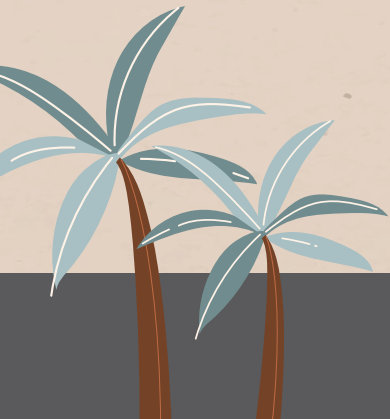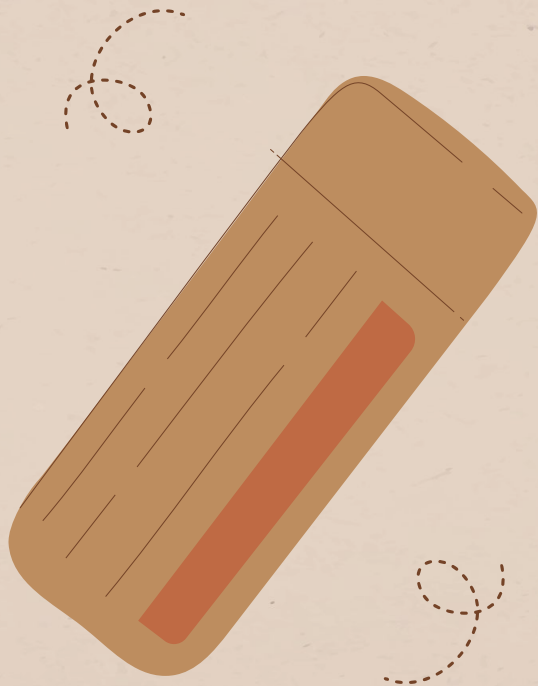# Travel Australia

Prepared for :
Tourism Australia
Senior Management

29 Jul 2022
Choo Wende

Agenda

1. Introduction
Background & Problem Statement

2. Data Analysis
Top Occuring Words & Sentiment Analysis

3. Classification Model
Model Selection & Evaluation

4. Conclusion
Summarise Findings & Recommendations

01

Introduction

Background & Problem Statement

Great Ocean Road

Travellers

Background

- Boost Australia Tourism Sector
- Promote roadtrips in Australia
- Identify travel related topics that are important to travellers.

Covid

Risk of further outbreak

Safety & Health

travellers' concern around
safety and health

Flight

availability of long-haul
international flights

02

Data Analysis

Top Occuring Words & Sentiment Analysis

Shortlisting Subreddits

01

r/travel

6.7M Members

02

r/roadtrips

1.9M Members

03

r/TravelHacks

353K Members

04

r/TravelNoPics

21K Members

05

r/AustraliaTravel

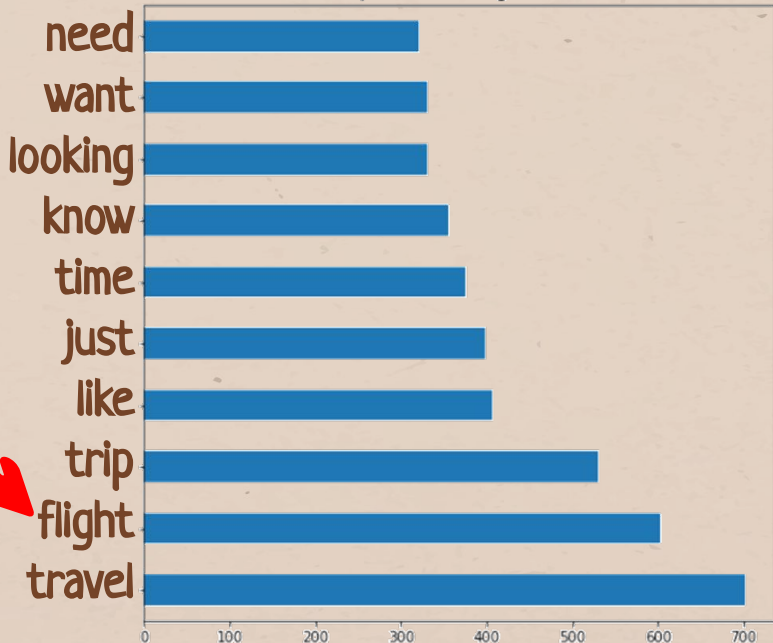732 Members

Top 10 recurring Words

r/TravelHacks

Top 10 Occuring Words

Word	
need	
want	
looking	
know	
time	
just	
like	
trip	
flight	
travel	

(x-axis: 0, 100, 200, 300, 400, 500, 600, 700)

r/roadtrip

Top 10 Occuring Words

Word	
days	
like	
car	
driving	
time	
day	
way	
drive	
road	
trip	

(x-axis: 0, 200, 400, 600, 800, 1000, 1200)

ROME
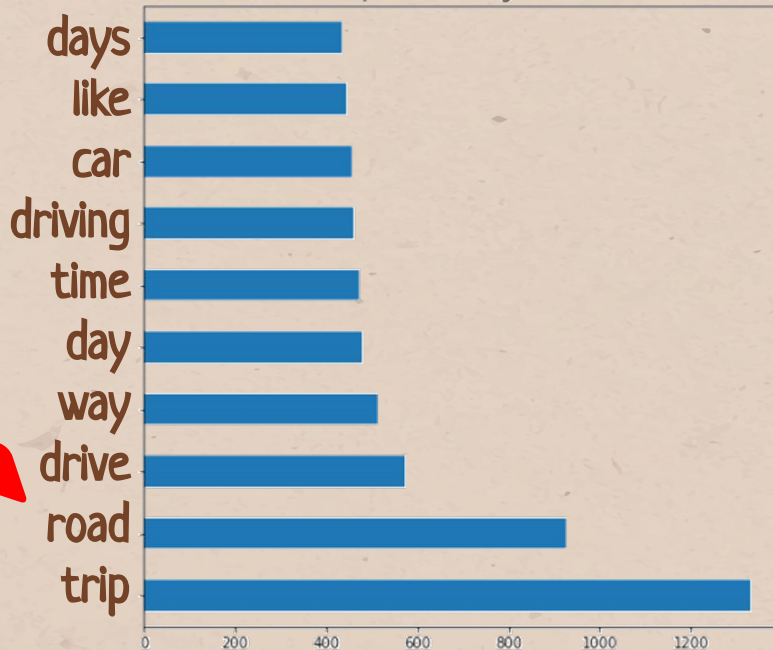
Sentiment Analysis

r/TravelHacks

r/roadtrip

- Incur extra cost of car rental
- Worried about ART test result

- Worried about safety of town concern about effects of long drive
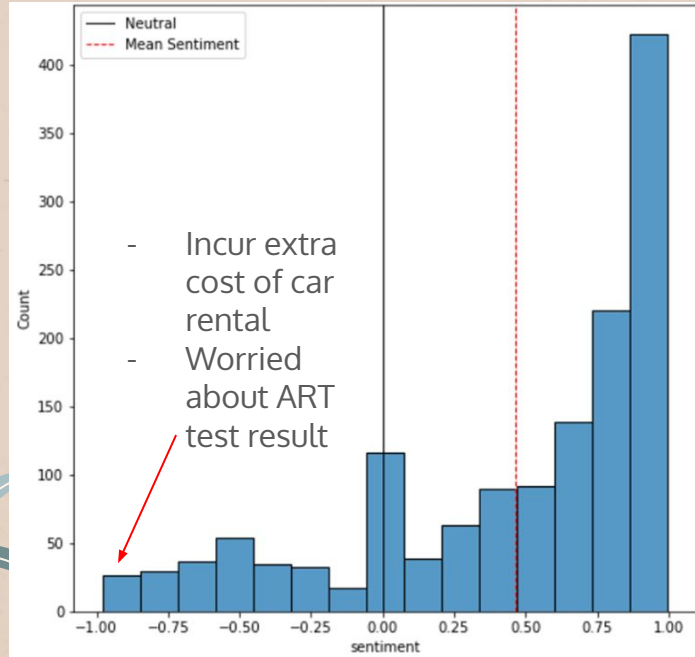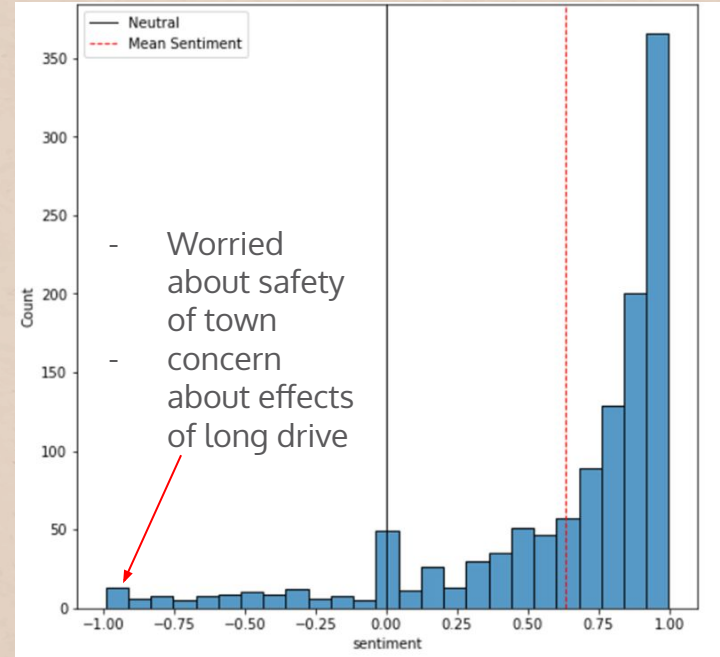
03

Classification Model

Model Selection & Evaluation

Methodology

Data Set

Training Set	Testing Set

- Train model
- Measure Accuracy

- Test trained model
- Measure Accuracy
- Compute true positive rate
- Compute true negative rate

"Model with the highest accuracy, true positive rate and true negative rate will be proposed."

Three models

01

Base model: Logistics Regression

- Use default count vectorizer
- Without optimisation

02

Logistics Regression (Optimised)

- Performed best with TF-IDF Vectorizer
- Hyperparameters Tuning

03

Naive-Bayes Model (Optimised)

- Performed best with Count Vectorizer
- Hyperparameters Tuning

Models Summary

Baseline Accuracy: <u>0.541</u>
(if predict all submissions from TravelHacks)

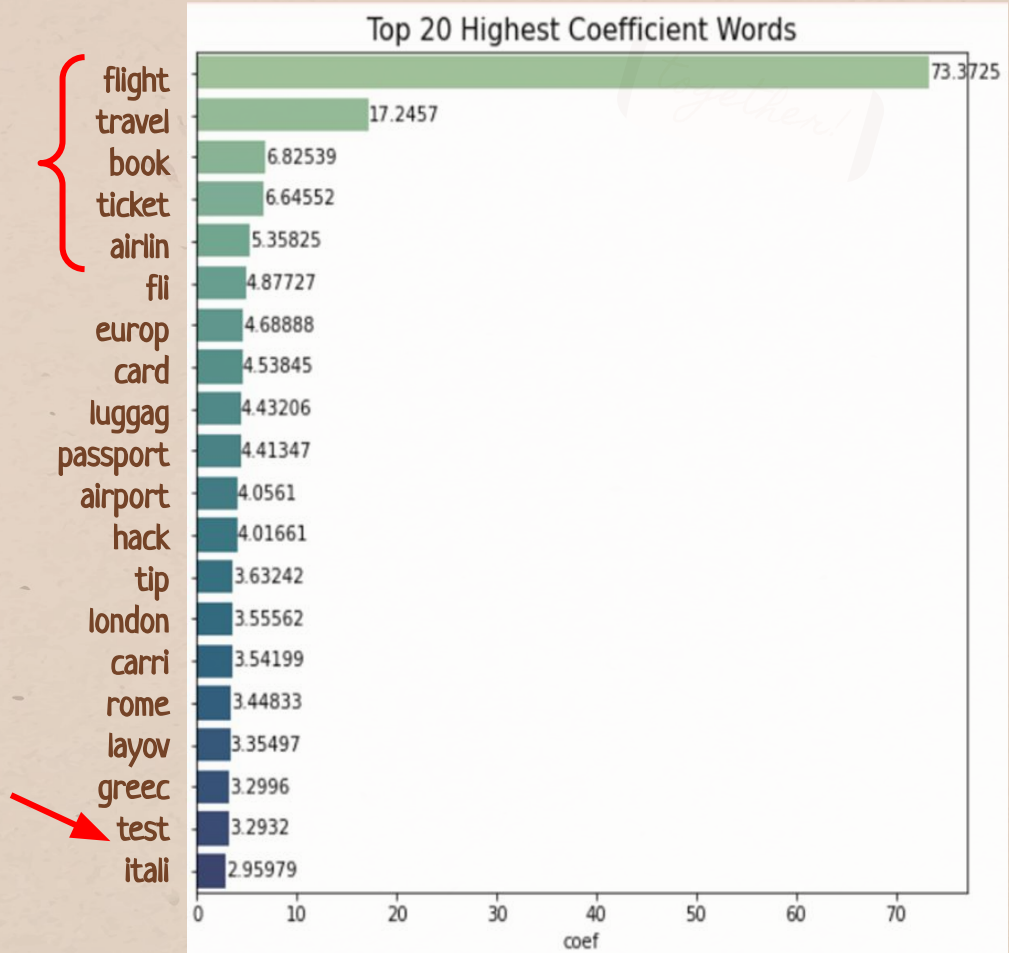| Model	Cross Validated Score	Training Accuracy	Accuracy	Sensitivity	Specificity	ROC
				True positive rate	True negative rate	
logreg_base	0.898408	0.997111	0.905769	0.928315	0.879668	0.964024
logreg_tvec (gs_logreg_model)	0.922237	0.953779	0.919231	0.935484	0.900415	0.969363
nb_cvec (gs_nb_model)	0.932595	0.932595	0.898077	0.870968	0.929461	0.959904

92 %
Of the variance in Subreddit classifications may be explained by model 2.

TravelHacks

The top five words for TravelHacks are 'flight', 'travel', 'book', 'ticket' and 'airline'.

Partner airlines to attract travellers and increase availability of flight.

Providing ART/PCR testing services at airport to meet test requirements

Top 20 Highest Coefficient Words

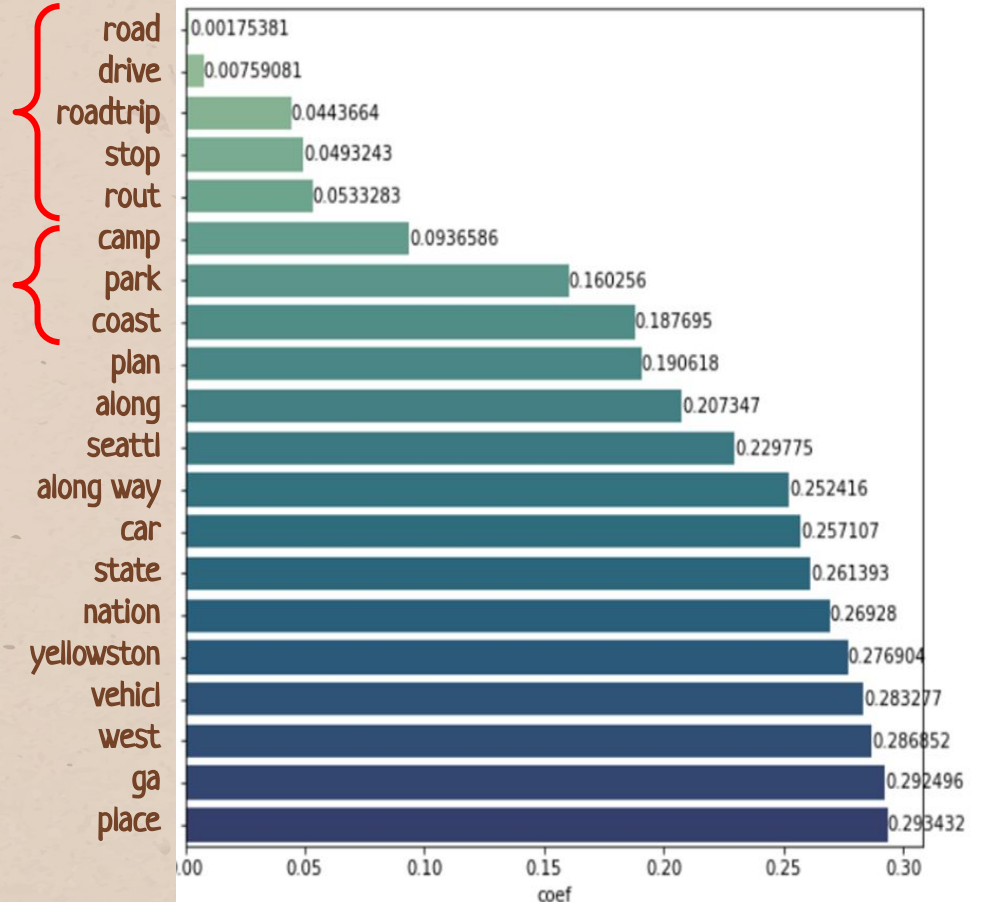| Word	coef
flight	73.3725
travel	17.2457
book	6.82539
ticket	6.64552
airlin	5.35825
fli	4.87727
europ	4.68888
card	4.53845
luggag	4.43206
passport	4.41347
airport	4.0561
hack	4.01661
tip	3.63242
london	3.55562
carri	3.54199
rome	3.44833
layov	3.35497
greec	3.2996
test	3.2932
itali	2.95979

roadtrip

The top five words for roadtrips are 'road', 'drive', 'roadtrip', 'stop' and 'route'.

Other words of interests:
'camp', 'park', 'coast'

Establish resting areas with amenities or camping areas around nature parks or along coastal roads, e.g. the Great Ocean Road.

20 Lowest Coefficient Words

Word	coef
road	0.00175381
drive	0.00759081
roadtrip	0.0443664
stop	0.0493243
rout	0.0533283
camp	0.0936586
park	0.160256
coast	0.187695
plan	0.190618
along	0.207347
seattl	0.229775
along way	0.252416
car	0.257107
state	0.261393
nation	0.26928
yellowston	0.276904
vehicl	0.283277
west	0.286852
ga	0.292496
place	0.293432

04

Conclusion

Summarise Findings & Recommendations

Covid

Risk of further outbreak

- Policy to recoup some of the costs incurred or allow trip to delay
- Provide testing services at airport

Safety & Health

travellers' concern around safety and health

- Review the typical roadtrip routes pass through safer areas
- Attractions/ Pit stop with amenities/ Camping areas

Flight

availability of long-haul international flights

- Strategic partnership with airline to offer discounted tickets and increase availability of flight.

Future Works

BE

Remove Advertisement/ Spam

- Some submissions are advertisement/ spam
- Noise that add to the inaccuracies.

Select post of certain length

- Some submissions are short
- May not be meaningful

Review Stopwords

- r/roadtrip is US centric
- Include US-related words

Thanks!

together!