COIMBATORE INSTITUTE OF TECHNOLOGY

COIMBATORE – 641014

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

INTERNSHIP I REPORT

01.08.2023 to 31.08.2023

SUBMITTED BY

JULIE K

71762108016

AUGUST 2023

TABLE OF CONTENTS

S.NO	DESCRIPTION	PAGE NO
1	**Self-Declaration**	3
2	**Evaluation Form**	4
3	**Offer Letter**	5
4	**Certificate of completion**	6
5	**Work Schedule**	7
6	**Work Done**	8
7	**Outcome of the Internship**	15
8	**Code and Snapshots**	16
9	**References**	19

COIMBATORE INSTITUTE OF TECHNOLOGY

COIMBATORE - 641014

DEPARTMENT OF INFORMATION TECHNOLOGY

SELF DECLARATION

Certified that this Internship in **Effitrac** is being carried out by JULIE K from 01.08.2023 to 31.08.2023.

Student Signature

COIMBATORE INSTITUTE OF TECHNOLOGY

COIMBATORE - 641014

DEPARTMENT OF INFORMATION TECHNOLOGY

EVALUATION FORM

	Evaluation Scheme	Marks
1	Internship Report with Certificate	/40
2.	Presentation of Report	/40
3.	Viva-voce Examination	/20
	Total	/100

Name and Signature of the Three Member Committee

1   Dr. Valliappan Raman

2   Dr. M. Prabhavathy

3   Dr. M. Karthiga

OFFER LETTER

effitrac

Coimbatore,

21st July, 2023.

Dear Ms. Julie K,

Sub: Internship Offer Letter

Effitrac Solutions is pleased to offer you an educational opportunity as **Internship Trainee - DS & ML Engineer**

The internship commences on August 1, 2023, and concludes on August 21, 2023. During this period, you are expected to dedicate your complete professional commitment and time to the Company, and we anticipate that you will carry out your responsibilities with dedication and excellence.

We are excited to welcome you to the Effitrac team as you embark on this promising chapter in your career. We wish you a fruitful and successful internship.

Warm congratulations and a warm welcome to our team!

For Effitrac Solutions India Private Limited

Logesh Velusamy
Chief Executive Officer

CERTIFICATE OF COMPLETION

WORK SCHEDULE

DATE	WORK DONE
01.08.2023 – 07.08.2023	Project initiation, data planning, initial exploration, and communication setup.
08.08.2023 – 14.08.2023	Intensive data collection, data preprocessing, and initial analysis commenced, laying the groundwork for deeper insights.
15.08.2023 – 21.08.2023	In-depth data analysis, including trend identification, scenario analysis, and synthesis of initial insights, progressed, shaping actionable recommendations.
22.08.2023 – 31.08.2023	Final analysis, report compilation, visualization and project conclusion wrapped up the comprehensive trend analysis project.

WORK DONE

TREND ANALYSIS ON AMAZON SALES DATA

Week 1: Data Conversion, Preliminary Exploration, and Dataset Overview

During the first week of the project, the primary focus was on converting the provided dataset from PostgreSQL to CSV format, conducting preliminary exploration to understand its structure and content, and providing an overview of the dataset. The following activities were undertaken:

1. Dataset Overview: Effitrac organization provided a dataset containing Amazon sales data stored in a PostgreSQL database. The dataset consisted of approximately 8,51,896 records encompassing various aspects of sales transactions, including order details, product information, customer demographics, and transactional records.

2. Data Extraction: SQL queries were utilized to extract the dataset from the PostgreSQL database, ensuring data integrity and accuracy during the extraction process. Careful consideration was given to retrieve the relevant tables and columns required for the trend analysis.

3. Data Conversion to CSV: The extracted dataset was converted from PostgreSQL format to CSV format using Python scripts. This conversion facilitated ease of manipulation and analysis, allowing for seamless integration with analytical tools and techniques.

4. Preliminary Data Exploration: A preliminary exploration of the converted CSV dataset was conducted to gain insights into its structure, format, and content. Sample data records were examined, and column headers were reviewed.

5. Dataset Overview Documentation: Detailed documentation was maintained to provide an overview of the dataset, including the number of records (approximately

8,51,896), key data fields, and data types. This documentation served as a reference for understanding the dataset's characteristics and scope.

6. Communication with Stakeholders: Regular communication was established with Effitrac organization to ensure alignment on data requirements, conversion process, and project timelines. Any discrepancies or issues identified during the dataset overview and exploration phase were discussed and resolved collaboratively.

By the end of the first week, the dataset containing approximately 8,51,896 records of Amazon sales data had been successfully converted from PostgreSQL to CSV format. Preliminary exploration provided insights into the dataset's structure and content, laying the groundwork for further analysis and interpretation in the subsequent weeks of the project.

Week 2: Data Preprocessing and Initial Analysis

In the second week of the project, the focus shifted towards preparing the collected data for analysis and conducting initial exploratory analysis. The following activities were undertaken:

1. Data Cleaning and Preprocessing: The collected datasets underwent thorough cleaning and preprocessing to address issues such as missing values, duplicates, outliers, and inconsistencies.

2. Feature Extraction: Relevant features for trend analysis were extracted from the cleaned datasets. This included variables such as purchase date, product category, sales volume, revenue, customer demographics, and geographical location.

3. Descriptive Analysis: Basic descriptive statistics and visualizations were generated to gain insights into the distribution, variability, and patterns present in the data. Summary statistics, frequency distributions, histograms, and scatter plots were utilized to explore key metrics such as sales trends over time, product performance, and customer behavior.

4. Identification of Data Challenges: Challenges or limitations encountered during the data preprocessing and exploratory analysis phase were identified and documented. This included issues related to data quality, completeness, consistency, and granularity. Strategies for addressing these challenges were proposed, such as data imputation, outlier detection, and data enrichment.

5. Documentation and Reporting: Detailed documentation was maintained throughout the data preprocessing and initial analysis phase to capture the steps taken, assumptions made, and insights gained. This documentation served as a reference for future analysis iterations and was compiled into a preliminary report summarizing the key findings and observations.

By the end of the second week, we completed the initial data preprocessing and exploratory analysis, laying the groundwork for more advanced analytical techniques in the subsequent phases of the project.

Week 3: Case Study on Advanced Analysis and Model Development

During the third week of the project, efforts were focused on conducting more advanced analysis and developing analytical models to extract actionable insights from the Amazon sales data. The following activities were undertaken:

1. Advanced Analytical Techniques: Various advanced analytical techniques are explored to uncover hidden patterns, trends, and relationships within the sales data. This included time series analysis, regression modeling, clustering, segmentation, and predictive modeling techniques such as machine learning algorithms.

2. Time Series Analysis: Time series analysis techniques are applied to analyze the temporal patterns and trends present in the sales data. This involved decomposing the time series into its components (trend, seasonality, and residual) and performing trend detection, seasonal adjustment, and forecasting using methods such as moving averages, exponential smoothing, and ARIMA modeling.

3. Regression Modeling: Regression models are developed to identify the factors driving sales performance and predict future sales trends. Multiple regression, logistic regression, and other regression techniques were applied to analyze the relationships between sales and various predictor variables such as marketing spend, pricing, product attributes, and external factors such as seasonality and economic indicators.

4. Machine Learning Algorithms: Advanced machine learning algorithms such as random forests, gradient boosting machines, and neural networks were trained on the Amazon sales data to perform predictive analytics and generate insights. Supervised and unsupervised learning techniques were applied to identify customer segments, recommend products, optimize pricing strategies, and personalize marketing campaigns.

5. Model Evaluation and Validation: The developed analytical model was rigorously evaluated and validated using appropriate performance metrics, cross-validation techniques, and hypothesis testing. Model accuracy, precision, recall, F1-score, and other evaluation criteria were assessed to ensure the reliability and robustness of the models.

6. Model Interpretation and Visualization: The results generated by the analytical models were interpreted and visualized to facilitate understanding and decision-making. Interactive dashboards, heatmaps, scatter plots, and other visualization techniques were utilized to communicate the insights gained from the analysis effectively.

By the end of the third week, we completed the development and validation of advanced analytical model, enabling the extraction of actionable insights from the Amazon sales data to inform strategic decision-making.

Week 4: Model Refinement and Report Compilation

In the final week of the project, efforts were directed towards refining the analytical models and compiling the findings into a comprehensive report. The following activities were undertaken:

1. Model Refinement: The analytical model developed in the previous phase was refined based on feedback from stakeholders, validation results, and additional data insights. Model parameters were fine-tuned and optimizations were made to improve performance and reliability.

2. Scenario Analysis: Sensitivity analysis and scenario planning were conducted to assess the impact of different variables and assumptions on sales performance and profitability. "What-if" scenarios were explored to evaluate the potential outcomes of alternative strategies and decisions.

3. Insights Synthesis: The findings and insights generated from the analysis were synthesized into actionable recommendations for key stakeholders. Strategic recommendations were formulated to optimize sales strategies, enhance customer engagement, and drive business growth.

4. Report Compilation: A comprehensive report summarizing the entire trend analysis project was compiled(this report), incorporating all the key findings, insights, methodologies, and recommendations. The report included sections on project background, objectives, methodology, data analysis, results, discussion, conclusions.

5. Visualization and Communication: The insights and recommendations were communicated to stakeholders through engaging visualizations, presentations, and interactive dashboards.

6. Feedback and Iteration: The final report was shared with effitrac organisation for review and feedback. Comments and suggestions was incorporated into the report, and iterative revisions were made as needed to ensure clarity, accuracy, and relevance.

By the end of the fourth week, we successfully completed the trend analysis on Amazon sales data and delivered a comprehensive report outlining the findings to drive strategic decision-making and business growth.

The 4-week trend analysis project on Amazon sales data involved a systematic and iterative approach to data collection, preprocessing, analysis, and reporting. By leveraging advanced analytical techniques and methodologies, valuable insights were generated to inform strategic decision-making and drive business growth. The comprehensive report compiled at the end of the project served as a valuable resource, providing actionable recommendations to optimize sales strategies, enhance customer engagement, and maximize profitability in the competitive e-commerce landscape.

SOME OF THE COMMON ERRORS ENCOUNTERED

- ❖ Data Quality Issues

  - Missing Values

  - Inconsistent Data Formats

  - Duplicate Records

- ❖ Analysis Challenges

  - Outliers

  - Overfitting

  - Sampling Bias

- ❖ Technical Challenges

  - Software Compatibility Issues

  - Memory and Computational Constraints

- ❖ Interpretation

  - Misinterpretation of Results
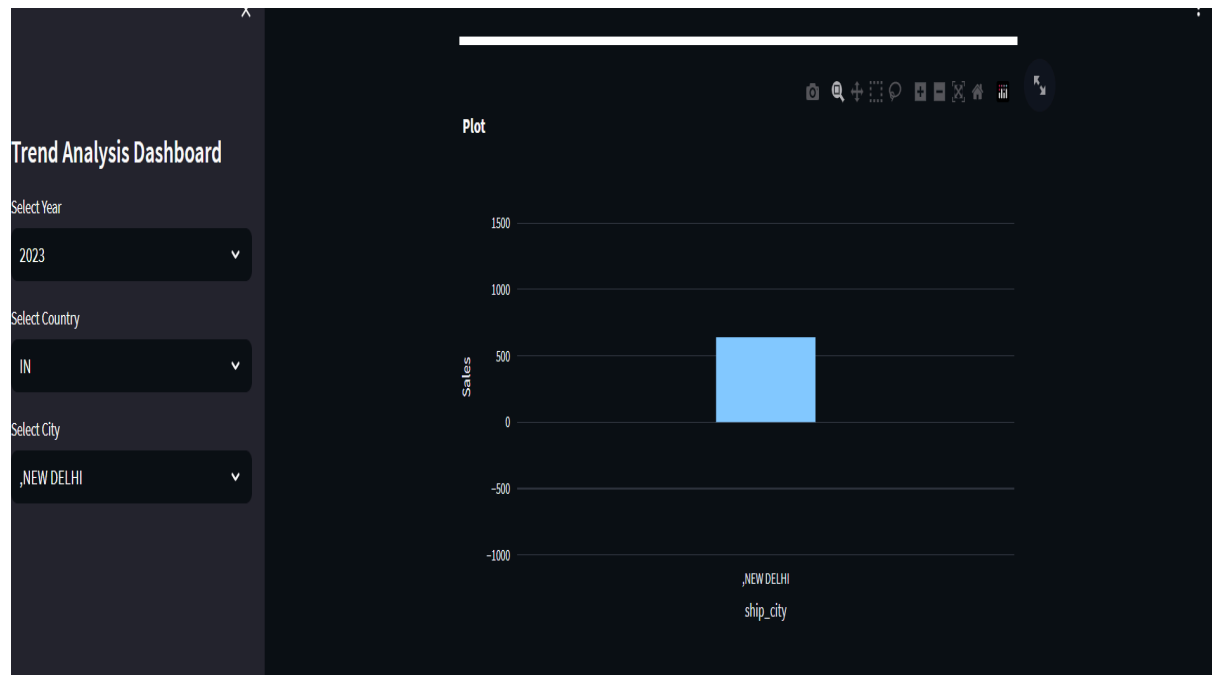
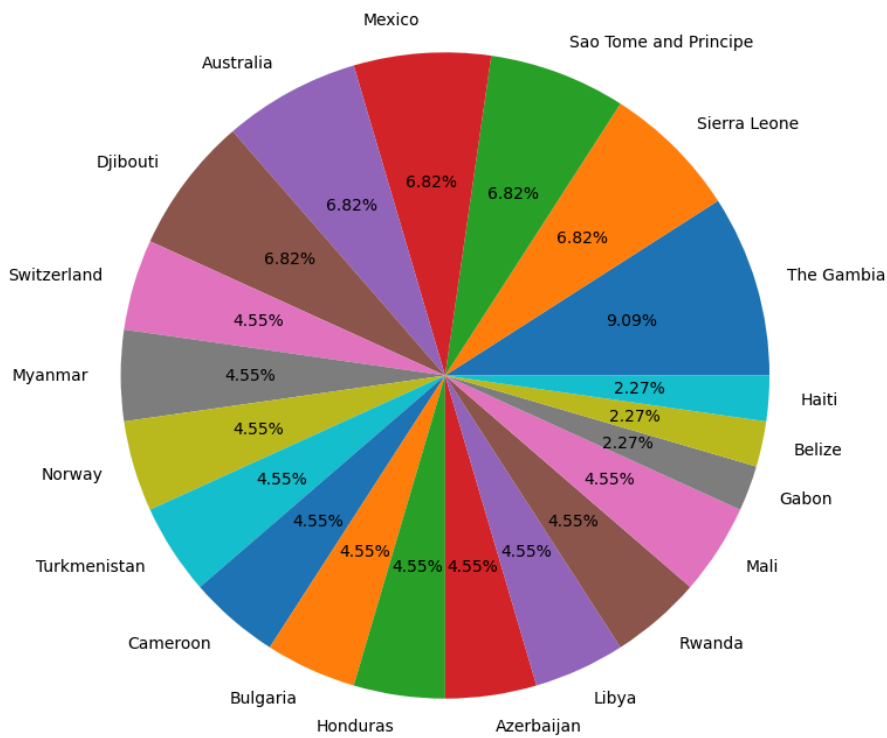| Input	Excepted Output	Actual Output	Result
Amazon sales data	Patterns and trends in the data of a specific year, country and city.	Visualization of trends and patterns of specific year, country and city.	Pass

OUTCOMES OF THE INTERNSHIP

The key outcomes are

- Data Handling Proficiency

- Advanced Analytical Skills Development

- Problem-Solving Abilities Demonstrated

- Technical Competence Exhibited

- Effective Communication Enhanced

- Practical Application of Learned Concepts

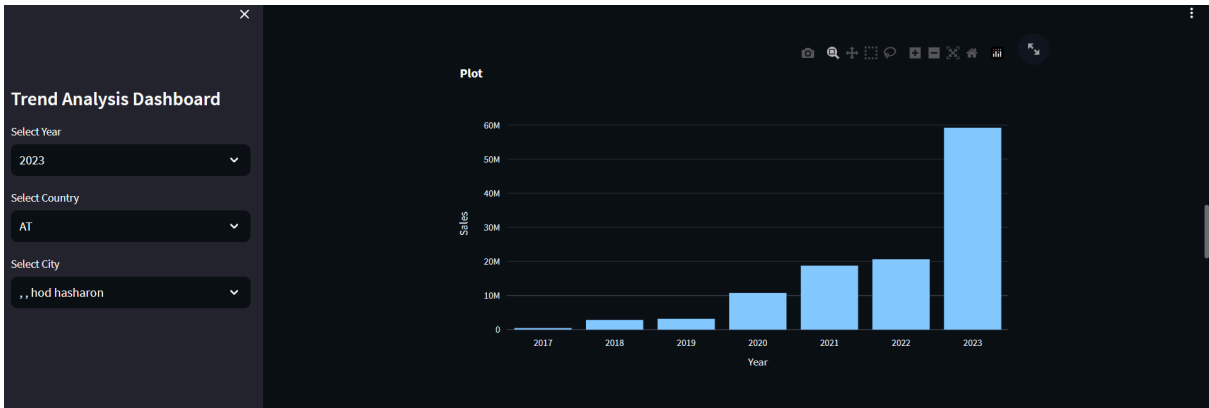- Commitment to Continuous Learning
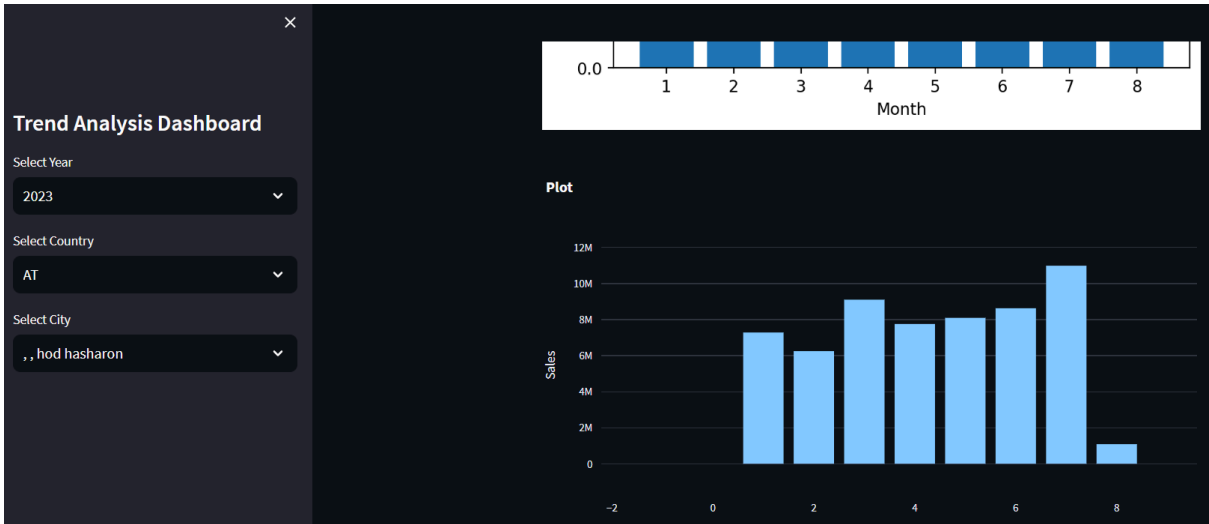
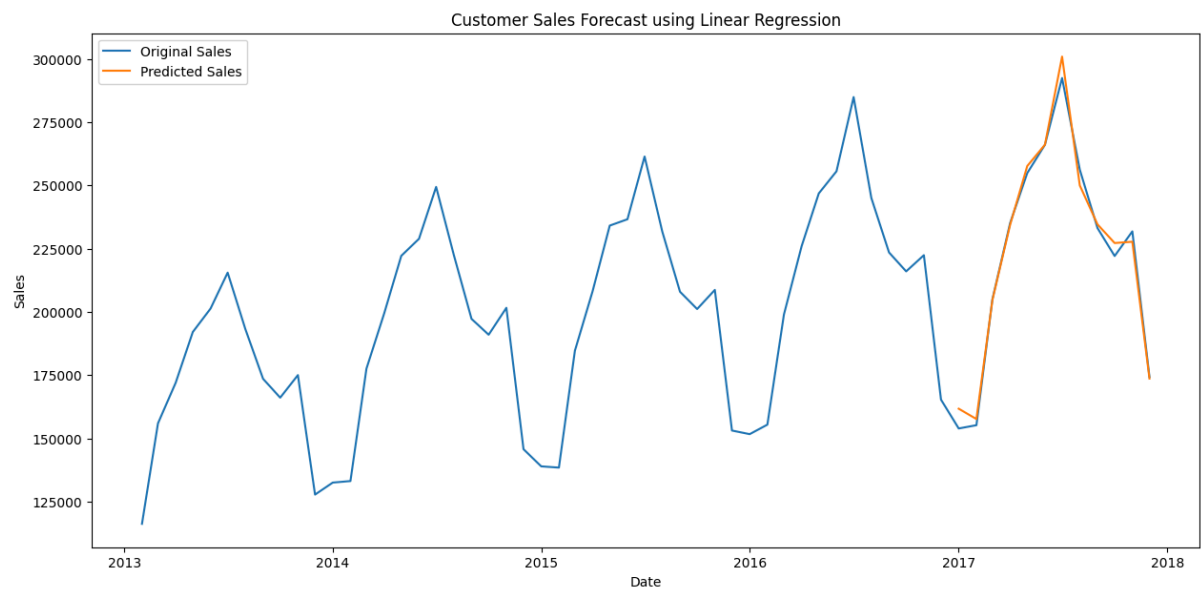- Successful Project Completion

Code and Snapshots

Pie chart of sales in top 20 countries:

Boxplot of total profit:
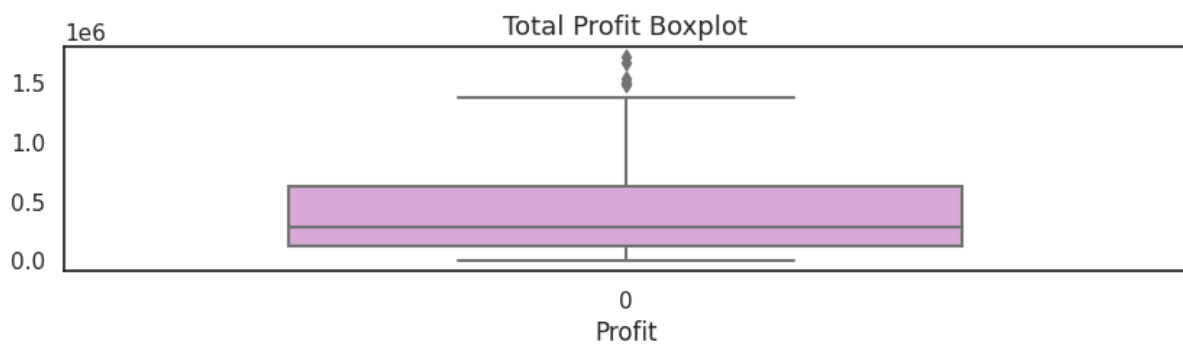
Linear regression of sales forecast:

Customer Sales Forecast using Linear Regression

Trend Analysis Dashboard

Select Year
2023

Select Country
AT

Select City
, , hod hasharon

Plot

Trend Analysis Dashboard

Select Year
2023

Select Country
IN

Select City
,NEW DELHI

Plot

Trend Analysis Dashboard

Select Year
2023

Select Country
IN

AT
AU
BB
BE
BG
BH
BL

Plot
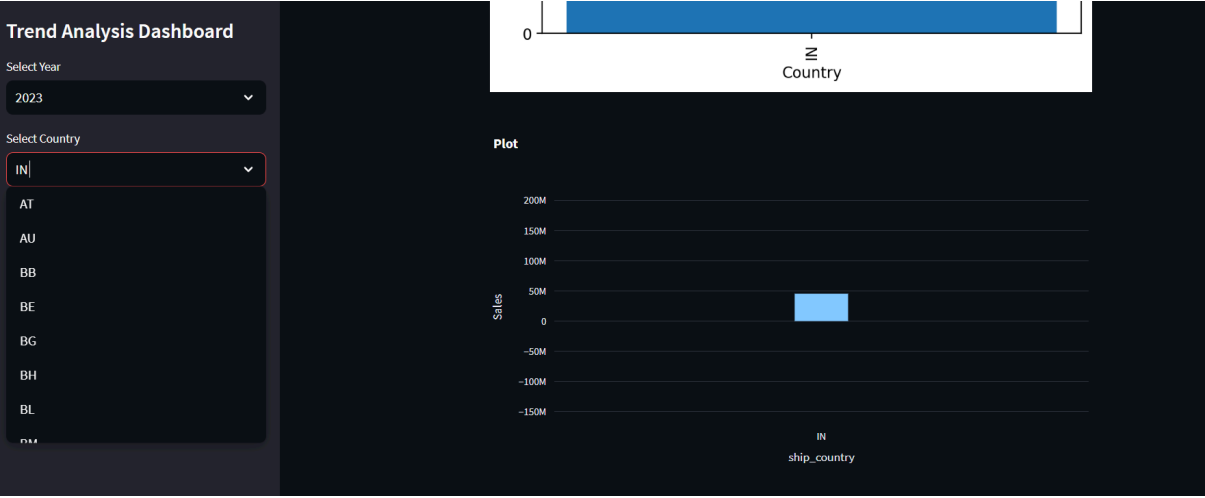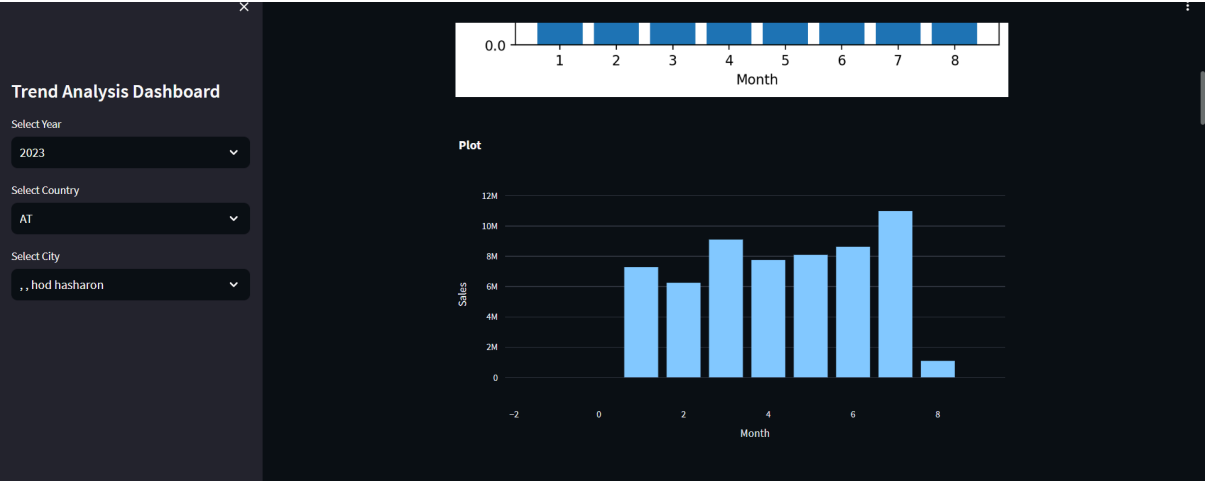
REFERENCES

- [What is Trend Analysis? Definition, Steps, Examples, Benefits and Best Practices (ideascale.com)](#)

- ▶ Sales Data Analysis using Python