

Capstone Project

Grocery Store Forecasting Challenge For Azubian

GROUP : REFORMATION



ABOUT THE PROJECT

The objective of this project was to create a model to forecast the number of products purchased per week per store over the next eight weeks, for grocery stores located in different areas in the same country.

IMPORTANCE

- Inventory Optimization
- Production Planning
- Resource Allocation
- Financial Planning
- Customer Satisfaction

BENEFITS

- Data-driven Decisions
- Reduce Costs
- Operational Efficiency
- Competitive Advantage

DESCRIPTION OF DATA

AZUBI AFRICA

This time series forecasting project involves daily sales data from 01/01/2001 to 13/08/2002, and the dataset is comprised of several files:

01 **train.csv**

Contains the target. This is the dataset that you used to train the model.

02 **holidays.csv**

Information about holidays

03 **test.csv**

Resembles Train.csv but without the target-related columns. This is the dataset on which the trained model will be applied to

04 **stores.csv**

Information about the different stores such as their locations

05 **dates.csv**

Information about the time periods with their associated date features

06 **SampleSubmission.csv**

Shows the submission format for this competition, with the 'ID' column mirroring that of Test.csv.

ISSUES WITH THE DATA & HOW WE FIXED THEM

AZUBI AFRICA

THE ISSUES

The data type for the date column in train, test, holidays, dates datasets are in numerical format

THE SOLUTIONS

Convert to Datetime format

stores dataset: the city, type & cluster are data type - int

Convert to string and make the categories more descriptive

The type column in holidays dataset is in a numerical format and the types of holidays do not have an ordinal relationship

Convert to string and make it more descriptive

After checking for unique values in yearofday column in the dates dataset, we found 366 unique values which showed that there was a leap year.

Create two new columns called "sin(dayofyear)" & "cos(dayofyear)". These new columns will help our machine learning models understand the cyclic nature of a year.

HYPOTHESIS & QUESTIONS

AZUBI AFRICA

HYPOTHESIS

- H0: Holidays have a big effect on sales, hence the sales data is seasonal.
H1: Holidays don't affect sales, hence sales data is stationary.

QUESTIONS ABOUT THE DATA

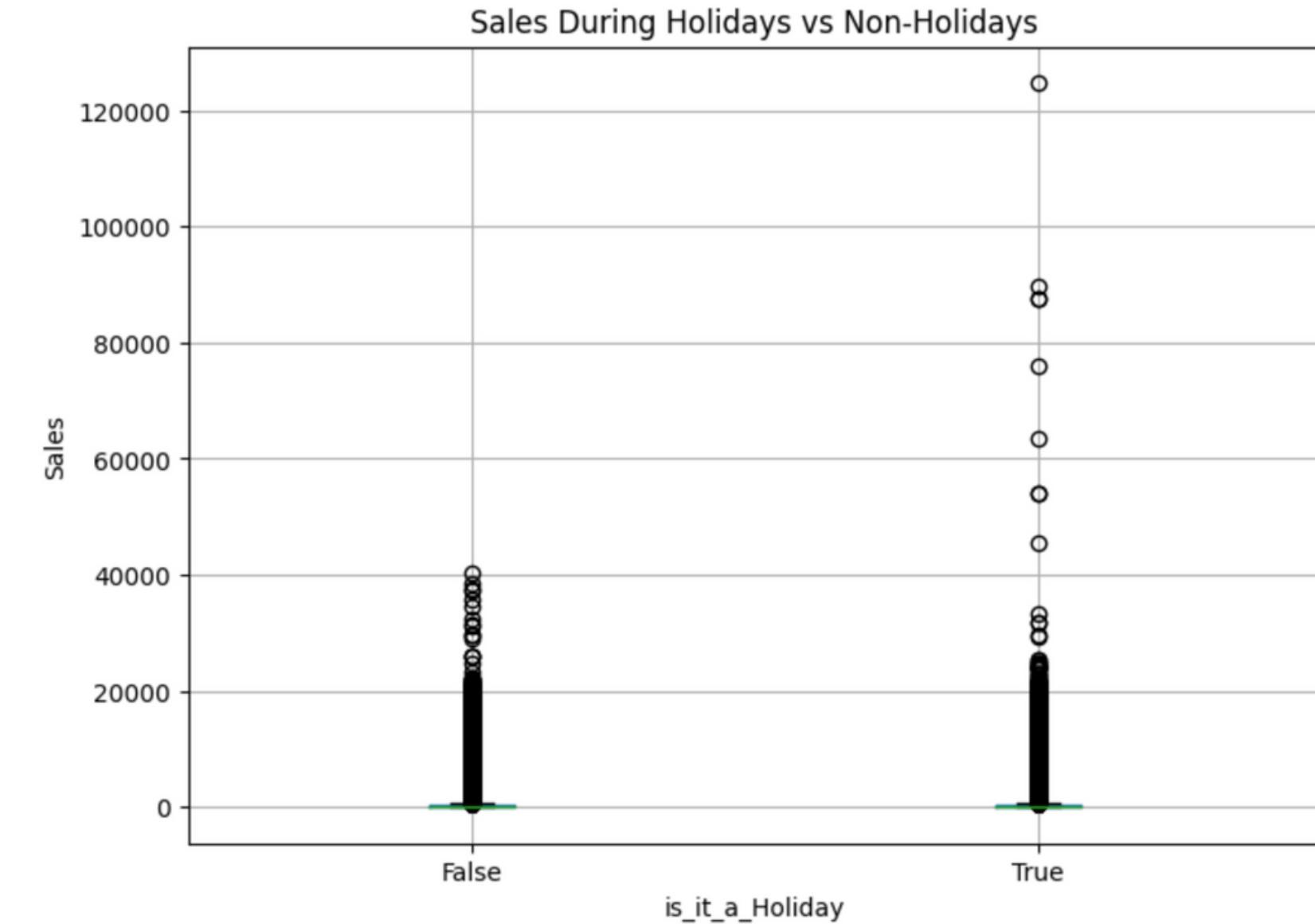
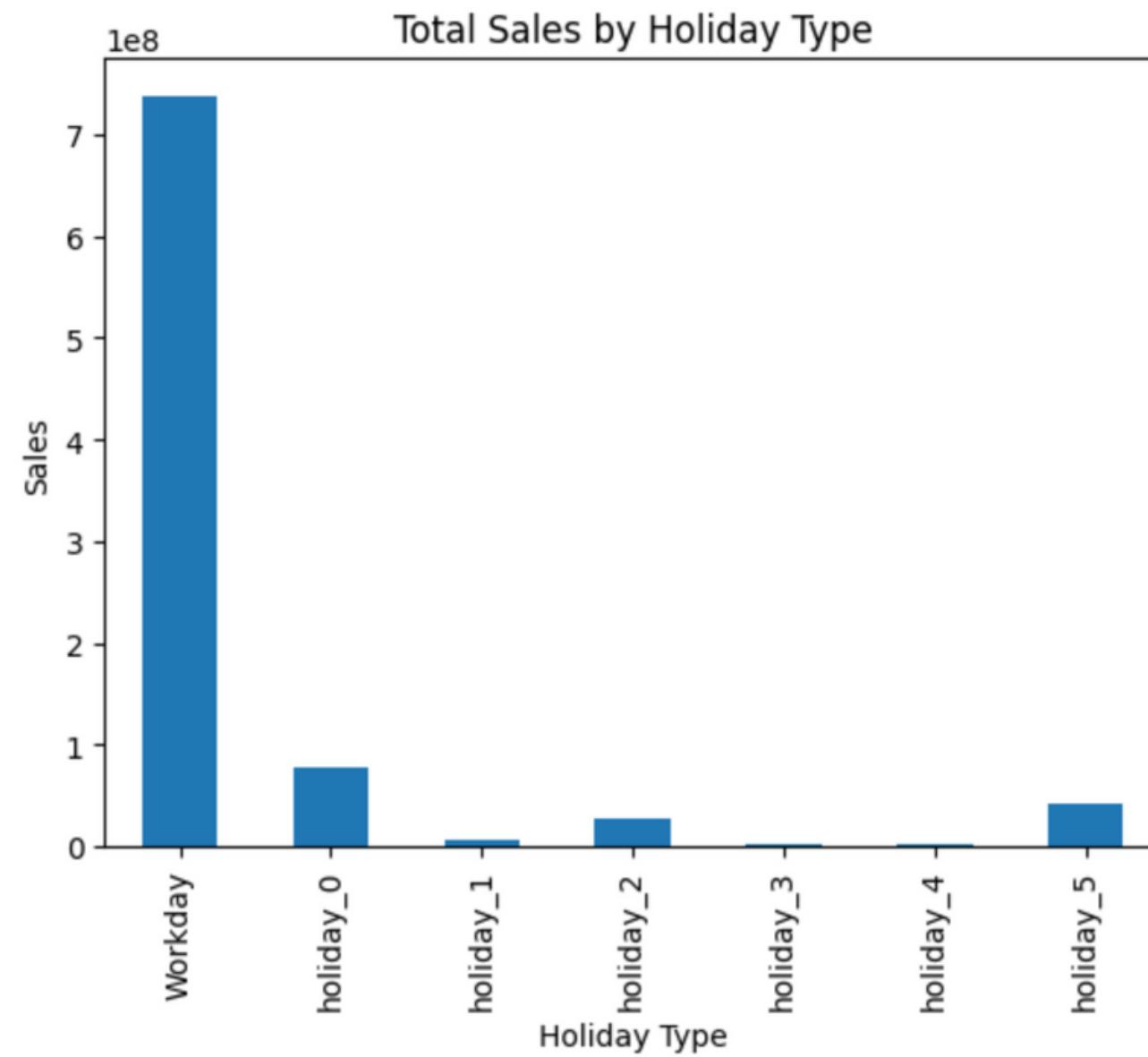
1. What is the distribution of sales?
2. What is the average sales for each category?
3. How do sales vary by promotion status?
4. Is there a relationship between sales and the number of transactions?
5. How do sales vary during holidays compared to non-holidays?

QUESTIONS ABOUT THE DATA

6. How do sales vary by holiday type?
7. What is the trend in sales over time?
8. How do sales vary across different store IDs?
9. Are there any seasonal patterns in sales?
10. How do sales vary across different combinations of category and promotion?

HYPOTHESIS VALIDATION

AZUBI AFRICA

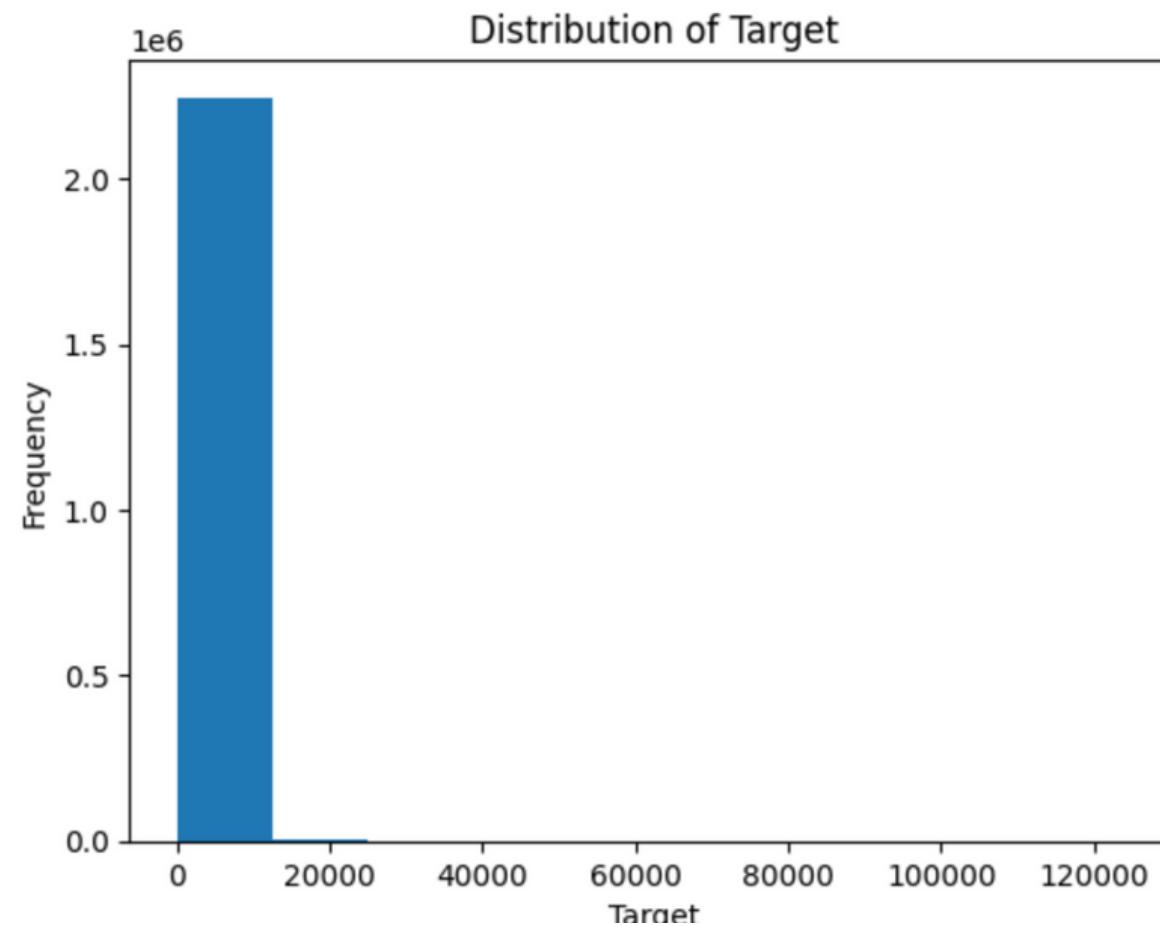


The hypothesis H1, which states that holidays don't affect sales and the sales data is stationary, is more likely to be

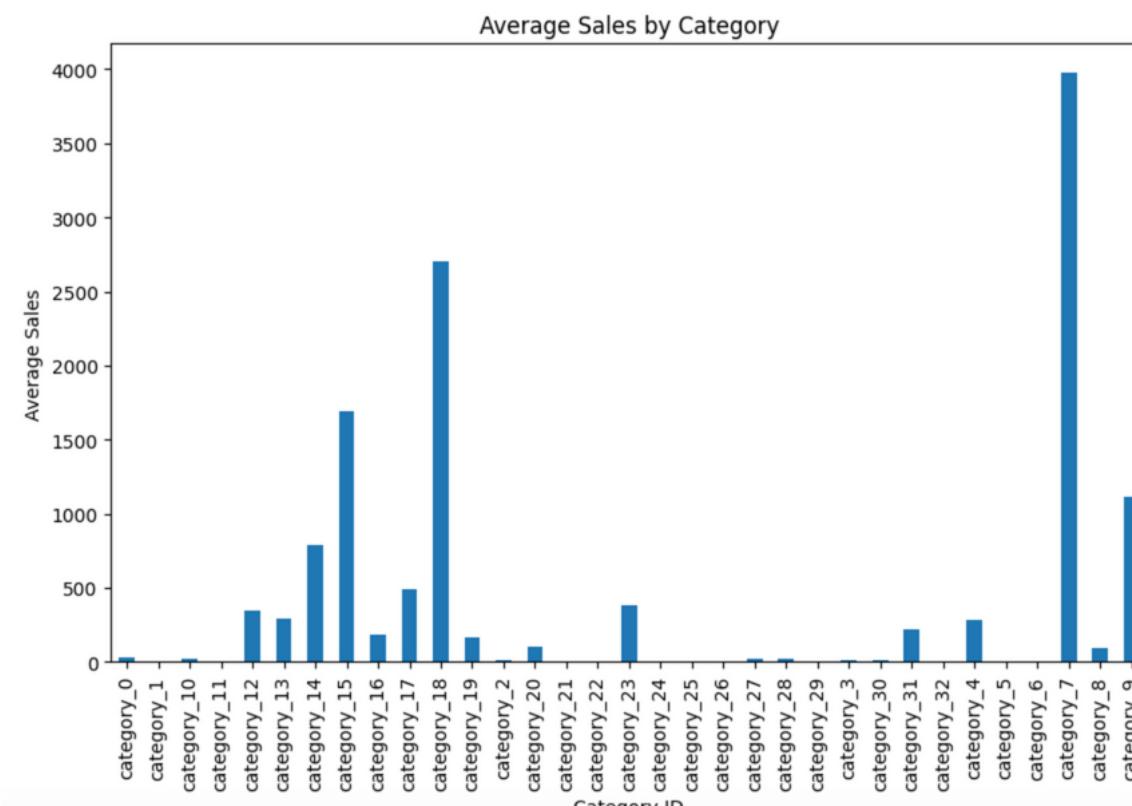
Based on the observation from the above, it is reasonable to conclude that the sales data is more likely stationary, and the presence or absence of holidays does not strongly influence the sales patterns.

ANSWERING THE QUESTIONS

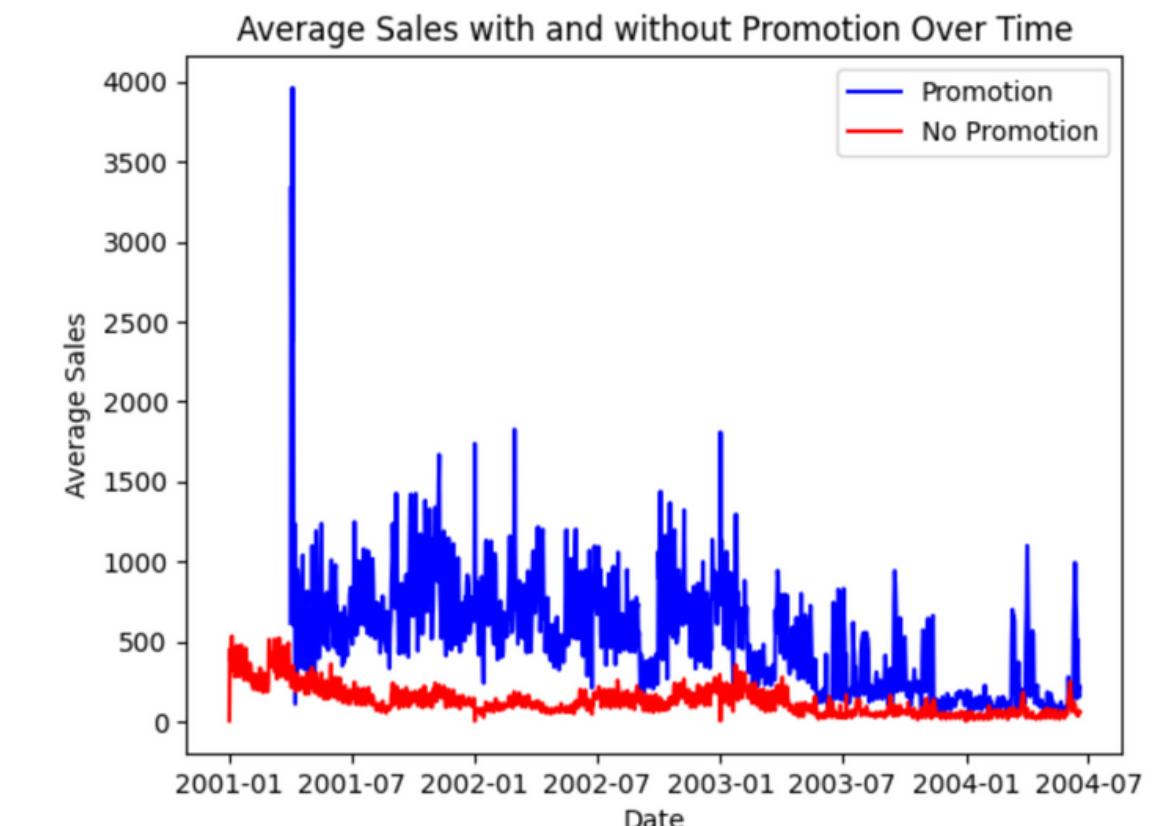
What is the distribution of Sales?



What is the average sales for each category?



How do sales vary by promotion status?

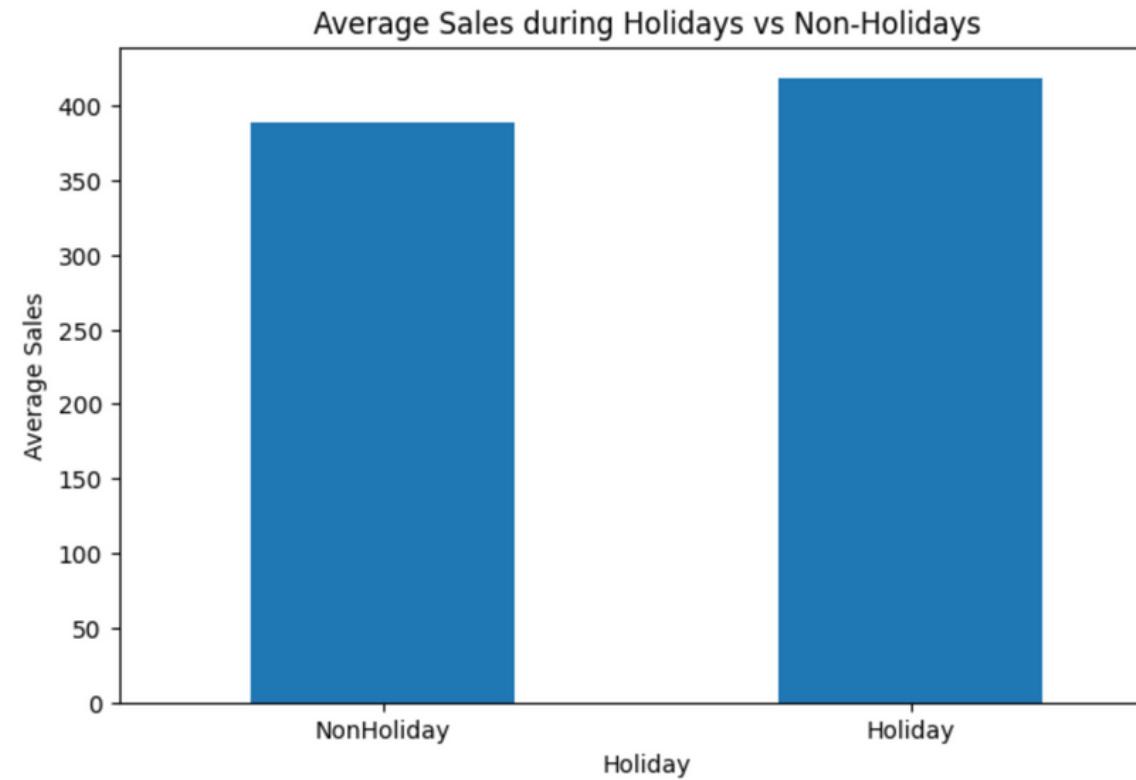


4. Is there a relationship between sales and the number of transactions?

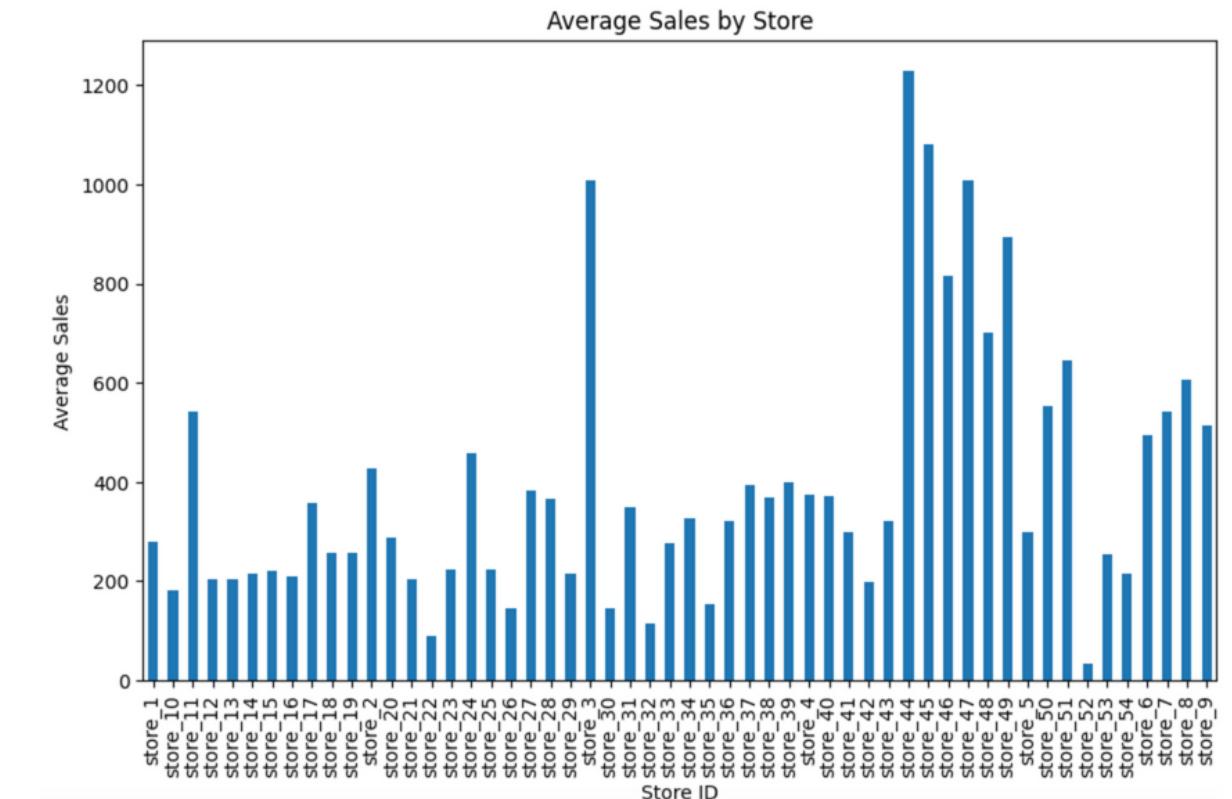
ANSWERING THE QUESTIONS

AZUBI AFRICA

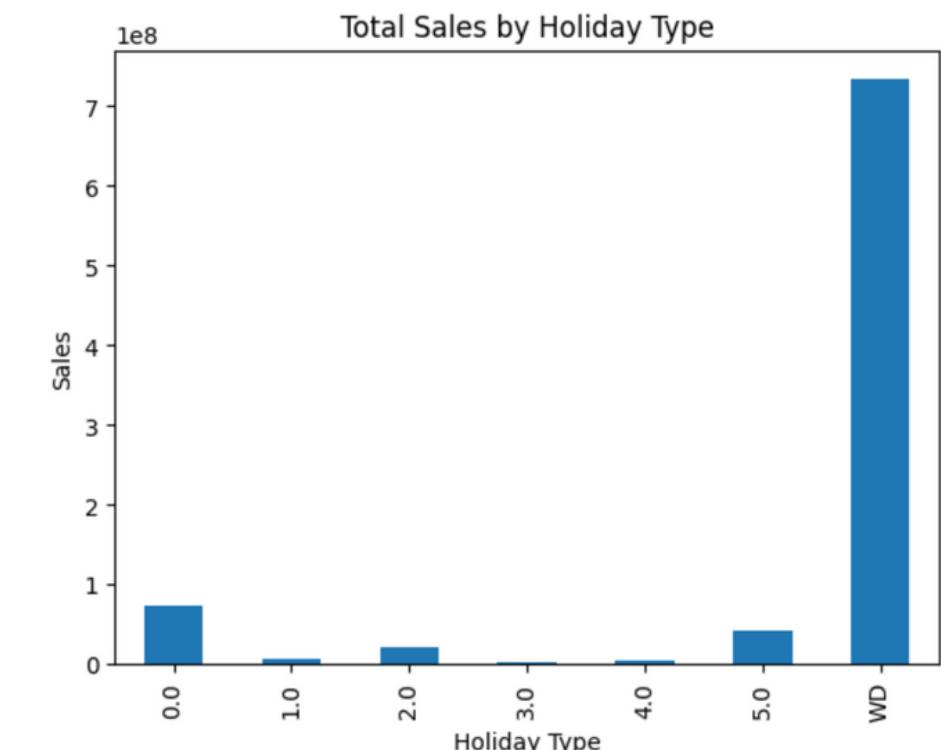
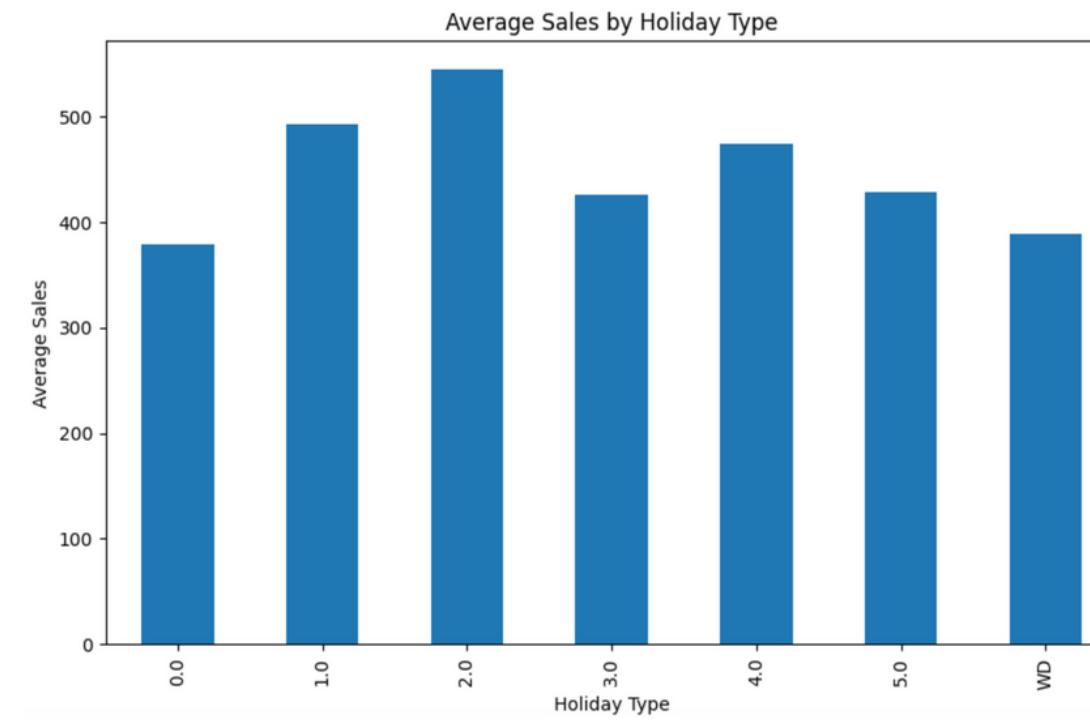
How does sales vary during holidays compared to non-holidays?



How do sales vary across different store IDs?



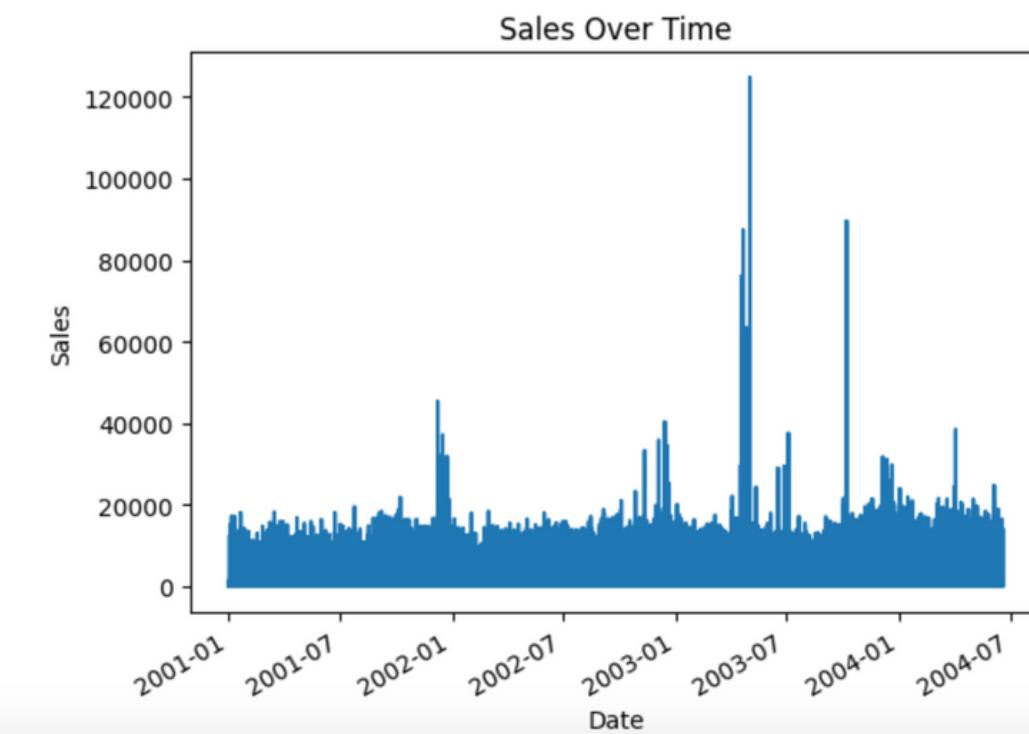
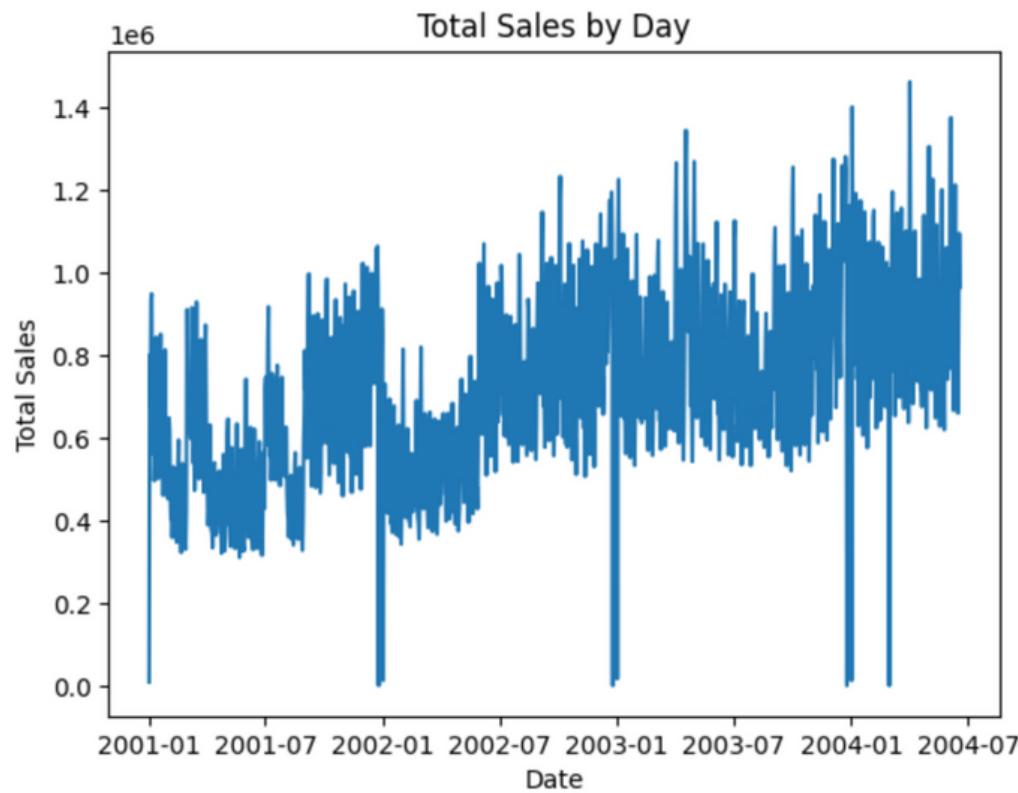
How do sales vary by holiday type?



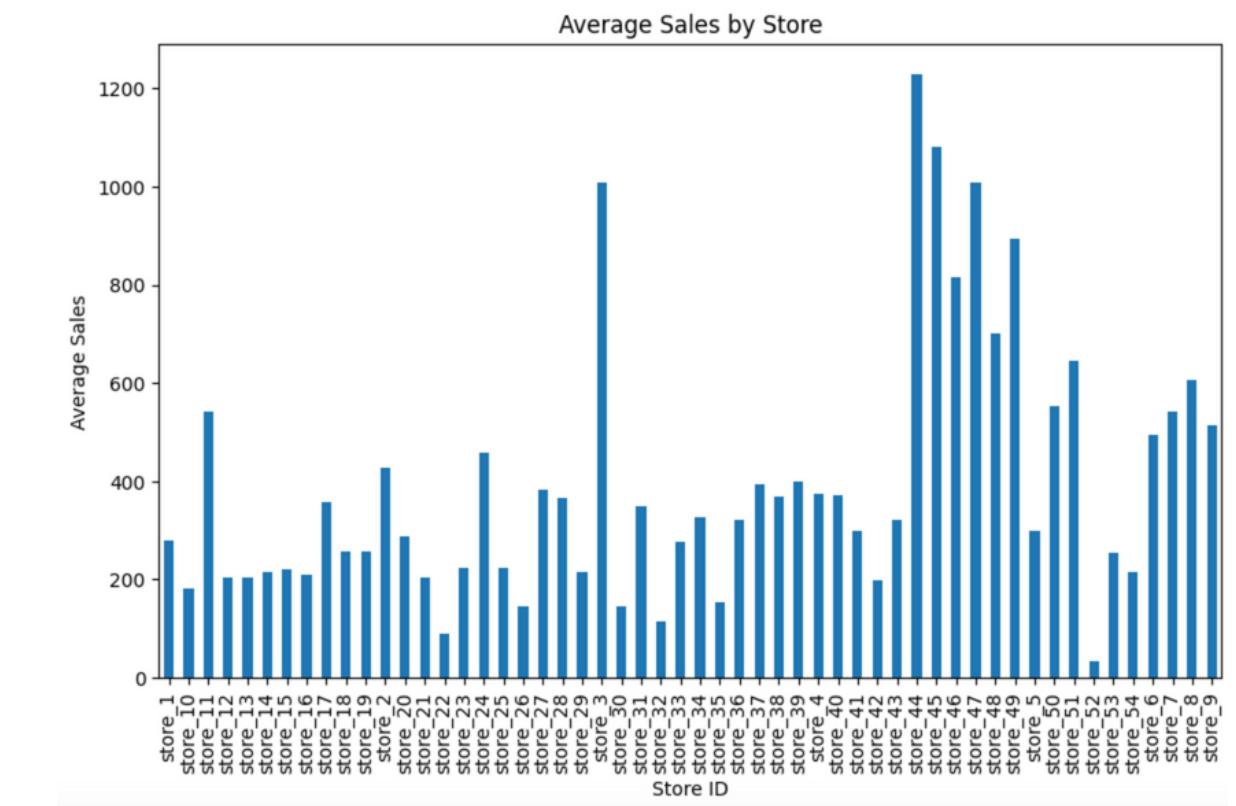
ANSWERING THE QUESTIONS

AZUBI AFRICA

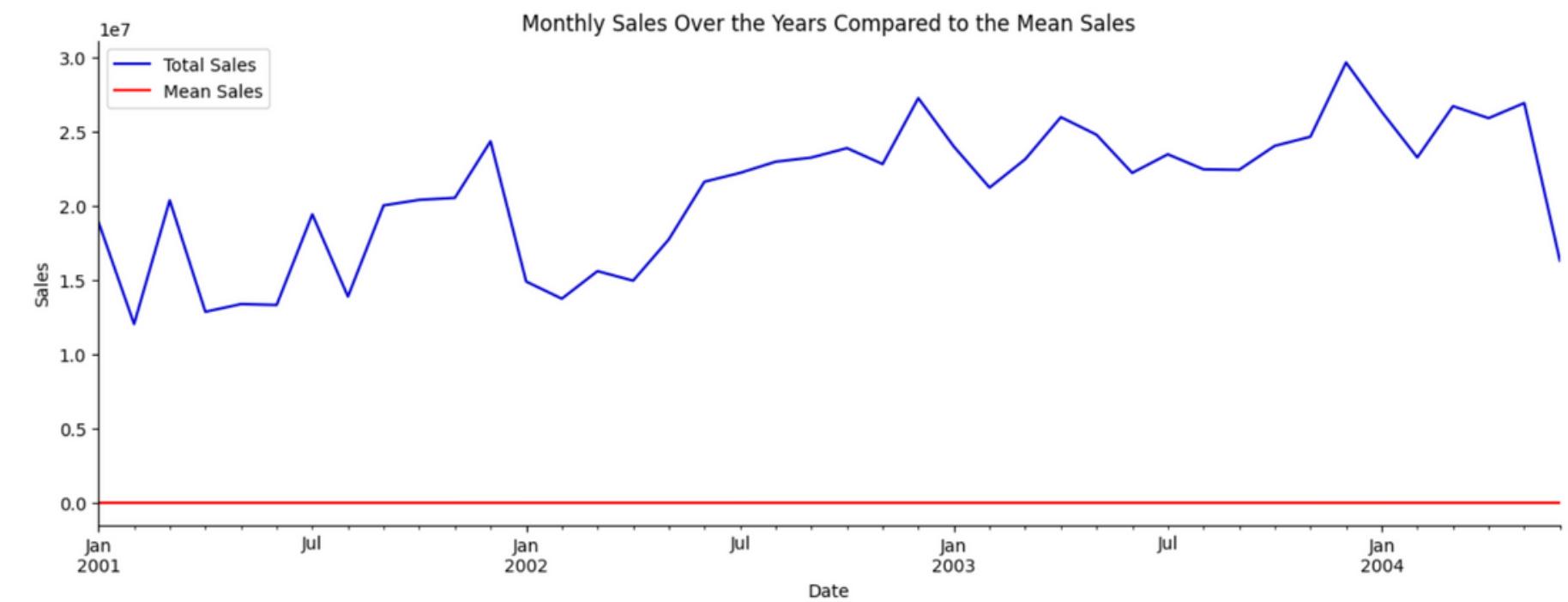
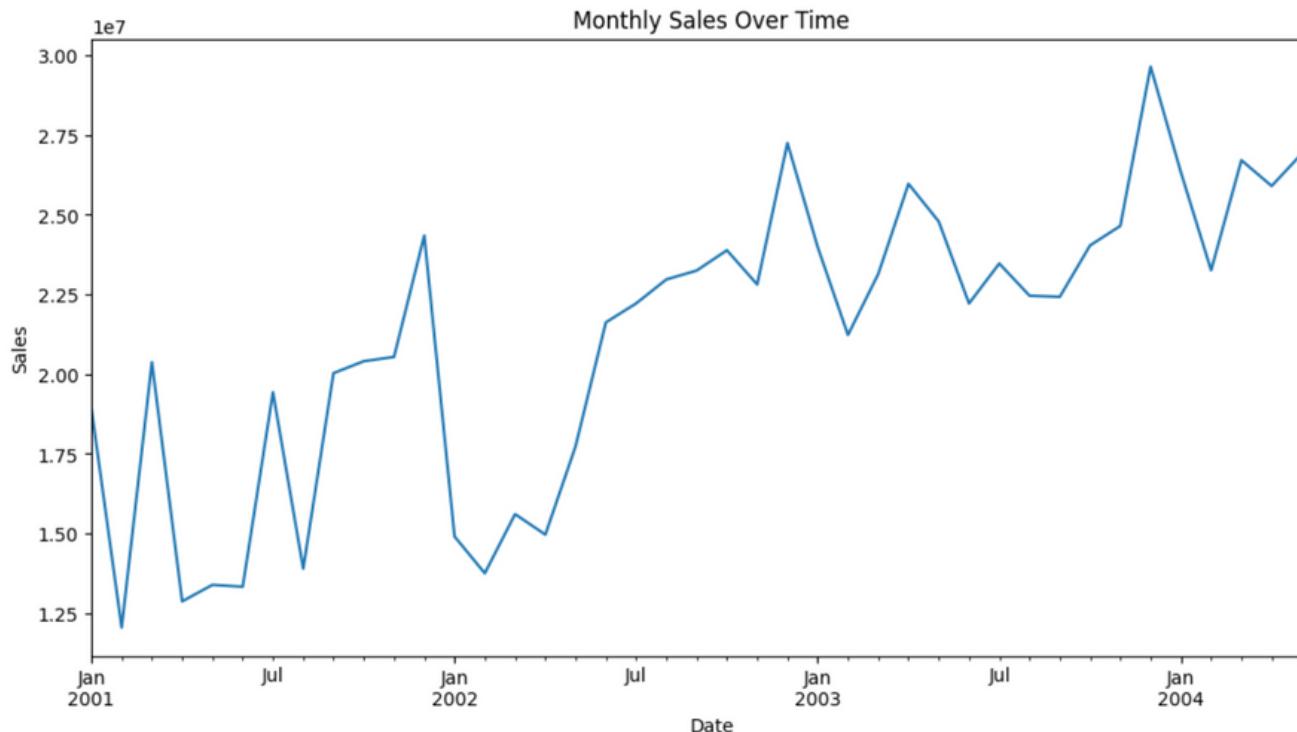
What is the trend in sales over time?



How do sales vary across different store IDs?



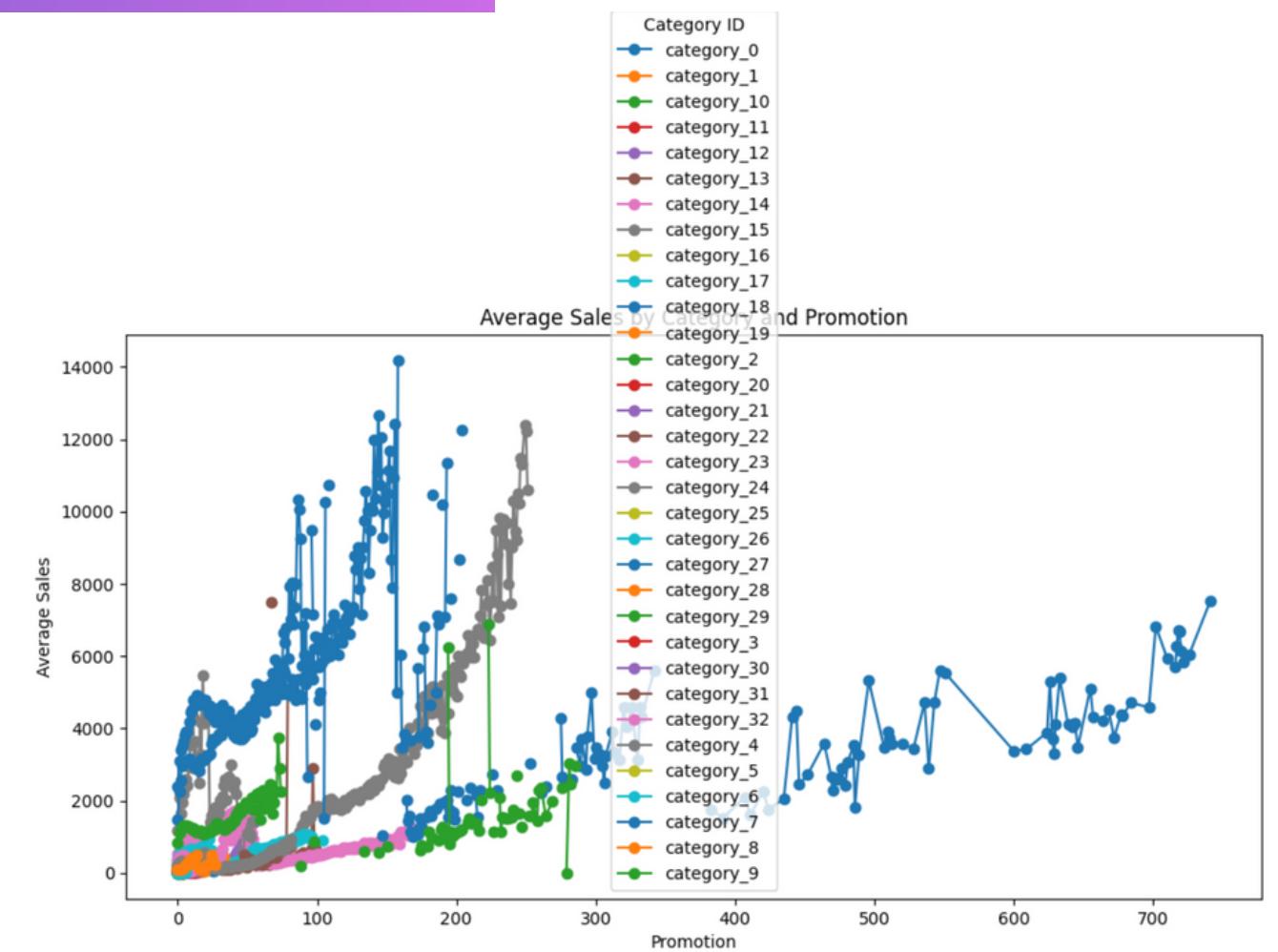
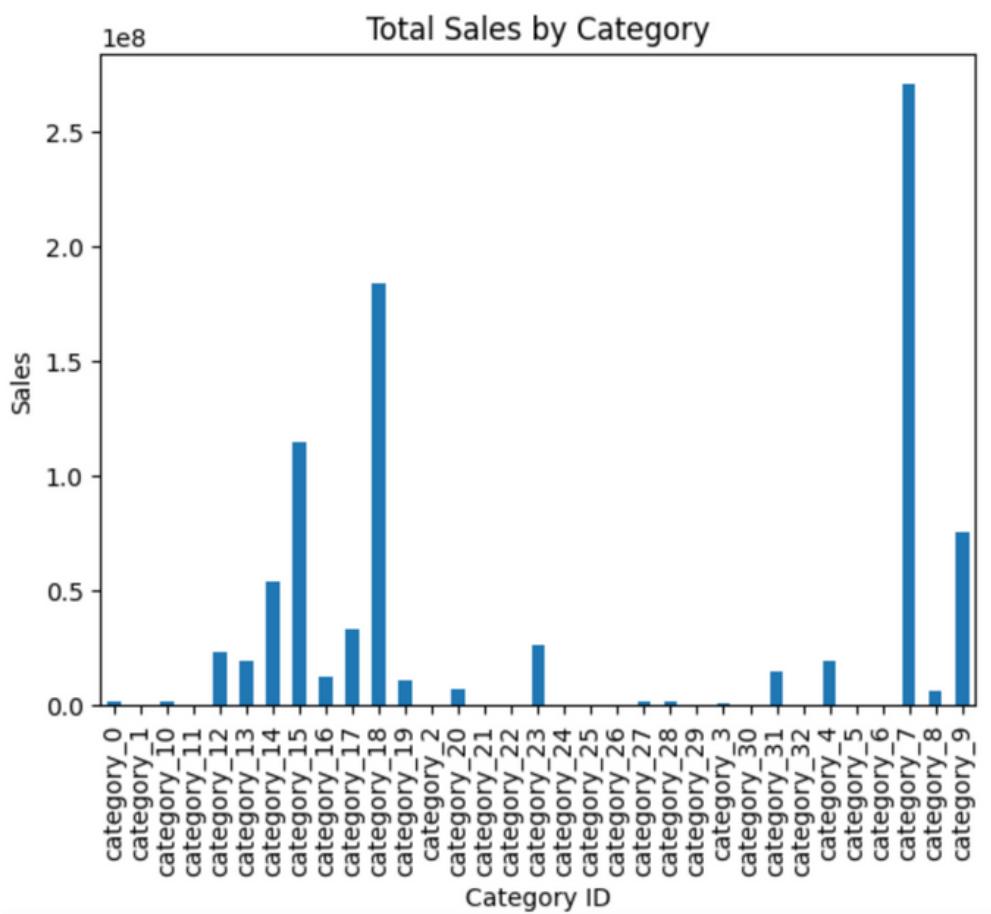
Are there any seasonal patterns in sales?



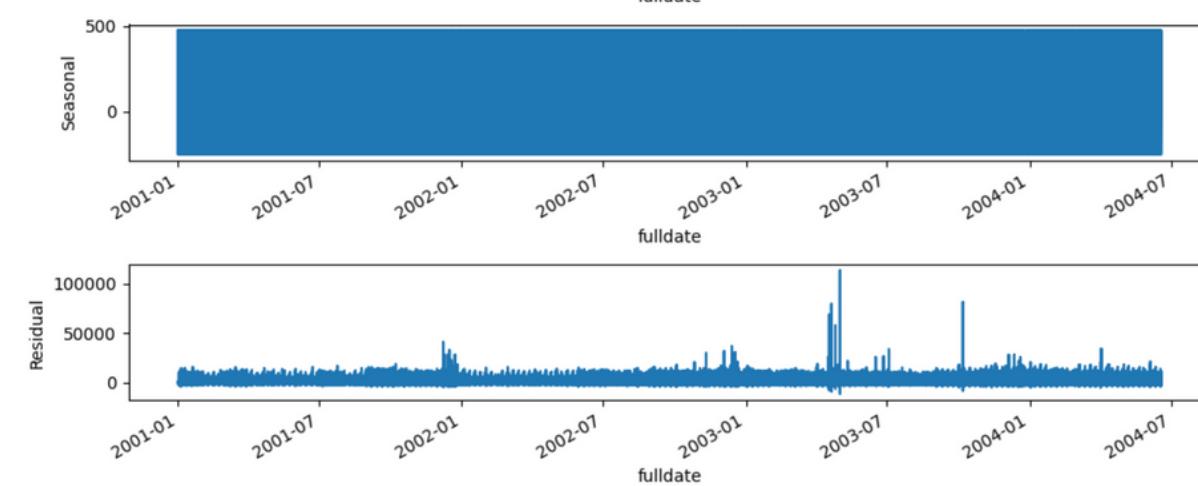
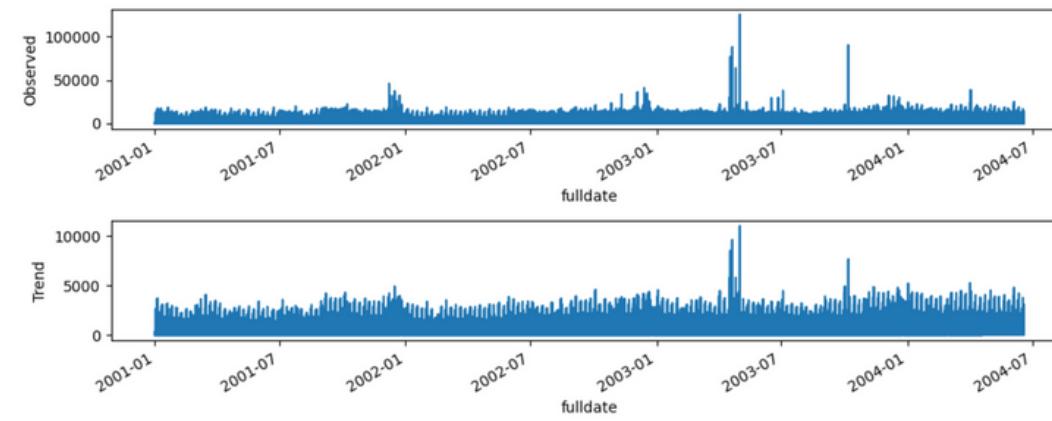
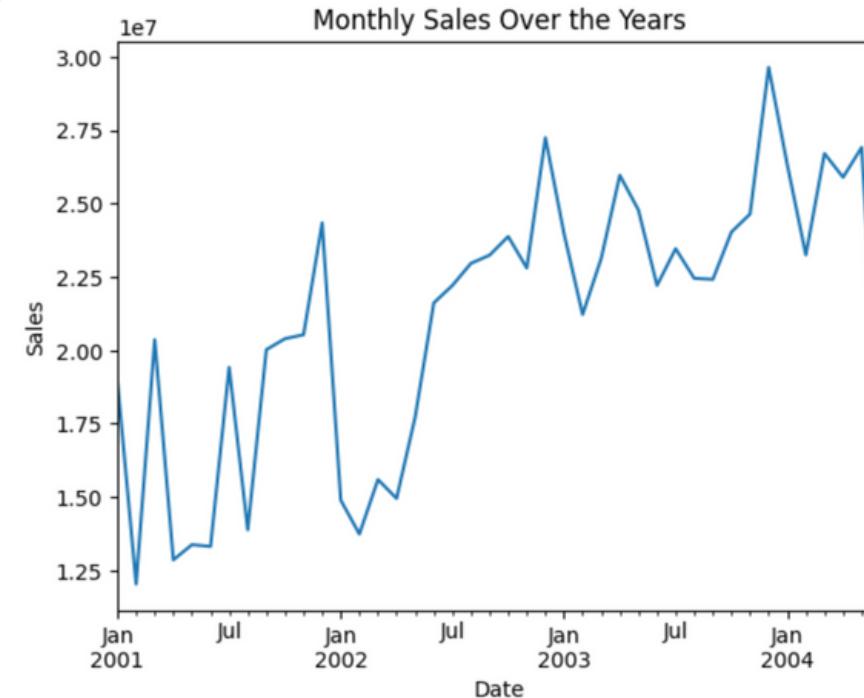
ANSWERING THE QUESTIONS

AZUBI AFRICA

How do sales vary across different combinations of category and promotion?



Monthly Statistics



MACHINE LEARNING MODELS

(TIME SERIES)

AZUBI AFRICA

MODELS

- DecisionTreeRegressor
- K-Nearest Neighbors (KNN)
- RandomForestRegressor
- Support Vector Regression (SVN)
- Gradient Boosting
- XGBoost
- Linear Regression

MODELS COMPARISON

	Model	MSE	MSLE	RMSE	RMSLE
0	DecisionTree	25529299.99	4.91	5052.65	2.22
1	KNN	33727331.33	5.60	5807.52	2.37
2	Random Forest	24470557.88	4.14	4946.77	2.04
3	SVR	65221316.88	12.05	8075.97	3.47
4	Gradient Boosting	32664295.38	12.05	5715.27	3.47
5	XGBoost	26561370.73	12.05	5153.77	3.47
6	Linear Regression	12152525365580938488123977883451392.00	12.05	110238493120964512.00	3.47

CONCLUSION

Based on the provided results, the model with the lowest values for these metrics is the Random Forest model.

Therefore, the Random Forest model is the best performing model among the ones listed, as it achieves the lowest MSE, MSLE, RMSE, and RMSLE values.

THE APP

AZUBI AFRICA

Navigation

Select an option

- Home
- Home
- About

Welcome

SEER

A Streamlit Sales Forecasting / Prediction App

Predict

Welcome

This is a Sales Forecasting App.



SEER- A Sales Forecasting APP

Enter the required information to forecast sales:

Date	Store_type
2023/06/22	Store_3

How many products are on promotion?	Store_id
0	Store_2

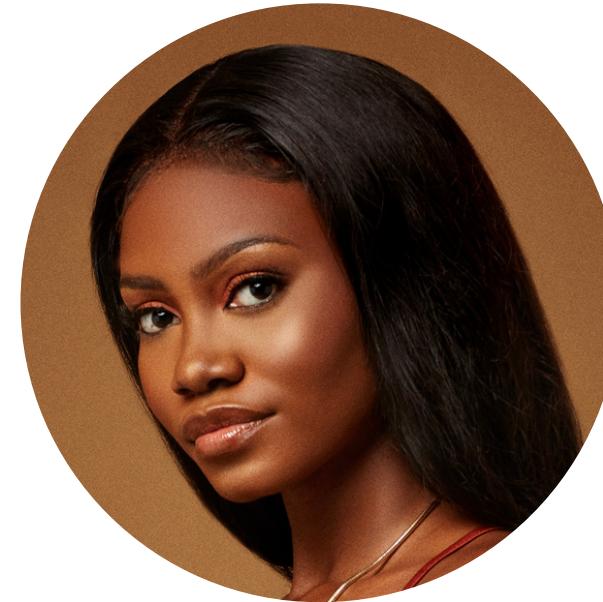
Category	City
Category_4	city_4

Cluster
cluster_0

Total sales for this week is: #657.0

Meet Our Team

AZUBI AFRICA



Violette Naa Adoley Allotey
violette.naa@azubiafrica.org

Gilbert Botchway
gilbert.botchway@azubiafrica.org

Alberta Cofie
alberta.cofie@azubiafrica.org



Emmanuel Ikogho
emmanuel.ikogho@azubiafrica.org

Adrien Nkurikiyumukiza
adrien.nkurikiyumukiza@azubiafrica.org

Twizeyimana Eric
twizeyimana.eric@azubiafrica.org

Quotes Today

"Predicting sales with machine learning: Where algorithms become modern-day fortune tellers."



AZUBI AFRICA

Thank You

GROUP: REFORMATION