Walmart weekly sales mid-term project

Yingmai Chen

I.Abstract

The project aims to find the relationship between weekly sales of walmart with other variables,and also want to find which variable is more correlated with weekly sales,and the project will include six parts:“Abstract”,“Introduction”,“method”,“results”,“discussion”,“appendix”. The first two parts are the descriptions of this project,and the method part would have two parts:EDA(Exploratory Data Analysis),and model.The results would be the comparison and illustration of the model.Discussion part would discuss some future questions about my project,appendix would include some poorly visualized graphs.

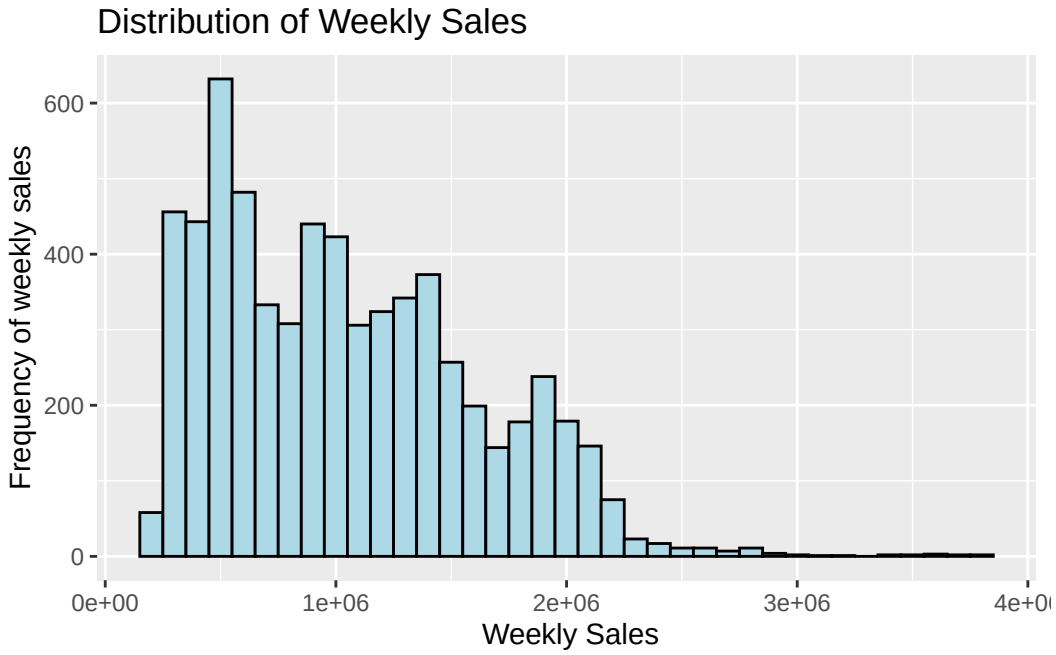
II.Introduction

In this study, we leveraged a dataset from Kaggle to identify and quantify various factors impacting weekly sales. The dataset includes key variables such as holidays, oil prices, temperature, unemployment rates, and the Consumer Price Index (CPI), we utilized regression analysis to assess the relative impact of these variables on sales volumes.Through this study, we hope to offer a more comprehensive and accurate sales forecasting model, helping businesses better understand market dynamics and make more effective business decisions.

III.Method

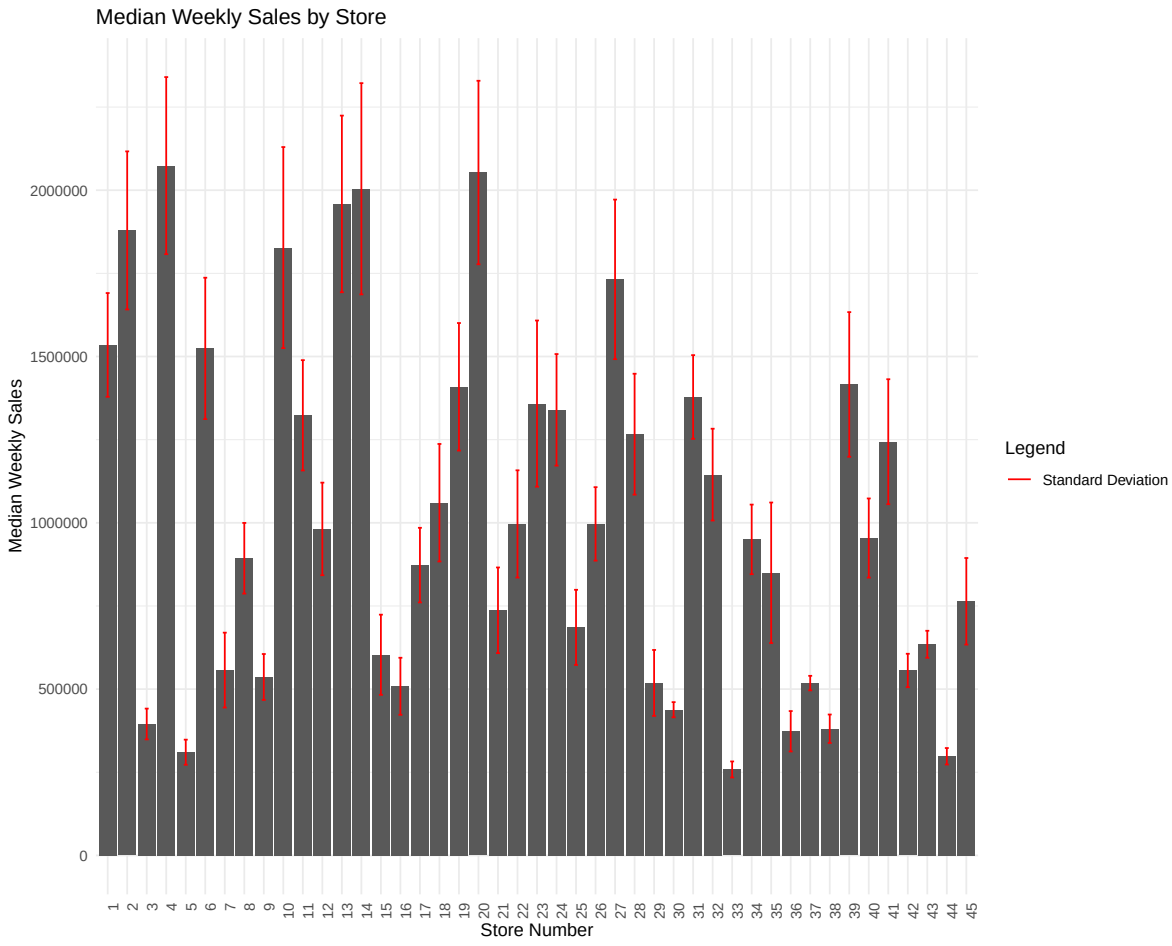
I check the summary of the data first and see there don’t have NA’s in the data,so we don’t need to deal with the missing value of data.

EDA



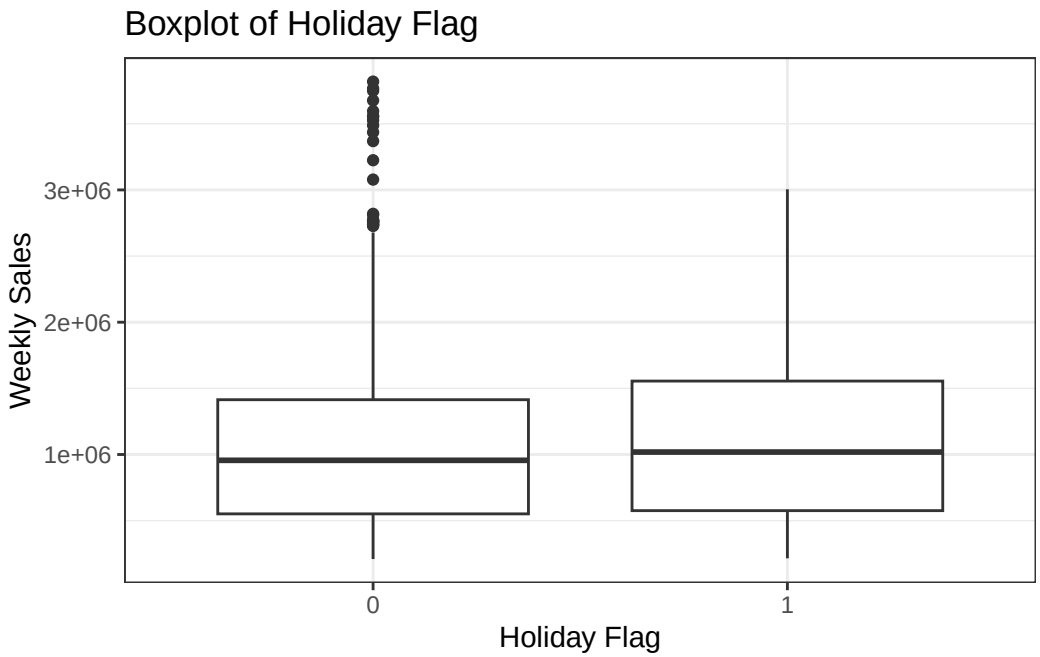
description

From this plot,we can see that there are fewer weeks with very high sales compared to weeks with low sales.This is typical for sales data where a small number of periods (like holiday seasons) might have exceptionally high sales.



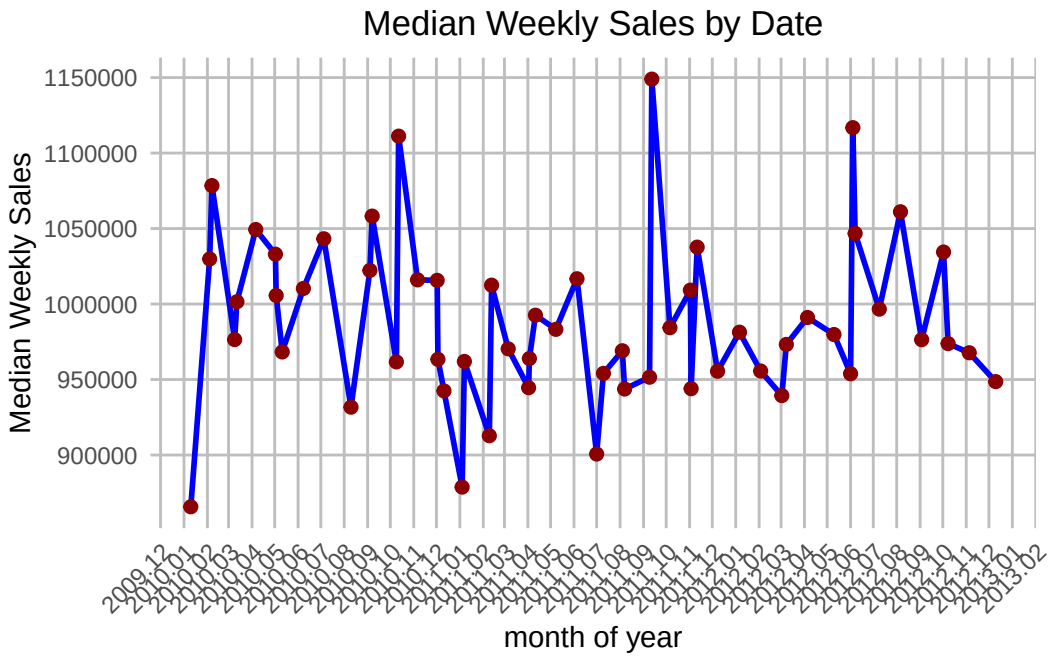
description

From this plot,we can see there exist obvious differences between each store,some has higher median of weekly sales, and some would have bigger error bar which means sales are more volatile.this suggests that we should trying multilevel model in the model part.



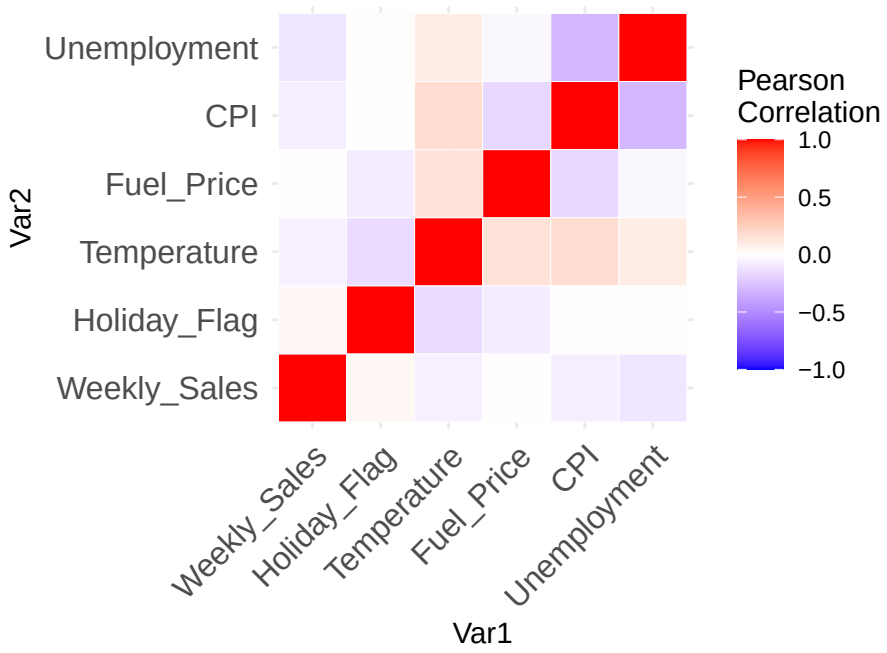
description

From this plot, we can see that the median of holiday weekly sales is a little bit higher than non-holiday weekly sale, and the numbers of outliers in non-holiday weekly sales are more than holiday-weekly sales,based on this,I think there would have some relationships between holidays and weekly sales.



description

From this plot,we can see that there is a recurring trend of lower sales at the beginning of each year. This dip in sales could be attributed to post-holiday season effects, where consumer spending typically drops following the end-of-year holidays,which means the date would have impact on weekly sales.



description

As we can see, from this plot,It is not obvious that these variables are correlated to weekly sales.

model

Because we can not easily draw a conclusion by doing EDA,so the next step for us shoud be modeling,

Null model

complete pooling model

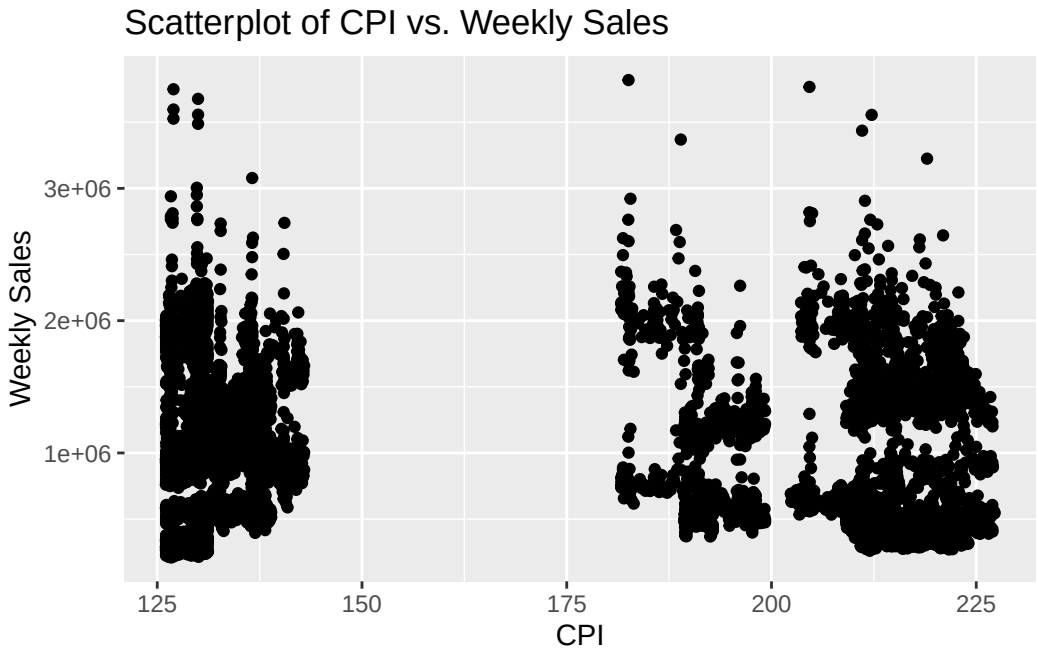
No pooling model

generalized linear Model

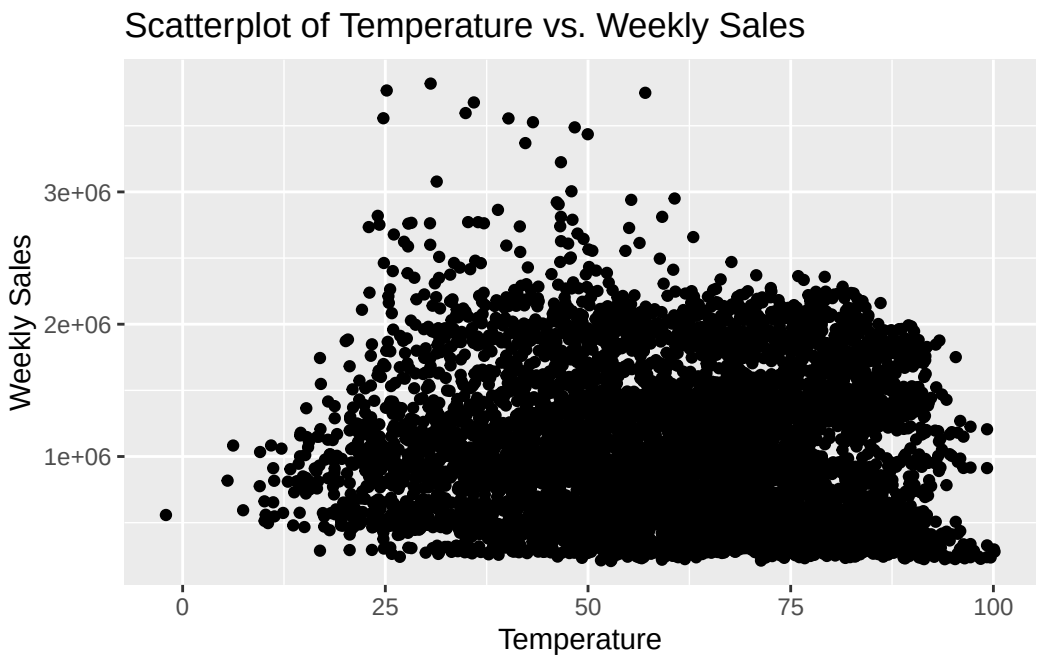
IV.result

V.discussion

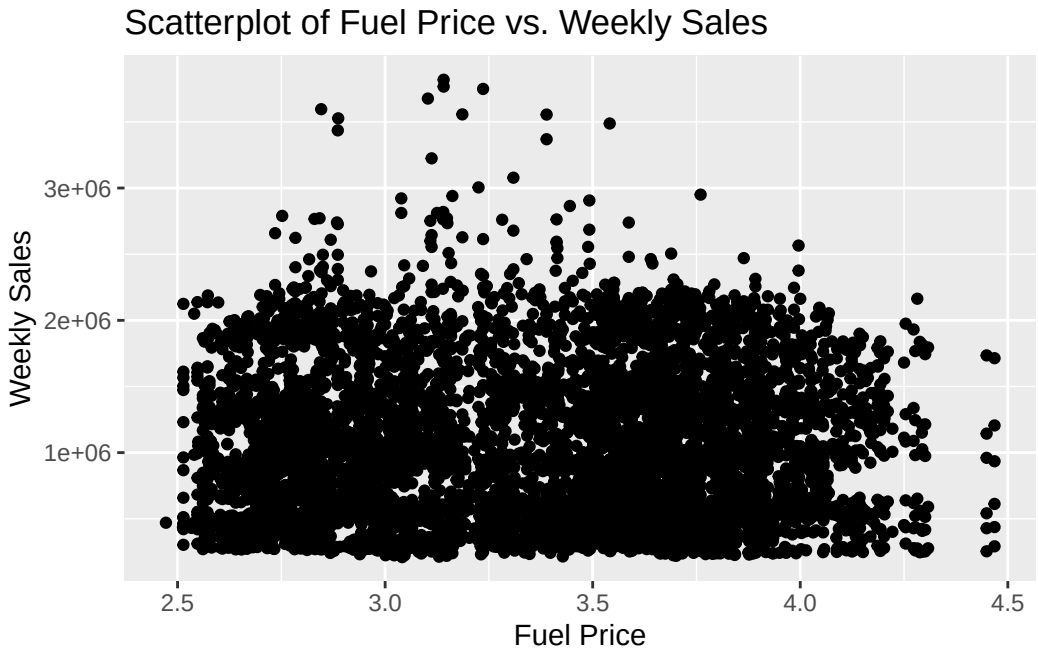
VI.appendix



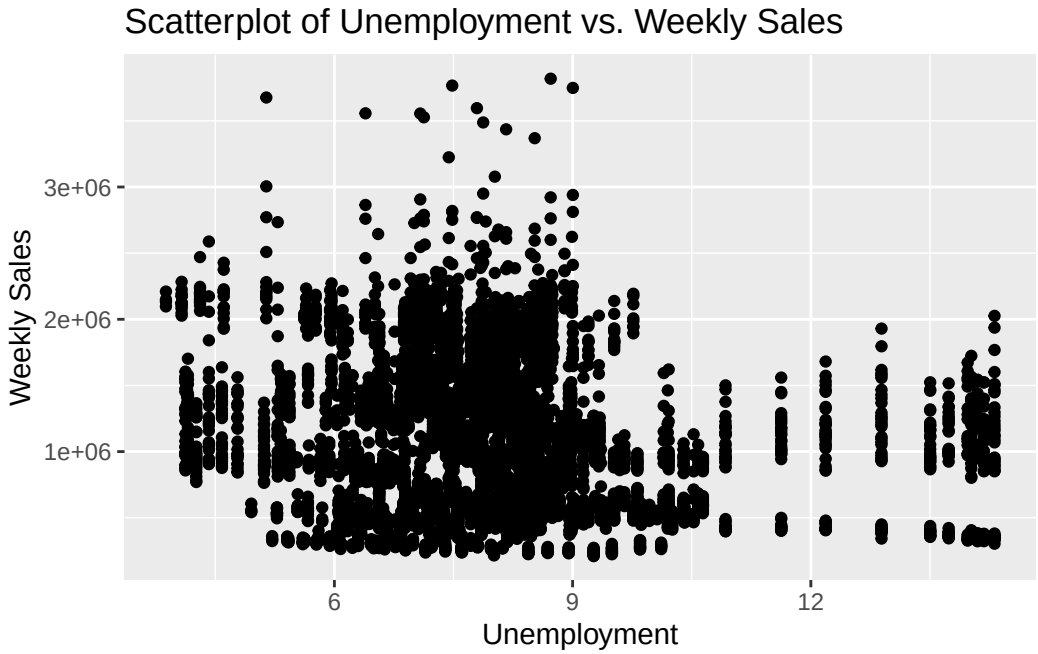
From this plot, we can hardly see if there exist relationships between CPI and weekly sales.



From this plot,we can see temperatures from 25 to 75 of the country tend to have much more numbers of higher weekly sales.



From this plot, we can hardly see if there exist relationships between fuel price and weekly sales.



From this plot,we can see the less of proportions of unemployment tends to have higher weekly sales.