Clothing Recommendation

Team members:
Maria Stella Vardanega, Risheng Guo,
Ting-Hsuan Chen, Jiazhi Jia, Kratik Gupta

Table of Contents

Business Problem

Dataset

EDA

The Methods

Analytical findings

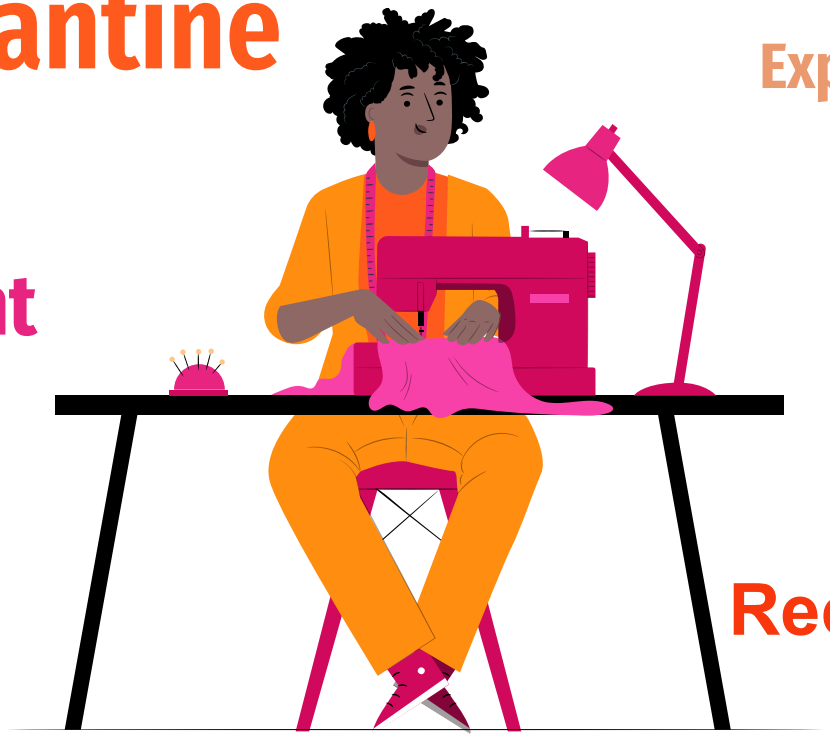
Conclusions

Business Problem

Return

Post - Quarantine

Expensive



Opportunity Cost

PARTY!!!

convenient

Disadvantage: can't try

Recommendation

Dataset

Columns

15

Rows

192,544

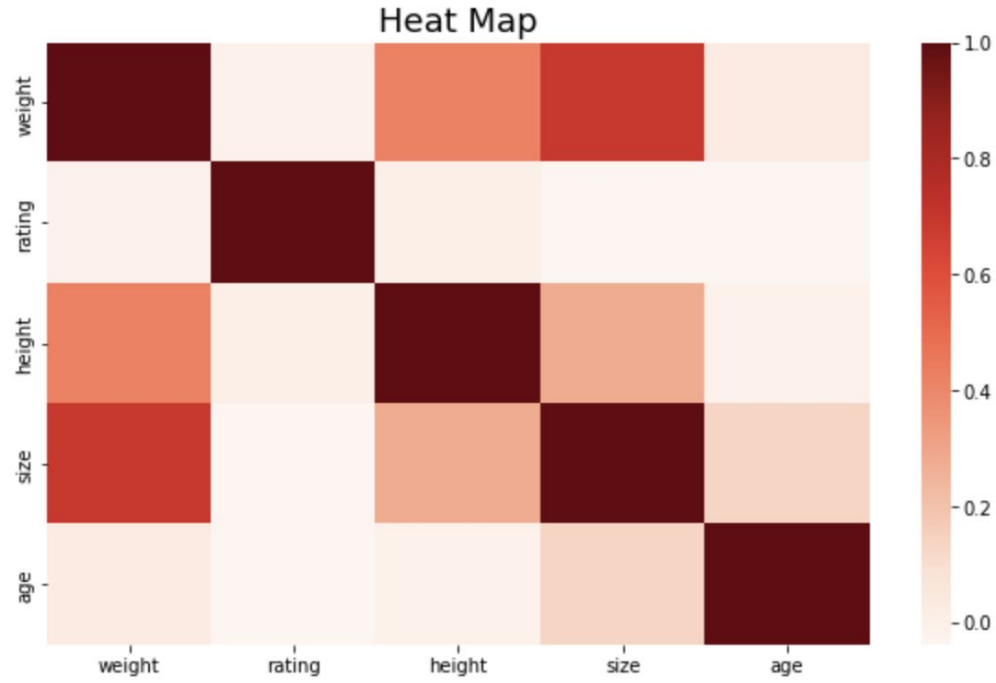
Confidential Source

Kaggle:
https://www.kaggle.com/rmisra/clothing-fit-dataset-for-size-recommendation?select=modcloth_final_data.json

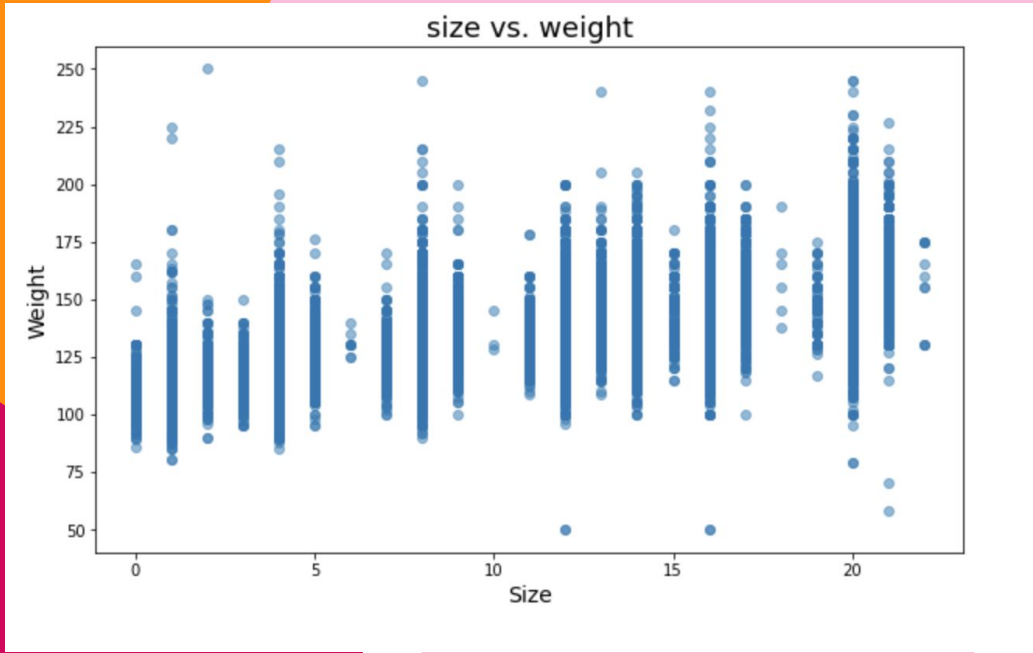
Our features

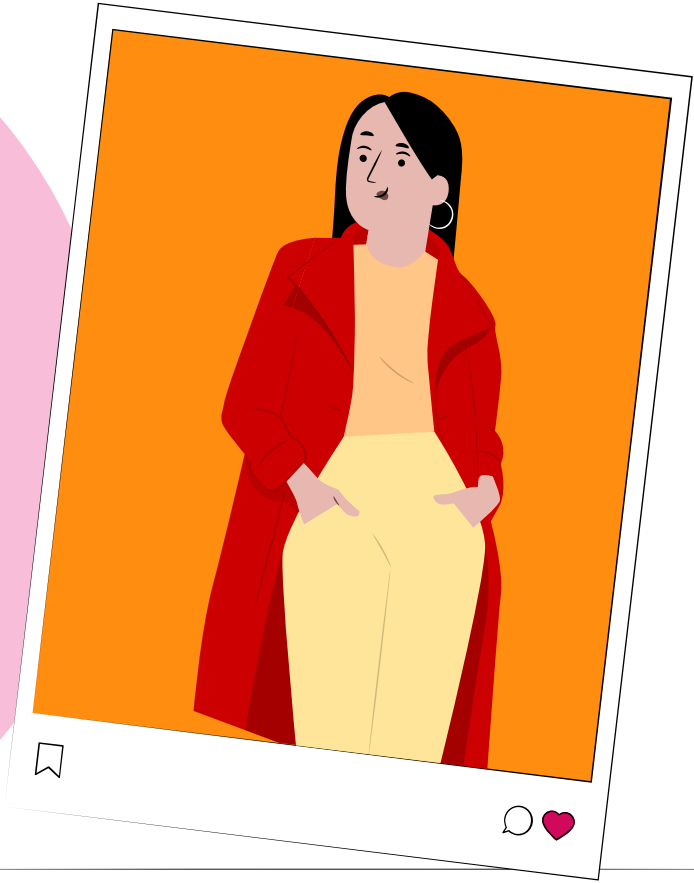
	fit	weight	body type
	user id	rating	review summary
	bust size	rented for	category
	item id	review_text	height
	size	age	review date

Exploratory Data Analysis

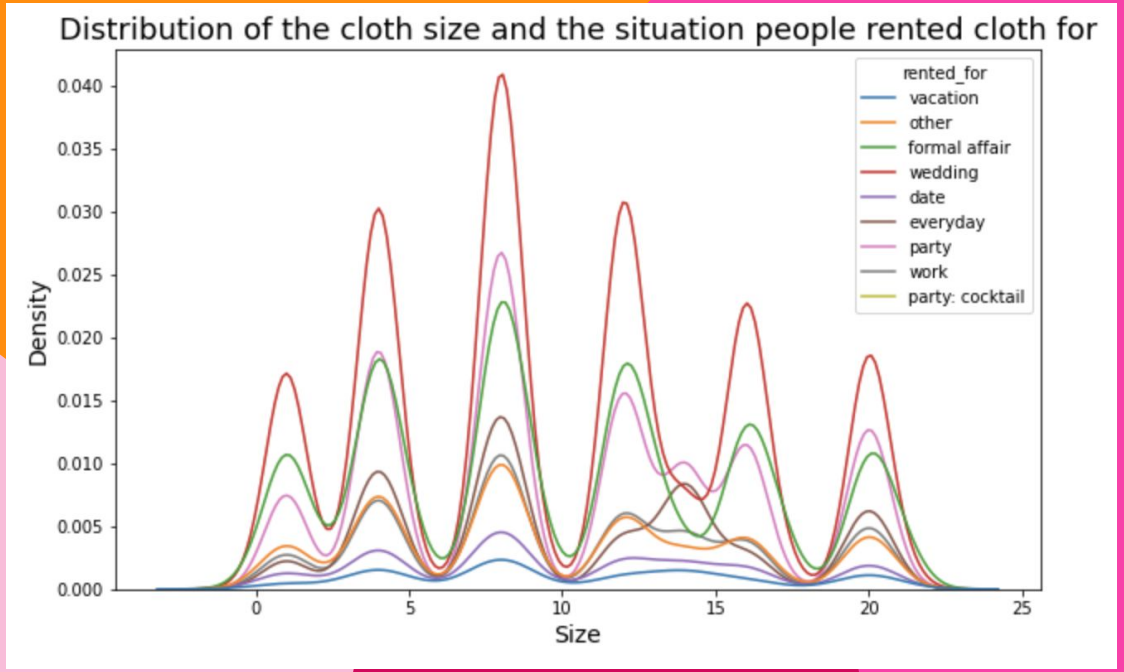


Exploratory Data Analysis

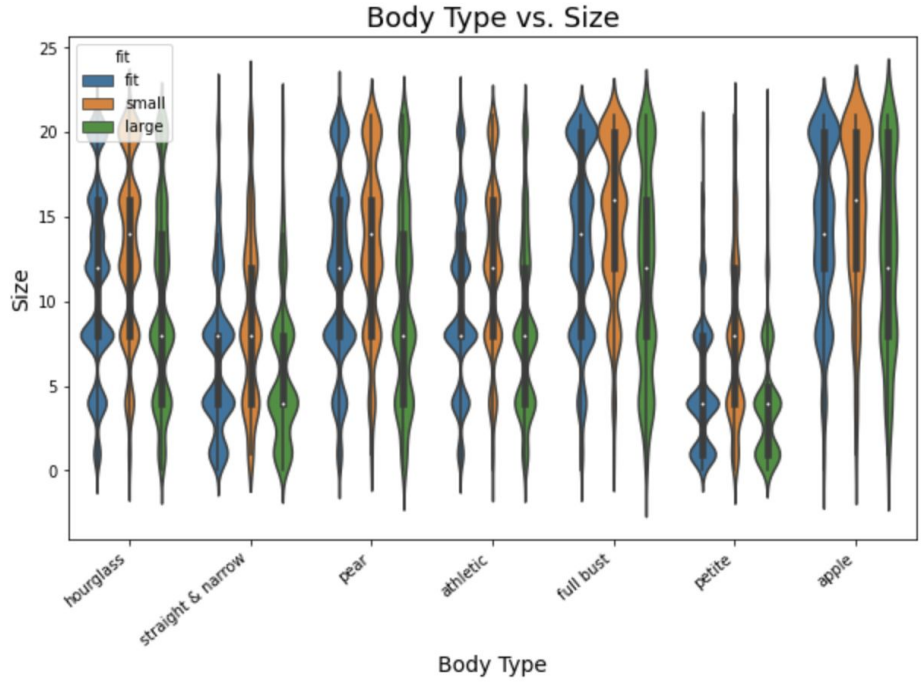




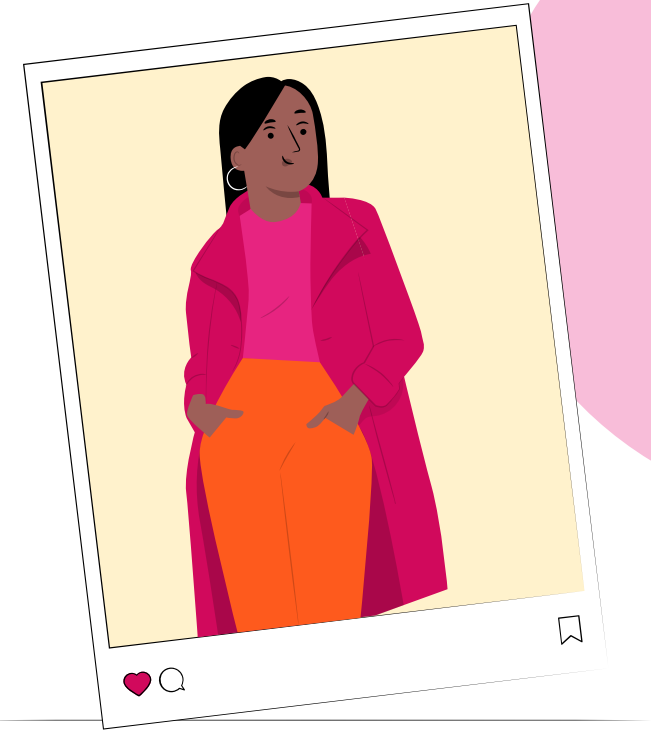
Exploratory Data Analysis

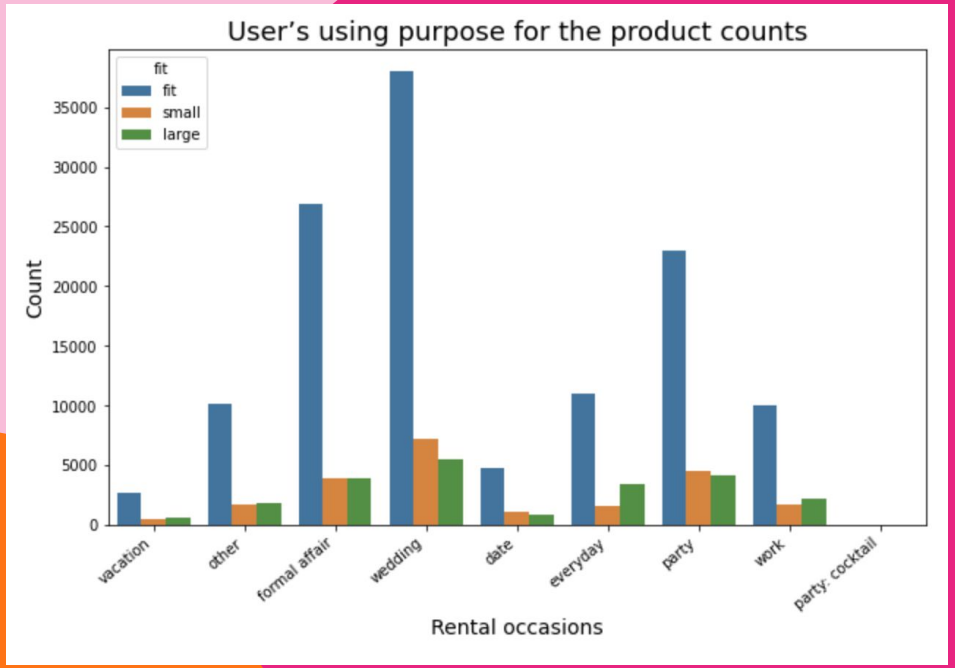


Exploratory Data Analysis

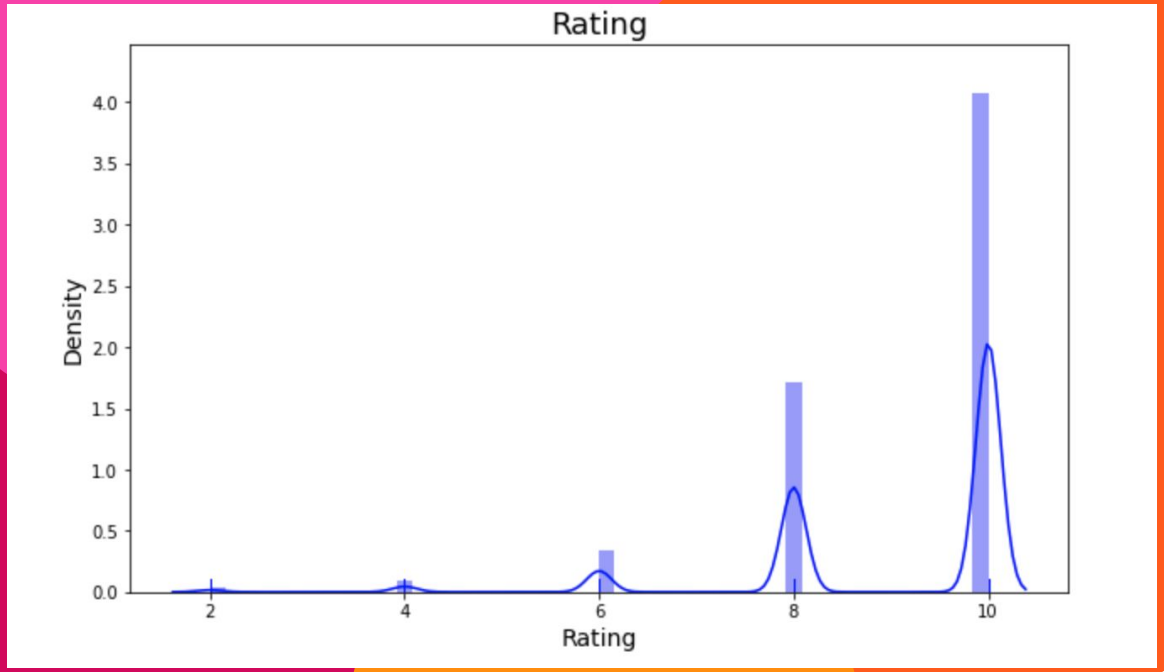


Exploratory Data Analysis





Exploratory Data Analysis



The Methods

Text analysis on reviews

afinn packages
TF-IDF Vectorizer
Spacy

Cluster analysis

Hierarchical Clustering
K-Means Clustering

Clothing Item and Size recommendation

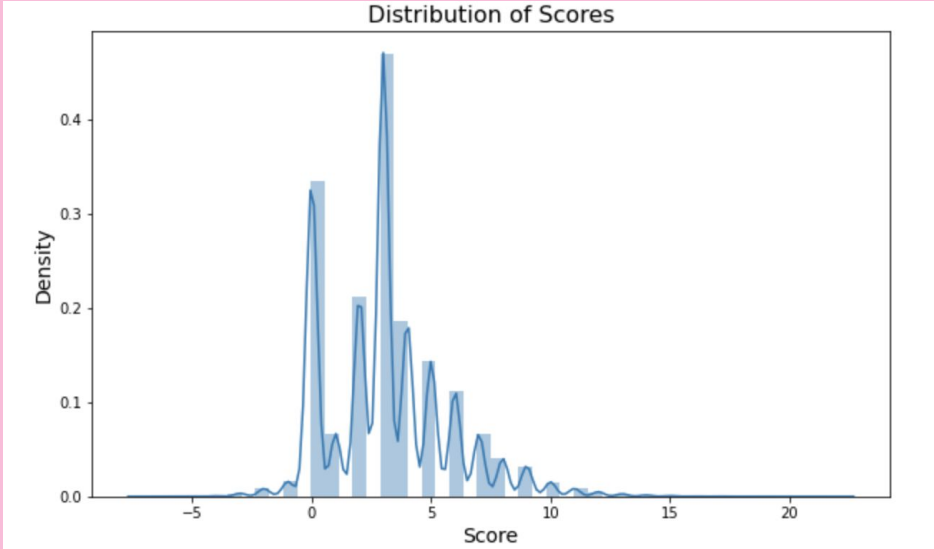
K-Nearest Neighbors
Random Forest Regressor

Analytical Findings

Text Analysis

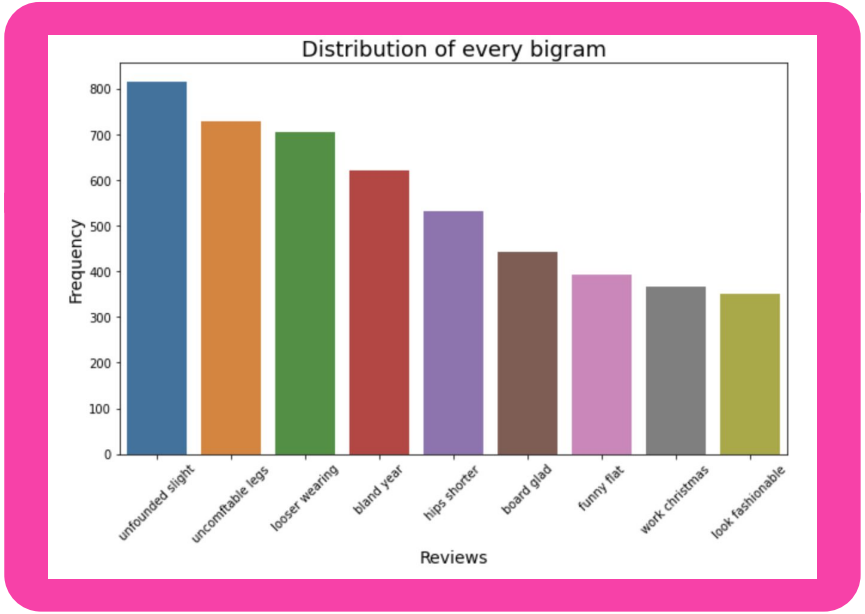
Most people tend to give a moderate review (afinn score between 0 to 5) to the clothes they rent.

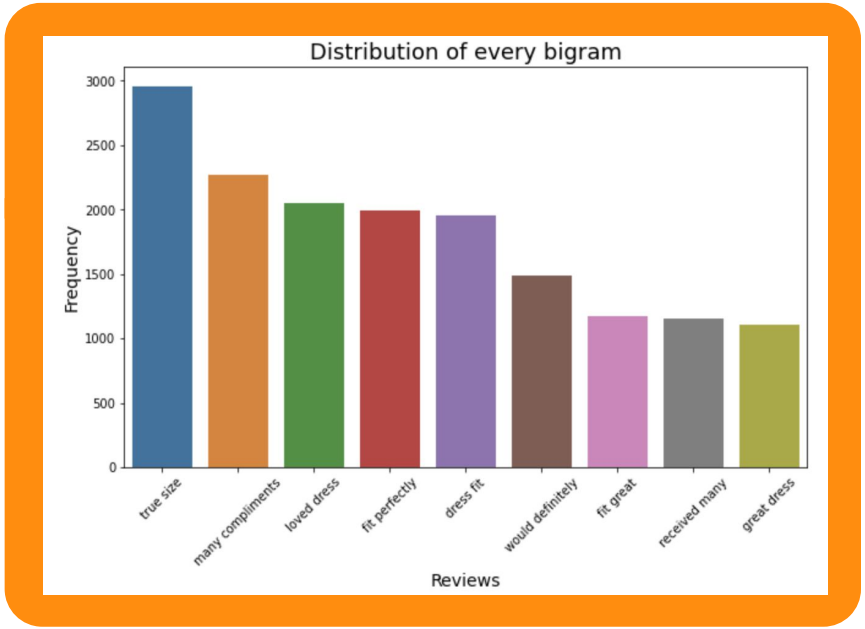
Our goal is to find out why some people would criticize or praise after receiving their clothes, and improve the quality of service.



Most frequent bigram in both situations

By using TF-IDF vectorizer to fit and transform the review text with a high (or low) score, we can get the frequency distribution of every bigram, which we think represent the opinions of customers.



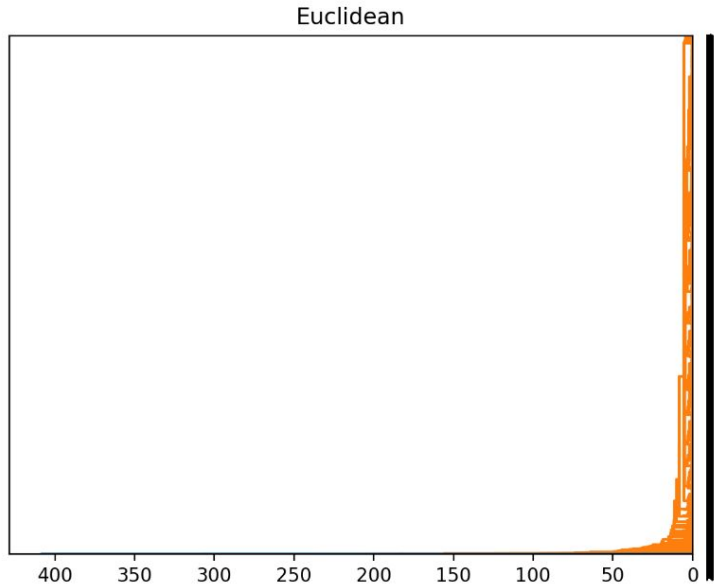


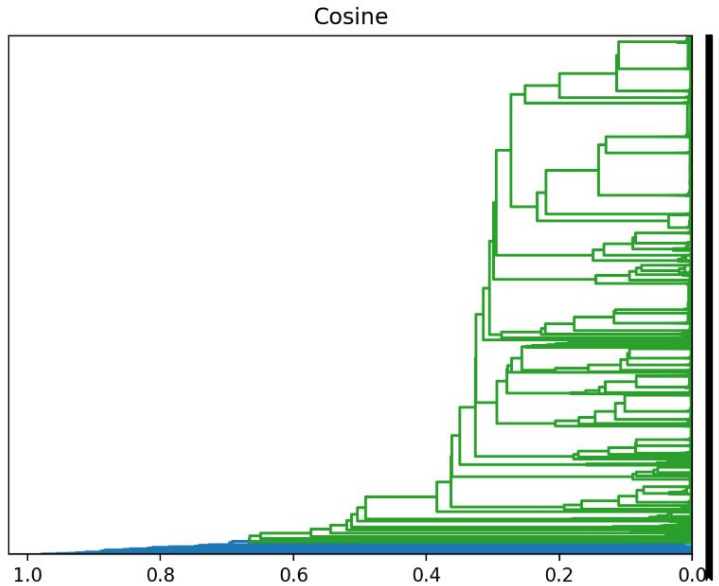
Word Cloud

The most important thing for customers is whether clothes fit them well. So it’s necessary to conduct a prediction on the size of clothes before new customers place orders.

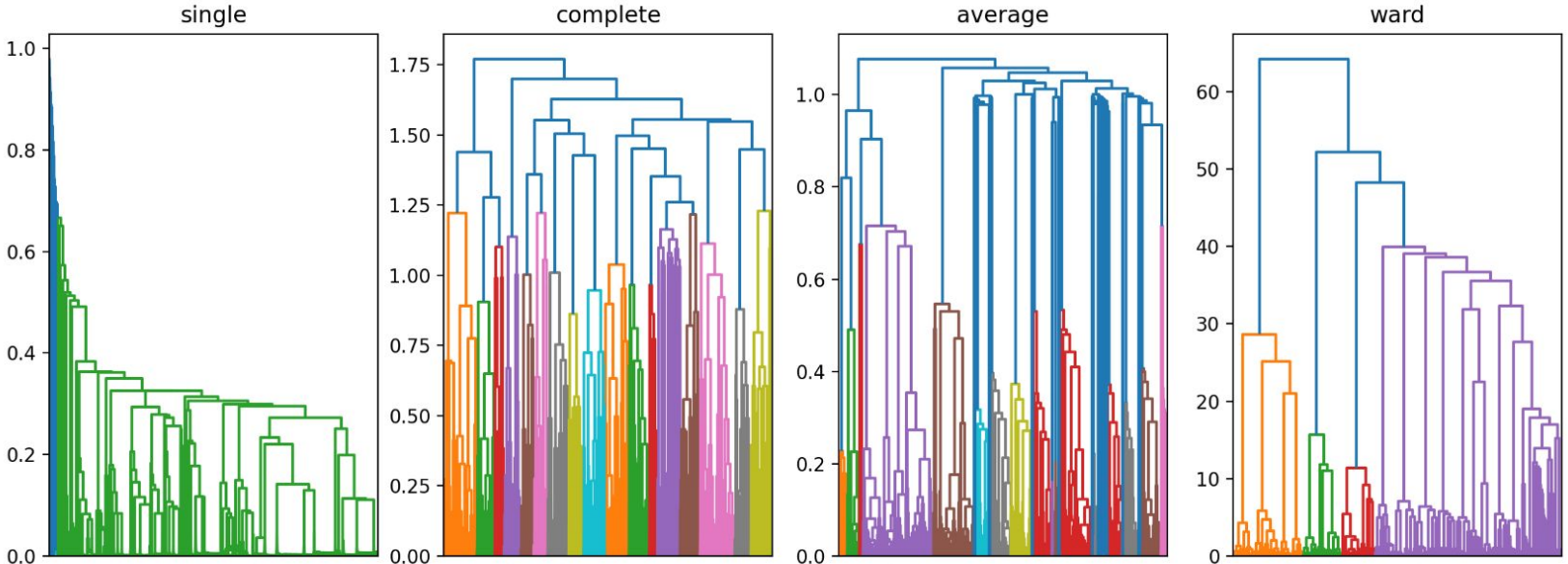
Hierarchical Clustering

Dendrograms





Dendrograms



Size Cluster Profiling

1

Size 14

Age: 38
Weight: 135
Height: 64
Body type: Hourglass

2

Size 16

Age: 33
Weight: 140
Height: 64
Body type: Hourglass

3

Size 12

Age: 40
Weight: 172
Height: 68
Body type: Full Bust

4

Size 4

Age: 32
Weight: 126
Height: 66
Body type: Hourglass

5

Size 10

Age: 34
Weight: 133
Height: 65
Body type: Hourglass

6

Size 9

Age: 32
Weight: 140
Height: 67
Body type: Hourglass

7

Size 20

Age: 34
Weight: 165
Height: 61
Body type: Full Bust

8

Size 12

Age: 36
Weight: 130
Height: 62
Body type: Hourglass

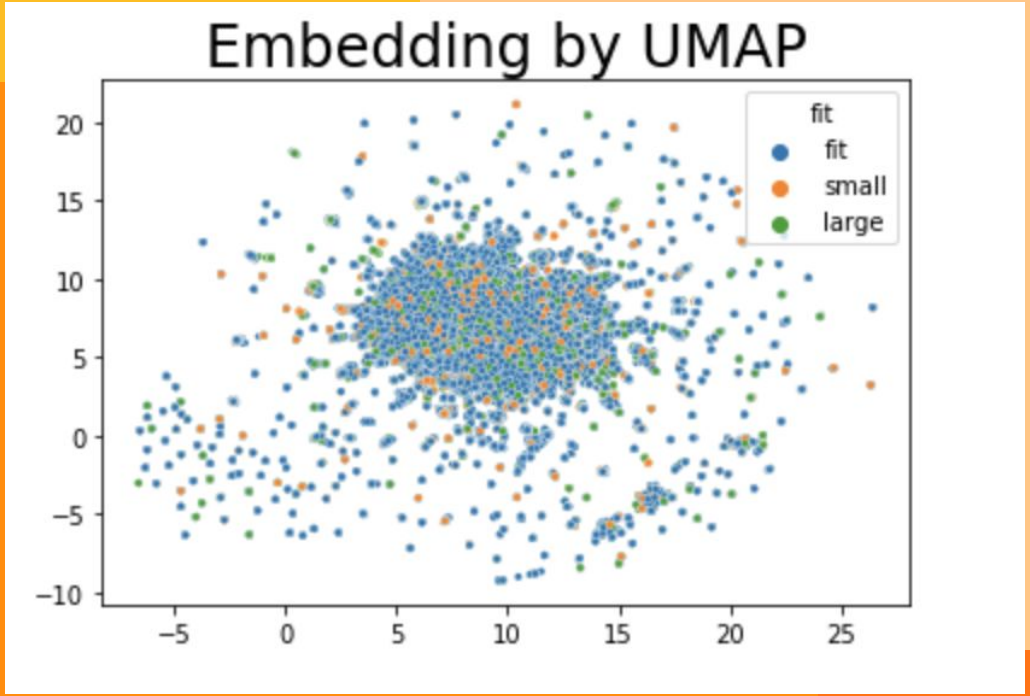
9

Size 8

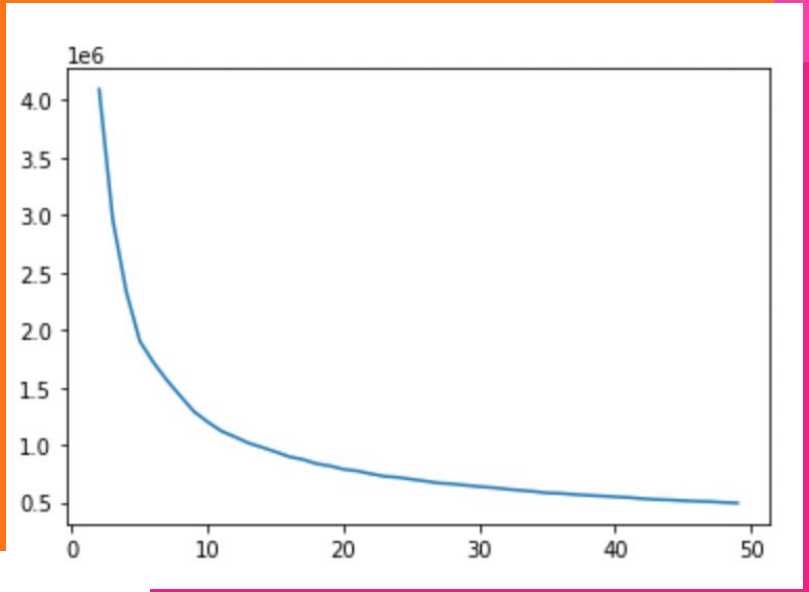
Age: 36
Weight: 117
Height: 64
Body type: Athletic

K-Means Clustering

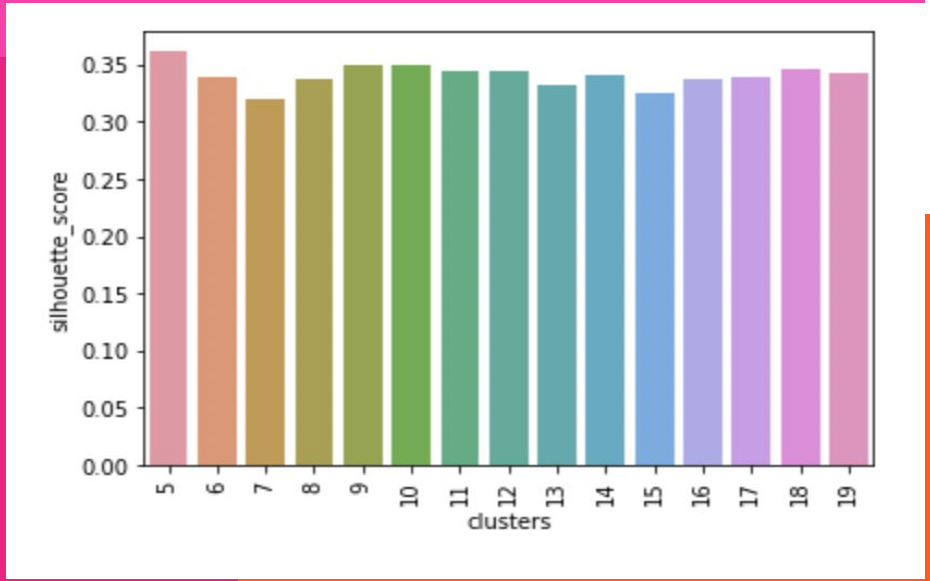
UMAP


Embedding by UMAP
fit
fit
small
large
20
15
10
5
0
−5
−10
−5
0
5
10
15
20
25

K-Means Clustering

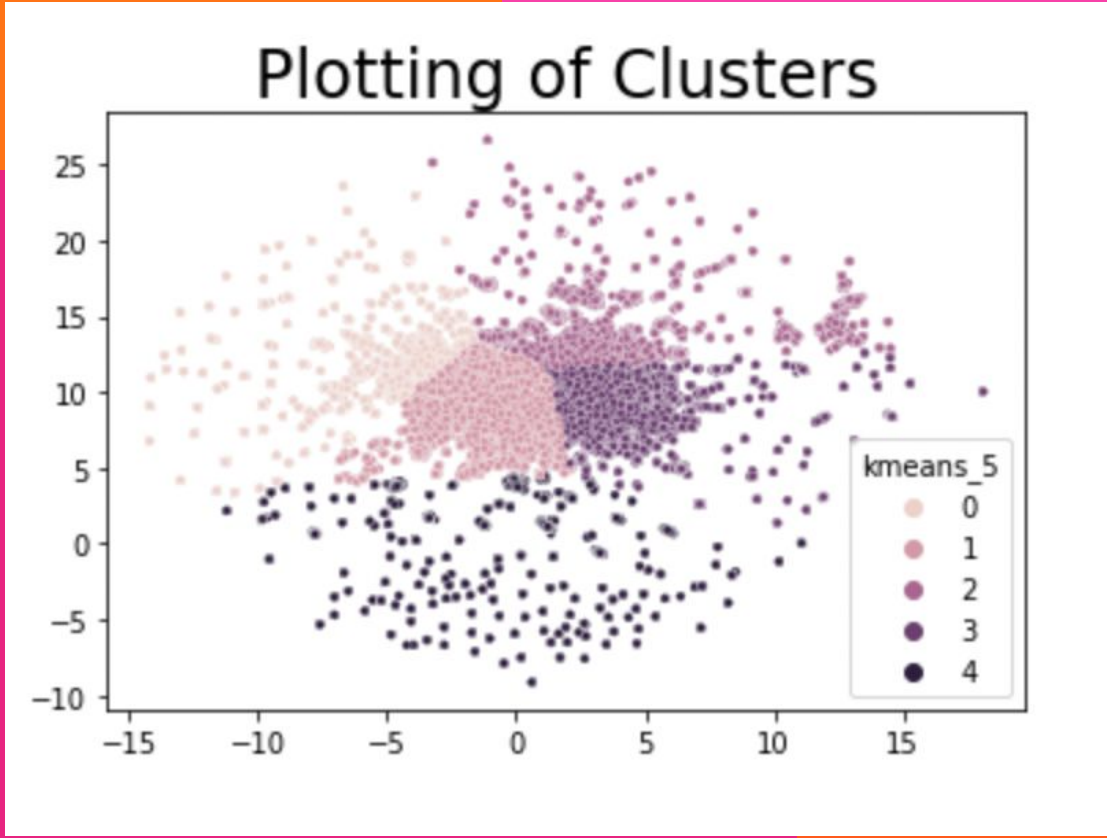


Inertia



Silhouette score

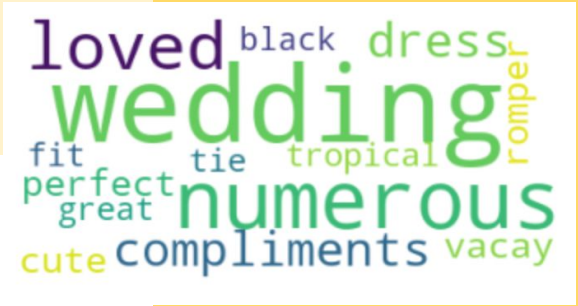
K-Means Clustering



Word Cloud for all clusters

Cluster 1

Cluster 2



Cluster 3

Cluster 4

Cluster 5

KNN for Recommendation

Content - based Filtering

```python
# getting distances and suggestions based on id
distances, suggestions = knn.kneighbors(data_txt.iloc[12345, :].values.reshape(1,-1), n_neighbors = 6)
```

Python

```python
data[data.index == suggestions[0,0]]
```

Python

	fit	user_id	bust_size	item_id	weight	rating	rented_for	review_text	body_type	review_summary	category	height	size	age	review_date
12345	fit	361493	34c	1795593	107.0	10.0	vacation	From the moment I saw this dress, I was in lov...	hourglass	In a word: Perfection!	maxi	61.0	4	31.0	2017-06-20

```python
data[data.index == suggestions[0,4]]
```

Python

	fit	user_id	bust_size	item_id	weight	rating	rented_for	review_text	body_type	review_summary	category	height	size	age	review_date
44347	fit	953990	34a	1974220	108.0	10.0	wedding	Definitely need fashion tape to keep dress fro...	hourglass	Dress was beautiful! Received many compliments...	gown	61.0	3	30.0	2017-01-22

Random Forest Regressor for recommendation

```
[ ]
def size_predict():
    age = float(input("age = "))
    weight = float(input("weight = "))
    height = float(input("height = "))
    return rfc.predict(np.array([height,weight,age]).reshape(1,3))
```

Conclusions and Recommendations

01	Size	Collect more consumer data
02	Size	Also use item fit to recommend sizes
03	Fit	Clearly display the fit of the clothing item

Thank you for listening!

Any questions?