Walmart Time Series Analysis

Matt Torok
Jesse Neumann

Buisness Case

- The purpose of this project is to predict future sales for 10 different Walmart stores across 3 different states

Business Value

- Being able to predict future sales will allow for Walmart to accurately send new shipments of products that will reduce excess inventory and reduce costs overall

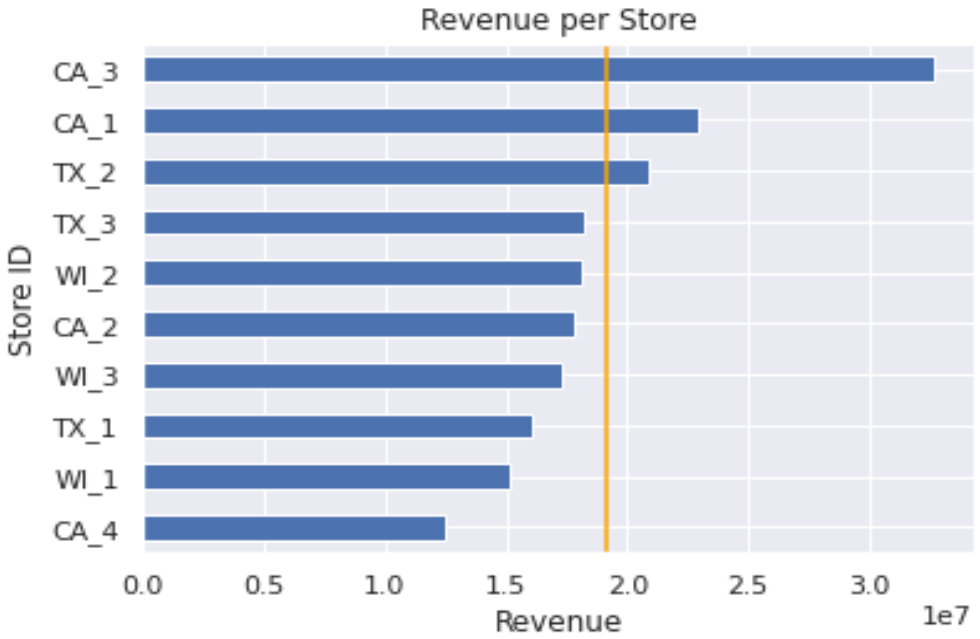
Goals

- In which state do the stores tend to be the most profitable?
- Which items tend to produce the most revenue at each store?
- Which categories generate the most revenue?
- Sales Predictions for Next 28 days

Dataset

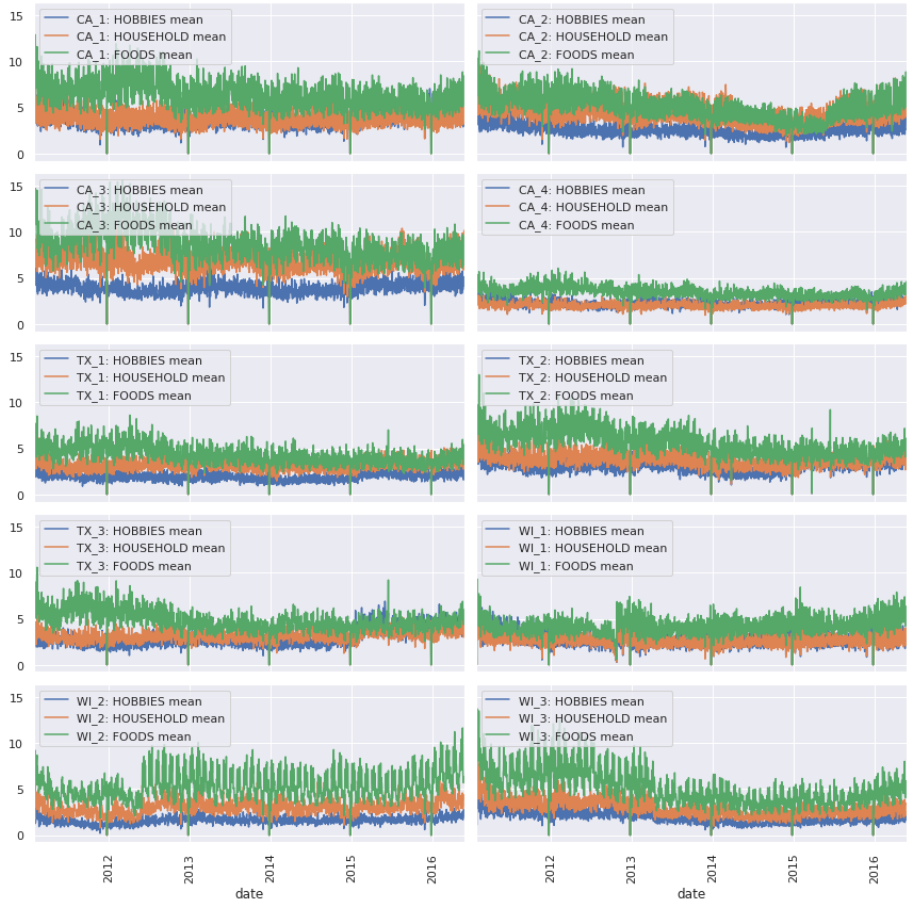
- The dataset for this project is available through the competition page on Kaggle at the link below:
- https://www.kaggle.com/c/m5-forecasting-accuracy

Recommendation #1



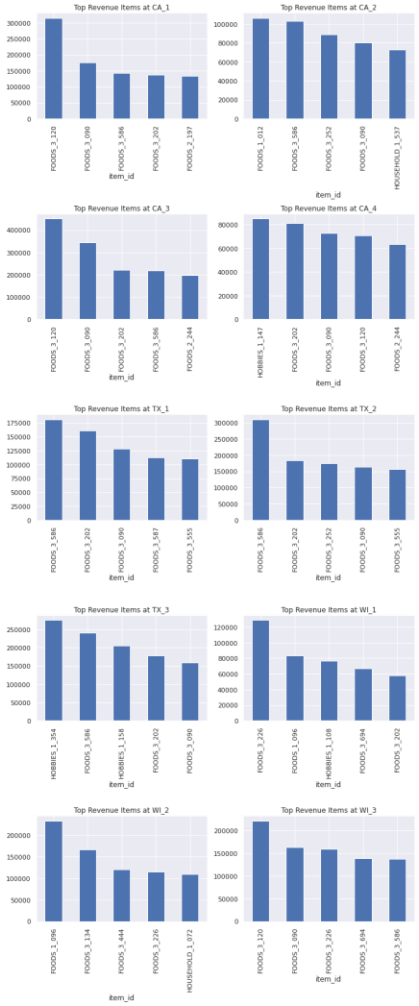
- California stores tend to be the most profitable so focus efforts on those stores

Recommendation #2



- Food tends to be the most profitable category across all stores so focus restocking efforts on that catagory

Recommendation #3



- We again see that food products consistently show up in the Top 5 products sold in each store

Predictions

id	F1	F2	F3	F4	F5
HOBBIES_1_001_C/	1.026352599	0.8133139468	0.7779641756	0.7779641756	0.7504477841
HOBBIES_1_002_C/	0.4258456717	0.3655587311	0.3655587311	0.3302089599	0.3525647849
HOBBIES_1_003_C/	0.5287141857	0.4774836983	0.4774836983	0.4774836983	0.4997287475
HOBBIES_1_004_C/	1.992021651	1.840086003	1.73767296	1.578966157	1.74258254
HOBBIES_1_005_C/	1.046190421	0.8739032523	1.110968261	1.253933819	1.210309946
HOBBIES_1_006_C/	1.121106511	1.026062061	1.149956031	1.023380214	1.074154191
HOBBIES_1_007_C/	0.3794426195	0.3524403581	0.3307822475	0.3191556788	0.3861695761
HOBBIES_1_008_C/	6.765903103	7.135941855	5.94957903	5.76822754	7.561314702
HOBBIES_1_009_C/	0.3641484127	0.4269537175	0.4052956069	0.4079774541	0.5266355387

- Sales predictions for the next 28 days for each individual time series

SARIMAX Model

SARIMAX Results

Dep. Variable:	sold	No. Observations:	1913
Model:	SARIMAX(1, 1, 1)x(1, 0, 1, 7)	Log Likelihood	-1774.683
Date:	Tue, 30 Jun 2020	AIC	3575.366
Time:	12:23:35	BIC	3647.531
Sample:	01-29-2011 - 04-24-2016	HQIC	3601.930
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
event_name_1	0.0078	0.005	1.505	0.132	-0.002	0.018
event_type_1	-0.0276	0.028	-0.986	0.324	-0.082	0.027
event_name_2	0.3899	0.940	0.415	0.678	-1.453	2.233
event_type_2	-0.8120	1.041	-0.780	0.436	-2.853	1.229
snap_CA	0.0032	0.035	0.090	0.928	-0.066	0.073
snap_TX	0.0113	0.035	0.320	0.749	-0.058	0.080
snap_WI	-0.0292	0.035	-0.843	0.399	-0.097	0.039
ar.L1	0.1063	0.615	0.173	0.863	-1.099	1.311
ma.L1	-0.9357	0.185	-5.060	0.000	-1.298	-0.573
ar.S.L7	-0.1687	0.949	-0.178	0.859	-2.030	1.692
ma.S.L7	0.0736	1.157	0.064	0.949	-2.195	2.342
var.measurement_error	0.3133	0.319	0.984	0.325	-0.311	0.938
sigma2	0.0558	0.336	0.166	0.868	-0.603	0.714

Ljung-Box (Q):	72.46	Jarque-Bera (JB):	6951.94
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	895.76	Skew:	2.21
Prob(H) (two-sided):	0.00	Kurtosis:	11.26

- RMSE = 0.588
- Corr = 0.519

Future Work

- Add holidays to FBProphet Model
- Investigate the effect of SNAP of sales data
- Figure out what time of year, certain products sell best

THANK YOU

- Questions?