Brewery Review Analysis

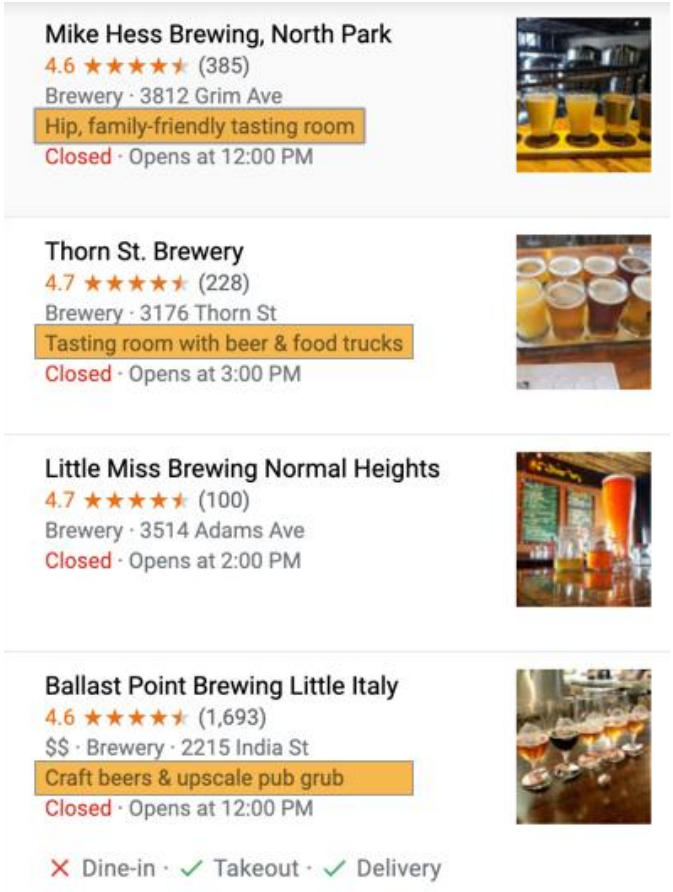
NLP and Named Entity Recognition (NER)

Chris J Mears • 01.04.2021
Springboard Data Science Career Track (April 2020 Cohort)

About Me

- Springboard Data Science Career Track April 2020 Cohort
- 10 Years in Tech as a Full-Stack Web Developer
- Creator and Maintainer of open-source dataset and API, Open Brewery DB

The Problem



Writing individual descriptions manually is time-intensive and inefficient.

How can we generate a short descriptive summarization of a business through text reviews?

Scope

DSM & Model Training

- Focus on the Data Science Method
- Train Named Entity Recognition model via spaCy

Breweries

- Simplify annotations
- Domain knowledge

Problem Breakdown

Manual Approach

- Human writing description
- Inefficient and extremely time-consuming

Rules-Based Approach

- Better, but still limited to the rules selected

spaCy

Solution: Named Entity Recognition

Statistical Model

- Use spaCy statistical model for training
- Every “decision” is a probability

Generalization

- Context aware
- Not memorization like pattern-recognition

The Data

Data Wrangling

- Scraped from Beer Advocate
  - Public data only
- Ethical scraping
  - Checked Terms of Use, off-hours, time-delay, user-agent, non-greedy

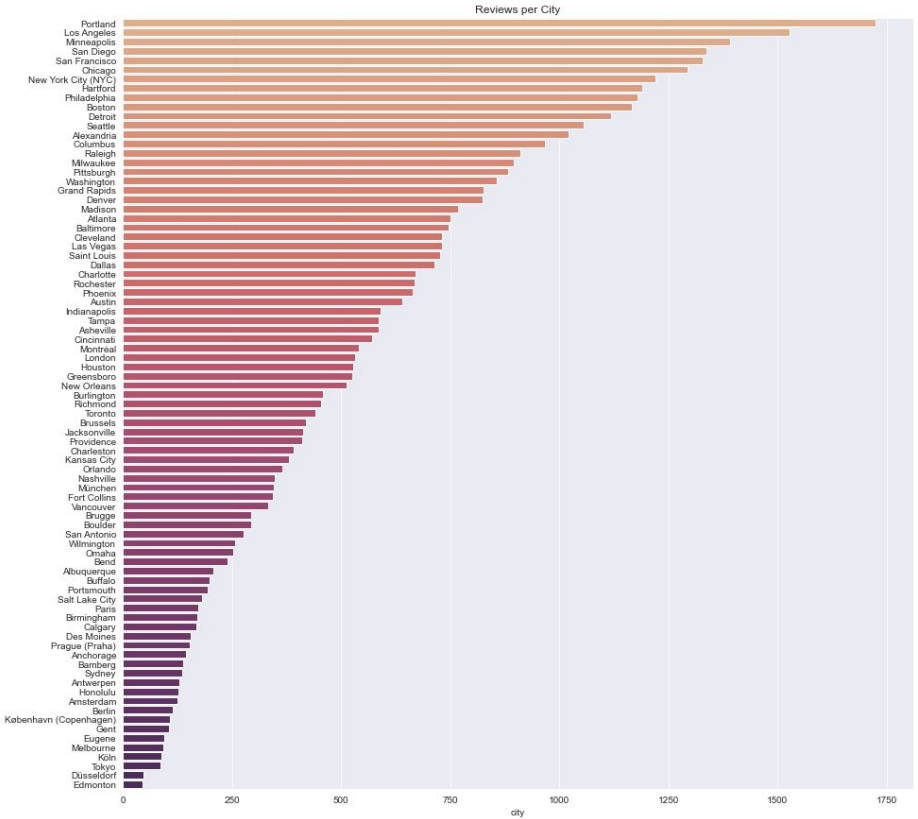
Data Wrangling

- 86 cities
- 4,984 breweries
- 45,417 reviews

Exploratory Data Analysis

check
town
around
owner
solid
side
make
long
big
area
review
two
good
outside
pretty
menu
good beer
tap
huge
service
night
will
store
awesome
order
cool
good selection
game
server
know
went
old
love
staff
although
food
stuff
brewpub
going
helpful
hard
far
left
bar
many
need
time
come
vibe
always
great
fantastic
guy
walk
several
first
lot
least
even
room
small
seemed
Overall
plus
right
shop
seem
etc
experience
beer bar
great place

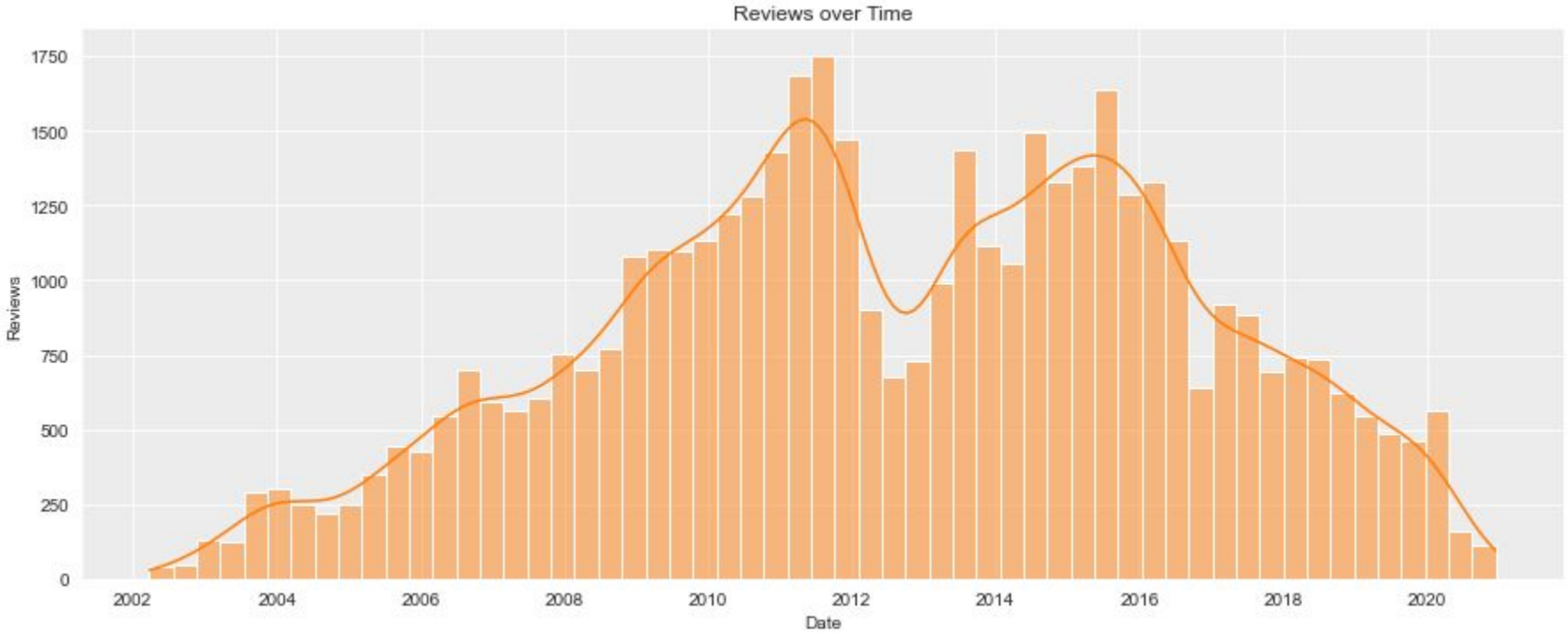
look
front
brewery
way
back
still
excellent
crowd
list
available
burger
spot
worth
go
find
probably
actually
inside
draft
option
wife
able
great beer
everything
really
average
feel
nice
hit
bad
stop
part
said
amazing
busy
work
street
beer
something
try
friend
couple
bartender
though
place
craft
selection
friendly
atmosphere
located
see
bit
sure
lunch
trip
people
definitely
little
interesting
knowledgeable
now
looking
came
favorite
downtown
wine
visit
quality
city
drink
take
never
enjoy
especially
table
pizza
better
u
want
price
got
decent
Belgian
usually
offer
found
offering
quite
think
maybe
dinner
restaurant
best
say
well
selection
pint
year
made
kind
location
bottle
give
much
one
space
wall
great selection
thing
pub
sit
large
along
new
tried
day
visited
beer
IPA
pretty good
another
plenty
brew
style
local
building
open



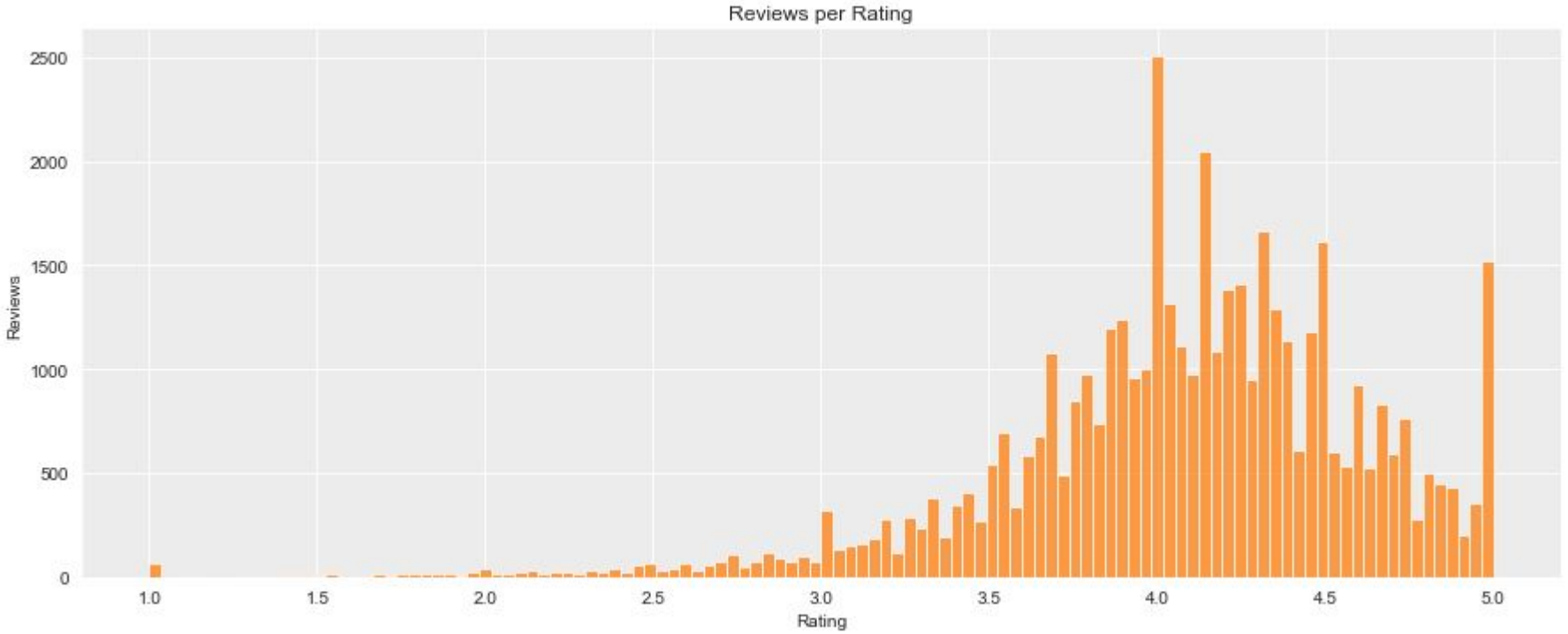
Top 5 Cities by Reviews

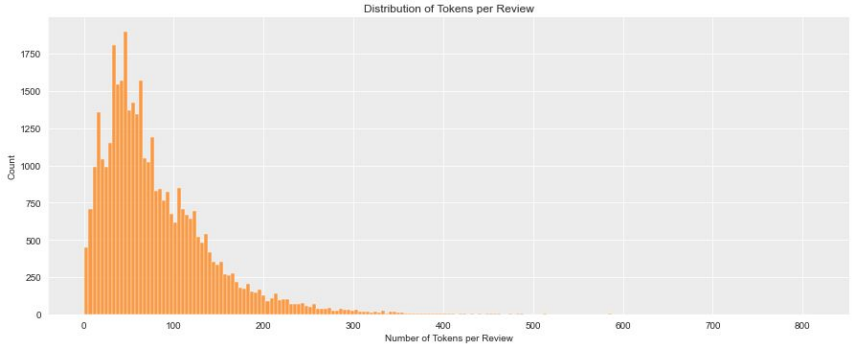
1. Portland, Oregon
2. Los Angeles, California
3. Minneapolis, California
4. San Diego, California
5. San Francisco, California

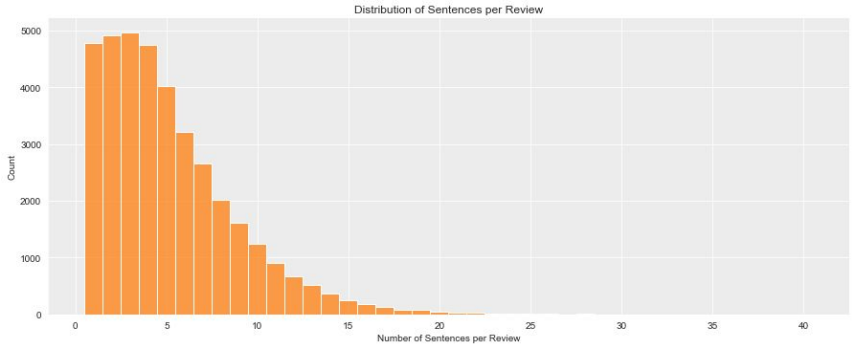
Reviews over Time



Reviews per Rating







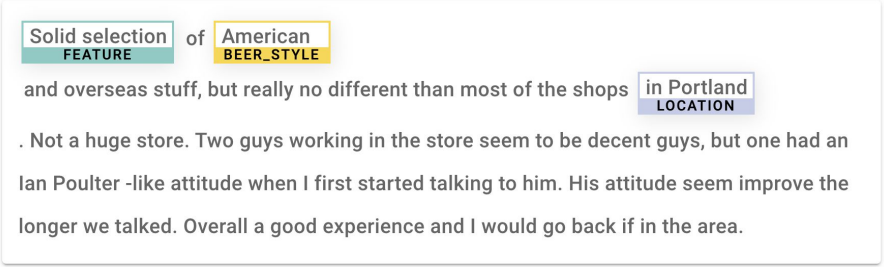
Token and Sentence Distribution over Reviews

- Most are short, but there's a long-tail of long-winded writers.

Annotation

Annotation Labels

- **FEATURE**
  - Brewery feature including "spacious", "selection", "dog-friendly", "patio", etc.
- **LOCATION**
  - City or general location, including landmarks. "San Diego", "in downtown", "near Petco Park".
- **BEER_STYLE**
  - Beer style (Belgian, IPA), character (hoppy, hazy), or specific beer name (Speedway Stout)
- **BREWERY**
  - Any brewery name. "Half Door," "Alesmith," "Dogfish Head"



Doccano

- Text Annotation Tool
- Open-Source Software

Model Training & Tuning

Model Training

- 200 annotation samples
- Convert
  - Doccano JSONL to NER JSONL to spaCy JSON
- 80% / 20% Train / Test Split
- Transfer Learning
  - **en-core-web-sm**
  - **en-core-web-md**

Evaluation

Model Evaluation

- **NER Loss**
  - Training loss for named entity recognizer. Should decrease, but usually not to 0.
- **NER P.**
  - NER precision on development data. Should increase.
- **NER R.**
  - NER recall on development data. Should increase.
- **NER F.**
  - NER F-score on development data. Should increase.

Model Evaluation

Metric	Blank Model	en-core-web-sm	en-core-web-md
Time	0.16 s	0.46 s	0.56 s
Words	4226	4226	4226
Words/s	27195	9212	7585
NER P	58.40	62.76	**62.94**
NER R	47.71	**59.48**	58.82
NER F	52.52	**61.07**	60.81

Conclusions

Tucked away in an old house near Petco. Loved the food, the beer was average at best. Their NE IPA BEER STYLE wasn't very good, but their dark beers were worth the pour. I might go back, it's not a bad choice but in San Diego LOCATION there's a lot of better places.

What a fantastic discovery on our first day in San Diego LOCATION . Knowledgeable staff with put being snobby. Wonderful setting in a old Victorian house FEATURE . Many of the beers were unfiltered yummy ness. Good selection FEATURE of local Honda too.

I'll start off by saying I really enjoyed my visit here. It's located in a former two story house near Gaslamp District and Petco Park. They have patio seating outside with a bar area inside and tables on two levels. The atmosphere here is unique FEATURE . It has a nice inviting feel. For me, it's somewhat similar to the same vibe I get when visiting Queenstown Public House near Little Italy.

The house beers were solid. I had Imperial Stout BEER STYLE , IPA BEER STYLE , and a Berliner BEER STYLE Weissbier. They also had several guest beers FEATURE on tap FEATURE . Beer and cocktail menu in a leather portfolio. Very nice FEATURE . I had their meat pie with braised short rib. Delicious! This is a nice little FEATURE place. I'll be back for sure.

During my San Diego LOCATION visit, I hit this place up two times. Located in a house with a nice sized porch that has plenty of seats. The bar is just inside and it is full of old wood and is dark, and gives you a cool pub feel. Definitely a younger crowd here too. There were about 10-12 brews on tap FEATURE , and they covered many styles including some sours BEER STYLE . The Double IPA BEER STYLE and high hopped pale BEER STYLE ales were very good FEATURE . Prices are reasonable and service was good. This is a must stop for anyone visiting the downtown area LOCATION .

Half Door is a recent entry to San Diego LOCATION 's magnificent beer scene. Located in a regular two-story clapboard house, a few blocks east of the touristy GasLight District, this place is a sheer delight. First, the building itself: an inviting house with a large terrace for outdoor

- Decent name recognition, but could still use more annotations and examples.
- Final Model
  - Transfer learning with en-core-web-sm existing model and ~50 epochs.
  - 300-500 annotation samples

Further Research

- More annotations with multiple passes for verification and consistency.
- Use context-free grammar or Markov chains to generate sentences.
    - Obtain the parts-of-speech from the named entities to construct from a script.
    - Example: {Feature - Adj} {Feature - Noun} featuring {Beer Style}. “Spacious brewery and taproom featuring hazy IPAs.”

Questions?