Group Project – Sentiment Analysis on Paris & Dubai using Rapid Miner on Twitter Data

Course Code & Name: ABM503D Web & Social Media Analytics
Submitted To: Dr. Nor Hasliza Md Saad

Submitted by "The Digital Knights"

Group Member Name	Metric Number	Student E-mail
Kareem Mustafa Mohamed Eldemerdash	*P-EM0112/22*	kareemmostafa15@student.usm.my
Fatma Ahmad Bakr Bakr Darwech	*P-EM0104/22*	fatma.bakr@student.usm.my
Ahmed Mohamed Ahmed Serwah	*P-EM0111/22*	ahmedserwah@student.usm.my
Ahmed Moharram Ibrahim	*P-EM0115/22*	ahmedmoharm1997@student.usm.my

Table of Contents

LIST OF FIGURES **2**

LIST OF TABLES **3**

ABSTRACT **3**

CHAPTER 1: INTRODUCTION **3**

1.1 Introduction: Data and Social Media Analysis 3
1.2 Research Statement 4
1.3 Research Objective 5
1.4 Research Questions 5
1.5 Significance of Study 5
1.6 Organization Report 5

CHAPTER 2: LITERATURE REVIEW **6**

2.1 Introduction 6
2.2 Sentiment Analysis 7
2.3 Topic Modelling 7

CHAPTER 3: METHODOLOGY **8**

3.1 Stage 1: Opinion Retrieval 9
3.2 Stage 2: Data pre-processing 10
3.3 Stage 3: Sentiment Analysis & Topic Modelling 11
3.3.1 Sentiment Analysis 11
3.3.2 Topic Modelling 14

CHAPTER 4: RESULTS AND DISCUSSION **15**

4.1 RQ1: What is the sentiment analysis of Dubai & Paris using Twitter data? 15
4.2 RQ2: What are the keywords used by the users to describe their travel experience and the topics discussed when it comes to these two cities? 16

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS **19**

5.1 Conclusion 19
5.2 Recommendations Based on the Findings 25
5.3 Limitations 25
5.4 Recommendations for Future Research 25

REFERENCES **26**

List of Figures

Figure 1 Architecture of Opinion Mining 9
Figure 2 Get Twitter Tweets Containing Keywords (Data) 9
Figure 3 Cleaning Data 10
Figure 4 Sentiment Analysis 11
Figure 5 Part 1 of Word Cloud Process – Frequent Words 12
Figure 6 Part 2 of Word Cloud Process - Top 20 Words 12
Figure 7 Part 1 of Word Cloud Process – Frequent Positive Words 13
Figure 8 Part 2 of Word Cloud Process - Positive Words Filter 13
Figure 9 Part 3 of Word Cloud Process - Top 20 Positive Words 13

Figure 10 The Process of Topic Modelling using the LDA Method 14
Figure 11 Dubai Sentiment Analysis 15
Figure 12 Paris Sentiment Analysis 15
Figure 13 Dubai Word Cloud 19
Figure 14 Dubai Positive Word Cloud 20
Figure 15 Dubai Negative Word Cloud 20
Figure 16 Dubai Neutral Word Cloud 21
Figure 17 Paris Word Cloud 21
Figure 18 Paris Positive Word Cloud 22
Figure 19 Paris Negative Word Cloud 22
Figure 20 Paris Neutral Word Cloud 23
Figure 21 Topic Modelling Visualization – Dubai 23
Figure 22 Representative Data for Keywords - Dubai 24
Figure 23 Topic Modelling Visualization – Paris 24
Figure 24 Representative Data for Keywords - Paris 25

List of Tables

Table 1 Top 20 Frequent Words 16
Table 2 Top 20 Positive Frequent Words 17
Table 3 Top 20 Negative Frequent Words 17
Table 4 Top 20 Neutral Frequent Words 18
Table 5 Dubai Topic Modelling 18
Table 6 Paris Topics Modelling 19

Abstract

Digital transformation changes the world and how people communicate and express there feelings. From this preceptive, sentiment analysis of unstructured text is rapidly increasing its importance. Many companies use it to analyze and attract customers, enter new markets and acquiring a bigger market share. This paper answers the question of what is the best city to travel to between Paris and Dubai based on people's opinion shared on Twitter. The data we have is scrubbed using RapidMiner of a total of 12000 rows. We used the Vader Approach through the data cleaning process applied before the analysis phase. We also used several RapidMiner analysis tools. This study can help any travel agency to understand what the twitter users' opinions about Paris and Dubai.

Chapter 1: Introduction

1.1 Introduction: Data and Social Media Analysis

Digital transformation is changing the world and everything changes with it. One of the significant changes is how we interpret opinions of unstructured text. By a single click nowadays we can express our feeling whether we are feeling angry, happy, or

sad. Facebook have changed the social life perspective. There are 17 billion location-tagged Facebook posts today. And regarding this huge amount of data; one captivating quote sums it up:

"Information is the oil of the 21st century, and analytics is the combustion engine." -Peter Syndergaard

Data can give meaningful insights that facilitate the decision-making process by answering several questions and through utilizing social media in hand, it's even more powerful. For instance, Twitter hashtags can provide insights about what twitter users' think about a specific product, person, topic, or service. And so, companies can get a sense about what their users think about the product or brand and how they want the product to be enhanced so they would have a better experience. In this paper, we are analyzing the users' opinions from Twitter towards specific cities Dubai, Paris for a travel agency in order to authorize an agreement with one of the two cities' tourism deputies. The sentiment analysis will help getting insights in how the Twitter users felt towards the two cities.The data will yields information towards a positive, negative or neutral scale of the opinions of about the two cities. The research will concentrate on sentiment analysis using the Vader approach to analyze the customer opinions towards the two cities; Dubai and Paris. These two cities are chosen because of their traction and the contrast of authenticity and history versus modernized and evolving. The main aim for this paper is to provide the travel agencies a comprehensive view about its customer behavior and opinions. We analyzed 6000 tweets for Dubai and the same amount for Paris. After the collection phase, we proceeded to clean the data to reduce any inconsistency in the data in terms of null/missing values and then we advanced to classify the tweets on a positive, negative or neutral scale. The sentiment analysis will treat the tweet that implicitly shows negative or positive feeling. However, neutral will be any tweet that don't shows positive or negative feeling. In this paper the data collected hinged on specific keywords to target the right customer segments and ensure the integrity of our research..

1.2 Research Statement

Sentiment analysis is mainly used by companies and individuals to understand the customers opinions about their products, services using data collected from various social media platform like Instagram, Facebook and twitter. These platforms allow users to post and comment their opinions. This in return can help companies and individuals to understand their weaknesses, develop better services and products.

This paper is to provide the **Tourism** industry a more comprehensive view about the opinions of their consumers to analyze customers preferring **Paris and Dubai** as a destination.

1.3 Research Objective

The study uses twitter as data source and RapidMiner as a software analysis tools and implement the data mining process to answer the following questions:

Q1: What is the better tourism destination; Paris or Dubai?

Q2: What are the keywords used by the users to describe their travel experience negatively or positively in Paris or Dubai?

1.4 Research Questions

I. What is the sentiment analysis of Dubai & Paris using Twitter data?
II. What are the keywords used by the users to describe their travel experience and the topics discussed when it comes to these two cities?

1.5 Significance of Study

The significance of the study is to get a better understanding of Twitter users when it comes to these two travel destinations and analyze their opinions. The aim of this research is to deliver insightful information about comparing between two of the most attractive cities to the tourism sector. This paper will help any company to understand customer opinions about these cities and build strong customer sentiment analysis databases.

1.6 Organization Report

This paper is mainly divided into five main chapters: the introduction, literature review, methodology, analysis and findings, and conclusion. The introduction briefly introduces the topic discussed, the objectives behind our research and our research questions. Chapter 2 gives a background about the industry and the topic discussed as well as sharing information about the techniques used. Chapter 3 identify the research methodology that was used to carry out this project and identify the process of opinion mining and the operators used in RapidMiner. The results and discussion in Chapter 4 are a summary of the findings from the opinion mining process. Finally, Chapter 5 discusses the results found in our research and our conclusion about the insights found in our analysis.

Chapter 2: Literature Review

2.1 Introduction

Tourism is an industry that transcends the traditional economic sectors. Economic, social, cultural, and environmental contributions are necessary. The issue with calling tourism an "industry" is that, unlike agriculture (tons of wheat) or drinks, it does not have the typical formal production function or an output that can be physically measured (liters of whisky). There isn't a universal design that encompasses every nation's sector. For instance, restaurants and shopping centers are popular tourist destinations in France and Italy but not in Russia. Even the foundational elements of the tourism sector, such lodging and transportation, can vary between nations. Tourism was a relatively fragmented sector until the 1950s; hotels, transportation providers, travel agencies, and tour guides all tended to operate separately from one another. Hotels made most of their money by renting out rooms. Seat sales were a part of the operations for both airlines and railroads. Of course, travel agencies sold vacations and travel, but in each case, they tended to run their companies relatively independently. From the middle of the 1950s onward, especially in the UK, the rise of tour operators started to transform the industry's fundamental business practices into more integrated ones. For instance, hotels started to recognize that customers wanted more than just a place to stay and started catering to their needs. Hotels started to build retail malls and later added secretarial services to enhance the amount of money visitors spent while staying at the hotel. Transport providers, especially those in the airline industry, viewed the sale of transportation services as being a crucial component of a much larger demand. Travelers might purchase insurance and make hotel reservations through airlines. By the 1980s, numerous airlines provided comprehensive travel services, such as booking vacations, getting medical care, renting a car, etc. From 1950 onward, several variables, including an increase in leisure time availability, a rise in paid holidays, the introduction of package tours, and the advancement of air travel, all worked together to create a considerably larger potential market for vacationing. In terms of socioeconomic groupings, this market was distinct from the period before 1950. Without the travel industry's growing specialization, long-haul travel would not have been able to develop as it has. Particularly with the advent of tour operators, there was a tendency to concentrate business either on the short-haul or long-haul markets. Within this trend, there was also a growing number of specialized tour

operators, who were typically trying to meet the need for tourists with greater per capita spending who wanted to see far-off places or partake in unique experiences but fewer tourists overall. One aspect that has contributed to changing the structure and nature of the travel business is specialization within the industry. (Lickorish & Jenkins , 1997) And so, we can expect people from all over the world travelling to destinations they wouldn't have thought would be possible. And with the data made accessible to travelers all around the world through the internet; they could get an insight about their destination. We will discuss more about the technologies used in our research and how it can help travel agencies to get a better understanding of their target customers.

2.2 Sentiment Analysis

There is a lot of study on user opinion data sentiment analysis, which mostly determines the polarities of user evaluations. These studies frequently use one of the three levels for sentiment analysis: the document level, phrase level, or attribute level. The literature review conducted in relation to sentiment analysis points to two categories of approaches, including machine learning and semantic orientation. Additionally, the nature language processing techniques (NLP) are utilized in this field, particularly in the detection of document sentiment. As a result, modern sentiment analysis is a field at the nexus of natural language processing and information retrieval. As such, it has some aspects in common with other activities including information extraction and text mining, computational linguistics, psychology, and predicative analysis. (Vinodhini & Chandrasekaran, 2012)

2.3 Topic Modelling

Social media opinion mining has grown to be a significant area of research. Topic modelling is a technique that is used in addition to sentiment analysis. One of the text analytics techniques, topic modeling's main objective is to extract information from a significant amount of text data. It will use an unsupervised machine learning technique to reflect the data in the set. The two major goals of topic modelling are to identify hidden themes in text data by grouping related words into appropriate clusters and to categorize the content into the identified theme. By assembling similar phrases into topics and recognizing social media trends, topic modelling has the effect of increasing classification. It frequently makes use of similarity measurements in recommender systems. Two fundamental topic models exist. Latent Dirichlet Allocation (LDA) and Probabilistic Latent Semantic Indexing

(PLSI) (LDA). PLSI is a method that uses matrix decomposition to identify the latent topic. LDA, meanwhile, employs Dirichlet over the latent topic and is a generative probabilistic corpus model. The fundamental idea is that documents are modelled as randomized mixes of latent themes, where each topic is characterized by a distribution across words. The PLSI has lower accuracy than the LDA but is simpler to train. The subject of modelling using the LDA technique has been extensively studied. A study by uses LDA to analyze the topic brought up during the Brexit debate on Twitter and discover a link between Brexit emotion and the value of the British pound sterling. The topic modelling approach is used by to mine Twitter to understand consumer sentiments toward vaccination. When Covid19 was investigated, the LDA approach was employed to interpret how the public perceived the non-pharmaceutical intervention (NPI) method. The results of this study defined problems associated with keywords in six nations and helped the government put NPI strategies and decision-making into practice. In addition, LDA-based topic modelling has been used to find Twitter product prospects. The subject of modelling using the LDA technique has been extensively studied. To determine the relationship between Brexit mood and the value of the British pound sterling, a study by employed LDA to look at the issue discussed during the Brexit on Twitter. Similar thinking is put into practice by where Twitter is mined using the topic modelling method to comprehend consumer sentiments around vaccination. When exploring the non-pharmaceutical intervention (NPI) method for Covid19, the LDA technique is employed to interpret public perception. The results of this study defined problems associated with keywords in six nations and helped the government put NPI strategies and decision-making into practice. In addition, LDA-based topic modelling has been utilized to quantify and find product potential among Twitter users. (Deraman, et al., 2021)

Chapter 3: Methodology

This chapter will illustrate the research methodology that was used to carry out this project and identify the process of opinion mining and the operators used in RapidMiner. There are four stages of the opinion mining process which are opinion retrieval, data pre-processing, topic modeling and sentiment analysis, and opinion summarization, as shown in Figure 1.

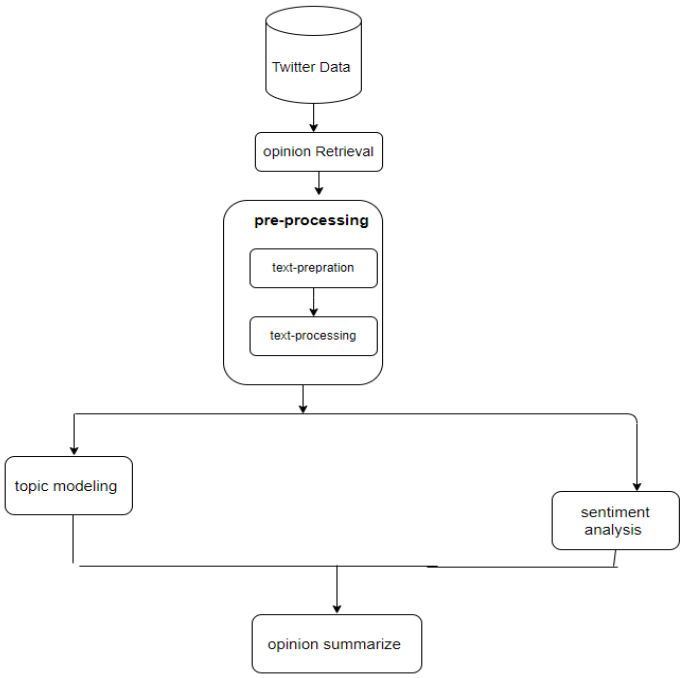

Figure 1 Architecture of Opinion Mining

3.1 Stage 1: Opinion Retrieval

The first step in the methodology of this project is to collect the social media data related to the topic.There are a few ways for collecting large scale social media data that can be used, and they include web scraping tools and APIs provided by various services. In this project, a scraper was used to extract tweets from Twitter related to the degree of acceptability and users' opinions on Dubai and Paris. The keywords explored were "Dubai Zoo", "The Dubai Fountain", "Jumeriah Mosque", "Burj Khalifa ", "Eiffel Tower", "Louvre Museum", "Arc De Triomphe ", and "Palace of Versailles". All of the collected data is saved in an XLSX file to be processed
We Perform the same process for each city. Figure 2 demonstrates how we imported Twitter users' tweets containing keywords and store them in excel file.

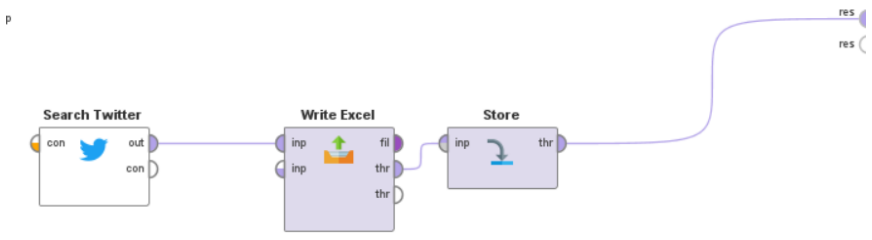

Figure 2 Get Twitter Tweets Containing Keywords (Data)

3.2 Stage 2: Data pre-processing

RapidMiner has been used to execute the data preprocessing for the raw data that have been collected from Twitter. At this stage, the raw data collected from Twitter will go through two processes which are text preparation and text pre-processing. Text preparation is the process of cleaning and transforming the raw data before processing and analysis. Figure 3 reveals the detailed process done through the various operators in RapidMiner. The standard operators such as, "Select Attributes", "subprocess " "trim ", "remove duplicate" , "replace the missing value" and "Process Documents from Data" were used in this process. The "Select Attributes" operator is used to select the attributes from the dataset and provides different filter types to make Attribute selection easy. "subprocess" operator introduces a process within a process. Whenever a Subprocess operator is reached during process execution, "Trim" operator removes leading and trailing spaces from the values of the selected nominal attributes. "remove duplicate" removes duplicate examples from an Example Set by comparing all examples with each other on the basis of the specified attributes. "replace the missing value" replaces missing values in Examples of selected Attributes by a specified replacement.

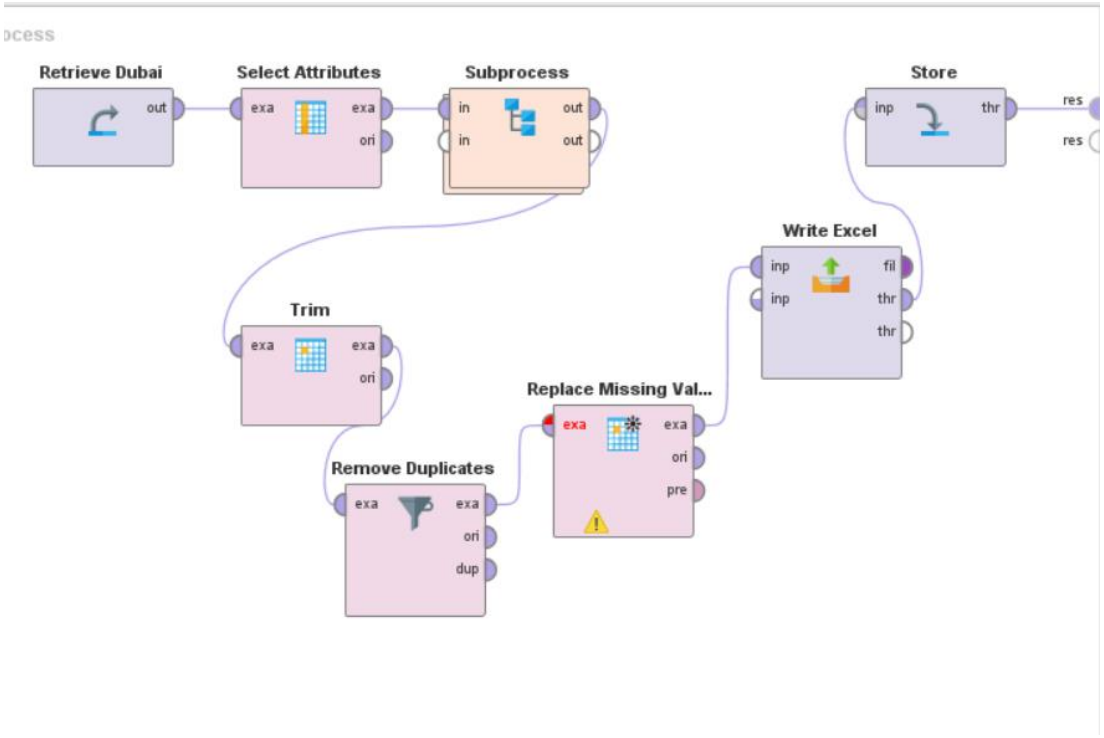

Figure 3 Cleaning Data

3.3 Stage 3: Sentiment Analysis & Topic Modelling

3.3.1 Sentiment Analysis

Figure 5 illustrates the process of sentiment analysis by using RapidMiner. The common operator tools such as “nominal to text ”, “generate attribute” are used before extracting the sentiment from the comments of the social media users. “Nominal to Text” operator changes the type of selected nominal attributes to text. It also maps all values of these attributes to corresponding string values. “Extract Sentiment” operator creates a sentiment score by applying either open-source sentiment dictionaries or proprietary API methods on an existing text attribute. There are options to expose additional results depending on the method chosen. “Generate Attributes” operator constructs new user-defined attributes using mathematical expressions. The operator is used to classify each of the text into positive, neutral and negative class by adding the “sentiment” attribute and using the function expression “if(Score > 0 , “positive” , if(Score < 0 , “negative” , if(Score == 0 , “neutral” , “ ”)))” to classify the text. After that, the result is saved in an excel file to extract the word from the comments for each of the classes by using the “Write Excel” operator for doing word cloud.

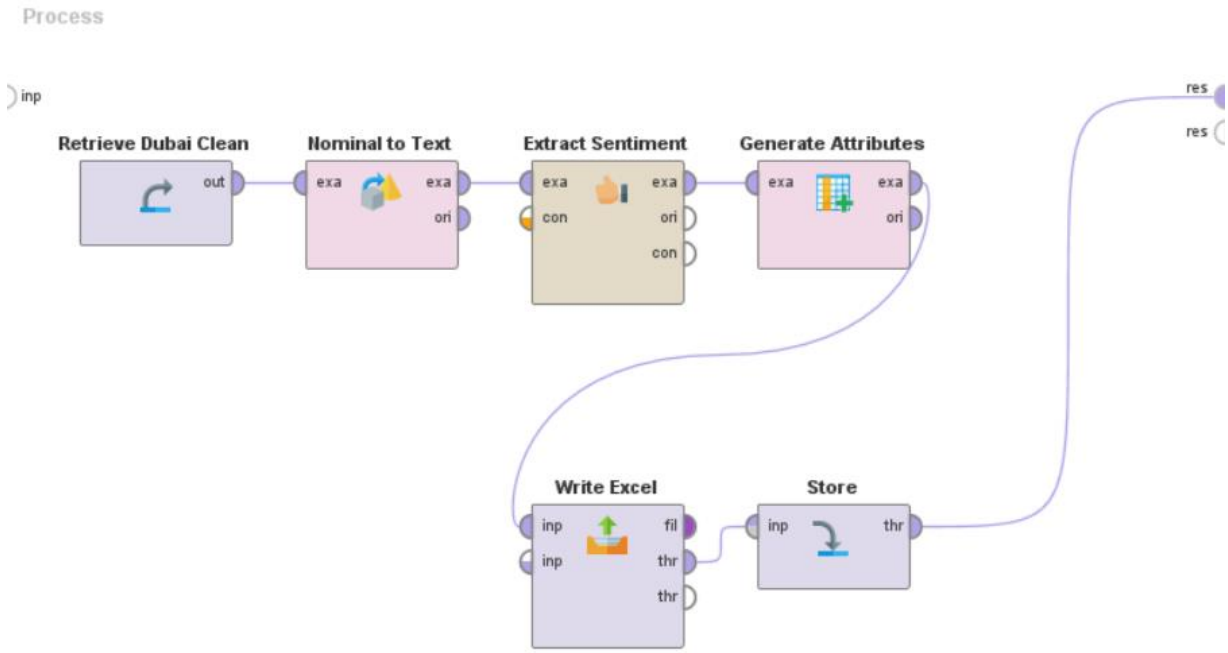

Figure 4 Sentiment Analysis

The word cloud process is shown in Figure 5 and Figure 6. standard operator tools such as “Nominal to Text” and “Process Documents from Data” which correspond

to text pre-processing and then saved in excel format for further process. After extracting the word from each comment and saved in an excel file, the "Sort" operator is used to sort each class by descending since this project want to find the most number of times the word that appeared for each class and the "Filter Example Range" operator is used to set the top 20 words. Then we filter the results according to their polarity as we explain in figure 5 through 9.

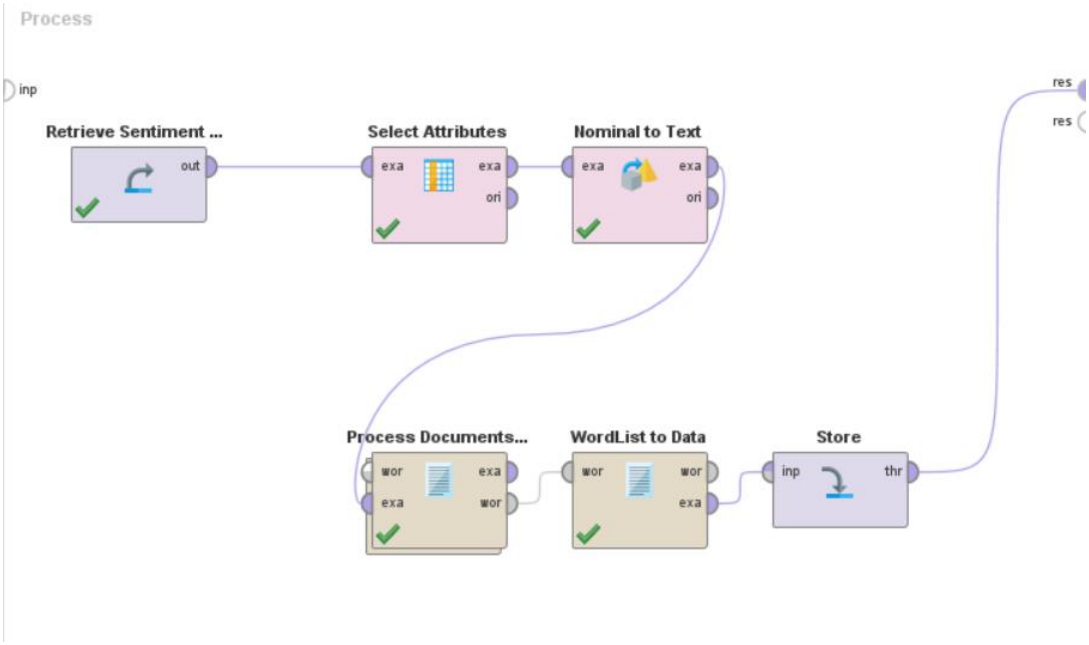

Figure 5 Part 1 of Word Cloud Process – Frequent Words

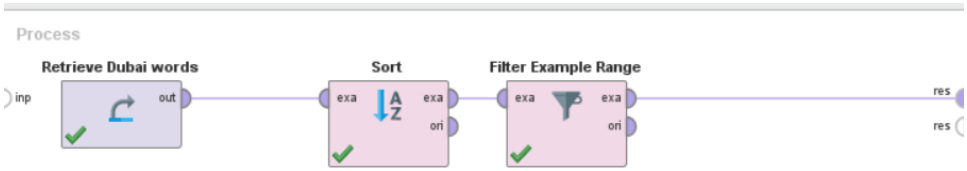

Figure 6 Part 2 of Word Cloud Process - Top 20 Words

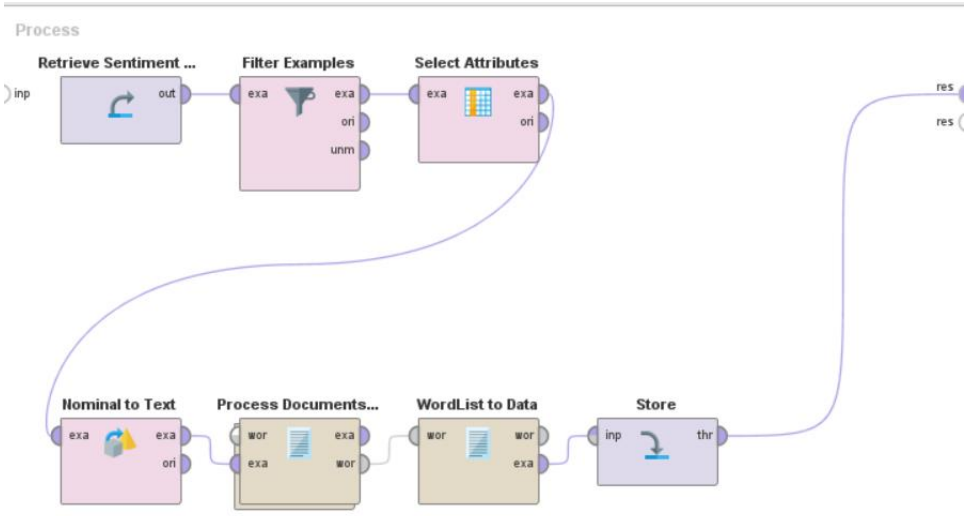

Figure 7 Part 1 of Word Cloud Process – Frequent Positive Words

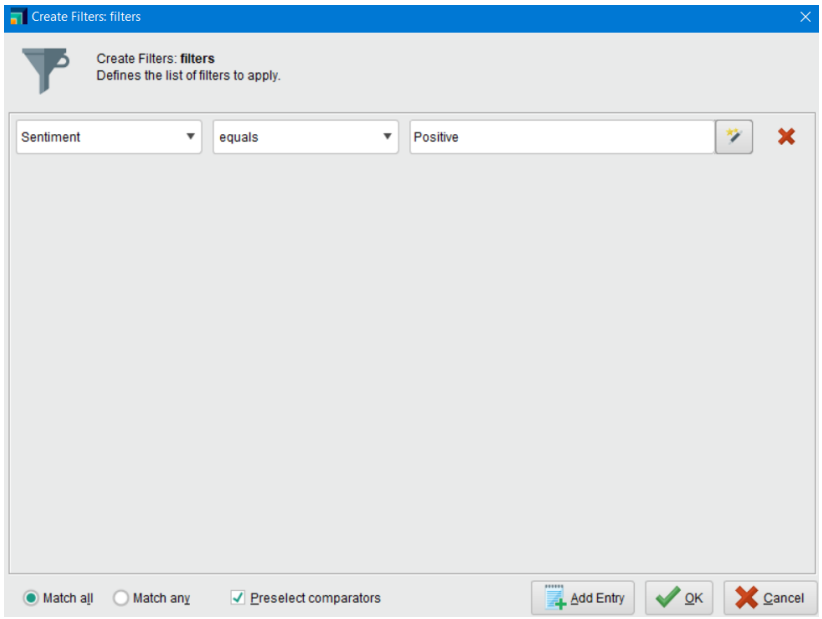

Figure 8 Part 2 of Word Cloud Process - Positive Words Filter

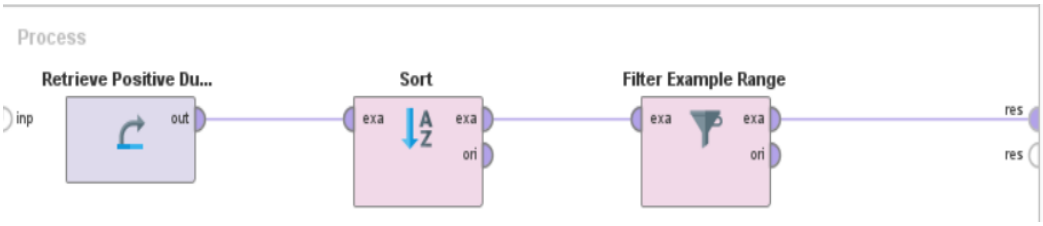

Figure 9 Part 3 of Word Cloud Process - Top 20 Positive Words

3.3.2 Topic Modelling

To find a the hidden topics discussed by social media users, we use Latent Dirichlet Allocation (LDA) method to Operators such as "Join" and "Select Attributes" are used and the subprocesses in the "Loop Collection" operator . After pre-processing, the "Extract Topics from Document (LDA)" operator is used to classify text documents to a particular topic. The parameter for topics and top words have been set manually to 10 and 5, respectively, which does not affect the outcome in identifing topics. The process for topic modeling using LDA for RapidMiner is shown in Figure 10.

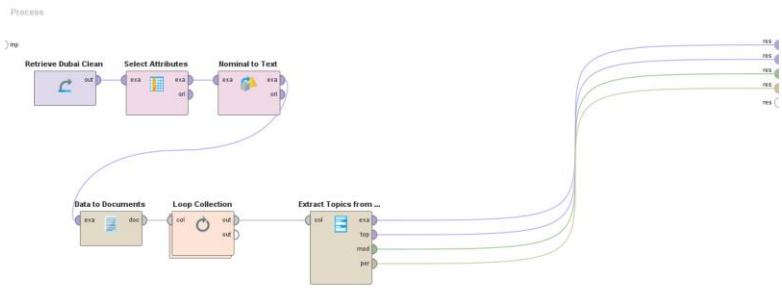

Figure 10 The Process of Topic Modelling using the LDA Method

Chapter 4: Results and Discussion

4.1 RQ1: What is the sentiment analysis of Dubai & Paris using Twitter data?

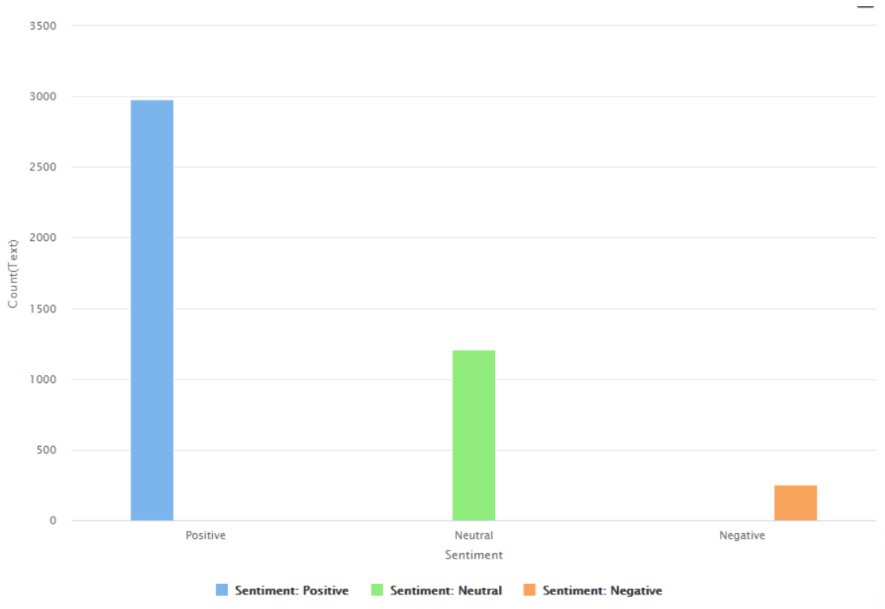

Figure 11 Dubai Sentiment Analysis

After analyzing the data extracted from Twitter and cleaning them; we have come to find that Dubai has a positive sentiment of almost 3000 rows, neutral sentiment of about 1200 rows and about 300 rows of negative sentiment as shown in the figure above.

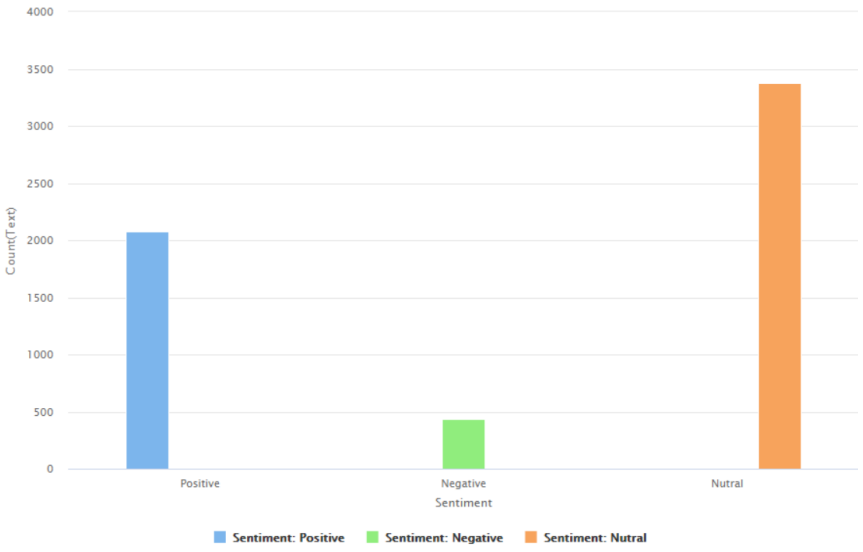

Figure 12 Paris Sentiment Analysis

On the other hand, Paris has almost 3500 rows of neutral sentiment, just above 2000 rows of positive sentiment and almost 500 rows of negative sentiment as shown in the figure above.

4.2 RQ2: What are the keywords used by the users to describe their travel experience and the topics discussed when it comes to these two cities?

Table 1 Top 20 Frequent Words

Dubai

Row No.	word	in documents	total
1	burj	16993	17045
2	khalifa	16981	17044
3	cruis	13992	14042
4	mission	13869	13869
5	film	13864	13866
6	imposs	13858	13858
7	kind	13857	13857
8	design	572	1026
9	naver	479	957
10	view	762	782
11	articl	765	766
12	taehyung	736	737
13	garner	716	716
14	smith	524	524
15	million	482	504
16	birthdai	477	483
17	project	473	474
18	cathedr	469	470
19	detail	462	462
20	dubai	408	457

Paris

Row No.	word	in documents	total
1	tower	2991	3153
2	eiffel	2993	3098
3	samothrac	1280	2558
4	nike	1442	1442
5	wing	1393	1393
6	marbl	1293	1293
7	victori	1287	1287
8	island	1284	1285
9	current	1279	1279
10	discov	1279	1279
11	sculptur	1277	1277
12	hellenist	1276	1276
13	pari	817	881
14	taehyung	514	608
15	triomph	583	588
16	prix	423	423
17	louvr	393	398
18	museum	365	372
19	stop	345	345
20	cool	344	344

Based on our findings; we have come to find that Twitter users uses mainly keywords about Burj Khalifa (A world class Touristic Attraction) as well as several historical touristic attractions in Paris such as the Eiffel Tower and the Louvre Museum as shown in the table above.

Table 2 Top 20 Positive Frequent Words

Dubai

Row No.	word	in documents	total
1	burj	2679	2714
2	khalifa	2669	2705
3	cruis	2286	2299
4	kind	2248	2248
5	mission	2248	2248
6	film	2247	2247
7	imposs	2246	2246
8	love	272	276
9	view	274	275
10	taehyung	255	255
11	articl	253	254
12	media	254	254
13	hundr	253	253
14	garner	252	252
15	thousand	252	252
16	dubai	100	126
17	smith	94	94
18	world	61	65
19	build	52	52
20	tallest	36	37

Paris

Row No.	word	in documents	total
1	tower	1097	1147
2	eiffel	1097	1118
3	pari	520	560
4	triomph	503	504
5	prix	417	417
6	cool	342	342
7	stop	339	339
8	artwork	334	334
9	louvr	227	228
10	museum	209	211
11	taehyung	99	190
12	franc	167	179
13	kiss	178	178
14	antonio	147	147
15	canova	147	147
16	cupid	147	147
17	psych	147	147
18	reviv	147	147
19	beauti	128	136
20	front	129	129

Twitter users have positively frequently used the word tallest and love when it came to describing or talking about Dubai. On the other hand, they used keywords such as kiss and cool to positively describe Paris.

Table 3 Top 20 Negative Frequent Words

Dubai

Row No.	word	in documents	total
1	khalifa	228	241
2	burj	228	235
3	hell	79	79
4	graffiti	78	78
5	design	33	63
6	cost	33	38
7	cathedr	36	36
8	nation	29	29
9	name	12	24
10	battl	20	20
11	pokemon	20	20
12	year	18	18
13	dubai	16	16
14	tallest	15	16
15	build	14	15
16	work	15	15
17	higher	13	13
18	appear	12	12
19	airport	8	11
20	million	6	11

Paris

Row No.	word	in documents	total
1	tower	378	474
2	eiffel	379	448
3	franc	68	70
4	pari	64	69
5	french	59	61
6	travel	51	52
7	forget	46	46
8	fuck	30	32
9	front	30	30
10	answer	29	29
11	lose	29	29
12	weird	29	29
13	mimic	27	27
14	black	23	23
15	louvr	22	23
16	design	21	21
17	world	20	21
18	dick	19	19
19	museum	18	18
20	reject	18	18

Twitter users have negatively frequently used the word tallest and higher when it came to describing or talking about Dubai. On the other hand, they used keywords such as reject and forget to positively describe Paris.

Table 4 Top 20 Neutral Frequent Words

Dubai

Row No.	word	in documents	total
1	burj	1184	1187
2	khalifa	1184	1185
3	naver	476	952
4	design	502	920
5	articl	507	507
6	view	479	496
7	million	468	482
8	taehyung	475	476
9	project	471	471
10	detail	460	460
11	birthdai	459	459
12	garner	459	459
13	cathedr	421	421
14	ghana	417	418
15	time	410	411
16	expens	400	400
17	paid	395	395
18	cost	91	120
19	world	103	104
20	tallest	96	97

Paris

Row No.	word	in documents	total
1	tower	1097	1147
2	eiffel	1097	1118
3	pari	520	560
4	triomph	503	504
5	prix	417	417
6	cool	342	342
7	stop	339	339
8	artwork	334	334
9	louvr	227	228
10	museum	209	211
11	taehyung	99	190
12	franc	167	179
13	kiss	178	178
14	antonio	147	147
15	canova	147	147
16	cupid	147	147
17	psych	147	147
18	reviv	147	147
19	beauti	128	136
20	front	129	129

Twitter users have neutrally frequently used the word cost and paid when it came to describing or talking about Dubai. On the other hand, they used keywords such as front and artwork to positively describe Paris.

Table 5 Dubai Topic Modelling

topic	cohere...	rank_1...	word-le...	exclusi...	docume...	eff_nu...	tokens	uniform...	corpus...	token-d...	allocati...
Topic_0	-5.577	0.919	5.200	0.416	5.700	31.245	3356	3.103	1.530	0.002	0.970
Topic_1	-5.576650461683215	0.986	5.400	0.461	6.172	9.395	5228	5.199	1.383	0.011	0.990
Topic_2	-0.162	0.981	5.600	0.538	6.021	8.617	4133	5.242	1.610	0.011	0.988
Topic_3	-0.051	0.988	6.200	0.724	5.535	8.063	2030	5.393	2.340	0.000	0.988
Topic_4	-32.090	0.916	5.400	0.643	4.956	29.386	1472	2.516	2.064	0.002	0.970
Topic_5	-0.128	0.969	5.400	0.346	7.018	7.005	7692	5.549	0.663	0.000	0.973
Topic_6	-41.776	0.952	5.600	0.426	5.237	60.812	1745	2.211	2.138	0.000	0.976
Topic_7	-0.130	0.976	5.200	0.345	7.065	7.010	8084	5.547	0.662	0.000	0.981
Topic_8	-20.490	0.955	5	0.499	4.613	77.842	1371	1.972	2.359	0.001	0.982
Topic_9	-23.360	0.952	5.600	0.633	5.550	10.491	1386	3.864	1.841	0.000	0.973

Table 5 above shows the results from the LDA Topic Modelling we have applied on Dubai Twitter Data.

Table 6 Paris Topics Modelling

topic	cohere...	rank_1...	word-le...	exclusi...	docume...	eff_nu...	tokens	uniform...	corpus...	token-d...	allocati...
Topic_0	-4.821	0.839	4.800	0.574	5.124	65.161	1676	3.072	2.694	0.000	0.941
Topic_1	-9.445	0.925	5.400	0.658	6.627	18.824	5249	3.977	1.540	0.000	0.961
Topic_2	-12.853	0.926	6	0.626	6.635	12.104	5081	4.695	1.669	0.000	0.953
Topic_3	-2.237	0.955	5.200	0.963	6.347	18.152	3977	4.653	2.389	0.000	0.963
Topic_4	-12.731	0.854	4.800	0.388	5.703	84.565	3168	2.575	1.894	0.000	0.933
Topic_5	-49.851	0.962	6.600	0.637	5.693	34.532	2766	4.363	2.334	0	0.971
Topic_6	-0.477	0.957	5	0.922	5.704	24.445	3188	4.862	2.520	0.000	0.969
Topic_7	-19.705	0.867	5	0.336	6.881	44.101	9606	2.244	1.089	0.000	0.940
Topic_8	-2.179	0.932	5	0.632	5.805	24.726	2973	4.708	2.225	0.000	0.955
Topic_9	-0.026	0.982	6	0.862	7.159	9.314	14067	6.038	1.281	0.011	0.985

Table 6 above shows the results from the LDA Topic Modelling we have applied on Paris Twitter Data.

Chapter 5: Conclusion and Recommendations

5.1 Conclusion

According to the results shown in the results section, Dubai of course has a more positive sentiment than Paris based on Twitter user's data. And so, it would be the better choice for travel agencies when it comes to setting up an agreement with the city deputies.

Figure 13 Dubai Word Cloud

Diving deeper into the frequent words used when it came to Dubai; we have concluded that it mainly was about Burj Khalifa and so it would be preferrable to have that touristic attraction as one of the sightseeing main events for travelers during their trip.

Figure 14 Dubai Positive Word Cloud

Twitter users even frequently used the Burj as a keyword when it came to positively describing their experience in Dubai.

Figure 15 Dubai Negative Word Cloud

They have also used other negatively impactful keywords to express their opinions about Dubai such as hell and graffiti.

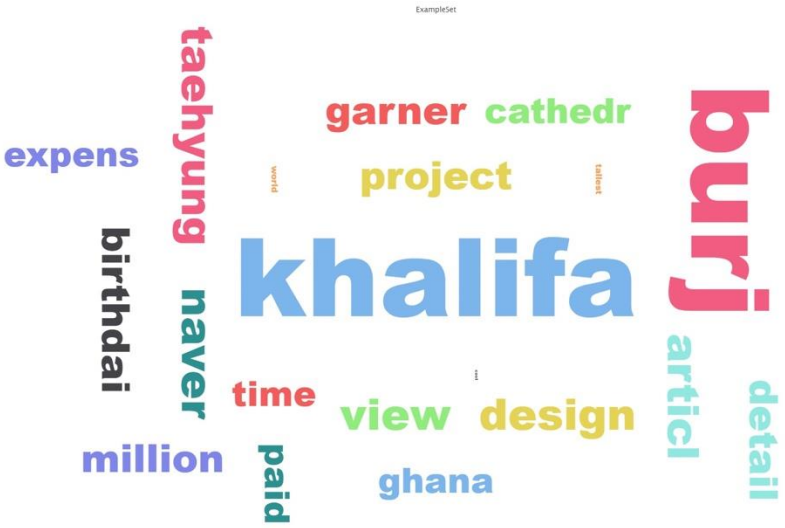

Figure 16 Dubai Neutral Word Cloud

One the other hand, keywords like view and design have been used to neutrally describe their experience in Dubai.

Figure 17 Paris Word Cloud

While looking into the frequent words used when it came to Paris; we have concluded that it mainly was about Eiffel Tower.

Figure 18 Paris Positive Word Cloud

Twitter users even frequently used keywords such as Eiffel, tower, pari when it came to positively describing their experience in Dubai.

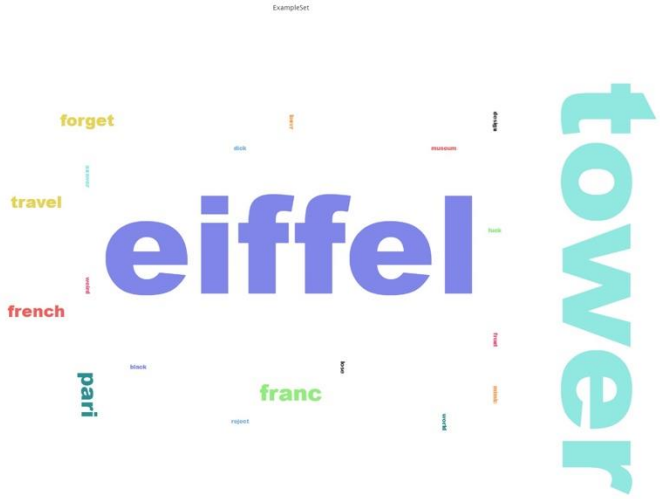

Figure 19 Paris Negative Word Cloud

On the other hand, Eiffel has come up frequently as well when Twitter users' have negatively described their experience in Paris.

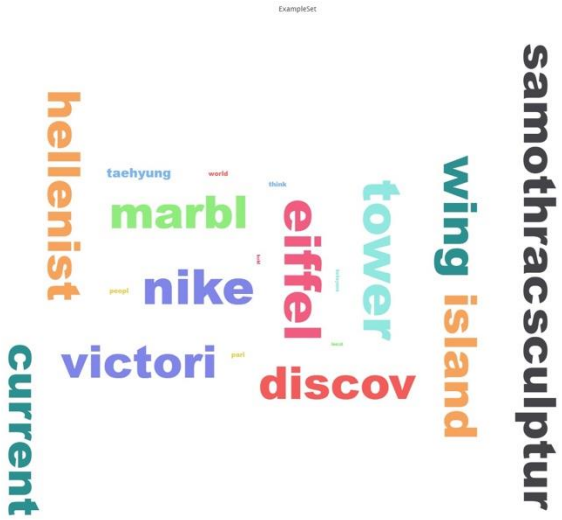

Figure 20 Paris Neutral Word Cloud

Another baffling insight is that people have used keywords such as wing and island frequently to neutrally describe Paris.

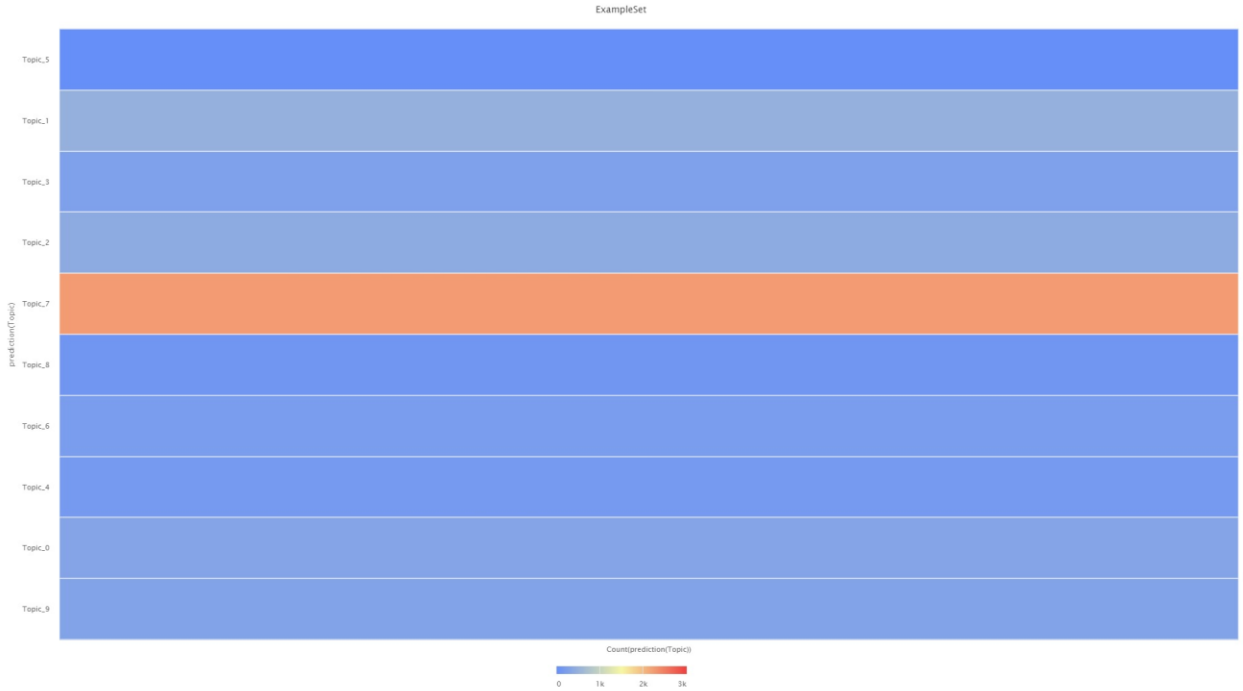

Figure 21 Topic Modelling Visualization – Dubai

Keyword	Major keywords and representative data
Dubai/ burg Khalifa	1- "Thanks for letting me stay in your flat in Burj Khalifa during my trip to Dubai bro and also thanks for letting me use your Maserati ??" 2- "Dubai Famous Land Mark Gift Tourist Souvenir Metal Burj Khalifa Tower on Crystal Base"
Film	1-"Tom Cruise sitting on top of the Burj Khalifa, without a harness of any kind, during filming of Mission Impossible Ghost"
Cruise	1-"Tom Cruise sitting on top of the Burj Khalifa, without a harness of any kind, during filming of Mission Impossible Ghost"

Figure 22 Representative Data for Keywords - Dubai

Keywords mainly used to discuss Topic 7 regarding Dubai were burj, khalifa, film, imposs, cruis and that was to discuss the premiere of the Mission Impossible Movie starring Tom Cruise. Travel agencies can use that to their advantage by marketing before the release of the next film about their packages which will include of course sightseeing of the famous Burj Khalifa as well as a chance to see the movie.

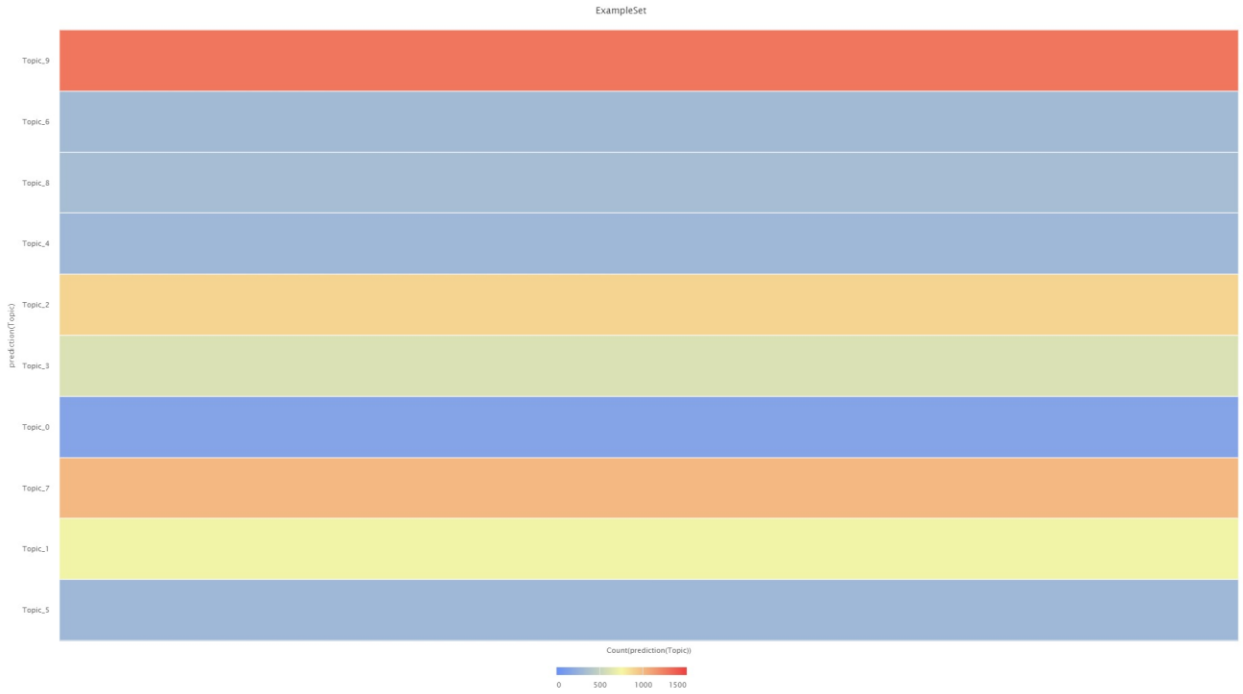

Figure 23 Topic Modelling Visualization – Paris

Keyword	Major keywords and representative data
island	"Winged Victory of Samothrace. Marble Hellenistic sculpture of Nike, discovered on the island of Samothrace. Currently "
Nike	" Winged Victory of Samothrace. Marble Hellenistic sculpture of Nike, discovered on the island of Samothrace. Currently"
wing	" Limestone statue of Assyrian Winged Bull (710 BC), known as Lamassu; from Palace at Khorasan (Citadel of Sargon II),
Samothrace.	"Winged Victory of Samothrace. Marble Hellenistic sculpture of Nike, discovered on the island of Samothrace. Currently "

Figure 24 Representative Data for Keywords - Paris

One the other hand, Twitter users have used keywords such as samothrac, victori, island, wing, nike to discuss Topic 1 regarding Paris. They are discussing different touristic monuments around the city.

5.2 Recommendations Based on the Findings

Based on our analysis as discussed above it would be a huge advantage to sign an agreement with Tourism deputies in Dubai as people are responding more positively to the evolving city. Another Marketing tactic that could help boost the launch of the services provided by the travel agencies is to sponsor or market before the premiere of the next Mission Impossible Movie to help keep the trend of the country going upwards as well as the positive sentiment of the users' sharing their opinions over Twitter.

5.3 Limitations

In our Research, we have come to find the number of people sharing about different topics when it came to each city limiting. As well as we were using the data of all Twitter users with no way of filtering or knowing if the person has visited the city or not, are they living in it and more insights about the users sharing their opinions.

5.4 Recommendations for Future Research

Our recommendation for future research is to send for travelers who have visited the city after the trip to share their experience in their own words and then assess the data collected using RapidMiner. This method could then help enhance our analysis

when compared with the data collected from the internet. This method as well would help us gain more descriptive insights about the demographics of the users sharing their opinions about their travel experience.

References

I. Deraman, N. A., Buja, A. G., Samah, K. A. F. A., Jono, M. N. H. H. J., Isa, M. A. M., & Saad, S. (2021). *A Social Media Mining Using Topic Modeling and Sentiment Analysis on Tourism in Malaysia During Covid19*. 1–9.

II. Lickorish, L., & Jenkins, C. L. (1997). *An Introduction to Tourism*. Taylor & Francis Group,.

III. Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment Analysis and Opinion Mining: A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*(6), 282–292.