Coffee Shops in San Francisco

Looking for meaningful, unconventional and unique experiences

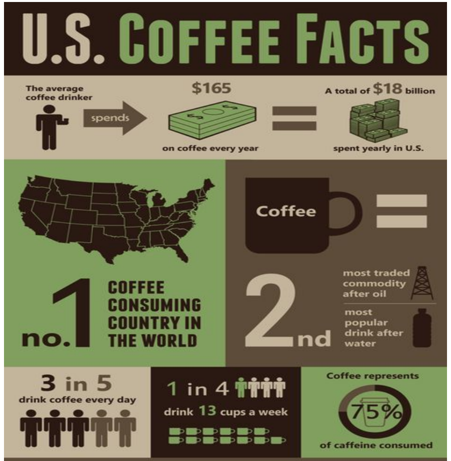
Introduction

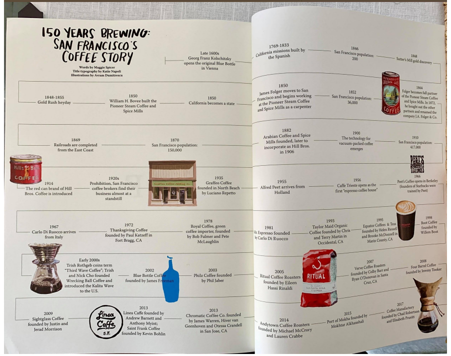
- Coffee in the **U.S.**
  - The average coffee drinker consumes **3.1 cups of coffee daily**
  - There are about **100 million coffee drinkers** in the U.S.
  - American workers spend about **$20 per week** on coffee

- Why **San Francisco**?

  > "()... you'll find an impressive number of coffee startups in San Francisco, and a new generation of tech-obsessed coffee-drinkers reshaping the city with every blink." (**Drift, Volume 7: San Francisco, July 15, 2018**)

- Which elements define a pleasurable or disappointing experiences?




150 YEARS BREWING: SAN FRANCISCO'S COFFEE STORY

Words by Maggie Spicer
Title typography by Katie Napoli
Illustrations by Avram Dumitrescu

Late 1600s
Georg Franz Kolschitzky opens the original Blue Bottle in Vienna

1769-1833
California missions built by the Spanish

1846
San Francisco population: 200

1848
Sutter's Mill gold discovery

1848-1855
Gold Rush heyday

1850
William H. Bovee built the Pioneer Steam Coffee and Spice Mills

1850
California becomes a state

1850
James Folger moves to San Francisco and begins working at the Pioneer Steam Coffee and Spice Mills as a carpenter

1852
San Francisco population: 36,000

1864
Folger becomes full partner of the Pioneer Steam Coffee and Spice Mills. In 1872, he bought out the other partners and renamed the company J.A. Folger & Co.

1869
Railroads are completed from the East Coast

1870
San Francisco population: 150,000

1882
Arabian Coffee and Spice Mills founded, later to incorporate as Hill Bros. in 1906

1900
The technology for vacuum-packed coffee emerges

1910
San Francisco population: 417,000

1914
The red can brand of Hill Bros. Coffee is introduced

1920s
Prohibition, San Francisco coffee brokers find their business almost at a standstill

1935
Graffeo Coffee founded in North Beach by Luciano Repetto

1955
Alfred Peet arrives from Holland

1956
Caffe Trieste opens as the first "espresso coffee house"

1966
Peet's Coffee opens in Berkeley (founders of Starbucks were trained by Peet)

1967
Carlo Di Ruocco arrives from Italy

1972
Thanksgiving Coffee founded by Paul Katzeff in Fort Bragg, CA

1978
Royal Coffee, green coffee importer, founded by Bob Fulmer and Pete McLaughlin

1981
Mr. Espresso founded by Carlo Di Ruocco

1993
Taylor Maid Organic Coffee founded by Chris and Terry Martin in Occidental, CA

1995
Equator Coffees & Teas founded by Helen Russell and Brooke McDonnell in Marin County, CA

1998
Boot Coffee founded by Willem Boot

Early 2000s
Trish Rothgeb coins term "Third Wave Coffee"; Trish and Nick Cho founded Wrecking Ball Coffee and introduced the Kalita Wave to the U.S.

2002
Blue Bottle Coffee founded by James Freeman

2003
Philz Coffee founded by Phil Jaber

2005
Ritual Coffee Roasters founded by Eileen Hassi Rinaldi

2007
Verve Coffee Roasters founded by Colby Barr and Ryan O'Donovan in Santa Cruz, CA

2008
Four Barrel Coffee founded by Jeremy Tooker

2009
Sightglass Coffee founded by Justin and Jerad Morrison

2013
Linea Caffe founded by Andrew Barnett and Anthony Myint; Saint Frank Coffee founded by Kevin Bohlin

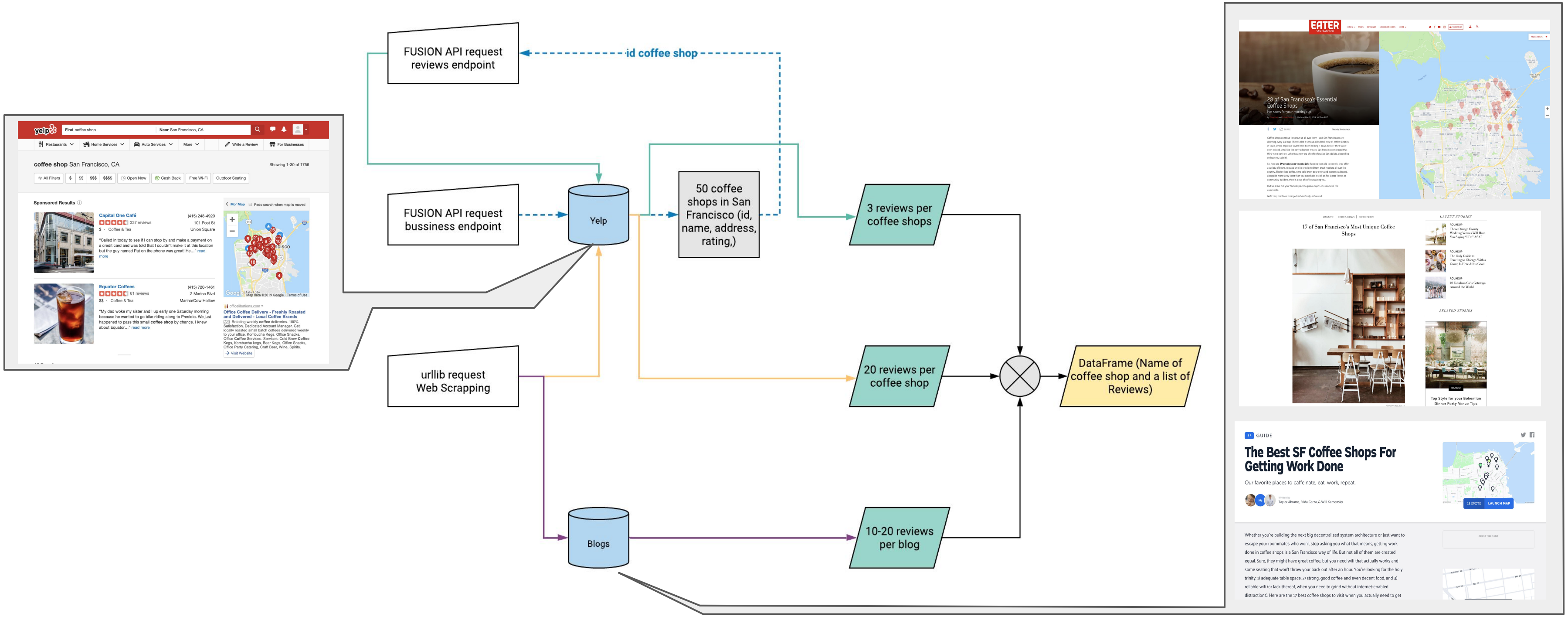
2013
Chromatic Coffee Co. founded by James Warren, Hiver van Geenhoven and Otessa Crandell in San Jose, CA

2014
Andytown Coffee Roasters founded by Michael McCrory and Lauren Crabbe

2015
Port of Mokha founded by Mokhtar Alkhanshali

2017
Coffee Manufactory founded by Chad Robertson and Elisabeth Prueitt

Acquisition



FUSION API request reviews endpoint

id coffee shop

FUSION API request bussiness endpoint

Yelp

50 coffee shops in San Francisco (id, name, address, rating,)

3 reviews per coffee shops

20 reviews per coffee shop

urllib request Web Scrapping

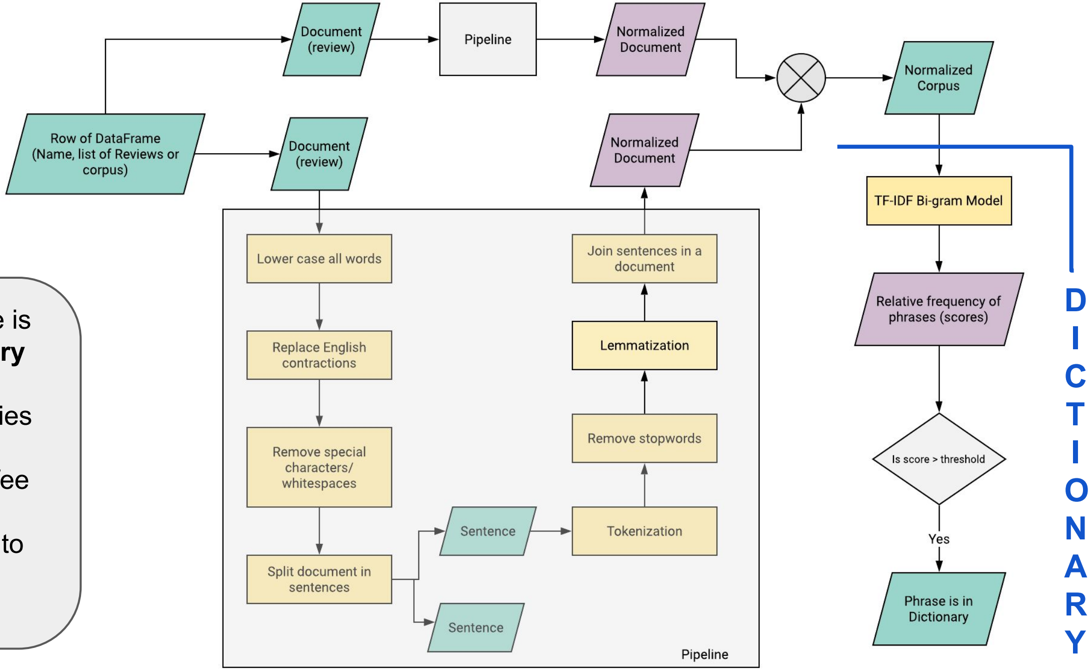
Blogs

10-20 reviews per blog

DataFrame (Name of coffee shop and a list of Reviews)

yelp
Find coffee shop
Near San Francisco, CA
coffee shop San Francisco, CA
Sponsored Results
Capital One Café
Equator Coffees
EATER
28 of San Francisco's Essential Coffee Shops
17 of San Francisco's Most Unique Coffee Shops
GUIDE
The Best SF Coffee Shops For Getting Work Done
Our favorite places to caffeinate, eat, work, repeat.
LAUNCH MAP

Wrangling and Pre-processing

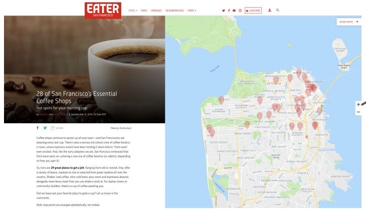


Why **Bi-grams**? We are looking for expressions as: **Pour over**, **latte art, cold brew**, **best coffee**, **amazing view**, **quiet place**, **seat available**, for instance.

The preprocessing pipeline is used to to build a **Dictionary of key-phrases**. They are labeled in different categories and sections related to the experience around the coffee shop.
The same pipeline is used to work with **normalized reviews** later.

A **threshold** is used for avoiding rescue expressions with lower relative frequency.

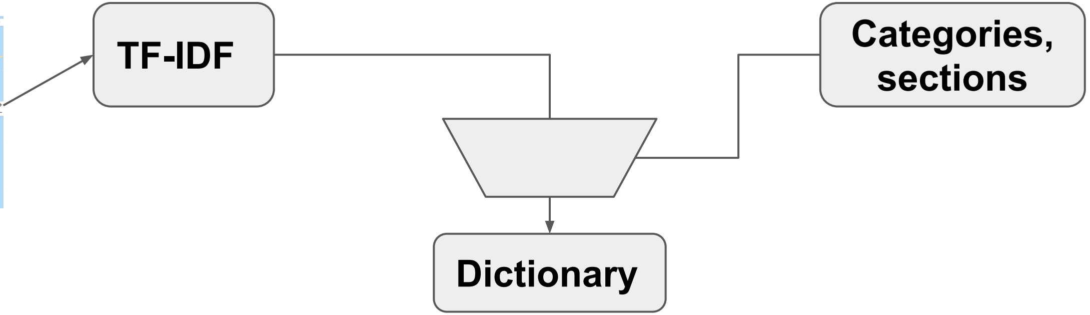
Initial findings: Analysis of the Dictionary

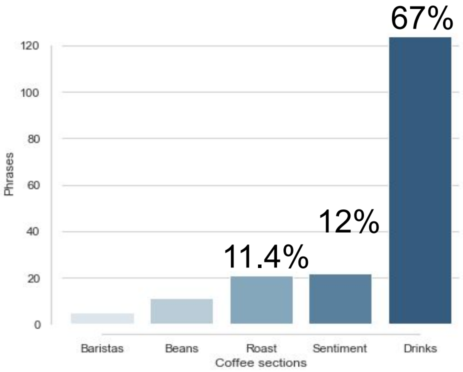


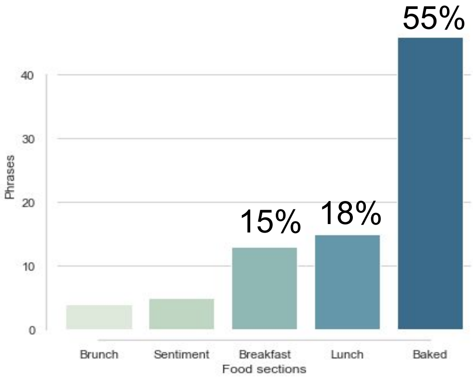
EATER

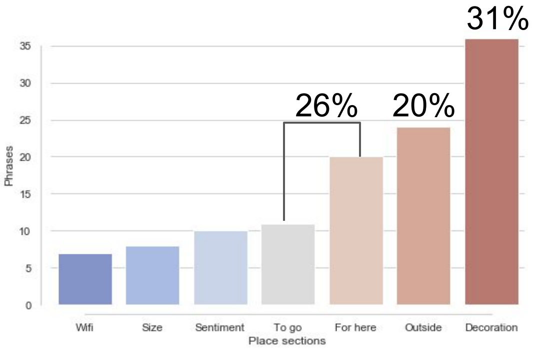
28 of San Francisco's Essential Coffee Shops

Hot spots for your morning cup



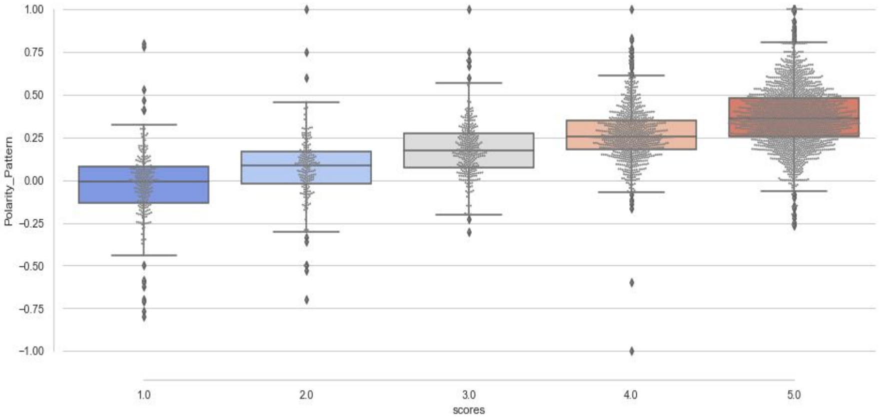




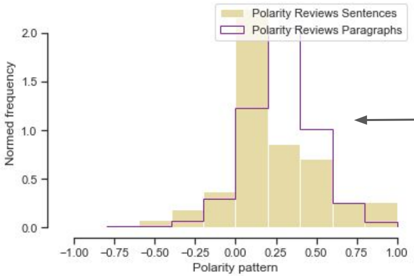


Initial findings: Sentiment Analysis of reviews

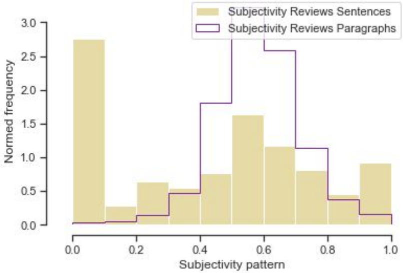
Could we measure how much **positive**, **negative** or **neutral** is the information of customer reviews?



How much the **polarity** and **subjectivity** change splitting reviews into sentences?



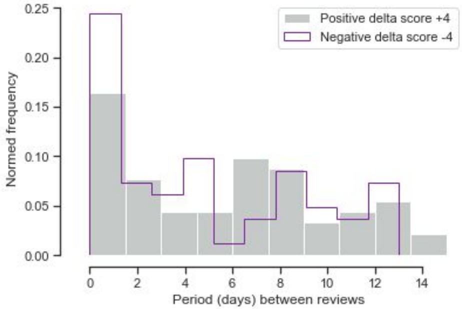
The **mean shifted at the left** when we split reviews.



Change the shape of the distribution: more neutral and subjective information.

Initial findings: Period between reviews

How long take to write the next review for a specific coffee shop?

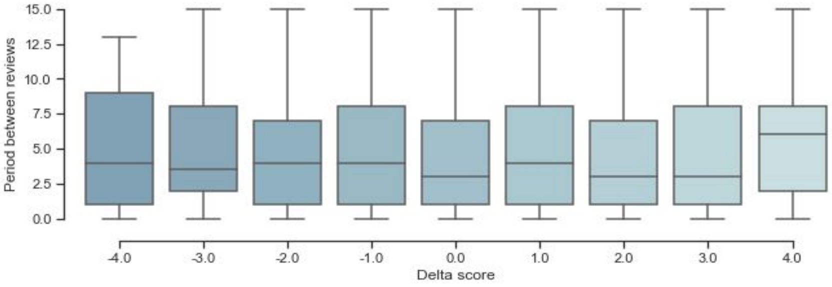


The difference between scores is **delta score** and the difference in days as the **period between reviews**

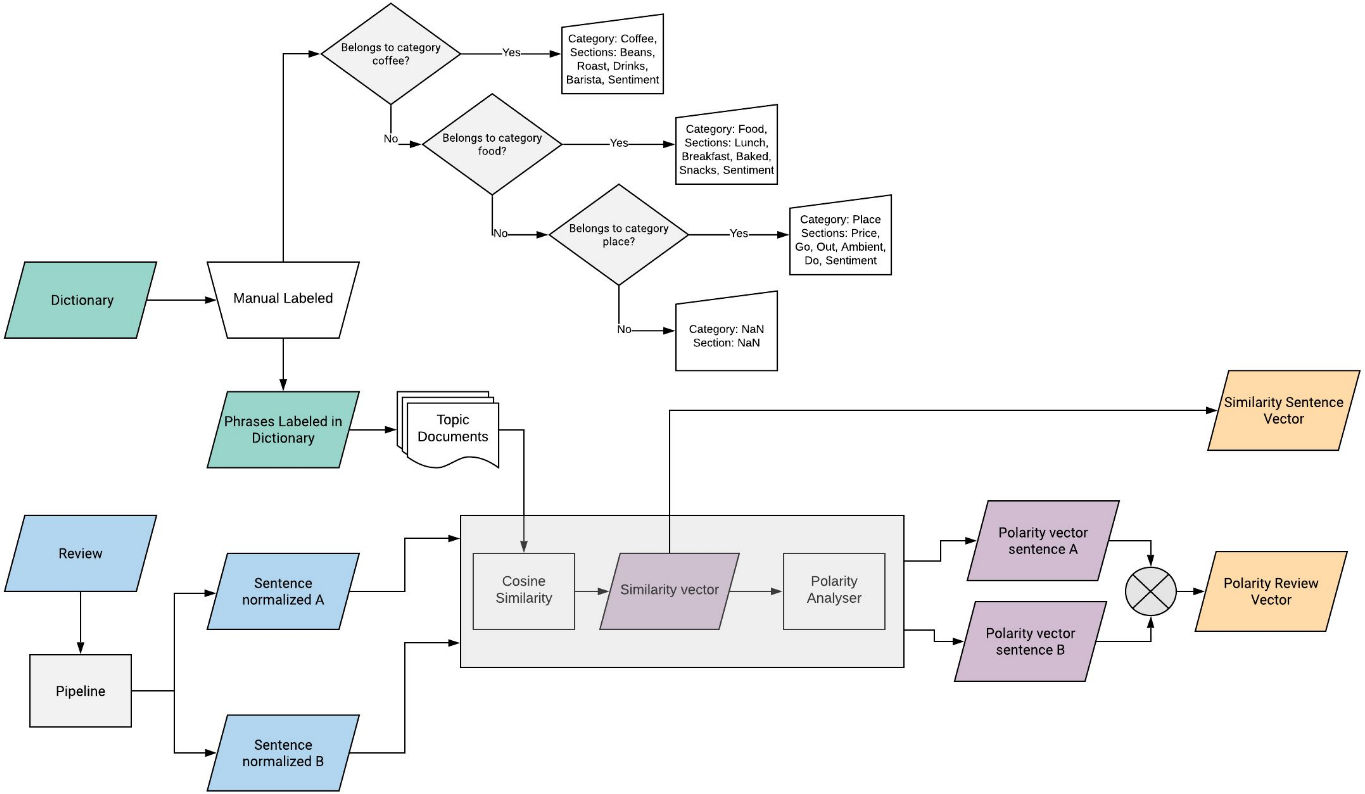
The mean for **positive delta score is 5.6 days** and for **negative is 4.92 days**

Period between all possible Delta Score:

Positive Delta Score of **+4** takes **more time** than the rest of delta scores

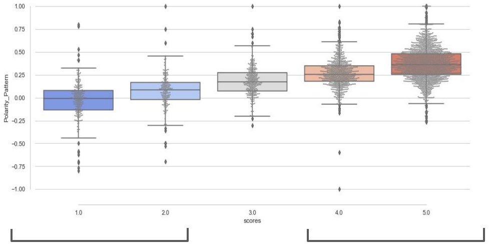


SML: How feature vectors are built?


Belongs to category coffee?
Yes
Category: Coffee, Sections: Beans, Roast, Drinks, Barista, Sentiment
No
Belongs to category food?
Yes
Category: Food, Sections: Lunch, Breakfast, Baked, Snacks, Sentiment
No
Belongs to category place?
Yes
Category: Place Sections: Price, Go, Out, Ambient, Do, Sentiment
No
Category: NaN Section: NaN
Dictionary
Manual Labeled
Phrases Labeled in Dictionary
Topic Documents
Similarity Sentence Vector
Review
Pipeline
Sentence normalized A
Sentence normalized B
Cosine Similarity
Similarity vector
Polarity Analyser
Polarity vector sentence A
Polarity vector sentence B
Polarity Review Vector

SML: Selecting features and output vectors

Output vectors



Negative reviews

Positive reviews

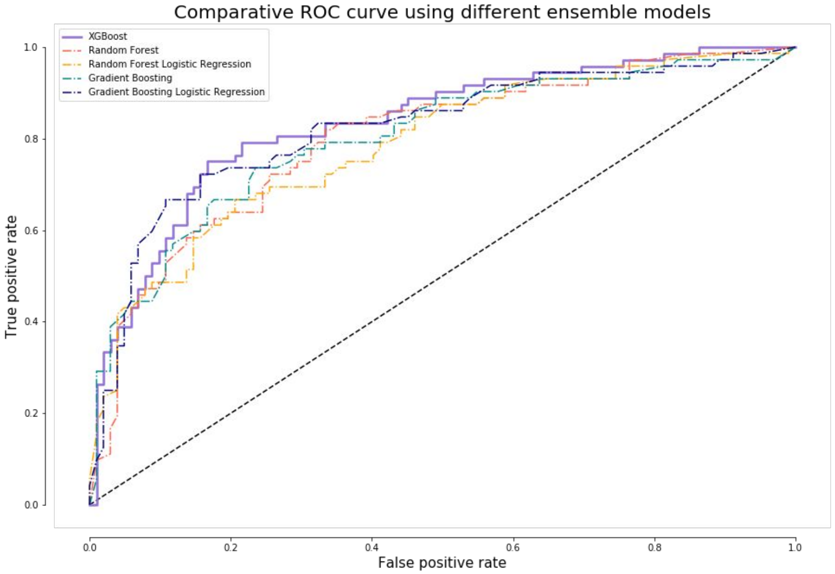
Selection of best feature vectors

6 features more strongly related to the output variable

Statistical **test chi-squared** with a **confidence level of 5%** (SelectKBest)

ID	SPECS	SCORE	ID	SPECS	SCORE
15	**food sentiment**	5.413	9	**price**	1.068
10	**place sentiment**	3.445	3	**barista**	1.065
4	**coffee sentiment**	2.620	2	**drinks**	0.850
13	**breakfast**	2.499	5	**go**	0.815
8	**ambient**	2.215	14	snacks	0.663
11	**baked**	1.955	12	lunch	0.562
6	**do**	1.552	0	beans	0.165
7	**out**	1.171	1	roast	0.086

SML: Training, tunning and testing models

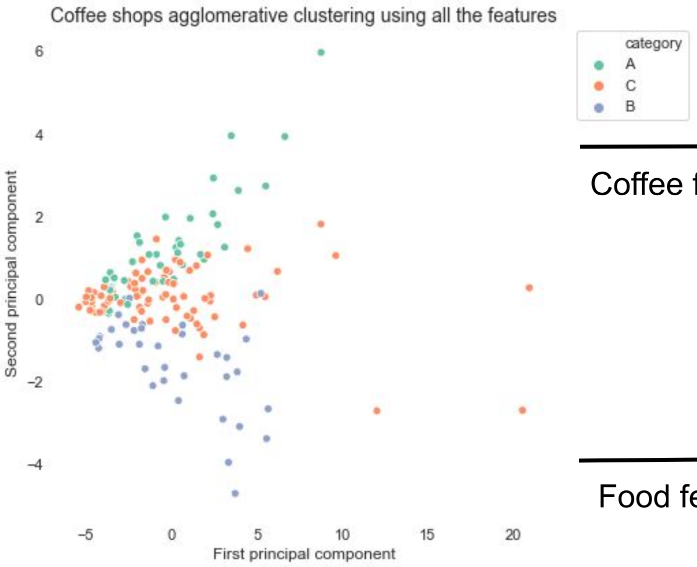


Model	AUC	Accuracy
XGBoost	**83.46%**	**77.01%**
Random Forest	79.71%	71.84%
Logistic Regression Random Forest	78.50%	71.26%
Gradient Boosting	80.22%	75.29%
Logistic Regression Gradient Boosting	**82.16%**	**73.56%**

XGBoost	TP Reviews	TN Reviews
PP Review	77	25
NP Review	15	57

UML: Agglomerative Clustering using all features

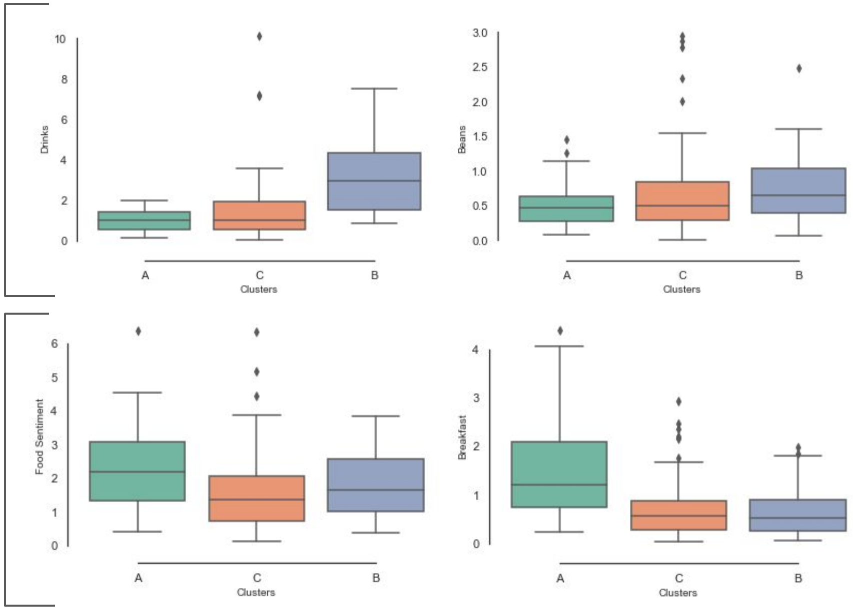
Bi-dimensional representation (PCA)



How **features** are **determining** these clusters?

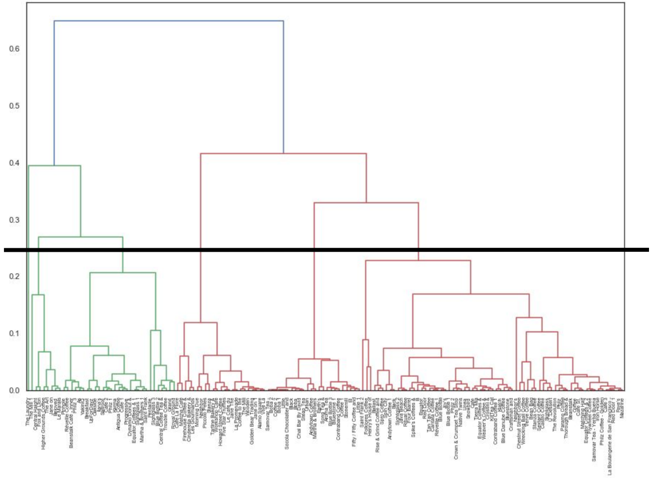
Coffee features

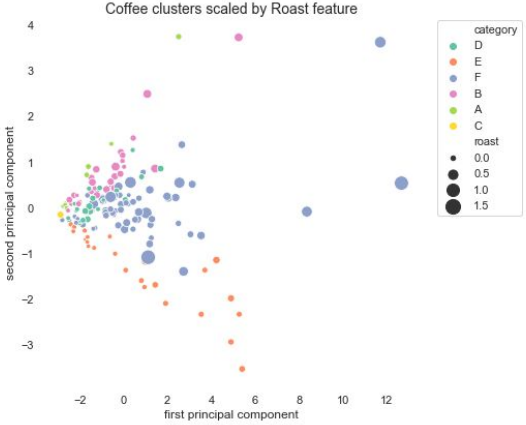
Food features



UML: Dendrograms splitting features

At distance of **0.25** coffee features divide coffee shops in **6 groups**



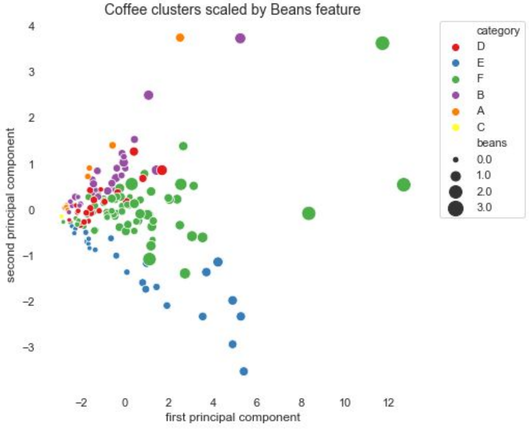


Cluster F: Andytown, Mazarine, Four Barrel, Red Door, Ritual Coffee, Saint Frank, Sightglass, Equator, Blue Bottle, Wrecking Ball, among others.

Cluster E, Mostly tea and chai stores and chocolatier business

Cluster F higher roasting and beans sum, followed by B and D. Clusters A and E, is the lowest.

Cluster A, Le Marais, Art's Cafe, some bakeries among others

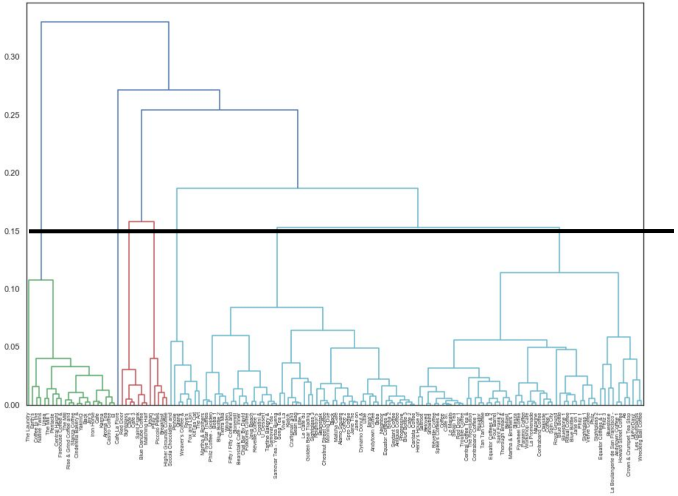


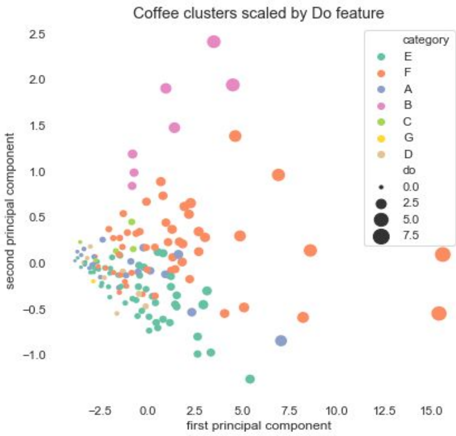
UML: Dendrograms splitting features

At distance of **0.15** coffee features divide coffee shops in **7 groups**

Cluster B higher Do values (feature about what to do in coffee shops): Matching Half, Red Door, Saint Frank, Sightglass (study, work)

Cluster C one of the lowest Do values. Piccolo Petes Urban, Higher Ground Coffee (bar styles, going to share with friends)





Conclusion

- **Blogs put more attention in:**
  - Drinks, baked items and decoration in their reviews.

- **Features most strongly vinculated with the score of customer reviews:**
  - **Sentiments** associated with **drinks**, **place** and **food**, **breakfast**, **ambient**, **outside**, **baked elements**, **baristas**, **available of seats**, **what to do** and **price**.
  - Less relationship: **quality** and **variety of beans**, **roasting,** little **snacks** and **lunch** options.

Conclusion

- **Supervised Learning:**
    - We evaluated different **Decision Trees** models to predict sentiment patterns in customer reviews. **XGBoost** got the best performance, with an AUC of 83.4% and an accuracy of 77 %.

- **Unsupervised Learning:**
    - Using **Agglomerative Clustering** and **dendrograms** as visualization tool, we found clusters with coffee shops distinguishing styles, what to do there and how much interested is the people in talking about **beans**, **roasting**, **drinks, places** and **food**.

Resources

- Yelp API (FUSION) and web scraping from Yelp
- *“28 of San Francisco's Essential Coffee Shops: Hot spots for your morning cup*” (Ellen Fort and Caleb Pershan). Available here.
- *“17 of San Francisco's Most Unique Coffee Shops* (Katie Bush). Available here.
- *“The Best SF Coffee Shops For Getting Work Done”* (Taylor Abrams, Frida Garza, and Will Kamensky). Available here.
- Coffee gives me superpowers (Ryoko Iwata). Published on April 7, 2015
- DRIFT San Francisco (A. Goldberg, Velasco, Lee, E. Goldberg and Spicer). Published on July 15, 2018
- XGBoost: The Excalibur for Everyone (Raghu Raj Rai). Towards Data Science. Available here
- A Beginner’s guide to XGBoost (George Seif). Towards Data Science. Available here
- Traditional Methods for Text Data (Dipanjan Sarkar). Towards Data Science. Available here
- Practical Statistics for Data Scientist (Peter Bruce and Andrew Bruce). O'REILLY, 2017.
- Text Classification is Your New Secret Weapon (Adam Geitgey, Medium). Available here
- A Practitioner’s Guide to Natural Language Processing (Part I) - Processing and Understanding Text (Dipanjan Sarkar). Towards Data Science. Available here