# M5 Forecasting – Accuracy

## Intermediate status report

Team: Economic Science Squad

Ismail Aigunov          iaaygunov@edu.hse.ru
Xianghao Li             xli@edu.hse.ru

# Introduction

Xianghao Li                    Mainly responsible for Introduction and Data.

Ismail Aigunov                 Mainly responsible for Exploratory and Descriptive Analysis

Repository: https://github.com/XHaoLi/Project-MLDM

**The M5 Competition**

**Background:** The Makridakis Open Forecasting Center (MOFC) at the University of Nicosia conducts cutting-edge forecasting research and provides business forecast training. The MOFC is well known for its Makridakis Competitions, the first of which ran in the 1980s. M5 competition is the fifth iteration.

The **objective** of M5 competition is using hierarchical sales data from Walmart to forecast daily sales for the next 28 days.
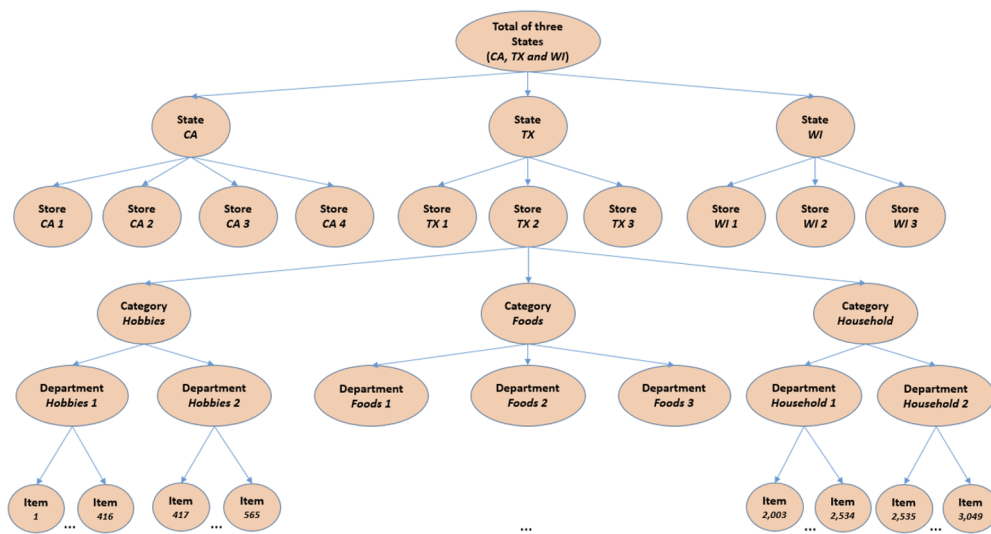
The **data**, covers stores in three US States (California, Texas, and Wisconsin) and includes item level, department, product categories, and store details.

M5 Forecasting - Accuracy: https://www.kaggle.com/competitions/m5-forecasting-accuracy/overview

# Data

The M5 dataset made available by **Walmart**, involves the unit sales of various products sold in the USA.

Organized in the form of **grouped time series**, involves the unit sales of **3,049 products**, classified in **3 product categories** (Hobbies, Foods, and Household) and **7 product departments**.  The products are sold across **ten stores**, located in **three States** (CA, TX, and WI).



3

# Data

The historical data range from **2011-01-29** to **2016-06-19**. The products have a (maximum) selling history of 1,941 days / 5.4 years (**test data of h=28 days not included**).

The dataset consists three (3) files: ***"calendar.csv"***, ***"sell_prices.csv"*** and ***"sales_train.csv"***.
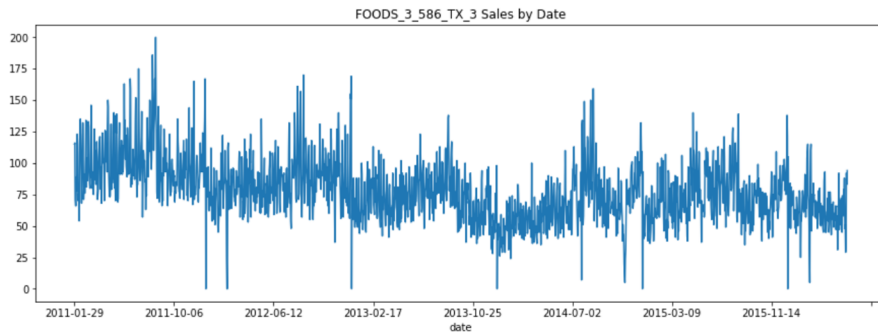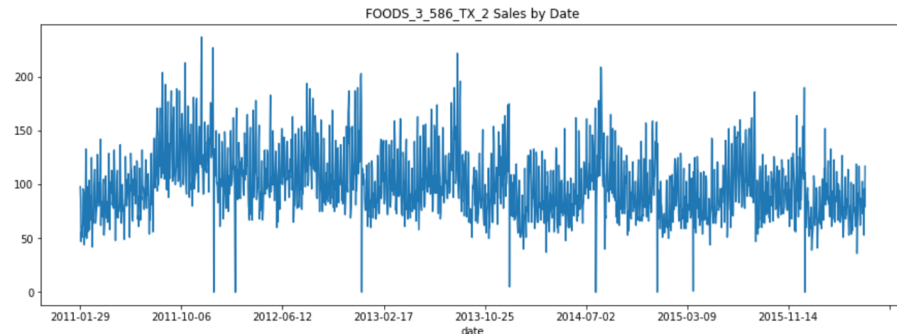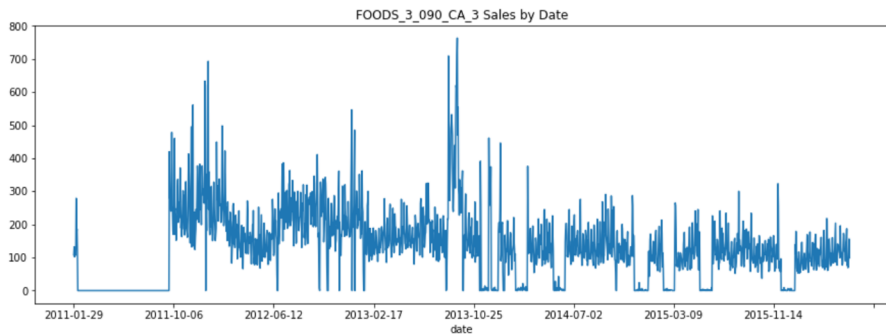
- ***"calendar.csv"*** - information about the dates the products are sold.

- ***"sell_prices.csv"*** - information about the price of the products sold per store and date.

- ***"sales_train.csv"*** - the historical daily unit sales data per product and store.

Details can be viewed in the file "Data.ipynb"
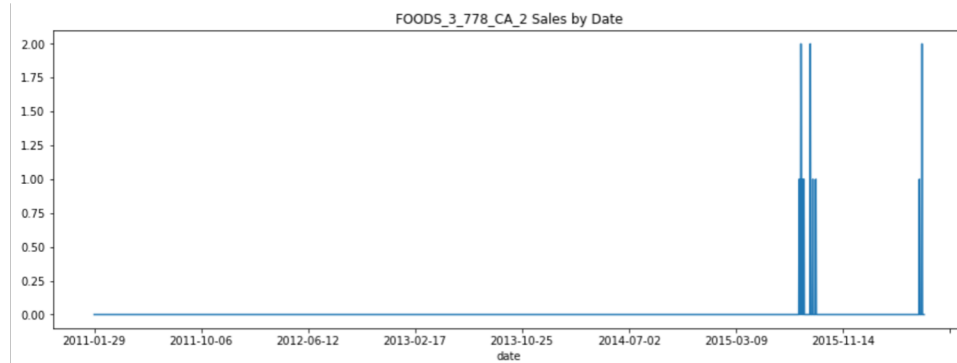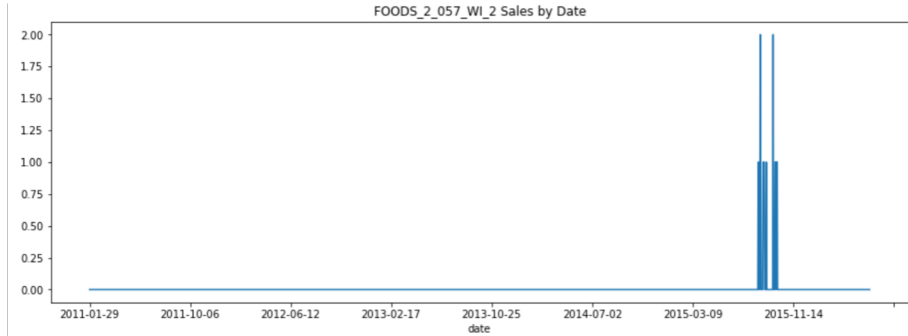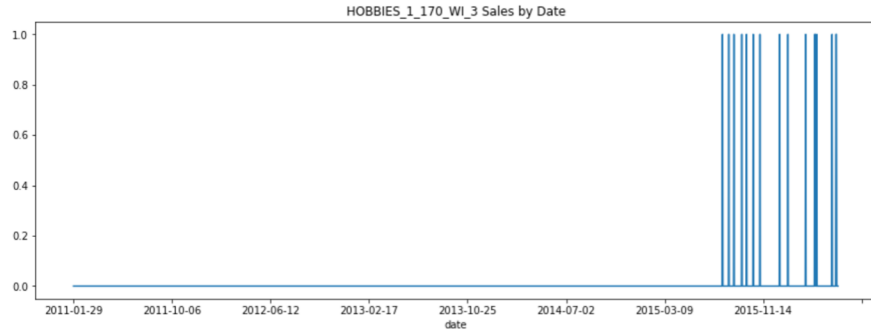
https://github.com/XHaoLi/Project-MLDM/blob/main/Data.ipynb

# Exploratory and Descriptive Analysis

# Top 3 units by Sales volume



FOODS_3_090_CA_3 Sales by Date



FOODS_3_586_TX_2 Sales by Date



FOODS_3_586_TX_3 Sales by Date
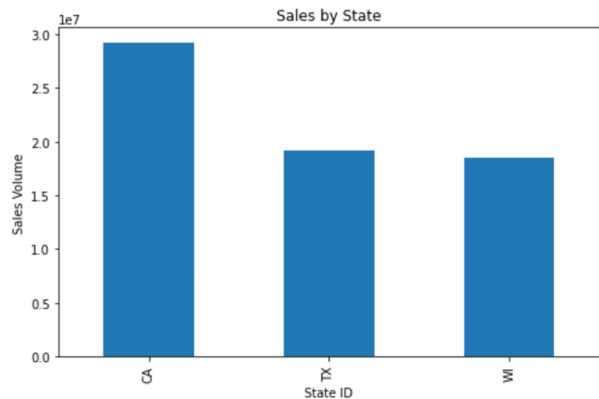
# Last 3 Units by Sales Volume

# Sales Volume by State, Stores and Categories



Total amount of sales account for 66 927 173 units