Walmart Weekly Sales MongoDB ETL Documentation

Data are extracted from 2 datasets i.e sales data-set.csv and Features data set, transformed using Pandas and loaded to MongoDB as ***weeklysales\_db*** under ***weeklysales table*** through PyMongo.

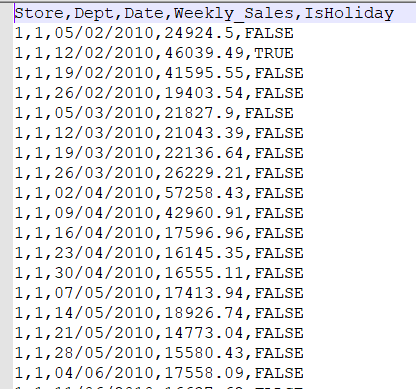
The objectives of ETL are,

* to upload a transformed dataset of Total Weekly Sales for a store at a given date with Markdown data.
* the dataset gives user the ability to query to find which store had best sales, whether the store had markdowns and etc

# Sales data-set.csv Description

Historical sales data, which covers to 2010-02-05 to 2012-11-01. Within this tab you will find the following fields:

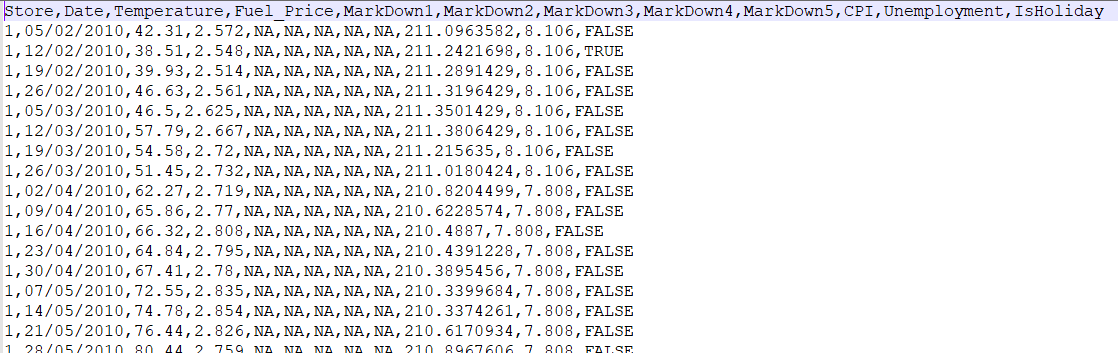
* *Store - the store number*
* *Dept - the department number*
* *Date - the week*
* *Weekly\_Sales - sales for the given department in the given store*
* *IsHoliday - whether the week is a special holiday week*



# Features data set.csv Description

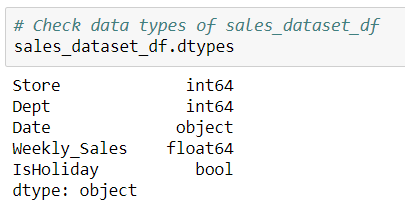
Contains additional data related to the store, department, and regional activity for the given dates.

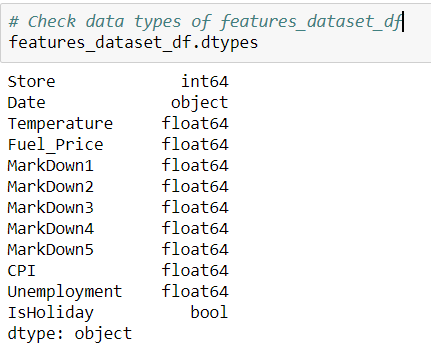
* *Store - the store number*
* *Date - the week*
* *Temperature - average temperature in the region*
* *Fuel\_Price - cost of fuel in the region*
* *MarkDown1 to 5 - anonymized data related to promotional markdowns. MarkDown data is only available after Nov 2011 and is not available for all stores all the time. Any missing value is marked with an NA*
* *CPI - the consumer price index*
* *Unemployment - the unemployment rate*
* *IsHoliday - whether the week is a special holiday week*



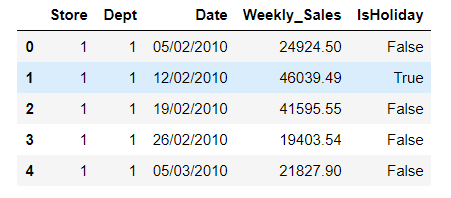
Data Extraction and Transformation

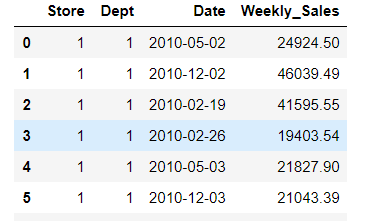
* Snapshot below shows data types of columns from 2 datasets.



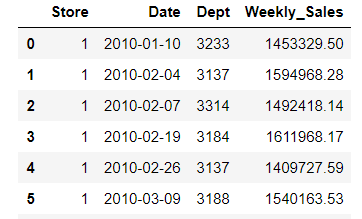
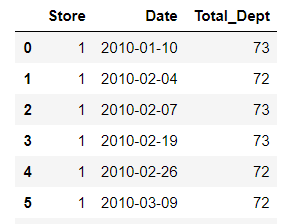


* Date in datasets aren’t a standard one and had to be converted,





- Calculate the total dept by date and Weekly\_Sales by date



# - Replace Nan column in MarkDownX with Zeros

# 

# 

# - Merge all dataframes and drop extra rows that has Nan

# 

# 

# - Rename MarkDownX columns to Mark\_Down\_X

# 

# Loading to MongoDB

* Convert merged dataframe to dictionary before uploading to MongoDB.

