In this part i have read the data from the csv file, which i have made with the help of web scraping this include the content of the post people made on the website of tripadvisor.

```python
In [77]: import pandas as pd
         t = pd.read_csv('replbodyCont.csv')
         t
```

Out[77]:

	selection1_reply_no_content_name
0	I would recommend going through immigration & ...
1	Aseem01: arriving at midnight in (ta && ta.que...
2	From the FAQ section [HKIA Skypier to (ta && t...
3	12 am is mid night, not early morning. With th...
4	-:- Message from TripAdvisor staff -:-TripAdvi...
5	>> Don't ferries from Skypier run all night? <...
6	-:- Message from TripAdvisor staff -:-This top...
7	Can any expert advise the best & cheap option ...
8	Take a direct bus from the airport.
9	A direct bus from HKIA to Shenzhen costs proba...
10	Thanks folks for the response.I will likely ch...
11	It's cheaper. But whether it's more "economica...
12	Everybody's time-money conversion factors are ...
13	Compare the total costs and convenience factor...
14	I had booked a special SIN- HK fare with Jetst...
15	Good for you but I saw Jetstar return fare fro...
16	Sounds like Air (ta && ta.queueForLoad ? ta.qu...
17	OP specifically stated flying Jetstar, not Air...
18	Hi,I'll be travelling with my parents, who qua...
19	They can purchase the Elder Octopus Card which...
20	I think you do have to prove your age with som...
21	I am a senior (66) and used my seniors card fr...
22	I was all prepared with my Australian Seniors ...
23	There aren't any conditions of purchase for Oc...
24	Someone at the info desk at the HK airport hel...
25	Ok great! Can we get a refund for our Octopus ...
26	You can get a refund at the airport by handing...
27	-:- Message from TripAdvisor staff -:-This top...
28	hi there - we are traveling from Auckland to P...
29	Hi Curtis..the price of iDevices can be obtain...
...	...
69549	You could be asked by (ta && ta.queueForLoad ?...
69550	Yes that was what I meant, sorry should have c...
69551	-:- Message from TripAdvisor staff -:-This top...
69552	HelloAm planning to travel to HK and (ta && ta...

selection1_reply_no_content_name

69553	It's difficult to make suggestions without kno...
69554	Welcome to HK. Suggest invest in a travel book...
69555	Check the links to the right, things to see/do...
69556	See Must See/Do in (ta && ta.queueForLoad ? ta...
69557	-:- Message from TripAdvisor staff -:-This top...
69558	As there are 5 of us arriving at HK airport we...
69559	What is their price? Usually two taxis are che...
69560	I knew some driver working free-lance for both...
69561	They quote $US80.I have to give cc details to ...
69562	Tipping is unusual for pre-arranged drivers in...
69563	Thanks Cookie2011We are happy with the price, ...
69564	Yes, price is good! You need to consider that ...
69565	PM me if you want contact of the one I use. Ve...
69566	Two taxis are cheaper. To which part/hotel in ...
69567	-:- Message from TripAdvisor staff -:-This top...
69568	We arrive in HK at 6.20 am, and then want to f...
69569	If your luggage is checked through to HCM, the...
69570	As usual, the best answers will probably emerg...
69571	I fly this route regularly and it is enough ti...
69572	To correct myself, I fly through HK regularly....
69573	Thanks to all. I will post in (ta && ta.queueF...
69574	Richard - I booked Cathay from Melbourne throu...
69575	MazI've never had only a one hour transfer, wh...
69576	Me too Richard- I've had plenty of 2 hour and ...
69577	I've had many one-hour transits at (ta && ta.q...
69578	Airlines always claim that 45 minutes is enoug...

69579 rows × 1 columns

```
In [ ]:  #removing duplicate
         t.drop_duplicates(subset='selection1_reply_no_content_name',keep = 'first', in
         place = True)
```

The data which is posted on the website are not all pure, that is there is chances that the data posted on the
website is duplicate of someone's earlier post, So, here i have used pandas to remove all the duplicate poste
from the scraped data. And printed pure the data as shown below.

In [79]: t

Out[79]:

	selection1_reply_no_content_name
0	I would recommend going through immigration & ...
1	Aseem01: arriving at midnight in (ta && ta.que...
2	From the FAQ section [HKIA Skypier to (ta && t...
3	12 am is mid night, not early morning. With th...
4	-:- Message from TripAdvisor staff -:-TripAdvi...
5	>> Don't ferries from Skypier run all night? <...
6	-:- Message from TripAdvisor staff -:-This top...
7	Can any expert advise the best & cheap option ...
8	Take a direct bus from the airport.
9	A direct bus from HKIA to Shenzhen costs proba...
10	Thanks folks for the response.I will likely ch...
11	It's cheaper. But whether it's more "economica...
12	Everybody's time-money conversion factors are ...
13	Compare the total costs and convenience factor...
14	I had booked a special SIN- HK fare with Jetst...
15	Good for you but I saw Jetstar return fare fro...
16	Sounds like Air (ta && ta.queueForLoad ? ta.qu...
17	OP specifically stated flying Jetstar, not Air...
18	Hi,I'll be travelling with my parents, who qua...
19	They can purchase the Elder Octopus Card which...
20	I think you do have to prove your age with som...
21	I am a senior (66) and used my seniors card fr...
22	I was all prepared with my Australian Seniors ...
23	There aren't any conditions of purchase for Oc...
24	Someone at the info desk at the HK airport hel...
25	Ok great! Can we get a refund for our Octopus ...
26	You can get a refund at the airport by handing...
28	hi there - we are traveling from Auckland to P...
29	Hi Curtis..the price of iDevices can be obtain...
30	There's also websites that compares apple pric...
...	...
69546	I may be wrong but I think it should be okay. ...
69547	G'day,but you will be carrying your passport o...
69548	The last time I bought Octopus cards at HK air...
69549	You could be asked by (ta && ta.queueForLoad ?...

	selection1_reply_no_content_name
69550	Yes that was what I meant, sorry should have c...
69552	HelloAm planning to travel to HK and (ta && ta...
69553	It's difficult to make suggestions without kno...
69554	Welcome to HK. Suggest invest in a travel book...
69555	Check the links to the right, things to see/do...
69556	See Must See/Do in (ta && ta.queueForLoad ? ta...
69558	As there are 5 of us arriving at HK airport we...
69559	What is their price? Usually two taxis are che...
69560	I knew some driver working free-lance for both...
69561	They quote $US80.I have to give cc details to ...
69562	Tipping is unusual for pre-arranged drivers in...
69563	Thanks Cookie2011We are happy with the price, ...
69564	Yes, price is good! You need to consider that ...
69565	PM me if you want contact of the one I use. Ve...
69566	Two taxis are cheaper. To which part/hotel in ...
69568	We arrive in HK at 6.20 am, and then want to f...
69569	If your luggage is checked through to HCM, the...
69570	As usual, the best answers will probably emerg...
69571	I fly this route regularly and it is enough ti...
69572	To correct myself, I fly through HK regularly....
69573	Thanks to all. I will post in (ta && ta.queueF...
69574	Richard - I booked Cathay from Melbourne throu...
69575	MazI've never had only a one hour transfer, wh...
69576	Me too Richard- I've had plenty of 2 hour and ...
69577	I've had many one-hour transits at (ta && ta.q...
69578	Airlines always claim that 45 minutes is enoug...

6796 rows × 1 columns

Now its time to cleane the data, cleaning means removing the links and ads from the post because there are lots of people who posts their add on the website, so here i have removed the all the links and non word character. to do this i have used natural language tool kit(nltk).

```
In [6]:  import nltk
         from nltk.corpus import stopwords
         # t['selection1_reply_no_content_name'][0:2]
```

In [90]:

```python
#stop words removing
# stop_words = set(stopwords.words('english'))
# for w in t['selection1_reply_no_content_name'][:5]:
#     if

result2 = []
for i in range(0, len(t['selection1_reply_no_content_name'])):
    try:
        txt = ' '.join(word for word in t['selection1_reply_no_content_name'][
i].split() if not word.startswith(("http",'(')) )   #removes link
        result2.append(txt)
    except:
        pass
a = ' '.join(word for word in result2)
result3 = ''
for k in a.split("\n"):
    result3 += re.sub(r"[^a-zA-Z0-9]+", ' ', k)

result3
# result2
```

Out[90]: 'I would recommend going through immigration getting to the ta queueForLoad t a queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 38315920 log autolink impression Macau Ferry Terminal taxi or night bus to take the next available ferry to Macau so you can get some sleep Much as HKIA is a great airport there is nothing whatsoever to do in the Skypier departure area you will have no access to the other airport areas including restaurants Aseem01 arriving at midnight in ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 38324674 log autolink impression Hong Kong and proceeding to Macau is a bit hard Have you planned budgeted one night of hotel on your arrival date 9 to 10 hours is a long time sitting on hard shell seats While you may proceed as zhuhai2007 suggested in 1 go through the immigration in Hong Kong and proceed to go to Macau likely by the time you reach Macau Ferry Terminal clear Macau immigration custom then it is 4 am Now what You may go to the hotel If your reservation is for the following day you would be sitting in their lobby area of the Hong Kong Airport for 8 hours for your room to become available From the FAQ section HKIA Skypier to ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 38327519 log autolink impression Macau read the thread from earlier this year about arriving at 1am need to travel directly to Macau tripadvisor com ShowTopic g664891 i12409 k50 12 am is mid night not early morning With the one year old I really think it s best to find a hotel in HK close to the airport and just bed down for the night Message from TripAdvisor staff TripAdvisor staff removed this post at the original author s request To review the TripAdvisor Forums Posting Guidelines please follow this link remo Don t ferries from Skypier run all night No Caspar Esq Not from Skypier First ferry 10 am last ferry 10 pmwww turbojet com hk eng schedule airport aspThe ferries from ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 38368609 log autolink impression Macau Ferry Terminal Sheung Wan to Macau do run all night Message from TripAdvisor staff This topic has been closed to new posts due to inactivity We hope you ll join the conversation by posting to an open topic or starting a new one To review the TripAdvisor Forums Posting Guidelines please follow this Can any expert advise the best cheap option to reach ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 36569763 log autolink impression Shenzhen from HK airport I am planning a trip next yr I have been reading that Bus 43 does stop at a transit railway stop for change of train to Shenzhen Please advise Take a direct bus from the airport A direct bus from HKIA to Shenzhen costs probably over HKD 150 The question was the cheapest mode It may take a little longer Take Airport bus A43 from HKIA towards Luen Wo Hui Alight at Stop 3 Sheung Shui Railroad Station Take ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 36570240 log autolink impression MTR for 1 stop towards Lo Wu Lok Ma Chau if you choose to Then walk over to Shenzhen border control A43 HKD 30 9MTR from Sheung Shui to Lo Wu HKD 21 0 Thanks folks for the response I will likely choose the local bus A43 option since it is more economical that way It s cheaper But whether it s more economical depends on how you value your time Everybody s time money conversion factors are different OP should know what works best for him her Taking a bus directly from the HKIA to ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 36601737 log autolink impression Shenzhen may not necessarily be faster It depends on the actual destination in Shenzhen At any rate taking a bus to Sheung Shui and change to MTR train should not take that much

longer than the direct HKIA Shenzhen bus You may want to cross the border at Lok Ma Chau instead of the more popular Lo Wu Then taxi bus metro in Shenzhen to your destination Compare the total costs and convenience factor by booking a flight directly to ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 36604320 log autolink impression Shenzhen vs via HKG You can fly China Eastern from SIN to SZX PVG for about HK 2 430 return in economy Fare as post at MU s web site I had booked a special SIN HK fare with Jetstar for S 230 USD 176 return No commercial airlines can beat that fare Edited 24 October 2011 21 54 Good for you but I saw Jetstar return fare from SIN to ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 36605457 log autolink impression HKG currently selling for as low as SGD 88 all in for mid February 2012 For those who do not know Jetstar is a no frill budget airline with 28 29 seat pitch on their single aisle Airbus aeroplanes Everything else checked bag and meal is extra except they have not yet start charging for using the lavatory onboard Edited 24 October 2011 22 11 Sounds like Air ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 36606250 log autolink impression Asia but does Jetstar manage to fly their schedule That would be unlike Air Asia OP specifically stated flying Jetstar not Air ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 36606937 log autolink impression Asia As the old saying goes you usually get what you have paid Hi I ll be travelling with my parents who qualify as senior citizens as they are above 65 yrs old Noted that there are senior citizen rates for travel on the trams etc Is there a Senior Citizen Octopus Card that they can apply for If not does that mean that everytime they go up the tram they have to show their passport and pay cash how to go about qualifying for the senior citizen fare Is there also senior citizen fares for travelling on the MTR Thanks They can purchase the Elder Octopus Card which is available to tourists as well as HK residents qualifying age is 65 The card will automatically give them concessionary fares where available usually half adult fare Additionally on Wednesdays and Saturdays until 31st August 2011 subject to review Elder Octopus Card holders are charged a maximum fare of HK 2 per journey on the ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 33442400 log autolink impression MTR not journeys to the mainland border stations of Lo Wu and Lok Ma Chau and until 31st January 2012 a maximum fare of HK 2 or 50 of the adult fare whichever is the lower on most bus journeys The Elder Octopus costs HK 70 of which HK 50 is a refundable deposit a small handling fee octopus com hk get your octopus index htmlI don t believe any proof of age is required for purchase For details of other discounts available to senior visitors to HK see hongkongextras com senior visitor discounts Edited 24 May 2011 02 39 I think you do have to prove your age with some kind of official ID when you purchase the senior card I am a senior and used my seniors card from Australia and my passport Got discoyunts lots of places I had no local id or card at all ta ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 33450793 log autolink impression Disneyland was the biggest saving Always ask for discount and tell them yiou are a senior anywhere just tell them Albon I was all prepared with my Australian Seniors card and passport when I went to buy my octopus card but they didn t even ask for it They took one look at me and gave me the Seniors Card There aren t any conditions of purchase for Octopus Cards which is why proof of entitlement isn t requested to purchase one You can also buy a Child Octopus for a child without the child being present and without any evidence of age There

are however conditions of use for Octopus Cards and if you use an Octopus Card for which you are not elegible in order to gain a concessionary fare its a breach of conditions of use and you risk being fined Bus drivers ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 33459254 log autolink impression MTR staff etc know when a concessionary Octopus is being used by the tone it sounds when swiped and by other means Someone at the info desk at the HK airport helped us get the Octopus cards They took my word that I was a senior You get on the ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 33461109 log autolink impression star Ferry free by showing your card at the gate It worked just great Ok great Can we get a refund for our Octopus cards at the HK departure airport If yes where can we find the refund counter Thanks You can get a refund at the airport by handing in your card at the ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 33469537 log autolink impression Airport Express Customer Service Desks There are two in Terminal 1 Arrivals Hall known as Meeters and Greeters Hall one in area A of the hall and one in area B of the hall and there is one in Terminal 2 on the Airport Express arrival platform Edited 25 May 2011 04 33 hi there we are traveling from Auckland to Paris via hong kong and we have a 3 hour stop over can you acess duty free areas as we would like to have a look at iphones ipad etc Is there a website we can check out hong kong prices on line cheers and look forward to your feedback and advice Curtis Hi Curtis the price of iDevices can be obtained by visiting Apple HK store There are two Apple resellers at HKIA DG Lifestyle and Nuance Watson HK is tax and duty free in this respect so price is same in airport duty free or at the Apple store at ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 40458774 log autolink impression IFC mall or downtown Apple resellers stores Even wine is tax free There s also websites that compares apple prices internationally such as www geekzone co nz forums asp Message from TripAdvisor staff TripAdvisor staff removed this post because it did not meet TripAdvisor s forum guidelines with regards to the participation of business representatives To review the TripAdvisor Forums Posting Guidelines please foll Anyway to answer your basic question Yes you will be able to access a range of shops in the transit area at HKIA A listing can be found on hongkongairport com eng shopping index htmlAny shop listed as being inside the restricted area is available to you without haveing to go through immigration Other shops are in the terminal pre immigration departure areas to access those you will have to go through immigration and then do your shopping and re enter through security 3 hours is enough to accomplish this if you really want to New Zealanders have visa free access so the only cost of doing this is your time spent in lines appreciate your help tax free wine sounds like a bonus to me thank you i will check out the site later on today appreciate your time to get back to me will check that site thank you you are a fountain of knowledge and I will pass this on to the family it sounds like it will be a busy 3 hours for us thank you By the way goods may be subject to tax in France If HK all goods are tax free for sure wine is tax free too I think it was referring to wine being duty free in HK since a while guess 2002 Anyway wine will be much cheaper in France Shops selling ipads etc can be found on both landside and airside Hi Please can anyone offer any advice on the best way to get to Wanchai from HK airport Are there buses trains that can take you there Or is it best to get a taxi I don t know how far the journey is but don t want to pay a fortune in taxi fares Many thanks Taxi from the airport costs about HK 280 plus HK 5 for each item of luggage in the boot Wanchai is also served by Cityflyer airport bus A11 which costs HK 40 and takes about 60 minutes You can find det

ails of which hotels are close to airport bus routes route details and nearest bus stop here nwstbus com hk routes You can also take ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 33993605 log autolink impression Airport Express train to Hong Kong Station mins then taxi HK 60 to Wanchai or some hotels in Wancahi are served by free Airport Express Shuttle Buses H1 and H2 from Hong Kong Station mtr com hk eng complom free bus htmlAirport Express single adult fare to Hong Kong Station is HK100 but if 2 or more people are travelling together discounted Group Tickets are available mtr com hk eng promotions 4persons html Should also mention that the list of hotels close to A11 bus route at the link in the above post isn t comprehensive If the place where you are staying isn t on the list suggest you advise exactly where you are staying for more specific advice We are flying in from the US and have 5 5 hour transit at HK airport Coming in at night and flying out just after midnight Any suggestions on where to freshen up and rest for a bit Lounge or hotel My only concern with the lounge is that my boys will be too noisy for other passengers You may want to try the pay in lounge email them to see if kids are welcome I don t see why not hongkongairport com eng passenger arrival t1 hello can you help me we would like to ride a train from shenzhen to hongkong our flight is 7pm how many hours does it take to hongkong from mtr terminal how many minutes to hk airport if we ride a taxi and lastly how much do we have to pay for the taxi thank you so much There is no direct train from the SZ border to HKIA There is an excellent journey planner on the ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 38316268 log autolink impression MTR website which shows all the interchanges It gives minimum journey time of 81 minutes I think it would be better to take a bus It all depends where you re going to in ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 38406495 log autolink impression Shenzhen but I usually take the Airport bus A43 to Shueng Shui MTR and then take the MTR to Lo Wu Since I usually stay in the Lo Wu district I walk across the border and then take the SZ metro As a border crossing I ve always found Lo Wu to be much less stressful than Huanggang Should have done that the other way around I normally allow about 1 5 hours to get from Lo Wu to the Airport and vice versa heard they are famous for their wife s biscuits and other chinese pastries only been to the one at causebay bay was wondering if they have any closer to mongkok anyone knows www kee wah com hkFlat 3 On G F Portion of the Open Yard Nos 25 27 29 31 Soy St KLN Tel 2770 1779business hour 0800 220010 Mong Kok East ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 13635527 log autolink impression MTR Station Tel 2628 9328business hour 0730 2200They even have 2 shops at HK airport hi I wonder for a family of 6 paxs what is the cheapest and best transport to use from HK airport to ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 14189570 log autolink impression Hart Avenue Tsimshatsui Kowloon I heard of the city buses and some friends mention take a cab is better for a big group thanks for the best view of the city take the airport bus only downside is if you have plenty of luggage and not everyone can help carry them taxi will be pretty expensive since you will have to take 2 Taxi only has seating for 5 and since you have six you would have to get 2 taxis although one taxi is pretty roomy You could also do a private charter or get a bus Another suggestion could be ta queueForLoad ta queueForLoad function f g document addEventListener DOMContentLoaded f function ta trackEventOnPage postLinkInline impression postLinks 14195832 log autolink impression Airport Express and the good thing about that is that kids rate is discounted and that they can check yo

u in early so no worries about check in at the airport A21 airport bus is com
fortable takes an hour has good views of the city and costs HK 33 per adult I
t will drop you 5 minutes walk from ta queueForLoad ta queueForLoad function
f g document addEventListener DOMContentLoaded f function ta trackEventOnPage
postLinkInline impression postLinks 14195885 log autolink impression Hart Ave
nue It has plenty of room for luggage on the lower deck if you have lots of l
uggage and you are considering 2 cabs also consider private transfers the cos
t is more or less the same as a cab ride or about 300 hkd to tst let me know
so i could send you a list of people i have used your hotel may have a bus th
at goes direct the eatons cost 120 hk each 1 way oct 24 2012 i took a ferry f
rom hk airport to shekou i paid 130hkd for the ferry ticket then a friend of
mine visited me here nov 8 2012 he took the same route but they charge him 22
0hkd i dont understand why fare is not fair its unfair D I thought the fare w
as about 280 RMB It is ridiculously expensive as it is quite a bit more than
the fare to Zhuhai When you go the other way you at least get the departure t
ax refund of 120 130 is the first class child price www xunlongferry com en l
ine info4 asp id 20 I will arrive with my family including a kid year old at
about 12 AM at HK airport However our ferry is from 0940 from the Skypier Are
there enough facilities at the Skypier to spend as many as 9 hours specially
with a kid '

after removing all the links from the post, the next task i have to do is to remove the stop words(is , there, where, how, are...)the word whish has no sense. and also i have to tockenize the sentence of the post. tocanization means breaking the sentens of the all the post into small tocken. tocken will help in the natural language processing.

In [61]:
```python
#removing stop words and tockenization
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
stop_words = set(stopwords.words('english'))

word_tokens = word_tokenize(result3)

filtered_sentence = [w for w in word_tokens if not w in stop_words]

filtered_sentence = []

for w in word_tokens:
    if w not in stop_words:
        filtered_sentence.append(w)

# print(word_tokens)
print(filtered_sentence)
```

['I', 'would', 'recommend', 'going', 'immigration', 'getting', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '38315920', 'log', 'autolink', 'impression', 'Macau', 'Ferry', 'Terminal', 'taxi', 'night', 'bus', 'take', 'next', 'available', 'ferry', 'Macau', 'get', 'sleep', 'Much', 'HKIA', 'great', 'airport', 'nothing', 'whatsoever', 'Skypier', 'departure', 'area', 'access', 'airport', 'areas', 'including', 'restaurants', 'Aseem01', 'arriving', 'midnight', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '38324674', 'log', 'autolink', 'impression', 'Hong', 'Kong', 'proceeding', 'Macau', 'bit', 'hard', 'Have', 'planned', 'budgeted', 'one', 'night', 'hotel', 'arrival', 'date', '9', '10', 'hours', 'long', 'time', 'sitting', 'hard', 'shell', 'seats', 'While', 'may', 'proceed', 'zhuhai2007', 'suggested', '1', 'go', 'immigration', 'Hong', 'Kong', 'proceed', 'go', 'Macau', 'likely', 'time', 'reach', 'Macau', 'Ferry', 'Terminal', 'clear', 'Macau', 'immigration', 'custom', '4', 'Now', 'You', 'may', 'go', 'hotel', 'If', 'reservation', 'following', 'day', 'would', 'sitting', 'lobby', 'area', 'Hong', 'Kong', 'Airport', '8', 'hours', 'room', 'become', 'available', 'From', 'FAQ', 'section', 'HKIA', 'Skypier', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '38327519', 'log', 'autolink', 'impression', 'Macau', 'read', 'thread', 'earlier', 'year', 'arriving', '1am', 'need', 'travel', 'directly', 'Macau', 'tripadvisor', 'com', 'ShowTopic', 'g664891', 'i12409', 'k50', '12', 'mid', 'night', 'early', 'morning', 'With', 'one', 'year', 'old', 'I', 'really', 'think', 'best', 'find', 'hotel', 'HK', 'close', 'airport', 'bed', 'night', 'Message', 'TripAdvisor', 'staff', 'TripAdvisor', 'staff', 'removed', 'post', 'original', 'author', 'request', 'To', 'review', 'TripAdvisor', 'Forums', 'Posting', 'Guidelines', 'please', 'follow', 'link', 'remo', 'Don', 'ferries', 'Skypier', 'run', 'night', 'No', 'Caspar', 'Esq', 'Not', 'Skypier', 'First', 'ferry', '10', 'last', 'ferry', '10', 'pmwww', 'turbojet', 'com', 'hk', 'eng', 'schedule', 'airport', 'aspThe', 'ferries', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '38368609', 'log', 'autolink', 'impression', 'Macau', 'Ferry', 'Terminal', 'Sheung', 'Wan', 'Macau', 'run', 'night', 'Message', 'TripAdvisor', 'staff', 'This', 'topic', 'closed', 'new', 'posts', 'due', 'inactivity', 'We', 'hope', 'join', 'conversation', 'posting', 'open', 'topic', 'starting', 'new', 'one', 'To', 'review', 'TripAdvisor', 'Forums', 'Posting', 'Guidelines', 'please', 'follow', 'Can', 'expert', 'advise', 'best', 'cheap', 'option', 'reach', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '36569763', 'log', 'autolink', 'impression', 'Shenzhen', 'HK', 'airport', 'I', 'planning', 'trip', 'next', 'yr', 'I', 'reading', 'Bus', '43', 'stop', 'transit', 'railway', 'stop', 'change', 'train', 'Shenzhen', 'Please', 'advise', 'Take', 'direct', 'bus', 'airport', 'A', 'direct', 'bus', 'HKIA', 'Shenzhen', 'costs', 'probably', 'HKD', '150', 'The', 'question', 'cheapest', 'mode', 'It', 'may', 'take', 'little', 'longer', 'Take', 'Airport', 'bus', 'A43', 'HKIA', 'towards', 'Luen', 'Wo', 'Hui', 'Alight', 'Stop', '3', 'Sheung', 'Shui', 'Railroad', 'Station', 'Take', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '36570240', 'log', 'autolink', 'impression', 'MTR', '1', 'stop', 'towards', 'Lo', 'Wu', 'Lok', 'Ma', 'Chau', 'choose', 'Then', 'walk', 'Shenzhen', 'border', 'control', 'A43', 'HKD', '30', '9

MTR', 'Sheung', 'Shui', 'Lo', 'Wu', 'HKD', '21', '0', 'Thanks', 'folks', 'response', 'I', 'likely', 'choose', 'local', 'bus', 'A43', 'option', 'since', 'economical', 'way', 'It', 'cheaper', 'But', 'whether', 'economical', 'depends', 'value', 'time', 'Everybody', 'time', 'money', 'conversion', 'factors', 'different', 'OP', 'know', 'works', 'best', 'Taking', 'bus', 'directly', 'HKIA', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '36601737', 'log', 'autolink', 'impression', 'Shenzhen', 'may', 'necessarily', 'faster', 'It', 'depends', 'actual', 'destination', 'Shenzhen', 'At', 'rate', 'taking', 'bus', 'Sheung', 'Shui', 'change', 'MTR', 'train', 'take', 'much', 'longer', 'direct', 'HKIA', 'Shenzhen', 'bus', 'You', 'may', 'want', 'cross', 'border', 'Lok', 'Ma', 'Chau', 'instead', 'popular', 'Lo', 'Wu', 'Then', 'taxi', 'bus', 'metro', 'Shenzhen', 'destination', 'Compare', 'total', 'costs', 'convenience', 'factor', 'booking', 'flight', 'directly', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '36604320', 'log', 'autolink', 'impression', 'Shenzhen', 'vs', 'via', 'HKG', 'You', 'fly', 'China', 'Eastern', 'SIN', 'SZX', 'PVG', 'HK', '2', '430', 'return', 'economy', 'Fare', 'post', 'MU', 'web', 'site', 'I', 'booked', 'special', 'SIN', 'HK', 'fare', 'Jetstar', 'S', '230', 'USD', '176', 'return', 'No', 'commercial', 'airlines', 'beat', 'fare', 'Edited', '24', 'October', '2011', '21', '54', 'Good', 'I', 'saw', 'Jetstar', 'return', 'fare', 'SIN', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '36605457', 'log', 'autolink', 'impression', 'HKG', 'currently', 'selling', 'low', 'SGD', '88', 'mid', 'February', '2012', 'For', 'know', 'Jetstar', 'frill', 'budget', 'airline', '28', '29', 'seat', 'pitch', 'single', 'aisle', 'Airbus', 'aeroplanes', 'Everything', 'else', 'checked', 'bag', 'meal', 'extra', 'except', 'yet', 'start', 'charging', 'using', 'lavatory', 'onboard', 'Edited', '24', 'October', '2011', '22', '11', 'Sounds', 'like', 'Air', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '36606250', 'log', 'autolink', 'impression', 'Asia', 'Jetstar', 'manage', 'fly', 'schedule', 'That', 'would', 'unlike', 'Air', 'Asia', 'OP', 'specifically', 'stated', 'flying', 'Jetstar', 'Air', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '36606937', 'log', 'autolink', 'impression', 'Asia', 'As', 'old', 'saying', 'goes', 'usually', 'get', 'paid', 'Hi', 'I', 'travelling', 'parents', 'qualify', 'senior', 'citizens', '65', 'yrs', 'old', 'Noted', 'senior', 'citizen', 'rates', 'travel', 'trams', 'etc', 'Is', 'Senior', 'Citizen', 'Octopus', 'Card', 'apply', 'If', 'mean', 'everytime', 'go', 'tram', 'show', 'passport', 'pay', 'cash', 'go', 'qualifying', 'senior', 'citizen', 'fare', 'Is', 'also', 'senior', 'citizen', 'fares', 'travelling', 'MTR', 'Thanks', 'They', 'purchase', 'Elder', 'Octopus', 'Card', 'available', 'tourists', 'well', 'HK', 'residents', 'qualifying', 'age', '65', 'The', 'card', 'automatically', 'give', 'concessionary', 'fares', 'available', 'usually', 'half', 'adult', 'fare', 'Additionally', 'Wednesdays', 'Saturdays', '31st', 'August', '2011', 'subject', 'review', 'Elder', 'Octopus', 'Card', 'holders', 'charged', 'maximum', 'fare', 'HK', '2', 'per', 'journey', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '33442400', 'log', 'autolink', 'impression', 'MTR', 'journeys', 'mainland', 'border', 'stations', 'Lo', 'Wu', 'Lok', 'Ma', 'Chau', '31st', 'January', '2012', 'maximum',

'fare', 'HK', '2', '50', 'adult', 'fare', 'whichever', 'lower', 'bus', 'journ
eys', 'The', 'Elder', 'Octopus', 'costs', 'HK', '70', 'HK', '50', 'refundabl
e', 'deposit', 'small', 'handling', 'fee', 'octopus', 'com', 'hk', 'get', 'oc
topus', 'index', 'htmlI', 'believe', 'proof', 'age', 'required', 'purchase',
'For', 'details', 'discounts', 'available', 'senior', 'visitors', 'HK', 'se
e', 'hongkongextras', 'com', 'senior', 'visitor', 'discounts', 'Edited', '2
4', 'May', '2011', '02', '39', 'I', 'think', 'prove', 'age', 'kind', 'officia
l', 'ID', 'purchase', 'senior', 'card', 'I', 'senior', 'used', 'seniors', 'ca
rd', 'Australia', 'passport', 'Got', 'discoyunts', 'lots', 'places', 'I', 'lo
cal', 'id', 'card', 'ta', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'functi
on', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'func
tion', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks',
'33450793', 'log', 'autolink', 'impression', 'Disneyland', 'biggest', 'savin
g', 'Always', 'ask', 'discount', 'tell', 'yiou', 'senior', 'anywhere', 'tel
l', 'Albon', 'I', 'prepared', 'Australian', 'Seniors', 'card', 'passport',
'I', 'went', 'buy', 'octopus', 'card', 'even', 'ask', 'They', 'took', 'one',
'look', 'gave', 'Seniors', 'Card', 'There', 'conditions', 'purchase', 'Octopu
s', 'Cards', 'proof', 'entitlement', 'requested', 'purchase', 'one', 'You',
'also', 'buy', 'Child', 'Octopus', 'child', 'without', 'child', 'present', 'w
ithout', 'evidence', 'age', 'There', 'however', 'conditions', 'use', 'Octopu
s', 'Cards', 'use', 'Octopus', 'Card', 'elegible', 'order', 'gain', 'concessi
onary', 'fare', 'breach', 'conditions', 'use', 'risk', 'fined', 'Bus', 'drive
rs', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'docum
ent', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEv
entOnPage', 'postLinkInline', 'impression', 'postLinks', '33459254', 'log',
'autolink', 'impression', 'MTR', 'staff', 'etc', 'know', 'concessionary', 'Oc
topus', 'used', 'tone', 'sounds', 'swiped', 'means', 'Someone', 'info', 'des
k', 'HK', 'airport', 'helped', 'us', 'get', 'Octopus', 'cards', 'They', 'too
k', 'word', 'I', 'senior', 'You', 'get', 'ta', 'queueForLoad', 'ta', 'queueFo
rLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoad
ed', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impressio
n', 'postLinks', '33461109', 'log', 'autolink', 'impression', 'star', 'Ferr
y', 'free', 'showing', 'card', 'gate', 'It', 'worked', 'great', 'Ok', 'grea
t', 'Can', 'get', 'refund', 'Octopus', 'cards', 'HK', 'departure', 'airport',
'If', 'yes', 'find', 'refund', 'counter', 'Thanks', 'You', 'get', 'refund',
'airport', 'handing', 'card', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'fu
nction', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f',
'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLi
nks', '33469537', 'log', 'autolink', 'impression', 'Airport', 'Express', 'Cus
tomer', 'Service', 'Desks', 'There', 'two', 'Terminal', '1', 'Arrivals', 'Hal
l', 'known', 'Meeters', 'Greeters', 'Hall', 'one', 'area', 'A', 'hall', 'on
e', 'area', 'B', 'hall', 'one', 'Terminal', '2', 'Airport', 'Express', 'arriv
al', 'platform', 'Edited', '25', 'May', '2011', '04', '33', 'hi', 'travelin
g', 'Auckland', 'Paris', 'via', 'hong', 'kong', '3', 'hour', 'stop', 'acess',
'duty', 'free', 'areas', 'would', 'like', 'look', 'iphones', 'ipad', 'etc',
'Is', 'website', 'check', 'hong', 'kong', 'prices', 'line', 'cheers', 'look',
'forward', 'feedback', 'advice', 'Curtis', 'Hi', 'Curtis', 'price', 'iDevice
s', 'obtained', 'visiting', 'Apple', 'HK', 'store', 'There', 'two', 'Apple',
'resellers', 'HKIA', 'DG', 'Lifestyle', 'Nuance', 'Watson', 'HK', 'tax', 'dut
y', 'free', 'respect', 'price', 'airport', 'duty', 'free', 'Apple', 'store',
'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document',
'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnP
age', 'postLinkInline', 'impression', 'postLinks', '40458774', 'log', 'autoli
nk', 'impression', 'IFC', 'mall', 'downtown', 'Apple', 'resellers', 'stores',
'Even', 'wine', 'tax', 'free', 'There', 'also', 'websites', 'compares', 'appl
e', 'prices', 'internationally', 'www', 'geekzone', 'co', 'nz', 'forums', 'as
p', 'Message', 'TripAdvisor', 'staff', 'TripAdvisor', 'staff', 'removed', 'po

st', 'meet', 'TripAdvisor', 'forum', 'guidelines', 'regards', 'participatio
n', 'business', 'representatives', 'To', 'review', 'TripAdvisor', 'Forums',
'Posting', 'Guidelines', 'please', 'foll', 'Anyway', 'answer', 'basic', 'ques
tion', 'Yes', 'able', 'access', 'range', 'shops', 'transit', 'area', 'HKIA',
'A', 'listing', 'found', 'hongkongairport', 'com', 'eng', 'shopping', 'inde
x', 'htmlAny', 'shop', 'listed', 'inside', 'restricted', 'area', 'available',
'without', 'haveing', 'go', 'immigration', 'Other', 'shops', 'terminal', 'pr
e', 'immigration', 'departure', 'areas', 'access', 'go', 'immigration', 'shop
ping', 'enter', 'security', '3', 'hours', 'enough', 'accomplish', 'really',
'want', 'New', 'Zealanders', 'visa', 'free', 'access', 'cost', 'time', 'spen
t', 'lines', 'appreciate', 'help', 'tax', 'free', 'wine', 'sounds', 'like',
'bonus', 'thank', 'check', 'site', 'later', 'today', 'appreciate', 'time', 'g
et', 'back', 'check', 'site', 'thank', 'fountain', 'knowledge', 'I', 'pass',
'family', 'sounds', 'like', 'busy', '3', 'hours', 'us', 'thank', 'By', 'way',
'goods', 'may', 'subject', 'tax', 'France', 'If', 'HK', 'goods', 'tax', 'fre
e', 'sure', 'wine', 'tax', 'free', 'I', 'think', 'referring', 'wine', 'duty',
'free', 'HK', 'since', 'guess', '2002', 'Anyway', 'wine', 'much', 'cheaper',
'France', 'Shops', 'selling', 'ipads', 'etc', 'found', 'landside', 'airside',
'Hi', 'Please', 'anyone', 'offer', 'advice', 'best', 'way', 'get', 'Wanchai',
'HK', 'airport', 'Are', 'buses', 'trains', 'take', 'Or', 'best', 'get', 'tax
i', 'I', 'know', 'far', 'journey', 'want', 'pay', 'fortune', 'taxi', 'fares',
'Many', 'thanks', 'Taxi', 'airport', 'costs', 'HK', '280', 'plus', 'HK', '5',
'item', 'luggage', 'boot', 'Wanchai', 'also', 'served', 'Cityflyer', 'airpor
t', 'bus', 'A11', 'costs', 'HK', '40', 'takes', '60', 'minutes', 'You', 'fin
d', 'details', 'hotels', 'close', 'airport', 'bus', 'routes', 'route', 'detai
ls', 'nearest', 'bus', 'stop', 'nwstbus', 'com', 'hk', 'routes', 'You', 'als
o', 'take', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g',
'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 't
rackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '33993605', 'l
og', 'autolink', 'impression', 'Airport', 'Express', 'train', 'Hong', 'Kong',
'Station', 'mins', 'taxi', 'HK', '60', 'Wanchai', 'hotels', 'Wancahi', 'serve
d', 'free', 'Airport', 'Express', 'Shuttle', 'Buses', 'H1', 'H2', 'Hong', 'Ko
ng', 'Station', 'mtr', 'com', 'hk', 'eng', 'complom', 'free', 'bus', 'htmlAir
port', 'Express', 'single', 'adult', 'fare', 'Hong', 'Kong', 'Station', 'HK10
0', '2', 'people', 'travelling', 'together', 'discounted', 'Group', 'Ticket
s', 'available', 'mtr', 'com', 'hk', 'eng', 'promotions', '4persons', 'html',
'Should', 'also', 'mention', 'list', 'hotels', 'close', 'A11', 'bus', 'rout
e', 'link', 'post', 'comprehensive', 'If', 'place', 'staying', 'list', 'sugge
st', 'advise', 'exactly', 'staying', 'specific', 'advice', 'We', 'flying', 'U
S', '5', '5', 'hour', 'transit', 'HK', 'airport', 'Coming', 'night', 'flyin
g', 'midnight', 'Any', 'suggestions', 'freshen', 'rest', 'bit', 'Lounge', 'ho
tel', 'My', 'concern', 'lounge', 'boys', 'noisy', 'passengers', 'You', 'may',
'want', 'try', 'pay', 'lounge', 'email', 'see', 'kids', 'welcome', 'I', 'se
e', 'hongkongairport', 'com', 'eng', 'passenger', 'arrival', 't1', 'hello',
'help', 'would', 'like', 'ride', 'train', 'shenzhen', 'hongkong', 'flight',
'7pm', 'many', 'hours', 'take', 'hongkong', 'mtr', 'terminal', 'many', 'minut
es', 'hk', 'airport', 'ride', 'taxi', 'lastly', 'much', 'pay', 'taxi', 'than
k', 'much', 'There', 'direct', 'train', 'SZ', 'border', 'HKIA', 'There', 'exc
ellent', 'journey', 'planner', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'f
unction', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f',
'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLi
nks', '38316268', 'log', 'autolink', 'impression', 'MTR', 'website', 'shows',
'interchanges', 'It', 'gives', 'minimum', 'journey', 'time', '81', 'minutes',
'I', 'think', 'would', 'better', 'take', 'bus', 'It', 'depends', 'going', 't
a', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document',
'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnP
age', 'postLinkInline', 'impression', 'postLinks', '38406495', 'log', 'autoli

nk', 'impression', 'Shenzhen', 'I', 'usually', 'take', 'Airport', 'bus', 'A4 3', 'Shueng', 'Shui', 'MTR', 'take', 'MTR', 'Lo', 'Wu', 'Since', 'I', 'usually', 'stay', 'Lo', 'Wu', 'district', 'I', 'walk', 'across', 'border', 'take', 'SZ', 'metro', 'As', 'border', 'crossing', 'I', 'always', 'found', 'Lo', 'Wu', 'much', 'less', 'stressful', 'Huanggang', 'Should', 'done', 'way', 'around', 'I', 'normally', 'allow', '1', '5', 'hours', 'get', 'Lo', 'Wu', 'Airport', 'vice', 'versa', 'heard', 'famous', 'wife', 'biscuits', 'chinese', 'pastries', 'one', 'causebay', 'bay', 'wondering', 'closer', 'mongkok', 'anyone', 'knows', 'www', 'kee', 'wah', 'com', 'hkFlat', '3', 'On', 'G', 'F', 'Portion', 'Open', 'Yard', 'Nos', '25', '27', '29', '31', 'Soy', 'St', 'KLN', 'Tel', '2770', '1779business', 'hour', '0800', '220010', 'Mong', 'Kok', 'East', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '13635527', 'log', 'autolink', 'impression', 'MTR', 'Station', 'Tel', '2628', '9328business', 'hour', '0730', '2200They', 'even', '2', 'shops', 'HK', 'airport', 'hi', 'I', 'wonder', 'family', '6', 'paxs', 'cheapest', 'best', 'transport', 'use', 'HK', 'airport', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '14189570', 'log', 'autolink', 'impression', 'Hart', 'Avenue', 'Tsimshatsui', 'Kowloon', 'I', 'heard', 'city', 'buses', 'friends', 'mention', 'take', 'cab', 'better', 'big', 'group', 'thanks', 'best', 'view', 'city', 'take', 'airport', 'bus', 'downside', 'plenty', 'luggage', 'everyone', 'help', 'carry', 'taxi', 'pretty', 'expensive', 'since', 'take', '2', 'Taxi', 'seating', '5', 'since', 'six', 'would', 'get', '2', 'taxis', 'although', 'one', 'taxi', 'pretty', 'roomy', 'You', 'could', 'also', 'private', 'charter', 'get', 'bus', 'Another', 'suggestion', 'could', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '14195832', 'log', 'autolink', 'impression', 'Airport', 'Express', 'good', 'thing', 'kids', 'rate', 'discounted', 'check', 'early', 'worries', 'check', 'airport', 'A21', 'airport', 'bus', 'comfortable', 'takes', 'hour', 'good', 'views', 'city', 'costs', 'HK', '33', 'per', 'adult', 'It', 'drop', '5', 'minutes', 'walk', 'ta', 'queueForLoad', 'ta', 'queueForLoad', 'function', 'f', 'g', 'document', 'addEventListener', 'DOMContentLoaded', 'f', 'function', 'ta', 'trackEventOnPage', 'postLinkInline', 'impression', 'postLinks', '14195885', 'log', 'autolink', 'impression', 'Hart', 'Avenue', 'It', 'plenty', 'room', 'luggage', 'lower', 'deck', 'lots', 'luggage', 'considering', '2', 'cabs', 'also', 'consider', 'private', 'transfers', 'cost', 'less', 'cab', 'ride', '300', 'hkd', 'tst', 'let', 'know', 'could', 'send', 'list', 'people', 'used', 'hotel', 'may', 'bus', 'goes', 'direct', 'eatons', 'cost', '120', 'hk', '1', 'way', 'oct', '24', '2012', 'took', 'ferry', 'hk', 'airport', 'shekou', 'paid', '130hkd', 'ferry', 'ticket', 'friend', 'mine', 'visited', 'nov', '8', '2012', 'took', 'route', 'charge', '220hkd', 'dont', 'understand', 'fare', 'fair', 'unfair', 'D', 'I', 'thought', 'fare', '280', 'RMB', 'It', 'ridiculously', 'expensive', 'quite', 'bit', 'fare', 'Zhuhai', 'When', 'go', 'way', 'least', 'get', 'departure', 'tax', 'refund', '120', '130', 'first', 'class', 'child', 'price', 'www', 'xunlongferry', 'com', 'en', 'line', 'info4', 'asp', 'id', '20', 'I', 'arrive', 'family', 'including', 'kid', 'year', 'old', '12', 'AM', 'HK', 'airport', 'However', 'ferry', '0940', 'Skypier', 'Are', 'enough', 'facilities', 'Skypier', 'spend', 'many', '9', 'hours', 'specially', 'kid']

Here i have lemetize the words of the sentence, lemitizatiom means converting the word into its original form, eg: 'happyyyyyyyy' is lemitize as happy

In [132]:
```python
from nltk.stem import PorterStemmer
# from nltk.stem import LancasterStemmer
#create an object of class PorterStemmer
porter = PorterStemmer()

stem_sentence=[]
for word in word_tokens:
    stem_sentence.append(porter.stem(word))
    stem_sentence.append(" ")
w = "".join(stem_sentence)
w
# stem_sentence
```

Out[132]: 'I would recommend go through immigr get to the ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 38315920 log autolink impress macau ferri termin taxi or night bu to take the next avail ferri to macau so you can get some sleep much as hkia is a great airport there is noth whatsoev to do in the skypier departur area you will have no access to the other airport area includ restaur aseem01 arriv at midnight in ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 38324674 log autolink impress hong kong and proceed to macau is a bit hard have you plan budget one night of hotel on your arriv date 9 to 10 hour is a long time sit on hard shell seat while you may proceed as zhuhai2007 suggest in 1 go through the immigr in hong kong and proceed to go to macau like by the time you reach macau ferri termin clear macau immigr custom then it is 4 am now what you may go to the hotel If your reserv is for the follow day you would be sit in their lobbi area of the hong kong airport for 8 hour for your room to becom avail from the faq section hkia skypier to ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 38327519 log autolink impress macau read the thread from earlier thi year about arriv at 1am need to travel directli to macau tripadvisor com showtop g664891 i12409 k50 12 am is mid night not earli morn with the one year old I realli think it s best to find a hotel in HK close to the airport and just bed down for the night messag from tripadvisor staff tripadvisor staff remov thi post at the origin author s request To review the tripadvisor forum post guidelin pleas follow thi link remo don t ferri from skypier run all night No cas par esq not from skypier first ferri 10 am last ferri 10 pmwww turbojet com hk eng schedul airport aspth ferri from ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 38368609 log autolink impress macau ferri termin sheung wan to macau do run all night messag from tripadvisor staff thi topic ha been close to new post due to inact We hope you ll join the convers by post to an open topic or start a new one To review the tripadvisor forum post guidelin pleas follow thi can ani expert advis the best cheap option to reach ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 36569763 log autolink impress shenzhen from HK airport I am plan a trip next yr I have been read that bu 43 doe stop at a transit railway stop for chang of train to shenzhen pleas advis take a direct bu from the airport A direct bu from hkia to shenzhen cost probabl over hkd 150 the question wa the cheapest mode It may take a littl longer take airport bu a43 from hkia toward luen Wo hui alight at stop 3 sheung shui railroad station take ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 36570240 log autolink impress mtr for 1 stop toward Lo Wu lok Ma chau if you choos to then walk over to shenzhen border control a43 hkd 30 9mtr from sheung shui to Lo Wu hkd 21 0 thank folk for the respons I will like choos the local bu a43 option sinc it is more econom that way It s cheaper but whether it s more econom depend on how you valu your time everybodi s time money convers factor are differ OP should know what work best for him her take a bu directli from the hkia to ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 36601737 log autolink impress shenzhen may not necessarili be faster It depend on the actual destin in shenzhen At ani rate take a bu to sheung shui and chang to mtr train should not take that much longer than the direct hkia shenzhen bu you may want to cross the border at lok Ma chau instead of the more popular Lo Wu then taxi bu metro in shenzhen to your destin compar the total cost and conveni factor by book a flight directli to ta queueforload ta queueforload function f g document addeven

tlisten domcontentload f function ta trackeventonpag postlinkinlin impress po stlink 36604320 log autolink impress shenzhen vs via hkg you can fli china ea stern from sin to szx pvg for about HK 2 430 return in economi fare as post a t MU s web site I had book a special sin HK fare with jetstar for S 230 usd 1 76 return No commerci airlin can beat that fare edit 24 octob 2011 21 54 good for you but I saw jetstar return fare from sin to ta queueforload ta queuefor load function f g document addeventlisten domcontentload f function ta tracke ventonpag postlinkinlin impress postlink 36605457 log autolink impress hkg cu rrent sell for as low as sgd 88 all in for mid februari 2012 for those who do not know jetstar is a no frill budget airlin with 28 29 seat pitch on their s ingl aisl airbu aeroplan everyth els check bag and meal is extra except they have not yet start charg for use the lavatori onboard edit 24 octob 2011 22 1 1 sound like air ta queueforload ta queueforload function f g document addeve ntlisten domcontentload f function ta trackeventonpag postlinkinlin impress p ostlink 36606250 log autolink impress asia but doe jetstar manag to fli their schedul that would be unlik air asia OP specif state fli jetstar not air ta q ueueforload ta queueforload function f g document addeventlisten domcontentlo ad f function ta trackeventonpag postlinkinlin impress postlink 36606937 log autolink impress asia As the old say goe you usual get what you have paid Hi I ll be travel with my parent who qualifi as senior citizen as they are abov 65 yr old note that there are senior citizen rate for travel on the tram etc Is there a senior citizen octopu card that they can appli for If not doe that mean that everytim they go up the tram they have to show their passport and p ay cash how to go about qualifi for the senior citizen fare Is there also sen ior citizen fare for travel on the mtr thank they can purchas the elder octop u card which is avail to tourist as well as HK resid qualifi age is 65 the ca rd will automat give them concessionari fare where avail usual half adult far e addit on wednesday and saturday until 31st august 2011 subject to review el der octopu card holder are charg a maximum fare of HK 2 per journey on the ta queueforload ta queueforload function f g document addeventlisten domcontentl oad f function ta trackeventonpag postlinkinlin impress postlink 33442400 log autolink impress mtr not journey to the mainland border station of Lo Wu and lok Ma chau and until 31st januari 2012 a maximum fare of HK 2 or 50 of the a dult fare whichev is the lower on most bu journey the elder octopu cost HK 70 of which HK 50 is a refund deposit a small handl fee octopu com hk get your o ctopu index htmli don t believ ani proof of age is requir for purchas for det ail of other discount avail to senior visitor to HK see hongkongextra com sen ior visitor discount edit 24 may 2011 02 39 I think you do have to prove your age with some kind of offici ID when you purchas the senior card I am a senio r and use my senior card from australia and my passport got discoyunt lot of place I had no local id or card at all ta ta queueforload ta queueforload fun ction f g document addeventlisten domcontentload f function ta trackeventonpa g postlinkinlin impress postlink 33450793 log autolink impress disneyland wa the biggest save alway ask for discount and tell them yiou are a senior anywh er just tell them albon I wa all prepar with my australian senior card and pa ssport when I went to buy my octopu card but they didn t even ask for it they took one look at me and gave me the senior card there aren t ani condit of pu rchas for octopu card which is whi proof of entitl isn t request to purchas o ne you can also buy a child octopu for a child without the child be present a nd without ani evid of age there are howev condit of use for octopu card and if you use an octopu card for which you are not eleg in order to gain a conce ssionari fare it a breach of condit of use and you risk be fine bu driver ta queueforload ta queueforload function f g document addeventlisten domcontentl oad f function ta trackeventonpag postlinkinlin impress postlink 33459254 log autolink impress mtr staff etc know when a concessionari octopu is be use by the tone it sound when swipe and by other mean someon at the info desk at the HK airport help us get the octopu card they took my word that I wa a senior y

ou get on the ta queueforload ta queueforload function f g document addeventl isten domcontentload f function ta trackeventonpag postlinkinlin impress post link 33461109 log autolink impress star ferri free by show your card at the g ate It work just great Ok great can we get a refund for our octopu card at th e HK departur airport If ye where can we find the refund counter thank you ca n get a refund at the airport by hand in your card at the ta queueforload ta queueforload function f g document addeventlisten domcontentload f function t a trackeventonpag postlinkinlin impress postlink 33469537 log autolink impres s airport express custom servic desk there are two in termin 1 arriv hall kno wn as meeter and greeter hall one in area A of the hall and one in area B of the hall and there is one in termin 2 on the airport express arriv platform e dit 25 may 2011 04 33 hi there we are travel from auckland to pari via hong k ong and we have a 3 hour stop over can you acess duti free area as we would l ike to have a look at iphon ipad etc Is there a websit we can check out hong kong price on line cheer and look forward to your feedback and advic curti Hi curti the price of idevic can be obtain by visit appl HK store there are two appl resel at hkia DG lifestyl and nuanc watson HK is tax and duti free in th i respect so price is same in airport duti free or at the appl store at ta qu eueforload ta queueforload function f g document addeventlisten domcontentloa d f function ta trackeventonpag postlinkinlin impress postlink 40458774 log a utolink impress ifc mall or downtown appl resel store even wine is tax free t here s also websit that compar appl price intern such as www geekzon co nz fo rum asp messag from tripadvisor staff tripadvisor staff remov thi post becaus it did not meet tripadvisor s forum guidelin with regard to the particip of b usi repres To review the tripadvisor forum post guidelin pleas foll anyway to answer your basic question ye you will be abl to access a rang of shop in the transit area at hkia A list can be found on hongkongairport com eng shop inde x htmlani shop list as be insid the restrict area is avail to you without hav e to go through immigr other shop are in the termin pre immigr departur area to access those you will have to go through immigr and then do your shop and re enter through secur 3 hour is enough to accomplish thi if you realli want to new zealand have visa free access so the onli cost of do thi is your time spent in line appreci your help tax free wine sound like a bonu to me thank y ou i will check out the site later on today appreci your time to get back to me will check that site thank you you are a fountain of knowledg and I will p ass thi on to the famili it sound like it will be a busi 3 hour for us thank you By the way good may be subject to tax in franc If HK all good are tax fre e for sure wine is tax free too I think it wa refer to wine be duti free in H K sinc a while guess 2002 anyway wine will be much cheaper in franc shop sell ipad etc can be found on both landsid and airsid Hi pleas can anyon offer ani advic on the best way to get to wanchai from HK airport are there buse train that can take you there Or is it best to get a taxi I don t know how far the journey is but don t want to pay a fortun in taxi fare mani thank taxi from t he airport cost about HK 280 plu HK 5 for each item of luggag in the boot wan chai is also serv by cityfly airport bu a11 which cost HK 40 and take about 6 0 minut you can find detail of which hotel are close to airport bu rout rout detail and nearest bu stop here nwstbu com hk rout you can also take ta queue forload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 33993605 log autol ink impress airport express train to hong kong station min then taxi HK 60 to wanchai or some hotel in wancahi are serv by free airport express shuttl buse H1 and H2 from hong kong station mtr com hk eng complom free bu htmlairport e xpress singl adult fare to hong kong station is hk100 but if 2 or more peopl are travel togeth discount group ticket are avail mtr com hk eng promot 4pers on html should also mention that the list of hotel close to a11 bu rout at th e link in the abov post isn t comprehens If the place where you are stay isn t on the list suggest you advis exactli where you are stay for more specif ad

vic We are fli in from the US and have 5 5 hour transit at HK airport come in at night and fli out just after midnight ani suggest on where to freshen up and rest for a bit loung or hotel My onli concern with the loung is that my boy will be too noisi for other passeng you may want to tri the pay in loung email them to see if kid are welcom I don t see whi not hongkongairport com eng passeng arriv t1 hello can you help me we would like to ride a train from shenzhen to hongkong our flight is 7pm how mani hour doe it take to hongkong from mtr termin how mani minut to hk airport if we ride a taxi and lastli how much do we have to pay for the taxi thank you so much there is no direct train from the SZ border to hkia there is an excel journey planner on the ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 38316268 log autolink impress mtr websit which show all the interchang It give minimum journey time of 81 minut I think it would be better to take a bu It all depend where you re go to in ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 38406495 log autolink impress shenzhen but I usual take the airport bu a43 to shueng shui mtr and then take the mtr to Lo Wu sinc I usual stay in the Lo Wu district I walk across the border and then take the SZ metro As a border cross I ve alway found Lo Wu to be much less stress than huanggang should have done that the other way around I normal allow about 1 5 hour to get from Lo Wu to the airport and vice versa heard they are famou for their wife s biscuit and other chines pastri onli been to the one at causebay bay wa wonder if they have ani closer to mongkok anyon know www kee wah com hkflat 3 On G F portion of the open yard no 25 27 29 31 soy St kln tel 2770 1779busi hour 0800 220010 mong kok east ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 13635527 log autolink impress mtr station tel 2628 9328busi hour 0730 2200they even have 2 shop at HK airport hi I wonder for a famili of 6 pax what is the cheapest and best transport to use from HK airport to ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 14189570 log autolink impress hart avenu tsimshatsui kowloon I heard of the citi buse and some friend mention take a cab is better for a big group thank for the best view of the citi take the airport bu onli downsid is if you have plenti of luggag and not everyon can help carri them taxi will be pretti expens sinc you will have to take 2 taxi onli ha seat for 5 and sinc you have six you would have to get 2 taxi although one taxi is pretti roomi you could also do a privat charter or get a bu anoth suggest could be ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 14195832 log autolink impress airport express and the good thing about that is that kid rate is discount and that they can check you in earli so no worri about check in at the airport a21 airport bu is comfort take an hour ha good view of the citi and cost HK 33 per adult It will drop you 5 minut walk from ta queueforload ta queueforload function f g document addeventlisten domcontentload f function ta trackeventonpag postlinkinlin impress postlink 14195885 log autolink impress hart avenu It ha plenti of room for luggag on the lower deck if you have lot of luggag and you are consid 2 cab also consid privat transfer the cost is more or less the same as a cab ride or about 300 hkd to tst let me know so i could send you a list of peopl i have use your hotel may have a bu that goe direct the eaton cost 120 hk each 1 way oct 24 2012 i took a ferri from hk airport to shekou i paid 130hkd for the ferri ticket then a friend of mine visit me here nov 8 2012 he took the same rout but they charg him 220hkd i dont understand whi fare is not fair it unfair D I thought the fare wa about 280 rmb It is ridicul expens as it is quit a bit more than the fare to zhuhai when you go the other way you at least get the departur tax refund of 120 130 is the first class child price www x

unlongferri com en line info4 asp id 20 I will arriv with my famili includ a
kid year old at about 12 AM at HK airport howev our ferri is from 0940 from t
he skypier are there enough facil at the skypier to spend as mani as 9 hour s
pecial with a kid '

here i have done Term frequency-inverse document frequency (TF-IDF) analysis. this is an analysis which will
help us to know that how much times a word is present in the whole post made on the tripadvisor website

In [63]:
```python
#Term frequency-inverse document frequency (TF-IDF) analysis
# for
wordSet = set(stem_sentence).union(set(stem_sentence))
wordDict = dict.fromkeys(wordSet, 0)
wordDict
```

In [64]:
```python
for word in stem_sentence:
    wordDict[word] += 1
wordDict
```

```
Out[64]:  {'whatsoev': 1,
           'postlinkinlin': 24,
           'then': 8,
           '88': 1,
           'ta': 73,
           '11': 1,
           'entitl': 1,
           'offer': 1,
           'etc': 4,
           'they': 13,
           'abl': 1,
           'room': 2,
           'new': 3,
           'chang': 2,
           'planner': 1,
           'read': 2,
           'HK': 26,
           'australia': 1,
           'do': 8,
           'with': 10,
           '1': 5,
           'sell': 2,
           'them': 5,
           'across': 1,
           'charter': 1,
           'queueforload': 48,
           'rout': 5,
           '4person': 1,
           'shui': 4,
           'done': 1,
           'understand': 1,
           'swipe': 1,
           'bay': 1,
           'insid': 1,
           'risk': 1,
           'platform': 1,
           '65': 2,
           'age': 4,
           'obtain': 1,
           'each': 2,
           'versa': 1,
           'macau': 10,
           'els': 1,
           'visitor': 2,
           'at': 30,
           'enough': 2,
           'stay': 3,
           '14195832': 1,
           'better': 2,
           'no': 6,
           '9328busi': 1,
           'hour': 12,
           'up': 2,
           'portion': 1,
           '43': 1,
           'shenzhen': 11,
           'discoyunt': 1,
```

```
'known': 1,
'thread': 1,
'Ok': 1,
'prove': 1,
'abov': 2,
'breach': 1,
'can': 22,
'pretti': 2,
'valu': 1,
'way': 6,
'departur': 4,
'hart': 2,
'friend': 2,
'walk': 3,
'stop': 6,
'requir': 1,
'pmwww': 1,
'130': 1,
'inact': 1,
'remov': 2,
'noth': 1,
'a21': 1,
'everytim': 1,
'mid': 2,
'for': 42,
'chines': 1,
'place': 2,
're': 2,
'flight': 2,
'over': 3,
'let': 1,
'you': 54,
'district': 1,
'websit': 3,
'promot': 1,
'email': 1,
'gain': 1,
'36601737': 1,
'australian': 1,
'six': 1,
'passport': 3,
'g664891': 1,
'2628': 1,
'9mtr': 1,
'after': 1,
'usd': 1,
'railway': 1,
'day': 1,
'10': 3,
'link': 2,
'rest': 1,
'go': 10,
'14195885': 1,
'howev': 2,
'regard': 1,
'next': 2,
'i': 6,
```

```
'normal': 1,
'budget': 2,
'becaus': 1,
'someon': 1,
'comfort': 1,
'feedback': 1,
'visit': 2,
'airbu': 1,
'pay': 4,
'class': 1,
'anyway': 2,
'open': 2,
'a43': 4,
'220hkd': 1,
'staff': 6,
'subject': 2,
'servic': 1,
'those': 2,
'transit': 3,
'20': 1,
'sin': 3,
'shueng': 1,
'access': 4,
'journey': 6,
'didn': 1,
'specif': 2,
'tone': 1,
'04': 1,
'guidelin': 4,
'use': 8,
' ': 3341,
'forward': 1,
'tri': 1,
'en': 1,
'wah': 1,
'custom': 2,
'sgd': 1,
'yiou': 1,
'roomi': 1,
'midnight': 2,
'hui': 1,
'htmlani': 1,
'12': 2,
'www': 3,
'fine': 1,
'best': 7,
'join': 1,
'13635527': 1,
'36569763': 1,
'as': 17,
'basic': 1,
'old': 4,
'say': 1,
'MU': 1,
'paid': 2,
'If': 5,
'note': 1,
```

```
'are': 25,
'tripadvisor': 10,
'citizen': 5,
'autolink': 24,
'8': 2,
'proceed': 3,
'good': 5,
'shuttl': 1,
'anywher': 1,
'mong': 1,
'OP': 2,
'peopl': 2,
'2': 9,
'zhuhai2007': 1,
'geekzon': 1,
'took': 4,
'other': 8,
'pax': 1,
'forum': 5,
'plu': 1,
'min': 1,
'complom': 1,
'deck': 1,
'2002': 1,
'it': 16,
'conveni': 1,
'down': 1,
'how': 6,
'tram': 2,
'downtown': 1,
'becom': 1,
'price': 5,
'some': 4,
'think': 4,
'serv': 2,
'cab': 3,
'fair': 1,
'one': 10,
'folk': 1,
'isn': 3,
'US': 1,
'two': 2,
'much': 6,
'get': 15,
'freshen': 1,
'detail': 3,
'welcom': 1,
'tel': 2,
'cross': 2,
'save': 1,
'the': 117,
'popular': 1,
'54': 1,
'is': 46,
'39': 1,
'look': 3,
'iphon': 1,
```

```
'2770': 1,
'fare': 17,
'quit': 1,
'if': 9,
'their': 5,
'visa': 1,
'an': 4,
'last': 1,
'ridicul': 1,
'fountain': 1,
'pleas': 5,
'176': 1,
'plan': 2,
'kln': 1,
'F': 1,
'caspar': 1,
'depend': 3,
'yr': 2,
'wonder': 2,
'lok': 3,
'long': 1,
'mainland': 1,
'81': 1,
'present': 1,
'more': 7,
'6': 1,
'believ': 1,
'reach': 2,
'eastern': 1,
'aseem01': 1,
'0730': 1,
'0940': 1,
'biscuit': 1,
'am': 5,
'230': 1,
'that': 20,
'bed': 1,
'B': 1,
'22': 1,
'today': 1,
'author': 1,
'automat': 1,
'pass': 1,
'mtr': 12,
'gave': 1,
'hall': 4,
'reserv': 1,
'web': 1,
'just': 4,
'while': 2,
'instead': 1,
'run': 2,
'exactli': 1,
'also': 8,
'aisl': 1,
'addit': 1,
'onli': 5,
```

```
'stress': 1,
'had': 2,
'lot': 2,
'check': 6,
'on': 22,
'togeth': 1,
'kowloon': 1,
'fortun': 1,
'drop': 1,
'albon': 1,
'9': 2,
'date': 1,
'asp': 2,
'eaton': 1,
'cost': 9,
'differ': 1,
'in': 35,
'saturday': 1,
'wanchai': 3,
'38406495': 1,
'430': 1,
'respect': 1,
'worri': 1,
'appreci': 2,
'secur': 1,
'station': 6,
'S': 1,
'downsid': 1,
'mani': 4,
'got': 1,
'aspth': 1,
'sleep': 1,
'schedul': 2,
'small': 1,
'36605457': 1,
'avail': 7,
'condit': 3,
'desk': 2,
'my': 8,
'recommend': 1,
'3': 5,
'i12409': 1,
'like': 7,
'G': 1,
'destin': 2,
'doe': 4,
'skypier': 6,
'120': 2,
'foll': 1,
'directli': 3,
'ha': 4,
'onboard': 1,
'word': 1,
'38327519': 1,
'ani': 8,
'chau': 3,
'so': 6,
```

```
'not': 14,
'spend': 1,
'ipad': 2,
'eng': 5,
'adult': 4,
'nearest': 1,
'great': 3,
'purchas': 5,
'fli': 5,
't1': 1,
'me': 7,
'29': 2,
'hk100': 1,
'remo': 1,
'arriv': 7,
'gate': 1,
'4': 1,
'sheung': 4,
'handl': 1,
'shell': 1,
'have': 24,
'such': 1,
'St': 1,
'need': 1,
'big': 1,
'particip': 1,
'shop': 7,
'boy': 1,
'a': 57,
'allow': 1,
'yet': 1,
'html': 1,
'option': 2,
'first': 2,
'ye': 2,
'octob': 2,
'hi': 2,
'mongkok': 1,
'could': 3,
'info': 1,
'charg': 3,
'except': 1,
'and': 50,
'No': 2,
'meeter': 1,
'famili': 3,
'due': 1,
'bag': 1,
'33450793': 1,
'both': 1,
'holder': 1,
'co': 1,
'40458774': 1,
'My': 1,
'which': 7,
'xunlongferri': 1,
'thi': 10,
```

```
's': 9,
'280': 2,
'mean': 2,
'famou': 1,
'februari': 1,
'htmli': 1,
'bonu': 1,
'DG': 1,
'repres': 1,
'should': 4,
'hand': 1,
'all': 8,
'same': 3,
'littl': 1,
'request': 2,
'bit': 3,
'asia': 3,
'meal': 1,
'mine': 1,
'total': 1,
'goe': 2,
'maximum': 2,
'300': 1,
'pastri': 1,
'parent': 1,
'bu': 23,
'Wu': 8,
'manag': 1,
'lavatori': 1,
'free': 12,
'night': 7,
'help': 4,
'or': 12,
'meet': 1,
'38368609': 1,
'k50': 1,
'child': 4,
'sit': 2,
'senior': 14,
'than': 3,
'wife': 1,
'watson': 1,
'border': 6,
'travel': 6,
'current': 1,
'sound': 4,
'well': 1,
'airsid': 1,
'Ma': 3,
'tst': 1,
'htmlairport': 1,
'too': 2,
'enter': 1,
'counter': 1,
'china': 1,
'includ': 2,
'show': 3,
```

```
'termin': 7,
'morn': 1,
'heard': 2,
'anoth': 1,
'whi': 3,
'group': 2,
'appli': 1,
'Lo': 8,
'duti': 4,
'januari': 1,
'view': 2,
'octopu': 14,
'list': 5,
'driver': 1,
'topic': 2,
'plenti': 2,
'hope': 1,
'resid': 1,
'vice': 1,
'state': 1,
'oct': 1,
'through': 5,
't': 10,
'Hi': 3,
'auckland': 1,
'resel': 2,
'your': 16,
'hongkongextra': 1,
'ask': 2,
'greeter': 1,
'earli': 2,
'low': 1,
'about': 11,
'restrict': 1,
'toward': 2,
'cheap': 1,
'Or': 1,
'cheaper': 2,
'deposit': 1,
'domcontentload': 24,
'cash': 1,
'convers': 2,
'vs': 1,
'franc': 2,
'by': 10,
'info4': 1,
'comprehens': 1,
'probabl': 1,
'rang': 1,
'pari': 1,
'luen': 1,
'I': 28,
'when': 5,
'boot': 1,
'60': 2,
'tsimshatsui': 1,
'mode': 1,
```

```
'thing': 1,
'index': 2,
'rmb': 1,
'carri': 1,
'On': 1,
'factor': 2,
'impress': 48,
'what': 4,
'start': 2,
'landsid': 1,
'hkd': 4,
'aeroplan': 1,
'prepar': 1,
'showtop': 1,
'choos': 2,
'0800': 1,
'air': 3,
'know': 6,
'kind': 1,
'work': 2,
'33': 2,
'unfair': 1,
'everyon': 1,
'ifc': 1,
'privat': 2,
'area': 9,
'hk': 8,
'transfer': 1,
'express': 6,
'H1': 1,
'At': 1,
'expens': 2,
'mention': 2,
'36606250': 1,
'biggest': 1,
'id': 2,
'airlin': 2,
'money': 1,
'post': 9,
'send': 1,
'pitch': 1,
'per': 2,
'0': 1,
'frill': 1,
'H2': 1,
'qualifi': 3,
'messag': 3,
'longer': 2,
'faster': 1,
'respons': 1,
'kee': 1,
'hard': 2,
'found': 3,
'cheer': 1,
'commerci': 1,
'control': 1,
'rate': 3,
```

```
'everyth': 1,
'yard': 1,
'ticket': 2,
'27': 1,
'31': 1,
'dont': 1,
'spent': 1,
'As': 2,
'later': 1,
'hkflat': 1,
'appl': 5,
'faq': 1,
'hello': 1,
'her': 1,
've': 1,
'zealand': 1,
'shekou': 1,
'21': 2,
'cheapest': 2,
'offici': 1,
'to': 99,
'com': 11,
'us': 2,
'hongkong': 2,
'we': 9,
'busi': 2,
'guess': 1,
'taxi': 12,
'alight': 1,
'necessarili': 1,
'special': 2,
'item': 1,
'130hkd': 1,
'facil': 1,
'150': 1,
'acess': 1,
'advic': 3,
'discount': 5,
'least': 1,
'14189570': 1,
'wednesday': 1,
'25': 2,
'buse': 3,
'cityfly': 1,
'suggest': 4,
'intern': 1,
'pvg': 1,
'knowledg': 1,
'interchang': 1,
'saw': 1,
'advis': 3,
'a11': 2,
'railroad': 1,
'line': 3,
'evid': 1,
'minut': 4,
'direct': 5,
```

```
'Wo': 1,
'minimum': 1,
'f': 48,
'would': 7,
'card': 17,
'less': 2,
'realli': 2,
'nz': 1,
'220010': 1,
'economi': 1,
'take': 20,
'august': 1,
'trackeventonpag': 24,
'review': 4,
'been': 3,
'idevic': 1,
'star': 1,
'causebay': 1,
'trip': 1,
'38316268': 1,
'metro': 2,
'noisi': 1,
'pre': 1,
'airport': 32,
'come': 1,
'who': 2,
'We': 2,
'33993605': 1,
'thank': 9,
'will': 15,
'proof': 2,
'usual': 4,
'find': 3,
'around': 1,
'but': 8,
'nwstbu': 1,
'36604320': 1,
'lower': 2,
'ID': 1,
'70': 1,
'transport': 1,
'clear': 1,
'whichev': 1,
'lastli': 1,
'hkg': 2,
'33442400': 1,
'of': 41,
'see': 3,
'refer': 1,
'30': 1,
'aren': 1,
'A': 3,
'postlink': 24,
'without': 3,
'passeng': 2,
'section': 1,
'kok': 1,
```

```
'he': 1,
'g': 24,
'everybodi': 1,
'closer': 1,
'disneyland': 1,
'tourist': 1,
'went': 1,
'actual': 1,
'be': 21,
'from': 31,
'By': 1,
'accomplish': 1,
'question': 2,
'alway': 2,
'citi': 3,
'hongkongairport': 2,
'hkia': 9,
'order': 1,
'24': 4,
'here': 2,
'unlik': 1,
'extra': 1,
'seat': 3,
'store': 3,
'36570240': 1,
'did': 1,
'document': 24,
'now': 1,
'Is': 3,
'zhuhai': 1,
'him': 2,
'40': 1,
'refund': 5,
'lobbi': 1,
'singl': 2,
'don': 5,
'consid': 2,
'concessionari': 3,
'31st': 2,
'tax': 7,
'site': 3,
'wancahi': 1,
'return': 3,
'28': 1,
'train': 6,
'2012': 4,
'back': 1,
'our': 3,
'hong': 8,
'elder': 3,
'compar': 2,
'eleg': 1,
'33459254': 1,
'origin': 1,
'half': 1,
'expert': 1,
'even': 3,
```

```
'time': 7,
'turbojet': 1,
'33461109': 1,
'50': 2,
'wan': 1,
'log': 24,
'7pm': 1,
'thought': 1,
'2200they': 1,
'luggag': 4,
'36606937': 1,
'mall': 1,
'earlier': 1,
'kong': 8,
'give': 2,
'buy': 2,
'out': 3,
'answer': 1,
'ride': 3,
'll': 2,
'anyon': 2,
'SZ': 2,
'want': 4,
'east': 1,
'econom': 2,
'restaur': 1,
'huanggang': 1,
'until': 2,
'there': 17,
'1779busi': 1,
'curti': 2,
'esq': 1,
'nuanc': 1,
'book': 2,
'nov': 1,
'avenu': 2,
'year': 3,
'D': 1,
'excel': 1,
'immigr': 6,
'2011': 5,
'1am': 1,
'jetstar': 5,
'38315920': 1,
'addeventlisten': 24,
'38324674': 1,
'szx': 1,
'where': 6,
'fee': 1,
'local': 2,
'lifestyl': 1,
'wa': 7,
'soy': 1,
'loung': 3,
'tell': 2,
'ferri': 12,
'most': 1,
```

```
        'close': 4,
        'hotel': 8,
        'follow': 3,
        'concern': 1,
        'may': 10,
        'sinc': 5,
        'To': 3,
        'It': 9,
        'wine': 5,
        'edit': 4,
        '33469537': 1,
        'kid': 4,
        'although': 1,
        '5': 6,
        'via': 2,
        'beat': 1,
        'whether': 1,
        'far': 1,
        'sure': 1,
        'function': 48,
        '02': 1,
        'AM': 1}
```

In [65]:
```python
from pandas import DataFrame as df
df([wordDict])
```

Out[65]:

	0	02	04	0730	0800	0940	1	10	11	...	ye	year	yet	yiou	you	your	yr	zealar
0	3341	1	1	1	1	1	1	5	3	1	...	2	3	1	1	54	16	2	

1 rows × 820 columns

◄ | ▓▓▓▓▓▓▓▓▓▓▓▓▓▓ | ► |
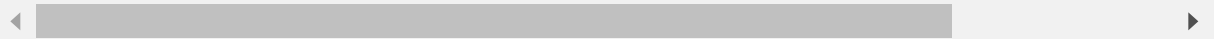
In [66]:
```python
def computeTF(wordDict, bow):
    tfDict = {}
    bowCount = len(bow)
    for word, count in wordDict.items():
        tfDict[word] = count/float(bowCount)
    return tfDict
tfBow = computeTF(wordDict, w)
tfBow
```

```
Out[66]:  {'whatsoev': 5.6041246357318986e-05,
           'postlinkinlin': 0.001344989912575655
           'then': 0.0004483299708585519,
           '88': 5.6041246357318986e-05,
           'ta': 0.004091010984084286,
           '11': 5.6041246357318986e-05,
           'entitl': 5.6041246357318986e-05,
           'offer': 5.6041246357318986e-05,
           'etc': 0.00022416498542927594,
           'they': 0.0007285362026451469,
           'abl': 5.6041246357318986e-05,
           'room': 0.00011208249271463797,
           'new': 0.00016812373907195696,
           'chang': 0.00011208249271463797,
           'planner': 5.6041246357318986e-05,
           'read': 0.00011208249271463797,
           'HK': 0.0014570724052902937,
           'australia': 5.6041246357318986e-05,
           'do': 0.0004483299708585519,
           'with': 0.0005604124635731899,
           '1': 0.00028020623178659493,
           'sell': 0.00011208249271463797,
           'them': 0.00028020623178659493,
           'across': 5.6041246357318986e-05,
           'charter': 5.6041246357318986e-05,
           'queueforload': 0.0026899798251513113,
           'rout': 0.00028020623178659493,
           '4person': 5.6041246357318986e-05,
           'shui': 0.00022416498542927594,
           'done': 5.6041246357318986e-05,
           'understand': 5.6041246357318986e-05,
           'swipe': 5.6041246357318986e-05,
           'bay': 5.6041246357318986e-05,
           'insid': 5.6041246357318986e-05,
           'risk': 5.6041246357318986e-05,
           'platform': 5.6041246357318986e-05,
           '65': 0.00011208249271463797,
           'age': 0.00022416498542927594,
           'obtain': 5.6041246357318986e-05,
           'each': 0.00011208249271463797,
           'versa': 5.6041246357318986e-05,
           'macau': 0.0005604124635731899,
           'els': 5.6041246357318986e-05,
           'visitor': 0.00011208249271463797,
           'at': 0.0016812373907195697,
           'enough': 0.00011208249271463797,
           'stay': 0.00016812373907195696,
           '14195832': 5.6041246357318986e-05,
           'better': 0.00011208249271463797,
           'no': 0.0003362474781439139,
           '9328busi': 5.6041246357318986e-05,
           'hour': 0.0006724949562878278,
           'up': 0.00011208249271463797,
           'portion': 5.6041246357318986e-05,
           '43': 5.6041246357318986e-05,
           'shenzhen': 0.0006164537099305089,
           'discoyunt': 5.6041246357318986e-05,
```

'known': 5.6041246357318986e-05,
'thread': 5.6041246357318986e-05,
'Ok': 5.6041246357318986e-05,
'prove': 5.6041246357318986e-05,
'abov': 0.00011208249271463797,
'breach': 5.6041246357318986e-05,
'can': 0.0012329074198610178,
'pretti': 0.00011208249271463797,
'valu': 5.6041246357318986e-05,
'way': 0.0003362474781439139,
'departur': 0.00022416498542927594,
'hart': 0.00011208249271463797,
'friend': 0.00011208249271463797,
'walk': 0.00016812373907195696,
'stop': 0.0003362474781439139,
'requir': 5.6041246357318986e-05,
'pmwww': 5.6041246357318986e-05,
'130': 5.6041246357318986e-05,
'inact': 5.6041246357318986e-05,
'remov': 0.00011208249271463797,
'noth': 5.6041246357318986e-05,
'a21': 5.6041246357318986e-05,
'everytim': 5.6041246357318986e-05,
'mid': 0.00011208249271463797,
'for': 0.0023537323470073975,
'chines': 5.6041246357318986e-05,
'place': 0.00011208249271463797,
're': 0.00011208249271463797,
'flight': 0.00011208249271463797,
'over': 0.00016812373907195696,
'let': 5.6041246357318986e-05,
'you': 0.003026227303295225,
'district': 5.6041246357318986e-05,
'websit': 0.00016812373907195696,
'promot': 5.6041246357318986e-05,
'email': 5.6041246357318986e-05,
'gain': 5.6041246357318986e-05,
'36601737': 5.6041246357318986e-05,
'australian': 5.6041246357318986e-05,
'six': 5.6041246357318986e-05,
'passport': 0.00016812373907195696,
'g664891': 5.6041246357318986e-05,
'2628': 5.6041246357318986e-05,
'9mtr': 5.6041246357318986e-05,
'after': 5.6041246357318986e-05,
'usd': 5.6041246357318986e-05,
'railway': 5.6041246357318986e-05,
'day': 5.6041246357318986e-05,
'10': 0.00016812373907195696,
'link': 0.00011208249271463797,
'rest': 5.6041246357318986e-05,
'go': 0.0005604124635731899,
'14195885': 5.6041246357318986e-05,
'howev': 0.00011208249271463797,
'regard': 5.6041246357318986e-05,
'next': 0.00011208249271463797,
'i': 0.0003362474781439139,

```
'normal': 5.6041246357318986e-05,
'budget': 0.00011208249271463797,
'becaus': 5.6041246357318986e-05,
'someon': 5.6041246357318986e-05,
'comfort': 5.6041246357318986e-05,
'feedback': 5.6041246357318986e-05,
'visit': 0.00011208249271463797,
'airbu': 5.6041246357318986e-05,
'pay': 0.00022416498542927594,
'class': 5.6041246357318986e-05,
'anyway': 0.00011208249271463797,
'open': 0.00011208249271463797,
'a43': 0.00022416498542927594,
'220hkd': 5.6041246357318986e-05,
'staff': 0.0003362474781439139,
'subject': 0.00011208249271463797,
'servic': 5.6041246357318986e-05,
'those': 0.00011208249271463797,
'transit': 0.00016812373907195696,
'20': 5.6041246357318986e-05,
'sin': 0.00016812373907195696,
'shueng': 5.6041246357318986e-05,
'access': 0.00022416498542927594,
'journey': 0.0003362474781439139,
'didn': 5.6041246357318986e-05,
'specif': 0.00011208249271463797,
'tone': 5.6041246357318986e-05,
'04': 5.6041246357318986e-05,
'guidelin': 0.00022416498542927594,
'use': 0.0004483299708585519,
' ': 0.18723380407980272,
'forward': 5.6041246357318986e-05,
'tri': 5.6041246357318986e-05,
'en': 5.6041246357318986e-05,
'wah': 5.6041246357318986e-05,
'custom': 0.00011208249271463797,
'sgd': 5.6041246357318986e-05,
'yiou': 5.6041246357318986e-05,
'roomi': 5.6041246357318986e-05,
'midnight': 0.00011208249271463797,
'hui': 5.6041246357318986e-05,
'htmlani': 5.6041246357318986e-05,
'12': 0.00011208249271463797,
'www': 0.00016812373907195696,
'fine': 5.6041246357318986e-05,
'best': 0.0003922887245012329,
'join': 5.6041246357318986e-05,
'13635527': 5.6041246357318986e-05,
'36569763': 5.6041246357318986e-05,
'as': 0.0009527011880744228,
'basic': 5.6041246357318986e-05,
'old': 0.00022416498542927594,
'say': 5.6041246357318986e-05,
'MU': 5.6041246357318986e-05,
'paid': 0.00011208249271463797,
'If': 0.00028020623178659493,
'note': 5.6041246357318986e-05,
```

```
'are': 0.0014010311589329747,
'tripadvisor': 0.0005604124635731899,
'citizen': 0.00028020623178659493,
'autolink': 0.0013449899125756557,
'8': 0.00011208249271463797,
'proceed': 0.00016812373907195696,
'good': 0.00028020623178659493,
'shuttl': 5.6041246357318986e-05,
'anywher': 5.6041246357318986e-05,
'mong': 5.6041246357318986e-05,
'OP': 0.00011208249271463797,
'peopl': 0.00011208249271463797,
'2': 0.0005043712172158709,
'zhuhai2007': 5.6041246357318986e-05,
'geekzon': 5.6041246357318986e-05,
'took': 0.00022416498542927594,
'other': 0.0004483299708585519,
'pax': 5.6041246357318986e-05,
'forum': 0.00028020623178659493,
'plu': 5.6041246357318986e-05,
'min': 5.6041246357318986e-05,
'complom': 5.6041246357318986e-05,
'deck': 5.6041246357318986e-05,
'2002': 5.6041246357318986e-05,
'it': 0.0008966599417171038,
'conveni': 5.6041246357318986e-05,
'down': 5.6041246357318986e-05,
'how': 0.0003362474781439139,
'tram': 0.00011208249271463797,
'downtown': 5.6041246357318986e-05,
'becom': 5.6041246357318986e-05,
'price': 0.00028020623178659493,
'some': 0.00022416498542927594,
'think': 0.00022416498542927594,
'serv': 0.00011208249271463797,
'cab': 0.00016812373907195696,
'fair': 5.6041246357318986e-05,
'one': 0.0005604124635731899,
'folk': 5.6041246357318986e-05,
'isn': 0.00016812373907195696,
'US': 5.6041246357318986e-05,
'two': 0.00011208249271463797,
'much': 0.0003362474781439139,
'get': 0.0008406186953597848,
'freshen': 5.6041246357318986e-05,
'detail': 0.00016812373907195696,
'welcom': 5.6041246357318986e-05,
'tel': 0.00011208249271463797,
'cross': 0.00011208249271463797,
'save': 5.6041246357318986e-05,
'the': 0.006556825823806321,
'popular': 5.6041246357318986e-05,
'54': 5.6041246357318986e-05,
'is': 0.0025778973324366732,
'39': 5.6041246357318986e-05,
'look': 0.00016812373907195696,
'iphon': 5.6041246357318986e-05,
```

```
'2770': 5.6041246357318986e-05,
'fare': 0.0009527011880744228,
'quit': 5.6041246357318986e-05,
'if': 0.0005043712172158709,
'their': 0.00028020623178659493,
'visa': 5.6041246357318986e-05,
'an': 0.00022416498542927594,
'last': 5.6041246357318986e-05,
'ridicul': 5.6041246357318986e-05,
'fountain': 5.6041246357318986e-05,
'pleas': 0.00028020623178659493,
'176': 5.6041246357318986e-05,
'plan': 0.00011208249271463797,
'kln': 5.6041246357318986e-05,
'F': 5.6041246357318986e-05,
'caspar': 5.6041246357318986e-05,
'depend': 0.00016812373907195696,
'yr': 0.00011208249271463797,
'wonder': 0.00011208249271463797,
'lok': 0.00016812373907195696,
'long': 5.6041246357318986e-05,
'mainland': 5.6041246357318986e-05,
'81': 5.6041246357318986e-05,
'present': 5.6041246357318986e-05,
'more': 0.0003922887245012329,
'6': 5.6041246357318986e-05,
'believ': 5.6041246357318986e-05,
'reach': 0.00011208249271463797,
'eastern': 5.6041246357318986e-05,
'aseem01': 5.6041246357318986e-05,
'0730': 5.6041246357318986e-05,
'0940': 5.6041246357318986e-05,
'biscuit': 5.6041246357318986e-05,
'am': 0.00028020623178659493,
'230': 5.6041246357318986e-05,
'that': 0.0011208249271463797,
'bed': 5.6041246357318986e-05,
'B': 5.6041246357318986e-05,
'22': 5.6041246357318986e-05,
'today': 5.6041246357318986e-05,
'author': 5.6041246357318986e-05,
'automat': 5.6041246357318986e-05,
'pass': 5.6041246357318986e-05,
'mtr': 0.0006724949562878278,
'gave': 5.6041246357318986e-05,
'hall': 0.00022416498542927594,
'reserv': 5.6041246357318986e-05,
'web': 5.6041246357318986e-05,
'just': 0.00022416498542927594,
'while': 0.00011208249271463797,
'instead': 5.6041246357318986e-05,
'run': 0.00011208249271463797,
'exactli': 5.6041246357318986e-05,
'also': 0.0004483299708585519,
'aisl': 5.6041246357318986e-05,
'addit': 5.6041246357318986e-05,
'onli': 0.00028020623178659493,
```

'stress': 5.6041246357318986e-05,
'had': 0.00011208249271463797,
'lot': 0.00011208249271463797,
'check': 0.0003362474781439139,
'on': 0.0012329074198610178,
'togeth': 5.6041246357318986e-05,
'kowloon': 5.6041246357318986e-05,
'fortun': 5.6041246357318986e-05,
'drop': 5.6041246357318986e-05,
'albon': 5.6041246357318986e-05,
'9': 0.00011208249271463797,
'date': 5.6041246357318986e-05,
'asp': 0.00011208249271463797,
'eaton': 5.6041246357318986e-05,
'cost': 0.0005043712172158709,
'differ': 5.6041246357318986e-05,
'in': 0.0019614436225061645,
'saturday': 5.6041246357318986e-05,
'wanchai': 0.00016812373907195696,
'38406495': 5.6041246357318986e-05,
'430': 5.6041246357318986e-05,
'respect': 5.6041246357318986e-05,
'worri': 5.6041246357318986e-05,
'appreci': 0.00011208249271463797,
'secur': 5.6041246357318986e-05,
'station': 0.0003362474781439139,
'S': 5.6041246357318986e-05,
'downsid': 5.6041246357318986e-05,
'mani': 0.00022416498542927594,
'got': 5.6041246357318986e-05,
'aspth': 5.6041246357318986e-05,
'sleep': 5.6041246357318986e-05,
'schedul': 0.00011208249271463797,
'small': 5.6041246357318986e-05,
'36605457': 5.6041246357318986e-05,
'avail': 0.0003922887245012329,
'condit': 0.00016812373907195696,
'desk': 0.00011208249271463797,
'my': 0.0004483299708585519,
'recommend': 5.6041246357318986e-05,
'3': 0.00028020623178659493,
'i12409': 5.6041246357318986e-05,
'like': 0.0003922887245012329,
'G': 5.6041246357318986e-05,
'destin': 0.00011208249271463797,
'doe': 0.00022416498542927594,
'skypier': 0.0003362474781439139,
'120': 0.00011208249271463797,
'foll': 5.6041246357318986e-05,
'directli': 0.00016812373907195696,
'ha': 0.00022416498542927594,
'onboard': 5.6041246357318986e-05,
'word': 5.6041246357318986e-05,
'38327519': 5.6041246357318986e-05,
'ani': 0.0004483299708585519,
'chau': 0.00016812373907195696,
'so': 0.0003362474781439139,

```
'not': 0.0007845774490024658,
'spend': 5.6041246357318986e-05,
'ipad': 0.00011208249271463797,
'eng': 0.00028020623178659493,
'adult': 0.00022416498542927594,
'nearest': 5.6041246357318986e-05,
'great': 0.00016812373907195696,
'purchas': 0.00028020623178659493,
'fli': 0.00028020623178659493,
't1': 5.6041246357318986e-05,
'me': 0.0003922887245012329,
'29': 0.00011208249271463797,
'hk100': 5.6041246357318986e-05,
'remo': 5.6041246357318986e-05,
'arriv': 0.0003922887245012329,
'gate': 5.6041246357318986e-05,
'4': 5.6041246357318986e-05,
'sheung': 0.00022416498542927594,
'handl': 5.6041246357318986e-05,
'shell': 5.6041246357318986e-05,
'have': 0.0013449899125756557,
'such': 5.6041246357318986e-05,
'St': 5.6041246357318986e-05,
'need': 5.6041246357318986e-05,
'big': 5.6041246357318986e-05,
'particip': 5.6041246357318986e-05,
'shop': 0.0003922887245012329,
'boy': 5.6041246357318986e-05,
'a': 0.003194351042367182,
'allow': 5.6041246357318986e-05,
'yet': 5.6041246357318986e-05,
'html': 5.6041246357318986e-05,
'option': 0.00011208249271463797,
'first': 0.00011208249271463797,
'ye': 0.00011208249271463797,
'octob': 0.00011208249271463797,
'hi': 0.00011208249271463797,
'mongkok': 5.6041246357318986e-05,
'could': 0.00016812373907195696,
'info': 5.6041246357318986e-05,
'charg': 0.00016812373907195696,
'except': 5.6041246357318986e-05,
'and': 0.0028020623178659494,
'No': 0.00011208249271463797,
'meeter': 5.6041246357318986e-05,
'famili': 0.00016812373907195696,
'due': 5.6041246357318986e-05,
'bag': 5.6041246357318986e-05,
'33450793': 5.6041246357318986e-05,
'both': 5.6041246357318986e-05,
'holder': 5.6041246357318986e-05,
'co': 5.6041246357318986e-05,
'40458774': 5.6041246357318986e-05,
'My': 5.6041246357318986e-05,
'which': 0.0003922887245012329,
'xunlongferri': 5.6041246357318986e-05,
'thi': 0.0005604124635731899,
```

```
'5': 0.0005043712172158709,
'280': 0.00011208249271463797,
'mean': 0.00011208249271463797,
'famou': 5.6041246357318986e-05,
'februari': 5.6041246357318986e-05,
'htmli': 5.6041246357318986e-05,
'bonu': 5.6041246357318986e-05,
'DG': 5.6041246357318986e-05,
'repres': 5.6041246357318986e-05,
'should': 0.00022416498542927594,
'hand': 5.6041246357318986e-05,
'all': 0.0004483299708585519,
'same': 0.00016812373907195696,
'littl': 5.6041246357318986e-05,
'request': 0.00011208249271463797,
'bit': 0.00016812373907195696,
'asia': 0.00016812373907195696,
'meal': 5.6041246357318986e-05,
'mine': 5.6041246357318986e-05,
'total': 5.6041246357318986e-05,
'goe': 0.00011208249271463797,
'maximum': 0.00011208249271463797,
'300': 5.6041246357318986e-05,
'pastri': 5.6041246357318986e-05,
'parent': 5.6041246357318986e-05,
'bu': 0.0012889486662183366,
'Wu': 0.0004483299708585519,
'manag': 5.6041246357318986e-05,
'lavatori': 5.6041246357318986e-05,
'free': 0.0006724949562878278,
'night': 0.0003922887245012329,
'help': 0.00022416498542927594,
'or': 0.0006724949562878278,
'meet': 5.6041246357318986e-05,
'38368609': 5.6041246357318986e-05,
'k50': 5.6041246357318986e-05,
'child': 0.00022416498542927594,
'sit': 0.00011208249271463797,
'senior': 0.0007845774490024658,
'than': 0.00016812373907195696,
'wife': 5.6041246357318986e-05,
'watson': 5.6041246357318986e-05,
'border': 0.0003362474781439139,
'travel': 0.0003362474781439139,
'current': 5.6041246357318986e-05,
'sound': 0.00022416498542927594,
'well': 5.6041246357318986e-05,
'airsid': 5.6041246357318986e-05,
'Ma': 0.00016812373907195696,
'tst': 5.6041246357318986e-05,
'htmlairport': 5.6041246357318986e-05,
'too': 0.00011208249271463797,
'enter': 5.6041246357318986e-05,
'counter': 5.6041246357318986e-05,
'china': 5.6041246357318986e-05,
'includ': 0.00011208249271463797,
'show': 0.00016812373907195696,
```

```
'termin': 0.0003922887245012329,
'morn': 5.6041246357318986e-05,
'heard': 0.00011208249271463797,
'anoth': 5.6041246357318986e-05,
'whi': 0.00016812373907195696,
'group': 0.00011208249271463797,
'appli': 5.6041246357318986e-05,
'Lo': 0.0004483299708585519,
'duti': 0.00022416498542927594,
'januari': 5.6041246357318986e-05,
'view': 0.00011208249271463797,
'octopu': 0.0007845774490024658,
'list': 0.00028020623178659493,
'driver': 5.6041246357318986e-05,
'topic': 0.00011208249271463797,
'plenti': 0.00011208249271463797,
'hope': 5.6041246357318986e-05,
'resid': 5.6041246357318986e-05,
'vice': 5.6041246357318986e-05,
'state': 5.6041246357318986e-05,
'oct': 5.6041246357318986e-05,
'through': 0.00028020623178659493,
't': 0.0005604124635731899,
'Hi': 0.00016812373907195696,
'auckland': 5.6041246357318986e-05,
'resel': 0.00011208249271463797,
'your': 0.0008966599417171038,
'hongkongextra': 5.6041246357318986e-05,
'ask': 0.00011208249271463797,
'greeter': 5.6041246357318986e-05,
'earli': 0.00011208249271463797,
'low': 5.6041246357318986e-05,
'about': 0.0006164537099305089,
'restrict': 5.6041246357318986e-05,
'toward': 0.00011208249271463797,
'cheap': 5.6041246357318986e-05,
'Or': 5.6041246357318986e-05,
'cheaper': 0.00011208249271463797,
'deposit': 5.6041246357318986e-05,
'domcontentload': 0.0013449899125756557,
'cash': 5.6041246357318986e-05,
'convers': 0.00011208249271463797,
'vs': 5.6041246357318986e-05,
'franc': 0.00011208249271463797,
'by': 0.0005604124635731899,
'info4': 5.6041246357318986e-05,
'comprehens': 5.6041246357318986e-05,
'probabl': 5.6041246357318986e-05,
'rang': 5.6041246357318986e-05,
'pari': 5.6041246357318986e-05,
'luen': 5.6041246357318986e-05,
'I': 0.0015691548980049316,
'when': 0.00028020623178659493,
'boot': 5.6041246357318986e-05,
'60': 0.00011208249271463797,
'tsimshatsui': 5.6041246357318986e-05,
'mode': 5.6041246357318986e-05,
```

```
'thing': 5.6041246357318986e-05,
'index': 0.00011208249271463797,
'rmb': 5.6041246357318986e-05,
'carri': 5.6041246357318986e-05,
'On': 5.6041246357318986e-05,
'factor': 0.00011208249271463797,
'impress': 0.0026899798251513113,
'what': 0.00022416498542927594,
'start': 0.00011208249271463797,
'landsid': 5.6041246357318986e-05,
'hkd': 0.00022416498542927594,
'aeroplan': 5.6041246357318986e-05,
'prepar': 5.6041246357318986e-05,
'showtop': 5.6041246357318986e-05,
'choos': 0.00011208249271463797,
'0800': 5.6041246357318986e-05,
'air': 0.00016812373907195696,
'know': 0.0003362474781439139,
'kind': 5.6041246357318986e-05,
'work': 0.00011208249271463797,
'33': 0.00011208249271463797,
'unfair': 5.6041246357318986e-05,
'everyon': 5.6041246357318986e-05,
'ifc': 5.6041246357318986e-05,
'privat': 0.00011208249271463797,
'area': 0.0005043712172158709,
'hk': 0.0004483299708585519,
'transfer': 5.6041246357318986e-05,
'express': 0.0003362474781439139,
'H1': 5.6041246357318986e-05,
'At': 5.6041246357318986e-05,
'expens': 0.00011208249271463797,
'mention': 0.00011208249271463797,
'36606250': 5.6041246357318986e-05,
'biggest': 5.6041246357318986e-05,
'id': 0.00011208249271463797,
'airlin': 0.00011208249271463797,
'money': 5.6041246357318986e-05,
'post': 0.0005043712172158709,
'send': 5.6041246357318986e-05,
'pitch': 5.6041246357318986e-05,
'per': 0.00011208249271463797,
'0': 5.6041246357318986e-05,
'frill': 5.6041246357318986e-05,
'H2': 5.6041246357318986e-05,
'qualifi': 0.00016812373907195696,
'messag': 0.00016812373907195696,
'longer': 0.00011208249271463797,
'faster': 5.6041246357318986e-05,
'respons': 5.6041246357318986e-05,
'kee': 5.6041246357318986e-05,
'hard': 0.00011208249271463797,
'found': 0.00016812373907195696,
'cheer': 5.6041246357318986e-05,
'commerci': 5.6041246357318986e-05,
'control': 5.6041246357318986e-05,
'rate': 0.00016812373907195696,
```

```
'everyth': 5.6041246357318986e-05,
'yard': 5.6041246357318986e-05,
'ticket': 0.00011208249271463797,
'27': 5.6041246357318986e-05,
'31': 5.6041246357318986e-05,
'dont': 5.6041246357318986e-05,
'spent': 5.6041246357318986e-05,
'As': 0.00011208249271463797,
'later': 5.6041246357318986e-05,
'hkflat': 5.6041246357318986e-05,
'appl': 0.0002802062317865949,
'faq': 5.6041246357318986e-05,
'hello': 5.6041246357318986e-05,
'her': 5.6041246357318986e-05,
've': 5.6041246357318986e-05,
'zealand': 5.6041246357318986e-05,
'shekou': 5.6041246357318986e-05,
'21': 0.00011208249271463797,
'cheapest': 0.00011208249271463797,
'offici': 5.6041246357318986e-05,
'to': 0.0055480833893745,
'com': 0.0006164537099305089,
'us': 0.00011208249271463797,
'hongkong': 0.00011208249271463797,
'we': 0.0005043712172158709,
'busi': 0.00011208249271463797,
'guess': 5.6041246357318986e-05,
'taxi': 0.0006724949562878278,
'alight': 5.6041246357318986e-05,
'necessarili': 5.6041246357318986e-05,
'special': 0.00011208249271463797,
'item': 5.6041246357318986e-05,
'130hkd': 5.6041246357318986e-05,
'facil': 5.6041246357318986e-05,
'150': 5.6041246357318986e-05,
'acess': 5.6041246357318986e-05,
'advic': 0.0001681237390719696,
'discount': 0.0002802062317865949,
'least': 5.6041246357318986e-05,
'14189570': 5.6041246357318986e-05,
'wednesday': 5.6041246357318986e-05,
'25': 0.00011208249271463797,
'buse': 0.0001681237390719696,
'cityfly': 5.6041246357318986e-05,
'suggest': 0.00022416498542927594,
'intern': 5.6041246357318986e-05,
'pvg': 5.6041246357318986e-05,
'knowledg': 5.6041246357318986e-05,
'interchang': 5.6041246357318986e-05,
'saw': 5.6041246357318986e-05,
'advis': 0.0001681237390719696,
'a11': 0.00011208249271463797,
'railroad': 5.6041246357318986e-05,
'line': 0.0001681237390719696,
'evid': 5.6041246357318986e-05,
'minut': 0.00022416498542927594,
'direct': 0.0002802062317865949,
```

```
'Wo': 5.6041246357318986e-05,
'minimum': 5.6041246357318986e-05,
'f': 0.0026899798251513113,
'would': 0.0003922887245012329,
'card': 0.0009527011880744228,
'less': 0.00011208249271463797,
'realli': 0.00011208249271463797,
'nz': 5.6041246357318986e-05,
'220010': 5.6041246357318986e-05,
'economi': 5.6041246357318986e-05,
'take': 0.00011208249271463797,
'august': 5.6041246357318986e-05,
'trackeventonpag': 0.0013449899125756557,
'review': 0.00022416498542927594,
'been': 0.00016812373907195696,
'idevic': 5.6041246357318986e-05,
'star': 5.6041246357318986e-05,
'causebay': 5.6041246357318986e-05,
'trip': 5.6041246357318986e-05,
'38316268': 5.6041246357318986e-05,
'metro': 0.00011208249271463797,
'noisi': 5.6041246357318986e-05,
'pre': 5.6041246357318986e-05,
'airport': 0.0017933198834342075,
'come': 5.6041246357318986e-05,
'who': 0.00011208249271463797,
'We': 0.00011208249271463797,
'33993605': 5.6041246357318986e-05,
'thank': 0.0005043712172158709,
'will': 0.0008406186953597848,
'proof': 0.00011208249271463797,
'usual': 0.00022416498542927594,
'find': 0.00016812373907195696,
'around': 5.6041246357318986e-05,
'but': 0.0004483299708585519,
'nwstbu': 5.6041246357318986e-05,
'36604320': 5.6041246357318986e-05,
'lower': 0.00011208249271463797,
'ID': 5.6041246357318986e-05,
'70': 5.6041246357318986e-05,
'transport': 5.6041246357318986e-05,
'clear': 5.6041246357318986e-05,
'whichev': 5.6041246357318986e-05,
'lastli': 5.6041246357318986e-05,
'hkg': 0.00011208249271463797,
'33442400': 5.6041246357318986e-05,
'of': 0.0022976911006500783,
'see': 0.00016812373907195696,
'refer': 5.6041246357318986e-05,
'30': 5.6041246357318986e-05,
'aren': 5.6041246357318986e-05,
'A': 0.00016812373907195696,
'postlink': 0.0013449899125756557,
'without': 0.00016812373907195696,
'passeng': 0.00011208249271463797,
'section': 5.6041246357318986e-05,
'kok': 5.6041246357318986e-05,
```

```
'he': 5.6041246357318986e-05,
'g': 0.0013449899125756557,
'everybodi': 5.6041246357318986e-05,
'closer': 5.6041246357318986e-05,
'disneyland': 5.6041246357318986e-05,
'tourist': 5.6041246357318986e-05,
'went': 5.6041246357318986e-05,
'actual': 5.6041246357318986e-05,
'be': 0.0011768661735036988,
'from': 0.0017372786370768885,
'By': 5.6041246357318986e-05,
'accomplish': 5.6041246357318986e-05,
'question': 0.00011208249271463797,
'alway': 0.00011208249271463797,
'citi': 0.00016812373907195696,
'hongkongairport': 0.00011208249271463797,
'hkia': 0.0005043712172158709,
'order': 5.6041246357318986e-05,
'24': 0.00022416498542927594,
'here': 0.00011208249271463797,
'unlik': 5.6041246357318986e-05,
'extra': 5.6041246357318986e-05,
'seat': 0.00016812373907195696,
'store': 0.00016812373907195696,
'36570240': 5.6041246357318986e-05,
'did': 5.6041246357318986e-05,
'document': 0.0013449899125756557,
'now': 5.6041246357318986e-05,
'Is': 0.00016812373907195696,
'zhuhai': 5.6041246357318986e-05,
'him': 0.00011208249271463797,
'40': 5.6041246357318986e-05,
'refund': 0.00028020623178659493,
'lobbi': 5.6041246357318986e-05,
'singl': 0.00011208249271463797,
'don': 0.00028020623178659493,
'consid': 0.00011208249271463797,
'concessionari': 0.00016812373907195696,
'31st': 0.00011208249271463797,
'tax': 0.0003922887245012329,
'site': 0.00016812373907195696,
'wancahi': 5.6041246357318986e-05,
'return': 0.00016812373907195696,
'28': 5.6041246357318986e-05,
'train': 0.0003362474781439139,
'2012': 0.00022416498542927594,
'back': 5.6041246357318986e-05,
'our': 0.00016812373907195696,
'hong': 0.0004483299708585519,
'elder': 0.00016812373907195696,
'compar': 0.00011208249271463797,
'eleg': 5.6041246357318986e-05,
'33459254': 5.6041246357318986e-05,
'origin': 5.6041246357318986e-05,
'half': 5.6041246357318986e-05,
'expert': 5.6041246357318986e-05,
'even': 0.00016812373907195696,
```

```
'time': 0.0003922887245012329,
'turbojet': 5.6041246357318986e-05,
'33461109': 5.6041246357318986e-05,
'50': 0.00011208249271463797,
'wan': 5.6041246357318986e-05,
'log': 0.0013449899125756557,
'7pm': 5.6041246357318986e-05,
'thought': 5.6041246357318986e-05,
'2200they': 5.6041246357318986e-05,
'luggag': 0.00022416498542927594,
'36606937': 5.6041246357318986e-05,
'mall': 5.6041246357318986e-05,
'earlier': 5.6041246357318986e-05,
'kong': 0.0004483299708585519,
'give': 0.00011208249271463797,
'buy': 0.00011208249271463797,
'out': 0.00016812373907195696,
'answer': 5.6041246357318986e-05,
'ride': 0.00016812373907195696,
'll': 0.00011208249271463797,
'anyon': 0.00011208249271463797,
'SZ': 0.00011208249271463797,
'want': 0.00022416498542927594,
'east': 5.6041246357318986e-05,
'econom': 0.00011208249271463797,
'restaur': 5.6041246357318986e-05,
'huanggang': 5.6041246357318986e-05,
'until': 0.00011208249271463797,
'there': 0.0009527011880744228,
'1779busi': 5.6041246357318986e-05,
'curti': 0.00011208249271463797,
'esq': 5.6041246357318986e-05,
'nuanc': 5.6041246357318986e-05,
'book': 0.00011208249271463797,
'nov': 5.6041246357318986e-05,
'avenu': 0.00011208249271463797,
'year': 0.00016812373907195696,
'D': 5.6041246357318986e-05,
'excel': 5.6041246357318986e-05,
'immigr': 0.0003362474781439139,
'2011': 0.0002802062317865949,
'1am': 5.6041246357318986e-05,
'jetstar': 0.0002802062317865949,
'38315920': 5.6041246357318986e-05,
'addeventlisten': 0.0013449899125756557,
'38324674': 5.6041246357318986e-05,
'szx': 5.6041246357318986e-05,
'where': 0.0003362474781439139,
'fee': 5.6041246357318986e-05,
'local': 0.00011208249271463797,
'lifestyl': 5.6041246357318986e-05,
'wa': 0.0003922887245012329,
'soy': 5.6041246357318986e-05,
'loung': 0.00016812373907195696,
'tell': 0.00011208249271463797,
'ferri': 0.0006724949562878278,
'most': 5.6041246357318986e-05,
```

```
            'close': 0.00022416498542927594,
            'hotel': 0.0004483299708585519,
            'follow': 0.00016812373907195696,
            'concern': 5.6041246357318986e-05,
            'may': 0.0005604124635731899,
            'sinc': 0.00028020623178659493,
            'To': 0.00016812373907195696,
            'It': 0.0005043712172158709,
            'wine': 0.00028020623178659493,
            'edit': 0.00022416498542927594,
            '33469537': 5.6041246357318986e-05,
            'kid': 0.00022416498542927594,
            'although': 5.6041246357318986e-05,
            '5': 0.0003362474781439139,
            'via': 0.00011208249271463797,
            'beat': 5.6041246357318986e-05,
            'whether': 5.6041246357318986e-05,
            'far': 5.6041246357318986e-05,
            'sure': 5.6041246357318986e-05,
            'function': 0.0026899798251513113,
            '02': 5.6041246357318986e-05,
            'AM': 5.6041246357318986e-05}
```

This is the final result of the Term frequency-inverse document frequency (TF-IDF) analysis. the below dataframe represents the frequency of the occurance of a word in the posts.

In [76]: `df([tfBow])`

Out[76]:

	0	02	04	0730	0800	0940	1	10	
0	0.187234	0.000056	0.000056	0.000056	0.000056	0.000056	0.00028	0.000168	0.00

1 rows × 820 columns
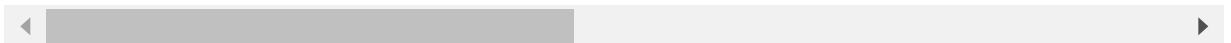
i have filter data at all the level, now the next thing i have to do is to do sentiment analyis of the posts of the people on the website. below are the code which is used to do sentiment analysis of the post and after analysis i have ploted the sentiment percentage with the help of pie chart, the sentiment is catoriged into very positive,positive, very negative, negative, neutral.

In [91]:
```python
#**************Sentiment Analysis*******************
from textblob import TextBlob, Word
from textblob.sentiments import NaiveBayesAnalyzer

text =result2
dataset = []
#reviewing
for i in range(len(text)):
    textB = TextBlob(text[i])
    sentiment = textB.sentiment.polarity
#     dataset.append(sentiment)
    if sentiment< -0.5:
        sentiment = "Very Negative"
        dataset.append(sentiment)
    elif -0.5<=sentiment< 0:
        sentiment = "Negative"
        dataset.append(sentiment)
    elif sentiment >=0.5:
        sentiment = "Very Positive"
        dataset.append(sentiment)
    elif 0<sentiment <0.5:
        sentiment = "Positive"
        dataset.append(sentiment)
    elif sentiment == 0:
        sentiment = "Neutral"
        dataset.append(sentiment)
# dataset
```

In [96]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
p = dataset.count("Positive")
vp = dataset.count("Very Positive")
n = dataset.count("Negative")
vn = dataset.count("Very Negative")
n0 = dataset.count("Neutral")
label = "Very Positive.",'Positive', "Very Negative",'Negative', "Neutral"
size = [vp,p, vn,n, n0]
explode = (0.0, 0.0, 0.1,0.0,0.0)
colors = ["yellowgreen", "lightcoral", "gold",'red','skyblue']
plt.pie(size, explode = explode, colors = colors, labels = label, autopct='%1.
1f%%', shadow=True,startangle = 140)
plt.axis("Equal")
plt.title("Sentiment Analysis")
plt.show()
```

Sentiment Analysis

Neutral

Negative

15.3%

16.9%

Very Positive.

Very Negative

10.2%

0.0%

57.6%

Positive

Now its time to plot the word cloud this word cloud will show the words which are used in the sentance, the words which is used most will be of bigger size.

In [136]:

```python
#wordcloud
#filtered_sentence - list of filtered words, without stopwords
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt

stopwords = set(STOPWORDS)
words = ''
for w in filtered_sentence:
    words += ' '+ str(w)

wordcloud = WordCloud(width = 800, height = 800,
                background_color ='white',
                stopwords = stopwords,
                min_font_size = 10).generate(words)

# plot the WordCloud image
plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)

plt.show()
```
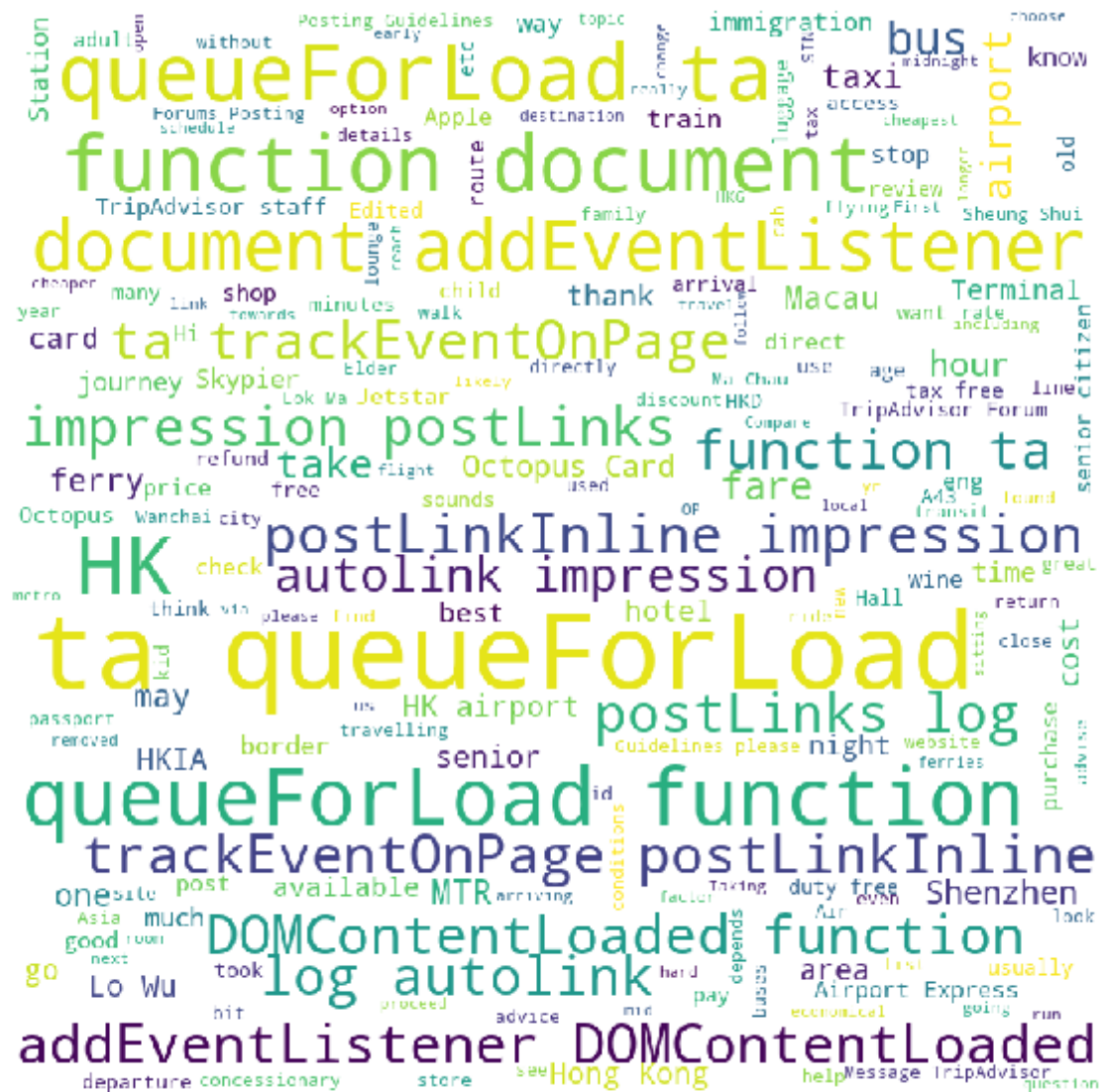
```
In [ ]:  #*********************************Some extra anallysis for bettere prediction**
         ***********************
```

All the task as suggested is done, now i have done some feature engineering to see that from which country most of the post is made, to do this i have ploted an word cloud for the locaton, in this word cloud the country name which appers bigger is in the post maximum of the time

In [3]:
```python
import pandas as pd
reploc = pd.read_csv('repLoc.csv')

# reploc.drop(reploc['selection1_reply_no_location']=='Hong Kong, China')
# dloc ={}
# dl = []
# for l in reploc['selection1_reply_no_location']:
#     if l != 'Hong Kong, China':
#         dl.append(l)

# dloc['selection1_reply_no_location'] = dl
# df(dloc)
```

In [4]:
```python
reloc = reploc[reploc.selection1_reply_no_location != 'Hong Kong, China ']
reloc1 = reloc[reploc.selection1_reply_no_location != 'Hong Kong, China']
reloc1
```

```
c:\users\anand\appdata\local\programs\python\python36\lib\site-packages\ipyke
rnel_launcher.py:2: UserWarning: Boolean Series key will be reindexed to matc
h DataFrame index.
```

Out[4]:

	selection1_reply_no_location
0	Christchurch, New...
1	Elk Grove...
3	Dublin, California
4	Elk Grove...
5	Macau, China
6	Macau, China
7	Macau, China
8	Macau, China
9	Macau, China
10	Macau, China
11	Macau, China
12	Macau, China
13	Macau, China
14	Singapore
15	Dublin, California
16	Elk Grove...
17	Singapore
18	Dublin, California
19	Elk Grove...
20	Singapore
31	Singapore, Singapore
32	New Forest, United...
33	Dublin, California
34	Melbourne, Australia
35	Sydney Australia
36	New Forest, United...
37	Nova Scotia
38	Singapore, Singapore
39	New Forest, United...
49	Auckland, New...
...	...
147536	Perth, Australia
147537	Elk Grove...
147538	Kuala Lumpur...
147539	Sydney, Australia

	selection1_reply_no_location
147541	Perth, Australia
147551	Mumbai, India
147552	Dublin, California
147565	UK
147567	Xiamen, China
147568	UK
147570	UK
147583	Brisbane, Australia
147585	Nha Trang, Vietnam
147586	Los Angeles...
147587	Los Angeles...
147588	Brisbane, Australia
147589	Melbourne, Australia
147590	Los Angeles...
147591	Melbourne, Australia
147592	Llanbrynmair...
147593	Tarentum...
147594	Ho Chi Minh City, Vietnam
147595	Ho Chi Minh City, Vietnam
147596	Ho Chi Minh City, Vietnam
147597	Ho Chi Minh City, Vietnam
147598	Ho Chi Minh City, Vietnam
147599	Ho Chi Minh City, Vietnam
147600	Ho Chi Minh City, Vietnam
147601	Ho Chi Minh City, Vietnam
147602	Ho Chi Minh City, Vietnam

75062 rows × 1 columns

```
In [5]: reloc1.to_csv('repLoc1.CSV')
```

I have mede an word cloud for the location to see which is mostly engeged with the H K airport

In [123]:
```python
#wordcloud for location
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt

stopwords = set(STOPWORDS)
words = ''
for w in reloc1['selection1_reply_no_location'][:]:
    words += ' '+ str(w)

wordcloud = WordCloud(width = 800, height = 800,
                background_color ='white',
                stopwords = stopwords,
                min_font_size = 10).generate(words)

# plot the WordCloud image
plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)

plt.show()
```
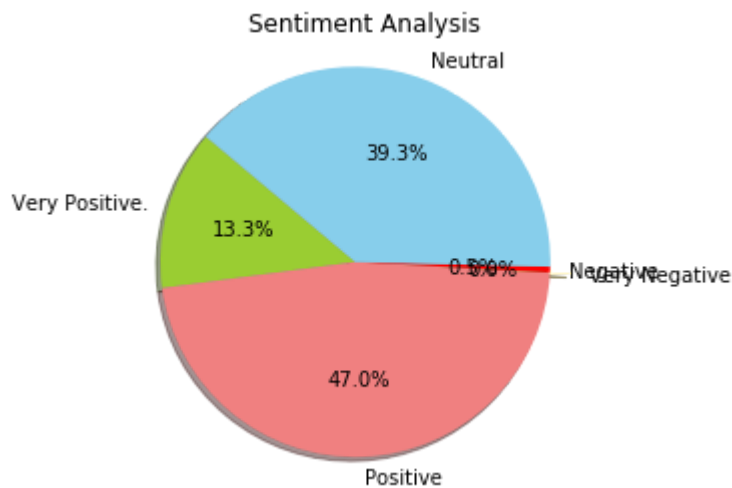
The e other feature engineering which i have done is that i have done an analysis of the title of the post which the people made on the trip advisor, i have done sentiment analysis and also ploted word cloud to see the point of concen of people.

In [127]:

```python
#******************reply topic**************************Analysis
repltopic = pd.read_csv('replTopic.csv')
# repltopic
result2 = repltopic['selection1_reply_no_topic'][:]
#**************Sentiment Analysis*******************
from textblob import TextBlob, Word
from textblob.sentiments import NaiveBayesAnalyzer

text =result2
dataset = []
#reviewing
for i in range(len(text)):
    textB = TextBlob(text[i])
    sentiment = textB.sentiment.polarity
#    dataset.append(sentiment)
    if sentiment< -0.5:
        sentiment = "Very Negative"
        dataset.append(sentiment)
    elif -0.5<=sentiment< 0:
        sentiment = "Negative"
        dataset.append(sentiment)
    elif sentiment >=0.5:
        sentiment = "Very Positive"
        dataset.append(sentiment)
    elif 0<sentiment <0.5:
        sentiment = "Positive"
        dataset.append(sentiment)
    elif sentiment == 0:
        sentiment = "Neutral"
        dataset.append(sentiment)
# dataset
```

In [128]:
```python
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
p = dataset.count("Positive")
vp = dataset.count("Very Positive")
n = dataset.count("Negative")
vn = dataset.count("Very Negative")
n0 = dataset.count("Neutral")
label = "Very Positive.",'Positive', "Very Negative",'Negative', "Neutral"
size = [vp,p, vn,n, n0]
explode = (0.0, 0.0, 0.1,0.0,0.0)
colors = ["yellowgreen", "lightcoral", "gold",'red','skyblue']
plt.pie(size, explode = explode, colors = colors, labels = label, autopct='%1.
1f%%', shadow=True,startangle = 140)
plt.axis("Equal")
plt.title("Sentiment Analysis")
plt.show()
```

Sentiment Analysis

Neutral

39.3%

Very Positive.

13.3%

0.5%0%   Negative egative

47.0%

Positive

word cloud for the people point of concern.

In [131]:
```python
stopwords = set(STOPWORDS)
words = ''
for w in repltopic['selection1_reply_no_topic'][:]:
    words += ' '+ str(w)

wordcloud = WordCloud(width = 800, height = 800,
                background_color ='white',
                stopwords = stopwords,
                min_font_size = 10).generate(words)

# plot the WordCloud image
plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)

plt.show()
```


In [ ]: