

INTERNSHIP PROJECT REPORT

RETAIL DATA ANALYSIS

Project head,

Ashok G Nair

Director-II-Enterprise Solutions (UST)

Submitted by,

Kavya Boban

UID:223811

DECLARATION

I hereby declare that the report of internship entitled ”**RETAIL DATA ANALYSIS**” submitted to UST Global, is uniquely prepared after the completion of four months internship under the supervision and guidance of Mr.Ashok G Nair, Project Head Director-II-Enterprise Solutions, UST Global Trivandrum.

I also confirm that the report is prepared for my academic requirement and not for any other purpose. It might not be used in the interest of the opposite party of the cooperation.

Date :11-08-2022

Place: Angamaly

KAVYA BOBAN

ACKNOWLEDGEMENT

Gratitude is a feeling which is more eloquent than words, more silent than silence. The internship opportunity I had with UST Global Trivandrum was a great chance for learning and professional development. Therefore, I consider myself a very lucky individual as I was provided with an opportunity to be a part of it. I am also grateful for having a chance to meet so many wonderful people and professionals who led me through this internship period.

Bearing in mind previously I am using this opportunity to express my deepest gratitude and special thanks to **Mr. Ashok G Nair**, Director-II-Enterprise Solutions UST Global who in spite of being extraordinarily busy with his duties, took time out to hear, guide, and keep me on the correct path and allow me to carry out my project his guidance at their esteemed organization and extending during the training.

I express my deepest thanks to **Ms. Gopika Bindu**, Associate II UST Global for taking part in useful decisions and giving necessary advice and guidance and arranging all facilities to make internship easier. I choose this moment to acknowledge her contribution gratefully.

I express my heartfelt gratitude to **Ms. Minu Sebastian** for being a well-wisher and mentor throughout the internship.

I Would like to thank all the people who worked with me **Dennis Mathew** and **Zanha Jamal** for their patience and help throughout this internship.

I perceive this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain my desired career objectives. Hope to continue cooperation with all of you in the future

Sincerely,

KAVYA BOBAN

ABSTRACT

Retail Data Analytics is used nowadays by shops in order to better predict the number of products, that might get sold and therefore to better estimate how much product should be produced. This is very important because the amount of sold products can vary mainly during the year. For example, people tend to buy more things before a festive season than during a normal, non-holiday week. So the revenue earned by the shop can increase if there is a good estimate about the seasons that can give a profit and plan accordingly. If a shop has too few products before the festive season, it will lose potential income. But if a shop has too many products, too much storage would be required and storage also costs money, so the company would again lose money. The retail data analysis can therefore be used in order to try to optimize the production of products, such that there is always an optimal amount available.

The proposed project is to predict the department-wide weekly sales for a store. This should then help to optimize the manufacturing process and therefore to increase income while lowering costs. It should be possible to feed in past sales data from a department and get the predicted weekly sales and provide recommended actions based on the insights and visualize the data for better understanding.

Contents

1	INTRODUCTION	5
2	PROOF OF CONCEPT	6
3	IMPLEMENTATION	7
3.1	Technology Used	7
3.2	Algorithms Used	8
3.3	Dataset	8
3.4	Modules	9
4	RESULT ANALYSIS	10
4.1	Results	10
4.2	Accuracy	11
5	CONCLUSION	12
5.1	Conclusion	12
6	APPENDIX	13
6.1	Screen Shots	13

Chapter 1

INTRODUCTION

People tend to buy more things before a festive season than during a normal, non-holiday week. So the revenue earned by the shop can increase if there is a good estimate about the seasons that can give a profit and plan accordingly.

If a shop has too few products before the festive season, it will lose potential income. But if a shop has too many products, too much storage would be required and storage also costs money, so the company would again lose money. Therefore the analysis is required for the smooth functioning of the shop without loss.

Retail Data Analytics is used nowadays by shops in order to better predict the number of products, that might get sold and therefore to better estimate how much product should be produced. This is very important because the amount of sold products can vary during the year.

Chapter 2

PROOF OF CONCEPT

The proposed system can therefore be used in order to optimize the production of products, such that there is always an optimal amount available. The proposed system of Retail data analysis and prediction of weekly sales can be used nowadays by shops in order to better predict the number of products, that might get sold and therefore to better estimate how much product should be produced.

In this project, with the help of a dataset that contains a huge amount of data was used in order to reach our goal, and also different kinds of visualization charts and graphs have been used in order to plot and get the best insight of the dataset, also used four different algorithms and found the one with the minimal error rate to predict the weekly sales of the data.

For the development of the front end the designing phase of the UI was done using Figma and later on developed the website using HTML, CSS, and javascript and the backend using flask and Jupiter notebook using the language python.

With the development of the system, the organization can easily predict the weekly sales for a store and visualize the data for better insight. This should then help to optimize the manufacturing process and thereby help to increase income while lowering costs.

Chapter 3

IMPLEMENTATION

3.1 Technology Used

Jupyter Notebook

Jupyter notebook is used for prediction using the algorithm implemented with the help of data science and machine learning, using the language python.

python

Python is the programming language used.

Visual Studio code

Visual Studio Code is used to link the app.py file with all other HTML and CSS and create a website with the help of a flask.

Flask (Python Framework)

Flask is a web framework. This means flask provides you with tools, libraries, and technologies that allow you to build a web application.

Figma

Figma is used for the designing of webpages.

3.2 Algorithms Used

Linear Regressor

Decision Tree Regressor

Random Forest

XGBoost Regressor

3.3 Dataset

The dataset from Kaggle is used for this project. The dataset contains historical sales data from 45 stores located in different regions. Each store is further divided into departments. The data itself is stored in an excel sheet. The excel sheet contains three tabs. The first tab contains the data from the stores. The second contains the features and the third contains the sales data.

1) Stores: There is data of 45 stores in total. Every store has its own type and size, which is also included in the excel sheet. The information contained in the excel sheet is anonymized. Table I contains the statistics on the size of the stores. In total there are three different types of stores (A, B, and C)

2) Features: The features are related to a store. Table II contains all available features and a short description of each one, while table III contains the statistics for some of the features. The data for Markdown1 - Markdown5 is very incomplete and has to be dropped or different methods for handling missing data have to be applied.

3) Sales: Each store also has historical sales data stored in the dataset. The sales data was collected from the fifth of February 2010 until the first of November 2012. Table IV contains all features related to the sales data and table V contains some statistics on the weekly sales.

3.4 Modules

1. Data collection
2. Data visualization
3. Data preprocessing
4. Data transformation
5. Model Training
6. Model Evaluation and Deployment

Chapter 4

RESULT ANALYSIS

4.1 Results

The result of the proposed project Retail Data Analysis using machine learning lies in developing a handy web app that can be successfully used by an admin or an employee of the organization to get an insight into the data that's been collected.

The visualization of data helps in better understanding of the dataset and helps in improving the profit of the organization by assigning the right amount of stock by calculating the weekly sales. The user of this application can save his or her time and also bring out the best.

	rmse	score	model
name			
LinearRegression	39.347376	0.486628	LinearRegression()
DecisionTreeRegressor	13.833565	0.936545	DecisionTreeRegressor()
RandomForestRegressor	10.985117	0.959986	(DecisionTreeRegressor(max_features='auto', ra...
XGBRegressor	7.727838	0.980198	XGBRegressor(base_score=0.5, booster='gbtree',...

Figure 4.1: rsme of the algorithms used
here in this project I used four different algorithms and found the error rate and its percentage and I am selecting the xgboost which has the least error rate to find the accuracy and train the model.

4.2 Accuracy

```
: model = xgb.XGBRegressor(objective = 'reg:squarederror', learning_rate = 0.2, max_depth = 10, n_estimators = 100)
  model.fit(X_train, y_train)

# make predictions on the test data

y_predict = model.predict(X_test)

result = model.score(X_test, y_test)

print("Accuracy : {}".format(result))

Accuracy : 0.9788490672060531
```

Figure 4.2: Accuracy of XGBoost

Chapter 5

CONCLUSION

5.1 Conclusion

This project is a general analysis of the retail data of 45 stores and their sales. It's a machine learning project that is done with four different algorithms and one with maximum accuracy and the least error rate is taken. this proposed system helps in understanding and getting an insight into the sales that are happening in these stores and predicts the weekly sales which will help in the stocks that are to be taken for that store. This proposed system helps the organization to save time and cost as well.

In an addition to the current project, the website is being further modified using the react.js, and making the login of the website is being modified using the google-sign-in method and WhatsApp login.

Chapter 6

APPENDIX

6.1 Screen Shots

Here I add some sample screenshots of our recommendation system which includes,

- Data Visualization
- Home/Login screen
- Admin - Dashboard
- Admin - Employee Details
- Admin - Customer Details
- Admin - Charts
- Admin - Stock

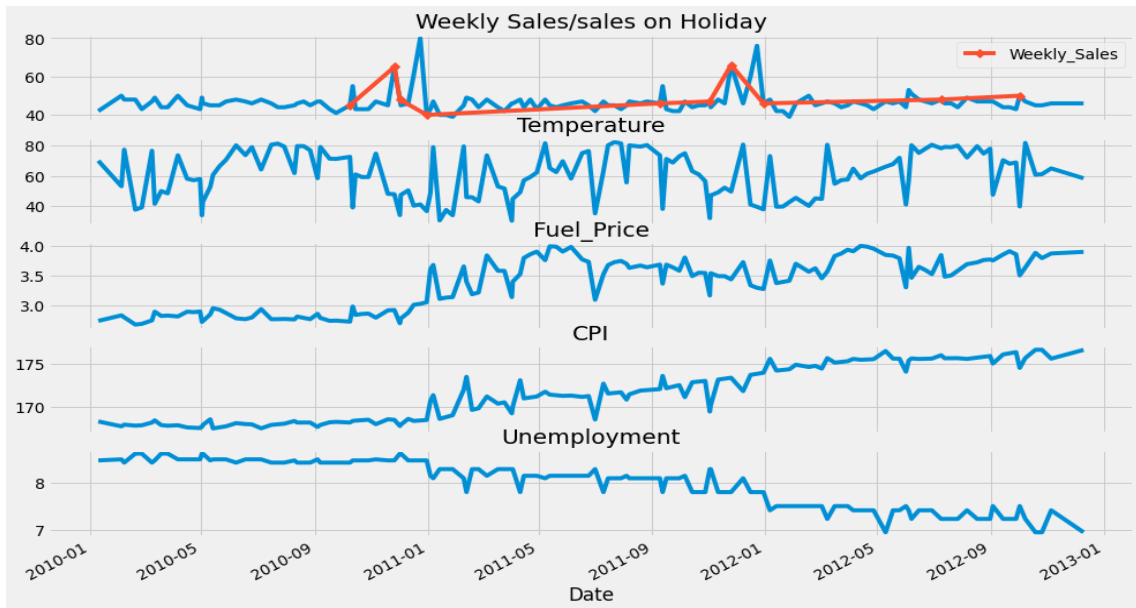


Figure 6.1: Visualization

Nov- Dec shows a spike in Weekly Sales, but over the year it is not increased.
 weeks nearby holiday shows peak
 Fuel Price and Consumer Price Index showed growth over the year.
 Unemployment decreased year after year.
 Temperature is showing a random walk

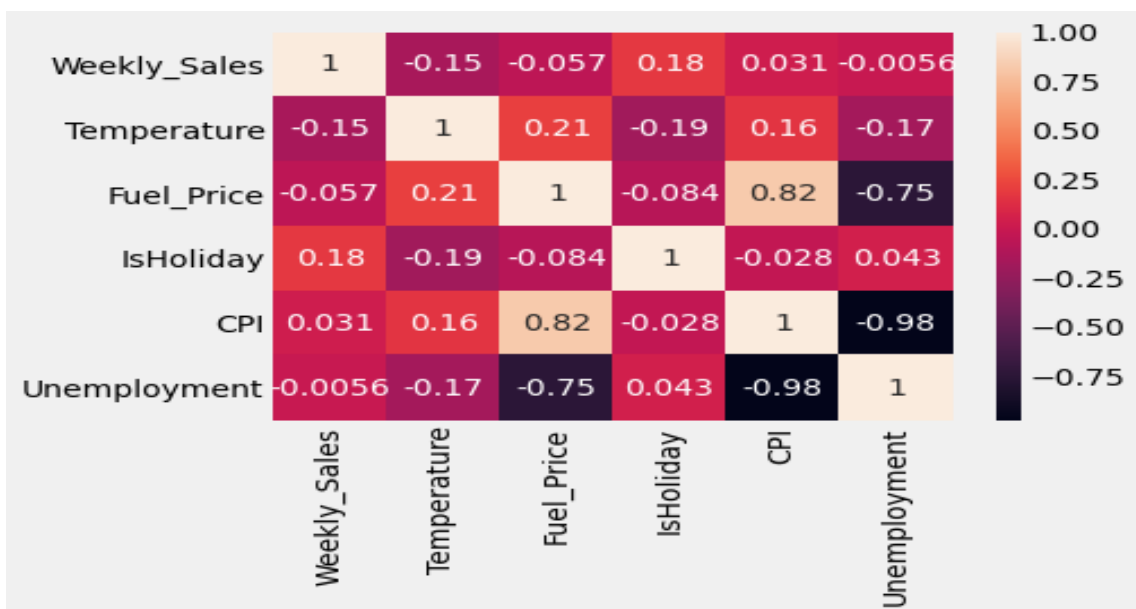


Figure 6.2: Visualization

Weekly sales don't show any high correlation with any other parameters.
 CPI and Unemployment show a negative correlation on the other hand Fuel Price shows a positive correlation
 Unemployment and Fuel price are also negatively correlated

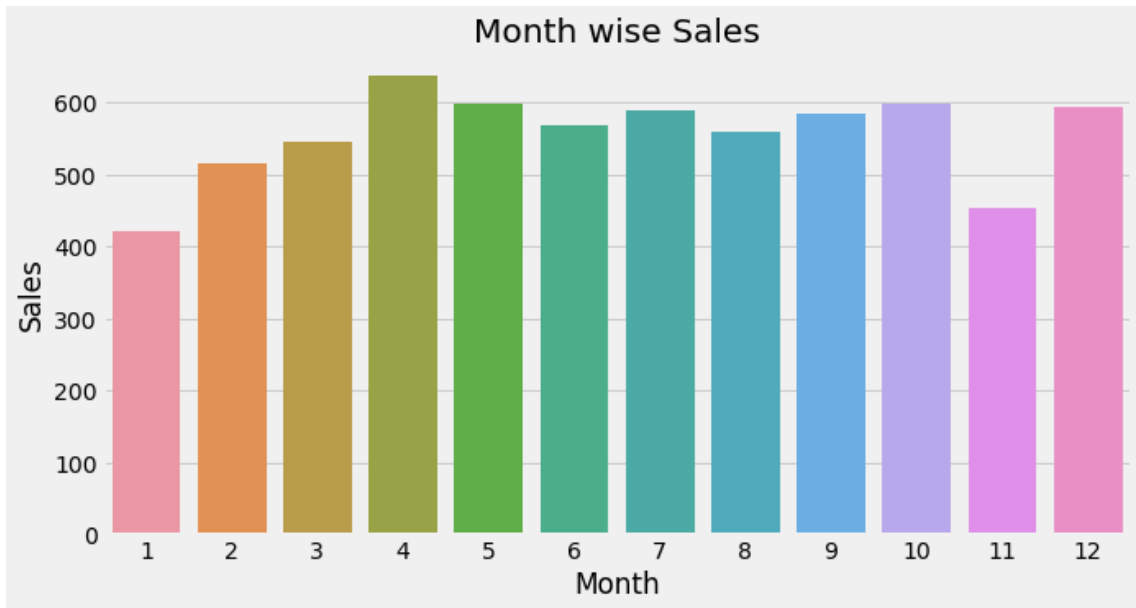


Figure 6.3: Visualization

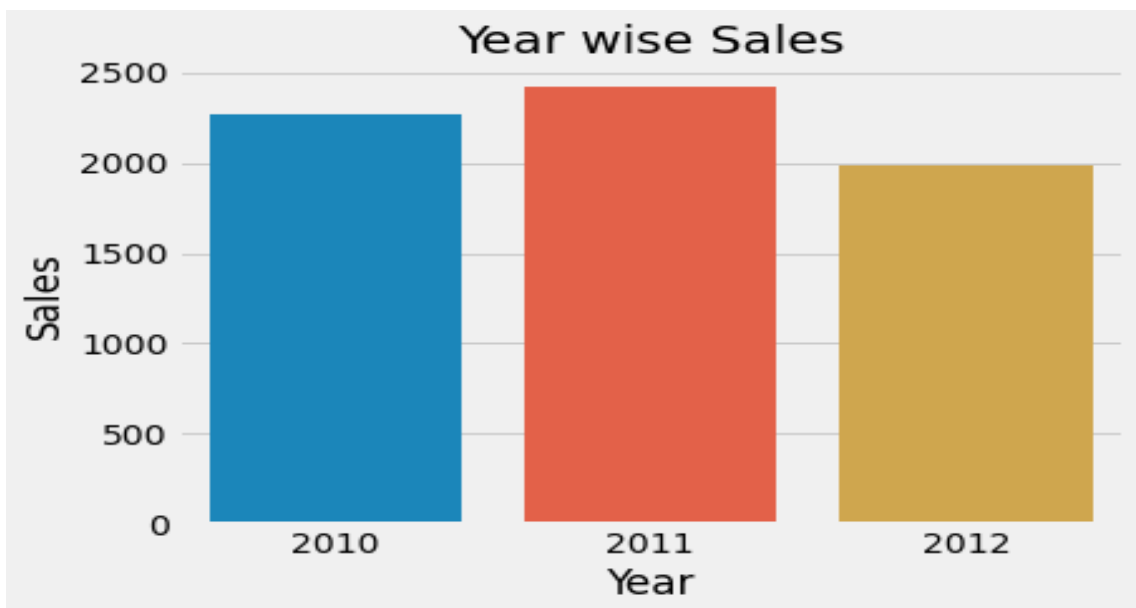


Figure 6.4: Visualization

It could be as Enough buying during offers and holidays (Black Friday Sales, Christmas, New year) and later sales go down and then again gain momentum in some months. In Nov also people show less buying can be waiting period for upcoming sales. The whole data for Dec-2013 I guess is not available so even after a spike in Dec the sales are not gone very high.

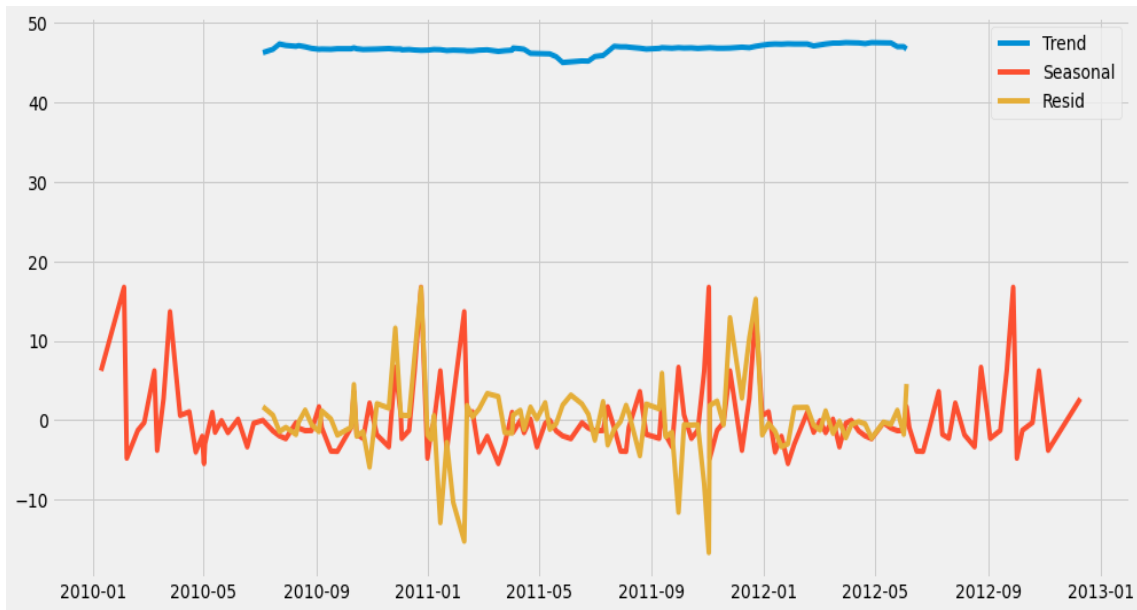


Figure 6.5: Visualization

It shows a sideways trend

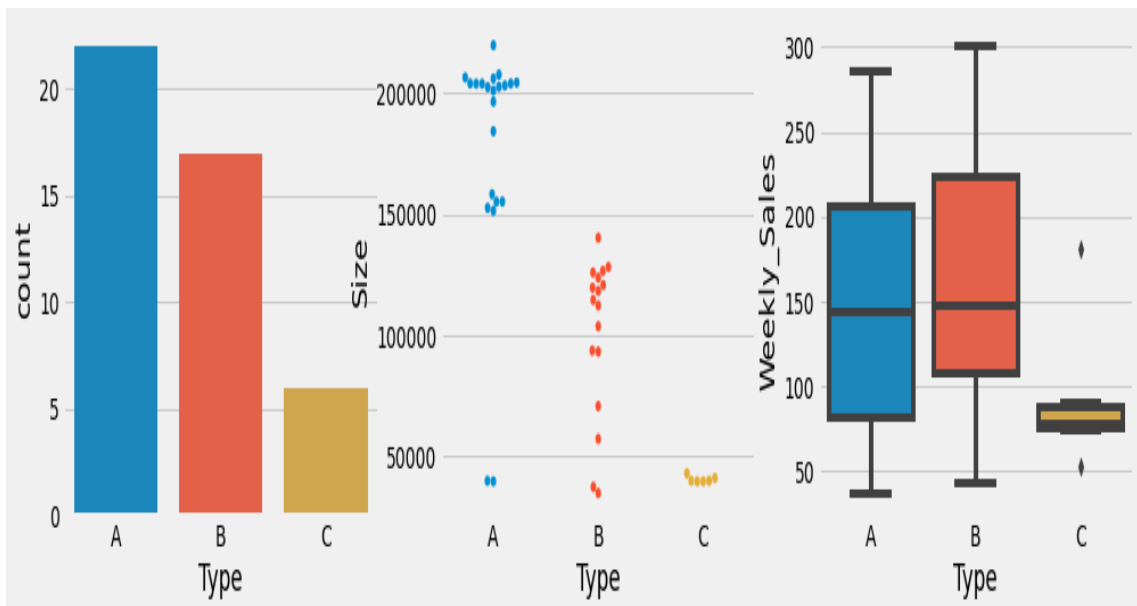


Figure 6.6: Visualization

Store Count is in the order of A, B and least is C

Size wise A store as most no of Item followed by B and than C

Even After less Count of Store and lesser Size B has more Weekly Sales than A and C at the least

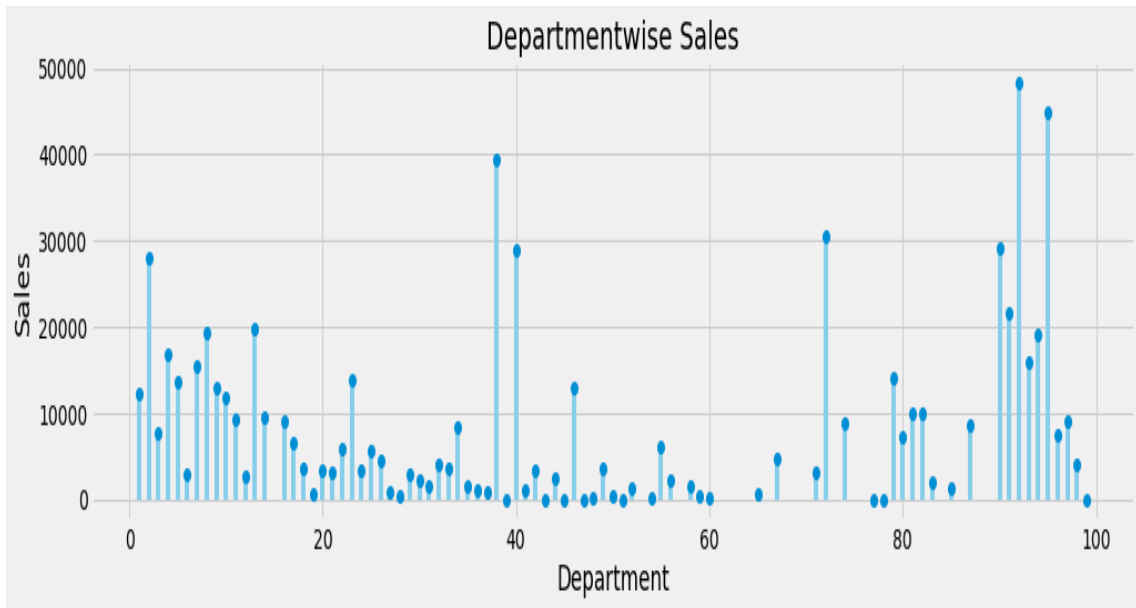


Figure 6.7: Visualization

Here we can see department 1-15, as well as 90-95, has shown higher weekly sales. some department as 38,40,72 has shown higher weekly sales.

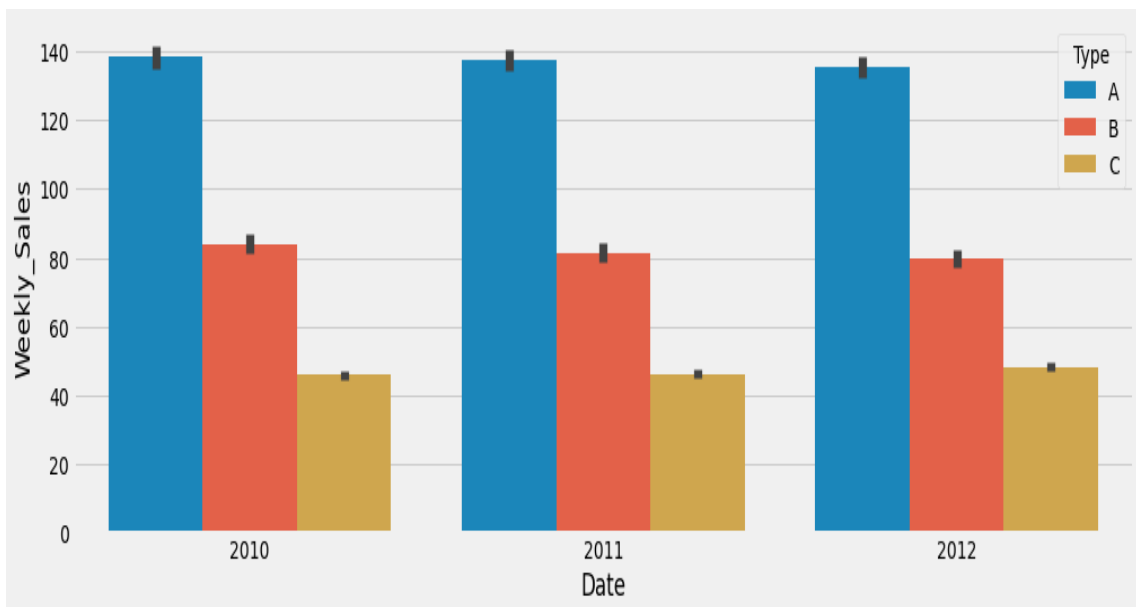


Figure 6.8: Visualization

Store Type A B has shown a slight decrease year on the basis in sales. Store C has shown a slight increase in sales.

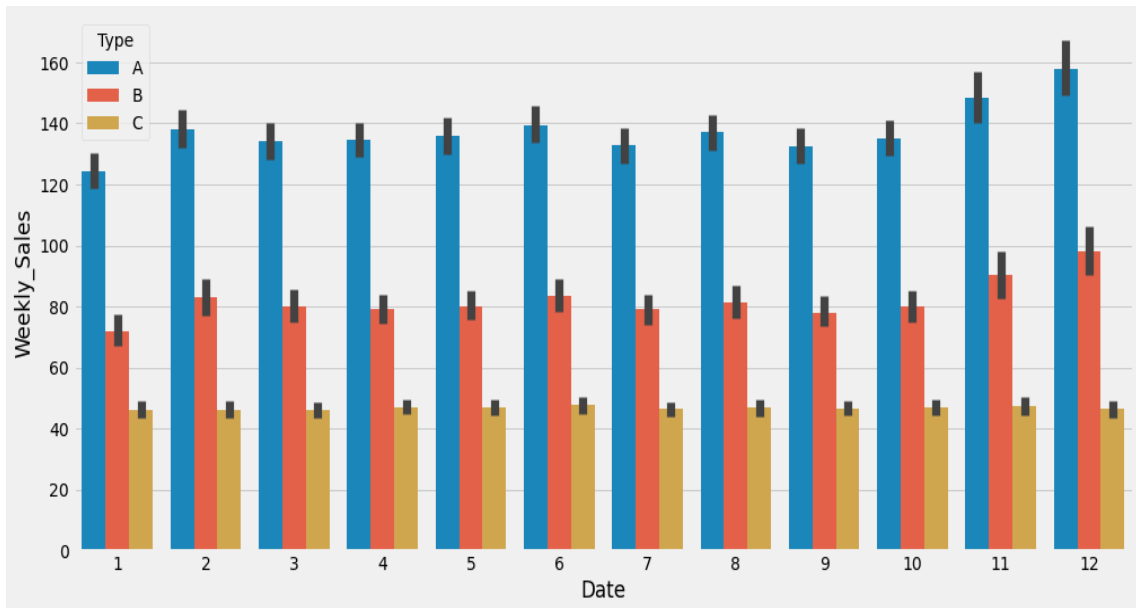


Figure 6.9: Visualization

Type A B store shows a spike in Nov Dec mostly due to holiday whereas Jan has shown a decrease in the sale.

Type C is consistent over every month and doesn't show much deviation

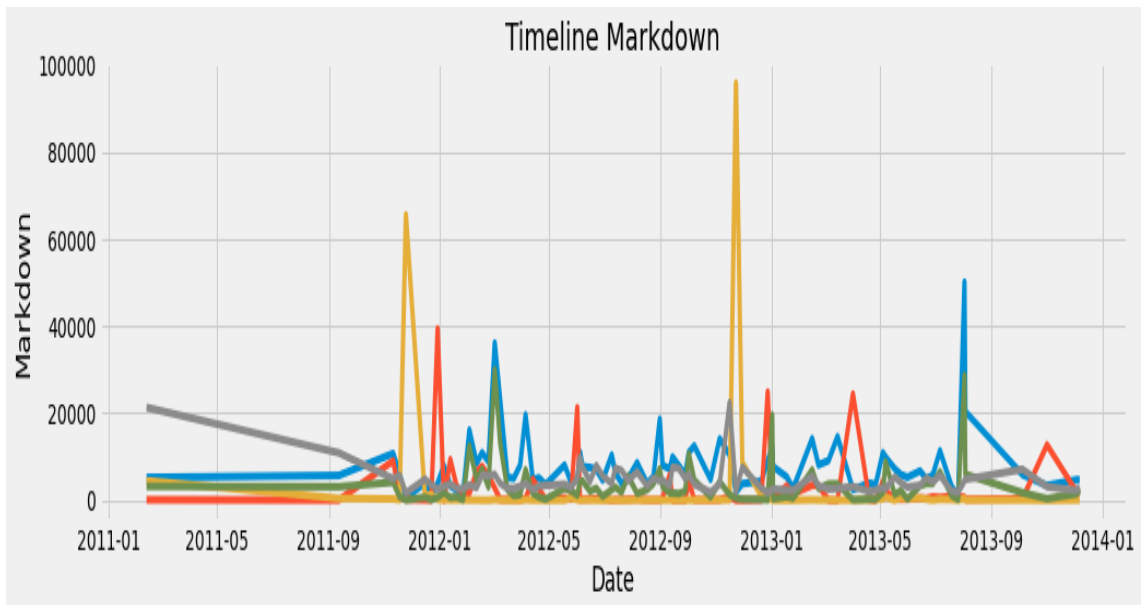


Figure 6.10: Visualization

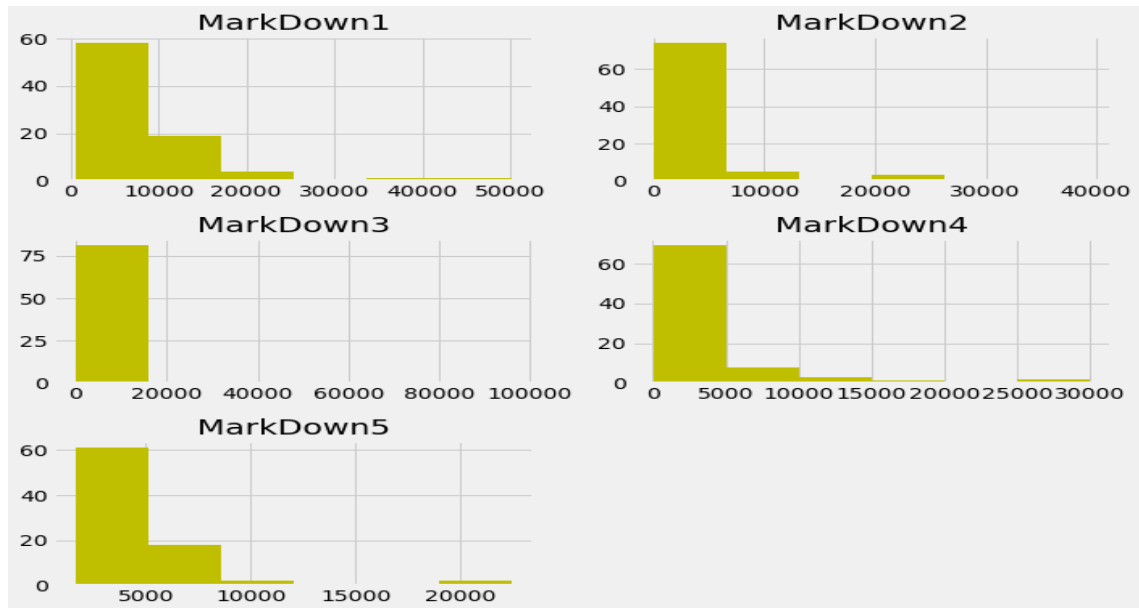


Figure 6.11: Visualization

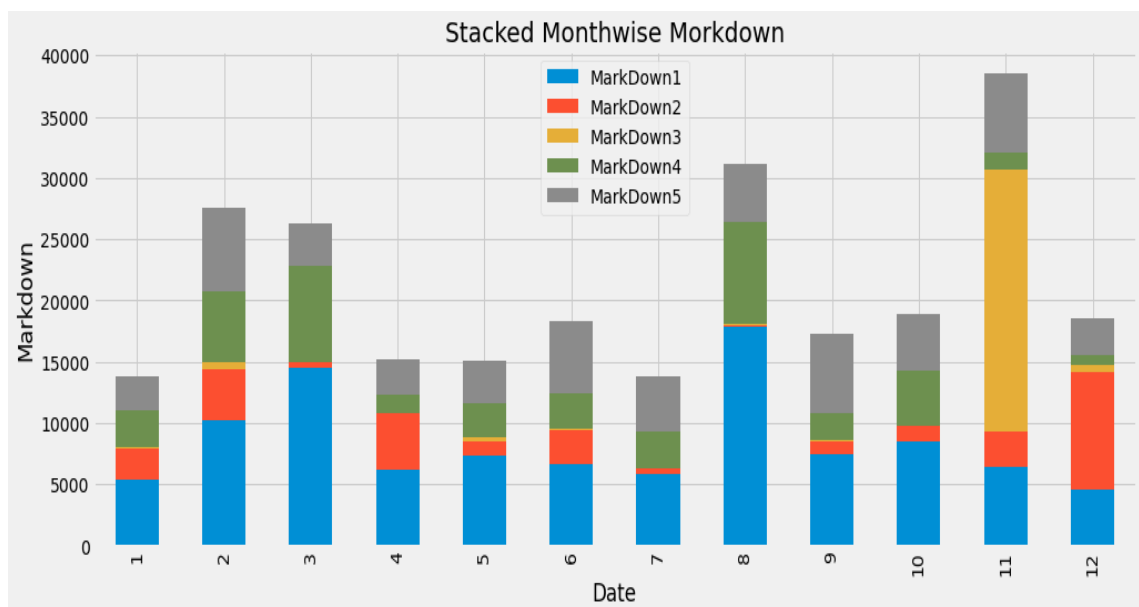


Figure 6.12: Visualization



Figure 6.13: Visualization

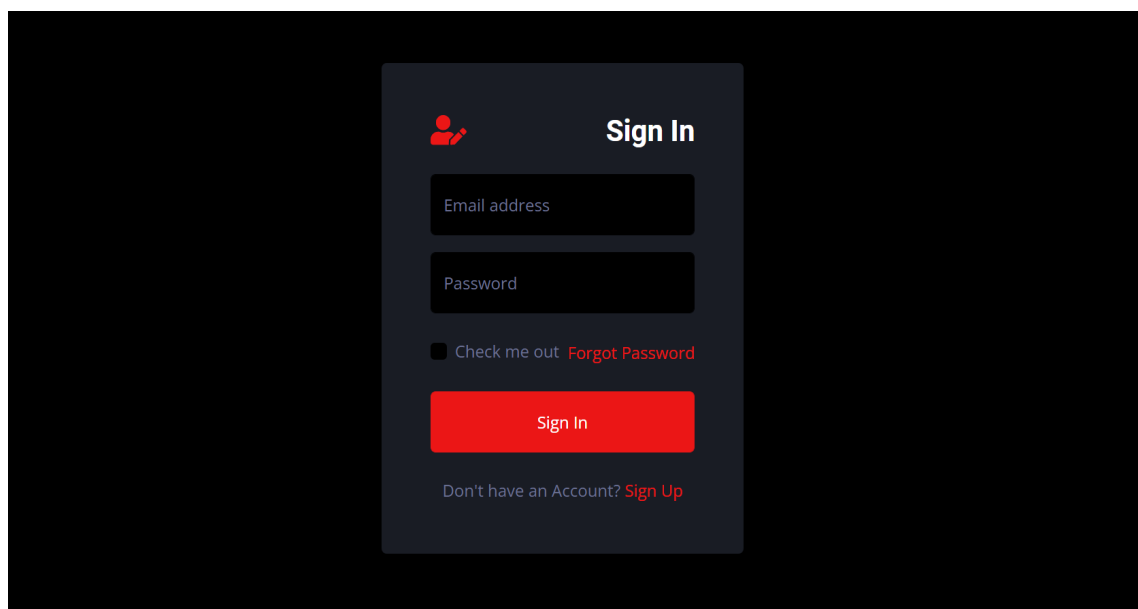


Figure 6.14: Login screen

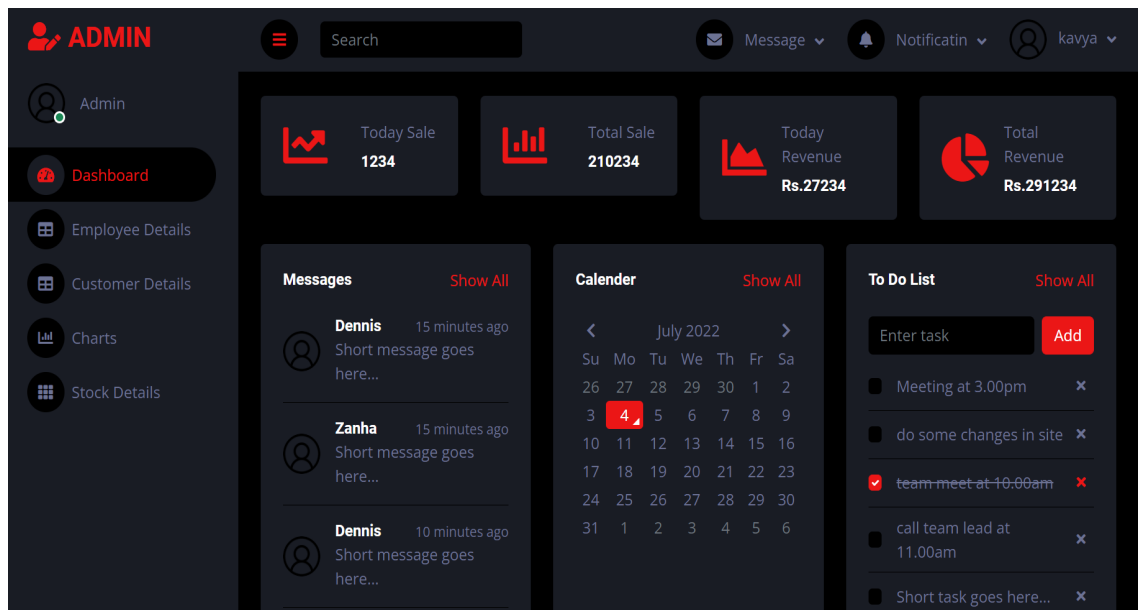


Figure 6.15: Admin - Dashboard

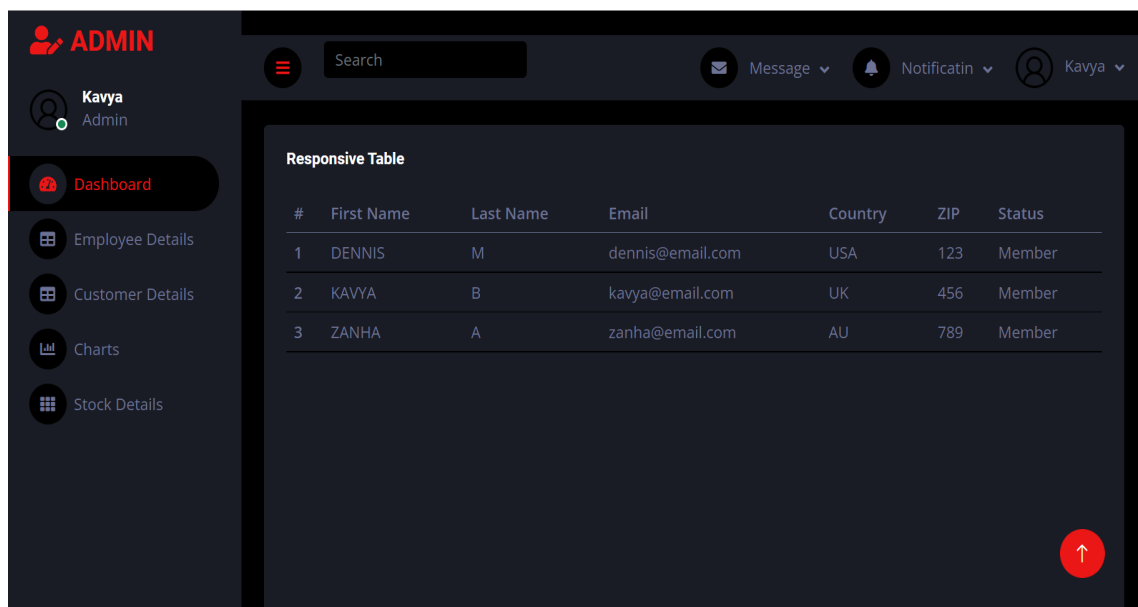
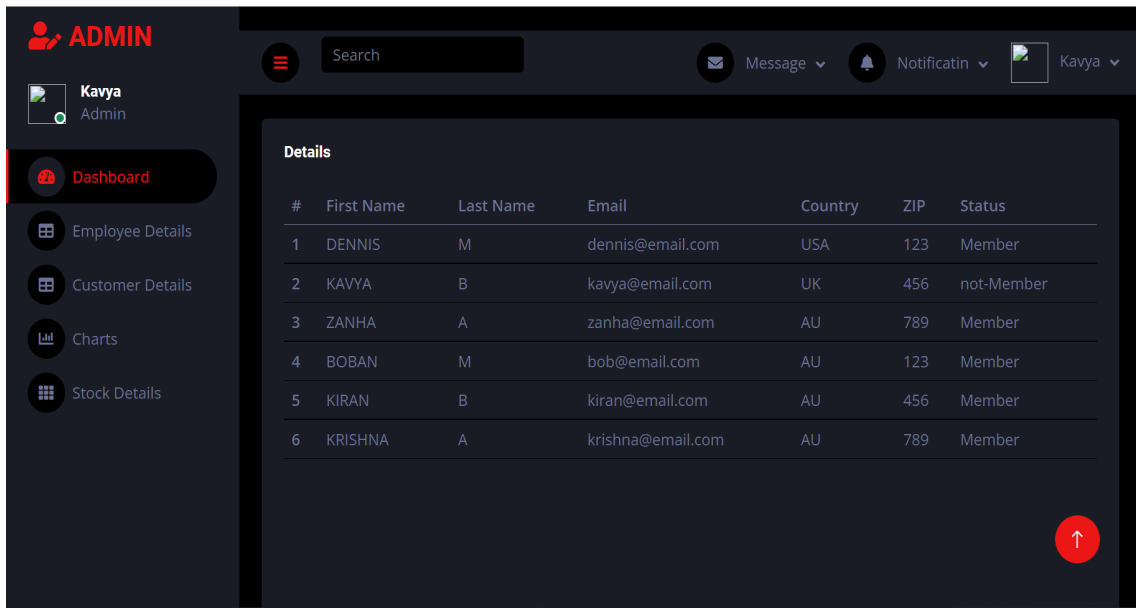


Figure 6.16: Admin - Employee Details



The screenshot shows the 'ADMIN' interface with a sidebar on the left containing 'Dashboard', 'Employee Details', 'Customer Details', 'Charts', and 'Stock Details'. The 'Dashboard' is selected. The main content area is titled 'Details' and contains a table with 6 rows of customer data. The table has columns for '#', 'First Name', 'Last Name', 'Email', 'Country', 'ZIP', and 'Status'. A red circular button with an upward arrow is located at the bottom right of the table.

#	First Name	Last Name	Email	Country	ZIP	Status
1	DENNIS	M	dennis@email.com	USA	123	Member
2	KAVYA	B	kavya@email.com	UK	456	not-Member
3	ZANHA	A	zanha@email.com	AU	789	Member
4	BOBAN	M	bob@email.com	AU	123	Member
5	KIRAN	B	kiran@email.com	AU	456	Member
6	KRISHNA	A	krishna@email.com	AU	789	Member

Figure 6.17: Admin - Customer Details

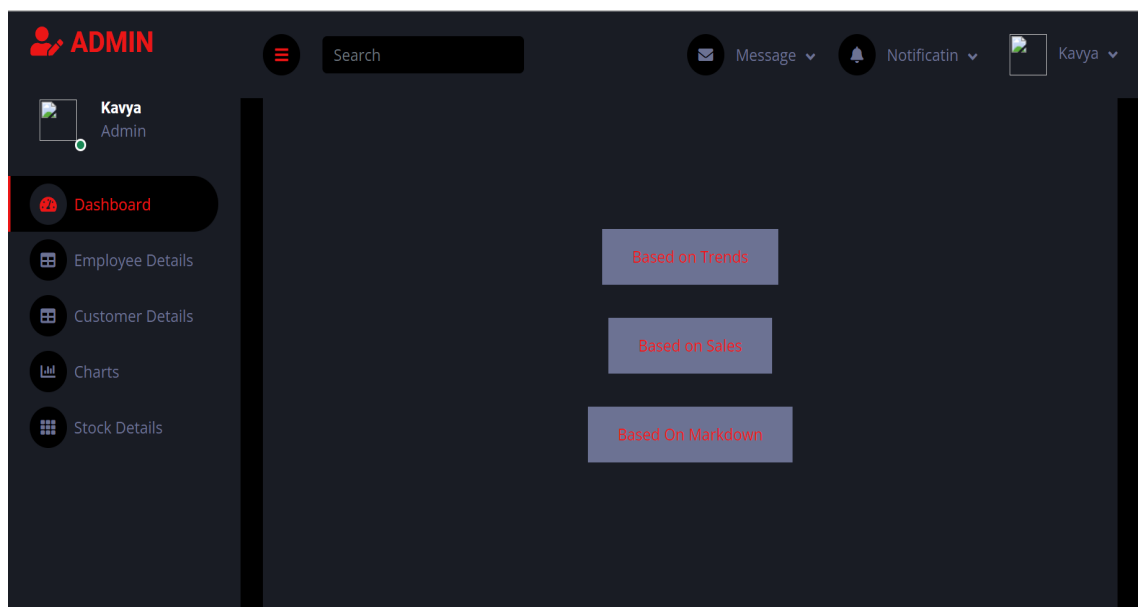


Figure 6.18: Admin - charts

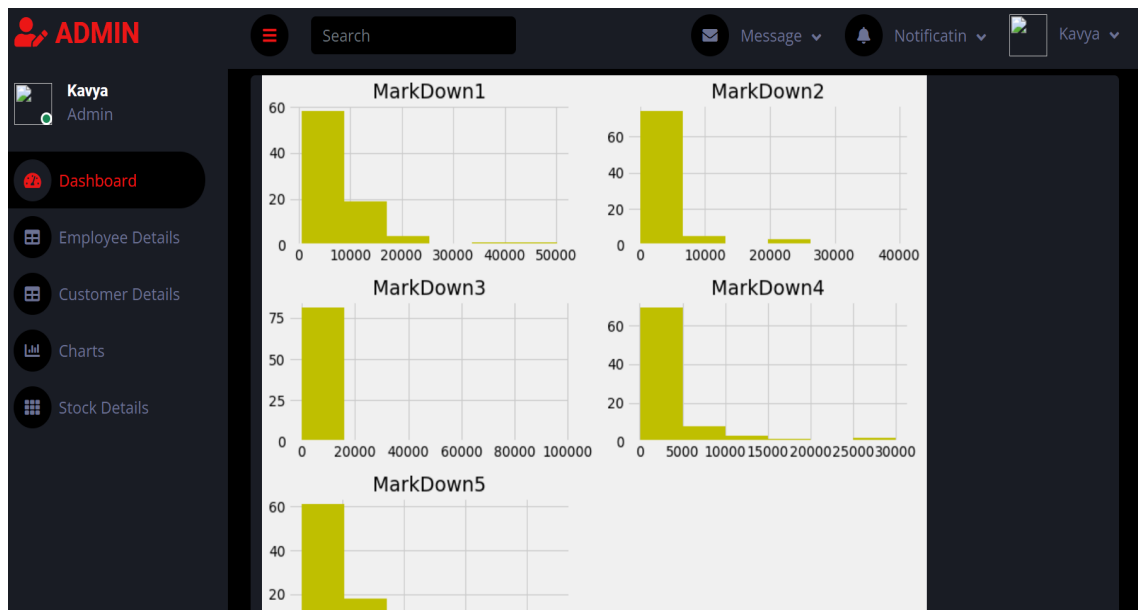


Figure 6.19: Admin - charts

The figure shows an admin dashboard with a dark theme, similar to Figure 6.19. The sidebar is on the left with the 'ADMIN' header and 'Kavya Admin' profile. The 'Dashboard' menu item is highlighted. The main content area contains a form for 'weekly sales prediction'. The form includes the following elements: a label 'Select a store from 1 to 45:' followed by a text input field; a label 'Select a Department from 1 to 98:' followed by a text input field; a label 'IsHoliday ? : ' followed by radio buttons for 'YES' and 'NO'; a label 'Select Date' followed by a date input field with the placeholder 'dd-mm-yyyy' and a calendar icon; a red 'Submit' button; and a placeholder for the prediction result, '{{prediction_text}}'.

Figure 6.20: Admin - weekly sales prediction

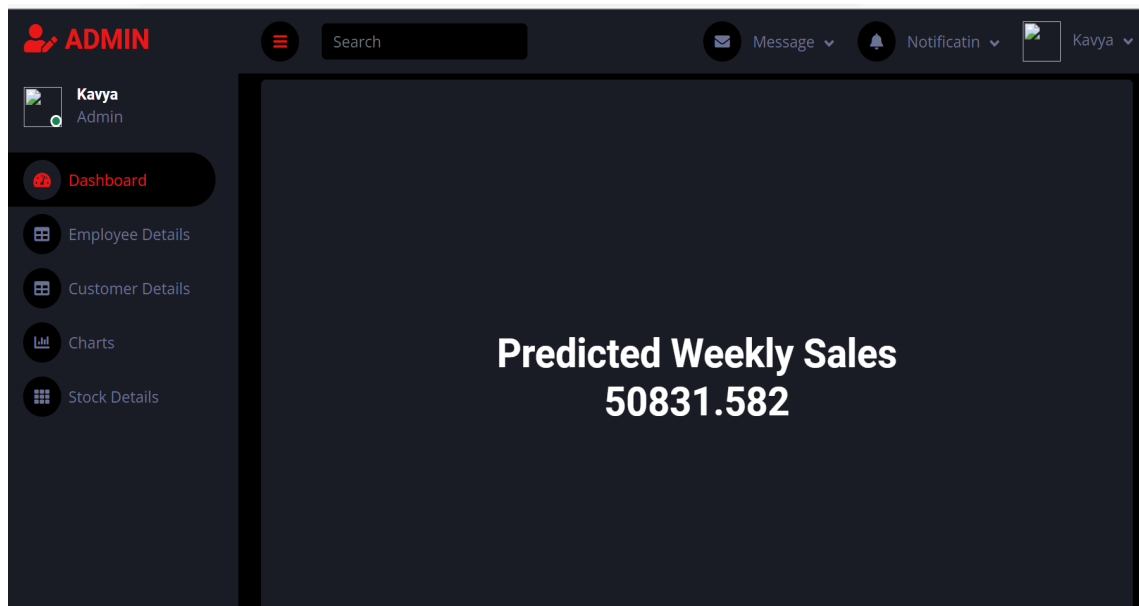


Figure 6.21: Caption