implementing ARIMA in a mobile application requires a mix of frontend and backend development skills, with an emphasis on integrating statistical models into the workflow. By using tools like Python for backend data processing, ARIMA modeling, and an intuitive mobile interface, users can easily perform time series forecasting directly from their mobile devices. The ARIMA model can be implemented on the backend using Python, taking advantage of its robust libraries for time series analysis. Start by installing the required dependencies. This backend configuration enables the mobile application to input time series data, apply the ARIMA model for processing, and generate forecast results that can then be visualized or further examined. The mobile app for ARIMA forecasting allows users to upload time series data, select or auto-tune ARIMA parameters, and trigger forecasting via a simple interface. The backend processes the data and returns the forecasted results, which are visualized through charts showing actual and predicted values along with confidence intervals. Advanced features like automatic parameter tuning, alerts, and performance optimization ensure the app handles various datasets efficiently.

Description of research experiment data: The dataset for this research was sourced directly from the “Elite Baby Production” company, encompassing a comprehensive analysis of data collected over a one-year period, from 2023 to 2024. This robust dataset serves as the foundation for conducting detailed statistical evaluations and deriving insightful trends related to the company's operations. The development of special purpose mobile applications, such as the Elite Baby Production (EBP) application created to facilitate tracking and monitoring of the company's progress, is still relevant today. The key technologies for this project are MySQL for database management, PHP for server-side scripting, Python for statistical data analysis, and Flutter for the mobile application interface. These technologies were chosen to provide reliable operation, secure administration and efficient data processing, and all were specifically designed to meet the requirements of the application. Figure 1 shows the architecture of how the GET and POST methods work and how the endpoints work.

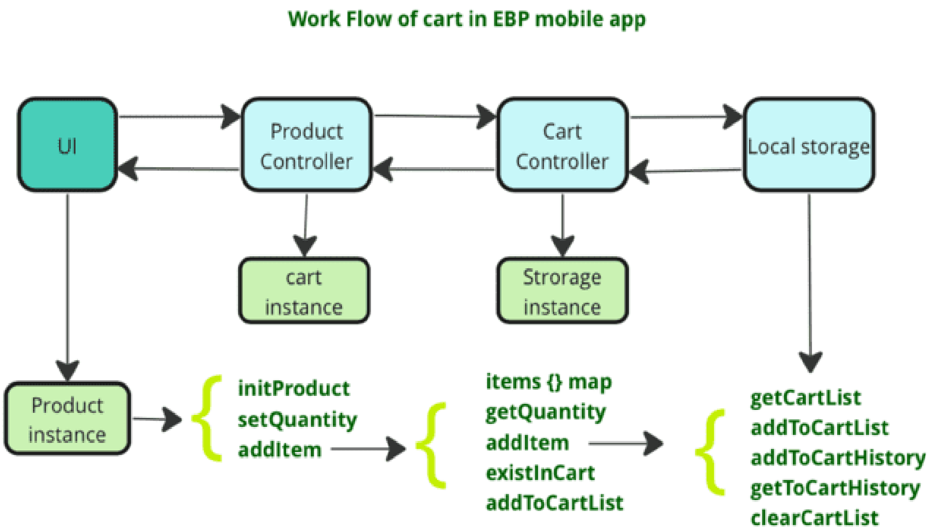


**Figure 1:** Architecture of methods.

The previously mentioned Laravel framework has an out-of-the-box connection to MySQL, which was chosen for this project because of its stability, reliability and extensive feature set that perfectly matches the project requirements. In addition, MySQL's compatibility with Eloquent ORM (object-relational mapping) for Laravel optimizes database operations, making it easier to manage database schemas, run complex queries and maintain data integrity. Laravel and MySQL integrate seamlessly, improving the development process and enabling faster and more secure application development. Here is what the final database table for an application looks like, shown in Figure 2.

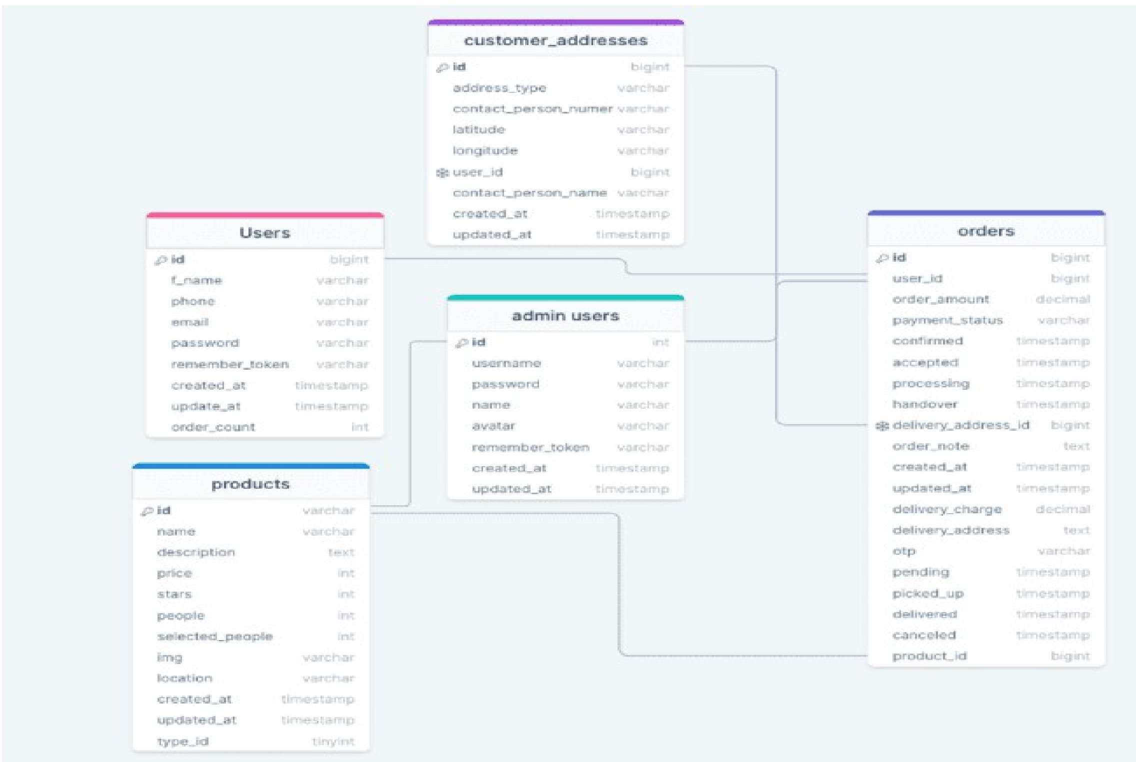


**Figure 2:** ER diagram.

Forecasting models for time series analysis: the ARIMA model consists of 3 components.

```
model = ARIMA(train, order=(5, 1, 0))
model_fit = model.fit()
print(model_fit.summary())
```

```
                               SARIMAX Results
==============================================================================
Dep. Variable:           order_amount   No. Observations:                 1168
Model:                 ARIMA(5, 1, 0)   Log Likelihood               -8335.870
Date:                Fri, 07 Jun 2024   AIC                          16683.740
Time:                        04:14:22   BIC                          16714.113
Sample:                    01-01-2020   HQIC                         16695.197
                         - 03-13-2023
```

```
import pandas as pd
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_absolute_error, mean_squared_error
import numpy as np
import matplotlib.pyplot as plt
data = pd.read_csv('order_Data.csv')
data.head()
```

| | id | user_id | order_amount | payment_status | order_status | confirmed | accepted | scheduled | processing | handover | failed | scheduled_at | delivery_address_id |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 52 | 258.223354 | paid | confirmed | False | False | False | True | True | False | 2020-01-01 04:00:00 | 1668 |
| 1 | 2 | 93 | 636.767083 | paid | handover | True | True | True | False | True | False | 2020-01-02 21:00:00 | 4106 |
| 2 | 3 | 15 | 499.331112 | unpaid | failed | False | False | True | False | False | True | 2020-01-03 21:00:00 | 1074 |
| 3 | 4 | 72 | 577.561434 | paid | scheduled | False | False | False | True | False | True | 2020-01-04 08:00:00 | 3991 |
| 4 | 5 | 61 | 839.725894 | unpaid | confirmed | True | False | True | False | False | False | 2020-01-05 07:00:00 | 4360 |

**Figure 3:** Forecasting models.

Choosing an appropriate statistical method is a very important step in data analysis [10]. This section discusses the design and implementation of a statistical analysis dashboard that visualizes user, product and order statistics as well as sales forecasts. The dashboard uses Laravel for internal data processing and Chart.js for external visualization. Statistical analysis helps to understand trends and patterns in the data, which is critical to making informed business decisions.

## 4. Results and discussion

A comprehensive statistical analysis of various key metrics related to users, products and orders over different periods is provided. This helps in tracking performance, identifying trends and forecasting future results as shown in Figures 4, 5, 6. Sales forecasting data, which is critical for predicting future trends, is displayed in plain text for easy and quick access to the dashboard. This approach ensures that users can easily understand the forecasted sales figures without the need for complex visualization. In addition, the front-end interface allows for dynamic updates and real-time data display, ensuring that charts and statistics reflect the most up-to-date information from the server side.

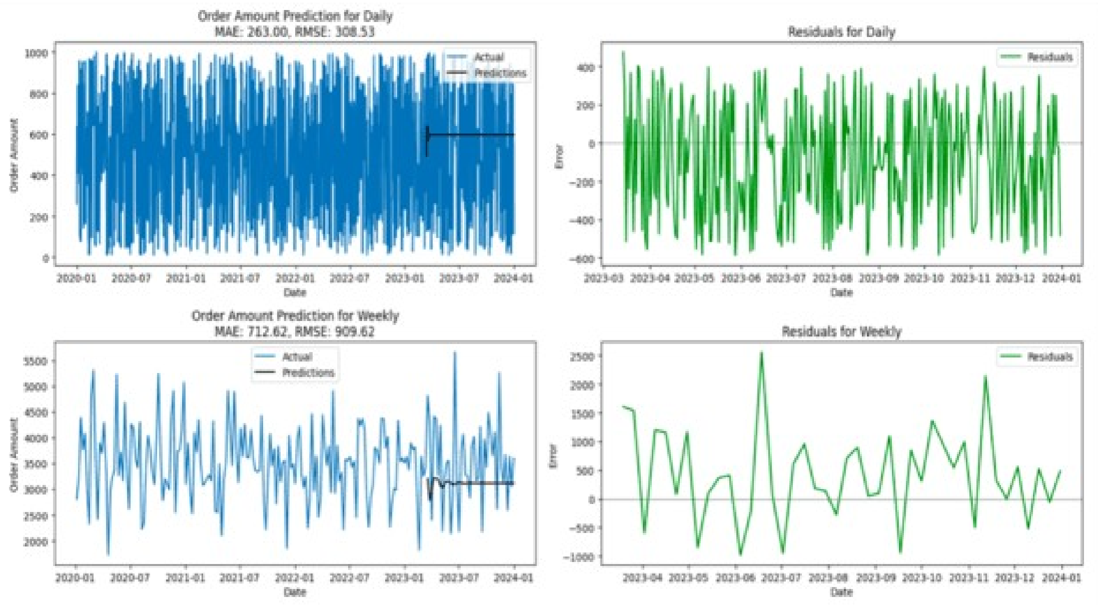


**Figure 4:** Daily order forecasting model.

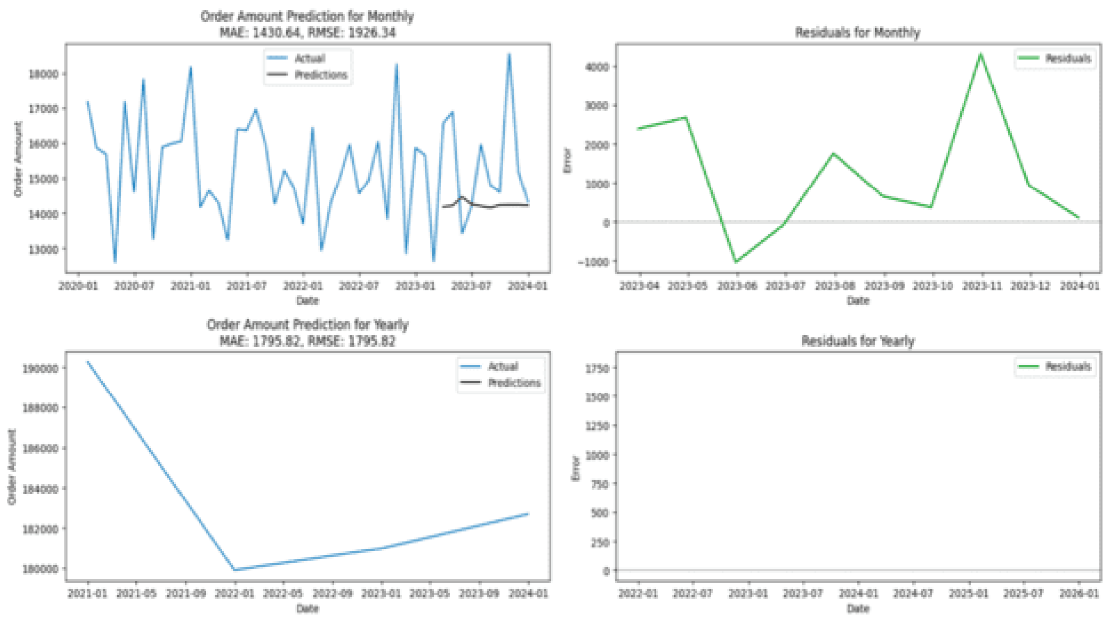


**Figure 5:** Annual order forecasting model.

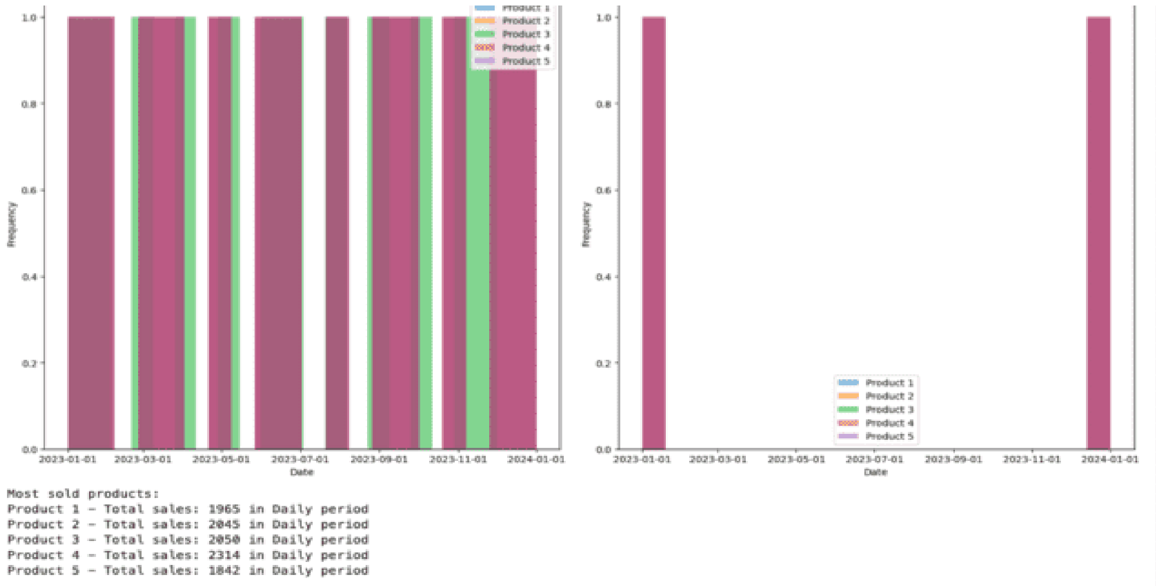


**Figure 6:** Accuracy results.

Let us analyse the results of modelling the sales of Elite Baby Production products based on synthetic data. The mean square error (MSE) is 79.02, indicating an average forecast deviation from actual values of 8.89 sales units (square root of 79.02). The R-squared is 0.38, which means that the model explains 38% of the variation in the data. Average daily sales are about 118 units. Minimum sales are 77 units and maximum sales are 187 units. The price is fixed at 4500 tenge per pack. The average daily income is about KZT 532,930.

Sales peak in April (133.42 units) with revenue of KZT 600,399 per day. The lowest sales are in September (103.19 units) with revenue of KZT 464,340 per day. High sales in spring and December, decline in summer and autumn. Total Sales (2 years): total Sales (2 years) is 86453 units, total Revenue (2 years) is KZT 389,038,619, average Daily Sales is 118.43 units, average Daily Revenue is KZT 532,930.

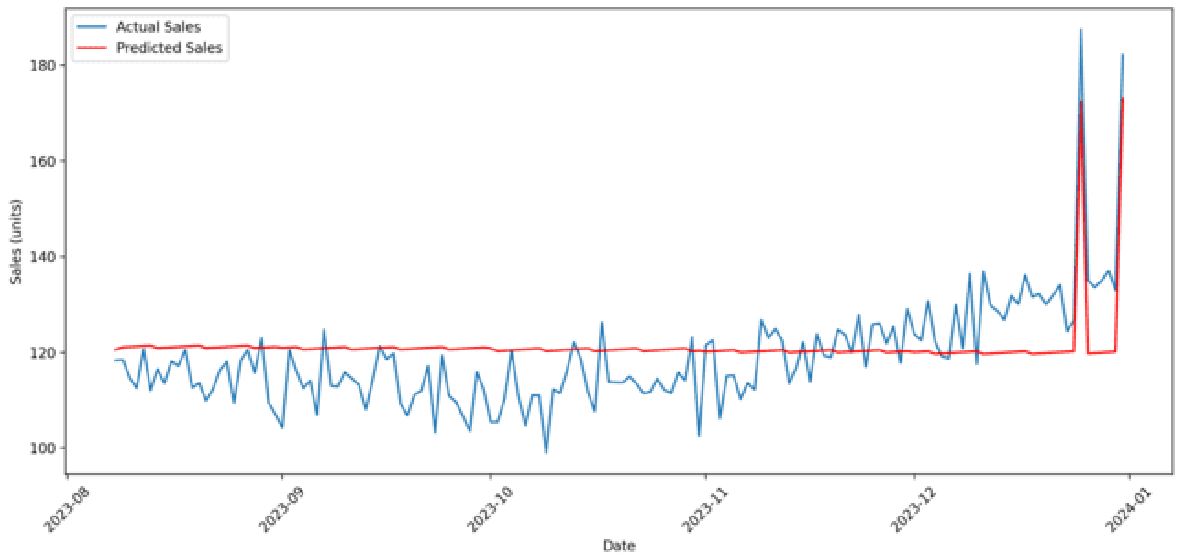


**Figure 7:** Actual and predicted sales by ARIMA model.

Figures 7, 8 show a comparison of actual and forecasted sales. The ARIMA model captures the general trend quite well.

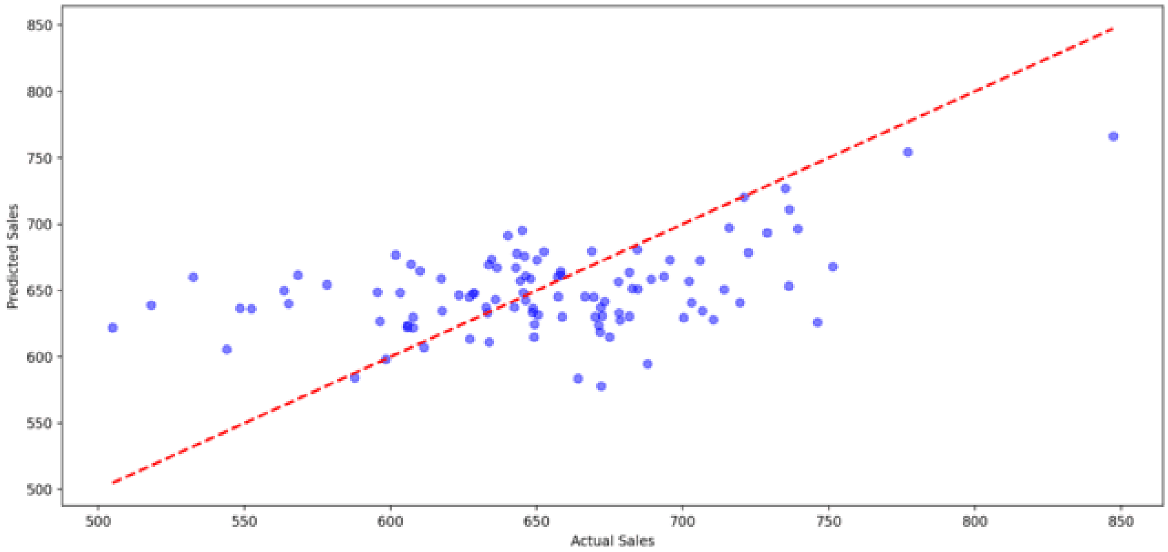


**Figure 8:** Actual and Predicted Sales.

Figure 9 shows the monthly trend of sales and revenue, clearly showing the seasonality of the business.

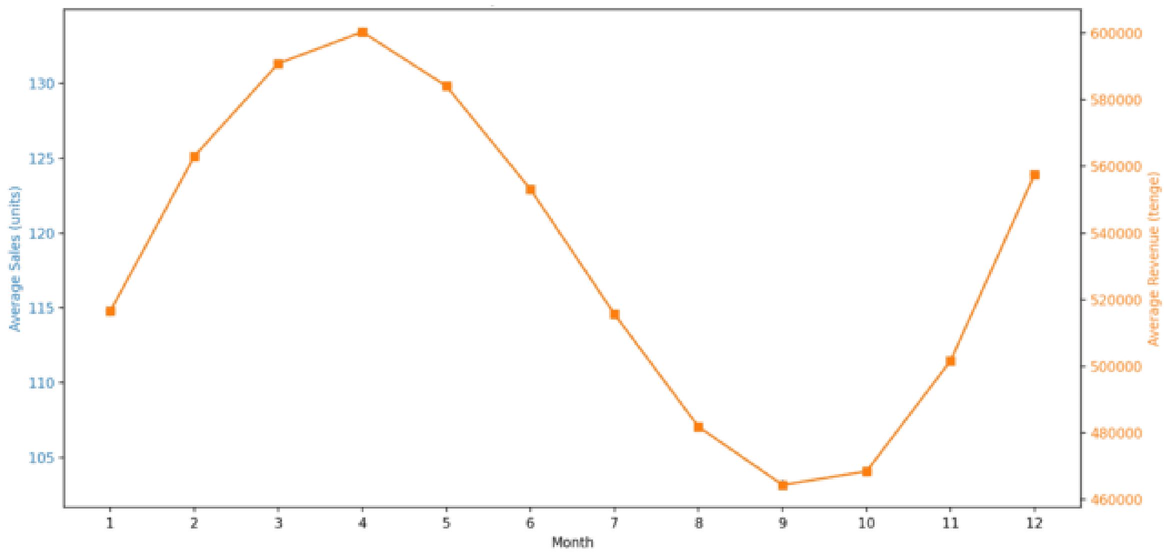


**Figure 9:** Monthly sales and revenue trend.

The implementation of sales forecasts will help Elite Baby Production to optimise its business processes, increase sales and profits, and improve customer satisfaction. By regularly analysing data and updating strategies based on new insights, the company will remain competitive in the organic market.

## 5. Conclusion

This study is a comprehensive analysis of the production and sales of environmentally friendly nappies of the company «Elite Baby Production», as well as the development of an innovative mobile application for statistical data analysis. The results of our research have significant theoretical and practical implication for the development of environmentally responsible business and application of modern technologies in enterprise management.

The analysis showed a clear seasonality in sales of eco-nappies with peaks in spring and December. This knowledge is critical for optimising production and inventory management.

With average daily sales of 118 units and revenue of KZT 532,930, the company demonstrates stable demand for its products, which confirms the potential of the organic baby products market.

The developed mobile application based on Flutter using MySQL, Laravel PHP and Python to analyse data significantly improves the efficiency of operations and the quality of decision-making in the company.

The study confirms the importance of the environmental aspect in marketing strategy, especially to stimulate sales during periods of low demand.

Adopting advanced data analytics technologies allows a company to respond more effectively to market changes and optimise its operations.

Further research could also look at integrating artificial intelligence technologies into a mobile app to improve demand forecasting and optimise the supply chain. In addition, benchmarking the performance of eco-friendly nappies against their traditional counterparts can provide valuable insights for the development of the industry as a whole.

In conclusion, this study demonstrates how a combination of environmental responsibility and technological innovation can create a competitive advantage in today's business landscape. Elite Baby Production serves as an example of how businesses can successfully integrate sustainability into their business models while improving operational efficiency through the adoption of advanced data analytics technologies.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References


[1] Gulzat T., Lyazat N., Siladi V., Gulbakyt S., Maxatbek S. Research on predictive model based on classification with parameters of optimization // Neural Network World. 2020. Vol. 5. P. 295–308. DOI: 10.14311/NNW.2020.30.020.

[2] Turken G., Naizabayeva L., Satymbekov M., Abdiakhmetova Z. Research and Development of Enterprise Data Warehouse Based on SAP BW // Modeling SIST 2023 IEEE International Conference on Smart Information Systems and Technologies, Proceedings. 2023. P. 5–9.

[3] Mishra P., Pandey C.M., Singh U., Gupta A. Scales of measurement and presentation of statistical data // Annals of Cardiac Anaesthesia. 2018. Vol. 21. P. 419–422.

[4] Sarmento R., Costa V. An Overview of Statistical Data Analysis // arXiv preprint arXiv:1908.07390. 2019. P. 2–3. DOI: 10.48550/arXiv.1908.07390.

[5] Ott R.L., Longnecker M. An Introduction to Statistical Methods and Data Analysis. Fifth Edition. DUXBURY, United States of America.

[6] Temirbekova Z., Naizabayeva L., Turken G., Abdiakhmetova Z., Satymbekov M. Identification of an algorithm for the analysis and study of urban road network trajectories // Eastern-European Journal of Enterprise Technologies. 2024. Vol. 2(3-128). P. 14–27.

[7] Hamed Taherdoost. Different Types of Data Analysis; Data Analysis Methods and Techniques in Research Projects Authors. International Journal of Academic Research in Management (IJARM), 2020, 9 (1), pp.1-9. ffhal-03741837.

[8] Rui Portocarrero Sarmento, Vera Costa, AnOverview of Statistical Data Analysis, August 2019, DOI:10.48550/arXiv.1908.07390, ResearchGate,

[9] Gideon, S.N.&Nwogu, V.U., Choosing the Appropriate Statistical Methods for Data Analysis in a Research: A Solution to the Puzzle Faced by Beginners, Propellers Journal of EducationVol. 2

No. 1, 1-8(June 2023) ISSN: 2971-639X (Online)ijvocter.comCitation:Gideon, S.N. & Nwogu, V.U. (2023). Propellers Journal of Education 2(1), pp. 1-8.

[10] Mishra P., Pandey C.M., Singh U., Keshri A., Sabaretnam M. Selection of Appropriate Statistical Methods for Data Analysis // Annals of Cardiac Anaesthesia. 2019. Vol. 22(3). P. 297–301. DOI: 10.4103/aca.ACA_248_18.