Grounds for Analysis

Natural Language Processing:
Predicting Scores and Recommending Coffee

Metis 2021 -Ethan Feldman

Agenda

Data and Streamlit 01
The grind

Natural Language Processing 02
Blooming and immersion

Topic Modeling 03
Extraction

Recommendations and Predictions 04
The First sip

01

Data

The Grind

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
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
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
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.

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.

- Can I predict the scores based on the blind assessment?

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.

- Can I predict the scores based on the blind assessment?

- Can I recommend other coffees from the blind assessment?

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

Score Predictor
Predict coffee scores from reviews
Generated Reviews
Based on your input coffee, I predict it to receive a score of:

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

Review

Stopwords
English+
"Keurig, coffee, cup..."

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

Topics

Topic 1

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

Topic 2

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

Topic 3

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

Topic 4

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

Topic 5

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

Topic 6

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

Topic 7

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

Topic 8

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

Topic 9

fruit, toned, cherry, chocolate, sweet, rich, floral, brandy, wine, low

Topics

Topic 1

Bright, Floral, Citrus

Topic 2

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

Topic 3

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

Topic 4

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

Topic 5

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

Topic 6

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

Topic 7

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

Topic 8

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

Topic 9

fruit, toned, cherry, chocolate, sweet, rich, floral, brandy, wine, low

Topics

Topic 1

Bright, Citrus, Floral

Topic 2

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

Topic 3

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

Topic 4

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

Topic 5

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

Topic 6

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

Topic 7

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

Topic 8

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

Topic 9

fruit, toned, cherry, chocolate, sweet, rich, floral, brandy, wine, low

Topics

Topic 1

Bright, Citrus, Floral

Topic 2

Chocolate, Dark, Woody

Topic 3

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

Topic 4

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

Topic 5

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

Topic 6

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

Topic 7

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

Topic 8

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

Topic 9

fruit, toned, cherry, chocolate, sweet, rich, floral, brandy, wine, low

Topics

Topic 1

Bright, Citrus, Floral

Topic 2

Chocolate, Dark, Woody

Topic 3

Tart, Sweet, Smooth

Topic 4

Cocoa, Nutty Clean

Topic 5

Sweet, Hazelnut, Pine

Topic 6

Juicy, Honey, Cacao

Topic 7

Red Berries

Topic 8

Nutty, Caramel, Wood

Topic 9

Cherry, Vinous, Chocolate

04

Recommendations and Predictions

The 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 Ratings

Ratings
Overall and
Subcategories

NMF Vector
As feature vector

Review Length
Word count and number
of different words

Sentiment Analysis
Vader
Positive, Negative,
Compound

RFR Model (from text)

Explains 68.8% of variability in rating

MAE = 1.7 points

RFR Model (from text)

Errors:

Negative reviews
Older reviews
Instant Coffee

Next Steps

- GloVe embedding in place of TF-IDF
- Scrape additional sources of coffee descriptions
- Add sentiment analysis to recommendation system
- Filter by time for recommendations and analysis
- Improve review generator (GloVe, POS)

Questions?

Appendix

Topics: Coffee Centroids

Topic 1

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

Topic 2

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

Topic 3

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

Topic 4

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

Topic 5

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

Topic 6

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

Topic 7

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

Topic 8

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

Topic 9

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

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

Reviews to Overall Score

Interesting Visualizations

Coffees reviewed have shifted groups over time

Origin and Roast Level by Group

Bright, Floral, Citrus

Chocolate, Dark, Woody

Tart, Sweet, Smooth

Cacao, Nutty, Clean

Sweet, Nut, Pine

Juicy, Honey, Cacao

Red Berries

Nutty, Caramel, Woody

Cherry, Vinuous, Chocolate

Reviews by Groups

Bright, Floral, Citrus

Chocolate, Dark, Woody

Tart, Sweet, Smooth

Cacao, Nutty, Clean

Sweet, Nut, Pine

Juicy, Honey, Cacao

Red Berries

Nutty, Caramel, Woody

Cherry, Vinuous, Chocolate

How do coffees tend toward topics on average?

Counter Culture Coffee Taster's Flavor Wheel

Body Descriptors

Light	Medium	Heavy
Watery Skim Milk Tea-like	Round 2% Milk Creamy	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 TASTERS FLAVOR WHEEL CREATED USING THE SENSORY LEXICON DEVELOPED BY WORLD COFFEE RESEARCH
© 2016 SCA AND WCR

Distribution of coffees to topic categories
1: Bright, Floral, Citrus
2: Chocolate, Dark, Woody
3: Tart, Sweet, Smooth
4: Cacao, Nutty, Clean
5: Sweet, Hazelnut, Pine
6: Juicy, Honey, Cacao
7: Red Berries
8: Nutty, Caramel, Woody
9: Cherry, Vinous, Chocolate
Number of Coffees Assigned
0
200
400
600
800
1000
1200
8
3
2
9
1
5
4
6
7
Topic Assignment

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

Streamlit App Demo

In Case of Video Failure

Streamlit

Next on list
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 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
Coffee Origin:	Alajuela, Costa Rica
Roast Level:	Medium-Light
Agtron:	58/80
Est. Price:	$19.00/12 ounces

Review Date:	February 2021	
Aroma:	9	8.59
Acidity /Structure :	8	8.24
Body:	9	8.06
Flavor:	9	8.72
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.souvenir-coffee.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

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.

JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
PHASE 1
Task 1
Task 2
PHASE 2
Task 1
Task 2

JANUARY
FEBRUARY
MARCH
APRIL
PHASE 1
Task 1
Task 2

...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

slidesgo