ISA 444: Business Forecasting

01: Introduction to Time Series Analysis and Forecasting

Fadel M. Megahed, PhD

Professor
Farmer School of Business
Miami University

@FadelMegahed
fmegahed
fmegahed@miamioh.edu
? Automated Scheduler for Office Hours

Spring 2025

Learning Objectives for Today's Class

- Describe the role of forecasting in business.
- Describe the key components of a **time-series** (**trend**, **seasonality**, **multiple seasonality**, and **cycles**).
- Explain the concept of data-generating process (DGP)
- Discuss limits of forecasting
- Understand key forecasting terminology

Course Motivation

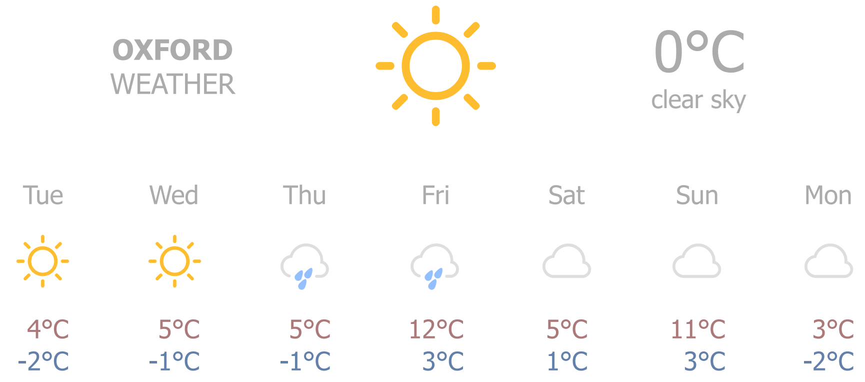
Forecasting Impacts Everything and Everyone

- **Businesses**: Sales forecasts set revenue targets, inventory projections optimize supply chains, and staffing plans ensure workforce readiness during demand fluctuations.
- **Gov.**: Tax revenue and social programs forecasts aid budgeting and resource allocation.
- **Individuals**: Financial forecasts support budgeting, saving, and retirement planning.
- **Weather**: Forecast inform agriculture, disaster prep, and daily decisions.

CBO's Baseline Budget Projections, by Category

	Actual, 2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	Total 2026–2030	Total 2026–2035
	In billions of dollars													
Revenues														
Individual income taxes	2,426	2,621	2,968	3,253	3,355	3,455	3,584	3,721	3,870	4,037	4,220	4,413	16,615	36,876
Payroll taxes	1,709	1,759	1,840	1,915	1,990	2,072	2,155	2,242	2,329	2,419	2,510	2,605	9,973	22,078
Corporate income taxes	530	524	495	469	462	453	450	456	468	493	503	517	2,329	4,767
Other[a]	253	259	277	298	301	310	360	416	439	456	474	496	1,545	3,827
Total	**4,918**	**5,163**	**5,580**	**5,935**	**6,108**	**6,290**	**6,549**	**6,834**	**7,106**	**7,405**	**7,708**	**8,031**	**30,463**	**67,548**
On-budget	3,658	3,859	4,217	4,516	4,637	4,760	4,959	5,180	5,389	5,623	5,860	6,114	23,088	51,254
Off-budget[b]	1,260	1,304	1,364	1,418	1,472	1,530	1,591	1,654	1,718	1,782	1,848	1,917	7,375	16,293

Source: Congressional Budget Office



Source: Weather Widget's Forecast

Microsoft's Missed Opportunity with Mobile Phones

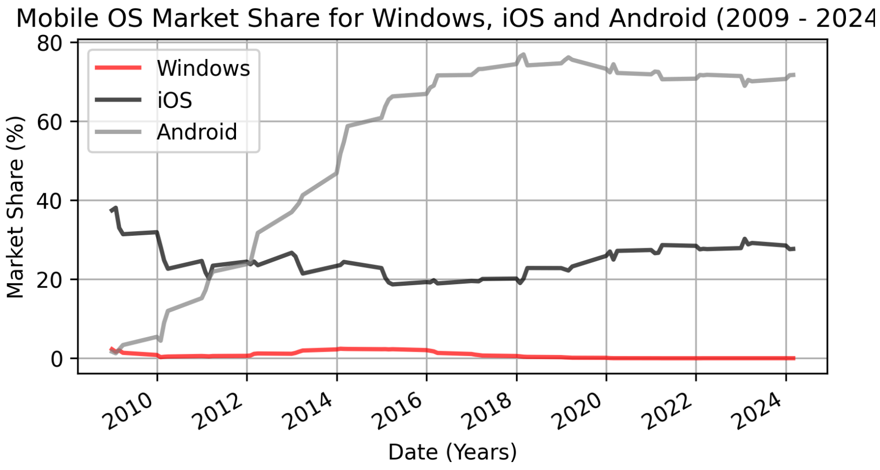
> **Q**: People get passionate when Apple comes out with something new—the iPhone; of course ... **Is that something you'd want them to feel about Microsoft?**

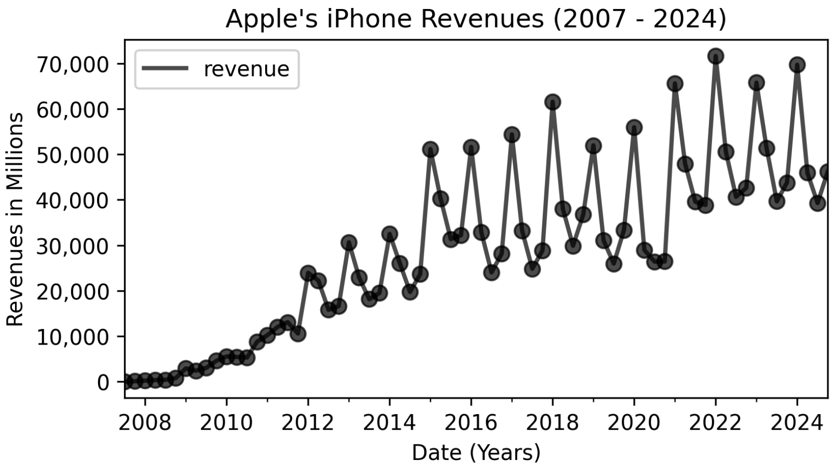
> **A**: It's sort of a funny question. **Would I trade 96% of the market for 4%?** *(Laughter.)*
>
> **I want to have products that appeal to everybody.** There's no chance that the iPhone is going to get any significant market share. **No chance.** It's a $500 subsidized item. They may make a lot of money.
>
> But if you actually take a look at the 1.3 billion phones that get sold, **I'd prefer to have our software in 60% or 70% or 80% of them**, than I would **2% or 3%, which is what Apple might get**.

Steve Ballmer, Former CEO of Microsoft in his infamous interview with USA Today 2007

Microsoft's Missed Opportunity: Back of a Napkin Calc.

- Since Q1 2009, **Windows' Mobile OS's market share <= 2.5%**, and is now at 0.02% (StatCounter).
- **Apple's Mobile iOS market share > 19%**, and is now at **27.69%** (StatCounter).
- Apple's **iPhone revenues** from 2007 to 2024 was **$2.037 trillion** (per statista).
- Assuming Microsoft could have captured just **5%** of Apple's market revenue → **$102 billion**.
- This estimate **excludes app store and brand value**, which will make the missed opportunity even larger.



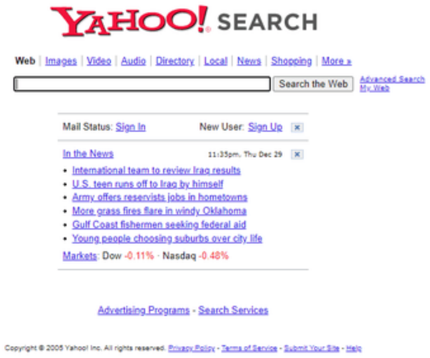


Other Real-World Forecasting Failures in Tech

IBM: Missed the PC revolution.

Blockbuster: Missed the streaming revolution.

Kodak: Missed the digital camera revolution.



Yahoo: Missed the search engine revolution.

Non-Tech Failures: Red Lobster's Endless Shrimp

Red Lobster’s endless shrimp deal was too popular, company says

By Ramishah Maruf, CNN

2 minute read · Published 5:10 PM EST, Wed November 29, 2023

More customers took advantage of Red Lobster's “Ultimate Endless Shrimp” than expected — the key reason for the chain’s roughly $11 million loss in the third quarter of 2023. Michael Nagle/Bloomberg/Getty Images

Non-Tech Failures: Target's Overestimation

Target's Stock Plunges 21% on Weak Sales Ahead of Holiday Season

The retailer's downbeat earnings report, which included lower profit and larger inventory, fell far short of Wall Street's expectations.

Listen to this article · 3:58 min Learn more

Share full article

Target's share price

Note: As of 4 p.m. Eastern on Nov. 20 · Source: FactSet · By The New York Times

Why Do These Stories Matter?

(1) Forecasting Errors = Real Money Lost

- **Microsoft**: $102 billion in potential mobile phone revenues.
- **Target: Overestimation** leads to unsold stock or overservicing (e.g., Target's Stock Plunging by 21% due to lower profit and larger inventories).
- **Red Lobster**: $11M in losses due to **underestimating** demand for their endless shrimp deal.

(2) Course Relevance:

This class will teach you how to identify **trends**, **seasonality**, and **cycles**, and **how to apply forecasting tools and models** so you can **avoid these pitfalls** in your future roles. The goal is to allow you to make **data-driven forecasts**, not just gut-based decisions, and more importantly, be able to **quantify the uncertainty** in your forecasts.

Can you Avoid Common Forecasting Mistakes?

08:00

Description | Your Logic | Your Sol | Class Results | Fadel's Logic | Fadel's Sol

- **Scenario**: You are a business analyst at a hotel chain.
- **Problem**: You are tasked with forecasting hotel occupancy for the next 11 months.
- **Data**: You have access to the hotel's monthly room occupancy data from 2022 to January of 2025. Download the file here.
- **Task: Without the use of any AI tools**, create a forecast for the hotel's occupancy for the next 11 months. The forecast can be made in Excel, R, or Python. **Document the process and the rationale behind your forecasts.**
- **Non-graded Class Activity:** Input your logic for the 11-month forecast, and your quantitative forecast for **February 2025** in the next 2 tabs, respectively.

Types of Data Over Time and the Components of a Time Series

Cross-Sectional Data

Cross-sectional data captures multiple variables at a single point in time for each observation; e.g., **all the variables within a given observation** in the DoL's LCA Disclosure Data for 2024 Q4 were collected simultaneously.

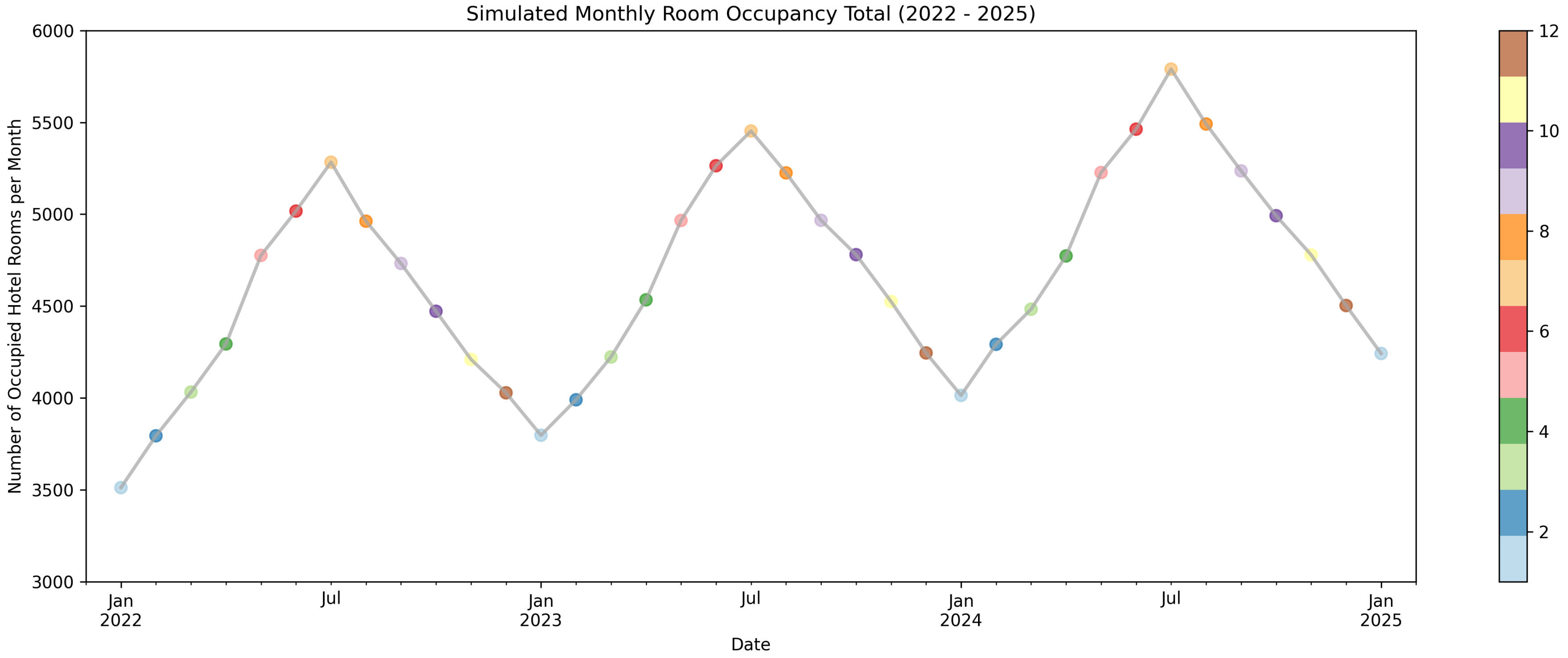
Show 4 entries Search:

Date	Job Title	Employer	WORKSITE_STATE	Salary
2022-04-14	Sr. Programmer/Analyst	Illumina, Inc.	OH	$142,294
2022-07-19	Senior SAP Application Consultant, FICO	NTT DATA Business Solutions Inc.	OH	$173,400
2022-08-10	Metallurgical Engineer	General Aluminum Mfg. Company	OH	$74,000
2023-03-13	Lecturer	The Ohio State University	OH	$50,944

Showing 1 to 4 of 184 entries Previous 1 2 3 4 5 … 46 Next

Time Series Data

Time series data captures a single variable at multiple points in time; e.g., the daily stock prices for Apple or our **simulated monthly hotel room occupancy** dataset.



Panel Data

Panel data captures multiple variables at multiple points in time for each observation; e.g., the Panel Study of Income Dynamics or the World Bank's World Development Indicators.

	iso3c	date	NY.GDP.MKTP.KD.ZG	SH.DYN.NMRT	SH.HIV.INCD.ZS	SH.MED.BEDS.ZS	SH.MED.I
1	CHN	2014	7.4	5.8		3.6	
2	CHN	2015	7	5.3		3.8	
3	CHN	2016	6.8	4.8		4.1	
4	CHN	2017	6.9	4.4		4.3	
5	CHN	2018	6.7	4.1		4.6	
6	CHN	2019	6	3.7		4.8	
7	CHN	2020	2.2	3.5		5	

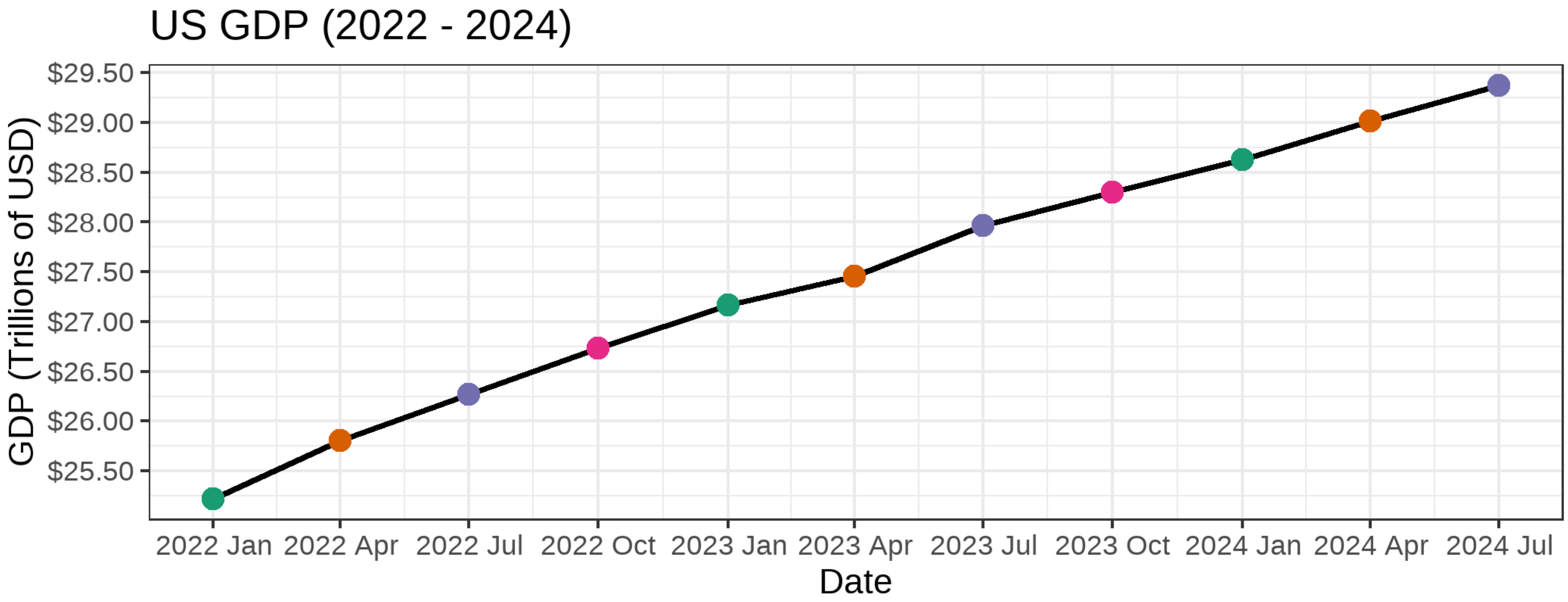
Showing 1 to 7 of 30 entries

Previous 1 2 3 4 5 Next

Source: Data queried from the World Bank Data using the wbstats in R. The printed results show a snapshot of 7 variables (out of a much larger panel dataset). You can think of panel data as a cross-sectional dataset with a longitudinal/time component.

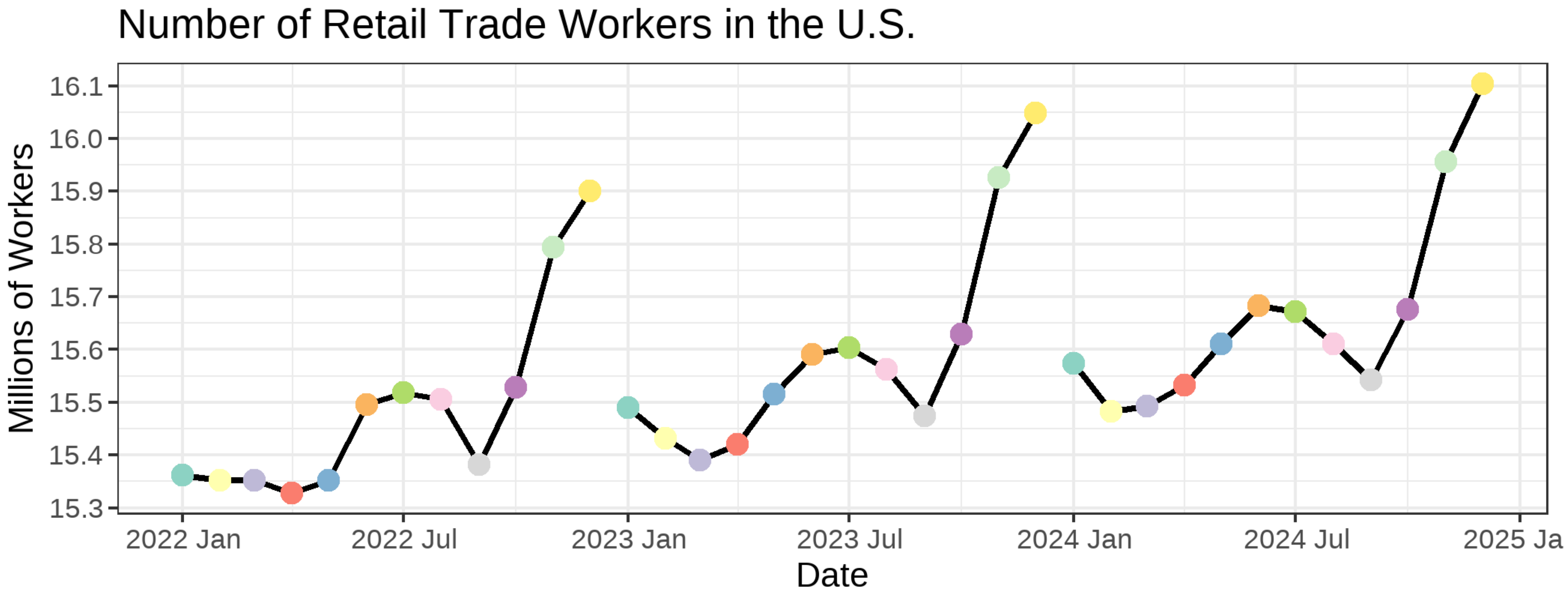
Components of Time Series Data: Trend

- **Trend**: A long-term increasing or decreasing pattern over time.
- **Example**: The US GDP has a long-term upward trend.



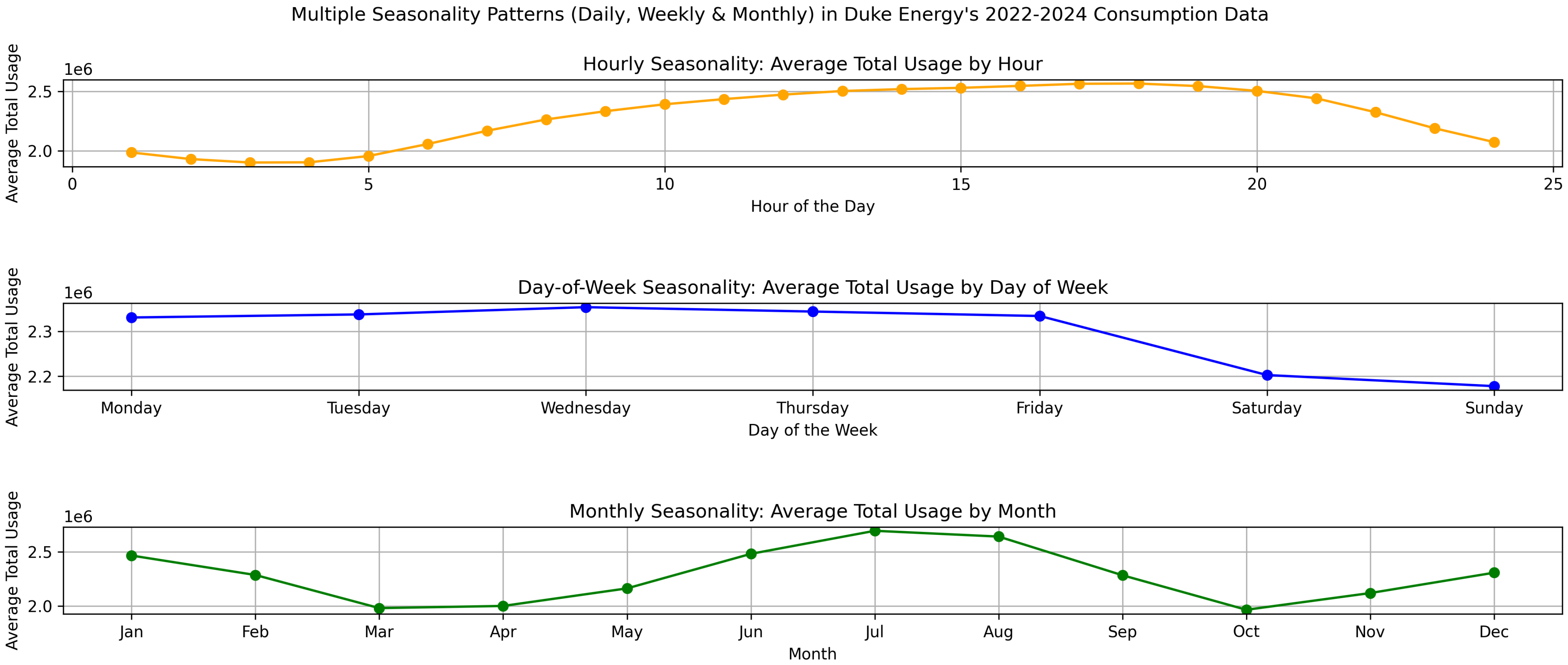
Components of Time Series Data: Seasonality

Seasonality refers to the property of a time series that displays REGULAR patterns that repeat at a constant frequency (*m*). For example, the number of retail trade workers has a seasonal pattern (with an upward trend).



Components of Time Series Data: Multiple Seasonality

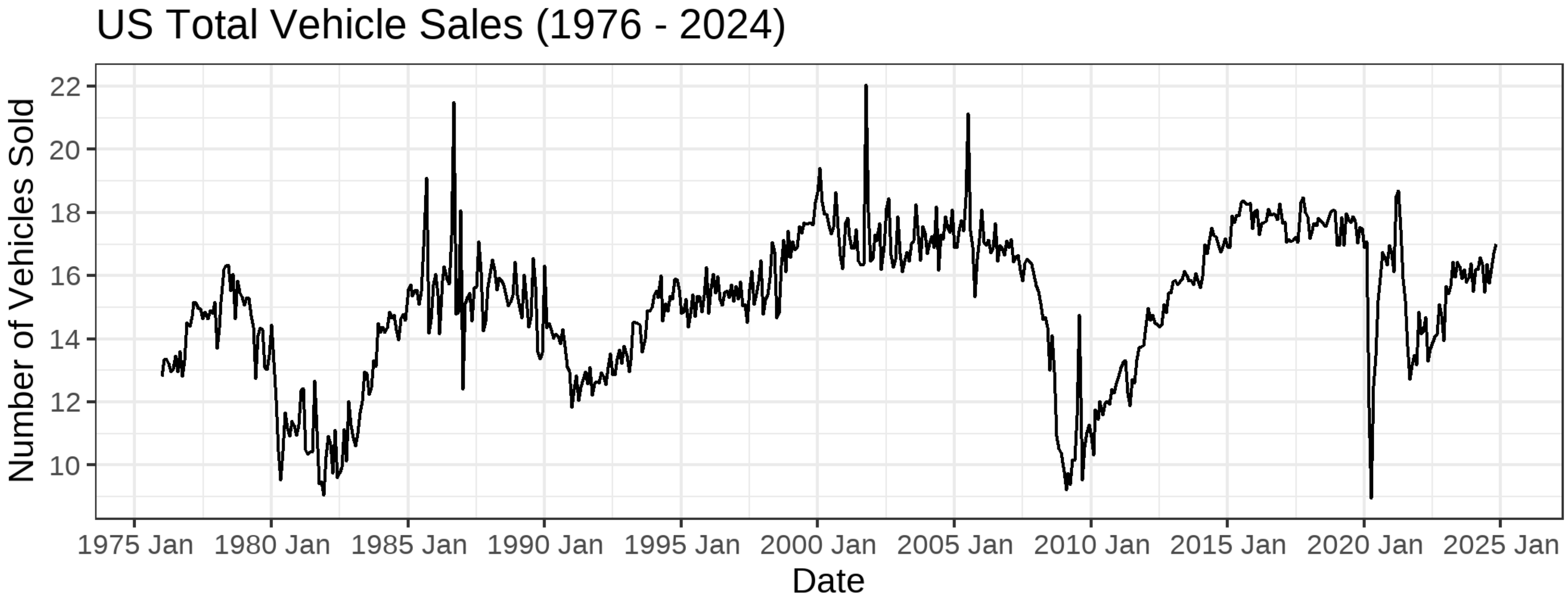
Multiple seasonality refers to the property of a time series that displays multiple seasonal patterns that repeat at different frequencies.



Data Source: Duke Energy, https://www.duke-energyohiocbp.com/Documents/LoadandOtherData.aspx, downloaded and charted on Jan 27, 2025

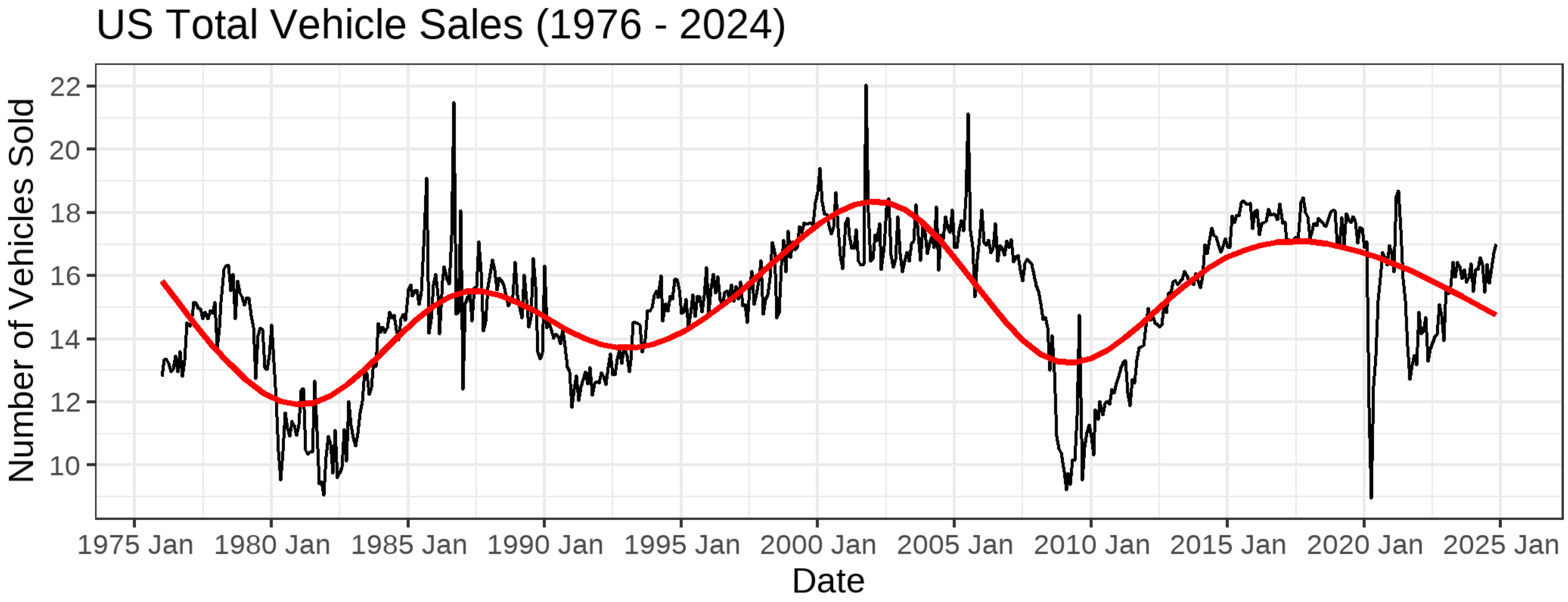
Components of Time Series Data: Cycles

Cycles refer to the property of a time series that displays irregular patterns that repeat at irregular frequencies. For example, the US Total Vehicle Sales has business cycles that are influenced by economic conditions and advancements in vehicle technologies.



Components of Time Series Data: Cycles

Cycles refer to the property of a time series that displays irregular patterns that repeat at irregular frequencies. For example, the US Total Vehicle Sales has business cycles that are influenced by economic conditions and advancements in vehicle technologies.



Kahoot Competition #01

To assess your understanding and retention of the topics covered so far, you will **compete in a Kahoot competition (consisting of 5 questions)**:

- Go to https://kahoot.it/
- Enter the game pin, which will be shown during class
- Provide your first (preferred) and last name
- Answer each question within the allocated 20-second window (**fast and correct answers provide more points**)

Winning the competition involves having as many correct answers as possible AND taking the shortest duration to answer these questions. The winner 🏆 of the competition will receive a **0.15 bonus on Assignment 01**. Good luck!!!

P.S: The Kahoot competition will have **no impact on your grade**. It is a **fun** way of assessing your knowledge, motivating you to ask questions about topics covered that you do not have a full understanding of it, and providing me with some data that I can use to pace today's class.

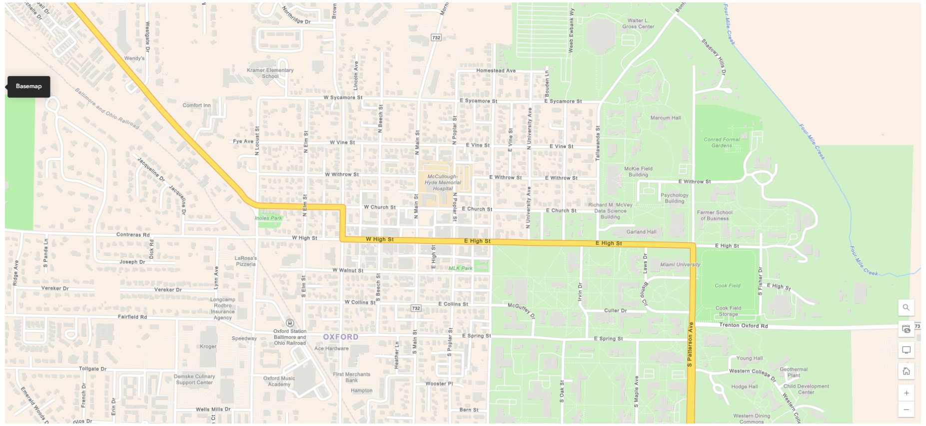
The Data Generating Process

The Idea of a Data Generating Process (DGP)

- A **time series** is defined as a **sequence of observations** recorded at regular time intervals.
- Any time series is generated by some kind of mechanism, which is often referred to as a **data generating process (GDP)**. For example, the hotel occupancy dataset is impacted by:
  - **season**, **holidays**, **economic conditions**, and **marketing campaigns**;
  - **number of rooms**, **room rates**, and **customer satisfaction**;
  - **weather**, **local events**, and **competition**; and
  - **number of rooms already booked**, **room cancellations**, and **no-shows**.
- The **DGP** is the **underlying theoretical mechanism** that generates the data we observe.
  - Accounts for both systematic patterns (e.g., trend, seasonality) and randomness.
- **But**: In real-world settings, there is often **no perfectly known** DGP.
  - Any formula or model we write is an approximation of the **unknowable "truth."**

Model vs. Reality – The Map Analogy for DGPs

- **A map ≠ the territory**:
  - We use maps to navigate, but they are always simplified.
  - Similarly, a forecast model ≠ reality—it is a *purposeful* simplification.
- **Different maps for different needs**:
  - A tourist map highlights landmarks, while a transportation map focuses on roads to inform navigation.
  - Each addresses *specific* questions, just as we build different forecasting models for different objectives.



Why Use DGPs If They Do not Actually Exist?

(1) Guiding Principle:

- Thinking in terms of a *hypothetical* DGP helps us design or select reasonable model structures.
- E.g., we incorporate domain insights: “Does our hotel occupancy data show strong seasonality?”

(2) Clarifying Assumptions:

- Even if the DGP is not known, stating assumptions (e.g., no trend, constant variance) makes our models testable and improvable.

(3) Iterative Refinement:

- As new data contradict our assumptions, we adjust our “map” of reality.
- In forecasting, we continually update models to capture changing conditions.

Key DGP Takeaways

- **All Models Are Wrong…**
  - ...but some are *useful* for forecasting, planning, or decision-making.
- **The DGP Is a Useful Fiction**
  - We talk about it to structure our thinking.
  - We never truly “see” it; we only see **data**.
- **Practical Implication**
  - A good model is *close enough* to guide accurate forecasts.
  - Remain aware of model limitations and be ready to adapt.

"What Can (and Can not) We Forecast?"

Rank these Scenarios in Terms of Forecastability

02:00

Description | Scenarios | Class Results

- **Rank** each scenario (in the next tab) from **easiest** (1) to **hardest** (6) to predict.
- **Submit** your ranking by clicking here

Perfect (or Near-Perfect) Forecasts

- **Examples**
  - **Sunset Times**:
    - Based on precise astronomical calculations.
    - We can predict sunset to the exact minute, *tomorrow* or even a year from now.
  - **Tides**:
    - Governed by well-modeled gravitational forces of the Moon and Sun.
    - Highly predictable for centuries into the future.
- **Why So Certain?**
  - These phenomena follow *deterministic* (or near-deterministic) physical laws.
  - Little to no stochastic “noise” in the process.

Partially Predictable – Weather & Markets

Weather

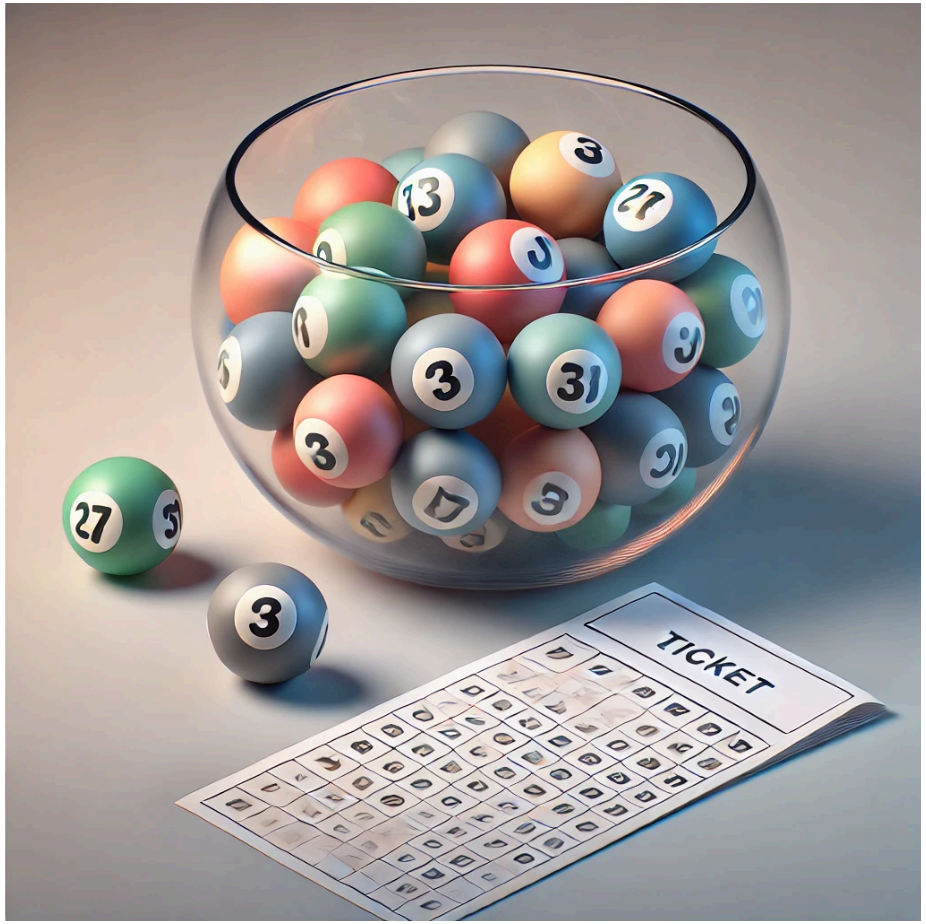
- *Tomorrow's Forecast*: Quite accurate (initial conditions + physical models).
- *1 Year Ahead*: Chaos and changing conditions degrade accuracy significantly.

S&P 500

- *Tomorrow's Close*: Some short-term signals exist, but accuracy is limited (especially if you are attempting to beat the market; accuracy is relatively high if you just want to be in the ball park of the `adjusted close`).
- *1 Year Ahead*: Many unknown macro shocks, making precise forecasts very uncertain.

Unpredictable – Lottery Numbers

- **No Predictable Pattern**
  - Draws are *engineered* to be random.
  - No matter how much data you collect, you can't *outpredict* chance.
- **Why "Un-forecastable"?**
  - The Data-Generating Process (DGP) is effectively *pure noise* by design.
  - No structural or deterministic component to model.



Relating It Back to the DGP

- **Different Types of DGPs**
  - **Deterministic (or nearly so)**: Sunset times, tidal schedules.
  - **Complex & Partly Stochastic**: Weather, financial markets.
  - **Pure Randomness**: Lottery draws.
- **Key Lesson**
  - *All* these processes have a DGP—some are more “knowable” than others.
  - *Forecastability* depends on how much of that DGP is deterministic vs. random and how well we can model it.

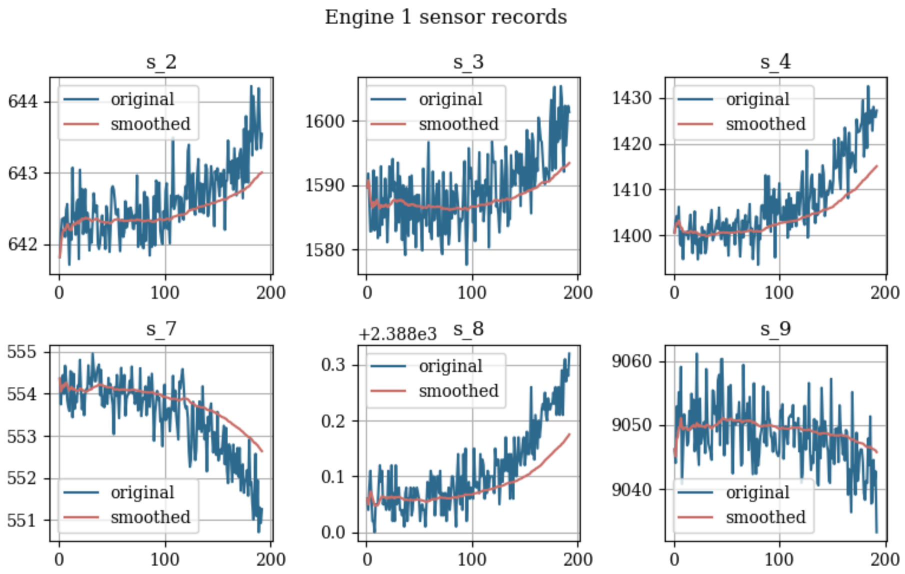
Key Forecasting Terms

Forecasting

- **Forecasting** is the process of using *historical data* and *patterns* to predict *future* values or events.
- The objective of most **time series analyses** is to provide forecasts of future values of the time series.

Explanatory Forecasting

- In addition to past data, **explanatory forecasting** uses *additional variables* to predict future values of the variable(s) of interest. To make the forecasts, we need to include both historical and future predictions for each of the explanatory variables. For example,
  - **Forecasting electricity demand** using weather forecasts, time of day, day of week, etc.



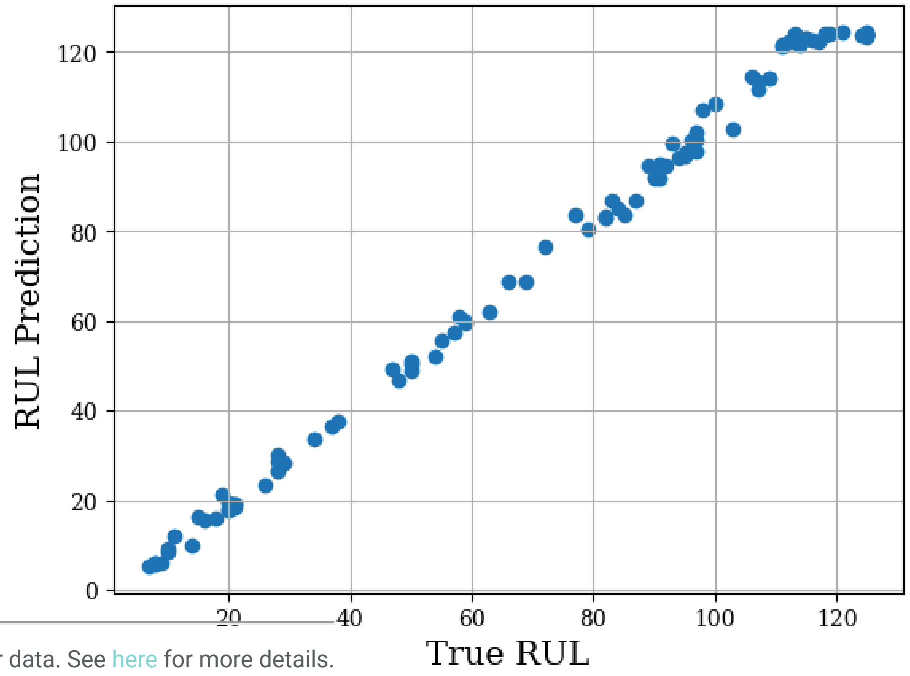


Image Sources: Nixtla's demo for forecasting the remaining useful life of an engine using exogenous sensor data. See here for more details.

Backtesting

- Backtesting is the practice of **evaluating** a forecasting model by applying it to **historical data** and comparing the predictions with the actual outcomes.
- A way to see how the model **would have performed** in the past.
- Backtesting is the time series equivalent of a **train-test split** in ML; **not random though**.

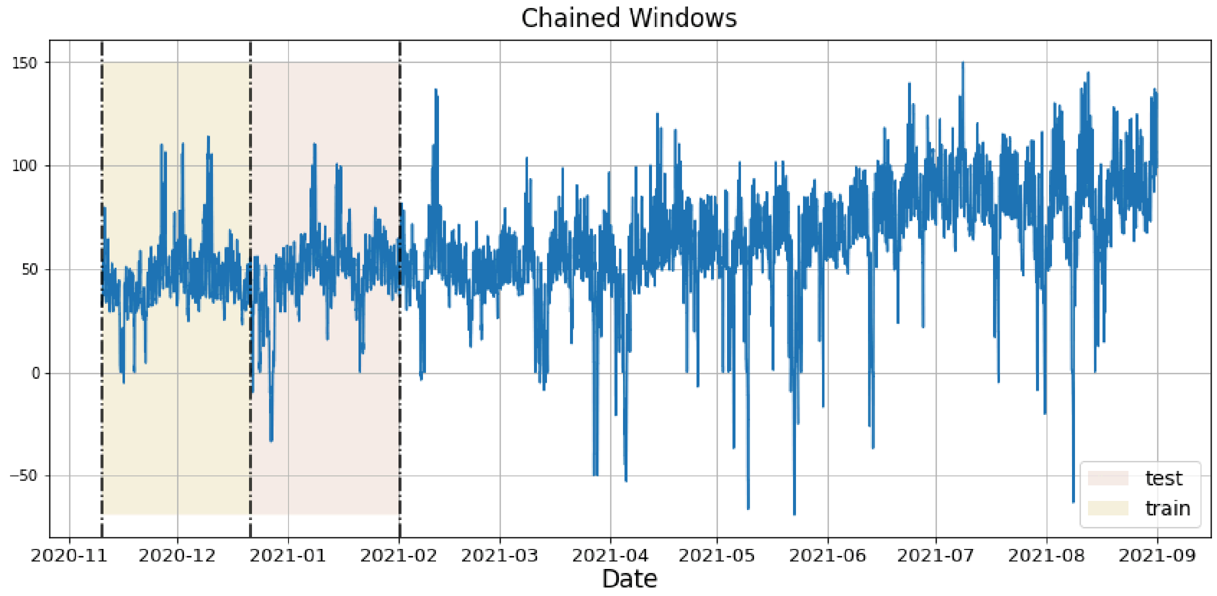


Image Source: Chained Windows from Nixtla's GitHub Repo for *statsforecast*.

Insample vs. Out-of-Sample Metrics

- **In-Sample Metrics**:
  - Metrics calculated on the **same data** used to train the model.
  - Can be misleading, as the model has already seen this data.
- **Out-of-Sample Metrics**:
  - Metrics calculated on **data not seen by the model** during training (i.e., test/holdout/out-of-sample data).
  - More reliable indicators of how the model will perform on new data.

Recap

Summary of Main Points

By now, you should be able to do the following:

- Describe the role of forecasting in business.
- Describe the key components of a **time-series** (**trend**, **seasonality**, **multiple seasonality**, and **cycles**).
- Explain the concept of data-generating process (DGP)
- Discuss limits of forecasting
- Understand key forecasting terminology

Review and Clarification

- **Class Notes**: Take some time to revisit your class notes for key insights and concepts.
- **Zoom Recording**: The recording of today's class will be made available on Canvas approximately 3-4 hours after the session ends.
- **Questions**: Please don't hesitate to ask for clarification on any topics discussed in class. It's crucial not to let questions accumulate.

📖 Required Readings 📖

🐍 Python Prep

- Getting Started with Conda
- Data Structures

🤖 LLM: Prep

- A Very Gentle Introduction to Large Language Models without the Hype

🎯 Assignment 🎯

- Go over your notes and complete Assignment 01 on Canvas.