Walmart Sales Forecasting

Pipitton Sanseeha

9/Jan/2023

Outline

- Executive summary
- Problem understanding
- Exploratory data analysis (EDA)
  - California
  - Texas
  - Wisconsin
  - Sales performance by state
  - Total
- Predictive analysis
- Conclusions

Executive summary

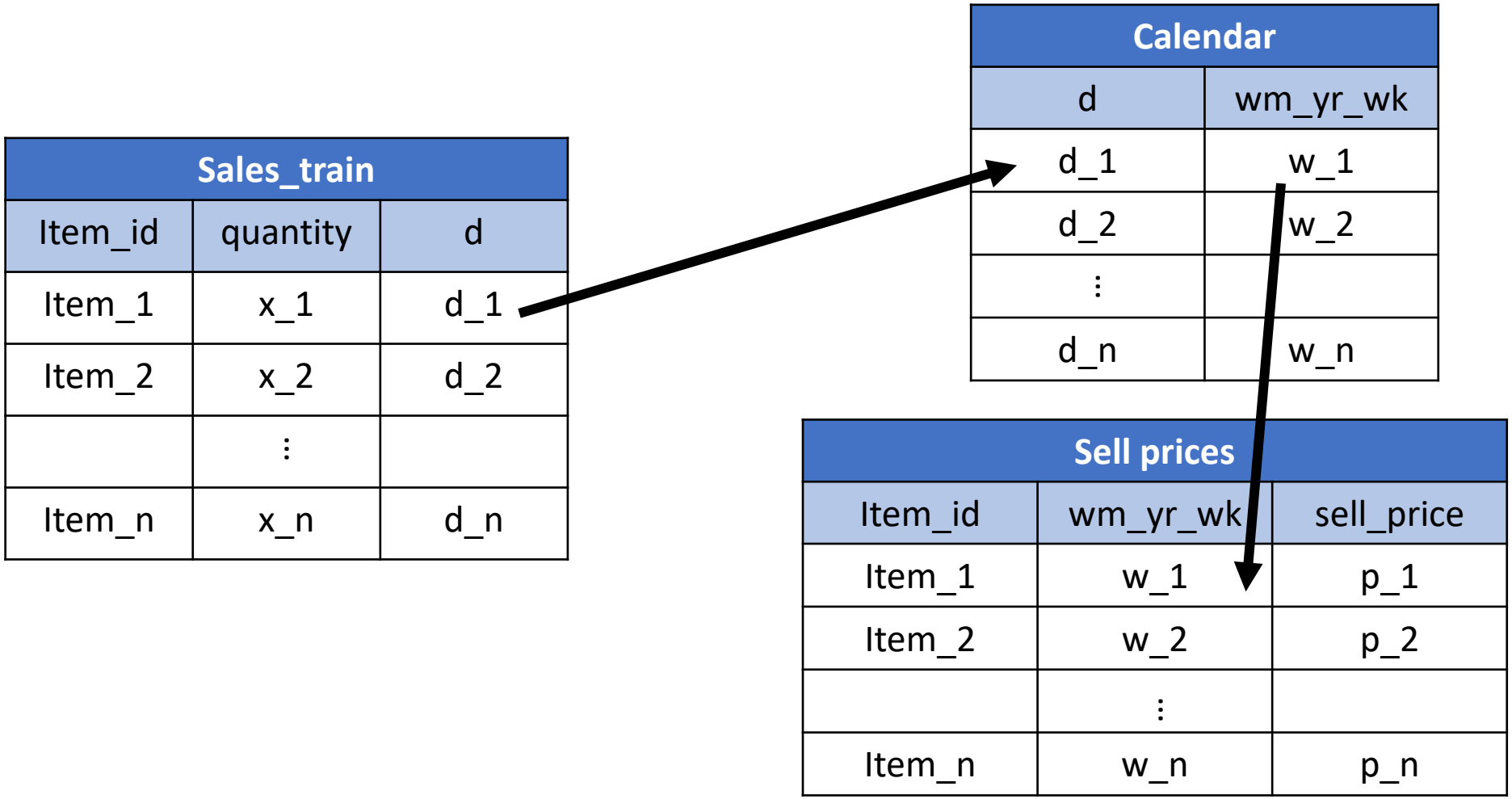
In this study, we examine Walmart sales in three states (California, Texas, and Wisconsin). First, we find business insight by doing exploratory data analysis. We found total sales depend on population density and month but do not depend on event. Finally, we employ a decision tree regression model to forecast revenues for the following 28 days. The model performs well.

Problem understanding

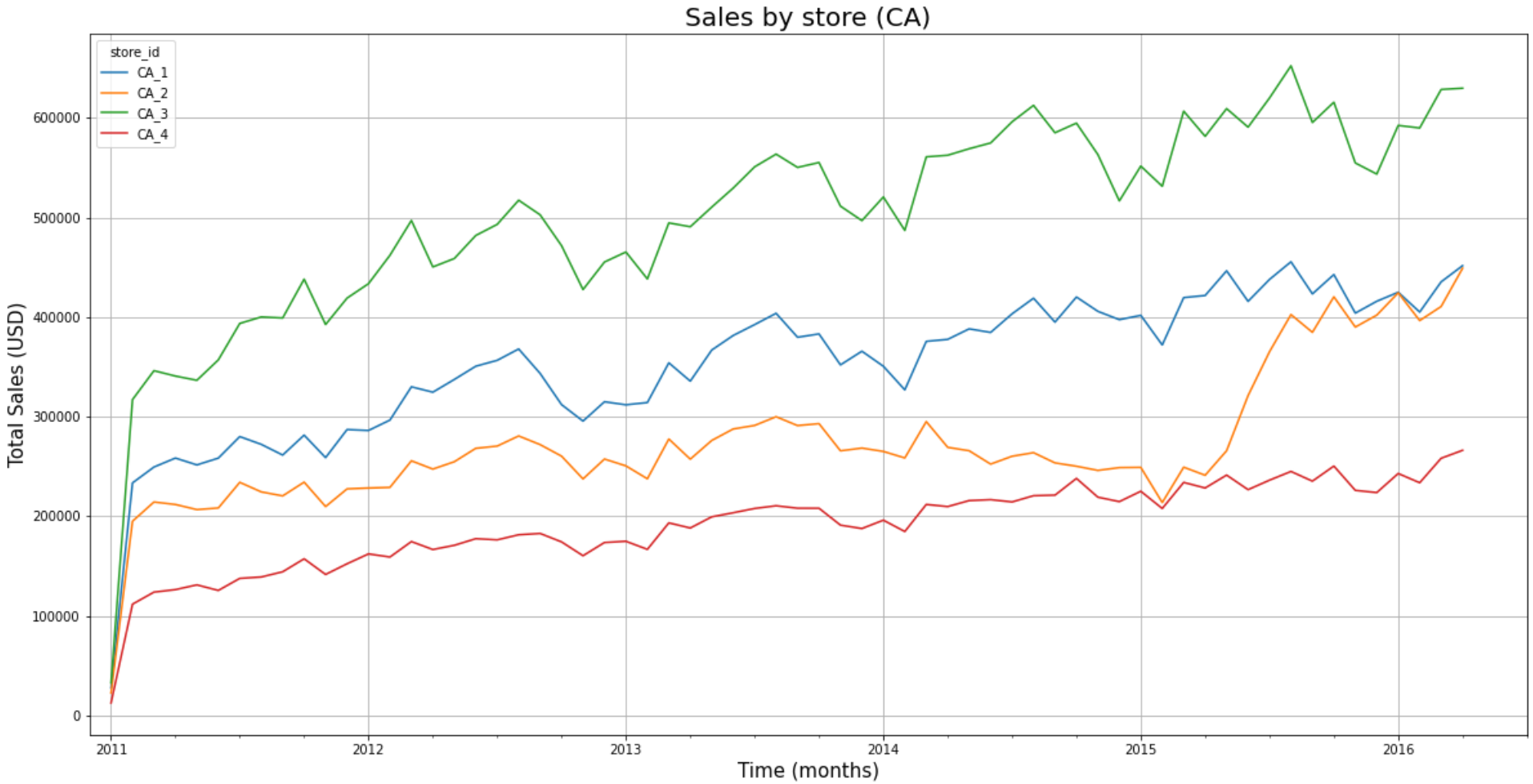
- Walmart is one of the notable retail corporation. They have provided sales data for stores in three states and includes item level, department, product categories, and store details.
- The goal of this study is to gain business insight and predict daily sales for the next 28 days.

Exploratory data analysis

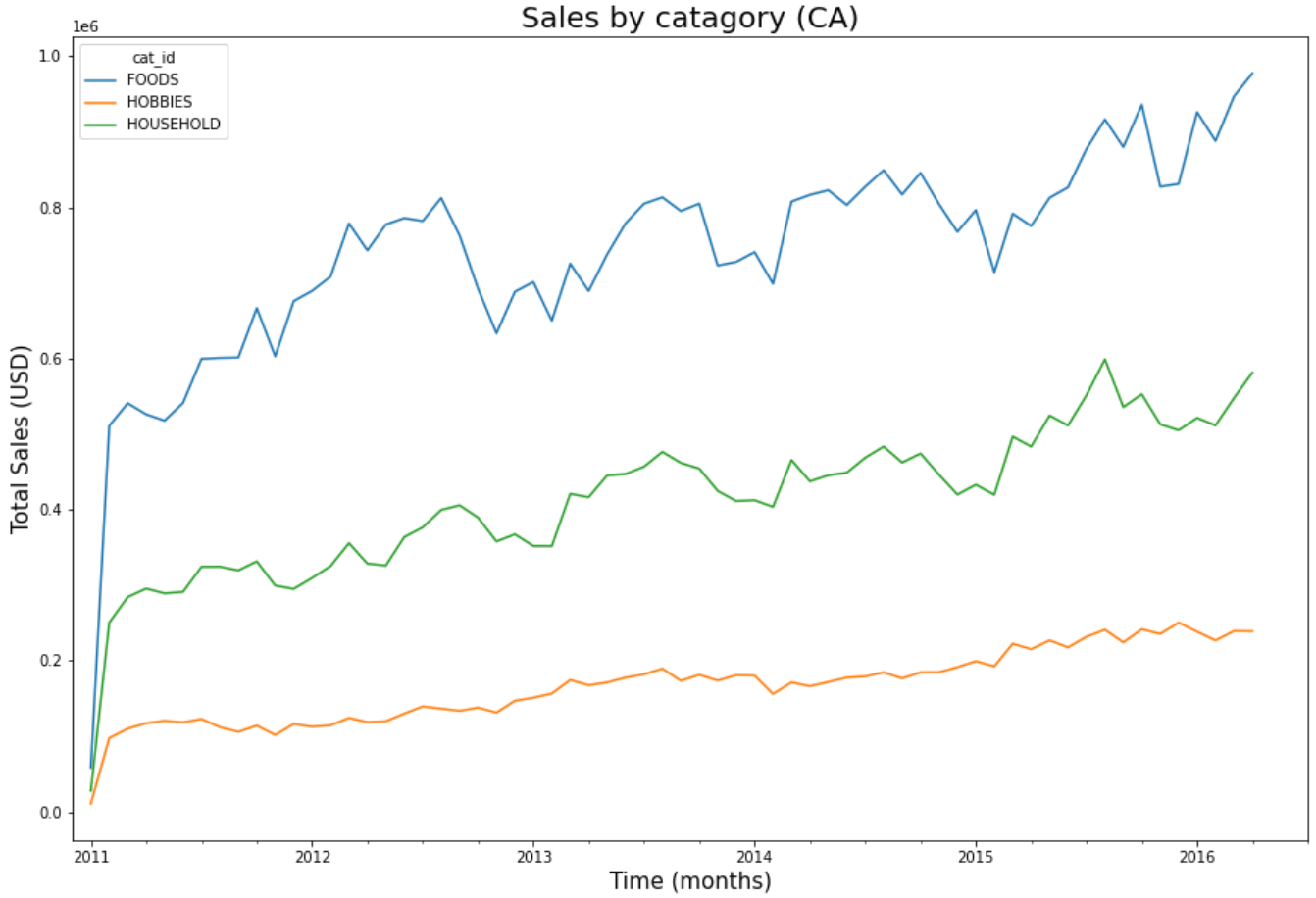
How to calculate total sales

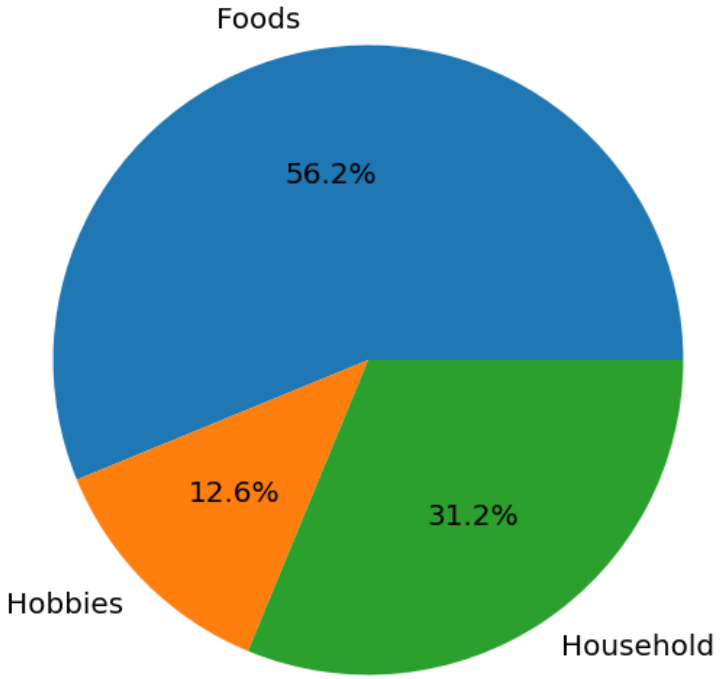


California (CA)

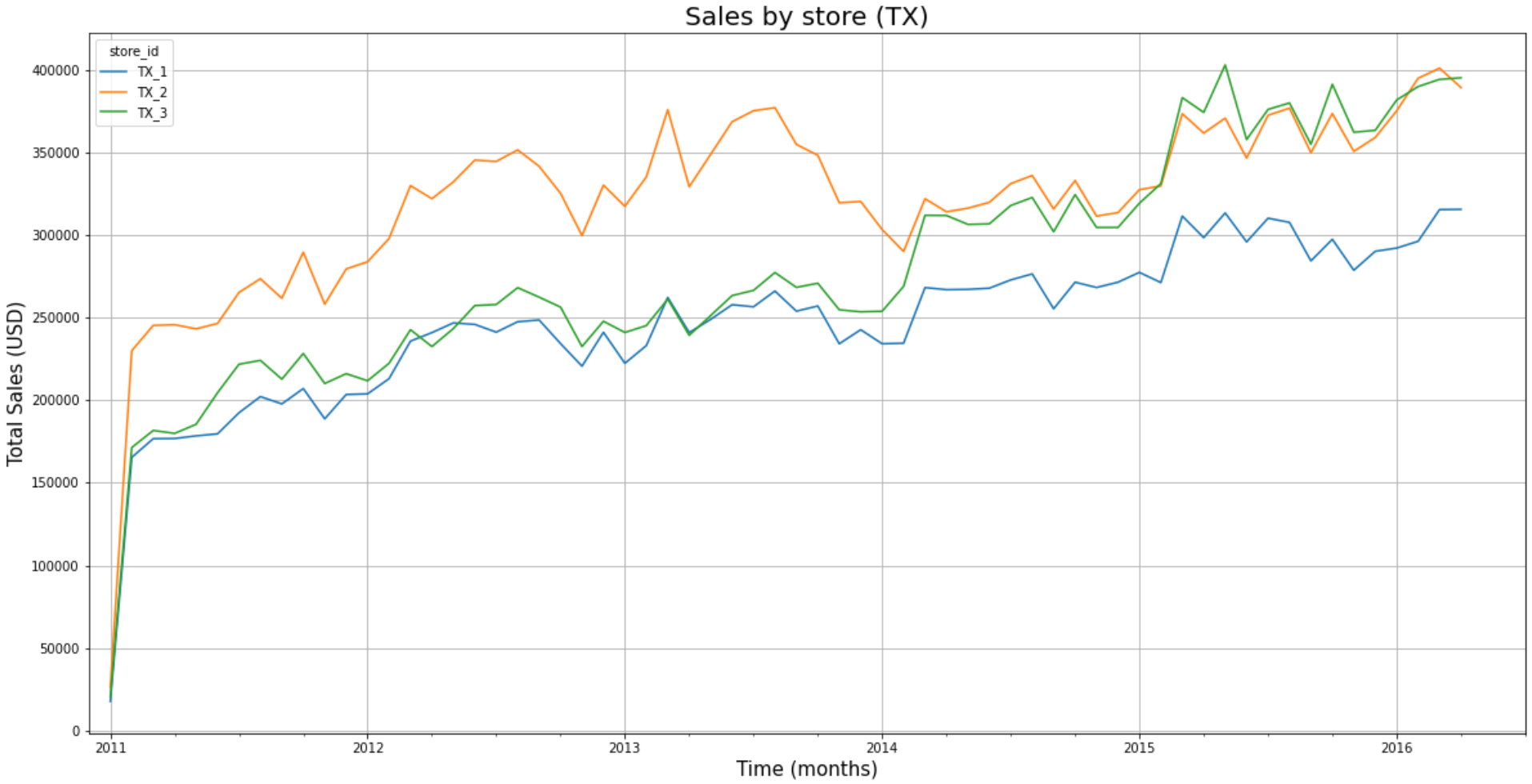


California (CA)

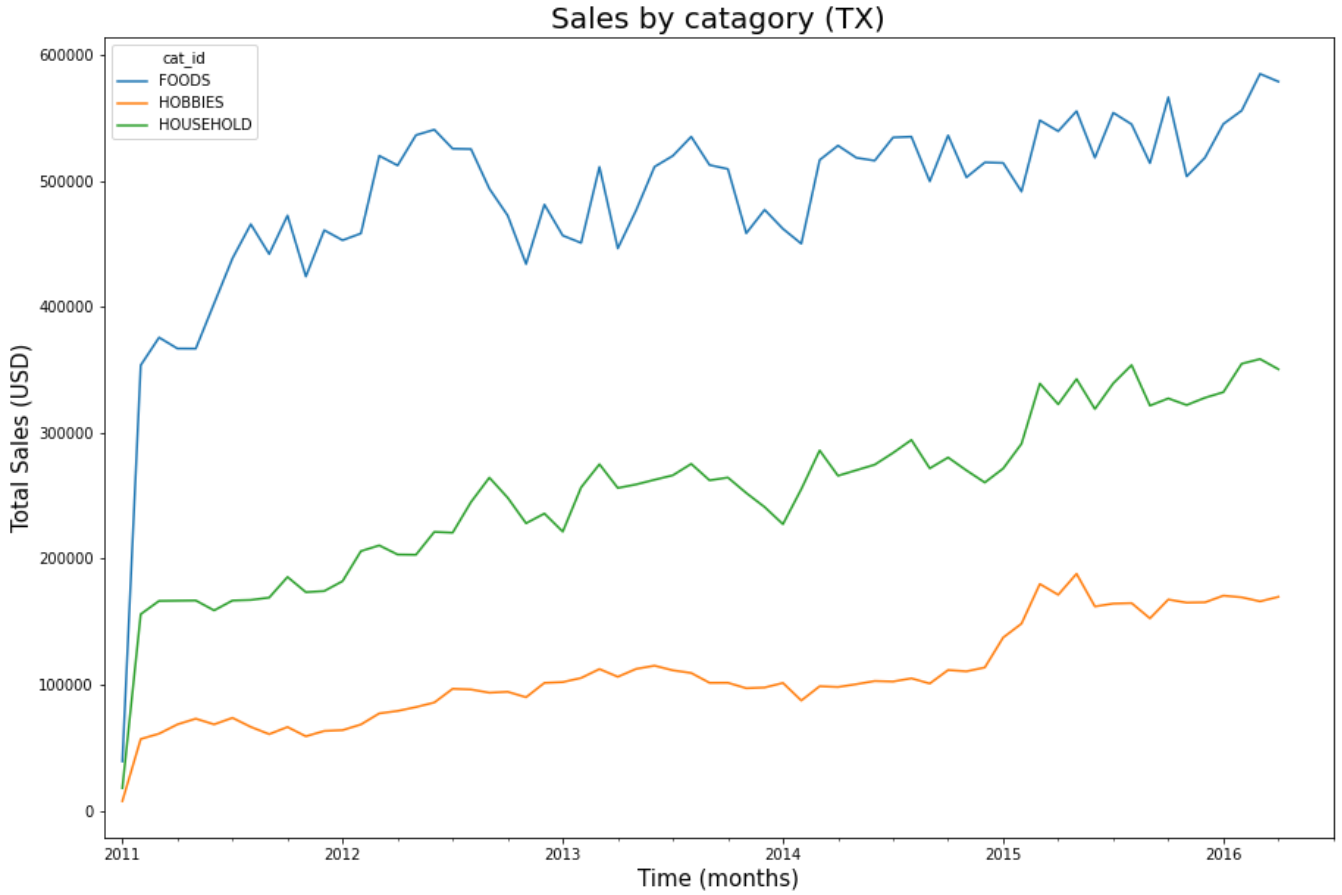


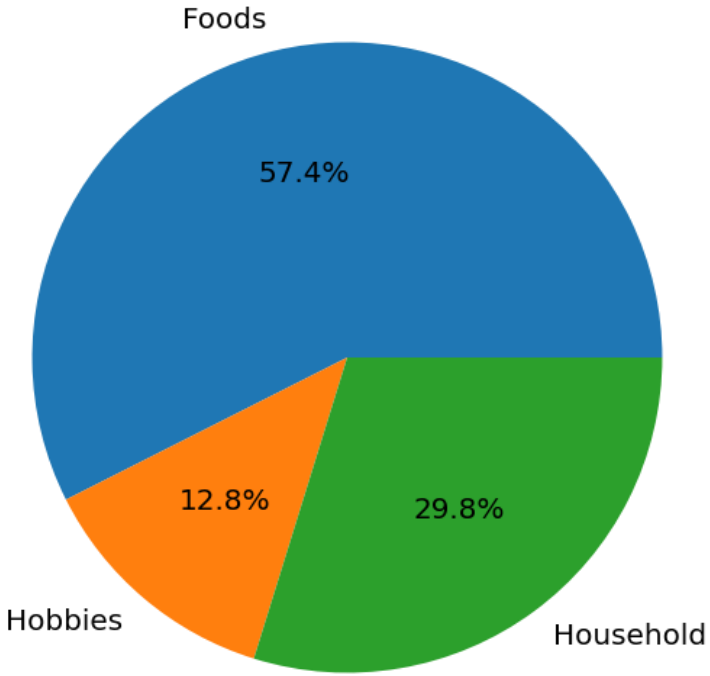


Texas (TX)


Sales by store (TX)
store_id
TX_1
TX_2
TX_3
Total Sales (USD)
0
50000
100000
150000
200000
250000
300000
350000
400000
2011
2012
2013
2014
2015
2016
Time (months)

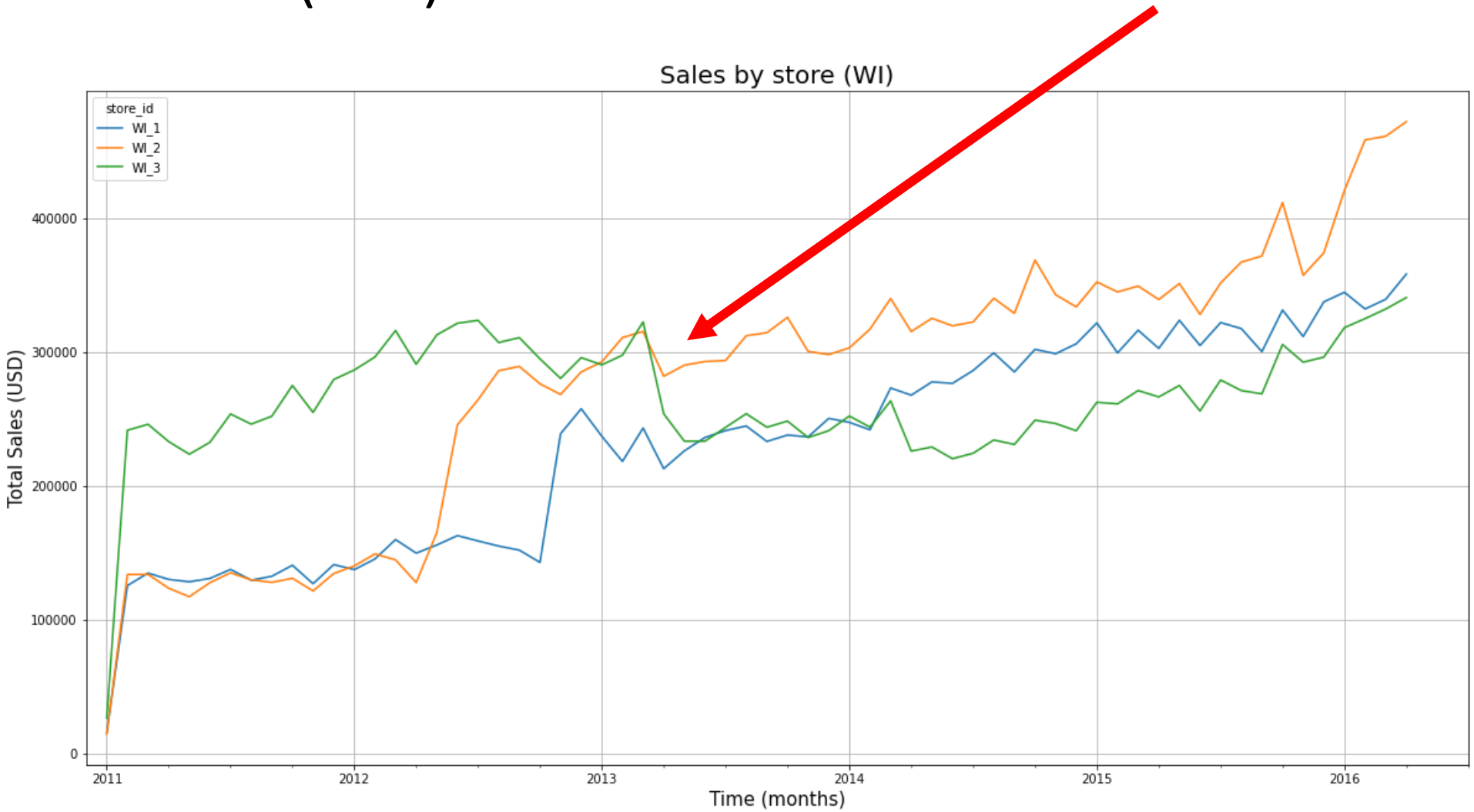
Texas (TX)


Sales by catagory (TX)
cat_id
FOODS
HOBBIES
HOUSEHOLD
Total Sales (USD)
600000
500000
400000
300000
200000
100000
0
2011
2012
2013
2014
2015
2016
Time (months)

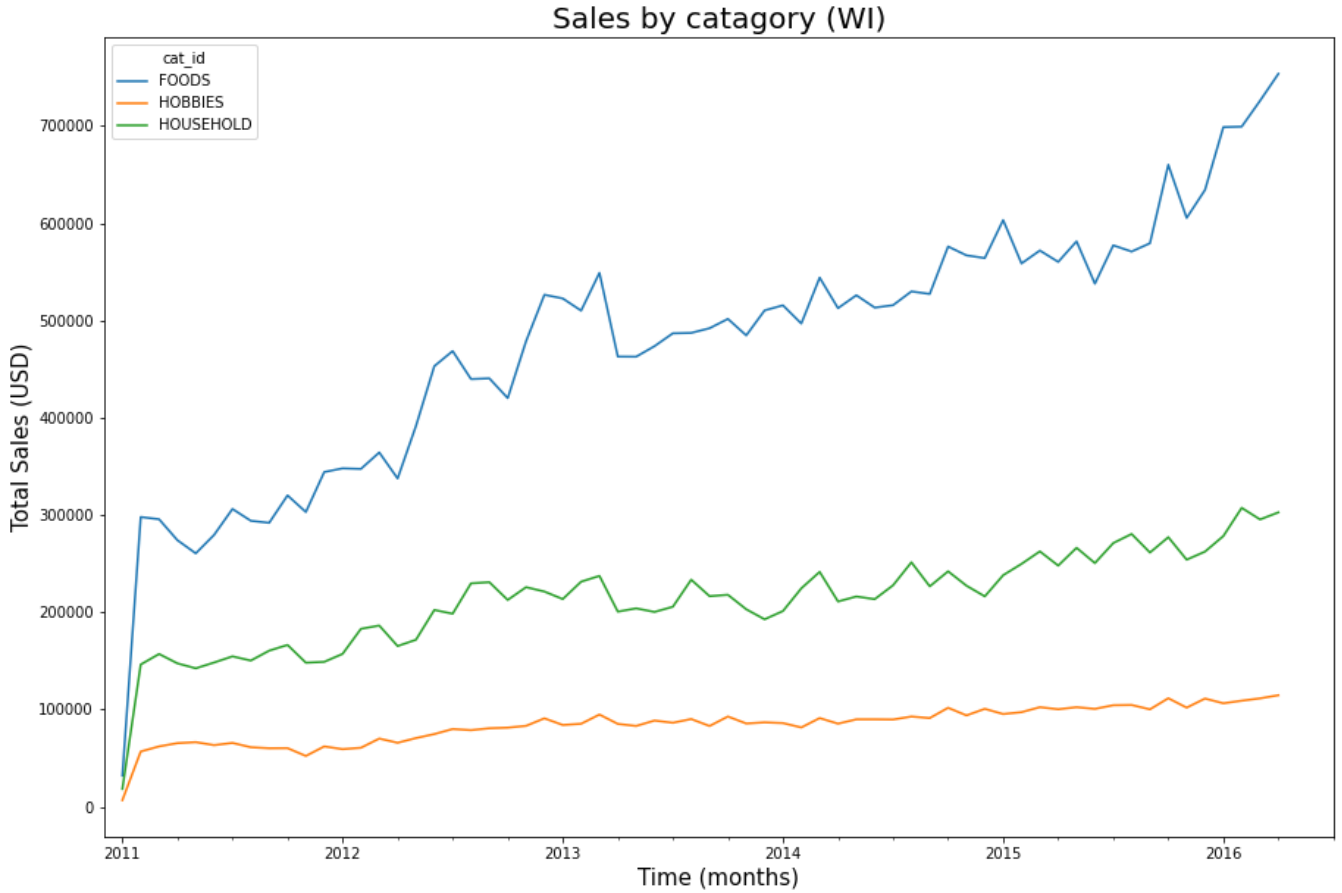

Foods
57.4%
12.8%
29.8%
Hobbies
Household

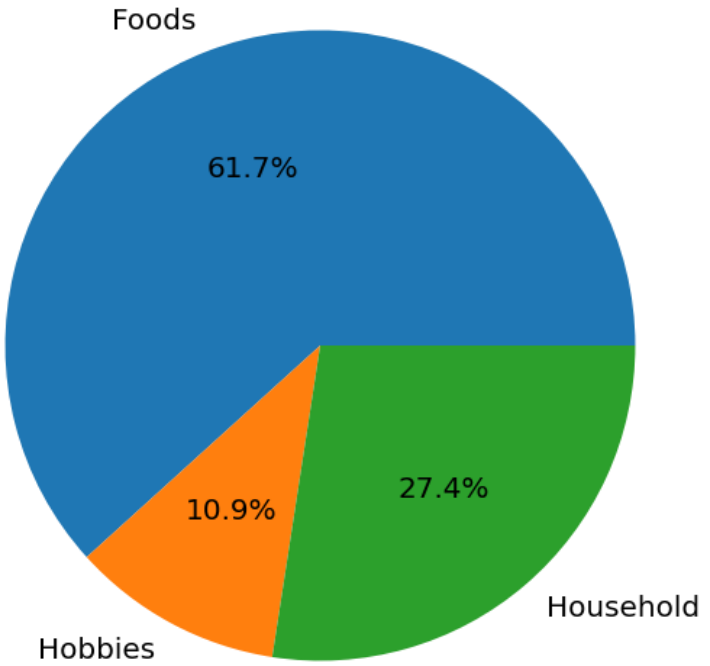
Wisconsin (WI)

Trend changed in 2013

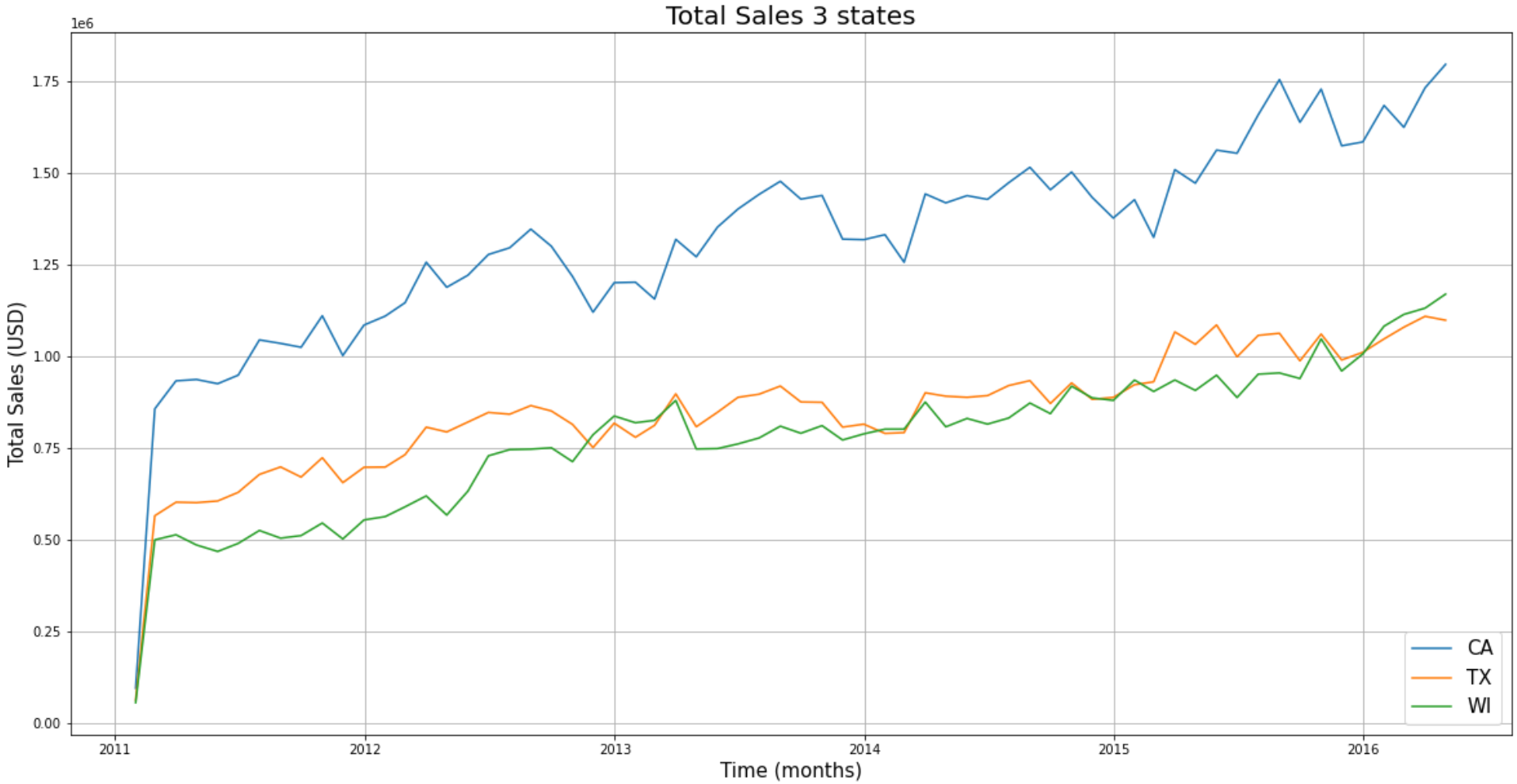


Wisconsin (WI)

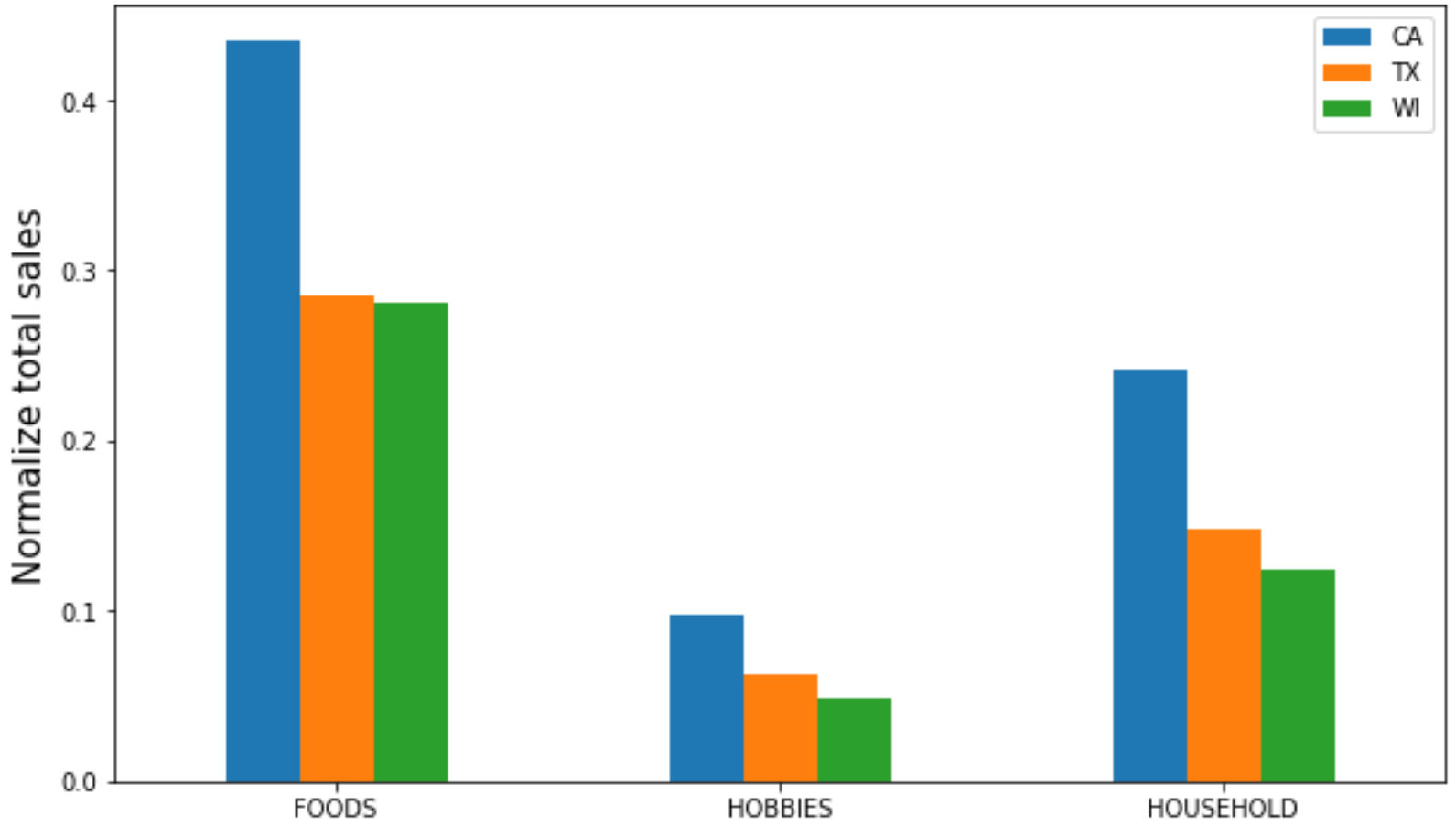




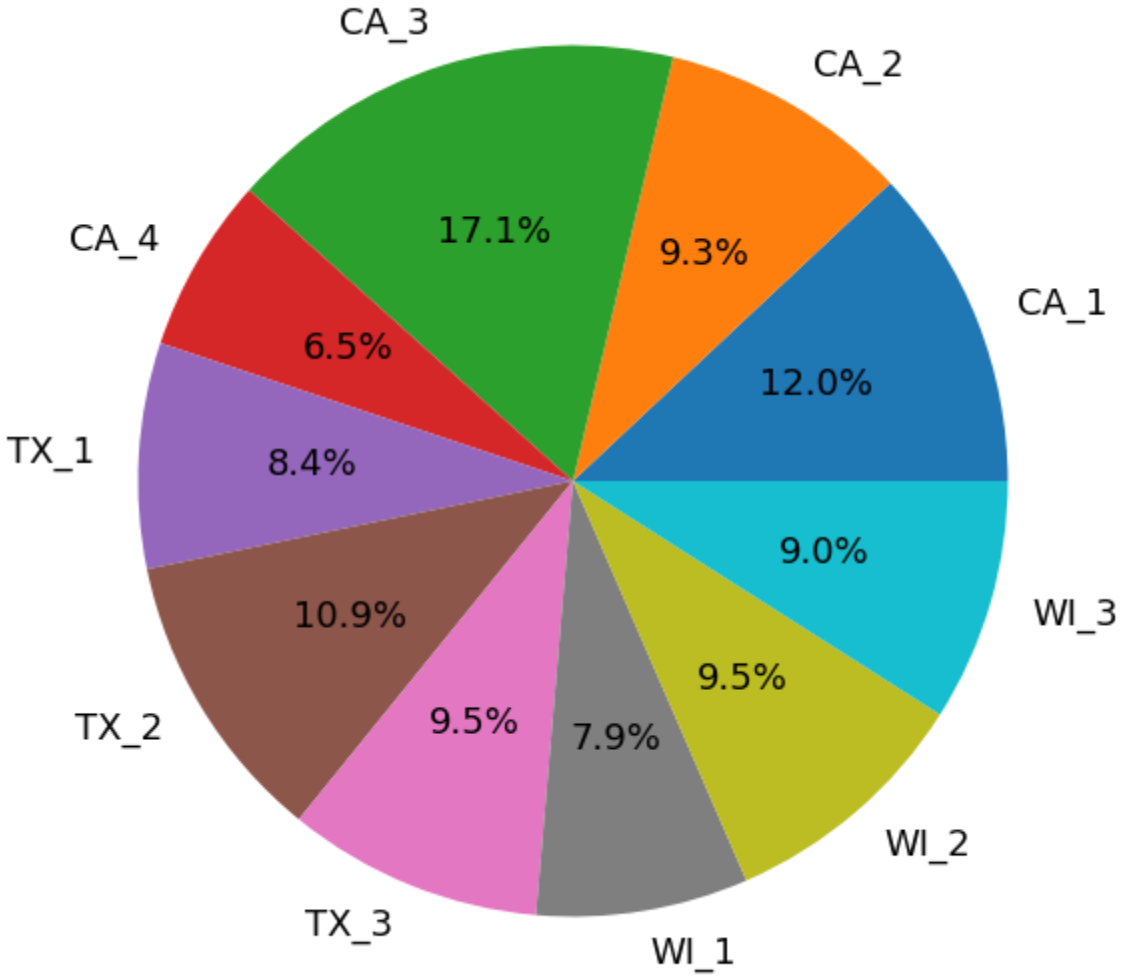
Sales performance by state



Sales performance by state

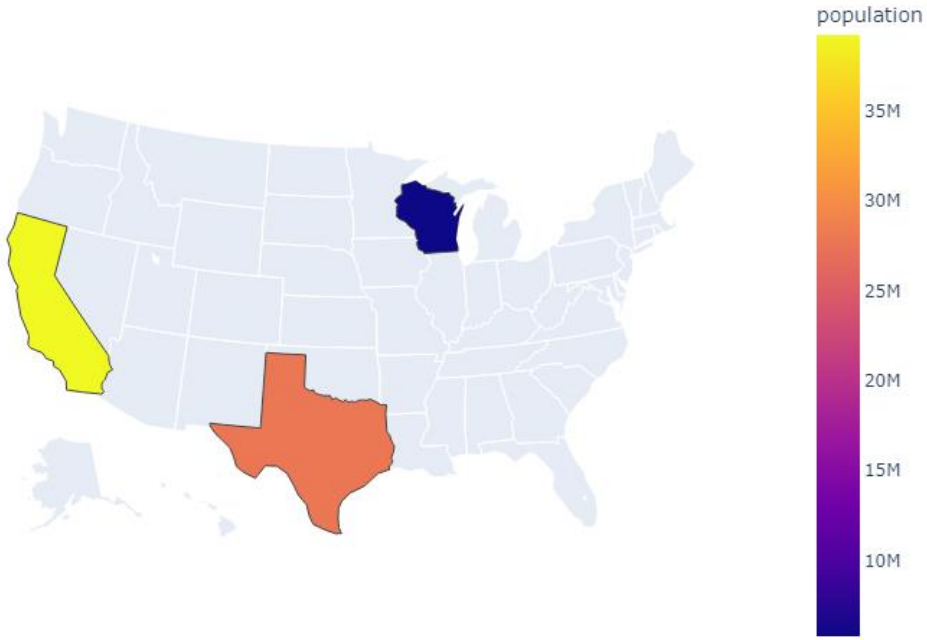


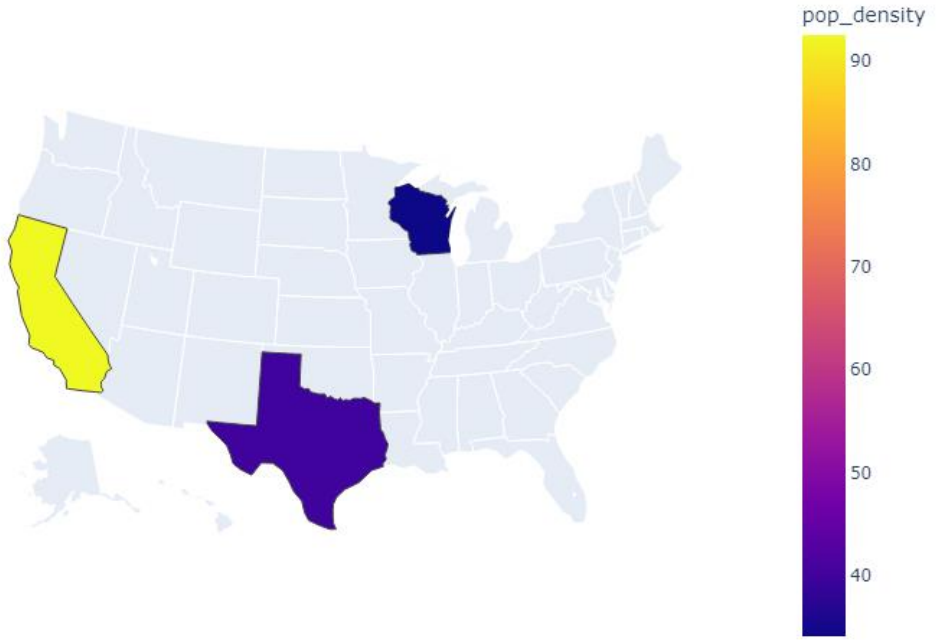
Sales performance by state



- CA_1 has the highest sales.
- CA_4 has the lowest sales.
- All stores in WI have contribute lower than 10% to the total sales.

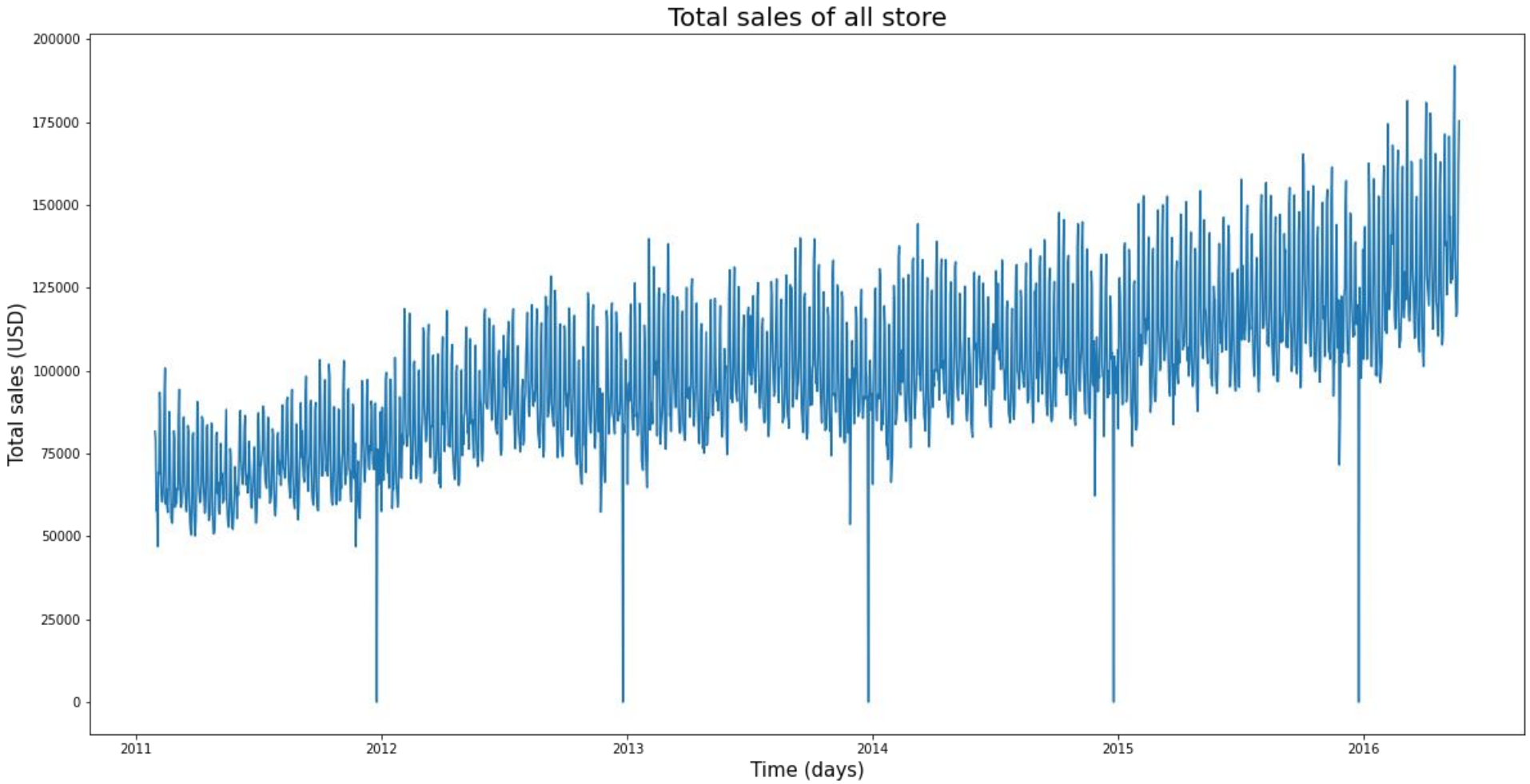
Population density



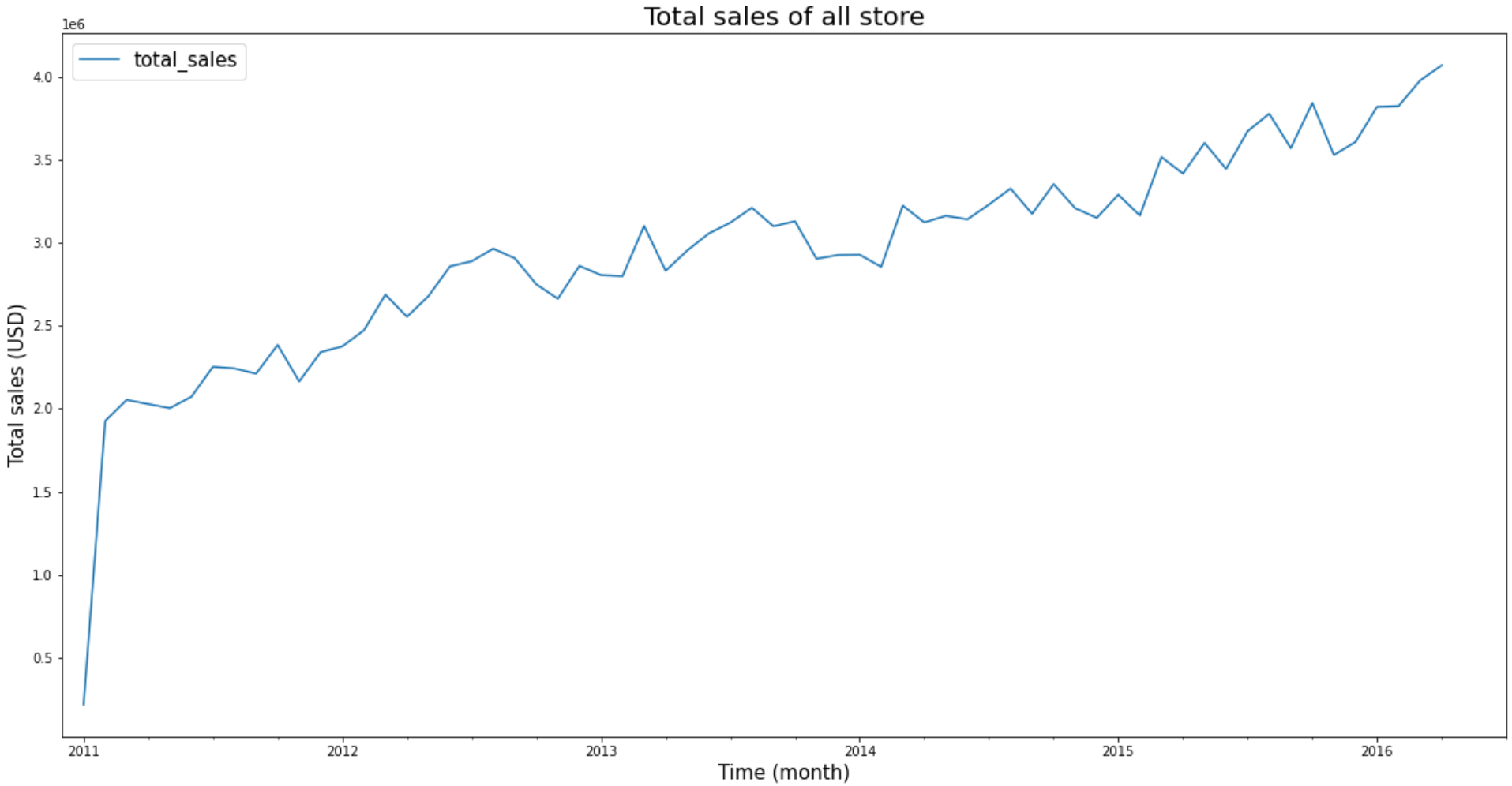


TX and WI have the same population density. This might be the reason that total sales in TX and WI are roughly the same.

Sales performance of all store



Sales performance of all store



Predictive analysis

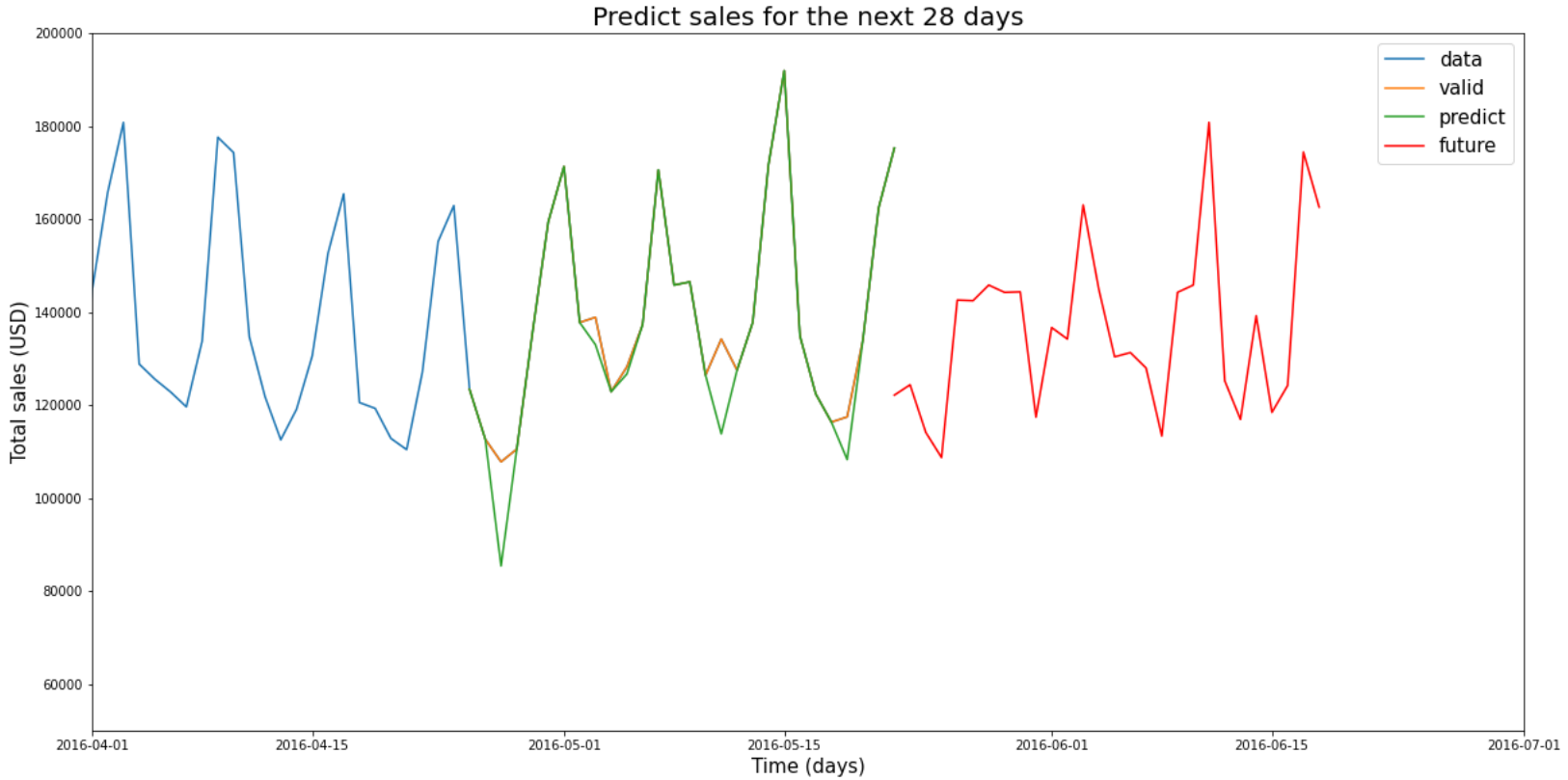
Predictive analysis

- We predict daily sales for the next 28 days by using a decision tree model.

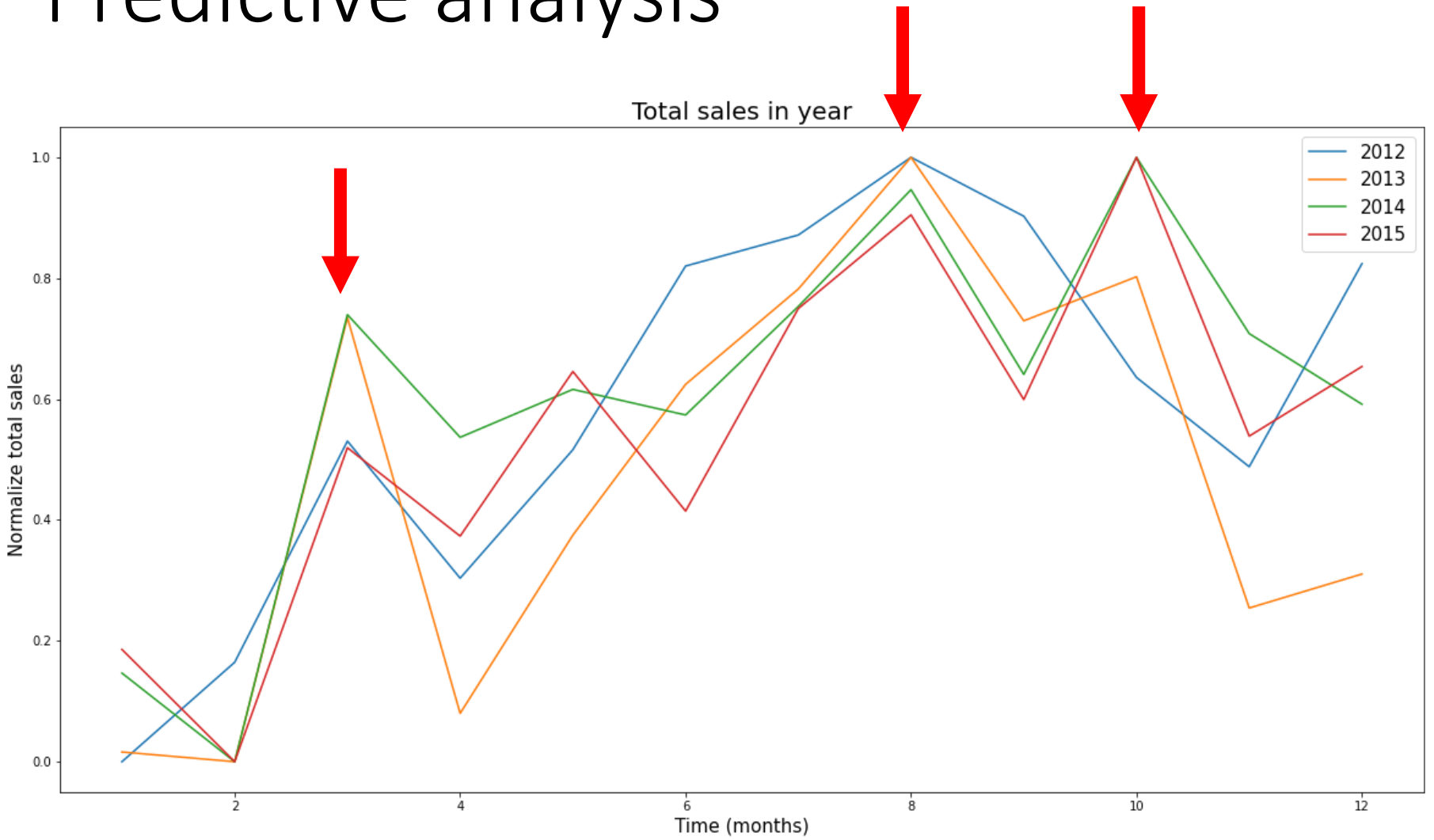
Predictive analysis

- The model has r2 score of 0.92

$$r2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

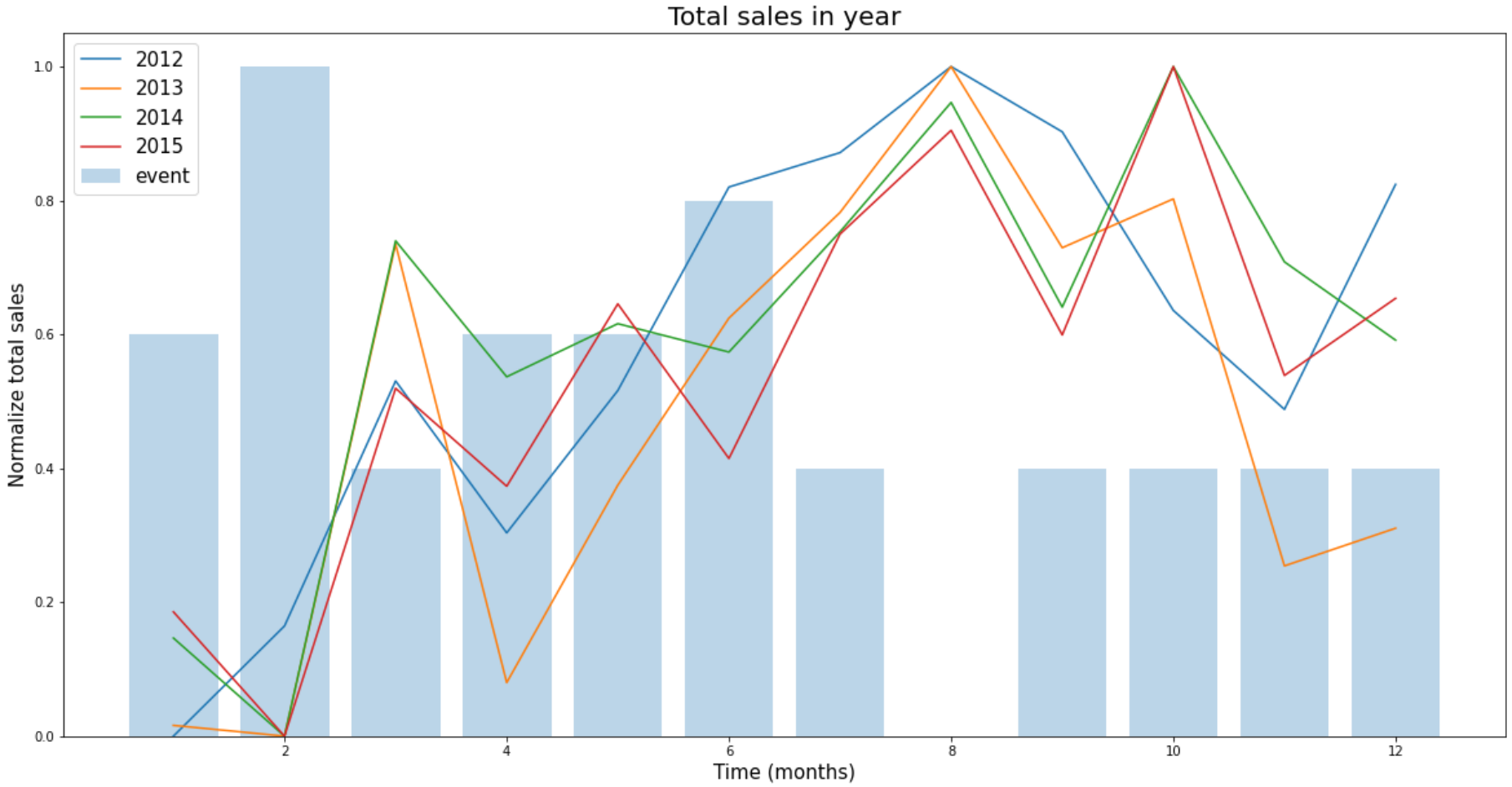


Predictive analysis



The data do not show a clear periodic trend.

Predictive analysis



There is no correlation between number of event and total sales.

Conclusions

Conclusions

- The highest sales are from CA.
- The food category has the highest sales, while hobbies and household items rank second and third, respectively.
- Total sales in each state depend on population density.
- Total sales increase every year.
- Total sales do not show a clear periodic function.
- Total sales do not depend on events.
- Total sales can be accurately predicted using our decision tree regression model

For the full report in Jupyter notebook

- https://github.com/pipitton-s/Coraline_assignment