Capstone Project

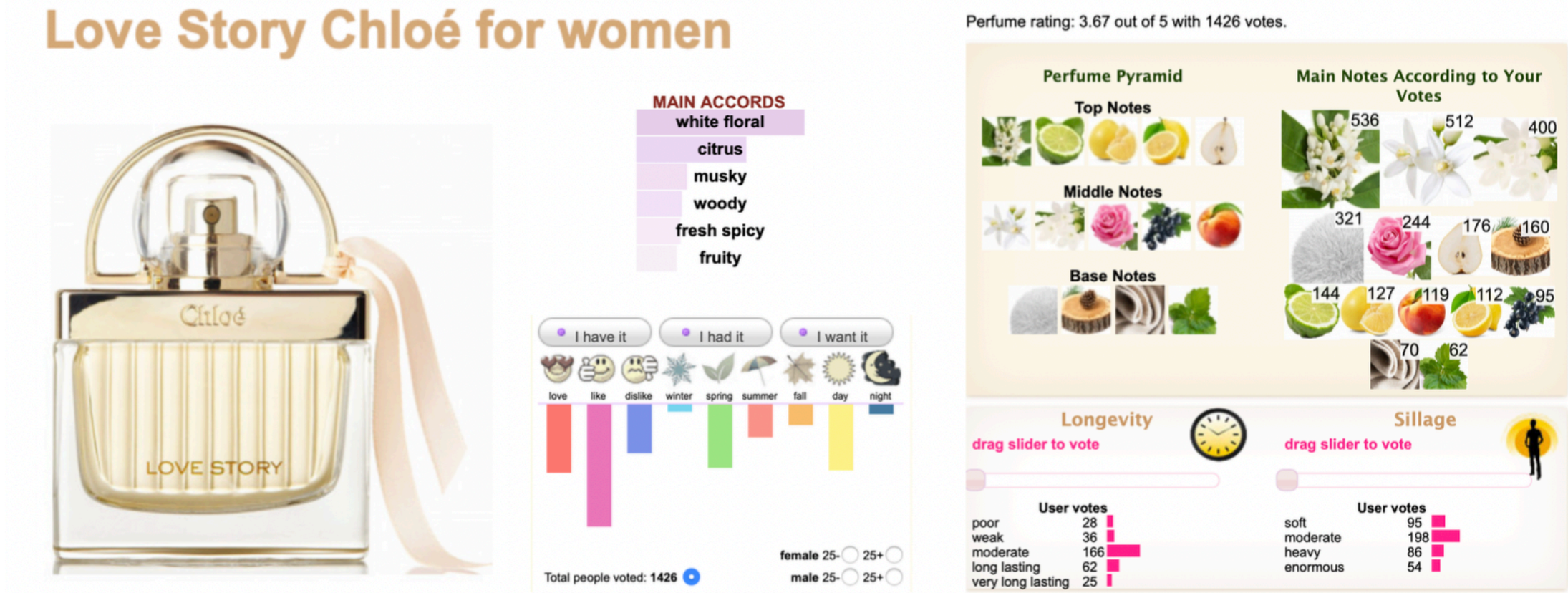
Perfumes Ratings

Worked by Kalina Zeligowska Serej

Fragrantica.com is a portal for perfume lovers.

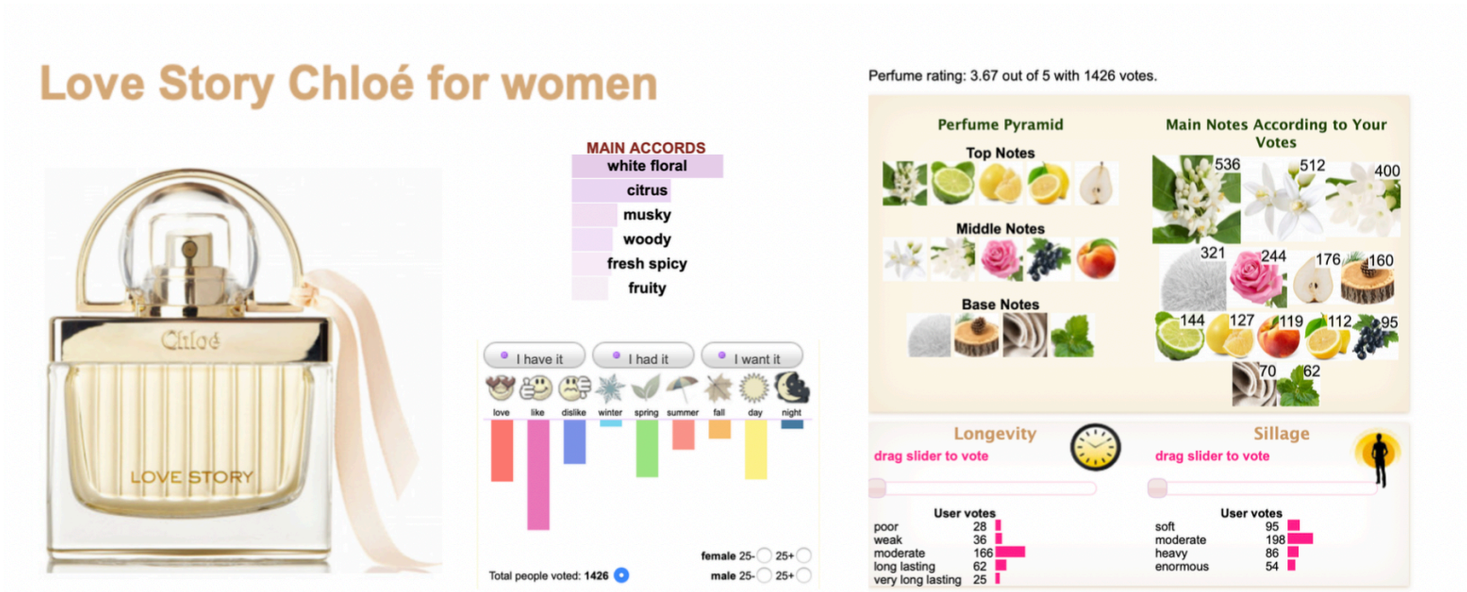
Is consists of three main segments :

- a catalogue of perfumes
- an independent magazine on this subject
- **a community - users who comment, discuss and assess perfumes**



The dataset which I worked on includes information scraped off fragrantica.com
I took it from kaggle.com as a ready-to-use csv file

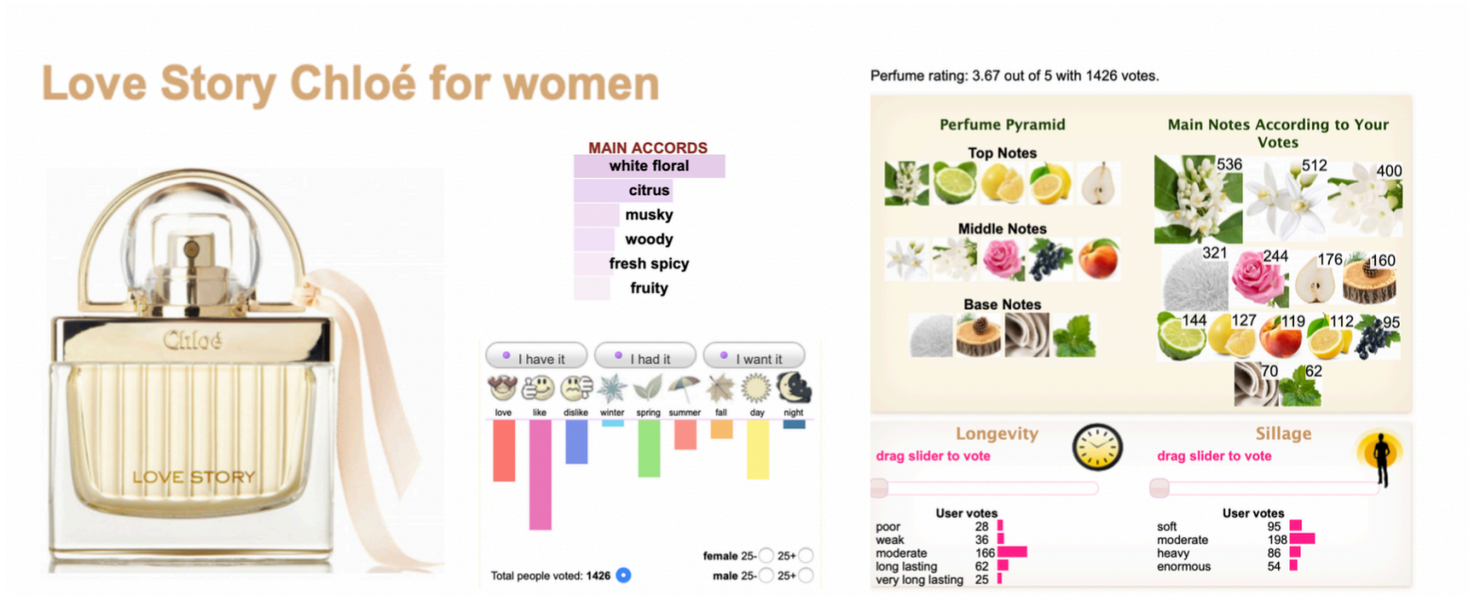
What is inside?



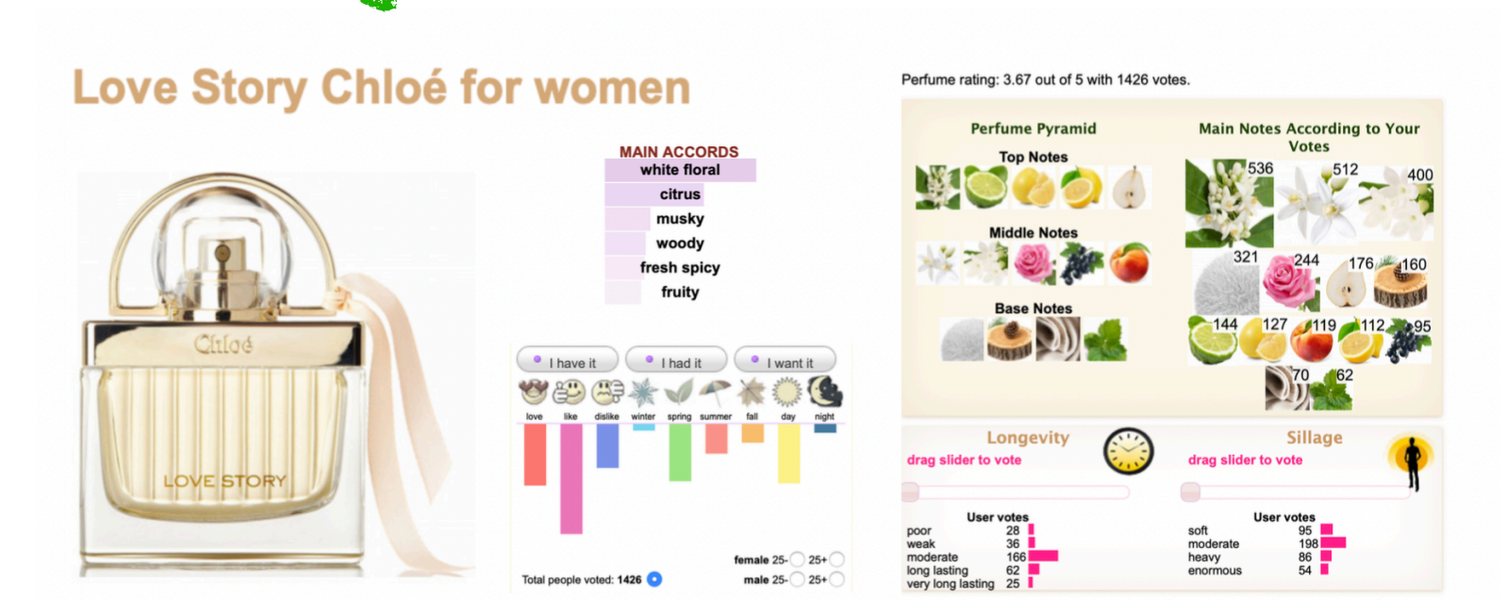
The dataset which I worked on includes information scraped off fragrantica.com
I took it from kaggle.com as a ready-to-use csv file

What is inside the data?

Each bottle of fragrance is described by several parameters

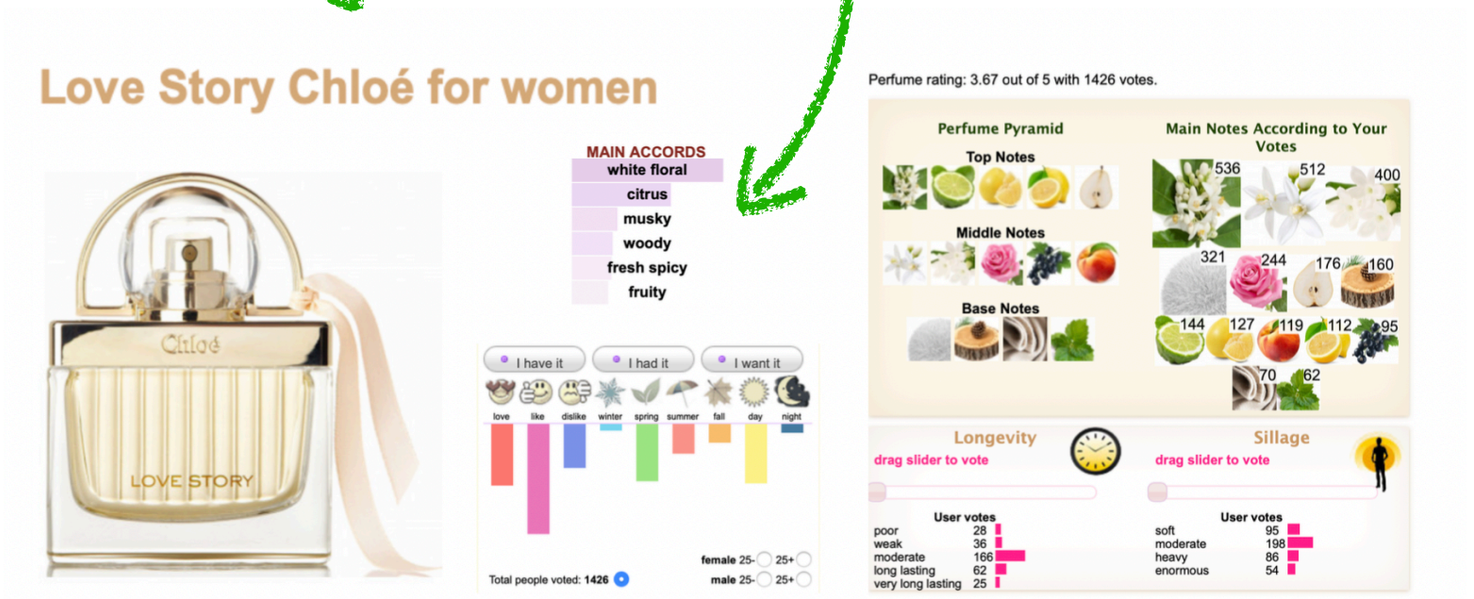


Title


Love Story Chloé for women
MAIN ACCORDS
white floral
citrus
musky
woody
fresh spicy
fruity
I have it
I had it
I want it
love
like
dislike
winter
spring
summer
fall
day
night
female 25- 25+
male 25- 25+
Total people voted: 1426
Perfume rating: 3.67 out of 5 with 1426 votes.
Perfume Pyramid
Top Notes
Middle Notes
Base Notes
Main Notes According to Your Votes
536
512
400
321
244
176
160
144
127
119
112
95
70
62
Longevity
drag slider to vote
User votes
poor 28
weak 36
moderate 166
long lasting 62
very long lasting 25
Sillage
drag slider to vote
User votes
soft 95
moderate 198
heavy 86
enormous 54

Main accords

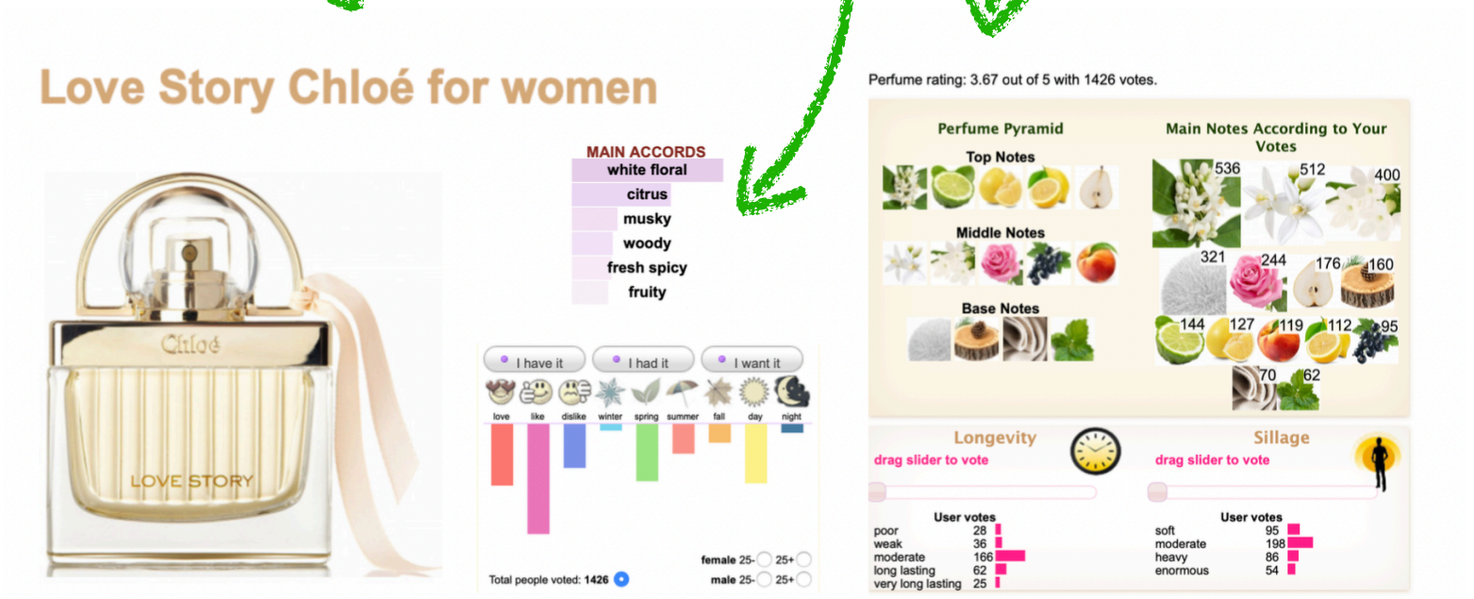
Title



Main accords

Title

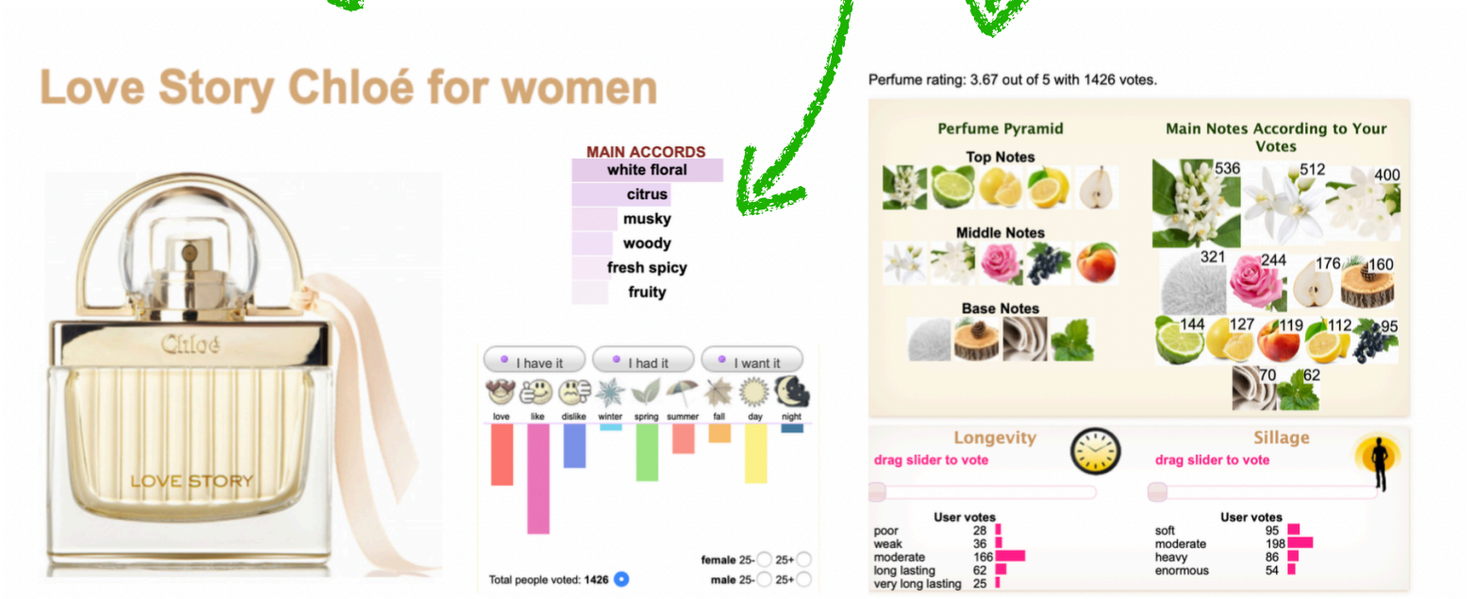
Rating score



Main accords

Title

Rating score

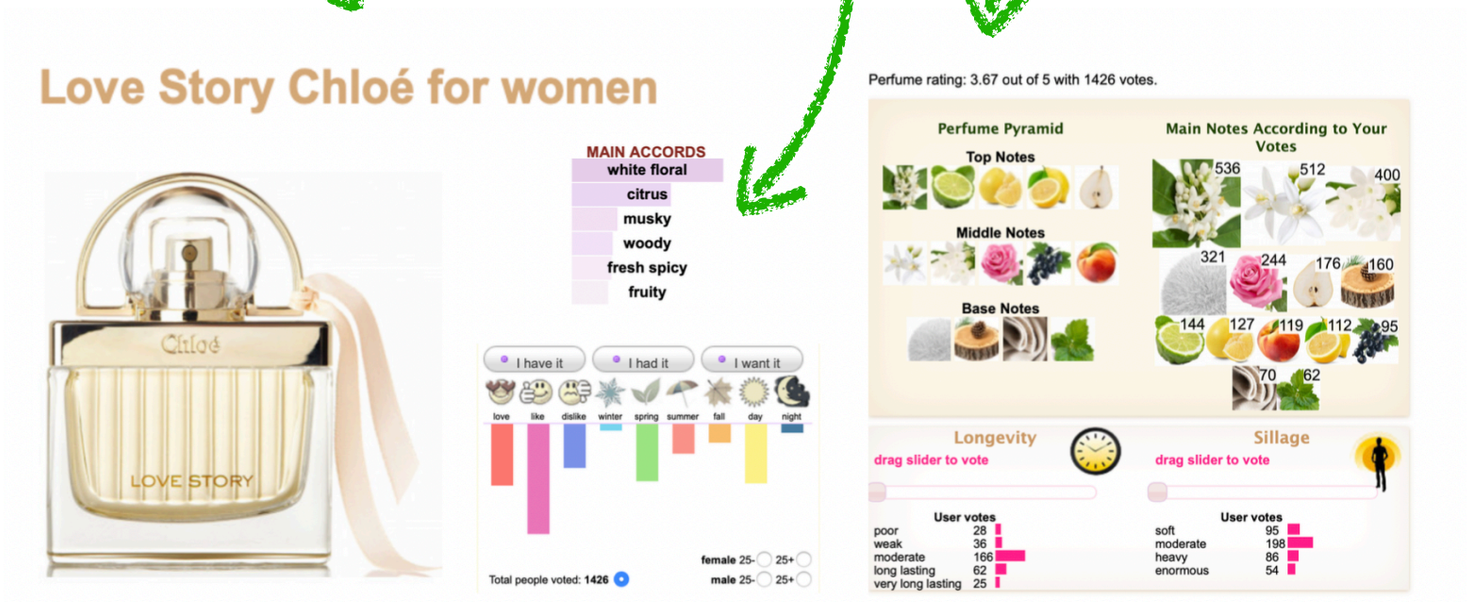


Votes

Main accords

Title

Rating score



Votes

Longevity

Main accords

Title

Rating score

Votes

Longevity

Sillage

Main accords

Title

Rating score

User’s preferences

Votes

Longevity

Sillage

Main accords

Title

Rating score

More about accords

User’s preferences

Votes

Longevity

Sillage

elements/ingredients of a fragrance

Main accords

Title

mean value of users votes

Rating score

More about accords

User's preferences

love/like/dislike
winter, spring, summer or autumn
cold or hot day
day or night
have/had/want

Votes

the number of votes submitted by users

Longevity

poor, weak, moderate,
long or very long

Sillage

soft, moderate,
heavy or enormous

What I used it for?

What I used it for?

I created

a statistical model

making predictions on perfumes ratings

The steps of my project are the following:

1. Data cleaning:

The biggest issues:

- correctly understanding the data and the way it was collected
- unifying all variables
- deciding which variables are valuable

2. Modelling

The biggest issues:

- models understanding
- checking multicollinearity

Modeling (so far):

1. Linear Regression Model as a baseline model
2. Lasso
3. Ridge

Modeling (so far):

1. **Linear Regression Model as a baseline model**
2. **Lasso**
3. **Ridge**

```
The initial R-Squared value for the linear model is: 0.0929
The Mean Squared Error for the linear model is: 0.1378
The Mean Absolute Error for the linear model is: 0.2839
```

Modeling (so far):

1. **Linear Regression Model as a baseline model**
2. **Lasso**
3. **Ridge**

```
The initial R-Squared value for the linear model is: 0.0929
The Mean Squared Error for the linear model is: 0.1378
The Mean Absolute Error for the linear model is: 0.2839
```

```
The R-Squared value for the lasso model is: 0.0
The Mean Squared Error for the lasso model is: 0.1519
The Mean Absolute Error for the lasso model is 0.301
```

Modeling (so far):

1. Linear Regression Model as a baseline model
2. Lasso
3. Ridge

```
The initial R-Squared value for the linear model is: 0.0929
The Mean Squared Error for the linear model is: 0.1378
The Mean Absolute Error for the linear model is: 0.2839
```

```
The R-Squared value for the lasso model is: 0.0
The Mean Squared Error for the lasso model is: 0.1519
The Mean Absolute Error for the lasso model is 0.301
```

```
The R-Squared value for the ridge model is: 0.0943
The Mean Squared Error for the ridge model is: 0.1376
The Mean Absolute Error for the ridge model is 0.2836
```

Modeling (so far):

1. Linear Regression Model as a baseline model
2. Lasso
3. Ridge

```
The initial R-Squared value for the linear model is: 0.0929
The Mean Squared Error for the linear model is: 0.1378
The Mean Absolute Error for the linear model is: 0.2839
```

```
The R-Squared value for the lasso model is: 0.0
The Mean Squared Error for the lasso model is: 0.1519
The Mean Absolute Error for the lasso model is 0.301
```

```
The R-Squared value for the ridge model is: 0.0943
The Mean Squared Error for the ridge model is: 0.1376
The Mean Absolute Error for the ridge model is 0.2836
```

So far, the Ridge model has the highest R^2 value so this can be used in future modelling.

Thank you!

Fell free to contact me:

kalina.zeligowska@gmail.com