Data Mining for Business Analytics
MSBA 511

Introduction

What is Data Mining?

Data mining is the process of analyzing large datasets to discover patterns and insights, enabling data-driven decision-making.

Netflix uses data mining to analyze user behavior, such as viewing history and preferences, to deliver personalized recommendations, enhance satisfaction, and optimize content production.

Spotify uses data mining to analyze songs and user behavior, delivering personalized playlists, discovering new artists, and providing insights for the music industry.

Amazon uses sentiment analysis to evaluate customer feedback, identify product strengths and weaknesses, address issues, and highlight helpful reviews to enhance satisfaction and loyalty.

Walmart uses market basket analysis to identify item associations and relationships that are frequently bought together, create bundles, and enhance cross-sell and upsell strategies.

Why Mine Data?

Rapid Growth of Data

The volume of data generated globally is increasing exponentially due to digital transactions, social media, IoT devices, and cloud computing.
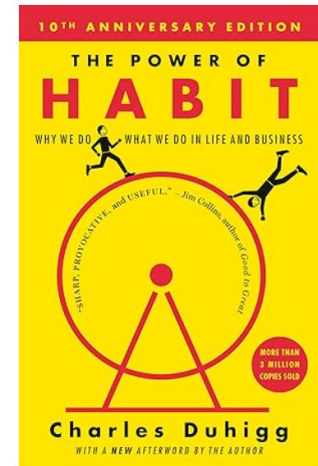
Technology Advancements

Modern data storage and processing technologies (cloud computing, big data frameworks) enable large-scale data analysis.

Competitive Advantage

Businesses that leverage data mining can uncover trends, optimize operations, and gain strategic advantages over competitors and create business value.

A Deep Dive in the Science of Habit

The [Power of Habit](#) by Charles Duhigg explores the science of habits in individuals, organizations, and societies.

In Chapter 7, Part I, the focus is on how companies use data mining to analyze consumer habits and behaviors to influence purchasing decisions.

How do you think retailers such as Target predict what customers want?

Duhigg, C. (2012). *The power of habit: Why we do what we do in life and business*. Random House.

How Target Knows What You Want Before You do

Every year, millions of Target shoppers' unknowingly share vast amounts of data about themselves by using their customer loyalty card, redeeming coupons, and using a credit card ultimately linking purchases to an individualized demographic profile.

Did you know:

- Half of all in-store sales are linked to a specific person.

- Almost all online sales.

- A quarter of online browsing.

Bringing it all together: Guest ID
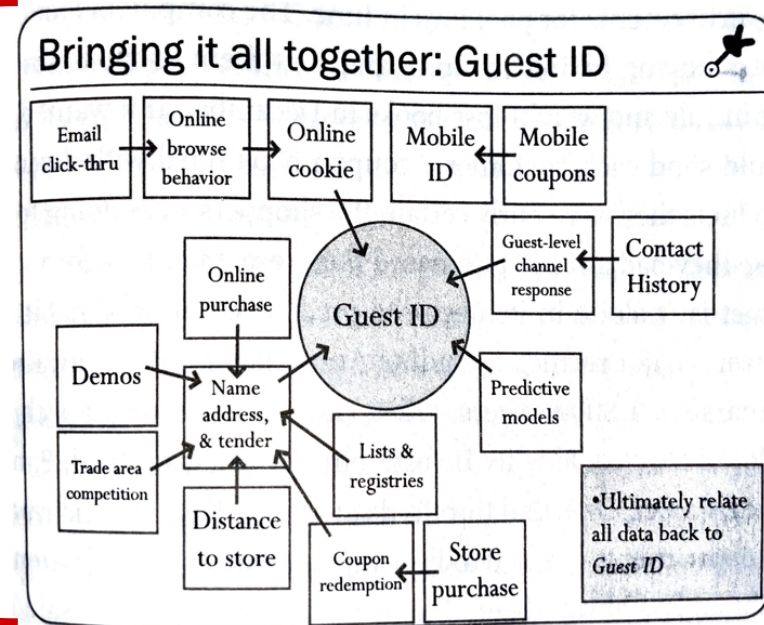
Pattern Deviation

Habits have a strong influence on purchase patterns and Target isn't alone in its desire to predict consumers' habits. Almost every major retailer has a predictive analytics departments devoted to figuring out these preferences.
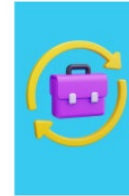
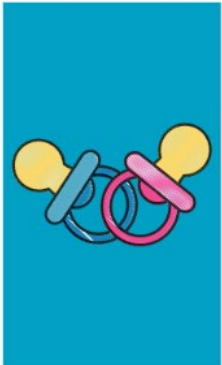**When and why do people deviate from their usual patterns?**

Buying a Home or Changing Where You Live

Getting Married or Divorced

Losing or Changing a Job

Having a Baby

What do all these examples have in common?

The Habit Loop

By using various data mining techniques on the mass amounts of data it collected, Target developed methods to identify major life events, especially pregnancy, that are key stages for influencing shopping behavior and was able to create highly targeted marketing campaigns.

Routine

Cue

Reward
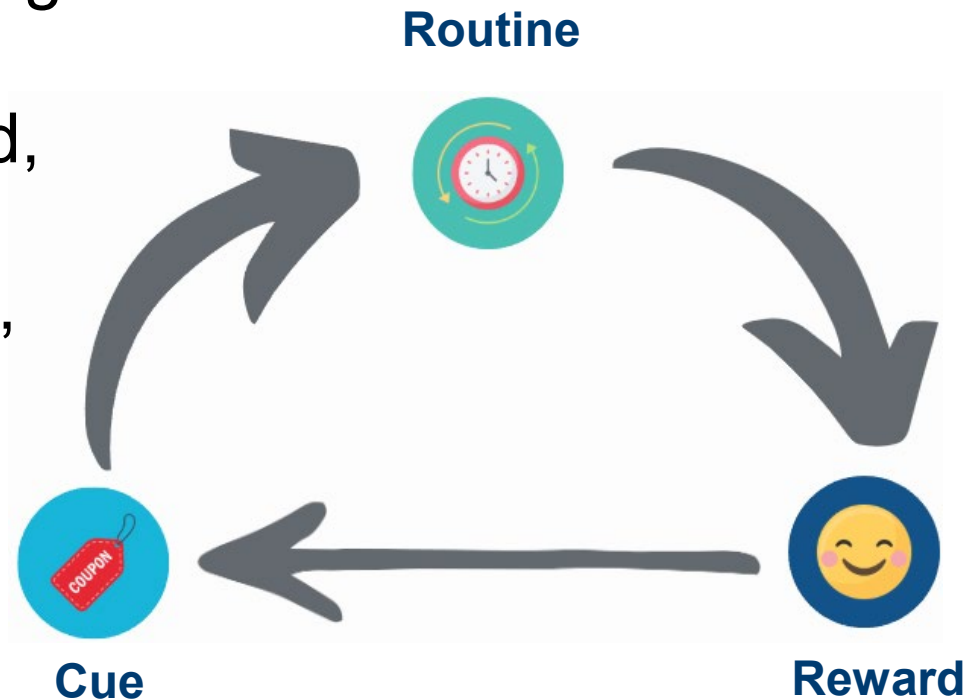
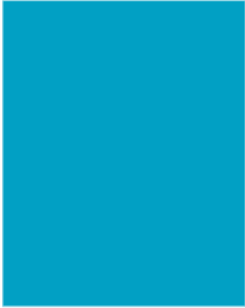# Did Target's Strategy Work?

Video reference: "*The Power of Habit: How Target Knows You Better Than You Do*" YouTube (April 2013). [Watch here.](#)
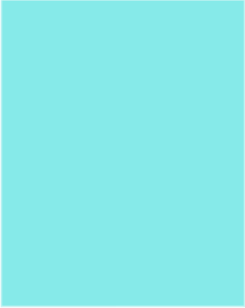
Insights vs. Discretion

How much should companies use consumer data without explicit consent?

How do you feel about companies using data mining and predictive analytics to influence your purchases?

Should companies be more transparent about the data they collect and how they use it?

Example 1 – Target Marketing

Amazon uses data mining techniques to identify and target customers with a coupon to incentivize a return visit and establish a new shopping routine.

Does it make sense for Amazon to provide a $15 off coupon to every customer that shops at an Amazon Fresh store?
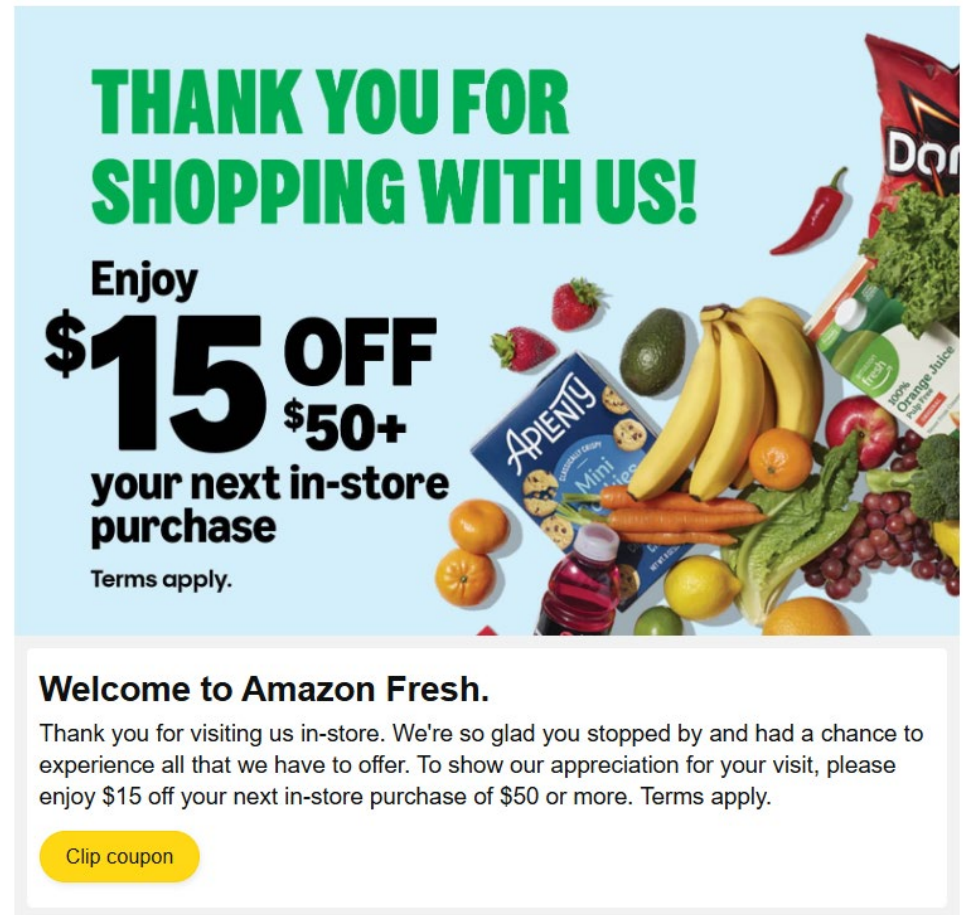
Techniques:

➤ **Classification**

➤ **Clustering**

Cue

amazon

THANK YOU FOR SHOPPING WITH US!

Enjoy

$15 OFF $50+

your next in-store purchase

Terms apply.

Welcome to Amazon Fresh.

Thank you for visiting us in-store. We're so glad you stopped by and had a chance to experience all that we have to offer. To show our appreciation for your visit, please enjoy $15 off your next in-store purchase of $50 or more. Terms apply.
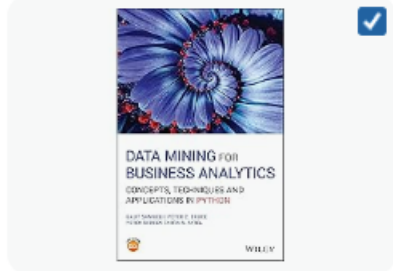
Clip coupon

Example 2 – What goes with what

Purchase recommendations for items that frequently go together before and after adding to the cart.

Techniques:

➢ **Association Rules aka Market Basket Analysis**

Gillette Fusion5 Men's Razor Blade Refills - 4ct

Edit delivery method in cart

Available deals

✓ **Target Circle Deal: $5 Target GiftCard with 4 select hair & personal care items**
Auto-applied
Expires Jan 25

Continue shopping

View cart & check out

Frequently bought together

Frequently also added

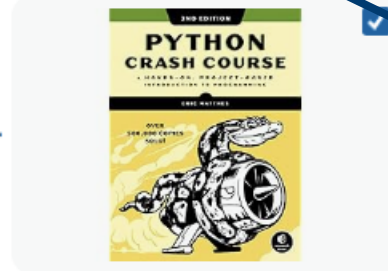**This item:** Data Mining for Business Analytics: Concepts, Techniques and Applications in...
$98³⁸

Python Crash Course, 3rd Edition: A Hands-On, Project-Based Introduction to Programming
$28⁹⁹

Python Crash Course, 2nd Edition: A Hands-On, Project-Based Introduction to Programming
$39⁹⁵

$23.99
ⓘ $5 Target GiftCard...
Gillette Fusion5 Value Pack Razor - Handle ..

$5.99
ⓘ $5 Target GiftCard...
Gillette PRO Men's Sensitive Shaving Ge...
Sponsored

Add to cart

Add to cart

Example 3 – Recommender Systems

In the prior example, the recommendation is at the product level. The recommendation is the same for everyone viewing the book Data Mining for Business Analytics.

In this example, HBO Max is making recommendations based on my viewing habits.

Techniques:

➤ **Collaborative Filtering**

Example 5 – Sports Analytics

On a basketball court, there's an area between the three-point line and the key known as the "mid-range." While it's not off-limits, you rarely see players taking shots from here anymore. This wasn't always the case!

Why do you think NBA teams have shifted away from the "mid-range" to the 3-Point shot?

Goodbye, Mid-Range Shot

Percentage of League Shot Attempts

So far, 35% of shot attempts this season were 3-pointers.

GSW wins the 2015 Finals.

Stephen Curry and Klay Thompson of GSW are first referred to as the "Splash Brothers."
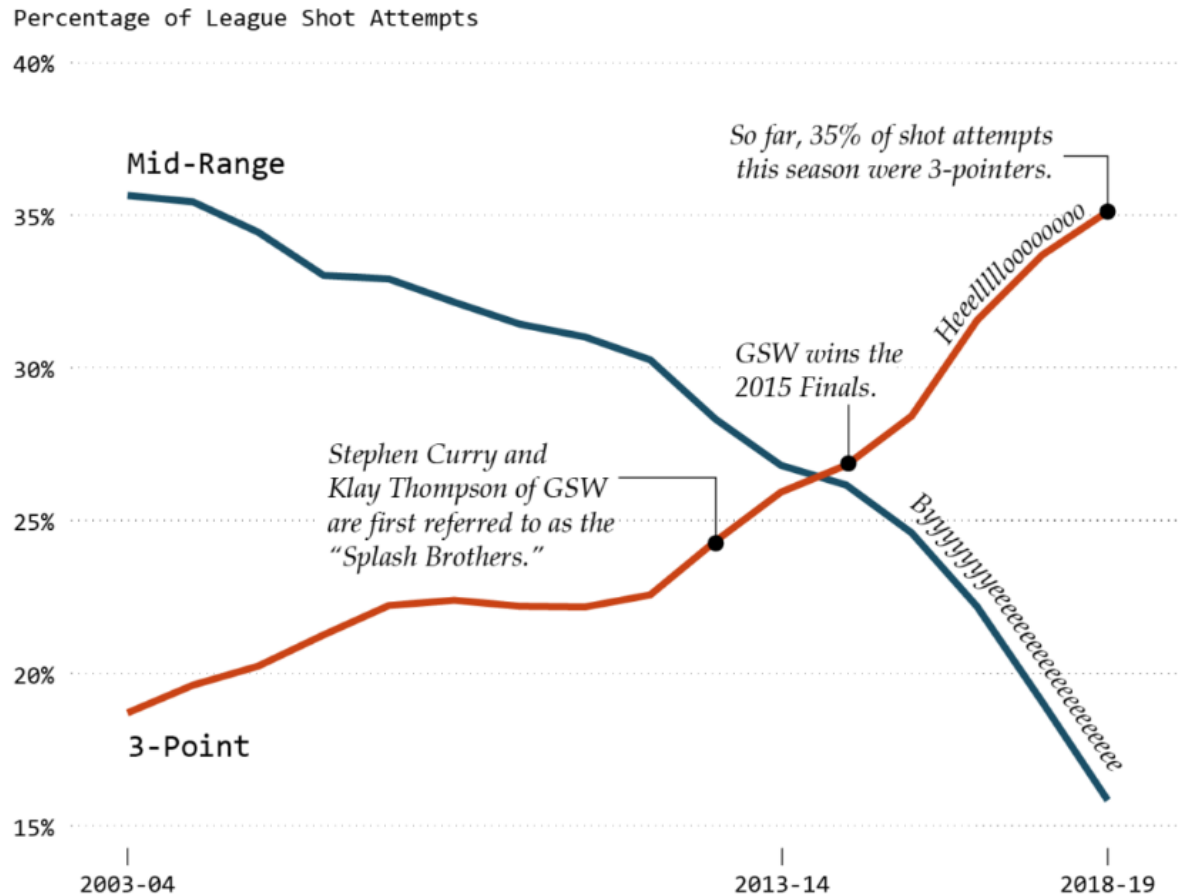
Source: NBA / By: FlowingData

Content and inspiration from Nathan Yau, "*Goodbye Mid-Range Shot*", FlowingData (January 15, 2019). Read here.

Data Driven Strategy

The game has changed, and teams have evolved their strategy based on data. The expected value of a 3-point shot is 1.07 points per attempt vs. only 0.81 points for mid-range shots.

2018-19 Regular Season

Average Points Per Shot

On average, 3-pointers produced 1.07 points.

.6 .8 1 1.2

Mid-range shots produced only 0.81 points, 24% less than 3-pointers.

Do you think this strategy makes sense for all teams regardless of what the data shows?
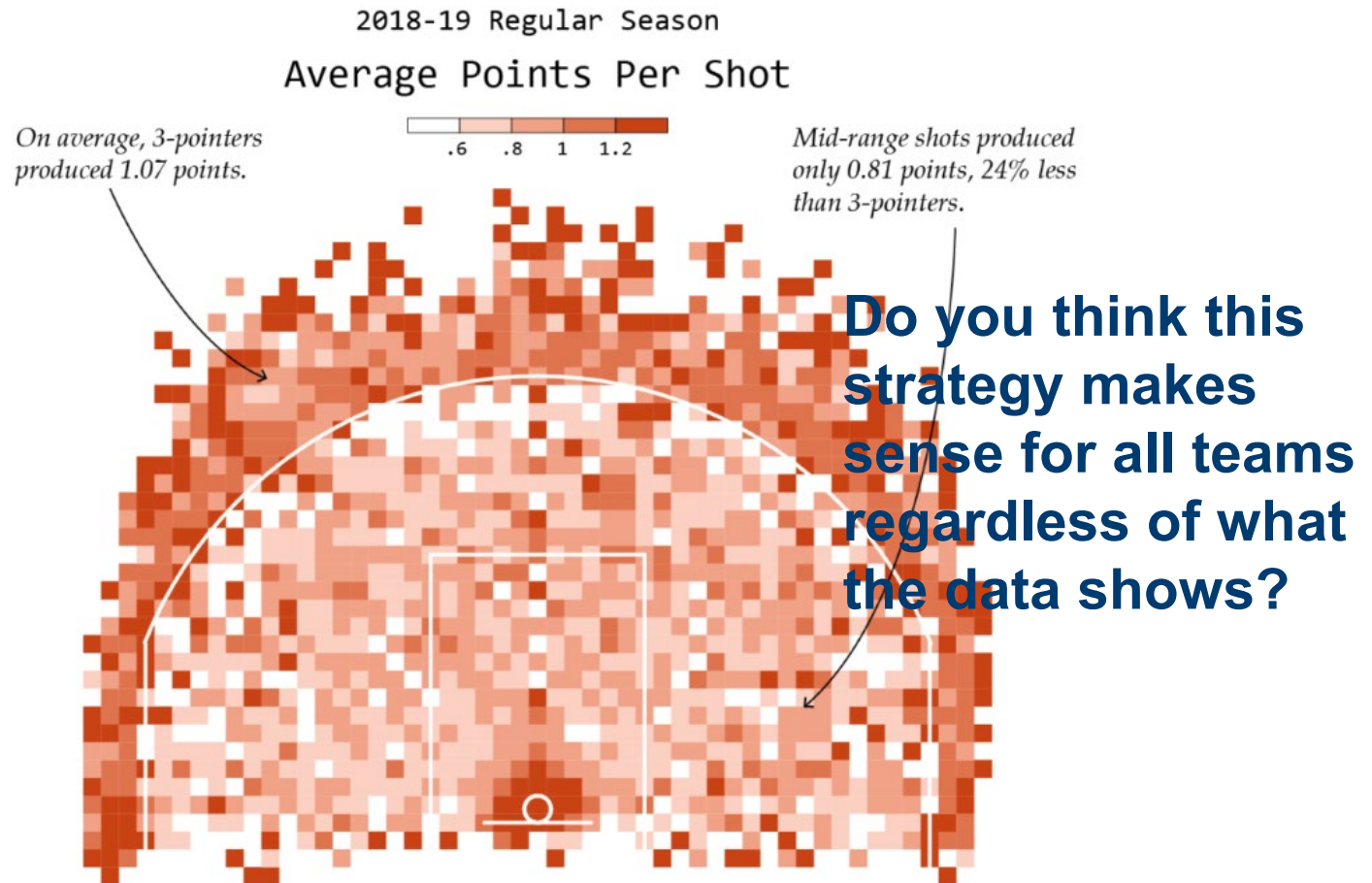
Core Ideas in Data Mining

Supervised Learning

Unsupervised Learning

Goal: Predict a single variable where the target value is **known**.

Goal: Segment data into meaningful groups; detect patterns where the target variable is **unknown**.

Methods:
- Classification
- Regression (Prediction)

Methods:
- Association Rules & Recommendation Systems
- Cluster Analysis
- Data & Dimension Reduction
- Data Exploration/Visualization

Steps in Data Mining

Obtain data (random sampling)

Specify task and choose technique

Assess and compare results

| Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 | Step 7 |

Define / understand business purpose

Explore, clean, pre-process, (reduce data)

Iterative implementation and tuning

Deploy solution