Hops
A Beer Recommender

Presented by Kelly Ochikubo

Pipeline

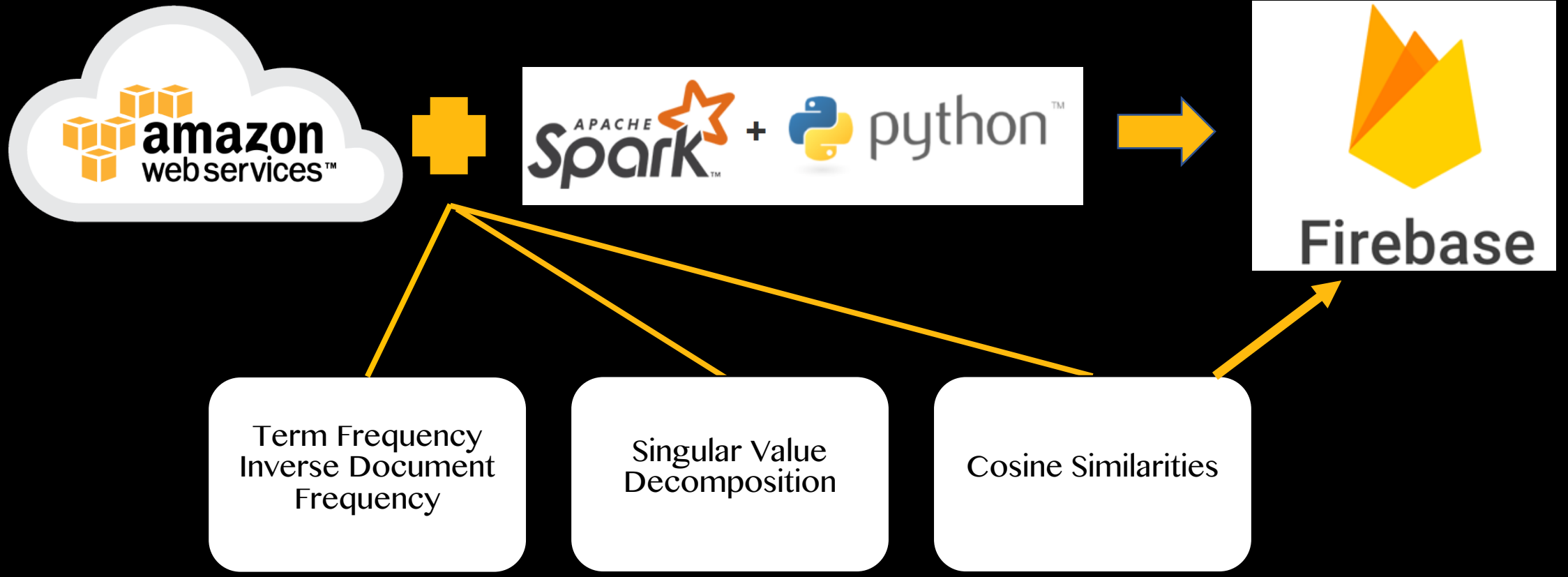
Data Source → Data Cleaning → Model → Storage → Web UI

amazon web services™

mongoDB

Pipeline

Data Source → Data Cleaning → Model → Storage → Web UI



Latent Semantic Indexing

Document 1

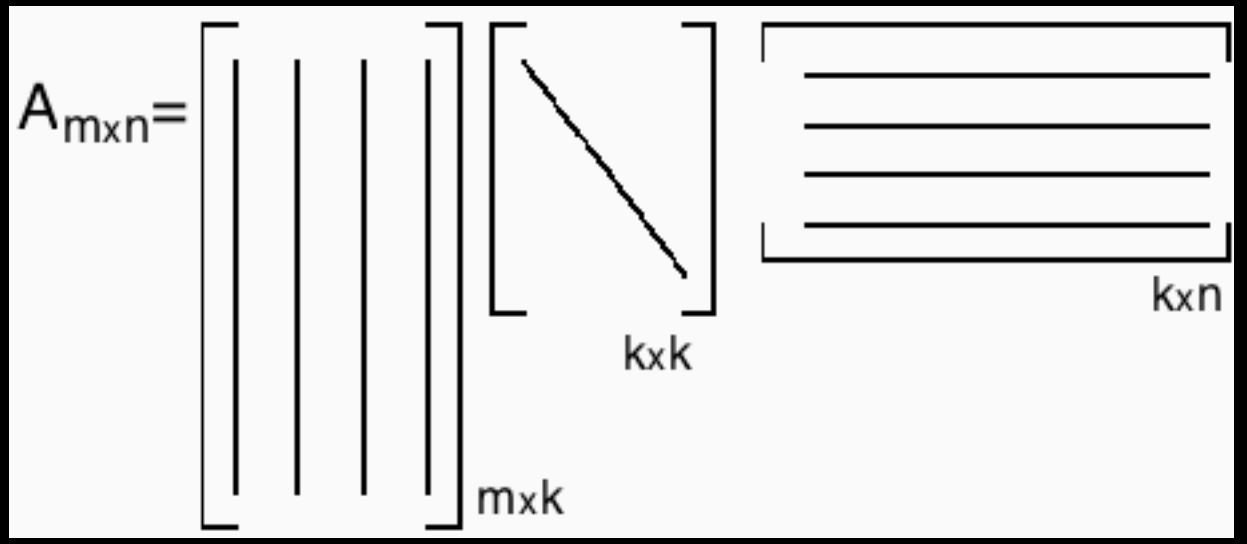
Term	Term Count
this	1
is	1
a	2
sample	1

Document 2

Term	Term Count
this	1
is	1
another	2
example	3

TF-IDF

+



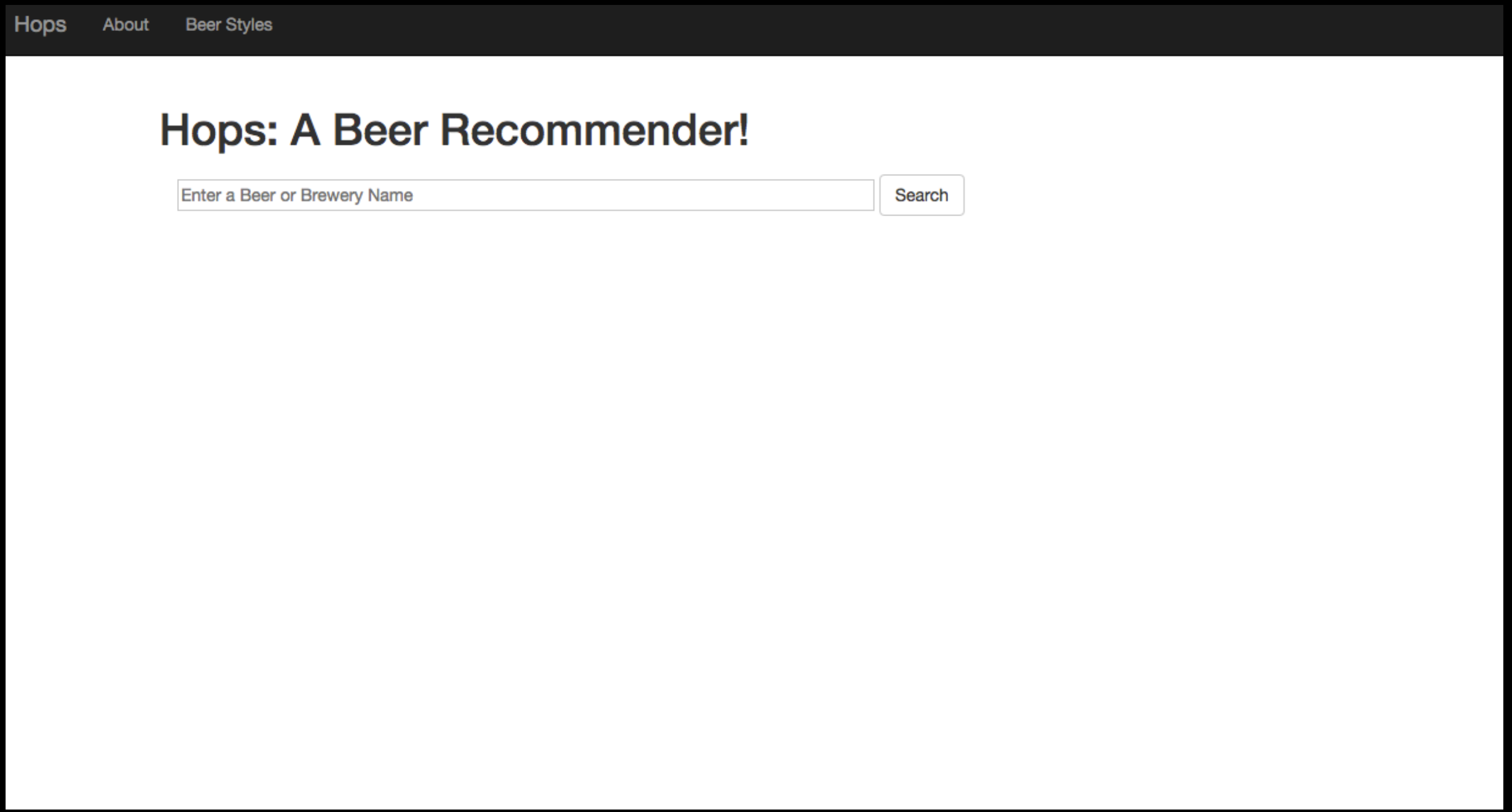
SVD

Pipeline

Data Source → Data Cleaning → Model → Storage → Web UI

Query Firebase database and return results to be displayed on the web app

Hops Web Application


Hops
About
Beer Styles
Hops: A Beer Recommender!
Enter a Beer or Brewery Name
Search

Hops: A Beer Recommender!

Kona

Search

- Lavaman Red Ale - **Kona** Brewing Co.
- Pipeline Porter - **Kona** Brewing Co.
- Lemongrass Luau - **Kona** Brewing Co.
- Fire Rock Pale Ale - **Kona** Brewing Co.
- Wailua Wheat - **Kona** Brewing Co.

Hops About Beer Styles

Hops: A Beer Recommender!

Wailua Wheat - Kona Brewing Co. Search

Wailua Wheat by Kona Brewing Co.

Style: American Pale Wheat Ale

Top Features: passion, wheat, kona, passionfruit, hawaii, tropical, hawaiian

Top 5 Beers Recommendations

Passion Fruit Prussia

4 Hands Brewing Co.

Style: Berliner Weissbier
Top Features: passion, berliner, lacto, passionfruit, tart, tartness, sour

LA-31 Grenade

Bayou Teche Brewery

Style: Fruit / Vegetable Beer
Top Features: passion, teche, bayou, louisiana, wheat, crawfish, skunked

5 Lizard

5 Rabbit Cerveceria

Style: Witbier
Top Features: lime, passion, passionfruit, coriander, witbier, wheat, rabbit

Tropical Pale Ale

Boulevard Brewing Co.

Style: American Pale Ale (APA)
Top Features: passion, passionfruit, grapefruit, tropical, boulevard, mango, juicy

LA-31 Passionne

Bayou Teche Brewery

Style: American Pale Wheat Ale
Top Features: passionfruit, passion, wheat, soapiness, invigorating, rudzud, tapered

Challenges & Takeaways

- Firebase
  - Limited ways to query database, limit to space on free tier
- Apache Spark
  - No PySpark implementation for SVD (only exists for Scala/Java)
    - Time to learn Scala!
  - To take maximum advantage of Spark's capabilities, one must think in "Map Reduce"
  - Troubleshooting can be difficult especially if using PySpark because the issue could lie either in your high level code or the source code
    - Documentation is your BFF (obviously!)

Future Work

- Model Improvements
  - Take into consideration other features (e.g. location)
- Extensions
  - Search beers by keywords (i.e. caramel, passionfruit)
  - Apply same LSA concept to beer styles and see what can be learned about the similarities and differences between them
  - Add visualizations to the web application to show user these similarities and differences

Application

- Text Mining/Natural Language Processing can be difficult in terms of the curse of dimensionality
- As the dimensionality increases, the volume of space increases so fast that the data becomes sparse
- Matrix Factorization techniques such as Singular Value Decomposition can be used to reduce dimensionality while maintaining as much “power” as possible

THANK YOU

KELLY.OCHIKUBO@GMAIL.COM