

PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA  
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH  
UNIVERSITY OF MOHAMED BOUDIAF - M'SILA

Faculty: Mathematics and Informatics

Department: Informatics

N°:.....



DOMAIN: Mathematics and Informatics

FIELD: Informatics

SUB-FIELD: INFORMATION SYSTEMS AND  
SOFTWARE ENGINEERING

A Dissertation Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Master

By:

- LEBCIR Messaouda
- MAGOURA Hadjira

**SUBJECT**

**A Web Scraping Application for Tourists'  
Assistance in Algeria**

**Board of Examiners:**

Said HAMANI	University of M'sila	President
Mahmoud BRAHIMI	University of M'sila	Supervisor
Said AMRI	University of M'sila	Examinator

**Academic year: 2021 / 2022**



**PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA**  
**MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH**  
**UNIVERSITY OF MOHAMED BOUDIAF - M'SILA**

**Faculty: Mathematics and Informatics**

**Department: Informatics**

**N°:.....**

**DOMAIN: Mathematics and Informatics**

**FIELD: Informatics**

**SUB-FIELD: INFORMATION SYSTEMS AND  
SOFTWARE ENGINEERING**

**A Dissertation Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Master**

**By:**

- LEB CIR Messaouda
- MAGOURA Hadjira

**SUBJECT**

**A Web Scraping Application for Tourists'**

**Assistance in Algeria**

**Board of Examiners:**

Said HAMANI

University of M'sila

President

Mahmoud BRAHIMI

University of M'sila

Supervisor

Said AMRI

University of M'sila

Examinator

**Academic year: 2021 / 2022**

# **Acknowledgement**

Praise be to Allah who enlightened us the path of science and knowledge and helped us to accomplish this work.

Firstly, we would like to thank our supervisor Dr. Mahmoud BRAHIMI for his guidance, supervision, and support to achieve this work.

Finally, we are grateful for our families for their unlimited motivation and support.

# Table of content

GENERAL INTRODUCTION .....	1
<b>CHAPTER I: TOURISM IN ALGERIA</b>	
1 Introduction .....	3
2 Definition of Tourism.....	3
2.1 Types of Tourism .....	3
2.2 Impact of Tourism on economy.....	4
3 Tourism in Algeria .....	5
3.1 Tourism types in Algeria .....	5
3.1.1 Saharan Tourism.....	5
3.1.2 Coastal Tourism. ....	6
3.1.3 Mountain and wildlife Tourism .....	6
3.1.4 Cultural Tourism .....	6
3.1.5 Religious Tourism .....	10
3.1.6 Medical and thermal Tourism.....	11
3.2 Infrastructure.....	12
3.2.1 Road transport.....	12
3.2.2 Ports and shipping .....	12
3.2.3 Air transport.....	12
3.2.4 Railways .....	13
3.2.5 Hotel potentialities .....	13
3.2.6 E-Tourism.....	14
3.3 Actual state of Tourism sector in Algeria.....	14
3.3.1 Tourism mouvement in Algeria .....	15
3.3.2 Tourism's contribution to the national economy.....	15
4 Conclusion.....	16
<b>CHAPTER II: BIG DATA</b>	
1 Introduction.....	17
2 Big Data .....	17
2.1 Big Data characteristics.....	18
2.2 Big Data in Tourism industry.....	19
2.2.1 What kind of data is collected .....	19
2.2.2 How can Tourism benefits from Big Data.....	20
2.3 Phases of Big Data Analytics lifecycle .....	21
3 Web Scraping.....	23
3.1 Why is Web Scraping used?.....	23
3.2 Is Web Scraping legal? .....	23
3.3 Web Scraping process.....	23

3.4 Web Scraping techniques .....	24
3.4.1 Regular expressions (Regex) .....	24
3.4.2 HTML DOM .....	24
3.4.3 XPath .....	24
3.5 Web Scraping tools .....	24
3.6 Limitations of Web Scraping .....	25
4 Conclusion .....	26

### **CHAPTER III: DESIGN AND ANALYSIS**

1 Introduction .....	28
2 Presentation of the project .....	28
3 Conceptual study of the project .....	28
3.1 UML .....	28
3.2 Characteristics of UML .....	29
3.3 UML diagrams .....	29
3.3.1 Structural diagrams .....	30
3.3.2 Behavioral diagrams .....	30
3.4 Project modeling .....	30
3.4.1 Use case diagram .....	30
3.4.2 Class diagram .....	32
3.4.3 Sequence diagram .....	33
3.4.4 Activity diagram .....	34
4 Conclusion .....	35

### **CHAPTER IV: IMPLEMENTATION**

1 Introduction .....	36
2 Presentation of the development environment .....	36
2.1 Developing languages .....	36
2.1.1 ASP.NET Core ( with C# code behind) .....	36
2.1.2 CSS 3 .....	36
2.1.3 HTML 5 .....	37
2.1.4 Javascript .....	37
2.2 Deployment frameworks .....	37
2.2.1 SQL Server .....	37
2.2.2 Entity Framework .....	37
2.2.3 MVC Architecture .....	38
2.2.4 IIS 10.0 Express .....	38
2.2.5 Web browser .....	38
2.3 Libraries .....	38
2.3.1 Selenium .....	38
2.4 Development environment .....	38
2.4.1 Visual Studio Code .....	38

3 Presentation of the web application .....	38
3.1 Application demonstration and results .....	39
4 Conclusion.....	46
GENERAL CONCLUSION .....	47
BIBLIOGRAPHY .....	48

# List of figures and tables

## List of Tables

Table 1.1 Capacity of hotels and similar establishments by type.....	12
Table 1.2 Tourists entry across Algerian borders .....	14
Table 1.3 Evolution of the share of the tourism sector in the GDP.....	15
Table 1.4 Employment in the tourism sector .....	15
Table 2.1 Web scraping tools .....	25

## List of Figures

Figure 1.1 Total contribution of tourism to world GDP and to employment.....	4
Figure 1.2 Timgad .....	6
Figure 1.3 Royal Mausleom of Mauritania - Tipaza.....	7
Figure 1.4 Djemila Cuicul .....	7
Figure 1.5 Tassili N'ajjer .....	7
Figure 1.6 Kalàat Beni Hammad-M'sila.....	8
Figure 1.7 Beni M'zab .....	8
Figure 1.8 The Casbah.....	8
Figure 1.9 Carpets.....	9
Figure 1.10 Pottery .....	9
Figure 1.11 Men Traditional Dressing.....	9
Figure 1.12 Women Traditional Dressing .....	9
Figure 1.13 Kouskous.....	10
Figure 1.14 Traditional sweets .....	10
Figure 1.15 Zawiya Tidjania .....	10
Figure 1.16 Great Mosque of Algeria .....	10
Figure 1.17 Tourist sites in Algeria .....	11
Figure 1.18 Development of tourist flow .....	14
Figure 2.1 Big Data tasks.....	17
Figure 2.2 The 5V of the Big Data .....	18
Figure 2.3 Data collected in Tourism industry.....	20
Figure 2.4 Big Data lifecycle.....	22
Figure 2.5 Web scraping .....	23
Figure 3.1 UML Diagrams .....	29
Figure 3.2 Use Case diagram of System interaction.....	31

Figure 3.3 Use Case diagram of User interaction .....	32
Figure 3.4 Class diagram .....	33
Figure 3.5 Sequence diagram .....	34
Figure 3.6 Activity diagram .....	35
Figure 4.1 ASP.NET vs ASP.NET Core .....	36
Figure 4.2 Entity Framework .....	37
Figure 4.3 Home page.....	39
Figure 4.4 List of all tours offers page .....	40
Figure 4.5 List of all hotels page .....	41
Figure 4.6 List of all restaurants page.....	42
Figure 4.7 Tour details page .....	43
Figure 4.8 Search page.....	44
Figure 4.9 About Algeria page .....	45
Figure 4.10 Tour datatable .....	46

## GENERAL INTRODUCTION

Tourism is one of the world's fastest growing industries in the world. It is a major source of income for many countries, it has a significant contribution in world GDP and also in the employment market. But in Algeria, despite its very big tourism potentials, which are rarely found in one country, from a coastal line that extends along 1200 km, and deserts classified as one of the most beautiful deserts in the world, which is able to make Tourism the wheel of the economic development, especially with the global trend to the Saharan tourism industry, which has become known as the tourism of the rich people. This kind of tourism contributes 70 percent of the tourism sector revenues in some Arab countries such as the United Arab Emirates, but unfortunately the Algerian tourism sector did not rise to the required level when compared to neighboring countries where it represents 1.1% of the Algerian market, compared 19% for Tunisia, and 14.9% for Morocco.

One of the aspects of the weakness of the Tourism sector in Algeria is the lack use of the new ICT technologies like websites and mobile applications dedicated to introduce tourism sites and resources in Algeria. Also, Information related to Algeria as a tourist destination can be accessed on various sites. For example, if the search is done through a search engine such as Google with the keyword "tourism in Algeria" then there will be many pages describing the tourist attraction in Algeria. But this will cause a difficulty in processing this Tourism data by humans.

Based on the above problems, we find that it is necessary to develop a tool that can collect this information and integrate it in one portal. Therefore, we will propose a Tourism portal website that will collect information from various Algerian tours and travel agencies websites which are then grouped according to the specified categories (tours, hotel, and restaurants). The presentation of information in this portal website will be more concise but complete and accurate. The technique used in making this portal is **Web scraping**.

To achieve this goal, we divided this dissertation in four chapters:

In the chapter one, we will present the concept of Tourism and mention the Tourism potentials in Algeria and the actual state of this sector.

The chapter two is dedicated to explain the concept of Big data with its different aspects, techniques and limitations. Also, we will present in this chapter the Web scraping technique with its specifications and its related technologies.

In the chapter three, we will provide our conceptual study of our application based on the use of UML as a modeling language.

In the last chapter we will present the implementation of this work with the set of technical tools used in this phase. Our practical work will be reserved for the development of a website portal using Web scraping allowing the collection, filtering and saving of data concerning the Tourism data available on the Algerian Tourism websites.

Finally, we will conclude this dissertation by a general conclusion and some suggestions to improve this work in the future.

# CHAPTER I

## TOURISM IN ALGERIA

---

### **1. Introduction:**

The aims of this chapter are to highlight the most important tourism potentialities acquired by Algeria, highlighting the most important tourist indicators by which the progress of the tourism sector in a given country can be measured, as well as the classification of the tourism sector in terms of the global ranking of international bodies and organizations.

### **2 Definition of Tourism:**

Tourism may be defined as the movement of the people from their normal place of residence to another place (with the intention to return) for a minimum period of twenty-four hours to a maximum of six months for the sole purpose of leisure and pleasure [1].

#### **2.1. Types of Tourism:**

Tourism has various types on the basis of the purpose of visit and alternative types. These are further divided into many types according to their nature. Here are most important types of Tourism:

- Adventure Tourism
- Bicycle Tours
- Coastal Tourism
- Cultural Tourism
- Ecotourism
- Medical Tourism
- Religious Tourism
- Rural Tourism
- Sports Tourism
- Sustainable Tourism
- Wildlife Tourism

Algeria enjoys many of these types of tourism, which we will discuss in the next sections.

## 2.2. Impact of Tourism on economy:

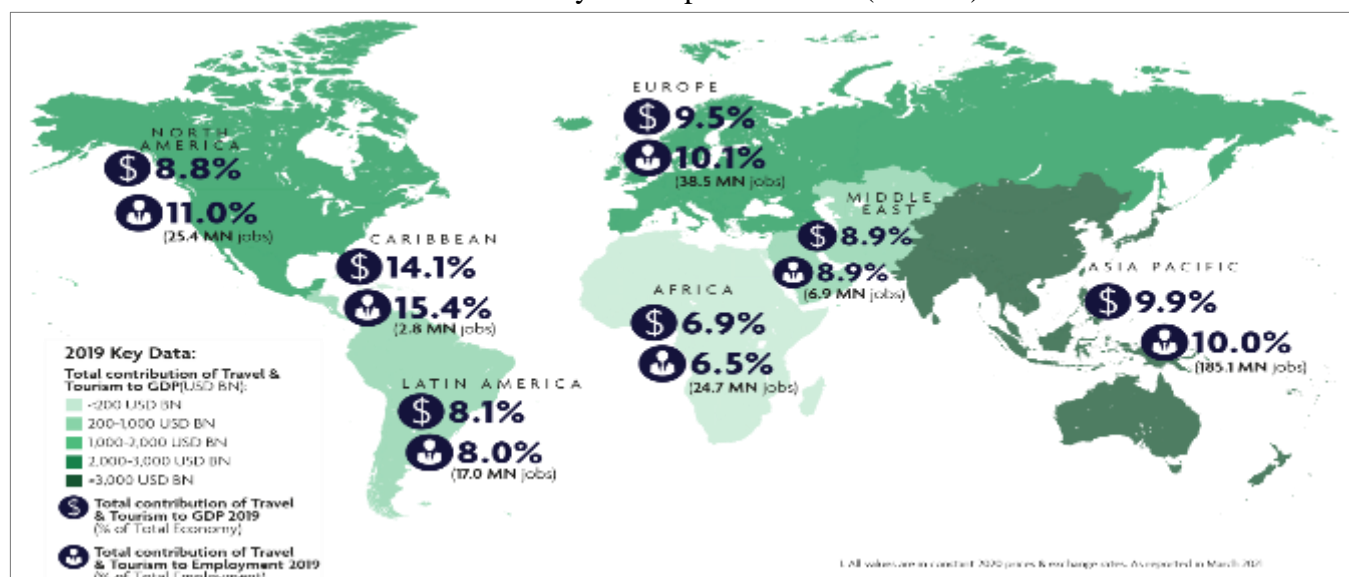
Tourism is big business; the World Tourism Organization estimates that it makes up about 10% of world GDP and creates one in every 11 jobs. It's also an industry on the rise, with huge future potential for global wealth and employment.

Tourism brings with it huge economic potential for a destination that wishes to develop their tourism industry. Employment, currency exchange, imports and taxes are just a few of the ways that tourism can bring money into a destination.

In recent years, tourism numbers have increased globally at exponential rates. There are a number of reasons for this growth including improvements in technology, increases in disposable income, the growth of budget airlines and consumer desires to travel further, to new destinations and more often.

Here are a few facts about the economic importance of the tourism industry globally: [2]

- The tourism economy represents 5 percent of world GDP
- Tourism contributes to 6-7 percent of total employment
- International tourism ranks fourth (after fuels, chemicals and automotive products) in global exports
- The tourism industry is valued at US\$ 1trillion a year
- Tourism accounts for 30 percent of the world's exports of commercial services
- 1.4billion international tourists were recorded in 2018 (UNWTO)
- In over 150 countries, tourism is one of five top export earners
- Tourism is the main source of foreign exchange for one-third of developing countries and one-half of less economically developed countries (LEDCs)



**Figure 1.1** Total contribution of tourism to world GDP and to employment [3]

### 3 Tourism in Algeria:

Algeria has a diverse landscape and lots to offer tourists, it has many charming cities with winding streets and stunning architecture, Mediterranean coast, lush landscapes and roman ruins to rival anywhere in the world. The main attraction in the country however is the Saharan region where the never-ending sand and the mysterious and lively cities are enough to indulge even the most seasoned tourist's imagination.

#### 3.1. Tourism types in Algeria:

Thanks to this diversity of natural, historical, and cultural resources, there are lot of Tourism types in Algeria:

##### 3.1.1. Saharan Tourism:

The area of the Algerian desert is about 2 million square kilometers which represents 83% of Algerian area, distributed over 14 desert regions [4]. The most important of them are:

- **Adrar:**

Located in the southwest of the Algerian desert, this region is known for the fusion of different cultures and also for its old castles.

- **Ilizi:**

Represents Tassili, which is located in the far southeast. This area is known as the National Park of Tassili, and in 1982 it was classified as a world heritage by the United Nations Educational, Scientific and Cultural Organization (UNESCO). [5]

- **The M'Zab Valley:**

A traditional human habitat, created in the 10th century by the Ibadites around their five Ksour (fortified cities), has been preserved intact in the M'Zab valley. Simple, functional and perfectly adapted to the environment, the architecture of M'Zab was designed for community living, while respecting the structure of the family. It is a source of inspiration for today's urban planners. [6]

- **Tamanraset:**

Specifically, the Ahaggar National Park which is located in the city of Tamanrasset, abounding with many livings natural evidence that is still passing through over millions of secrets of human, animal and plant existence, it has been classified as the largest museums open to nature in the world, extending over an area of 50000 km. Its rocks contain animals, plants and also forest remains, in which has more than 350 species of plants, water sources, geological sites, mines, archaeological sites, and the remains of cemeteries, where the Tuareg practice a pastoral Bedouin life.

### 3.1.2. Coastal Tourism:

The Algerian coast extends over an area of 1200 km, and is characterized by its height and rocky formation, and there are several rare tourist spaces in it. Among those areas, we find: El Kala, Taqzirt, Sidi Faraj, Tenes, Beni Saf...etc. [7]

### 3.1.3. Mountain and wildlife Tourism:

Mountainous areas are characterized by the presence of the Tell Atlas Hills and the Desert Atlas series, which give opportunities for discovery, hunting and skiing. Famous tourist highlands we find Chr ea national park and the ski resort of Tikjda.

### 3.1.4. Cultural Tourism:

Cultural heritage contains the antiquities that constitute previous eras in the past, and it is a broad concept that includes all aspects of life, and we can observe cultural heritage in old buildings, remains of antiquities, handicrafts, traditional dresses and gastronomy.

➤ **Historical sites:** Algeria has important cultural and historical sites. There are seven UNESCO World Heritage sites including: [8]

#### 1. Timgad:

Timgad lies on the northern slopes of the Aur s Mountains and was created as a military colony by the Emperor Trajan in A.D. 100. With its square enclosure and orthogonal design based on the cardo and decumanus, the two perpendicular routes running through the city, it is an excellent example of Roman town planning.



**Figure 1.2** Timgad

#### 2. Tipaza:

On the shores of the Mediterranean, Tipaza was an ancient Punic trading post conquered by Rome and turned into a strategic base for the conquest of the kingdoms of Mauretania. It comprises a unique group of Phoenician, Roman, Paleo-Christian and Byzantine ruins alongside indigenous monuments such as the Kbor er Roumia, the great royal mausoleum of Mauretania, sometimes known as the “tomb of the Christian woman”.



**Figure 1.3** Royal Mausoleom of Mauritania - Tipaza

3. Djemila:

Situated 900 meters above sea-level, Djemila, or Cuicul, with its forum, temples, basilicas, triumphal arches and houses, is an interesting example of Roman town planning adapted to a mountain location.



**Figure 1.4** Djemila Cuicul

4. Tassili N'ajjer:

This site is home to one of the most important groupings of prehistoric cave art in the world. More than 15,000 drawings and engravings record the climatic changes, animal migrations and evolution of human life on the edge of the Sahara from 6,000 B.C. to the first centuries of the present era.



**Figure 1.5** Tassili N'ajjer

#### 5. Kalâat Béni-Hammade:

In a mountainous site of extraordinary beauty, the ruins of the first capital of the Hammadid emirs, founded in 1007 and demolished in 1152, provide an authentic picture of a fortified Muslim city. The mosque, whose prayer room has 13 aisles with eight bays, is one of the largest in Algeria.



**Figure 1.6** Kalâat Beni Hammad-M'sila

#### 6. The M'Zab Valley:

A traditional human habitat created in the 10th century by the Ibadites around their five ksour (fortified cities), has been preserved intact in the M'Zab valley. Simple, functional and perfectly adapted to the environment, the architecture of M'Zab was designed for community living while conforming to the structure of the family. It is a source of inspiration for today's urban planners.



**Figure 1.7** Beni M'zab

#### 7. The Casbah:

The Casbah is a unique kind of medina or Islamic city. It stands in one of the finest coastal sites on the Mediterranean where a Carthaginian trading-post was established in the 4th century B.C. The Kasbah contains the remains of the citadel, old mosques and palaces as well as the vestiges of a traditional urban structure associated with a deep-rooted sense of community.



**Figure 1.8** The Casbah

➤ **Handicrafts:** the nomenclature of craft activities and trades includes 339 craft activities for a number 58 473 jobs in 2020 [9]. The diversity of techniques, natural resources added to ancestral know-how explains this great artisanal variety, here are some field of activities:

1. Jeweler.
2. Pottery and ceramics.
3. Copperwares.
4. Carpets and weavings.
5. Carved wood.
6. Traditional dresses:



**Figure 1.9** Carpets



**Figure 1.10** Pottery

➤ **Traditional dresses:** Algerian traditional dresses are a selection from local traditions and the influence of various civilizations that have lived on Algeria, and since it is a meeting point of civilizations, Algeria has a variety of dress and costumes like Burnous and Gandoura for men, El Hayek, Chedda dress, Touareg dress, Kabyle dress, Karakou, and Bedroune for women.



**Figure 1.11** Men Traditional Dressing



**Figure 1.12** Women Traditional Dressing

➤ **Gastronomy:** Algerian cuisine offers a variety of dishes depending on the region and the season, but vegetables and cereals remain at its core. Most of the Algerian dishes are centered around bread, meats (lamb, beef or poultry), olive oil, vegetables and fresh herbs. Vegetables are often used for salads, soups, tajines, couscous and sauce-based dishes. Of all the Algerian traditional dishes available, the most famous one is couscous, recognized as a national dish, in addition to: Chakhchoukha, Mhadjeb, Tajines, Bourek, Chorba, etc. Also, Algeria is famous by its variety of delicious sweets like Makroud, Baklawa, Griwech, Mkhabez, etc



**Figure 1.13** Kouskous



**Figure 1.14** Traditional sweets

### 3.1.5. Religious Tourism:

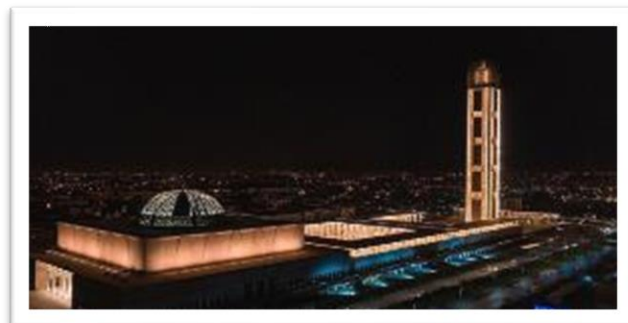
Algeria has many religious places and shrines that attract thousands of visitors from different countries, especially African ones. Zawiyas (religious institutions) are receiving great attention from those who follow these paths and who are influenced by their divine spiritual approach. Among those Zawiyas we find:

- Zawiya Bekaidia in Oran.
- Zawiya Tidjania in Laghouat.
- Zawiya Sidi Abderrahmene Thàalibi in Algiers.
- Zawiya El Hamel in M'sila.

Also, Algeria has many Mosques having significant historical value, for example we find the Great Mosque of Tlemcen, the Great Mosque of Algiers, and the Ketchaoua Mosque in Algiers which is a UNESCO World Heritage site [10]. There are also historical churches that can be classified as tourist attraction sites like the Notre Dame d’Afrique church in Algiers and Saint Augustin Basilica in Annaba.



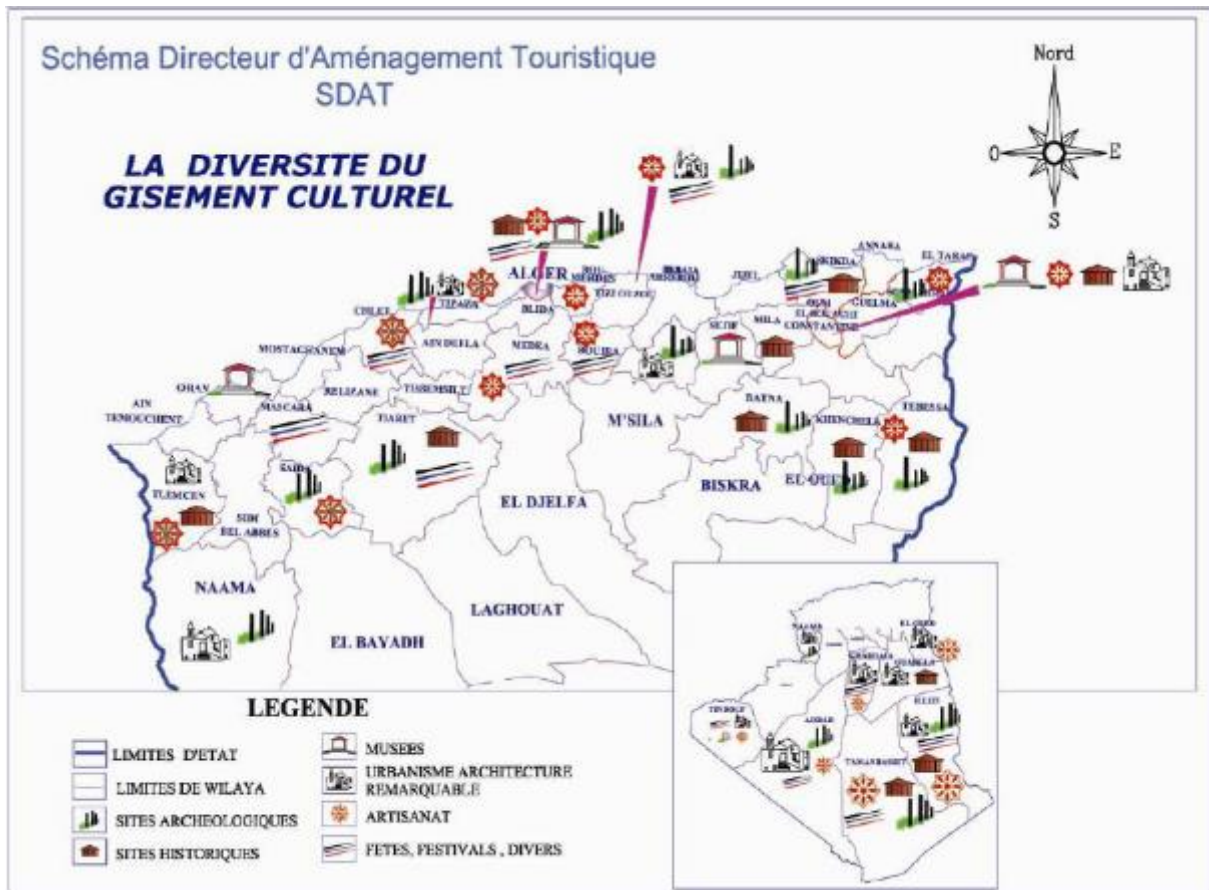
**Figure 1.15** Zawiya Tidjania



**Figure 1.16** Great Mosque of Algeria

### **3.1.6. Medical and thermal Tourism:**

As regard to thermalism, the thermal assessment emerges with a census of 202 sites open to exploitation, most of them are situated in the north of the country [11]. Among these thermal springs, some present a national vocation and a therapeutic interest of first rate which require an improvement considering the needs expressed on the subject. The most famous of these springs are: Righ Hammam, Ain Defla, Hammam Bouhanifia in Mascara, Hammam guergour in Setif, Hammam al-Salihin in Biskra, Hammam Dabbagh and Shallala in the province of Guelma...etc.



**Figure 1.17** Tourist sites in Algeria [12]

### 3.2. Infrastructure:

#### 3.2.1. Road transport:

The Algerian Road network is considered one of the largest networks in the African continent, with an estimated length of 112,696 km, of which 29,280 km of national roads, 26,626 km of state roads, 62,100 km of town roads.

#### 3.2.2. Ports and Shipping:

95% of Algeria's external foreign trade takes place through ports that's why the numbers of ports increase to 52 ports in 2013, among those ports there are: 11 commercial ports including: Algiers, Oran, Annaba, Skikda, Arzew, 2 oil ports, and 37 ports for fishing.

#### 3.2.3. Air transport:

Air transport is considered the most important means of transportation in revitalizing the international tourism movement towards Algeria, Algeria owns 35 airports, 13 of which are international, Algiers airport is the most important, attracting 6 million passengers annually. Algeria Airlines dominates the air transport market, which has registered since its opening for competition 8 other private companies.

**3.2.4. Railways:**

The railway network in Algeria is estimated at 2,150 km, as it has recently witnessed electrification of some sections to place high-speed trains close to linking the country's big cities. The railway network is operated by the National Railways Transport Corporation (SNTF). This network is equipped with more than 200 stations covering mainly the north of the country, including:

299 km of electrified railways.

305 km double rail.

1085 km narrow rail.

**3.2.5. Hotel potentialities:**

The capacity of accommodation and hospitality in Algeria increased from 76,000 beds in 1994 to 127,614 beds in 2020. The number of hotels at the end of the same year reached 1,449 classified hotels, and the number of tourism projects in the process of completion reached 804 projects. Also, the number of tourism and travel agencies reached 3,546 agencies.

Classification	2017		2018	
	Number	Beds number	Number	Beds number
Hotel (*****)	13	6734	13	6734
Hotel (****)	23	4508	24	4746
Hotel (***)	59	5678	62	5886
Hotel (**)	48	4565	52	5185
Hotel (*)	159	11335	162	11684
Hotel (without *)	160	8533	162	8590
Tourist accommodation (**)	2	384	2	384
Tourist accommodation (*)	1	313	1	313
Hostel (**)	2	93	2	93
Hostel (*)	1	30	1	30
Youth Hostel (**)	1	16	1	16
Youth Hostel (*)	1	20	1	20
Holiday village (***)	1	274	1	274
Tourism apartment	5	91	5	91
Tourist Residence	10	426	10	426
Mountain lodge	6	170	6	170

To be classified	601	59713	659	64727
Other hotel structures	196	9381	204	9786
<b>Total</b>	<b>1289</b>	<b>112264</b>	<b>1368</b>	<b>119155</b>

**Table 1.1** Capacity of hotels and similar establishments by type [13]

As you can notice, the accommodation capacity is by far the most detrimental factor to the development of tourism in Algeria. The number of hotel beds is small that it can't meet local demand, let alone accommodating more international tourists. In this situation the government should work to find the appropriate solutions for this problem in order to eventually bring the number of accommodation establishments to an equilibrium point where the supply meets the demand.

### **3.2.6. E-Tourism in Algeria:**

The tourists appeal to the sites of touristic information and especially to the sites that also allow bookings for accommodation or transport services. In Algeria, there are many official tourism websites for promoting and selling tourism products of Algeria such as:

- The website of Tourism Ministry: <https://www.mta.gov.dz/> and all organizations that related with it (National Office of Tourism: [www.ont.dz](http://www.ont.dz) and National Office of Algerian Tourism: <https://onatdz.com/>)
- Touring Algeria: <https://touring-algeria.com/>
- Group-htt (group of governmental hotels and spa resorts): <http://groupe-htt.com/>

But all these websites don't reflect the big potentialities of Algeria tourism destination. Also, these websites are weak from design, information, presentation and overall impression.

### **3.3. Actual state of tourism sector in Algeria:**

Algeria ranked **116th** in the world out of **140** countries in the tourism and travel competitiveness indicator [14], according to the tourism and travel competitiveness report issued by the World Economic Forum Davos which is submitted every two years.

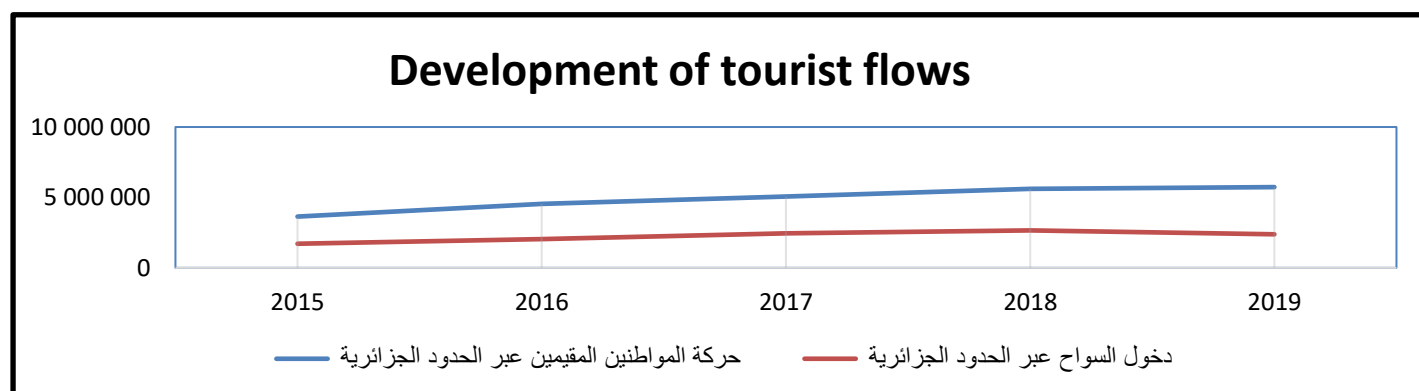
The Algerian government needs to invest in developing the appropriate infrastructure to enable tourism industry to compete internationally. However, Algeria still lacks the proper infrastructure for example, according to the previous table there were only 13 five-star hotels in 2018, while the number of unclassified hotels was estimated at 160, which reflects the weakness of accommodation in Algeria.

### 3.3.1. Tourism movement in Algeria:

Tourism movement in Algeria witnesses a fluctuation as the statistics confirms and this returns to many reasons, these are some statistics concerning Tourism in Algeria in the recent years.

	2015	2016	2017	2018	2019
<b>Total foreigners</b>	1 083 121	1 322 712	1 708 375	2 018 753	1 933 778
<b>Growth rate</b>		22,12%	29,16%	18,17%	-4,21%
<b>Algerian expatriates</b>	626 873	716 732	742 410	638 360	437 278
<b>Growth rate</b>		14,33%	3,58%	-14,02%	-31,50%
<b>Total</b>	1 709 994	2 039 444	2 450 785	2 657 113	2 371 056
<b>Growth rate</b>		19,27%	20,17%	8,42%	-10,77%

**Table 1.2:** Tourists entry across Algerian borders [15]



**Figure 1.18** Development of tourist flow [16]

### 3.3.2. Tourism's contribution to the national economy:

Various countries considered Tourism as a source for national income and succeeded in doing such as Spain, France, Switzerland and even some Arab countries like Tunisia, Morocco and Egypt.

Years	2015	2016	2017	2018	2019	2020
<b>Share of tourism in GDP (%)</b>	1,3	1,4	1,6	1,7	1,8	1.10

**Table 1.3:** Evolution of the share of the tourism sector in the GDP [17]

Years	2015	2016	2017	2018	2019
Number of jobs	265 803	270 317	300 000	308 027	320 000

**Table 1.4:** Employment in the tourism sector [17]

#### **4. Conclusion:**

Algeria has great tourism resources, whether natural or historical, but the lack of focus and the interest on this sector make it a marginal sector and reduce benefiting from it as one of the important GDP sources. One of the aspects of this neglect to this sector is the weakness use of the new TIC technologies like websites and mobile applications dedicated to introduce tourism sites and resources in Algeria. Therefore, we will develop a tourist portal that highlights the beauty of the Algerian destination by using Big Data Analysis and web scraping techniques. In the next chapter, we will discuss the Big Data and its role in improving Tourism industry.

# CHAPTER II

## BIG DATA

---

### **1. Introduction:**

The evolution of computer systems leads companies to process more and more data from ever more varied sources. The growth of processed data exceeds the limits of traditional technologies. We speak of petabyte “billiard of bytes” or even of zettabyte “trillion of bytes”. Following this evolution, the notion of "Big data" is strongly imposed to deal with this avalanche of data. Web scraping is one of the techniques to fetch a large volume of public data from websites. In this chapter, we will present the field of Big data with its concepts, technologies and applications, focusing on the task of web scraping. Therefore, we will see how this field can strongly contribute to the evolution of tourism through various applications ranging from simple tasks of collection and presentation of information to decision-making support.

### **2. Big Data:**

Big data is a term that describes large, hard-to-manage volumes of data – both structured and unstructured – that inundate businesses on a day-to-day basis. But it’s not just the type or amount of data that’s important, it’s what organizations do with the data that matters. Big data can be analyzed for insights that improve decisions and give confidence for making strategic business moves [18].

Big data refers to data that is so large, fast or complex that it’s difficult or impossible to process using traditional methods. The act of accessing and storing large amounts of information for analytics has been around for a long time. But the concept of big data gained momentum in the early 2000s when industry analyst Doug Laney articulated the now-mainstream definition of big data as the three V’s: Volume, Velocity and Variety. These 3V’s and with the evolution of big data practicals are now extended to 5V’s: Volume, Velocity, Variety, Veracity and Value.



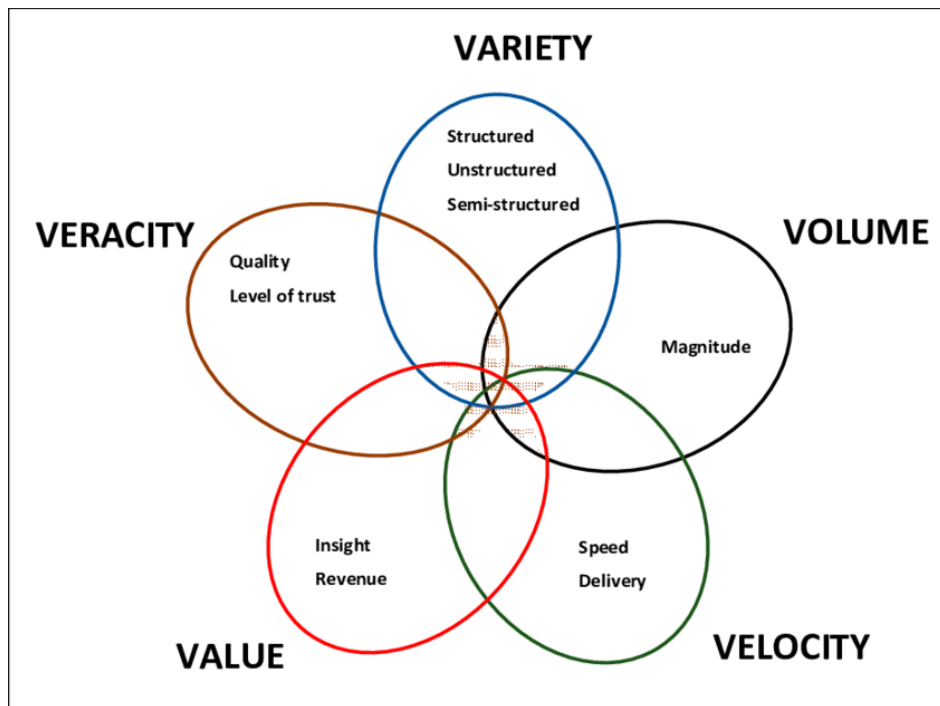
**Figure 2.1** Big Data tasks

### 2.1. Big Data characteristics:

Big data is a collection of data from many different sources and is often described by five characteristics: volume, value, variety, velocity, and veracity: [19]

- **Volume:** the size and amounts of big data that companies manage and analyze
- **Value:** the most important “V” from the perspective of the business, the value of big data usually comes from insight discovery and pattern recognition that led to more effective operations, stronger customer relationships and other clear and quantifiable business benefits
- **Variety:** the diversity and range of different data types, including unstructured data, semi-structured data and raw data
- **Velocity:** the speed at which companies receive, store and manage data – e.g., the specific number of social media posts or search queries received within a day, hour or other unit of time
- **Veracity:** the “truth” or accuracy of data and information assets, which often determines executive-level confidence.

The additional characteristic of variability can also be considered:



**Figure 2.2** The 5V of the Big Data

## 2.2. Big Data in Tourism industry:

There has long been a plethora of data on tourism but turning it into useful information and using it to inform wise decisions have been a persistent challenge. While people generate a lot of data on the internet purposely or accidentally there are several other alternative Data sources that the travel industry can use to make better decisions. These datasets can help the travel industry in seeing their “blind spots” and make more analytics-driven decisions.

Companies and destination marketing organizations in the tourism sector can benefit from Big Data in many ways. That includes marketing campaigns, offering packages sorted to visitor’s likely interests or passenger profiles, to focus on getting customers in from new source markets.

These companies can identify their potential customers while increasing the efficiency and the quality of services given at every stage in the planning process and customer journey. Big Data can also be used as a predictive tool to forecast which new products, hotels, and tours might work well in their destination.

### 2.2.1. What kind of Data is collected?

The travel industry collects massive amounts of data about its customers. This data comes from a variety of sources, both internal and external. It can be structured, as with the fields and records in a database, or unstructured free-form data. In the travel industry, big data can include information about: [20]

- **Destination Information:** Places to visit, rules & regulations, etc.,
- **Hotels & Restaurant Data:** Customer Reviews & Ratings, Pricing, Service Details, etc.,
- **Transportation Data** — Air Tickets, Route Maps, Traffic Data
- **Events Data:** Sports, Festivals, Summits etc.,
- **Government Schemes Data**
- **Tourist Data:** Inflow & Outflow of Tourists, Place of Origin, Age, Languages Spoken, etc.,
- **Social Data:** Travelers social media posts with geotags, brand tags, travel-related hashtags etc.,
- **Travel & Hospitality News**
- **Miscellaneous Data:** Web search, website visits, passport & visa application requests, international education and employment, etc.,



**Figure 2.3** Data collected in Tourism industry

### 2.2.2. How can Tourism industry benefits from Big Data?

Tourism boards and companies in the tourism sector can benefit from data of this type in many ways. That includes pinpointing marketing campaigns, offering packages tailored to visitors' likely interests, and deciding which countries to focus on winning customers in. Here are some ways that Big Data can benefit travel industry: [21]

- **Tourism market research:** Tourism trends, tourist habits & behaviors, the impact of certain schemes/campaigns on the influx of tourists, etc can be analyzed.
- **Revenue management:** Hotels and restaurants can have information around occupancy rates, current bookings, prices, etc and combine it with external data like climate conditions, school vacations, local events, flight information, etc to effectively forecast the demands and make strategies for maximizing the revenue. For example, sensibly hiking the prices.
- **Customer experience:** By monitoring the brand image using social media reviews and also analyze competitor pricing strategies and intelligently price the services.
- **Strategic marketing:** In the travel industry, it might be difficult to get the right marketing done because potential customers are varied in who they are, where they come from, and what they are looking for. But big data helps tourism companies to adopt a more strategic approach in their marketing efforts, targeting the right people in the right way.
- **Targeted marketing:** Diversity, leisure, business travel, meeting relatives, etcetera are the leading factors influencing people to travel. If you know the intent behind travel, you can personalize your marketing strategies to clock more sales. The more qualified data you have, the better you can execute targeted personalized marketing.

### 2.3. Phases of Big Data analytics lifecycle:

Big Data analysis differs from traditional data analysis primarily due to the volume, velocity and variety characteristics of the data being processes. To address the distinct requirements for performing analysis on Big Data, a step-by-step methodology is needed to organize the activities and tasks involved with acquiring, processing, analyzing and repurposing data. Big Data analytics involves mainly four important phases that are carried out in a cycle. These four phases are:

- **Big Data Collection:** In data collection phase, data from different sources comes with different formats: structured, semi structured, and unstructured. Depending on the business scope

of the analysis project and nature of the business problems being addressed, the required datasets and their sources can be internal and/or external to the enterprise.

- **Big Data preprocessing:** Big data preprocessing refers to a series of operations such as “cleaning, filling, smoothing, merging, normalization, consistency check” and other operations on the collected raw data before data analysis, in order to improve the data Quality lays the foundation for later analysis work. Data preprocessing mainly includes four parts

- Data cleaning
- Data integration
- Data conversion
- Data specification

- **Big Data storage:** Once the data is processed, it sometimes needs to be stored in a database. Big data technologies offer plenty of alternatives regarding this point. The most common alternative is using the Hadoop File System for storage that provides users a limited version of SQL, known as HIVE Query Language. This allows most analytics task to be done in similar ways as would be done in traditional BI data warehouses, from the user perspective. Other storage options to be considered are MongoDB, Redis, and SPARK.

- **Big Data analysis and mining:** Finally, the analytics phase comes up with new information and valued knowledge to be used by decision makers. The created knowledge is considered as sensitive information especially in a competition environment. Organizations take care of their sensitive information to be far away from their rivals. Further, they aware of their sensitive data (e.g., client personal data) not to be publicly released.



**Figure 2.4** Big Data lifecycle

In this dissertation, we will perform the first phase of Big Data lifecycle which is **Data collection phase**. We will use the **Web scraping technique** in order to gather tourism data (formal data) from the Algerian tourism websites. So let's first see what Web Scraping is and how to use it.

### 3. Web scraping:

Web scraping is an automated method used to extract large amounts of data from websites. The data on the websites are unstructured. Web scraping helps collect these unstructured data and store it in a structured form. There are different ways to scrape websites such as online Services, APIs or writing your own code [22].



**Figure 2.5** Web scraping

#### 3.1 Why is Web scraping used?

Web Scraping is used for getting data. Access to relevant data, having methods to analyze it and performing intelligent actions based on analysis can make a huge difference in the success and growth of most businesses in the modern world. Data collection and analysis is important even for government, non-profit and educational institutions.

The following are few of the many uses of Web Scraping:

- In eCommerce, Web Scraping is used for competition price monitoring.
- In Marketing, Web Scraping is used for lead generation, to build phone and email lists for cold outreach.
- In Real Estate, Web Scraping is used to get property and agent/owner details.
- Web Scraping is used to collect training and testing data for Machine Learning projects.

#### 3.2 Is Web scraping legal?

Scraping data which is already available in public domain is legal as long as you use the data ethically. If websites wish to prevent web scraping, they can employ techniques like CAPTCHA forms and IP banning.

### 3.3 Web scraping process:

When you run the code for web scraping, a request is sent to the URL that you have mentioned. As a response to the request, the server sends the data and allows you to read the HTML or XML page. The code then, parses the HTML or XML page, finds the data and extracts it.

To extract data using web scraping with python, you need to follow these basic steps:

1. Find the URL that you want to scrape
2. Inspecting the Page
3. Find the data you want to extract
4. Write the code
5. Run the code and extract the data
6. Store the data in the required format

### 3.4 Web scraping techniques:

In order to do automatic or semi-automatic extractions, it is necessary to identify the data of interest in the analyzed documents analyzed in order to separate them from the entire content. Here is a non-exhaustive list of techniques that can be used [23]:

#### 3.4.1 Regular Expression (Regex):

Regular Expression (Regex) is a formula with a specific pattern that describes a set of words above several alphabets. Regex can be used to match certain character patterns in a set of strings. There are two types of regular expressions namely ordinary characters and metacharacters.

#### 3.4.2 HTML DOM:

Hyper Text Markup Language Document Object Model (HTML DOM) is a standard for getting, changing, adding, or deleting HTML elements. DOM performance is by defining objects and properties of all HTML elements, with methods to access them. With DOM, JavaScript can access all elements in an HTML document. HTML DOM uses programming languages to access objects, usually JavaScript. All HTML elements are treated as objects. The programming interface is the method and property of each object.

#### 3.4.3 XPath:

XPath is the main element in the XSLT standard (Stylesheet Language Transformation). XPath can be used to navigate elements and attributes in eXtensible Markup Language (XML) documents. XPath is a language for selecting nodes in XML documents, can also be used with HTML. The most useful XPath expression is the location path. A path location at least uses one step location to identify a set of nodes in the document. The simplest location

path is one that selects the document root node. This road is just a slash "/". The symbol is the root of the Unix system file and also the root node of a document.

### 3.5 Web scraping tools:

There are several software and techniques for doing web scraping. The following list illustrates only "low-level" tools (i.e., which do not provide for a more or less explicit purpose, but which integrate scraping into a more extensive programming logic).

Tool	Language	Description
Rvest	R	R package that allows you to easily import data from web pages
Goutte	PHP	PHP Library for Web Scraping/crawling
JQuery	JavaScript	Library that allows to do scraping on the client side
Scrapy	Python	Python library for Web Scraping/crawling
HtmlAgilityPack	C#	HTML parser written in C# to read/write DOM and supports plain XPATH or XSLT

**Table 2.1** Web scraping tools

### 3.6 Limitations of the Web scraping [24]:

#### 1. Learning curve

Even the easiest scraping tool takes time to master. Some tools, like Apify, still require coding knowledge to use. Some non-coder friendly tools may take people weeks to learn. To scrape websites successfully, knowledge about XPath, HTML, AJAX is necessary. So far, the easiest way to scrape websites is to use prebuilt web scraping templates to extract data within clicks.

#### 2. The structure of websites change frequently

Scraped data is arranged according to the structure of the website. Sometimes you revisit a site and will find the layout changed. Some designers constantly update the websites for better UI, some may for the sake of anti-scraping. The change could be as small as a position change of a button, or a drastic change of overall page layout. Even a minor change can mess up your data. As the scrapers are built according to the old site, you have to adjust your crawlers every few weeks to get correct data.

#### 3. It is not easy to handle complex websites

Here comes another tricky technical challenge. If you look at web scraping in general, 50% of websites are easy to scrape, 30% are moderate, and the last 20% are rather tough to scrape from. Some scraping tools are designed to pull data from simple websites that apply numbered navigation. Yet nowadays, more websites are starting to include dynamic elements such as AJAX. Big sites like Twitter apply infinite scrolling, and some websites need users to click on the "load more" button to keep loading the content. In this case, users require a more functional scraping tool.

#### **4. To extract data on a large scale is way harder**

Some tools are not able to extract millions of records, as they can only handle a small-scale scraping. This gives headaches to e-commerce business owners who need millions of lines of regular data feeds straight into their database. Cloud-based scrapers like Octoparse and Web Scraper perform well in terms of large-scale data extraction. Tasks run on multiple cloud servers. You get rapid speed and gigantic space for data retention.

#### **5. A web scraping tool is not omnipotent**

What kinds of data can be extracted? Mainly texts and URLs. Advanced tools can extract texts from source code (inner & outer HTML) and use regular expressions to reformat it. For images, one can only scrape their URLs and convert the URLs into images later. If you are curious about how to scrape image URLs and bulk download them, you can have a look at [How to Build an Image Crawler Without Coding](#).

What's more, it is important to note that most web scrapers are not able to crawl PDFs, as they parse through HTML elements to extract the data. To scrape data from PDFs, you need other tools like Smallpdf and PDFelements.

#### **6. Your IP may get banned by the target website**

Captcha annoys. Does it ever happen to you that you need to get past a captcha when scraping from a website? Be careful, that could be a sign of IP detection. Scraping a website extensively brings heavy traffic, which may overload a web server and cause economic loss to the site owner. To prevent getting blocked, there are many tricks. For example, you can set up your tool to simulate the normal browsing behavior of a human.

#### **7. There are even some legal issues involved**

Is web scraping legal? A simple “yes” or “no” may not cover the whole issue. Let's just say... it depends. If you are scraping public data for academic uses, you should be fine. But if you scrape private information from sites clearly stating any automated scraping is disallowed, you may get yourself into trouble. LinkedIn and Facebook are among those who clearly state that “we don't welcome scrapers here” in their robots.txt file/terms and service (ToS). Mind your acts while scraping.

#### **4. Conclusion:**

In this chapter, we have approached the field of Big Data with its potentialities which can strongly contribute to the evolution of tourism. We have also focused on Web Scraping which is the basic concept of our work, therefore we will present in the next chapter our conceptual study based on web scraping.

## CHAPTER III

# DESIGN AND ANALYSIS

---

### 1. Introduction:

In this chapter, we will provide a presentation of our work by explaining the main aim of it, then we will perform the conceptual study of the project, we have chosen the UML (Unified Modeling Language) approach to accomplish the modeling of our work.

### 2. Presentation of the project:

The main goal of this project is to build a firm foundation for online public tourism data collection for the field of tourism in Algeria. The amount of data is very crucial to data analytics process, therefore a web scraping tool which can automatically collect the public tourism data from the Internet is our choice in this project.

This project will focus on scraping the public tourism data from Algerian agencies websites. Through observation, it is noticed that many agencies' websites have provided recommending services for several categories of tourism information.

This system will target 3 main travel categories which are **guided tours**, **hotels**, and **restaurants** as tourists will always concern about the questions such as where to go, where to stay, and what to visit. As this project is currently solely intended for Algerian tourism aspect, therefore the websites used for scraping are the Algerian travel agencies websites.

### 3. Conceptual study of the project:

#### 3.1. UML:

UML (Unified Modeling Language) is defined as a graphical and textual modeling language intended to understanding and describing needs, specifying and documenting systems, it is a formal and standardized language which, thanks to its graphic representation, makes it possible to design solutions and facilitates their understanding. [25]

UML has an essential standard for analysis and design activities, and in particular makes it possible to:

- Understand and describe needs
- Specify a system
- Establish the software architecture

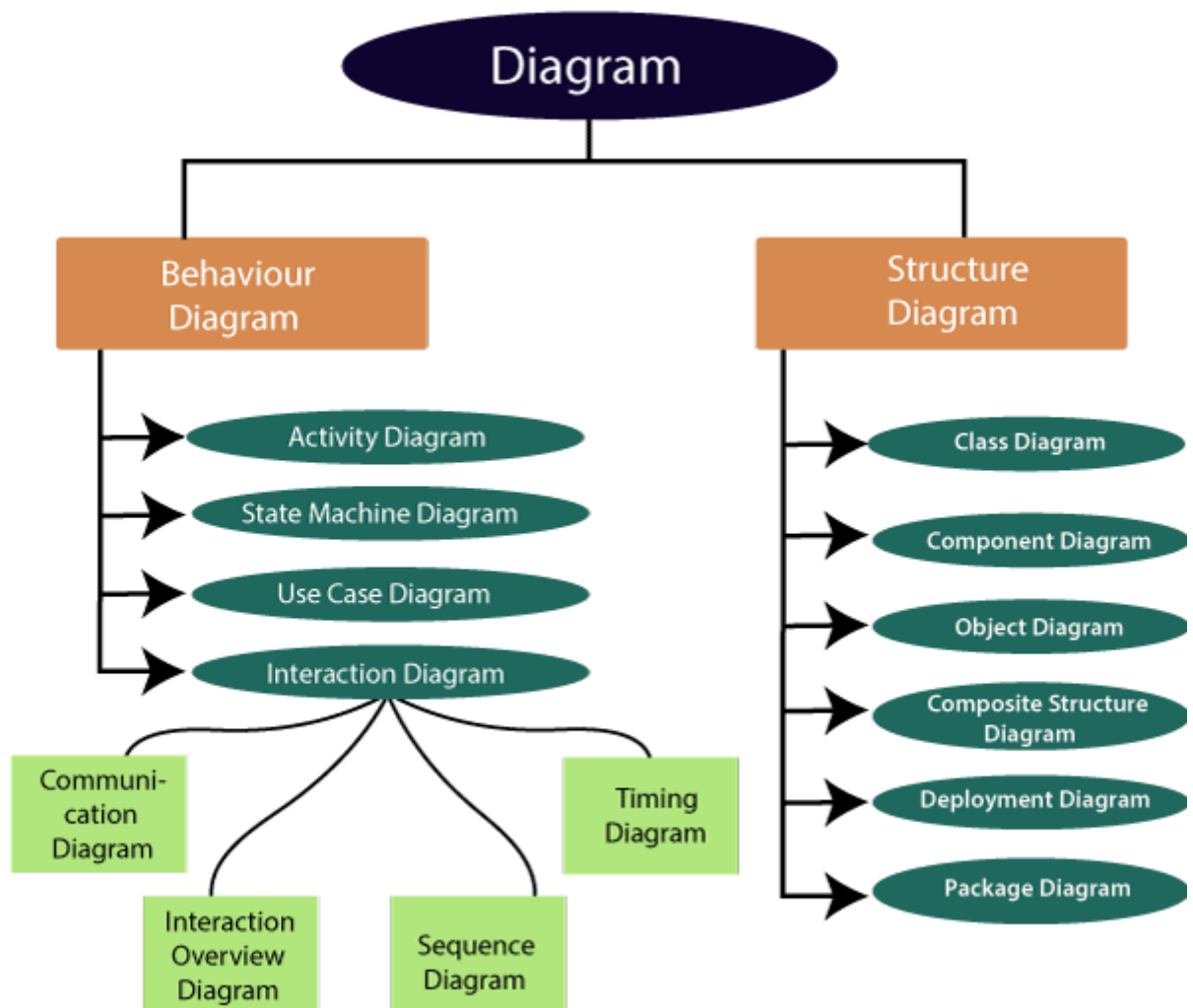
### 3.2. Characteristics of UML:

The UML has the following features: [25]

- It is a generalized modeling language.
- It is distinct from other programming languages like C++, Python, etc.
- It is interrelated to object-oriented analysis and design.
- It is used to visualize the workflow of the system.
- It is a pictorial language, used to generate powerful modeling artifacts.

### 3.3. UML diagrams:

The UML diagrams are categorized into **structural diagrams**, **behavioral diagrams**, and also interaction **overview diagrams**. The diagrams are hierarchically classified in the following figure:



**Figure 3.1** UML Diagrams [25]

### **3.3.1. Structural Diagrams:**

Structural diagrams depict a static view or structure of a system. It is widely used in the documentation of software architecture. It embraces class diagrams, composite structure diagrams, component diagrams, deployment diagrams, object diagrams, and package diagrams. It presents an outline for the system. It stresses the elements to be present that are to be modeled. UML has the following types of behavioral diagrams: [25]

- Class diagram
- Object diagram
- Component diagram
- Deployment diagram

### **3.3.2. Behavioral diagrams:**

Any system can have two aspects, static and dynamic. So, a model is considered as complete when both the aspects are fully covered. Behavioral diagrams basically capture the dynamic aspect of a system. Dynamic aspect can be further described as the changing/moving parts of a system. UML has the following five types of behavioral diagrams [26]

- Use case diagram
- Sequence diagram
- Collaboration diagram
- State chart diagram
- Activity diagram

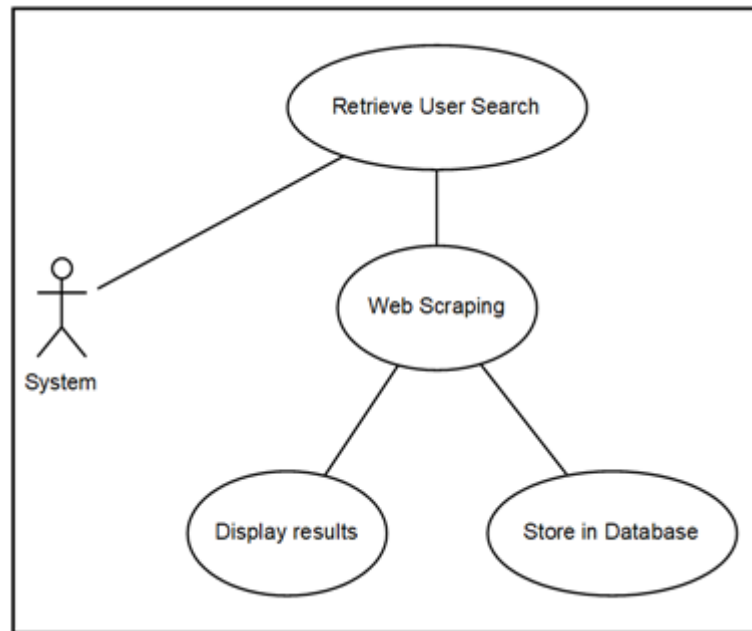
## **3.4. Project modeling:**

To model our project, we will use four diagrams, which are:

### **3.4.1. Use Case diagram:**

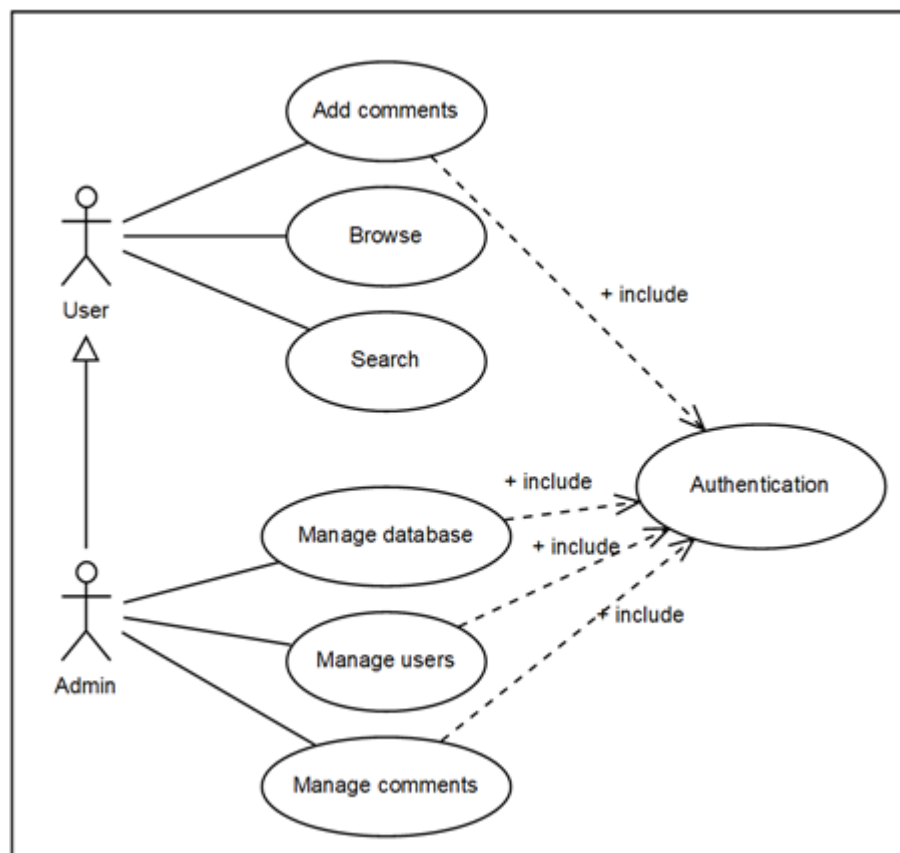
Use case diagrams model the functionality of a system using actors and use cases. Use cases are a set of actions, services, and functions that the system needs to perform. The "actors" are people or entities operating under defined roles within the system. [27]

- Use Case diagram of System interaction:



**Figure 3.2** Use Case diagram of System interaction

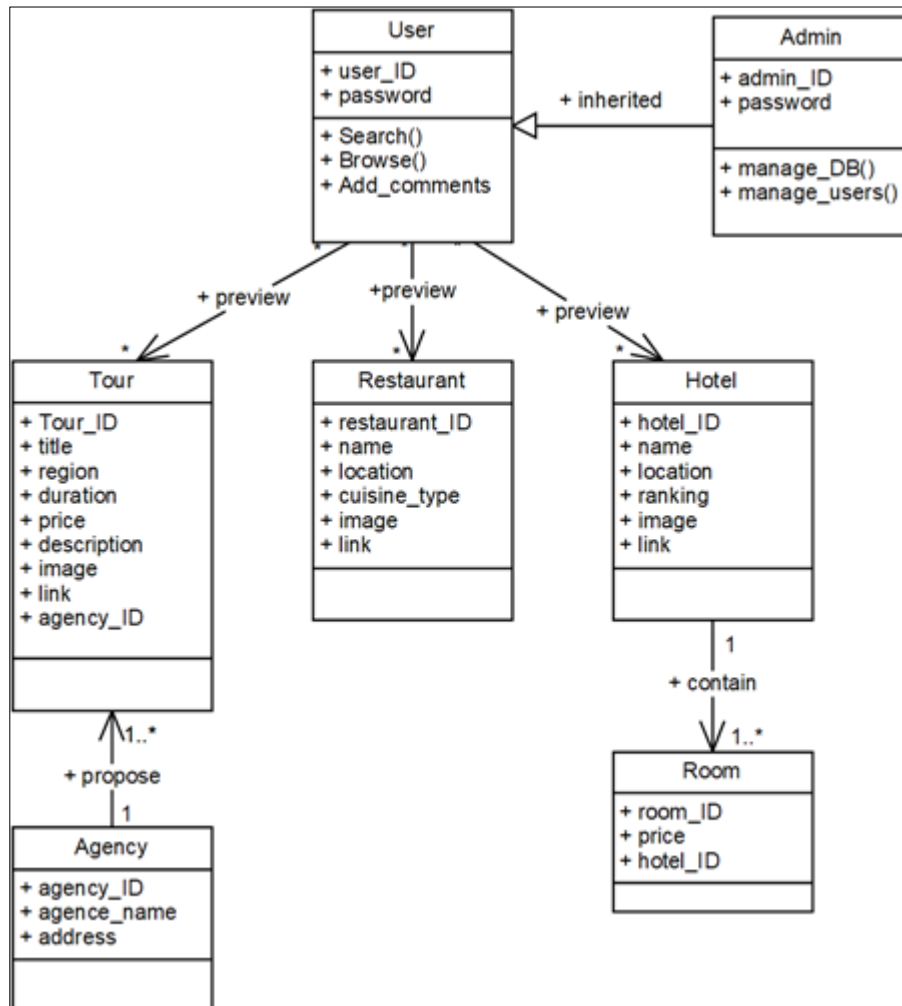
- Use Case diagram of User interaction:



**Figure 3.3** Use Case diagram of User interaction

### 3.4.2. Class diagram:

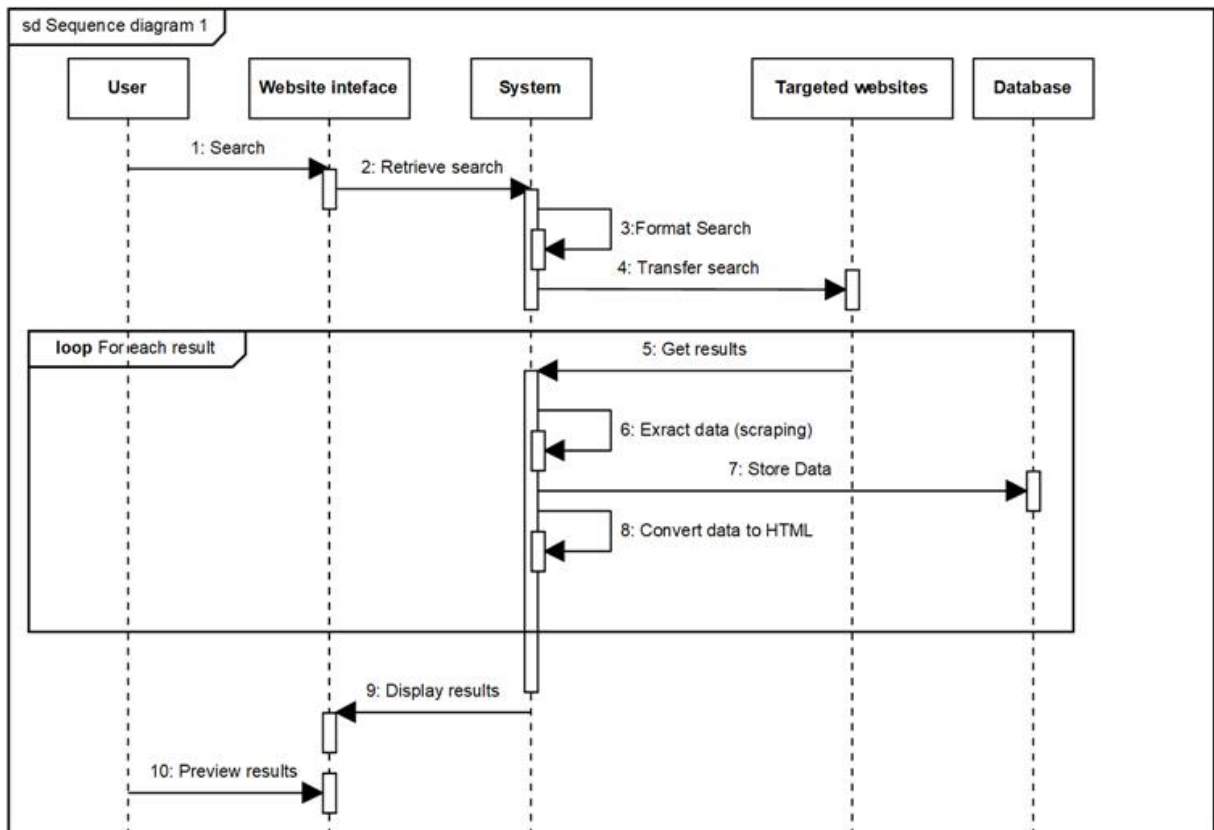
Class diagram is a static diagram. It gives an overview of a software system by displaying classes, attributes, operations, and their relationships. This diagram includes the class name, attributes, and operation in separate designated compartments.



**Figure 3.4** Class diagram

### 3.4.3. Sequence diagram:

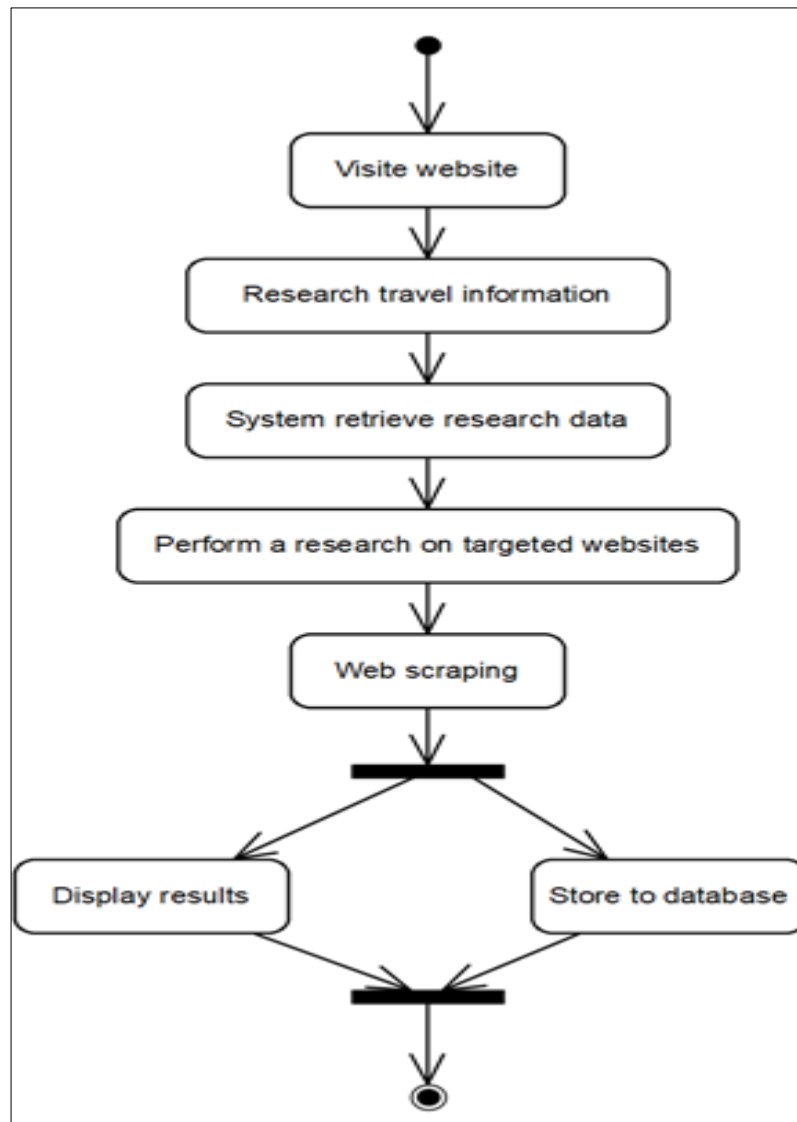
Sequence diagrams show a detailed flow for a specific use case or even just part of a specific use case. They are almost self-explanatory; they show the calls between the different objects in their sequence and can show, at a detailed level, different calls to different objects.



**Figure 3.5** Sequence diagram

#### 3.4.4. Activity diagram:

. Activity diagram models the flow of control from one activity to the other. With the help of an activity diagram, we can model sequential and concurrent activities. It visually depicts the workflow as well as what causes an event to occur.



**Figure 3.6** Activity diagram

#### **4. Conclusion:**

We have presented in this chapter our conceptual study using UML as a modeling language. In what follows, we will present the implementation phase with the different used tools and techniques.

# CHAPTER IV

## IMPLEMENTATION

### 1. Introduction:

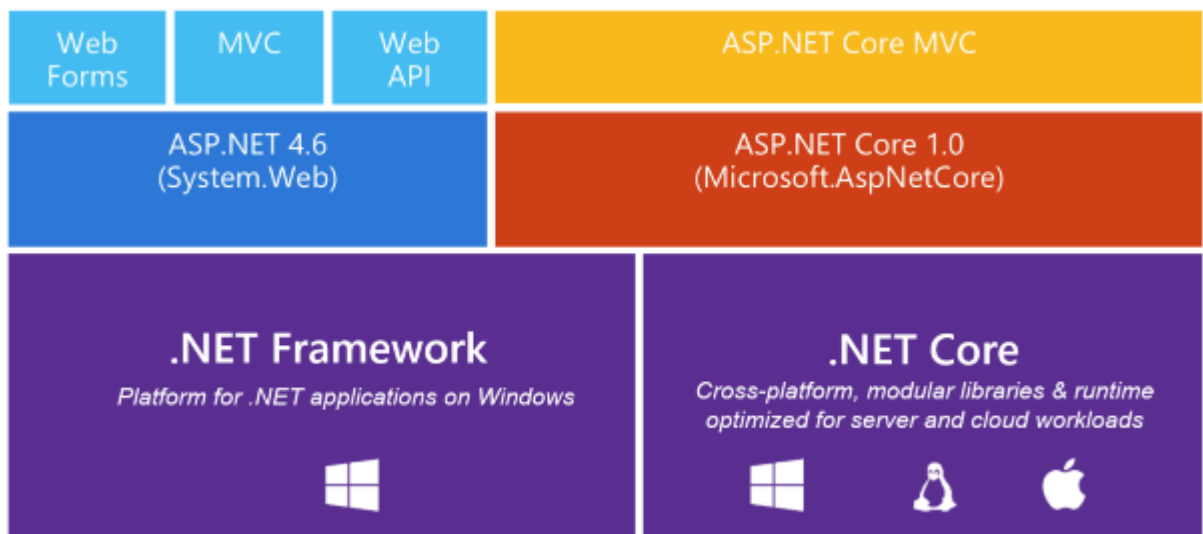
After finishing the design step of our work, we approach in this chapter the implementation phase with its different development steps. We also present the adopted technical environment and the different tools used in this phase. Finally, some interfaces of our result will be illustrated by the end of this chapter.

### 2. Presentation of the development environment:

#### 2.1. Developing languages:

##### 2.1.1. ASP.NET Core (with C# code behind):

ASP.NET Core is the open-source version of ASP.NET, which runs on macOS, Linux, and Windows. ASP.NET Core was first released in 2016 and is a re-design of earlier Windows-only versions of ASP.NET [28].



**Figure 4.1** ASP.NET vs ASP.NET Core

##### 2.1.2. CSS 3:

Cascading Style Sheets (CSS) is a stylesheet language used to describe the presentation of a document written in HTML or XML (including XML dialects such as SVG, MathML or XHTML). CSS describes how elements should be rendered on screen, on paper, in speech, or on other media [29].

##### 2.1.3. HTML 5:

Hypertext Markup Language revision 5 (HTML5) is markup language for the structure and presentation of World Wide Web contents. HTML5 supports the traditional HTML and XHTML-style syntax and other new features in its markup, New APIs, XHTML and error handling.

#### 2.1.4. Javascript:

JavaScript is a dynamic programming language that's used for web development, in web applications, for game development, and lots more. It allows you to implement dynamic features on web pages that cannot be done with only HTML and CSS [30].

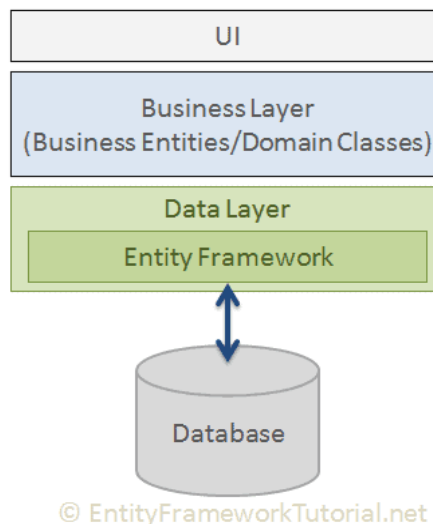
### 2.2. Deployment frameworks:

#### 2.2.1. SQL Server:

Microsoft SQL Server is a relational database management system (RDBMS) that supports a wide variety of transaction processing, business intelligence and analytics applications in corporate IT environments.

#### 2.2.2. Entity Framework:

Entity Framework is an object-relational mapper (O/RM) that enables .NET developers to work with a database using .NET objects. It eliminates the need for most of the data-access code that developers usually need to write [31]. It supports LINQ queries, change tracking, updates, and schema migrations. EF Core works with many databases, including SQL Database (on-premises and Azure), SQLite, MySQL, PostgreSQL, and Azure Cosmos DB. The following figure illustrates where the Entity Framework fits into the application:



**Figure 4.2** Entity Framework

### **2.2.3. MVC Architecture:**

The Model-View-Controller (MVC) architectural pattern separates an application into three components: Models, Views, and Controllers. This pattern helps to achieve separation of concerns. In this pattern, user requests are routed to a Controller. A Controller invokes the Model to perform user actions or retrieve data. The controller then passes this Model to a View and it is returned to the user.

### **2.2.4. IIS 10.0 Express:**

Internet Information Services is the default web server for systems Windows Servers, primarily designed to host Active Server applications pages (ASP).

### **2.2.5. Web browser:**

As it is a web application, any browser can be used, but for compatibility reasons with some JavaScript libraries used we recommend Google Chrome.

## **2.3. Libraries:**

### **2.3.1. Selenium:**

Selenium is a library and tool used for automating web browsers to do a number of tasks. One of such is web-scraping to extract useful data and information that may be otherwise unavailable.

Selenium can also be used to emulate human actions such as clicks on a link or button. To use this solution, you need to go to NuGet and install Selenium.WebDriver package and (to use Headless Chrome) Selenium.WebDriver.ChromeDriver.

## **2.4. Development environment:**

### **2.4.1. Visual Studio Code:**

Visual Studio Code is a streamlined code editor with support for development operations like debugging, task running, and version control. It aims to provide just the tools a developer needs for a quick code-build-debug cycle and leaves more complex workflows to fuller featured IDEs, such as Visual Studio IDE.

## **3. Presentation of the web application:**

Our project named “**Discover Algeria**” focuses on scraping the public tourism data from tourism and travel websites especially the Algerian ones. Even though most travel websites have provided limited data and weak services for the categories of tourism, we tried to gather tourism information from best Algerian websites in term of availability of tourism data.

In this project we targeted 3 main travel categories which are **guided tours**, **hotels**, and **restaurants** in order to meet the needs of tourists. As this project is currently solely intended for Algerian tourism aspect, therefore we used these tour and travel websites for scraping tourism data:

- <http://essendilene-voyages.com/>
- <http://www.algerie-tours.com/>
- <https://www.dunevoyages.com/>
- <http://www.abirvoyages.com/>
- <https://onatdz.com/>
- <https://timboo-voyage.com/>
- <https://touring-algeria.com/>
- <http://www.timgad-voyages.com/>
- <http://www.leadertours-dz.com/>
- <https://www.tripadvisor.com/>

So, our application “Discover Algeria” does the following tasks:

- Scrape tourism data (Tours / Hotels / Restaurants) from these websites.
- Store scraped data into SQL Server database.
- Display these data in the web pages.
- Repeat the Web scraping task every 24h (scheduled task) and update the database.

The user of “Discover Algeria” could:

- Browse tourism data (Tours / Hotels / Restaurants).
- Perform custom research according to his needs.
- Comment and add reviews.

### **3.1 Application demonstration and results:**

We present in the following some scenarios of use of our application illustrated by relative graphical interfaces:

- Main interface:

From this Home page, the user will be able to consult the top tourism data by category (Tours, Hotels, and Restaurants)

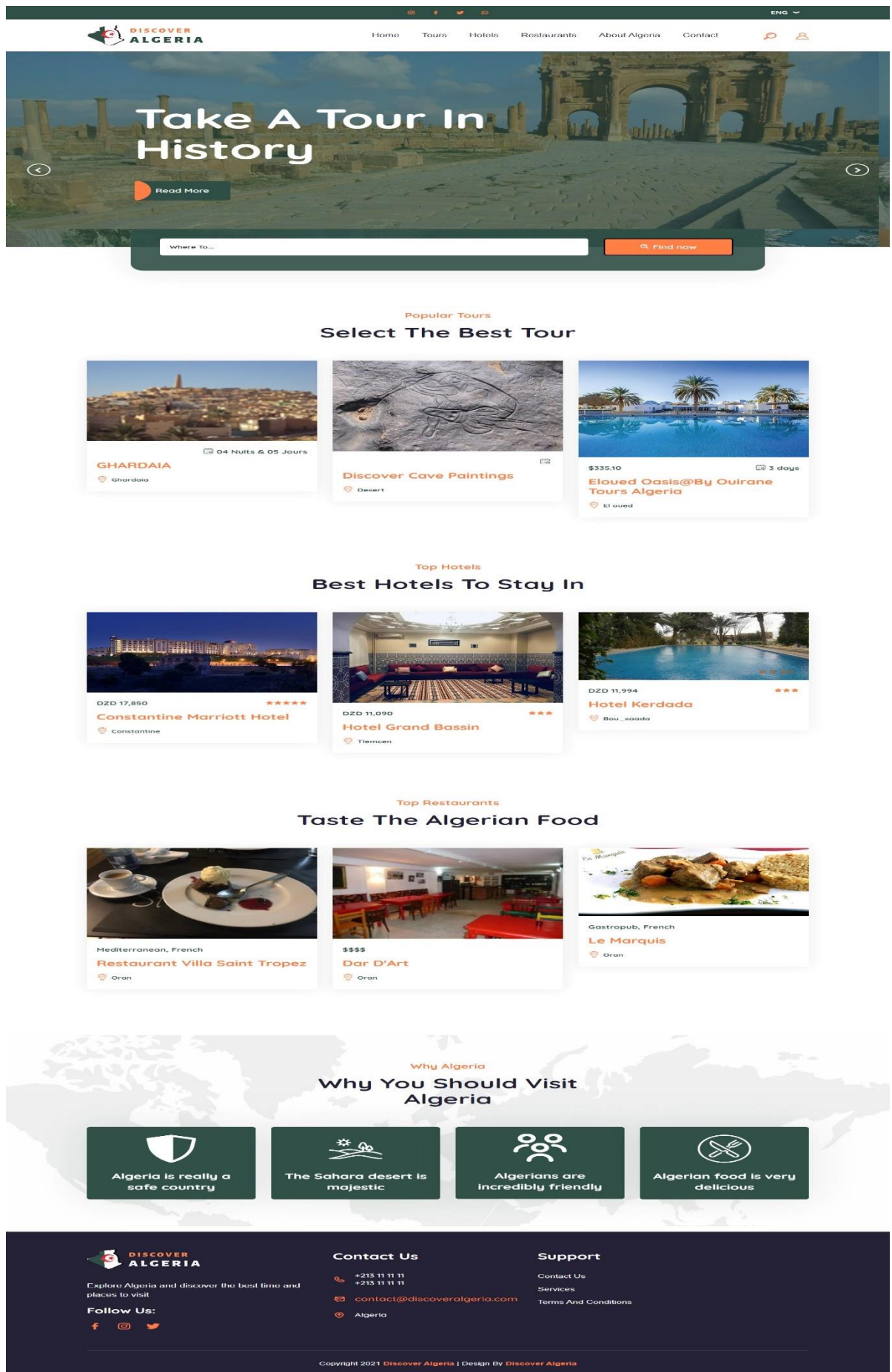


Figure 4.3 Home page

- List of all tours:

This page displays all the scraped tours offers:

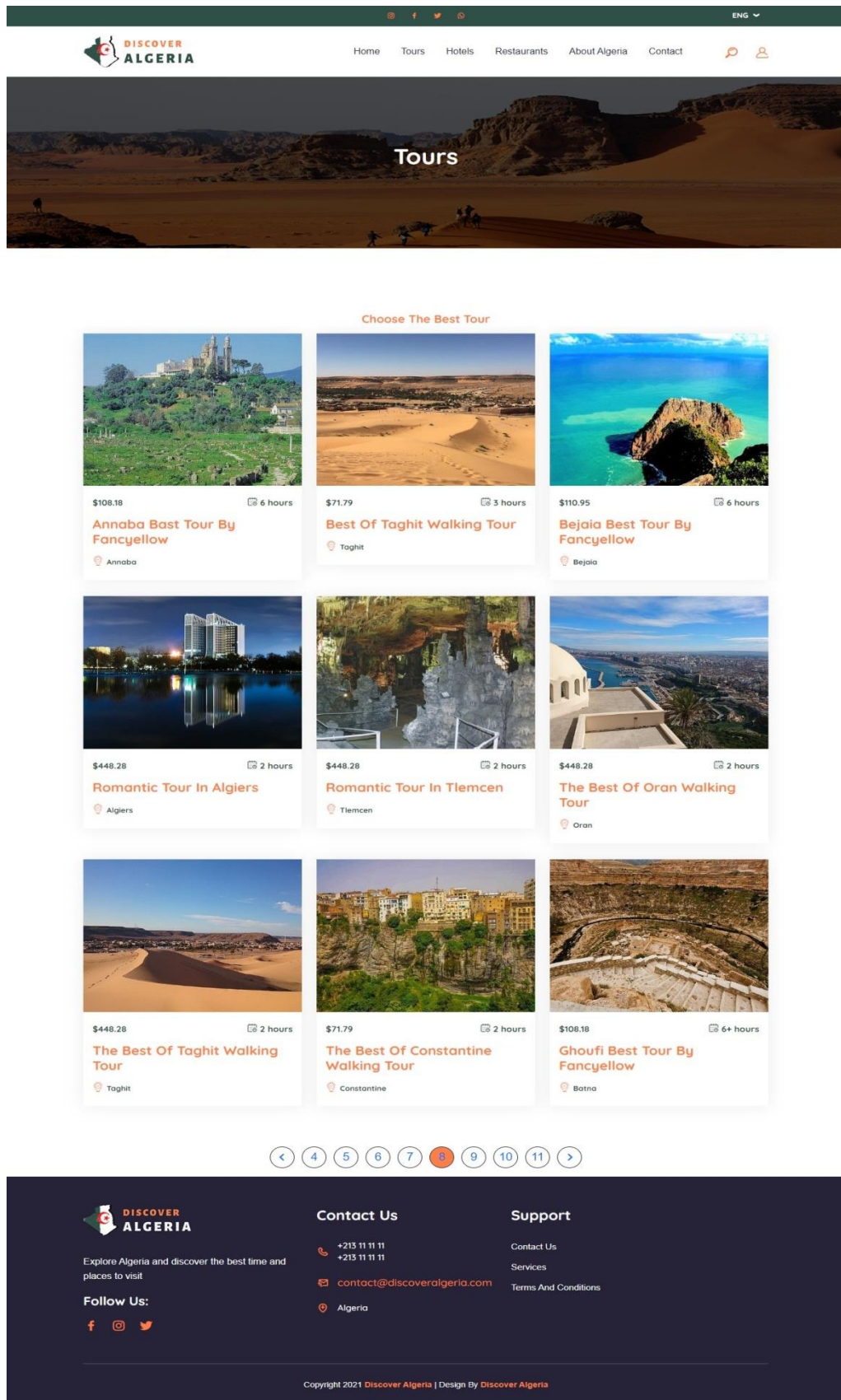


Figure 4.4 List of all tours offers

- List of all hotels:

The screenshot displays the 'Hotels' section of the Discover Algeria website. The page layout includes a dark green navigation bar with the 'DISCOVER ALGERIA' logo and menu items: Home, Tours, Hotels, Restaurants, About Algeria, and Contact. A search icon and user profile icon are also present. Below the navigation is a large hero image of a desert landscape with the word 'Hotels' centered in white. The main content area is a grid of hotel cards under the heading 'Choose The Best Hotel'. Each card features a photograph of a hotel room, the hotel name, price in DZD, a three-star rating, and the location. The hotels listed are: Hotel El Bey Constantine (Constantine, DZD 7,392), Hocine Hotel (Constantine, DZD 11,271), Hotel Suisse (Algiers, DZD 9,041), Hotel Tadj El Mouada (Setif, DZD 15,300), Hotel Le Majestic (Annaba, DZD 9,981), Hotel ABC (Algiers, DZD 15,612), Ibis Setif (Setif, DZD 12,633), Hotel Les Ambassadeurs (Oran, DZD 8,279), and Rue D'or Hotel (Setif, DZD 8,039). A pagination bar at the bottom of the grid shows page numbers 1 through 8, with page 5 selected. The footer contains the Discover Algeria logo, contact information (+213 11 11 11, contact@discoveralgeria.com, Algeria), support links (Contact Us, Services, Terms And Conditions), and social media icons (Facebook, Instagram, Twitter). Copyright information for 2021 is also present.

Figure 4.5 List of all hotels

- List of all restaurants:

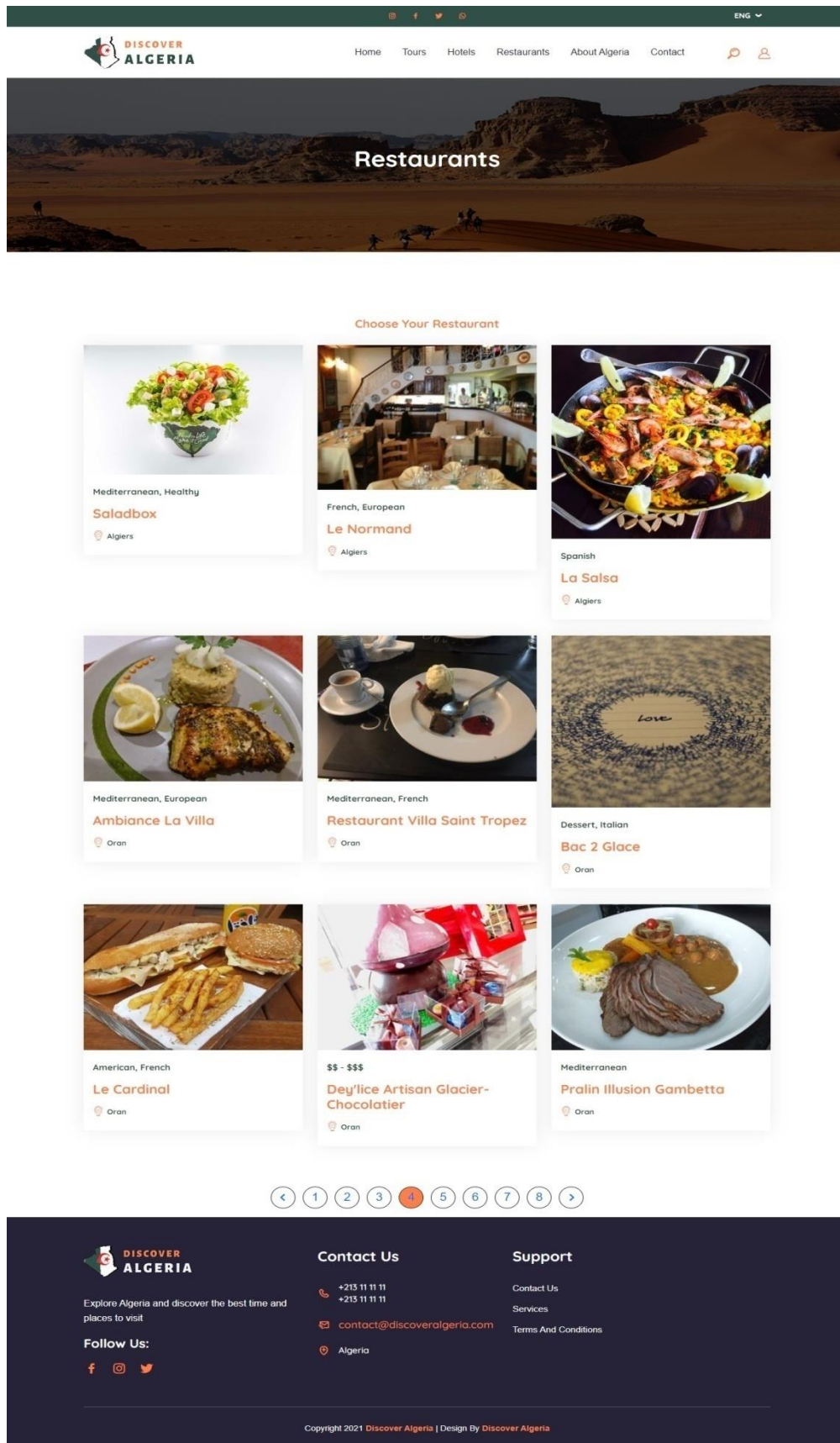


Figure 4.6 List of all restaurants

- Tour details:

This page displays the tour details if it is available, by clicking on ‘See More Details’ button it will redirect you to the original page of the tour offer

The screenshot displays the 'Discover Algeria' website interface. At the top, there is a dark green navigation bar with social media icons and a language dropdown set to 'ENG'. Below this is a white header with the 'DISCOVER ALGERIA' logo and a menu including Home, Tours, Hotels, Restaurants, About Algeria, and Contact. A search and user profile icon are also present.

The main content area features a large banner image of a desert landscape with the text 'Tour Details' overlaid. Below the banner, the 'The Best of Constantine Walking Tour' is highlighted with a large image of a cliffside town. Key details include:
 

- Location:** Constantine
- Duration:** 2 hours
- Price:** \$71.79

 Action buttons for 'Information', 'Travel Plan', and 'Our Gallery' are provided. An 'Overview' section begins with the text: 'Constantine is also recognized as Qasentina and it's the capital of Constantine Province in the northern part of Algeria...'. A 'See More Details' button is located at the bottom of the overview section.

A 'Popular Tours' sidebar on the right lists:
 

- Méharée Saharan spaces (\$109.27)
- Ghardaia & Elmnia@By Ouirane Tours-Algeria (3 days, \$252.16)
- TLEMCCEN (3 Jours Et 2 Nuits)
- Crossing Hoggar-Tassili

The footer contains the 'Discover Algeria' logo, contact information (+213 11 11 11, contact@discoveralgeria.com, Algeria), and support links (Contact Us, Services, Terms And Conditions). A copyright notice for 2021 is at the bottom.

**Figure 4.7** Tour Details

- Search page:

In this page the user can perform a search on the available attractions sites ( tours, hotels , restaurants) in a specific region:

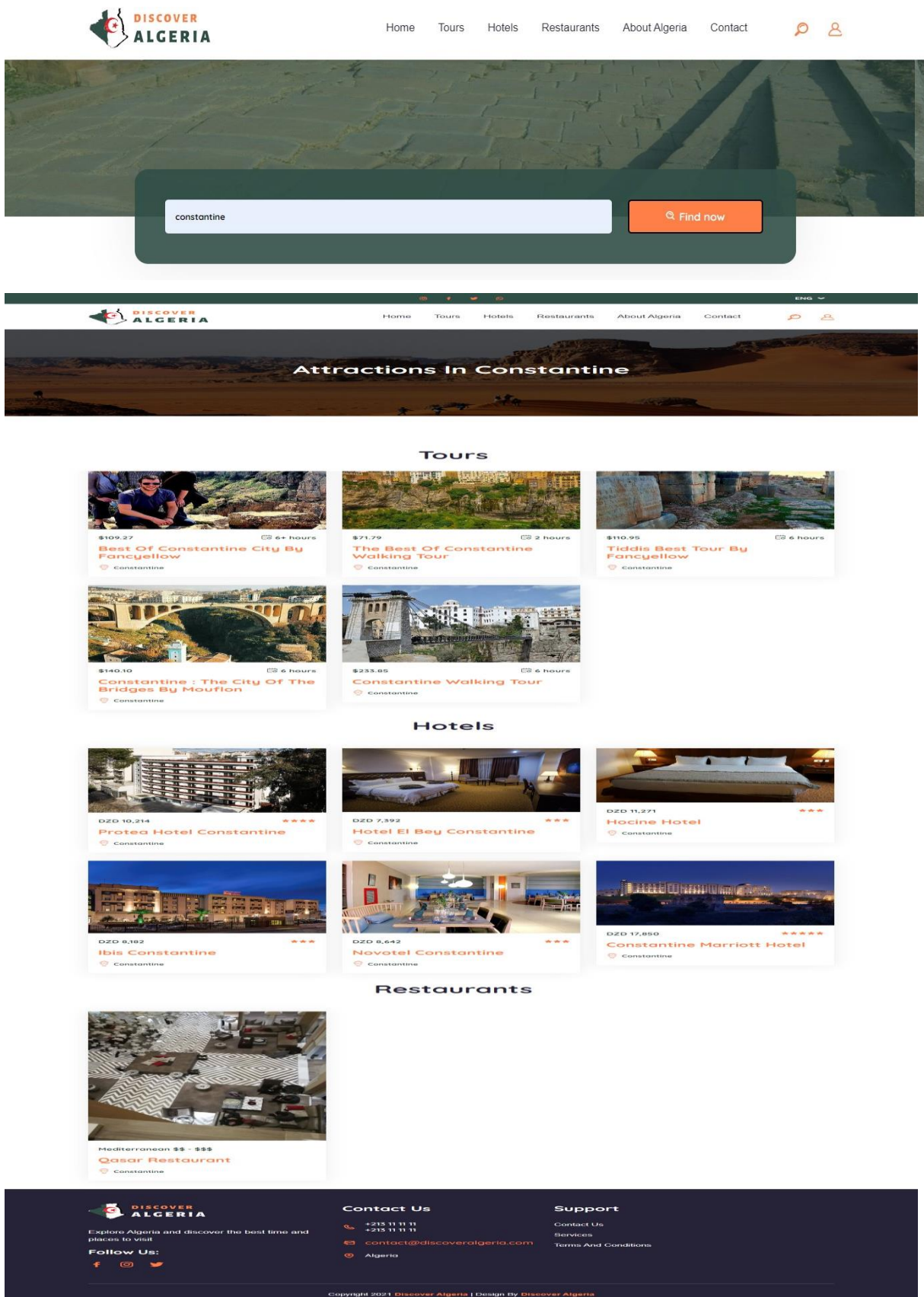


Figure 4.8 Search page

- About Algeria page:

This is a static page that contains photos and videos about Algeria, its natural potentials, its tourist sites, and its various traditions and handicrafts.

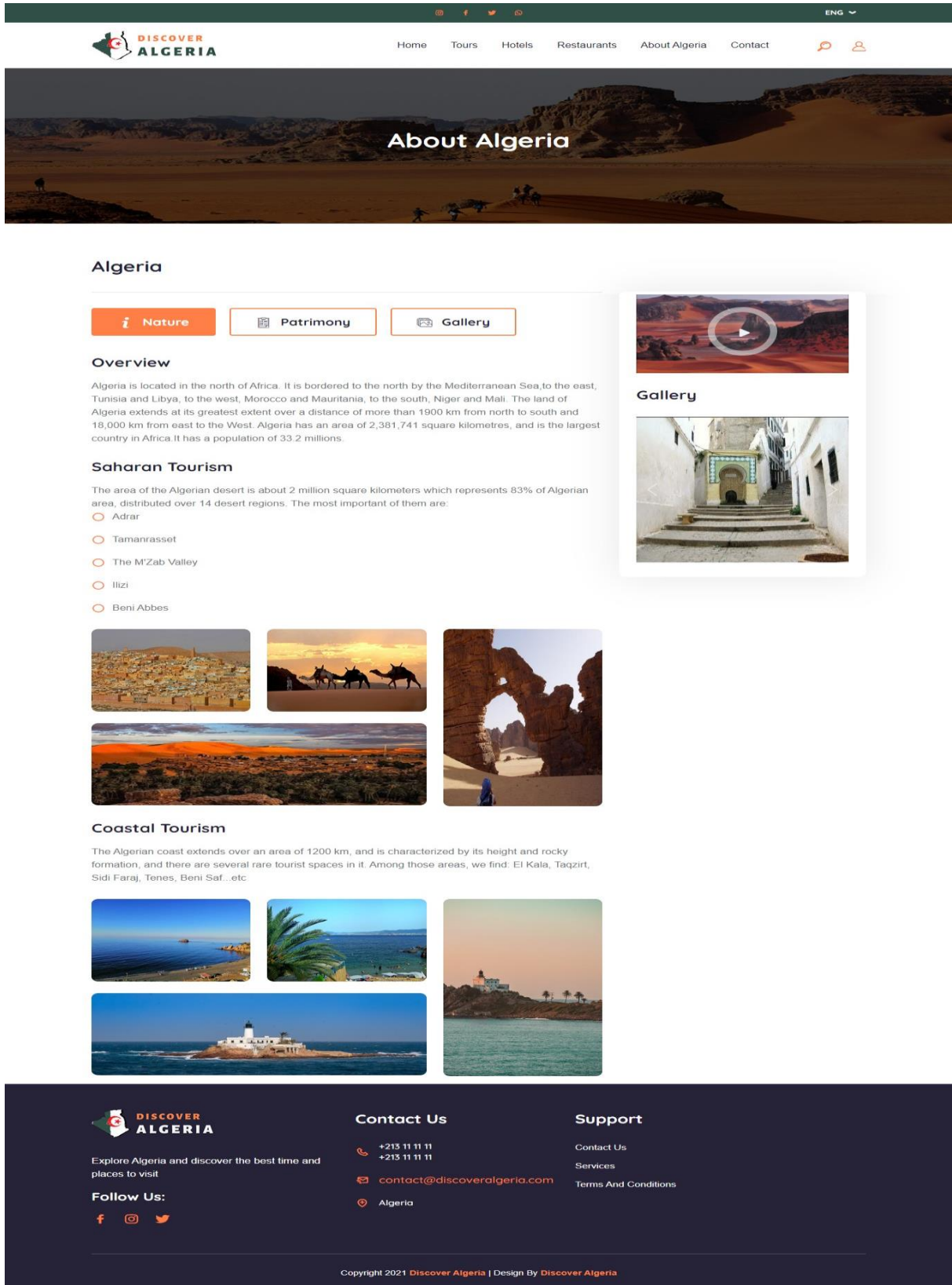


Figure 4.9 About Algeria

- Storing results in SQL Server tables:

The scraped data are stored in SQL Server datatables, for example this is Tour datatable:

tourID	tourUniqueKey	title	region	description	price	duration	image	link
24	website/Spéci...	Spécial Algérois & Kabylie	Alger/Blida/Ka...	Programme 7 J...		7 JOURS/ 6 Nuits	https://onatdz...	https://onatdz.com/boutique/sejours...
25	website/st0fil...	A la découverte zones humi...	Djanet	Voyage organis...	DZD75,000	7JOURS	data:image/svg...	https://timboo-voyage.com/voyage/...
26	website/st0fil...	A la découverte de Tiemcen	Tiemcen	Voyage organis...	DZD23,500	3JOURS	data:image/svg...	https://timboo-voyage.com/voyage/...
27	website/st0fil...	Tissemisilt, laissez la nature en...	Tissemisilt	Voyage organis...	DZD17,500	3JOURS	data:image/svg...	https://timboo-voyage.com/voyage/...
28	website/st0fil...	Djanet & Tamanrasset	Tamanrasset	Voyage organis...	DZD60,000	7JOURS	data:image/svg...	https://timboo-voyage.com/voyage/...
29	website/st0fil...		NULL		DZD75,000	7JOURS	https://timboo...	https://timboo-voyage.com/voyage/...
30	website/st0fil...	Connexion à Djanet	Djanet	Voyage organis...	DZD75,000	5JOURS	https://timboo...	https://timboo-voyage.com/voyage/...
35	website/TIM...	TIMIMOUN - CIRCUIT 04	Timimoun	Voyage organis...		9 jours et 8 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
36	website/TIM...	TIMIMOUN - CIRCUIT 03	Timimoun	Voyage organis...		10 jours et 9 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
37	website/CHAS...	CHASSE TOURISTIQUE	Oum El Bouagh...	Voyage organis...		Programmes d...	http://www.tim...	http://www.timgad-voyages.com/cir...
38	website/AURÉ...	AURÉS OASIS - CIRCUIT 01	Aurès	Voyage organis...		6 jours et 5 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
39	website/AURÉ...	AURÉS OASIS - CIRCUIT 02	Aurès	Voyage organis...		10 jours et 9 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
40	website/AURÉ...	AURÉS OASIS - CIRCUIT 03	Aurès	Voyage organis...		10 jours et 9 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
41	website/AURÉ...	AURÉS OASIS - CIRCUIT 04	Aurès	Voyage organis...		14 jours et 13 n...	http://www.tim...	http://www.timgad-voyages.com/cir...
42	website/RAND...	RANDONNÉE PÉDESTRE - CIR...	Desert	Voyage organis...		10 jours et 9 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
43	website/RAND...	RANDONNÉE PÉDESTRE - CIR...	Desert	Voyage organis...		9 jours et 8 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
44	website/RAND...	RANDONNÉE PÉDESTRE - CIR...	Desert	Voyage organis...		6 jours et 5 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
45	website/TASSI...	TASSILI DU HOGGAR - CIRCUI...	Tassili	Voyage organis...		8 jours et 7 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
46	website/TASSI...	TASSILI DU HOGGAR - CIRCUI...	Tassili	Voyage organis...		8 jours et 7 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
47	website/TASSI...	TASSILI DU HOGGAR - CIRCUI...	Tassili	Voyage organis...		8 jours et 7 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
48	website/RAND...	RANDONNÉE PÉDESTRE - CIR...	Desert	Voyage organis...		10 jours et 9 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
49	website/RAND...	RANDONNÉE PÉDESTRE - CIR...	Desert	Voyage organis...		9 jours et 8 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
50	website/TIMI...	TIMIMOUN - CIRCUIT 02	Timimoun	Voyage organis...		9 jours et 8 nuit...	http://www.tim...	http://www.timgad-voyages.com/cir...
51	website/TIMI...	TIMIMOUN - CIRCUIT 01	Timimoun	Voyage organis...		11 jours et 10 n...	http://www.tim...	http://www.timgad-voyages.com/cir...

Figure 4.10 Tour datatable

#### 4. Conclusion:

In this chapter, we have presented the different steps of the implementation phase. We have also presented the various tools and technologies used during this phase. Some result and interfaces are also presented at the end of this chapter.

## GENERAL CONCLUSION

In this dissertation, we presented a method for extracting Tourism data from a set of Algerian touristic agencies websites and use these data to build a portal website to expose different tourism offers and attractions in Algeria. The proposed system has successfully achieved its main functionality which is identifying the intended tourism data on the travel websites in order to scrape it down and store it into SQL Server database.

This project is served to build a foundation for data collection phase of Big Data analytics. With the realization of this project, a huge amount of Algerian tourism data (structured data) can be easily obtained and therefore can be used in future projects.

The main problem we faced during the realization of this work is the weakness and the lack of tourism data in Algerian travel agencies websites. For example, some websites don't provide enough information about their tourist offers like the price, the duration and even the date of the offer. Some other tourist agencies didn't update their websites for a long time ago. Other tour and travel agencies don't even have a website, they have just Facebook pages.

For future improvement, we suggest to scrape Tourism data from other sources like social media pages (Facebook, Instagram, Twitter, etc.), magazines, newspapers, brochures, personal blogs, and different tourism sources on the internet. In this case other Big Data technologies like Hadoop for example should be used to deal with these non-structured data and perform big data analytics and visualize the trends and patterns. Another suggestion, is to develop a mobile application that does the same work we did in our project "Discover Algeria".

Finally, we hope that this work represents a good introducing portal of Tourism in Algeria, and it contributes in developing this sector.

## BIBLIOGRAPHY

- [1] Tourism, <https://tourismnotes.com/travel-tourism/>, Accessed on 15/05/2022
- [2] Economic impacts of tourism, <https://tourismteacher.com/economic-impacts-of-tourism/>, Accessed on 15/05/2022.
- [3] WTTC Economic impacts report 2019
- [4] خليف مصطفى غرايبة، السياحة الصحراوية، ط 1، دار قنديل للنشر والتوزيع، عمان، 2009، ص 214
- [5] UNESCO World Heritage Center <https://whc.unesco.org/en/decisions/5276/> Accessed on 09/03/2022.
- [6] UNESCO World Heritage Center <https://whc.unesco.org/en/list/188> Accessed on 09/03/2022.
- [7] National Tourism Office of Algeria, **Algérie Balnéaire: guide touristique.**
- [8] UNESCO World Heritage Center <https://whc.unesco.org/ar/list/?iso=dz> Accessed on 16/03/2022.
- [9] Algerian Ministry of Tourism and Handicrafts
- [10] UNESCO World Heritage Center <https://whc.unesco.org/en/list/565> Accessed on 14/03/2022
- [11] الديوان الوطني للسياحة، الحمامات المعدنية متنوع خاص، مجلة الجزائر سياحة، مطبعة الديوان الجزائر، العدد 33، بدون سنة نشر، ص 14
- [12] Algerian Ministry of Tourism and Handicrafts.
- [13] Algerian National office of statistics.
- [14] Algerian Ministry of Tourism and Handicrafts
- [15] Algerian Ministry of Tourism and Handicrafts
- [16] World Economic Forum Davos, **Tourism and Travel Competitiveness Report 2019-2020**
- [17] Algerian Ministry of Tourism and Handicrafts

- [18] Big Data, what is it and why it matters, [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html), Accessed on 21/05/2022.
- [19] What are the 5Vs of the Big Data, <https://www.teradata.com/Glossary/What-are-the-5-V-s-of-Big-Data>, Accessed on 22/03/2022
- [20] What is Big Data in Tourism, <https://medium.com/dataseries/what-is-big-data-in-tourism-86b9910cc818>, Accessed on 23/05/2022.
- [21] Big Data in Travel Industry, <https://data-flair.training/blogs/big-data-in-travel-industry/>, Accessed on 22/03/2022
- [22] <https://www.edureka.co/blog/web-scraping-with-python/> Accessed on 21/05/2022.
- [23] Comparison of Web Scraping Techniques: Regular Expression, HTML DOM and Xpath, January 2019, Rohmat Ganawan and al.
- [24] Web scraping limitations, <https://www.octoparse.com/blog/web-scraping-limitations>, Accessed on 22/05/2022.
- [25] Grady Booch, James Rumbaugh, Ivar Jacobson, Le guide de l'utilisateur UML, 2000 (ISBN 2-212-09103-6)
- [26] UML Tutorials, <https://www.javatpoint.com/uml>, Accessed on 28/03/2022
- [27] Use Case Diagram, <https://www.smartdraw.com/use-case-diagram/>. Accessed on 23/03/2022
- [28] What is ASP.NET Core, <https://dotnet.microsoft.com/en-us/learn/aspnet/what-is-aspnet-core>, Accessed on 22/05/2022.
- [29] <https://medium.com/beginners-guide-to-mobile-web-development/whats-new-in-css-3-dcd7fa6122e1>, Accessed on 22/05/2022
- [30] <https://www.freecodecamp.org/news/what-is-javascript-definition-of-js/>, Accessed on 22/05/2022
- [31] Entity Framework tutorial, <https://www.entityframeworktutorial.net/what-is-entityframework.aspx> Accessed on 22/05/2022

## ملخص

يهدف هذا العمل الى إنشاء بوابة إلكترونية لمساعدة السياح الراغبين في استغلال عطلم في الجزائر . تقدم هذه البوابة بشكل عام مجال السياحة في الجزائر وتجمع كميات كبيرة من المعلومات المتعلقة بالعروض السياحية والرحلات المنظمة والفنادق والمطاعم الجزائرية. يعتمد العمل على تقنيات استخراج وتخزين بيانات الويب (كشط الويب) التي تشكل المهام الأولى التي تمارس في مجال البيانات الضخمة وهذا من مواقع وكالات السياحة الجزائرية من أجل اقتراح عروض على المقاس للسياح لمساعدتهم على وضع خططهم السياحية.

**كلمات مفتاحية:** السياحة، الجزائر، كشط الويب، البيانات الضخمة، UML

---

### **Abstract:**

The goal of this work is to create an electronic portal to assist tourists who want to exploit their holidays in Algeria. This portal presents a general overview of the tourism sector in Algeria and gathers a large amount of information related to tourist offers, organized trips, hotels and restaurants in Algeria. The work is based on web data extraction and archiving techniques (web scraping) which constitute the first tasks carried out in the Big Data discipline and this from the websites of Algerian travel agencies in order to suggest filtered offers to tourists in order to help them in their tourism projects.

**Key words:** Tourism, Algeria, Web scraping, big data, UML.

---

### **Résumé**

L'objectif de ce travail est de créer un portail web pour l'assistance des touristes qui veulent exploiter leurs vacances en Algérie. Ce portail, introduit en général le domaine de tourisme en Algérie et rassemble des grandes masses informationnelles concernant les offres touristiques, les voyages organisés, les hôtels et les restaurants en Algérie. Le travail est basé sur les techniques d'extraction et de stockage de données web (web scraping) qui constituent les premières tâches exercées dans la discipline du Big data et cela à partir des sites web des agences touristiques algériennes afin de proposer des offres filtrées aux touristes pour les aider à établir leurs plans touristiques.

**Mots clés :** Tourisme, Algérie, Grattage web, Big data, UML.