# Walmart Sales Forecasting

690 competitors

## Entities:

Stores (45), Department (81)

## Data:

Stores (Id, Type, Size)

Dept (DeptId)

Date (Week)

Holiday (Boolean)

Temperature (Avg) by Region

Fuel\_Price by Region

Markdown Data (5 categories, anonymized)

CPI

Unemployment rate

**Target:** Sales by Dept/Week

Benchmarks:

Seasonal naïve: 2943.93191/3025.89776 on public/private

Auto.arima with diff at 1 and 52 -> about 2700

## Takeaways:

Winner’s innovation was to learn seasonal/holiday patterns across stores by using SVD to filter out noise

Combination models win

Time series based models primarily

Strong seasonality -> Same week last year good predictor

Holiday adjustments are key, and have no automated solution!

* Domain knowledge was key to winning

Features not used much

ML models used in ensembles mainly, but not necessary to obtain good solution

## Notable writeups

#1st Place David Thaler (2237.70796/**2301.48**)

* Combination forecast
  + Simple models
    - TSLM: trend + seasonal dummy
    - Average based models for store/dept and week/dept
    - Seasonal naïve
    - Average of these after holiday shift: 2425/2499
  + 5 more complex models
* Time series models, does not use features, except for holiday adjustments
* STL Function used to model seasonality

#2nd Place Srihari Jaganathan (2310.37/**2371.42**)

* Combination forecast
  + Arima
  + Unobserved components
  + RF
  + KNN -> Best single model
  + TSLM
  + PC Regression

#3rd Place James King (2299.60/**2394.70**)

* Last year as base
  + Take weighted average of 2 closest weeks last year
* Line up holidays
* Account for growth rates/trend by store/dept
* Adjust for warm days

#4th Place Giulio (2355.27/**2424.24**)

* Similar to others, mostly manual adjustments to last year
* Used GBM’s, but not good results

#5th Place Bluefool (2358.69/**2427.05**)

* Lining up days
* Adjust for easter/christmas
* Predict growth from last year using CV
* Used temperature and fuel diff

#6th Place Timothee Henry (2374.43/**2430.87**)

* Many adjustments manually
* Previous year as base
* Holiday adjustment
* Growth adjustment

#8th Place Breakfast Pirate (2406.61/**2457.78**)

* Lined holidays up
* LR for each store and department (about 3600 models)
  + Avg sales by store/dept/week
  + Markdown4
  + Feature to account for start/end of month
* Christmas adjustment

#9th Place Neil Summers (2433.32/**2498.07**)

* Detrend
* LR with L1 to fit holidays
* Seasonality adjustment
* LR with L1 on Unemployment, CPI, Fuel Price to predict deseasonalized, detrended data

#10th Place Giba (2436.80/**2516.71**)

* Combination forecast
  + Median LR with obs from prev 2 years
  + GLM with last year obs + features -> higher errors than simple LR by itself
  + GBM with last year obs + features -> Higher errors than simple LR by itself