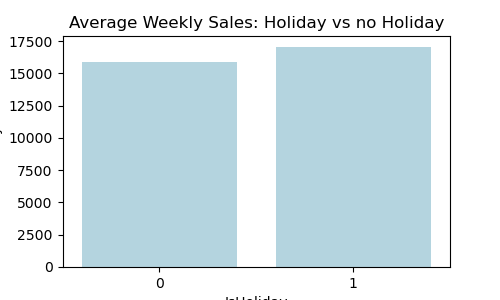
**Key Insights about Walmart Sales Data Analysis:**

The average weekly sales for holiday weeks is higher than the average weekly sales for non-holiday weeks.



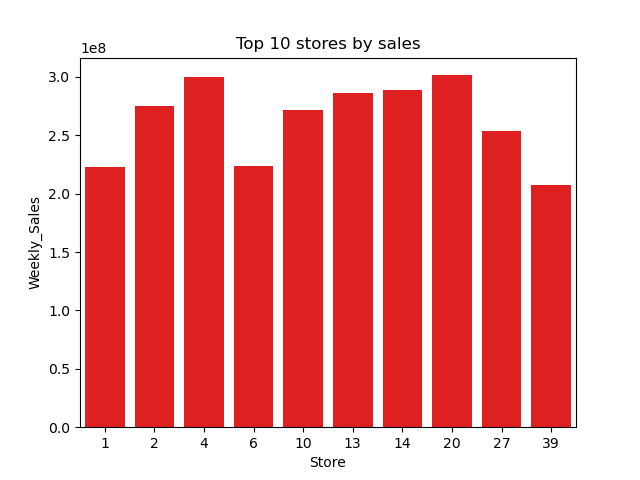
**Departments:**

Departments 92, 95, and 38 generate most weekly sales and have highest weekly sales among all other departments.

Also, these top 3 departments mentioned above, have statistically different weekly sales on the average. (p-value < 0.05).

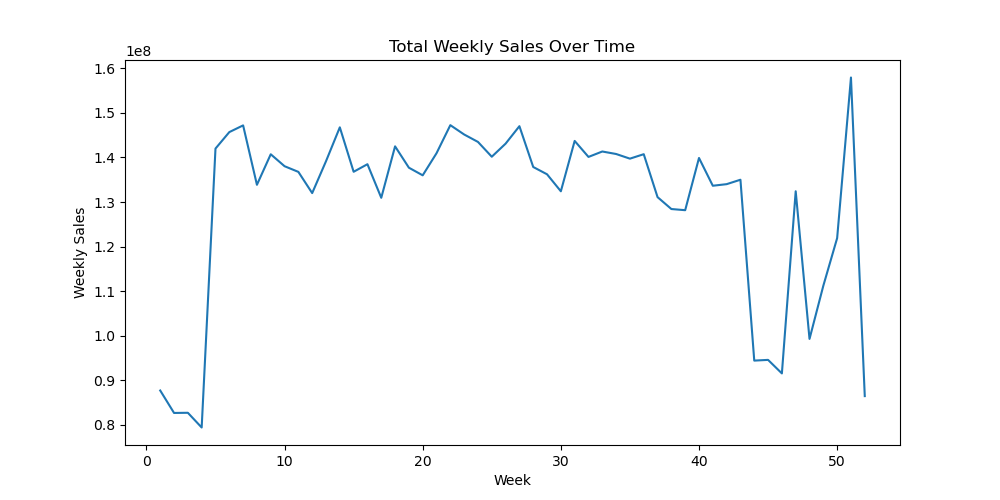
**Stores:**

Store number 20, 4, and 14 generate the most weekly sales among all other stores.



**Overall weekly sales trend:**

Weekly sales tend to increase at the start of the weeks of the year.



**Type-A, Type-B, and Type-C stores:**

The average weekly sales for all three types of stores are found to be statistically different from zero, that means that the average weekly sales for 3 types of stores are not same. (p-value < 0.05).

**Predictive Modeling Summary**

To forecast department-level weekly sales across Walmart stores, I implemented a supervised machine learning approach using both **Linear Regression** and **Random Forest Regressor**. The dataset was first preprocessed by merging train.csv, features.csv, and stores.csv using left joins on Store and Date, ensuring that relevant features like temperature, fuel price, store type, and markdown data were aligned properly. Null values in the markdown columns were replaced with zeros, and categorical variables (such as Type) were converted to dummy variables. Feature engineering included creating binary indicators for holiday promotions (Promo\_Holiday) and aggregating markdown information.

The dataset was then split into input features (X) and target variable (Weekly\_Sales). I first trained a **Linear Regression** model, which achieved an R² score of **0.974** and a Mean Absolute Error (MAE) of approximately **14,560**, establishing a strong baseline. To improve predictive performance and capture non-linear relationships, I trained a **Random Forest Regressor** with 100 trees. This model showed improved accuracy and robustness to outliers. Feature importance analysis revealed that **Department**, **Store**, and **Size** were the most influential variables driving weekly sales, while markdowns and holiday flags contributed less than expected — likely due to missing values or limited overlap with sales spikes.

After validating the model, predictions were made on the unseen test.csv dataset, and the results were concatenated with the original test data to produce the final output containing the predicted weekly sales.