everqry

Meaning and Insight Through Search Analytics

Context

Adobe purchased the stock media provider Fotolia in 2015 for $800m

fotolia

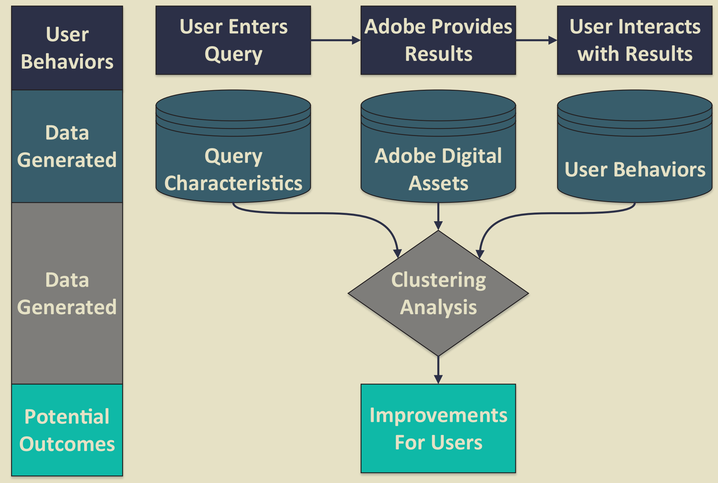
St Adobe Stock

Problem

Adobe wants to know how their customers interact with this new service, but users generate massive amounts of data that can be difficult to understand.

80 Million Searches — *Performed By* → **3 Million** Users — *Interacting With* → **50 Million** Images and Videos

Data Flow

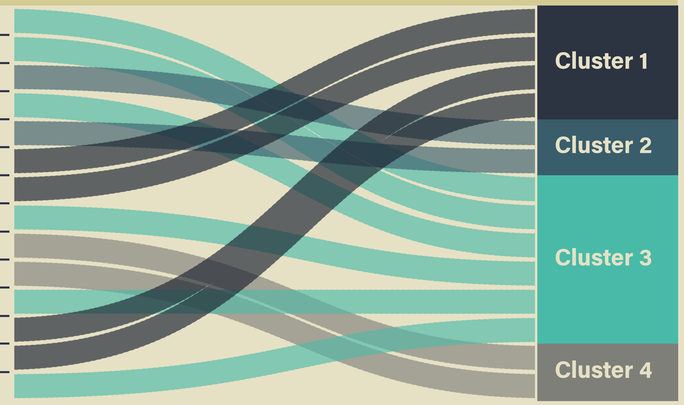


Approach

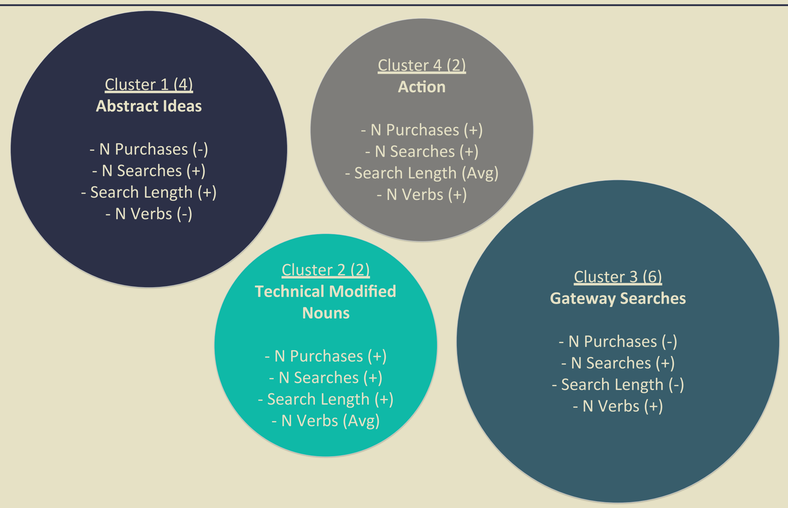
- Raw query data aggregated in Hive
- Content engagement variables generated
- Query variables generated using natural language processing
- K-means clustering algorithm performed in Python
- Process refined and repeated until meaningful clusters are formed

Query	Purchase Rate	Number of Searches	...	Search Length	Number of Verbs
Bird	9%	14K	...	4	0
Working Out	13%	15 K	...	11	1
Backgroung Logo	24%	5 K	...	15	0
Theater	14%	13 K	...	7	0
Yoga Silhoute	27%	5 K	...	14	0
Student Loans	12%	15 K	...	13	0
Perfect Day	11%	13 K	...	11	0
Strategy	3%	14 K	...	8	0
Man Fishing	31%	16 K	...	11	1
Flying Plane	26%	17 K	...	12	1
Skyline	6%	15 K	...	7	0
Job Interview	17%	15 K	...	13	0
Commerce	15%	15 K	...	8	0
Cloud	18%	13 K	...	5	0

Clustering



Results



Hypotheses and Validation

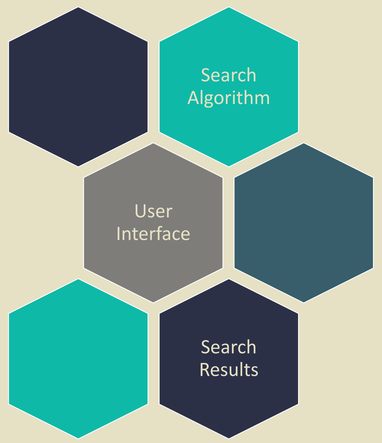
Example Hypothesis:

Cluster 3: Gateway Queries represent short and simple queries where a user is beginning a search use a generic term, and looking at results to help formulate their subsequent searches. Gateway searches are common but are rarely result in purchases.

Validation:

Hypotheses are validated using “user traces” where a customer’s pruchasing and download history is viewed in conjunction with their movement from one search to another.

Outcomes



Alex Filipkowksi MSIM ◆ Fernando Centurion MSIM
Jorge Retamales MSIM ◆ Grant Woods MSIM

Andy Edmonds
Project Sponsor

Information School
UNIVERSITY of WASHINGTON