

TEAM 1

AIR FRANCE

IS BACK!



09 00 47 28

RAVEL ST



AIRFRANCE
ARRIVED
7-NOV-202



"France is in the Air"



You chose us because
you know we deliver
and you trust our
expertise.

-TEAM 1



AIRMAN
DEPARTED
7-NOV-2021

MEET THE TEAM

- Benedicta Oruerio
- Mudabbir Khawaja
- Pieter B. Krommenhoek
- Shambhavi Raj
- Uresha Ojha



UNDERSTANDING THE DATA



- Air France was founded in 1933 as a merger of five French airlines.
- The airline grew due to more international travels after WW II.
- It has a history of bankruptcies and low returns.



09 00 47

TRAVEL STAMP



TEAM 1

No data is reliable unless
analyzed by a data pirate.

PROCESS

MASSAGING/CLEANING UP THE DATA

- We first massaged and cleaned the data
- Secondly, broke down each search phrase into keywords.
- Lastly took out the space and other strings to make each string a stand-alone word

word <chr>	freq <dbl>
cheap	787
flight	775
paris	749
france	736
ticket	604
air	483
flights	464
airline	396
international	381
airfare	379

1–10 of 10 rows

PROCESS

MASSAGING/CLEANING UP THE DATA



- We created a word cloud from the data to show most frequently searched keywords
 - We set a condition of a minimum frequency of 4 that should not be displayed
 - We can see that Cheap, Paris and Tickets are the top searched keywords

- To find the relative association of the top 3 keywords:

1) Cheap

2) Paris

3) Ticket

```
findFreqTerms(dtm, lowfreq = 5)

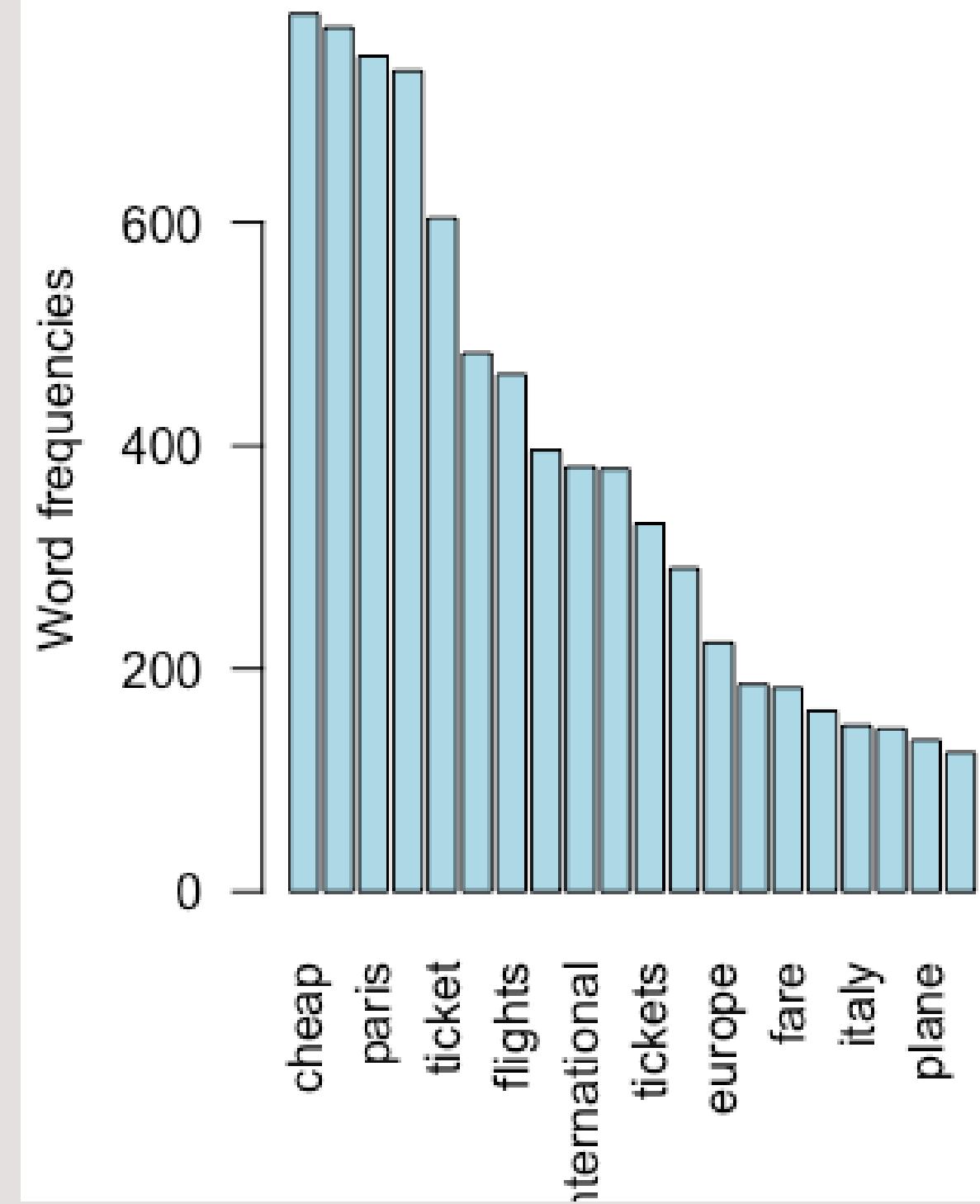
#correlation between words
findAssocs(dtm, terms = "cheap", corlimit = 0.05)
findAssocs(dtm, terms = "paris", corlimit = 0.05)
findAssocs(dtm, terms = "ticket", corlimit = 0.05)

barplot(d[1:20,]$freq, las = 2, names.arg = d[1:20,]$word,
        col ="lightblue", main ="Most frequent words",
        ylab = "word frequencies")

#filtering first words
first_words <- d[1:20,]

##### Analysis #####
#Optimize ROA
#Return on advertisement spend
#Revenue/Cost acquisition
```

Most frequent words



Descriptive
Analysis

BUSINESS INSIGHTS

Code snippets describe how we derived the business insights...

REGRESSION INSIGHTS

Clicks and Impressions = Full Business Success.
Strong correlation.

CONFUSION MATRIX INSIGHTS

99% Business Success Rate.
The prediction shows zero false predictions.

(Dashboard with insights)

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-4.88242	2.85362	-1.711	0.0871	.
clicks	101.38733	21.78235	4.655	3.25e-06	***
`Avg. cost per click`	-0.22739	1.05827	-0.215	0.8299	
Impressions	-11.28841	2.42448	-4.656	3.22e-06	***
Amount	-0.01189	0.04122	-0.288	0.7730	

Confusion Matrix and statistics

Prediction	Reference	
	0	1
0	3432	0
1	0	1078

Accuracy : 1
95% CI : (0.9992, 1)
No Information Rate : 0.761
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 1
McNemar's Test P-Value : NA



Thank you for having
us onboard

www.datapiratesteam1.com