

Predictive Retail Analytics



Nicole Aguilera
Dr. Brett Werner
DSC 630
June 5, 2021

Project Statement

Through a retail data analytics data set, the model created through this project will be used to predict sales for a store for the following year, separated by department



Variables

Variable	Description
Date	MM/DD/YY – Date of the start of the week in question
Weekly_Sales	Total weekly sales for the week
Is_Holiday	Whether the week is a holiday week (1) or not (0)



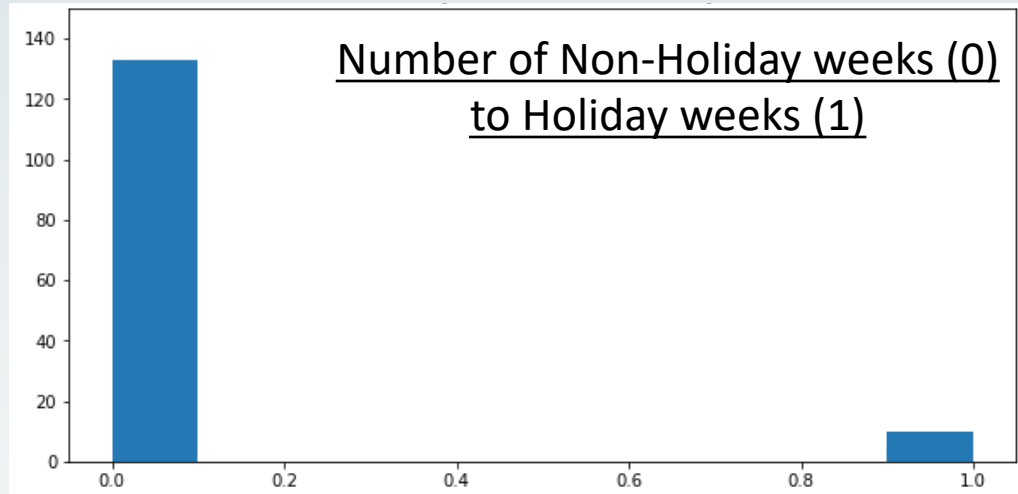
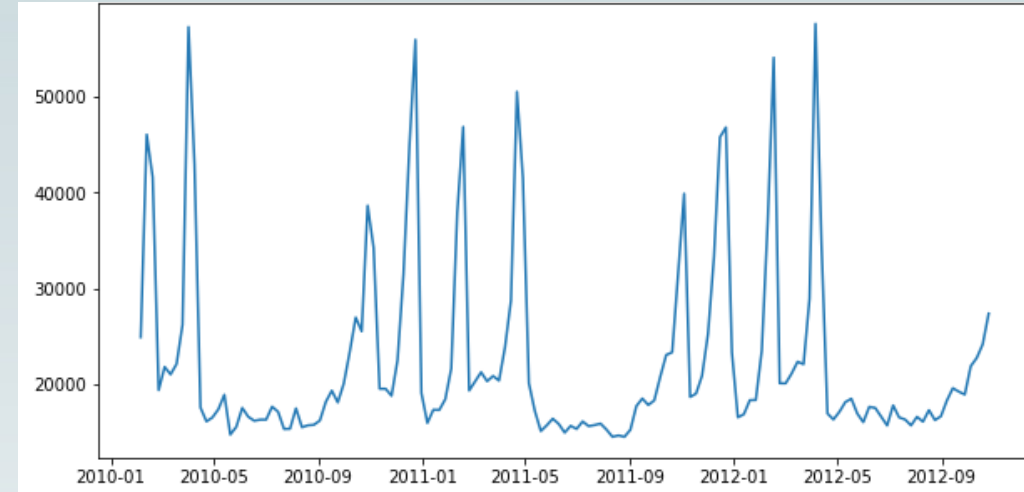
Exploratory Data Analysis

Weekly Sales from
February 2010 to October 2012

Summary Statistics

	Weekly_Sales	Is_Holiday
count	143.000000	143.000000
mean	22513.322937	0.069930
std	9854.349032	0.255926
min	14537.370000	0.000000
25%	16494.630000	0.000000
50%	18535.480000	0.000000
75%	23214.215000	0.000000
max	57592.120000	1.000000

Data set reviews 3
years of sales data to
analyze predictions



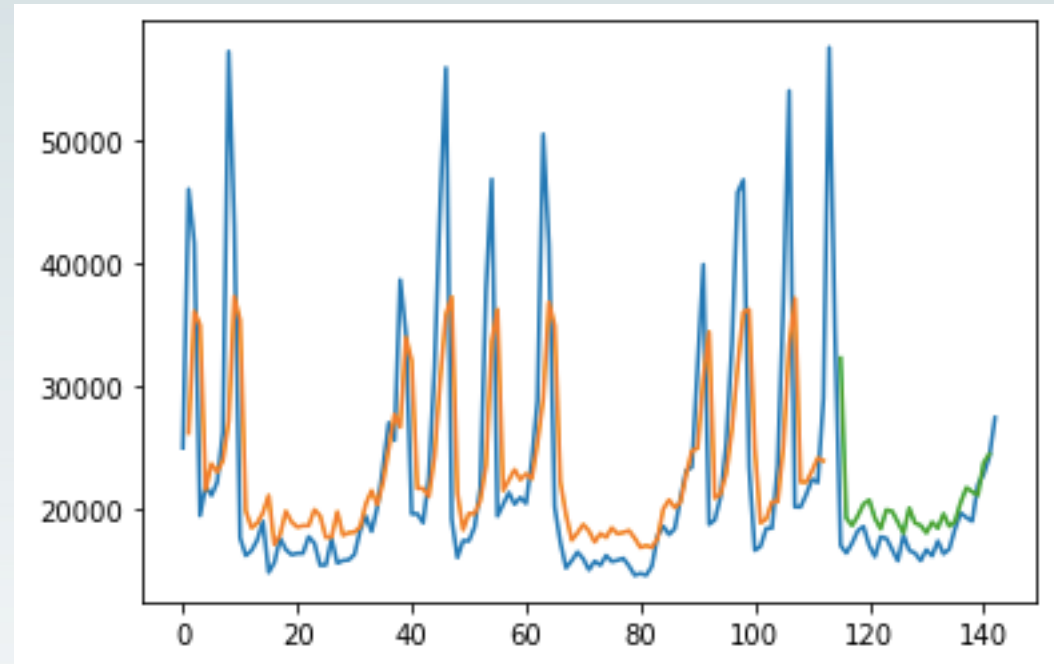
LSTM Regression Model

- Long Short Term Memory
- Time series prediction of Sales through neural network model
- Train score was able to predict within \$7,846.19
- Test score was able to predict within \$3,665.01



```
Epoch 95/100  
112/112 - 0s - loss: 0.0341  
Epoch 96/100  
112/112 - 0s - loss: 0.0340  
Epoch 97/100  
112/112 - 0s - loss: 0.0342  
Epoch 98/100  
112/112 - 0s - loss: 0.0339  
Epoch 99/100  
112/112 - 0s - loss: 0.0338  
Epoch 100/100  
112/112 - 0s - loss: 0.0336
```

Model Results



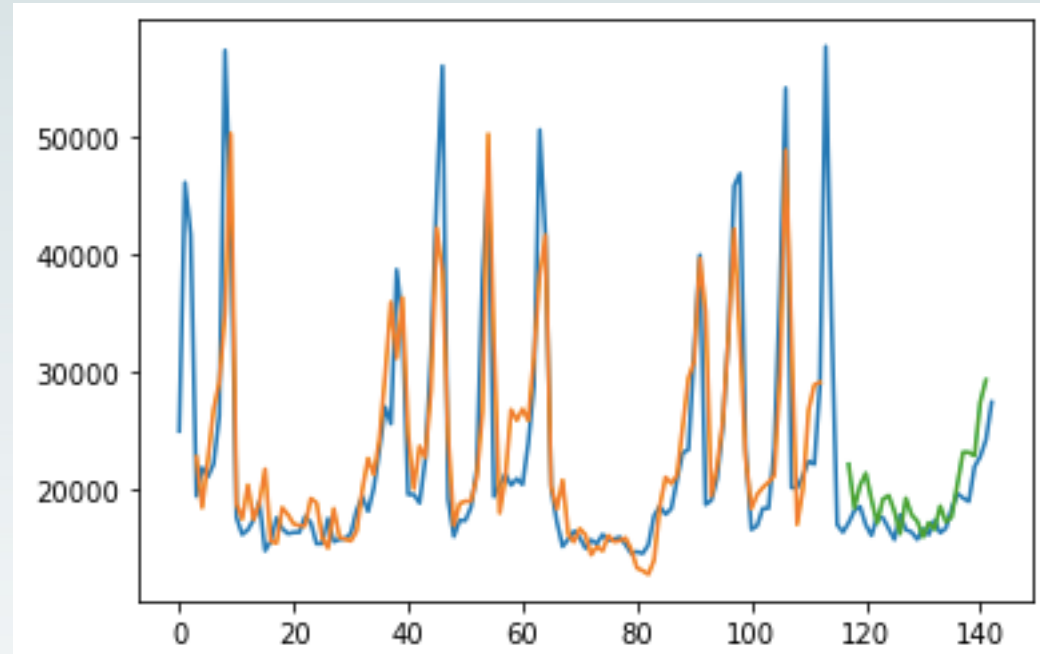
LSTM Regression with Time Steps Model

- Time series prediction of Sales using past observations as input features, instead of separate features
- Train score was able to predict within \$5,090.35
 - Test score was able to predict within \$2,655.37



```
Epoch 95/100  
110/110 - 0s - loss: 0.0146  
Epoch 96/100  
110/110 - 0s - loss: 0.0146  
Epoch 97/100  
110/110 - 0s - loss: 0.0145  
Epoch 98/100  
110/110 - 0s - loss: 0.0144  
Epoch 99/100  
110/110 - 0s - loss: 0.0142  
Epoch 100/100  
110/110 - 0s - loss: 0.0145
```

Model Results



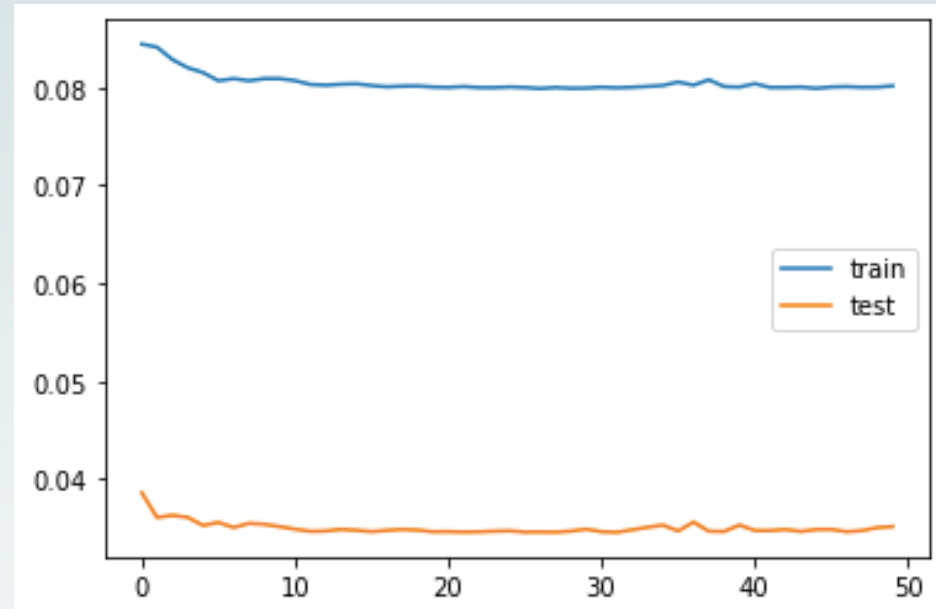
Data Set
Train Score
Test Score

LSTM Regression Adding Holiday Variable

- Multivariate time series forecasting
- Model can predict results with about 3.52% error
- Loss/train results keeps consistent around 8%



Model Results



```
Epoch 45/50  
2/2 - 0s - loss: 0.0799 - val_loss: 0.0349  
Epoch 46/50  
2/2 - 0s - loss: 0.0800 - val_loss: 0.0349  
Epoch 47/50  
2/2 - 0s - loss: 0.0801 - val_loss: 0.0346  
Epoch 48/50  
2/2 - 0s - loss: 0.0800 - val_loss: 0.0348  
Epoch 49/50  
2/2 - 0s - loss: 0.0800 - val_loss: 0.0351  
Epoch 50/50  
2/2 - 0s - loss: 0.0801 - val_loss: 0.0352
```

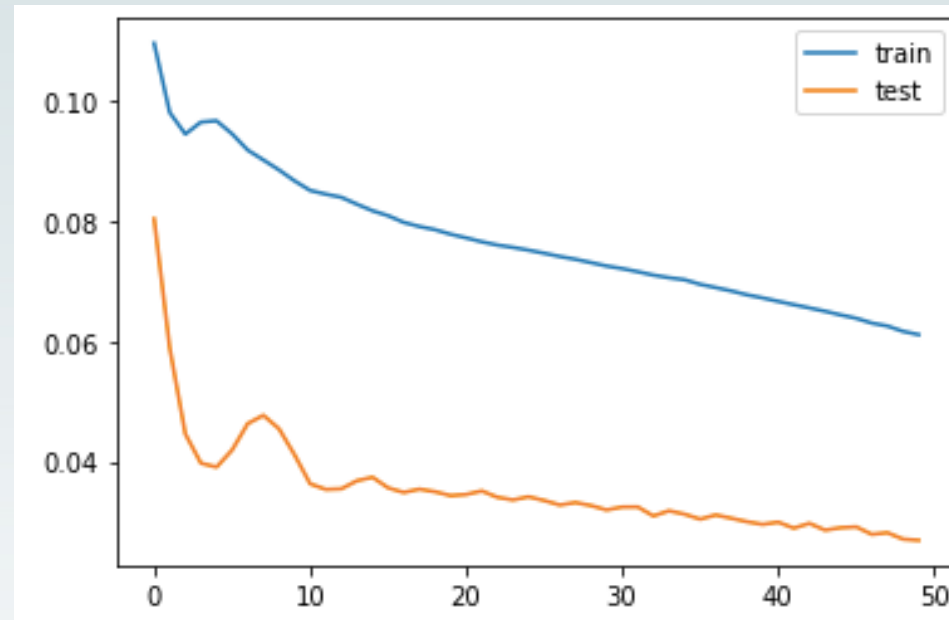
LSTM Regression with Two Departments and Holiday Weeks

- Multivariate time series forecasting
- Model can predict results with about 2.70% error
- Loss is 6.11%



```
Epoch 45/50  
2/2 - 0s - loss: 0.0644 - val_loss: 0.0290  
Epoch 46/50  
2/2 - 0s - loss: 0.0638 - val_loss: 0.0292  
Epoch 47/50  
2/2 - 0s - loss: 0.0630 - val_loss: 0.0279  
Epoch 48/50  
2/2 - 0s - loss: 0.0625 - val_loss: 0.0282  
Epoch 49/50  
2/2 - 0s - loss: 0.0616 - val_loss: 0.0271  
Epoch 50/50  
2/2 - 0s - loss: 0.0611 - val_loss: 0.0270
```

Model Results



Conclusion

- Low error margins overall
- Univariate LSTM models mimicked the data, proving to be accurate
- Multivariate LSTM models had test and training sets results that seemed to be far apart, only 0.05 difference
- Best model was the univariate LSTM time model
- Add ARIMA model to be considered for multivariate analysis in the future



Works Cited

- Brownlee, J. (2016). *Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras*. Machine Learning Mastery. <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/>
- Brownlee, J. (2017). *Multivariate Time Series Forecasting with LSTMs in Keras*. Machine Learning Mastery. <https://machinelearningmastery.com/multivariate-time-series-forecasting-lstms-keras/>
- Retail Data Analytics. (2017). Kaggle. <https://www.kaggle.com/manjeetsingh/retaildataset>

