Use of Data Web on Trip Advisor as an opportunity to develop indicators of tourism development in Albania

PhD. BRUNILDA LIÇAJ

Tourism Department, Faculty of Business, University" A. Moisiu" Durres, Albania
Email: brunilda.licaj@yahoo.it

PhD. LEIDA MATJA
Tourism Department, Faculty of Business, University" A. Moisiu" Durres, Albania
Email: lmatja@hotmail.com

The paper seeks to examine the phenomenon called "Big Data" and the possibility of using data drawn from Trip Adviser, aiming at understanding and developing indicators about hotels in the Albanian territory. After a careful study of the literature, the focus is concentrated on the examination of cases in Albania aiming to be transformed into an information database structure.

The paper takes in consideration Albanian hotel evaluations during the last 5 years, aiming to derive the approach towards identifying the problems and improving the service. The paper will continue with an econometric model that will be used to evaluate Albanian accommodation in many of its categories.

Through visitor's evaluation and by using this methodology we intend to find solutions and take actions not only by private institutions but also and state institutions, in order to improve the hotelier touristic product in the future.

Keywords: web scraping, tourism, big date, trip advisor, information,

JEL Classification: Z 320 Tourism and Development

I. Tourism in Albania Overview

The tourism sector in Albania experienced a late development by having thus a marginal role in the Albanian economy. The socialist system (1944-1990) of that time based its policy on the ideology of Karl Marx's theory according to which ―the services have influence in the distribution of prosperity, but they do not have influence in GDP configuration" As in every planned economy, the central government was the authority that should care of everything. The first development traces of Tourism in Albania seemed to be at the beginning of 20th century.

Tourism is one of the most strategic sectors regarding the Albania's economic growth. Plenty of natural, cultural and historical assets, territories panoramas as well as the favorable geographic position offer endless opportunities for its development. In the recent years, tourism sector has been considered as one of the priority sectors in the country's economic development. The aim of Albania's inclusion among the favorite destinations in the Mediterranean tourism offer of the international tourism market is accompanied by a demand for increasing accommodation capacity and improvement of the service.

According to INSTAT data, the number of accommodation units for the 2015 has been 1116 units. Referring to booking.com, the number of accommodation structures registered on this website in Albania for the 2016 year is 1356. Among them, 1192 are hotels and 164 are accommodation structures of different categories.

Table1. Accommodation structures among years

Economic Activity	2013	2014	2015	2016	2016/ 2013 %
Hotels and similar structures	852	793	938	1082	27%
Accommodation structure of holidays	62	58	151	160	158%
Camping, Beach chair and caravan spaces	8	5	9	11	37%
Other accommodation structures	24	20	18	20	17%
TOTAL	**946**	**876**	**1116**	**1273**	**35%**

Source: INSTAT

Table 2.The number of beds and overnights

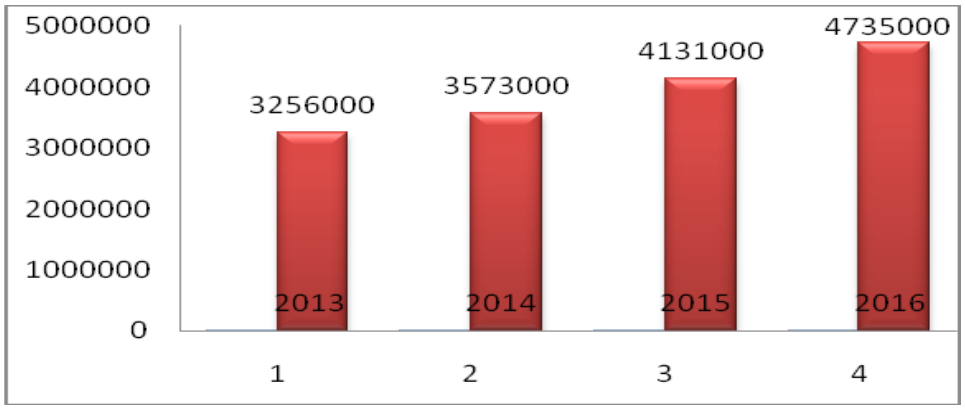
	2013	2014	2015	2015/ 2014 %	2016
No of Beds	29737	32878	41885	27%	65,000
No of overnights	62	58	151	72%	1,158,320

Source: INSTAT and evaluation at Ministry of Economic Development, Trade and Entrepreneurship, (MZHETTS)[1]

[1]Author: B.Licaj, part of the work group regarding statistic calculations

INSTAT's data also show that the number of beds for the 2015 year has been 41885 and the number of overnights has been 791000. The number of visitors for 2013-2016 has had a growth in number of a 45.4% of foreign citizens. During 2016, based on data of the Department of Border and Migration, in the General Directorate of State Police, regarding the incomings and outgoings of foreign nationals in our country, there have been entering about 473554111 foreign citizens.

Graphic 1: Foreign visitors number during 2013-2016



Source: General Directorate of State Police, Ministry of Economic Development, Trade and Entrepreneurship, MZHETTS

The biggest part has been occupied by citizens of Kosovo, Macedonia, then Greeks, Montenegrins, Italians, English, German tourists, etc. The number of visitors for personal use only in 2016 has been increased in number by 14%, while daily visitors make up a 11% of visitors in total. Meanwhile, the number of visitors visiting Albania for businesses has had an increase in number of 39% compared to the previous year.What is noticeable in this flow is the positive contribution to the real growth rate of GDP during 2016. All the important branches have led to a positive contribution, but the group "Trade, hotels and restaurants, transports" has been increased by 0.86 percentage points.

Table 3. Balance of payments regarding tourism sector

	2013	2014	2015	2016
Incomes (million euro)	1106	1283	1353	1,528
Expenditures	1113	1196	1117	1,139

Source: Bank of Albania

Table 4. Net Sales regarding 'Accommodation and Food Service'

Net Sales(Million ALL)	2013	2014	2015	2016
Accommodation and Food Service	32176	32490	35506	37,636

The figures listed above reflect the growing trend in the tourism sector, which has also been accompanied by important qualitative changes in the visitor service, an aspect of which is the accommodation sector, an important part of the study.

III. Literature Review

Big DATA application is known as one of the trends of economic developments and not only the recent years. They are necessary data when it comes to society, companies or governments that through computer techniques use can assure very necessary information without the need of extra expenditures and a relatively long time.

Big Data is today, a very popular buzzword (a Google search provides more than 62 million items containing the expression). Big data offer clear objective i.e using information generated by customers to provide key solution to target audience thus resulting in profitable value proposition. [2]

The Big Data Term is often confused with "big data analytic" giving this way the idea of something messy. In fact, it is the efficiency use of many data for a final objective. Siemens and Long (2011) define big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze[3]. Research by MGI and McKinsey's Business Technology Office examines the state of digital data and documents the significant value that can potentially be unlocked.

Russom (2011) writes that for data to be classified as big data it must possess the three Vs: Volume, Variety, and Velocity[4]. Big data is not just large, but it is varied. Google Marissa Mayer defined data by three Ss: Speed, Scale, and Sensors (where 'sensors' refers to new types of data).

The use of Big Data is often used for:

- Understanding customer tastes and preferences
- Personal customer experience
- Enhancing growth of travel companies with Big Data Analytic

RocaSaldavella and Telefonica published the first report over tourism in Spain by using BIG DATA as an example that offers endless opportunities. The research was made by crossing data through payment

[2] https://www.kelltontech.com/kellton-tech-blog/how-big-data-has-transformed-tourism-industry

[3] Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review, 46(5), 30–32,

[4] Russom, P. (2011). TDWI Best Practices Report: Big Data Analytics (Best Practices) (pp. 1–35). The Data Warehouse Institute (TDWI). Retrieved from http://tdwi.org/research/2011/09/best-practices-report-q4- big-data-analytics.aspx?tc=page0

carats and the users of mobile phones in October, 2012 in Madrid and Barcelona by assuring a multiple statistical process and by also protecting the customer's privacy. Big Data helps in creating an idea over the psychology of the tourist, in the market segmentation (relating to its expenditures), but it also helps the accommodation units to save the costs of market research. Yang art (2014) use traffic volume data of a destination marketing organization to predict hotel demand, showing an improvement in the error reduction more traditional forecasting models. Whereas Xiang (2010) analyzed the review of a certain number of tourists in hotel by accompanying it with consumer's pleasure, etc.

IV. Methodology and Objectives

This work has been based on the survey of visitors' ratings in the Albanian accommodation structures during the last 5 years and the use of their data in the BIG DATA version. Regarding the realization of the work, it was taken into consideration the review of visitors in the Trip Adviser, which at the moment of the survey reached 60,758, for all HORECA businesses.

Table5. Evaluation of quality through traveler rating and Travel Type

	Number	Traveler rating					Traveler type				
Evaluation	of review	Excellenc	Very good	Average	Poor	Terrible	Families	Couples	Solo	Business	Friends
4.6	45	18	9	1	0	0	16	8	1	0	0
4.7	21	9	0	1	2	3	5	3	2	2	2
4.6	91	10	15	4	1	8	10	16	2	7	8
4.8	112	46	9	8	0	7	22	12	2	9	7
4.3	57	9	14	4	0	0	2	13	2	8	1
4.7	81	24	10	10	1	5	17	10	3	8	9
4.7	117	51	16	4	2	1	9	20	7	15	12
4.9	276	138	56	4	8	0	11	56	36	56	36
4.7	112	75	11	3	0	0	5	16	6	48	12
4.9	211	106	16	2	0	0	12	29	15	33	25
4.7	238	165	112	19	30	2	40	34	29	47	18
4.8	452	165	112	19	5	2	29	48	20	178	17
4.5	82	30	11	3	0	1	8	14	13	13	6
4.9	163	62	33	9	1	0	9	16	5	54	14
5	92	48	11	0	0	0	2	8	4	37	6
4.8	460	165	112	19	5	2	29	48	20	178	17
4.7	580	125	141	69	12	10	31	67	37	183	40
4.8	108	28	20	8	0	1	20	25	0	1	10
4.3	111	32	20	9	2	3	29	14	2	4	3
4.2	238	58	64	13	5	2	13	63	17	8	34
4	133	43	11	8	3	4	35	30	0	2	6
4.3	82	14	22	4	3	0	11	17	4	9	8
3.8	49	8	10	2	1	1	6	6	0	2	6
4.9	76	38	9	2	1	1	10	17	2	9	8
4.6	36	1	10	14	2	3	2	10	3	4	9
3.9	21	7	8	3	2	1	9	8	0	2	3
4.2	54	20	11	0	0	1	12	12	0	7	5
4.3	38	11	0	1	1	7	5	5	2	1	2
4.6	115	18	30	8	3	1	6	23	4	13	8
3.8	61	5	8	5	2	2	0	8	1	7	5

Source: Personal data from authors

The paper was focused only on 30 hotels distributed throughout Albanian territory in the cities of Tirana, Durres, Vlora, Shkodra, Korca, Pogradec, Gjirokastra and Saranda.

The selection of hotels was not a casual one but it was taken from the rank that the visitors left in their feedback. Initially, it was found the average value of evaluation for each hotel, and then it was observed the evaluation according to the traveler rating and type for each of them.

There are 4312 reviews from different clients, families, businessmen or companies. The aim is to find a correlation between the hotel's assessment and the client typology that is accommodated into them.

Econometric results

1. Excellence Evaluation

Regression Analysis

		(Note: No intercept in the model. Interpret R^2 and R with caution.)	
R^2	0.925		
Adjusted R^2	0.910	n	30
R	0.962	k	5
Std. Error	21.652	Dep. Var.	**Excellence**

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	144,057.0067	5	28,811.4013	61.46	3.10E-13
Residual	11,719.9933	25	468.7997		
Total	155,777.0000	30			

Regression output					*confidence interval*	
variables	*coefficients*	*std. error*	*t (df=25)*	*p-value*	*95% lower*	*95% upper*
(No Intercept)						
Families	1.4715	0.4428	3.323	.0027	0.5596	2.3833
Couples	-0.4365	0.6922	-0.631	.5340	-1.8621	0.9891
Solo	2.2021	0.8479	2.597	.0155	0.4558	3.9484
Business	0.3612	0.1170	3.086	.0049	0.1202	0.6023
Friends	0.8542	1.1526	0.741	.4655	-1.5195	3.2280

Regression equation, if the type of the visitor affects the EXCELLENCE evaluation results as below:

$$\mathbf{y=1.5x_1 -0.4x_2 +2.2x_3 +0.4x_4 +0.9x_5 \text{ and } r^2 =92.5\% \ (1)}$$

This implies that if the familiar visitor changes with one unit and other types of visitor remain unchanged, the Excellence evaluation changes with 1.5. If the couple visitor changes with one unit and other types of visitors remain the same, then the EXCELLENCE evaluation changes with 0.4 on the opposite. We notice that only the familiar and business visitor is significant in the equation because they have a p-value <0.05. The other types of visitors are insignificant and finally we notice that only 92.5 % of the changes

in the excellence evaluation from a hotel to the other one come as a result of the change of the type of the visitor and 7.5% come from other external indicators.

2. 'Good' Evaluation:

Regression Analysis

		(Note: No intercept in the model. Interpret R^2 and R with caution.)	
R^2	0.945		
Adjusted R^2	0.934	n	30
R	0.972	k	5
Std. Error	12.410	Dep. Var.	**Very good**

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	66,376.8624	5	13,275.3725	86.20	6.12E-15
Residual	3,850.1376	25	154.0055		
Total	70,227.0000	30			

Regression output					*confidence interval*	
variables	*coefficients*	*std. error*	*t (df=25)*	*p-value*	*95% lower*	*95% upper*
(No Intercept)						
Families	0.4467	0.2538	1.760	.0906	-0.0759	0.9694
Couples	0.7050	0.3967	1.777	.0877	-0.1121	1.5221
Solo	1.5702	0.4860	3.231	.0034	0.5693	2.5711
Business	0.2991	0.0671	4.459	.0002	0.1609	0.4373
Friends	-1.0499	0.6606	-1.589	.1246	-2.4104	0.3107

Regression equation, if the type of the visitor affects the VERY GOOD evaluation results as below:

y=0.4x1 +0.7x2 +1.6x3 +0.3x4 -1.1x5 and r2 =94.5% (2)

The correlation shows that there exists a weak connection between variables due to R=0.201

But if we notice, only the Solo and Business visitor is significant for our equation because they have an p-value of < 0.05. The other types are insignificant. Finally, we notice that only 94.5 of the changes in VERY GOOD evaluation from a hotel into the other come as an opposite result.

3. ***Average*** Evaluation:

Regression Analysis

		(Note: No intercept in the model. Interpret R^2 and R with caution.)	
R^2	0.736		
Adjusted R^2	0.683	n	30
R	0.858	k	5
Std. Error	8.526	Dep. Var.	**Average**

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*

Regression	5,056.8801	5	1,011.3760	13.91	1.51E-06
Residual	1,817.1199	25	72.6848		
Total	6,874.0000	30			

Regression output					*confidence interval*	
variables	*coefficients*	*std. error*	*t (df=25)*	*p-value*	*95% lower*	*95% upper*
(No Intercept)						
Families	0.2769	0.1743	1.588	.1248	-0.0822	0.6360
Couples	-0.2167	0.2725	-0.795	.4341	-0.7780	0.3447
Solo	-0.0478	0.3339	-0.143	.8872	-0.7355	0.6398
Business	0.1235	0.0461	2.679	.0129	0.0285	0.2184
Friends	0.5988	0.4538	1.319	.1990	-0.3359	1.5335

Regression equation, if the type of the visitor affects the AVARAGE evaluation results as below:

y=0.3x1 -0.2x2 -0.1x3 +0.1x4 +0.6x5 and r2 =73.6% (3)

But if we notice, no one of the types of visitors is important regarding our equation because they have e p-value of > 0.05. As a result, they are all insignificant. It is also noticed that only 73.6 of the changes in the AVARAGE evaluation from a hotel to the other come as a result of the change of the type of the visitor and 26.4 % come as a result of other external indicators.

4. Poor' Evaluation:

Regression Analysis

		(Note: No intercept in the model. Interpret R² and R with caution.)	
R²	0.771		
Adjusted R²	0.725	n	30
R	0.878	k	5
Std. Error	3.372	Dep. Var.	**Poor**

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	955.8082	5	191.1616	16.82	2.70E-07
Residual	284.1918	25	11.3677		
Total	1,240.0000	30			

Regression output					*confidence interval*	
variables	*coefficients*	*std. error*	*t (df=25)*	*p-value*	*95% lower*	*95% upper*
(No Intercept)						
Families	0.2213	0.0689	3.210	.0036	0.0793	0.3633
Couples	-0.0838	0.1078	-0.778	.4441	-0.3058	0.1382
Solo	0.6778	0.1320	5.133	2.64E-05	0.4059	0.9497
Business	-0.0394	0.0182	-2.162	.0404	-0.0769	-0.0019
Friends	-0.1941	0.1795	-1.081	.2898	-0.5637	0.1755

Regression equation, if the type of the visitor affects the POOR evaluation results as below:

y=0.2x1 -0.1x2 +0.7x3 -0.03x4 -0.2x5 and r2 =77.1%

At the equation, we notice that only familiar and solo visitors are important for our equation because they have a p-value of <0.05. The other types of visitors are insignificant. Only 77.1% of the changes in the POOR evaluation from a hotel to the other come as a result of the change of the type of the visitor and 22.9 % come as a result of other external factors.

Conclusions.

Tourism and technology development are an already known binom. In the evaluation context of the BIG DATA, scientific researchers can make various and specific researches in different branches by bringing this way in real time many information according to what marketers are looking for in the time when trends go on with big steps.

Just some specific programs are enough in order to accumulate data not only from Trip Adviser but from each specific web to do researches which in terrain would be difficult to take place.

This study could be a handbook to help managers to lead and modify the current work practices so that they can measure and then increase customer satisfaction based on the control, assessment and improvement of the service quality.

References

1. Buhalis, D., & Foerste, M. (2015). SoCoMo marketing for travel and tourism: Empowering co-creation of value. *Journal of Destination Marketing & Management*, 4(3), 151-161.
2. How big data has transformed tourism industry https://www.kelltontech.com/kellton-tech-blog/how-big-data-has-transformed-tourism-industry
3. Big data aplicado al sector del turismo un caso practico https://www.telefonica.com/en/web/public-policy/-/big-data-aplicado-al-sector-del-turismo-un-caso-practico
4. Lam, C., & McKercher, B.(2013). The tourism data gap: The utility of official tourism information for the hospitality and tourism industry/ Tourism Management Perspectives, 6, 82-94
5. Russom, P. (2011). TDWI Best Practices Report: Big Data Analytics (Best Practices) (pp. 1–35). The Data Warehouse Institute (TDWI).
6. Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE Review, 46(5), 30–32,

7. Yang, Y., Pan, B., & Song, H.(2014). Predicting hotel demand using destination marketing organization's web traffic data. Journal of Travel Research, 53(4), 433-477