ISA 444: Business Forecasting

01: Introduction to Time Series Analysis and Forecasting

Fadel M. Megahed, PhD

Endres Associate Professor
Farmer School of Business
Miami University

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

Fall 2023

Learning Objectives for Today's Class

- Describe **course motivation** and **structure**.
- Explain the differences between **cross sectional**, **time series** and **panel** datasets.
- Describe the **components of time series** datasets.
- Explain the **forecasting steps**.

Course Motivation and Structure

The Analytics Journey: Pre-analytics

Pre-Analytics/Data Management: where one attempts to **extract** the needed *data* for analysis.

The Analytics Journey: Pre-analytics

Pre-Analytics/Data Management: where one attempts to **extract** the needed *data* for analysis.

Show 6 entries Search:

	DATE	RSCCASN
1	Jan 1992	6,938
2	Feb 1992	7,524
3	Mar 1992	8,475
4	Apr 1992	9,401
5	May 1992	9,558
6	Jun 1992	9,182

Showing 1 to 6 of 372 entries

Previous 1 2 3 4 5 … 62 Next

The Analytics Journey: Descriptive

03:00

Descriptive Analytics: where one attempts to **understand** the data through **visualizations**, **descriptive statistics**, and **exploratory statistical models**.

Activity | Viz 1 | Viz 2 | Viz 3 | Viz 4 | Key Points

> Over the next 3 minutes, please identify what you have learned from the charts in each tab.

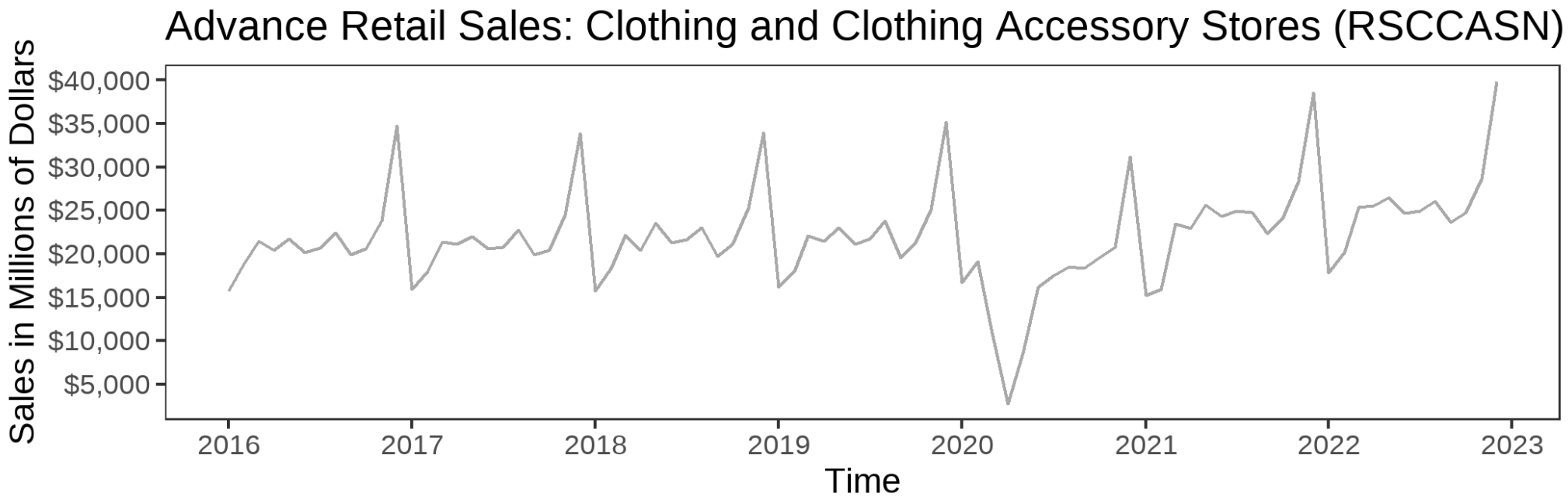
- Write down your answers in the last tab (it is editable).
- Discuss your answers with your neighboring classmates.
- Be prepared to share these answers with class.

The Analytics Journey: Descriptive

03:00

Descriptive Analytics: where one attempts to **understand** the data through **visualizations**, **descriptive statistics**, and **exploratory statistical models**.

Activity | Viz 1 | Viz 2 | Viz 3 | Viz 4 | Key Points



The Analytics Journey: Descriptive

03:00

Descriptive Analytics: where one attempts to **understand** the data through **visualizations**, **descriptive statistics**, and **exploratory statistical models**.

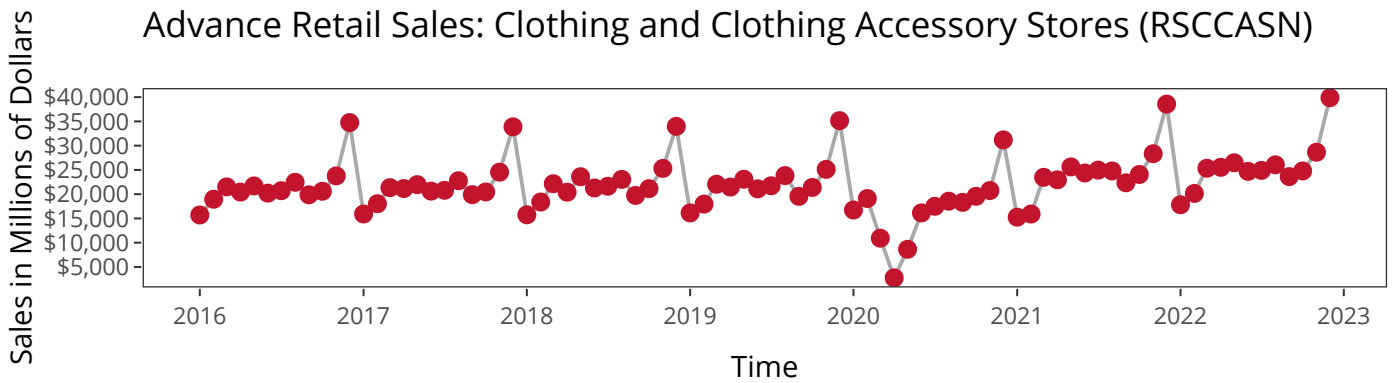
Activity | Viz 1 | Viz 2 | Viz 3 | Viz 4 | Key Points

The Analytics Journey: Descriptive

03:00

Descriptive Analytics: where one attempts to **understand** the data through **visualizations**, **descriptive statistics**, and **exploratory statistical models**.

Activity | Viz 1 | Viz 2 | Viz 3 | Viz 4 | Key Points

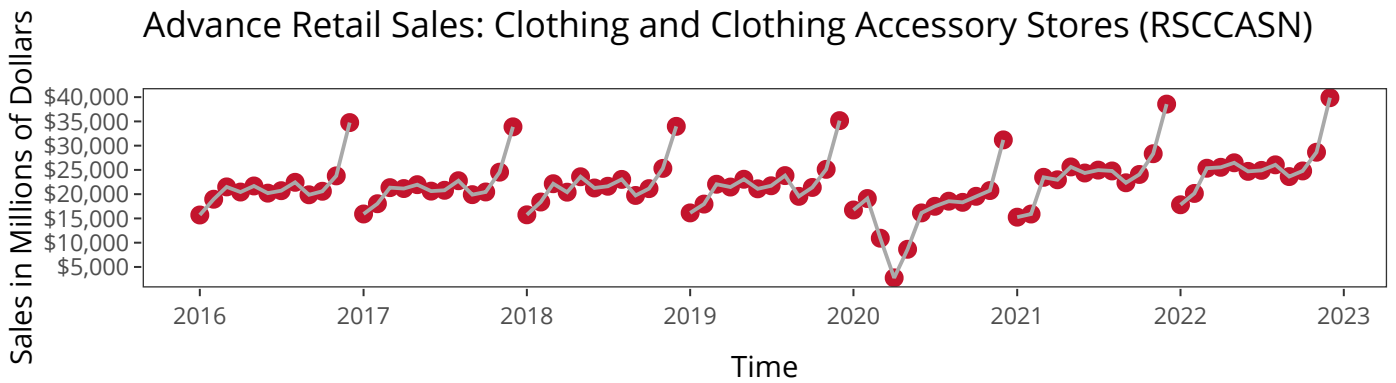


The Analytics Journey: Descriptive

03:00

Descriptive Analytics: where one attempts to **understand** the data through **visualizations**, **descriptive statistics**, and **exploratory statistical models**.

Activity | Viz 1 | Viz 2 | Viz 3 | Viz 4 | Key Points



The Analytics Journey: Descriptive

03:00

Descriptive Analytics: where one attempts to **understand** the data through **visualizations**, **descriptive statistics**, and **exploratory statistical models**.

Activity | Viz 1 | Viz 2 | Viz 3 | Viz 4 | Key Points

Main Insight(s): (Insert below)

- Edit me
- ...
- ...

The Analytics Journey: Descriptive

Descriptive Analytics: where one attempts to **understand** the data through **visualizations**, **descriptive statistics**, and **exploratory statistical models**.

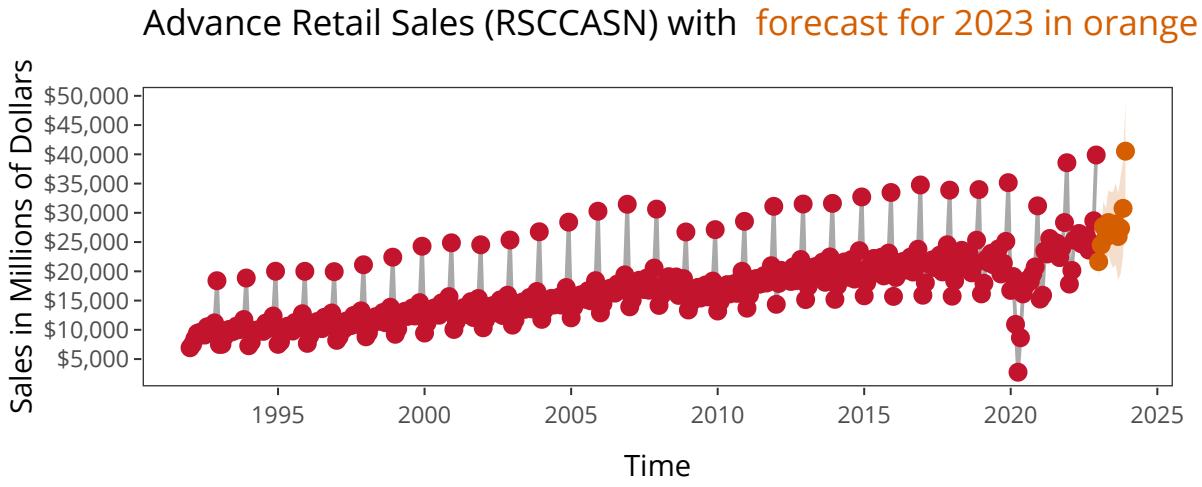
Show 6 entries Search:

Month	Seasonal Factor
Jan	-4,419.70
Feb	-2,845.96
Mar	-756.82
Apr	-1,228.00
May	-90.66
Jun	-1,100.75

Showing 1 to 6 of 12 entries Previous 1 2 Next

The Analytics Journey: Predictive

Predictive Analytics: where **statistical** and **machine learning** models are used to help us utilize independent variable[s] to predict an outcome variable of choice.

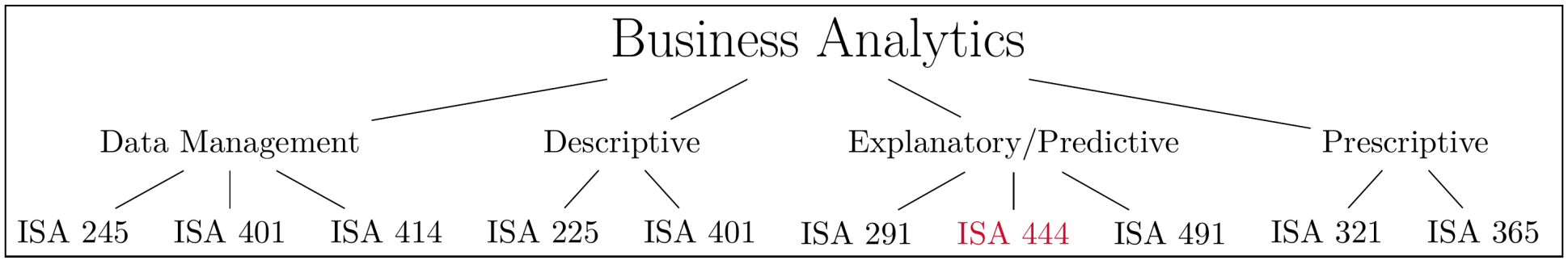


The Analytics Journey: Prescriptive

Prescriptive Analytics: where **mathematical models** are used to make recommendations for business actions.

- Our **overarching goal** behind data/business analytics, is to **make informed decisions based on what we have learned from the data**. Hence, this stage is where we build on what we learned during the *descriptive* and *predictive* stages to make more informed decisions.
- Imagine that you are L Brands or Express Inc.; how would you use the information pertaining to the **U.S. trends in clothing sales for staffing and operational decisions**?

How does our Curriculum at Miami University Prepare you for this Journey?



My take on the courses within the business analytics major/minor at Miami University

Why are time series data different?

- Time series (TS) data **cannot be** analyzed using standard regression techniques.
- In regression, you assume that your data are **randomly sampled, resulting in independent observations.**
- In TS, we want to capitalize on the **structure of the data**, i.e., observations are typically **correlated over time (autocorrelated)** to build our models.

Course Objectives

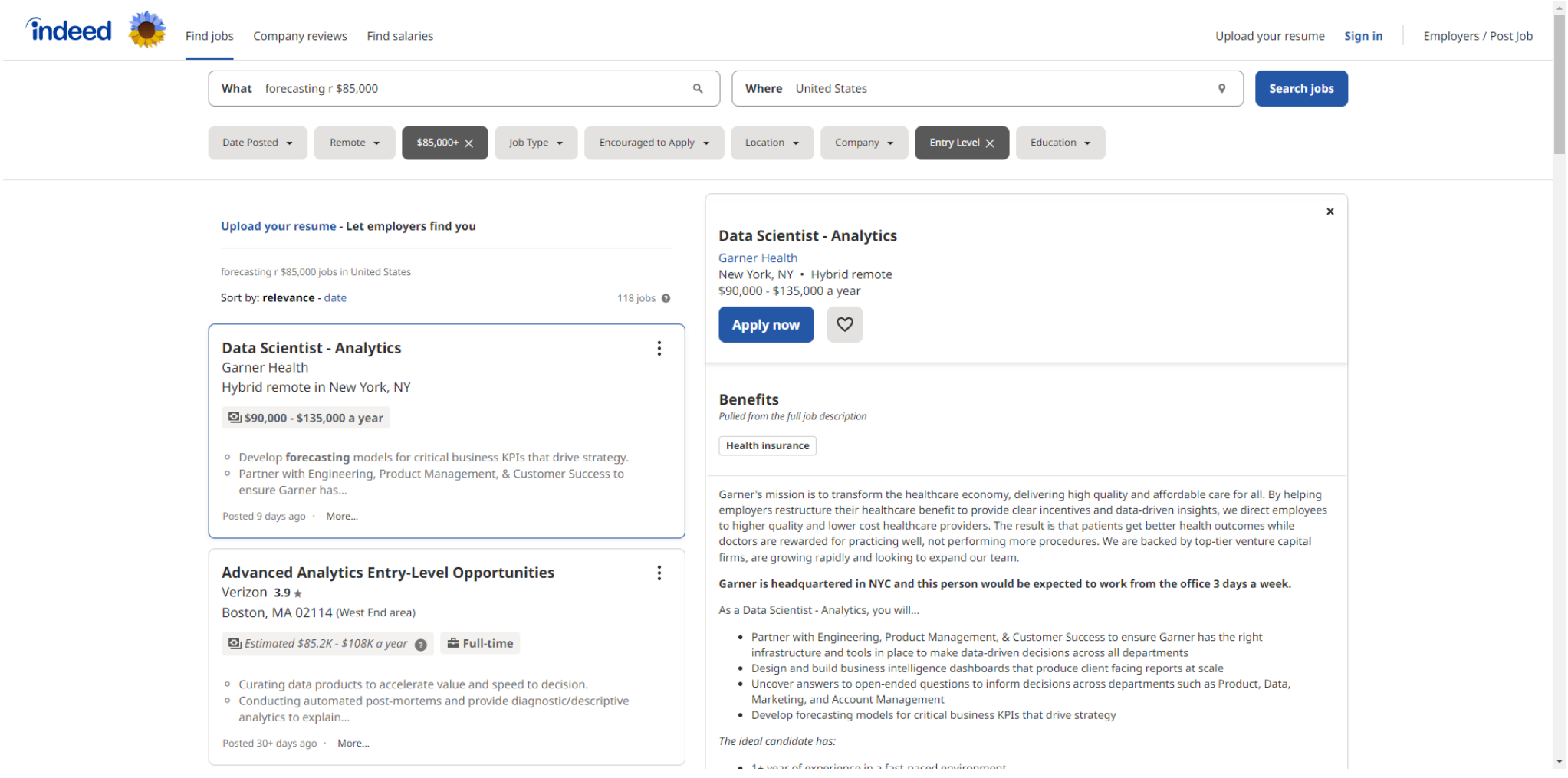
By the **end of this course**, you should be able to:

- Explain the purpose of forecasting in a business setting.
- Use the basic tools of forecasting including plots, summary measures, transformations, measures of forecast accuracy, and prediction intervals.
- Use an appropriate smoothing-based method for time series analysis and forecasting.
- Use an appropriate ARIMA model to forecast a time series.
- Use simple and multiple linear regression models to forecast a time series.

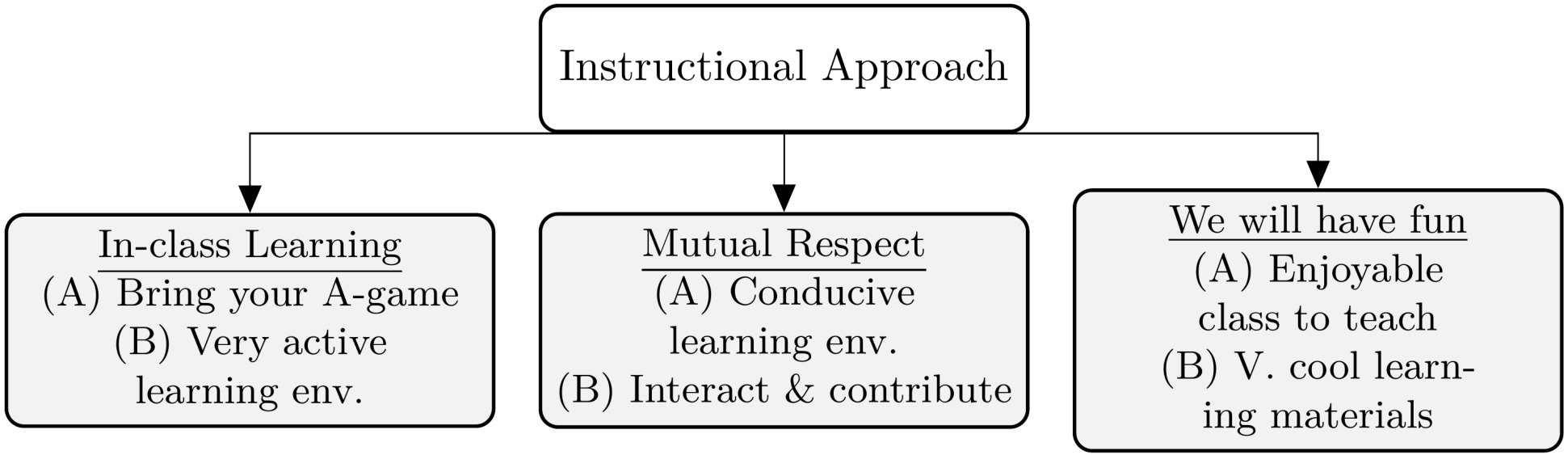
Why Should You Care?

- **Sales forecasting:** Businesses use time series analysis to forecast future sales based on historical sales data, in order to make informed decisions about inventory management, staffing, and marketing efforts.
- **Financial forecasting:** Companies use time series analysis to forecast future financial performance, such as revenue, expenses, and profits, in order to make strategic business decisions and plan for future growth.
- **Demand forecasting:** Manufacturers and retailers use time series analysis to forecast demand for their products, in order to optimize production and inventory levels.
- **Resource planning:** Businesses use time series analysis to forecast resource usage, such as energy consumption, in order to plan for future infrastructure needs and reduce costs.

Why Should You Care?

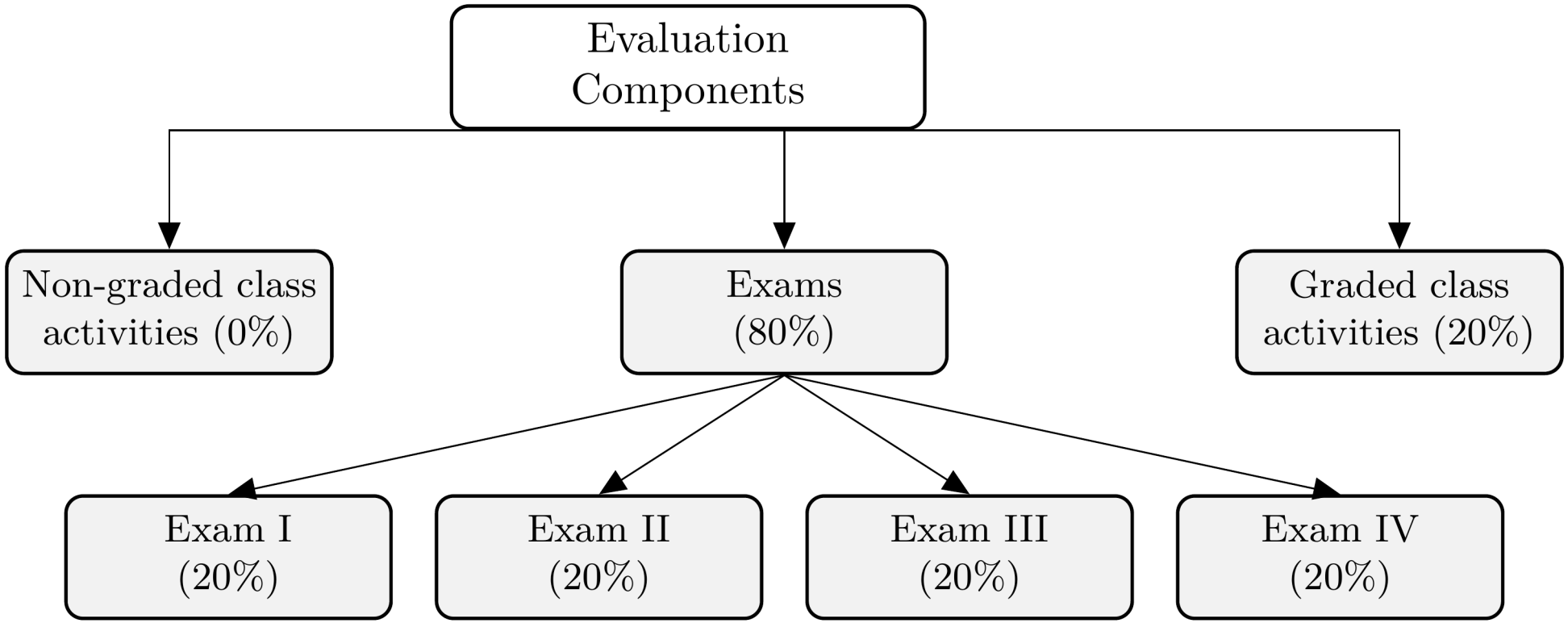


Instructional Approach



An overview of the instructional approach for ISA 444.

How will I Evaluate your Learning?



An overview of the evaluation components for ISA 444.

Types of Data Over Time

Cross Sectional Data

Cross Sectional Data: Measurements on multiple units, recorded in a single time period.

Example 1: H1B 2020-2022 Data for Senior Data Scientists at Netflix

	START DATE	JOB TITLE	BASE SALARY	LOCATION
1	2021-08-11	SENIOR DATA SCIENTIST	118,955	LOS GATOS, CA
2	2021-09-09	SENIOR DATA SCIENTIST	143,291	LOS GATOS, CA
3	2021-06-14	SENIOR DATA SCIENTIST	143,291	LOS GATOS, CA
4	2021-06-14	SENIOR DATA SCIENTIST	143,291	LOS GATOS, CA
5	2021-10-18	SENIOR DATA SCIENTIST	143,562	LOS GATOS, CA

Showing 1 to 5 of 23 entries Previous 1 2 3 4 5 Next

Data scraped from https://h1bdata.info/index.php?em=Netflix+Inc&job=Senior+Data+Scientist&city=LOS+GATOS&year=All+Years on August 28, 2023 using the rvest. The printing was limited to those individuals who started on/after January 01, 2020, with the filters specified in the URL.

Cross Sectional Data

Cross Sectional Data: Measurements on multiple units, recorded in a single time period.

Example 2: NBA 2022-2023 Leaders - Top Players in PTS/Game

	Player			Pos		Age	Tm	G	FG
1	Joel Embiid	C	28	PHI	66	11	0.548	0.573	33.1
2	Luka Dončić	PG	23	DAL	66	10.9	0.496	0.56	32.4
3	Damian Lillard	PG	32	POR	58	9.6	0.463	0.564	32.2
4	Shai Gilgeous-Alexander	PG	24	OKC	68	10.4	0.51	0.531	31.4
5	Giannis Antetokounmpo	PF	28	MIL	63	11.2	0.553	0.572	31.1

Showing 1 to 5 of 679 entries

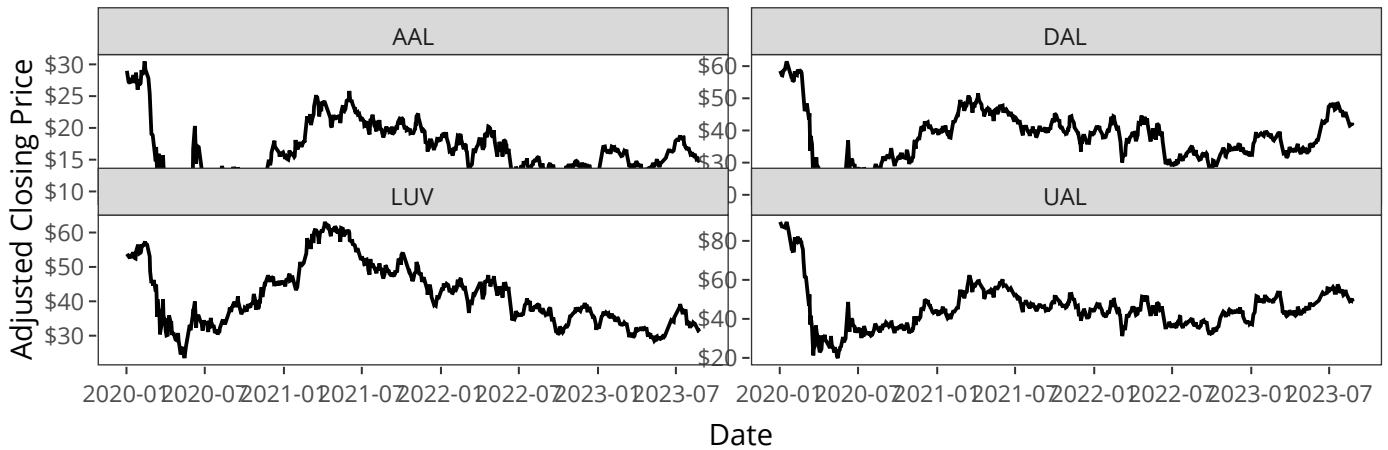
Previous 1 2 3 4 5 ... 136 Next

Source: Data scraped from https://www.basketball-reference.com/leagues/NBA_2023_per_game.html{Basketball-Reference on August 28, 2023 using the rvest 🤗. The printing was limited to the selected variables.

Time Series Data

Time Series Data: Comparable measurements recorded on a single (or a few) variables over time (usually a long period of time).

Example: Stock prices of U.S. Airlines



Panel Data

Panel Data: Cross sectional measurements (usually many variables) repeated over time (usually over a few time periods).

Example: World Bank's Data

	iso3c	date	NY.GDP.MKTP.KD.ZG	SH.DYN.NMRT	SH.HIV.INCD.ZS	SH.MED.BED
1	CHN	2020	2.2	3.4		
2	CHN	2021	8.4	3.2		
3	CHN	2022	3	3.2		
4	EGY	2020	3.6	10.4		
5	EGY	2021	3.3	10		

Showing 1 to 5 of 9 entries Previous 1 2 Next

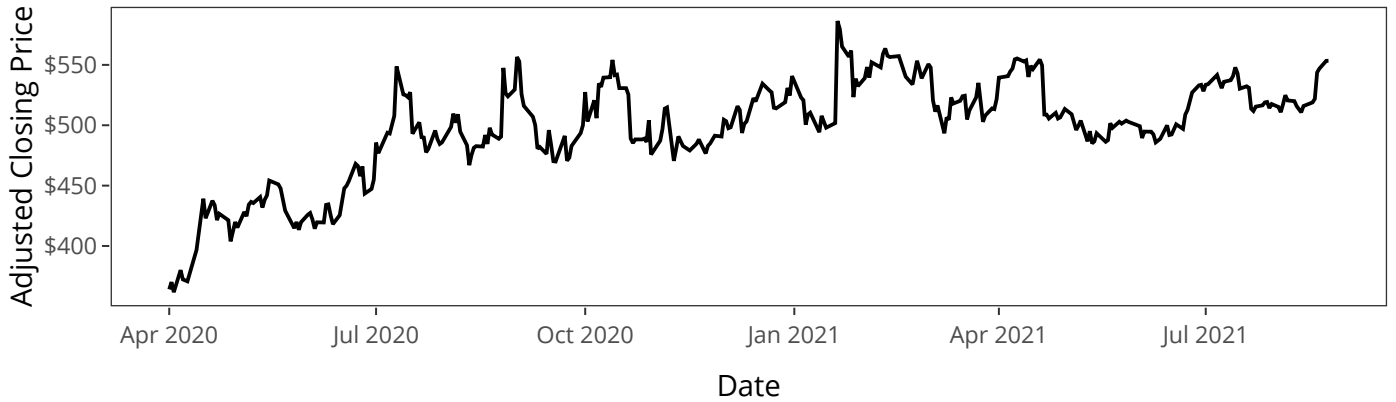
Source: Data queried from the World Bank Data using the wbstats in R. The printed results show a snapshot of 7 variables (out of a much larger panel dataset). You can think of panel data as a cross-sectional dataset with a longitudinal/time component.

Components of a Time Series

Trend

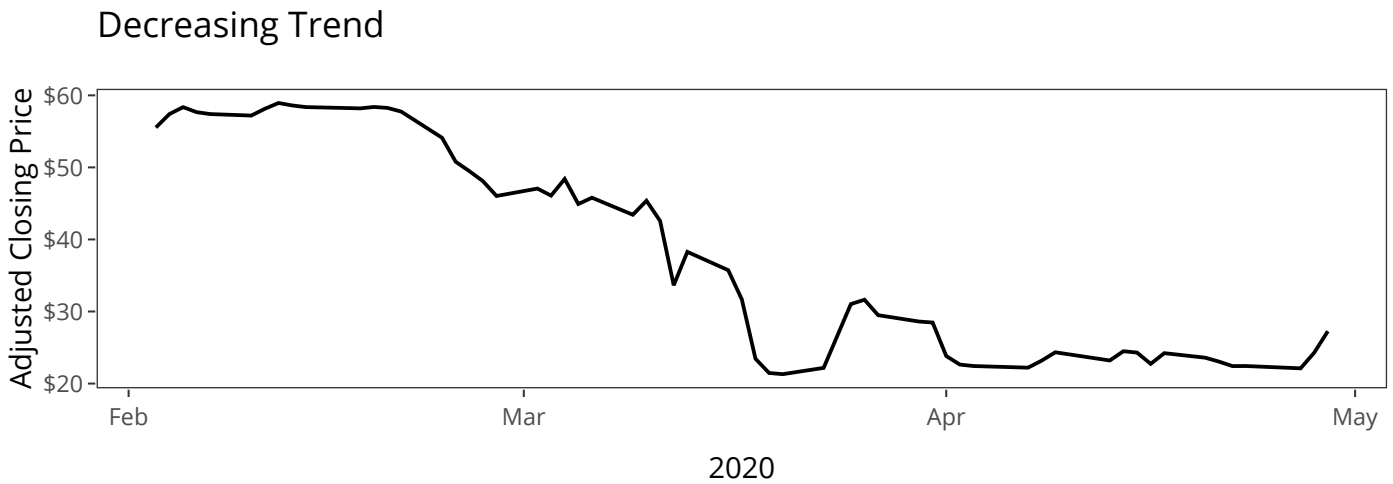
A **trend** is an increasing or decreasing pattern over time.

An increasing trend: rise of $NFLX during the COVID-19 Pandemic



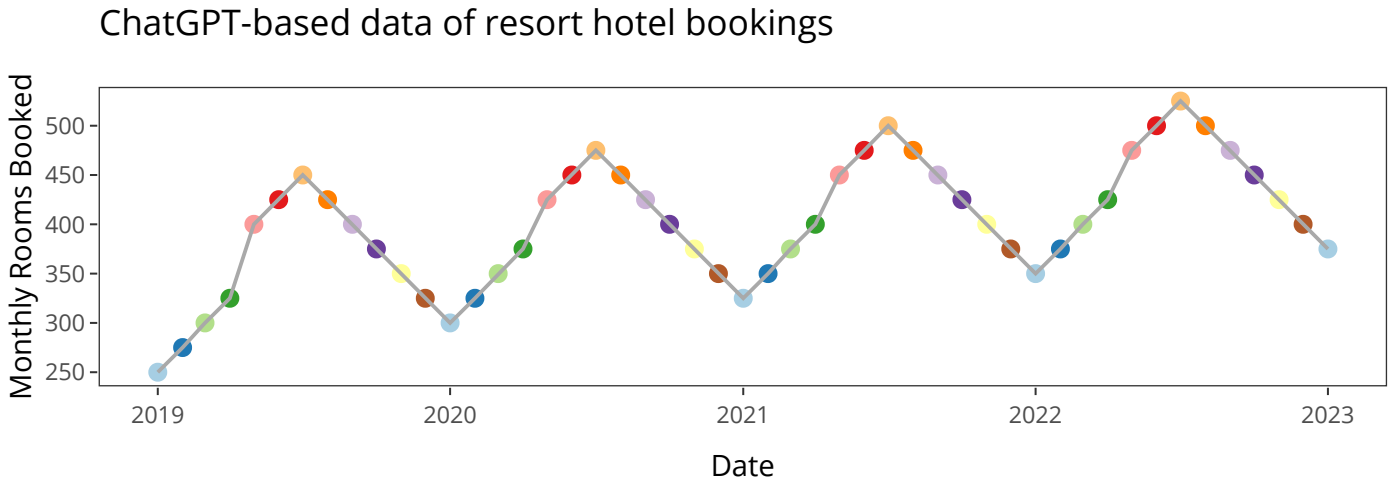
Trend

A **trend** is an increasing or decreasing pattern over time.



Seasonality

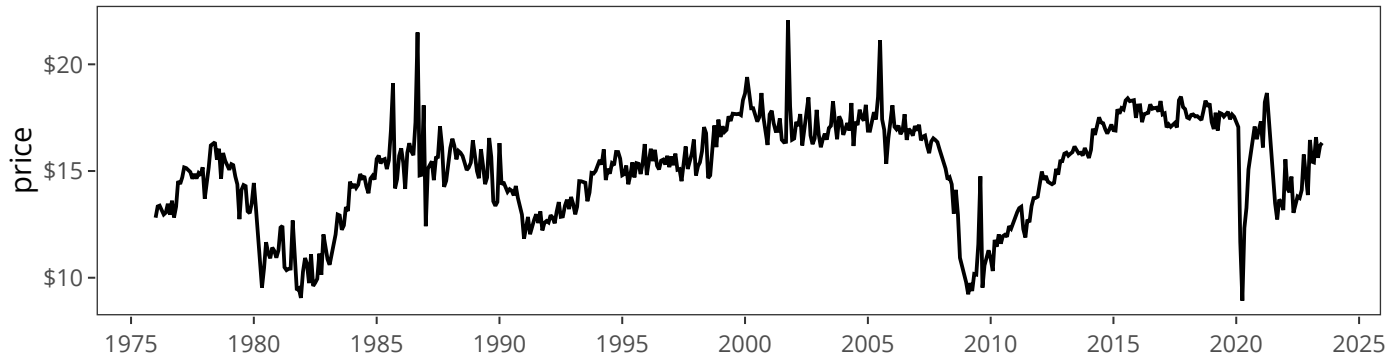
Seasonality refers to the property of a time series that displays REGULAR patterns that repeat at a constant frequency (*m*).



Cycle

Cyclical fluctuations are somewhat irregular (unknown duration).

The cyclical nature of auto sales



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)**:

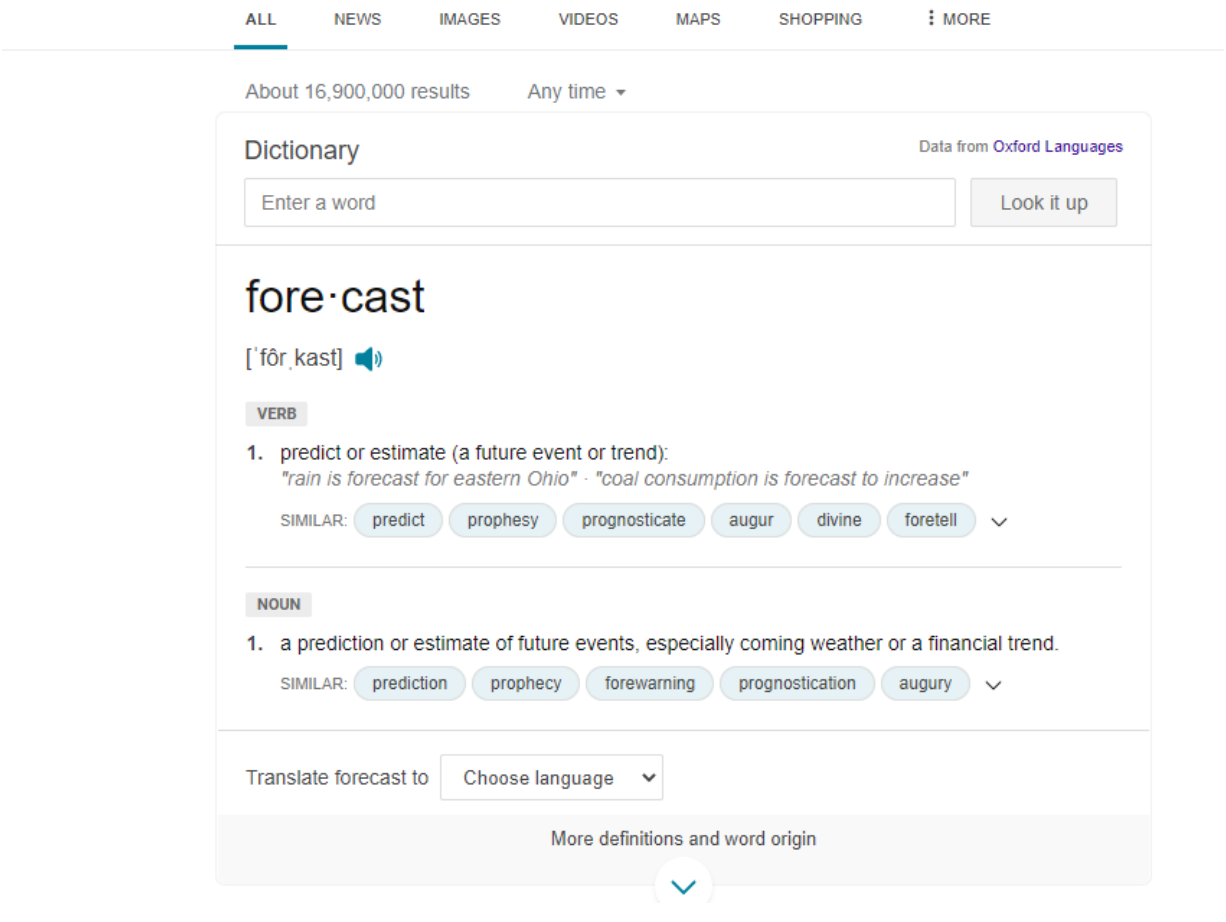
- Go to https://kahoot.it/
- Enter the game pin, which will be shown during class
- Provide your first (preferred) and last name
- Answer each question within the allocated 20-second window (**fast and correct answers provide more points**)

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 gift of their choosing ($10 Starbucks gift card or a large chocolate bar). 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.

Forecasting and Steps

So What is Forecasting?



Working Definitions

Forecasting: **The process of predicting a future event. The objective of most time series analyses is to provide forecasts of future values of the time series.**

Time Series: **A time series is a sequence of observations on a variable measured at successive points in time or over successive periods of time. In the simplest cases, observations are evenly spaced at regular intervals such as hourly, daily, weekly, monthly, or yearly, or at any other regular interval**

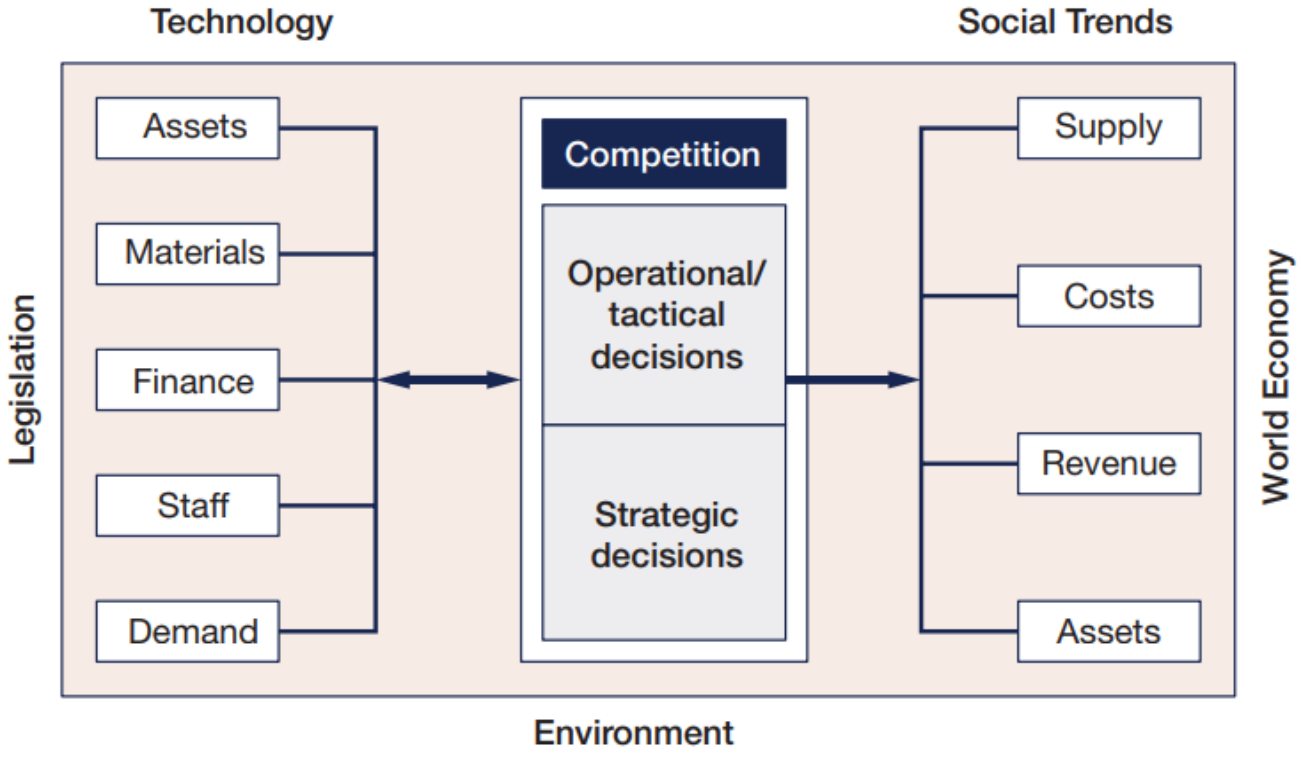
Source: Both definitions are based on Dr. Allison Jones-Farmer's lecture notes, Miami University, Spring 2020.

Why do we Forecast? (PIVASE)

- **Purpose:** What do we hope to achieve by generating the forecast? That is, what plans are dependent upon the results of the forecasting exercise? How far ahead do we wish to forecast? We refer to this period as the forecasting horizon.
- **Information:** What do we know that may help us in forecasting. And when will we know it? Detailed data is only useful if it is available in timely fashion.
- **Value:** How valuable is the forecast? What would you pay for perfect knowledge?
- **Analysis:** From analyzing the data can we develop a model that captures its characteristics? And how does it perform on new (hold-out sample) data?
- **System:** What models and software are needed to meet the needs of the organization?
- **Evaluation:** How do we know whether a particular forecasting exercise was effective and what the potential is for improvement?

From: Ord, K., Fildes, R., & Kourentzes, N. (2017). Principles of Business Forecasting (2nd ed., p. 3-6).

Why do Businessess Forecast?



An Organization's Forecasting Needs

From: Ord, K., Fildes, R., & Kourentzes, N. (2017). Principles of Business Forecasting (2nd ed., p. 7).

Forecasting Step by Step

On the basis of our preliminary discussion of **PIVASE**, we can identify **seven major steps in the forecasting process:**

- Define the forecasting and planning problem, the forecast horizon and decide the value of better forecasts.
- Determine the resources to be devoted to providing the forecasts.
- Collect relevant information, whether from a survey, from company records, or from information generated by other agencies (e.g., government figures).
- Conduct an initial analysis of the data.
- Select an appropriate forecasting method.
- Generate forecasts.
- Evaluate the forecasting exercise by checking forecasts against actual outcomes.

From: Ord, K., Fildes, R., & Kourentzes, N. (2017). Principles of Business Forecasting (2nd ed., p. 16).

Recap

Summary of Main Points

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

- Describe **course motivation** and **structure**.
- Explain the differences between **cross sectional**, **time series** and **panel** datasets.
- Describe the **components of time series** datasets.
- Explain the **forecasting steps**.

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 📖

📈 Time-Series Prep

- Read through Chapter 01 of our reference book

R Prep

- Workflow: Basics
- Names and Values
- Vectors
- Subsetting

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.