# Building Optimized and Hyper-local Product Assortments: A Nonparametric Choice Approach

Vivek Farias (MIT) Srikanth Jagabathula (NYU) Devavrat Shah (MIT)

Special thanks to the Celect team: Akshay Anand, John Andrews, Juan Chaneton, Todd Harris, Ying Zong Huang, Vin Sachidananda, Vig Sachidananda and Greg Yu

# BUSINESS CONTEXT AND PROBLEM

# An introduction to Celect, Inc.

- Overview
  - Predictive analytics platform (SaaS) focused on the retail sector
  - Based in Boston, MA; Venture-backed

- Awards & Recognitions
  - MIT Computer Science and Artificial Intelligence (CSAIL) Top 50 Greatest Innovations
  - DEMO Fall 2014 – DEMO God Winner

# Celect, Inc.: sample customers

# Celect: predictive analytics solutions for retail

our focus


Image credit: Celect, Inc.

# Assortment problem at a Celect client

- Mid-size retailer (~$3B in revenue; close to 300 stores)
- Key decision: Merchandizing function (involving millions of $s)

# Assortment problem at a Celect client

1. Determine assortment
2. Determine purchase quantities

Michael Kors

$20M

Chaps

$40M

Coach

$30M

# Assortment problem at a Celect client

1. Determine assortment
2. Determine purchase quantities
3. Allocate to stores

Michael Kors

$20M

Chaps

$30M

Store 1

Chaps

$10M

Coach

$30M

Store 2

# Assortment problem at a Celect client

1. Determine assortment
2. Determine purchase quantities
3. Allocate to stores

Strategic initiative: hyper-localize product allocation to

identify revenue growth opportunities
minimize inventory waste

# Assortment problem at a Celect client

Strategic initiative: hyper-localize product allocation to

identify revenue growth opportunities

minimize inventory waste

Requires answering the following challenging questions:

1. Michael Kors (MK) bags were sold out at the Minneapolis store last year. What should be the allocation this year?
2. MK handbags were never sold at the Lancaster, PA, store. Should we introduce them?
3. What is the effect of introducing a new brand on existing brands?

# Financial impact

- Mid-size retailer (~$3B in revenue; close to 300 stores)
- Celect task: hyper-localize product assortments

6.65% additional (YoY) revenue growth with *same* investment

$$\text{Potential gross profit impact} = \frac{\text{revenue impact}}{\text{gross margin}} \sim 20\%$$

Potential dollar impact ~ 6.65% x $3B ~ $200M

# How decisions are made today

- combining managerial expertise with trends from data

Increase spend on Athleisure wear, trends suggest it will be big

200 platform sandals were sold last year, 100 two years ago, so we forecast 300 this year

# How decisions are made today

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | R | S | T | U | V | W | X | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 8 DONUT 15 | $80 | 0.0% | $346 | 0.0% | 234 | $1 | 8 DONUT 15 | $5,652 | 2.7% | $2,364 | 0.3% | 0 | 18.6 | $55 | 6909% | 8 DONUT 15 | $6,322 | 0.7% | $362 | 0.0% | 0 | 0.2 | 8 D |
| 16 | 8 PINCH 06 | $234,234 | 5.6% | $23,242 | 0.0% | 236 | $2,296 | 8 PINCH 06 | $3,426 | 1.6% | $36 | 0.0% | 0 | 0.4 | $2 | -100% | 8 PINCH 06 | $236 | 0.0% | $2,362 | 0.1% | 0 | 43.0 | 8 P |
| 17 | 8 PONCHO LJ | $23,432 | 0.6% | $34,643 | 0.1% | 2364 | $230 | 8 PONCHO LJ | $54,642 | 26.0% | $2,346 | 0.3% | 0 | 6.4 | $205 | -11% | 8 PONCHO LJ | $23,623 | 2.7% | $2,362 | 0.1% | 0 | 0.4 | 8 P |
| 18 | X | $3,453,434 | 83.1% | $34,543 | 0.1% | 362 | $33,857 | X | $326 | 0.2% | $0 | 0.0% | 0 | 0.0 | $2 | -100% | X | $236 | 0.0% | $3,626 | 0.1% | 0 | 66.1 | X |
| 45 | Not Applicable | $234 | 0.0% | $45,322,323 | 96.5% | 0 | $2 | Not Applicable | $2,634 | 1.3% | $0 | 0.0% | 0 | ### | $205 | 8854% | Not Applicable | $23,623 | 2.7% | $263 | 0.0% | 0 | 0.0 | No |
| 46 | total | $4,157,896 | 100.0% | $46,978,818 | 100.0% | 3222 | $40,764 | total | $210,306 | 100.0% | $812,011 | 100.0% | 12 | 48.6 | $7,711 | -81% | total | $886,712 | 100.0% | $2,948,539 | 100.0% | 30 | 14.3 | tot |
| 47 | | | | | | | $0 | | | | | | | | | | | | | | | | | |
| 48 | ATTRIBUTE 2 | | | | | | $0 | ATTRIBUTE 2 | | | | | | | | | ATTRIBUTE 2 | | | | | | | AT |
| 49 | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | $/store | COMP % | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | |
| 50 | 8 ROBE UN | $234,225 | 31.0% | $0 | | 236 | $2,296 | 8 ROBE UN | $23,623 | 0.9% | $236 | 0.0% | 0 | 0.0 | $0 | -100% | 8 ROBE UN | $26 | 0.0% | $2,362 | 0.1% | 0 | ### | 8 R |
| 51 | A.KANE IY | $345 | 0.0% | $0 | | 236 | $3 | A.KANE IY | $2,362 | 0.1% | $326,426 | 8.1% | 0 | 0.0 | $2 | -33% | A.KANE IY | $262 | 0.5% | $234,632 | 8.9% | 0 | ### | A.K |
| 52 | A.CHOPPED CR | $24,525 | 3.2% | $0 | | 32462 | $240 | A.CHOPPED CR | $2,633 | 0.1% | $234 | 0.0% | 0 | 0.0 | $205 | -15% | A.CHOPPED CR | $23,623 | 40.8% | $263 | 0.0% | 0 | 0.0 | A.C |
| 53 | 8 BARBIE XI | $5,645 | 0.7% | $0 | | 346 | $55 | 8 BARBIE XI | $23,463 | 0.9% | $3,642,324 | 90.9% | 0 | 0.0 | $31 | -44% | 8 BARBIE XI | $3,564 | 6.2% | $2,634 | 0.1% | 0 | 3.2 | 8 B |
| 54 | 8 PT BREAK EO | $254,634 | 33.7% | $0 | | 362 | $2,496 | 8 PT BREAK EO | $2,363,232 | 87.4% | $3,462 | 0.1% | 0 | 0.0 | $23 | -99% | 8 PT BREAK EO | $2,623 | 4.5% | $23,642 | 0.9% | 0 | 38.8 | 8 P |
| 55 | 8 TRBL MAKER NI | $235,253 | 31.2% | $0 | | 263 | $2,306 | 8 TRBL MAKER NI | $3,623 | 0.1% | $3,264 | 0.1% | 0 | 0.0 | $21 | -99% | 8 TRBL MAKER NI | $2,362 | 4.1% | $263 | 0.0% | 0 | 0.5 | 8 T |
| 56 | A.MUMMy SCARF VN | $564 | 0.1% | $0 | | 26 | $6 | A.MUMMy SCARF VN | $234,623 | 8.7% | $32,642 | 0.8% | 0 | 0.0 | $221 | 3904% | A.MUMMy SCARF VN | $25,462 | 44.0% | $23,623 | 0.9% | 0 | 4.0 | A.M |
| 57 | X | $0 | 0.0% | $0 | | 0 | $0 | X | $26,323 | 1.0% | $0 | 0.0% | 0 | | $0 | | X | $0 | 0.0% | $236 | 0.0% | 0 | | X |
| 58 | X | $0 | 0.0% | $0 | | 0 | $0 | X | $23,636 | 0.9% | $0 | 0.0% | 0 | | $0 | | X | $0 | 0.0% | $2,345,623 | 89.0% | 0 | | X |
| 87 | 0 | $0 | 0.0% | $0 | | 92 | $0 | 0 | $263 | 0.0% | $0 | 0.0% | 0 | | $0 | | 0 | $0 | 0.0% | $236 | 0.0% | 0 | | |
| 88 | Not Applicable | $0 | 0.0% | $0 | | cl | $0 | Not Applicable | $24 | 0.0% | $0 | 0.0% | 0 | | $0 | | Not Applicable | $0 | 0.0% | $2,626 | 0.1% | 0 | | No |
| 89 | | $755,191 | 100.0% | $0 | 0.0% | 34023 | $7,404 | | $2,703,805 | 100.0% | $4,008,588 | 100.0% | 0 | 0.0 | $504 | -93% | | $57,922 | 100.0% | $2,636,140 | 100.0% | 0 | ### | |
| 90 | harry pic | $469,478 | | $0 | | $499 | $4,603 | | $27,246 | | $3,500 | | $0 | 0.0 | $21 | | | $2,388 | | $2,625 | | $0 | 4.7 | |
| 91 | ATTRIBUTE 3 | | | | | | $0 | ATTRIBUTE 3 | | | | | | | | | ATTRIBUTE 3 | | | | | | | AT |
| 92 | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | $/store | COMP % | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | |
| 93 | 8 L/S 67 | $45,245 | 35.9% | $1,315,931 | 61.8% | 18 | $444 | 8 L/S 67 | $95,989 | 3.8% | $208,566 | 21.6% | 4 | ### | $6,508 | 1367% | 8 L/S 67 | $748,433 | 66.6% | $2,790,195 | 69.3% | 25 | 16.0 | 8 L |
| 94 | 8 3/4 SLV 70 | $34,634 | 27.5% | $0 | 0.0% | 0 | $340 | 8 3/4 SLV 70 | $236 | 0.0% | $0 | 0.0% | 0 | 0.0 | $5 | -99% | 8 3/4 SLV 70 | $585 | 0.1% | $5,265 | 0.1% | 1 | 38.7 | 8 3 |
| 95 | 8 S/S 66 | $45,674 | 36.2% | $319,441 | 15.0% | 4 | $448 | 8 S/S 66 | $2 | 0.0% | $319,441 | 33.2% | 4 | 30.1 | $563 | 26% | 8 S/S 66 | $64,770 | 5.8% | $189,362 | 4.7% | 4 | 12.6 | 8 S |
| 96 | 8 S/LESS 68 | $452 | 0.4% | $492,432 | 23.1% | 13 | $4 | 8 S/LESS 68 | $88,227 | 3.5% | $435,512 | 45.2% | 6 | ### | $2,696 | 60737% | 8 S/LESS 68 | $310,031 | 27.6% | $1,042,618 | 25.9% | 13 | 14.5 | 8 S |
| 97 | X | $0 | 0.0% | $0 | 0.0% | 0 | $0 | X | $2,362 | 0.1% | $0 | 0.0% | 0 | | $0 | | X | $0 | 0.0% | $0 | 0.0% | 0 | | X |
| 98 | X | $0 | 0.0% | $0 | 0.0% | 0 | $0 | X | $263 | 0.0% | $0 | 0.0% | 0 | | $0 | | X | $0 | 0.0% | $0 | 0.0% | 0 | | X |
| 105 | Not Applicable | $0 | 0.0% | $0 | 0.0% | 0 | $0 | Not Applicable | $2,362,362 | 92.7% | $0 | 0.0% | 0 | | $0 | | Not Applicable | $0 | 0.0% | $0 | 0.0% | 0 | | No |
| 106 | | $126,005 | 100.0% | $2,127,805 | 100.0% | 35 | $1,235 | | $2,549,440 | 100.0% | $963,519 | 100.0% | 14 | 72.6 | $9,772 | 691% | | $1,123,819 | 100.0% | $4,027,440 | 100.0% | 43 | 15.4 | |
| 107 | LS + 3/4 | $79,879 | 63.4% | $1,315,931 | 61.8% | 18 | $783 | | $26,362 | 3.8% | $208,566 | 21.6% | 4 | 70.8 | $6,513 | 732% | LS + 3/4 | $749,018 | 66.6% | $2,795,460 | 69.4% | 26 | 16.0 | |
| 108 | S/LESS + SS | $46,126 | 36.6% | $811,874 | 38.2% | 17 | | | $265 | 3.5% | $754,953 | 78.4% | 10 | | $3,259 | | S/LESS + SS | $374,801 | 33.4% | $1,231,980 | 30.6% | 17 | | |
| 109 | ATTRIBUTE 4 | | | | | | $0 | ATTRIBUTE 4 | | | | | | | | | ATTRIBUTE 4 | | | | | | | AT |
| 110 | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | $/store | COMP % | | SLS $ | % CONT | INV $ | % CONT | STYLE CT | | |
| 111 | 8 SWAYZE ROUGE | $457,456 | 89.4% | $0 | | 0 | $4,485 | 8 SWAYZE ROUGE | $0 | | $0 | | 0 | 0.0 | $0 | -100% | 8 SWAYZE ROUGE | $0 | 0.0% | $0 | 0.0% | 0 | | 8 S |
| 112 | 8 SQUARE DNC KB | $54,346 | 10.6% | $0 | | 0 | $533 | 8 SQUARE DNC KB | $0 | | $0 | | 0 | 0.0 | $0 | -100% | 8 SQUARE DNC KB | $0 | 0.0% | $0 | 0.0% | 0 | | 8 S |

cv | cllpln | retplan | dd sh | lat plan | online | +

[Image credit: Celect, Inc.]

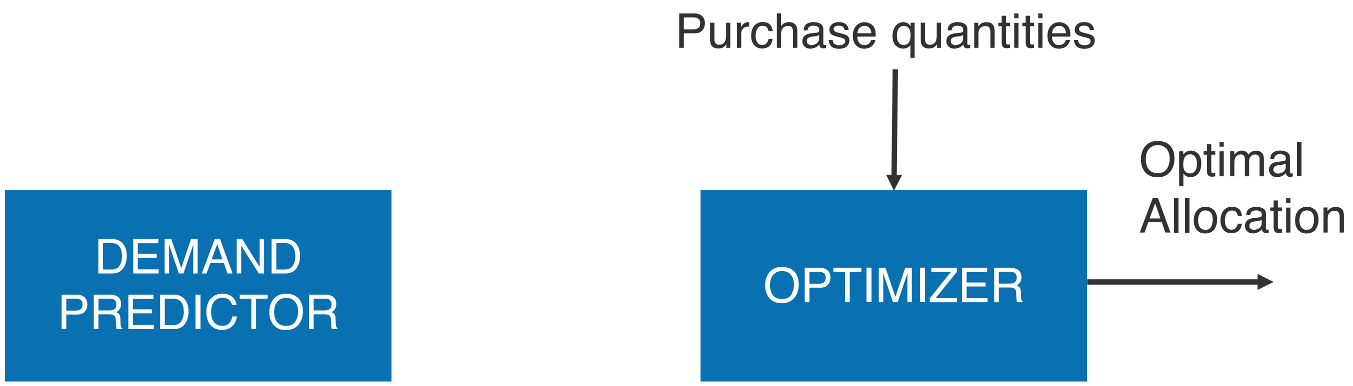
# A BETTER WAY…
# OUR APPROACH

# A better way: automated decision support system

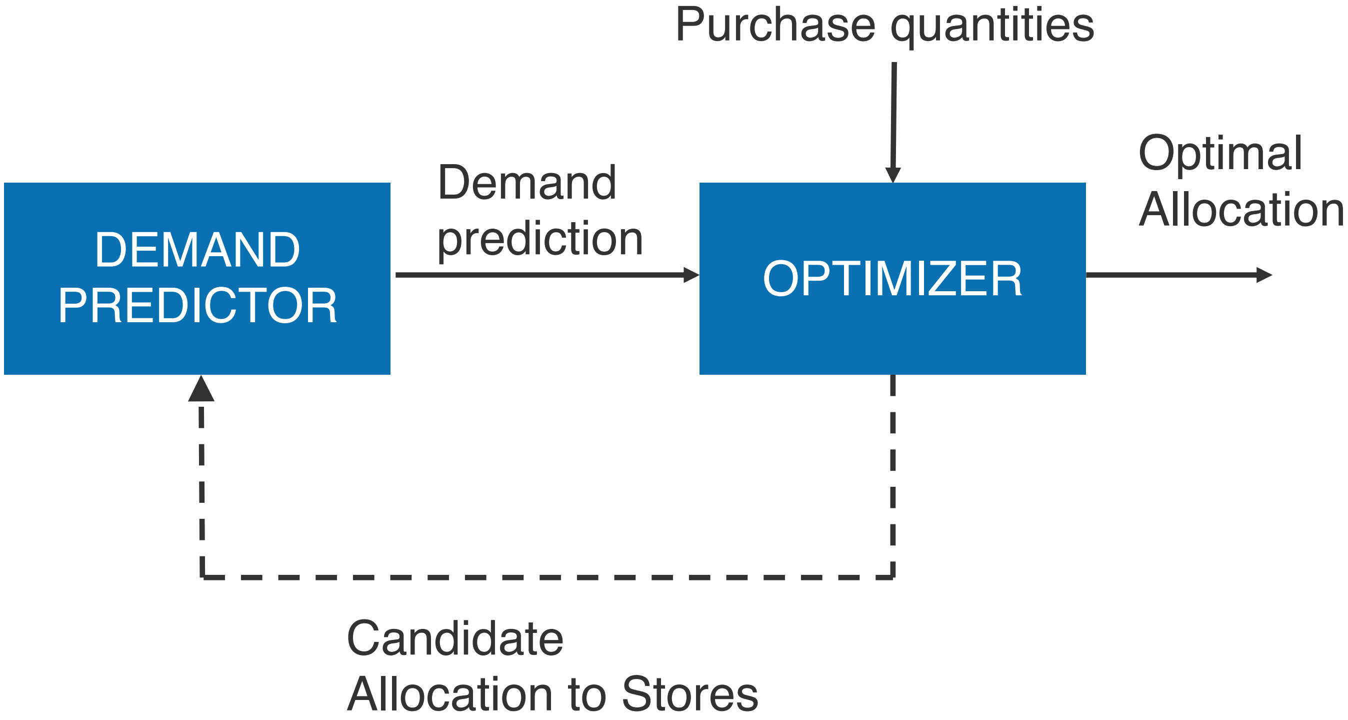
DEMAND PREDICTOR

OPTIMIZER

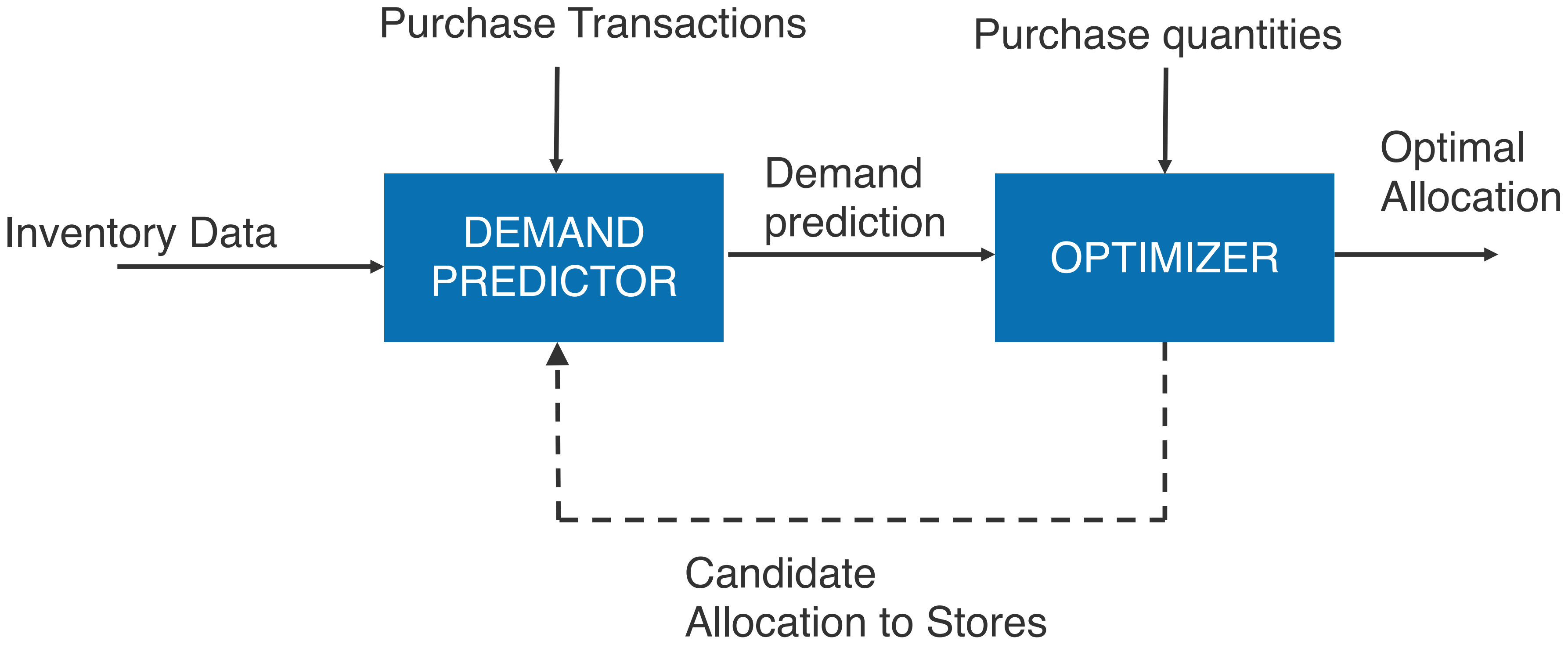
# A better way: automated decision support system



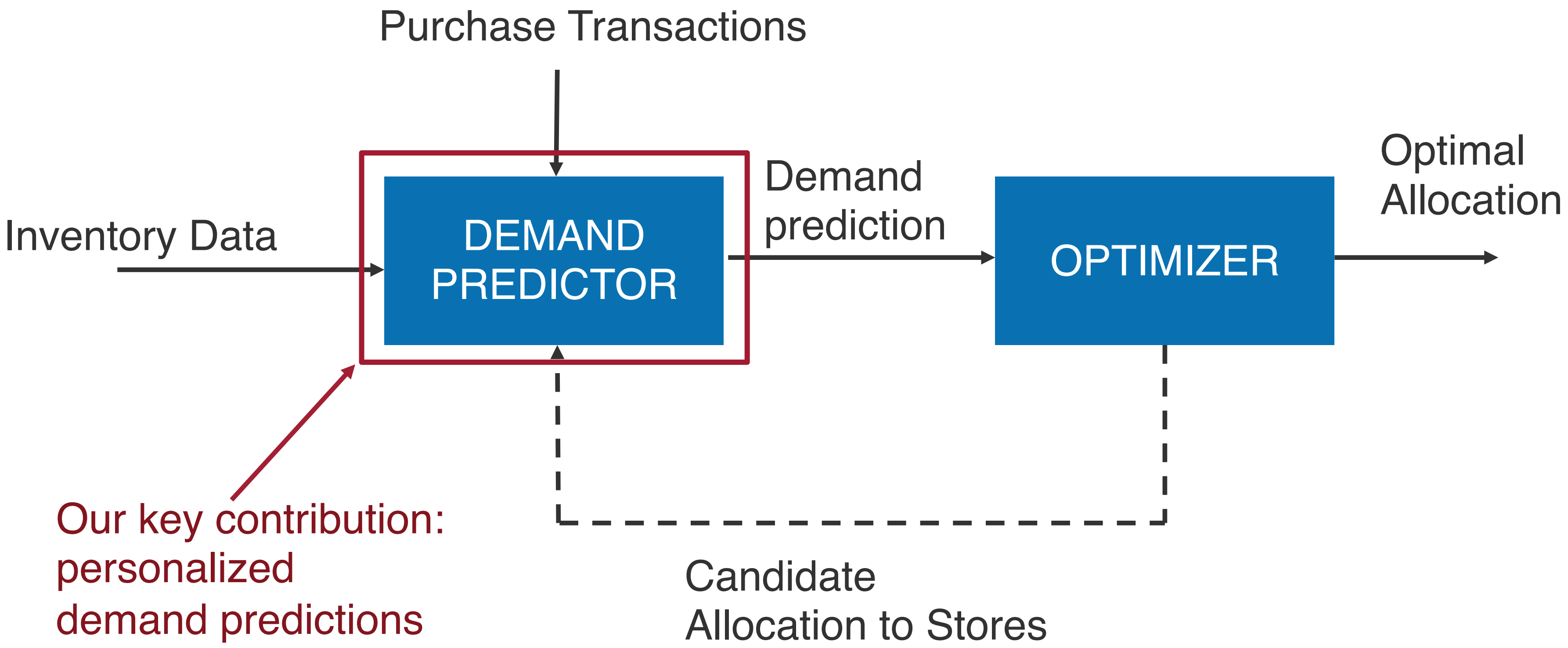
# A better way: automated decision support system



# A better way: automated decision support system



# A better way: automated decision support system



# Transaction data indicate product selection

transactions spreadsheet

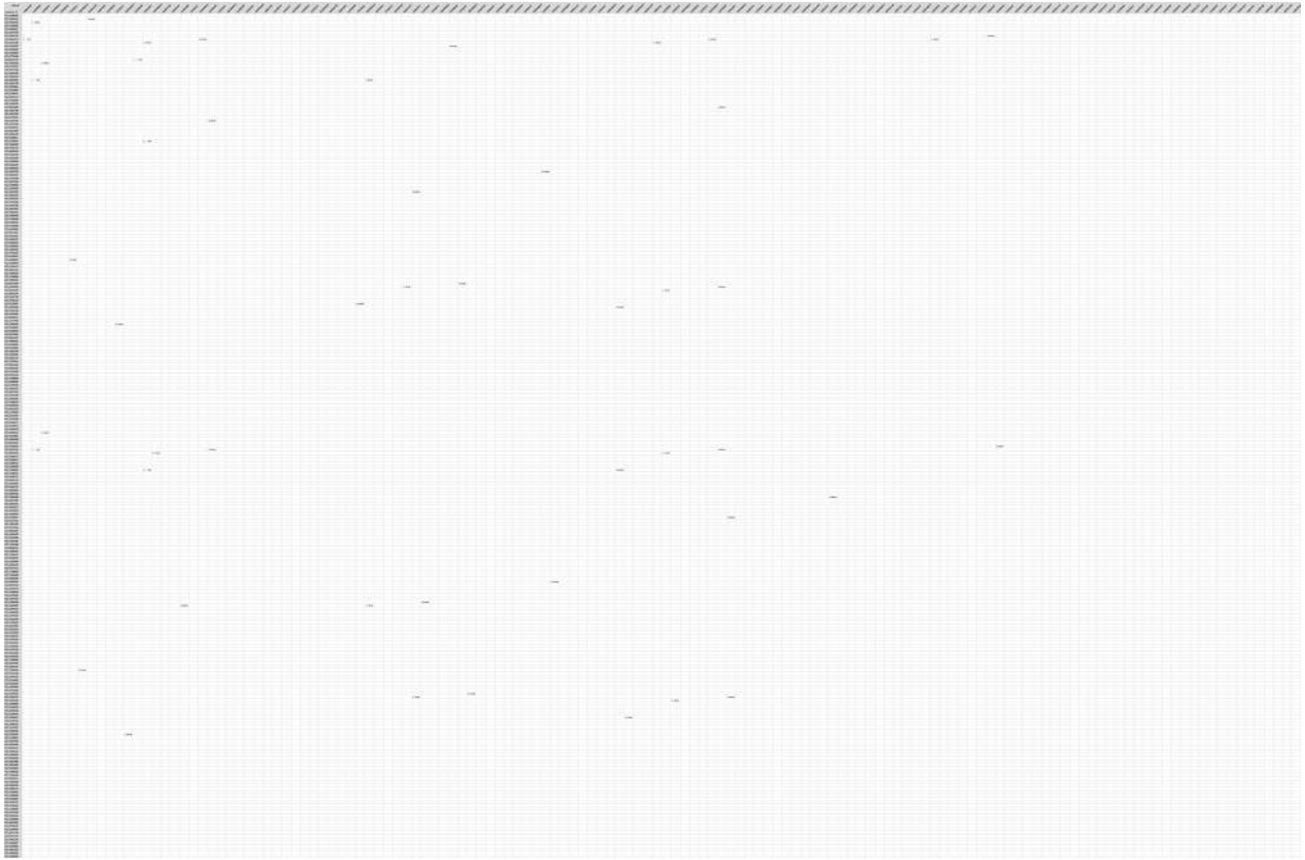
| SKU_ID<br>Customer ID | IAND4928 | KSOR8258 | KALE8350 | NKWN8322 | KSOR8259 | KALE8351 | NKWN8323 | HBSW1 |
|---|---|---|---|---|---|---|---|---|
| CID_7380399 | | | | | | $ 35.99 | | |
| CID_2247645 | | | | | | | | $ 170.00 |
| CID_9224163 | | $ 99.95 | | | | | | |
| CID_7735325 | | | | | | | | |
| CID_5267909 | | | | $ 5.99 | | | | |
| CID_4811544 | | | | | | | | |
| CID_6377106 | | | | $ 129.35 | | | | |
| CID_2917747 | $ 7.99 | | | | | | | |
| CID_4596598 | | | | | | | | |

[Image credit: Celect, Inc.]

# But data about each customer is sparse

Zoomed-out transactions spreadsheet

Lots of empty cells



[Image credit: Celect, Inc.]

# Build customer choice model that leverages "context"

Data provide not only what was chosen

but also what was not chosen

Fit a discrete choice model personalized to each customer

# Demand predictions: nonparametric choice model

# Demand predictions: nonparametric choice model

choose black bag

30% 30% 7% 7% 20% 6%

= 30% + 30% = 60%

# Demand predictions: nonparametric choice model


= 30% + 30% + 7% = 67%

# Demand predictions: nonparametric choice model

## Nonparametric preference-list model

No parametric assumptions needed

## parametric utility model

$\text{utility}_1 + \text{noise}_1$

$\text{utility}_2 + \text{noise}_2$

$\text{utility}_3 + \text{noise}_3$

Requires parametric assumptions

[MNL, Nested Logit, Mixed Logit, etc.]

# Demand predictions: nonparametric choice model

## Nonparametric preference-list model

No parametric assumptions needed

Data ‘picks’ the (simplest) model; more robust to over/under-fit

## parametric utility model

$\text{utility}_1 + \text{noise}_1$

$\text{utility}_2 + \text{noise}_2$

$\text{utility}_3 + \text{noise}_3$

Requires parametric assumptions

[MNL, Nested Logit, Mixed Logit, etc.]

Prone to over/under-fit issues

# Demand predictions: nonparametric choice model

## Nonparametric preference-list model

No parametric assumptions needed

Data ‘picks’ the (simplest) model; more robust to over/under-fit

Better predictive accuracy
20% improvement across all data-regimes on car sales data [Farias, Jagabathula, Shah ‘13]

## parametric utility model

$\text{utility}_1 + \text{noise}_1$

$\text{utility}_2 + \text{noise}_2$

$\text{utility}_3 + \text{noise}_3$

Requires parametric assumptions
[MNL, Nested Logit, Mixed Logit, etc.]

Prone to over/under-fit issues

Poor predictive accuracy

# Personalize the population-level choice model

60%

list 1

40%

list 2

# Personalize the population-level choice model

# Personalize the population-level choice model

$$\frac{w_1 \times 60\%}{w_1 \times 60\% + w_2 \times 40\%}$$

60%

list 1

$w_1$: similarity weight with list 1
$w_2$: similarity weight with list 2

$$\frac{w_2 \times 40\%}{w_1 \times 60\% + w_2 \times 40\%}$$

40%

list 2

# Personalize the population-level choice model

$$\frac{w_1 \times 60\%}{w_1 \times 60\% + w_2 \times 40\%}$$

$$w_1 > w_2$$

$$\frac{w_2 \times 40\%}{w_1 \times 60\% + w_2 \times 40\%}$$

60%

list 1

$w_1$: similarity weight with list 1
$w_2$: similarity weight with list 2

40%

list 2

# Technology newsworthiness…

- Academic recognition in machine learning, OM, and RM
  - Best student paper awards: NIPS 2008, MSOM 2010
  - INFORMS Revenue Management and Pricing Section Prize
  - 2016 Best OM paper in *Management Science*
- Patents filed through MIT and licensed by Celect, Inc.
  - US9201968 B2 (Dec 1, 2015) and US9251527 B2 (Feb 2, 2016)
- Broader press recognition


The New York Times
Business Day
Technology
WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS
Search Technology
Go
Inside Technology
Internet | Start-Ups | Business Computing | Companies

Is It Time to Retire the 5-Star Rating System?

By COLLEEN TAYLOR of GIGAOM
Published: July 14, 2011


Connect with us
cnet
Search CNET
Reviews | News | Video | How To | Deals
CNET › Tech Culture › MIT prof: Netflix has its recommendations wrong

MIT prof: Netflix has its recommendations wrong

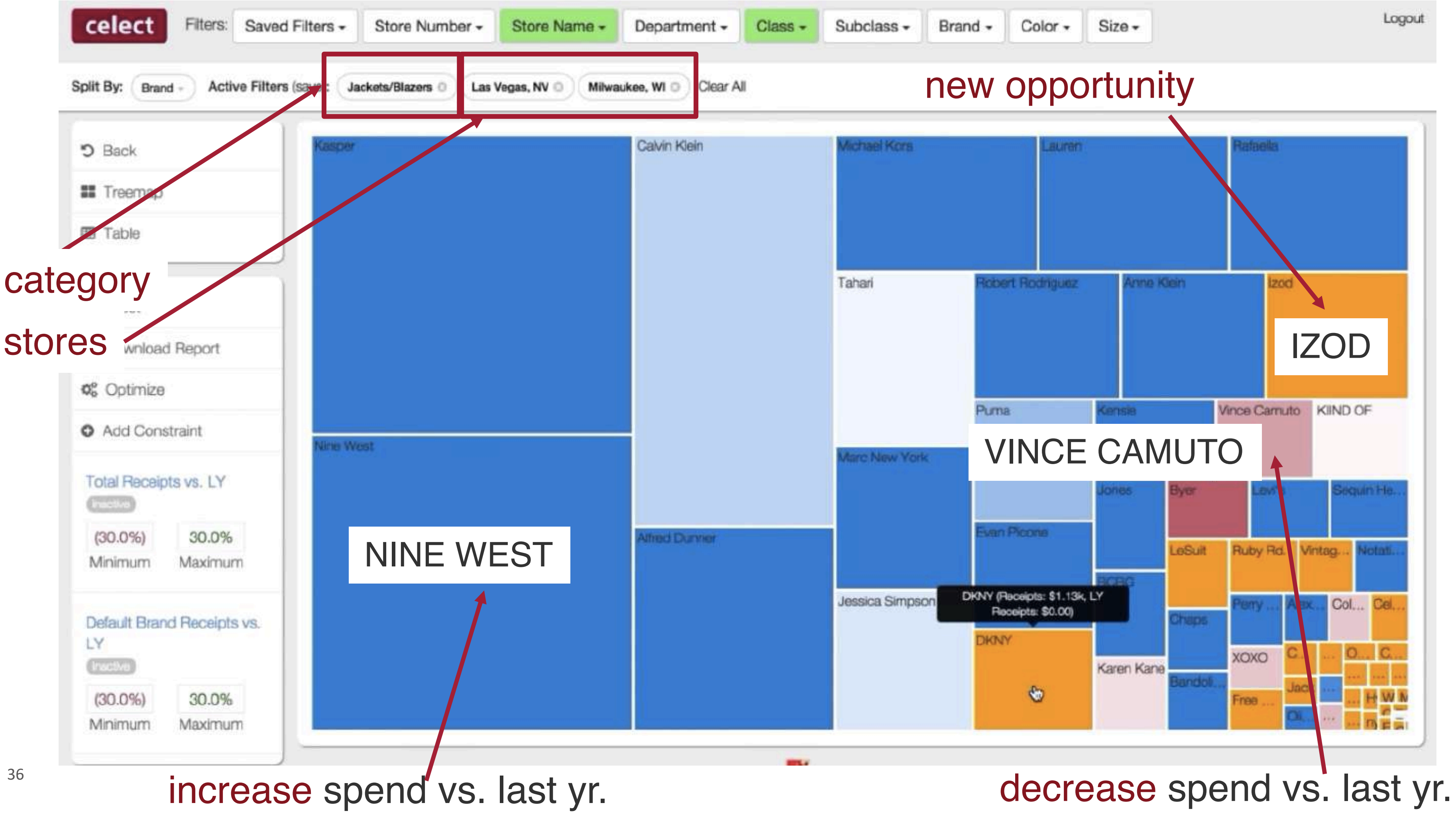
An MIT engineering professor says that absolute recommendation scales, such as star ratings, are not terribly helpful and that comparing two products is a far more reliable system.

# IMPLEMENTATION AND IMPACT

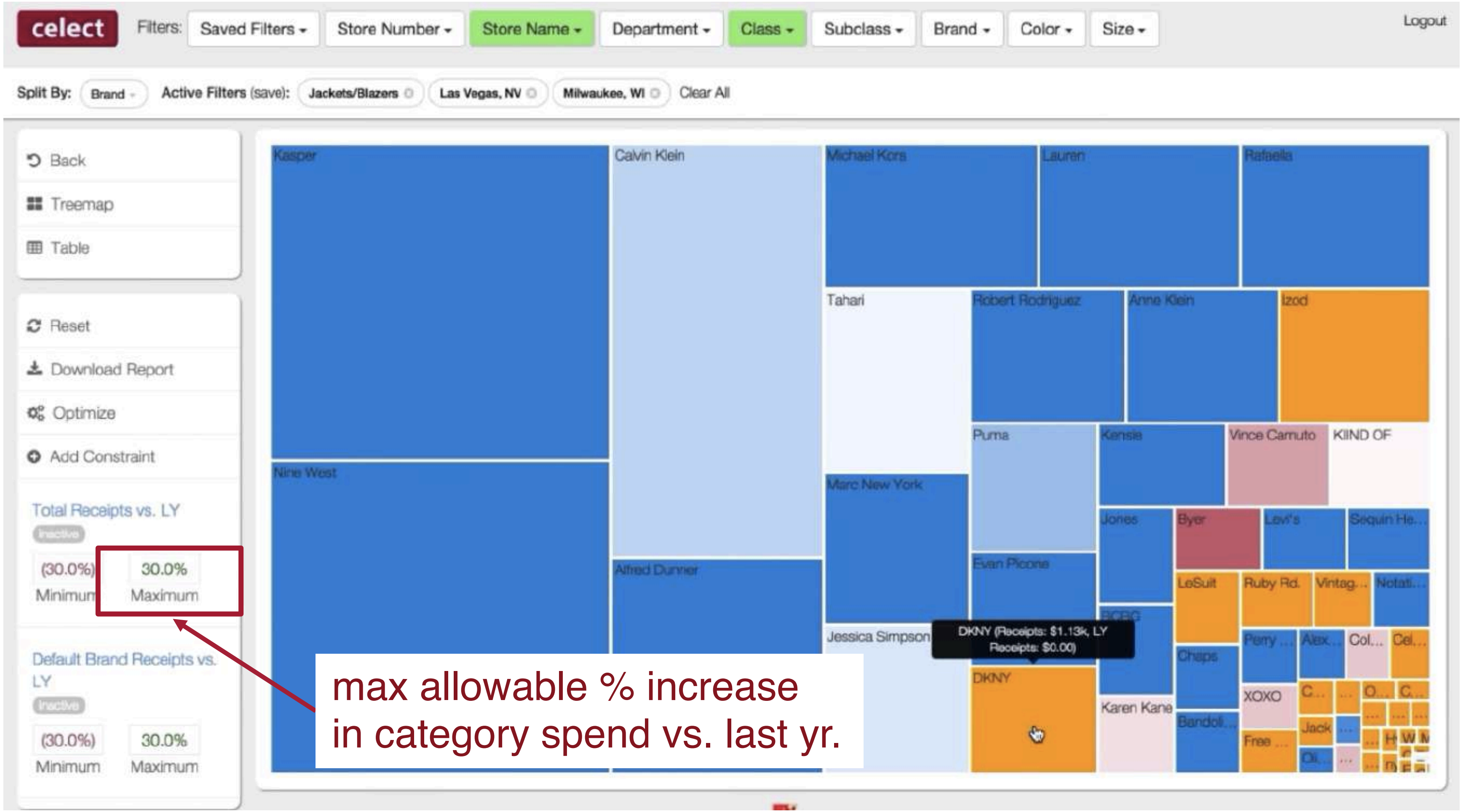
# Implementation challenges

- Big-data challenge
  - personalized predictions for millions of customers over a few million SKUs
  - contrast with Netflix Prize: 480K users over 17K movies
  - addressed by
    - designing graph-based estimation algorithms
    - distributed implementation on cloud
- Adoption challenge
  - easy-to-use GUI based decision support tool
  - gained trust for "counter-intuitive" recommendations by first providing
    - similar to merchant's own recommendations
    - different from merchant's recommendations but in line with their intuition

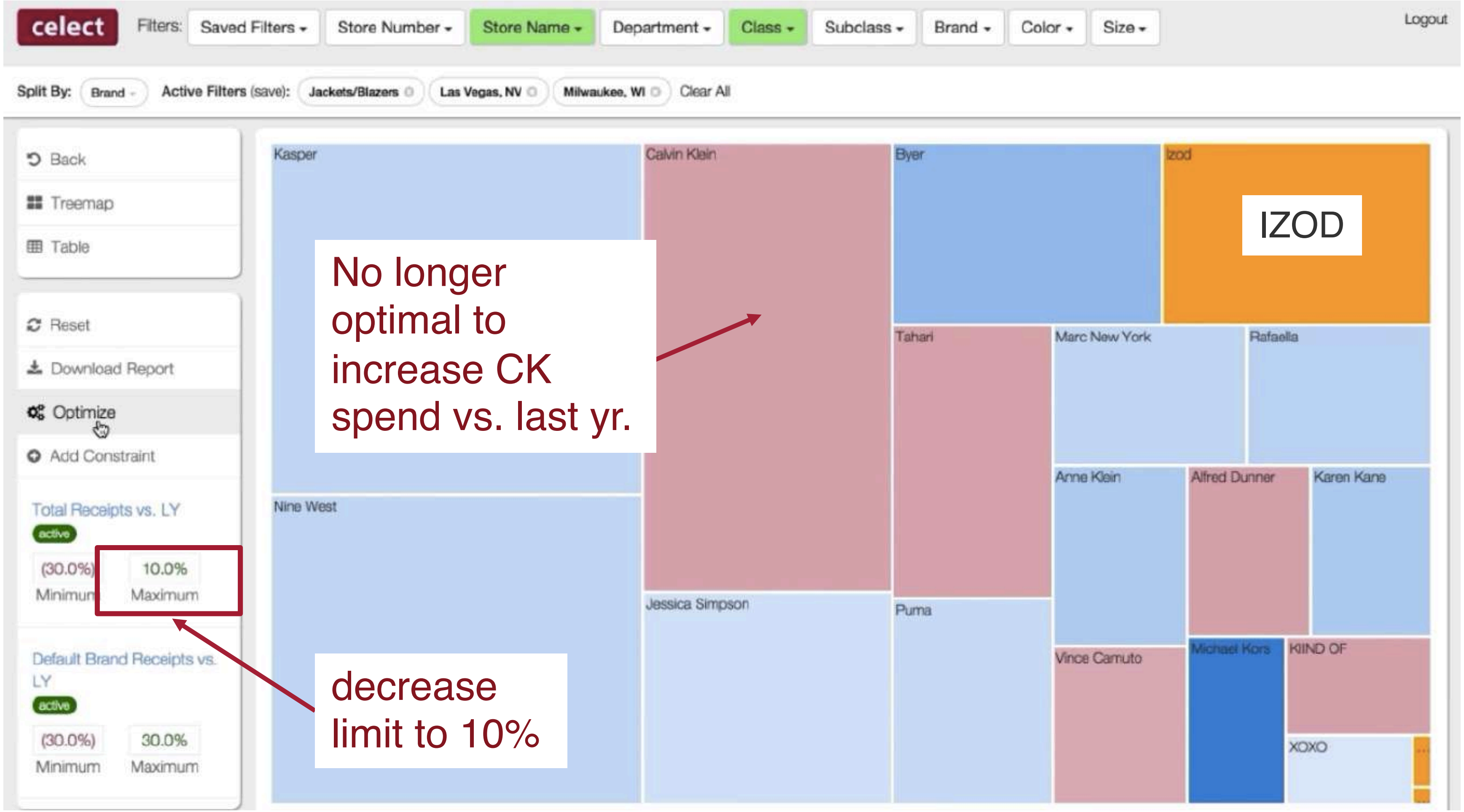
[Image credit: Celect, Inc.]



[Image credit: Celect, Inc.]



[Image credit: Celect, Inc.]



# 6.65% additional revenue growth with *same cost*

| | 10 test stores | 10 control stores | |
|---|---|---|---|
| Aug 03 – Sep 27 | 5.71% | -2.91% | test period |
| Jun 15 – Jul 26 | -0.54% | -2.51% | trailing 6 weeks |

test stores net relative
5.71 - -0.54 = 6.25%

control stores net relative
-2.91 - -2.51 = -0.40%

# 6.65% additional revenue growth with *same cost*

| | 10 test stores | 10 control stores | |
|---|---|---|---|
| Aug 03 – Sep 27 | 5.71% | -2.91% | test period |
| Jun 15 – Jul 26 | -0.54% | -2.51% | trailing 6 weeks |

adjusted net relative
6.25 - -0.40 = +6.65%

# Financial impact summary

**7%**

Increased revenue-growth at *same inventory* investment

We predicted 6.45%; so improvement predictable

**20%**

Potential gross profit impact (revenue impact/gross-margin)

**3%**

Spillover effects to other departments due to Celect assorted departments

# Insights

## Our approach better at identifying growth opportunities

E.g., Growth opportunity for an “urban” brand of handbag

- traditional: no opportunity at store A because store A not urban
- our tech: opportunity because some “urban” customers at store A

# Insights

## Our approach better at identifying growth opportunities

E.g., Growth opportunity for an “urban” brand of handbag

- traditional: no opportunity at store A because store A not urban
- our tech: opportunity because some “urban” customers at store A
  - When implemented, our recommendation did very well, exceeding expectations

# Insights

## Our approach better at identifying growth opportunities

E.g., Growth opportunity for an “urban” brand of handbag

traditional: no opportunity at store A because store A not urban

our tech: opportunity because some “urban” customers at store A

## Identified “halo” effect of higher-end products

Store 1: 4% lift in high-end accompanied 35% lift in moderate
Store 2: 7% lift in high-end accompanied 45% lift in moderate

# Transportability: data-driven tech applicable to other retailers

# Key learnings/Conclusions

- Accurate demand predictions biggest challenge to assortment opt.
- Wealth of data makes it a fertile ground for “Big Data” tech.
- Increasing willingness among retailers to try new solutions

Ideal for further academic inquiry from marketers, statisticians, and the machine learning community to obtain practical solutions