Sales Forecasting Report

Data from reduced_train.csv

Store	Dept	Date	Weekly_Sales
1	1	2010-02-05	24924.5
1	1	2010-02-12	46039.49
1	1	2010-02-19	41595.55
1	1	2010-02-26	19403.54
1	1	2010-03-05	21827.9

Data from reduced_features.csv

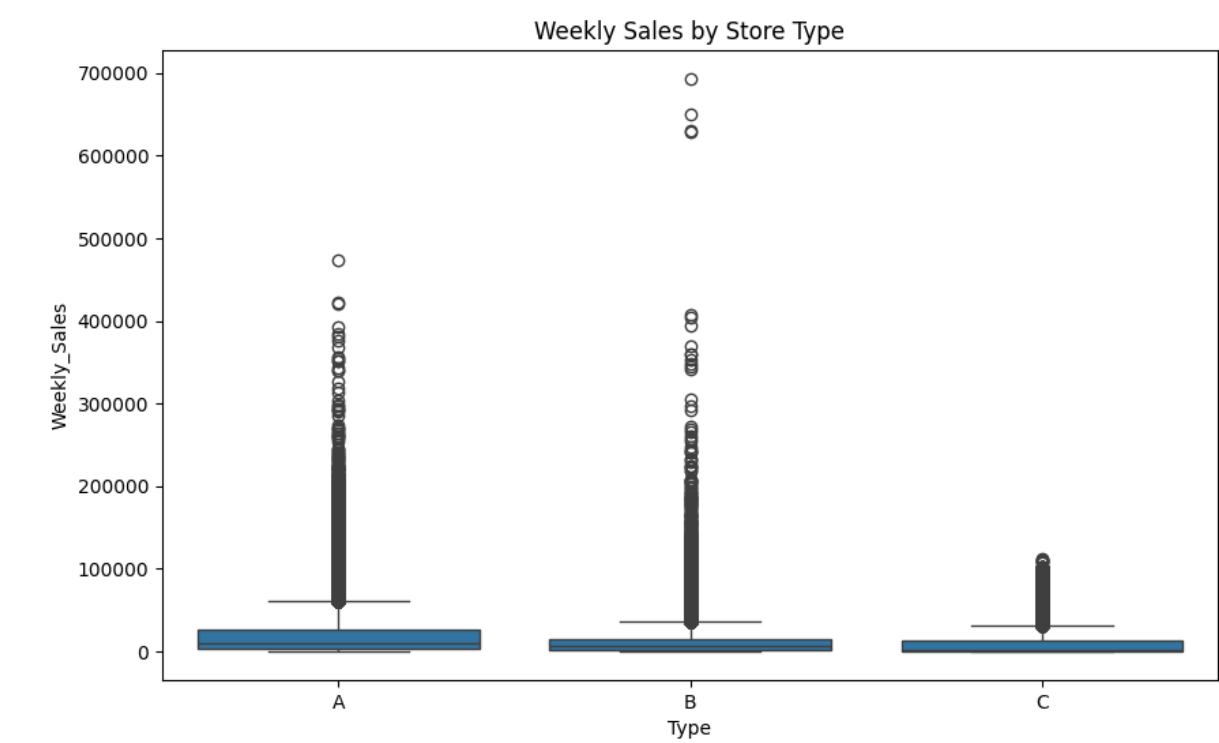| Store	Date	Temperature	Fuel_Price
1	2010-02-05	42.31	2.572
1	2010-02-12	38.51	2.548
1	2010-02-19	39.93	2.514
1	2010-02-26	46.63	2.561
1	2010-03-05	46.5	2.625

Chat History:

Q: what the columns that are present in my file?

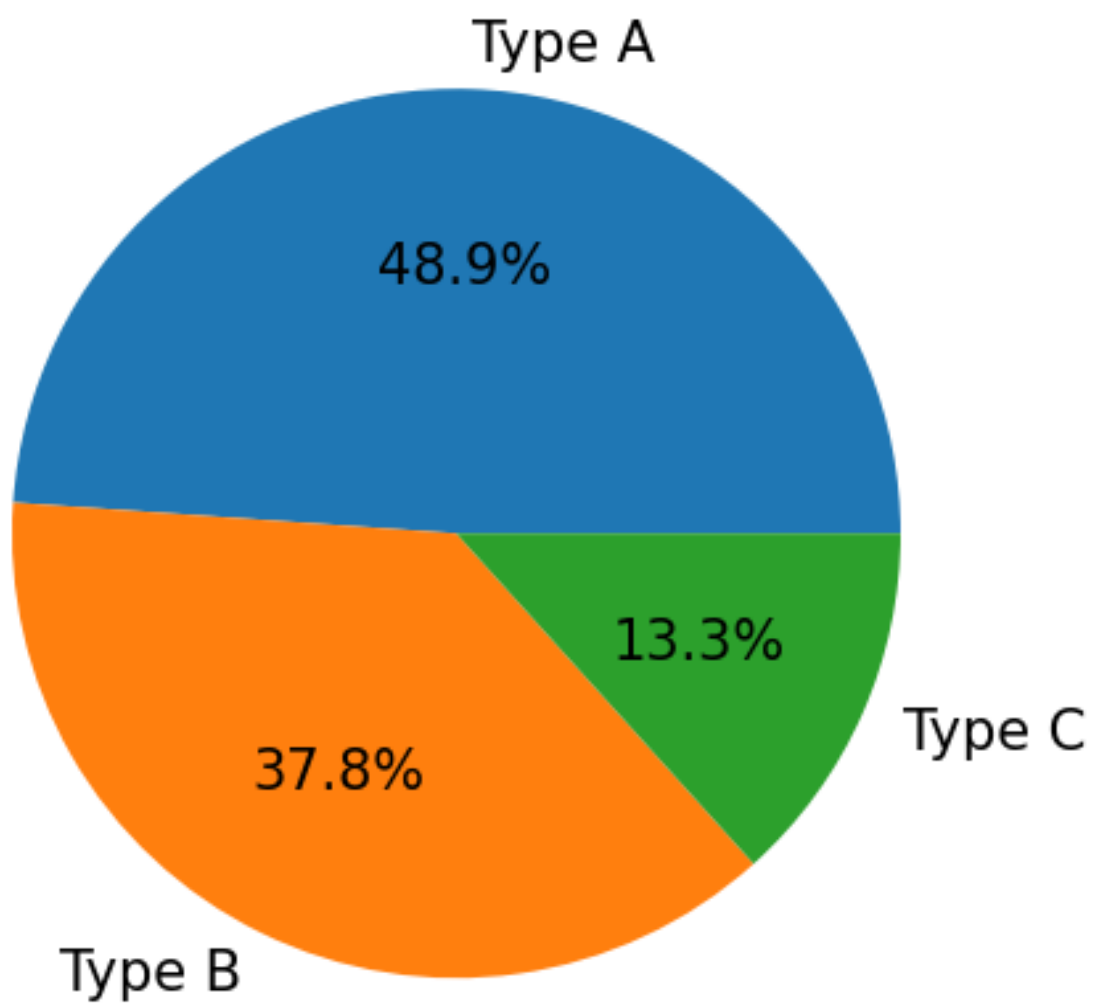A: Based on the given pieces of context, it appears that the file contains the following columns:

* Store (appears twice)

* Dept (appears twice)

* Date (appears twice)

* Weekly_Sales

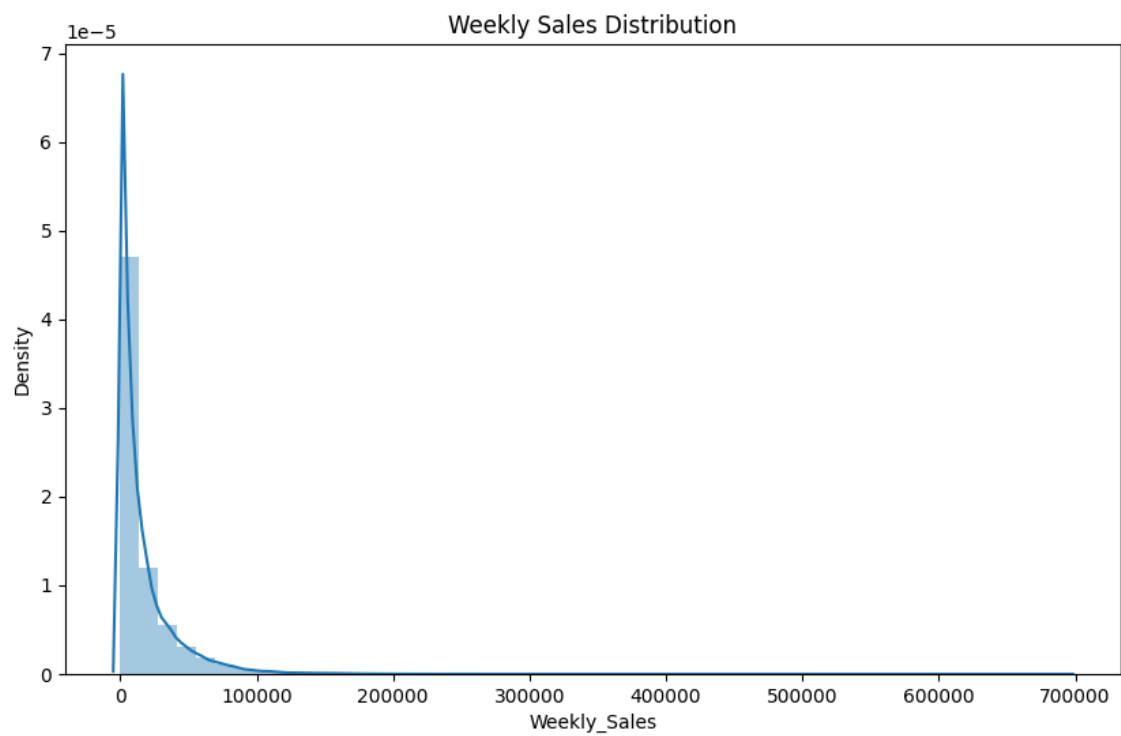Please let me know if you have any other questions.
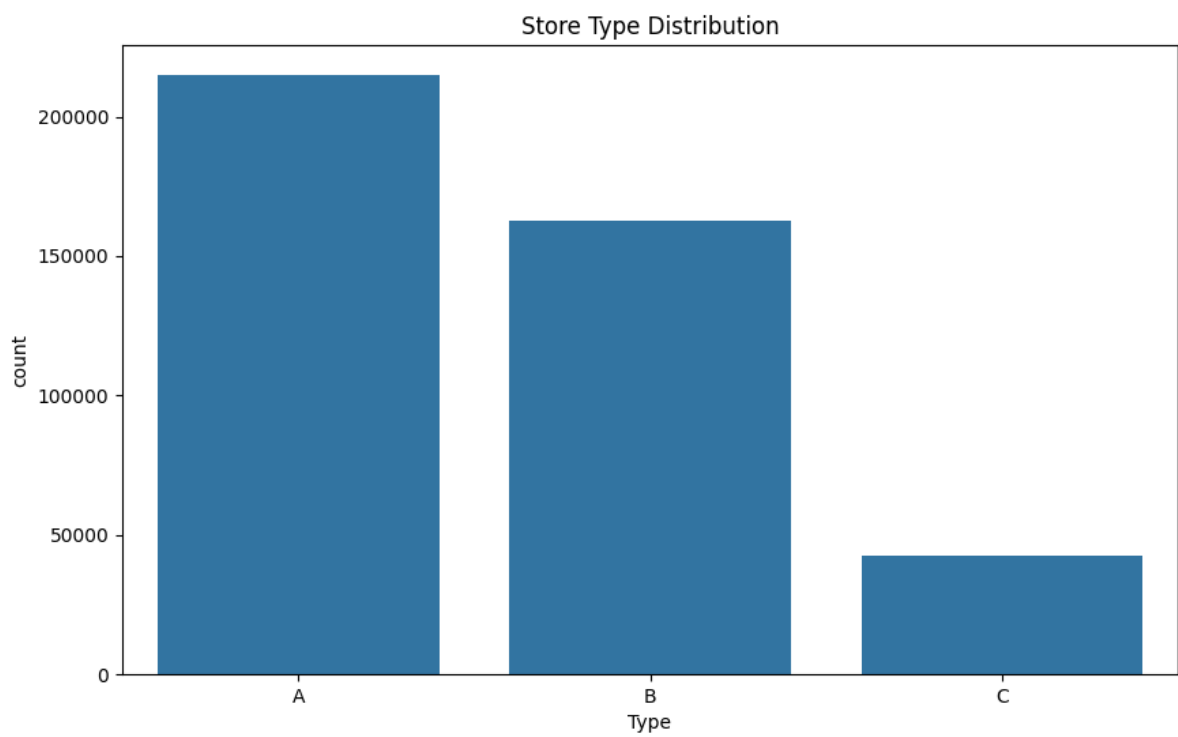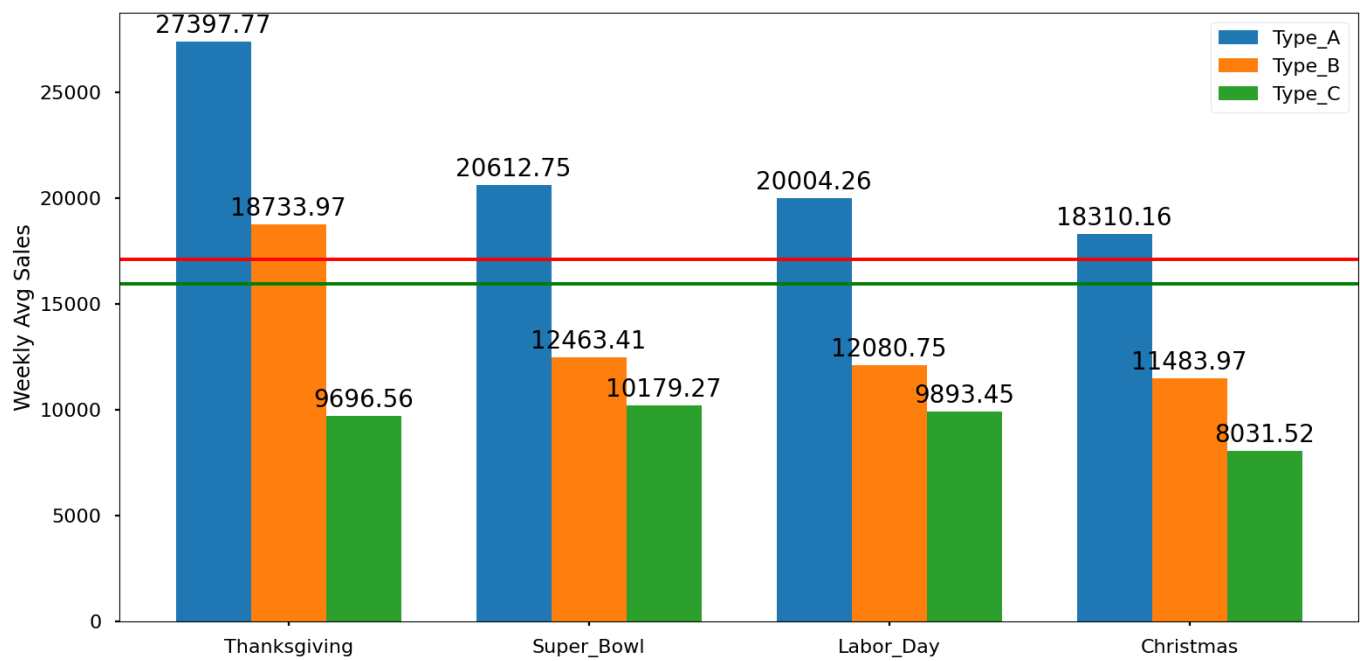

Exploratory data analysis:

sales_by_store_type.png

types_of _stores.png
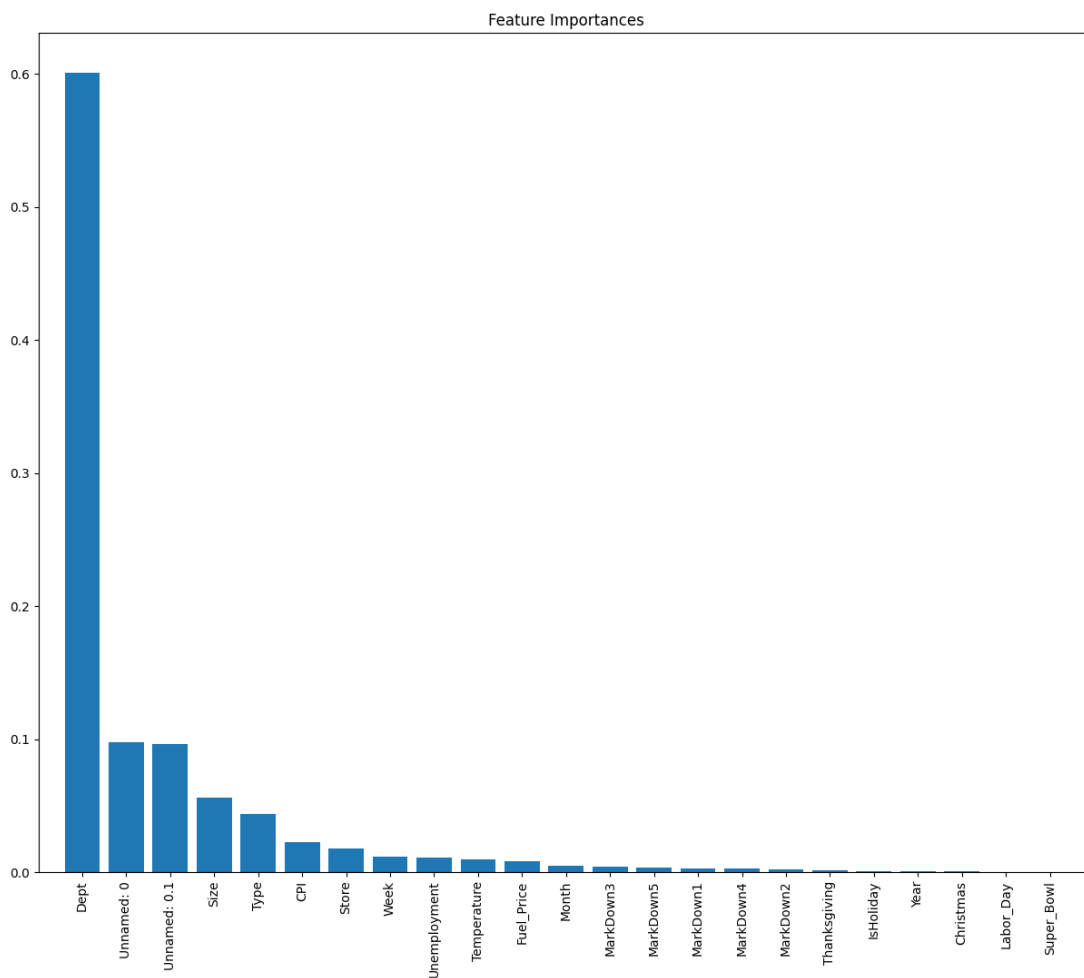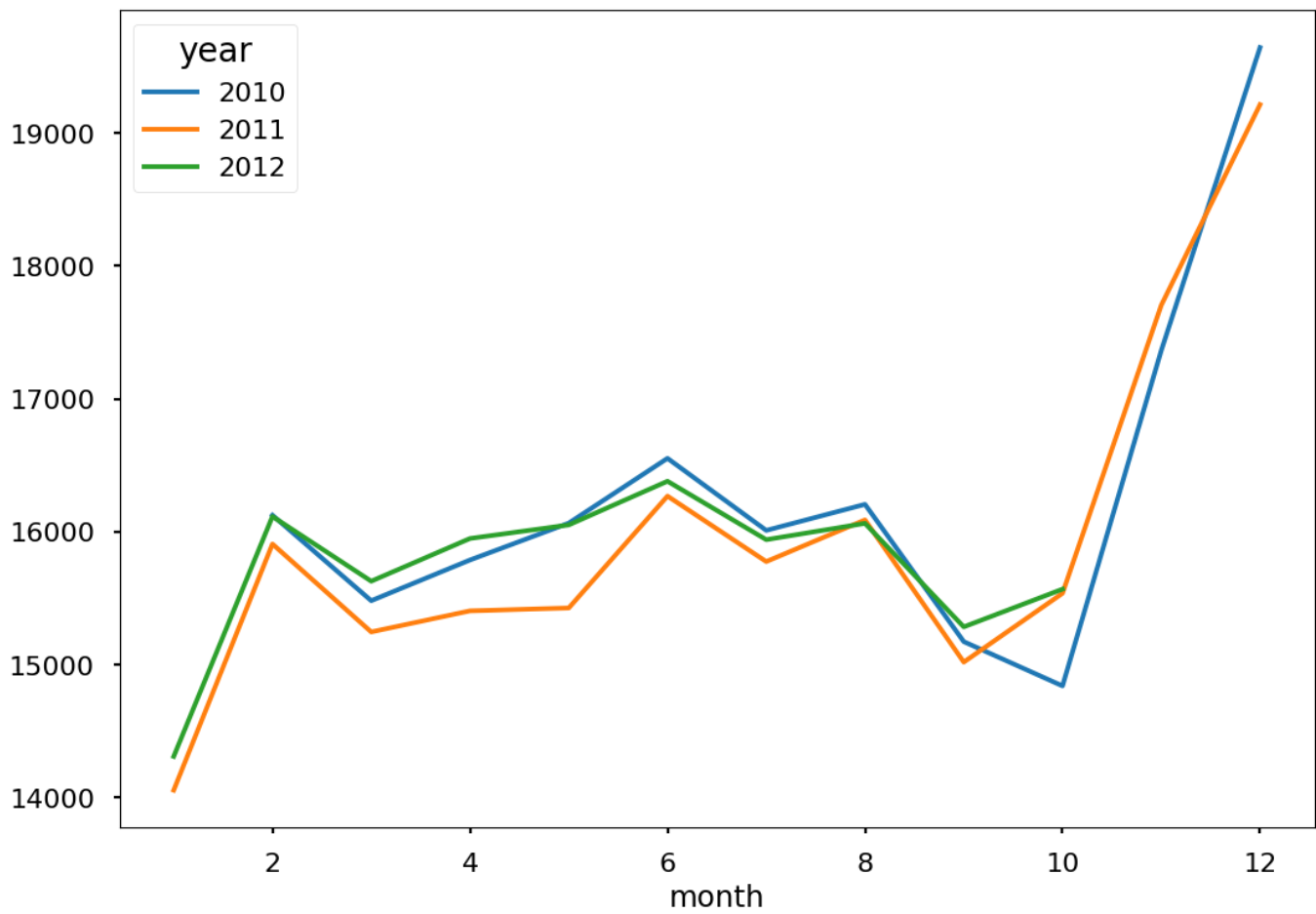
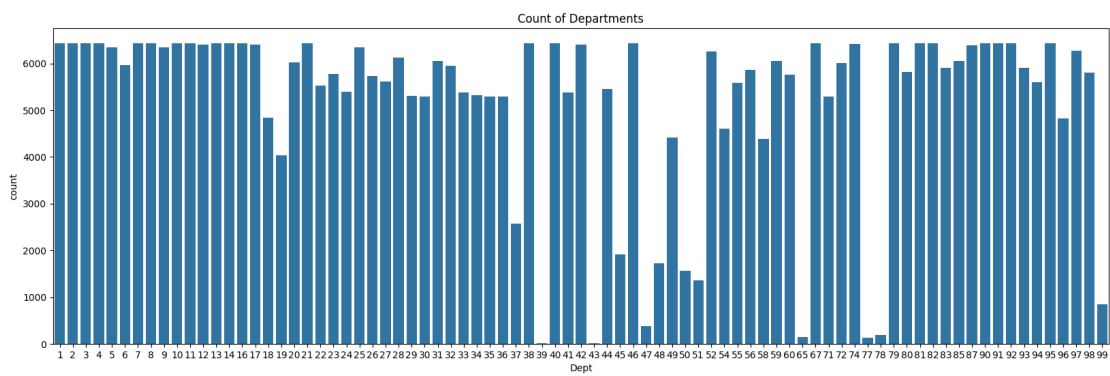weekly_sales_distribution.png

store_type_distribution.png
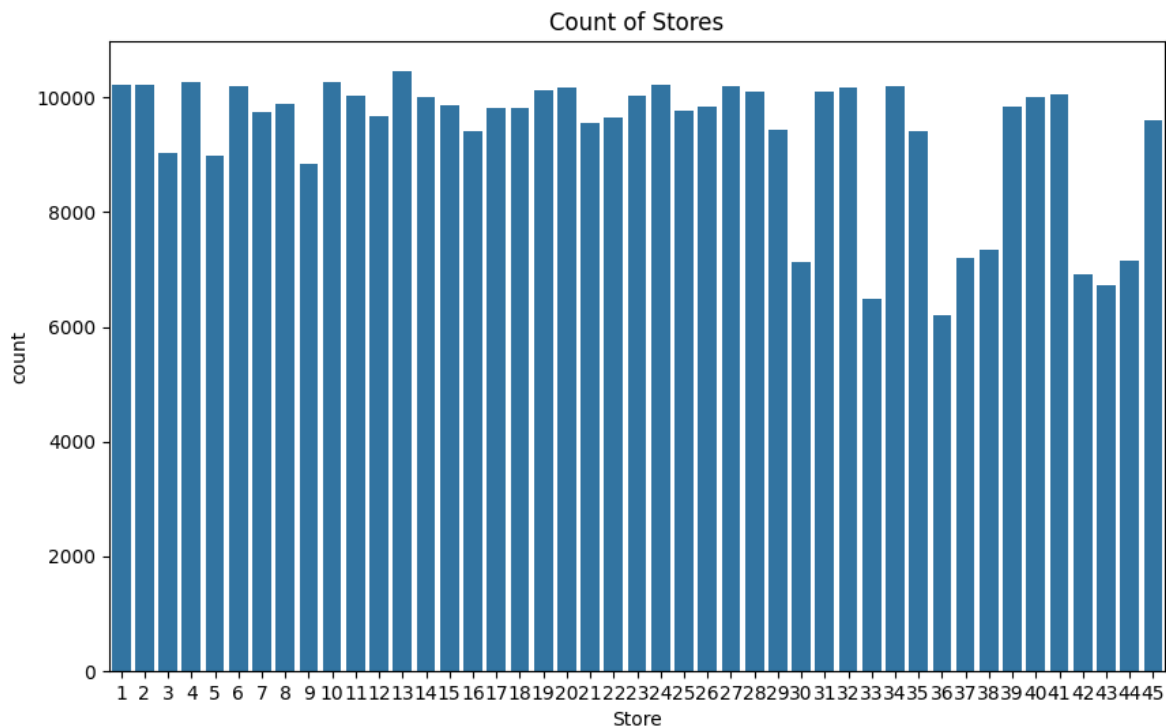
sales_analysis_with_holidays.png

Feature Importances

feature_importances.png

sales_data.png

dept_count.png

store_count.png

Results:

Model	WMAE
RandomForest	13693.063053913578
GradientBoosting	7864.814057426034
XGBoost	7459.980518710376
Lasso	12343.5881535082
RandomForest with feature selection	1801.9445760608