



Leveraging AWS and Machine Learning to Power Search at Zocdoc

Pedro Rubio

Head of Search Engineering

Brian d'Alessandro

Head of Data Science

Agenda

- How we're built - People and Architecture
- How we're built - the Data
- Questions

Problem Statements:

1. Patients need to find and book with a doctor, and,
2. Patients don't often know what kind of doctor they need.



Neurodevelopmental Disabilities Specialists and other providers in 10010

Sort by: default order

Mon
Apr 3Tue
Apr 4Wed
Apr 5

Dr. Mark Rybakov, DO

Child and Adolescent Psychiatrist



"I would highly recommend Dr. Rybakov. He saw me promptly with almost no wait..."

28 W 44th Street, New York, NY 10036

Within 1.5 miles

-

8:30 am

-

9:00 am

-

11:00 am

-

more



Dr. Randall Ross, MD

Child and Adolescent Psychiatrist



"Dr. Ross is a very laid back and personable practitioner. He is very attentive and..."

15 W 14th St, New York, NY 10011

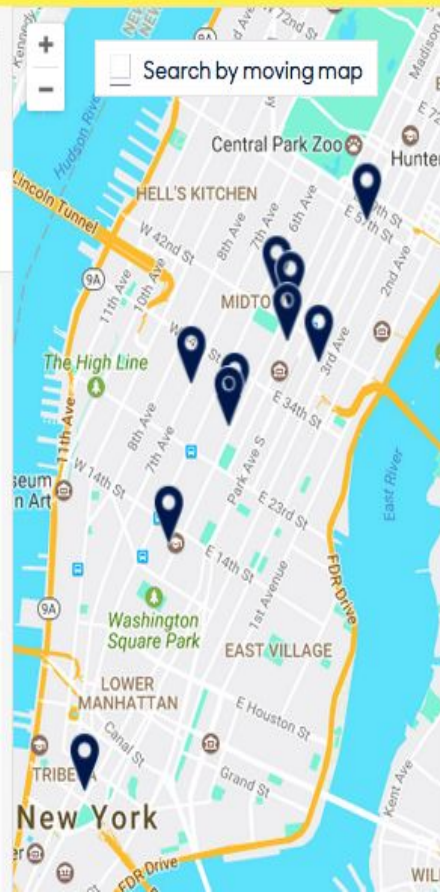
-

6:00 pm

-

-

-



How we're built

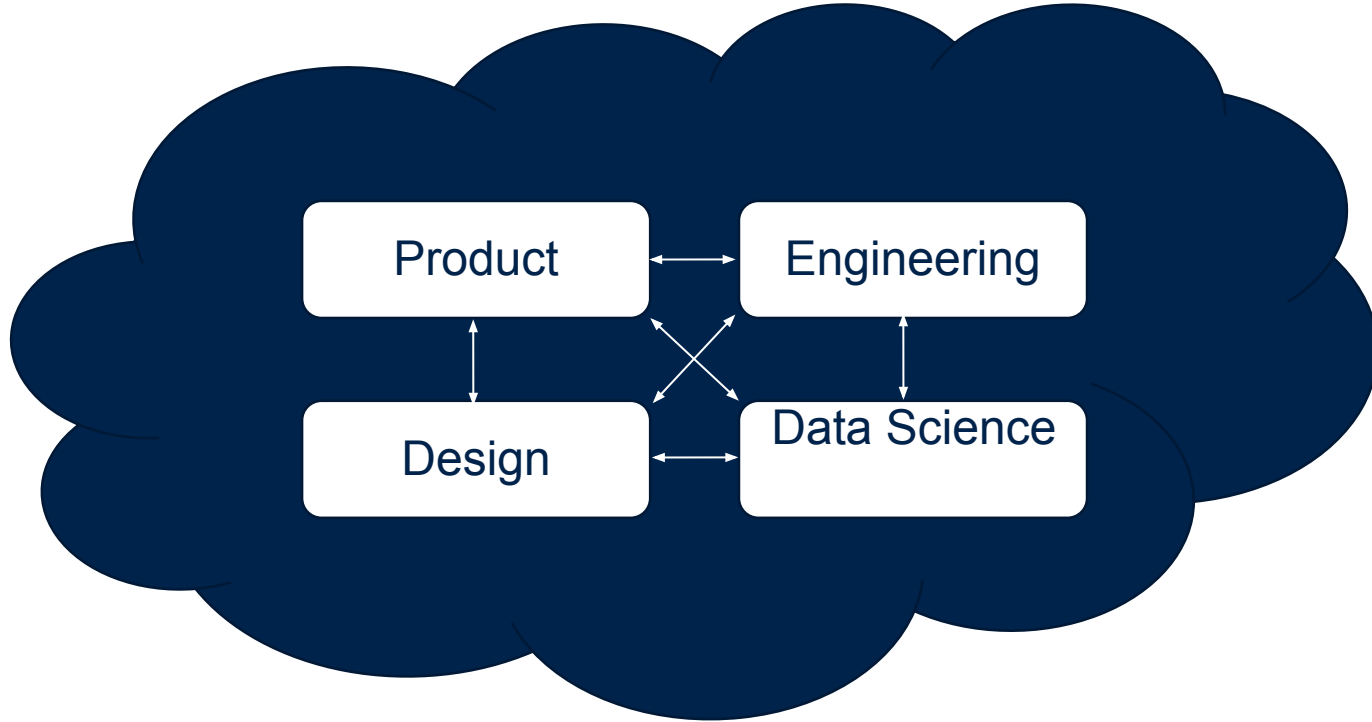
And solving “what the patient means”

Core Optimization Problems for ZD Search

- Cross team collaboration enabling maximum iteration speed
- Deliver recommendations < 200 ms
- Patient satisfaction

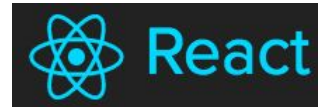
(And our architecture plays a big role here!)

The Search Team



Zocdoc Tech Stack

- NodeJS, ES6, Babel, React
- AWS - Cloudformation, Docker, ECR, EC2, ELB
- Kinesis / Firehose - S3
 - reporting to data-lake
- Monitoring with Datadog
- Routes with Express



Our Legacy Search



List your practice on Zocdoc

Sign In/Join

Blurry? Find a doctor.

Specialty Choose a Specialty	in Zip Code or City	Insurance I'll choose my insurance later	Find
Popular Specialties			
Primary Care Doctor			
OB-GYN (Obstetrician-Gynecologist)			
Dermatologist			
Dentist			
Ear, Nose & Throat Doctor (ENT)			
Eye Doctor			
Psychiatrist			

Free Text (Patient Powered) Search



List your practice on Zocdoc

Sign In/Join

Blurry? Find a doctor.

specialty, symptom, condition, treatment, or doctor name

New York, NY

UnitedHealthcare - UnitedHealthcare ...



Types of Intent

pedro rubio

 **Noel Rubio**
Internist · Chicago, IL

Name of doctor

mrj

MRI - Knee

Medical procedure

cardiologist|

Cardiologist (Heart Doctor)

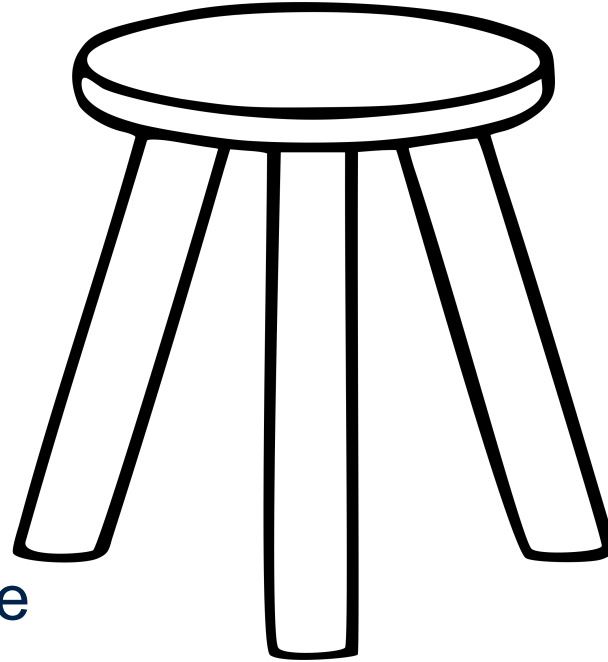
Specialty

math agony|

Are you looking for?
Neurodevelopmental Disabilities Specialist

Symptom

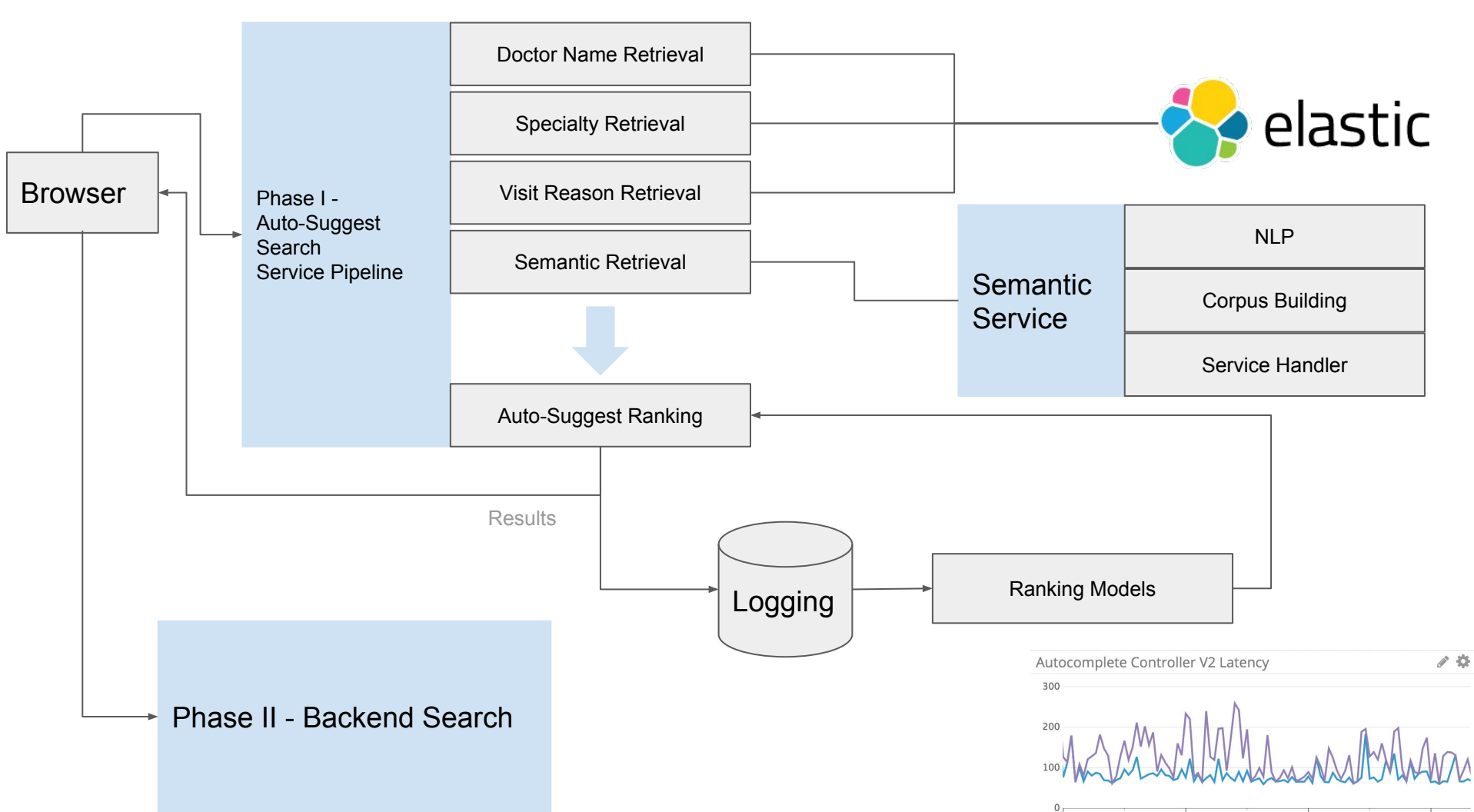
Intent Parsing



Architecture

Design

Machine
Learning



Solving for the Long Tail

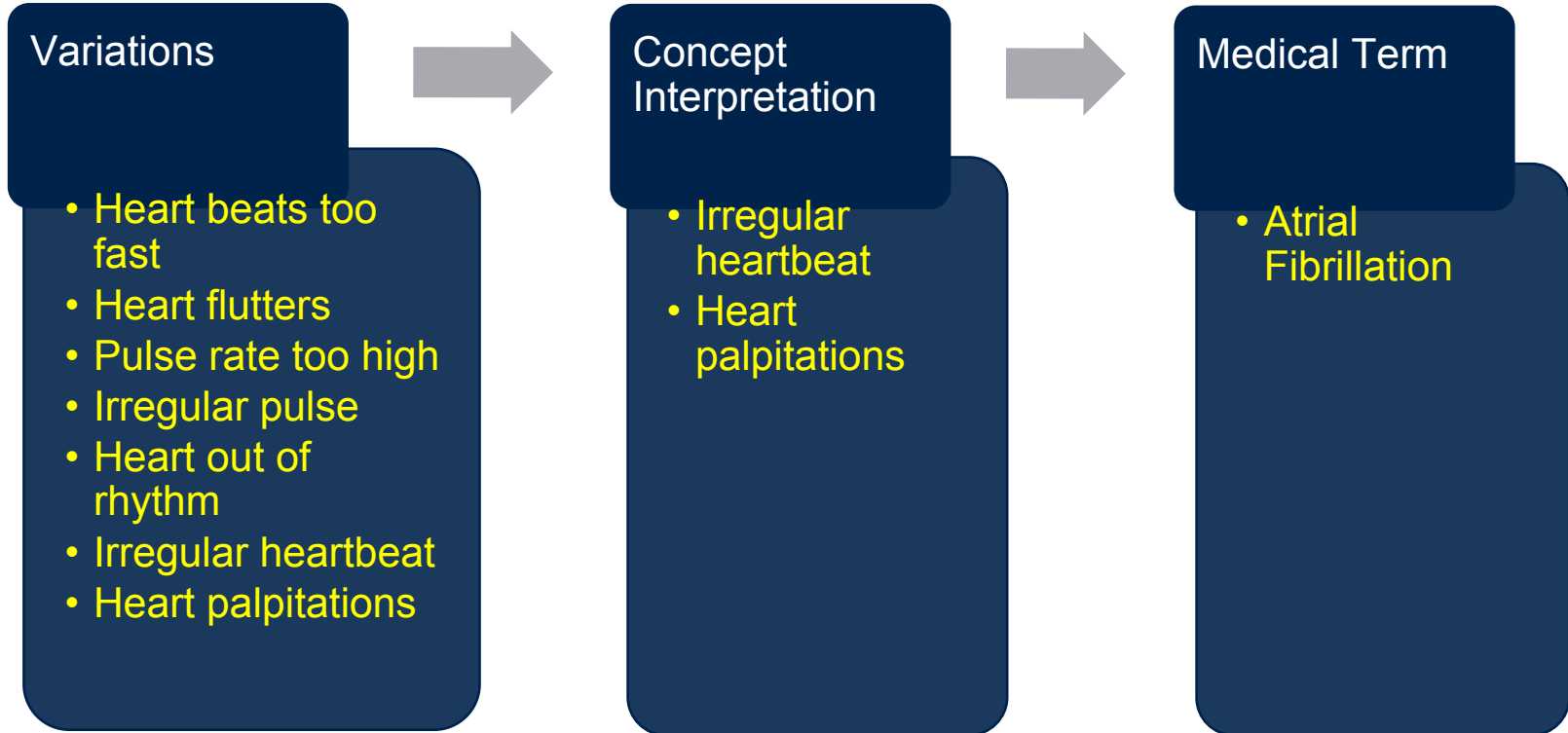
The structured queries comprise a reasonable percent of traffic, but are a minority of total search terms we service.



We use Natural Language Processing (NLP) algos to map unstructured terms into our structured search set.

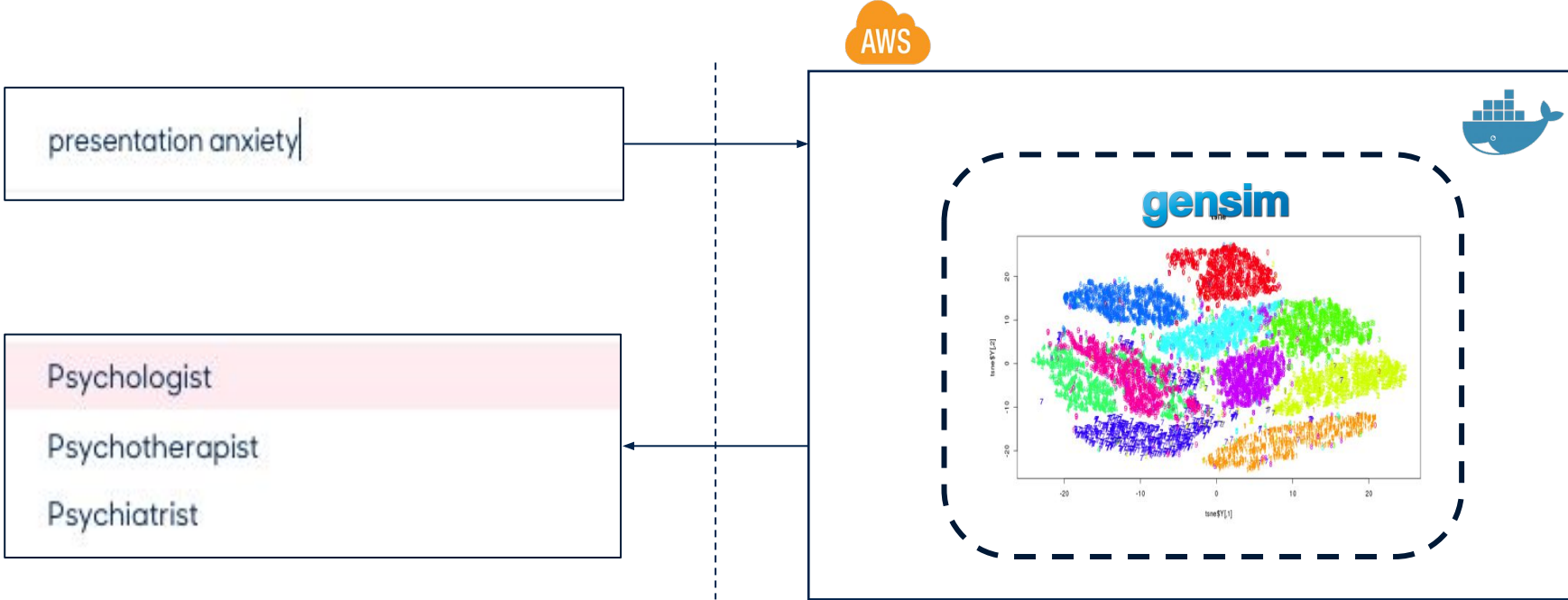
Specialties = $O(10^2)$, Procedures = $O(10^3)$, Names = $O(10^6)$, Other = $O(10^7)$

Different Representations of Same Concept



ZocDoc Semantic Service

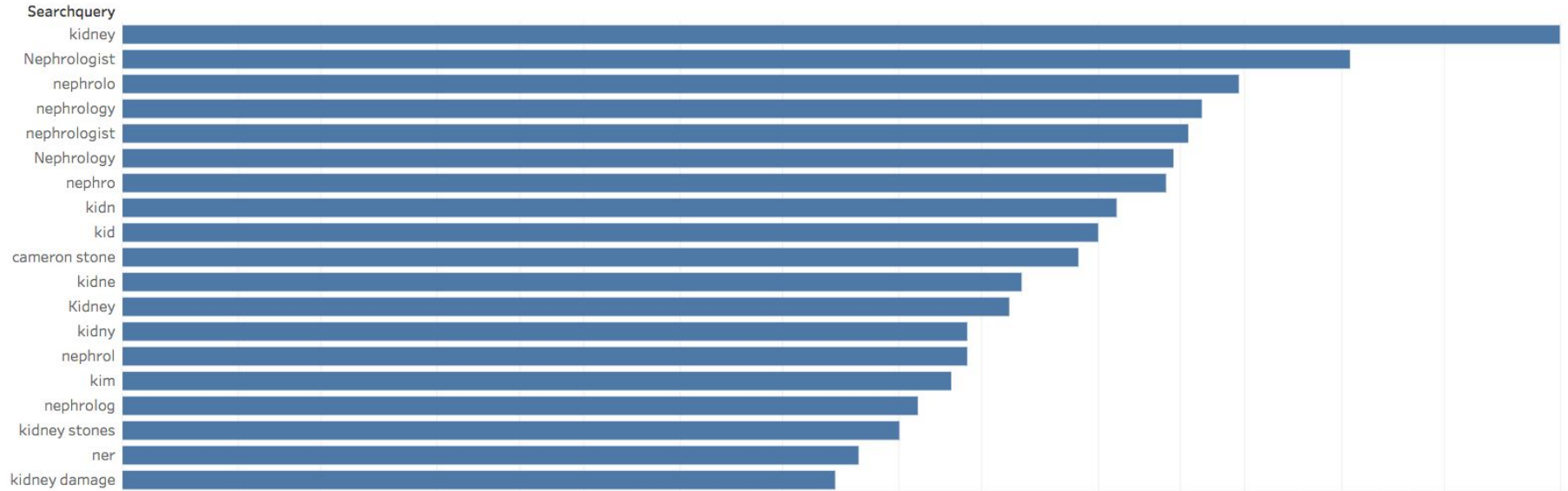
f("presentation anxiety") = [{Specialty = "Psychologist", Relevance = 0.8}, ..., {Specialty = "Psychiatrist", Relevance = 0.7}]



Early Results (And Why You Need to Always Experiment)

Query	Errant Behavior	Desired Behavior	Underlying Issue
flu sho	returns " shoulder surgeon " as the first specialty; returns "flu shot" as the first visit reason	return "pcp" as the first specialty; do not return "shoulder surgeon"	fuzzy string matching on one word of a two-word query
cold	returns " colorectal surgeon " as the first specialty	do not return "colorectal surgeon"	fuzzy string matching on a common word
internal	returns " sleep medicine " as the first specialty	return "internist" as the first specialty	not fuzzy enough matching
broken toe	returns " toxicologist " as the first specialty; returns "broken tooth" as the first visit reason	do not return "toxicologist"; do not return "broken tooth"; return Podiatrist among results	fuzzy string matching on one word
teeth pain	returns " teeth whitening " as the first specialty	do not return "teeth whitening"	exact string matching on one word
irregular heartbeat	returns " irregular menstruation " as the first specialty	do not return "irregular menstruation"	exact string matching on one word

Searches that Lead to “Nephrology”

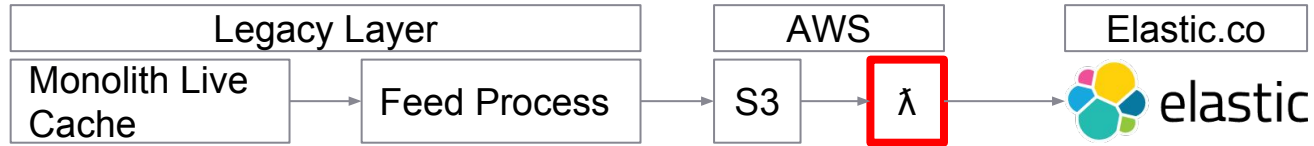


Many patients don't know what a Nephrologist is. They don't need to know to find one now.

How We're Built - The Data

Data - Indexing so we can Search

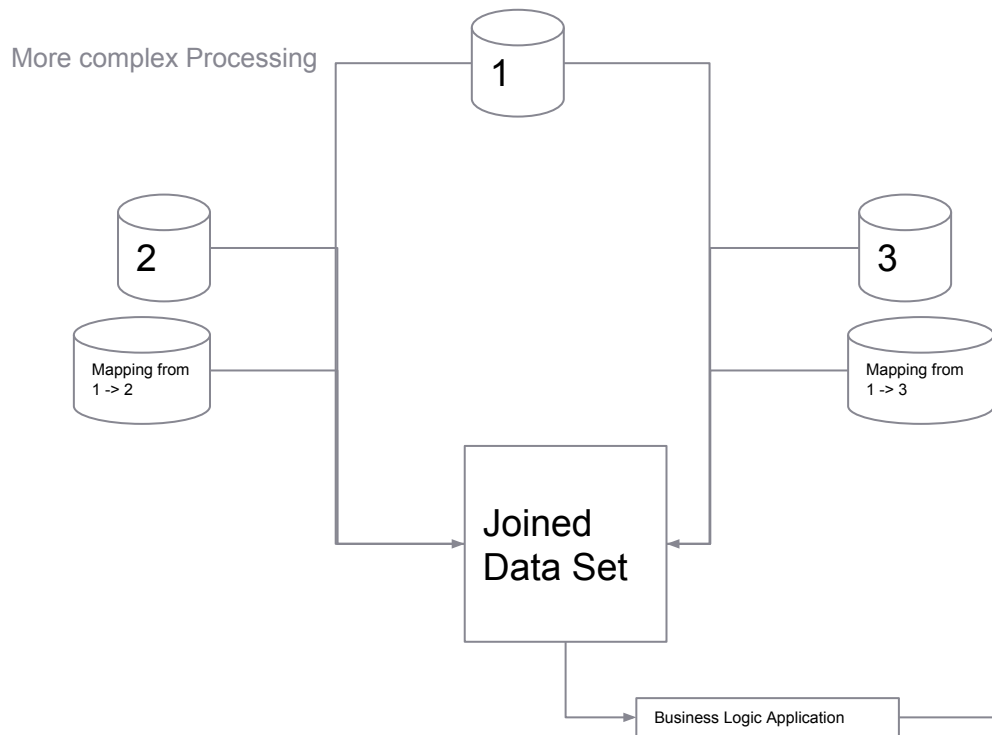
Lesson Learned with Indexing Data



Lambda's act as a mini ETL layer getting the documents ready for our retrieval stage.

- Lambda memory max 1500mb
- Our data much larger
- Manage state in S3 and Elasticsearch
- Complex “stateless” ETL process that transforms this data into the data that we need in Elasticsearch
- Load piecemeal into Elasticsearch
- At the very end, swap alias to use newly uploaded indexes

New ETL - Spark



- Spark - ETL
 - Get over 1500mb limit
 - Get over 5 minute runtime limit
- Easily add more data-sets
- Currently in Databricks
- Plan to migrate to EMR (Elastic Map Reduce)



Data - Event Data So we can
Learn

The Marketplace

The screenshot displays a list of three primary care doctors. Each entry includes a circular profile picture, the doctor's name and title, a star rating, a patient testimonial, the doctor's affiliation and address, and the distance from the user. Availability is shown in a grey box for the first and third doctors, and a specific time slot is shown for the second doctor.

Doctor Name	Title	Rating	Testimonial	Availability	Address	Distance
Dr. Robert Samuelson, MD, FCCP	Primary Care Doctor	★★★★★	"I had a life-long issue that had been previously diagnosed, but the diagnosis..."	Next availability: Mon, Apr 10	Weill Cornell Medicine 38 East 32nd Street, New York, NY 10016	Within 0.5 mile
Dr. Irina Linetskaya, MD	Primary Care Doctor	★★★★★	"Great and knowledgeable doctor!"	4:10 pm	314 West 14th St, New York, NY 10014	Within 1.5 miles
Dr. Anne Nolte, MD	Primary Care Doctor	★★★★★	"My favorite doctor. I usually dislike Manhattan offices because I feel so..."	Next availability: Mon, May 8	Gianna Center, New York - Female Patients Only 15 E. 40th St., New York, NY 10016	Within 1 mile

Goal:

Make it as easy as possible to match the user to the right doctor.

Considerations:

- **How to weight distance vs. availability vs. experience vs. reviews?**
- Does Dr. take this type of patient?
- Are we meeting regulatory requirements?

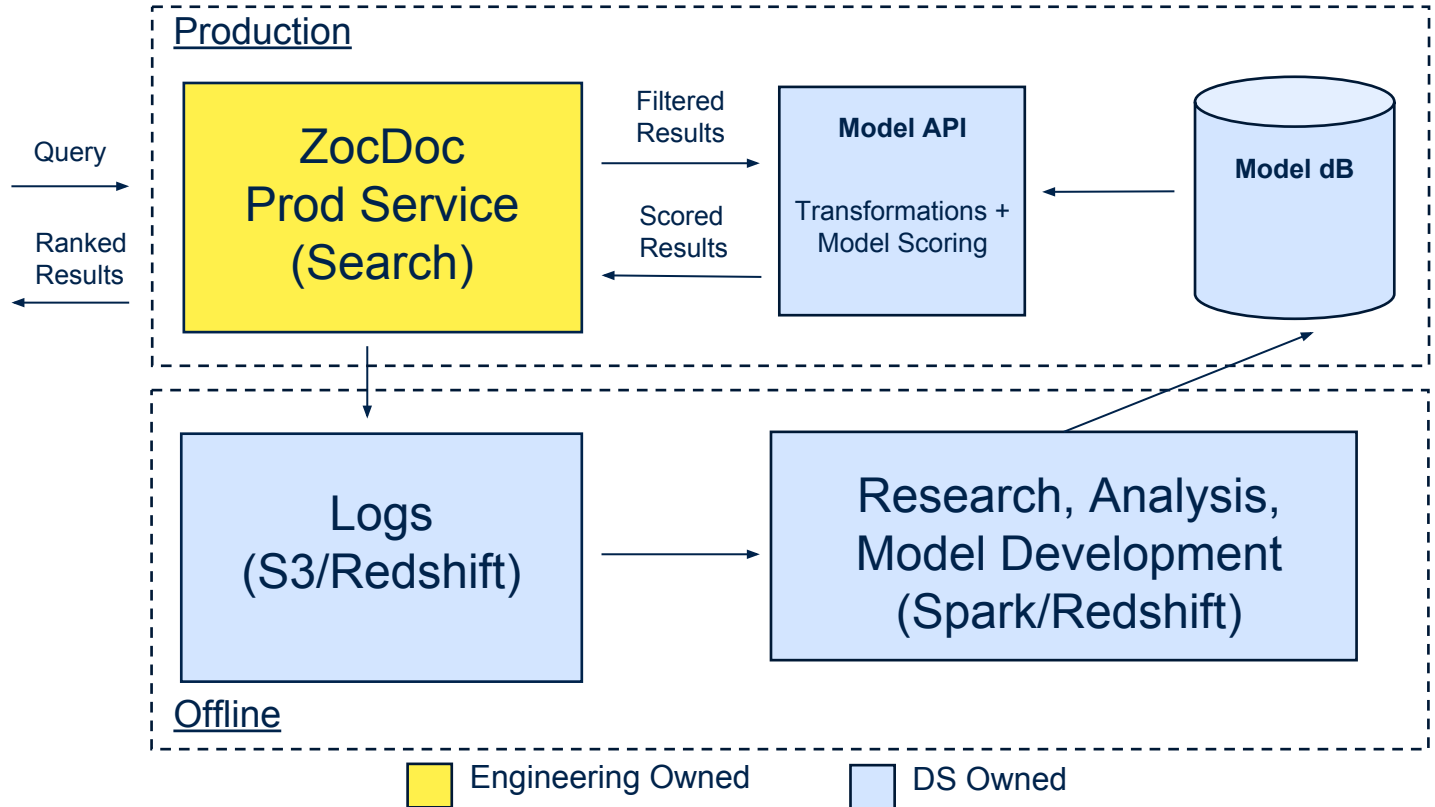
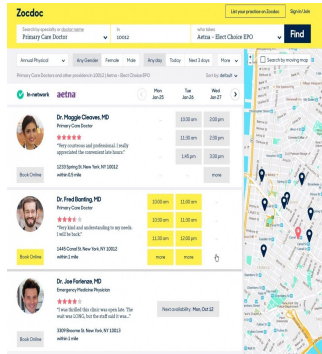
Organizational Optimization

Optimize: algo iteration speed

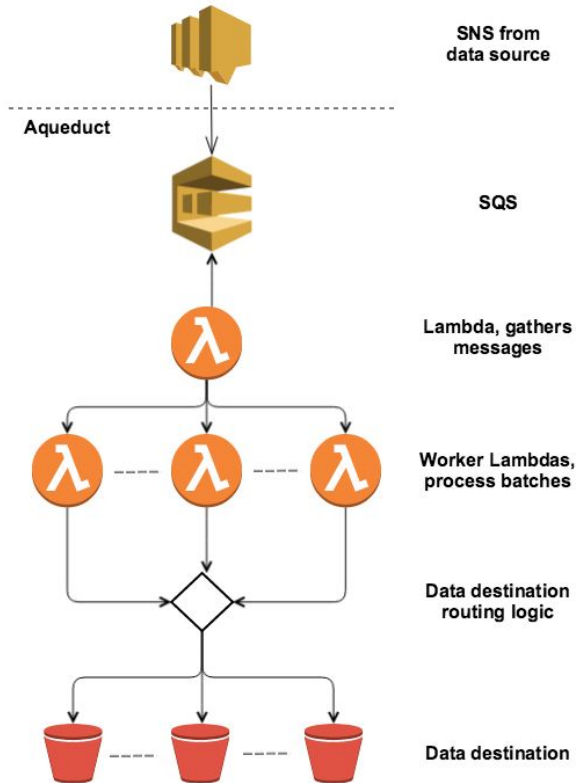
Subject to:

- Org too small to justify full time data scientists within search
- Throwing models over the wall to be implemented doesn't work

Agile Machine Learning



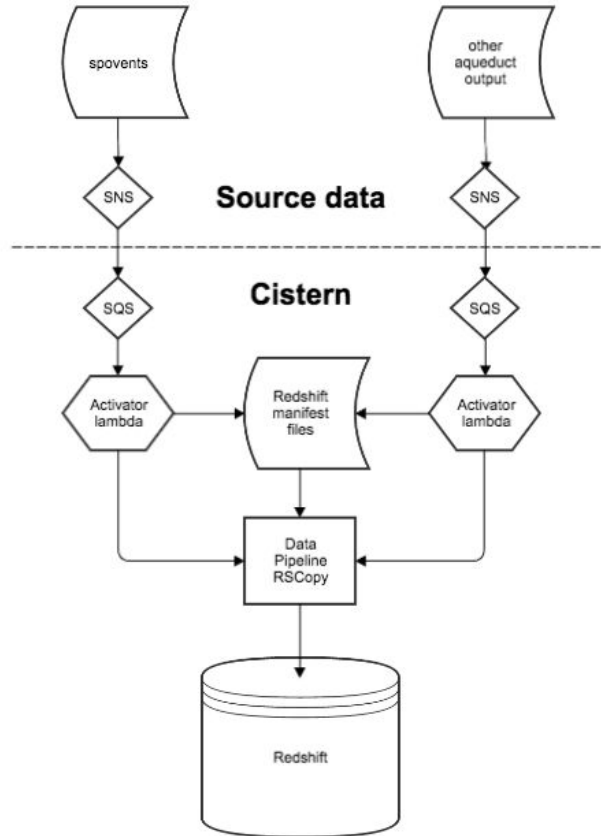
Aqueduct: Filling the Data Lake



Some Data Lake principles:

- Allow producers to easily push data
- Allow data format changes
- Smart ETL to make consumption very easy

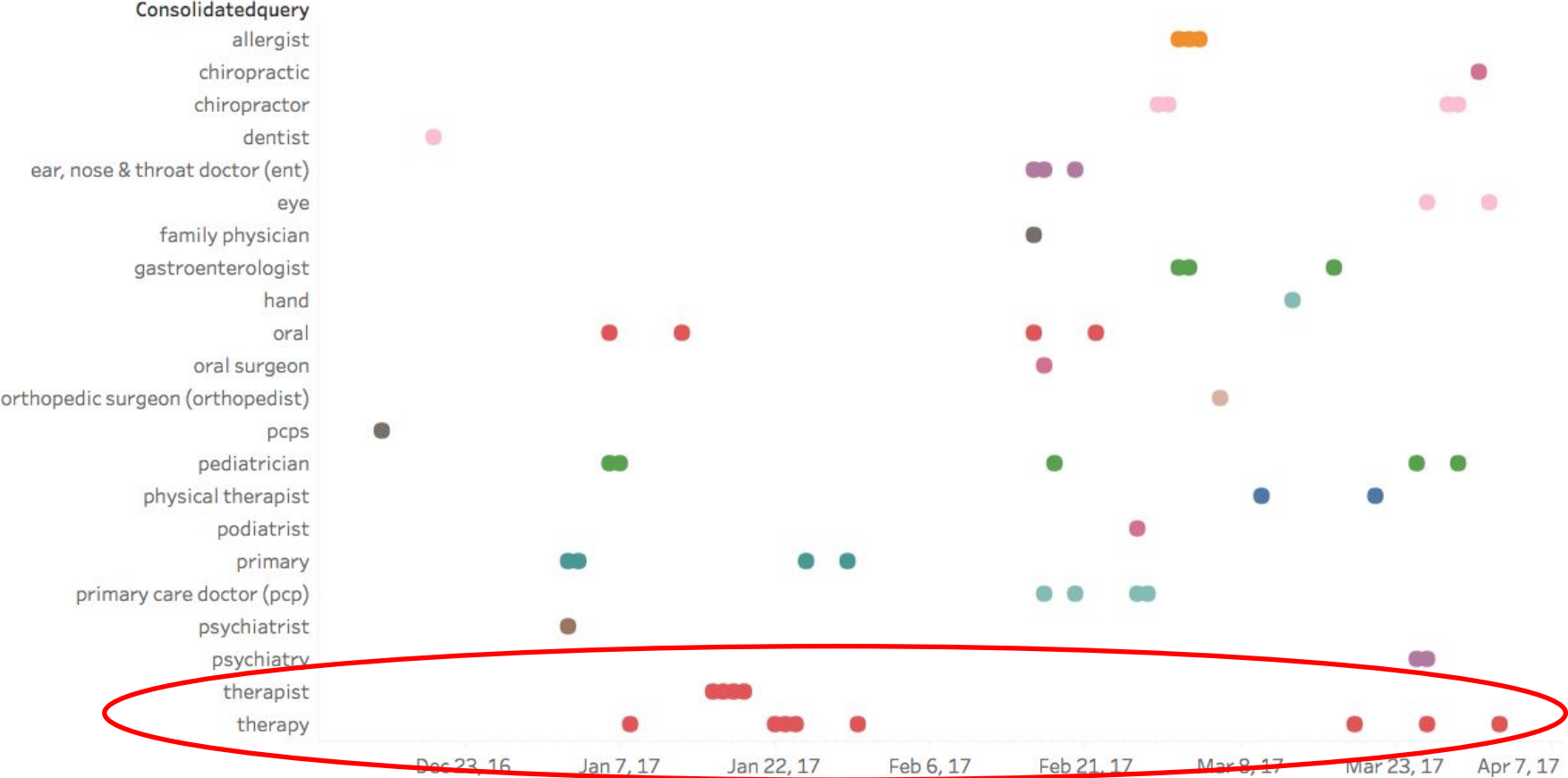
Cistern: Making Datalake Drinkable



- “Raw” data lake good for exploratory research (we use Spark)
- “Clean” data lake better for analytics and quick exploration

Data - Insights

Search Queries on Days with Unusually High Volume



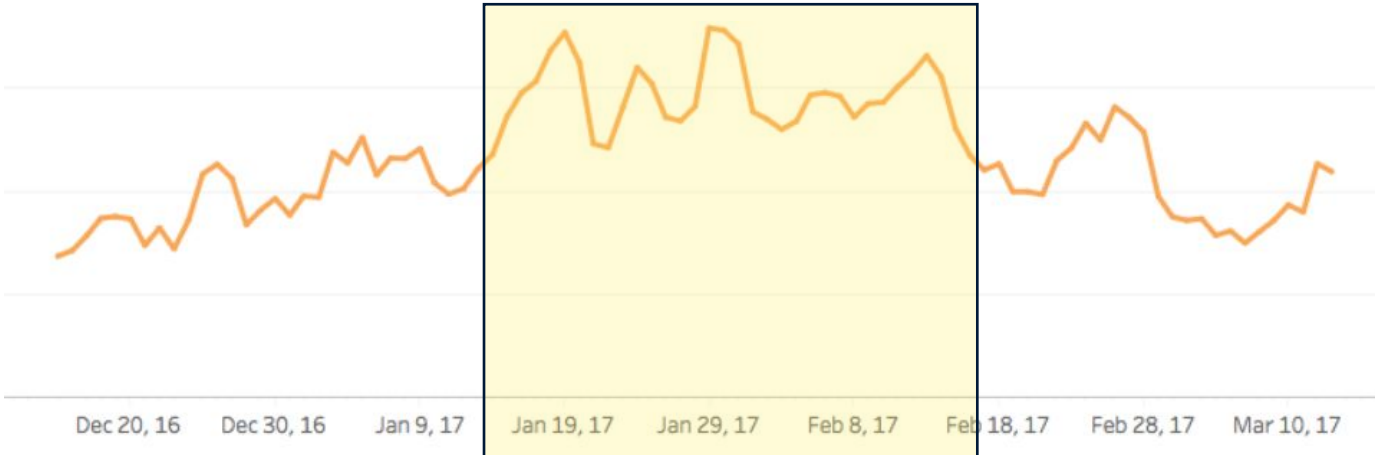


U.S. | [NEW YORK](#) | [METRO MONEY](#)

New Yorkers Seek Ways to Cope With New World Order

In left-leaning New York City, stress-relief specialists—from acupuncturists to barkeepers and therapists—say the fledgling Trump administration has triggered a surge in demand.

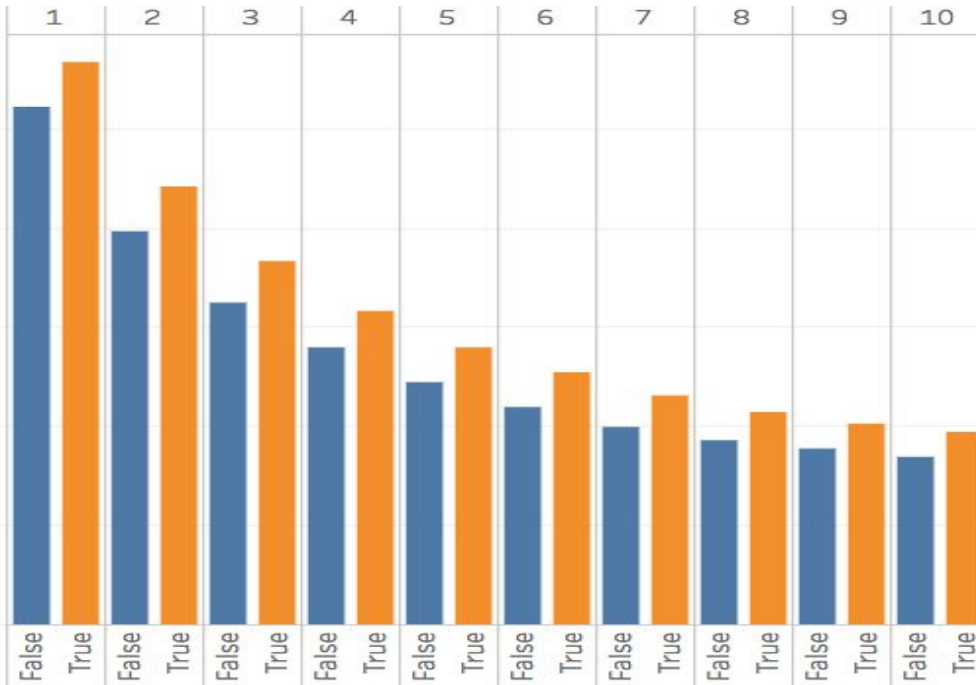
Searches for Therapy/Therapist on Zocdoc



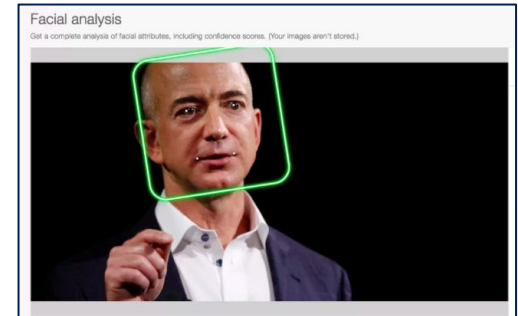
We've got our fingers on the pulse of public health trends.

How Much is that Smile Worth?

Click Conversion by Search Rank and DrIsSmiling



We're exploring AWS Rekognition to research what drives user interest in Dr. profiles.





Psst. Want a job?

Zodoc has won workplace awards from Crain's NYC, Modern Healthcare, and Fortune.

[Check our careers page](#)



Thank you and Questions!