# Strategically leveraging Amazon reviews to be victorious in the eCommerce battlefield

"Your most unhappy customers are your greatest source of learning." -- Bill Gates

Why Review Mining Matters?

- Intelligence: Reviews tell you what your audience likes and helps companies to learn more about their target demographics.
- Integrity: Amazon actively defends the integrity of its reviews by prosecuting those involved with publishing fake reviews.
- Flexibility: Apply customized filtering criteria to capture differences & trends across score ranges, product categories & locations.
- Scalability: Process vast amounts of data efficiently and with open source programs such as Python, SQL, Apache Spark & MLlib.

Process Flow
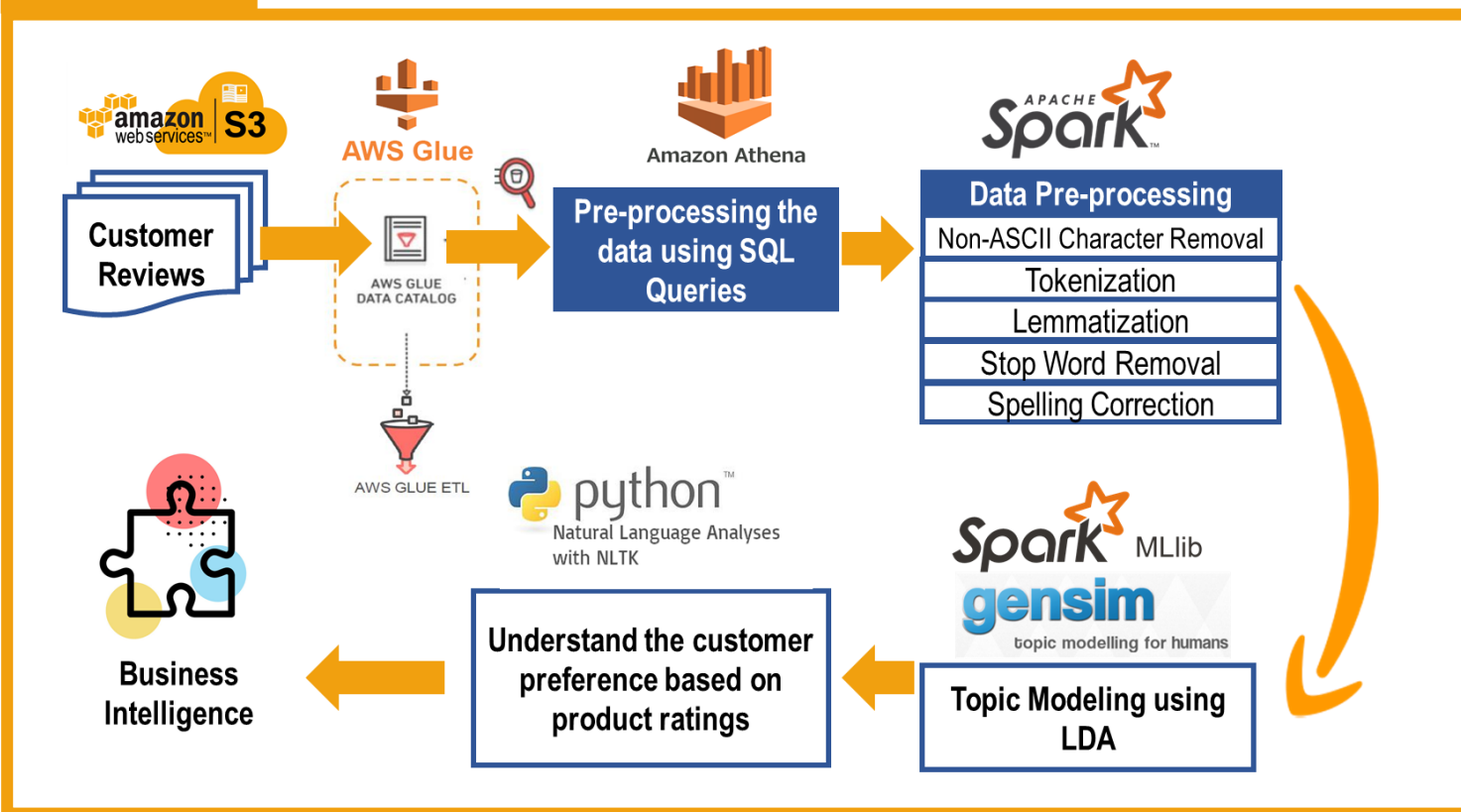
Process Explanation:

We used 160 million Amazon review data spanning from 2003 to 2015 as the raw data set.

Each review is tokenized by splitting into individual words (and bigrams). Stop words are removed and then each word is lemmatized and stemmed. Then the words are turned into spelling correction.

Next, we applied topic modeling using LDA to extracts key topics and themes from a large corpus of text.

(Latent Dirichlet Allocation (LDA) is a generative and probabilistic model that can be used to automatically group words into topics and documents into a mixture of topics.)

Contact: Shakti Kothari - kotha045@umn.edu, Robert Lu - lu000175@umn.edu, Rutwik Rath - rath0126@umn.edu, Rakesh Gupta - Gupta487@umn.edu, Jonas Kimtai - kimt0001@umn.edu, Adam Chang - chan1773@umn.edu

Sample Results

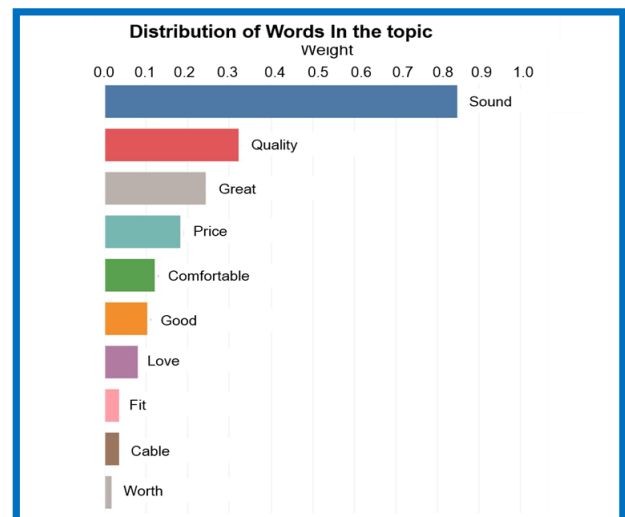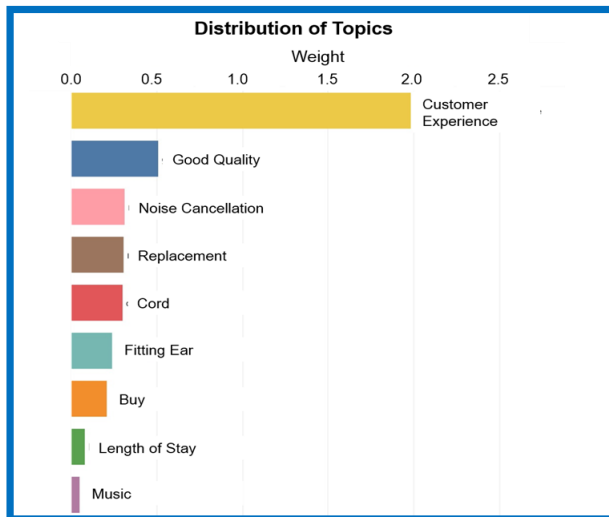Amy TOP 500 REVIEWER

★★★★★ Great sound quality and amazing noise cancellation
December 7th 2014
Color: Black | Style: Headphones only | **Verified Purchase**

> "This headset is of **great quality** and **sound**. It **fits** well to my ears and doesn't fall off even while I am running. The best thing about this headset is its **noise cancellation** feature. I can hear to even the most intricate **beats** of the music. Its surreal to get such high quality headsets at such great **price**, Its totally **worth** it"

Distribution of Topics
Weight

Topic	
Customer Experience	~2.0
Good Quality	~0.5
Noise Cancellation	~0.3
Replacement	~0.25
Cord	~0.25
Fitting Ear	~0.2
Buy	~0.2
Length of Stay	~0.2
Music	~0.15

Distribution of Words In the topic
Weight

Word	
Sound	~0.85
Quality	~0.3
Great	~0.28
Price	~0.2
Comfortable	~0.13
Good	~0.12
Love	~0.1
Fit	~0.08
Cable	~0.07
Worth	~0.05

Target Audiences and Business Impacts

eCommerce Merchants

- Optimize the product mix on their platforms;
- Identify and resolve any existing customer service issues;
- Detect any shipping and delivery issues with their product in a timely manner

Product Developers

- Identify any issues with existing products;
- Provide a better estimator of the product return rate;
- Better design next iteration of existing products

Marketers

- Create better targeted ad that highlight the strong selling points of a product

Contact: Shakti Kothari - kotha045@umn.edu, Robert Lu - lu000175@umn.edu, Rutwik Rath - rath0126@umn.edu, Rakesh Gupta - Gupta487@umn.edu, Jonas Kimtai - kimt0001@umn.edu, Adam Chang - chan1773@umn.edu