

#### Agenda

Coffee Review Intro **Data** 

Natural Language Processing Cleaning, preparing, and embedding

Topic Modeling
Ion-negative Matrix Factorization

Visualizations
Seeing insights 04

**Applet** ecommendations, Modeling, and more







www.coffeereview.com









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:	0
		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.



Costa Rica Cloza Estate

Roaster Location:	Berkeley, California	Review Date:	February
Coffee Origin:	Alajuela, Costa Rica	Review Date.	2021
Roast Level:	Medium-Light	Aroma:	9
Agtron:	58/80	Acidity	Q
Est. Price:	\$19.00/12 ounces	/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.souvenircoffee.com for more information.



#### 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

8

Aroma:

Acidity

/Structure:

Body:

Flavor: 9

Aftertaste:

**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.

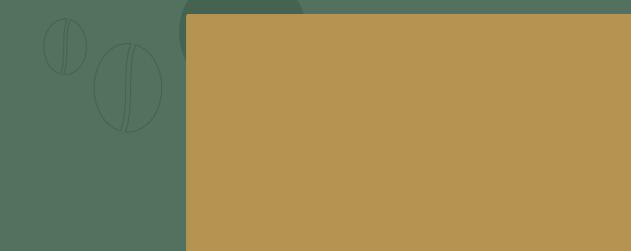


Costa Rica Cloza Estate

Roaster Location:	Berkeley, California	Review Date:	February
Coffee Origin:	Alajuela, Costa Rica		2021
Roast Level:	Medium-Light	Aroma:	9
Agtron:	58/80	Acidity	Q
Est. Price:	\$19.00/12 ounces	/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.souvenircoffee.com for more information.



**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

**Blind Reviews** 

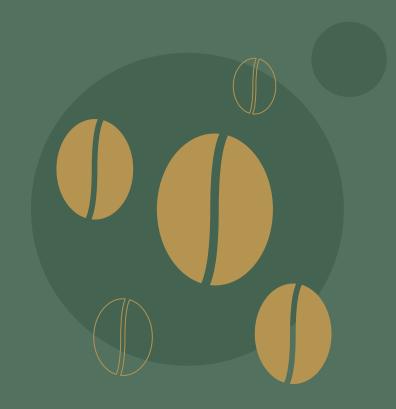




## Natural Language Processing









#### 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..."

#### **Minimum Usage**

Occur in multiple reviews (spelling errors mostly)

#### TF-IDF

Frequency in a review

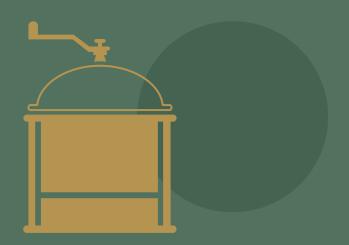
Frequency in all reviews



### **Topic Modeling**

Non-negative Matrix Factorization





#### **NMF**



#### **General Overview**

- Create potential clusters of reviews
- Assign each review a score for assignment to clusters
- Find meaning in clusters to understand topics
- By-product: Reduce each review to a smaller feature space

#### **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







#### Topic 1

flowers, honey, silky acidity, like, bright, saturated, complex, lemon, orange

#### Topic 4

cocoa, toned, powder nib, cedar, structure, crisp, nut, sweet, date

#### Topic 7

black, currant, cherry savory, red, pungent, blackberry, syrupy, deeply, berry

#### Topic 2

chocolate, dark, cedar, milk, small, chocolaty, deep, syrupy, flowers, scorched

#### Topic 5

fresh, cut, fir, lightly, syrupy, drying, gently, hazelnut, round, acidity

#### Topic 8

wood, body, aromatic, nut, sweetness, rather, sweet, hints, tones, caramel



#### Topic 3

structure, tart, sweet, zest, richly, savory, almond, satiny, toned, sweetly

#### Topic 6

cacao, nib, roasted, drying, lively, juicy, crisp, honeysuckle, acidity, sandalwood

#### Topic 9

#### Topics

#### Topic 1

Bright, Floral, Citrus

#### Topic 4

cocoa, toned, powder nib, cedar, structure, crisp, nut, sweet, date

#### Topic 7

black, currant, cherry savory, red, pungent, blackberry, syrupy, deeply, berry

#### Topic 2

chocolate, dark, cedar, milk, small, chocolaty, deep, syrupy, flowers, scorched

#### Topic 5

fresh, cut, fir, lightly, syrupy, drying, gently, hazelnut, round, acidity

#### Topic 8

wood, body, aromatic, nut, sweetness, rather sweet, hints, tones,



#### Topic 3

structure, tart, sweet, zest, richly, savory, almond, satiny, toned, sweetly

#### Topic 6

cacao, nib, roasted, drying, lively, juicy, crisp, honeysuckle, acidity, sandalwood

#### Topic 9



#### Topic 1

Bright, Citrus, Floral

#### Topic 4

cocoa, toned, powder nib, cedar, structure, crisp, nut, sweet, date

#### Topic 7

black, currant, cherry savory, red, pungent, blackberry, syrupy, deeply, berry

#### Topic 2

chocolate, dark, cedar, milk, small, chocolaty, deep, syrupy, flowers, scorched

#### Topic 5

fresh, cut, fir, lightly, syrupy, drying, gently, hazelnut, round, acidity

#### Topic 8

wood, body, aromatic, nut, sweetness, rather, sweet, hints, tones,

#### Topic 3

structure, tart, sweet, zest, richly, savory, almond, satiny, toned, sweetly

#### Topic 6

cacao, nib, roasted, drying, lively, juicy, crisp, honeysuckle, acidity, sandalwood

#### Topic 9



#### Topics

#### Topic 1

Bright, Citrus, Floral



cocoa, toned, powder nib, cedar, structure, crisp, nut, sweet, date

#### Topic 7

black, currant, cherry savory, red, pungent, blackberry, syrupy, deeply, berry

#### Topic 2

Chocolate, Dark, Woody

#### Topic 5

fresh, cut, fir, lightly, syrupy, drying, gently, hazelnut, round, acidity

#### Topic 8

wood, body, aromatic, nut, sweetness, rather, sweet, hints, tones,



#### Topic 3

structure, tart, sweet, zest, richly, savory, almond, satiny, toned, sweetly

#### Topic 6

cacao, nib, roasted, drying, lively, juicy, crisp, honeysuckle, acidity, sandalwood

#### Topic 9

#### Topics

#### Topic 1

Bright, Citrus, Floral



Cocoa, Nutty

#### Topic 7

**Red Berries** 

#### Topic 2

Chocolate, Dark, Woody

#### Topic 5

Sweet, Hazelnut, Pine

#### Topic 8

Nutty, Caramel Wood

#### Topic 3

Tart, Sweet, Smooth

#### Topic 6

Juicy, Honey, Cacao

#### Topic 9

Cherry, Vinous, Chocolate



#### **Topics: Coffee Centroids**



Caribou: Guatemala Temple: Amaro, E Flamenco: Yirgacheffe, E



Nongfu: Brazil/Guat. JBC: Guatemala El Gran Cafe: Guatemala

#### Topic 7

Terroir: Kirinyaga, K PT's: South-central, K Fratello: Nyeri, K

#### Topic 2

Simon Hsieh: Colombia Magnolia: Brazil (blend) The Beanery: n/a

#### Topic 5

Wonderstate: Worka, E JBC: Yirgacheffe, E Kunshan: Guat/Colo/Bra

#### Topic 8

Diedrich: Java, Indonesia Tully's: n/a Tully's: n/a

#### Topic 3

Baba Java: Yemen Paradise: Guatemala Java Blend: Yirgacheffe, E

#### Topic 6

Topeca: Sidamo, E BeanFruit: Colombia Greater Goods: Yirgacheffe, E

#### Topic 9

CofFeeling: Central, K De Clieu: Kenya Terroir: South-Central, K



# **Q4**Visualizations

#### Origin and Roast Level by Group



#### Reviews by Groups

Bright, Floral, Citrus



Cacao, Nutty, Clean



Red Berries



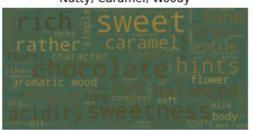
Chocolate, Dark, Woody



Sweet, Nut, Pine



Nutty, Caramel, Woody



Tart, Sweet, Smooth



Juicy, Honey, Cacao

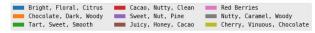


Cherry, Vinuous, Chocolate

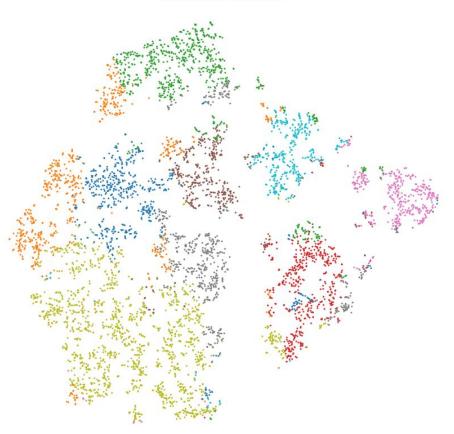


#### All Coffee Reviews Clustered by Category

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



Blind Reviews from Coffee-Review.com





Recommendations, Predictions, and More





#### **Coffee Recommendations**

#### Find Similar Reviews











#### **NMF Vector**

Creates a five 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

#### **Reviews to Overall Score**







**NMF Vector** As feature vector



Review Length
Word count and number
of different words



Sentiment Analysis
Vader
Positive, Negative,
Compound

#### **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

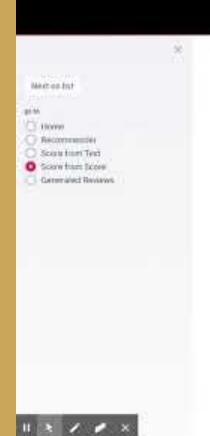


Costa Rica Cloza Estate

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:	0
		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.



#### Score Predictor (if you have the details)

故

With subcetegory scores as shown above, I predict your naffee to be review overall act. VI.......



# Questions?





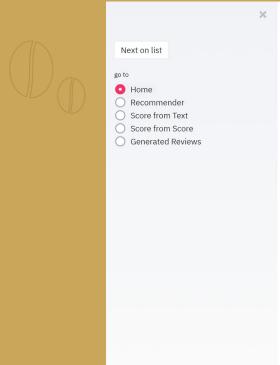
Costa Rica Cloza Estate

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:	0
		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.

#### **Streamlit**



#### 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
- 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



#### **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 desribed 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 descripton 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 descripton 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
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.souvenircoffee.com for more information.

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

8.598.24

8.06

8.72

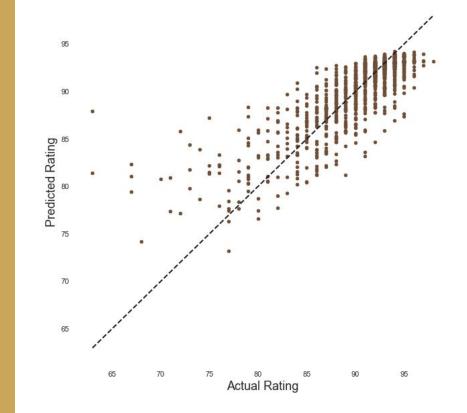
8.21

# RFR Model (from text)

 $r^2 = .688$ 

MAE = 1.7 points

Can we predict coffee ratings from reviews using a RandomForest

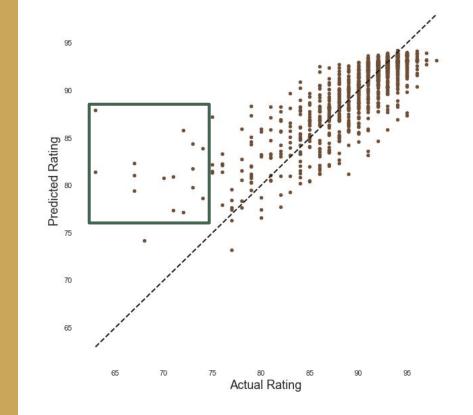


# RFR Model (from text)

#### Errors:

Negative reviews
Older reviews
Instant Coffee

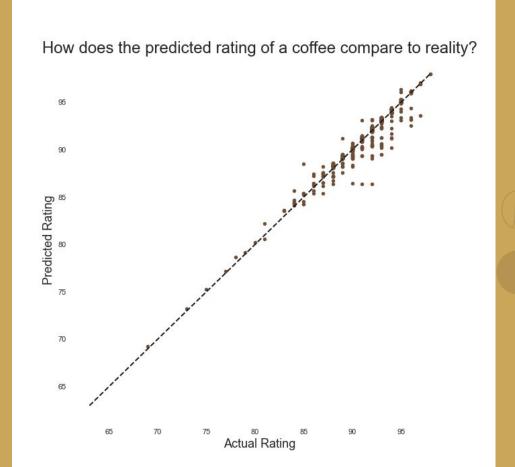
Can we predict coffee ratings from reviews using a RandomForest



# Linear Model (from scores)

 $r^2 = .943$ 

MAE = .36 points



# Linear Model (from scores)

#### Coefficients

Aroma = 1.04

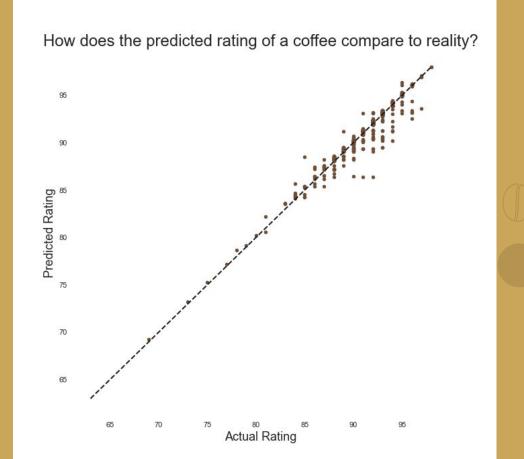
Body = 0.75

Flavor = 1.00

Aftertaste = 1.14

Acidity = 0.89

Intercept = 51.65



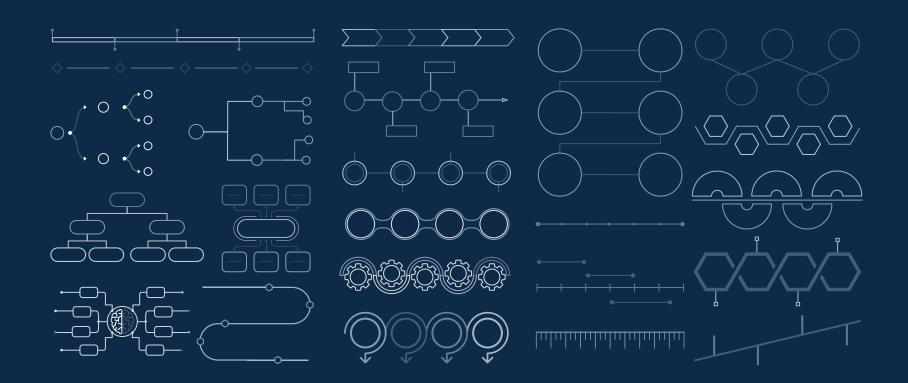
#### Use our editable graphic resources...

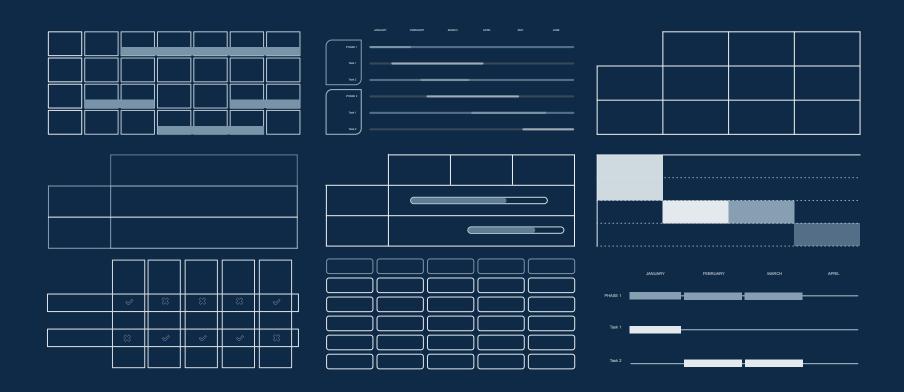
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want.

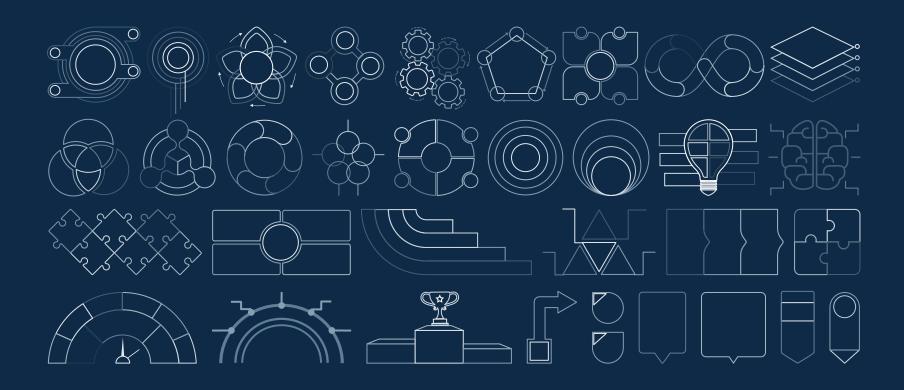
Group the resource again when you're done. You can also look for more infographics on Slidesgo.

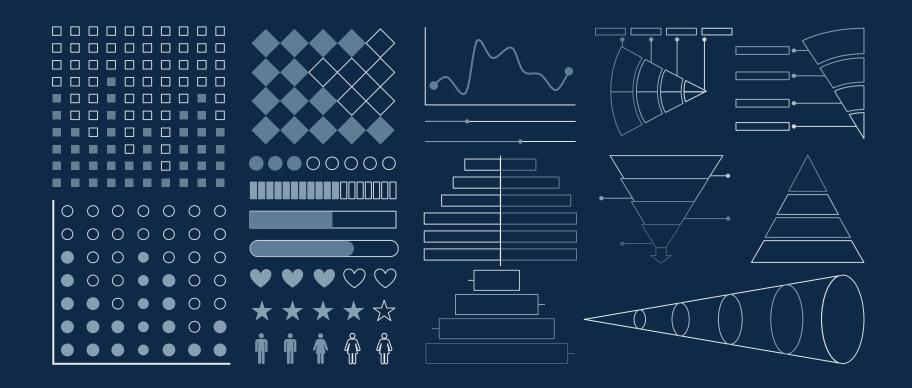












#### ...and our sets of editable icons

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**

# R 門為合品句品

#### **Medical Icons**



#### **Business Icons**

## (S) # # # # # # [S] 西野鱼属加拿里 學為參彩眉光會遍

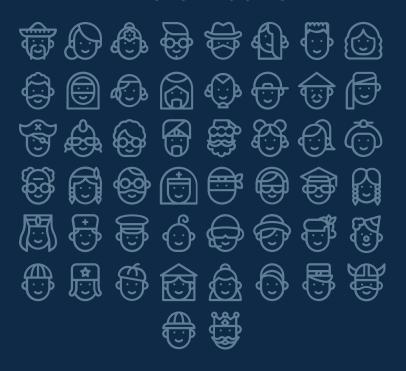
#### **Teamwork Icons**



#### Help & Support Icons

## 

#### **Avatar Icons**



#### Creative process Icons



#### Performing Arts Icons



#### **Nature Icons**



#### SEO & Marketing Icons



### **slides**go