Data Science Use Case Seminars

Working with Time Series in Ad Tech and e-Commerce

November 24, 2021

Outline

- ➔ **My Background / Use Cases**
- ➔ Demo - Part I - Data Exploration
- ➔ Time Series Modeling
- ➔ Demo - Part II - Modeling
- ➔ Optimization
- ➔ Demo - Part III - A/B Testing
- ➔ Simpson's Paradox in Time
- ➔ Q&A / Discussion

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

Data Science Experience

Fit prediction and style matching services for online clothing retailers

Ad Optimization

Data and analytics consulting for large media and entertainment companies

PhD/Teaching in Psychology and Statistics

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 baking or cooking something special for the holidays?

Example

Anyone baking or cooking something special for the holidays?

Strategies to Increase Revenue:

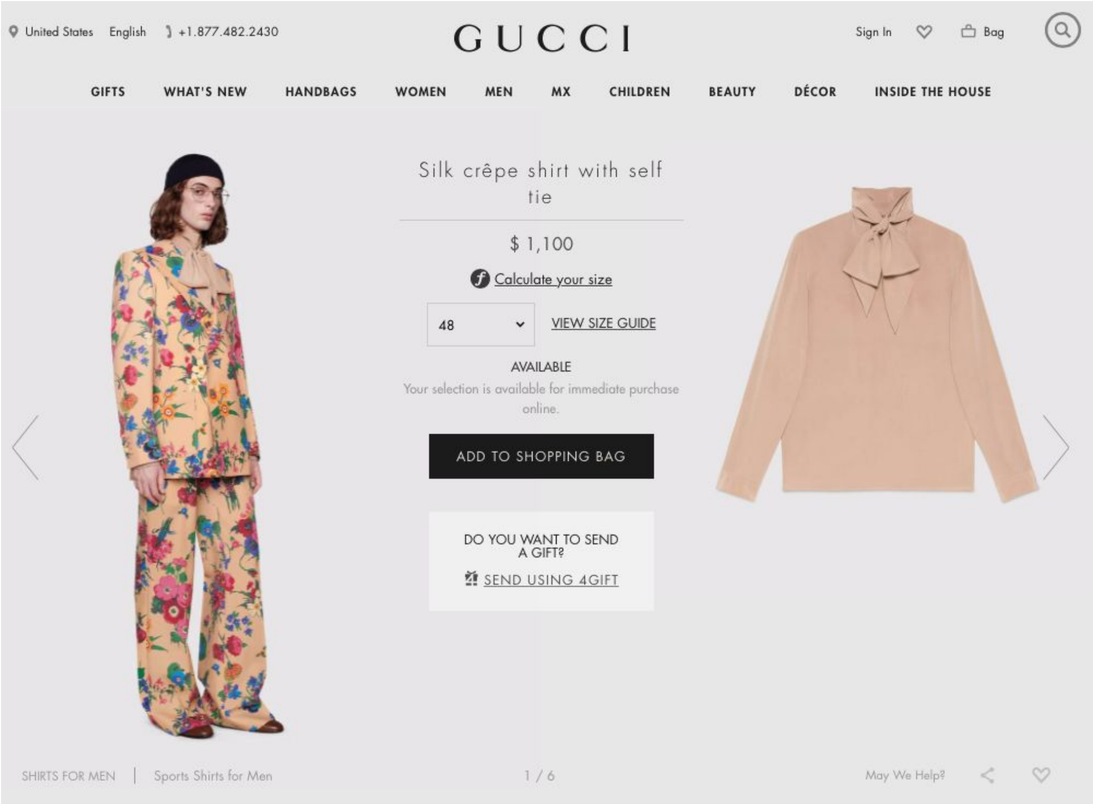
➔ place higher valued ads
➔ place “clickier” ads
➔ modify engagement
➔ increase traffic

What about e-Commerce?

Pretty much every aspect of an online shopping experience can/will be optimized.

Example

Fashion sites work to make it easy for customers to make a purchase and reduce the likelihood that an item is returned.



Outline

- My Background / Use Cases
- **Demo - Part I - Data Exploration**
- Time Series Modeling
- Demo - Part II - Modeling
- Optimization
- Demo - Part III - A/B Testing
- Simpson's Paradox in Time
- Q&A / Discussion

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

Useful Metrics in Ad Tech & e-Commerce

Visit = Someone came to the site

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

Revenue Per Visit **(RPV)** = Revenue / Visits

Conversion Rate = Purchases / Visits

Demo - Part I

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

Let's take a look at the data...

Getting a Pet Rabbit? 4 Things to Know First

BY JESSLYN SHIELDS   MAY 9, 2019

Rabbits are cute and cuddly, but they require veterinary care just as a dog or cat does. RALPH ORLOWSKI/GETTY IMAGES

Imagine having a pet bunny to snuggle on the couch while you watch Netflix, or maybe to hop around in your yard, posing for the cameras with your children on Easter morning. All that sounds pretty adorable, if you're into that kind of thing. And yet, like any pet, rabbits are also a much bigger commitment than you might realize. So, what do you need to know before picking up a rabbit and commencing couch

Outline

- My Background / Use Cases
- Demo - Part I - Data Exploration
- **Time Series Modeling**
  - **Intro**
  - ARIMA
  - Exponential Smoothing
- Demo - Part II - Modeling
- Optimization
- Demo - Part III - A/B Testing
- Simpson’s Paradox in Time
- Q&A / Discussion

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

What is special about time-series?

Time Series

Source:
https://www.dailyfx.com/forex/fundamental/article/special_report/2018/11/30/Brexit-Impact-on-GBP-How-the-Pound-Might-Move-After-Parliamentary-Vote.html

Time Series

What is special about time-series?

- Order matters
- Autocorrelation
- Trends
- Seasonality

Source:
https://www.dailyfx.com/forex/fundamental/article/special_report/2018/11/30/Brexit-Impact-on-GBP-How-the-Pound-Might-Move-After-Parliamentary-Vote.html

What Kinds of Models Are Time Series Models?

What kinds of methods can be used to model time series?

- 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

- ➔ My Background / Use Cases
- ➔ Demo - Part I - Data Exploration
- ➔ **Time Series Modeling**
  - ◆ Intro
  - ◆ **ARIMA**
  - ◆ Exponential Smoothing
- ➔ Demo - Part II - Modeling
- ➔ Optimization
- ➔ Demo - Part III - A/B Testing
- ➔ Simpson's Paradox in Time
- ➔ Q&A / Discussion

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

Stationary vs Non-stationary Time Series

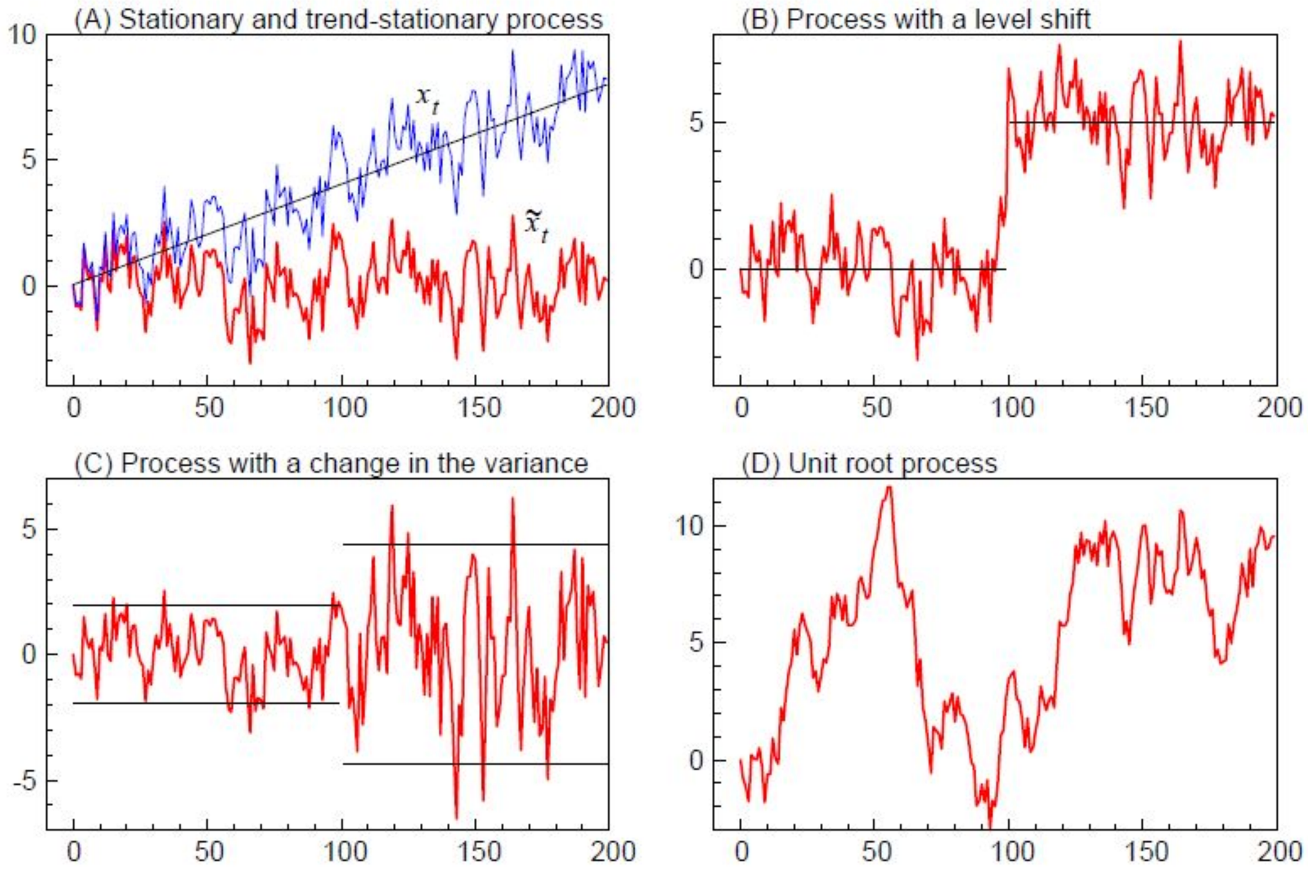


Figure 1: Simulated examples of non-stationary time series.

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

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

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

- My Background / Use Cases
- Demo - Part I - Data Exploration
- **Time Series Modeling**
  - Intro
  - ARIMA
  - **Exponential Smoothing**
- Demo - Part II - Modeling
- Optimization
- Demo - Part III - A/B Testing
- Simpson's Paradox in Time
- Q&A / Discussion

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

Simple Exponential Smoothing

The forecast is simply a weighted average of past observations:

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

Often useful to express in terms of components:

$$\begin{aligned} \text{Forecast equation} \qquad & \hat{y}_{t+h|t} = \ell_t \\ \text{Smoothing equation} \qquad & \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}, \end{aligned}$$

Adding in Trend and Seasonality

We can add in extra components for trend and seasonality.

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

The b_t component estimates the slope.

The s_t component subtracts away the overall level and trend and what's left tells us about seasonality.

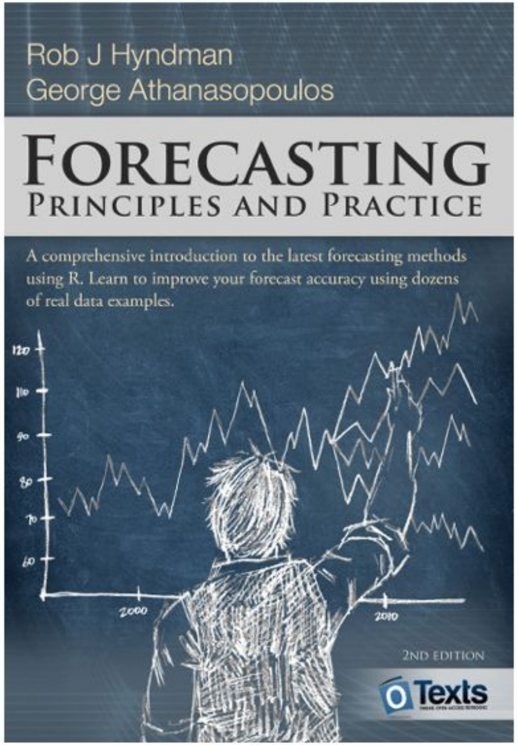
Adding in Trend and Seasonality

We can add in extra components for trend and seasonality.

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

The b_t component estimates the slope.

The s_t component subtracts away the overall level and trend and what's left tells us about seasonality.



https://otexts.com/fpp2/

Outline

- ➔ My Background / Use Cases
- ➔ Demo - Part I - Data Exploration
- ➔ Time Series Modeling
- ➔ **Demo - Part II - Modeling**
- ➔ Optimization
- ➔ Demo - Part III - A/B Testing
- ➔ Simpson's Paradox in Time
- ➔ Q&A / Discussion

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

Outline

- My Background / Use Cases
- Demo - Part I - Data Exploration
- Time Series Modeling
- Demo - Part II - Modeling
- **Optimization**
- Demo - Part III - A/B Testing
- Simpson's Paradox in Time
- Q&A / Discussion

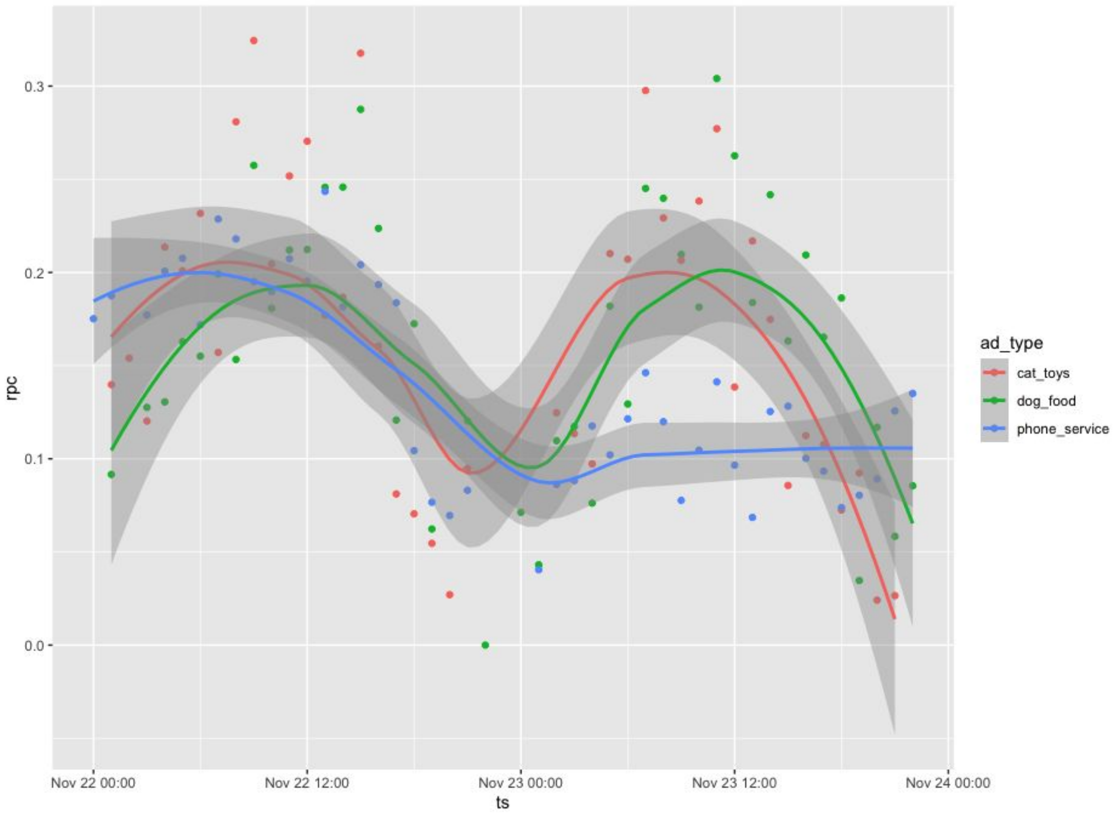
```
# Demo Materials:
git clone https://github.com/alextrickey/ts_demo
```

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

- → My Background / Use Cases
- → Demo - Part I - Data Exploration
- → Time Series Modeling
- → Demo - Part II - Modeling
- → Optimization
- → **Demo - Part III - A/B Testing**
- → Simpson's Paradox in Time
- → Q&A / Discussion

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

Are we making money?

We predicted our new feature / optimization would improve some key performance indicator (e.g. Conversion Rate, RPV, etc.).

But things don’t always go as planned.

We need to monitor/test it via:

- A/B Testing
- Maintaining a hold out set that does not receive the new feature

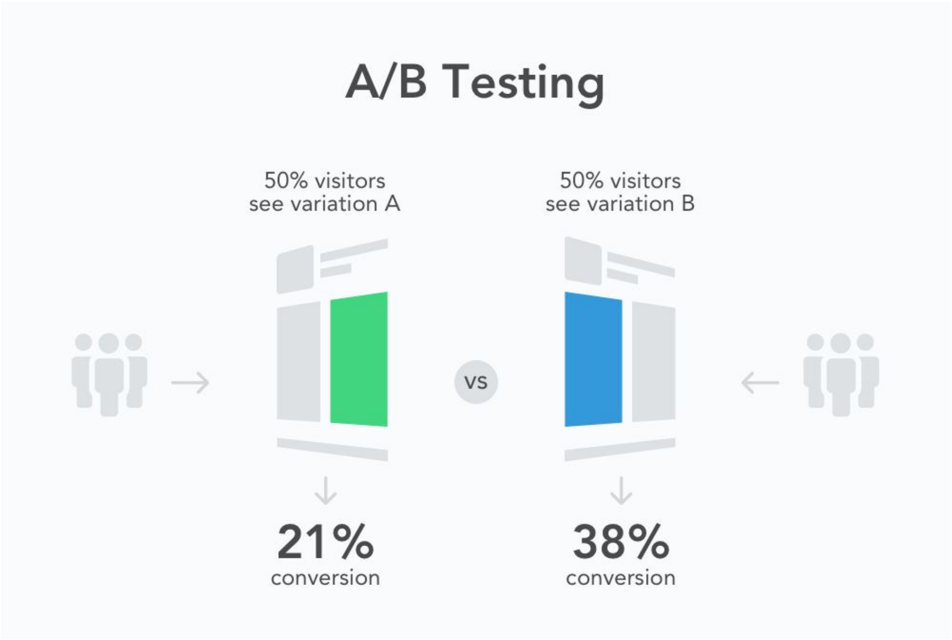


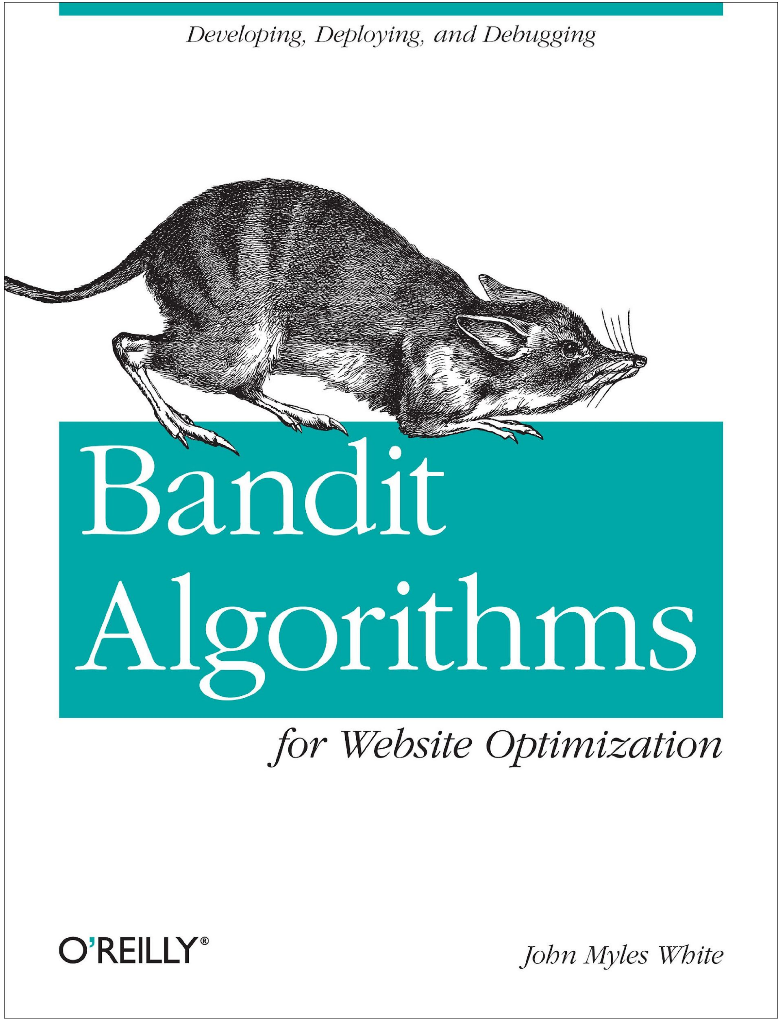
Image Source: https://appradar.com/academy/aso-advanced/app-store-listing-a-b-testing

Demo - Part III

Suppose, we implemented our forecasting strategy and decide to use a Bandit Algorithm to choose with ad to play based on the forecasts.

We rolled it out on 20% of the traffic.

Should we send it more traffic?

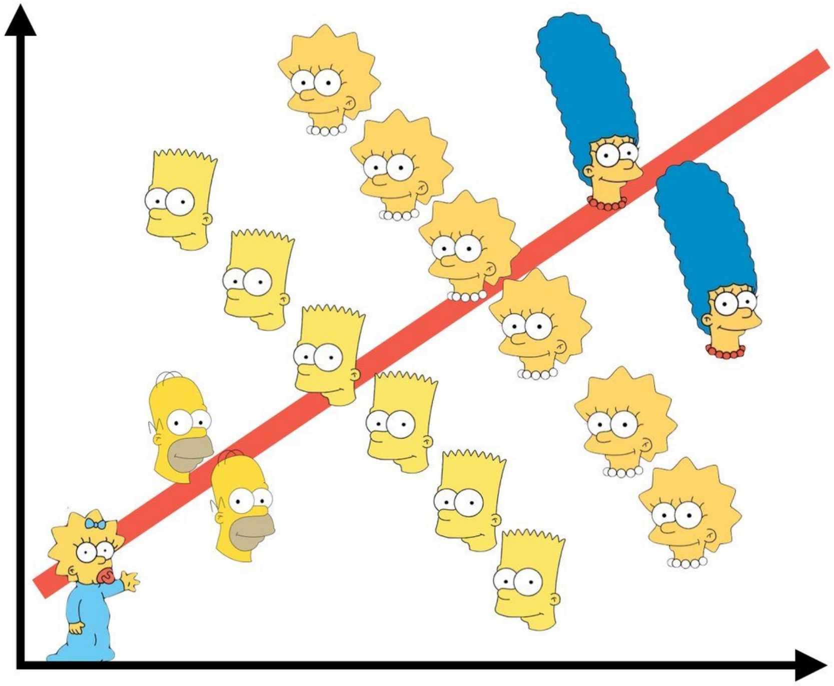


Outline

- My Background / Use Cases
- Demo - Part I - Data Exploration
- Time Series Modeling
- Demo - Part II - Modeling
- Optimization
- Demo - Part III - A/B Testing
- **Simpson's Paradox in Time**
- Q&A / Discussion

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

Simpson's Paradox



"**Simpson's paradox** is a phenomenon in probability and statistics, in which a trend appears in several different groups of data but disappears or reverses when these groups are combined."

- Wikipedia Definition

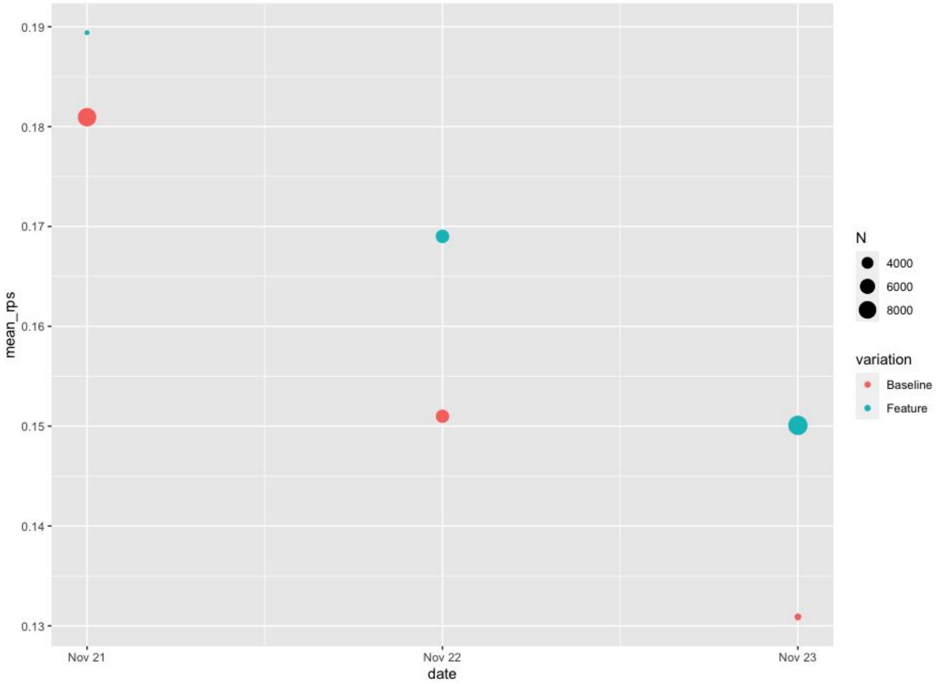
Source: https://www.analyticsindiamag.com/understanding-simpsons-paradox-and-its-impact-on-data-analytics/

Ok, but which trend do we believe?

We have to understand what is causing the different trends.

What are some possible explanations for the trend we see on the right?

What are our nexts steps/what might we do differently next time?



Thank You!

Questions???

- ➔ My Background / Use Cases
- ➔ Demo - Part I - Data Exploration
- ➔ Time Series Modeling
- ➔ Demo - Part II - Modeling
- ➔ Optimization
- ➔ Demo - Part III - A/B Testing
- ➔ Simpson’s Paradox in Time
- ➔ **Q&A / Discussion**

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