

# Walmart sales forecasting



## Part I: Annual sales forecasting with regression models

Qiwei Lu

Oct 2020

## WANT TO BECOME WALMART'S SUPPLIER? THE LINE STARTS HERE...

- Everyone wants to sell through Walmart, the question is: can you keep up with the most advanced inventory management system Walmart has?
- Do you know how much inventory is enough for next year?

## DATA OVERVIEW: SOURCE

- Source: M5 Walmart sales forecasting competition on Kaggle

# DATA SET OVERVIEW: DETAILS

For 3 categories:  
FOODS,  
HOBBIES,  
HOUSEHOLD

} FOODS 1-3,  
HOBBIES 1-2,  
HOUSEHOLD 1-2

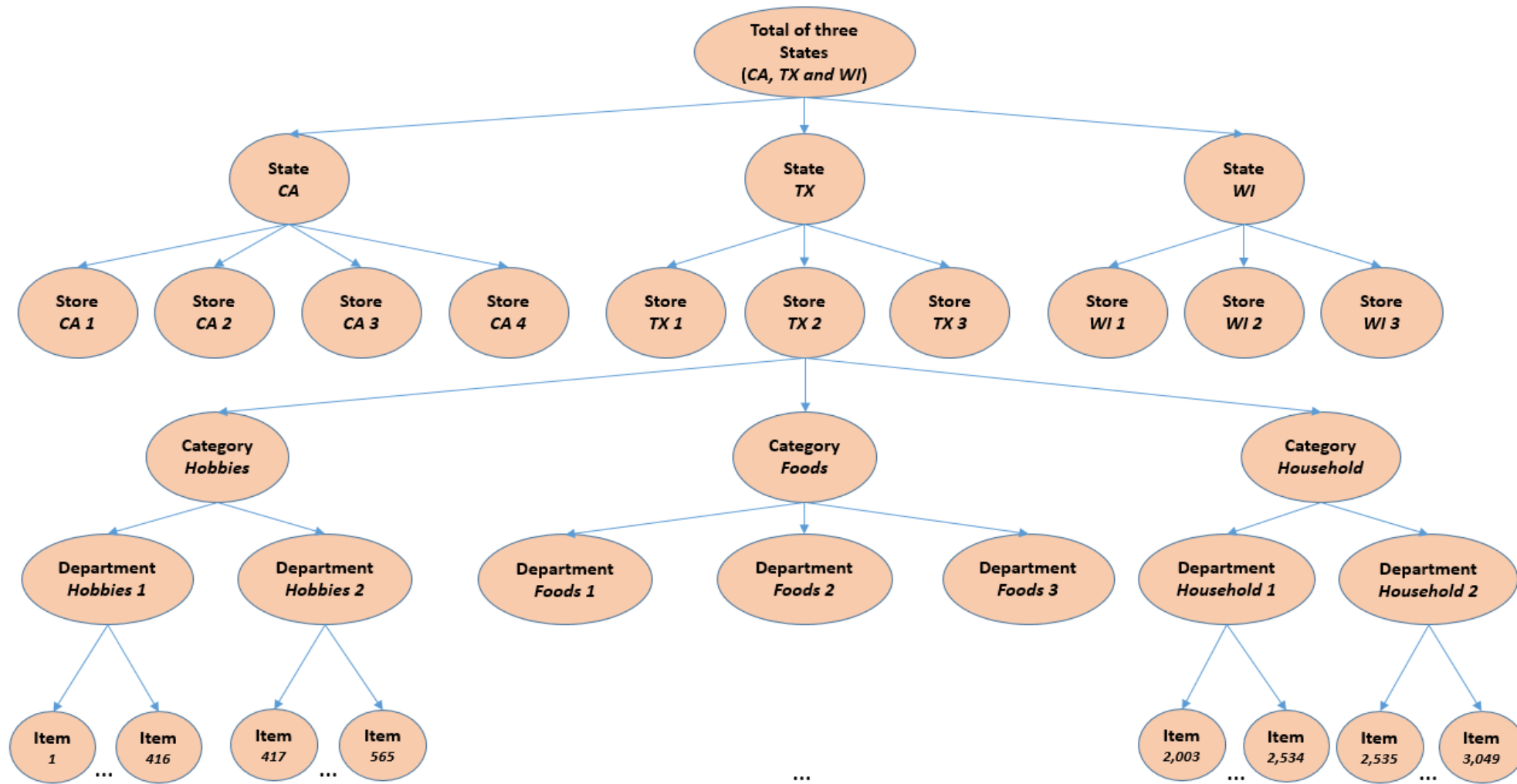
- 3 files including 30,490 records for 3,049 unique products in 10 different stores

Daily sale  
number and  
weekly price are  
given

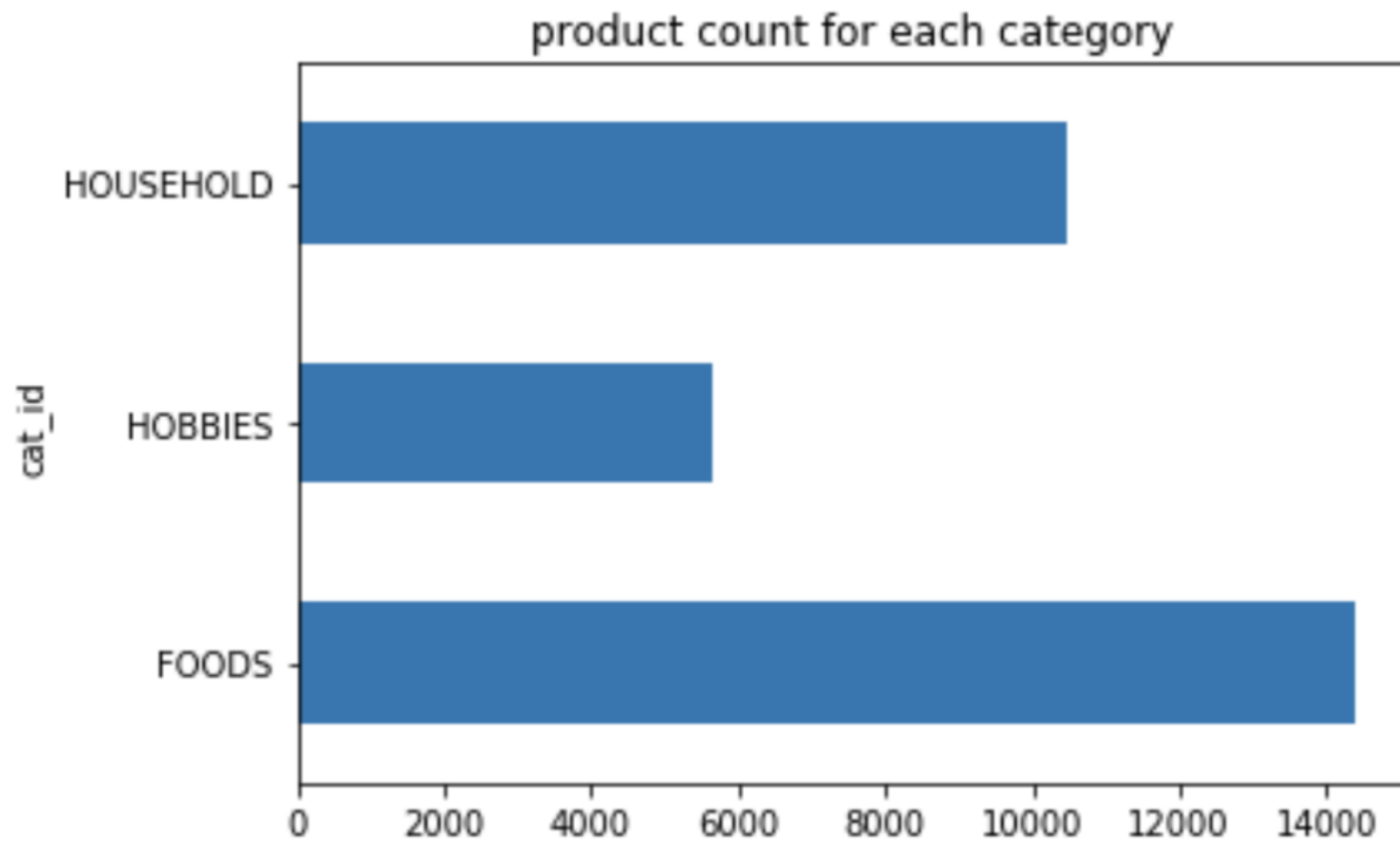
Across 3  
states: CA,  
TX,WI

Historical data range from **2011-01-29** to **2016-06-19(5.5 years)**

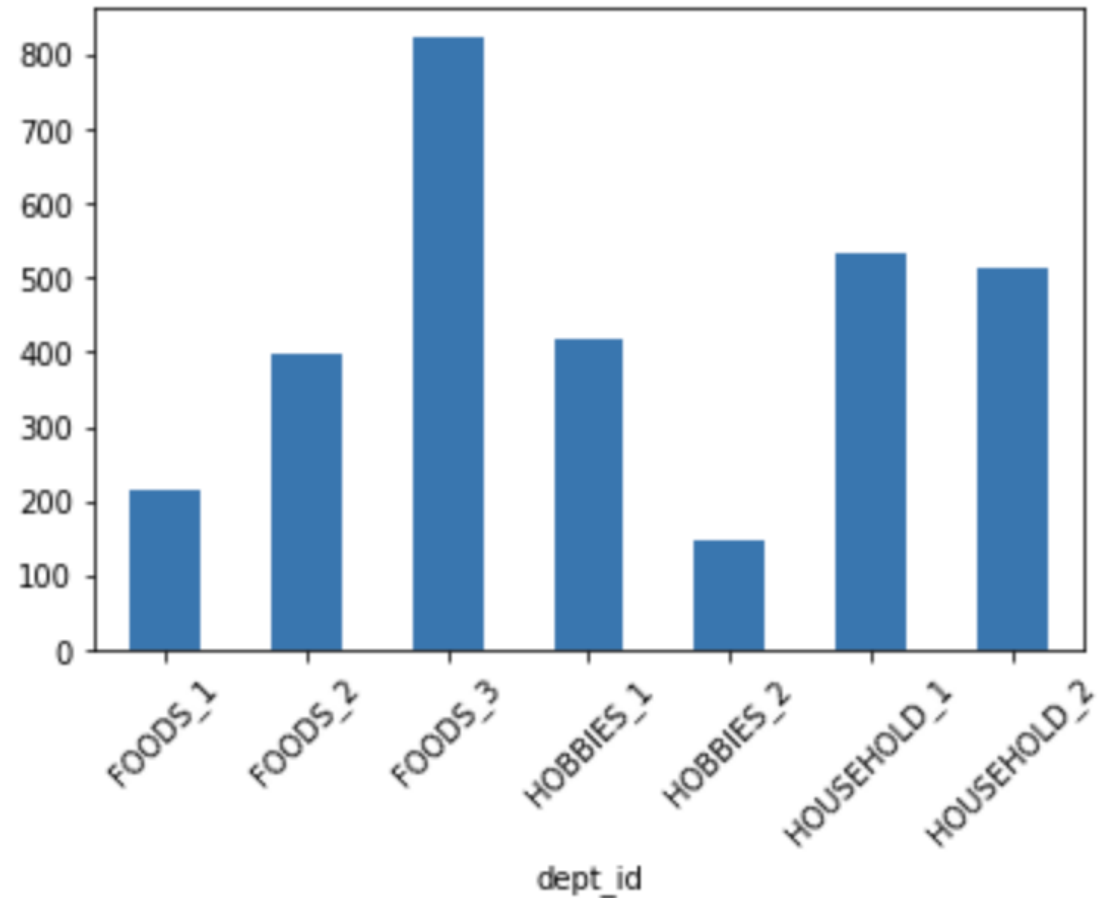
# DATA SET OVERVIEW: STRUCTURE



# DATA WRANGLING AND VISUALIZATION – PRODUCT COUNTS



# DATA WRANGLING AND VISUALIZATION – PRODUCT COUNTS

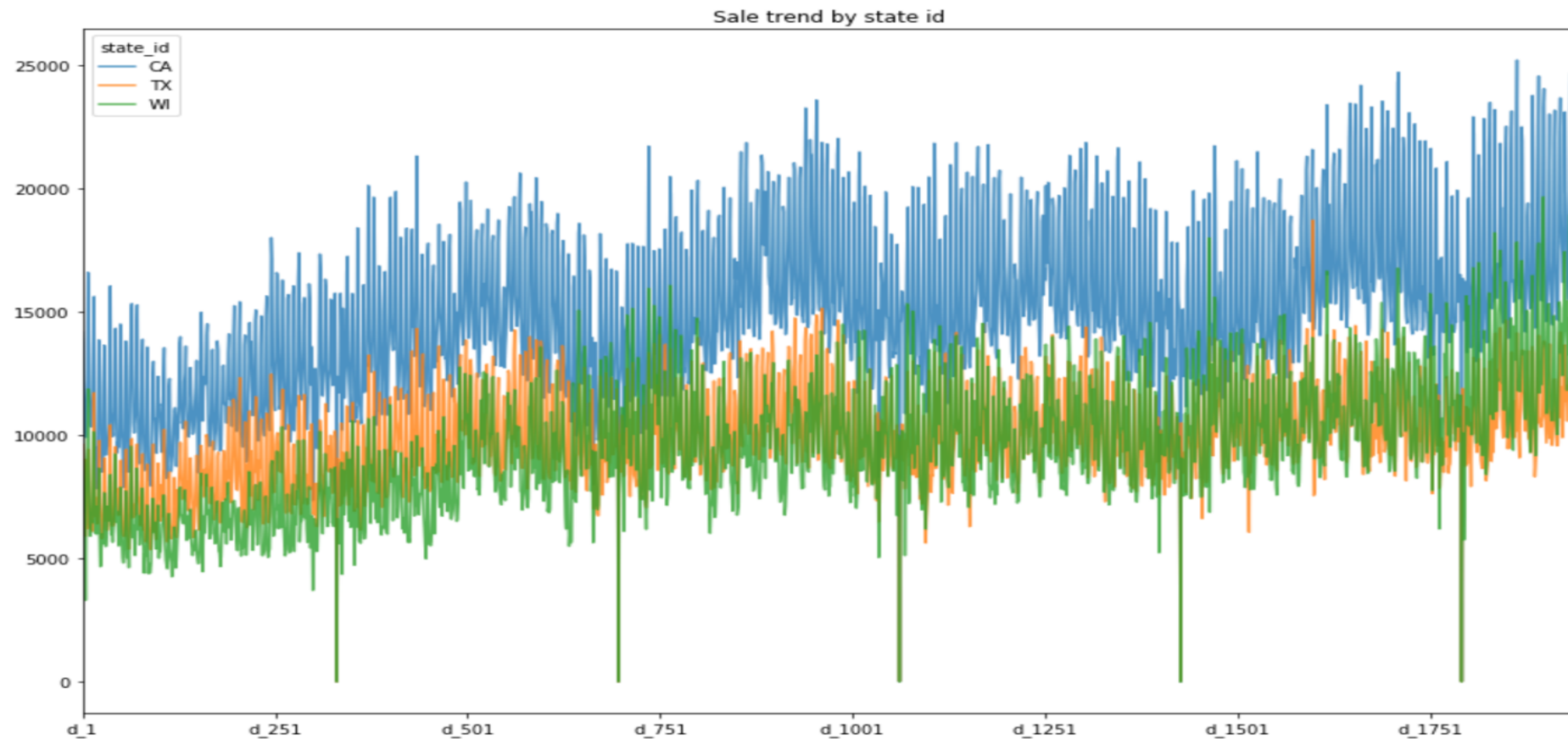


# DATA WRANGLING AND VISUALIZATION – SALES TREND

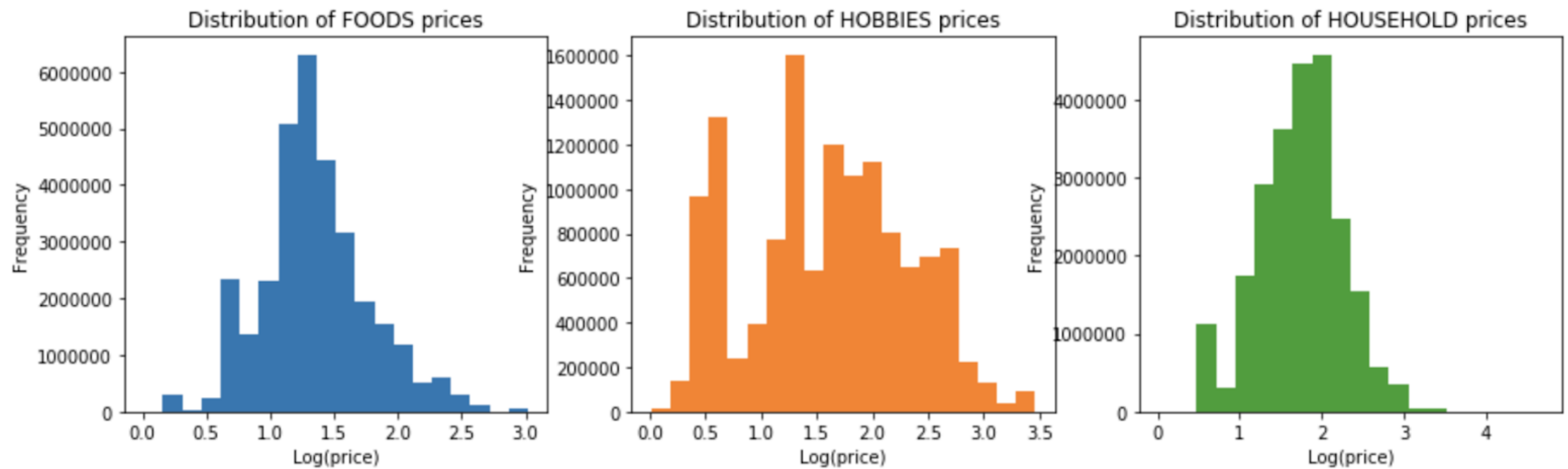




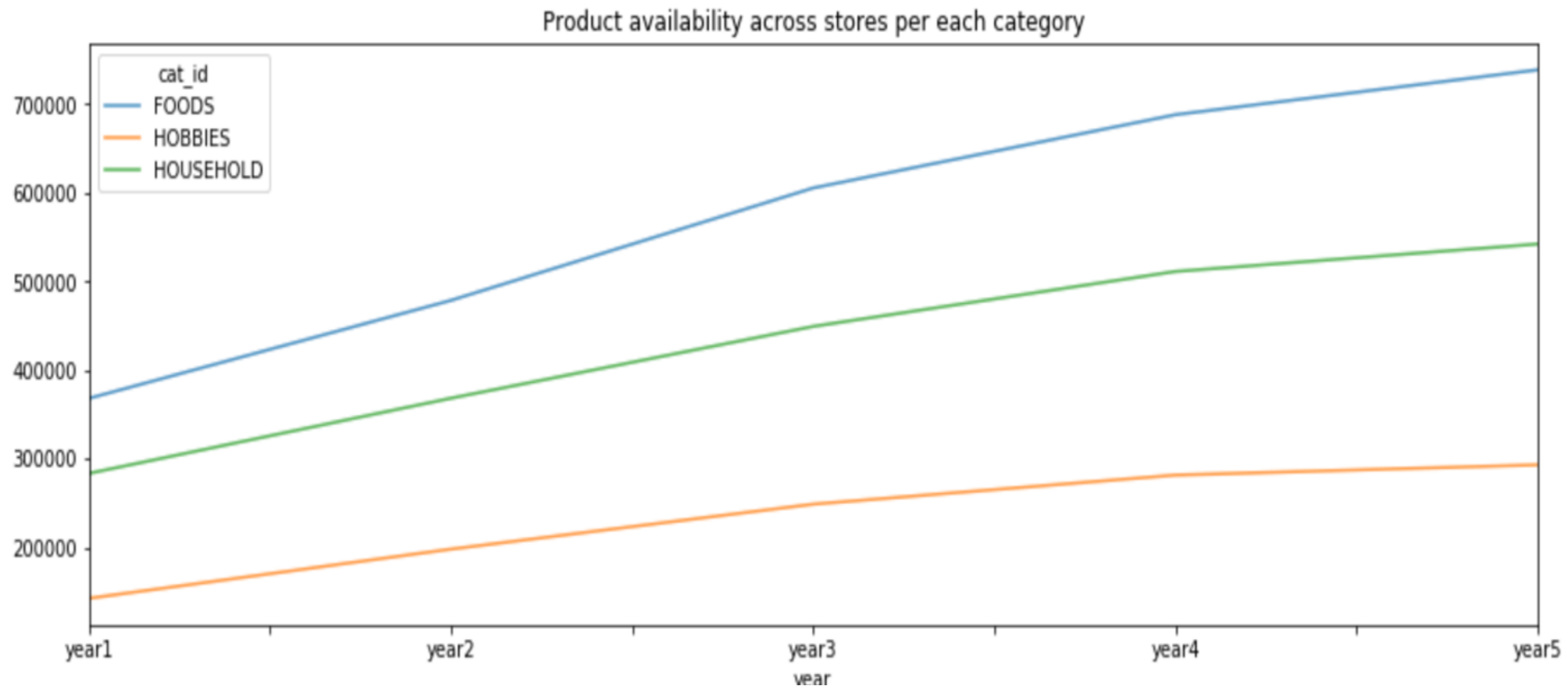
# DATA WRANGLING AND VISUALIZATION – SALES TREND



# DATA WRANGLING AND VISUALIZATION – PRICE DISTRIBUTION



# DATA WRANGLING AND VISUALIZATION – PRODUCT AVAILABILITY



# ASSUMPTIONS

- 1.About SNAP\*: we assume the product is eligible to be purchased by SNAP as long as it is FOODS
- 2.About sale of '0': we assume restock is always on time. Customers can have access as long as there is inventory
- 3.About when the price is set: we will assume that the prices are set on the previous year end and the suppliers know if their products will be sold at each store by that time as well.

\*Supplemental Nutrition Assistance Program (SNAP) provides a monthly supplement for purchasing nutritious food. If you qualify, you'll get a debit card to use for groceries.








## FEATURE DESIGN

- current year' min, max and average price and previous year's min, avg, max price
- previous year's total selling days for each store for the year (which is certain and happened already), and if in current years, the product will be continue (if the price count for current year is over 0 then yes, equals 0 then no)
- previous year's sale by store\_id and current year's sales number as target.
- department id and category id
- if the item can be purchased by SNAP (we assume all the food item is eligible)

# MODEL SELECTION AND TRAINING


- Linear Regressor
- Ridge Linear Regressor
- Lasso Regression
- Support Vector Linear Regressor & Support Vector Polynomial Regressor
- Decision Tree Regressor
- Random Forest Regressor
- Adaboost Regressor
- Neural Networks
- **Metrics for performance evaluation:** Root-mean-square error (RMSE) &  $R^2$  (coefficient of determination) score as reference for linear models

## CROSS VALIDATION RESULT– DROP THE WEAK MODELS

	Linear	Ridge	Lasso 	SVM Linear 	SVM Polynomial 	Decision Tree 	Random Forest 	AdaBoost 	Neural Networks 
rmse_ave	2132.75	2132.38	2131.70	2935.50	6523.93	2571.51	1999.71	4691.14	2018.95
rmse_std	1655.92	1656.22	1657.56	1434.29	1086.02	1688.46	1872.53	1364.84	1716.49
R2	0.93	0.93	0.93	0.90	0.60	0.91	0.93	0.78	0.93

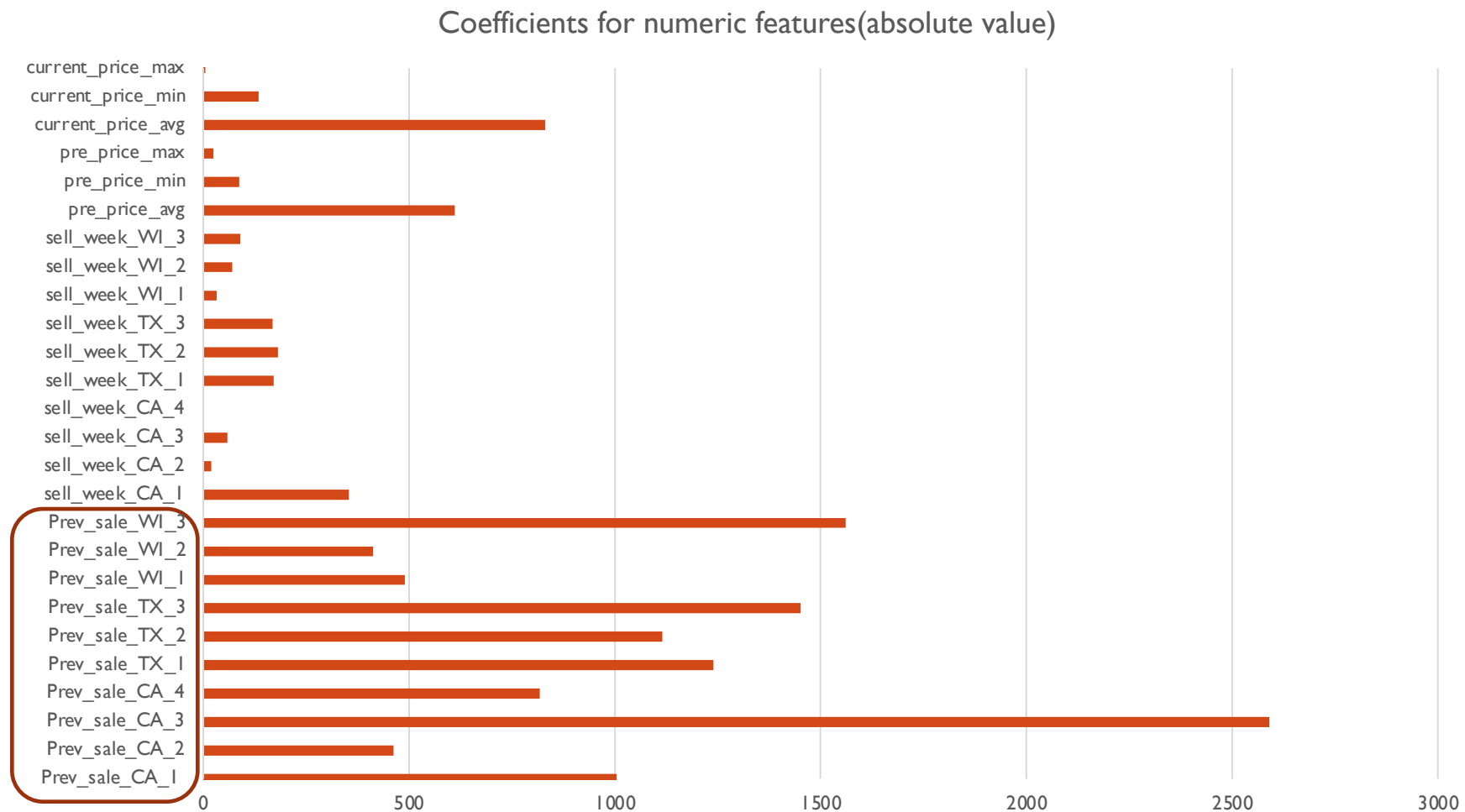
Benchmark: we assume the target is the same as previous year's sales, the Root-mean-square deviation (RMSE) is 2545.92

## TESTING RESULT

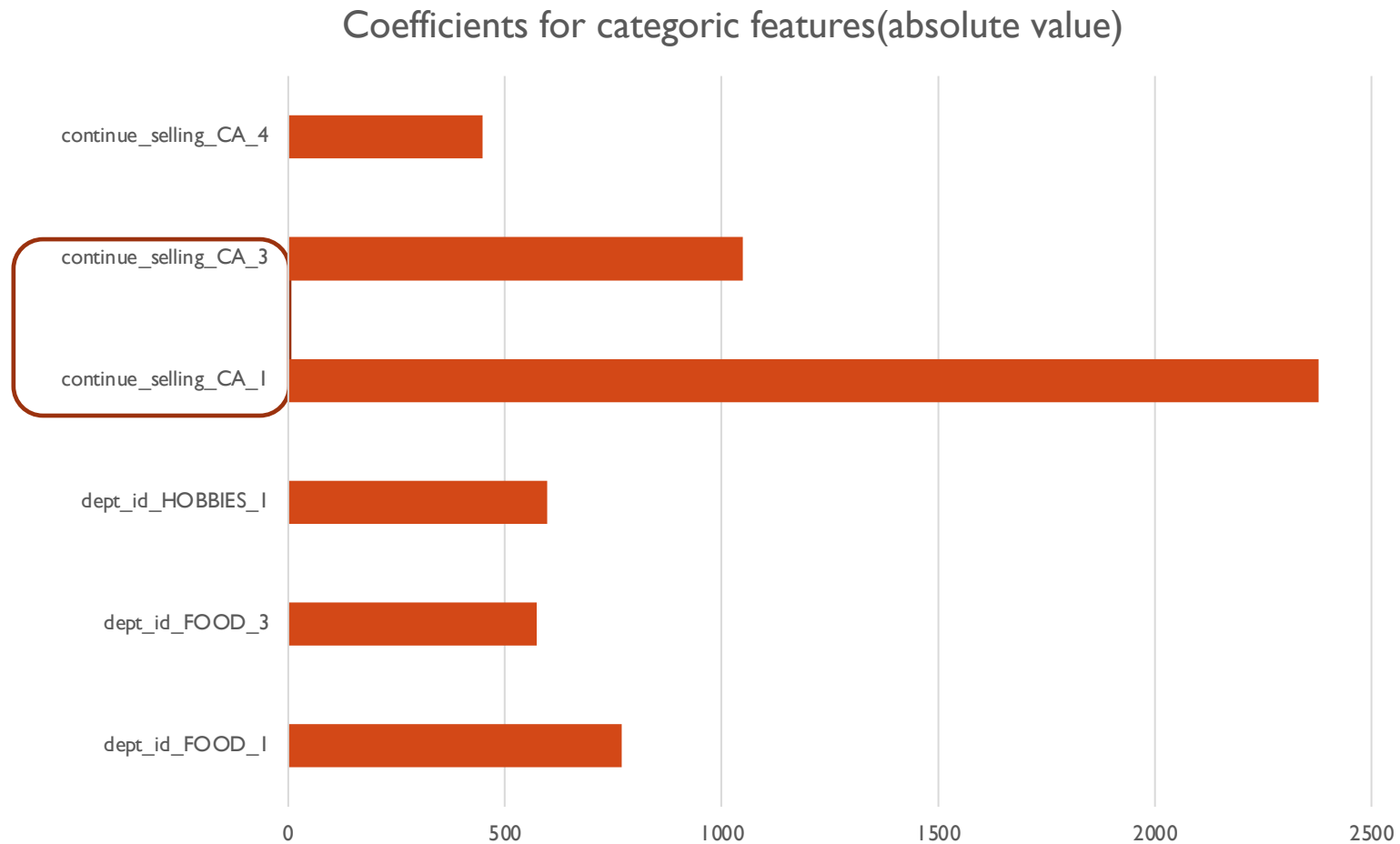
	Lasso 	Random Forest	Neural Networks
RMSE	2279.70	2810.62	7463.27
R <sup>2</sup>	0.94	0.91	0.34



# FEATURE IMPORTANCE – NUMERIC FEATURES



# FEATURE IMPORTANCE – CATEGORIC FEATURES



## APPLICATION TO THE REAL WORLD

- Potential and current Walmart's suppliers should make production plan based on the category of their products and use previous sales number as reference. For suppliers that sell in CA, knowing the next year's selling plan is helpful too.

## Q&A

**WAL★MART®**



**Thank You**