

Data Science and Technology Entrepreneurship

Course Recap

Week 14
Sameer Maskey

Announcements

- ▶ Friday Open Office hours + Co-work with teams
 - ▶ Come and work together
 - ▶ Ask questions
 - ▶ Dates :
 - ▶ April 19 - 4:30 - 6:30 pm
 - ▶ April 26 - 4:30 - 6:30 pm
 - ▶ **May 3 - 4:30 - 6:30 pm**
 - ▶ Location
 - ▶ CS conference room

Announcements

- ▶ **Final Pitch/Demo Guidelines**
 - ▶ 4 min long
 - ▶ Judges will provide feedback in judging time
 - ▶ 12 concurrent demos
 - ▶ 2 laptops
 - ▶ 1 for slides
 - ▶ 1 for demo

Announcements

- ▶ **Final Pitch/Demo Guidelines**
 - ▶ Executive Summary
 - ▶ 1-3 page executive summary
 - ▶ Make 15+ copies that judges can take home

Announcements

▶ Executive Summary

- ▶ Please write a 1-3 page Executive Summary. You can use parts of your Final Report and Business Canvas for the executive summary. You can use either of the following template: <http://www.entrepreneur.com/formnet/form/653> OR <http://www.entrepreneur.com/formnet/form/459>

▶

Announcements

- ▶ Final DSTE Survey
 - ▶ 10 Questions
 - ▶ Provide feedback so that I can improve the course for Spring 2014
 - ▶ <https://www.surveymonkey.com/s/DSTE2013finalsurvey>

Announcements

- ▶ Assignment 4 is out
- ▶ Due May 5th Sunday @6pm
- ▶ 4 parts
 - ▶ Part I - Final Report
 - ▶ I have written down 16 points or so that should guide you on how to write the Final Report
 - ▶ (Important) – Please provide the final business model canvas.
 - ▶ Write a paragraph on individual role and contribution of each team member to the business.

Announcements

▶ Assignment 4

▶ Part II - Final Technical Report

- ▶ System Architecture Diagram
- ▶ Overall choices you made

▶ Part III - MVP

- ▶ Last iteration of your MVP
- ▶ Please write about the changes you have made from the last MVP

▶ Part IV - MVP Field Test

- ▶ Similar to last homeworks - customer validation of your MVP with at least 5 customers

Announcements

▶ Final Presentation

▶ Date : May 7th (Tuesday)

- ▶ Yes it's on **Tuesday**
- ▶ First day of reading week

▶ Location - Uris 142

▶ Schedule

- ▶ 12:45 - 1:00 pm - Registration
- ▶ 1:00 - 1:15 - Overview of the class and projects
- ▶ 1:15 - 1:30 - Mentors/Advisors Introductions
- ▶ 1:30 - 2:30 - Student Presentation (4 min each)
- ▶ 2:30 - 3:15 - Coffee and Sandwiches
- ▶ 3:15 - 4:15 - 12 concurrent Student Demos (judges score the demos)
- ▶ 4:15 - 4:30 - Winners announcement

Topics for Today

- ▶ Course Recap
- ▶ Survey
- ▶ Analyzing team effort

Team

- ▶ Imagine a company is willing to buy you for 100,000\$
- ▶ Split the money and give it to individual members
- ▶ Each person should send me what they believe is the right split

Customer/Market Risk vs Invention Risk



Web 2.0	Enterprise Software	Enterprise Hardware	Comm Hardware	Comm Software	Consumer Electronics	Game Software	Semicon	Electronic Design Automation	Cleantech	Med Dev / Health Care	Life Science Biotech
---------	---------------------	---------------------	---------------	---------------	----------------------	---------------	---------	------------------------------	-----------	-----------------------	----------------------

Source : Steve Blank

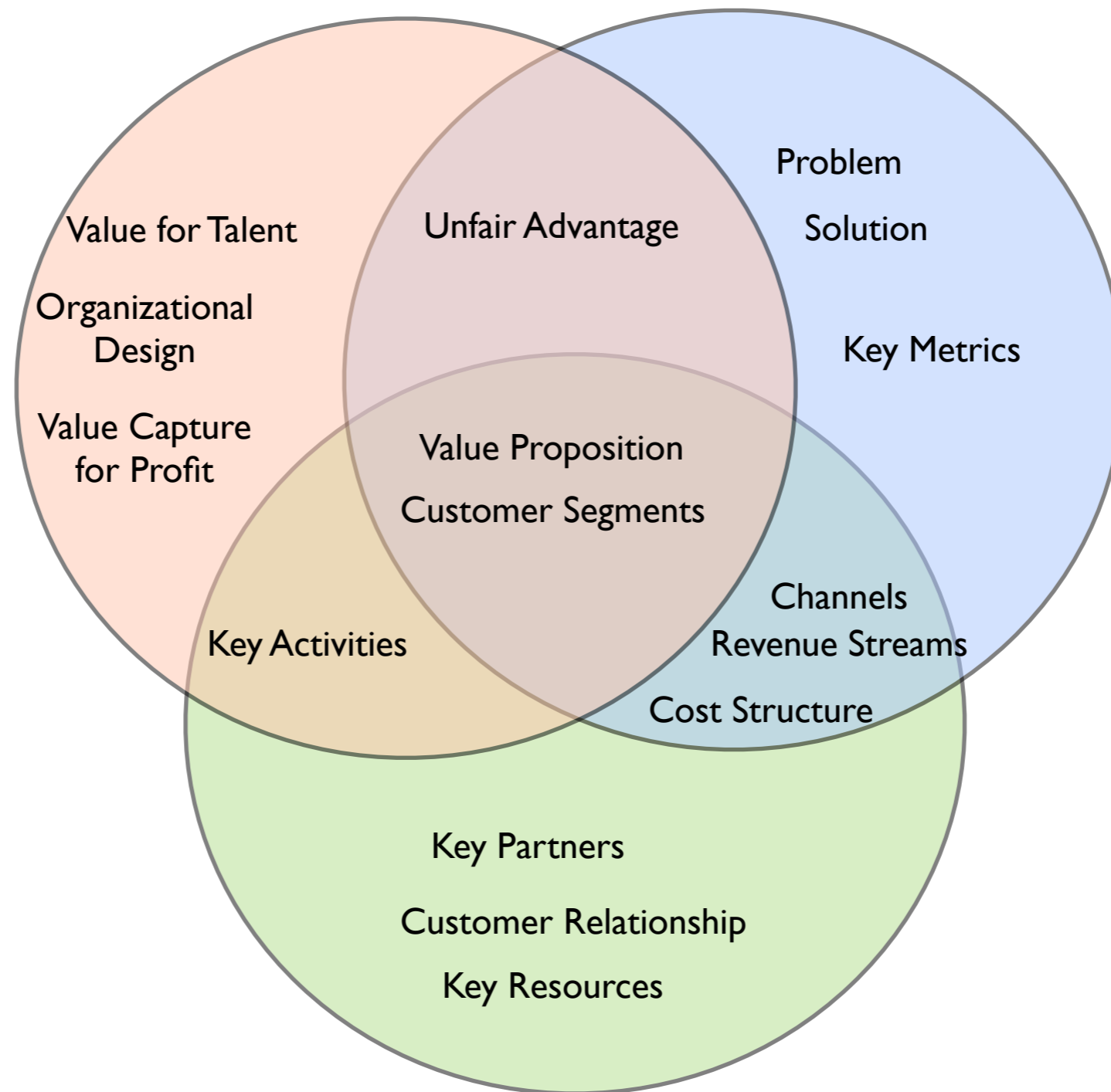
Where does your startup fall?

Lean Canvas [Maurya, A]

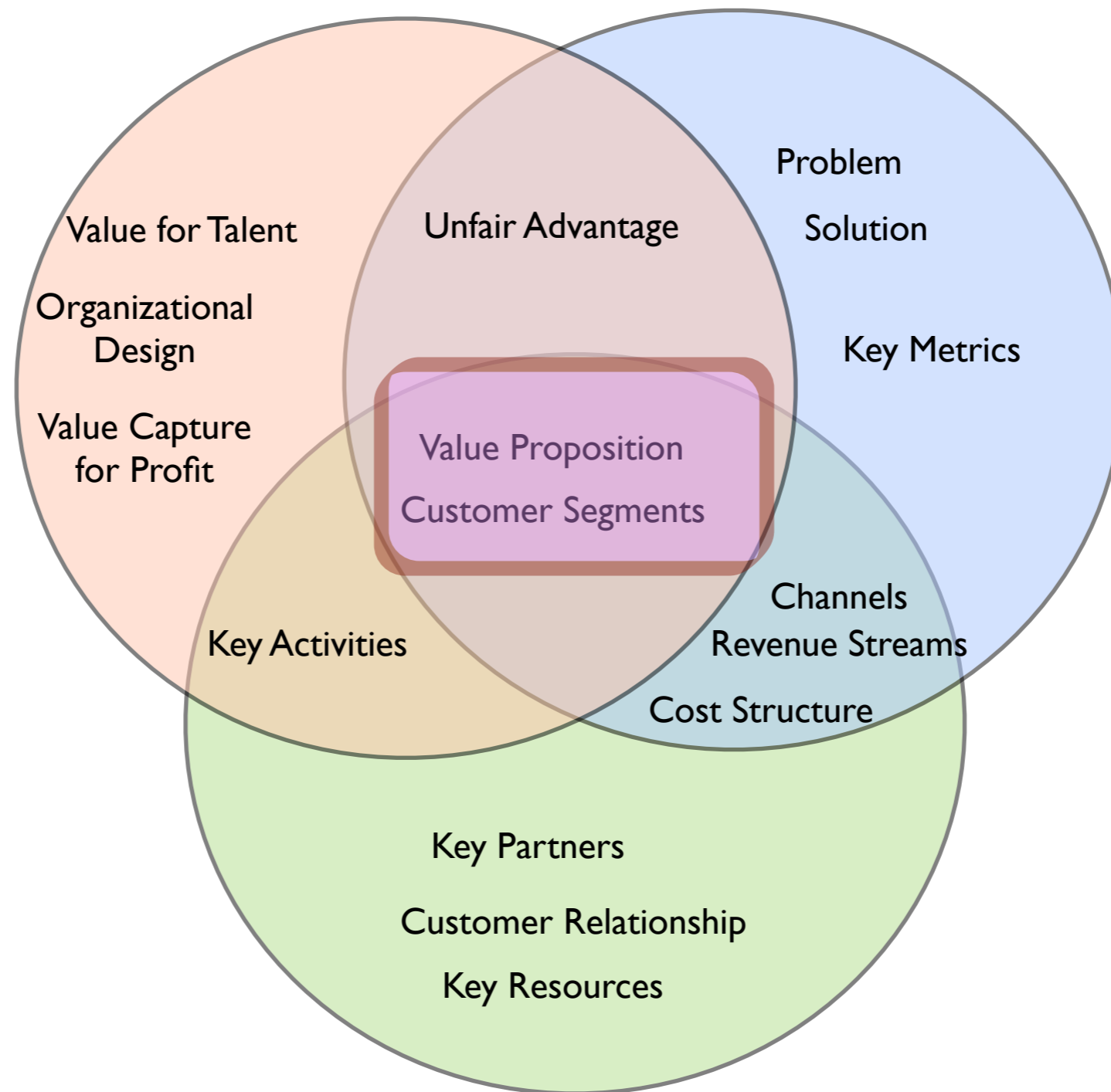
1 PROBLEM Top 3 problems	4 SOLUTION Top 3 features	3 UNIQUE VALUE PROPOSITION Single, clear, compelling message that states why you are different and worth buying	5 UNFAIR ADVANTAGE Can't be easily copied or bought	2 CUSTOMER SEGMENTS Target customers
	8 KEY METRICS Key activities you measure		9 CHANNELS Path to customers	
7 COST STRUCTURE Customer Acquisition Costs Distributing Costs Hosting People, etc.		6 REVENUE STREAMS Revenue Model Lifetime Value Revenue Gross Margin		

Lean Canvas is adapted from The Business Model Canvas (<http://www.businessmodelgeneration.com>) and is licensed under the Creative Commons Attribution-Share Alike 3.0 Un-ported License.

Common Concept Across Frameworks

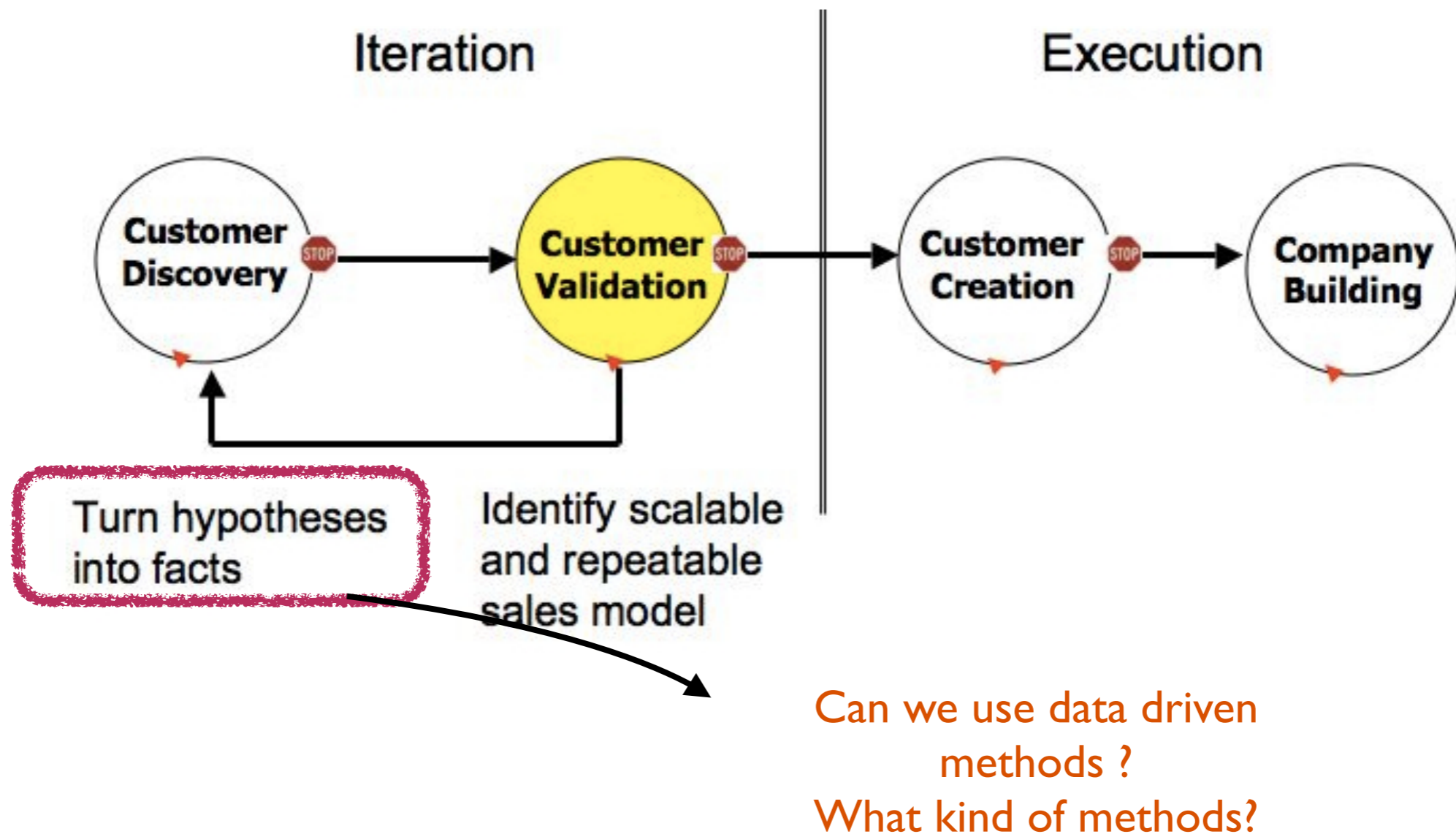


Common Concept Across Frameworks

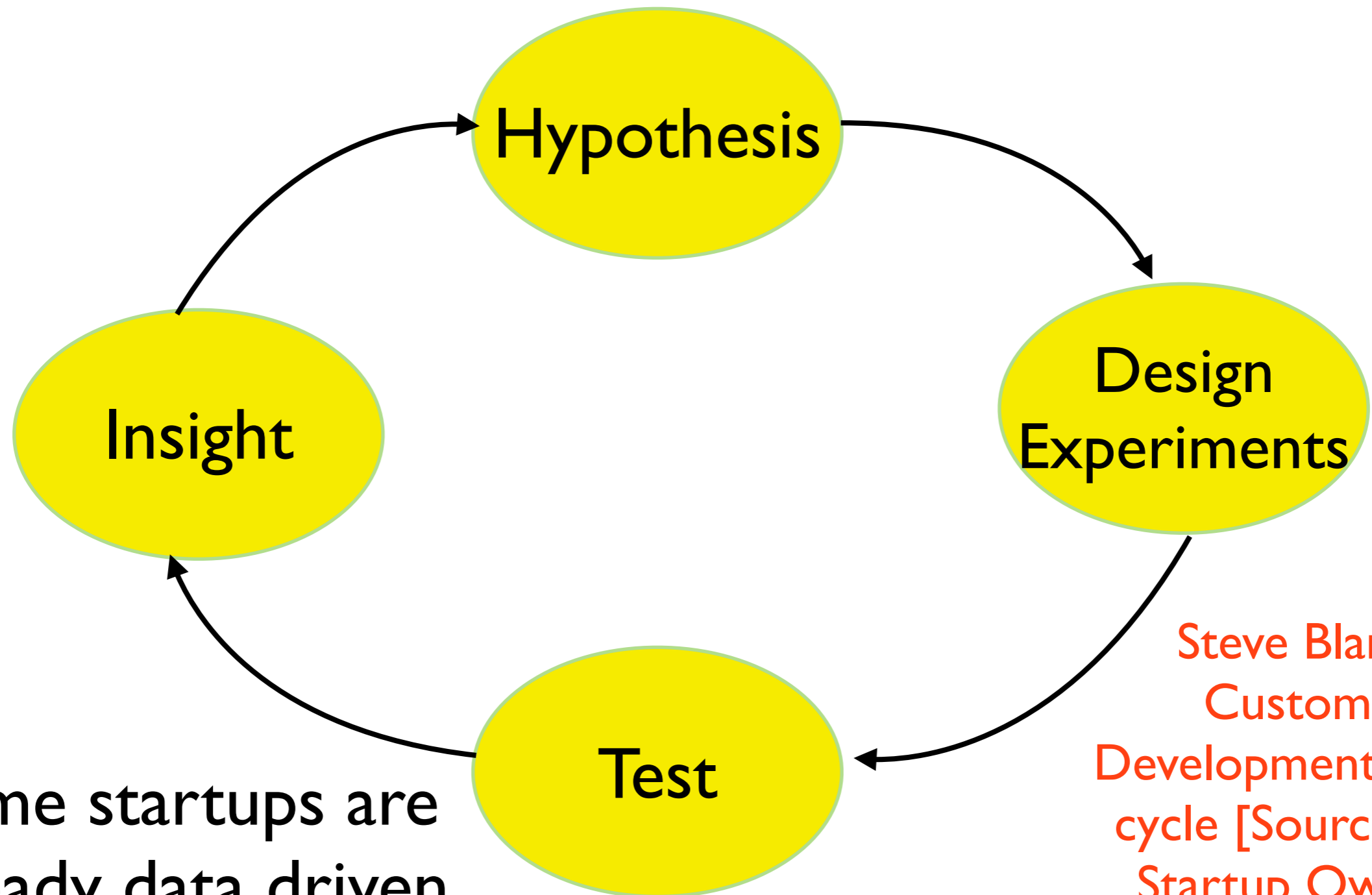


Customer Discovery Process

- Source : Startup Owner's Manual - Steve Blank and Bob Dorf



Data Driven Decision and Startups



Some startups are already data driven

Steve Blank's
Customer
Development Insight
cycle [Source :The
Startup Owner's
Manual]

Value Proposition & Customer Segments

- ▶ This is common theme across all business model frameworks
- ▶ 2 important questions to ask yourself
- ▶ Who are your customers?
- ▶ What is the value proposition for them?

Customer Discovery with MVP

Phase 1 : Set of Hypotheses about your business
(Problem?, Solution? Value Proposition?)

Phase 2 : Set of Hypotheses about your business
(Test your hypotheses by talking to customers)

Phase 3 : Build MVP and test MVP with customers
(Does your MVP solves the problem customer want?)

Phase 4 : Analyze results of your Phase 3
(Ready to signup paying customers?)

Multiple MVPs

- ▶ Multiple MVPs can be used to test competing hypotheses
- ▶ Example :
 - ▶ MVP with pay per use model
 - ▶ MVP with pay per month model
- ▶ If it is not difficult to build multiple MVPs then build them and test them with customers

MVP and Sales

**Value
Proposition**

**Customer
Validation**

Channels

**Revenue
Model**

Validation Process for Sales

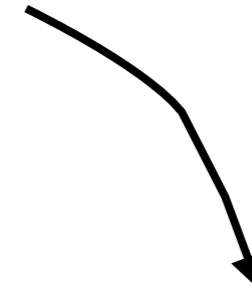
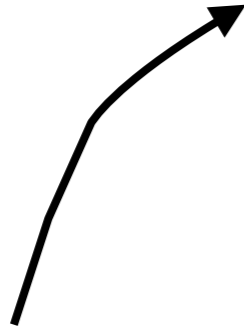
Phase I
Get ready to sell

Phase 2
**Get out of the building
and sell**

Phase 3
Develop Positioning

Phase 4
**Verify metrics that
matter**

Steve Blanks'
4 phases



Ask yourself

- ▶ Is the business scalable?
- ▶ Sales roadmap
 - ▶ Is it repeatable?
 - ▶ Is it scalable?
- ▶ Sales funnel predictable?

What are your validation checkpoints?

▶ B2B

- ▶ 3 meetings required for sale
- ▶ 10% responds to email
- ▶ 50% of respondent gives us meeting
- ▶ 50% of meeting result in contract signing meeting
- ▶ 50% of them use prototype
- ▶ 50% of them start paying
- ▶ Number of orders per month?
- ▶ Buys extra features

What are your validation checkpoints?

▶ B2C

- ▶ 40% response rate to marketing email
- ▶ 20% sign up
- ▶ 30% are repeat visitors?
- ▶ 20% of customer will pay for the service?
- ▶ 10% will pay for more features?

What kind of Data Science Methods can you Use?

- ▶ As you run experiments for validation of your hypotheses you start generating data points
- ▶ You also generate a lot of data points from user engagements, sales engagements, etc
- ▶ How can you use some these data points?

Data to Insights

- ▶ Data to Insight

- ▶ Data to Scores

- ▶ example: predict the likelihood of selling a watch

- ▶ Data to Classes

- ▶ example: predict what watches to buy from wholesaler

- ▶ Data to Clusters

- ▶ example: cluster customers based on their preferences

Machine Learning and Business

- ▶ Methods to analyze data that are all useful in decision making for businesses in general
- ▶ Data to Scores
- ▶ Data to Classes
 - ▶ Discriminative Methods
 - ▶ Generative Methods
- ▶ Data to Clusters

Use of Data Science Methods in Business

▶ LinkedIn

- ▶ People You May Knows
- ▶ Identified number of connections it takes for a long-term engagement

▶ Netflix

- ▶ Signup process
- ▶ Encourage to add movies to your queue
- ▶ Once you add certain number of movies likelihood of you being a long term customer goes up

Use of Data Science Methods in Business

▶ Zynga

- ▶ Monitors their users constantly
- ▶ Analyze how users interact with games to find out what makes a game successful

▶ Financial Services

- ▶ Fraud detection

▶ OkCupid

- ▶ Marketing Analytics with viral blogs
- ▶ Facial attitude and new contacts blog

Data Science and Technology Startups

Data Science and Technology Startups

- ▶ Technology Startups can generate a lot of data

Data Science and Technology Startups

- ▶ Technology Startups can generate a lot of data
- ▶ For example a web startup with 500K users can generate a lot of data every user action is stored
 - ▶ Visits
 - ▶ Click through rates
 - ▶ Search logs
 - ▶ User generated content
 - ▶ Time spent on individual pages
 - ▶ Mouse movement behavior
 - ▶ Many more individual data points

Data Science and Technology Startups

- ▶ Technology Startups can generate a lot of data
- ▶ For example a web startup with 500K users can generate a lot of data every user action is stored
 - ▶ Visits
 - ▶ Click through rates
 - ▶ Search logs
 - ▶ User generated content
 - ▶ Time spent on individual pages
 - ▶ Mouse movement behavior
 - ▶ Many more individual data points
- ▶ Mining this large set of data generated every day for identify various types of pattern about users could lead to increased engagement

Data Science and Technology Startups

- ▶ Technology Startups can generate a lot of data
- ▶ For example a web startup with 500K users can generate a lot of data every user action is stored
 - ▶ Visits
 - ▶ Click through rates
 - ▶ Search logs
 - ▶ User generated content
 - ▶ Time spent on individual pages
 - ▶ Mouse movement behavior
 - ▶ Many more individual data points
- ▶ Mining this large set of data generated every day for identify various types of pattern about users could lead to increased engagement



Data Science and Technology Startup

- ▶ Let's do another hypothetical example
- ▶ Using data science for a web startup that sells products online
- ▶ Want to increase the click through rate on related items?
- ▶ In other words, want to build a very simple minded recommendation engine

Example : User Data and Engagement

- ▶ Assume you are running a shopping site and you want to produce top 5 items to recommend like Amazon

Inspired by Your Shopping Trends



Bluetooth USB 2.0 Micro Adapter Dongle
Generic
★★★★☆ (817)
~~\$19.99~~ \$4.49



RF Wireless Laser Pointer with Page...
Generic
★★★★☆ (179)
\$6.87



Satechi SP400 Smart-Pointer 2.4Ghz RF...
★★★★☆ (359)
~~\$49.99~~ \$34.99



August LP103R Red Laser Presentation...
★★★★☆ (8)
~~\$16.99~~ \$9.95



Logitech Wireless Presenter R400
★★★★☆ (216)
~~\$49.99~~ \$36.63

▶ [View your shopping cart](#)

- ▶ Data you have

- ▶ Click through rates and data item

User Data History

► Data from User's history

Product Clicks




 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches

4

5

2

User Data History




Product	Clicks
 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	4
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	5
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	2

- ▶ Can you use this data to build a simple model that can predict the number of clicks for a new product?
- ▶ Imagine you have such data for millions of users

Process Raw Data

Product

Clicks




 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	, 4
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	, 5
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	, 2

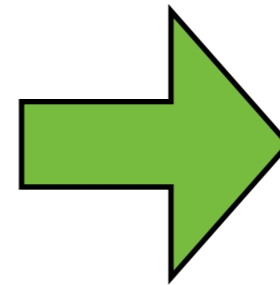
Raw Data

Process Raw Data

Product

Clicks

 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	4
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	5
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	2






Raw Data

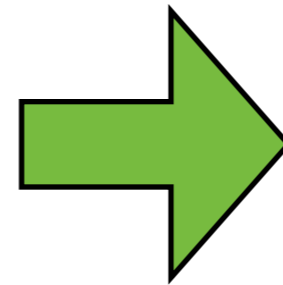
Process Raw Data

Product

Clicks

 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	4
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	5
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	2

Process






Raw Data

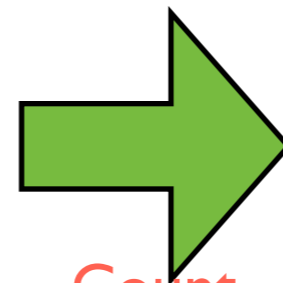
Process Raw Data

Product

Clicks

 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	4
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	5
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	2

Process




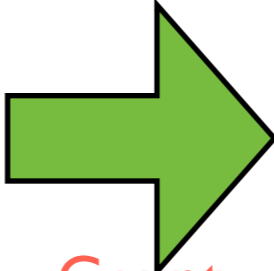


Count
number of word
"Watch" in
Product Description

Raw Data

Process Raw Data

Product

Clicks




 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	4	<p>Process</p>  <p>Count number of word "Watch" in Product Description</p>	
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	5		
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	2		

Raw Data

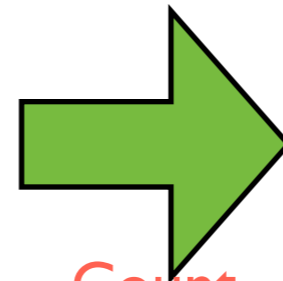
Process Raw Data

Product

Clicks

 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	4	
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	5	
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	2	

Process



Count
number of word




“Watch” in
Product Description

Raw Data

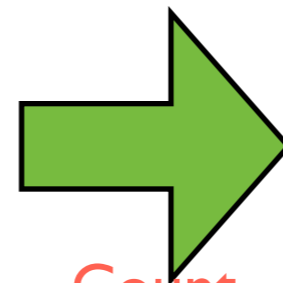
Process Raw Data

Product

Clicks

 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	4	
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	5	
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	2	

Process



Count
number of word
"Watch" in
Product Description

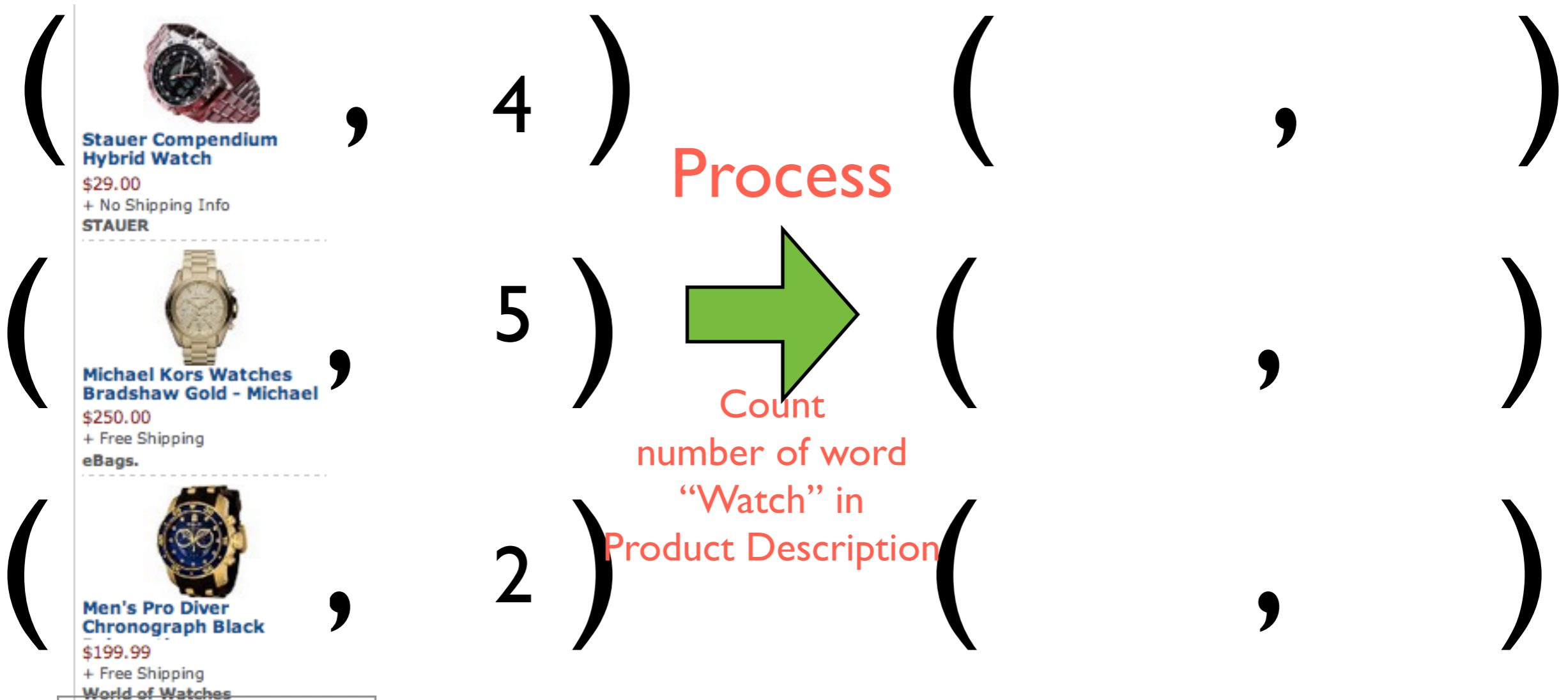
Raw Data

Process Raw Data

Product

Clicks

Product



Raw Data

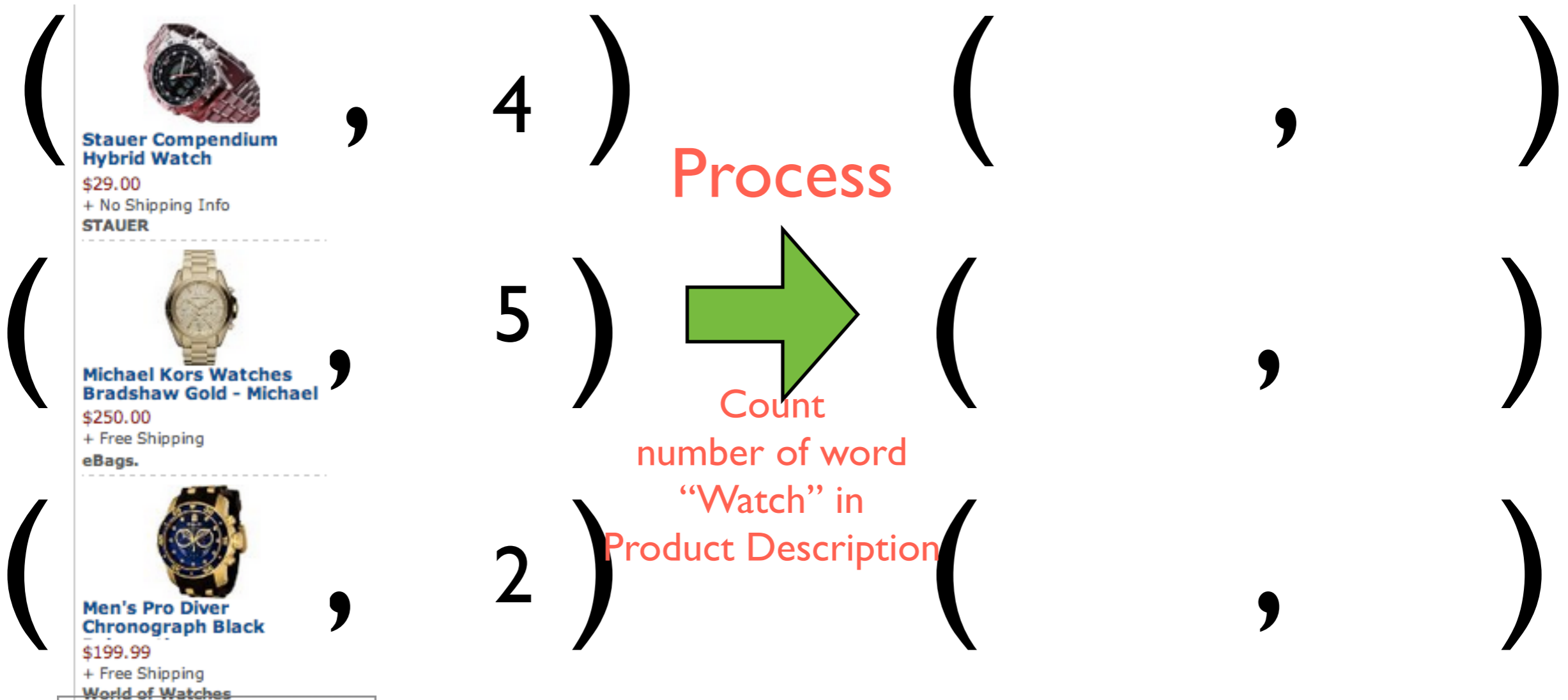
Process Raw Data

Product

Clicks

Product

Clicks



Raw Data

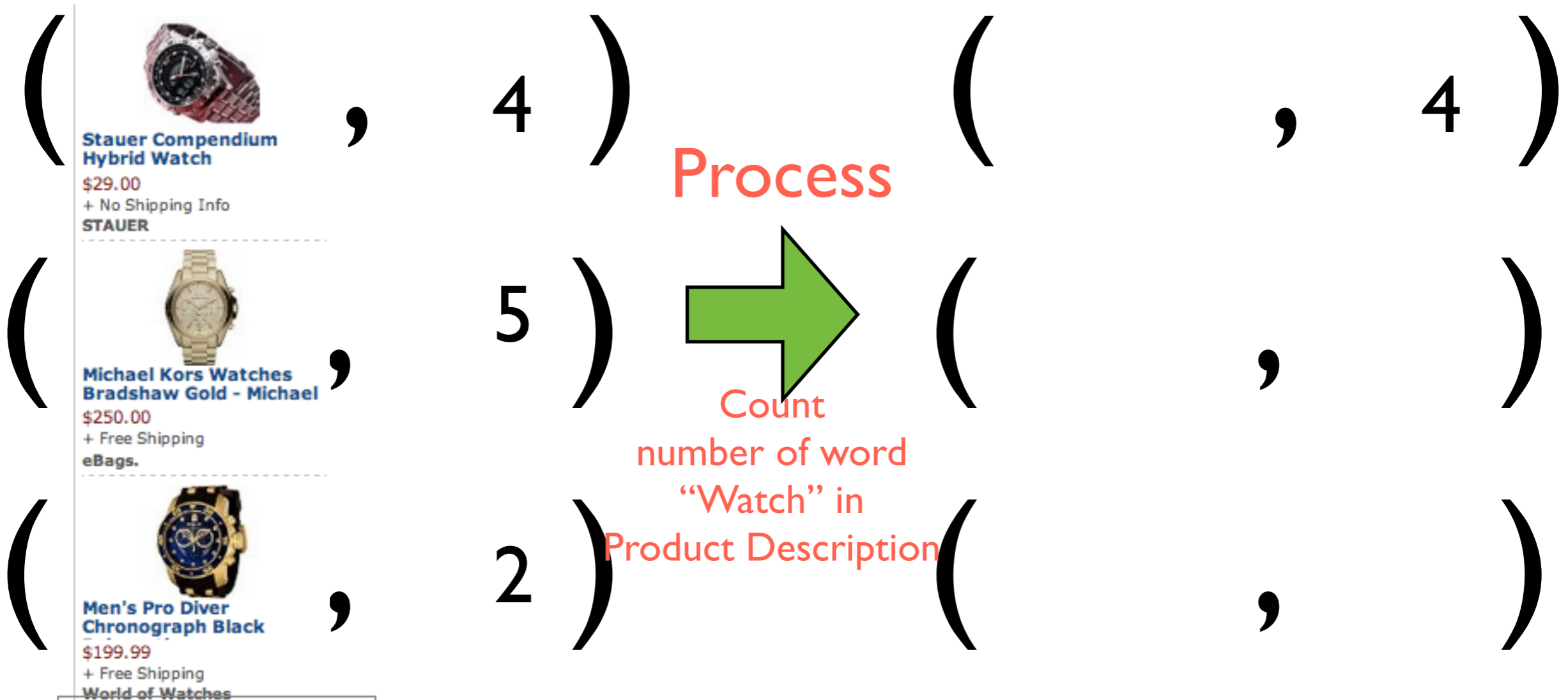
Process Raw Data

Product

Clicks

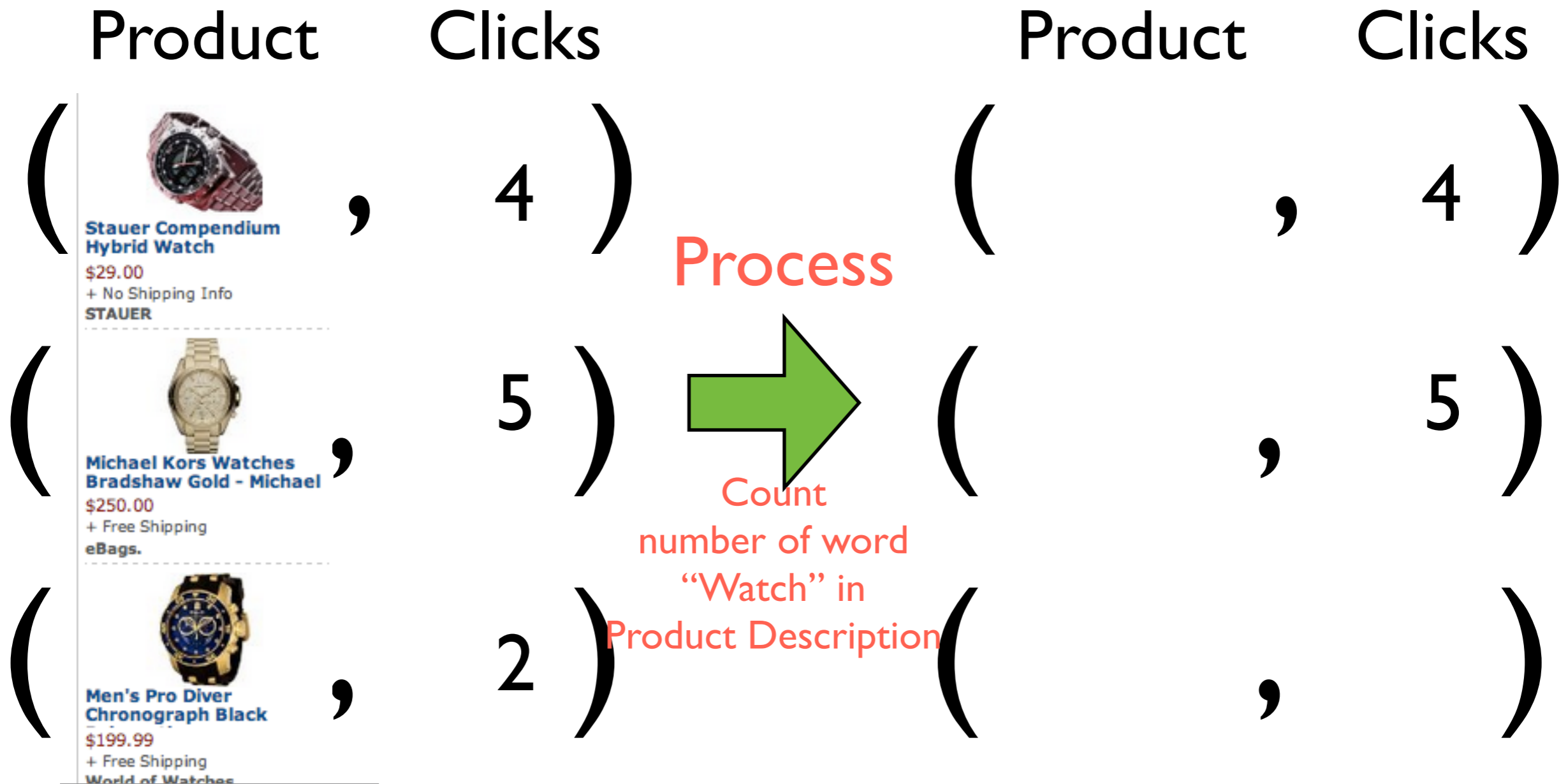
Product

Clicks



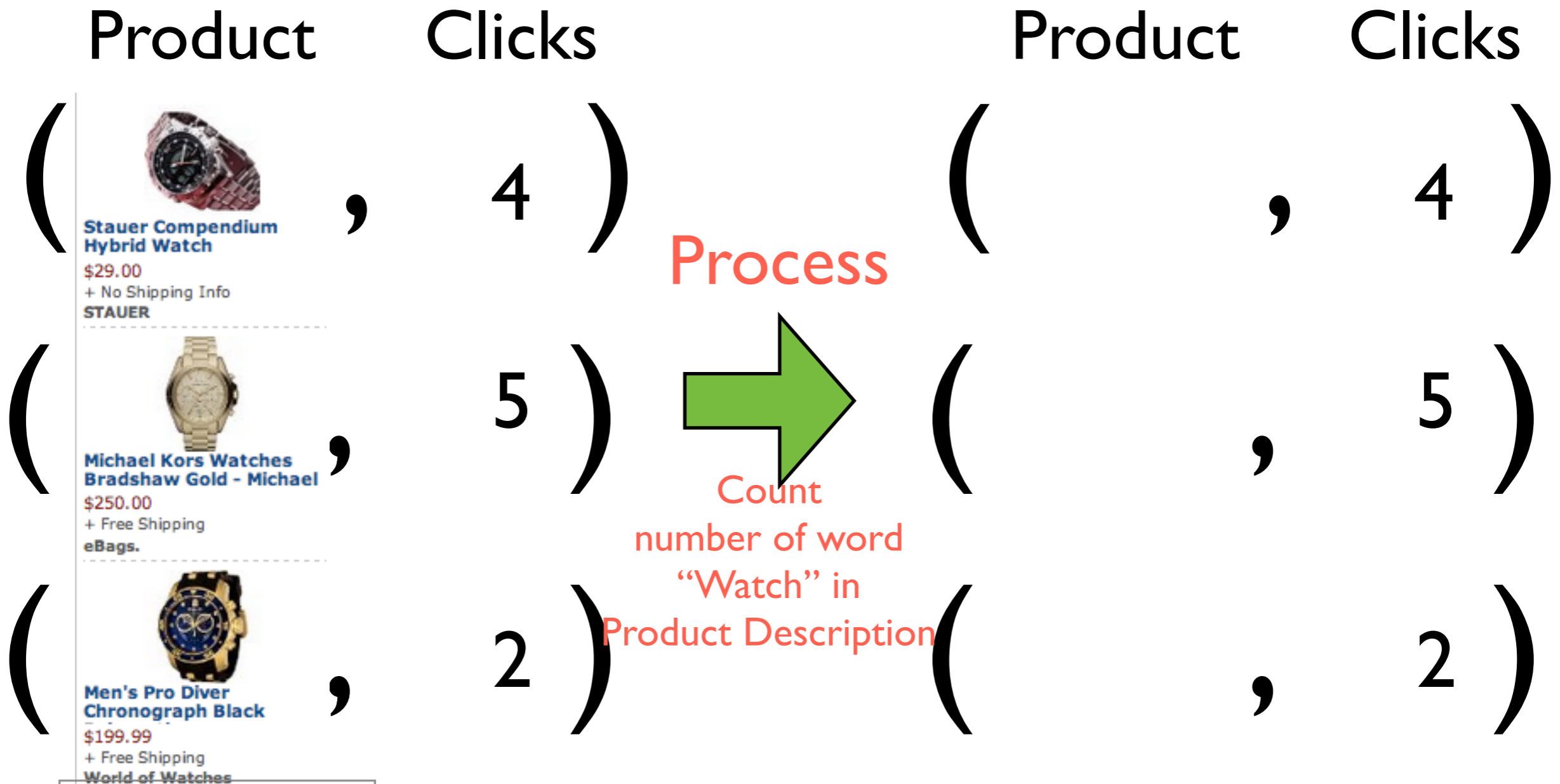
Raw Data

Process Raw Data



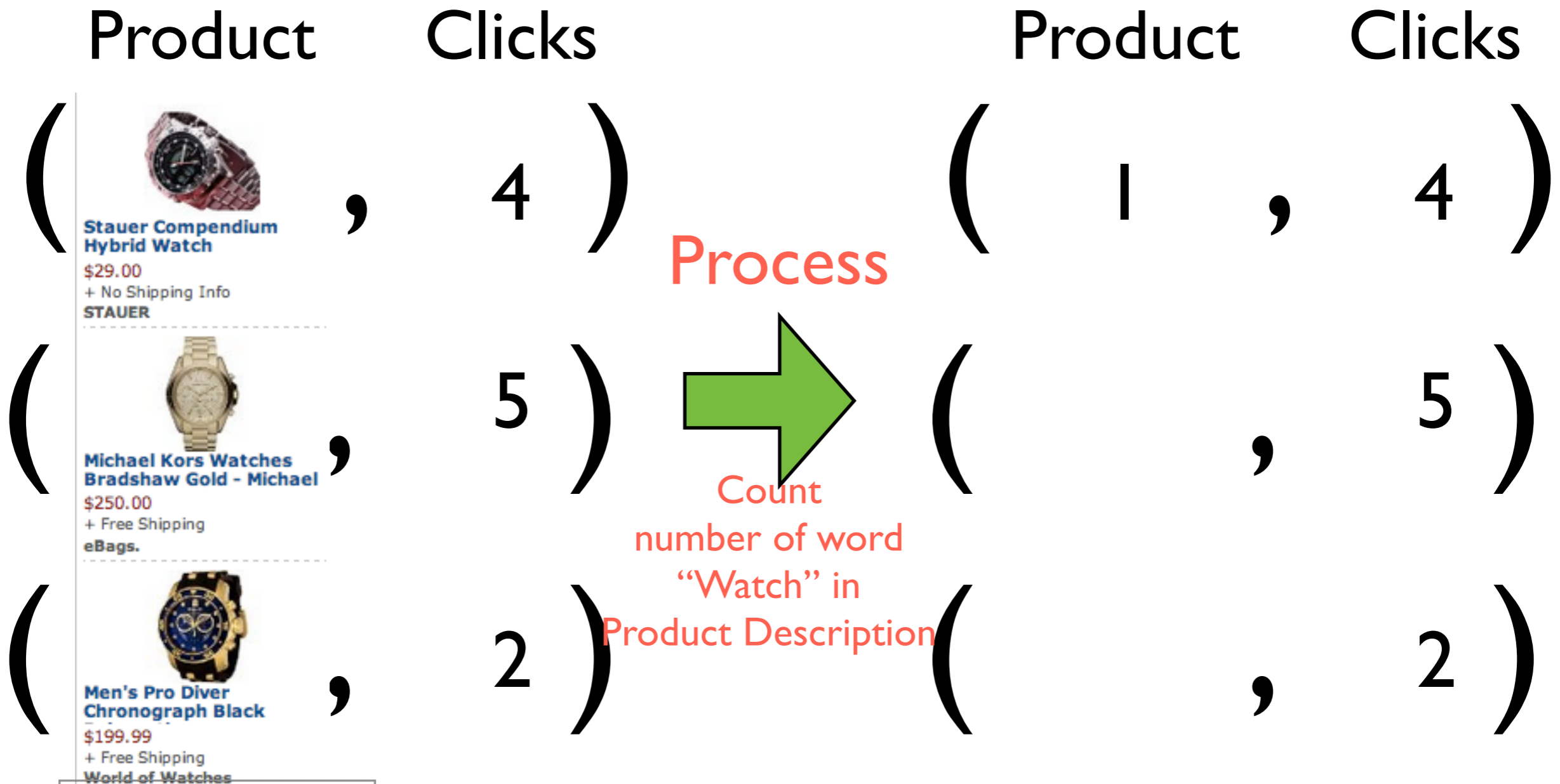
Raw Data

Process Raw Data



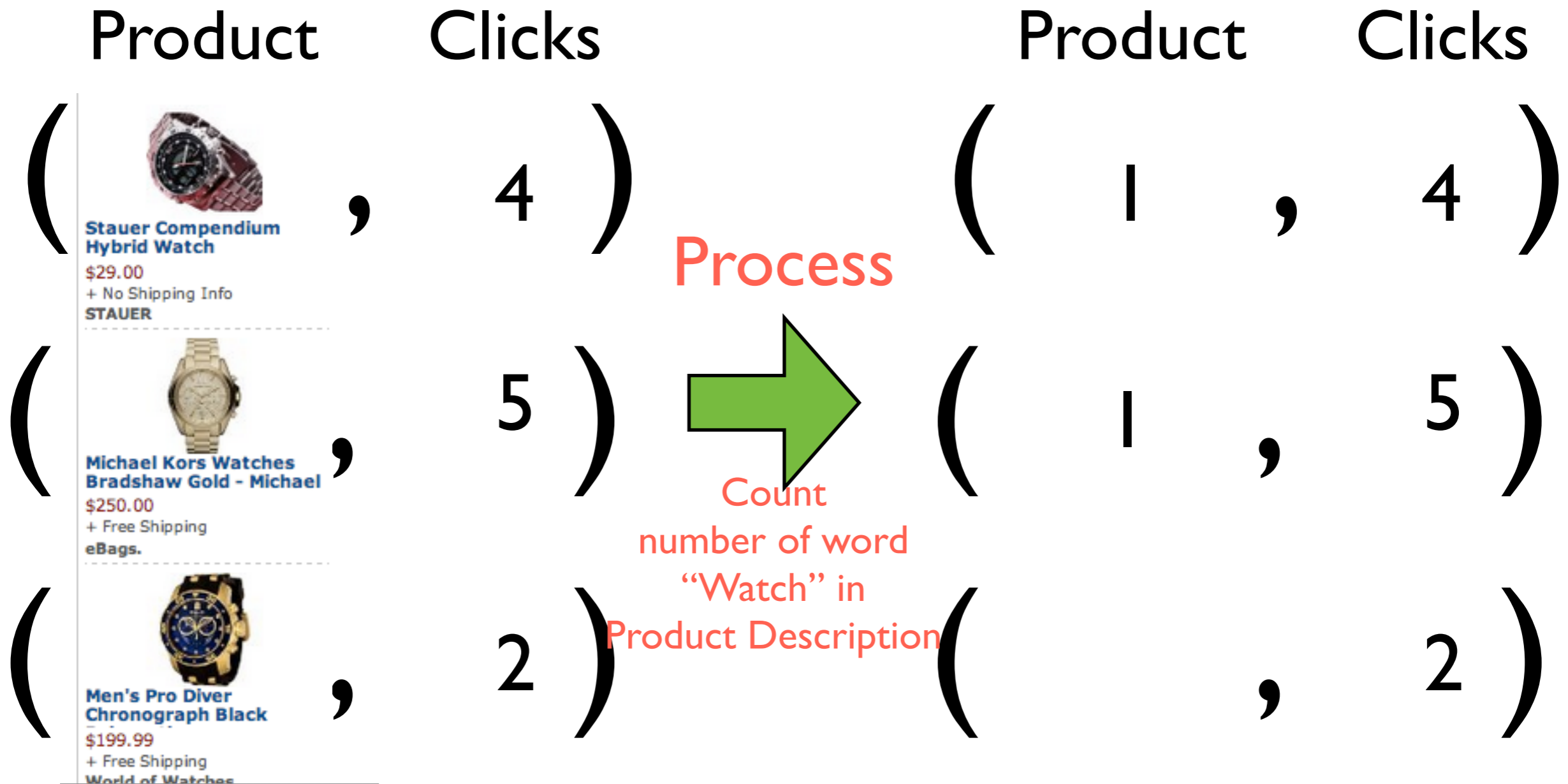
Raw Data

Process Raw Data



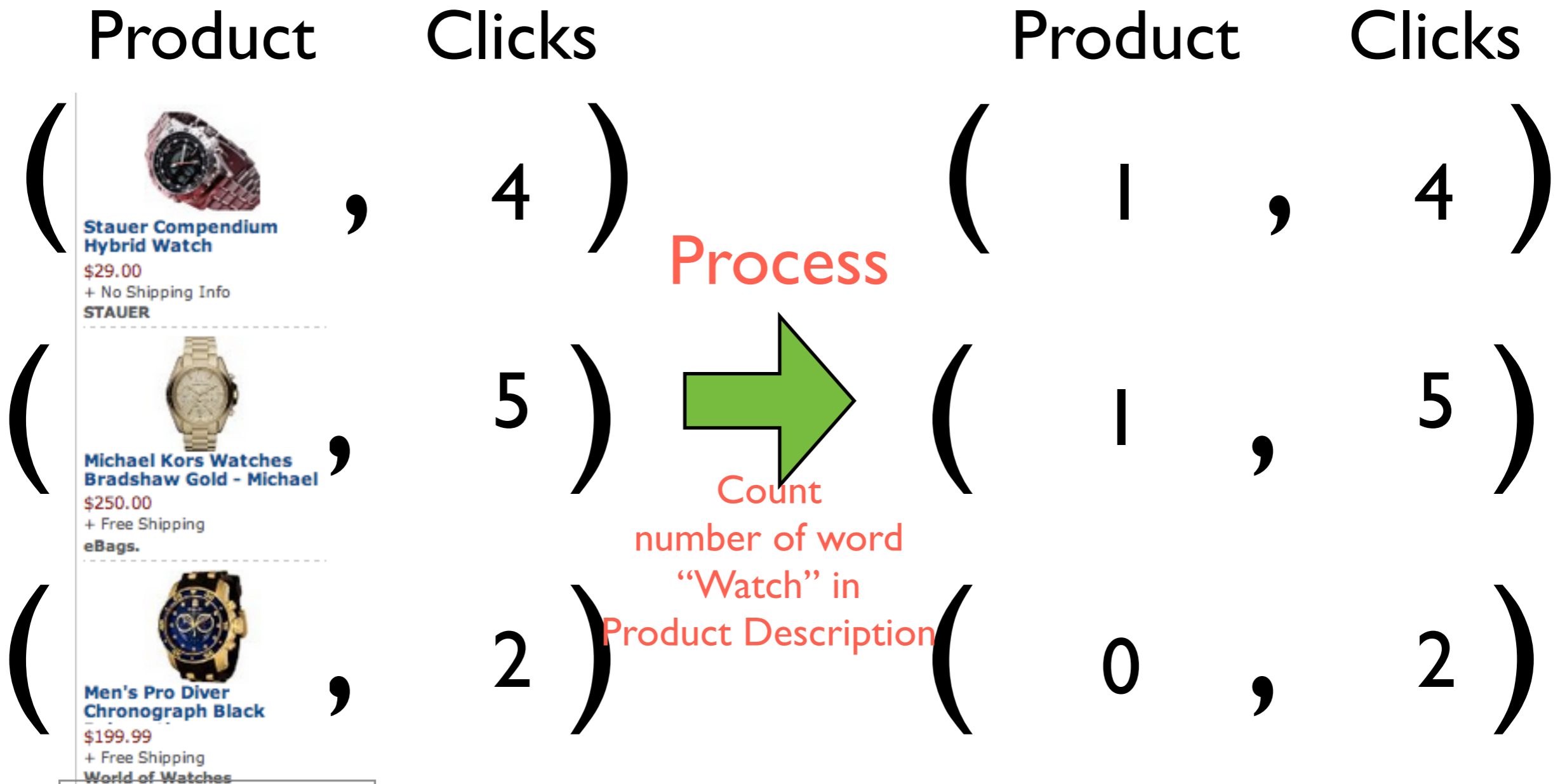
Raw Data

Process Raw Data



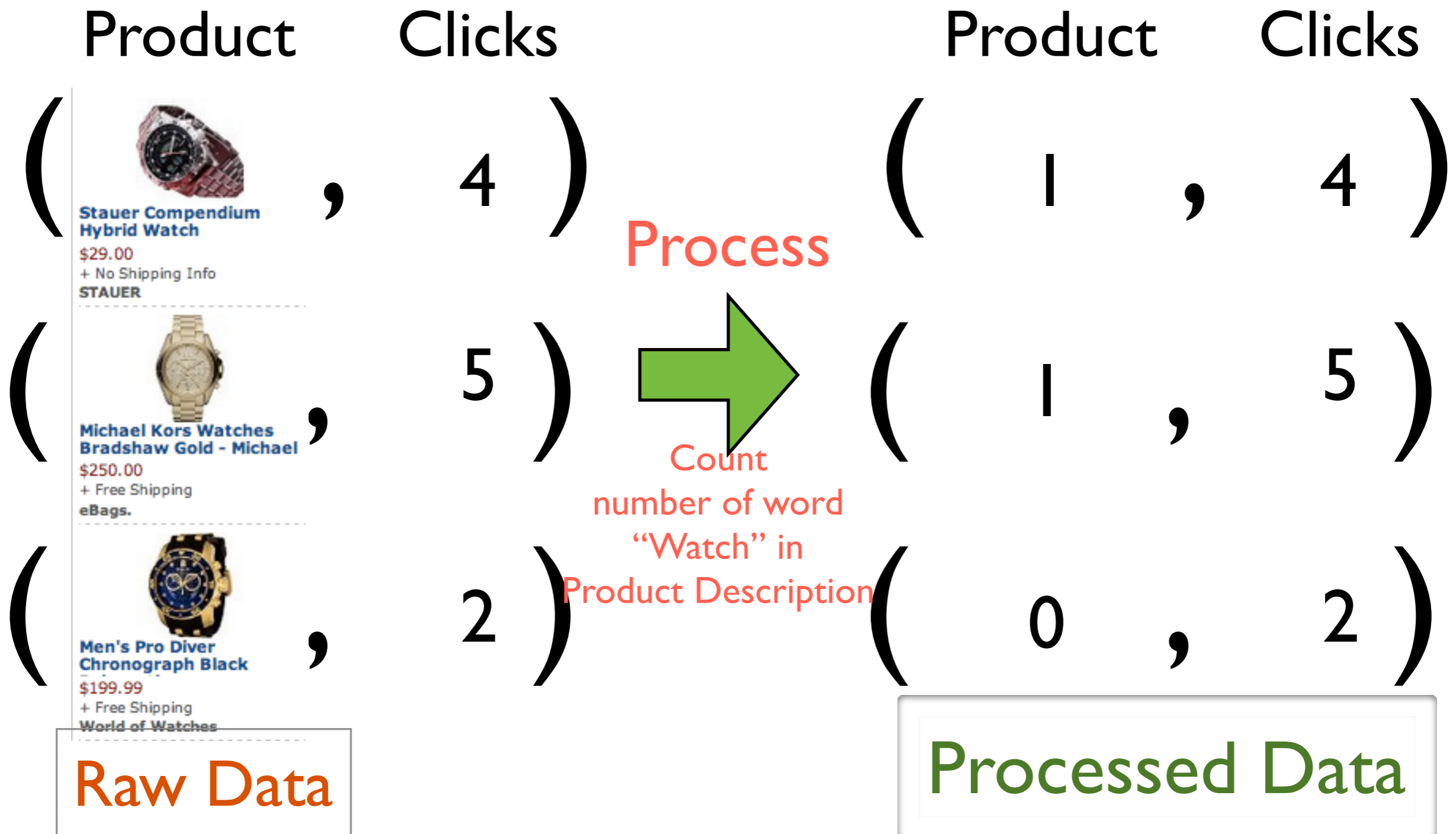
Raw Data

Process Raw Data



Raw Data

Process Raw Data



Process Raw Data


(x_i, y_i)


Product


Clicks

Product

Clicks

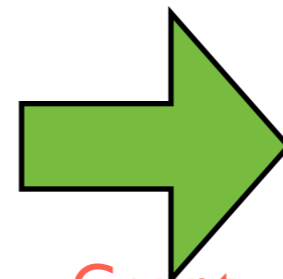
(, 4)

(, 5)

(, 2)

Raw Data

Process



Count
number of word

“Watch” in
Product Description

(1, 4)

(1, 5)

(0, 2)

Processed Data

Process Raw Data


(x_i, y_i)

Product


Clicks

Product


Clicks

(, 4)

Stauer Compendium Hybrid Watch
\$29.00
+ No Shipping Info
STAUER

(, 5)

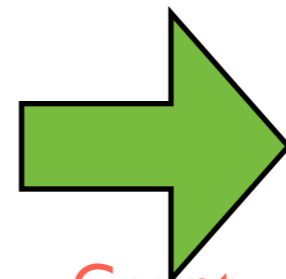
Michael Kors Watches Bradshaw Gold - Michael
\$250.00
+ Free Shipping eBags.

(, 2)

Men's Pro Diver Chronograph Black
\$199.99
+ Free Shipping
World of Watches

Raw Data

Process



Count number of word

"Watch" in Product Description

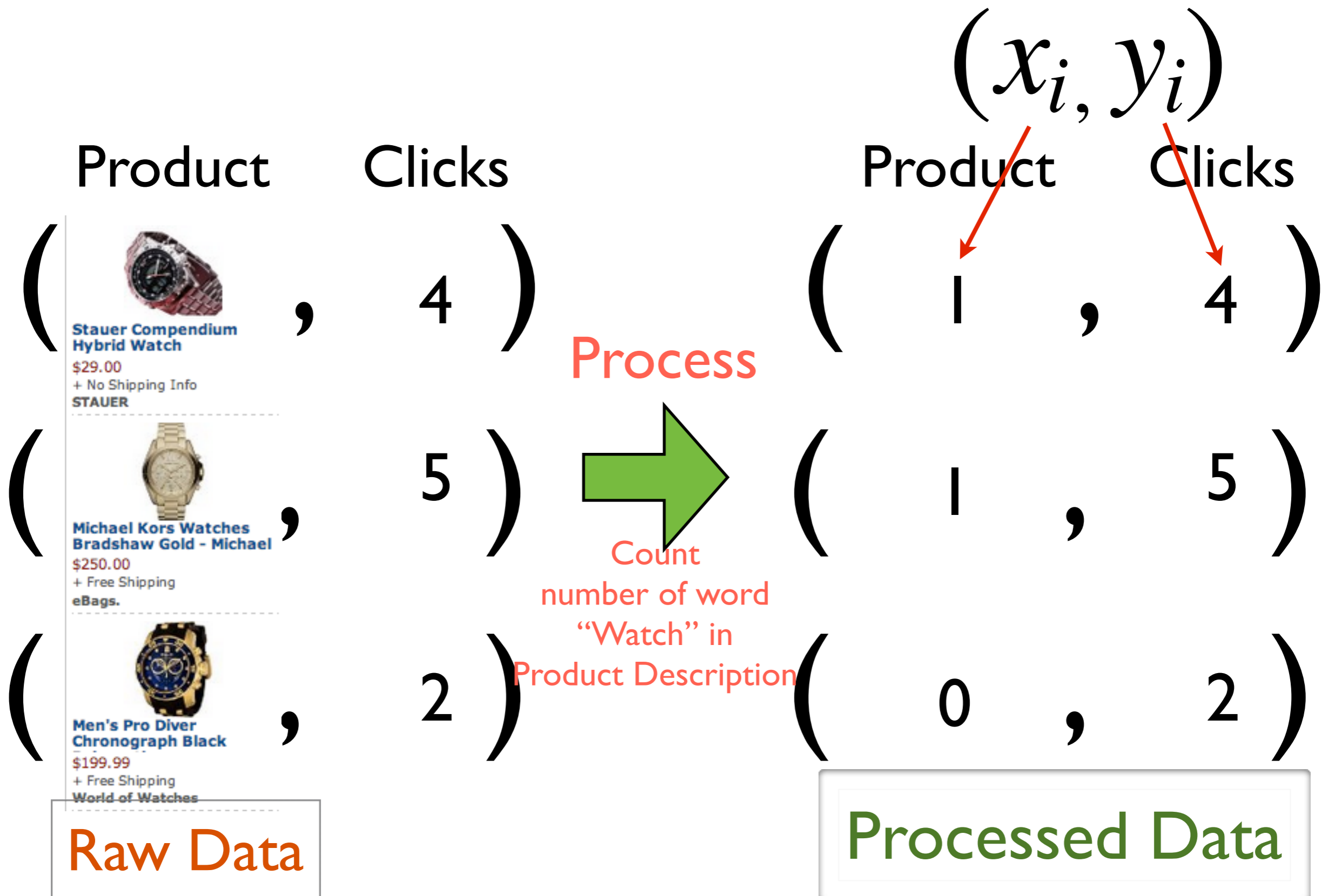
(1, 4)

(1, 5)

(0, 2)

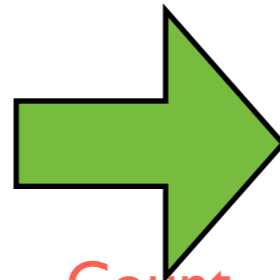
Processed Data

Process Raw Data



Data to Predicted Scores

Process



Count
number of word
"Watch" in
Product Description

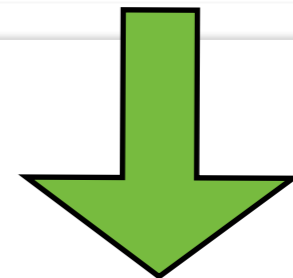


Casio Men's PRW2500-1
Pathfinder Triple Sensor
Tough Solar Digital Multi-
Funtion Pathfinder ...

Raw Data

(| , ?)

Processed Data

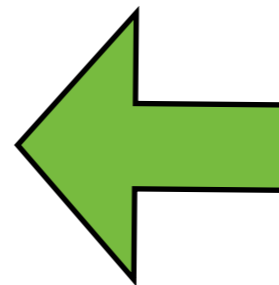


$$f(x) = \theta_0 + \theta_1 x$$




$$\theta_0 = 0 ; \theta_1 = 2$$

Our Model

2

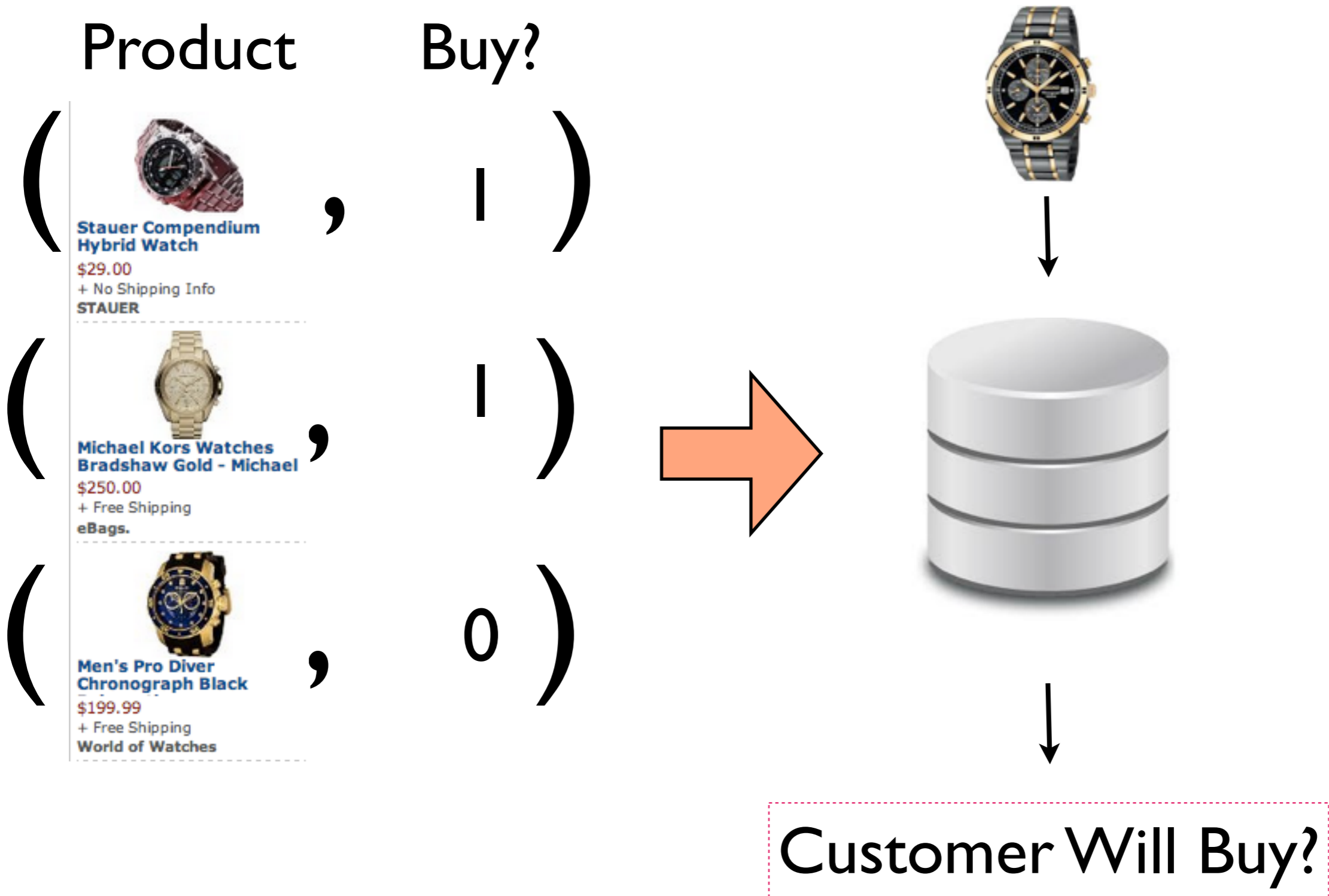


Sales Data

Product	Buy?
 Stauer Compendium Hybrid Watch \$29.00 + No Shipping Info STAUER	1
 Michael Kors Watches Bradshaw Gold - Michael \$250.00 + Free Shipping eBags.	1
 Men's Pro Diver Chronograph Black \$199.99 + Free Shipping World of Watches	0

1 = Bought
0 = Didn't Buy

Sales Prediction Model



Zoolaster can potentially buy more watches from wholesaler that have higher potential of selling online

Data to Classification

- ▶ Given a set of features

$$X=(x_1, x_2, x_3, \dots, x_n)$$

- ▶ we want to predict Y

$$Y=\{0,1\}$$

How about x ?
How do we get them?

Data to Classification

- ▶ Given a set of features

$$X=(x_1, x_2, x_3, \dots, x_n)$$

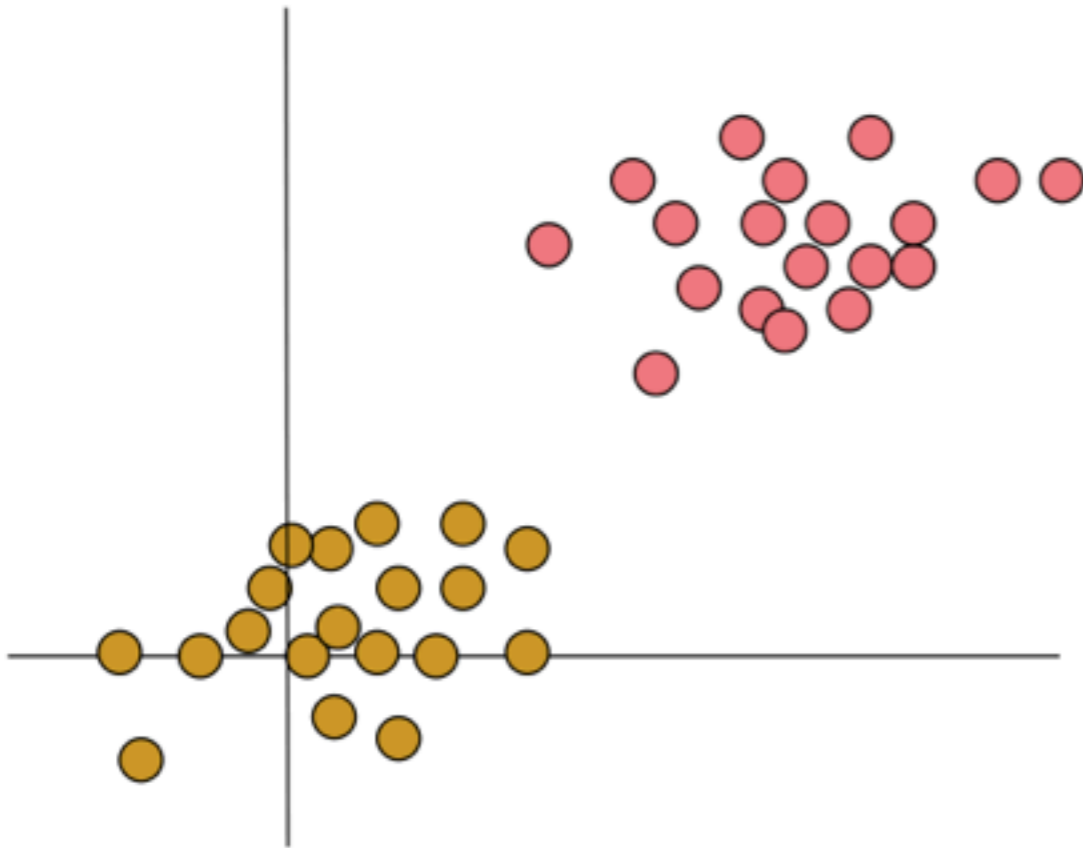
- ▶ we want to predict Y

How about x?
How do we get them?

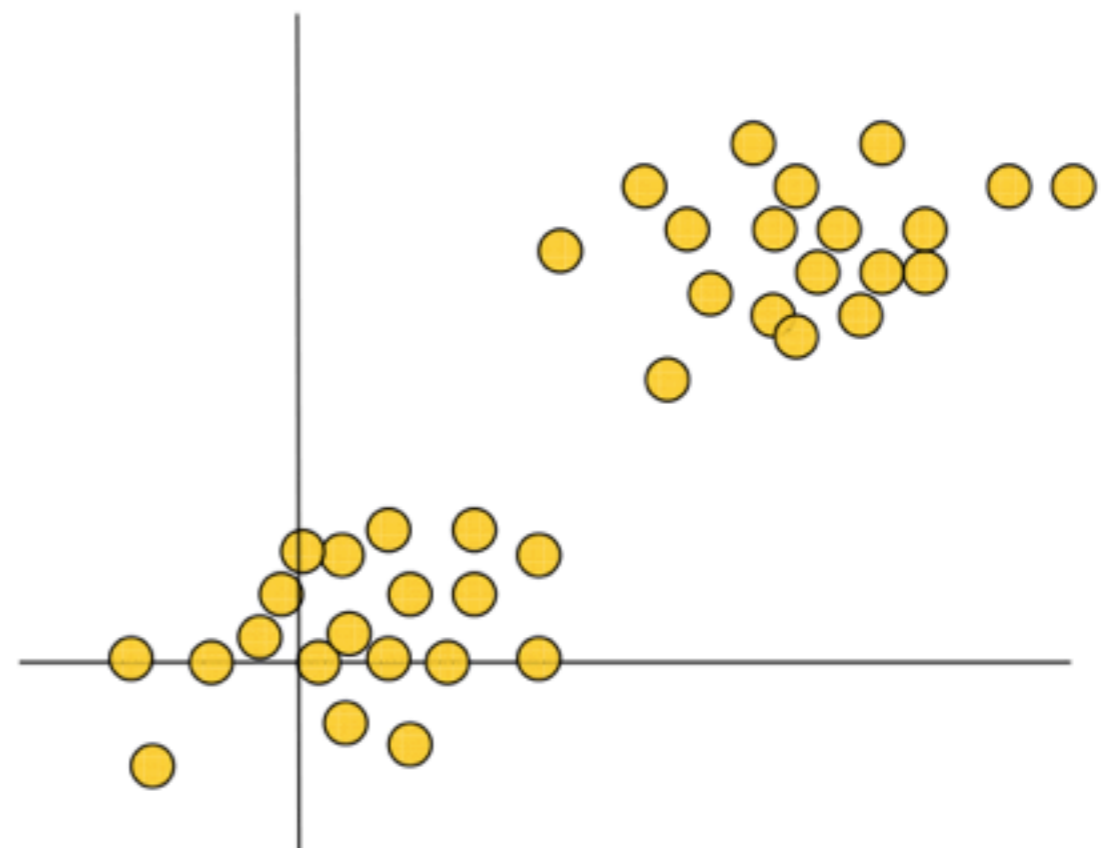
$$Y=\{0,1\}$$



Data to Clusters



Supervised Training
of Classification Algorithm



Unsupervised Training
of Clustering Algorithm