How can Google Trends and sentiment analysis of TripAdvisor and Facebook predict visitor numbers to the United Kingdom and Canada from 2010 - 2015?

User Configuration Manual

MSc Research Project - Data Analytics

Supervisor: Mr. Tony Delaney

Emmet Hutchin
Emmet.hutchin@student.ncirl.ie

National College of Ireland
Project Submission Sheet – 2015/2016
School of Computing

National College of Ireland

Student Name:	Hutchin Emmet
Student ID:	x15018695
Programme:	Data Analytics
Year:	2016
Module:	MSc Reseach Project
Lecturer:	Mr. Tony Delaney
Submission Due Date:	22/08/2016
Project Title:	How can Google Trends and sentiment analysis of TripAdvisor and Facebook predict visitor numbers to the United Kingdom and Canada from 2010 - 2015?
Word Count:	3825

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	22nd August 2016

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

User Configuration Manual

Introduction

This user configuration manual is designed to allow users to follow and recreate the implementation process used to implement the MSc Research Project. It will cover the extraction of the hotel list and reviews from TripAdvisor, comments from Facebook and the index data from Google Trends. It will then show how sentiment analysis was carried out on the reviews and comments, how the optimal lag for the independent variables and finally how the Linear and Multiple Regression were carried out and analysed.

Hotels

Canada

TripAdvisor does not provide a complete list of hotels for Canada and does not provide an API for research purposes, however, it does provide a list of hotels by province (Figure 1). Using the rvest package in R, the name and URL of each hotel was scraped from the website and stored in separate files for each province which were then amalgamated to obtain the total number of hotels in Canada (for the full code see Appendix 1). An example of the code is shown in Figure 2.

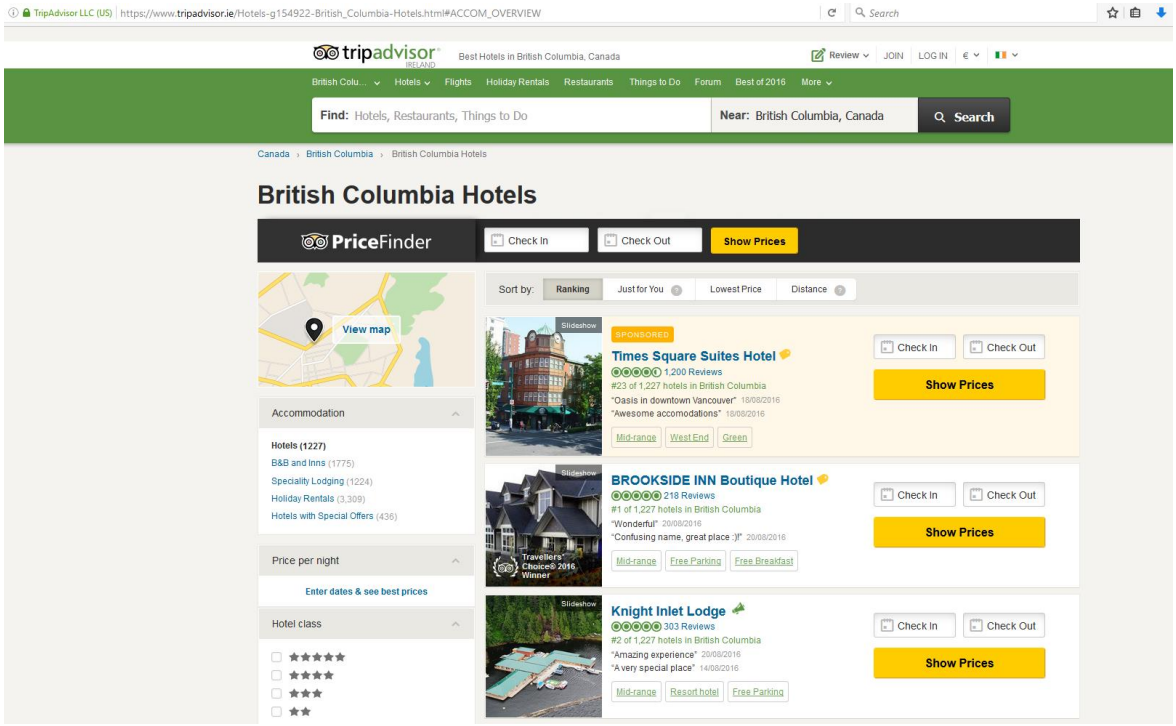


Figure 1 – Example of the list of Hotels on TripAdvisor.

```r
split.value <- "g3563882"

file1 <- "Yukonhotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g155045-Yukon-Hotels.html#ACCOM_OVERVIEW"
split.value <- "g155045"

########################

url <- read_html(hotel.url)

max.page.list <- ".last"
max.page2.list <- html_nodes(url, max.page.list) %>%
  html_text()

max.page2.list <- as.integer(max.page2.list)

max.page <- max.page2.list

max.counter <- (max.page * 30) - 30
page.counter <- 0


while(page.counter <= max.counter)
{

  if (page.counter == 0)
  {
    url <- read_html(hotel.url)

  }

  else
  {

    hotel.url.split <- strsplit(hotel.url, split.value)
    rejoined.url <- paste(hotel.url.split[[1]][1],split.value,"-oa",page.counter,hotel.url.split[[1]][2], sep="")
    url <- read_html(rejoined.url)
  }
```

Figure 2 – Example code for Hotel Name extraction.

The sample size for the hotels was obtained for each province based on the number of visitors that began their visits in that province (Figure 3)

	A	B	C	D	E	F	G	H	I	J
1	Access by point of entry 2015									
2	http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot-dec2015_en-lowres.pdf									
3										
4	Total visitors	17,782,949	Sample of hotels to be used	0.1	Total No. of hotels	5413				
5										
6	Province/Territory	Visitors	% of visitors	No. of hotels to smaple	% of hotels					
7	Ontario	8,267,015	46.48843676	2516	252					
8	Quebec	2,622,135	14.74522027	798	80					
9	Nova Scotia	217,476	1.222946768	66	7					
10	New Brunswick	312,972	1.759955562	95	10					
11	Manitoba	228,393	1.284337035	70	7					
12	British Columbia	4,925,916	27.70022003	1499	150					
13	Prince Edward Island	310	0.001743243	0	0					
14	Saskatchewan	85,866	0.482855796	26	3					
15	Alberta	912,319	5.130302066	278	28					
16	Newfoundland and Labrador	60,644	0.341023303	18	2					
17	Yukon	149,903	0.842959174	46	5					
18										
19										

Figure 3 – Numbers of hotels per province for analysis.

The hotels were selected randomly from each provincial file using the rand function in Excel and then amalgamated into a sample hotel file using R.

R then looped through every hotel in the file to obtain the maximum number of review pages for each hotel as TripAdvisor only show 10 reviews per page.

The UK

The process was similar for the UK except TripAdvisor breaks the UK up by town for England, Scotland, Wales and Northern Ireland. The pages of towns in each of these had to be looped through to obtain a full list of towns, this was then looped through to scrape a list of hotels in each town and this list was amalgamated and a sample of hotels taken in a similar manner to Canada.

Facebook

This list of hotels was used to obtain the Facebook ID and URL of each hotel that has a Facebook page using the rFacebook package in R (see appendix 2)

Google Trends

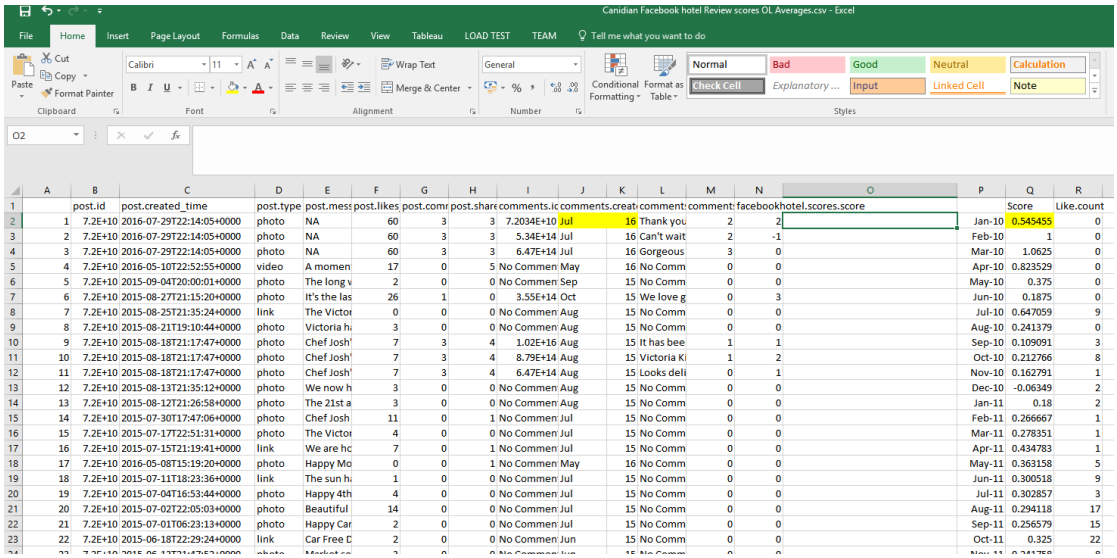
The hotel list was also used to extract Google Trends data for each hotel (if it existed) using the gtrendsR package in R (see appendix 3) and the data was stored in a csv file.

Sentiment Analysis

The hotel review data from TripAdvisor was extracted using the code in Appendix 4 into individual files and then amalgamated into one file. The file containing the Facebook ID's was passed into the program in Appendix 5 to download each comment and its associated post, like count, date etc. into individual files and again these were amalgamated into a single file.

The sentiment analysis score was calculated by using the code in appendix 6 for the Facebook and TripAdvisor files. This program sets the path for the relevant lexicon, cleans the data of punctuation, carriage return symbols etc. and analyses the reviews/comments for positive and negative words and gives an overall score per comment.

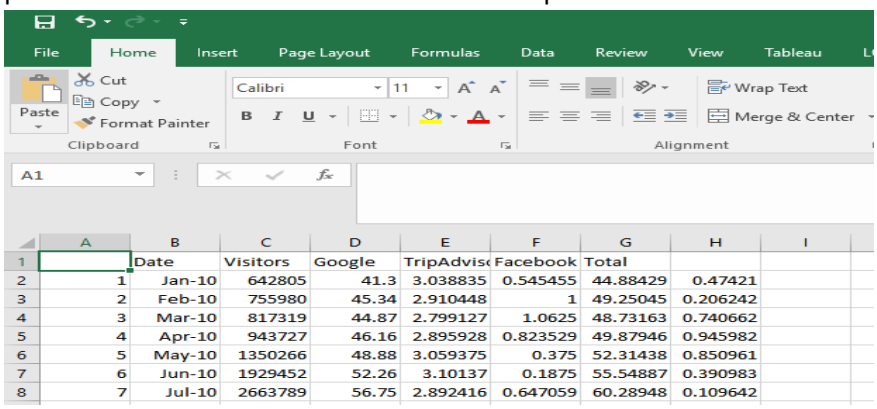
These CSV files are then manipulated in Excel to change the data format to match the data format in the visitor data files i.e. month and year. The AVERAGEIFS function was then used to calculate the average score per month, see Figure 4.



	post.id	post.created_time	post.type	post.mess	post.likes	post.comr	post.share	comments.ic	comments.creat	comment	comment	facebookhotel.scores.score		Score	Like.count
2	1	7.2E+10	2016-07-29T22:14:05+0000	photo	NA	60	3	3	7.2034E+10	Jul	16	Thank you	2	2	Jan-10	0.545455	0
3	2	7.2E+10	2016-07-29T22:14:05+0000	photo	NA	60	3	3	5.34E+14	Jul	16	Can't wait	2	-1	Feb-10	1	0
4	3	7.2E+10	2016-07-29T22:14:05+0000	photo	NA	60	3	3	6.47E+14	Jul	16	Gorgeous	3	0	Mar-10	1.0625	0
5	4	7.2E+10	2016-05-10T22:52:55+0000	video	A momen	17	0	5	No Commen	May	16	No Comm	0	0	Apr-10	0.823529	0
6	5	7.2E+10	2015-09-04T20:00:01+0000	photo	The long v	2	0	0	No Commen	Sep	15	No Comm	0	0	May-10	0.375	0
7	6	7.2E+10	2015-08-27T21:15:20+0000	photo	It's the las	26	1	0	3.55E+14	Oct	15	We love g	0	3	Jun-10	0.1875	0
8	7	7.2E+10	2015-08-25T21:35:24+0000	link	The Victor	0	0	0	No Commen	Aug	15	No Comm	0	0	Jul-10	0.647059	9
9	8	7.2E+10	2015-08-21T19:10:44+0000	photo	Victoria hi	3	0	0	No Commen	Aug	15	No Comm	0	0	Aug-10	0.241379	0
10	9	7.2E+10	2015-08-18T21:17:47+0000	photo	Chef Josh'	7	3	4	1.02E+16	Aug	15	It has bee	1	1	Sep-10	0.109091	3
11	10	7.2E+10	2015-08-18T21:17:47+0000	photo	Chef Josh'	7	3	4	8.79E+14	Aug	15	Victoria Ki	1	2	Oct-10	0.212766	8
12	11	7.2E+10	2015-08-18T21:17:47+0000	photo	Chef Josh'	7	3	4	6.47E+14	Aug	15	Looks deli	0	1	Nov-10	0.162791	1
13	12	7.2E+10	2015-08-13T21:35:12+0000	photo	We now h	3	0	0	No Commen	Aug	15	No Comm	0	0	Dec-10	-0.06349	2
14	13	7.2E+10	2015-08-12T21:26:58+0000	photo	The 21st a	3	0	0	No Commen	Aug	15	No Comm	0	0	Jan-11	0.18	2
15	14	7.2E+10	2015-07-30T17:47:06+0000	photo	Chef Josh	11	0	1	No Commen	Jul	15	No Comm	0	0	Feb-11	0.266667	1
16	15	7.2E+10	2015-07-17T22:51:31+0000	photo	The Victor	4	0	0	No Commen	Jul	15	No Comm	0	0	Mar-11	0.278351	1
17	16	7.2E+10	2015-07-15T21:19:41+0000	link	We are hc	7	0	1	No Commen	Jul	15	No Comm	0	0	Apr-11	0.434783	1
18	17	7.2E+10	2016-05-08T15:19:20+0000	photo	Happy Mo	0	0	1	No Commen	May	16	No Comm	0	0	May-11	0.363158	5
19	18	7.2E+10	2015-07-11T18:23:36+0000	link	The sun hi	1	0	0	No Commen	Jul	15	No Comm	0	0	Jun-11	0.300518	9
20	19	7.2E+10	2015-07-04T16:53:44+0000	photo	Happy 4th	4	0	0	No Commen	Jul	15	No Comm	0	0	Jul-11	0.302857	3
21	20	7.2E+10	2015-07-02T22:05:03+0000	photo	Beautiful	14	0	0	No Commen	Jul	15	No Comm	0	0	Aug-11	0.294118	17
22	21	7.2E+10	2015-07-01T06:23:13+0000	photo	Happy Car	2	0	0	No Commen	Jul	15	No Comm	0	0	Sep-11	0.256579	15
23	22	7.2E+10	2015-06-18T22:29:24+0000	link	Car Free D	2	0	0	No Commen	Jun	15	No Comm	0	0	Oct-11	0.325	22

Figure 4 – Facebook average sentiment analysis score per month

The Visitor data, Google Trend, TripAdvisor and Facebook data broken down by month was then placed in a file and the total score of the independent variables was calculated. This file was then used in the regression analysis (see Figure 5)



	A	B	C	D	E	F	G	H	I
1		Date	Visitors	Google	TripAdvis	Facebook	Total		
2	1	Jan-10	642805	41.3	3.038835	0.545455	44.88429	0.47421	
3	2	Feb-10	755980	45.34	2.910448	1	49.25045	0.206242	
4	3	Mar-10	817319	44.87	2.799127	1.0625	48.73163	0.740662	
5	4	Apr-10	943727	46.16	2.895928	0.823529	49.87946	0.945982	
6	5	May-10	1350266	48.88	3.059375	0.375	52.31438	0.850961	
7	6	Jun-10	1929452	52.26	3.10137	0.1875	55.54887	0.390983	
8	7	Jul-10	2663789	56.75	2.892416	0.647059	60.28948	0.109642	

Figure 5 – Amalgamated data by month

The appropriate time lag for each independent variable was then calculated by the linear regression code and when it was determined the regression analysis was carried out using the code in appendix 7 which calculates the correlation, R-squared values, coefficients, Variance Inflation Factor etc. and stores them in text files and the scatter plots etc. as jpg files.

Appendix 1 – Hotel extraction code for Canada:

```
install.packages(c("devtools", "rjson", "bit64", "httr", "plyr", "ggplot2", "doBy", "XML", "base64enc",
"Rfacebook", "rvest" , "stringr" ))
install.packages(c("xlsx"))
install.packages(c("plyr"))
setwd("/Users/ Dropbox/MSc Data Analytics/Thesis/Hotels")
#devtools allows us to install from github
library(devtools)

#these are various libraries that we use throughout this example
library(plyr)
library(httr)
library(doBy)
library(Rfacebook)
library(rvest)
library(ggplot2)
library(stringr)
library(xlsx)


file1 <- "BChotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g154922-British_Columbia-
Hotels.html#ACCOM_OVERVIEW"
split.value <- "g154922"

file1 <- "Ontariohotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g154979-Ontario-Hotels.html#ACCOM_OVERVIEW"
split.value <- "g154979"

file1 <- "Quebechotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g155025-Quebec-Hotels.html#ACCOM_OVERVIEW"
split.value <- "g155025"

file1 <- "Nova_Scotiahotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g154967-Nova_Scotia-
Hotels.html#ACCOM_OVERVIEW"
split.value <- "g154967"

file1 <- "New_Brunswickhotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g154956-New_Brunswick-
Hotels.html#ACCOM_OVERVIEW"
split.value <- "g154956"

file1 <- "Manitobahotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g154950-Manitoba-
Hotels.html#ACCOM_OVERVIEW"
split.value <- "g154950"

file1 <- "Prince_Edward_Islandhotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g155022-Prince_Edward_Island-
Hotels.html#ACCOM_OVERVIEW"
```

```
split.value <- "g155022"

file1 <- "Saskatchewanhotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g155038-Saskatchewan-
Hotels.html#ACCOM_OVERVIEW"
split.value <- "g155038"

file1 <- "Albertahotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g154909-Alberta-Hotels.html#ACCOM_OVERVIEW"
split.value <- "g154909"

file1 <- "Newfoundlandhotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g3563882-
Newfoundland_Newfoundland_and_Labrador-Hotels.html#ACCOM_OVERVIEW"
split.value <- "g3563882"

file1 <- "Yukonhotels.csv"
hotel.url <- "https://www.tripadvisor.ie/Hotels-g155045-Yukon-Hotels.html#ACCOM_OVERVIEW"
split.value <- "g155045"
########################
url <- read_html(hotel.url)

max.page.list <- ".last"
max.page2.list <- html_nodes(url, max.page.list) %>%
 html_text()
max.page2.list <- as.integer(max.page2.list)

max.page <- max.page2.list
max.counter <- (max.page * 30) - 30
page.counter <- 0
while(page.counter <= max.counter)
{
   if (page.counter == 0)
 {
   url <- read_html(hotel.url)

 }
   else
 {
   hotel.url.split <- strsplit(hotel.url, split.value)
   rejoined.url <- paste(hotel.url.split[[1]][1],split.value,"-oa",page.counter,hotel.url.split[[1]][2],
sep="")
   url <- read_html(rejoined.url)
 }
 #Adoped from https://github.com/hadley/rvest/blob/master/demo/tripadvisor.R
 hotel.name <- ".property_title "
 hotel.name2 <- html_nodes(url, hotel.name) %>%
   html_text()

 hotel.page <- ".property_title "
 web.page <- html_nodes(url, hotel.page) %>%
```

```
   html_attr("href")

  price.range <- ".clickable_tags"
  price.range2 <- html_nodes(url, price.range) %>%
   html_text()

  #
  #http://zevross.com/blog/2015/05/19/scrape-website-data-with-the-new-r-package-rvest/
  #
  to_remove<-paste(c("Northeast","North", "South", "East", "West", "Calgary","Place
LaRue","Beltline", "Business District","Airport Business Area","Business Area","Airport","Core","St.
James Industrial","Winnipeg"," Halifax", "Halifax", "South End","Ville-Marie", "Green", "Pool", "Free
Parking", "Downtown", "Casino", "Pets Allowed", "Beach", "Free Breakfast", "West End", "Resort
hotel", "Fairview", "City Center", "James Bay", "East Cambie", "Mt. Pleasant", "North Shore",
"Newton", "Aberdeen", "Guildford", "Willingdon Heights", "Whalley", "Central City", "Burnside",
"Valleyview", "Gastown", "Kitsilano", "Fleetwood", "Renfrew", "Willowbrook", "Yorkville","
Toronto", "Toronto", "Central", "Central London", "The Annex","Cote-des-Neiges-Notre-Dame-de-
Grace","Rosemont-La Petite-Patrie","Le Plateau-Mont-Royal","Mile-End","District du Lac-
Beuchamp","Maizerets","Quebec City","City","District des Promenades","District des
Riverains","Dorval/ /TrainStation","Montcalm","Sillery"," "), collapse="|")
  price.range2<-gsub(to_remove, "", price.range2)

  data.frame(hotel.name2,web.page, price.range2, stringsAsFactors = FALSE)
   hotelfile <- data.frame(hotel.name2,web.page, price.range2, stringsAsFactors = FALSE)

  if (page.counter == 0)
  {
    write.csv(hotelfile, file = file1)
  }
  else
  {
    hotel <- read.csv (file1,stringsAsFactors = FALSE)
    hotelWithoutX <- hotel[, -grep("X$", colnames(hotel))]
    bind <- rbind(hotelWithoutX,hotelfile)
    write.csv(bind, file = file1)
  }
   page.counter <- page.counter + 30
}

#To add https://www.tripadvisor.com to the link scraped from TripAdvisor.

web <- read.csv (file1,stringsAsFactors = FALSE)
link <- paste("https://www.tripadvisor.com",web$web.page, sep = "")
data.frame(web$hotel.name2,link,web$price.range2, stringsAsFactors = FALSE)
web.link <- data.frame(web$hotel.name2,link,web$price.range2,stringsAsFactors = FALSE)
write.csv(web.link, file = file1)

############
############
BChotels <- read.csv ("BChotels.csv",stringsAsFactors = FALSE)
Ontariohotels <- read.csv ("Ontariohotels.csv",stringsAsFactors = FALSE)
```

```
Quebechotels <- read.csv ("Quebechotels.csv",stringsAsFactors = FALSE)
Nova_Scotiahotels <- read.csv ("Nova_Scotiahotels.csv",stringsAsFactors = FALSE)
New_Brunswickhotels <- read.csv ("New_Brunswickhotels.csv",stringsAsFactors = FALSE)
Manitobahotels <- read.csv ("Manitobahotels.csv",stringsAsFactors = FALSE)
Prince_Edward_Islandhotels <- read.csv ("Prince_Edward_Islandhotels.csv",stringsAsFactors =
FALSE)
Saskatchewanhotels <- read.csv ("Saskatchewanhotels.csv",stringsAsFactors = FALSE)
Albertahotels <- read.csv ("Albertahotels.csv",stringsAsFactors = FALSE)
Newfoundlandhotels <- read.csv ("Newfoundlandhotels.csv",stringsAsFactors = FALSE)
Yukonhotels <- read.csv ("Yukonhotels.csv",stringsAsFactors = FALSE)

bind <- rbind(BChotels, Ontariohotels, Quebechotels, Nova_Scotiahotels, New_Brunswickhotels,
Manitobahotels, Prince_Edward_Islandhotels, Saskatchewanhotels, Albertahotels,
Newfoundlandhotels, Yukonhotels)
write.csv(bind, file = "All hotels Canada.csv")

###################
#Binds sample hotels
BChotels_sample <- read.csv ("BChotels_sample.csv",stringsAsFactors = FALSE)
Ontariohotels_sample <- read.csv ("Ontariohotels_sample.csv",stringsAsFactors = FALSE)
Quebechotels_sample <- read.csv ("Quebechotels_sample.csv",stringsAsFactors = FALSE)
Nova_Scotiahotels_sample <- read.csv ("Nova_Scotiahotels_sample.csv",stringsAsFactors = FALSE)
New_Brunswickhotels_sample <- read.csv ("New_Brunswickhotels_sample.csv",stringsAsFactors =
FALSE)
Manitobahotels_sample <- read.csv ("Manitobahotels_sample.csv",stringsAsFactors = FALSE)
Saskatchewanhotels_sample <- read.csv ("Saskatchewanhotels_sample.csv",stringsAsFactors =
FALSE)
Albertahotels_sample <- read.csv ("Albertahotels_sample.csv",stringsAsFactors = FALSE)
Newfoundlandhotels_sample <- read.csv ("Newfoundlandhotels_sample.csv",stringsAsFactors =
FALSE)
Yukonhotels_sample <- read.csv ("Yukonhotels_sample.csv",stringsAsFactors = FALSE)

bind <- rbind(BChotels_sample, Ontariohotels_sample, Quebechotels_sample,
Nova_Scotiahotels_sample, New_Brunswickhotels_sample, Manitobahotels_sample,
Saskatchewanhotels_sample, Albertahotels_sample, Newfoundlandhotels_sample,
Yukonhotels_sample)
write.csv(bind, file = "All hotels_sample Canada.csv")

###################
#To find the maximum number of pages for the hotel reviews
##
opening.file <- read.csv ("All hotels_sample Canada.csv",stringsAsFactors = FALSE)
closing.file <- "All hotels_sample Canada max.csv"

hotel.name <- opening.file$web.hotel.name2
hotel.url <- opening.file$link
price.range <- opening.file$web.price.range2
rand <- opening.file$Rand
row.counter <- 1
max.row <- NROW(hotel.url)
```

```
#Loop through and count all the pages to obtain the maximum number of review pages.
while(row.counter <= max.row)
{
  url <- read_html(hotel.url[row.counter])
#TripAdvisor uses several ways to desigate the last page of reviews.
  max.page <- ".taLnk:nth-child(8)"
  max.page2 <- html_nodes(url, max.page) %>%
    html_text()
    to_remove <-paste("Safari", collapse = "|")
  max.page2 <-gsub(to_remove,"", max.page2[1])
  max.page2 <- as.integer(max.page2)
    if(is.na(max.page2))
  {
    max.page <- ".pageNum.taLnk"
    max.page2 <- html_nodes(url, max.page) %>%
      html_text()
    max.page2 <- as.integer(max.page2)
    max.page2 <- max(max.page2)
  }
   if(length(max.page2) == 0)
  {
    max.page2 <- 1
  }
    if(max.page2 == -Inf)
  {
    max.page2 <- 1
  }
    h.name <- hotel.name[row.counter]
  h.url <- hotel.url[row.counter]
  h.range <- price.range[row.counter]
  h.rand <- rand[row.counter]
  web.link <- data.frame(h.name,h.url,h.range,max.page2,h.rand, stringsAsFactors = FALSE)

  if(row.counter == 1)
  {
    write.csv(web.link, file = closing.file)
  }
   else
  {
    hotel.max <- read.csv (closing.file,stringsAsFactors = FALSE)
    hotelWithoutX.max <- hotel.max[, -grep("X$", colnames(hotel.max))]
    bind <- rbind(hotelWithoutX.max,web.link)
    write.csv(bind, closing.file)
  }
   row.counter <- row.counter + 1

}
```

Appendix 2 – Facebook ID extraction

```
fb_oauth <- fbOAuth(app_id="111111111111", app_secret="1111111111111",
extended_permissions = TRUE)
save(fb_oauth, file="fb_oauth")
load("fb_oauth")

etoken <- "1212121213212121"
#pages <- searchPages( string="Sandman Signature Langley Hotel", token=fb_oauth, n=1000 )

setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels")
#opening.file <- read.csv ("All hotels_sample Canada max 1.csv",stringsAsFactors = FALSE)
#opening.file <- read.csv ("All hotels Canada.csv",stringsAsFactors = FALSE)
#opening.file <- read.csv ("UK Sample Hotels max 1.csv",stringsAsFactors = FALSE)
opening.file <- read.csv ("Combined Hotels UK.csv",stringsAsFactors = FALSE)

closing.file <- "Facebook Name and ID"
#country <- "Canada"
country <- "United Kingdom"
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/UK/Facebook 2")
#setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook 2")
#For UK
hotel.counter <- opening.file$X.1
hotel.name <- opening.file$hotel.name2
#hotel.TripAdvisor.link <- opening.file$link
hotel.TripAdvisor.link <- opening.file$web.page

#For Canada
#hotel.counter <- opening.file$X
#hotel.name <- opening.file$web.hotel.name2
#hotel.TripAdvisor.link <- opening.file$link

row.counter <-1
max.counter <- NROW(hotel.counter)
#max.counter <- 12
while(row.counter <= max.counter)
{
  counter <- hotel.counter[row.counter]
  h.name <- hotel.name[row.counter]
  link <- hotel.TripAdvisor.link[row.counter]
  link <- as.character(link)

  closing.file1 <- paste(counter," ",h.name," ",closing.file,".csv", sep="")

  split <- strsplit(link,"Reviews-")
  split2 <- split[[1]][2]
  split3 <- strsplit(split2,"-")
  split4 <- strsplit(split3[[1]][2],"_")

  city <- split4[[1]][1]

  #tryCatch http://stackoverflow.com/questions/14748557/skipping-error-in-for-loop
```

```
 tryCatch({
 pages <- searchPages( string=h.name, token=etoken, n=1000 )
 #pages <- searchPages( string=h.name, token=fb_oauth, n=1000 )
 },error=function(e){})

 fb.page.counter <- 1
 max.fb.counter <- NROW(pages$id)

 while(fb.page.counter <= max.fb.counter)
 {
  #tryCatch http://stackoverflow.com/questions/14748557/skipping-error-in-for-loop
  #tryCatch({
  fb.id <- pages$id[fb.page.counter]
  fb.link <- pages$link[fb.page.counter]
  fb.city <- pages$city[fb.page.counter]
  fb.country <- pages$country[fb.page.counter]
  fb.name <- pages$name[fb.page.counter]

  check.file.exists.flag <- 0
  ###################

  name.check <- h.name[h.name %in% (fb.name)]

  h.name.length <- length(name.check)
  ##################
  if(!is.na(fb.name) && !is.na(fb.city) && !is.na(fb.country))
  {
   to_remove<-paste(c(" hotel"," Hotel"," Lodge"," lodge" ," The", " A45", " (M6/J21)"),
collapse="|")
    fb.name2<-gsub(to_remove, "", fb.name)

    if(h.name %in% fb.name2 && city == fb.city && country == fb.country)
    {
     facebook.file <- data.frame(h.name,fb.name,link,fb.id,fb.link,fb.city,fb.country,stringsAsFactors
= FALSE)

      if(check.file.exists.flag == 0)
      {
       write.csv(facebook.file, file = closing.file1)
       check.file.exists.flag <- 1
      }
      else
      {
       fb <- read.csv (closing.file1,stringsAsFactors = FALSE)
       fbWithoutX <- hotel[, -grep("X$", colnames(fb))]
       bind <- rbind(fbWithoutX,facebookfile)
       write.csv(bind, file = closing.file1)
      }
    }
    else
    {
```

```
      do.this <- "Do Nothing"
     }
    }

    fb.page.counter <- fb.page.counter + 1
     }
    row.counter <- row.counter + 1
}
###################
#To merge all of the files in each countries directory
#Based on https://stat.ethz.ch/pipermail/r-help/2003-April/032077.html
###################
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook 1")
all.the.files <- list.files()
all.the.data <- lapply( all.the.files,  read.csv, header=TRUE)
DATA <- do.call("rbind", all.the.data)
write.csv (DATA,"Facebook IDs Canada.csv")
#################
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/UK/Facebook 2")
all.the.files <- list.files()
all.the.data <- lapply( all.the.files,  read.csv, header=TRUE)
DATA <- do.call("rbind", all.the.data)
write.csv (DATA,"Facebook IDs UK.csv")
```

Appendix 3 – Google Trend extraction data

```
###################################
#Basted on https://www.r-bloggers.com/download-and-plot-google-trends-data-with-r/
###################################
install.packages("gtrendsR")
library(gtrendsR)
######
#Delete user ID and password
user <- "xxxxxxxxxxxxx" #Google user ID
psw <- "xxxxxxxxxx" #Google password
gconnect(user, psw)
########
#Canada
#setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels")
#opening.file <- read.csv("All hotels_sample Canada max.csv",header=TRUE, sep=",")
#master.file <- "Google Trends data - Canada.csv"
#setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Google Trends")
#########
# setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook 2")
# opening.file <- read.csv("Facebook IDs Canada.csv",header=TRUE, sep=",")
# master.file <- "Google Trends data - Canada.csv"
# setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Google Trends 2")
#########
#UK
#setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels")
#opening.file <- read.csv("UK Sample Hotels max.csv",header=TRUE, sep=",")
#master.file <- "Google Trends data month - UK.csv"
#setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/UK/Google Trends")
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels")
opening.file <- read.csv("Combined Hotels UK.csv",header=TRUE, sep=",")
master.file <- "Google Trends data month - UK.csv"
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/UK/Google Trends 2")


#For Canada
#hotel.counter <- opening.file$X.1
#hotel.name <- h.name.car
#hotel.TripAdvisor.link <- opening.file$link
#hotel.TripAdvisor.link <- opening.file$h.url

str(opening.file)
#For UK
h.name.car <- as.character(opening.file$hotel.name2)
 hotel.counter <- opening.file$X.1
 hotel.name <- h.name.car
 hotel.TripAdvisor.link <- opening.file$web.page
 #hotel.TripAdvisor.link <- opening.file$h.url

row.counter <- 1
max.counter = NROW(opening.file$X)
```

```
check.file.exists.flag <- 0
#check.file.exists.flag <- 1

while(row.counter <= max.counter)
{
  h.name <- hotel.name[row.counter]
  link <- hotel.TripAdvisor.link[row.counter]
   link <- as.character(link)
  split <- strsplit(link,"Reviews-")
  split2 <- split[[1]][2]
  split3 <- strsplit(split2,"-")
  split4 <- strsplit(split3[[1]][2],"_")
  city <- split4[[1]][1]
  search.term <- paste(h.name, " ", city, sep="")

  closing.file <- paste(row.counter," ","Google Trends Month", " ",h.name,".csv", sep="")

  tryCatch({
  #hotel.search <- gtrends(c(query = h.name), start_date =as.Date("2010-01-01"), end_date =
as.Date("2015-12-31"))
  hotel.search <- gtrends(c(query = search.term), start_date =as.Date("2010-01-01"), end_date =
as.Date("2015-12-31"))

  start.date <- hotel.search$trend[1] #Start Date
  end.date <- hotel.search$trend[2] #End Date
  hotel <- hotel.search$trend[3] #Hotel Name

  query.results <- data.frame(start.date,end.date,hotel)

  write.csv(query.results, closing.file)

  if(check.file.exists.flag == 0)
  {
    write.csv(query.results, master.file)
    check.file.exists.flag <- 1
  }
  else
  {
    g.t <- read.csv(master.file,stringsAsFactors = FALSE)
    g.tWithoutX <- g.t[, -grep("X$", colnames(g.t))]
    temp.file <- read.csv(closing.file, header=TRUE, sep=",")
    temp.file.tWithoutX <- temp.file[, -grep("X$", colnames(temp.file))]
    bind <- cbind(g.tWithoutX,temp.file.tWithoutX[3])
    write.csv(bind, master.file)
  }
  },error=function(e){})
  row.counter <- row.counter + 1
}
```

Appendix 4 – Trip Advisor Review Data

```
######################
#To obtain the review data
#Based on https://github.com/hadley/rvest/blob/master/demo/tripadvisor.R
######################
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels")
opening.file <- read.csv ("UK Sample Hotels max.csv",stringsAsFactors = FALSE)
closing.file <- "TripAdvisor Reviews UK"
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/UK")

hotel.review.name <- opening.file$h.name
hotel.review.url <- opening.file$h.url
maximum.page <- opening.file$max.page2

row.counter <- 865
max.row <- NROW(hotel.review.url)

while(row.counter <= max.row)
{
 hotel.url <- hotel.review.url[row.counter]
 max.page <- maximum.page[row.counter]

 max.counter <- (max.page * 10) - 10
 page.counter <- 0

 while(page.counter <= max.counter)
 {

  if (page.counter == 0)
  {
   url <- hotel.url
  }
  else if (page.counter <100)
  {
   hotel.url.split <- strsplit(hotel.url, "Reviews-")
   url <- paste(hotel.url.split[[1]][1],"Reviews-or0",page.counter,"-",hotel.url.split[[1]][2], sep="")
   #url
  }
  else
  {
   hotel.url.split <- strsplit(hotel.url, "Reviews-")
   url <- paste(hotel.url.split[[1]][1],"Reviews-or",page.counter,"-",hotel.url.split[[1]][2], sep="")
   #url
  }

  reviews <- url %>%
   read_html() %>%
   html_nodes("#REVIEWS .innerBubble")

  if(length(reviews) > 0)
  {
```

```
    id <- reviews %>%
      html_node(".quote a") %>%
      html_attr("id")

    quote <- reviews %>%
      html_node(".quote span") %>%
      html_text()

    rating <- reviews %>%
      html_node(".rating .rating_s_fill") %>%
      html_attr("alt") %>%
      gsub(" of 5 stars", "", .) %>%
      as.integer()

    date <- reviews %>%
      html_node(".rating .ratingDate") %>%
      html_text()

    review <- reviews %>%
      html_node(".entry .partial_entry") %>%
      html_text()

    #hotel name
    name.h <- hotel.review.name[row.counter]
    #hotel url
    url.h <- hotel.url

    hotelfile <- data.frame(name.h,url.h, id, quote, rating, date, review, stringsAsFactors = FALSE)
        closing.file1 <- paste(closing.file,row.counter,".csv", sep="")

    if (page.counter == 0)
    {
      write.csv(hotelfile, file = closing.file1)
    }
        else
    {
      hotel <- read.csv (closing.file1,stringsAsFactors = FALSE)
      hotelWithoutX <- hotel[, -grep("X$", colnames(hotel))]
      bind <- rbind(hotelWithoutX,hotelfile)
      write.csv(bind, file = closing.file1)
    }

   }
   page.counter <- page.counter + 10
  }
  row.counter <- row.counter + 1

}
```

Appendix 5 – Facebook Comment Extraction

```
fb_oauth <- fbOAuth(app_id="1213645", app_secret="lGNLGNSADSGDet", extended_permissions =
TRUE)
save(fb_oauth, file="fb_oauth")
load("fb_oauth")
etoken <- "laegeralngaerlgralgakrlg"
#Canada
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook 1")
opening.file <- read.csv ("Facebook IDs Canada.csv",stringsAsFactors = FALSE)
country <- "Canada"

fb.row.counter <- 1
fb.max.counter <- NROW(opening.file$X.1)


while(fb.row.counter <= fb.max.counter)
{
 fb.id <- opening.file$fb.id[fb.row.counter]
 fb.name <- opening.file$fb.name[fb.row.counter]

 fb_page <- getPage(page=fb.id, token=etoken, n=1500)

 page.file.name <- paste(fb.name, " ", fb.id," ", "all pages Review.csv", sep="" )
 setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook Page Reviews 1")
 write.csv(fb_page, page.file.name)
 setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook Reviews 1")

 page.count <- 1
 max.count <- NROW(fb_page$from_id)


 while (page.count <= max.count)
 {
  tryCatch({
   post <- getPost(post=fb_page$id[page.count], n=2000, token=etoken)

    if(length(post$comments$created_time) == 0)
    {
     post.id <- post$post$from_id
     post.created_time <- post$post$created_time
     post.type <- post$post$type
     post.message <- post$post$message
     post.likes_count <- post$post$likes_count
     post.comments_count <- post$post$comments_count
     post.shares_count <- post$post$shares_count
     comments.id <- "No Comment"
     comments.created_time <- post$post$created_time
     comments.message <- "No Comment"
     comments.likes_count <- 0
```

```
      temp.post.nocomment <- data.frame(post.id, post.created_time, post.type, post.message,
post.likes_count, post.comments_count, post.shares_count, comments.id, comments.created_time,
comments.message, comments.likes_count)

      closing.file <- paste(fb.name, " ", fb.id," ", "Reviews"," ",page.count,".csv", sep="" )
      write.csv(temp.post.nocomment, closing.file)

      page.count <- page.count + 1
    }
    else
    {
      post.id <- post$post$from_id
      post.created_time <- post$post$created_time
      post.type <- post$post$type
      post.message <- post$post$message
      post.likes_count <- post$post$likes_count
      post.comments_count <- post$post$comments_count
      post.shares_count <- post$post$shares_count
      comments.id <- post$comments$from_id
      comments.created_time <- post$comments$created_time
      comments.message <- post$comments$message
      comments.likes_count <- post$comments$likes_count

      temp.post.commnet <- data.frame(post.id, post.created_time, post.type, post.message,
post.likes_count, post.comments_count, post.shares_count, comments.id,comments.created_time,
comments.message, comments.likes_count)

      closing.file <- paste(fb.name, " ", fb.id," ", "Reviews"," ",page.count,".csv", sep="" )
      write.csv(temp.post.commnet, closing.file)

      page.count <- page.count + 1
    }
  },error=function(e){})

  }
 page.count <- 1
 fb.row.counter <- fb.row.counter + 1
 }
```

Appendix 6 – Sentiment Analysis score

```
#these are the packages we need for this example - executing this line will install them

install.packages(c("devtools", "rjson", "bit64", "httr", "plyr", "ggplot2", "doBy", "XML", "base64enc",
"Rfacebook", "rvest" ))
install.packages(c("xlsx"))
install.packages( "stringr")
install.packages(c("plyr"))
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels")
#devtools allows us to install from github
library(devtools)

#these are various libraries that we use throughout this example
library(plyr)
library(httr)
library(doBy)
library(Rfacebook)
library(rvest)
library(ggplot2)
library(stringr)
library(xlsx)

######################
#File Clean up
######################
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/TripAdvisor Reviews")

all.the.files <- list.files()

all.the.data <- lapply( all.the.files,  read.csv, header=TRUE)

DATA <- do.call("rbind", all.the.data)

write.csv (DATA,"TripAdvisor Consolidated Canada.csv")

######################

setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook Reviews 1")

all.the.files <- list.files()

all.the.data <- lapply( all.the.files,  read.csv, header=TRUE)

DATA <- do.call("rbind", all.the.data)

write.csv (DATA,"Facebook Consolidated Reviews Canada.csv")
################
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/TripAdvisor Reviews")

hotelfile <- read.csv("TripAdvisor Consolidated Canada.csv", stringsAsFactors = FALSE)
hotelfileWithoutX <- hotelfile[, -grep("X$", colnames(hotelfile))]
hotelfileWithoutX.1 <- hotelfileWithoutX[, -grep("X.1$", colnames(hotelfileWithoutX))]
```

```
hotelfile <- hotelfileWithoutX.1

hotelfile$date <- gsub("Reviewed ", "",hotelfile$date)
hotelfile$date <- gsub("\n", "",hotelfile$date)
hotelfile$review <- gsub("\n", "",hotelfile$review)

write.csv(hotelfile,"TripAdvisor Review.csv")
hotel.reviews <- read.csv("TripAdvisor Review.csv", header=TRUE, sep=",")
########################################
#Based on http://analyzecore.com/2014/05/11/twitter-sentiment-analysis-based-on-affective-
lexicons-in-r/
########################################
#here we read in the two dictionaries that have been downloaded -
#hu.liu.pos = scan('C:/Users/Dropbox/MSc Data Analytics/Thesis/Lexicons/Opinion Lexicon/positive-
words.txt', what='character', comment.char=';')
#hu.liu.neg = scan('C:/Users/Dropbox/MSc Data Analytics/Thesis/Lexicons/Opinion
Lexicon/negative-words.txt', what='character', comment.char=';')

hu.liu.pos = scan('C:/Users/Dropbox/MSc Data Analytics/Thesis/Lexicons/Harvard General
Inquirer/positive-words.txt', what='character', comment.char=';')
hu.liu.neg = scan('C:/Users/Dropbox/MSc Data Analytics/Thesis/Lexicons/Harvard General
Inquirer/negative-words.txt', what='character', comment.char=';')



#our first function
score.sentence <- function(sentence, date, pos.words, neg.words) {
  #here some basic cleaning
  sentence = gsub('[[:punct:]]', '', sentence)
  sentence = gsub('[[:cntrl::]]', '', sentence)
  sentence = gsub('\\d+', '', sentence)
  sentence = tolower(sentence)

  #basic data structure construction
  word.list = str_split(sentence, '\\s+')
  words = unlist(word.list)

  #here we count the number of words that are positive and negative
  pos.matches = match(words, pos.words)
  neg.matches = match(words, neg.words)

  #throw away those that didn't match
  pos.matches = !is.na(pos.matches)
  neg.matches = !is.na(neg.matches)

  #compute the sentiment score
  score = sum(pos.matches) - sum(neg.matches)

  return(score)
}
```

```
#our second function that takes an array of sentences and sentiment analyses them
score.sentiment <- function(sentences, date, pos.words, neg.words) {
  require(plyr)
  require(stringr)

  #here any sentence/tweet that causes an error is given a sentiment score of 0 (neutral)
  scores = laply(sentences, function(sentence, date, pos.words, neg.words) {
    tryCatch(score.sentence(sentence, date, pos.words, neg.words ), error=function(e) 0)
  }, date, pos.words, neg.words)

  #now we construct a data frame
  scores.df = data.frame(score=scores, text=sentences, date=date)

  return(scores.df)
}

#our third function, that communicates with the file and then scores each of the reviews

collect.and.score <- function (hotelreview, date, pos.words, neg.words) {
  score = score.sentiment(hotelreview, date, pos.words, neg.words)

  return (score)
}

hotel.scores = collect.and.score(hotel.reviews$review, hotel.reviews$date, pos.words, neg.words)
headline.scores = collect.and.score(hotel.reviews$quote, hotel.reviews$date, pos.words, neg.words)
overall.scores.ol <- data.frame(headline.scores$date, headline.scores$score, headline.scores$text,
hotel.scores$score, hotel.scores$text)
#write.csv(overall.scores.ol, "Canidian hotel Review scores OL.csv")
write.csv(overall.scores.ol, "Canidian hotel Review scores HGI.csv")

################
#Facebook
###############
setwd("/Users/Dropbox/MSc Data Analytics/Thesis/Hotels/Canada/Facebook Reviews 1")
facebook.file <- read.csv("Facebook Consolidated Reviews Canada.csv", header=TRUE, sep=",")
#str(facebook.file)

facebook.fileWithoutX.1 <- facebook.file[, -grep("X.1$", colnames(facebook.file))]
#facebook.fileWithoutX.2 <- facebook.fileWithoutX.1[, -grep("X.2$",
colnames(facebook.fileWithoutX.1))]
facebook.fileWithoutX <- facebook.fileWithoutX.1[, -grep("X$", colnames(facebook.fileWithoutX.1))]
facebook.file <- facebook.fileWithoutX

facebook.file$post.message <- gsub("\n", "",facebook.file$post.message)
facebook.file$comments.message <- gsub("\n", "",facebook.file$comments.message)

facebookhotel.scores = collect.and.score(facebook.file$comments.message,
facebook.file$comments.created_time, pos.words, neg.words)
facebookhotel.scores.ol <- data.frame(facebook.file,facebookhotel.scores$score)
```

```
#write.csv(facebookhotel.scores.ol, "Canidian Facebook hotel Review scores OL.csv")
write.csv(facebookhotel.scores.ol, "Canidian Facebook hotel Review scores HGI.csv")
```

Appendix 7 – Regression code

```
install.packages(c("devtools", "rjson", "bit64", "httr", "plyr", "ggplot2", "doBy", "XML", "base64enc",
"Rfacebook" ))
install.packages("psych")

#http://stackoverflow.com/questions/35207624/package-pbkrtest-is-not-available-for-r-version-3-
2-2 (accessed 20/8/2016)
#to allow the package pbkrtest to be loaded as it's needed for car
install.packages("lme4")
library(lme4)
packageurl <- "https://cran.r-project.org/src/contrib/Archive/pbkrtest/pbkrtest_0.4-4.tar.gz"
install.packages(packageurl, repos=NULL, type="source")
library(pbkrtest)
install.packages("car")
####
library(plyr)
library(httr)
library(doBy)
library(Quandl)
library(twitteR)
library(Rfacebook)
library(devtools)
library(psych)
library(stats)
library(car)
#######
install.packages("gvlma")
library(gvlma)
#######
#The Linear Regression and Normalisation code is based on Lantz-Machine Learning

setwd("/Users/Dropbox/MSc Data Analytics/Thesis/UK Visitor data source/Data")

#Average Values - Time lag
time.lag <- "seven"
training.data.file <- paste("HGI UK Monthly Data Average time lag ",time.lag, " training.csv", sep="")
test.data.file <- paste("HGI UK Monthly Data Average time lag ", time.lag, " test.csv", sep ="")
results.file.1 <- paste("HGI R Plot Data Average time lag ",time.lag ,".jpeg", sep ="")
results.file.2 <- paste("HGI R Panals Data Average time lag ", time.lag,".jpeg", sep ="")
results.file.3 <- paste("HGI Data Average time lag ",time.lag,".txt", sep ="")
results.file.4 <- paste("HGI Corolation Average time lag ",time.lag,".txt", sep ="")
#Average Values - Time lag
#time.lag <- "seven"
#training.data.file <- paste("OL UK Monthly Data Average time lag ",time.lag, " training.csv", sep="")
#test.data.file <- paste("OL UK Monthly Data Average time lag ", time.lag, " test.csv", sep ="")
#results.file.1 <- paste("OL R Plot Data Average time lag ",time.lag ,".jpeg", sep ="")
#results.file.2 <- paste("OL R Panals Data Average time lag ", time.lag,".jpeg", sep ="")
#results.file.3 <- paste("OL Data Average time lag ",time.lag,".txt", sep ="")
#results.file.4 <- paste("OL Corolation Average time lag ",time.lag,".txt", sep ="")
#results.file.5 <- paste("OL Test Stats Average time lag ",time.lag,".txt", sep ="")
#results.file.6 <- paste("OL Residuals Plot Google Average time lag ", time.lag,".jpeg", sep ="")
```

```
results.file.7 <- paste("OL Residuals Plot TripAdvisor Average time lag ", time.lag,".jpeg", sep ="")
results.file.8 <- paste("OL Residuals Plot Facebook Average time lag ", time.lag,".jpeg", sep ="")
results.file.9 <- paste("OL component + residual plot Average time lag ", time.lag,".jpeg", sep ="")
results.file.10 <- paste("OL Ceres plots Average time lag ", time.lag,".jpeg", sep ="")

#Training Data
Can_Visitors <- read.csv(training.data.file, header = TRUE, sep = ",")

#Test/Prediction Data
Can_Visitors_p <- read.csv(test.data.file, header = TRUE, sep = ",")

#str(Can_Visitors)
#str(Can_Visitors_p)

summary(Can_Visitors$Total)

setwd("/Users/Dropbox/MSc Data Analytics/Thesis/UK Visitor data source/LR Results")

pairs(Can_Visitors[c("Visitors", "Google", "TripAdvisor", "Facebook", "Total")])
pairs.panels(Can_Visitors[c("Visitors", "Google", "TripAdvisor", "Facebook", "Total")])

jpeg(results.file.1)
pairs(Can_Visitors[c("Visitors", "Google", "TripAdvisor", "Facebook", "Total")])
dev.off()

jpeg(results.file.2)
pairs.panels(Can_Visitors[c("Visitors", "Google", "TripAdvisor", "Facebook", "Total")])
dev.off()

vis_modelsg <- lm(Visitors ~ Google, data = Can_Visitors)
vis_modelsg
summary(vis_modelsg)

vis_modelsg.res <- resid(vis_modelsg)
plot(Can_Visitors$Google, vis_modelsg.res, ylab="Residuals", xlab="Google", main="Googel Trends
Residuals")

jpeg(results.file.6)
plot(Can_Visitors$Google, vis_modelsg.res, ylab="Residuals", xlab="Google", main="Googel Trends
Residuals")
dev.off()

vis_modelst <- lm(Visitors ~ TripAdvisor, data = Can_Visitors)
vis_modelst
summary(vis_modelst)

vis_modelst.res <- resid(vis_modelst)
plot(Can_Visitors$TripAdvisor, vis_modelst.res, ylab="Residuals", xlab="TripAdvisor",
main="TripAdvisor Residuals")
```

```
jpeg(results.file.7)
plot(Can_Visitors$TripAdvisor, vis_modelst.res, ylab="Residuals", xlab="TripAdvisor",
main="TripAdvisor Residuals")
dev.off()

vis_modelsf <- lm(Visitors ~ Facebook, data = Can_Visitors)
vis_modelsf
summary(vis_modelsf)

vis_modelsf.res <- resid(vis_modelsf)
plot(Can_Visitors$Facebook, vis_modelsf.res, ylab="Residuals", xlab="Facebook", main="Facebook
Residuals")

jpeg(results.file.8)
plot(Can_Visitors$Facebook, vis_modelsf.res, ylab="Residuals", xlab="Facebook", main="Facebook
Residuals")
dev.off()

vis_model <- lm(Visitors ~ Google + TripAdvisor + Facebook, data = Can_Visitors)
vis_model
summary(vis_model)

#VIF
vif(vis_model) # variance inflation factors
#Test for Multi-collinearity
sqrt(vif(vis_model)) > 2

# Evaluate Nonlinearity
# component + residual plot
crPlots(vis_model)
# Ceres plots
ceresPlots(vis_model)

jpeg(results.file.9)
crPlots(vis_model)
dev.off()

jpeg(results.file.10)
ceresPlots(vis_model)
dev.off()


# Test for Autocorrelated Errors
durbinWatsonTest(vis_model)

# Global test of model assumptions
gvmodel <- gvlma(vis_model)
summary(gvmodel)

vis_model2 <- lm(Visitors ~ Total, data = Can_Visitors)
vis_model2
```

```
summary(vis_model2)

vis_model3 <- lm(Visitors ~ Google * TripAdvisor * Facebook, data = Can_Visitors)
vis_model3
summary(vis_model3)

#VIF
vif(vis_model) # variance inflation factors
#Test for Multi-collinearity
sqrt(vif(vis_model)) > 2

# Evaluate Nonlinearity
# component + residual plot
crPlots(vis_model)
# Ceres plots
ceresPlots(vis_model)

# Test for Autocorrelated Errors
durbinWatsonTest(vis_model)

# Global test of model assumptions
gvmodel <- gvlma(vis_model)
summary(gvmodel)

sink(results.file.5)
"VIF"
vif(vis_model) # variance inflation factors
"Test for Multi-collinearity"
sqrt(vif(vis_model)) > 2
"Test for Autocorrelated Errors"
durbinWatsonTest(vis_model)
"Global Summary"
summary(gvmodel)
sink()



sink(results.file.3)
summary(vis_modelsg)
summary(vis_modelst)
summary(vis_modelsf)
summary(vis_model)
summary(vis_model2)
summary(vis_model3)
sink()


vis_predsg <- predict(vis_modelsg,Can_Visitors_p)
one <- cor(vis_predsg,Can_Visitors_p$Visitors)

vis_predst <- predict(vis_modelst,Can_Visitors_p)
```

```
two <- cor(vis_predst,Can_Visitors_p$Visitors)

vis_predsf <- predict(vis_modelsf,Can_Visitors_p)
three <- cor(vis_predsf,Can_Visitors_p$Visitors)

vis_pred <- predict(vis_model,Can_Visitors_p)
four <- cor(vis_pred,Can_Visitors_p$Visitors)

vis_pred2 <- predict(vis_model2,Can_Visitors_p)
five <- cor(vis_pred2,Can_Visitors_p$Visitors)

vis_pred3 <- predict(vis_model3,Can_Visitors_p)
six <- cor(vis_pred3,Can_Visitors_p$Visitors)

corrolation <- data.frame(one,two,three,four,five,six,stringsAsFactors = FALSE)

sink(results.file.4)
"All"
corrolation
"Google"
cor(vis_predsg,Can_Visitors_p$Visitors)
"TripAdvisor"
cor(vis_predst,Can_Visitors_p$Visitors)
"Facebook"
cor(vis_predsf,Can_Visitors_p$Visitors)
"Google, TripAdvisor & Facebook"
cor(vis_pred,Can_Visitors_p$Visitors)
"Total"
cor(vis_pred2,Can_Visitors_p$Visitors)
"Combined"
cor(vis_pred3,Can_Visitors_p$Visitors)
sink()

setwd("/Users/Dropbox/MSc Data Analytics/Thesis/UK Visitor data source/Data")
```

Appendix 8 - Canadian Visitor Data Sources

http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_1_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_2_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_3_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_4_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_5_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_6_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_7_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_8_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_9_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_10_en.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_11_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2010_12_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2011_01_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/TourismSnapshot_2011_02_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_03_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_04_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_05_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_06_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_07_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_08.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_09_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_10_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_11_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2011_12_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_01.pdf

http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_02_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_03_eng.pdf
http://en.destinationcanada.com/sites/default/files/content/homepage/homepage_slider/tourismsnapshot_2012_04_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_05_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_06_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_07_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_08_eng.pdf
en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_09_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_10_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_11_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2012_12_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_01_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_02_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_03_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_04_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_05_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_06_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_07_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_08_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_09_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_10_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_11_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2013_12_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_01_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_02_eng.pdf

http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_03_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_04_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_05_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_06_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_07_eng_ver2.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_08_eng_lowres.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_09_eng_lowres.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_10_eng_lowres.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_11_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2014_12_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_01_eng_lowres.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_02_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_03_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_04_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_05_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_06_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_07_v2_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_08_eng_lowres.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_09_eng.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_10_eng_lowres.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot_2015_11_eng_lowres.pdf
http://en.destinationcanada.com/sites/default/files/pdf/Research/Stats-figures/International-visitor-arrivals/Tourism-monthly-snapshot/tourismsnapshot-dec2015_en-lowres.pdf

Appendix 9 – UK Visitor Data Sources

2010 -
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-269251

2011 - 2012

http://webarchive.nationalarchives.gov.uk/20160105160709/http://ons.gov.uk/ons/rel/ott/overseas-travel-and-tourism---monthly-release/march-2014/tsd--march-2014.html

2013 - 2015

http://webarchive.nationalarchives.gov.uk/20160105160709/http://ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-391944

Jun-15

"http://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/overseastravelandtourism/2015-08-21"

Jul-15

"http://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/overseastravelandtourism/2015-09-18"

Aug-15

"http://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/overseastravelandtourism/2015-10-23"

Sep-15

"http://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/overseastravelandtourism/provisionalresultsforseptember2015"

Oct-15

http://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/overseastravelandtourism/provisionalresultsforoctober2015

Nov-15

http://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/overseastravelandtourism/provisionalresultsfornovember2015

Dec-15

http://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/bulletins/overseastravelandtourism/provisionalresultsfordecember2015