

machine learning conference



Identifying Miscategorized Products

M. Hidayath Ansari [hidayata@](#)
Applied Scientist II | Browse

Subhadeep Chakraborty [sdeep@](#)
Sr SDE | Browse

Customer Experience – Shopping Funnel



Customer Experience – Shopping Funnel



Customer Experience – Shopping Funnel



Customer Experience – Shopping Funnel



Customer Experience – Shopping Funnel



Customer Experience – Category Landing Page

prime

FREE One-Day

Get **FREE One-Day Shipping** on qualifying orders over \$35

Show results for

< Appliances

< Parts & Accessories

< Humidifier Parts & Accessories

Absorption Sleeves

Refine by

Part Type

Generic

Color

Delivery Day

Get It by Tomorrow

Amazon Prime

prime

prime

FREE One-Day

Featured Brands

Opaceluuk

Aprilaire

Homecube

ERP

Kissbuty

Trion

StarryBay

See more

Avg. Customer Review

& Up

& Up

Humidifier Absorption Sleeves

Welcome to the Humidifier Absorption Sleeves Store, where you'll find great prices on a wide range of different humidifier absorption sleeves for your home.

Best sellers

Opaceluuk OP-USB-HUMI01 Portable Personal Mini...

★★★★☆ 197

\$23.99 prime

Aprilaire 510 Replacement Filter

★★★★☆ 32

\$67.90

TONERONE 400ml Ultrasonic Aroma Essential Oil...

★★★★☆ 178

\$28.99 prime

USB Car Diffuser, Homecube Ultrasonic Cool Mist...

Homecube

★★★★☆ 42

\$9.99 prime

Kissbuty 140 Packs Shelf Pins 5mm Spoon Shape...

★★★★☆ 13

\$9.69 prime

Replacement Drain Pump For 27001233 Fits Maytag...

ERP

★★★★☆ 9

\$15.45

StarryBay Aroma Diffuser Essential Oil Diffuser...

★★★★☆ 18

\$25.99 prime

Most gifted

AMAZON CONFIDENTIAL

7

Customer Experience: Sort by Price

1-60 of over 2,000 results for Electronics : Car & Vehicle Electronics : Vehicle Electronics Accessories : Radar Detectors Report a search issue

Sort by Price: Low to High

☐ prime | FREE One-Day

Get FREE One-Day Shipping on qualifying orders over \$35

Show results for

< Electronics

< Car & Vehicle Electronics

< Vehicle Electronics Accessories

Radar Detectors

Refine by

Delivery Day

☐ Get It by Tomorrow

Amazon Prime

☐ prime

☐ prime | FREE One-Day

Feature Keywords

☐ Laser

☐ 360 Degree

☐ Voice Alerts

☐ Undetectable

☐ Camera Detector

☐ GPS

☐ Wireless

See more

Featured Brands

☐ Uniden

☐ Escort


☐ Cobra

☐ GPS-Radar-Shop


☐ stellardeals

☐ Garmin


☐ Bushnell




Radar Detector
Free
★★★★★ 1




For Samsung Galaxy S9 Plus 6.2inch,Sunfei Shockproof Heavy Duty Stand Case Skin Cover
\$0⁰¹ - \$1⁴³




DondPO Womens Fashion Lace up Short Sleeve T-Shirt Sexy Deep V-Neck Sexy Blouse Tops Casual Summer T-Shirt Loose...
\$0⁰¹ \$19.99




DondPO 2018 Newest Blouse, Ladies' Bandages Sleeveless Vest Top High Low Tank Top Notes Strappy Tank Tops
\$0⁰¹ \$15.99




DondPO Fashion Bohemian Womens Rhinestone Geometric Type Gem Pendant Statement Necklace
\$0⁰¹ \$9.99




DondPO Women's Self Tie Back V Neck Crop Cami Top Camisole Cotton Blended Stripe Shirts
\$0⁰¹ \$15.99




DondPO 2018 Newest Blouse, Women Ladies Star T-Shirt Summer Casual Short Sleeve Stripe Round Neck Floral Knot...
\$0⁰¹ \$15.99




DondPO Exquisite Sequins Multilayer Chain Turquoise Beads Necklace with Feather Pendant
\$0⁰¹ \$9.99




DondPO Unisex Men Man Stainless Steel Cross Pendant Chain Necklace for Men Women
\$0⁰¹ \$9.99



DondPO New Man Wolf Tooth Necklace Titanium Steel Domineering Pendant Jewelry For Cool Men
\$0⁰¹ \$9.99



DondPO Mother's Day New GiftsNew Retro Peace Necklace Pendant Black Leather Cord Choker Charm Collar Necklaces
\$0⁰¹ \$9.99



Watches for Men On Sale! Paymenow Men Multi-Function LED Electronic Watches Digital Watches Fashion Outdoor Spor...
\$0⁰¹

“Radar Detectors”

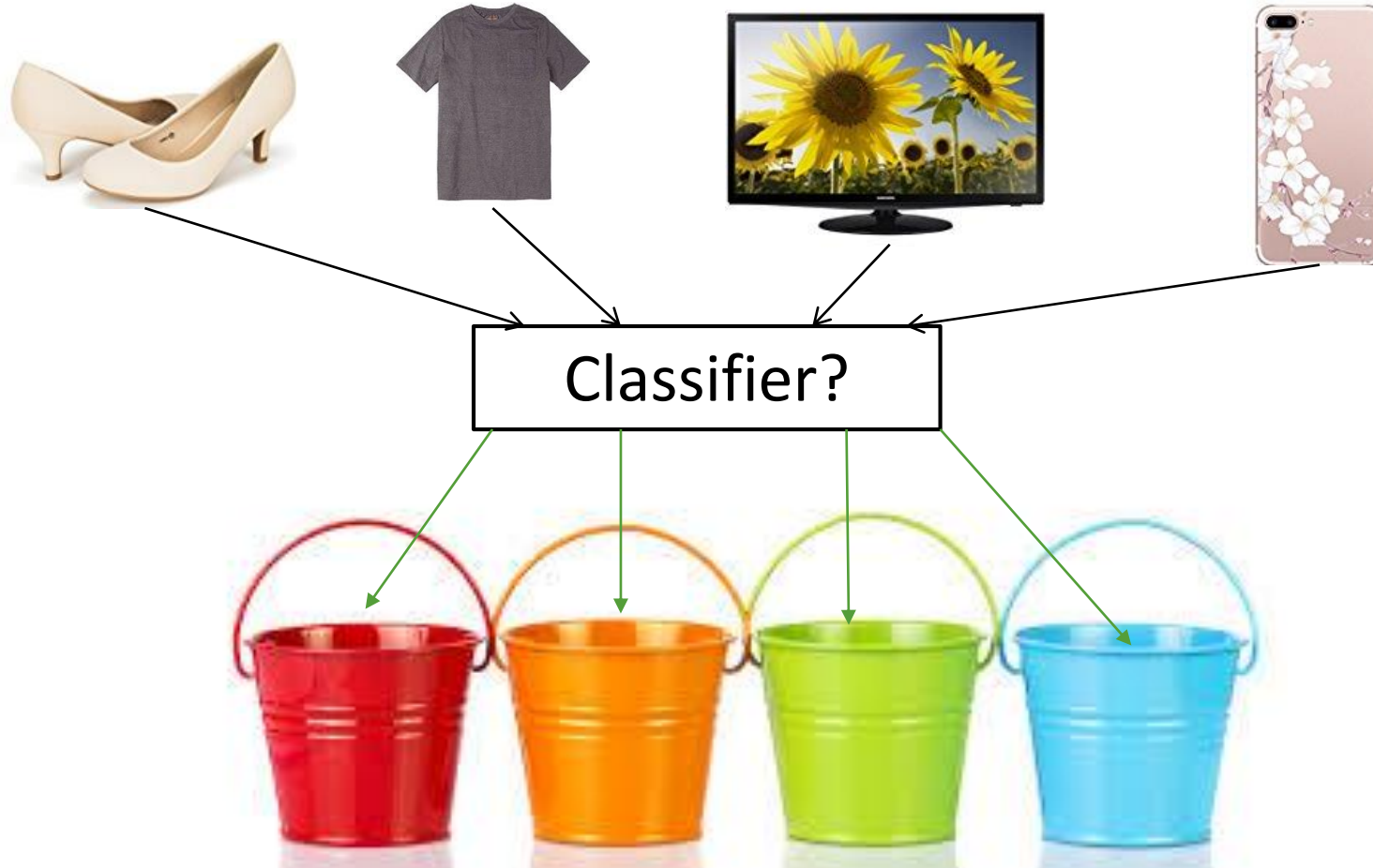
Product Discovery

Amazon Browse: Enabling customers' discovery experience



Without having to weed through irrelevant products

Finding Product Miscategorizations



BUT ...

Product Miscategorizations as Outliers

Product Miscategorizations as Outliers

“Which of these doesn’t belong?”

Product Miscategorizations as Outliers

“Which of these doesn’t belong?”



Product Miscategorizations as Outliers

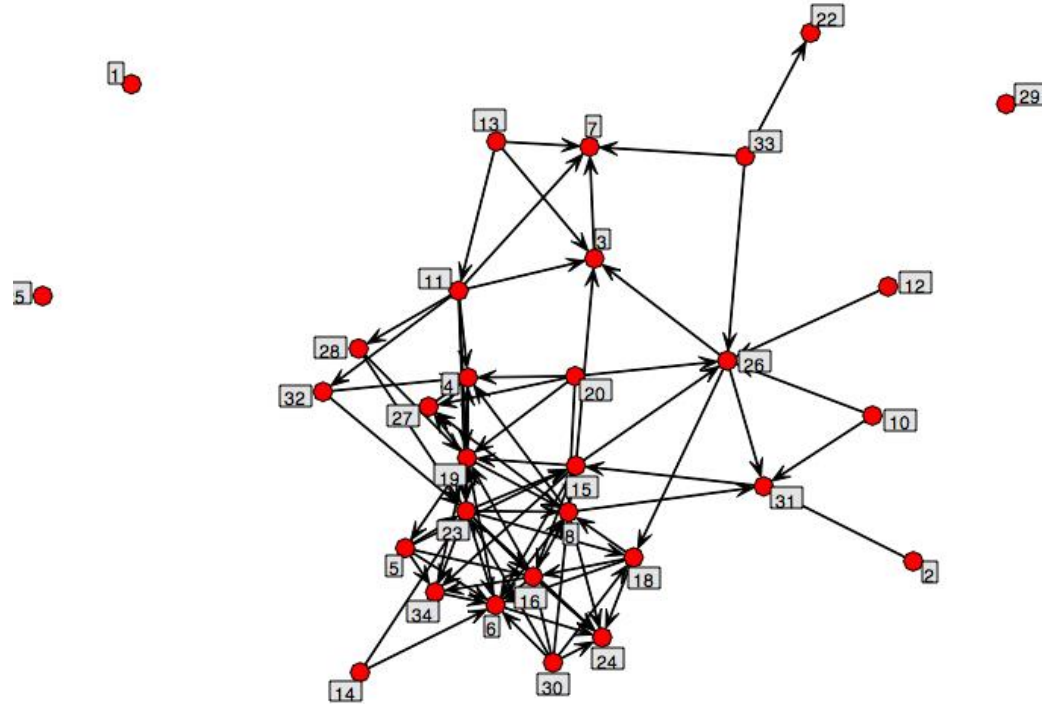
“Which of these doesn't belong?”



Unsupervised/Lightly supervised methods

PageRank

- Finds the most “central”/“important” nodes
- Conversely ...

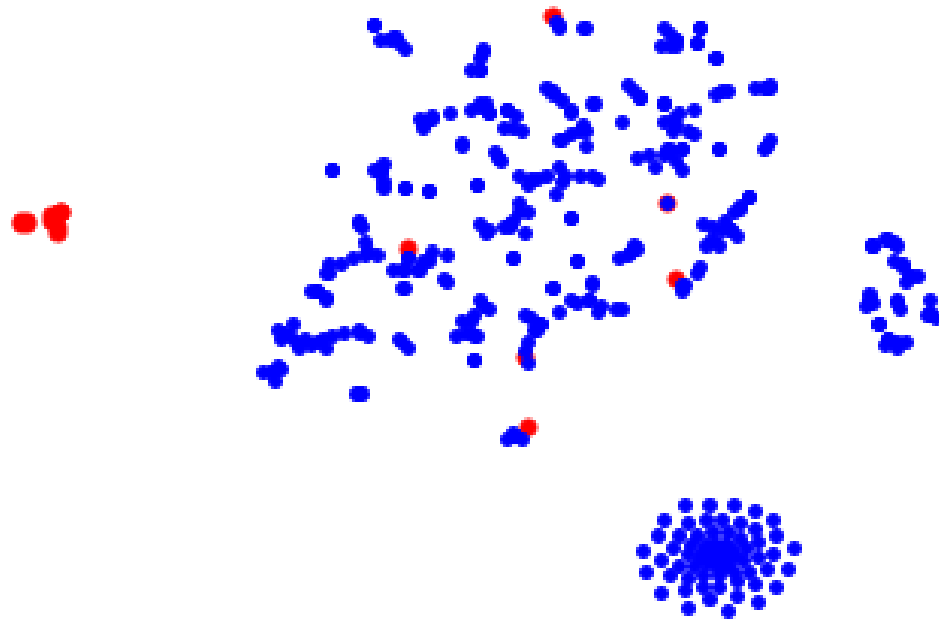


Outrank – Example with ASINs



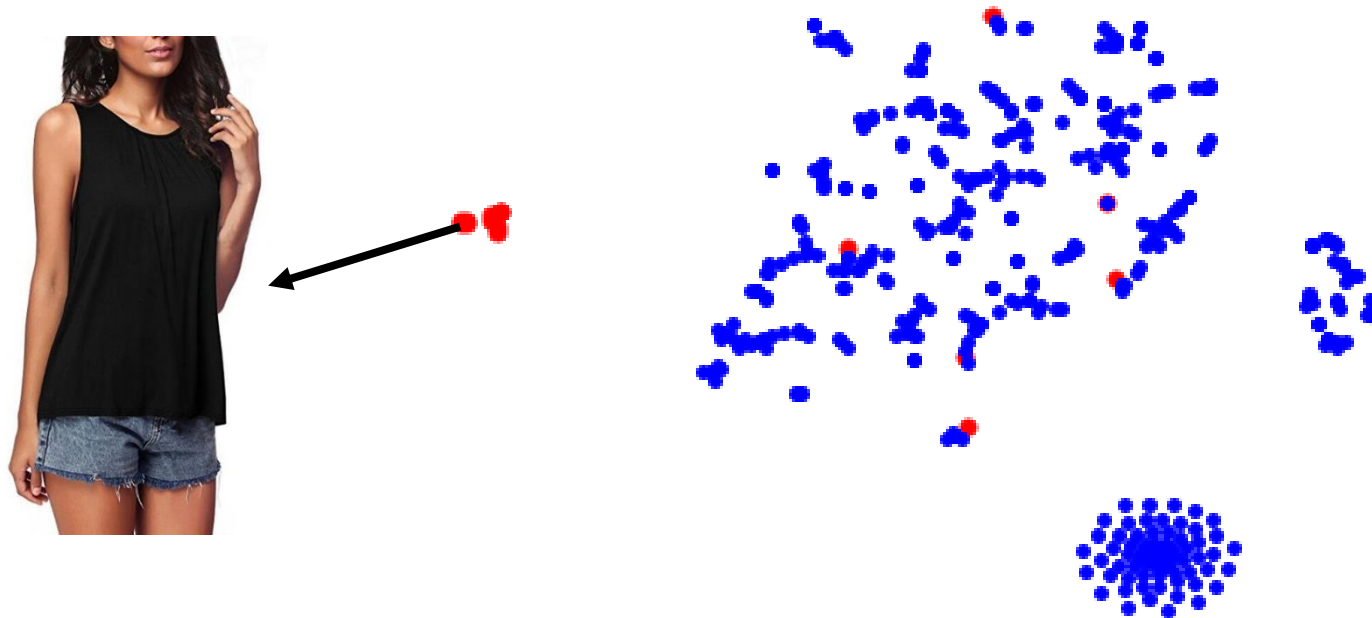
Dishwashers (dimensionally reduced)

Outrank – Example with ASINs



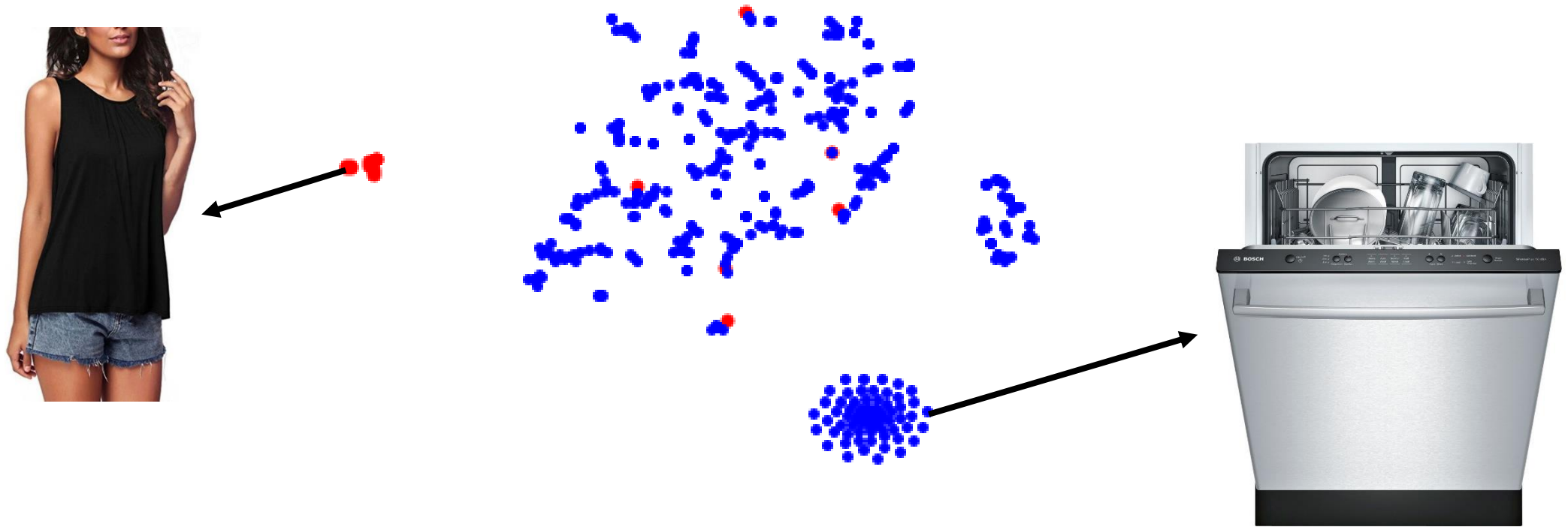
Dishwashers (dimensionally reduced)

Outrank – Example with ASINs



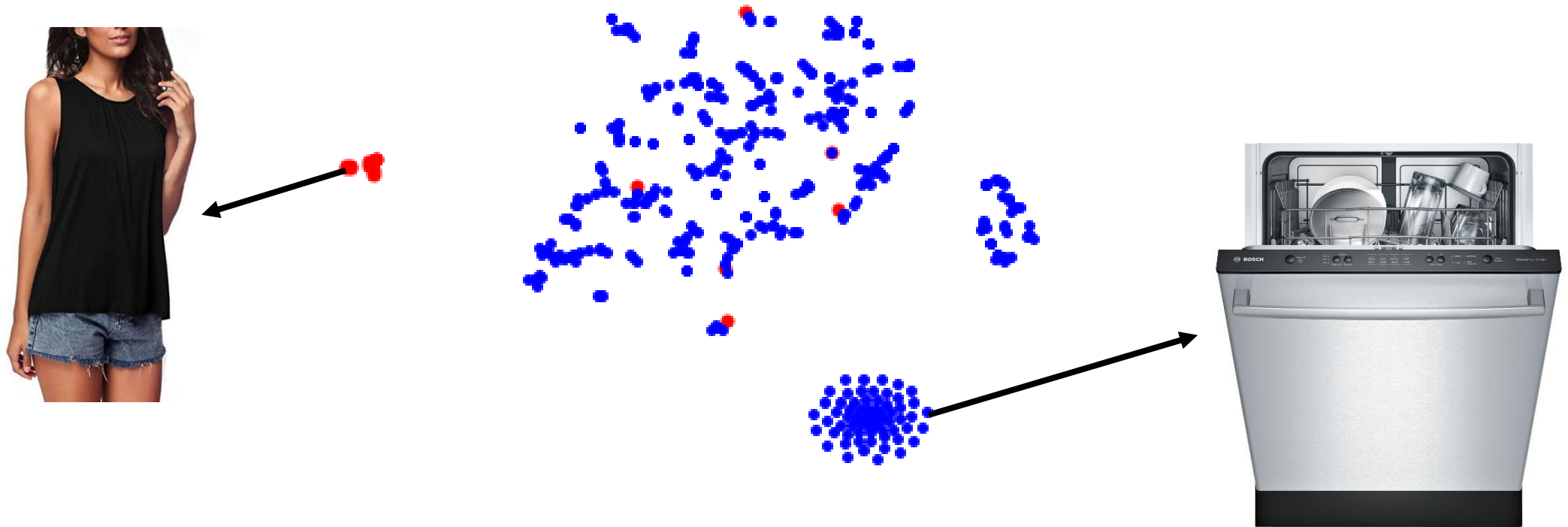
Dishwashers (dimensionally reduced)

Outrank – Example with ASINs



Dishwashers (dimensionally reduced)

Outrank – Example with ASINs

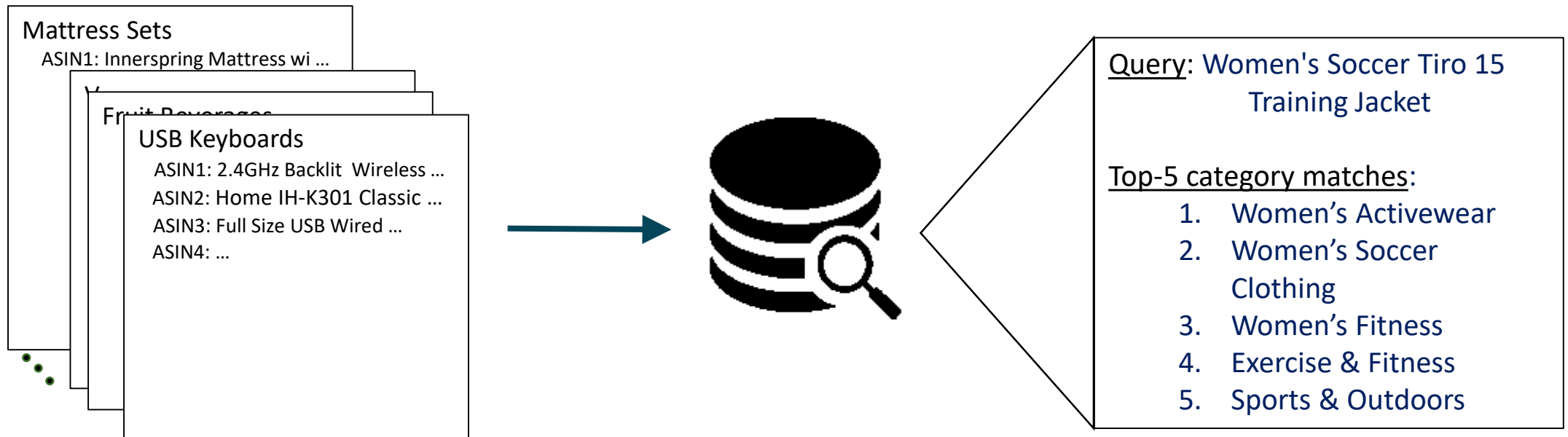


Still, Outrank has a high false positive rate
Dishwashers (dimensionally reduced)

Moonesinghe, H. D. K., and Pang-Ning Tan. "Outrank: a graph-based outlier detection framework using random walk." *International Journal on Artificial Intelligence Tools* 17.01 (2008): 19-36.

Miscategorizations – Using Search

Idea: Use a search index for product categories to find miscategorized products

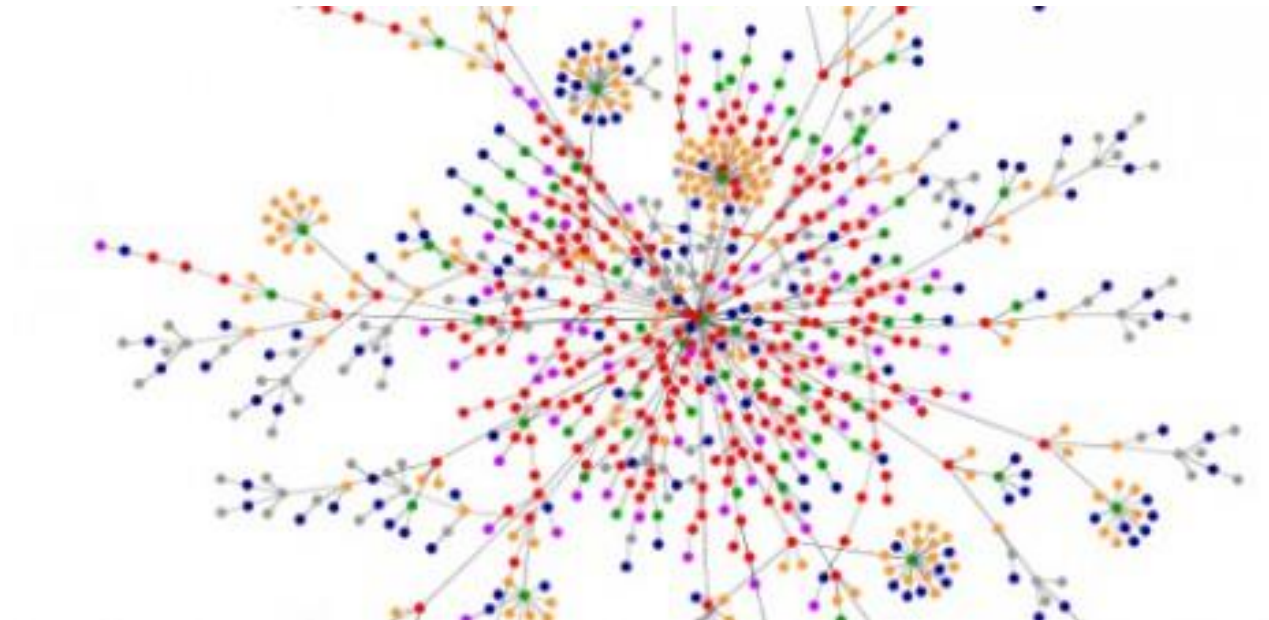


Singular outliers – Results & Impact

Marketplace	Assignments removed	Audited miscat Precision	Irrelevancy Reduction	OPS impact through irrelevancy reduction
US	2,319,105	91.8%	24% → 9%	\$15,925,972
UK	2,168,596	97.2%	20% → 14%	\$4,886,300

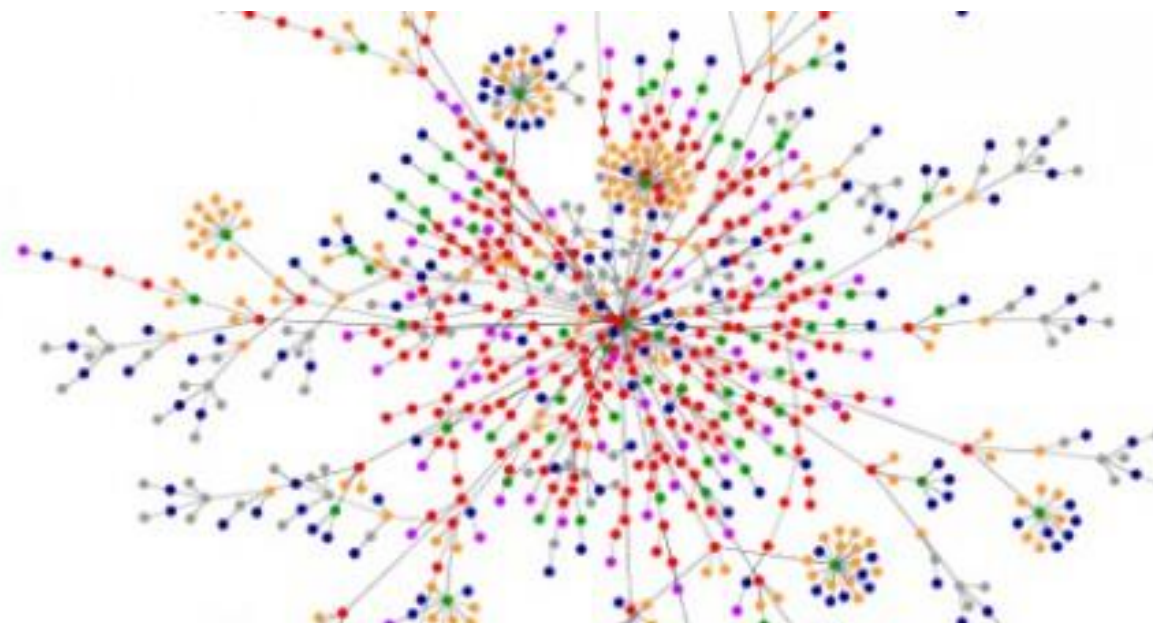
Results for bestseller ASINs

Clump outliers – Using expert input



Clump outliers – Using expert input

1. Select “exemplars”
2. Get judgments
3. Propagate judgments



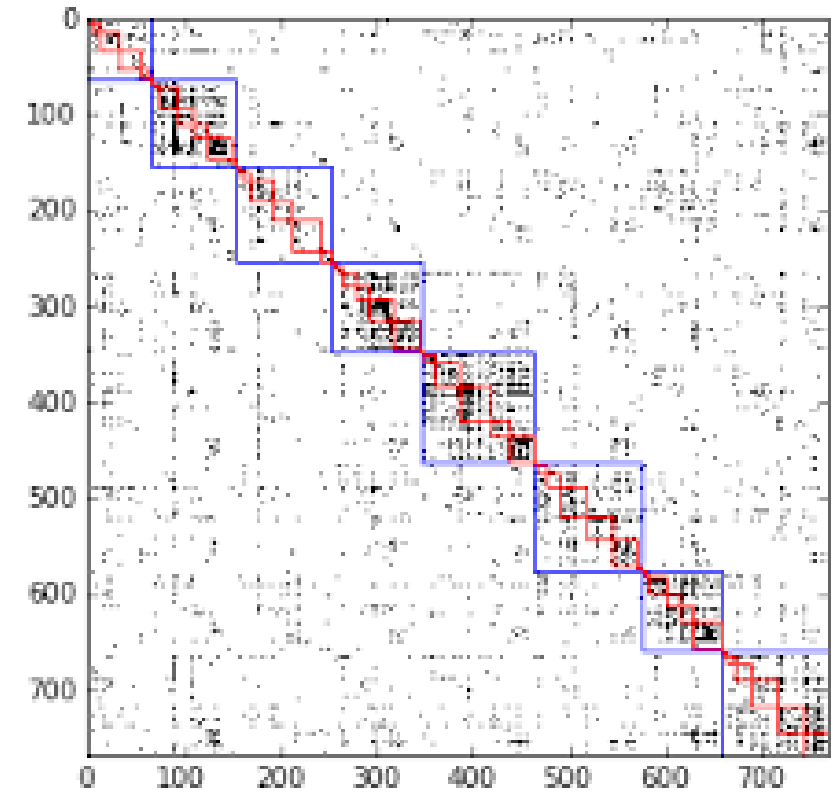
Clump outliers – Affinity Matrix

Constructing a graph adjacency/ASIN affinity matrix:

1. Vectorize ASINs using averaged GloVe
2. Find the 250 nearest neighbors of each ASIN
3. Construct matrix using (Euclidean) distances to nearest neighbors
4. Apply Gaussian kernel to each non-zero entry:

$$A_{ij} = \exp\left(\frac{-\text{dist}(\mathbf{x}_i, \mathbf{x}_j)}{2\sigma^2}\right)$$

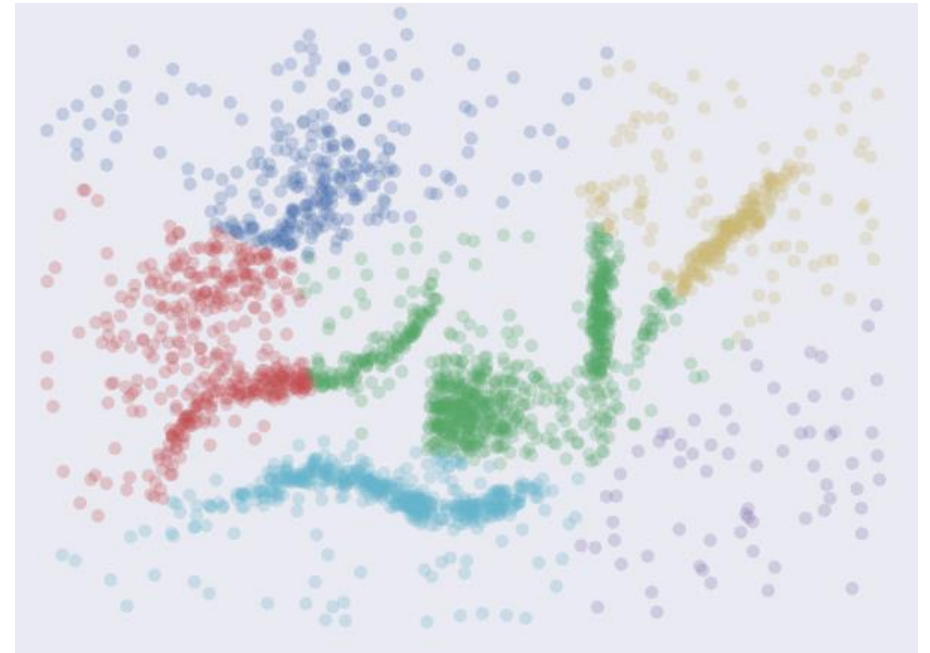
5. Symmetrize matrix



Clump outliers – Exemplar Selection

Choosing a set of “representative” ASINs:

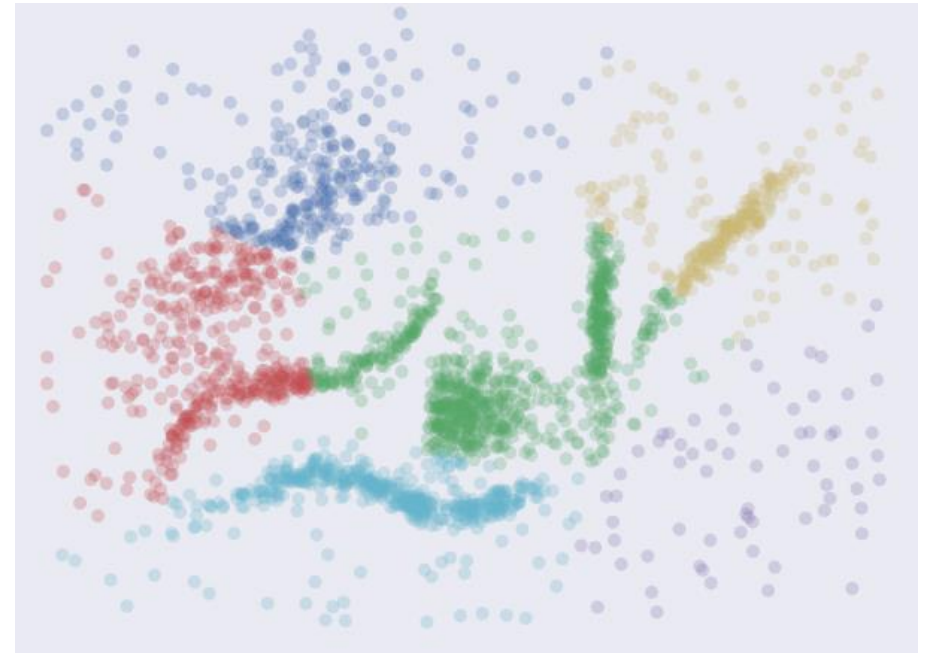
1. Spectral clustering using affinity matrix
2. Compute the “degree vector”
3. Select ASIN with highest “degree” in each cluster



Clump outliers – Exemplar Selection

Choosing a set of “representative” ASINs:

1. Spectral clustering using affinity matrix
2. Compute the “degree vector”
3. Select ASIN with highest “degree” in each cluster

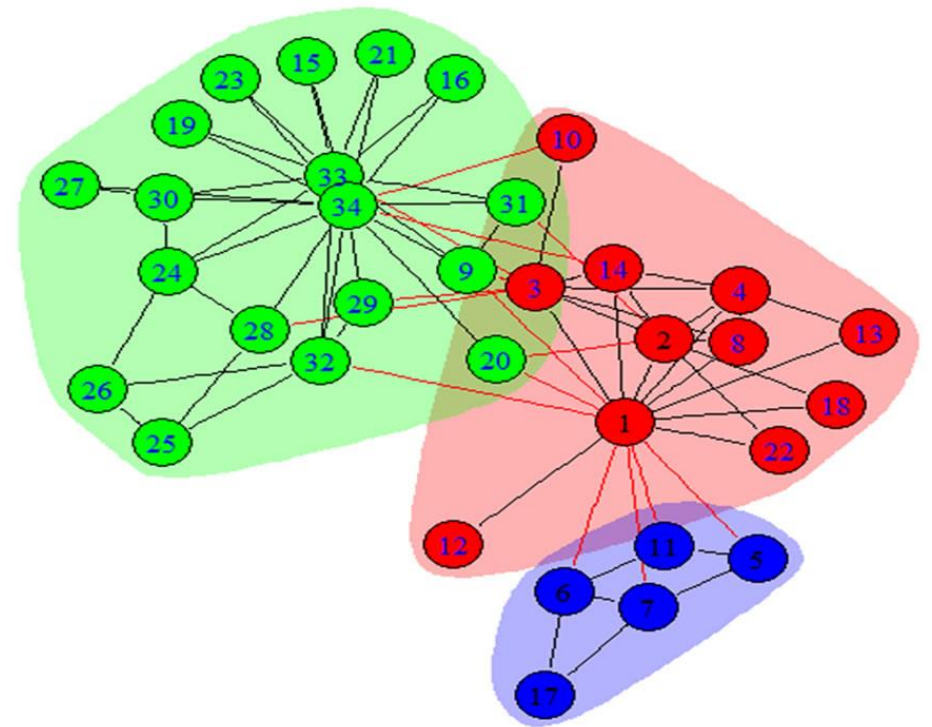


Obtain judgments for selected ASINs

Clump outliers – Judgment Propagation

Propagating expert judgments:

1. Construct $S = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$
2. Iterate $F(t + 1) = \alpha SF(t) + (1 - \alpha)Y$ until convergence
(for α between 0 and 1)
3. Obtain labels for ASINs from the limit of $F(t)$ by applying thresholds



Clump outliers – Results

Dataset	Total ASINs	Negative exemplars	Negative Propagations	Miscat Precision	Miscat Recall
TVs	19,345	20/50	5,117	98%	77%
Storage Benches	2,123	19/20	1,647	100% (est.)	-
Radar Detectors	15,225	13/19	9,427	100% (est.)	-
Cooktops	22,266	40/50	19,263	100% (est.)	-
Dart Cases & Wallets	28,992	98/100	27,151	100% (est.)	-
50 UK categories	1,748,936	2,372/5000	818,176	95.8% (audit)	-

Towards a Pristine Catalog

Questions?