

# Task

To capture details from the following url:

URL:

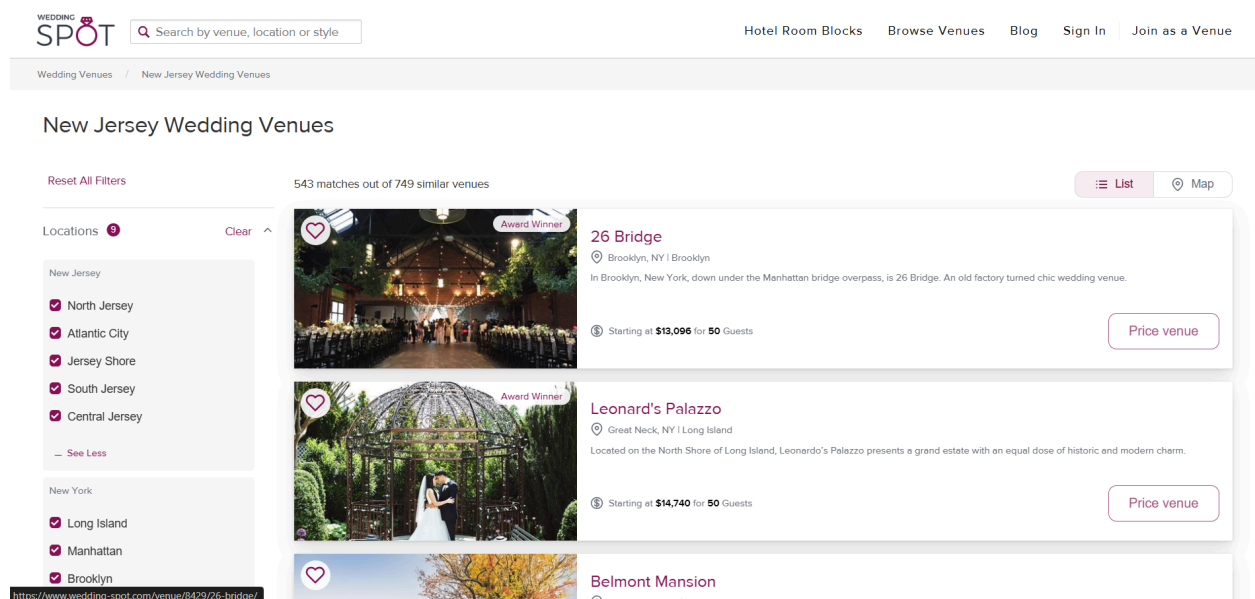
<https://www.wedding-spot.com/wedding-venues/?pr=new%20jersey&r=new%20jersey%3anorth%20jersey&r=new%20jersey%3aatlantic%20city&r=new%20jersey%3ajersey%20shore&r=new%20jersey%3asouth%20jersey&r=new%20jersey%3acentral%20jersey&r=new%20york%3along%20island&r=new%20york%3amanhattan&r=new%20york%3abrooklyn&r=pennsylvania%3aphiladelphia&sr=1>

**Use either Scrappy (Python) or Crawlee (Node.js) for this assignment. We've provided enough time so you can get familiar with the basics of these frameworks if needed.**

## Steps to follow:

### 1. Start at the listing page

Visit the provided link and you'll land on a page showing multiple venue listings.



### 2. Handle Pagination

There are multiple listing pages, so you need to handle the pagination for scraping more than 36 detail pages.

### 3. Visit Each Venue's Detail Page

The scraper needs to visit each detail page of the venue.

For example the first detail page from the image above would be:

<https://www.wedding-spot.com/venue/1400/leonards-palazzo/>

WEDDING SPOT Search by venue, location or style

Hotel Room Blocks Browse Venues Blog Sign In Join as a Venue

Wedding Venues / New York Wedding Venues / Long Island Wedding Venues / Leonard's Palazzo




Photo by: BRI JOHNSON LLC Report

**B** Leonard's Palazzo

**C** CALL VENUE FAVORITE

**D** Venue Highlights

- 4x award winner
- Indoor and outdoor spaces
- Upgraded rentals included

STARTING AT \$14,740 FOR 100 GUESTS

Price this venue to get your customized estimate, no strings attached

PRICE THIS VENUE

Venue details

WEDDING SPOT Search by venue, location or style

Hotel Room Blocks Browse Venues Blog Sign In Join as a Venue

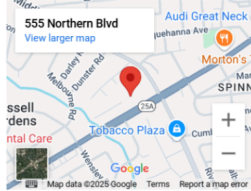
**Venue details**

**Style:** Ballrooms, Modern

**Guest capacity:** Accommodates up to 1000 guests **E**

**Services:** Ceremony / Reception Indoor / Outdoor

**Location:** 555 Northern Boulevard Great Neck, NY 11021 **F**



**Awards:**

- Best of 2016
- Best of 2017
- BEST OF 2020

**Description:**

Located on the North Shore of Long Island, Leonardo's Palazzo presents a grand estate with an equal dose of historic and modern charm. An ideal location for a fairytale wedding and reception, this lavish venue is sure to impress all of your friends and family on your wedding day. Adorned with dazzling details everywhere you look, the Palazzo exudes a real sense of magic and romance, inside and out. Designed with a posh Banquet Hall, this opulent space includes 17th century themed murals, crystal chandeliers, and extraordinary architectural elements. For an open air location to say 'I do', discover the lovely atmosphere of the Garden and Patio space. Embellished with cherub fountains covered in vines, the Garden and Patio presents the perfect space for unforgettable photos with your love. If you imagined a wedding day fit for a king and a queen,

Price this venue to get your customized estimate, no strings attached

PRICE THIS VENUE

**Interested in this venue?**

Don't miss out! Message this venue now to check their availability and receive more information.

No commitment required.

CONTACT VENUE

#### 4. Data to Scrape

**Extract the following fields:**

- A - Url (url of the detail page)
- B - Venue Name
- C - Phone
- D - Venue Highlights
- E - Guest Capacity
- F - Address

**Here is the sample data for guidance:**

[https://docs.google.com/spreadsheets/d/1KS2\\_pli7b2C0LG3VYGKYIJdkwzPCiIr-hFhPgPs-20/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1KS2_pli7b2C0LG3VYGKYIJdkwzPCiIr-hFhPgPs-20/edit?usp=sharing)

\* While selecting the elements, please do not use Beautiful soup. This assignment requires you to use CSS or XPATH selectors.

#### **Submission Requirements:**

1. A complete set of code files for the scraping task (either a zip file or a github link)
2. Match the structure of the sample data and ensure accuracy..
3. A Readme file with instructions on setting up the project, running it and viewing the data.
4. A sample of the scraped data(JSON, CSV or Excel file). Try to scrape upto 4 or 5 listing pages.

**Deadline:** Submit your project within 14 days of receiving this mail.

Ensure that your submission is thorough and meets all the specific requirements.

If you encounter any issues with the task or have any questions, do not hesitate to reach out via email. We are here to assist and can provide additional reading materials and support as needed.