EGYPTOPIA

Egyptopia Data Collection Report

Faculty of Computers and Data Science.

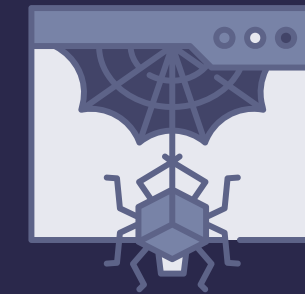Alexindria University.

How the Data Was Collected ?

Data Collection with Business-Driven Design

A relational database schema was developed to support the tourism domain, with data manually collected and curated to ensure accuracy and consistency. The dataset included information on places, events, and Egyptian food and drinks, covering aspects such as descriptions, locations, event details, and classifications.

Custom-built functions were used to convert CSV files into SQL statements, enabling smooth and repeatable data integration. This process resulted in a well-structured and meaningful dataset, forming a reliable foundation for building dashboards, recommendation systems, or APIs, while maintaining a strong focus on data quality.
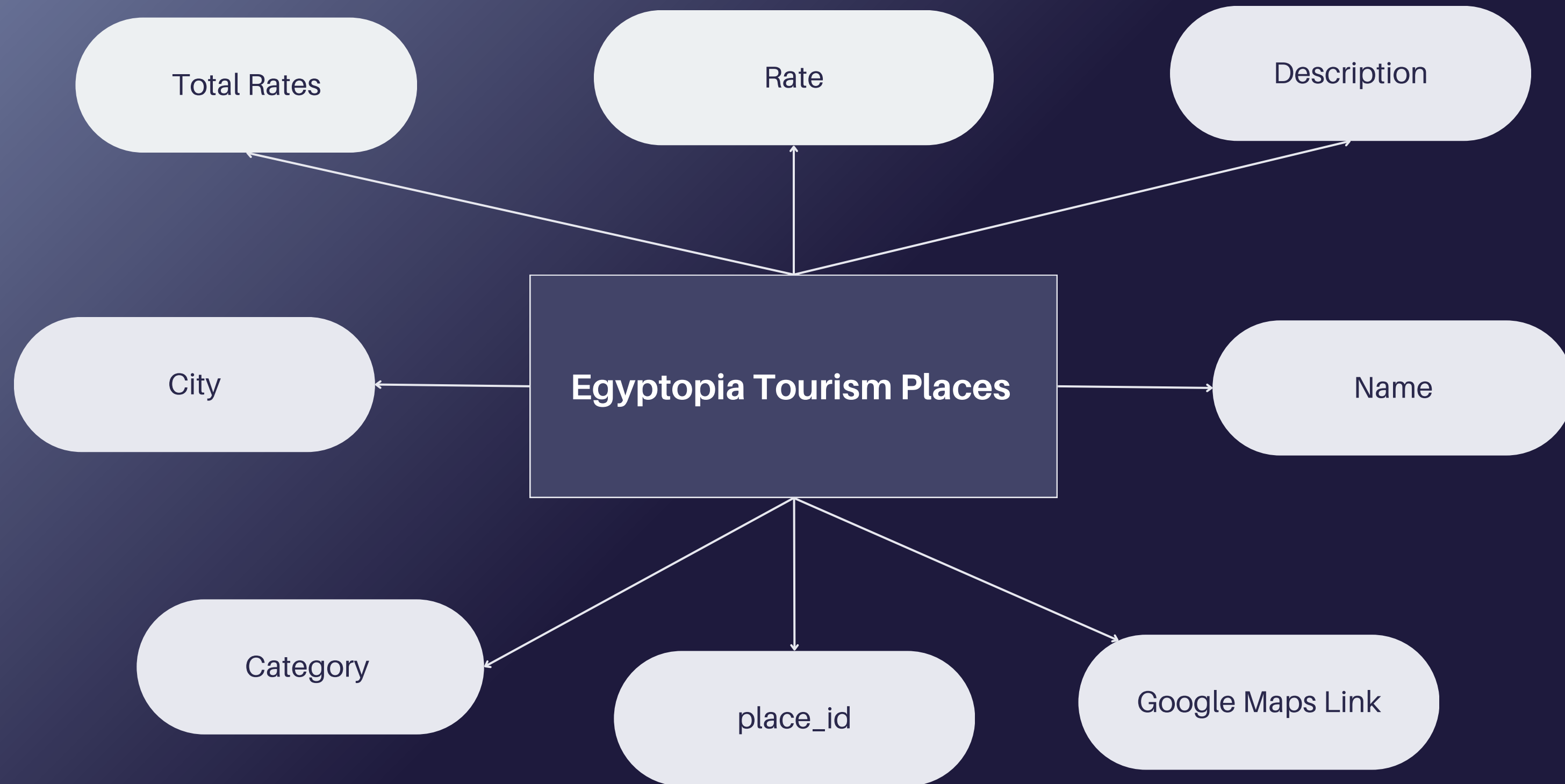
Web Scraping

Tourism data was collected from websites like GetYourGuide using Selenium with Python. The tool was used to simulate user actions (like scrolling and clicking) to load hidden content on dynamic web pages. After identifying the needed elements (such as activity names, prices, and locations), the data was extracted, cleaned, and saved in a structured format. This method allowed for up-to-date, scalable, and accurate data collection, which was later integrated into the project's database for use in recommendations and other features.
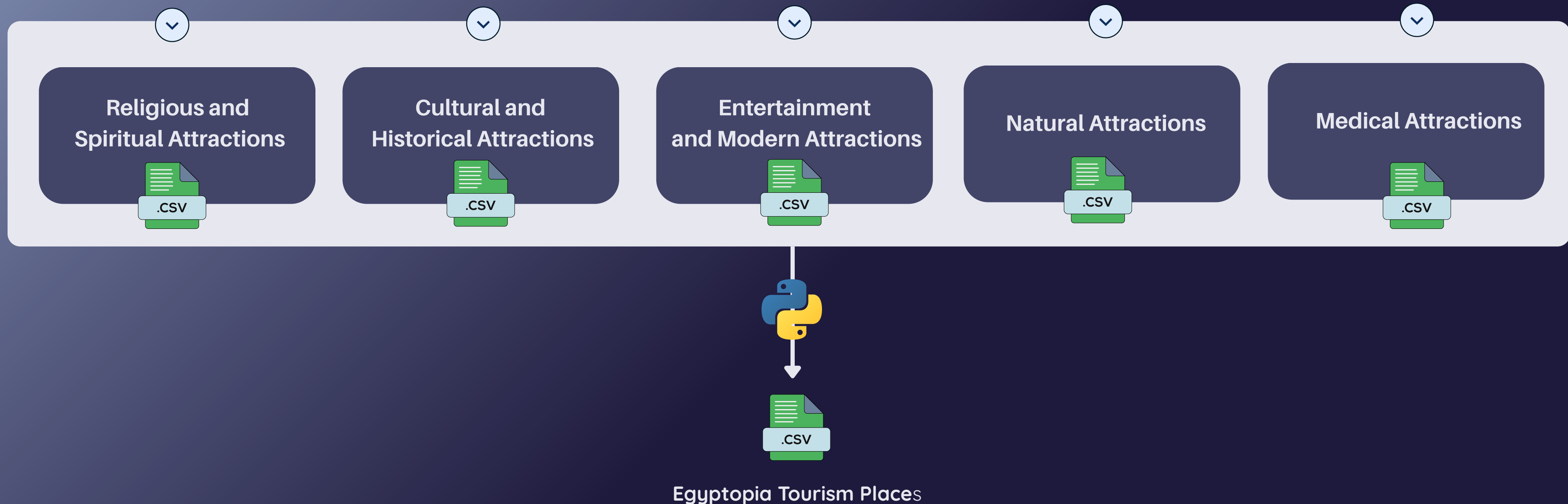
1.Egyptopia: Places Dataset
Defined Attributes of Egyptopia Tourism Places

EGYPTOPIA

Total Rates

Rate

Description

City

Egyptopia Tourism Places

Name

Category

place_id

Google Maps Link

Data Collection with Business-Driven Design

Egyptopia: Places Dataset "Datasets Merging & Features Extraction"

Religious and Spiritual Attractions	Cultural and Historical Attractions	Entertainment and Modern Attractions	Natural Attractions	Medical Attractions
.CSV	.CSV	.CSV	.CSV	.CSV

.CSV

Egyptopia Tourism Places

The individual CSV files were merged into a single dataset, with a new column named "Tourism Type" added based on the original file names, each representing a specific type of tourism. Additionally, a city enumeration process was applied to generate a corresponding "city_id" column, ensuring consistency with the schema of the system's relational database.

Data Collection with Business-Driven Design

Egyptopia Tourism Places Final Version

Extracted Features

	place_id	Name	Category	Google Maps Link	City	Rate	Total Rates	Description	Tourism Type	city_id
0	1	Dream Park	Theme Park	https://maps.app.goo.gl/HKuc4iMGP3xxkuzT7	Giza	4.2	25565	If you're looking for a place filled with exci...	Entertainment and Modern Attractions	0
1	2	Magic Land	Theme Park	https://maps.app.goo.gl/pagQ93SuUb5smLFT7	Giza	4.1	100	Magic Land is a fantastic theme park that offe...	Entertainment and Modern Attractions	0
2	3	Sindbad Amusement Park	Theme Park	https://maps.app.goo.gl/Y2oiuubB3HZFshM69	Cairo	3.1	67	Sindbad Amusement Park is a fantastic destinat...	Entertainment and Modern Attractions	1
3	4	Zed Park	Theme Park	https://maps.app.goo.gl/2Ya6VcWJUCakgbgy6	Giza	4.2	3337	Zed Park is one of Egypts most modern and vibr...	Entertainment and Modern Attractions	0
4	5	Family Park	Theme Park	https://maps.app.goo.gl/zkstQZsVx3x32cES8	Cairo	4.4	24568	Family Park is a unique amusement and recreati...	Entertainment and Modern Attractions	1

Data Collection with Business-Driven Design

Egyptopia: Places Dataset MetaData Overview

Toursim Type	Cultural & Historical Attractions	Entertainment & Modern Attractions	Natural Attractions	Religious & Spiritual Attractions	Medical Attractions
Distinct	9x Categories	5x Categories	6x Categories	5x Categories	9x Categories
Categories	HistoricalSite,Temple, Fortress, Library Museum, Palace Tomb,Theater, Cultural Center	Beach, Natural Reserve, Island, Mountain, Aquarium	Water Park, Zoo, Tower, Shopping, Garden	Mosque, Church, Monastery, Synagogue, Museum	Hot Spring, Natural Reserve, Island, Mountain, Aquarium
Total Places	116	71	65	46	21
Total Images	920	389	633	320	115

City	Cairo	Giza	Alexandria	Marsa Matrouh	Hurghada	Sharm El Sheikh	Luxor	Aswan
Counts	89	32	29	18	16	15	15	14
Percentage	28%	10%	9%	6%	5%	5%	5%	4%
City	Sinai	New Valley	Red Sea	Marsa Alam	Fayoum	North Coast	Dahab	Suez
Counts	12	11	10	8	6	6	5	5
Percentage	4%	3%	3%	3%	2%	2%	2%	2%
City	Assiut	Minya	Sohag	El Gouna	Beni Suef	Beheira	Qena	Sharqia
Counts	5	5	4	3	3	3	3	3
Percentage	2%	2%	1%	1%	1%	1%	1%	1%

Data Collection with Business-Driven Design

Egyptopia: Places Dataset MetaData Overview

Category	Museum	Beach	Shopping	Natural Reserve	Tomb
Count	34	33	26	21	18
Percentage	11%	10%	8%	7%	6%
Category	Historical Site	Temple	Church	Mosque	Palace
Count	17	16	14	12	12
Percentage	5%	5%	4%	4%	4%
Category	Theme Park	Monastery	Mountain	Island	Garden
Percentage	11	10	9	9	9
Percentage	3%	3%	3%	3%	3%
Category	Fortress	Water Park	Tower	Healing Oases & Sand Therapy	Synagogue
Count	9	7	7	7	6
Percentage	3%	2%	2%	2%	2%
Category	Hot Spring	Library	Zoo	Rehabilitation & Wellness Center	Cultural Center
Count	6	6	4	5	5
Percentage	2%	2%	1%	2%	2%

Data Collection with Business-Driven Design

EGYPTOPIA

2. Food & Drinks Dataset

A structured dataset was created to represent Egypt's rich culinary culture through 50 iconic food and drink items. Each entry includes attributes such as item ID, Arabic and English names, classification (e.g., appetizers, main dishes, hot drinks), type (food or drink), and a detailed description. . This dataset provides a comprehensive foundation for showcasing Egypt's diverse food heritage within the application.

Main Dishes	Desserts	Appetizers & Side Dishes	Cold Drinks	Hot Drinks
16	12	8	8	6

3. Events Dataset

The events dataset includes a curated collection of recurring static events that are culturally and historically significant across Egypt. These events are fixed in time and place, such as the Sun Alignment at Abu Simbel Temple, the Pyramids Marathon, and the Cairo International Book Fair. Each record in the dataset contains detailed attributes such as event name, description, location, city, date, ticket price, organizer, and a Google Maps link, allowing for precise representation within the system.

Importantly, the system is designed to allow administrators and partnered organizers to add new static events directly. This ensures that major annual occurrences—like the Sun Alignment—are always featured and updated, while also making room for new entries that may emerge from official tourism bodies or cultural organizations.

This flexible and structured approach supports the application's goal of promoting authentic Egyptian experiences, enhancing both the richness of the dataset and the relevance of the information provided to users.

EGYPTOPIA

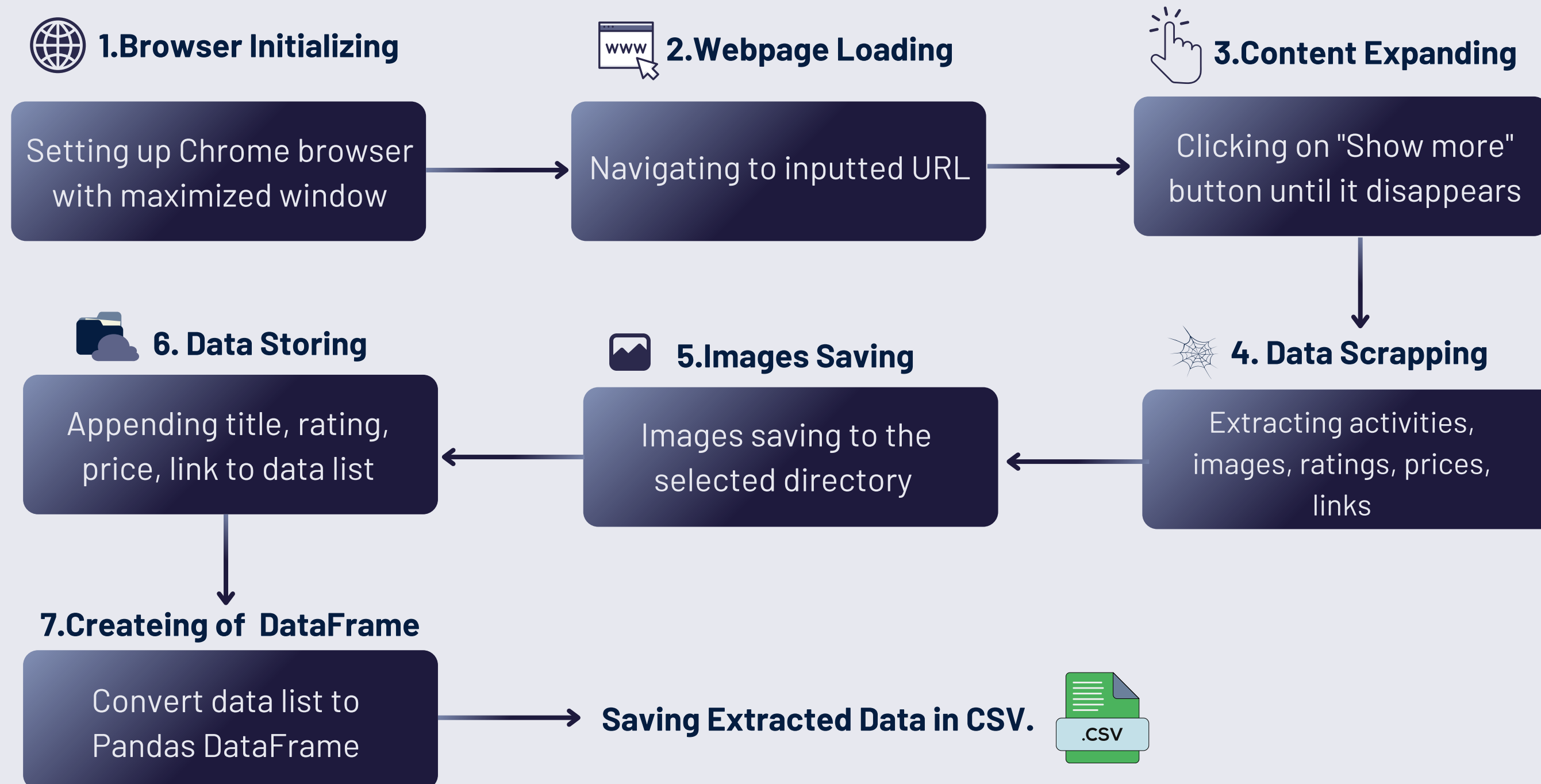**SCRAPING DATA** → **DATA PROCESSING** → **UPLOADING DATA** → .CSV

To overcome the challenge of collecting rich and accurate tourism data in Egypt, automated web scraping using Selenium with Python was employed. This approach enabled the extraction of detailed activity data from dynamic tourism platforms like GetYourGuide, covering experiences such as safaris, balloon tours, and diving. Unlike manual collection, this method provided a scalable, real-time, and structured solution. The scraped data supported more accurate modeling of the system's database schema and aligned closely with the project's goal of promoting tourism in Egypt through updated and diverse offerings.
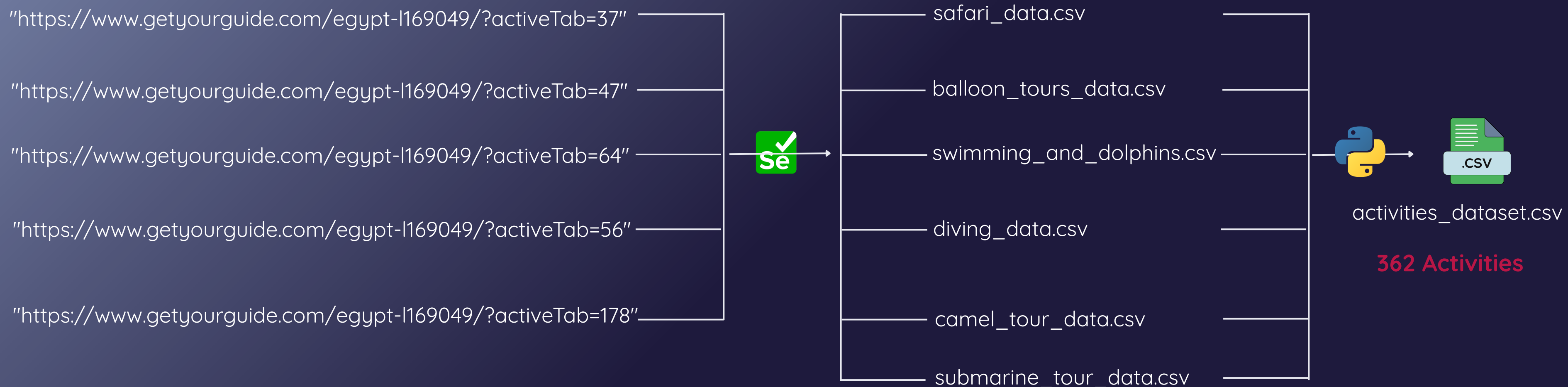
Scrapping Script Logic

EGYPTOPIA

1. Browser Initializing

Setting up Chrome browser with maximized window

2. Webpage Loading

Navigating to inputted URL

3. Content Expanding

Clicking on "Show more" button until it disappears

6. Data Storing

Appending title, rating, price, link to data list

5. Images Saving

Images saving to the selected directory

4. Data Scrapping

Extracting activities, images, ratings, prices, links

7. Createing of DataFrame

Convert data list to Pandas DataFrame

Saving Extracted Data in CSV.

.CSV

Data Collection with Web Scraping

4. Egyptopia: Activities Dataset Metadata Overview

Activity Type	Count	Average Price	Average Rating	Percentage
Ballon Tours	66	27746.7	4.44	18.23%
Camel Tours	84	3963.65	4.39	23.20%
Diving	59	3002.68	4.47	16.30%
Safari	110	4611.35	4.47	30.39%
Submarine Tours	15	1341.47	4.32	4.14%
Swimming With Dolphins	28	3852.93	4.53	7.73%

Summary of Key Insights from the Egyptopia Tourism Dataset

EGYPTOPIA

In conclusion, this dataset presents a comprehensive and nuanced representation of Egypt's tourism landscape. It comprises 320 distinct tourist sites, systematically categorized into five key segments: Cultural & Historical, Entertainment & Modern, Natural, Religious & Spiritual, and Medical Attractions. This is further enriched by a curated collection of 2,377 high-quality images, enhancing the descriptive value of the dataset.

Moreover, the dataset includes 362 documented activities—ranging from Safaris and Camel Tours to Diving and Balloon Rides—reflecting the diversity and depth of experiential tourism available across the country. The consistently high visitor ratings across these activities indicate a strong level of tourist satisfaction and engagement.

In addition, the integration of 50 iconic food and beverage items, spanning traditional Main Dishes to locally favored Hot Drinks, provides valuable insights into Egypt's rich culinary heritage.

Collectively, these components form a robust foundation for in-depth tourism analysis and offer significant potential for promoting Egypt's cultural, natural, and gastronomic assets to a global audience in a data-driven and impactful manner.