Data Science Use Case Seminars

Applications in Advertising Technology, Etc.

Alex Trickey
May 8, 2024

Outline

Part I: Intro

- ➔ **My Path in Data Science**
- ➔ Ad Tech Background
- ➔ Time Series Modeling

Part II: Demo

- ➔ Data Exploration
- ➔ Modeling
- ➔ Optimization
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

Academics & Teaching

Dual BS Math, Psych [2009]

Undergrad Thesis used ANN to model EEG responses to painful stimuli.

PhD in Quantitative Psychology [2015]

Research: Heuristic Optimization and Structural Equation Models

Teaching: Statistics, Psychology/Neuroscience

Data Science Roles

Data and Analytics Consultant [2015-2017]

Example Projects:

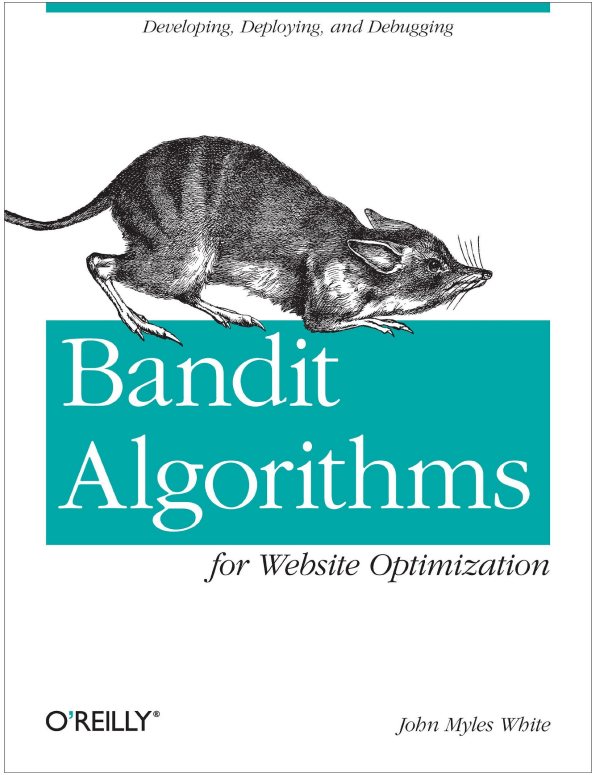
- Data(base) modeling for data migration
- Dashboard to find bottlenecks in video editing/delivery system
- Map film titles across distinct sources

Data Science Roles

SYSTEM1

Ad Tech Data Scientist [2019-2023]

- Modeled clickstream/revenue data
- Optimized which ads are shown
- Used bandits to manage the explore-exploit dilemma
- Reported on performance

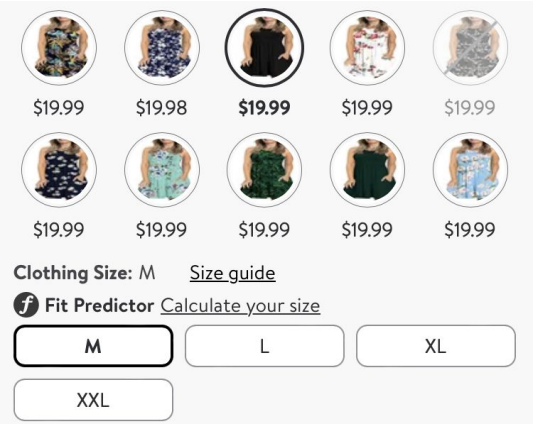


Data Science Roles

Senior Data Scientist in Fashion Tech [2019-2023]

- Developed fit prediction algorithms and their production modeling jobs

- Example Projects: model improvements, multi-person profile handling, codebase migration, backtesting framework, analytics advising, bug investigations



Data Science Roles

UTOPUS INSIGHTS

Squad Lead [2023-2024]

- Led team in development of alerting systems to detect and forecast mechanical failures in wind turbines to enable repairs and avoid costly component replacements and extended downtime

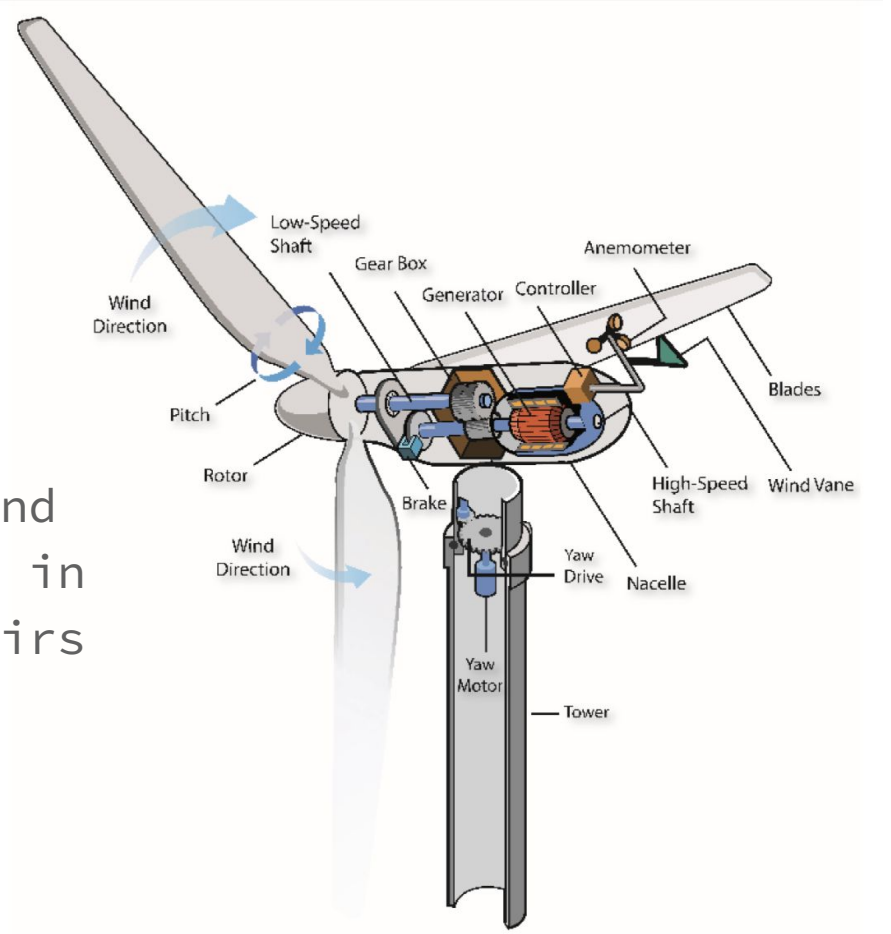
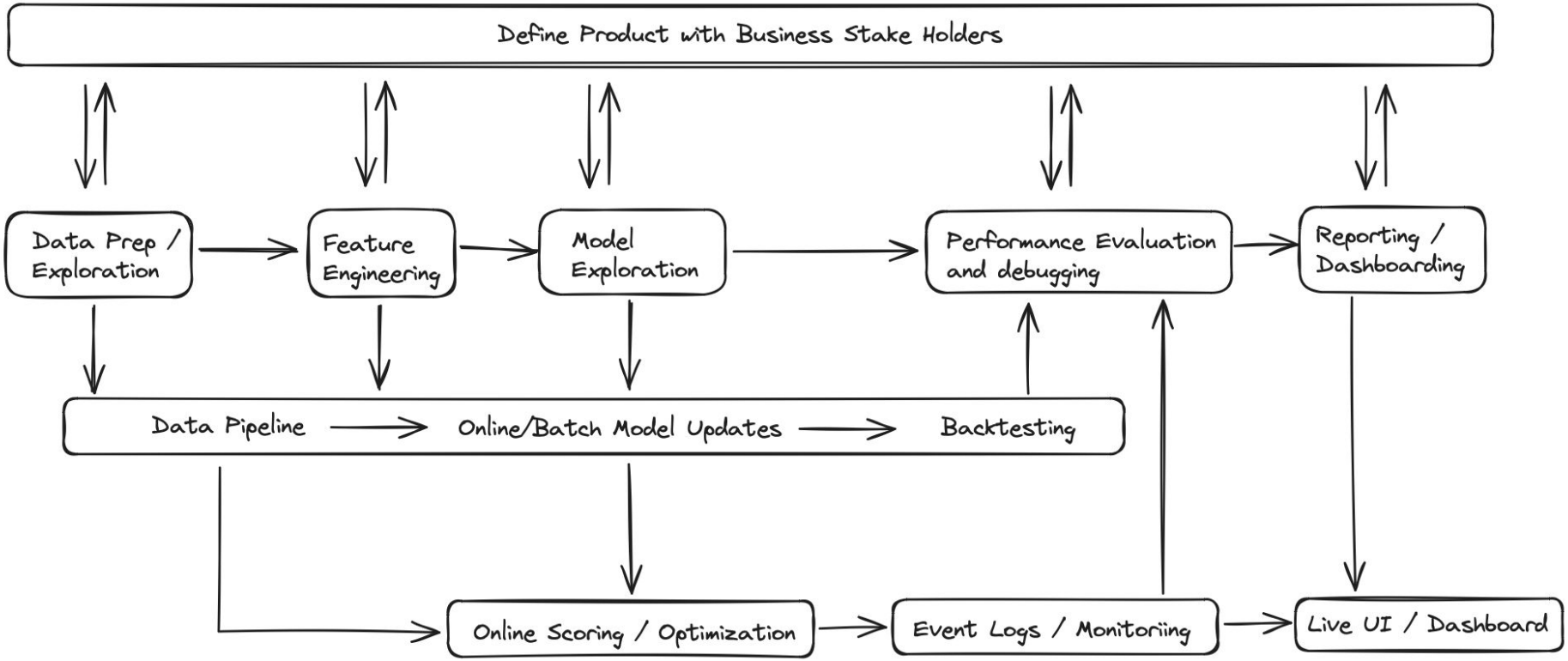
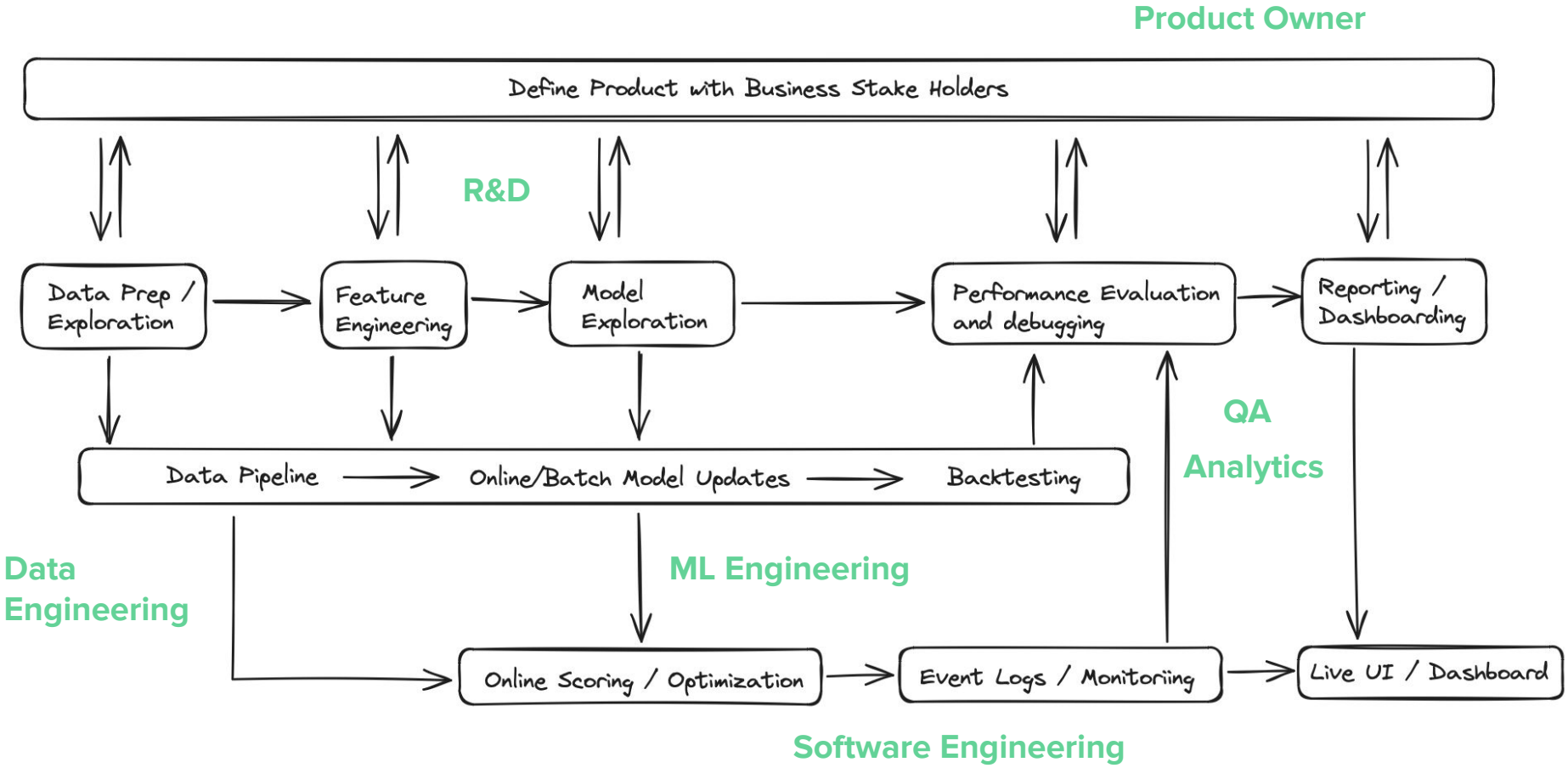


Image Reference:
https://www.mdpi.com/2227-9717/8/9/1066

What even is Data Science?



Specializations / Related Roles



Outline

Part I

- ➔ My Path in Data Science
- ➔ **Ad Tech Background**
- ➔ Time Series Modeling

Part II: Demo

- ➔ Data Exploration
- ➔ Modeling
- ➔ Optimization
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

What Is Ad Tech?

Advertising technology

The set of tools, analyses, algorithms, strategies, etc used to target and serve ads on the internet

What kinds of companies are in the “Ad Tech” space?

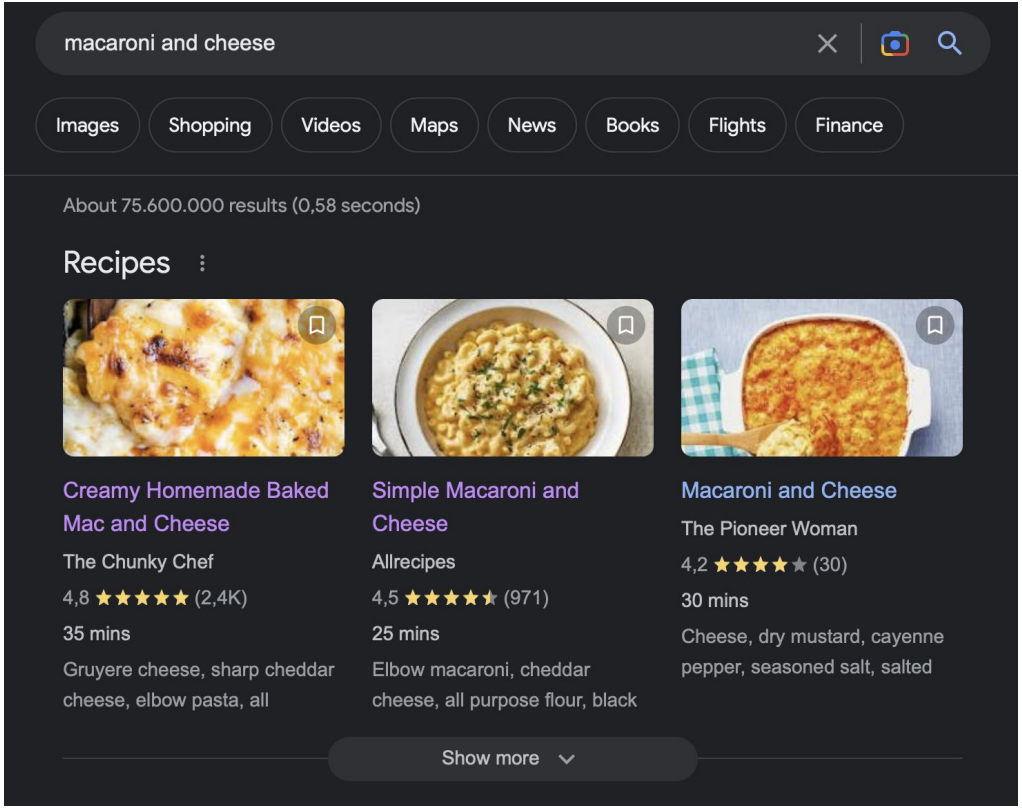
Search Engines: Google, Yahoo, Bing

Media and Merchandising: Facebook, Amazon, BuzzFeed, websites that monetize via ads, etc.

Ad Trading, Tools and Optimization: Critio, MediaMath, AdRoll, The Trade Desk, Rubicon, etc.

Example

Anyone cook with online recipes?

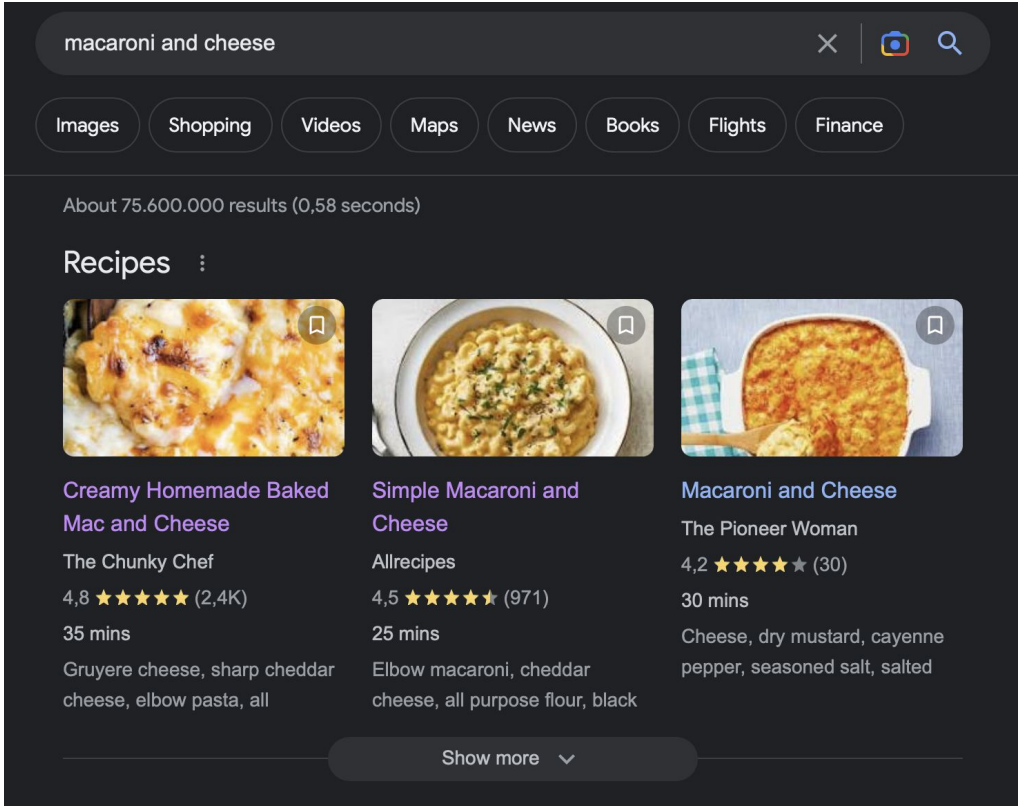


Example

Anyone cook with online recipes?

Strategies to Increase Revenue:

- ➔ place higher valued ads
- ➔ place “clickier” ads
- ➔ modify engagement
- ➔ increase traffic



Useful Metrics

Impression = An object of interest (e.g. an ad) was displayed
Click = An object of interest was clicked

Click-Through-Rate **(CTR**) = Clicks / Impressions
Revenue Per Click **(RPC)** = Revenue / Clicks
Revenue Per Impression **(RPI)** = Revenue / Impressions

Outline

Part I

- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ **Time Series Modeling**
  - ◆ **Intro**
  - ◆ ARIMA
  - ◆ Exponential Smoothing

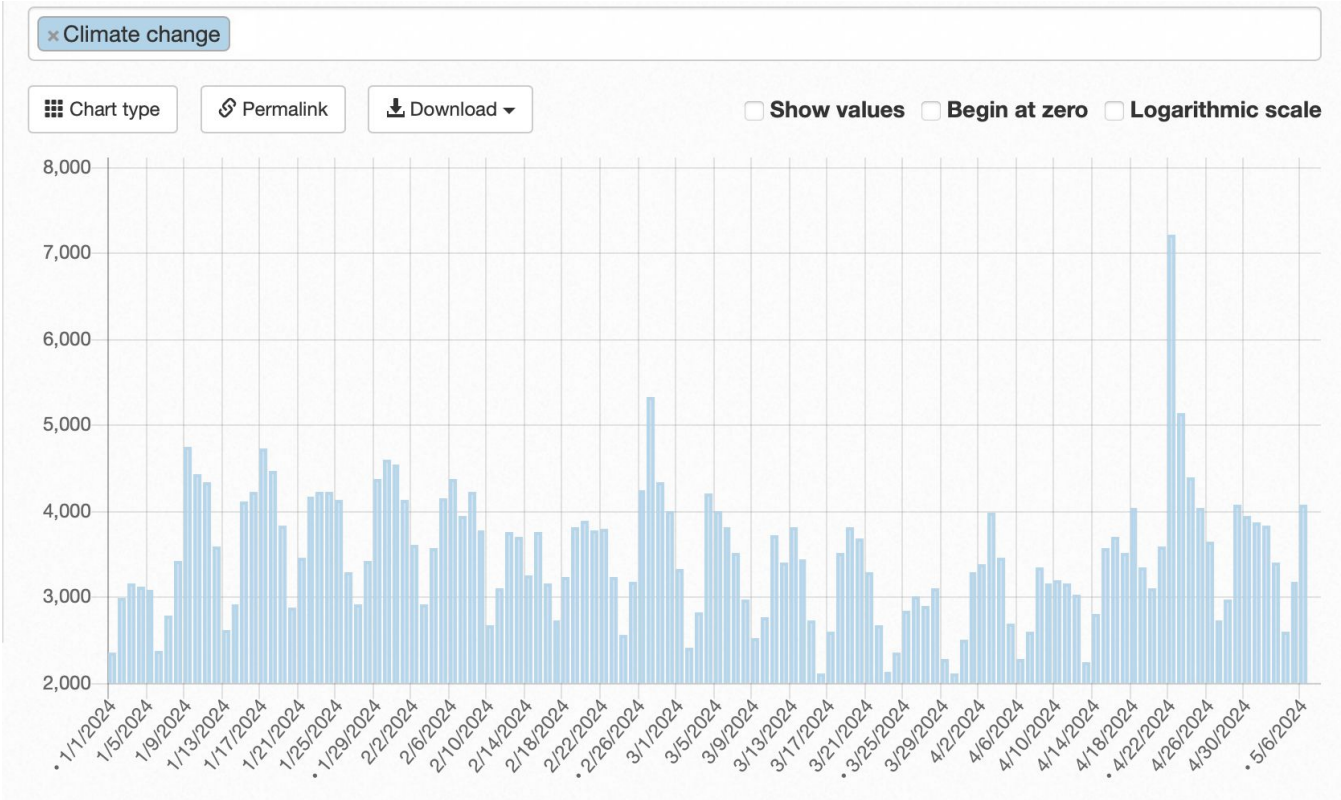
Part II: Demo

- ➔ Data Exploration
- ➔ Modeling
- ➔ Optimization
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

What is special about time-series?

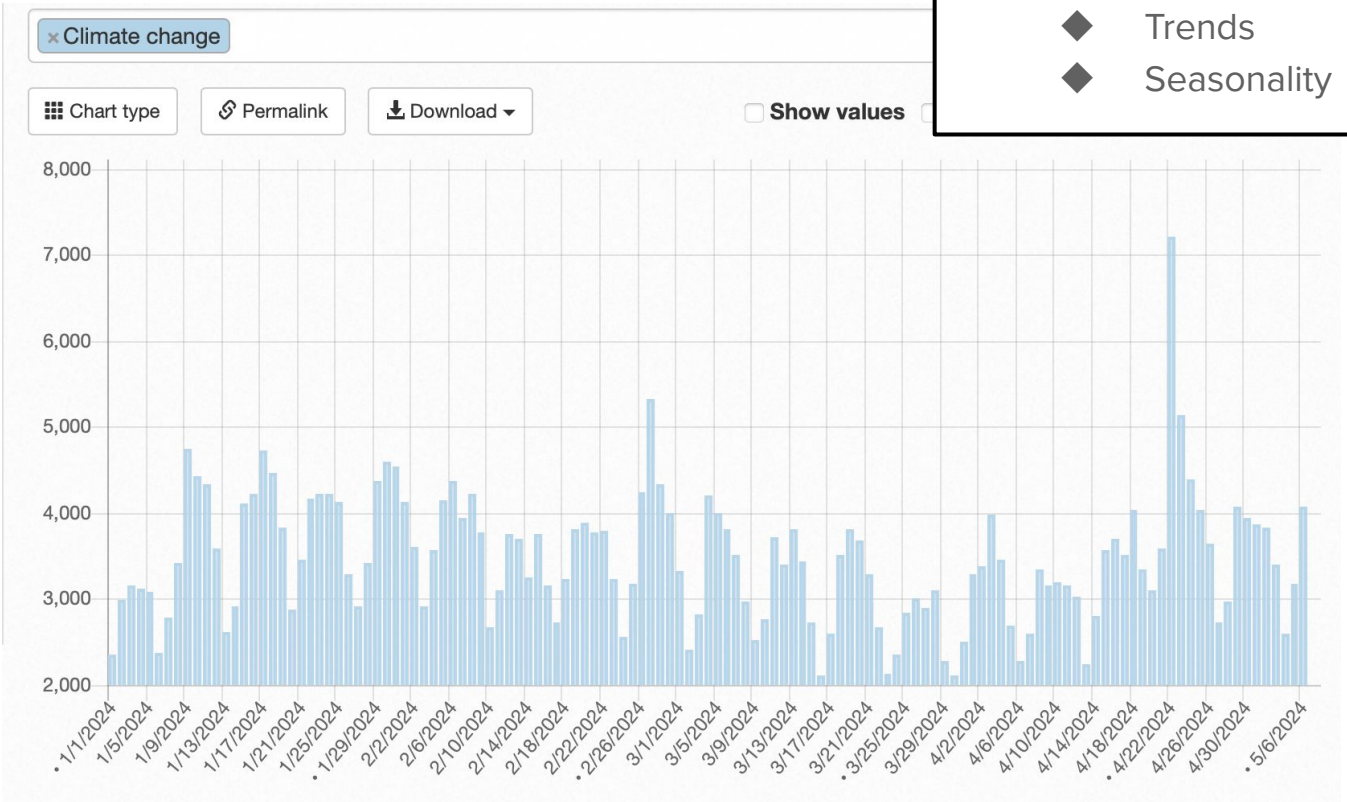
Time Series



Time Series

What is special about time-series?

- Order matters
- Autocorrelation
- Trends
- Seasonality



How Can Time Series Be Modeled?

There are many strategies:

- Classical Time Series Models
- Fourier / Spectral Analysis
- Signal Processing
- Neural Networks / Deep Learning
- Structural / Hierarchical Models
- And many more: https://en.wikipedia.org/wiki/Time_series#Tools

We should choose based on the problem we are trying to solve.

Outline

Part I: Intro

- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ **Time Series Modeling**
  - ◆ Intro
  - ◆ **ARIMA**
  - ◆ Exponential Smoothing

Part II: Demo

- ➔ Data Exploration
- ➔ Modeling
- ➔ Optimization
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

Stationary vs Non-stationary Time Series

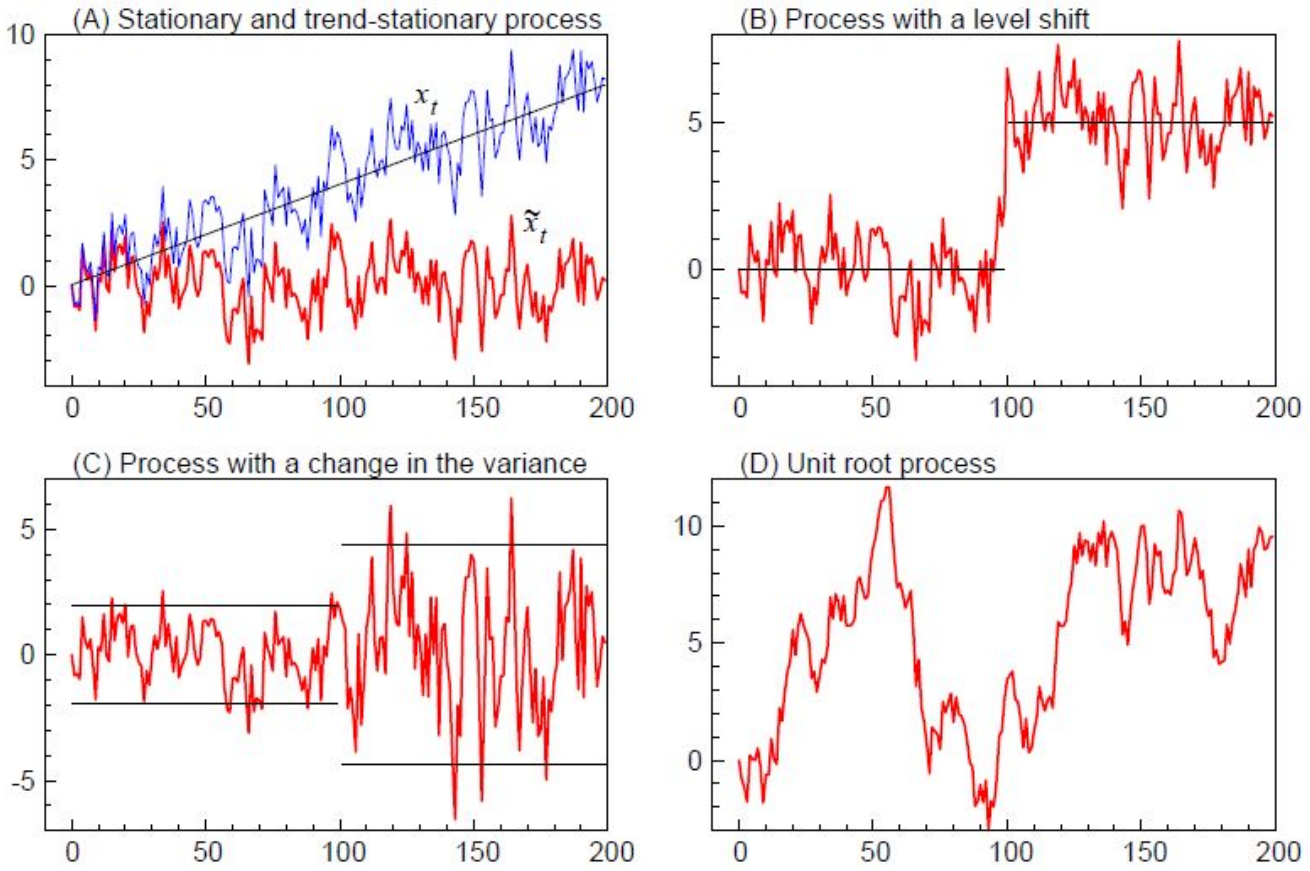


Figure 1: Simulated examples of non-stationary time series.

Source: https://stats.stackexchange.com/questions/4142/using-the-distance-of-p-value-from-alpha

Stationary vs Non-stationary Time Series

The mean, variance and autocorrelation of stationary data do not change over time.

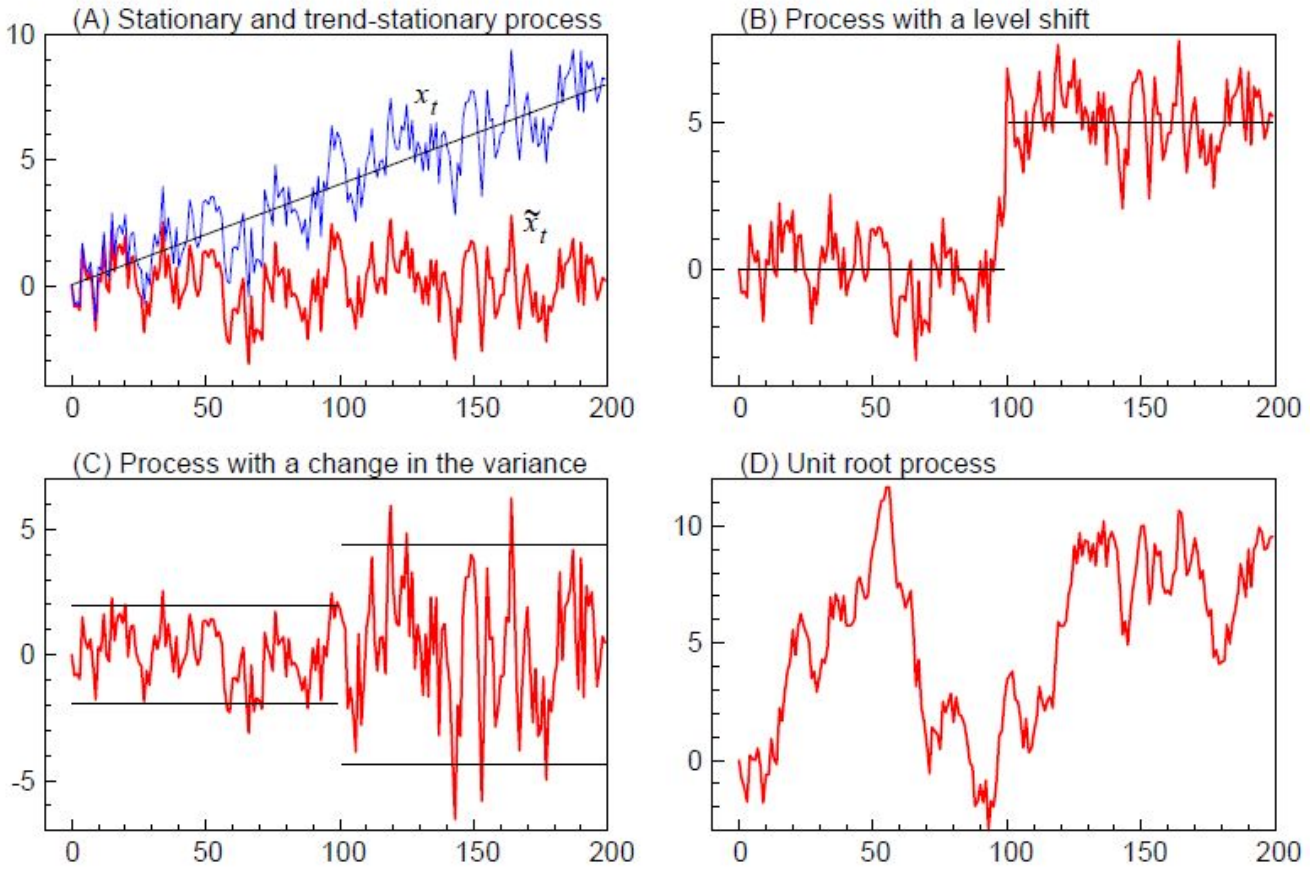


Figure 1: Simulated examples of non-stationary time series.

Image source: https://stats.stackexchange.com/questions/4142/using-the-distance-of-p-value-from-alpha

Models of Stationary Data

Autoregressive Models (AR)

- Current value depends on preceding values
- E.g. Temperature

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

Moving Average Model (MA)

- Current value depends on previous errors
- E.g. Demand for a product

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Models of Stationary Data

Autoregressive Models (AR)

- Current value depends on preceding values
- E.g. Temperature

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

AR + MA = ARMA: $X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$

Moving Average Model (MA)

- Current value depends on previous errors
- E.g. Demand for a product

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

Modeling Non-Stationary Data

ARIMA

- ➔ ARMA with “differencing” transformations to make it stationary.
- ➔ Differencing:
  - ◆ Literally subtracting the previous value from the next one:

$$y'_t = y_t - y_{t-1}$$

  - ◆ Can also be used to account for seasonality:

$$y'_t = y_t - y_{t-m}$$

Outline

Part I: Intro

- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ **Time Series Modeling**
  - ◆ Intro
  - ◆ ARIMA
  - ◆ **Exponential Smoothing**

Part II: Demo

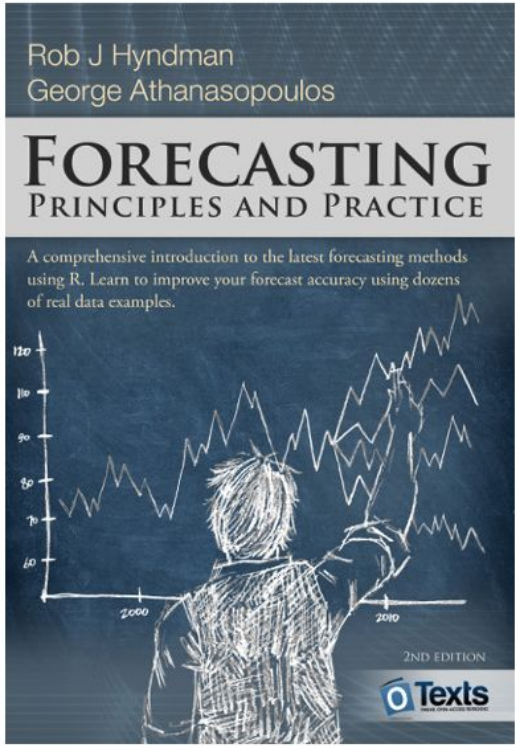
- ➔ Data Exploration
- ➔ Modeling
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

Simple Exponential Smoothing

The forecast is simply a weighted average of past observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots,$$



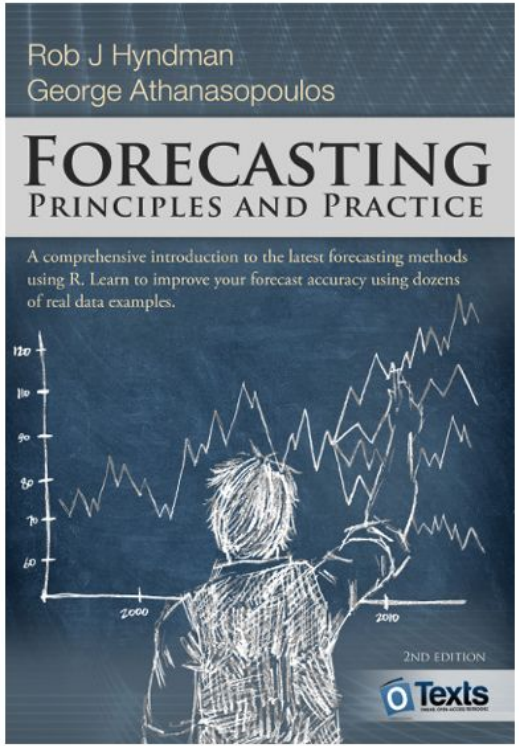
https://otexts.com/fpp2/

Simple Exponential Smoothing

The forecast is simply a weighted average of past observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots,$$

Since alpha is between 0 and 1, the weights decay exponentially as time points get older.



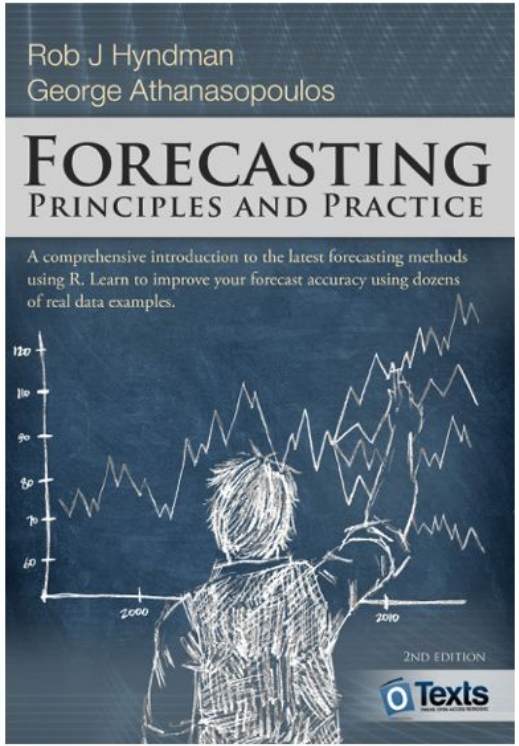
Simple Exponential Smoothing

The forecast is simply a weighted average of past observations:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \cdots,$$

This can also be written in component form:

$$\hat{y}_{t+h|t} = \ell_t$$
$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1},$$



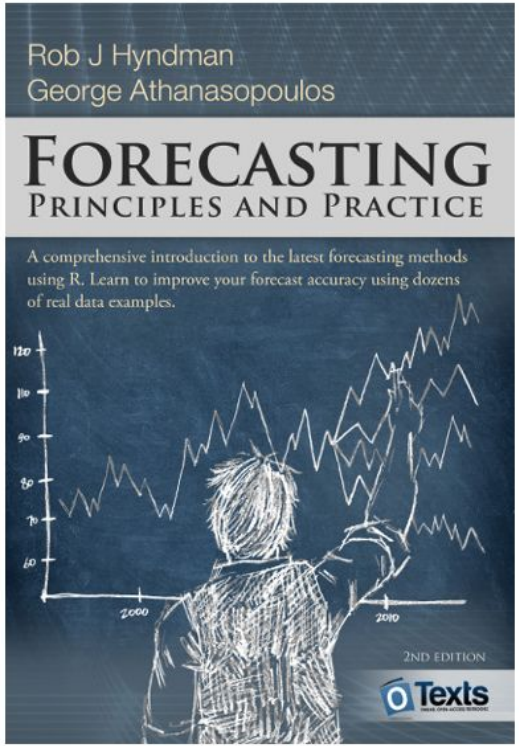
https://otexts.com/fpp2/

Adding a Trend

Simple exponential smoothing doesn't model trend or seasonality.

We can add a slope component to model trend:

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$



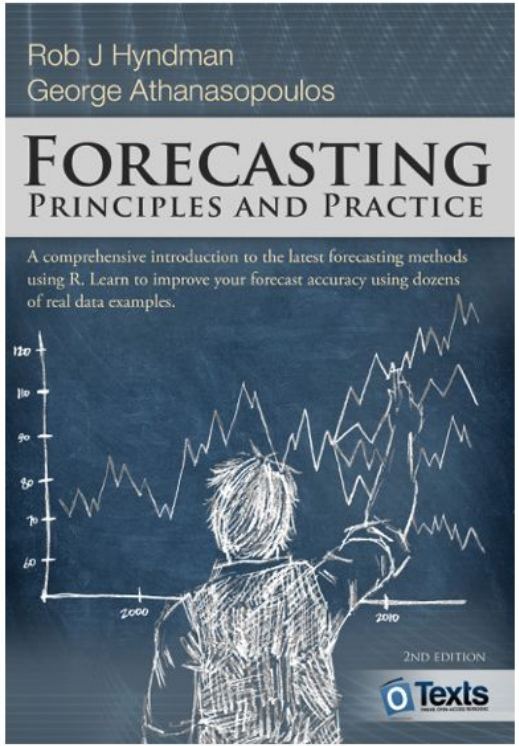
https://otexts.com/fpp2/

Adding a Trend

Simple exponential smoothing doesn't model trend or seasonality.

We can add a slope component to model trend:

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$$



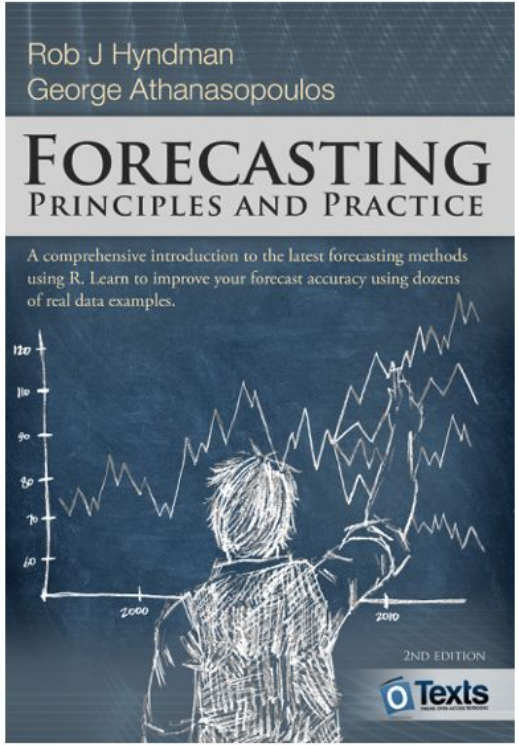
https://otexts.com/fpp2/

Adding a Trend

Simple exponential smoothing doesn't model trend or seasonality.

We can add a slope component to model trend:

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1},$$



https://otexts.com/fpp2/

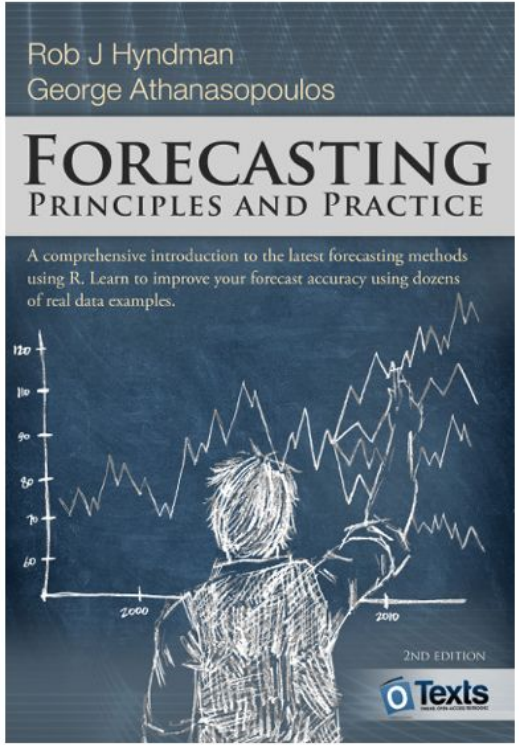
Adding Seasonality

Seasonality can be added similarly:

$$\begin{aligned}\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\ \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},\end{aligned}$$

The s_t component is estimated using what's left after removing the overall level and trend.

This is the Holt-Winters Additive Model.



https://otexts.com/fpp2/

Outline

Part I: Intro

- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ Time Series Modeling

Part II: Demo

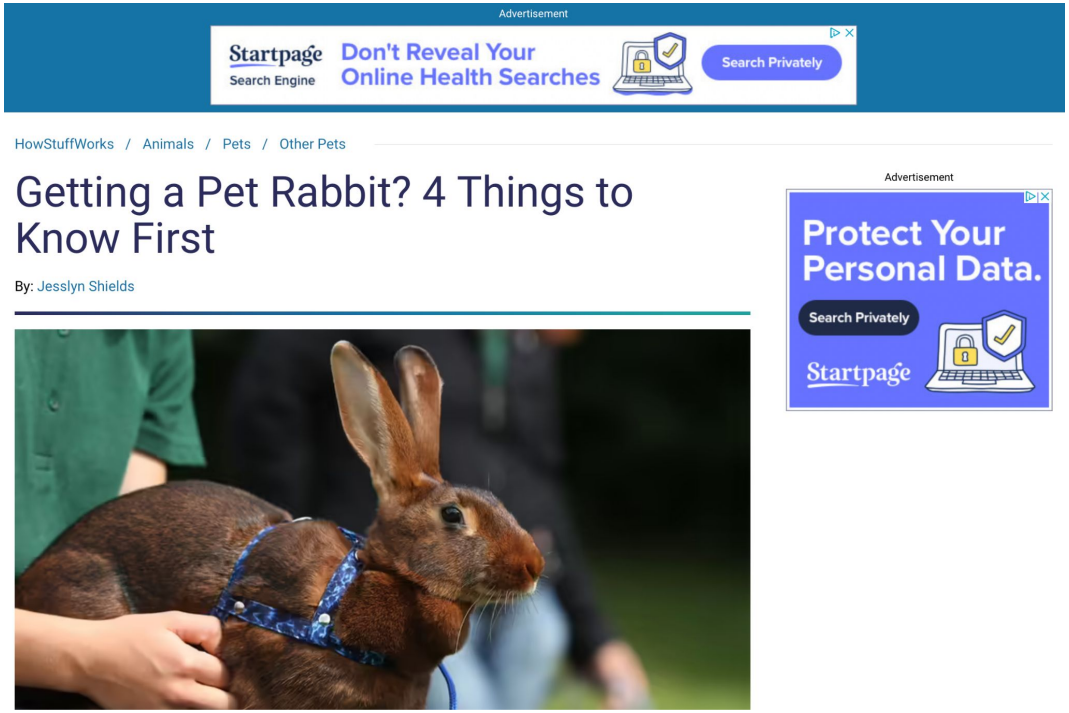
- ➔ **Data Exploration**
- ➔ Modeling
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

Demo - Introduction

We want to decide what type of ad to place on this gem of a website.

We have click and revenue data for three different options. We can use time series models to inform our selection.



Outline

Part I: Intro

- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ Time Series Modeling

Part II: Demo

- ➔ **Data Exploration**
- ➔ Modeling
- ➔ Optimization
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

Outline

Part I: Intro

- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ Time Series Modeling

Part II: Demo

- ➔ Data Exploration
- ➔ **Modeling**
- ➔ Optimization
- ➔ A/B Test

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

Outline

Part I: Intro

- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ Time Series Modeling

Part II: Demo

- ➔ Data Exploration
- ➔ Modeling
- ➔ **Optimization**
- ➔ A/B Test

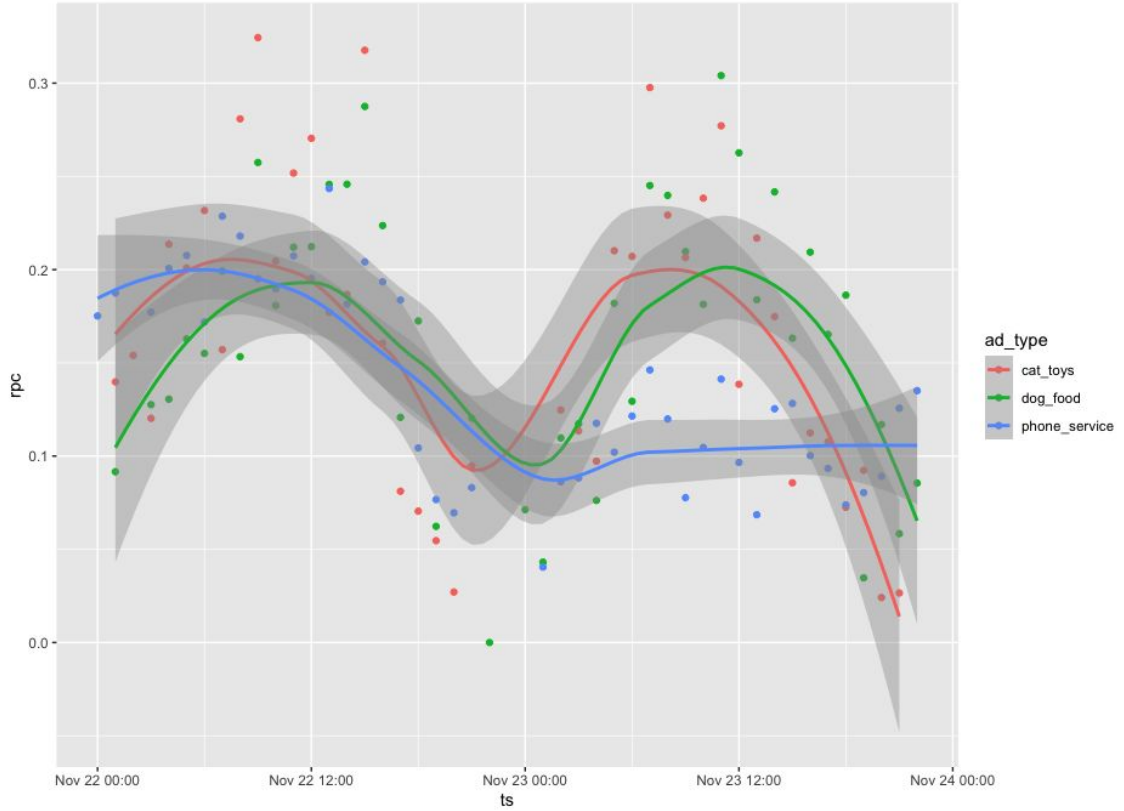
```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

Optimization Considerations

Which ad might we want to show when?

How often might we want to update the model?

What will the training data look like next time we run the forecast?



Goodhart's Law

When a measure becomes a target it ceases to be a good measure.

The new optimization/feature will shape the data.

When you exploit one option, you will lose information about the alternatives. [Explore-Exploit Dilemma]

Related Topics: Reinforcement Learning, Bandit Algorithms

Outline

Part I: Intro

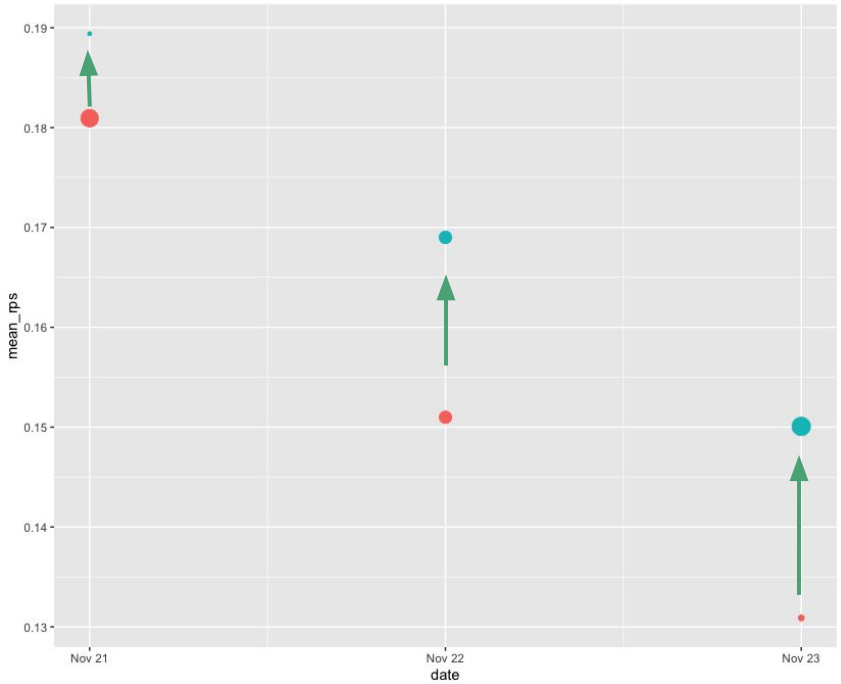
- ➔ My Path in Data Science
- ➔ Ad Tech Background
- ➔ Time Series Modeling

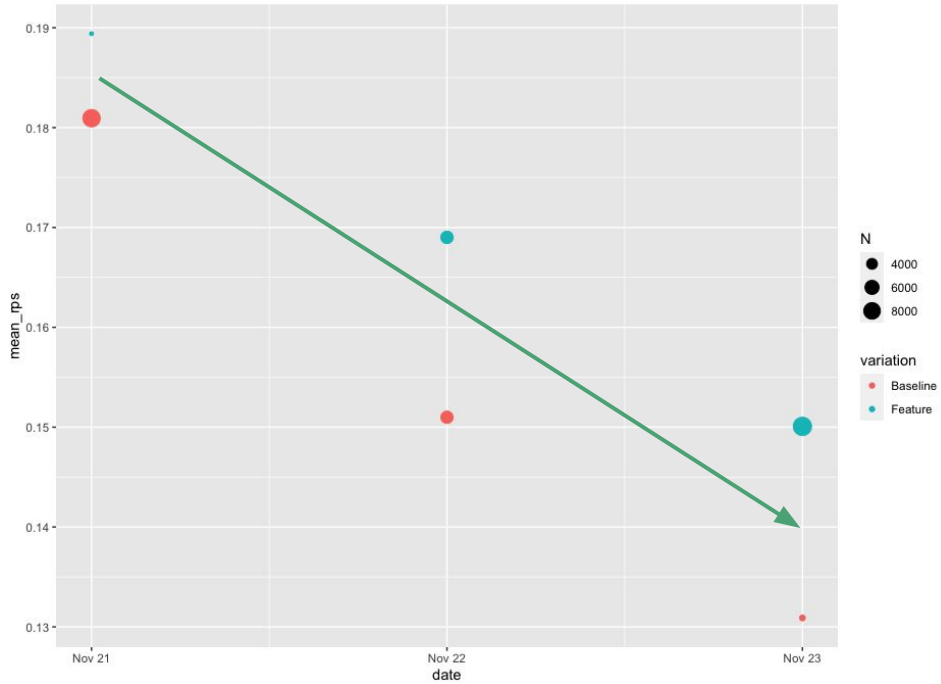
Part II: Demo

- ➔ Data Exploration
- ➔ Modeling
- ➔ Optimization
- ➔ **A/B Test**

```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

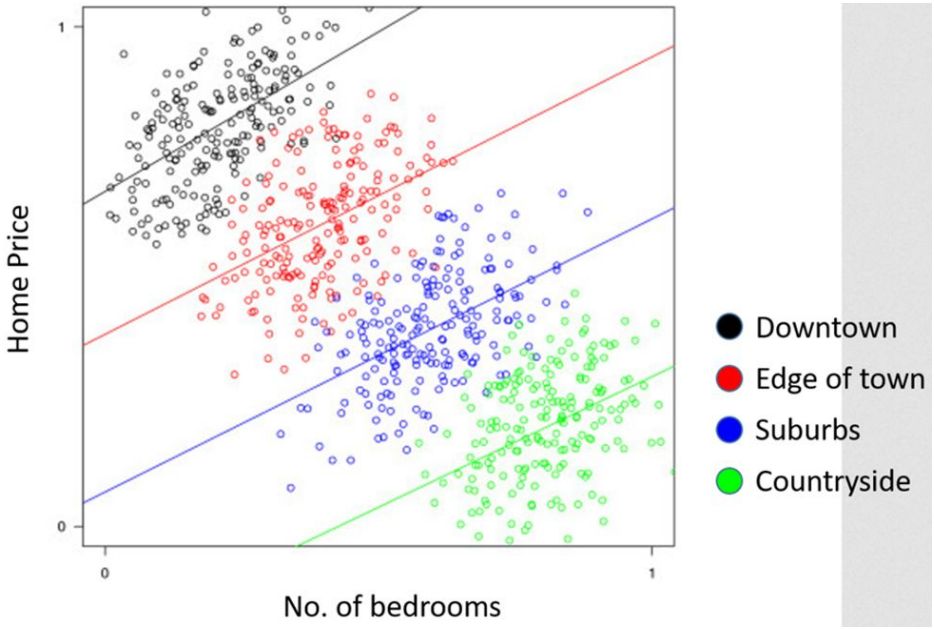
There are two opposite trends...





What is the effect of sending sending more traffic to the new feature?

Simpson's Paradox



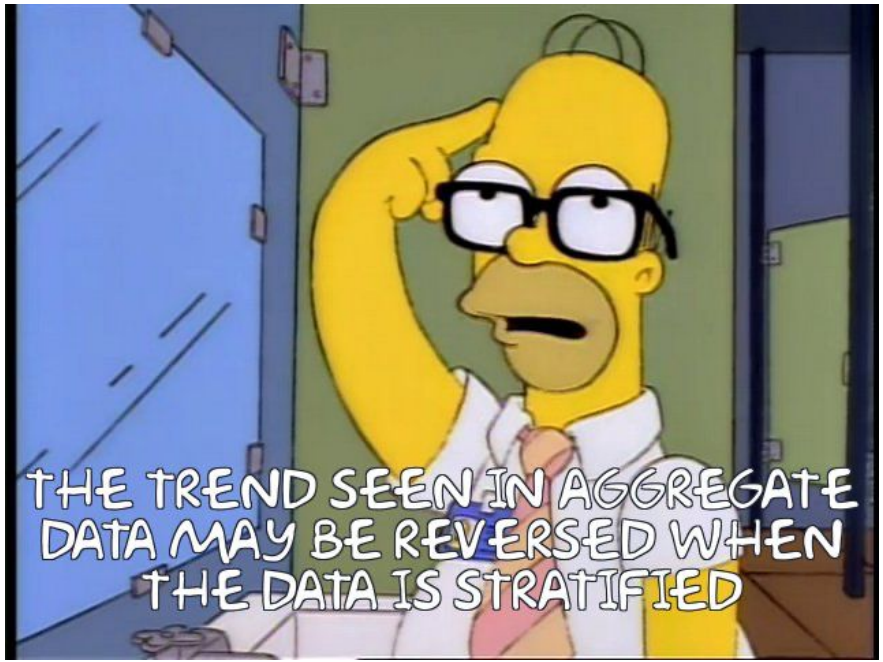


Image sources: https://trends.ufm.edu/en/article/tax-burdens-per-capita-income-simpsons-paradox/
https://www.methodsman.com/blog/pasta-bmi-and-simpsons-paradox

Thank You!