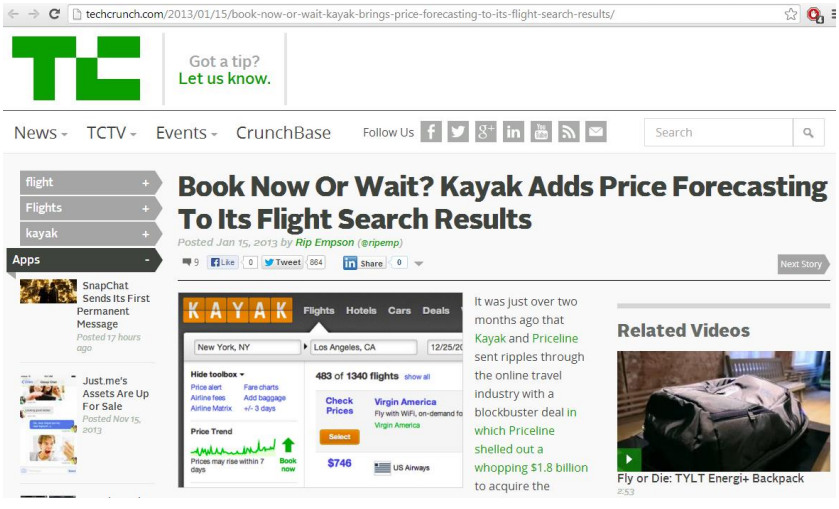
Forecasting Analytics

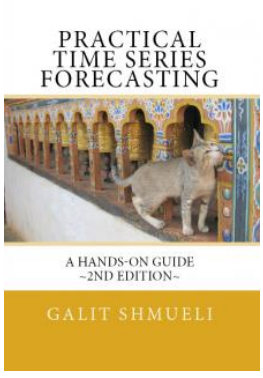
Professor Casey Lichtendahl (Classes 1-5)

Professor Galit Shmuéli (Classes 6-10)

What You Can Learn



Technical stuff: Outside of class



ForecastingBook.com

Deliverables

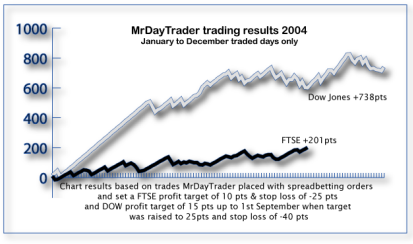
- 3 Assignments
- 1 Group Project

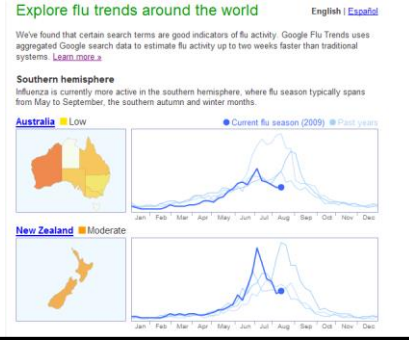
LMS Indian School of Business ISB

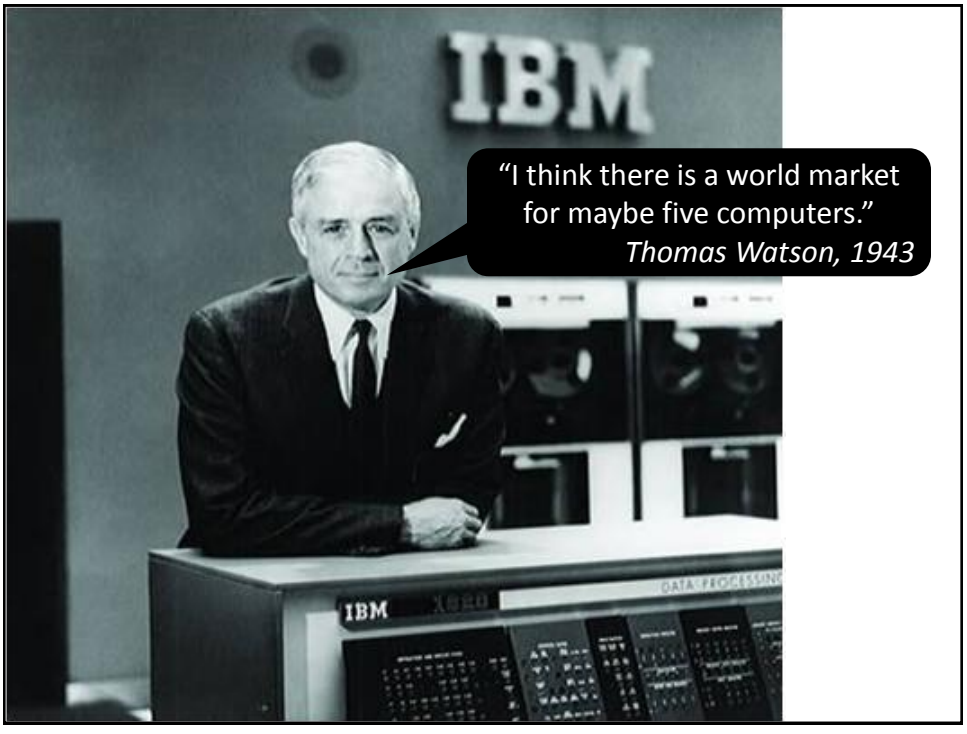
WHY FORECAST?

"Forecast"= predict the future value of a time series

Time Series Everywhere







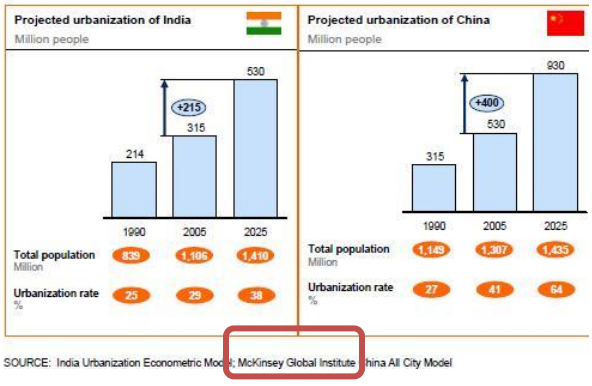
Who generates forecasts?

Governments
NGOs

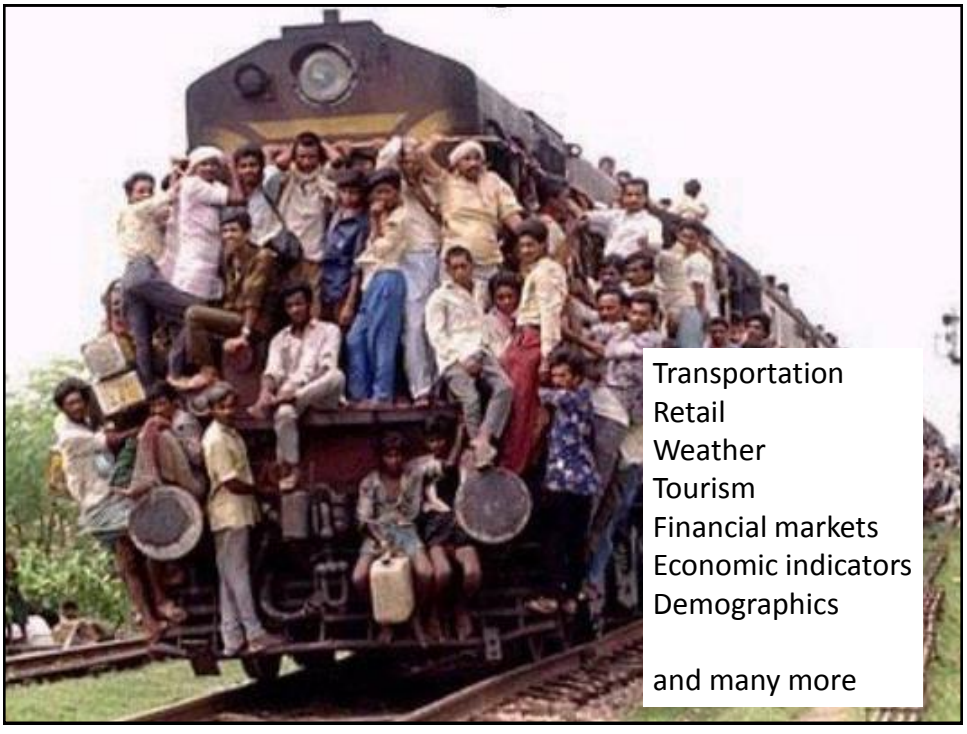
Corporates
Private sector

Consulting firms

China is more urbanized than India today and will urbanize more quickly



SOURCE: India Urbanization Econometric Mod[illegible]; McKinsey Global Institute China All City Model



Transportation
Retail
Weather
Tourism
Financial markets
Economic indicators
Demographics

and many more

Forecasting Retail Sales of $1 Billion

Zappos

Company

Zappos.com is an online shoe and clothing shop currently based in Las Vegas, Nevada. In July 2009, the company announced it would be acquired by Amazon.com in an all-stock deal worth about $1.2 billion. Wikipedia

Customer service: +1 800-927-7671

CEO: Tony Hsieh

Founder: Nick Swinmurn

Founded: 1999

"We have to plan well in order to keep our fulfillment center operating efficiently!"

– Amanda Nevins, Zappos.com

- two fulfillment centers in Shepherdsville, KY (near UPS hub)
- automated warehouse 24/7, not the most efficient way to run a warehouse, but it gets the orders out to our customers as quickly as possible
- a lot of customers order as late as midnight Eastern Standard Time and their orders show up on their doorstep 8 hours later
- 1 out of every 60 OVERNIGHT packages shipped by UPS is a Zappos box
- **we primarily use historical data, trends, and buyer/ financial experience to set financial and operating plans using excel**

The Forecast Analyst

Provides decision makers with **forecasts** of

- Sales
- Cash flow
- Inventory
- Cost
- Workload
- Demand, etc.

The **value and limitation** of the forecasts

International Institute of Forecasters

Institute of Business Forecasting & Planning

Time series vs. cross-sectional data

Example: Ridership on Amtrak Trains

US Railway company

Monthly ridership, Jan 1991 to Mar 2004 (Amtrak.xls)

Basic Notation

t=1,2,3… = time period index

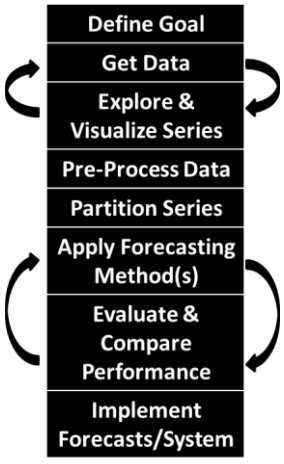
Y_t = value of the series at time period *t*

F_{t+k} = forecast for time period *t+k*, given data until time *t*

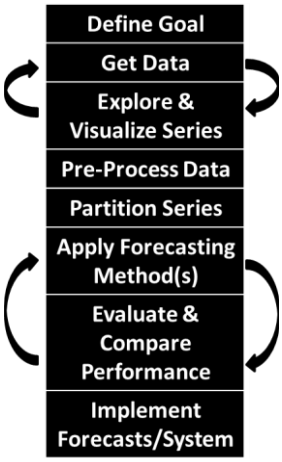
e_t = forecast error for period *t*

Month	Ridership
1/1/1991	1709
2/1/1991	1621
3/1/1991	1973
4/1/1991	1812
5/1/1991	1975
6/1/1991	1862
7/1/1991	1940
8/1/1991	2013
9/1/1991	1596
10/1/1991	1725
11/1/1991	1676
12/1/1991	1814
1/1/1992	1615
2/1/1992	1557
3/1/1992	1891
4/1/1992	1956
5/1/1992	1885
6/1/1992	1623
7/1/1992	1903
8/1/1992	1997
9/1/1992	1704
10/1/1992	1810

The Forecasting Process



GOAL DEFINITION



#1: Is the goal descriptive or predictive?

Descriptive = *time series analysis, retrospective, explanatory*

Predictive = *time series forecasting, prospective*

#2: What is the forecast horizon?

How far into the future? (k in F_{t+k})

Rolling forward or at single time point?

One-time forecasting, or ongoing task?

Implications:
How much data is needed?
Choice of forecasting methods
Expected levels of accuracy
Performance evaluation
Model deployment

#3: How will the forecast(s) be used?

Who are the stakeholders?

Numerical or event forecast?

Cost of over-prediction and under-prediction

Will forecasts undergo “adjustments”?

#4: Forecasting expertise and automation

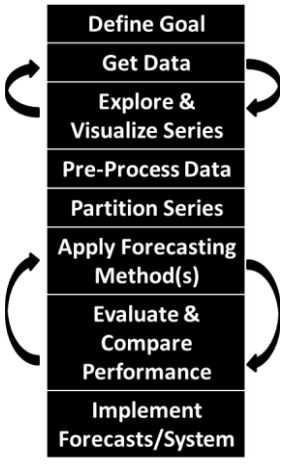
In-house forecasting or consultants?

How many series? How often?

One-time forecasting or ongoing?

Data and software availability

DATA COLLECTION



#1: Data Quality

Typically small sample, so need good quality

Data same as series to be forecasted

#2: Temporal Frequency

Should we use real-time ticket collection data?

Balance between signal & noise

Sometimes: aggregation, modeling, aggregated forecasts, then disaggregation

#3: Series Granularity

Coverage of the data

- Geographical, population, time,…

Should be aligned with goal!

Implications: zero/low counts vs. mixed populations

Particular routes

Particular populations (senior citizens, children)

Particular days (weekends)

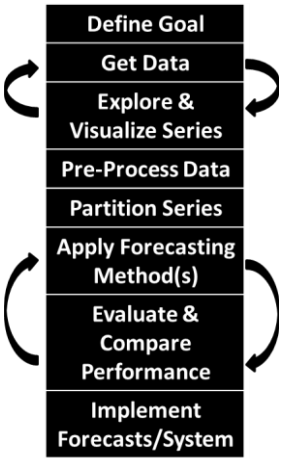
#4: Domain Expertise

Necessary information source!

Affects modeling process from start to end

Level of communication/coordination between forecasters and domain experts

EXPLORE & VISUALIZE SERIES



Time Series Components

Systematic part

- Level
- Trend
- Seasonal patterns

Non-systematic part

- "Noise"

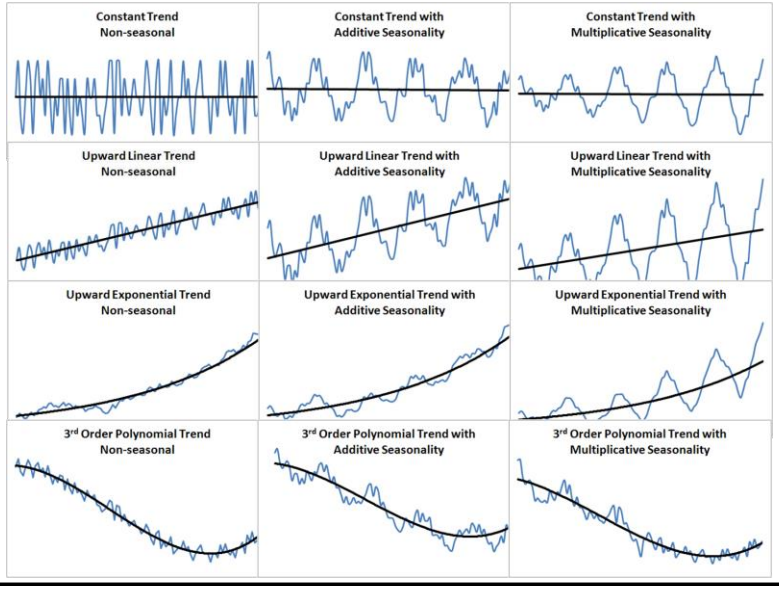
Additive:

Y_t = Level + Trend + Seasonality + Noise

Multiplicative:

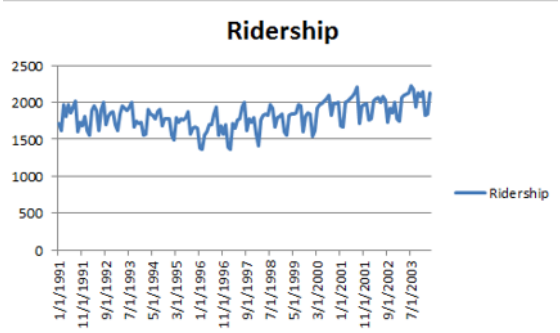
Y_t = Level x Trend x Seasonality x Noise

Time series components



The Amtrak ridership series contains which components? (choose 1 or more)

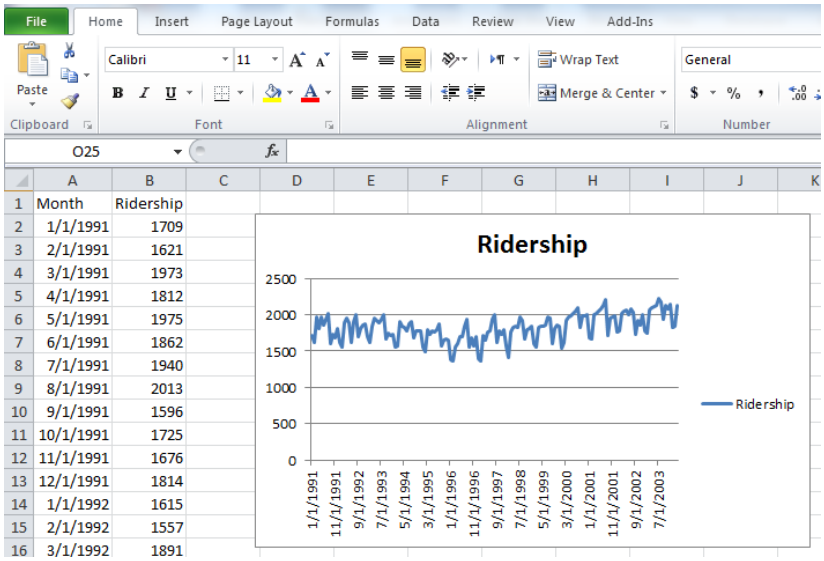
1. Level
2. Trend
3. Seasonality
4. Noise



Visualizing Time Series

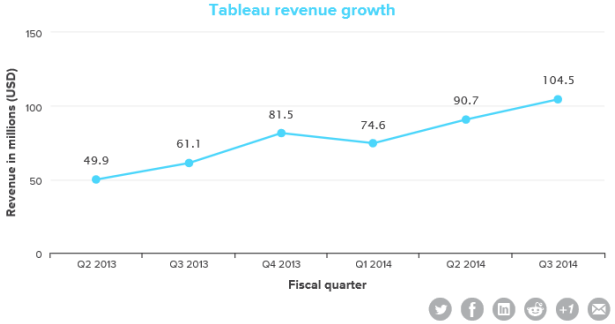
- Why visualize? Why interactively?
- Introduction to Tableau software
- Visualizing past data, levels, trendlines, seasonality, forecasts

Microsoft Excel

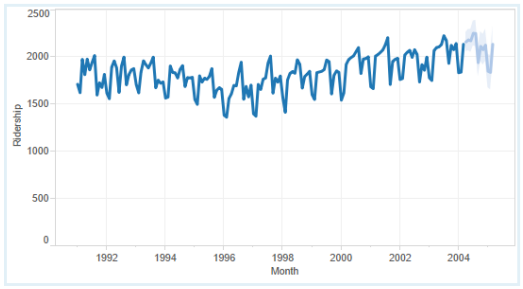


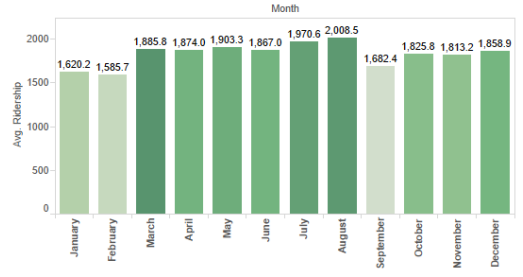
	A	B
1	Month	Ridership
2	1/1/1991	1709
3	2/1/1991	1621
4	3/1/1991	1973
5	4/1/1991	1812
6	5/1/1991	1975
7	6/1/1991	1862
8	7/1/1991	1940
9	8/1/1991	2013
10	9/1/1991	1596
11	10/1/1991	1725
12	11/1/1991	1676
13	12/1/1991	1814
14	1/1/1992	1615
15	2/1/1992	1557
16	3/1/1992	1891

Tableau



Interactive Visualization: Tableau Dashboard





Forecast Indicator

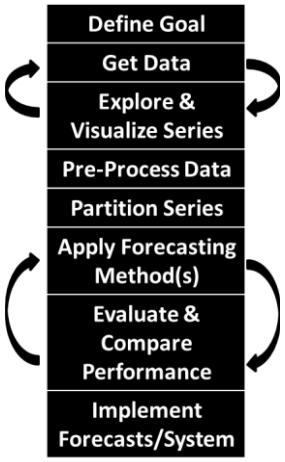
Month of Month	Actual	Estimate
January 2003	1,778.0	
February 2003	1,749.5	
March 2003	2,066.5	
April 2003	2,098.9	
May 2003	2,104.9	
June 2003	2,129.7	
July 2003	2,223.3	
August 2003	2,174.4	
September 2003	1,931.4	
October 2003	2,121.5	
November 2003	2,076.1	
December 2003	2,140.7	
January 2004	1,831.5	
February 2004	1,838.0	
March 2004	2,132.4	
April 2004		2,157.3
May 2004		2,178.3
June 2004		2,169.7
July 2004		2,248.0
August 2004		2,247.1
September 2004		1,939.0
October 2004		2,110.4
November 2004		2,075.9
December 2004		2,123.2
January 2005		1,847.7
February 2005		1,835.0
March 2005		2,132.2

Interactive Visualization

- Zoom and pan
- Filtering
- Change of scale
- Temporal hierarchies and aggregation
- Color
- Dashboards

Supports **detecting patterns** and **exceptions**

PRE-PROCESS DATA



Before modeling, we must **detect** potential challenges

- Missing values
- Unequally-spaced series
- Extreme values
- Time span (how far back?)

Good visualization is very useful

Before next class

Send team names
(5 students) to AA