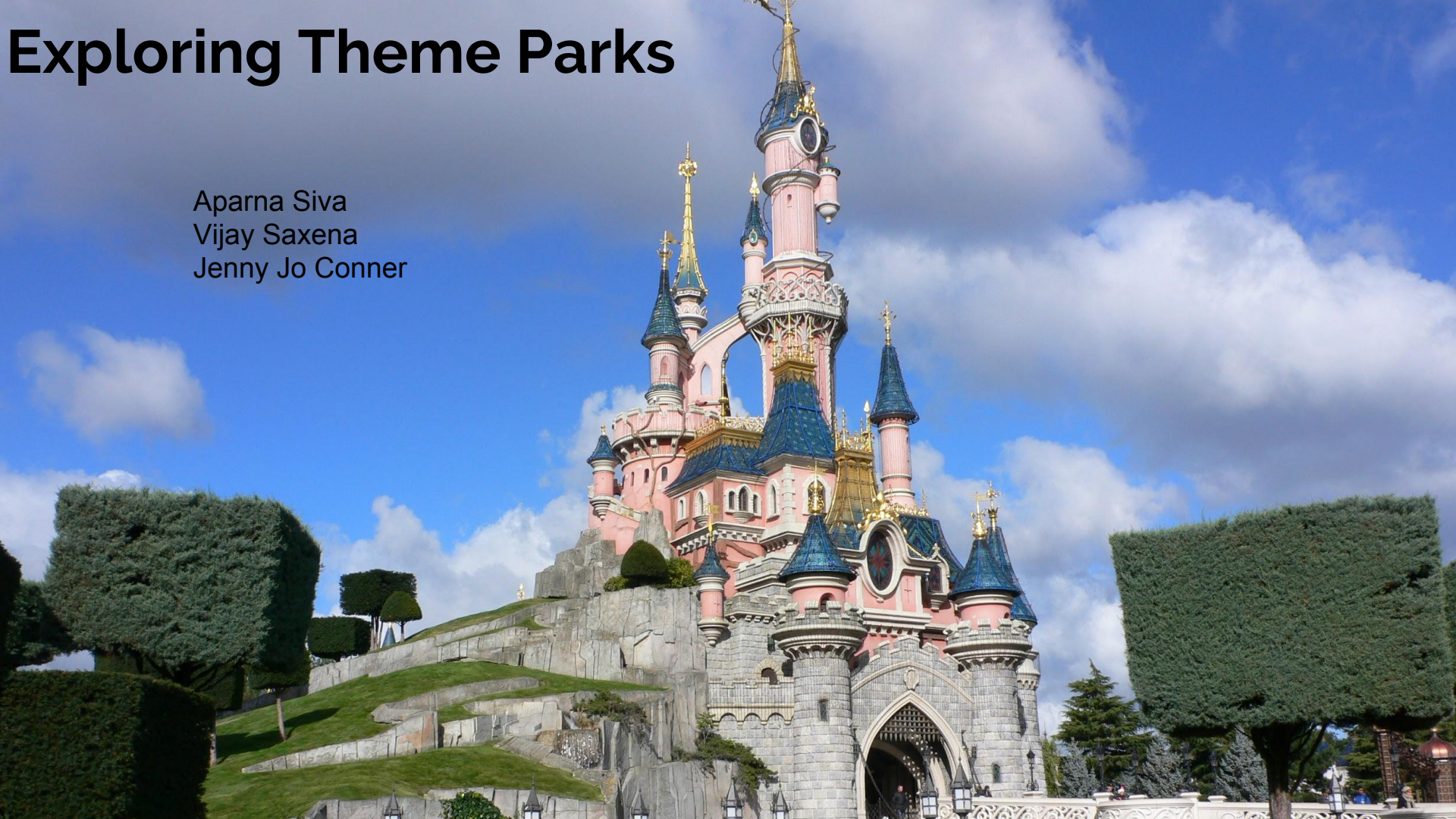
Exploring Theme Parks



Aparna Siva
Vijay Saxena
Jenny Jo Conner

Step 1

- Using Pandas, Beautiful soup, Matplotlib, lxml (HTML parser), CSS selector, re
- Select the website to be processed/scraped
  - Verify the legality

Step 2

- UTF-8 encode data
- Set parameters for Beautiful Soup to output page text
- Capture data using CSS selector
- Create DataFrame using Pandas, Data cleanup
- Translate data-type object to integers

Step 3

- Matplot
  - Visual validation
- Output to CSV
  - UTF-8 encode

Step 4

- Visualization
  - Dashboard (html)
  - Data analysis (D3)
  - Map visualization (Leaflet)

```
# Importing regex( regular expressions search , module )
import re
new_list = []
# pattern searching from the raw text after scrapping = "\xa0"-----> non ascii characters , that appeared in text
for summary in summaries:
    text = re.sub(r'\xa0','', summary)
    new_list.append(text)
```

```
summaries
```

```
["We didn't\xa0know if we were laughing because of its zippy track or the prospect of flying out of the coaster car (nervous laughter, of course), but this was not our favorite of the bunch.\xa0",
 "The good: It is adjacent to Busch Gardens' Serengeti Overlook Restaurant, which offers an all-you-can-eat buffet, fruity cocktails and views of African wildlife.\xa0The bad: You absolutely do not want to ride this spinning, backwards coaster on a full stomach. It will turn into a curse, indeed.\xa0",
```