

The background of the slide features a blurred photograph of the Disneyland Paris castle at dusk or night. The castle is illuminated with warm lights, and the sky is dark with some clouds. Overlaid on the bottom right of the image is the Disneyland Park logo, which consists of the word "Disneyland" in a large, stylized blue font with a white outline, and "Park" in a smaller, bold blue font below it.

**Disneyland**  
Park

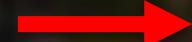
# Understanding Visitors

NLP Approach using Python

# Opinion Avenue



Google Maps



Disneyland Park X

The stars always try to give me and my family good service."

[Write a review](#)

---

Reviews

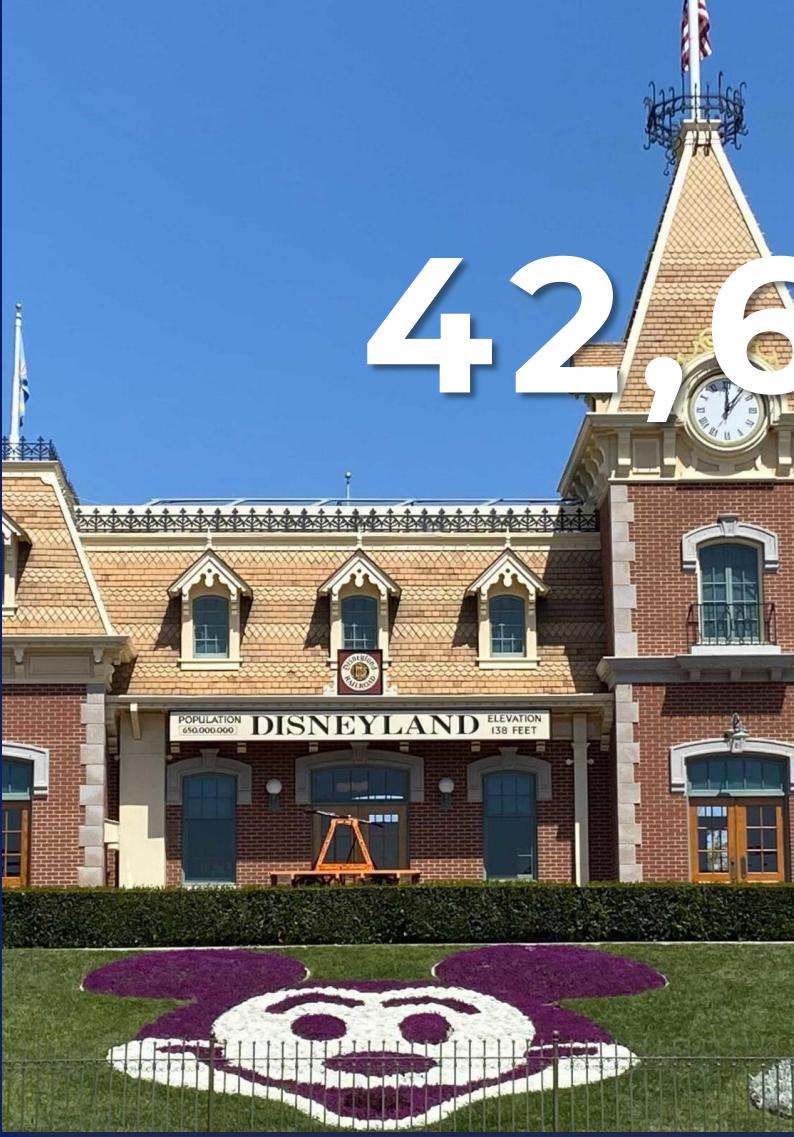
All fast pass 1822 star wars land 569

mickey 262 haunted mansion 194 +6

Eric Yang  
Local Guide · 175 reviews · 931 photos ...

a week ago NEW

Fantastic place, first time visiting in over a decade. I highly recommend buying tickets, making the reservation, and buying the parking all in advance. If you're just going to Disneyland Park, I recommend parking in the Toy Story parking ... [More](#)

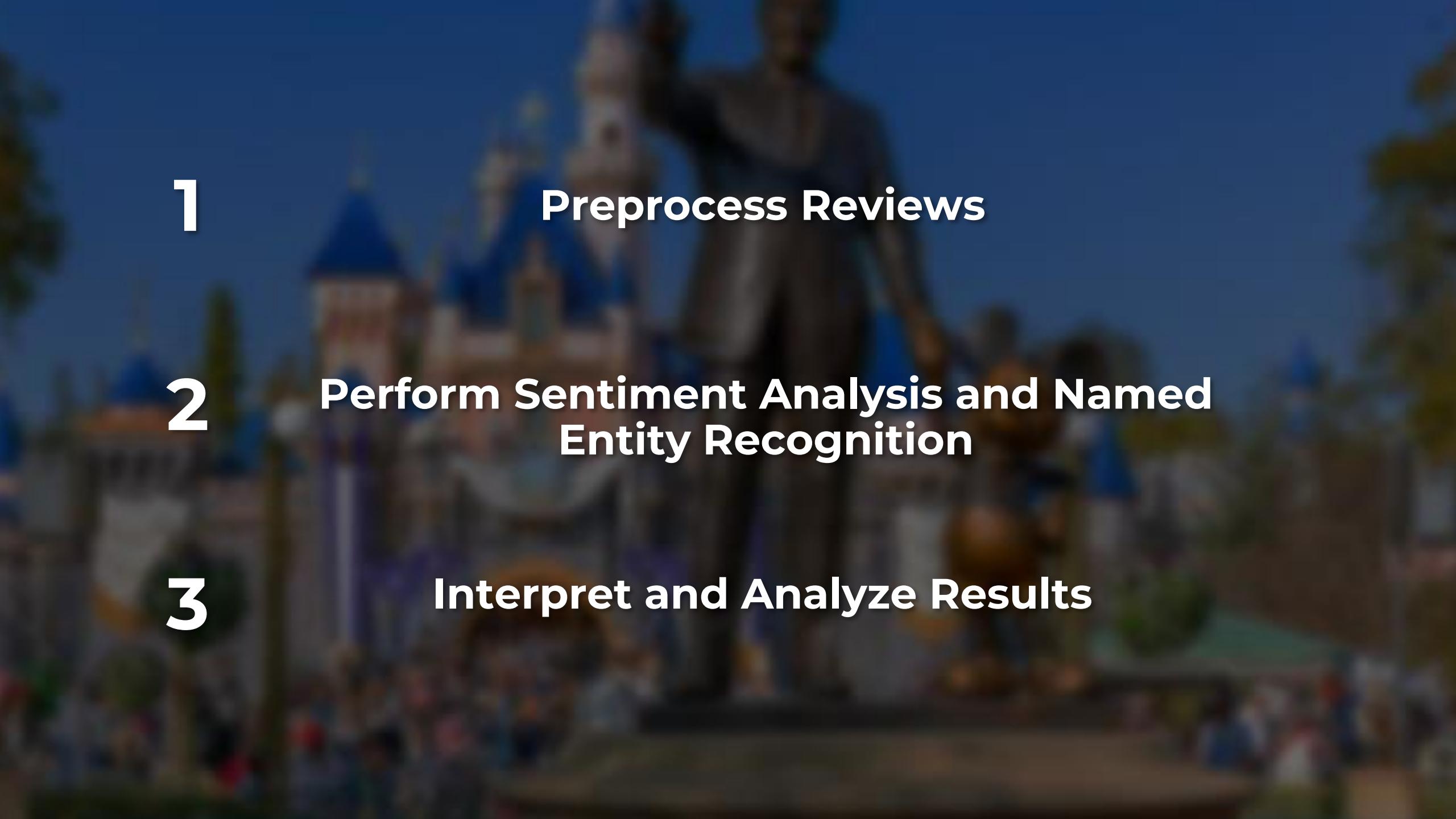


42,632 reviews

Disneyland  
Park

DISNEYLAND  
PARIS

HONG KONG  
DISNEYLAND

The background of the slide is a blurred photograph of a city skyline at night. The buildings are dark silhouettes against a bright sky, with some windows and architectural details visible through the blur.

**1**

**Preprocess Reviews**

**2**

**Perform Sentiment Analysis and Named Entity Recognition**

**3**

**Interpret and Analyze Results**

# 1. Preprocess Reviews

Tools Used



# Data Preprocessing

## Explore Data

- Understand Columns
- Determine Distribution
- Generate Reports

## Feature Engineering

- Split Columns
- Aggregate New Columns
- Remove Columns
- Dealing with Null Values

## Remove Noise

- URLs
- Extra spaces
- Duplicates
- Separate Contractions
- Emojis

## Normalization

- Lemmatization
- Stemming
- Lowercase Conversion
- Tokenization

# Columns

- **Review ID**
- **Rating 1-5**
- **Quarter**
- **Reviewer Location**
- **Review**
- **Park Location**
- **Year**
- **Month**
- **Quarter**
- **Sentiment**

# Review Column Sample

its been while since last time we visit hk disneyland yet this time we only stay in tomorrowland aka marvel land now they have iron man experience d newly open ant man d wasp ironman great feature so exciting especially whole scenery of hk hk central area to kowloon antman changed by previous buzz lightyear more or less same but am expecting to have something most however my boys like it space mountain turns into star wars this 1 is great for cast members staffs felt bit minus point from before just dun feel like its disney brand seems more local like ocean park or even worst they got no smiling face but just want to you to enter attraction leave hello this is supposed to be happiest place on earth brand but just really do not feel it bakery in main street now have more attractive delicacies disney theme sweets these are good points last they also have starbucks now inside the theme park

## 2. Perform Sentiment Analysis and Named Entity Recognition

Tools Used



# Natural Language Processing

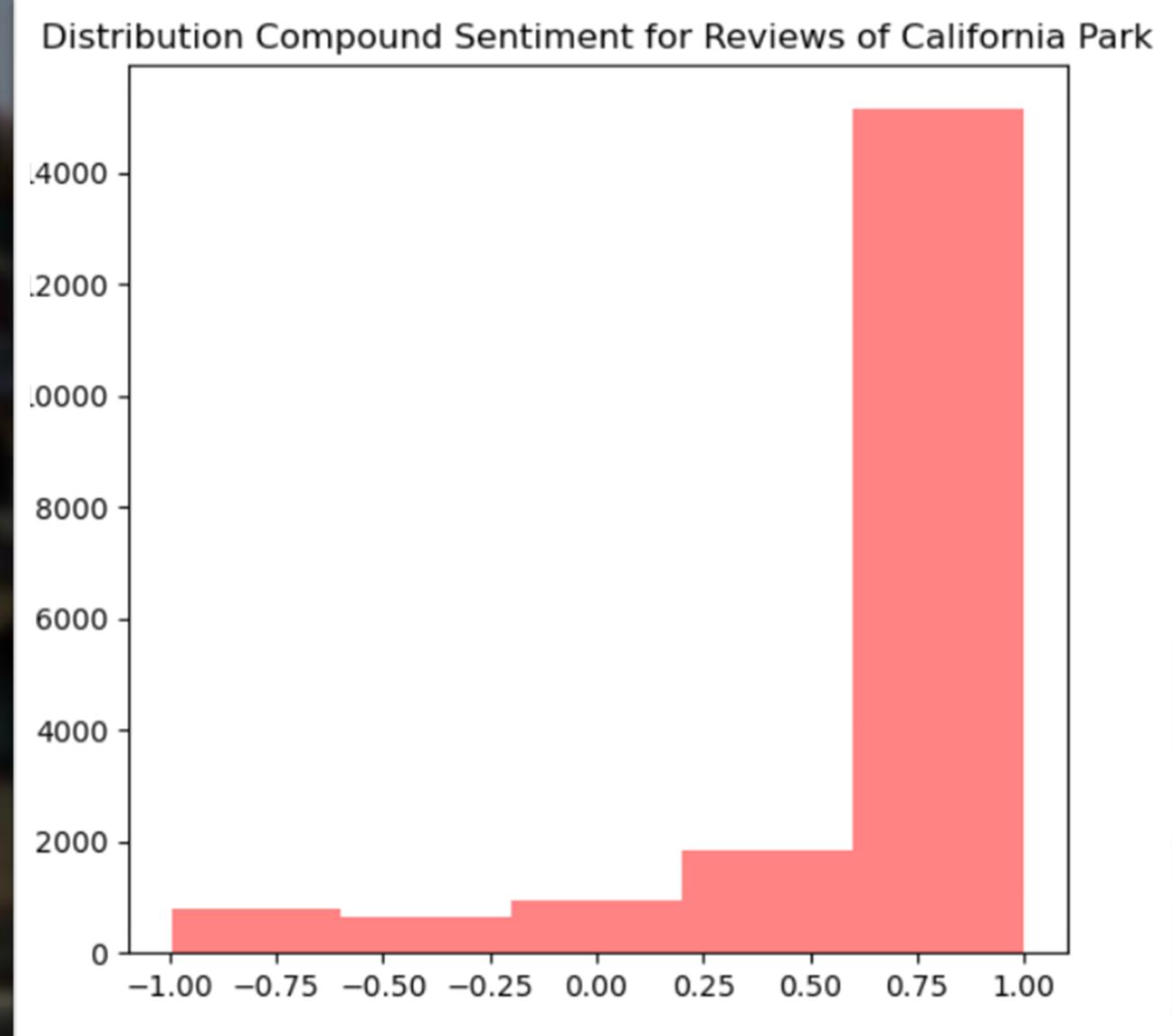
## Sentiment Analysis

- Vader from NTLK package
- Pre-trained Model for Social Media Text
- Tokenization Available
- Categorize each review

## Named Entity Recognition

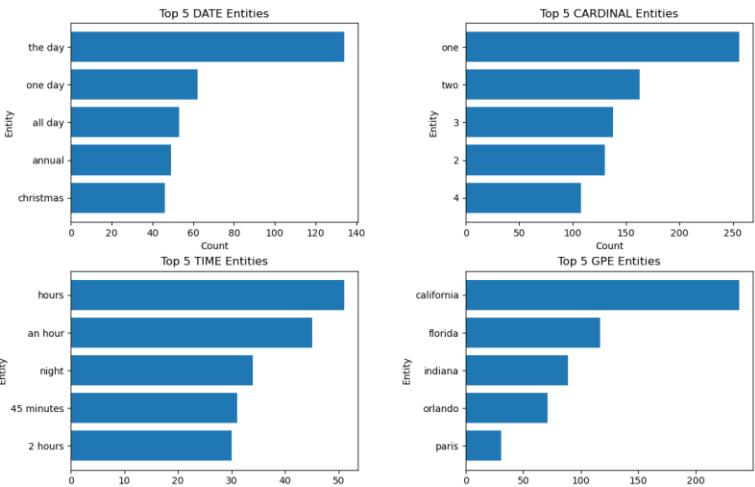
- NER from SPACY Package
- Categorize Words of Review based on Entities
- Top 5 words categorized by park
- Categories include Person, Organization, Place/Location, Date, Ordinal, Numerical

- -1 = Most Negative
- 0 = Neutral
- 1 = Most Positive

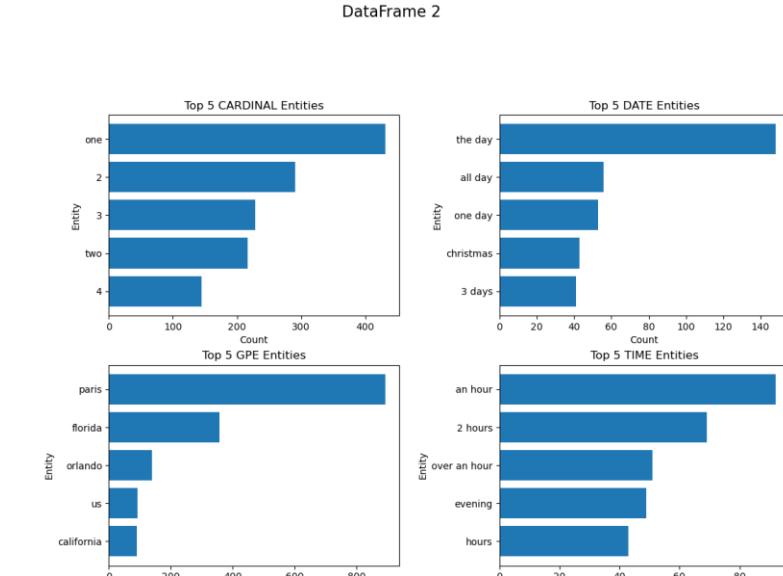


# Keywords Generated by NER and are Plotted

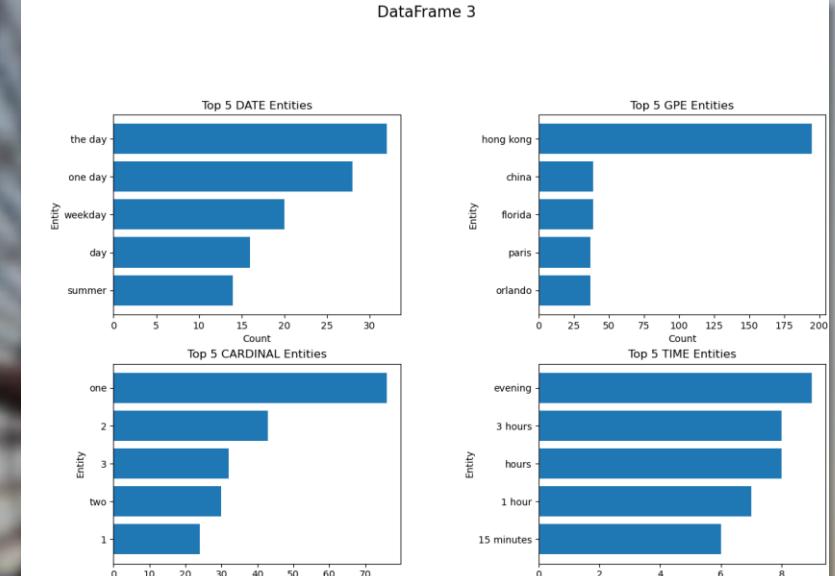
DataFrame 1



DataFrame 2



DataFrame 3



Disneyland  
Park

Disneyland  
PARIS

HONG KONG  
Disneyland

# **3. Interpret and Analyze Results**

# California Disneyland Keywords

- **Indiana Jones Ride**
  - Based on some news articles, it seemed like the ride based on 2019 was due for a refurbishment. It seems the ride had some mechanical issues which could have led to visitors leaving negative reviews about the ride.
- **Peter Pan Ride**
  - High number of reviews mentioning the character name in reference to the ride, mainly due to its wait time.

# Disneyland Paris Keywords

- Pirates of the Caribbean ride
- Peter Pan ride in Paris
- It is showing a couple of reviews mentioning Crockett which may refer to the Disney Davy Crockett Ranch.

# Disneyland Hong Kong Keywords

- It seems like for Hong Kong, not much inferences or conclusions can be made which is fine, which is why other NLP techniques available and not all is lost.
- One most notable thing is with the Person Entity where Cinderella is mentioned which can refer to the Cinderella Carousel.

# Conclusion

- NLP Technique useful unsupervised method of generating insights for reviews
- Other NLP methods may also be useful:
  - Topic Modeling
  - Summarization