











# Aspect Based Sentiment Analysis

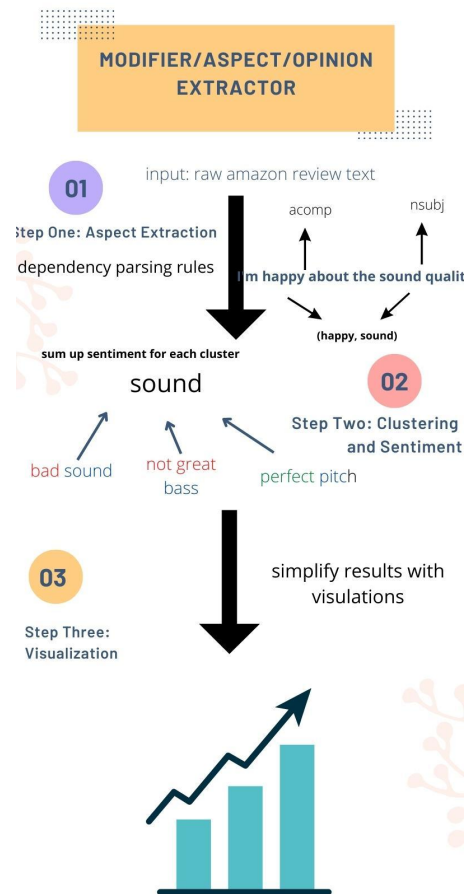
Making sense of messy amazon text  
data with a simple tool

Aspect Category	Sentiment
 Customer Service	
 Value	
 quality	
 design	

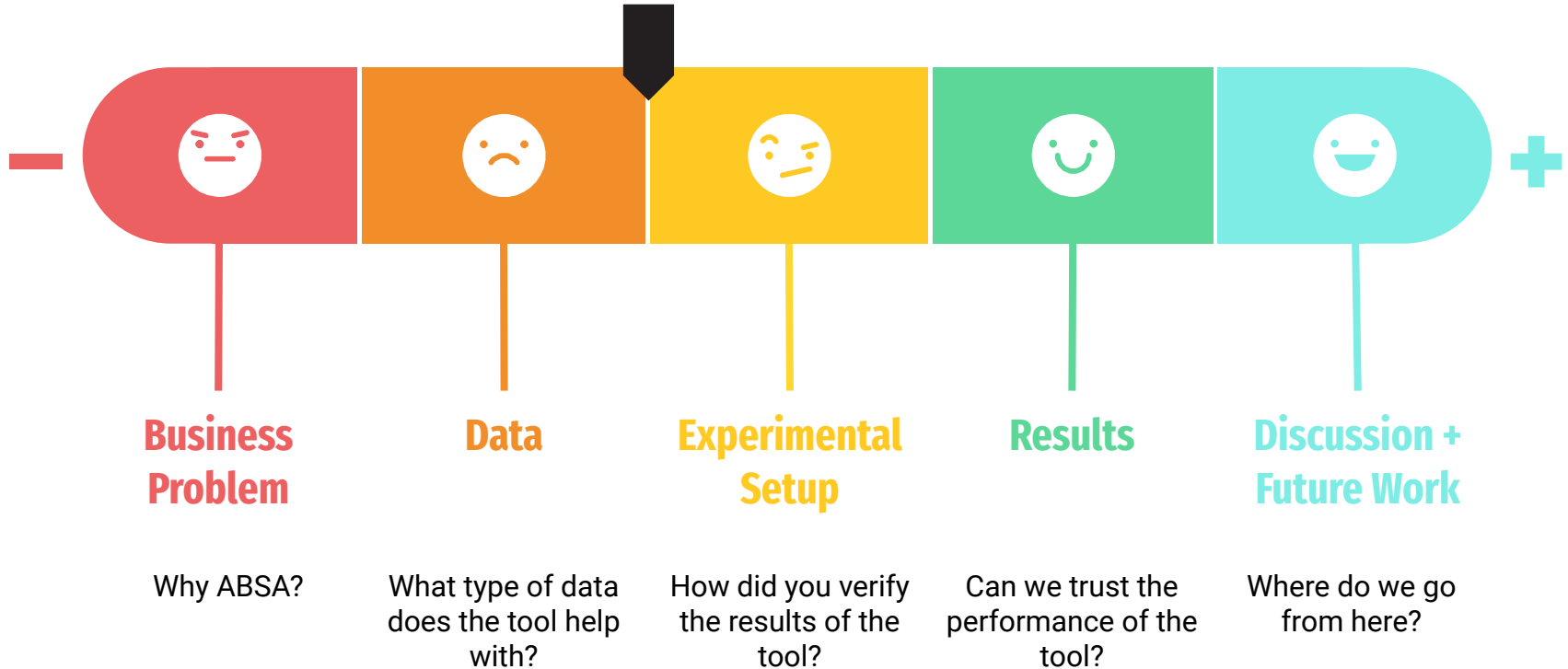
Before we begin...

# Summary

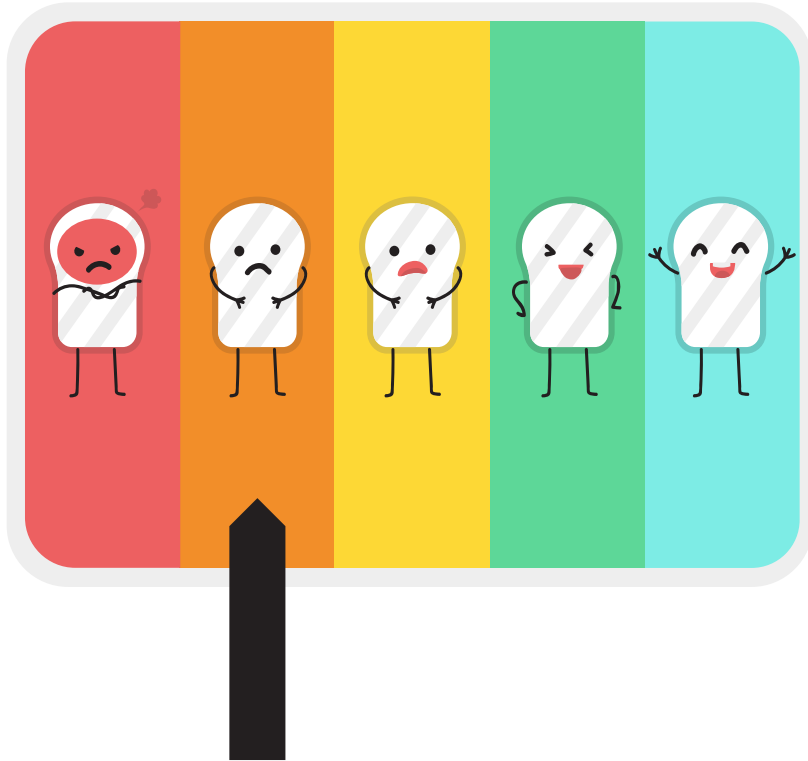
- Creating an out of box aspect/opinion/sentiment triplet extractor using spaCy large english model and parsing logic/pos tagging capabilities
  - Clustering aspects to simple categories with unsupervised machine learning
  - Visualizing total sum of sentiment in clustered aspect graphs
- Verifying model performance with crowdsourced labels from Amazon Turk
- Verifying experimental setup



# Outline



# Why Aspect Based Sentiment Analysis (ABSA)



**E-commerce**

Today, most e-commerce website designs include a section where their customers can post reviews for products or services

**Informal Reviews**

There is potentially a disconnect from the amazon review ratings, and the overall sentiment of the body text explaining the review

**Large Internet Text Data**

It is often difficult to efficiently get useful data from a large collection of text data

**Simplified**

Transformation of informal review data can help make informative decisions

# The Data

The Amazon Customer Reviews (Product Reviews) contains over 130+ million customer reviews available to researchers in TSV files in the amazon-reviews-pds S3 bucket in AWS US East Region, as per the provided readme file. The reviews were collected from 1995 to 2015. See the provided link for associated metadata. This project focuses on the dataset given by pulling “[https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon\\_reviews\\_us\\_Electronics\\_v1\\_00.tsv.gz](https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Electronics_v1_00.tsv.gz)” from the S3 bucket.

```
In [4]: 1 df = pd.read_csv('data/df_electronics.tsv', sep='\t')
        2 df.groupby('product_id').count().sort_values(by='star_rating').tail(20)
```


Out[4]:

	marketplace	customer_id	review_id	product_parent	product_title	product_category	star_rating	helpful_votes	total_votes	vine	verified_purch
	product_id										
	B004LTEUDO	3997	3997	3997	3997	3997	3997	3997	3997	3997	
	B004HHICKC	4213	4213	4213	4213	4213	4213	4213	4213	4213	
	B001TH7GSW	4866	4866	4866	4866	4866	4866	4866	4866	4866	
	B008KVUAGU	5015	5015	5015	5015	5015	5015	5015	5015	5015	
	B003WGRUQQ	5072	5072	5072	5072	5072	5072	5072	5072	5072	
	B002MAP7TU	5295	5295	5295	5295	5295	5295	5295	5295	5295	
	B001GTT0VO	5580	5580	5580	5580	5580	5580	5580	5580	5580	
	B0052SCU8U	5756	5756	5756	5756	5756	5756	5756	5756	5756	
	B00316263Y	5813	5813	5813	5813	5813	5813	5813	5813	5813	
	B00D5Q75RC	6062	6062	6062	6062	6062	6062	6062	6062	6062	
	B004QK7H18	6536	6536	6536	6536	6536	6536	6536	6536	6536	
	B00F5NE2KG	6688	6688	6688	6688	6688	6688	6688	6688	6688	
	B0019EHU8G	7586	7586	7586	7586	7586	7586	7586	7586	7586	
	B000WYVBRO	7835	7835	7835	7835	7835	7835	7835	7835	7835	
	B0001FTVEK	8793	8793	8793	8793	8793	8793	8793	8793	8793	
	B0012S4APK	9359	9359	9359	9359	9359	9359	9359	9359	9359	
	B003EM8008	9766	9766	9766	9766	9766	9766	9766	9766	9766	
	B0002LSR78	11166	11166	11166	11166	11166	11166	11166	11166	11166	

Product\_id “B0001FTVEK”, or Sennheiser-RS120-Wireless-Headphones - was chosen to showcase the triplet extractor as it had a large amount of verified reviews and a pair of headphones seemed like a reasonable choice for aspect based sentiment analysis.

# The Data/Experimental Setup

- The aspect/opinion pairs are created through unsupervised machine learning, more specifically a k-means clustering algorithm by scikit-learn with 4 clusters. This means that some “new” data was generated with this project. Below you can see an example of an html sheet a Turk Worker was assigned when labeling aspect/opinion pairs for this project.
- In total, 410 workers submitted 6107 non-null aspect/opinion pairs for sentiment intensity pertaining to 1438 unique aspects. Duplicate pairs of aspect/opinion pairs were included to inspect variance of submission from human labels and machine labels for each opinion pair. No qualifications or screening was put in place before the workers were chosen, but I did review sections of the data and accept or reject what seemed reasonable.
- All aspects/opinion pairs generated from Product\_id “B0001FTVEK” from previous slide

 **Previewing Answers Submitted by Workers**  
This message is only visible to you and will not be shown to Workers.  
You can test completing the task below and click "Submit" in order to preview the data and format of the submitted results.

InstructionsShortcutsWhat sentiment does this text convey?

Instructions

×

the noun and adjective should typically describe a pair of headphones from an amazon review  
Choose the primary sentiment conveyed by each modifier/aspect pair, from very positive to very negative

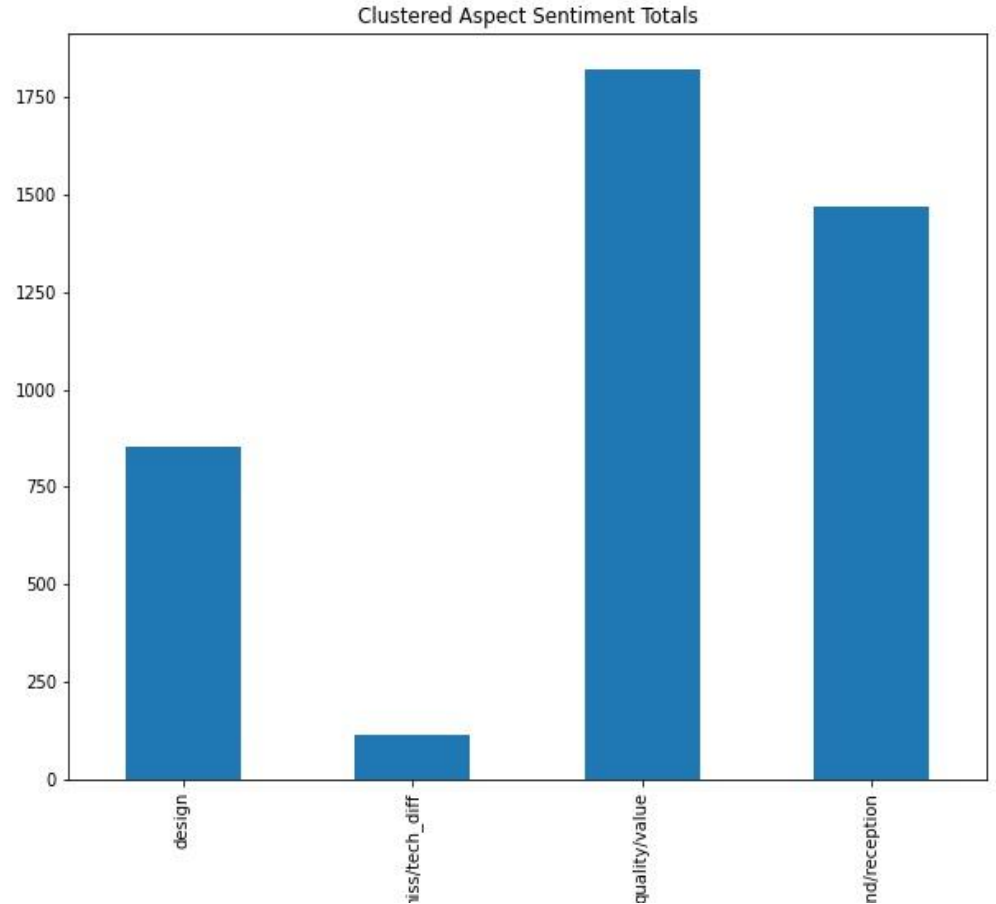
\$(Modifier) \$(Aspect)

Select an option

Very Positive	1
Positive	2
Neutral	3
Negative	4
Very Negative	5
N/A	6

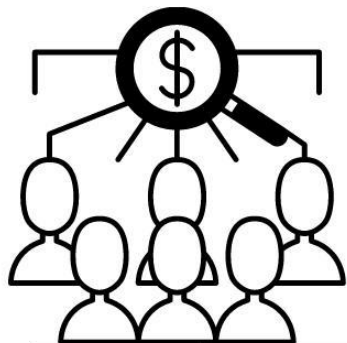
# Results

- a lot of positive sentiment quality/value and sound/reception categories as compared to the design and hiss/tech\_diff categories.
- The hiss category is low enough that it should be the major focus for the company to funnel resources in response to customer demand for improving their product.
- focus first on fixing the hiss
- Look at improving Design next such as batteries and cradle



# Results

## Comparison of Turk Data to Extractor Data



**VS**



	Neutral	Positive	Very Positive	Negative	Very Negative
Precision	36.20%	51.64%	35.76%	35.52%	23.86%
Accuracy	23.58%	7.76%	3.68%	2.78%	0.41%

Human labeled tuples	Sum_of_squares
('wireless', 'headphones')	232.75
('second', 'set')	4.75
('clear', 'sound')	26
('sound', 'quality')	870
('great', 'headphones')	19
('second', 'pair')	11.2
('good', 'sound')	60.75
('rechargeable', 'batteries')	4.666666667
('good', 'range')	16.66666667
('great', 'quality')	58.75
('Sound', 'quality')	4.5
('great', 'product')	34.75
('great', 'range')	18.66666667
('great', 'sound')	60.66666667
('good', 'quality')	83
('long', 'time')	11.2



# Discussion and Future Work

1) Due to a lack of funding from grants and some other issues in experimental design, I feel the most appropriate next step would be to repeat the experiment with more carefully collected labeled data. Increasing the reward per label and the requirements for submission (such as proving a proficiency in english) I believe can substantially improve the quality of the human labeled data and reduce the variance in this data significantly. Sourcing reliable test data is the number one priority in continuing future work for this project.

2) Integrate into AWS for scaling

3) integrate into a SQL server for data management

4) incorporate a flash based UI for viewing results at scale



# Thank You!

Dylan Dey

[Ddey2985@gmail.com](mailto:Ddey2985@gmail.com)

