



Grounds for Analysis

Natural Language Processing:
Predicting Scores and Recommending Beans



Agenda

Data and Streamlit
The grind

Natural Language Processing
Blooming and immersion

Topic Modeling
Extraction

The Modeling Process
First sip

01

02

03

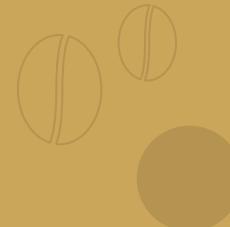
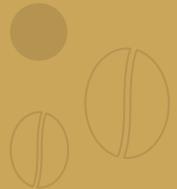
04



01

Data

The Grind



93

Souvenir Coffee

Costa Rica Cloza Estate

Roaster Location: Berkeley, California

Review Date: February

Coffee Origin: Alajuela, Costa Rica

2021

Roast Level: Medium-Light

Aroma: 9

Agtron: 58/80

Acidity 8

Est. Price: \$19.00/12 ounces

/Structure :

Body: 9

Flavor: 9

Aftertaste: 8

Blind Assessment: Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and cup. Very sweet in structure with brisk acidity; plush, syrupy-smooth mouthfeel. Sweetly nut-toned finish supported by freesia-like floral tones.

Notes: Produced by Kathia Zamora of Cloza Estate from trees of the Caturra and Catuaí varieties of Arabica and processed by the washed method. Souvenir Coffee, based in Berkeley, California, goes by the motto, “People first, then coffee.” Visit www.souvenir-coffee.com for more information.

The Bottom Line : A pleasing, delicately spice-toned, richly nutty Costa Rica cup, classically balanced.

93

Souvenir Coffee

Costa Rica Cloza Estate

Roaster Location: Berkeley, California

Review Date: February

2021

Coffee Origin: Alajuela, Costa Rica

Aroma: 9

Roast Level: Medium-Light

Acidity 8

Agtron: 58/80

/Structure :

Est. Price: \$19.00/12 ounces

Body: 9

Flavor: 9

Aftertaste: 8

Blind Assessment: Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and cup. Very sweet in structure with brisk acidity; plush, syrupy-smooth mouthfeel. Sweetly nut-toned finish supported by freesia-like floral tones.

Notes: Produced by Kathia Zamora of Cloza Estate from trees of the Caturra and Catuaí varieties of Arabica and processed by the washed method. Souvenir Coffee, based in Berkeley, California, goes by the motto, “People first, then coffee.” Visit www.souvenir-coffee.com for more information.

The Bottom Line : A pleasing, delicately spice-toned, richly nutty Costa Rica cup, classically balanced.

93

Souvenir Coffee

Costa Rica Cloza Estate

Roaster Location: Berkeley, California

Coffee Origin: Alajuela, Costa Rica

Roast Level: Medium-Light

Agtron: 58/80

Est. Price: \$19.00/12 ounces

Review Date: February
2021

Aroma: 9

Acidity /Structure : 8

Body: 9

Flavor: 9

Aftertaste: 8

Blind Assessment: Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and cup. Very sweet in structure with brisk acidity; plush, syrupy-smooth mouthfeel. Sweetly nut-toned finish supported by freesia-like floral tones.

Notes: Produced by Kathia Zamora of Cloza Estate from trees of the Caturra and Catuaí varieties of Arabica and processed by the washed method. Souvenir Coffee, based in Berkeley, California, goes by the motto, “People first, then coffee.” Visit www.souvenir-coffee.com for more information.

The Bottom Line : A pleasing, delicately spice-toned, richly nutty Costa Rica cup, classically balanced.

93

Souvenir Coffee

Costa Rica Cloza Estate

Roaster Location: Berkeley, California

Review Date: February

2021

Coffee Origin: Alajuela, Costa Rica

Aroma: 9

Roast Level: Medium-Light

Acidity 8

Agtron: 58/80

/Structure :

Est. Price: \$19.00/12 ounces

Body: 9

Flavor: 9

Aftertaste: 8

Blind Assessment: Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and cup. Very sweet in structure with brisk acidity; plush, syrupy-smooth mouthfeel. Sweetly nut-toned finish supported by freesia-like floral tones.

Notes: Produced by Kathia Zamora of Cloza Estate from trees of the Caturra and Catuaí varieties of Arabica and processed by the washed method. Souvenir Coffee, based in Berkeley, California, goes by the motto, “People first, then coffee.” Visit www.souvenircoffee.com for more information.

The Bottom Line : A pleasing, delicately spice-toned, richly nutty Costa Rica cup, classically balanced.

Blind Assessment: Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and cup. Very sweet in structure with brisk acidity; plush, syrupy-smooth mouthfeel. Sweetly nut-toned finish supported by freesia-like floral tones.

5959 Reviews



Next coffee

get

- Home
- Recommended
- Score from Test
- Score from Score
- Generated Reviews

Score Predictor

Predict coffee scores from reviews

Please provide a couple sentences describing of the flavor, acid level, aroma, aftertaste and body of your coffee.

SL, syrupy smooth mouthfeel. Sweetly nut-toned finish supported by freshie-like floral tones.

Based on your most coffee, I predict it to receive a score of: 90.00

An aroma score of out of 100: 8.50

An acidity score of out of 100: 8.50

An aftertaste score of out of 100: 8.50

A flavor score of out of 100: 8.50

A body score of out of 100: 8.50



© 2018 All rights reserved.

Natural Language Processing

02

Blooming and Immersion



Goal

Turn the words in a review into a numerical embedding

Capture word choice, frequency with a review, uniqueness across reviews

Avoid common words and errors

Cleaning



Stopwords
English+
“Keurig, coffee,
cup...”



Review



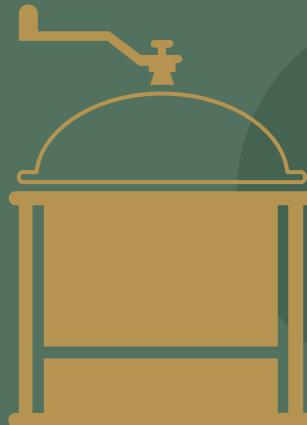
Minimum Usage
Occur in multiple reviews
(spelling errors mostly)



TF-IDF
 $\frac{\text{Frequency in a review}}{\text{Frequency in all reviews}}$

03

Topic Modeling Extraction



Non-negative Matrix Factorization (NMF)

Process

- Perform NMF fit and transform
 - Across various numbers of topics
 - Across numerous word embeddings
 - Visualize using scree plots
- Inspect most important words in topics to ascertain meaning

A stylized illustration of a coffee cup and saucer. The cup is dark teal with a light teal handle and base. It has vertical lines on its side and a small circular logo on the base. The saucer is a lighter shade of teal and has a similar vertical line pattern. The cup and saucer are centered on a large, semi-transparent teal circle.

04

The Modeling Process

First Sip

Coffee Recommendations

Find Similar Reviews

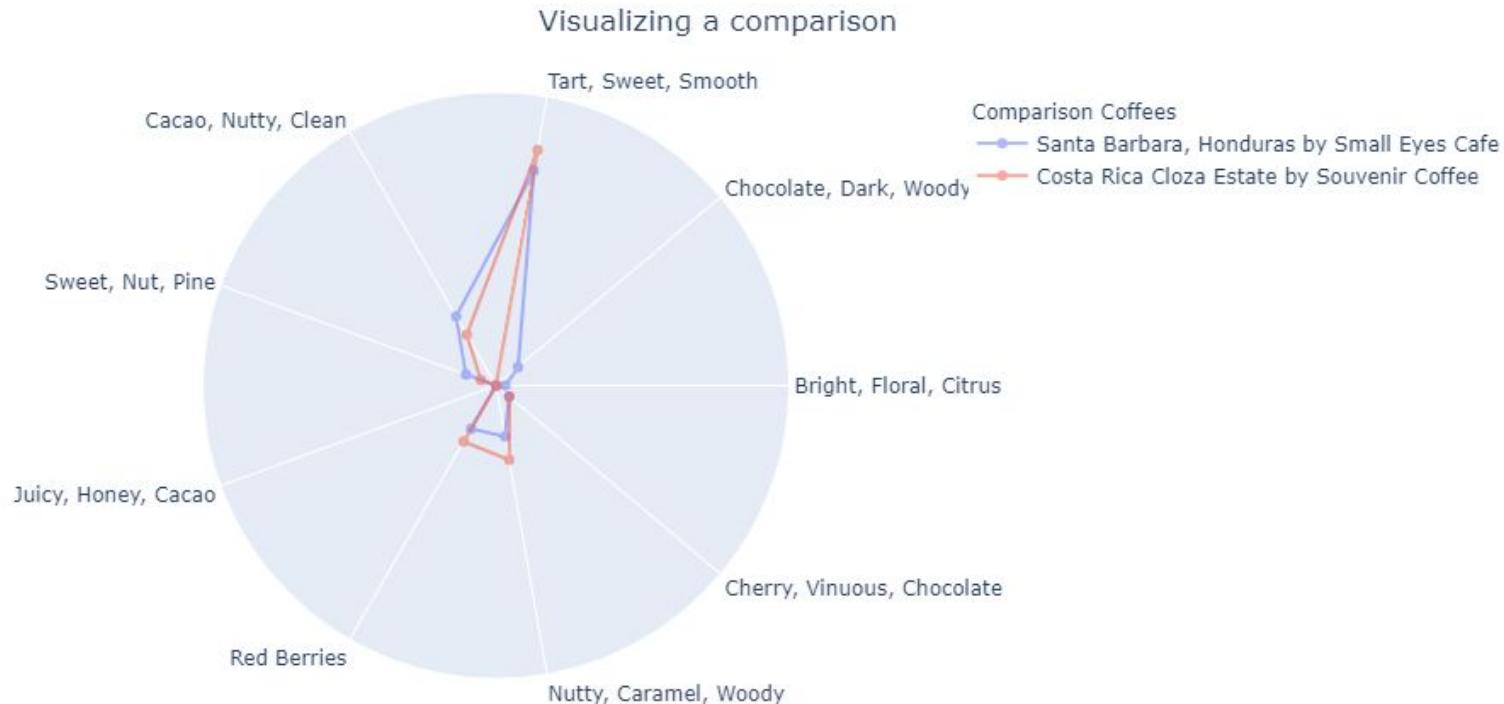


NMF Vector
Creates a nine dimensional representation of each review

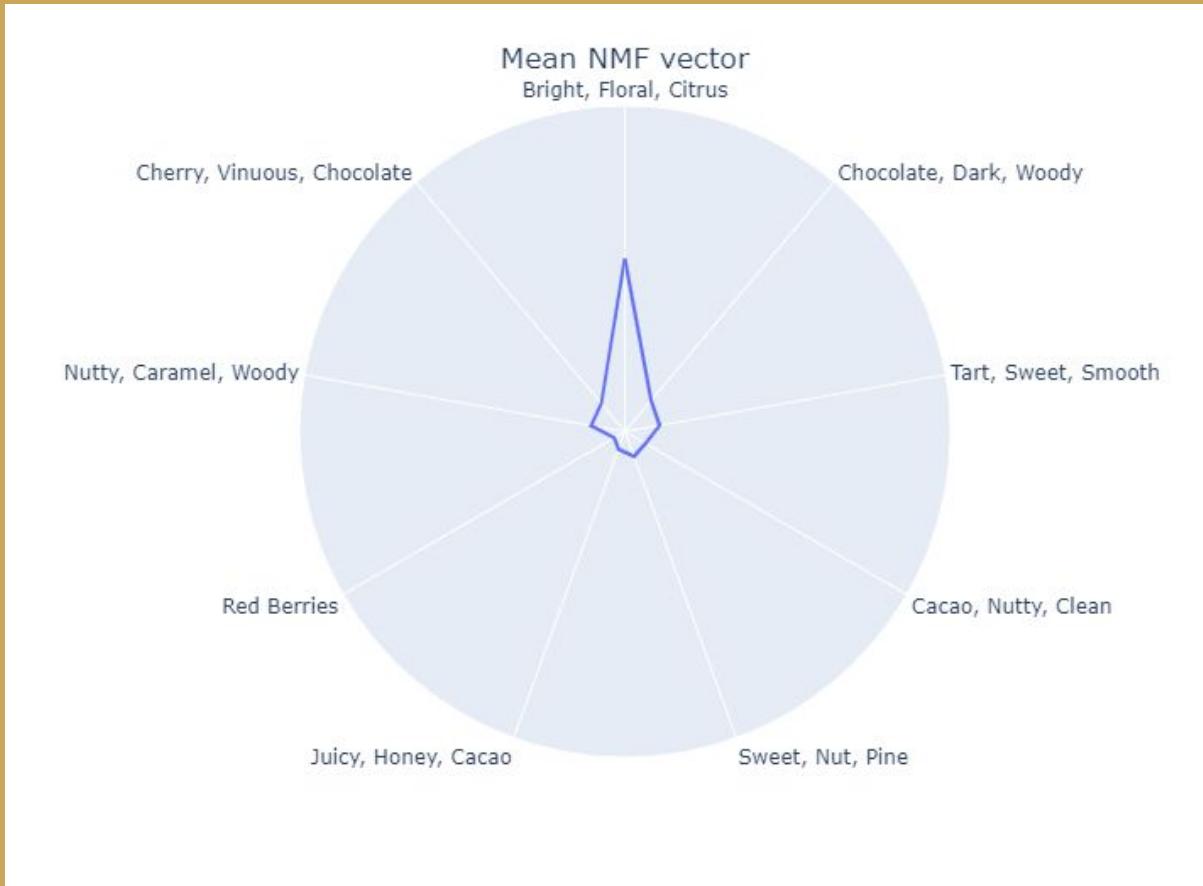
Pairwise Distance
Use cosine similarity to compare reviews to each other

Recommendation
Pick the review that is most like the input

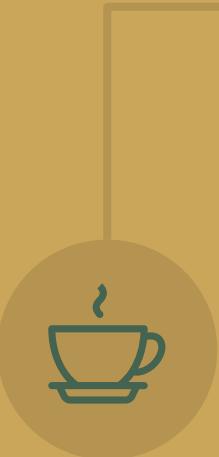
How do those recommendations appear?



How do coffees tend toward topics on average?



Reviews to Subcategories

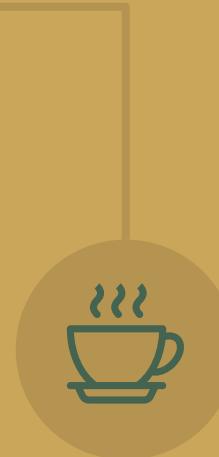


NMF Vector
As feature vector

Aroma
Body
Flavor
Aftertaste
Acidity



Review Length
Word count and number
of different words



Sentiment Analysis
Vader
Positive, Negative,
Compound

Reviews to Overall Score

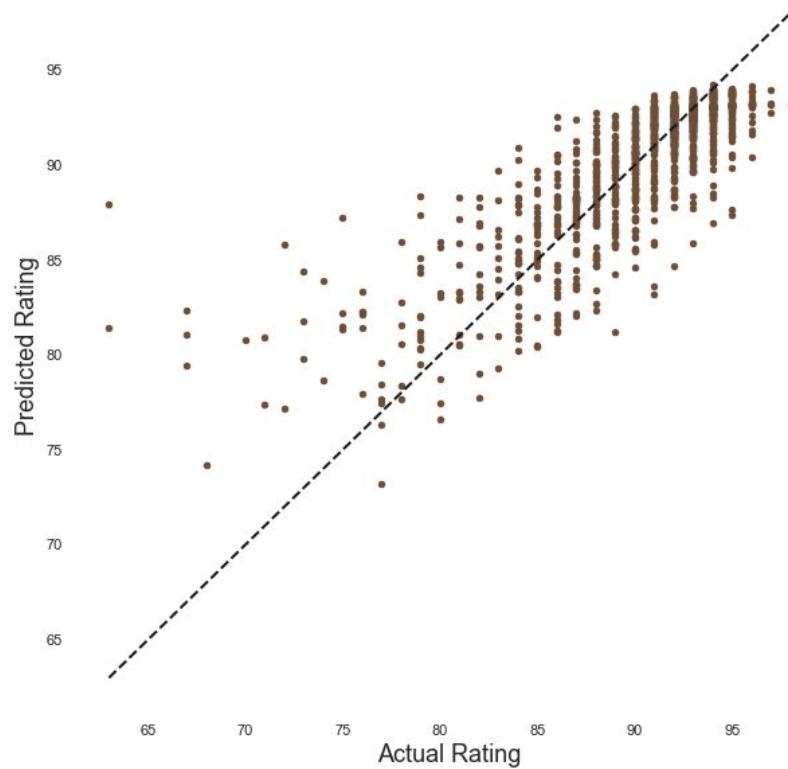


RFR Model (from text)

$r^2 = .688$

MAE = 1.7 points

Can we predict coffee ratings from reviews using a RandomForest?

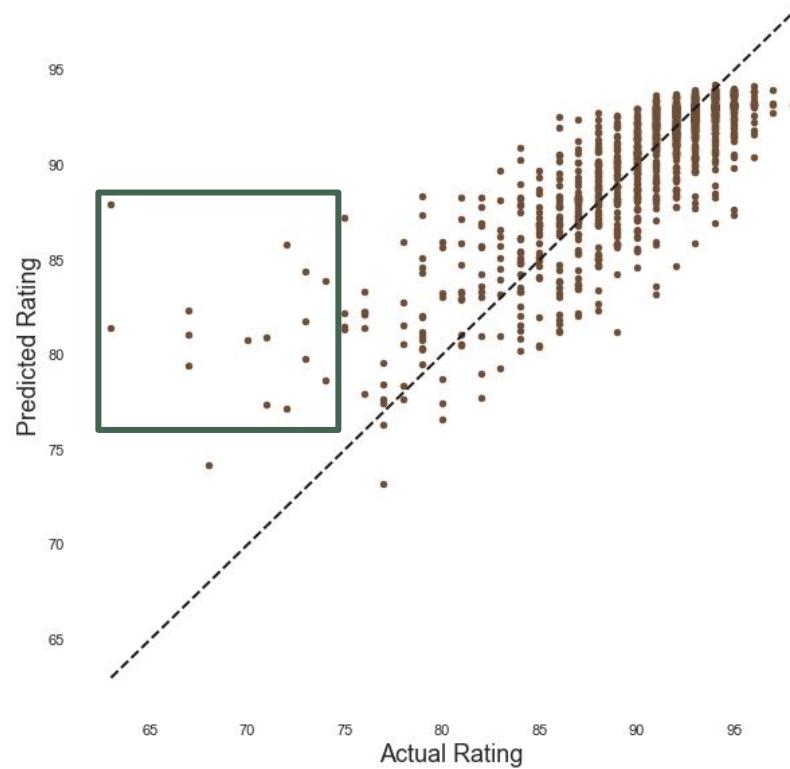


RFR Model (from text)

Errors:

Negative reviews
Older reviews
Instant Coffee

Can we predict coffee ratings from reviews using a RandomForest?



Questions?

Appendix

Subcategories to Overall Score

Overall Score



NMF Vector
As feature vector



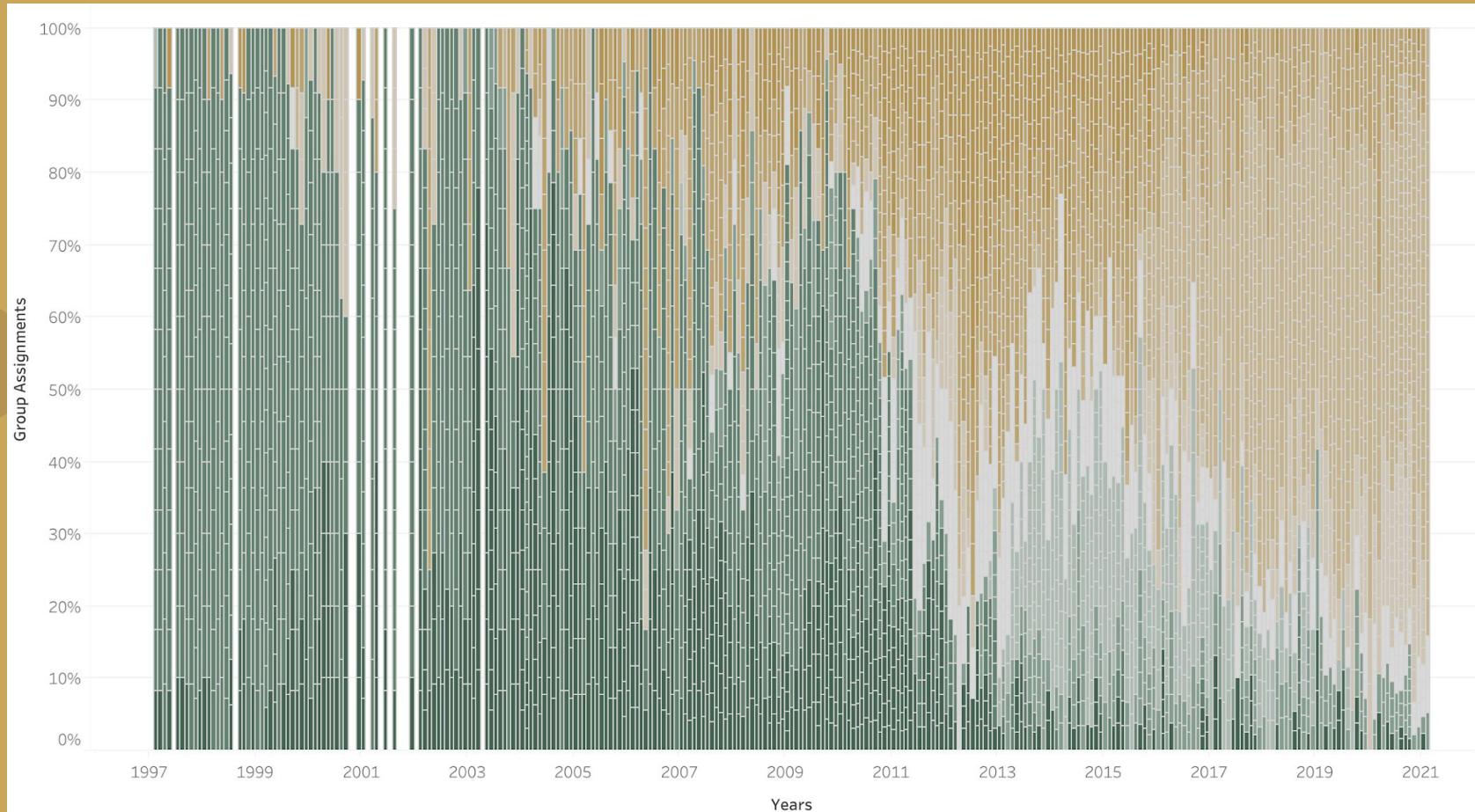
Review Length
Word count and number
of different words



Sentiment Analysis
Vader
Positive, Negative,
Compound

Interesting Visualizations

Coffees reviewed have shifted groups over time



Origin and Roast Level by Group

Bright, Floral, Citrus



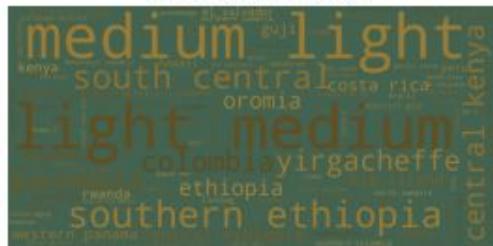
Cacao, Nutty, Clean

Chocolate, Dark, Woody



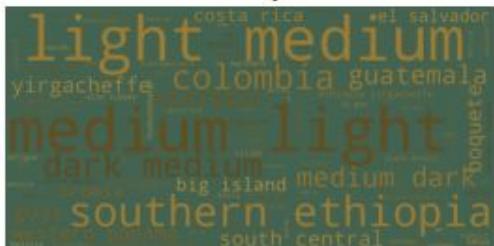
Sweet, Nut, Pine

Tart, Sweet, Smooth



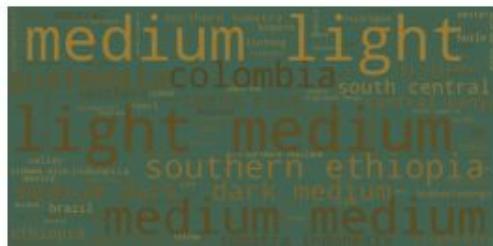
Juicy, Honey, Cacao

Costa Rica



Red Berries

• Page 718



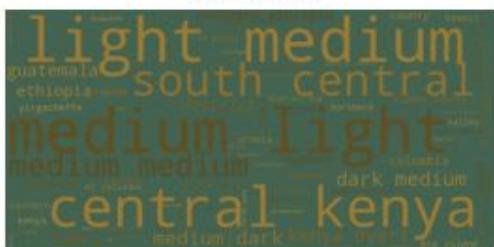
Nutty, Caramel, Woody

卷之三十一



Cherry, Vinyous, Chocolate

ANSWER



Reviews by Groups

Bright, Floral, Citrus



Cacao, Nutty, Clean

Chocolate, Dark, Woody



Tart, Sweet, Smooth



Cacao, Nutty, Clean

Sweet, Nut, Pine

Juicy, Honey, Cacao

Red Berries

Nutty, Caramel, Woody

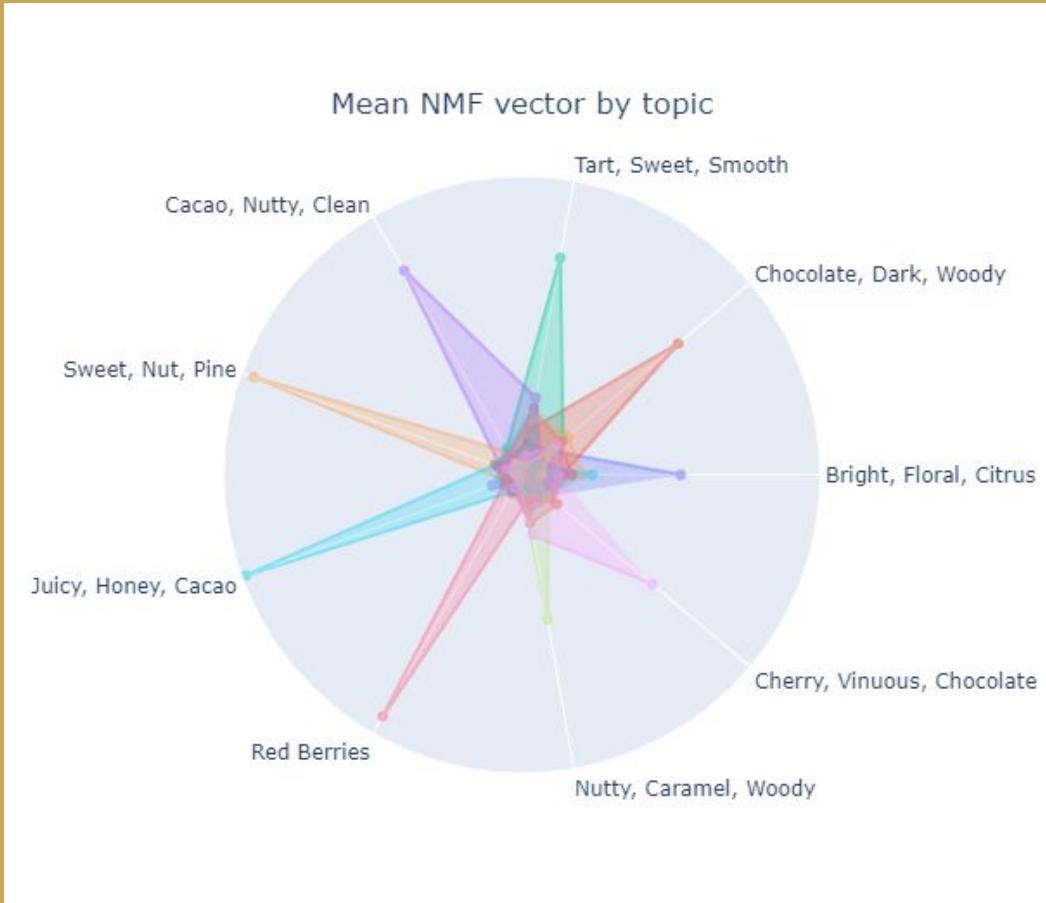
Cherry, Vinous, Chocolate

black currant
dark chocolate long
flower pungent sweet hints
rich chocolate black cherry

rich simple sweet tone
rather sandy caramel profile
like a character hints
aromatic wood flower
dry balanced
acidity soft milk
body fruity

rich fruit
complex **sweet** **acidity**
pungent balanced delicate nut
floral dark chocolate
cherry - deep stone
hints chocolate-toned
tart fruity notes

How do coffees tend toward topics on average?



Counter Culture Coffee Taster's Flavor Wheel



Body Descriptors

Light	Watery Skim Milk Tea-like	Medium	Round 2% Milk Creamy	Heavy	Full Whole Milk Chewy
-------	---------------------------------	--------	----------------------------	-------	-----------------------------

Adjectives & Intensifiers for Coffee

Crisp Bright Vibrant Tart	Muted Dull Mild	Wild Unbalanced Sharp Pointed	Structured Balanced	Dense Deep Complex	Soft Faint Delicate	Juicy Syrupy	Dry Astringent	Lingering Coating Dirty	Quick Clean
------------------------------------	-----------------------	--	------------------------	--------------------------	---------------------------	-----------------	-------------------	-------------------------------	----------------

Visit counterculturecoffee.com/tasterswheel to learn more. | ©Counter Culture Coffee 2020

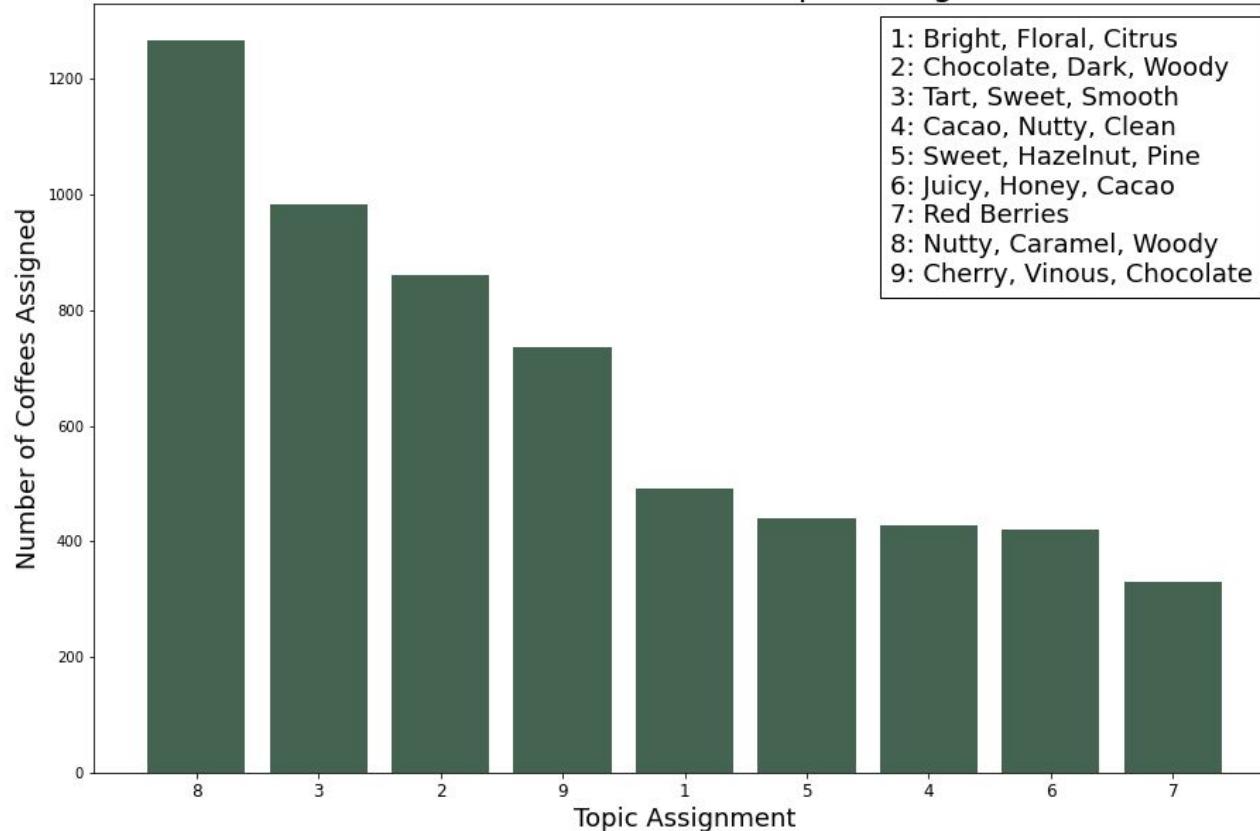
SCA

Coffee Taster's Flavor Wheel



COFFEE TASTER'S FLAVOR WHEEL® CREATED USING THE SENSORY
LICKBOOK DEVELOPED BY WORLD COFFEE RESEARCH.
© 2014 AKA AND WCR

Distribution of coffees to topic categories

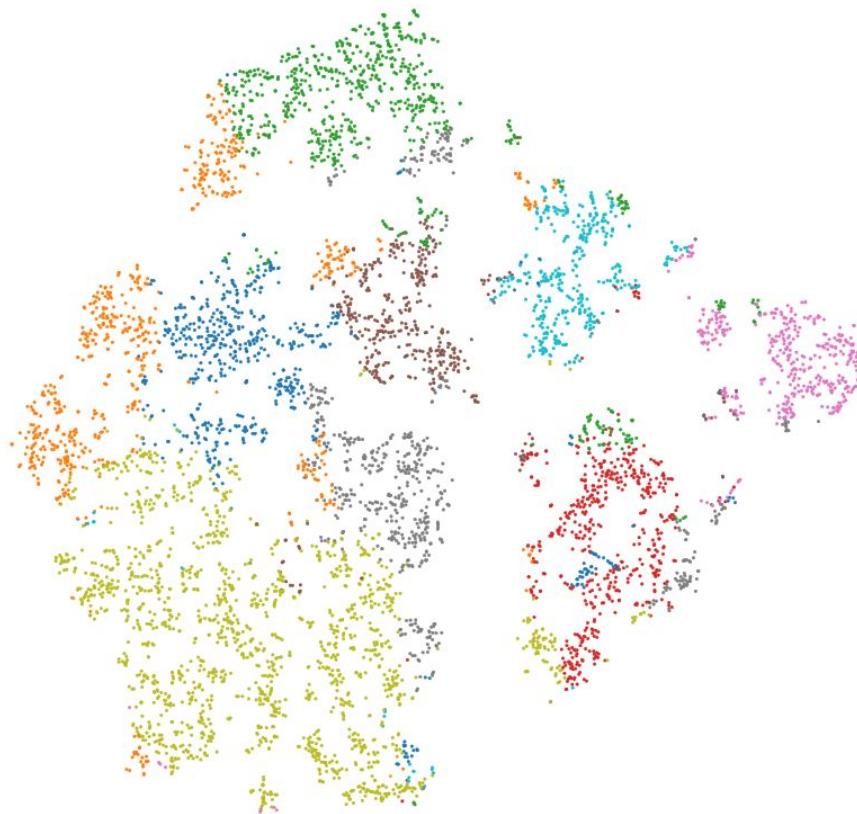


All Coffee Reviews Clustered by Category

Unsupervised topic modeling with NMF based on textual content + 2D-embedding with t-SNE:

■ Bright, Floral, Citrus	■ Cacao, Nutty, Clean	■ Red Berries
■ Chocolate, Dark, Woody	■ Sweet, Nut, Pine	■ Nutty, Caramel, Woody
■ Tart, Sweet, Smooth	■ Juicy, Honey, Cacao	■ Cherry, Vinuous, Chocolate

Blind Reviews from Coffee-Review.com



Streamlit App Demo

In Case of Video Failure

Streamlit

Next on list x

go to

- Home
- Recommender
- Score from Text
- Score from Score
- Generated Reviews

Welcome to my data analysis app for coffee reviews!

On the side bar on the left you will find a few different applications

1. This is the **Home Page**
2. Use the **Recommender** app to get a coffee recommendation based on your flavor description
3. Use the **Score from Text** app to generate a prediction for overall and subcategory score based on a coffee's description
4. Use the **Score from Scores** app to generate an overall score prediction based on subcategory scores
5. Use the **Generated Reviews** app to create a computer generated review for a coffee depending on different flavor attributes

Made with [Streamlit](#)

Applet Recommendation

Coffee Recommender

Get a new coffee recommendation

Give a couple sentences here of how you describe your ideal coffee. Try to include as much as you can about your desired flavor profile.

Excited to recommend a coffee for you!

Applet Recommendation

Coffee Recommender

Get a new coffee recommendation

Give a couple sentences here of how you describe your ideal coffee. Try to include as much as you can about your desired flavor profile.

Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and

Based on your input coffee, I recommend you try:

Small Eyes Cafe who roast a bean from Santa Barbara, Honduras .

It could be described as:

Crisply sweet, nut-toned, richly floral. Almond butter, lilac, lemon verbena, red apple, oak in aroma and cup. Sweet structure with soft, round acidity; very full, syrupy-smooth mouthfeel. Lemon verbena supports the resonantly nut-toned finish.

Score Predictions

Score Predictor

Predict coffee scores from reviews

Provide a couple sentence description of the flavors, acid level, aroma, aftertaste, and body of your coffee.

|

Excited to predict the score of your coffee!

Score Prediction

Score Predictor

Predict coffee scores from reviews

Provide a couple sentence description of the flavors, acid level, aroma, aftertaste, and body of your coffee.

Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and

Based on your input coffee, I predict it to receive a score of: **92.32**

An aroma score of (out of 10): **8.59**

An acidity score of (out of 10): **8.24**

An aftertaste score of (out of 10): **8.06**

A flavor score of (out of 10): **8.72**

A body score of (out of 10): **8.21**

92.3

93

Souvenir Coffee

Costa Rica Cloza Estate

Roaster Location: Berkeley, California

Review Date: February

2021

Coffee Origin: Alajuela, Costa Rica

8.59

Roast Level: Medium-Light

8.24

Agtron: 58/80

8.06

Est. Price: \$19.00/12 ounces

8.72

Aroma: 9

Acidity /Structure : 8

Body: 9

Flavor: 9

Aftertaste: 8

8.21

Blind Assessment: Crisp, balanced, richly nut-toned. Nutella, red apple, freesia, agave syrup, cedar in aroma and cup. Very sweet in structure with brisk acidity; plush, syrupy-smooth mouthfeel. Sweetly nut-toned finish supported by freesia-like floral tones.

Notes: Produced by Kathia Zamora of Cloza Estate from trees of the Caturra and Catuaí varieties of Arabica and processed by the washed method. Souvenir Coffee, based in Berkeley, California, goes by the motto, “People first, then coffee.” Visit www.souvenircoffee.com for more information.

The Bottom Line: A pleasing, delicately spice-toned, richly nutty Costa Rica cup, classically balanced.

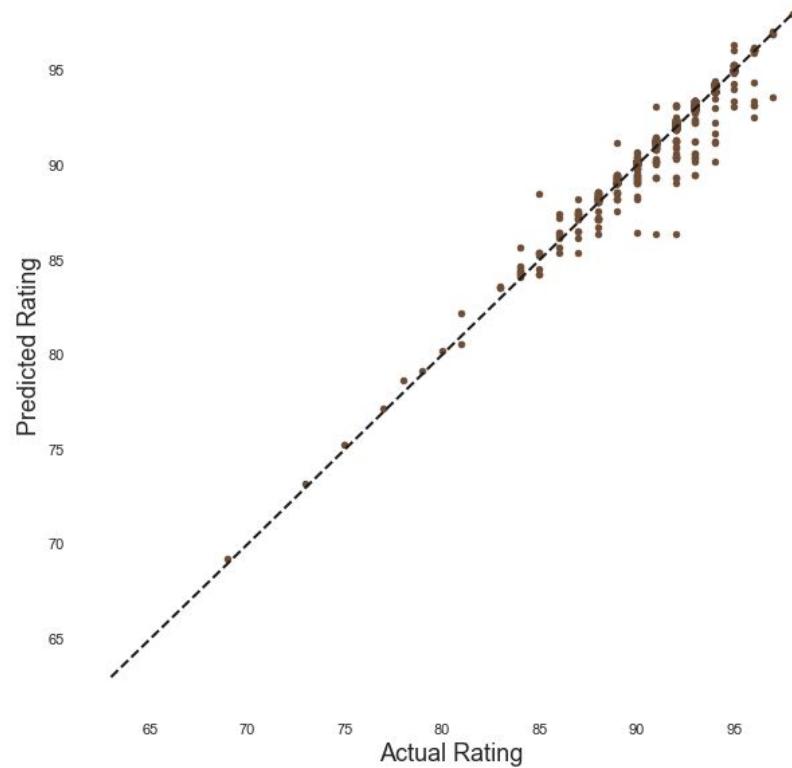
More details on Model Predictions And Errors

Linear Model (from scores)

$r^2 = .943$

MAE = .36 points

How does the predicted rating of a coffee compare to reality?



Linear Model (from scores)

Coefficients

Aroma = 1.04

Body = 0.75

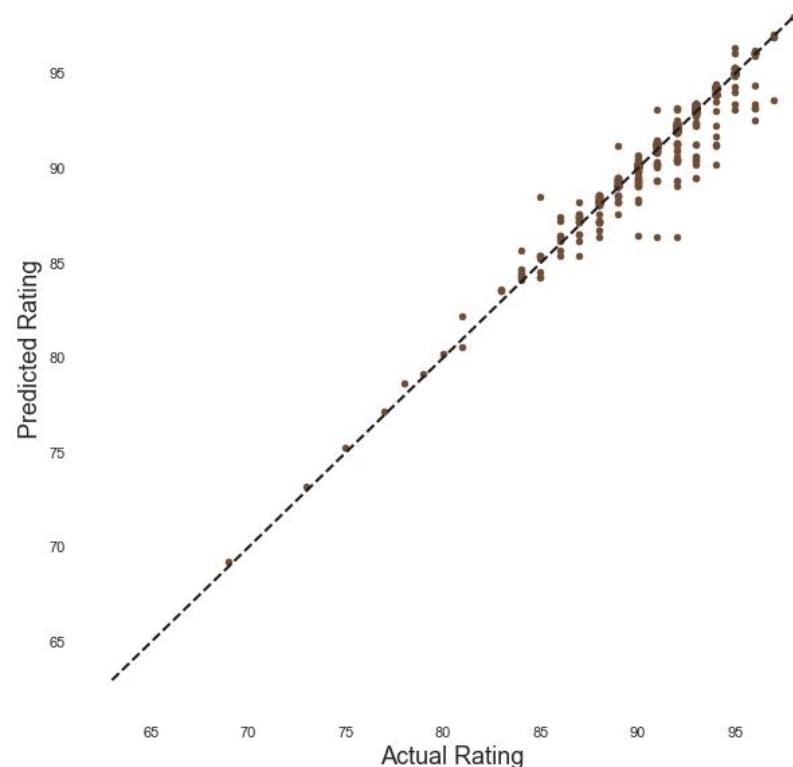
Flavor = 1.00

Aftertaste = 1.14

Acidity = 0.89

Intercept = 51.65

How does the predicted rating of a coffee compare to reality?



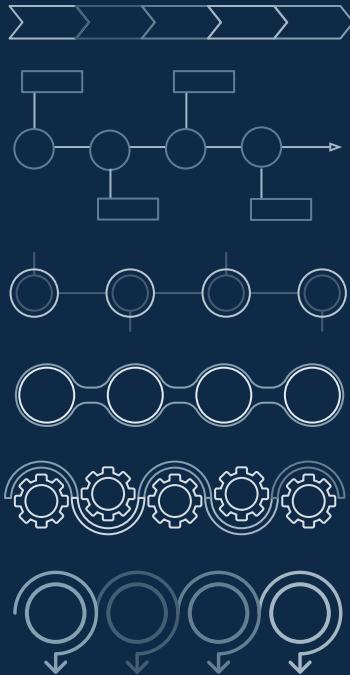
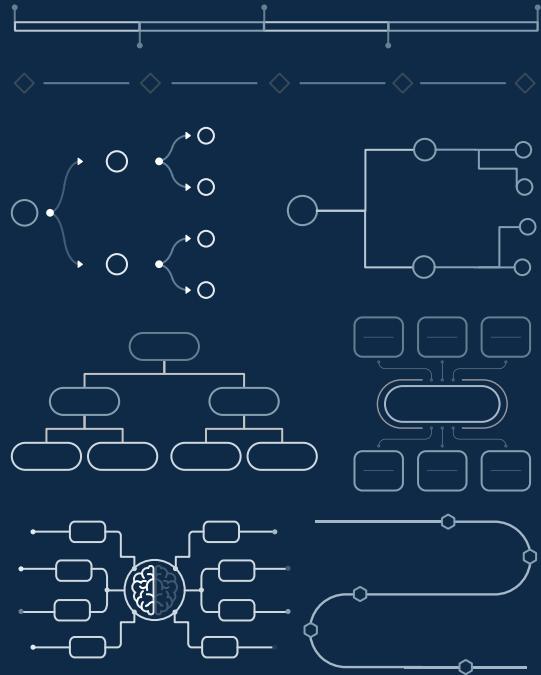
Use our editable graphic resources...

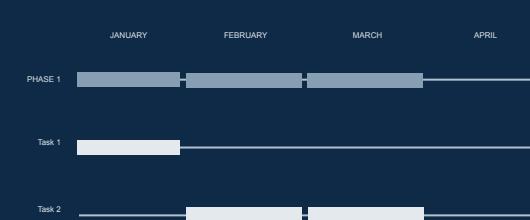
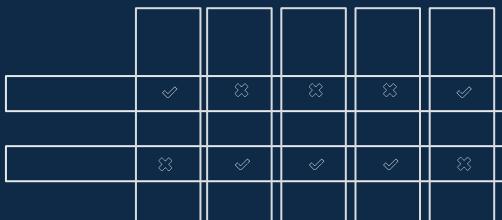
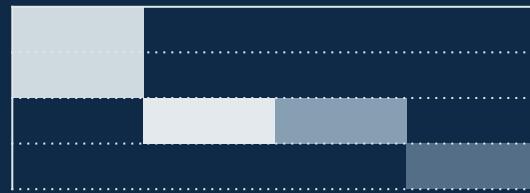
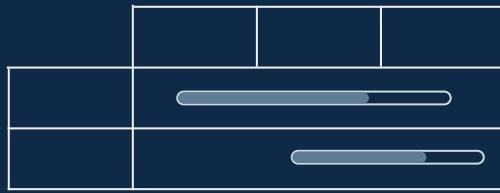
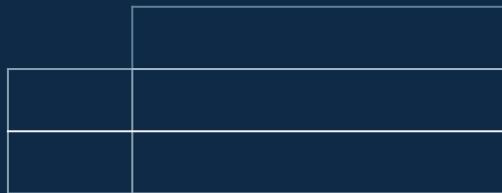
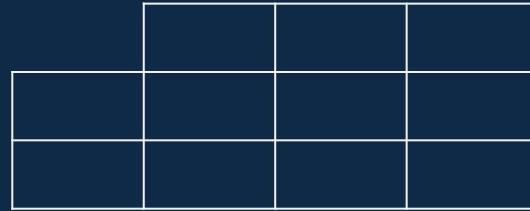
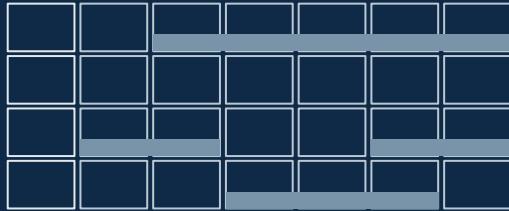
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want.

Group the resource again when you're done. You can also look for more infographics on Slidesgo.

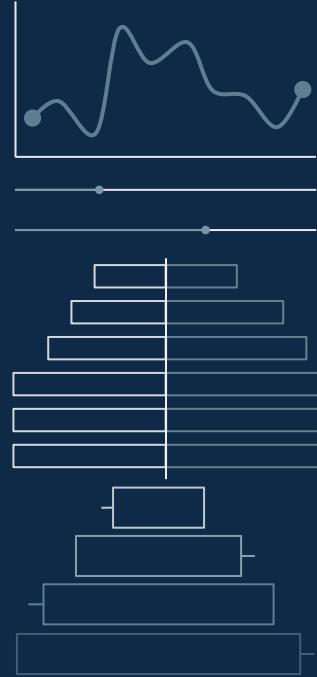
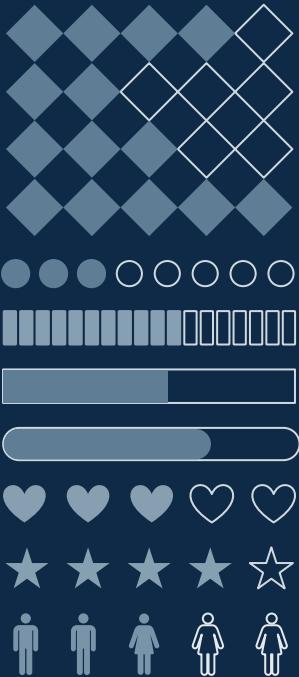
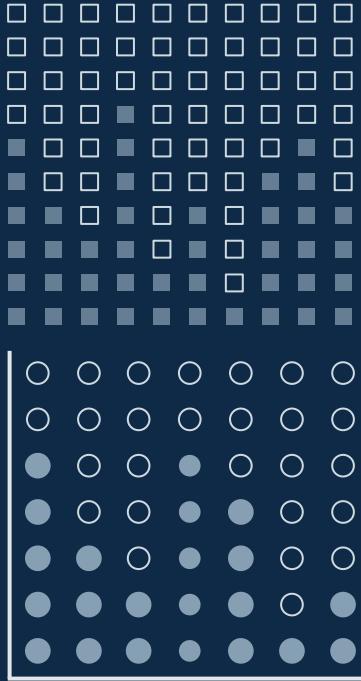












...and our sets of editable icons

You can resize these icons, keeping the quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.



Educational Icons



Medical Icons



Business Icons



Teamwork Icons



Help & Support Icons



Avatar Icons



Creative process Icons



Performing Arts Icons



Nature Icons



SEO & Marketing Icons



