# 40.012 MANUFACTURING AND SERVICE OPERATIONS

## Assignment 1

**Michael Hoon**
1006617

May 21, 2024

# Question 1

The four elements of production and operations strategy is as follows:

- Time Horizon
- Focus
- Evaluation
- Consistency

## Time Horizon

In a typical manufacturing process, the time horizon in production and operations strategy is categorized into short, medium, and long-term decision horizons.

- **Short-Term**: Spans a period of hours or days, and focuses on immediate operational decisions such as workforce scheduling, inventory management (purchasing), and meeting current customer demands. For example, a car manufacturer like General Motors might plan a short-term strategy to ramp up production for a new model's launch.
- **Medium-Term**: Spans weeks or months, and includes tactical decisions like capacity distribution planning, manpower development, and medium-term forecasting strategies. An electronics manufacturer might plan for the introduction of new technologies and the scaling up of production lines in response to expected market trends (based on forecasts).
- **Long-Term**: Project spans one year (or longer) and involves strategic decisions such as investment in new facilities, long-term partnerships and sales channels, and research and development. A pharmaceutical company, for instance, might invest in a new research facility to develop innovative drugs expected to be needed in the next decade, as well as marketing strategies to sell their product.

## Focus

The focus element determines areas where the company aims to excel:

- **Process Technologies**: Investing in advanced manufacturing technologies to improve efficiency and reduce costs.
- **Market Demands**: Tailoring operations to meet changing customer preferences and market trends.
- **Production Volume**: Deciding whether to produce in large volumes for economies of scale or in smaller batches for customization.
- **Quality Level**: Prioritizing high quality to differentiate products in the market, and obtain a larger share of consumers.

- **Manufacturing Tasks**: Aligning operations with specific tasks such as just-in-time production or lean manufacturing practices.

## Evaluation

Evaluation involves assessing performance through various metrics, such as:

- **Cost**: Analyzing production costs to identify areas for improvement and cost reduction. For example, a food processing plant might evaluate the cost per unit of production to find ways to minimize waste and improve efficiency.
- **Quality**: Measuring product quality to ensure it meets the required standards and customer expectations.
- **Profitability**: Assessing the financial performance to ensure the company remains profitable.
- **Customer Satisfaction**: Gathering feedback to understand customer satisfaction levels and identify areas for improvement.

## Consistency

Consistency ensures alignment across various aspects of production and operations, such as:

- **Professionalism**: Maintaining high standards of professionalism in all operations.
- **Product Proliferation**: Managing a diverse product range effectively without compromising operational efficiency.
- **Manufacturing Tasks**: Making manufacturing tasks explicit and ensuring they are aligned with overall strategy.

# Question 2

## (a)

The Product Process Matrix (PPM) is used to analyze the relationship between the type of product being produced and the type of process being used to produce it. It helps us understand the implications of business choices regarding product design and process selection. The PPM comprises of the Product Life Cycle (stages a product goes through from start-up to decline), and the Process Life Cycle (different methods or approaches used to transform inputs into outputs - job shop, assembly, flow etc.). We describe this in Figure 1 as a (grid) matrix with the product types listed on one axis and process types on the other. Each cell describes a combination of a product type and a process type, with elements in the main diagonal representing efficient processes.

## (b)

To give an example, consider the Product-Process Matrix shown below:

| | low | med | large |
|---|---|---|---|
| job-shop | Customised Artisanal Furniture | | |
| flow/batch | | Consumer Electronics | Mass-Produced Toys (Lego) |
| assembly | | Automated Food Production Line | |
| continuous flow | | | Automated Pharmaceutical Manufacturing |

Table 1: Product Process Matrix Example

## (c)

The main diagonal represents the alignment between the type of product and the type of process. Each cell on the diagonal indicates that the product and process types are well-suited to each other. The matrix helps companies understand the trade-offs between product variety and process flexibility. Operating on the **main diagonal is advantageous** as it is the most optimal and efficient utilisation of resources in a company, to match the requirements of its products, which reduces any inefficiencies in the production process. Furthermore, by using processes that align with the characteristics of their products, companies can ensure higher quality outcomes. For example, standardized processes are better suited for producing standardized products, resulting in consistent quality.

Operating **off-diagonal is disadvantageous** as there is a misalignment between the resources available and the requirements of the products being produced. For example, using

a high-volume, standardized process to produce highly customized products can lead to inefficiencies and higher costs. Furthermore, mismatched processes can result in lower product quality due to difficulties in meeting the specific requirements of the products. For instance, using a batch process to produce standardized products may lead to variations in quality between batches.

# Question 3

An Industrial Revolution is defined as the process of change from an agrarian and handicraft economy to one dominated by industry and machine manufacturing. These technological changes introduced novel ways of working and living and fundamentally transformed society [2]. There are currently only a total of 4 Industrial Revolutions known to man.

## First Industrial Revolution

Occurred from 1760 to 1840, this phase is characterised by the mechanization of textile production, the development of steam power, and the growth of iron and coal industries. It brought about significant changes in agriculture, transportation, and communication.

## Second Industrial Revolution

Occurred from 1870 to 1914, this period is characterized by advancements in steel production, electricity, and chemical industries. It led to the widespread adoption of mass production techniques, the rise of big businesses, and the development of transportation infrastructure such as railroads.

## Third Industrial Revolution

This period occurred from the late $20^{\text{th}}$ Century to the present. This phase is also known as the Digital Revolution or Information Age, and is marked by the proliferation of computers, telecommunications, and the internet. It has transformed economies and societies through automation, digitization, and the rise of information technology.

## Fourth Industrial Revolution

The Fourth Industrial Revolution is a recent term describing rapid technological advancement in the $21^{\text{st}}$ Century. A large part of this phase of industrial change is the joining of technologies such as Artificial Intelligence, Genetic Editing, and Advanced Robotics that blur the lines between the physical, digital, and biological worlds. Also known as "Industry 4.0", fundamental shifts are taking place in how the global production and supply network operates through ongoing automation of traditional manufacturing and industrial practices, largely fuelled by "Artificial Intelligence (AI)" and "Internet of Things (IoT)".

## Fifth Industrial Revolution

In fact, AI is a key component of the Fourth (and potentially the Fifth) Industrial Revolution, and its integration into various sectors is driving significant changes in how industries operate, transforming both manufacturing and service operations. AI is mainly characterised by **Automation and Robotics** (in Manufacturing processes, enabling smarter robots that can work alongside humans), **Data Analysis and Predictive Maintenance** (Algorithms can

analyse vast amounts of data from industrial processes), and **Enhanced Production Processes** (optimising production lines, managing supply chains, and improving Quality Control). As for service operations, AI can analyze customer behavior and predict future trends, enabling businesses to tailor their services, improve customer satisfaction, and increase operational efficiency. On a more personal angle, I am currently working with applications of Large Language Models (LLMs) in industries, and there have been major advancements recently in deploying such models for database integrations in Supply Chain industries. This will introduce a pipeline allowing for Natural Language user queries over a database (instead of SQL), allowing users to seamlessly ask questions about their comprehensive databases.

A summary of the Industrial Revolutions is given in the figure below [4]:

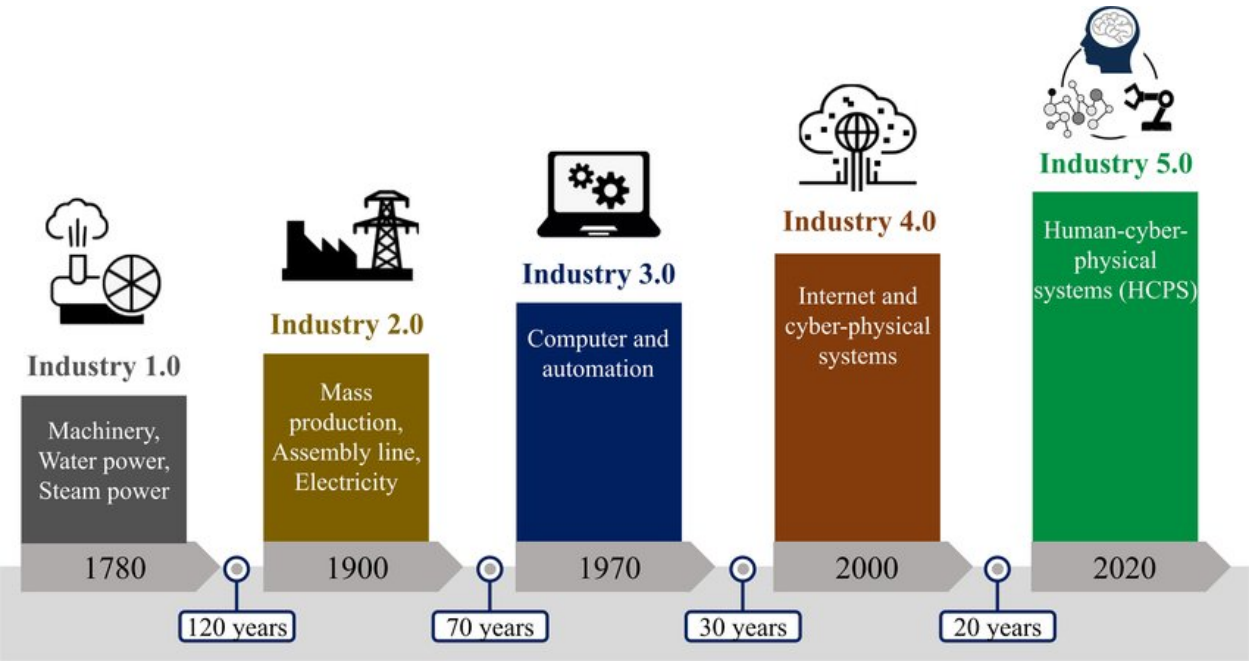


Figure 1: Industial Revolutions

# Question 4

Given a 75% learning curve, as the cumulative production doubles, the time required to produce a unit decreases to 75% of the time it took to produce the previous batch. Let $Y(u)$ be the number of labour minutes required for the $u^{\text{th}}$ unit. Since the first unit takes 5 minutes, we have:

$$Y(u) = 5u^{-b}$$

We also have that the time for the $2u^{\text{th}}$ unit is 75% of the $u^{\text{th}}$ unit, thus:

$$\begin{aligned}
\frac{Y(2u)}{Y(u)} &= 0.75 \\
\frac{5(2u)^{-b}}{5(u)^{-b}} &= 0.75 \\
-b \ln 2 &= \ln(0.75) \\
\therefore b &= -\frac{\ln(0.75)}{\ln 2} \\
&= 0.415037
\end{aligned}$$

The time required to produce the $100^{\text{th}}$ and $1000^{\text{th}}$ unit is thus:

$$\begin{aligned}
Y(100) &= 5 \times (100)^{-0.415037} \\
&= 0.739426 \\
&\approx \boxed{0.739 \text{ minutes}} \\
&\approx 44.4 \text{ seconds}
\end{aligned}$$

$$\begin{aligned}
Y(1000) &= 5 \times (1000)^{-0.415037} \\
&= 0.284352 \\
&\approx \boxed{0.284 \text{ minutes}} \\
&\approx 17.1 \text{ seconds}
\end{aligned}$$

# Question 5

The last two digits of my student ID is 17, and is correspondingly the letter "Q". Unfortunately, the only country in the entire world that starts with the letter "Q" is **Qatar** (in Asia). The 12th and 22nd letter of the alphabet is "L" and "V" respectively, and will be assigned "Luxembourg" (Europe) and "Venezuela" (South America). The 4th country is assigned as "Singapore". As much as possible, I have tried to select countries which adhere to "preferably in different continents" and "major global economies". We will study the GDP compositions of these 4 countries qualitatively and quantitatively.

## Qualitative Analysis

### Qatar

The state of Qatar is a country in West Asia, and its economy is one of the highest in the world based on GDP per capita, driven by its vast natural gas reserves. They primarily export to Japan, South Korea, India, and China. Recently, the country has strategically diversified its economy away from oil by investing heavily in liquefied natural gas (LNG) production and exportation.

During a press conference held on Sunday, February 25 2024, at QatarEnergy's headquarters in Doha, the company's CEO and Qatar's Minister of State for Energy Affairs disclosed the firm's plans to proceed with a new LNG expansion project, called the North Field West (NFW), to further raise the country's LNG production capacity by almost 85% from current production levels before the end of this decade [14]. By expanding into LNG, Qatar reduces its dependence on oil revenues, thereby diversifying its economy. This diversification helps mitigate the risks associated with fluctuations in oil prices, providing a more stable economic foundation.

### Luxembourg

Luxembourg has a highly developed and diverse economy, characterized by a strong focus on financial services, steel production, and technology. It is known for its favorable business environment, political stability, and strategic location within the European Union. Banking is the largest sector in the Luxembourg economy.

Having played an instrumental role in the creation of the Eurodollar markets in the 1960s, Luxembourg's banks have long-standing expertise in serving corporate clients, notably in the provision of international loans (bilateral and syndicated) and treasury services [9].

### Venezuela

Venezuela is the 25th largest producer of oil in the world and the 8th largest member of the Organization of the Petroleum Exporting Countries (OPEC). Venezuela also manufactures and exports heavy industry products such as steel, aluminum, and cement. Venezuela's oil revenues accounted for a significant portion of government revenue and GDP. However, the

country has experienced severe economic instability and contraction due to political issues [3].

In recent years, Venezuela has suffered economic collapse, with output shrinking significantly and rampant hyperinflation contributing to a scarcity of basic goods, such as food and medicine. It has now been classified as a "Petrostate", where the government is highly dependent on fossil fuel income, power is concentrated, and corruption is widespread [3].

### Singapore

Singapore boasts a highly developed and diversified economy, characterized by a strong focus on trade, finance, and manufacturing. The country has positioned itself as a global hub for trade and finance, leveraging its strategic location, efficient infrastructure, and business-friendly policies.Singapore's economy is driven by exports, particularly electronics, chemicals, and biomedical products, which contribute significantly to GDP and employment [12].

In the decades after independence, Singapore rapidly developed from a low-income country to a high-income country. GDP growth in the city-state has been among the world's highest, at an average of about 7% since independence and topping 9.2% in the first 25 years [12]. With rapid industrialization in the 1960s catapulting the island nation's development trajectory, manufacturing became the main driver of growth.

In 2023, the main drivers of GDP growth were Information and Communications, Transportation and Storage, and Other Service Industries (from 2023 study by Ministry of Trade and Industry, Singapore).

## Quantitative Analysis

### Qatar

Qatar's GDP values are given in the Table below:

| | |
|---|---|
| Population, million | 2.9 |
| GDP, current US$ billion | 156.8 |
| GDP per capita, current US$ | 54069.0 |

Table 2: Qatar GDP values (2023)

Qatar's real GDP growth is given in the figure below.

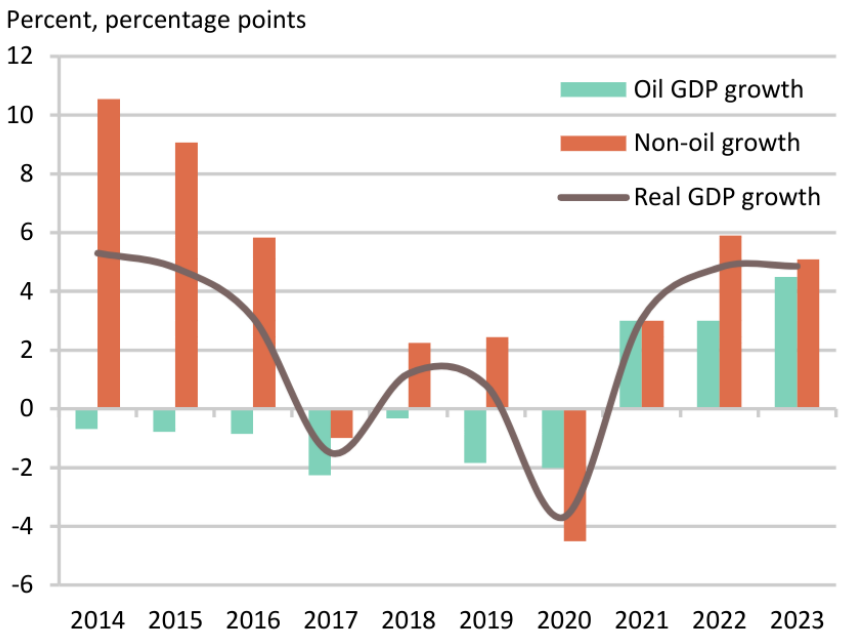


Figure 2: Qatar Real GDP Growth over years

**Luxembourg**

Luxembourg's GDP values are given in the table below:

| | |
|---|---|
| Population, million | 0.653 |
| GDP, current US$ billion | 81.64 |
| GDP per capita, current US$ | 125,006 |

Table 3: Luxembourg GDP values (as of 2022)

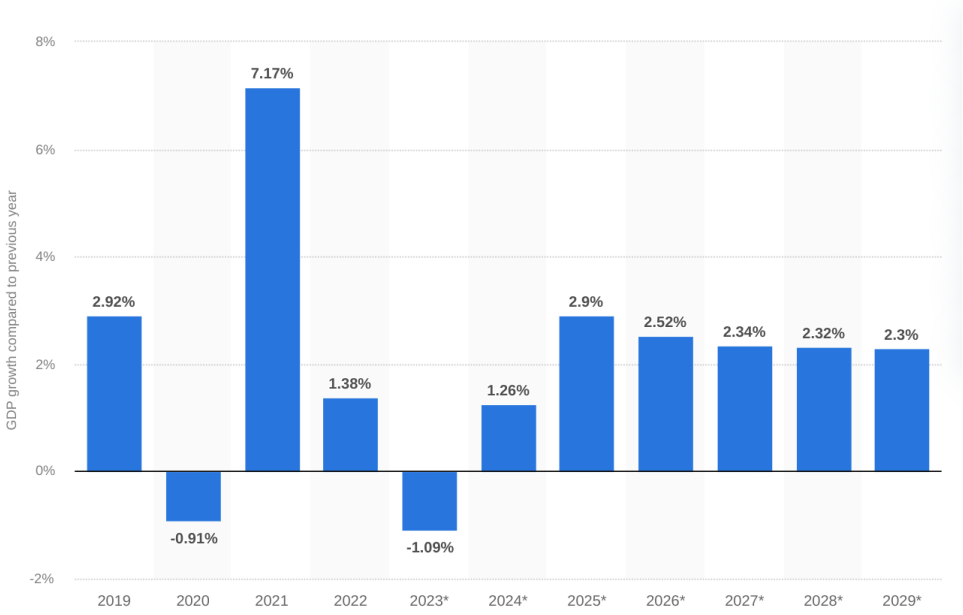


Figure 3: Luxembourg real GDP growth rate from Statista

## Venezuela

| | |
|---|---|
| Population, million | 29.4 |
| GDP, current US$ billion | 105.88 |
| GDP per capita, current US$ | 3867.44 |

Table 4: Venezuela's GDP values (as of 2024, from Statista)

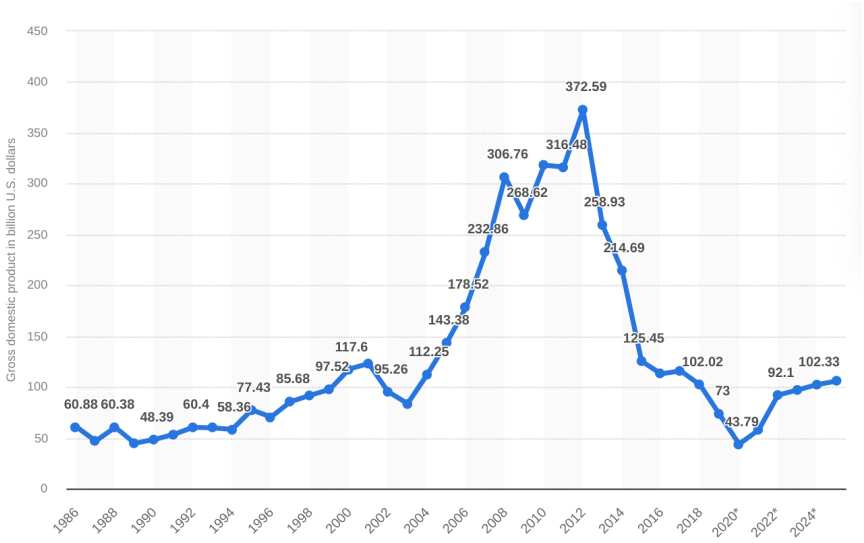


Figure 4: Venezuela GDP in current prices (from Statista)

**Singapore**

| | |
|---|---|
| Population, million | 6.01 |
| GDP, current US$ billion | 673 |
| GDP per capita, current US$ | 113,779 |

Table 5: Singapore's GDP values (as of 2023, from MTI Singapore)

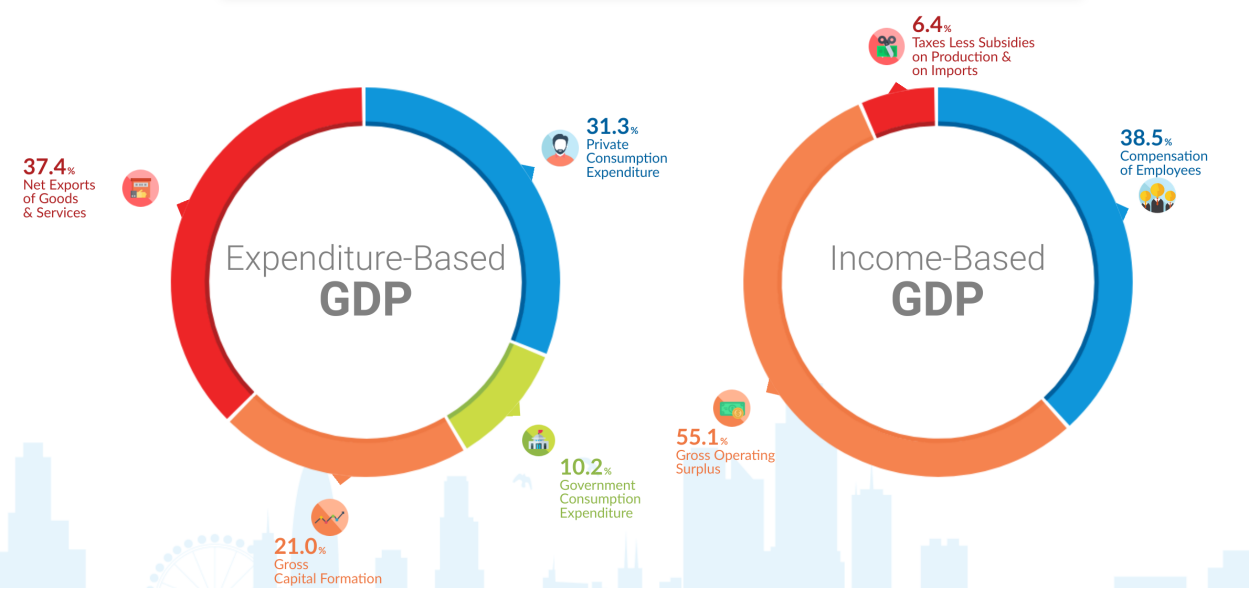


Figure 5: Singpore GDP Breakdown, from SingStat


Figure 6: Singpore GDP, from SingStat

# Question 6

## Some Background

The M-Competition is a series of competitions aimed at comparing the accuracy of different forecasting methods, and is named after the founder of the International Institute of Forecasters (IIT), Spyros Makridakis. The M1 competition was held in 1982, evaluating the performance of 24 forecasting methods on 1001 time series. Over the years, the competitions became increasingly more complex, featuring larger sets of time series with the M4 in 2018 expanding to 100,000 time series. It is only at this point where contemporary Machine Learning (ML) methods were introduced and tested, as these methods (especially Deep Learning (DL)) require larger datasets to achieve optimal performance.

These models have many parameters that need to be trained, and large datasets help to **avoid overfitting and improve generalisation**. With the M1 dataset that only has 1001 time series and around 100 observations each, it is simply **insufficient** and **inefficient** to run large ML models. Classic statistical methods such as ARIMA, Exponential Smoothing, and Linear Regression techniques work exceptionally well here to capture the patterns and trends in data with relatively few observations. This is further proven by a paper titled "Monash Time Series Forecasting Archive" by a team at Monash University [6], which experimented with different models for each of the M competition datasets:

| Dataset | SES | Theta | TBATS | ETS | (DHR-)ARIMA | PR | CatBoost | FFNN | DeepAR | N-BEATS | WaveNet | Transformer | Prophet | Informer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M1 Yearly | 4.938 | 4.191 | **3.499** | 3.771 | 4.479 | 4.588 | 4.427 | 4.355 | 4.603 | 4.384 | 4.666 | 5.519 | 5.633 | - |
| M1 Quarterly | 1.929 | 1.702 | 1.694 | **1.658** | 1.787 | 1.892 | 2.031 | 1.862 | 1.833 | 1.788 | 1.700 | 2.772 | 2.136 | - |
| M1 Monthly | 1.379 | 1.091 | 1.118 | **1.074** | 1.164 | 1.123 | 1.209 | 1.205 | 1.192 | 1.168 | 1.200 | 2.191 | 1.712 | - |
| M3 Yearly | 3.167 | **2.774** | 3.127 | 2.860 | 3.417 | 3.223 | 3.788 | 3.399 | 3.508 | 2.961 | 3.014 | 3.003 | 4.152 | - |
| M3 Quarterly | 1.417 | **1.117** | 1.256 | 1.170 | 1.240 | 1.248 | 1.441 | 1.329 | 1.310 | 1.182 | 1.290 | 2.452 | 1.672 | - |
| M3 Monthly | 1.091 | 0.864 | **0.861** | 0.865 | 0.873 | 1.010 | 1.065 | 1.011 | 1.167 | 0.934 | 1.008 | 1.454 | 1.375 | - |
| M3 Other | 3.089 | 2.271 | 1.848 | **1.814** | 1.831 | 2.655 | 3.178 | 2.615 | 2.975 | 2.390 | 2.127 | 2.781 | 4.694 | - |
| M4 Yearly | 3.981 | **3.375** | 3.437 | 3.444 | 3.876 | 3.625 | 3.649 | - | - | - | - | - | 5.256 | - |
| M4 Quarterly | 1.417 | 1.231 | 1.186 | **1.161** | 1.228 | 1.316 | 1.338 | 1.420 | 1.274 | 1.239 | 1.242 | 1.520 | 1.758 | - |
| M4 Monthly | 1.150 | 0.970 | 1.053 | **0.948** | 0.962 | 1.080 | 1.093 | 1.151 | 1.163 | 1.026 | 1.160 | 2.125 | 1.367 | - |
| M4 Weekly | 0.587 | 0.546 | 0.504 | 0.575 | 0.550 | 0.481 | 0.615 | 0.545 | 0.586 | **0.453** | 0.587 | 0.695 | 1.049 | - |
| M4 Daily | 1.154 | 1.153 | 1.157 | 1.239 | 1.179 | 1.162 | 1.593 | **1.141** | 2.212 | 1.218 | 1.157 | 1.377 | 3.698 | - |
| M4 Hourly | 11.607 | 11.524 | 2.663 | 26.690 | 13.557 | **1.662** | 1.771 | 2.862 | 2.145 | 2.247 | 1.680 | 8.840 | 1.776 | - |

Figure 7: Time Series Evaluations by Monash University

Evidently from Figure 7, we see that for the **M1-M3 datasets** [6], traditional statistical models such as TBATS (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components.), ETS (Exponential Smoothing), and Theta seem to dominate the performance metrics (lower is better - using sMAPE, symmetric Mean Absolute Percentage Error). Conversely, we see that contemporary Deep Learning methods such as

Transformers, CatBoost (Categorical Gradient Boosting Trees) and Prophet (A DL-based time series forecasting method introduced by Meta) tend to **perform significantly better** than traditional methods.

## M5-Competition Details

If we are keen to experiment with DL-based time series forecasting methods, we will need to perform analysis on the M4 and beyond datasets. **I will be choosing to use the M5 dataset**, as in addition to the large amount of time series data available to train the model on (M4), M5 further involves hierarchical data, requiring models to predict sales at various levels of aggregation. For this competition, hierarchical sales data from Walmart is provided, and we are required to forecast 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. In addition, it has explanatory variables such as price, promotions, day of the week, and special events [1].

The M5 Competition uses a Weighted Root Mean Squared Scaled Error (RMSSE) metric for evaluation of the performance of models, designed to be scale invariant and symmetric:

$$RMSSE = \sqrt{\frac{\frac{1}{h}\sum_{t=n+1}^{n+h}(y_t - \hat{y}_t)^2}{\frac{1}{n-1}\sum_{t=2}^{n}(y_t - y_{t-1})^2}} \tag{1}$$

## Exploratory Data Analysis

As with every project on forecasting and data analytics, we need to first conduct Exploratory Data Analysis (EDA) on our dataset to have an initial idea of the type of data we are working with, before selection of an appropriate model.

### Data Loading

We start first by loading the data. As shown below, we are provided the following 3 datasets, `sales_train.csv`, `sell_prices.csv` and `calendar.csv` respectively:

| id | item_id | dept_id | cat_id | store_id | state_id | d_1 | d_2 | d_3 | d_4 |
|---|---|---|---|---|---|---|---|---|---|
| HOBBIES_1_001_CA_1_validation | HOBBIES_1_001 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 0 | 0 |
| HOBBIES_1_002_CA_1_validation | HOBBIES_1_002 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 0 | 0 |
| HOBBIES_1_003_CA_1_validation | HOBBIES_1_003 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 0 | 0 |
| HOBBIES_1_004_CA_1_validation | HOBBIES_1_004 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 0 | 0 |
| HOBBIES_1_005_CA_1_validation | HOBBIES_1_005 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 0 | 0 |
| HOBBIES_1_006_CA_1_validation | HOBBIES_1_006 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 0 | 0 |
| HOBBIES_1_007_CA_1_validation | HOBBIES_1_007 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 0 | 0 |
| HOBBIES_1_008_CA_1_validation | HOBBIES_1_008 | HOBBIES_1 | HOBBIES | CA_1 | CA | 12 | 15 | 0 | 0 |
| HOBBIES_1_009_CA_1_validation | HOBBIES_1_009 | HOBBIES_1 | HOBBIES | CA_1 | CA | 2 | 0 | 7 | 3 |
| HOBBIES_1_010_CA_1_validation | HOBBIES_1_010 | HOBBIES_1 | HOBBIES | CA_1 | CA | 0 | 0 | 1 | 0 |

Figure 8: Training Sales Dataframe

| store_id | item_id | wm_yr_wk | sell_price |
|---|---|---|---|
| CA_1 | HOBBIES_1_001 | 11325 | 9.58 |
| CA_1 | HOBBIES_1_001 | 11326 | 9.58 |
| CA_1 | HOBBIES_1_001 | 11327 | 8.26 |
| CA_1 | HOBBIES_1_001 | 11328 | 8.26 |
| CA_1 | HOBBIES_1_001 | 11329 | 8.26 |
| CA_1 | HOBBIES_1_001 | 11330 | 8.26 |
| CA_1 | HOBBIES_1_001 | 11331 | 8.26 |
| CA_1 | HOBBIES_1_001 | 11332 | 8.26 |
| CA_1 | HOBBIES_1_001 | 11333 | 8.26 |
| CA_1 | HOBBIES_1_001 | 11334 | 8.26 |

Figure 9: Sales Prices Dataframe

| date | wm_yr_wk | weekday | wday | month | year | d | event_name_1 | event_type_1 | event_name_2 | event_type_2 | snap_CA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011-01-29 | 11101 | Saturday | 1 | 1 | 2011 | d_1 | NA | NA | NA | NA | 0 |
| 2011-01-30 | 11101 | Sunday | 2 | 1 | 2011 | d_2 | NA | NA | NA | NA | 0 |
| 2011-01-31 | 11101 | Monday | 3 | 1 | 2011 | d_3 | NA | NA | NA | NA | 0 |
| 2011-02-01 | 11101 | Tuesday | 4 | 2 | 2011 | d_4 | NA | NA | NA | NA | 1 |
| 2011-02-02 | 11101 | Wednesday | 5 | 2 | 2011 | d_5 | NA | NA | NA | NA | 1 |
| 2011-02-03 | 11101 | Thursday | 6 | 2 | 2011 | d_6 | NA | NA | NA | NA | 1 |
| 2011-02-04 | 11101 | Friday | 7 | 2 | 2011 | d_7 | NA | NA | NA | NA | 1 |
| 2011-02-05 | 11102 | Saturday | 1 | 2 | 2011 | d_8 | NA | NA | NA | NA | 1 |

Figure 10: Calendar Data

The data comprises 3049 individual products from 3 categories and 7 departments, sold in 10 stores in 3 states. The hierachical aggregation captures the combinations of these factors. We note that there are also no missing values as well, although there are a significant number of zeroes, indicating days where there are no sales at all.

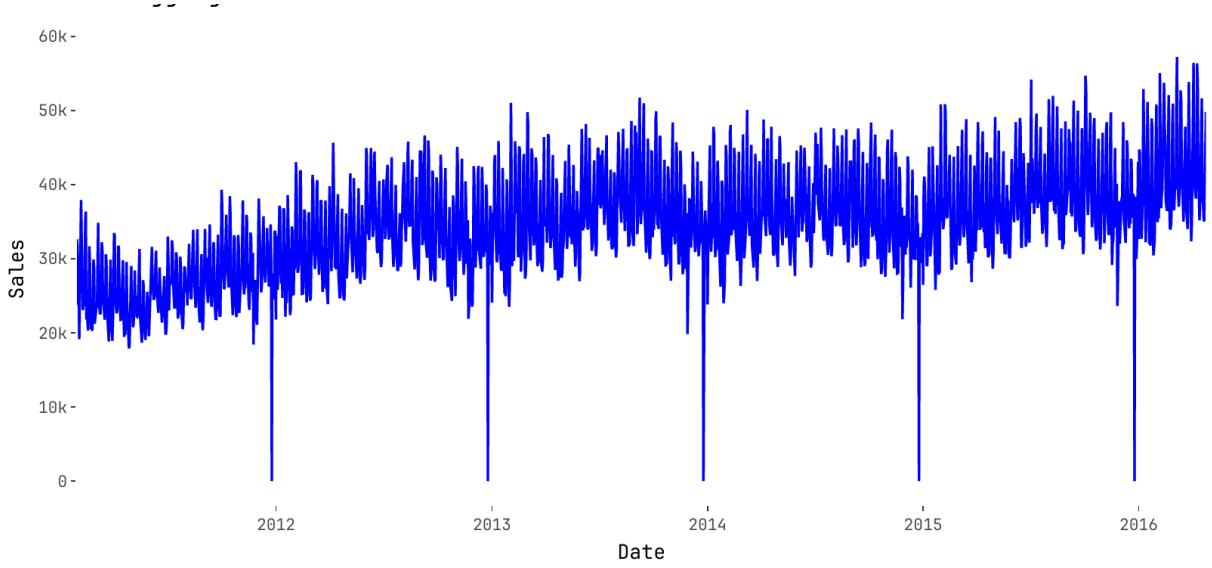


Figure 11: Aggregate Sales of all items across all years

Plotting the aggregate time series over all items, stores, categories, departments and sales, we see that there is strong weekly seasonality and a general upward trend in sales over the years.

### Seasonality Analysis

Using a smoothed LOESS (local regression) technique, we can model the upward trend of the data as shown in Figure 12 below:

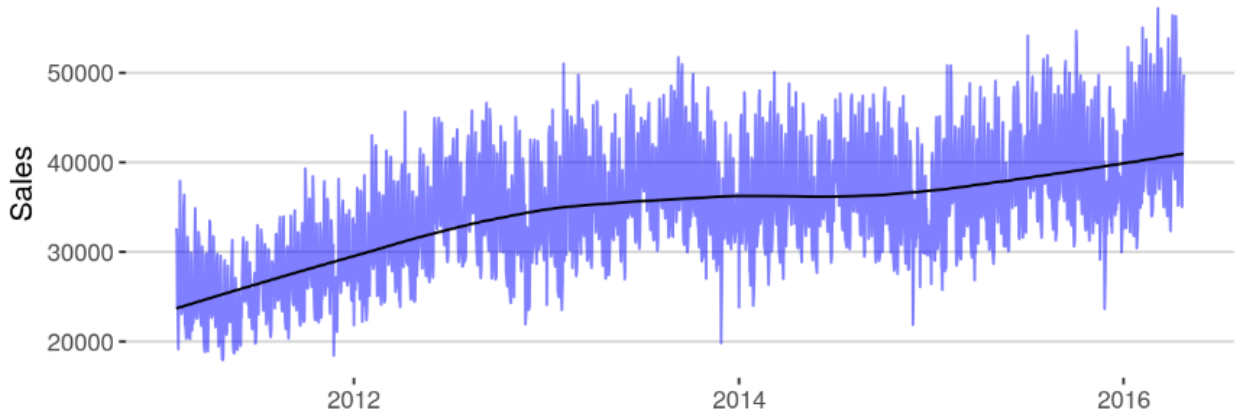


Figure 12: Total Sales with LOESS Smoothing

Additionally, to visualise the daily seasonal trend in sales, we can use a heatmap to plot the Days of the week against the Month of the year for all data points:

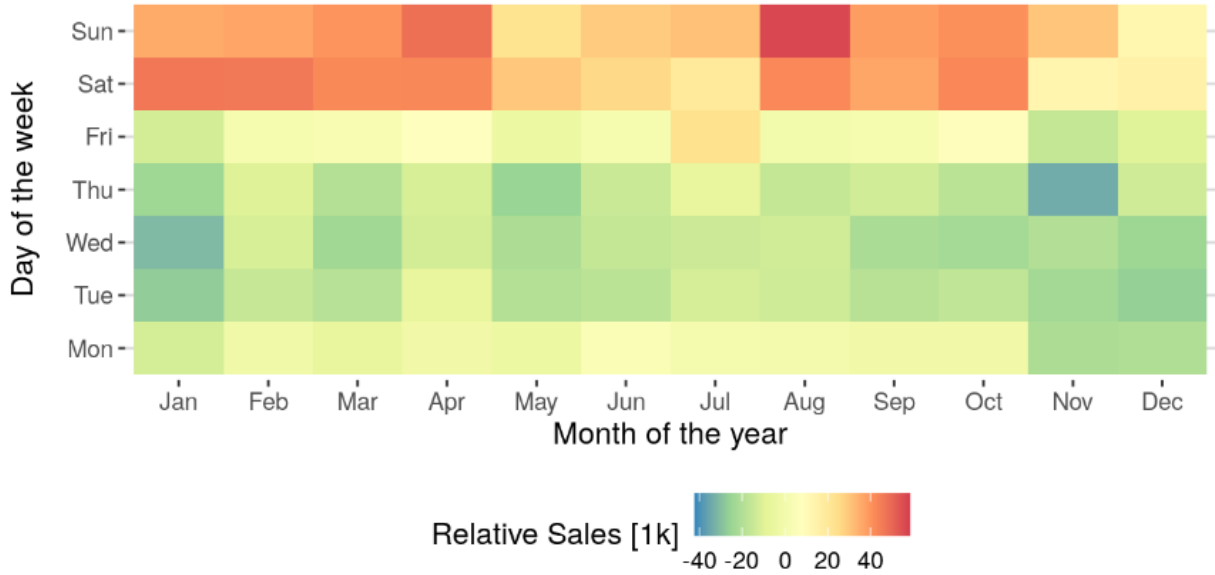


Figure 13: Seasonality Heatmap

Notably, there are weekend spikes in the data and a significant dip in sales during November and December. We can further analyse on a deeper level by considering state-level and category-level seasonalities:

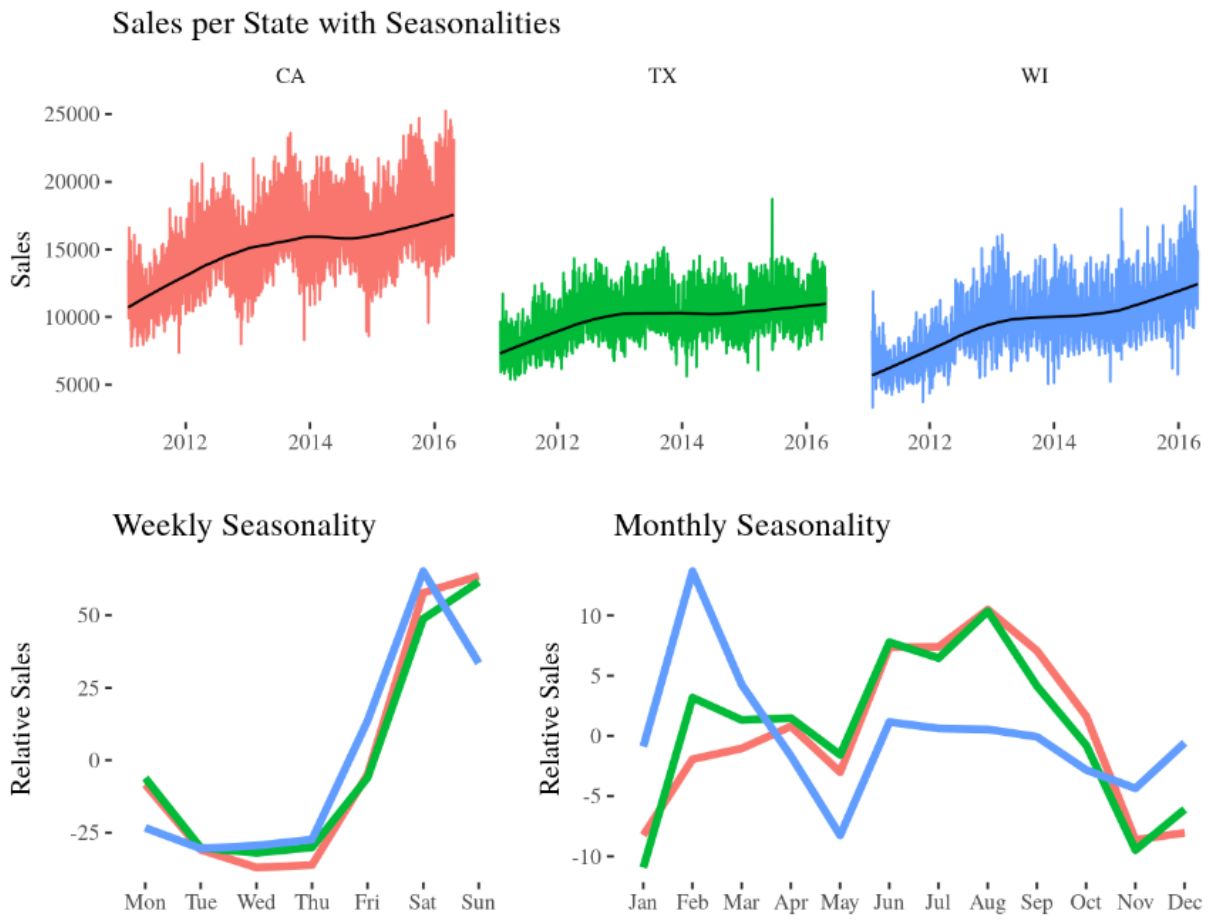


Figure 14: State-Level Seasonality Trends

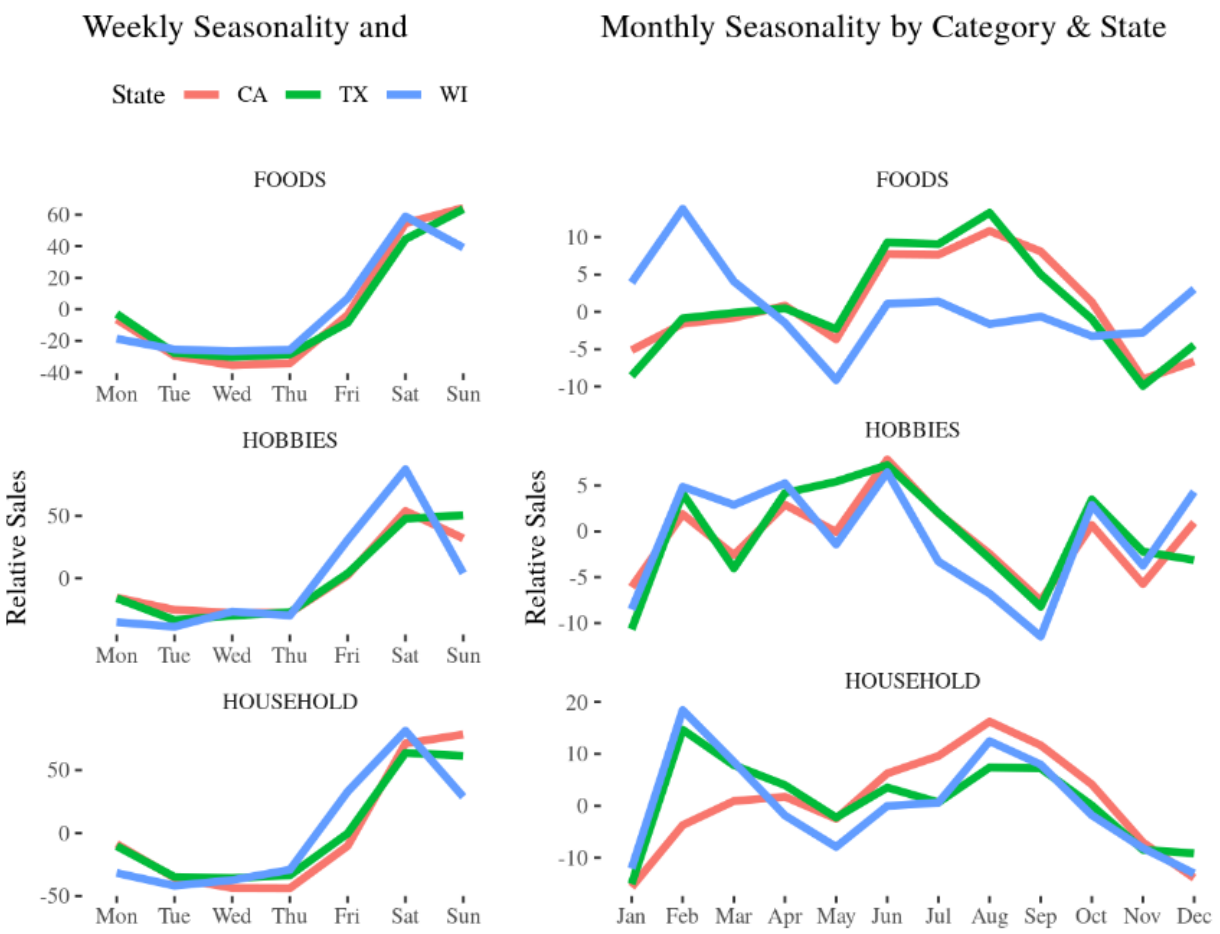


Figure 15: Category-Level Seasonality Trends

We can see that there is no consistent trends between seasonality of the different categories, as expected. Here, the individual time series is scaled by its global mean to allow for better comparison between the 3 states. More information about seasonality and the EDA process can be found in this hyperlink.

### Explanatory Variables

Next, we proceed to analyse the two explanatory variables given: item prices and calendar events. We first cover item prices over the years, as we have information for each item ID, including the `category`, `department`, and `store` IDs. The figure below shows a facet grid with overlapping Kernel Density Estimate (KDE) plots for price distributions for each of the years.

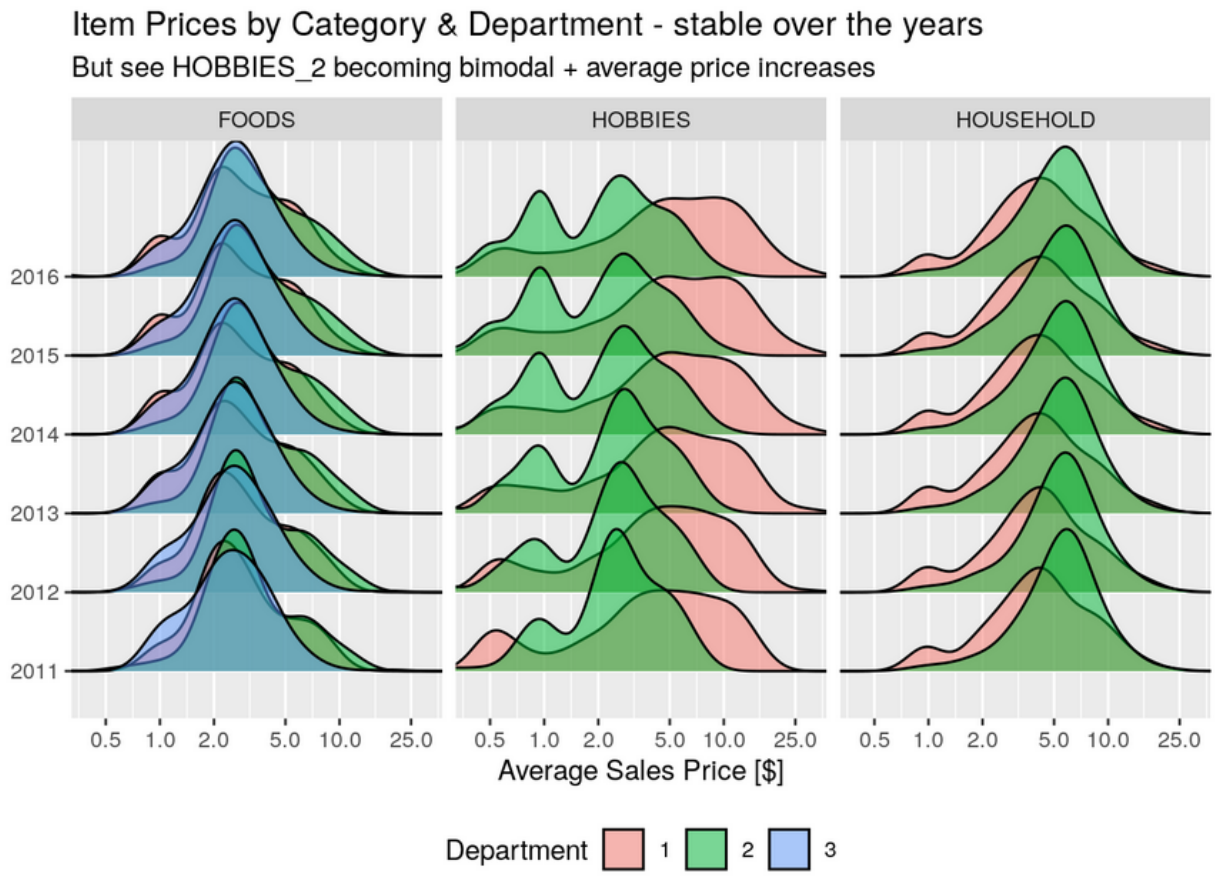


Figure 16: Item Prices by Category and Department

From Figure 16, we see that the item prices generally remain stable over the years (excluding inflation), with Hobbies exhibiting a **bimodal graph**, with increasing average price. Similarly, if we plot for each of the IDs (note the logarithmic scale on the $x$-axis):

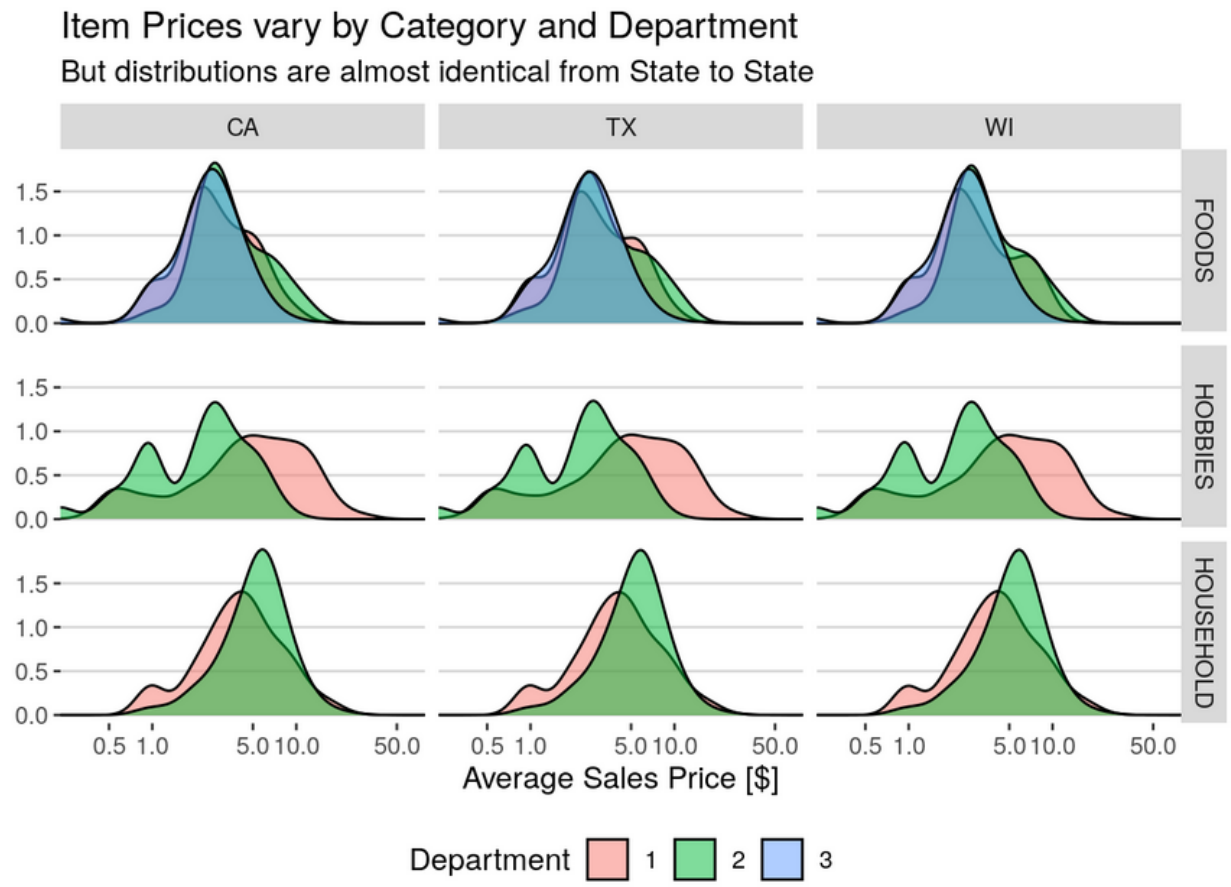


Figure 17: Item Prices vary by Category and Department

We can see that the distributions are mostly identical between states, and Food on average seems to be cheaper than Household items.

## Deep Learning Methods

With the EDA process complete, we can now begin to build an appropriate model for forecasting. There are a multitude of models available to choose from with vastly different architectures, including the Recurrent Neural Network (RNNs), Transformer Models (Temporal Fusion Transformer / Informer), Neural Basis Expansion Analysis Time Series (N-BEATS), and Sequence to Sequence (Seq2Seq) models. A quick search on the Kaggle leaderboard shows that the top 10 submissions **do not in fact use contemporary DL methods such as RNNs (LSTM) or Transformer models**. Unsurprisingly, most of the better submissions use a combination of techniques that include traditional statistical methods, in conjuction with "older" Machine Learning models such as **Gradient Boosting frameworks** (specifically the LightGBM variant).

### Gradient Boosting Frameworks

Gradient Boosting methods such as LightGBM perform exceptionally well on forecasting tasks as they are particularly well-suited for structured (tabular) data, and is not prone to overfitting even with multiple feature engineering techniques (creating rolling means, lagged values etc). They are also robust to both categorical and numerical data, which the M5 dataset includes such as the IDs and sale prices. An important distinction between other forecasting datasets is that the M5 is specifically designed to have hierarchical features (products within stores) as mentioned previously, and LightGBM is able to forecast at multiple levels, giving a significant edge over other models such as Transformers. We can see that successful models tend to incorporate a mix of Cross-Validation, Hierarchical Modelling, and Feature Engineering techniques, which LightGBM is able to capture via ensembling methods.

Although it is not a "contemporary AI/DL" technique per se, ensembling methods evidently perform better than other methods here. A typical structure of the ensemble technique used by Gradient Boosting models is shown in Figure 18 here [13]:

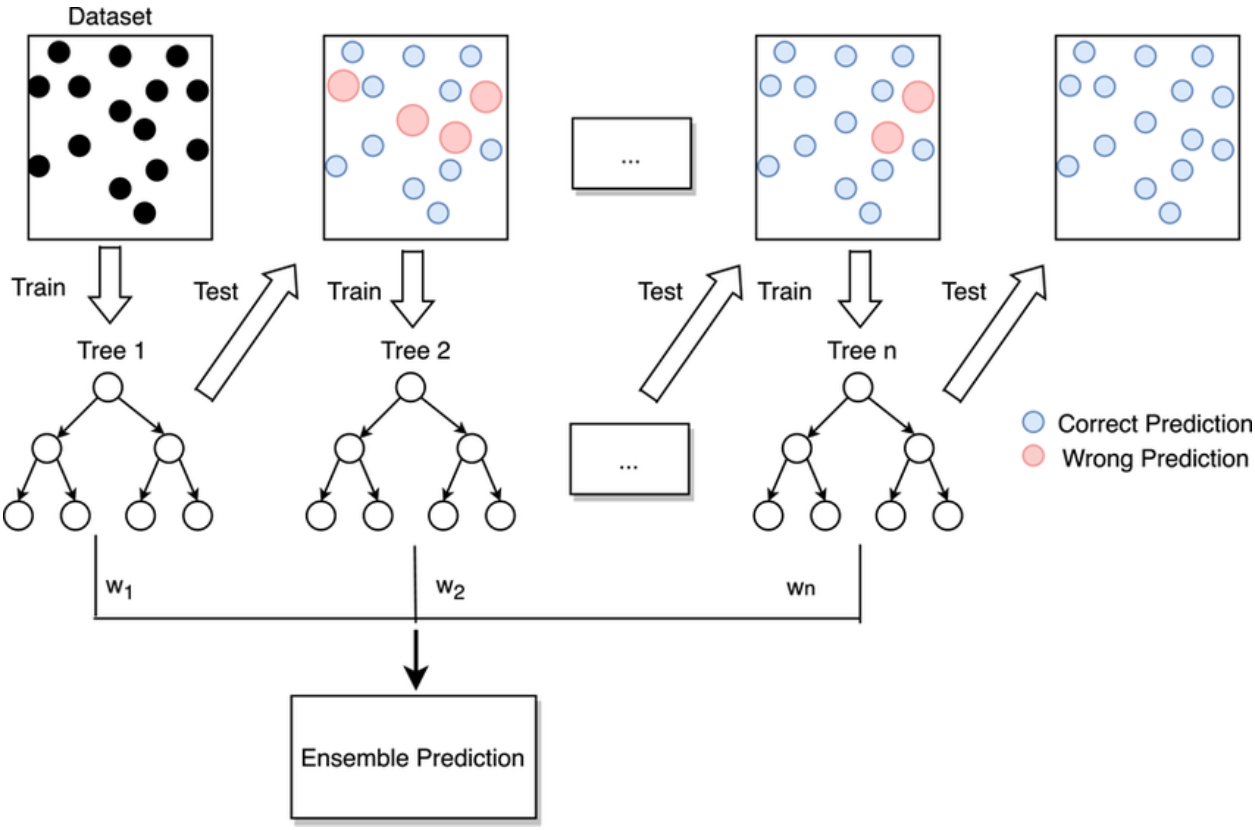


Figure 18: Gradient Boosting Algorithm

LightGBM is a particular variant of these models which is optimised for speed and efficiency, handling large datasets effectively. This technique trains models sequentially, to correct the errors of the combined ensemble using gradient descent-like methods to optimise a loss function.

## Actual Forecasting

Referencing techniques used by various sources online on building LightGBM models, the steps are detailed below.

### Core of the Model

We start with a LightGBM model with features based on past transactions up to the previous week. This means that to predict the 28 days ahead, the algorithm predicts week by week for 4 weeks, and uses the predictions from the previous week to generate the new features to predict the following week. However, as we have to forecast for 28 days, 4 weeks, the first week will use the true value from the previous week. The second week uses the true values for the features that are with a lag $> 7$, but has to use the predicted values for the features with a lag $< 7$, increasing the uncertainty, and leading to less accurate forecast for the following weeks.

To lower the impact, all the features based on the previous week are rolling features using at least a window of 7 days, to smooth the predictions. Thus, the predictions are not used directly as a lag value, which would give it too much importance when training the algorithm and being too inaccurate to predict, but they are used to construct rolling features over a period of 7, 14, 28 days. This technique is further explained in the hyperlink here.

### Feature Engineering

Additional features were created based on the following:

- Lag values for lag $> 28$ days, and rolling function with different window size
- Rolling mean and rolling standard deviation for the lag values $< 28$ days with different window size
- Features based on calendar such as day of the week, weekend, holidays, month, special event
- Release date of the item
- Price features, min, max, mean, sd, normalized, difference in price with previous day, number of shop with that price
- Mean encoding
- Various statistics

### Hyperparameter Tuning

In many of the winning solutions of the Kaggle competition, it is noted that the distribution of actual sales resembles that of a Tweedie distribution with variance power $p = 1.5$, thus an appropriate loss function to choose for the LightGBM model for hyperparameter tuning is the Tweedie loss function.

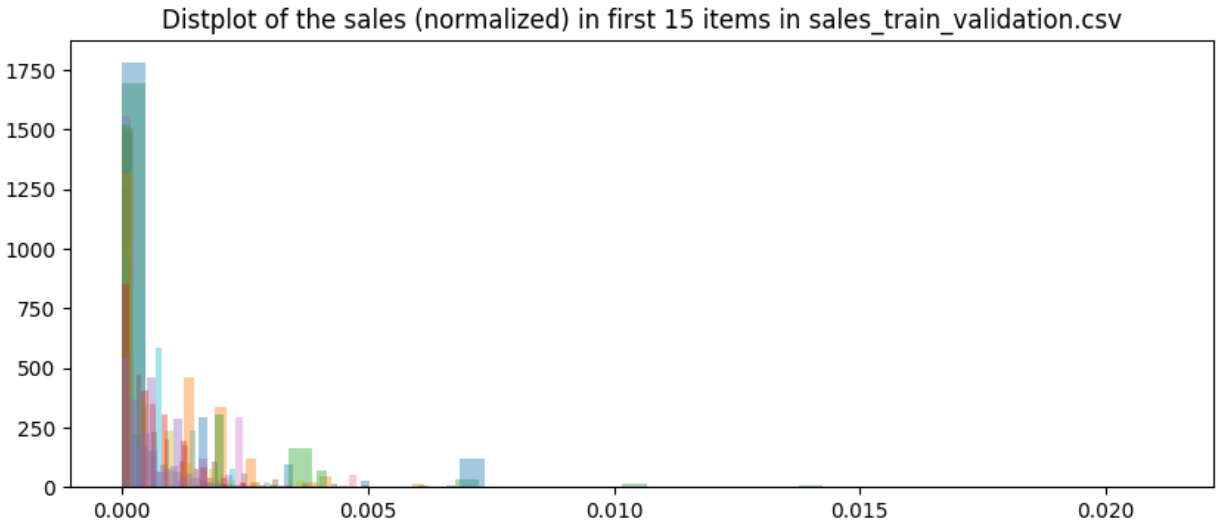


Figure 19: Distribution of plot of sales for first 15 items, resembling a Tweedie Distribution

LightGBM computes the Tweedie Loss Function as follows:

$$L(\hat{y}) = -y \cdot \frac{\hat{y}^{1-p}}{1-p} + \frac{\hat{y}^{2-p}}{2-p} \tag{2}$$

### Cross Validation

Although Cross-Validation (CV) is a widely-known technique to improve and generalise ML models (specifically the k-fold CV), it requires a significant amount of training time to perform (This method does not include CV techniques due to performance constraints). It is still important to acknowledge the importance of this as it provides significant improvements to models. The CV technique works by partitioning the dataset into subsets, training the model on a training set, and evaluating it on the validation set. This process is then repeated multiple times to ensure that the model performs well on different portions of the data and not just on a specific subset. The k-fold variant in particular divides the dataset into $k$ equally-sized folds, trained $k$ times ($k-1$ folds for training and 1 for validation), and evaluated over $k$ iterations.

**Results**

After running the model, we obtain the following forecasts, for the State, Category, and Item-levels accordingly:

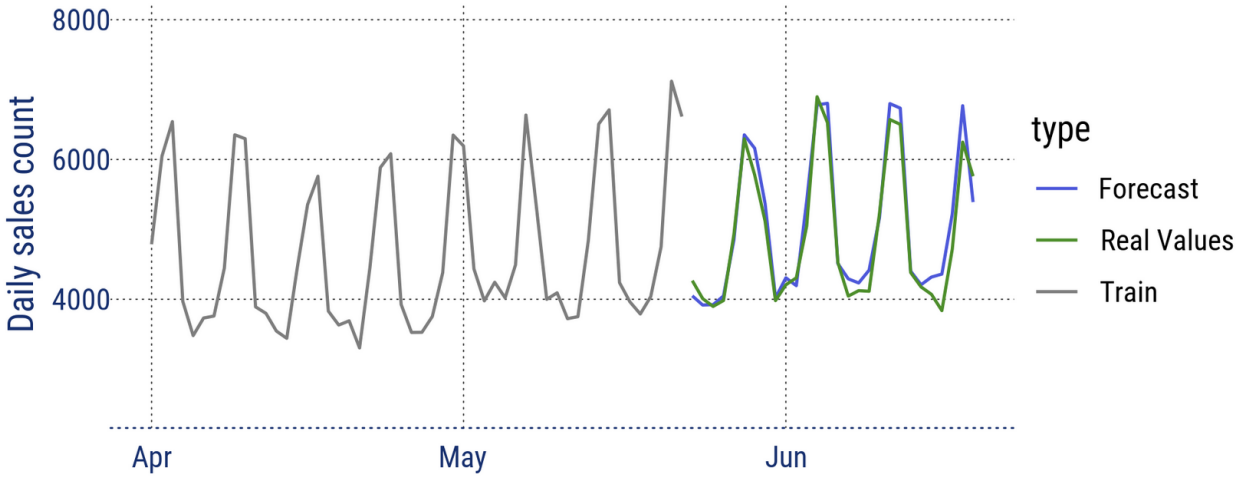


Figure 20: Forecast for all Stores in California

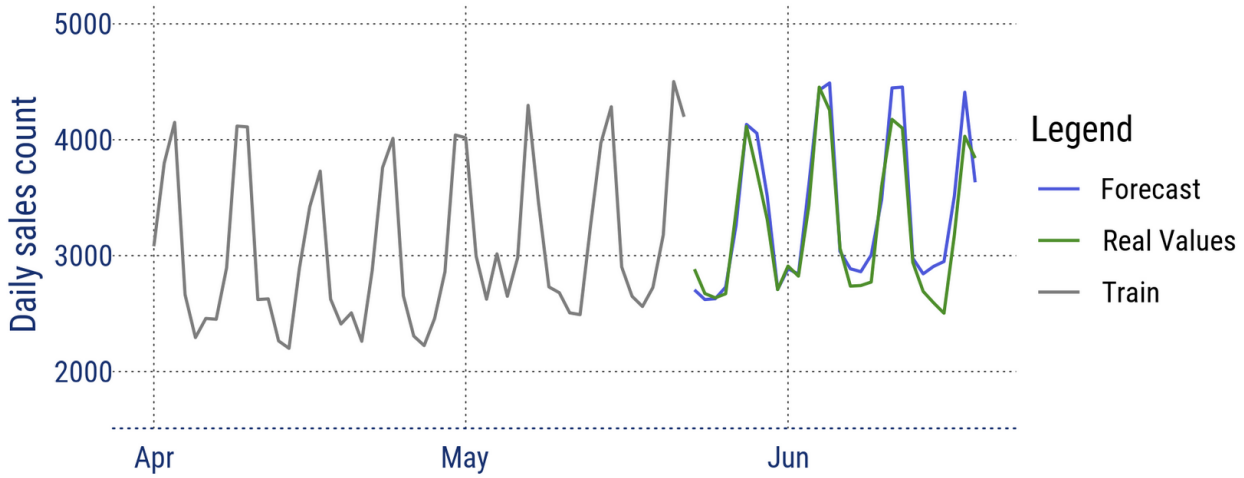


Figure 21: Forecast for Food category in store CA_2

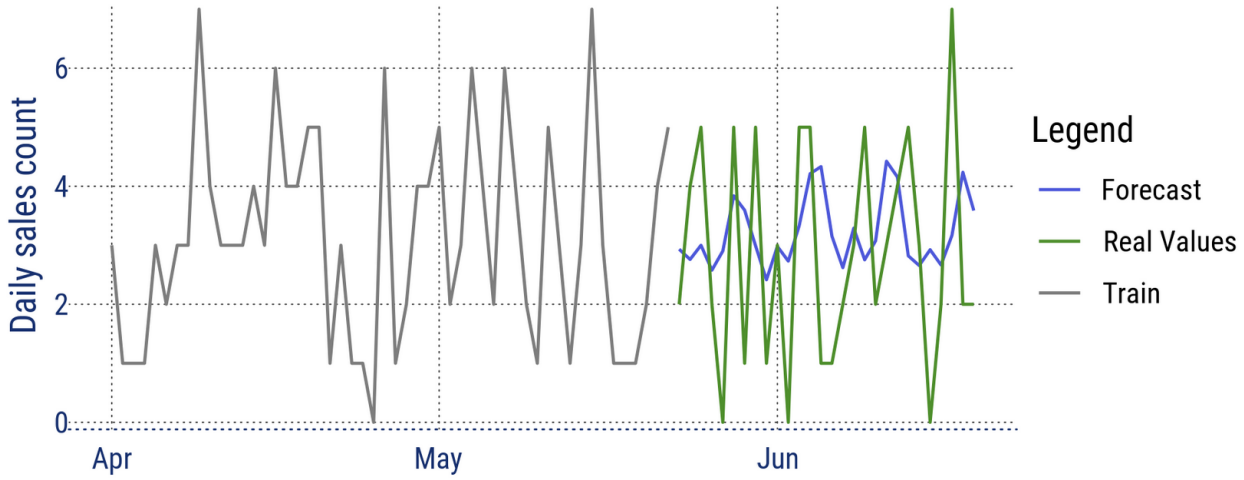


Figure 22: Forecast for specific item FOODS_3_065 in all stores in California

From the figures, we see that the forecast for all stores in California and a specific category seemed to perform relatively well, but the individual item forecast performed rather poorly (as expected, due to high variations). As for performance measures, this particular model obtained a **WMRSSE score of 0.48734**, which is vastly superior to that of traditional statistical methods such as ARIMA or ETS shown in Figure 7.

Further improvements to this model could be achieved via techniques such as CV as mentioned previously (with appropriate computing resources), and an automated hyperparameter tuning process (either via Grid Search or Bayesian Optimisation). A detailed discussion about time series methods post-competition can also be found in this link.

## Other Forecasting Techniques

Some other interesting winning solutions provided on Kaggle include the LSTM model (using a multi-layered approach), and Seq2Seq models. This is especially interesting as LSTMs and Seq2Seq models were not originally designed for time series forecasting (initially developed for Natural Language Processing (NLP) and Machine Translation), but they have now been specifically adapted due to their ability to handle sequential data and capture temporal dependencies in time series data [10].

An even more interesting result is that Generative Pre-Trained Transformers (GPTs) have also recently been developed and fine tuned for time series analysis [5] [7] (more information on Nixtla.io)

Furthermore, there have also been recent (a few weeks ago actually) advancements in a new form of neural network: the Kolmogorov-Arnold Network (KAN) [8]. These networks have also been adapted to experiment with time series forecasting, in the paper titled "Kolmogorov-Arnold Networks (KANs) for Time Series Analysis" [11].

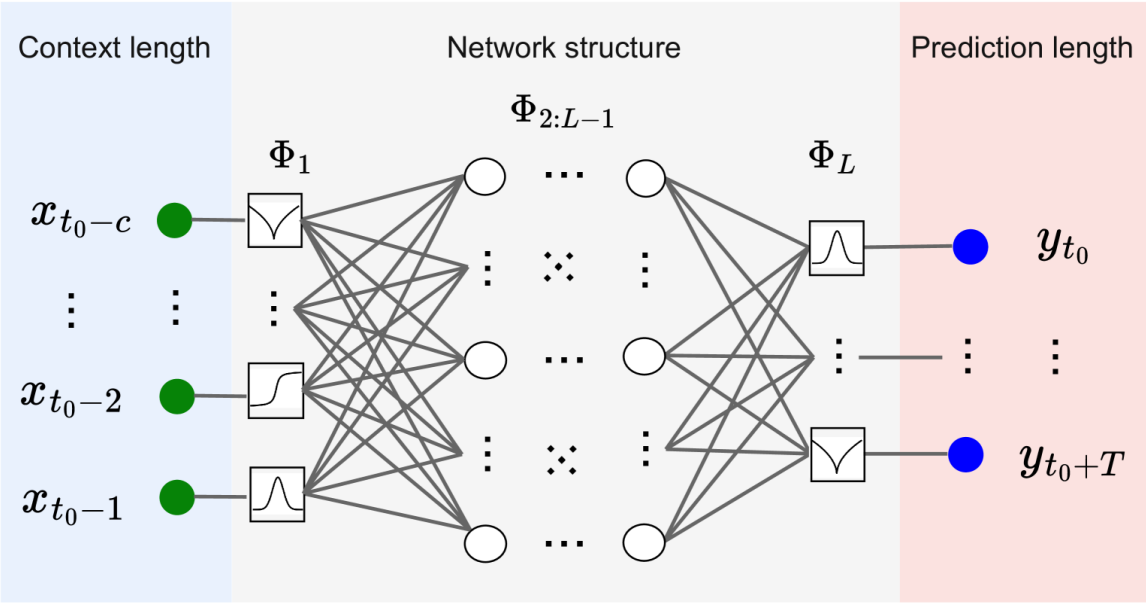


Figure 23: KAN Network Architecture for forecasting. Learnable activations are represented inside a square box.

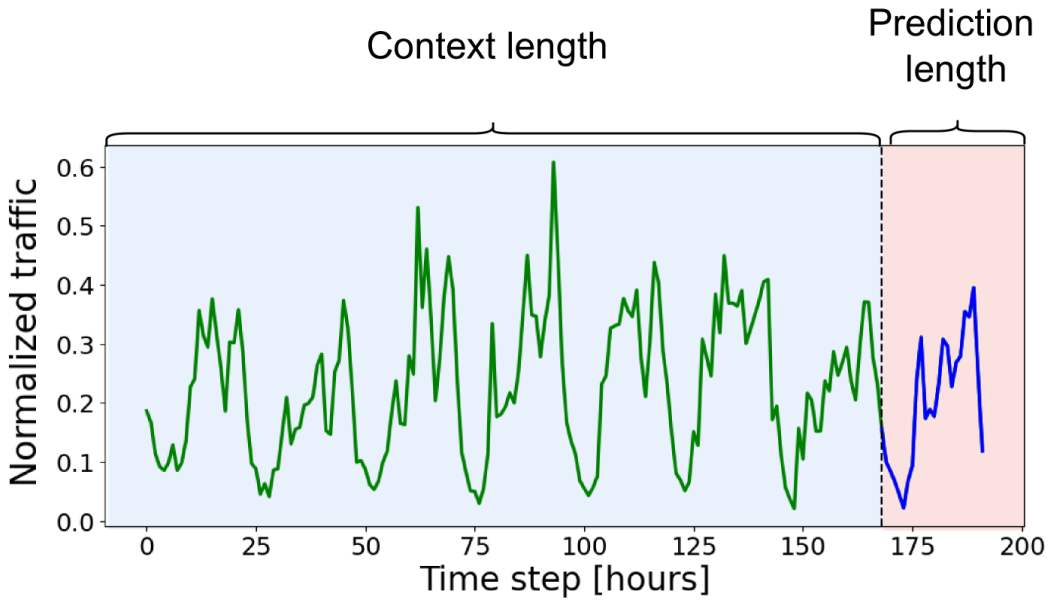


Figure 24: Example Time Series Data Prediction via KANs

# Code References

The code used for Question 6 was ran on a GPU instance in a Kaggle kernel.

Code for the EDA Process can be found in this hyperlink.

Code for the LightGBM Model can be found in this hyperlink.

# References

[1] Spyros Makridakis vangelis Addison Howard, inversion. M5 forecasting - accuracy, 2020. Accessed from `https://kaggle.com/competitions/m5-forecasting-accuracy` on 2024-04-01.

[2] Editors of Encyclopaedia The Britannica. Industrial revolution, Apr 2024. Accessed from `https://www.britannica.com/event/Industrial-Revolution` on 2024-04-01.

[3] Amelia Cheatham and Diana Roy. Venezuela: The rise and fall of a petrostate. Accessed from `https://www.cfr.org/backgrounder/venezuela-crisis` on 2024-04-01.

[4] Xiao Chen, Martin Eder, Asm Shihavuddin, and Dan Zheng. A human-cyber-physical system toward intelligent wind turbine operation and maintenance. *Sustainability*, 13:561, 01 2021.

[5] Azul Garza and Max Mergenthaler-Canseco. Timegpt-1, 2023.

[6] Rakshitha Godahewa, Christoph Bergmeir, Geoffrey I. Webb, Rob J. Hyndman, and Pablo Montero-Manso. Monash time series forecasting archive. In *Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.

[7] Wenlong Liao, Fernando Porte-Agel, Jiannong Fang, Christian Rehtanz, Shouxiang Wang, Dechang Yang, and Zhe Yang. Timegpt in load forecasting: A large time series model perspective, 2024.

[8] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks, 2024.

[9] For Finance Luxembourg. Banking - luxembourg financial centre, Aug 2023. Accessed from `https://www.luxembourgforfinance.com/en/financial-centre/banking/` on 2024-04-01.

[10] Christopher Olah. Understanding lstm networks, Aug 2015.

[11] Cristian J. Vaca-Rubio, Luis Blanco, Roberto Pereira, and Màrius Caus. Kolmogorov-arnold networks (kans) for time series analysis, 2024.

[12] Group World Bank. The world bank in singapore. Accessed from `https://www.worldbank.org/en/country/singapore/overview` on 2024-04-01.

[13] Tao Zhang, Wuyin Lin, Andrew Vogelmann, Minghua Zhang, Shaocheng Xie, Yi Qin, and Jean-Christophe Golaz. Improving convection trigger functions in deep convective parameterization schemes using machine learning. *Journal of Advances in Modeling Earth Systems*, 13, 05 2021.

[14] Melisa Čavčić. New gas expansion project to magnify qatar's lng production by nearly 85 Accessed from `https://www.offshore-energy.biz/new-gas-expansion-project-to-magnify-qatars-lng-production-by-nearly-85/` on 2024-04-01.