Code: https://www.kaggle.com/rivuletnriver/code

The methods I used during the process were: checking nan values and distribution of numerical features, merging training data with features and stores, creating visualization on some of the features using heatmap, pairplot, and direct line plot of feature & target vs time. And some of the findings include the following:

1. Features don't seem to have any correlation with each other, which is good for training

2. Markdown and CPI values tends to be near zero when it's not holiday

3. Most features have significantly larger std when it is holiday

4. Not strong correlation with weekly_sales for any of these features

5. MarkDown 1 and MarkDown 4 are somewhat positively correlated

6. Size of store seems to be weakly correlated with Weekly Sales, Markdown 1 and 5

7. Weekly Sales has a weak, negative correlation with Fuel Price after 2010-09

8. Unemployment rate starts to drop significantly at around 2011-01, so sth happened there, since it is not an coincidence that Weekly sales reached such a high bar

9. CPI is simply increasing