Datathon
Project Proposal

Team 118 - Sky Five

SKY FIVE
DATATHON 2023

Membership - Sky Five

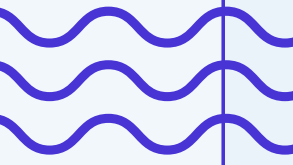|	
Nguyễn Minh Nhật (Leader)	Problem statement and solution presentation
Trần Trọng Nghĩa	User Interface design
Đỗ Khải Hưng	Technological solutions
Nguyễn Quốc Trung	Technological solutions
Võ Hoàng Hoa Viên	Data visualization

SKY FIVE
DATATHON 2023
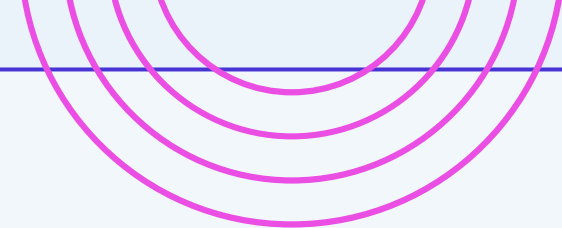
Table of contents

01

Introduction

Introducing the problem

Businesses must forecast future sales and inventory to efficiently manage goods, save costs and inform strategic decision-making.

Businesses must assess the efficiency and determine the impacts of proposed business strategies.
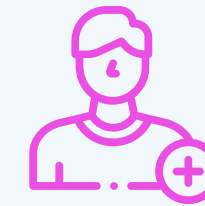
Introducing MVP

Data normalization

Data cleaning and normalization ensure only relevant features and patterns are retained.

Sales & inventory prediction

A model performs forecasts that can adapt to different market situations.

Strategy evaluation

The model can be adjusted in accordance to business strategies.

Opportunities

Apply everywhere

Build a catch-all model from national retail data to apply to small enterprises.

Simulate strategies

Allow businesses to realize the potential of their stategies.

Save money

Predict market changes to reduce storage and logistics expenses.

02

Problem
Statement

Identify the problem

- Sales forecasting
- Inventory quantity management
- Strategy simulation

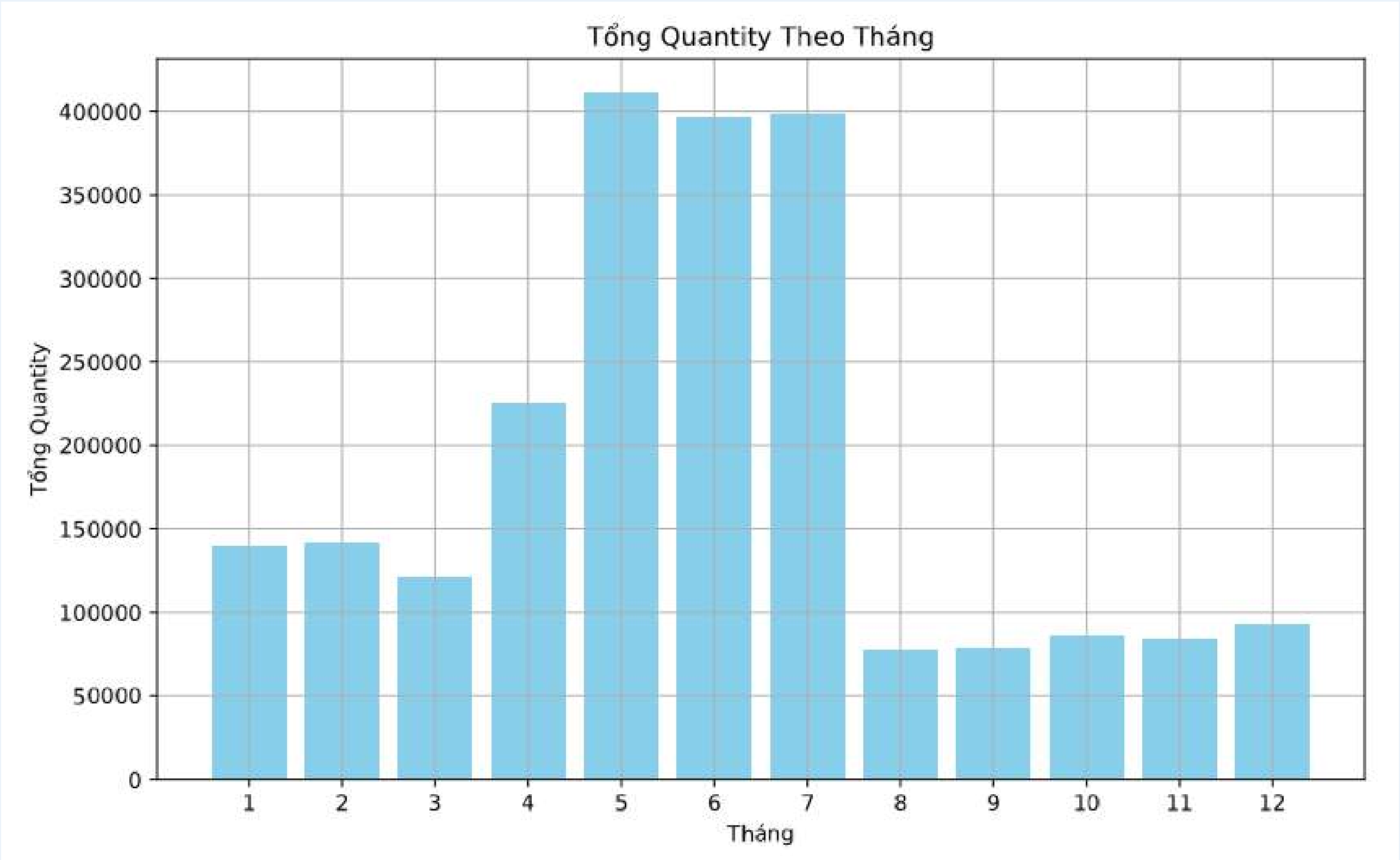# Dataset overview: Sales and Inventory Data of Vietnam Retailers

The dataset consists of 3 components:
- Inventory data area: stocks of goods across plants.
- Master data area: relevant entity information.
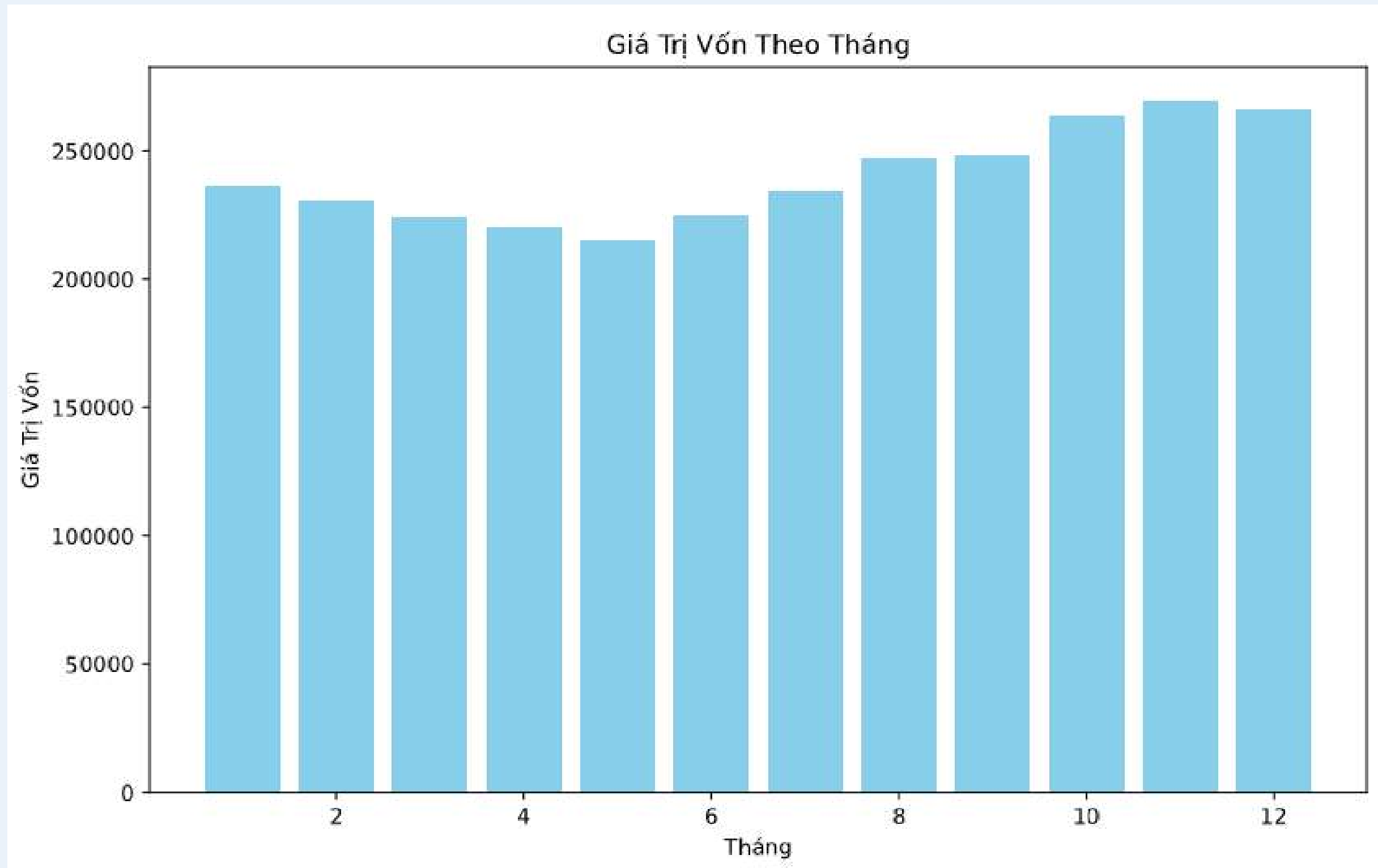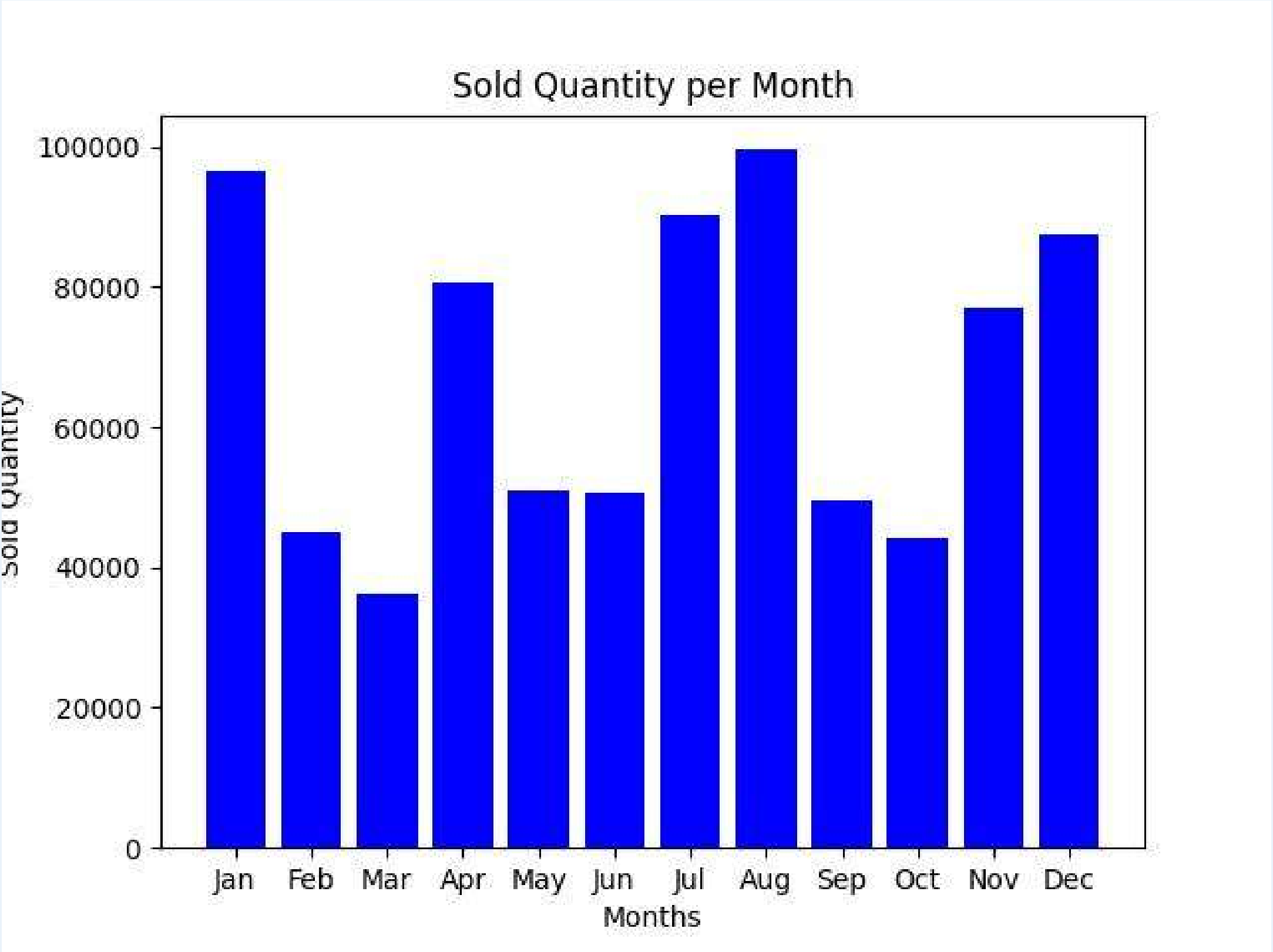- Sales data area: monthly sales data in 2022.

Inventory data

Monthly inventory quantity, 2022

Inventory data

Monthly inventory value, 2022

Sale data

Monthly sales, 2022

Sale data
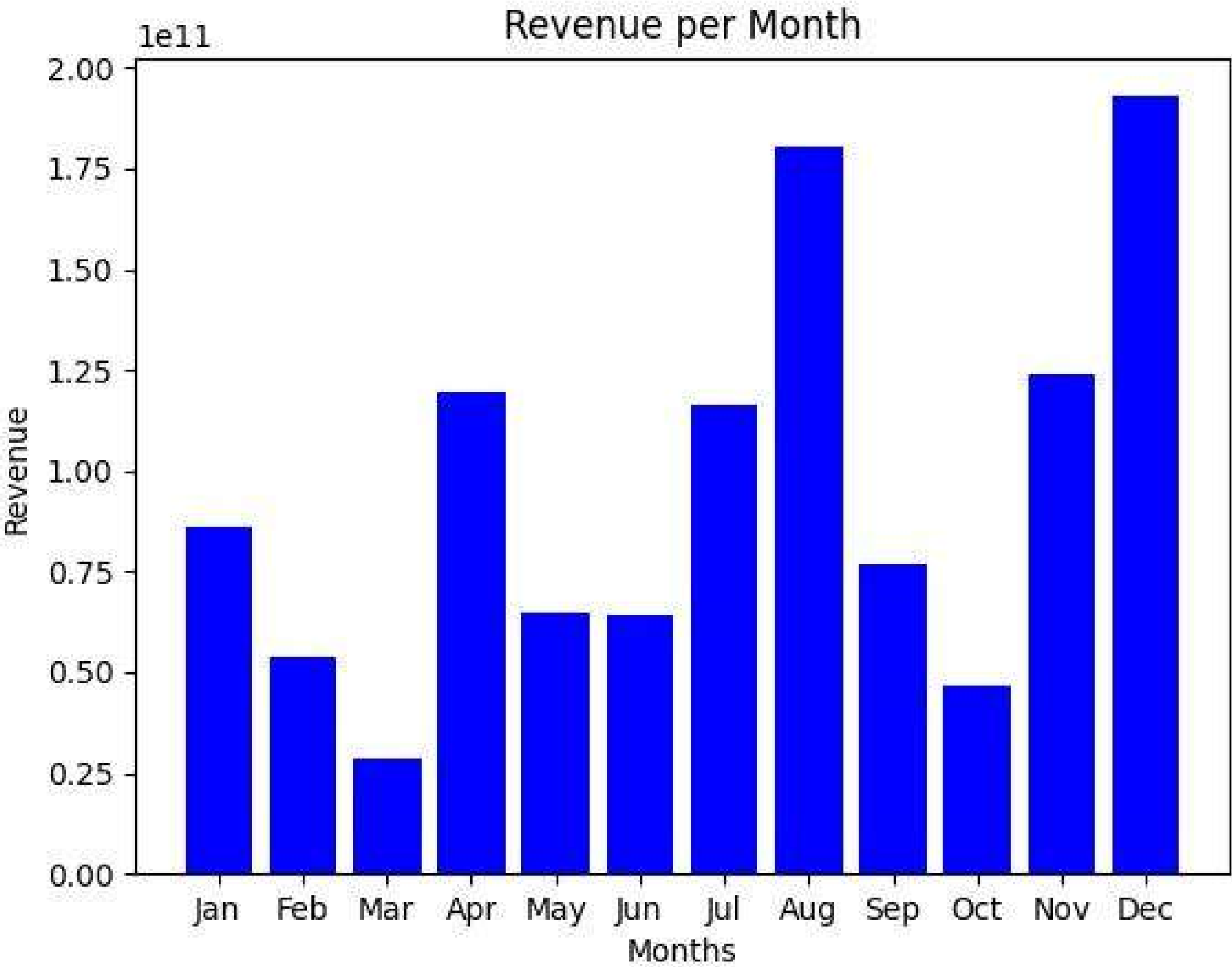
Monthly revenues, 2022

Sale data

Tổng Giá Vốn và Tổng Giá Bán Theo Tháng

Legend:
- Cost Price
- Net Price

Y-axis: Tổng Giá Trị (1e10)

X-axis: Tháng (1–12)
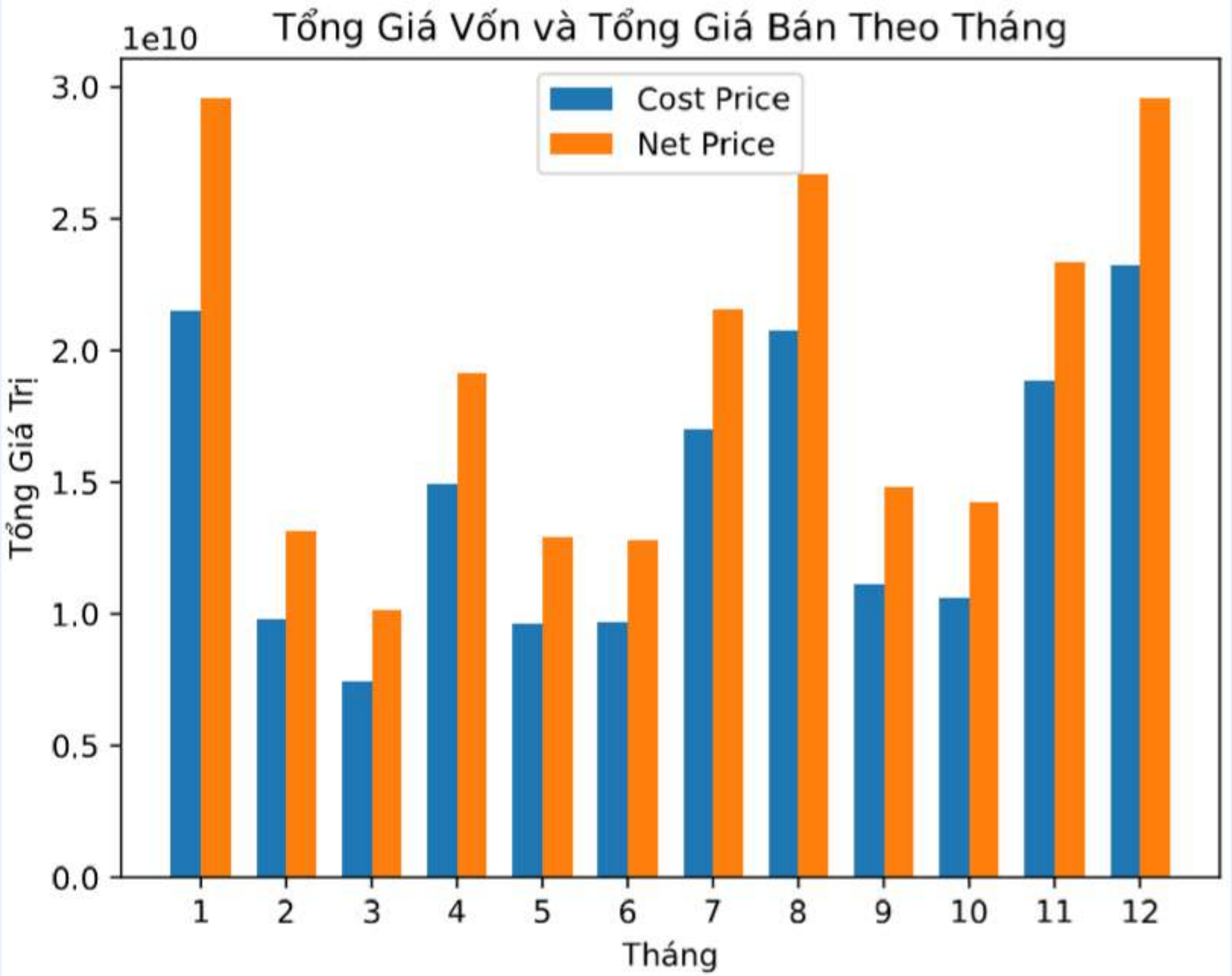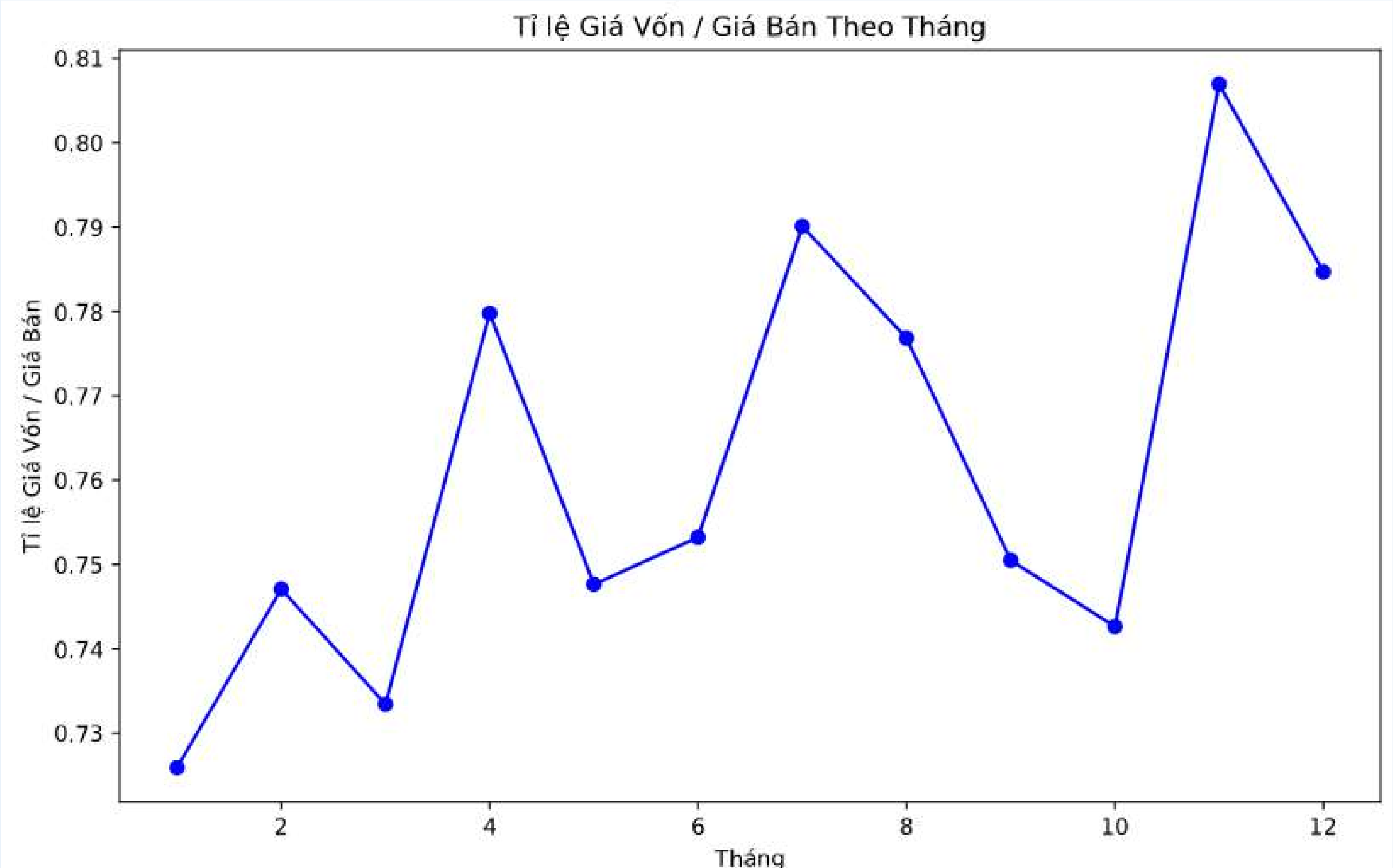
Monthly cost and net price, 2022

Sale data

Monthly cost price to net price ratio, 2022

Comments on data

The data illustrates a segment of fashion retail sales in Vietnam.

Regarding inventory data: During the months of May, June, and July, the inventory levels surpass those of other months. This is attributed to the fact that a majority of shoppers tend to make purchases at the commencement or conclusion of the year.

The concentration of holidays in the initial and final months of the year is significantly higher compared to the middle months.

Towards the year-end, stores frequently lower promotional prices in anticipation of the new year, resulting in a substantial reduction in inventory due to a higher number of customers.

○

Comment on data

- Per observation, monthly revenue shares the same trend as monthly sales revenue:
  - From January to March: Gradual decrease.
  - From March to April: Sharp increase.
  - From May to June: Steady.
  - From June to August: Sharp increase.
  - From August to December: Gradual increase.

Challenge

- **Market volatility**: difficulties predicting volatility.
- **Limited data**: the model is calculated based on data from 2022 to 2023, so there is a lack of connectivity between economic periods, especially in the context of post-pandemic economic fluctuations. More data that accounts for interconnectivity between periods is needed to produce reliable judgments.

Challenge

- The data only represents a portion of fashion retail sales in Vietnam, so the proposed model is susceptible to bias when dealing with data on other goods and data at regional / international scale.

Challenge

- **Security and privacy:** Protecting an enterprise's business data is a top priority. Before processing data, private information must be removed while still retaining the dataset's representativeness and usability.
- NLP models employed to extract user requirements and match them with the model's parameters are difficult to fine-tune.

3. Solution Overview

Solution Overview

Prediction model

- Employ ARIMA model for time series analysis

Strategies simulation

- Utilize a rule set to determine which parameters are updated following input feedback.

Data visualization

- Connect to database, define relationships and objects.
- Join data tables.
- Convert plain texts into data tables using ORM.

Forecasting

- Use the Bard NLP model to import photos and give reviews.
- Highlight areas in the data with fluctuating sales or inventory.
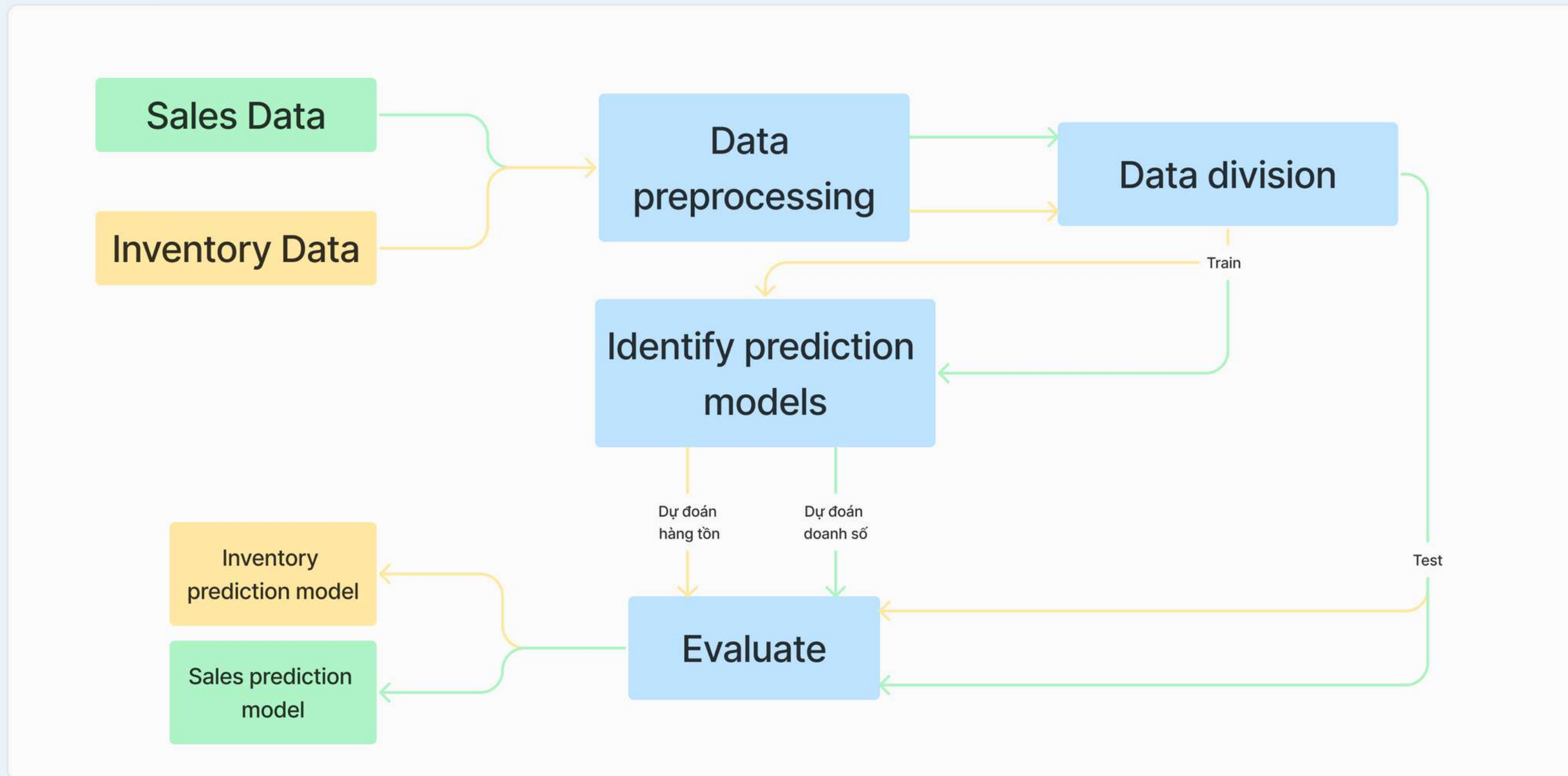
04

• • • • •

Methodologies

ARIMA model

- Step 1: Import the model from a pre-built Python library.
- Step 2: Identify the optimal model parameters using Autocorrelation Function, Information Criterion, and Grid Search.
- Step 3: Implement the chosen model with the determined parameters.
- Step 4: Reassess the model's effectiveness using proposed metrics on the test set.
- Step 5: Prepare guidance documents and officially announce the adopted model.

Prediction Model

Prediction Model

Sales Data → Data preprocessing → Data division

Inventory Data → Data preprocessing

Data division → Train → Identify prediction models

Identify prediction models:
- Dự đoán hàng tồn → Evaluate
- Dự đoán doanh số → Evaluate

Data division → Test → Evaluate

Evaluate → Inventory prediction model

Evaluate → Sales prediction model

Prediction Model

- Step 1: Gather data on retail sales and inventory in Vietnam for the period 2022-2023.
- Step 2: Preprocess the collected inventory data by eliminating irrelevant information and emphasizing key details.
- Step 3: Split the data into two subsets:
  - Train: Used for constructing prediction models.
  - Test: Employed to assess and refine the developed model.
- Step 4: Assess the model's effectiveness by revisiting proposed metrics, using the test set.
- Step 5: Compose documentation and launch.

MVP
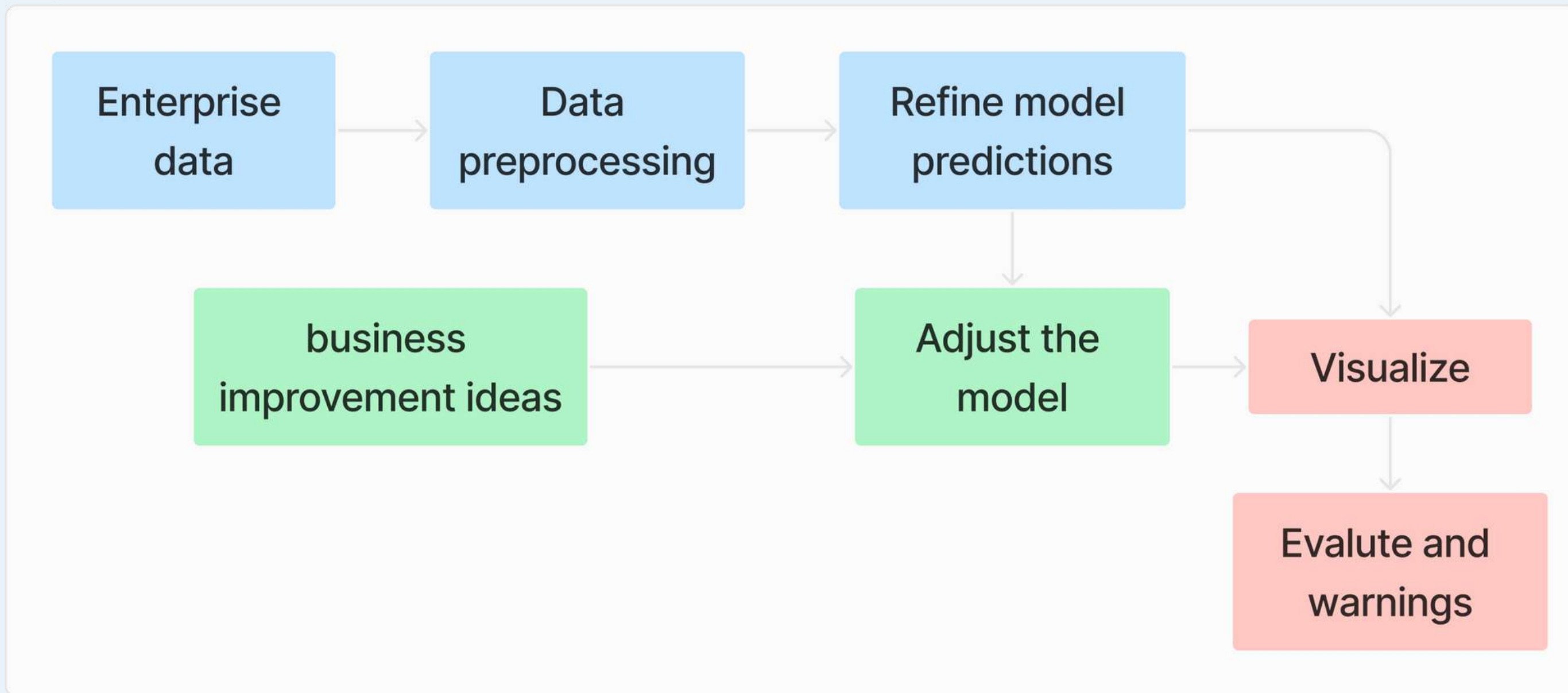
MVP

- Step 1: Acquire data submitted by the business.
- Step 2: Preprocess data.
- Step 3: Identify and estimate parameters for the prediction model.
- Step 4: Receive user improvement requests and standardize these requests into a set of rules.
- Step 5: Modify the parameters of the prediction model based on user feedback.
- Step 6: Visualize predicted data over a defined future period.
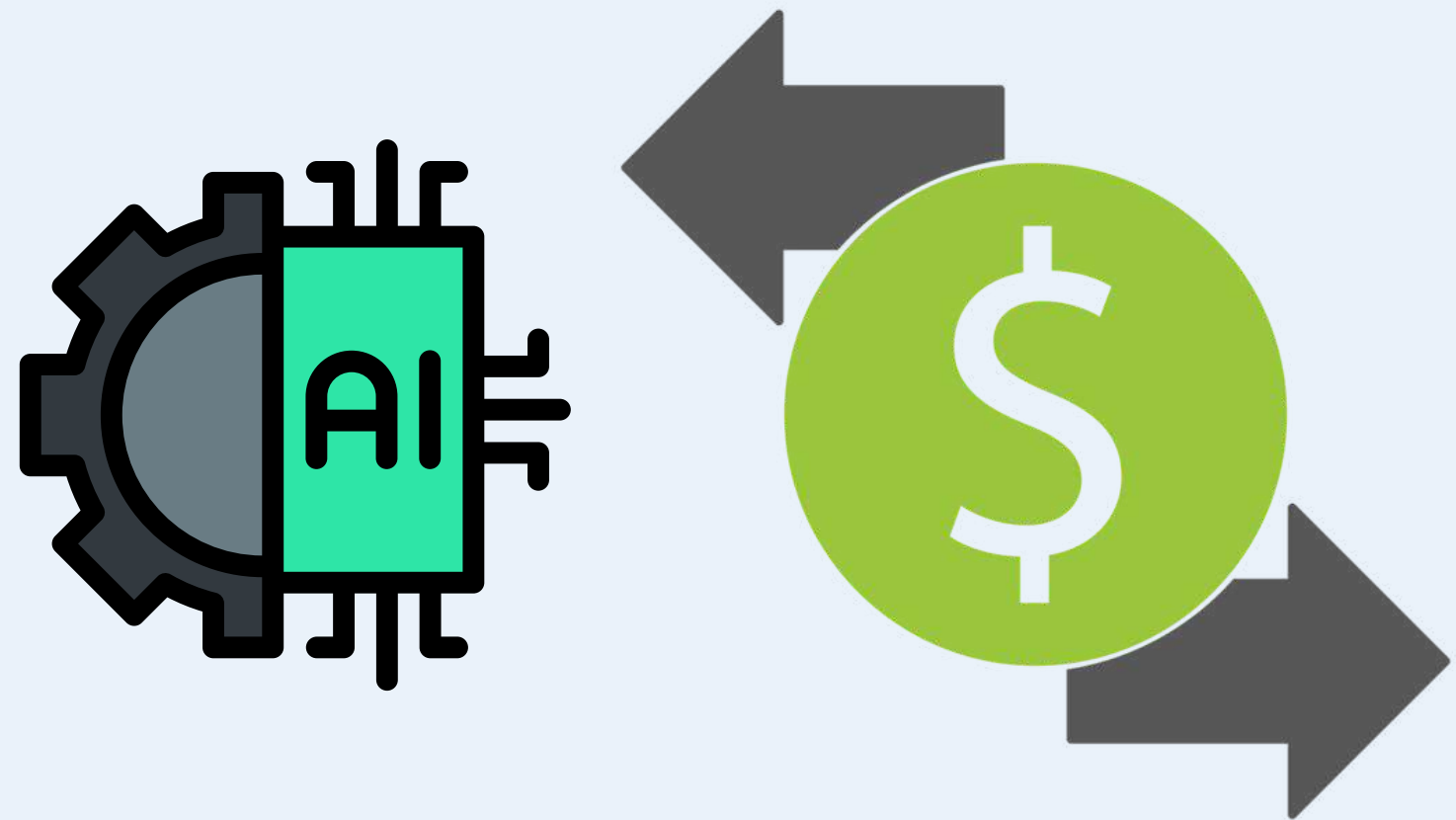- Step 7: Emphasize crucial patterns that require attention.

05

Core Functionalities

Providing commercial prediction model

Building and offering a prediction model for other organizations seeking to use or enhance it, with options for fees or subscriptions.
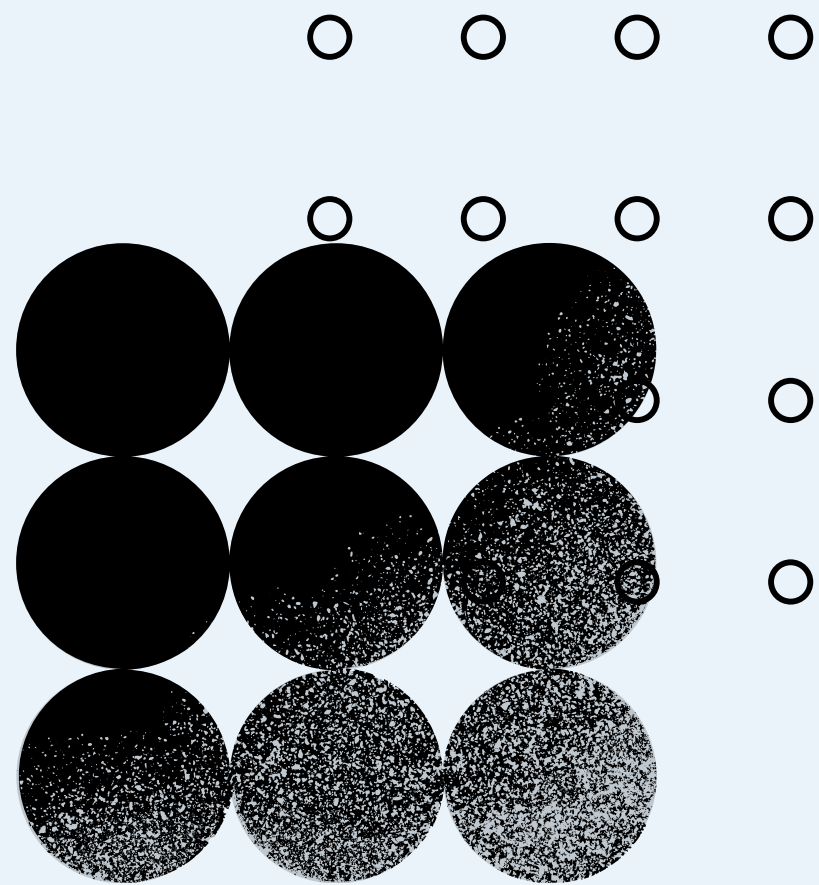
Understand data insight

Provide graphs to demonstrate data visualization after processing

Leaders will gain a comprehensive understanding of the business.

Modify parameter according to business demand

→ Adjust to create a new model based on the original model

→ Export new model for specification purpose

Give advice on business strategy

Apply NLP for analyzing data prediction and expected result

Suggest the optimal strategy to maximize profit.

06

User Interface

Screen 1

- Add Data Button: Clicking this button triggers an Open dialog box, enabling businesses to select and download sales and inventory data.
- Visualize Button: Visualizes previously added data (requires the inclusion of sales and inventory data).
  - Visual Chart Selection Bar:
  - Visualize Sale Data: Revenue.
  - Visualize Sale Data: Sold Quantity.
  - Visualize Inventory: Total Quantity.
  - Visualize Inventory: Total Cost Price and Net Price.
  - Visualize Inventory: Ratio of Cost Price to Net Price.
  - Visualize Inventory: Total Capital Value.
- Add Prompt Button: Allows businesses to upload prompts requesting system improvements. Clicking this button opens a suggestion dialog box for businesses to fill out the prompt parameters as per the suggestion form.
- Sales Forecast Button: Forecasts sales in the upcoming months. Clicking this button transitions to Screen 2.
- Inventory Forecast Button: Forecasts inventory in the upcoming months. Clicking this button transitions to Screen 3.
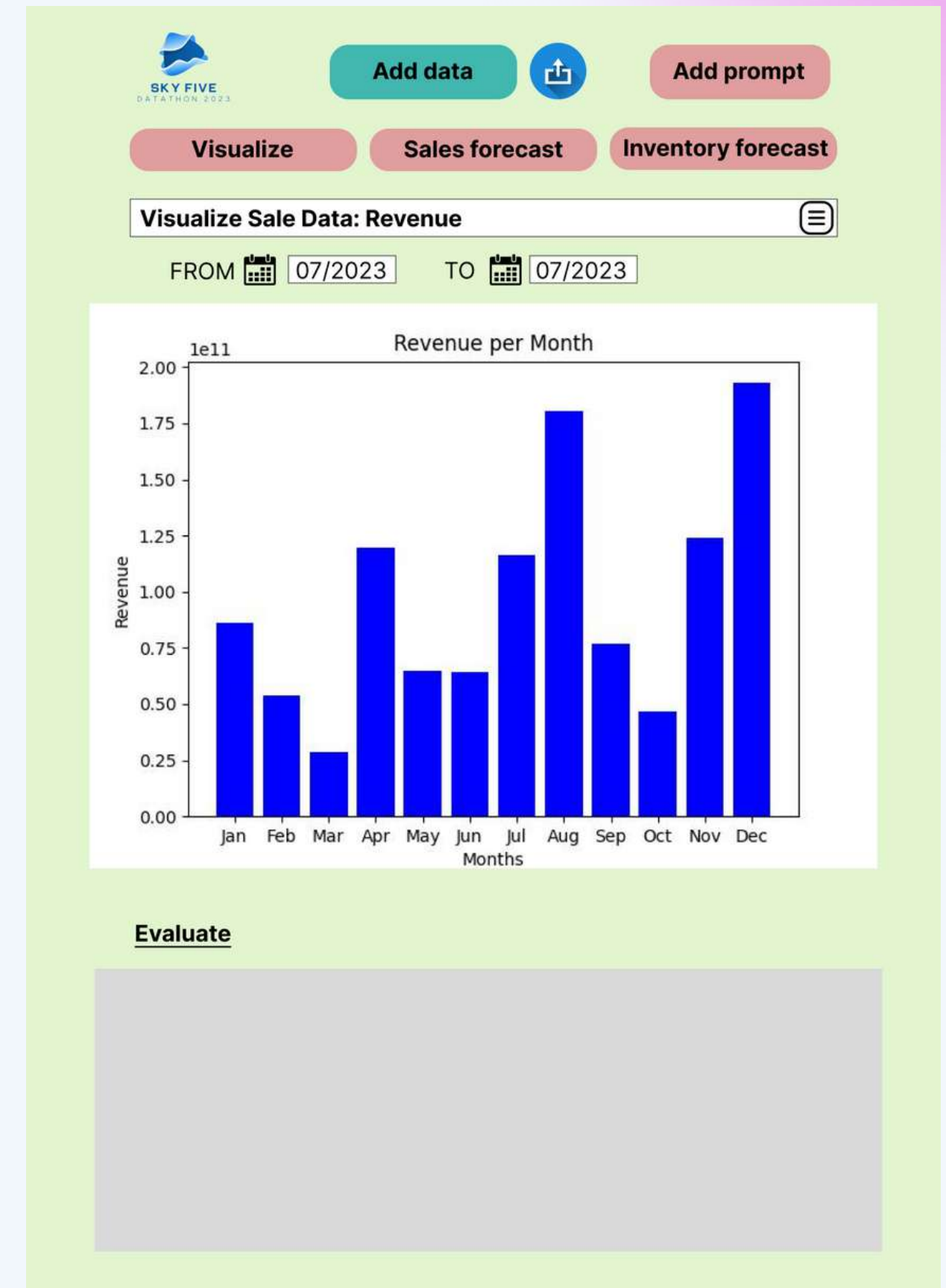- Each visualized chart has an Evaluate section to evaluate each chart.
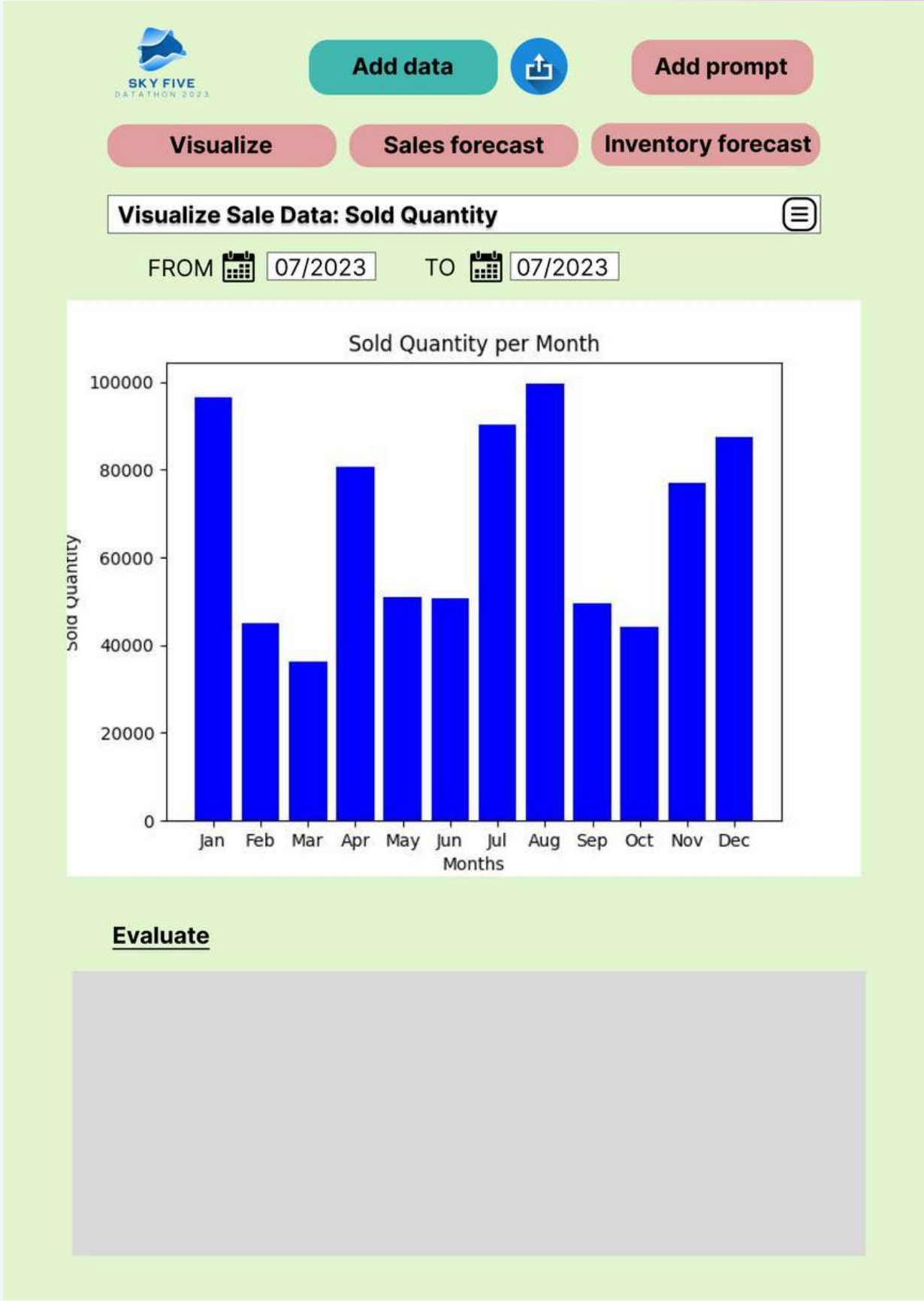
Visualize Sales Data: Revenue.
Enable the selection of a time period (month) and subsequently presents sales revenue based on the chosen months.

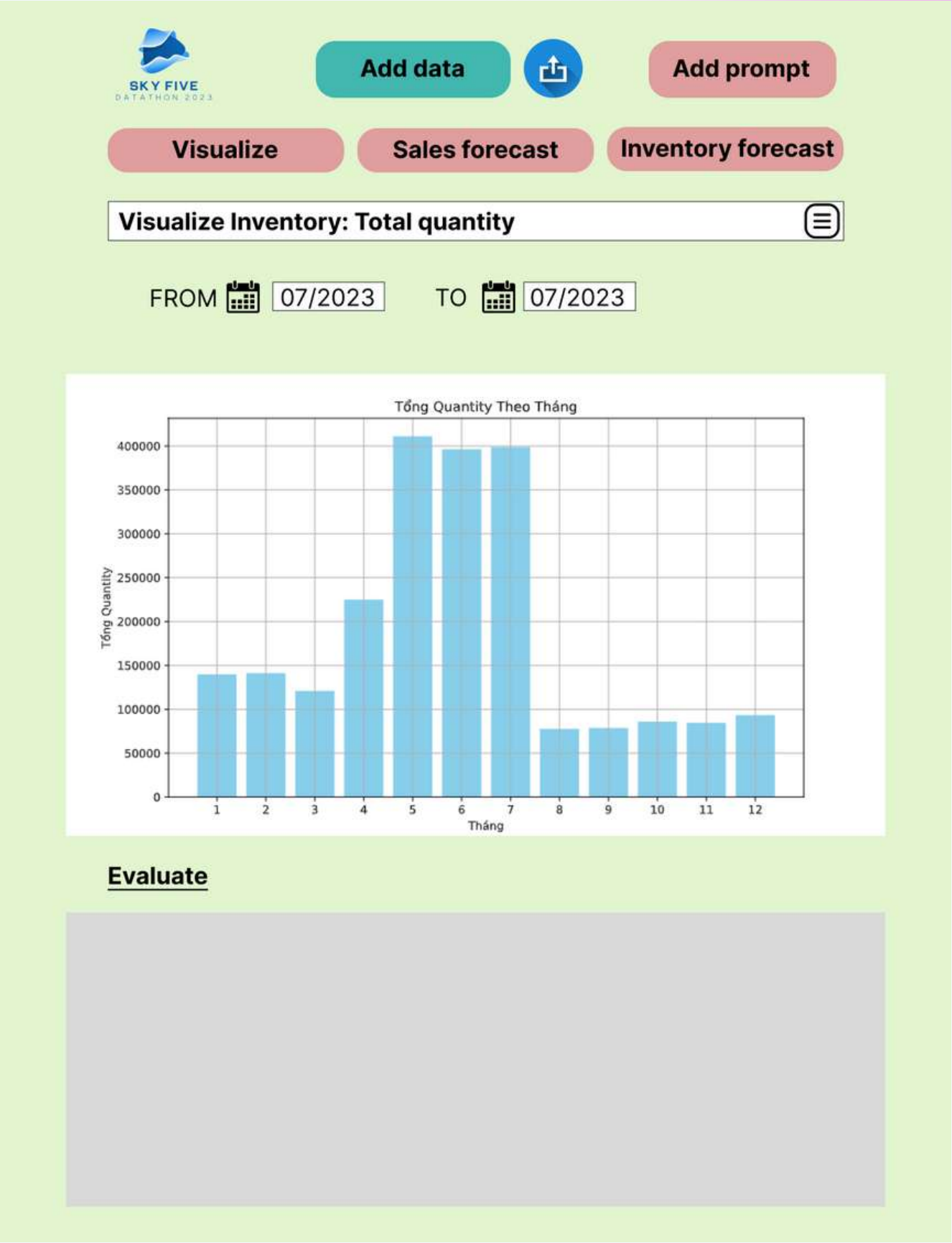$$Revenue = \Sigma \left(Quantity \times NetPrice \right)$$

Visualize Sales Data: Sold Quantity.
Allow the selection of a period (month) and display the total quantity for the chosen period.
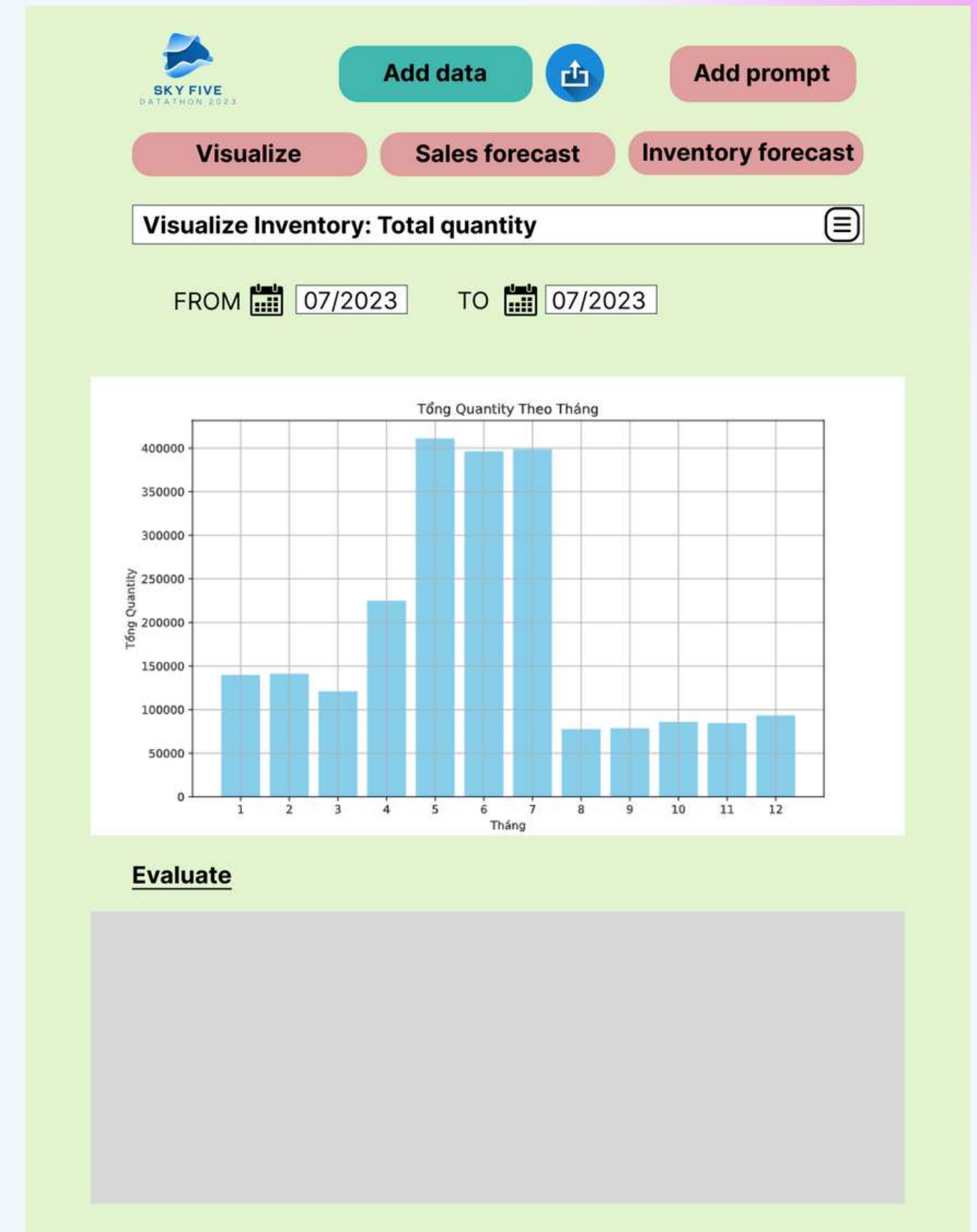
Visualize Inventory: Total Quantity.
Enable the selection of a time period
(month) and display the total quantity for
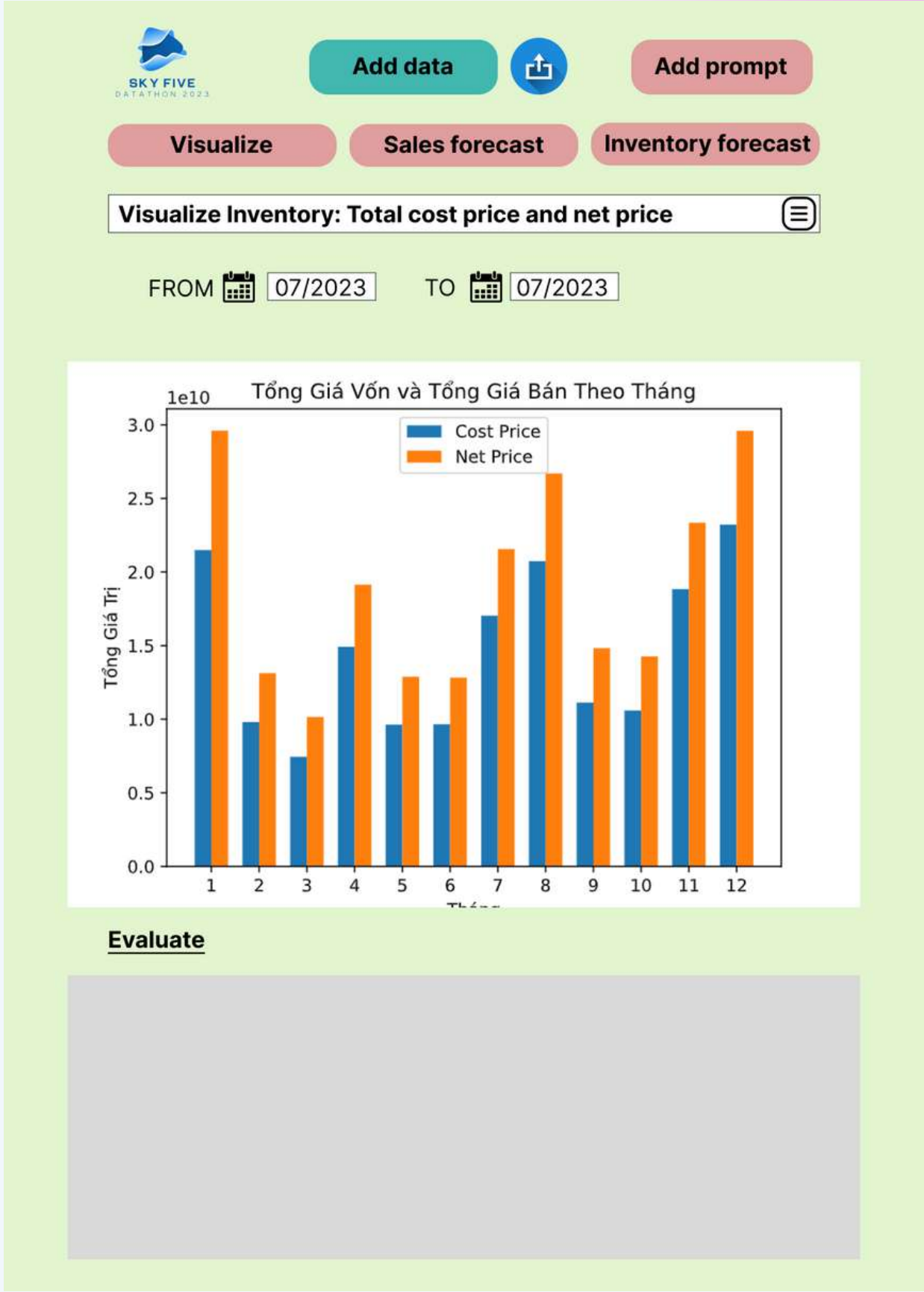the chosen months.

Visualize Inventory: Total Quantity.
Permits the selection of a time period (month) and showcases the total quantity for the chosen months.
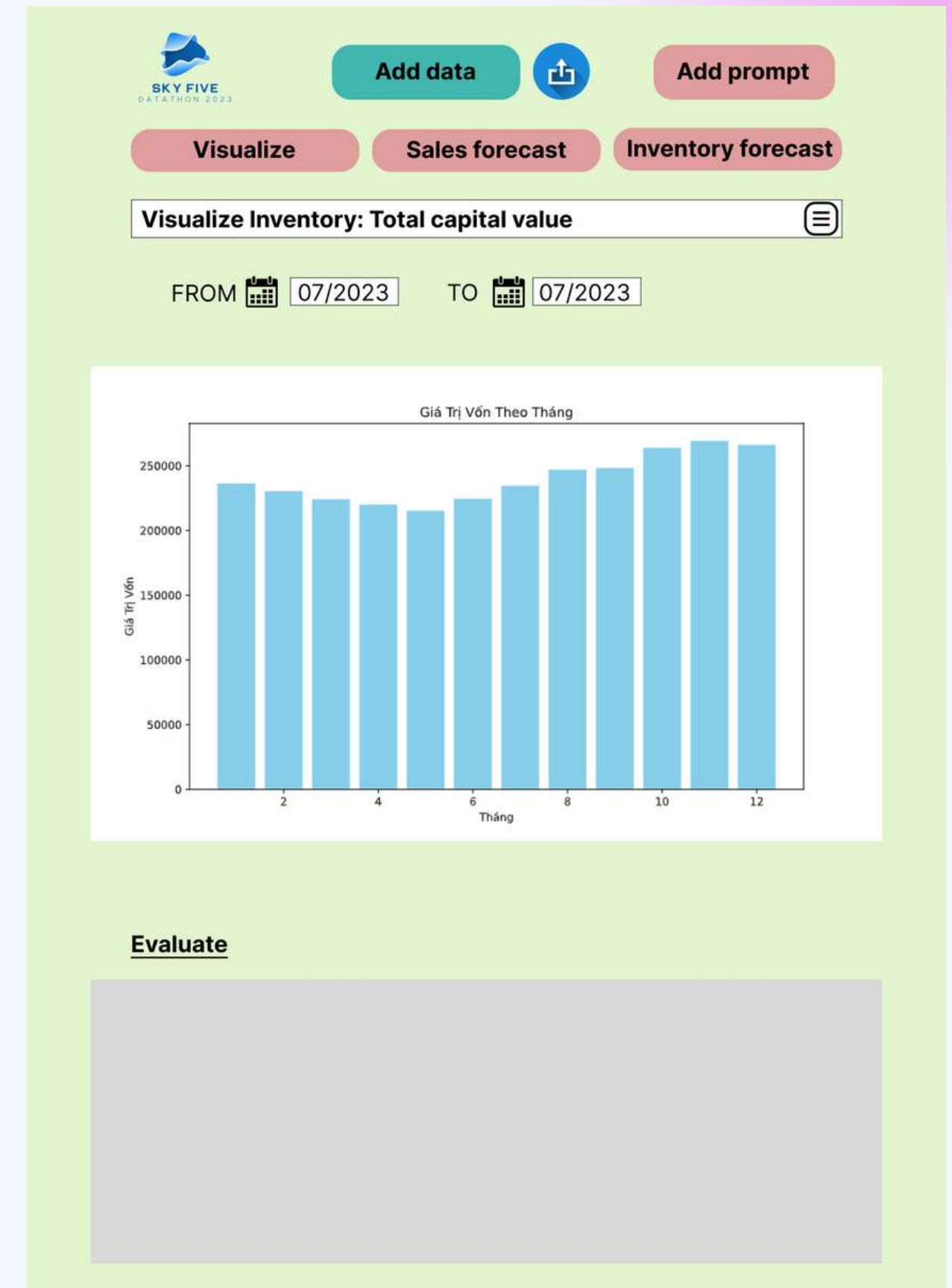
Visualize Inventory: Total Cost Price and Net Price. Enables the selection of a time period (month) and presents the total cost price and total selling price for the chosen months.
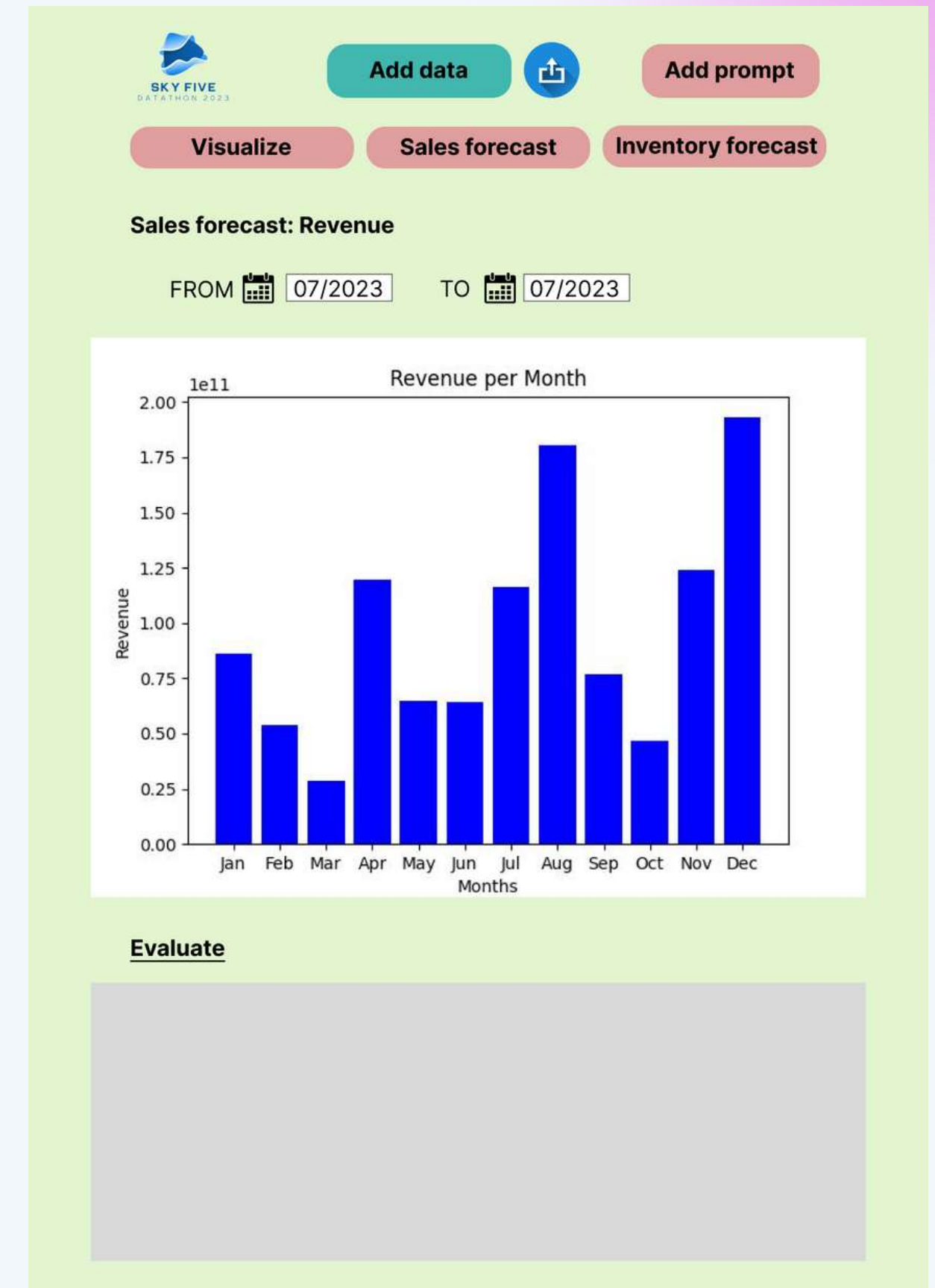
- Visualize Inventory: Total capital value.
- Allow to choose a time period (month), then display the capital value according to the selected months.
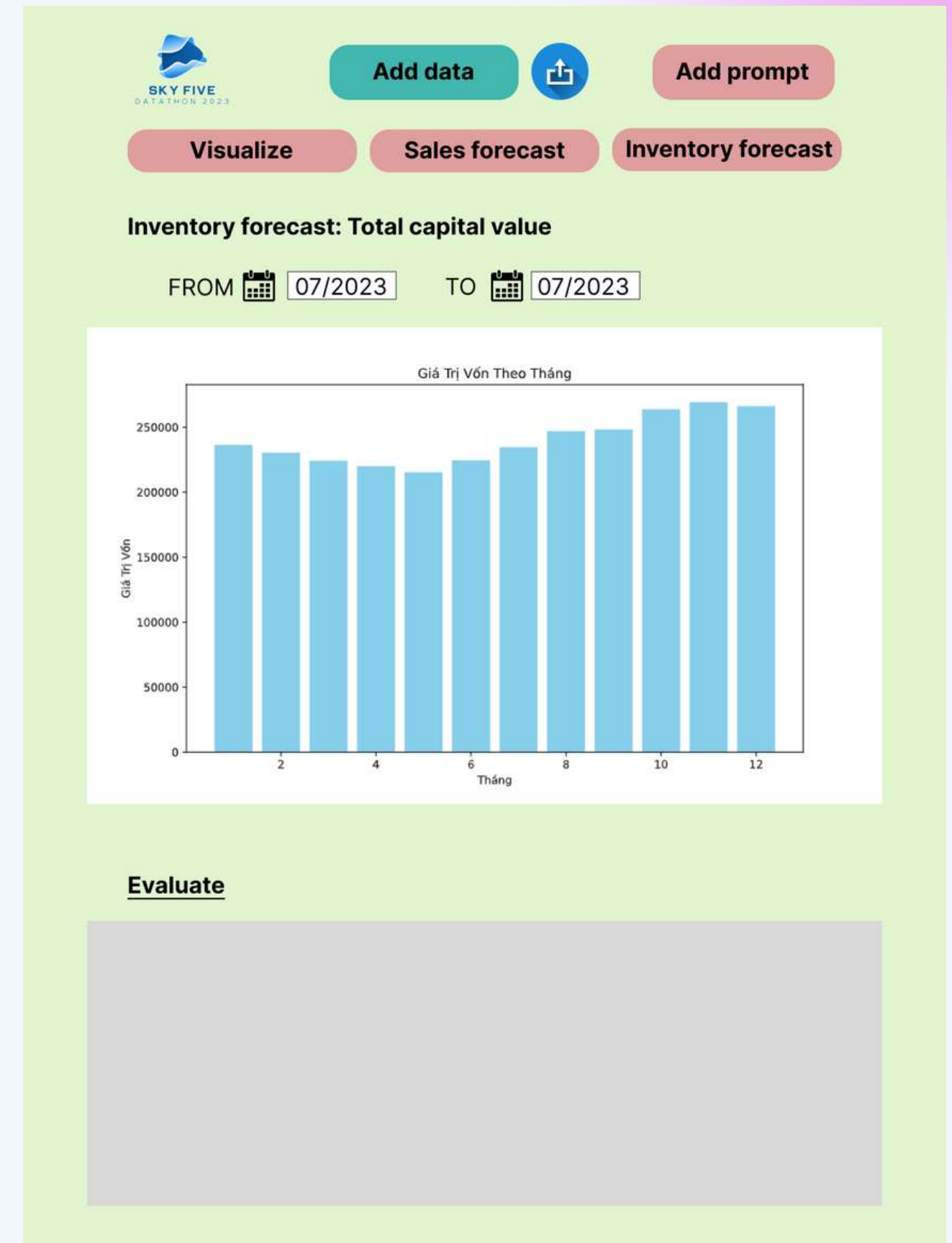
Screen 2

- Sales forecast: Allow users to select a future period (monthly) to forcast sales.
- Evaluate: Evaluate sales for selected months.

Screen 3

- Inventory forecast: Allow users to select a future period (monthly) to forecast the capital value of inventory.
- Evaluate: Evaluate inventory for selected months.

07

Roadmap

Roadmap

Activities	20/11	27/11	4/12	11/12	16/12	17/12
Determine parameters for the prediction model.	■	■				
Connect to SQL DBMS	■	■				
Research the ARIMA model.	■	■				
Search for additional data.		■	■			
Determine a set of rules to adjust parameters.		■	■			
Design the MVP		■	■			
Data preprocessing.			■	■		
Draw visual notices on the chart.			■	■		
Deploy the ARIMA model.			■	■		
Build the MVP (40%).				■	■	
Complete the MVP .					■	■

08

Limitations and Future Potentials

Limitations

- The model is limited by the narrow time interval and the sector coverage of the input data.
- Sets of rules to determine parameters must be kept up-to-date.
- Users need business expertise and domain knowledge to effectively construct input data.

Potentials

- Use ChatGPT's API to identify parameters and features that need to be updated on the fly.
- Build a system to automatically crawl annual data in retail and other sectors.