Walmart Sales Forecast – Big data

Overview

My data is taken from a Kaggle competition challenge where I am supposed to make a 12-month weekly forecast give data from the previous almost two years starting from February 2010 to October 2012. I am however not needed to make the forecast at this time but to put it in three databases and eventually import it to my Hadoop cluster that I have created.

RDBMS

My choice of database management software is MySQL, PostreSQL, and Microsoft SQL Server in which I put Store1, Store2, and Store4 respectively. In each of them I uploaded ten tables. I utilized the import data tools in MySQL Workbench for MySQL and in Dbeaver for both Postgres and MS SQLserver.

Data

The data totals 6.7 GB and is from 45 stores from different parts of the US with each store having many departments. The stores are chosen at random and are anonymous. Their locations can be inferred from the available features data but the exact store cannot be known. I chose to focus on the first 10 departments from the first 3 Class A stores numbered 1,2 and 4. The class A stores are the super-centers. I then gave each store its own database and then gave the ten departments a table each.

The following are the department represented by the tables:

1. Candy and Tobacco
2. Health and Beauty
3. Stationery
4. Paper Goods
5. Media and Gaming
6. Cameras
7. Toys
8. Pets
9. Sporting Goods
10. Automotive

Benefit

The project aims to assist the management make decisions for product stockpiling based on limited historical data, taking into account things such as holidays, weather, consumer price index, and the unemployment rate among other factors.