Walmart Sales Forecasting

With 56% of U.S. grocery market-share, 4,743 U.S. stores, 240M customers worldwide, and $559B in worldwide revenue it's easy to say that Walmart is huge (Statista 2021, April 27), and with Walmart's huge size comes huge amounts of data. Thus, in 2014 they created a public competition for data-scientists to predict store sales using a selection of their anonymized datasets. *Awesome.*

But it's not a worthwhile competition if it's easy: holiday markdown events and a lack of what Walmart describes as "complete/ideal historical data" will make the data interesting and difficult to predict with, especially since only three years' worth of data, from 2010 to 2013, are provided. I will have to mold my model to use few, yet important, holiday events' records to accurately predict future holidays, though sales can vary so much.

So, using Walmart's provided training datasets I will create a model to predict store sales on the provided test dataset. This will mean predicting sales per store, per respective store's departments, per date, and if that date is a holiday. For example: "What were the sales for Store 1, Department 4 on November 23rd, 2012, which was a holiday (Thanksgiving)?".

To judge whether models are successful or not Walmart provided a scoring criterion for entrants, the weighted mean absolute error of predictions vs actual. 1st on the Kaggle leaderboard has a score of 2,301, the median (344th) 3,797, and 3rd from last (686th) 24,8479. Being a novice data-scientist, and within context of the leaderboard, I believe a score less than 3,500 is a well-set measure of success. More than that, it would likely mean great returns in real life for Walmart, if my model were implemented. We'll see once the predictions are fleshed out.

My deliverables include a slide deck to present my project and findings, a project report to go into more details, the code of my project in the form of Jupyter Notebooks, and a GitHub repository containing all previous deliverables.

References

Statista. (2021, April 27). *Walmart - Statistics & Facts*.

https://www.statista.com/topics/1451/walmart/