Project 3 Data Engineering

Syndy Escajeda - Min Pak - Flora Aka - Maria Paula Ardila

Project Summary



- If someone could travel to any of the concert locations in the US or Canada for the "Las Mujeres No Lloran World Tour," which one would be the most budget-friendly concert location, and what is the budget range per concert?

			SQL C	UERY COME	TICKETMASTER API	HOTELS.COM SCRAPPING		TICKETMASTER API		HOTELS.COM SCRAPPING	
city	concert_date	checkin_date	min_budget	max_budget	ticket_min_price	min_hotel	hotel_name_for_min_price	ticket_max_price		hotel_name_for_max_price	
Montreal	2024-12-10	2024-12-09	146.5	4786.5	119.5	27	Hygie Pod Hotel - Hostel	4344.5	442	Fairmont The Queen Elizabeth Gold Experience	
Charlotte	2024-11-23	2024-11-22	201.5	1078.5	129.5	72	Royal Inn Charlotte Airport Hotel	429.5	649	The Ritz-Carlton, Charlotte	
Washington	2024-11-25	2024-11-24	204.5	999.5	129.5	75	Motto by Hilton Washington DC City Center	429.5	570	Waldorf Astoria Washington DC	
Phoenix	2024-11-07	2024-11-06	210.5	1254.5	144.5	66	Econo Lodge Phoenix North I-17	484.5	770	The Global Ambassador	
Miami	2024-11-20	2024-11-19	219.5	1161.5	145.5	74	Up Midtown	485.5	676	Four Seasons Hotel Miami	
Boston	2024-12-08	2024-12-07	221.5	1203.5	119.5	102	Temple Place Suites	399.5	804	Raffles Boston	
Dallas	2024-11-17	2024-11-16	222.5	1175.5	149.5	73	Serene Dallas Condo w/ Pool - Walkable Location!	491.5	684	Hôtel Swexan	
Miami	2024-11-21	2024-11-20	222.5	1197.5	145.5	77	Up Midtown	485.5	712	Four Seasons Hotel Miami	
Chicago	2024-12-14	2024-12-13	230.5	931	144.5	86	Ohio House Motel	485	446	Trump International Hotel & Tower Chicago	
Toronto	2024-11-30	2024-11-29	236.5	1113.5	139.5	97	The Rex Hotel	459.5	654	The Hazelton Hotel Toronto	
Brooklyn	2024-12-05	2024-12-04	253.5	1392.5	149.5	104	NY Moore Hostel	489.5	903	The Wall Street Hotel	
San Antonio	2024-11-16	2024-11-15	254.5	842.5	189.5	65	WoodSpring Suites San Antonio UTSA - Medical Center	499.5	343	La Cantera Resort & Spa	
Palm Desert	2024-11-03	2024-11-02	277.5	4404.75	201.5	76	Rodeway Inn & Suites Thousand Palms - Rancho Mirage	564.75	3840	Villa Verbena by Avantstay Expansive Estate w/ B-ball Court, Pool/spa & Entertainer's Yard	
Palm Desert	2024-11-02	2024-11-01	286.5	4425.75	201.5	85	Rodeway Inn & Suites Thousand Palms - Rancho Mirage	564.75	3861	Villa Verbena by Avantstay Expansive Estate w/ B-ball Court, Pool/spa & Entertainer's Yard	
Brooklyn	2024-12-06	2024-12-05	310.5	2986.5	149.5	161	Super 8 by Wyndham Brooklyn Fulton Street/Broadway Junction	489.5	2497	Holiday Inn Express New York City- Wall Street, an IHG Hotel	
Inglewood	2024-11-09	2024-11-08	334.5	866.5	249.5	85	Royal Century Hotel	589.5	277	Four Points by Sheraton Los Angeles Westside	

Ticketmaster API



Discovery API - V 2.0

- The Ticketmaster Discovery API allows to search for events, attractions, or venues
- To run a successful API call, we had to pass the API Key in the "apikey" query parameter
- To get a list of all events for Shakira's Tour in the USA and Canada we had to use the following parameters:

slidesmania.com



The JSON response has dictionary of events

```
# Get to the dictionary that cotains the concerts information concerts = events['_embedded']['events']
pprint(concerts)
```

 On the dictionary of each event we found the concert id, dates, price range (min and max) and the venue info

```
{'name': 'Shakira - Las Mujeres Ya No Lloran - World Tour',
 'type': 'event',
 'id': 'G5eVZbFMb0bre',
 'dates': {'start': {'localDate': '2024-11-23',
   'localTime': '19:30:00',
   'dateTime': '2024-11-24T00:30:00Z',
   'dateTBD': False.
   'dateTBA': False,
   'timeTBA': False,
   'noSpecificTime': False},
  'timezone': 'America/New_York',
  'status': {'code': 'cancelled'},
  'spanMultipleDays': False},
 'priceRanges': [{'type': 'standard',
   'currency': 'USD',
   'min': 129.5,
   'max': 429.5}].
```

```
'_embedded': {'venues': [{'name': 'Spectrum Center ',
   'type': 'venue',
  'id': 'KovZpZA6AEIA',
   'test': False.
  'url': 'https://www.ticketmaster.com/spectrum-center-tickets-charlotte/venue/369357',
  'locale': 'en-us',
  'images': [{'ratio': '16_9',
    'url': 'https://s1.ticketm.net/dbimages/18670v.jpg',
    'width': 205,
    'heiaht': 115.
     'fallback': False}],
   'postalCode': '28202',
   'timezone': 'America/New York'.
  'city': {'name': 'Charlotte'},
   'state': {'name': 'North Carolina', 'stateCode': 'NC'},
  'country': {'name': 'United States Of America'. 'countryCode': 'US'}.
   'address': {'line1': '333 East Trade Street'},
   'location': {'longitude': '-80.839921', 'latitude': '35.225222'},
```

slidesmania.com

Ticketmaster API



 Using a for loop, we created a list of dictionaries for concerts and venues:

```
dict_of_concerts = {}
for concert in concerts:
   concert dict = {}
    if (concert['dates']['status']['code'] != 'offsale'):
       concert id = concert['id']
       concert name = concert['name']
       concert dict['Concert Name'] = concert name
       concert_date = concert['dates']['start']['localDate']
       concert_dict['Concert_Date'] = concert_date
       ticket sale status = concert['dates']['status']['code']
       concert_dict['Ticket_Sale_Status'] = ticket_sale_status
           ticket_min_price = concert['priceRanges'][0]['min']
           concert_dict['Ticket_Min_Price'] = ticket_min_price
           ticket max price = concert['priceRanges'][0]['max']
           concert dict['Ticket Max Price'] = ticket max price
        except (KevError):
            concert dict['Ticket Min Price'] = 'Concert is in presale. General pricing is not available'
            concert_dict['Ticket_Max_Price'] = 'Concert is in presale. General pricing is not available'
            ticket limit = concert['ticketLimit']['info']
            concert dict['Ticket Limitation'] = ticket limit
            concert dict['Ticket Limitation'] = 'No Limit'
        venue id = concert[' embedded']['venues'][0]['id']
       concert dict['Venue_ID'] = venue_id
       dict of concerts [concert id] = concert dict
```

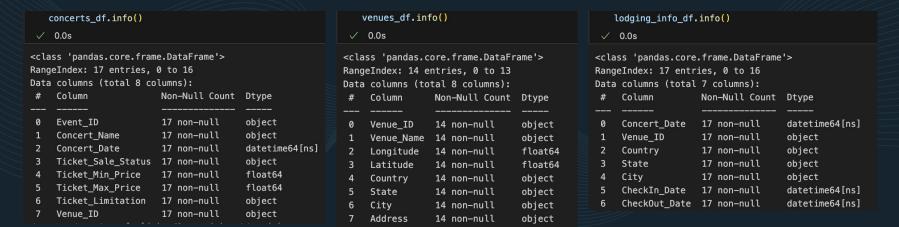
```
{'G5eVZbFMb0bre': {'Concert_Name': 'Shakira - Las Mujeres Ya No Lloran - World Tour',
  'Concert Date': '2024-11-23'.
  'Ticket Sale Status': 'cancelled',
  'Ticket Min Price': 129.5,
  'Ticket Max Price': 429.5,
  'Ticket_Limitation': 'No Limit',
  'Venue ID': 'KovZpZA6AEIA'}.
 'G5dIZbFwIrg08': {'Concert_Name': 'Shakira - Las Mujeres Ya No Lloran - World Tour',
  'Concert Date': '2024-11-16',
  'Ticket Sale Status': 'cancelled',
  'Ticket Min Price': 189.5,
  'Ticket_Max_Price': 499.5,
  'Ticket_Limitation': 'There is an overall 6 ticket limit for this event.',
  'Venue ID': 'KovZpZAJJEdA'}.
 'vvG1VZbFU3Rv5A': {'Concert Name': 'Shakira - Las Mujeres Ya No Lloran - World Tour',
  'Concert Date': '2024-11-20',
  'Ticket Sale Status': 'cancelled',
  'Ticket_Min_Price': 145.5,
  'Ticket_Max_Price': 485.5,
  'Ticket Limitation': 'There is a six (6) ticket limit for this event.'.
  'Venue ID': 'KovZpZAJtFaA'},
 'vv170ZbFGkMDd40B': {'Concert Name': 'Shakira - Las Mujeres Ya No Lloran - World Tour',
  'Concert Date': '2024-11-09',
  'Ticket_Sale_Status': 'cancelled',
  'Ticket_Min_Price': 249.5,
  'Ticket_Sale_Status': 'rescheduled',
  'Ticket Min Price': 139.5,
  'Ticket Max Price': 459.5,
  'Ticket_Limitation': 'There is an overall 6 ticket limit for this event.',
  'Venue ID': 'KovZpZAFFE1A'}}
```

slidesmania.cor

Ticketmaster API



- At the end of the process we created and export the CSVs of the following Data Frames:
 - concerts_df & venues_df
 - O lodging_info_df Filtered df after joining concerts and Venues (for the project we defined the CheckIn_Date as the day before the concert & the CheckOut_Date as the day after the concert)



Ticketmaster API

On Friday Oct 18th Shakira decided to change the dates of the tour in the US and Canada









INDIO, CALIFORNIA - APRIL 12: (FOR EDITORIAL USE ONLY) Shakira performs with
Bizarrap at the Sahara Tent during the 2024 Coachella Valley Music and Arts Festival at

Colombian singer Shakira announced on her social media that due to the high demand for her concerts, she will postpone the Las Mujeres Ya No Lloran tour to May 2025.

Trending Stories



Alleged squatter could face 15 years in prison for occupying vacant Queens home



CRIME AND COURTS
Rudy Giuliani ordered to turn
over his N.Y. apartment and
valuables to the former
Georgia election workers he
defamed



IRS announces new federal income tax brackets for 2025



What is Prop 1? New York set to vote on state constitutional amendment on Election Day

	Event_ID	Concert_Name
	vv17OZbFGkU9i-pl	Shakira - Las Mujeres Ya No Lloran - World Tour
	1Av0ZbFGkTIDVAd	Shakira - Las Mujeres Ya No Lloran - World Tour
	vvG1lZbFTpcNCA	Shakira - Las Mujeres Ya No Lloran - World Tour
	vvG1VZbFU3Rv5A	Shakira - Las Mujeres Ya No Lloran - World Tour
	G5dlZbFwlrgO8	Shakira - Las Mujeres Ya No Lloran - World Tour
	vvG1lZbFSkKfhJ	Shakira - Las Mujeres Ya No Lloran - World Tour
	G5eVZbFMbObre	Shakira - Las Mujeres Ya No Lloran - World Tour
	vvG1VZbFX9ZdVo	Shakira - Las Mujeres Ya No Lloran - World Tour
	vv170ZbFGkMDd4OB	Shakira - Las Mujeres Ya No Lloran - World Tour
	vv178ZbFGkR05dgA	Shakira - Las Mujeres Ya No Lloran - World Tour
	vvG1YZbFSrWgPY	Shakira - Las Mujeres Ya No Lloran - World Tour
	1AvfZbFGkRnDJls	Shakira - Las Mujeres Ya No Lloran - World Tour
	1AdZZbFGkww3sSf	Shakira - Las Mujeres Ya No Lloran - World Tour
	vv177ZbFGkwEuVoY	Shakira - Las Mujeres Ya No Lloran - World Tour
14	1AYZke8GkdyGvh8	Shakira - Las Mujeres Ya No Lloran - World Tour
	1Ad7ZbFGkmFYsMD	Shakira - Las Mujeres Ya No Lloran - World Tour

1AvZZbFGkRDEKcW

Shakira - Las Muiere

	Ticket_Limitation	Ticket_Max_Price	Ticket_Min_Price	Ticket_Sale_Status	Concert_Date	_Name
ŀ	There is an overall 6 ticket limit for this ev	259.50	59.5	rescheduled	2025-05-22	Ya No ld Tour
Ko	There is an overall 6 ticket limit for this ev	484.50	144.5	rescheduled	2025-06-22	Ya No ld Tour
	There is an overall 6 ticket limit for this ev	564.75	201.5	cancelled	2024-11-03	Ya No ld Tour
K	There is a six (6) ticket limit for this event.	485.50	145.5	cancelled	2024-11-20	Ya No ld Tour
Ko	There is an overall 6 ticket limit for this ev	499.50	189.5	cancelled	2024-11-16	Ya No ld Tour
	There is an overall 6 ticket limit for this ev	564.75	201.5	cancelled	2024-11-02	Ya No ld Tour
Ko	No Limit	429.50	129.5	cancelled	2024-11-23	Ya No ld Tour
K	There is a six (6) ticket limit for this event.	485.50	145.5	cancelled	2024-11-21	Ya No ld Tour
Ko	There is an overall 6 ticket limit for this ev	589.50	249.5	cancelled	2024-11-09	Ya No ld Tour
	There is an overall 6 ticket limit for this ev	485.00	144.5	cancelled	2024-12-14	Ya No ld Tour
Ko	There is an overall 6 ticket limit for this ev	491.50	149.5	cancelled	2024-11-17	Ya No ld Tour
	There is an overall 6 ticket limit for this ev	429.50	129.5	cancelled	2024-11-25	Ya No Id Tour
	There is a 6 ticket limit on this event	489.50	149.5	cancelled	2024-12-05	Ya No ld Tour
	There is an overall 6 ticket limit for this ev	399.50	119.5	cancelled	2024-12-08	Ya No Id Tour
	There is a 6 ticket limit on this event	489.50	149.5	cancelled	2024-12-06	Ya No ld Tour
	ll y a une limite de 6 billets pour cet événem	4344.50	119.5	rescheduled	2025-05-20	Ya No ld Tour
Ko	There is an overall 6 ticket limit for this ev	459.50	139.5	rescheduled	2025-05-26	Ya No ld Tour

Venue_ID

ovZ917A25V

ovZpZAE617A KovZ917A-Tn

ovZpZAJtFaA

ovZpZAJJEdA

KovZ917A-Tn

ovZpZA6AEIA

ovZpZAJtFaA

ovZpZAEkn6A

KovZpa2M7e

ovZpZAJ67eA

KovZpaKuJe

KovZ917AtP3
KovZpa2gne

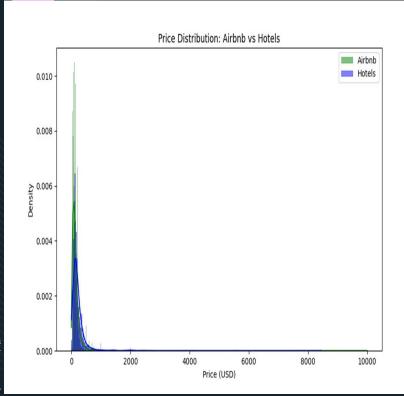
KovZ917AtP3

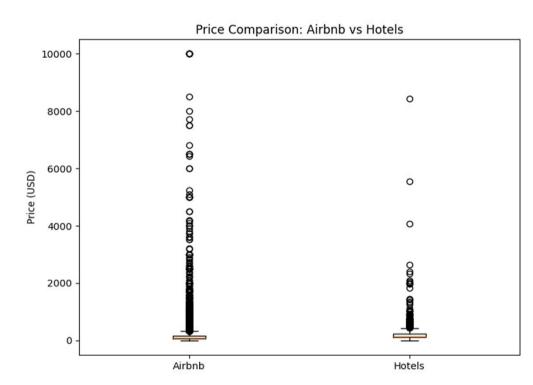
KovZpauRJe

vZpZAFFE1A

desmania.com

Airbnb vs hotel pricing comparison





Scrapping Process

- 1) Input data gathered from Ticketmaster.
- 2) Create list of urls that will be scraped.
- 3) Loop thru urls and save that data into a dictionary of list.
- 4) Clean data.
- 5) Convert data into a dataframe to be uploaded into Postgres for analysis.



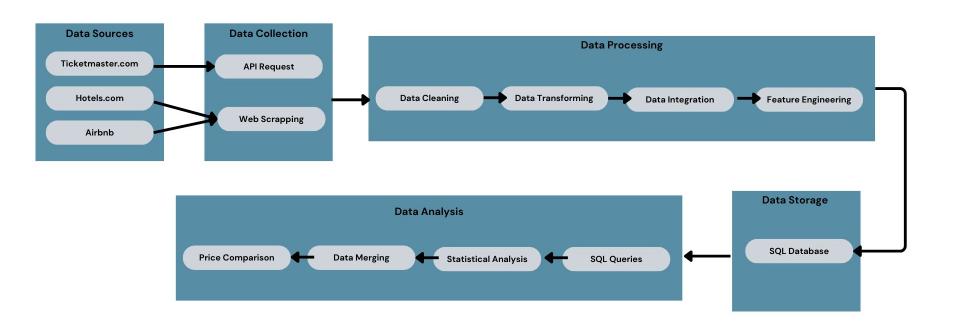
Code Demo

urllib.parse — Parse URLs into components

Source code: Lib/urllib/parse.py

This module defines a standard interface to break Uniform Resource Locator (URL) strings up in components (addressing scheme, network location, path etc.), to combine the components back into a URL string, and to convert a "relative URL" to an absolute URL given a "base URL."

Project Pipeline



ERD

compiled hotel prices

Ow VARCHAR(255) Hotel Price FLOAT City VARCHAR(255) State VARCHAR(255) VARCHAR(255) Country CheckIn Date DATE CheckOut Date DATE

lodging info ORIGINAL DATES

Concert Date DATE Venue ID VARCHAR(255) VARCHAR(255) Country State VARCHAR(255) City VARCHAR(255) CheckIn Date DATE CheckOut Date DATE

venues info ORIGINAL DATES

Ov VARCHAR(255) Venue ID VARCHAR(255) Venue Name Longitude FLOAT Latitiude FLOAT VARCHAR(255) Country VARCHAR(255) State City VARCHAR(255) Address

VARCHAR(255)

concerts info ORIGINAL DATES

Event ID Ow VARCHAR(255) Concert Name VARCHAR(255) Concert Date DATE VARCHAR(255) Ticket Sale Status Ticket Min Price FLOAT FLOAT Ticket Max Price Ticket Limitation VARCHAR(255) Venue_ID VARCHAR(255)

slidesmania.com

SQL Queries

```
CREATE TABLE compiled_hotel_prices (
    Hotel VARCHAR NOT NULL,
   Price INT NOT NULL,
   CheckIn_Date DATE NOT NULL,
    CheckOut_Date DATE NOT NULL,
   City VARCHAR NOT NULL,
    State VARCHAR NOT NULL,
   Country VARCHAR NOT NULL
CREATE TABLE concerts info ORIGINAL DATES (
   Event_ID VARCHAR NOT NULL,
   Concert Name VARCHAR NOT NULL,
   Concert_Date DATE NOT NULL,
   Ticket_Sales_Status VARCHAR NOT NULL,
   Ticket_Min_Price FLOAT NOT NULL,
   Ticket Max Price FLOAT NOT NULL,
   Ticket_Limitation VARCHAR NOT NULL,
   Venue ID VARCHAR NOT NULL
CREATE TABLE venues info ORIGINAL DATES (
    Venue ID VARCHAR NOT NULL,
    Venue_Name VARCHAR NOT NULL,
   Longitude FLOAT NOT NULL,
   Latitude FLOAT NOT NULL,
   Country VARCHAR NOT NULL,
   State VARCHAR NOT NULL,
   City VARCHAR NOT NULL,
   Address VARCHAR NOT NULL
CREATE TABLE lodging_info_ORIGINAL_DATES (
   Concert Date DATE NOT NULL.
   Venue_ID VARCHAR NOT NULL,
   Country VARCHAR NOT NULL,
   State VARCHAR NOT NULL.
   City VARCHAR NOT NULL.
   CheckIn_Date DATE NOT NULL,
   CheckOut_Date DATE NOT NULL
```

```
-- guery to find the min budget and max budget
SELECT sindys_results.city,
       sindys results.concert date.
       sindys results.checkin date, -- Include checkin date in the outer query
       ticket_min_price + min_hotel as min_budget,
       ticket_max_price + max_hotel as max_budget,
       sindys results.ticket min price,
       sindys results.min hotel,
       CHP_min.hotel as hotel_name_for_min_price,
       sindys results.ticket max price.
       sindys results.max hotel,
       CHP_max.hotel as hotel_name_for_max_price
    SELECT compiled_hotel_prices.city,
        lodging info ORIGINAL DATES.concert date,
        lodging_info_ORIGINAL_DATES.checkin_date,
        concerts_info_ORIGINAL_DATES.ticket_min_price,
        concerts info ORIGINAL DATES.ticket max price,
        MIN(compiled_hotel_prices.price) AS min_hotel,
        MAX(compiled hotel prices.price) AS max hotel
    FROM compiled hotel prices
    LEFT JOIN lodging info ORIGINAL DATES
    ON compiled_hotel_prices.checkin_date = lodging_info_ORIGINAL_DATES.checkin_date
    LEFT JOIN concerts info ORIGINAL DATES
    ON lodging info ORIGINAL DATES. venue id = concerts info ORIGINAL DATES. venue id
    GROUP BY lodging_info_ORIGINAL_DATES.concert_date,
             lodging_info_ORIGINAL_DATES.checkin_date,
             compiled hotel prices.city,
             concerts_info_ORIGINAL_DATES.ticket_min_price,
             concerts info ORIGINAL DATES.ticket max price
) as sindys results
```

Query Outcome Structure

Subquery

- compiled_hotel_prices
- lodging_info
- concerts_info

Group by

- Concert date
- CheckIn date
- City
- Ticket min price
- Ticket max price

LEFT JOIN compiled_hotel_prices as CHP_min ON sindys_results.min_hotel = CHP_min.price AND sindys_results.checkin_date = CHP_min.checkin_date

LEFT JOIN compiled_hotel_prices as CHP_max ON sindys_results.max_hotel = CHP_max.price AND sindys_results.checkin_date = CHP_max.checkin_date

ORDER BY min_budget

Outer join to get the hotel name for the min and the max hotel price



"Data… don't lie"

				QUERY COME	TICKETMASTER API	HOTELS.COM SCRAPPING		TICKETMASTER API	HOTELS.COM SCRAPPING	
city	concert_date	checkin_date	min_budget	max_budget	ticket_min_price	min_hotel	hotel_name_for_min_price	ticket_max_price	max_hotel	hotel_name_for_max_price
Montreal	2024-12-10	2024-12-09	146.5	4786.5	119.5	27	Hygie Pod Hotel - Hostel	4344.5	442	Fairmont The Queen Elizabeth Gold Experience
Charlotte	2024-11-23	2024-11-22	201.5	1078.5	129.5	72	Royal Inn Charlotte Airport Hotel	429.5	649	The Ritz-Carlton, Charlotte
Washington	2024-11-25	2024-11-24	204.5	999.5	129.5	75	Motto by Hilton Washington DC City Center	429.5	570	Waldorf Astoria Washington DC
Phoenix	2024-11-07	2024-11-06	210.5	1254.5	144.5	66	Econo Lodge Phoenix North I-17	484.5	770	The Global Ambassador
Miami	2024-11-20	2024-11-19	219.5	1161.5	145.5	74	Up Midtown	485.5	676	Four Seasons Hotel Miami
Boston	2024-12-08	2024-12-07	221.5	1203.5	119.5	102	Temple Place Suites	399.5	804	Raffles Boston
Dallas	2024-11-17	2024-11-16	222.5	1175.5	149.5	73	Serene Dallas Condo w/ Pool - Walkable Location!	491.5	684	Hôtel Swexan
Miami	2024-11-21	2024-11-20	222.5	1197.5	145.5	77	Up Midtown	485.5	712	Four Seasons Hotel Miami
Chicago	2024-12-14	2024-12-13	230.5	931	144.5	86	Ohio House Motel	485	446	Trump International Hotel & Tower Chicago
Toronto	2024-11-30	2024-11-29	236.5	1113.5	139.5	97	The Rex Hotel	459.5	654	The Hazelton Hotel Toronto
Brooklyn	2024-12-05	2024-12-04	253.5	1392.5	149.5		NY Moore Hostel	489.5	903	The Wall Street Hotel
San Antonio	2024-11-16	2024-11-15	254.5	842.5	189.5	65	WoodSpring Suites San Antonio UTSA - Medical Center	499.5	343	La Cantera Resort & Spa
Palm Desert	2024-11-03	2024-11-02	277.5	4404.75	201.5	76	Rodeway Inn & Suites Thousand Palms - Rancho Mirage	564.75	3840	Villa Verbena by Avantstay Expansive Estate w/ B-ball Court, Pool/spa & Entertainer's Yard
Palm Desert	2024-11-02	2024-11-01	286.5	4425.75	201.5	85	Rodeway Inn & Suites Thousand Palms - Rancho Mirage	564.75	3861	Villa Verbena by Avantstay Expansive Estate w/ B-ball Court, Pool/spa & Entertainer's Yard
Brooklyn	2024-12-06	2024-12-05	310.5	2986.5	149.5	161	Super 8 by Wyndham Brooklyn Fulton Street/Broadway Junction	489.5	2497	Holiday Inn Express New York City- Wall Street, an IHG Hotel
Inglewood	2024-11-09	2024-11-08	334.5	866.5	249.5	85	Royal Century Hotel	589.5	277	Four Points by Sheraton Los Angeles Westside

Ethical Considerations

Hotels.com - Terms of Service / Ticketmaster.com - Terms of Use

- Allows for personal and non-commercial use
- Prohibits data scraping on the website
- Educational use, allowed

Ticketmaster website blocks or pauses traditional web scraping and also prohibits the use of any software to go around the blocking or pausing security measures

Only way to get ethically or legally get Ticketmaster data is from their API

Questions?