



Data Science Immersive: Capstone Project

European Hotel Reviews

Generating Insights from Hotel guests for Business Owners

Booking.com

Executive Summary



Problem:

- Businesses do not have tools to quickly gather insights from reviews.
- High costs involved to monitor and act upon responses.



Solution:

- Develop tools to quickly identify Sentiment from reviews.
- Fast and efficient method to identify what's good and bad.



Impact:

- Deliver efficient business responses towards consumer needs.
- Drives focus in tackling business opportunities.



Business Problem

WHY HOTEL REVIEWS

Feedback: information enhances business decisions

Reviews - reflect the business image

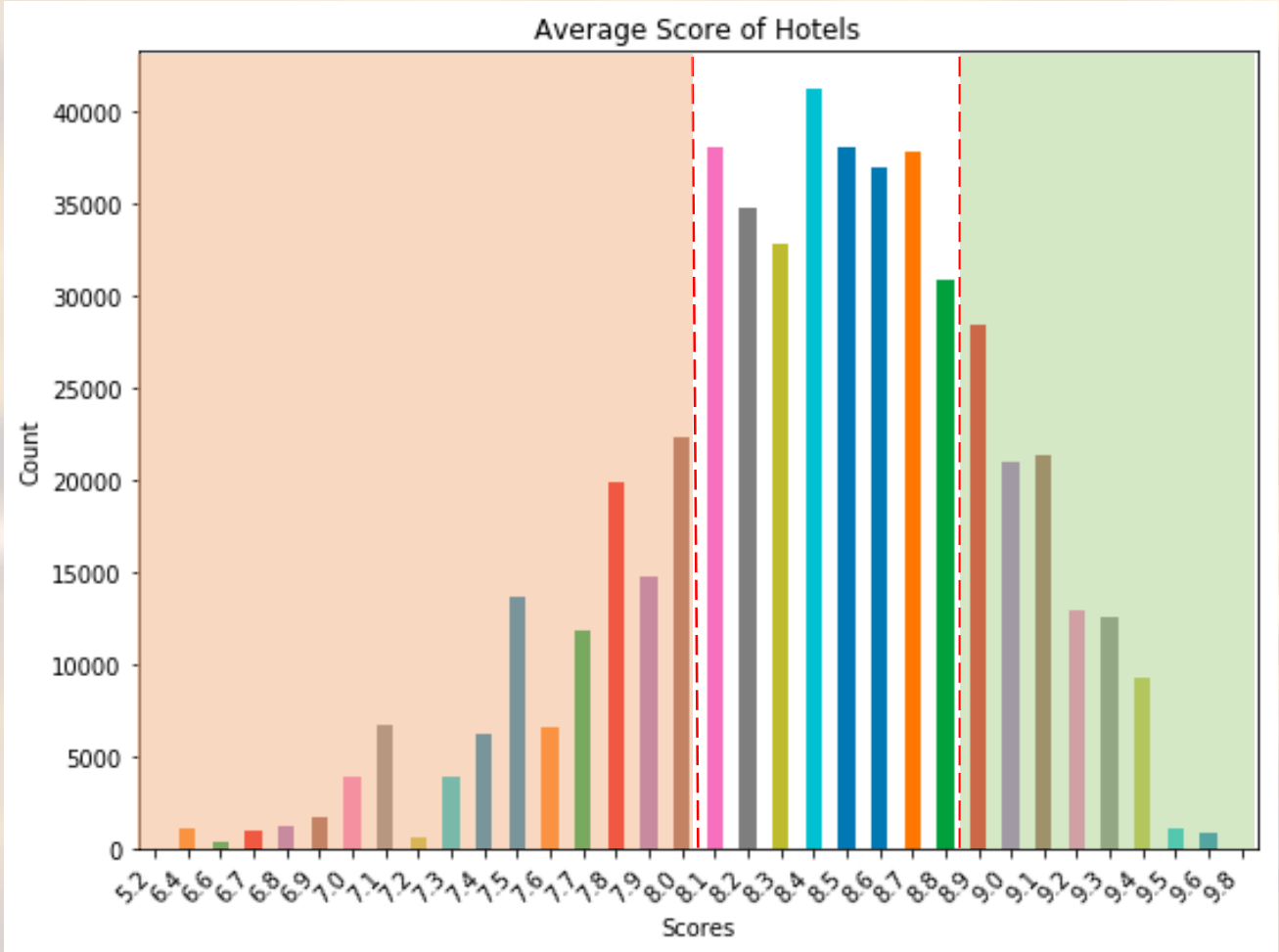
OBJECTIVE

To develop a tool to extract useful information from reviews

For Business Owners

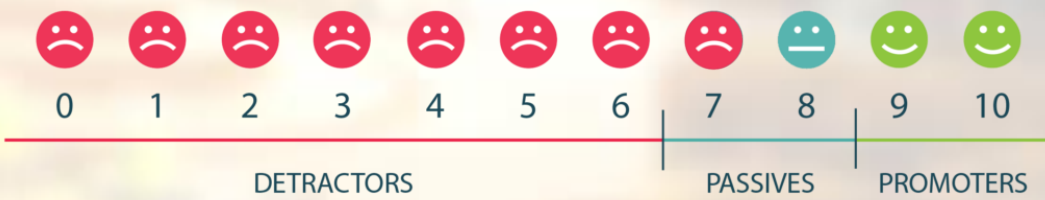


Understanding the data



IQR: 8.1 to 8.8

- Positive: > 8.8 (20.9%)
- Negative: < 8.1 (22.6%)



Outcomes

SENTIMENT ANALYSIS

Identify if reviews are positive or negative

EFFECTIVE VISUALIZATION

Understanding the trends, impact factors



Sentiment Analysis

Lexicon Algorithm

map words to sentiment by building a 'dictionary of sentiment' for identifying polarity

Example: VADER

Machine Learning

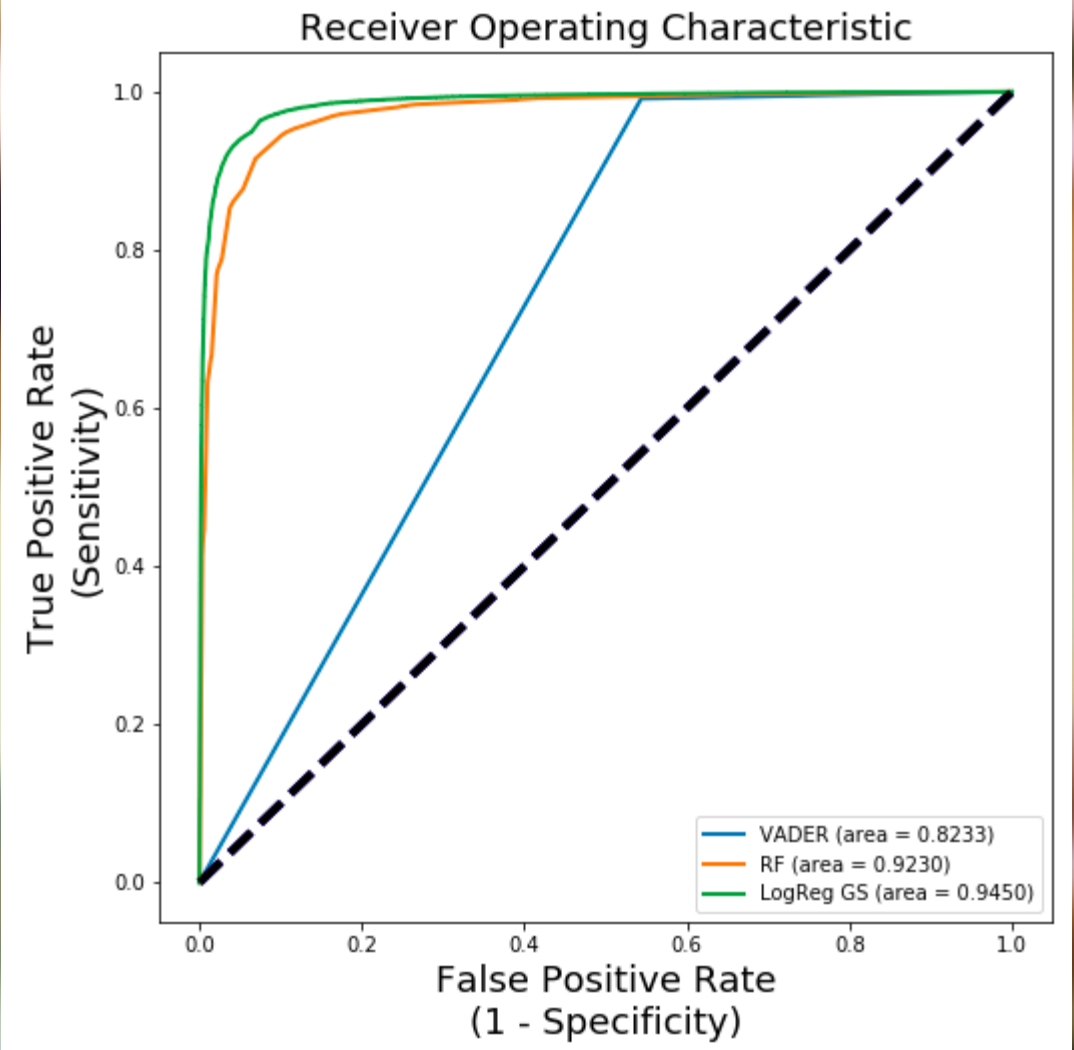
previously labeled data to determine the sentiment of never-before-seen sentences

Example: Logistic Regression



Classification

Tests	ROC AUC Score	Precision	Recall
Baseline	0.5000	-	-
VADER	0.8233	0.667	0.991
Random Forest	0.9230	0.915	0.924
Multinomial	0.9404	0.964	0.912
Logistic Regression	0.9429	0.929	0.952
Logistic Regression (HT)	0.9450	0.933	0.952



Results Analysis



Top 5 Negative Words: [('nothing', 33), ('room', 26), ('small', 15), ('bit', 14), ('pool', 8)]

Top 5 Positive Words: [('staff', 150), ('location', 85), ('great', 80), ('excellent', 54), ('friendly', 49)]

Generate insights by:

Specific Hotel

State

Country of
Origin (Guests)

Next Steps

1

Implementation to site features
(Chat bot for interaction, software tool)

2

More data (users, tags) can be collected to further
develop into solutions: Recommenders





END