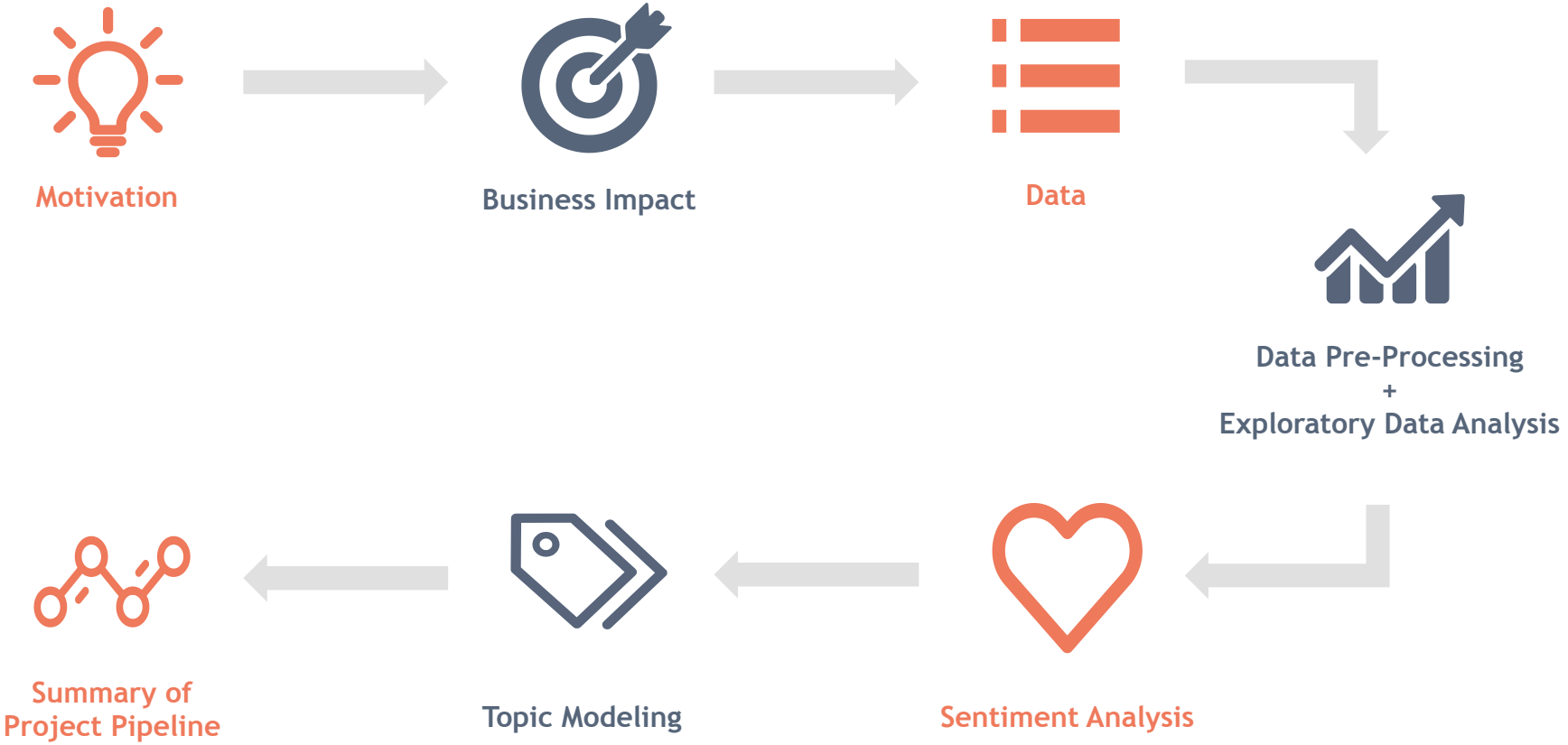
DSI-7
Capstone Project

Optimising Customer Service
&
Operational Efficiency in the Hotel Industry

By: Vincent Kwan Wen Seng

Overview


Motivation
Business Impact
Data
Data Pre-Processing
+
Exploratory Data Analysis
Sentiment Analysis
Topic Modeling
Summary of
Project Pipeline

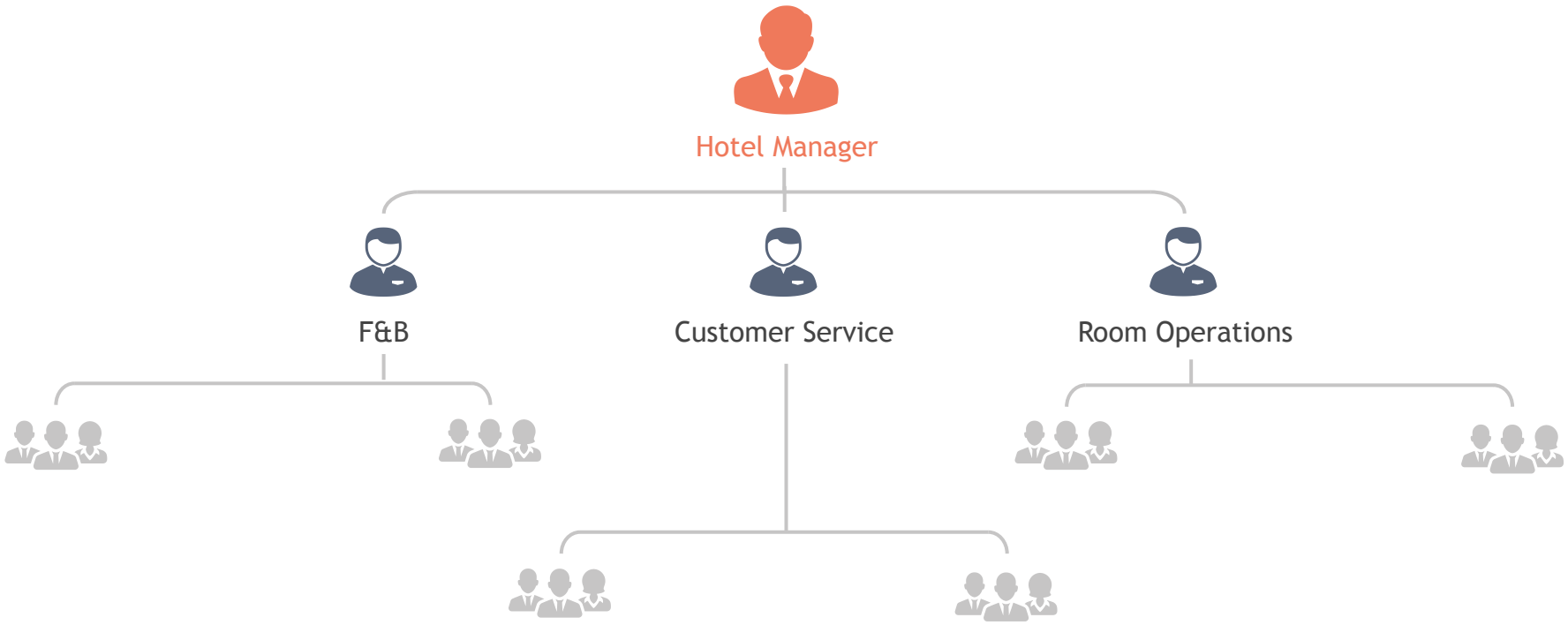
Motivation

Motivation

Can Data Science empower hoteliers with the ability to quantify customer satisfaction?

Can it help to optimise customer service delivery & allow managers to focus on areas where it matter most?

Motivation



Business Impact

- Remove man-hours required to process thousands of reviews across different sites
- Facilitate business units in decision-making
- Quantify customers' satisfaction. Maintain quality customer service
- Improve customer experience by addressing concerns effectively & in good time

Natural Language Processing:

Sentiment Analysis

+

Topic Modeling

Dataset

The Dataset (Mandarin Orchard Hotel)

Python Web Scraping – BeautifulSoup, Xpath, Selenium

Observations

20K

Features

7

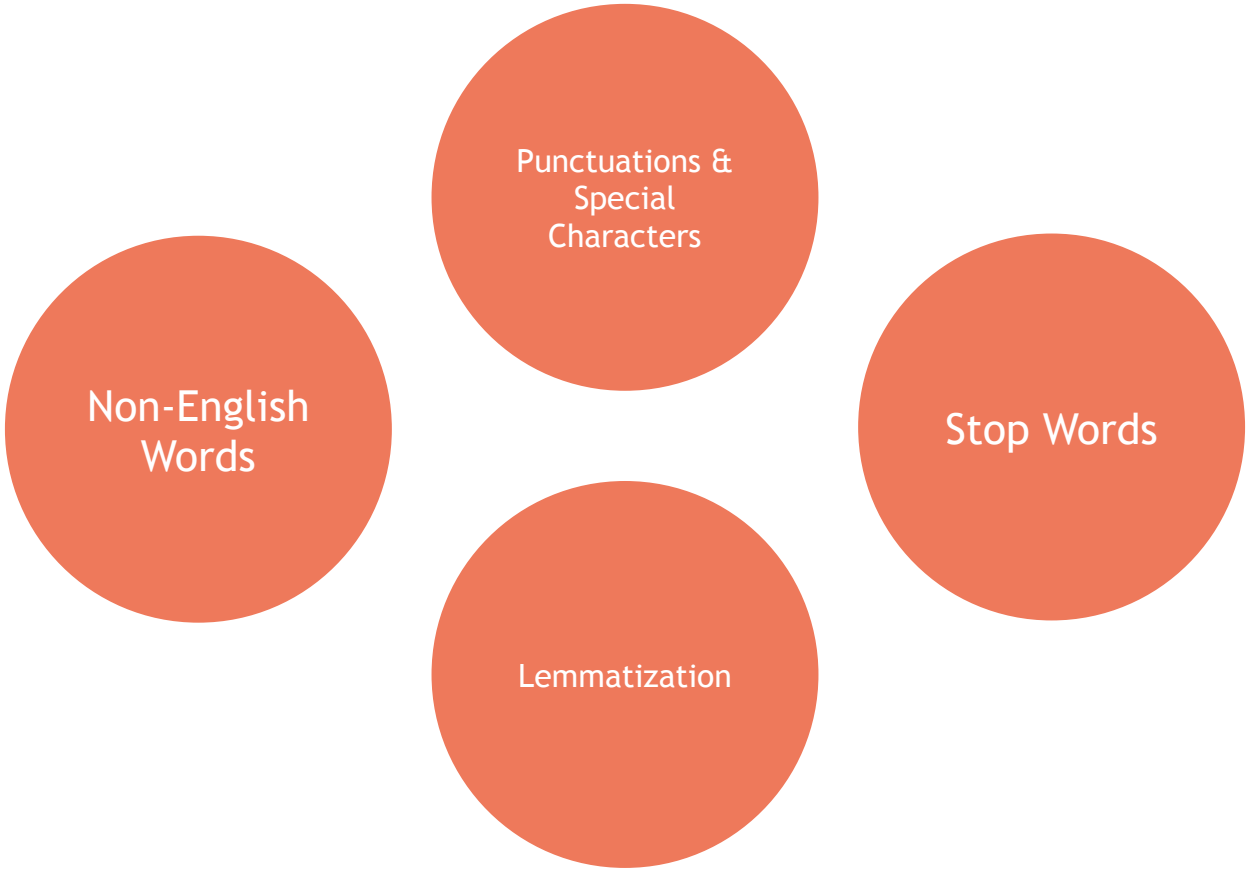
- Reviews & Dates
- Hotel's Response
- Purpose of Stay
- Reviewer's Nationality

Data Pre-processing

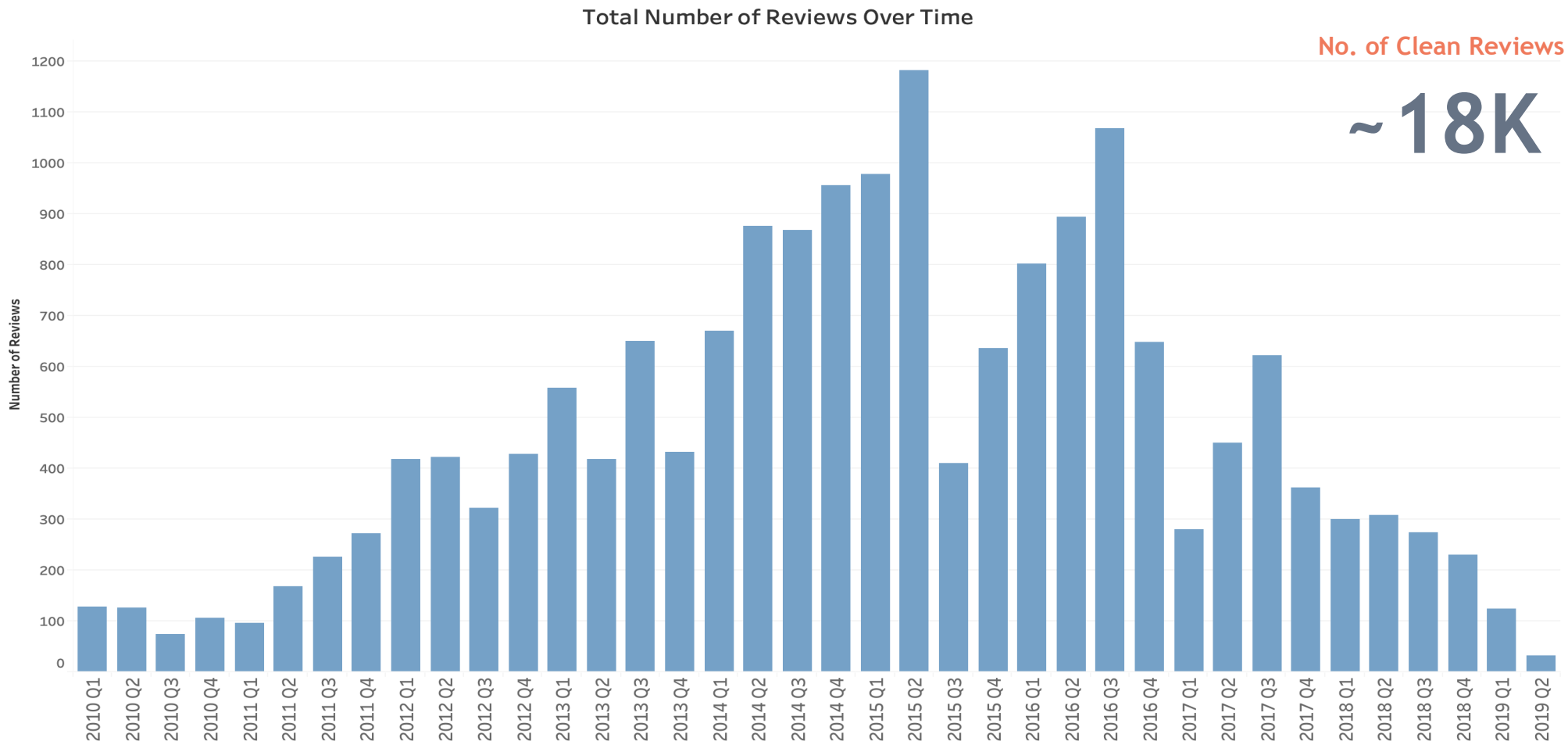
&

Exploratory Data Analysis

Data Pre-processing



Exploratory Data Analysis



No. of Clean Reviews

~18K

Exploratory Data Analysis

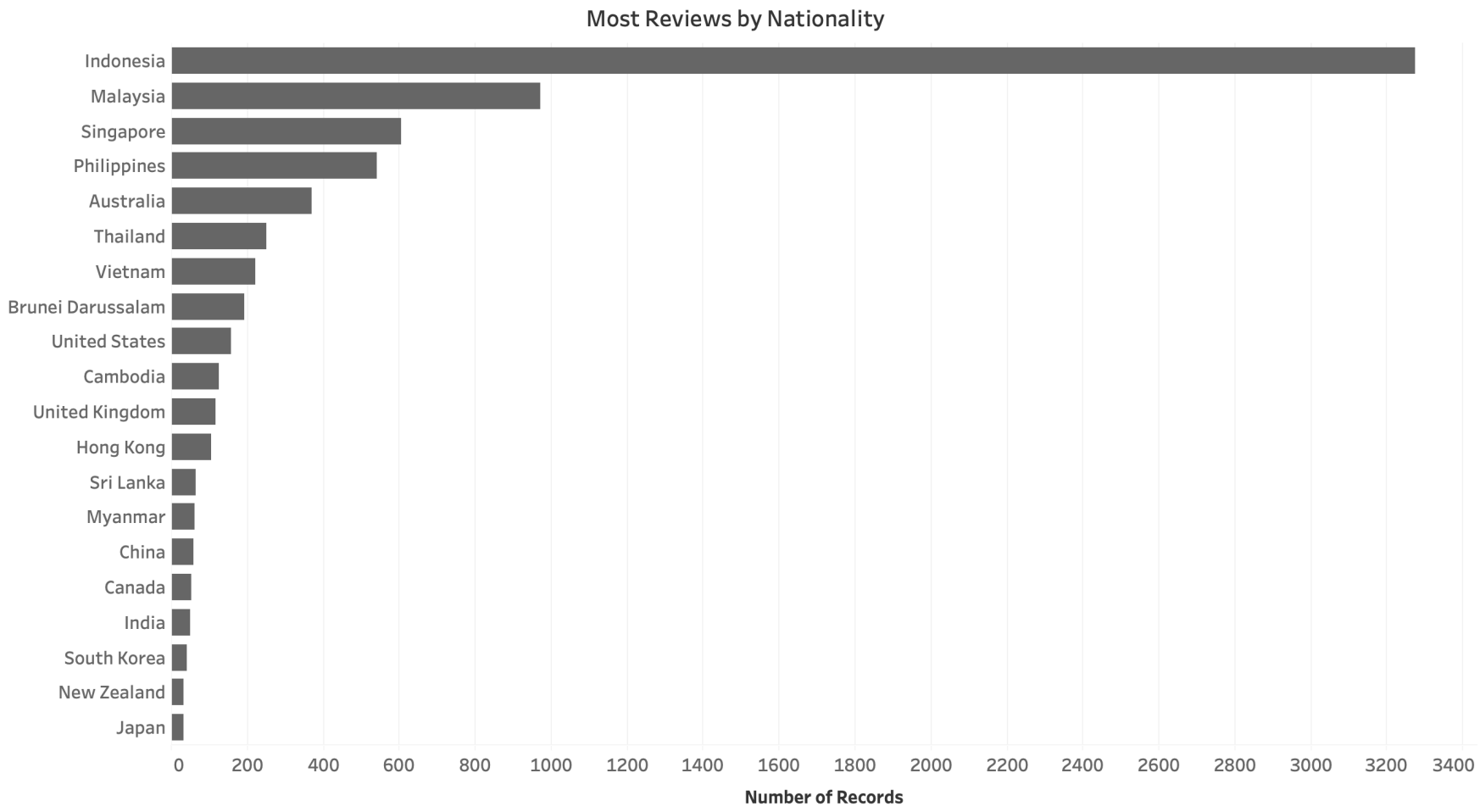
Leisure

74%

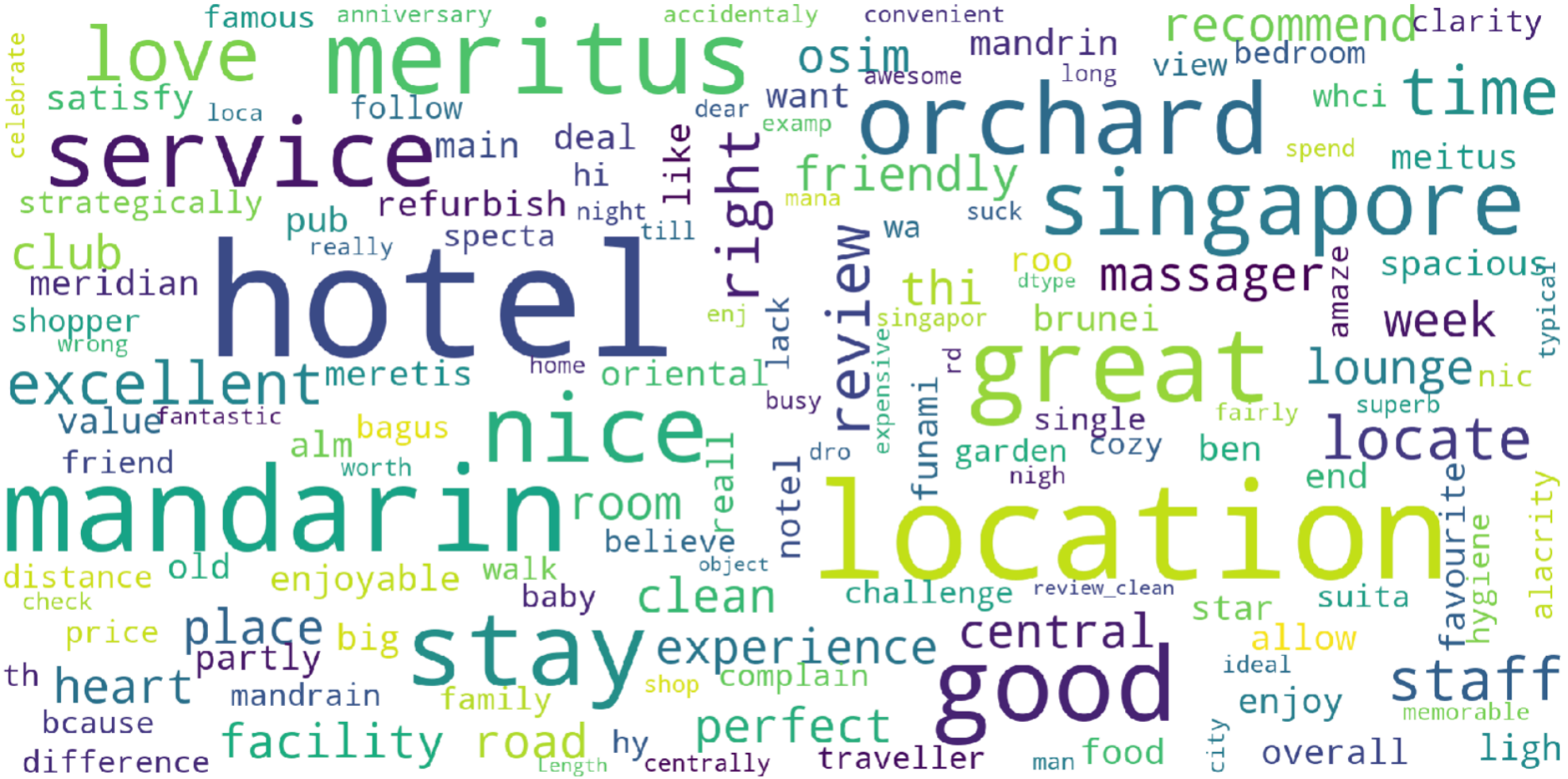
Business

26%

Exploratory Data Analysis



Exploratory Data Analysis



Sentiment Analysis

Sentiment Analysis

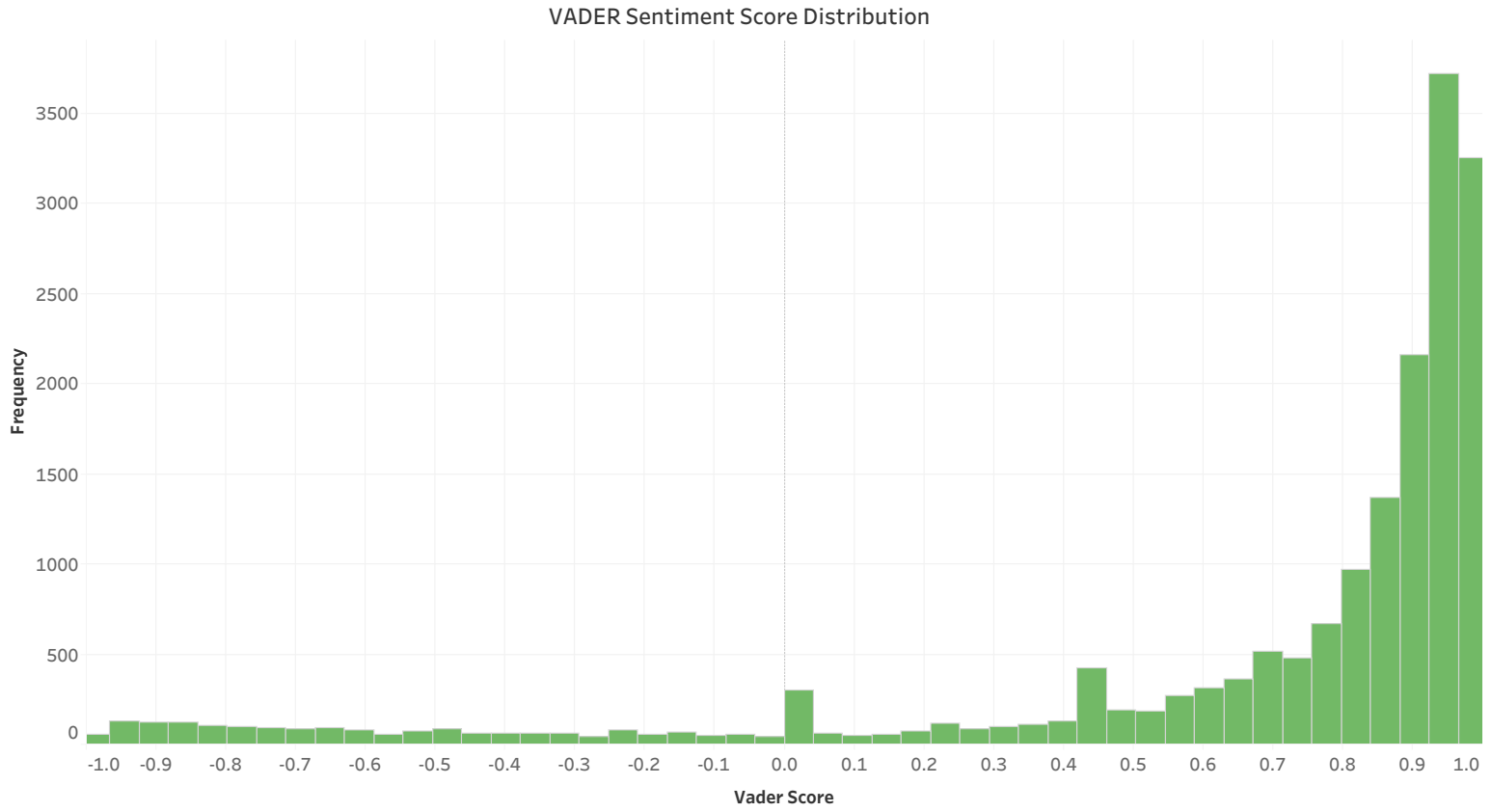
- **TextBlob**
- **VADER** - (Valence Aware Dictionary and sEntiment Reasoner)

Sentiment Analysis

VADER

+ve:	15654 (88%)
Neutral:	366 (2%)
-ve:	1778 (10%)

Tolerance of Bad Review can be adjusted



Sentiment Analysis

"UNorganised, old + dilapidated hotel. LOUSY ! Stayed in this hotel for a night and then moved to Hyatt which is just down the road and about 50 dollar more. It worth every penny at the hyatt. The meritius mandarin looked so old and the ***toilet they have is DISGUSTING****. The flooring and the marble furnished they have are old and looks uncleaned, and this is at the "premier room" , whic i paid $340 sing ++ for one night. Terrible, the hotel is so un-organized, you will have to go through a set of shops before getting on the set of lifts that took me to my room.* ***Absolutely rubbish and very poor management, the environment made me feels that i am in a FLEA market rather than a 5 star hotel****. If you are thinking to stay in Orchard Road, pick either Marriot or Hyatt. They are both way better !"*

TextBlob: -0.1 **VADER: -0.9**

Topic Modeling

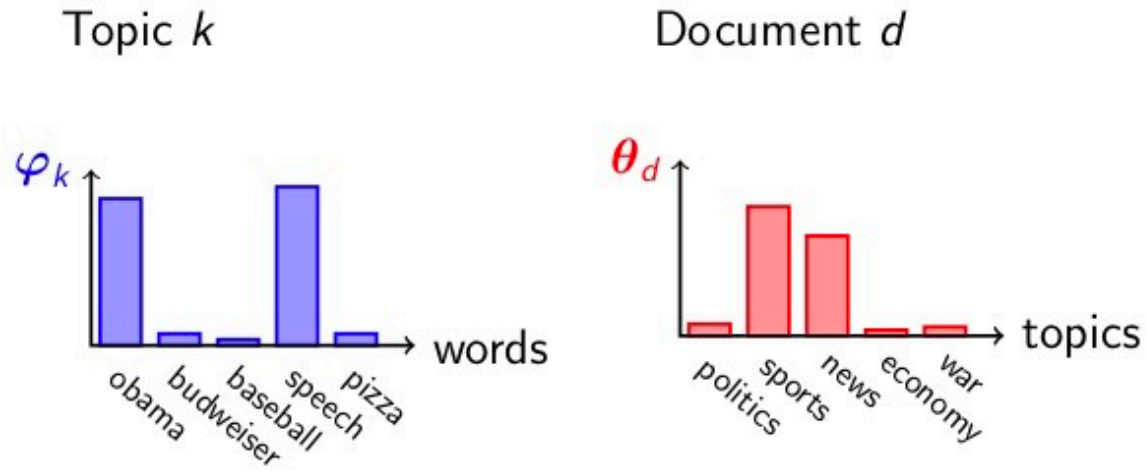
Topic Modeling

gensim

Latent Dirichlet Allocation (LDA)

Hidden

A Type of Probability Distribution



Topic Modeling

gensim

Latent Dirichlet Allocation (LDA)

Input:

1. **Dictionary of Tokens**
2. **Document Term Matrix (BoW)**

Part-of-Speech:	SpaCy Nouns Only
Stop Words:	SpaCy + Custom
Lemmatization:	SpaCy

3.

Number of Topics:	3 to 6
Passes:	100, 250, 500
Ngrams:	Unigram

Output:

Each document (review) is assigned a set of probabilities of whether it belongs to topic 1, 2, 3,…, k

Topic Modeling

Latent Dirichlet Allocation (LDA)

How to evaluate?

- Coherence (‘u_mass’)
- Interpret
- Visualise (pyLDAViz)

Topic Modeling

Topic 1: Customer Svc/Check-in/Reception

- Room, check, time, staff, service, guest, hour, luggage, stay, reception

Topic 2: Room

- Room, bed, bathroom, floor, shower, stay, water, service, tower, wing

Topic 3: F&B/Breakfast

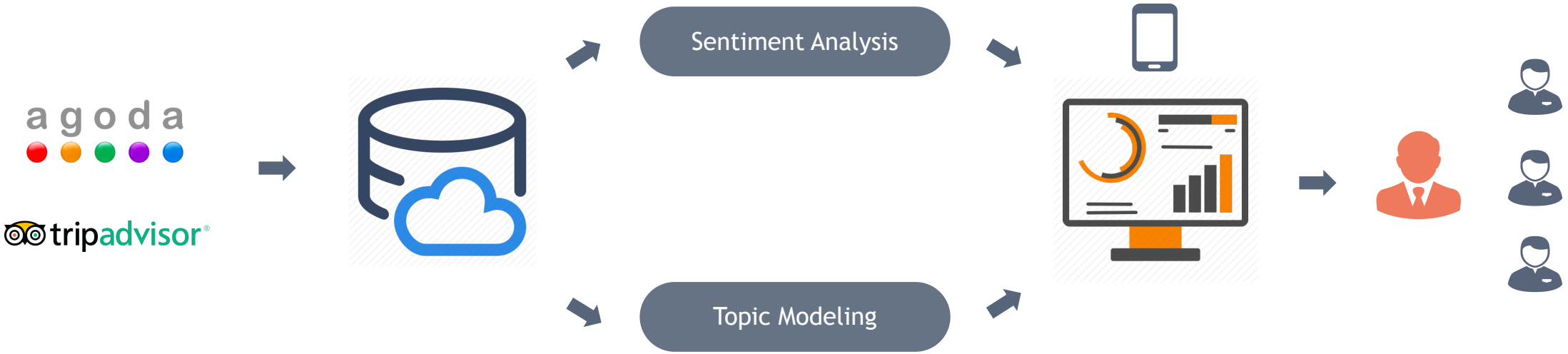
- Breakfast, staff, service, room, food, stay, club, time, buffet, restaurant

Topic 4: Shopping/Restaurants/Location

- Shopping, room, mall, place, staff, food, restaurant, area, heart, service

Coherence = -2.456

Summary of Project Pipeline



Dashboard

Limitations

VADER Sentiment Analysis

- Does the job but not exactly state-of-the-art
- For this dataset, it tends to over-rate sentiments at the positive end
- Lexicon-based (order doesn't matter)– will struggle on figure-of-speech

Topic Modeling (LDA)

- May incorrectly assign topics for short documents (probabilistic)

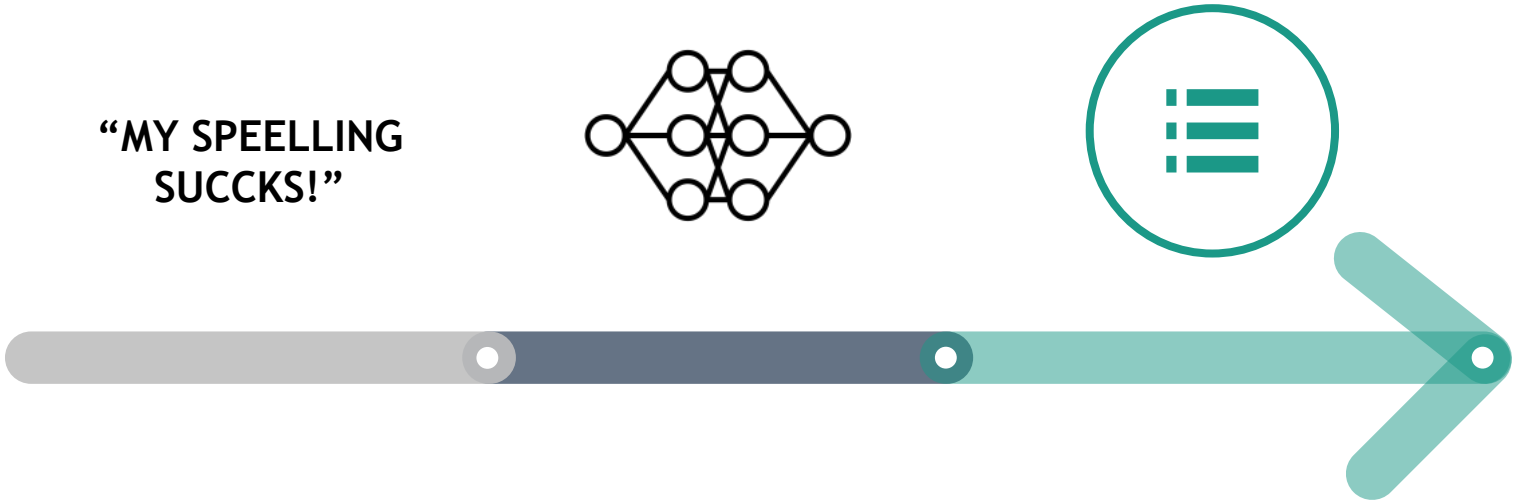
Other Business Applications

E-commerce Product Reviews

On-site Customer Service Feedback

Financial Market Prediction

Next Steps



DATA PRE-PROCESSING

TextBlob:
Spelling Correction
Translate non-English words

TRY ADVANCED MODELS

Sentiment Analysis:
Stanford NLP (pre-trained)

Topic Modelling:
Deep-Learning Word-Embedding Models

‘Mother of All Models’

Build one that can generalise for entire hotel industry

Thank You!