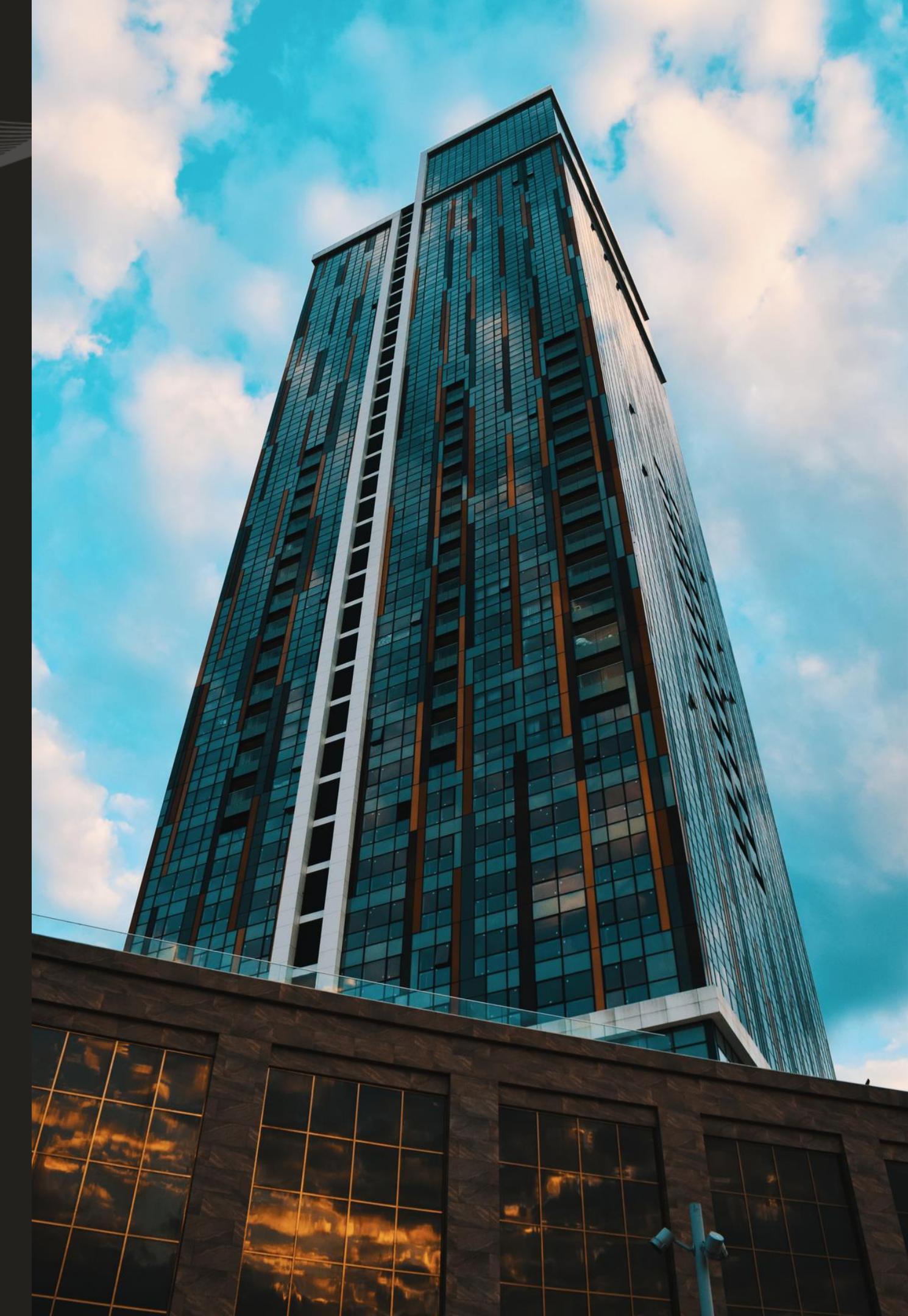


LUXURY HOTELS

Final Project – Sentiment Analysis

Alexander Meneses - C0900272
Christian Jimenez - C0891438
Magda Diaz - C0900484
Maira Ospino - C0895237
Johana Saenz - C0903415
Paing Hein - C0907275



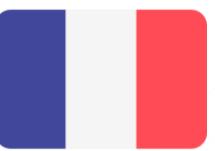
Overview

- Introduction 01
- Objective 02
- Business Problem 03
- Code 04
- Data Preprocessing 05
- Sentiment Analysis 06
- Exploratory Data Analysis 07
- Financial Viability 08



Project Introduction

Customers' reviews of Luxury Hotels
in Expedia



France



Mexico

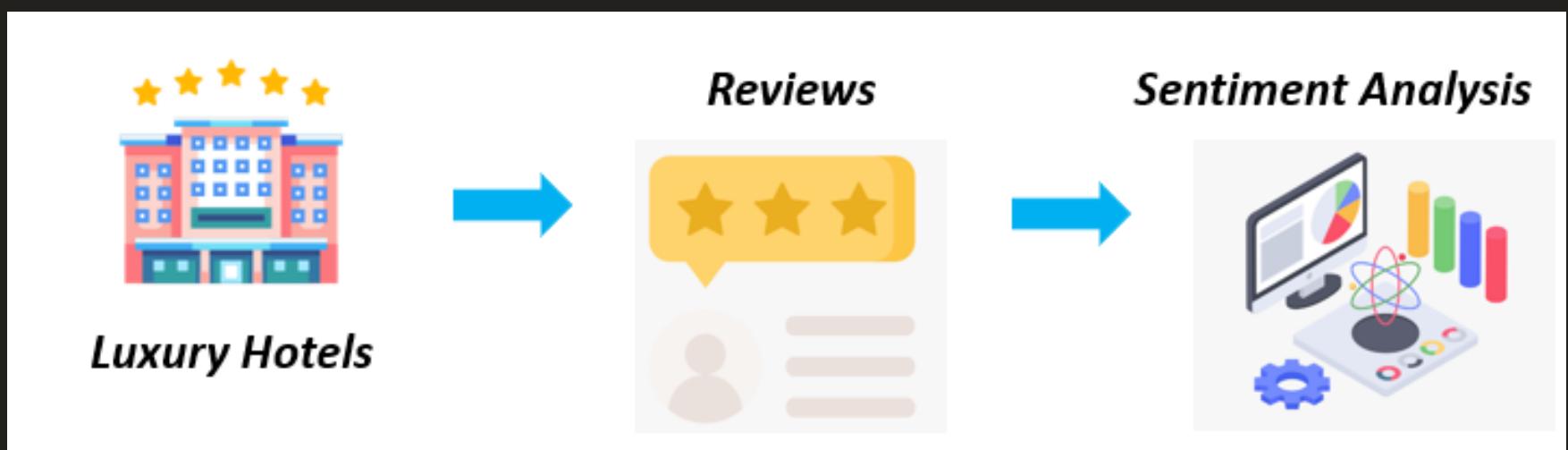


Switzerland



Objective

- Sentiment analysis to see patterns and trends
- Management decision for improvement



Business Problem

"Luxury hotels use to face an absence of a strategy to identify and act on negative consumer feedback."

Project Focus

- Collecting reviews made by online users to identify improvement opportunities.
- Text Analysis: Use NLP techniques with tools like Python's NLTK, spaCy, or specialized platforms.
- Sentiment Scoring: Score reviews (positive, neutral, negative) using ML or rule-based methods.
- Identify Key Themes: Extract common themes like cleanliness, service, location, or amenities.
- Creating action plans to address the negative feedback by improving the found areas.



Code

- Library Installation & Import
- Download NLTK Resources
- Data Loading and Preview
- Data Cleaning
- Data Preprocessing
- Sentiment Analysis
- Exploratory Data Analysis
- Data Visualization
 - Sentiment Distribution Visualization
 - Sentiment by Country Visualization
 - Word Cloud Generation for Each Sentiment

Code

Install and Import Libraries

```
# Step 2: Import necessary libraries

import pandas as pd # For data manipulation and analysis
import matplotlib.pyplot as plt # For creating static, animated, and interactive visualizations
import seaborn as sns # For statistical data visualization
from nltk.corpus import stopwords # For common words removal in text processing
from nltk.tokenize import word_tokenize # For splitting text into individual words
from textblob import TextBlob # For text processing tasks like sentiment analysis and translation
from googletrans import Translator # For text translation using Google Translate API
import nltk # Main NLTK module for downloading additional text processing resources
import re # For regular expressions in text searching and manipulation
```

Step 3: Download necessary NLTK resources

```
nltk.download('punkt') # Tokenizer models
nltk.download('stopwords') # Stopword lists
```

Data Cleaning

```
# Step 6: Identify and drop or fill null values

print("Null values in each column:")
print(df.isnull().sum())

# Drop rows where 'Review' or 'Rating' is null
df_cleaned = df.dropna(subset=['Review', 'Rating'])

# Ensure 'Date' is in datetime format
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

# Fill missing 'Date' with a placeholder date
df_cleaned['Date'].fillna('2023-01-01', inplace=True)

# Fill missing 'City' In 'Blank'
df_cleaned['City'].fillna(' ', inplace=True)

# Fill missing 'City' In 'Blank'
df_cleaned['Hotel_name'].fillna(' ', inplace=True)
```

[7] # Step 7: Text Cleaning Function

```
def clean_text(text):

    # Remove HTML tags
    text = re.sub(r'<.*?>', '', text)

    # Remove special characters and digits
    text = re.sub(r'[^a-zA-Z\s]', '', text)

    # Convert to lowercase
    text = text.lower()

    # Remove extra whitespace
    text = re.sub(r'\s+', ' ', text).strip()

    return text

# Apply text cleaning
df['cleaned_review'] = df['Review'].apply(clean_text)

df = df.drop_duplicates(subset=['Review', 'Hotel_name', 'Date'])

df.head()
```

Code

Translation

```
✓ [9] # Step 9: Initialize the translator
      translator = Translator()

✓ [10] # Step 10: Function to translate text to English
       def translate_to_english(text):
           try:
               translated = translator.translate(text, dest='en')
               return translated.text
           except Exception as e:
               return text # Return the original text if translation fails

✓ [11] # Step 11: Translate reviews to English
       review_column_name = 'Review' # Change this to the actual column name if different
       df['review_english'] = df[review_column_name].apply(translate_to_english)
```

Data Preprocessing

Raw Text → "Magnificent, very well placed ... very high -end service"

Tokenization Magnificent , very well placed ... very high -end service

Normalization magnificent , very well placed ... very high -end service

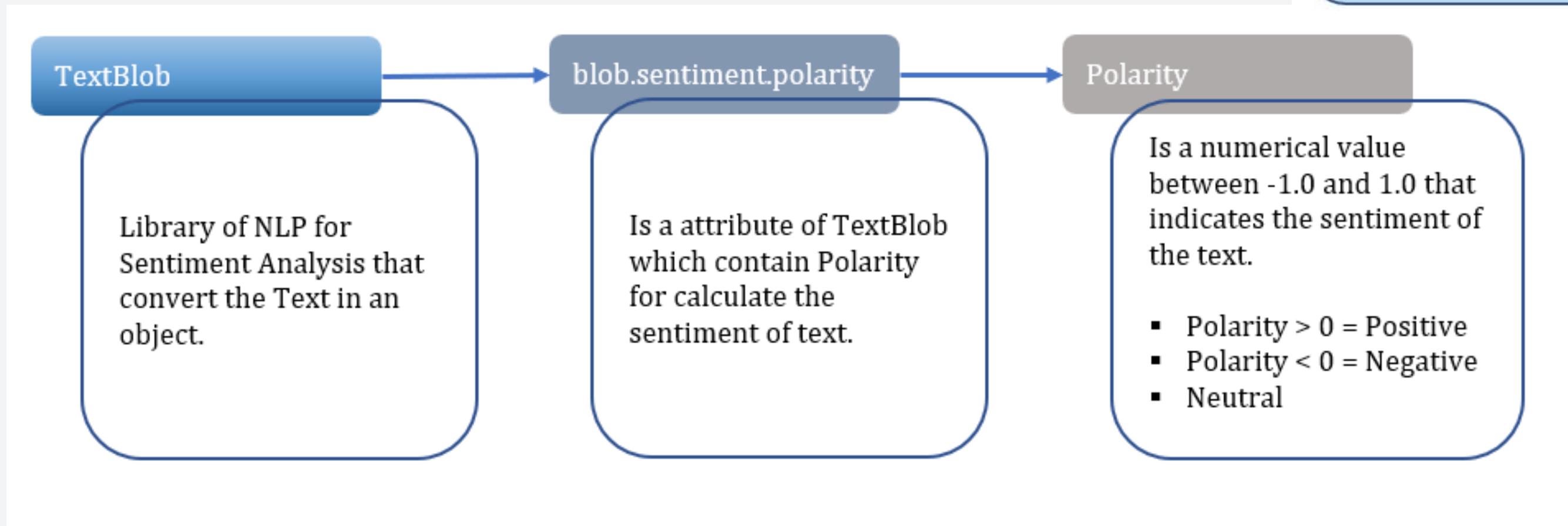
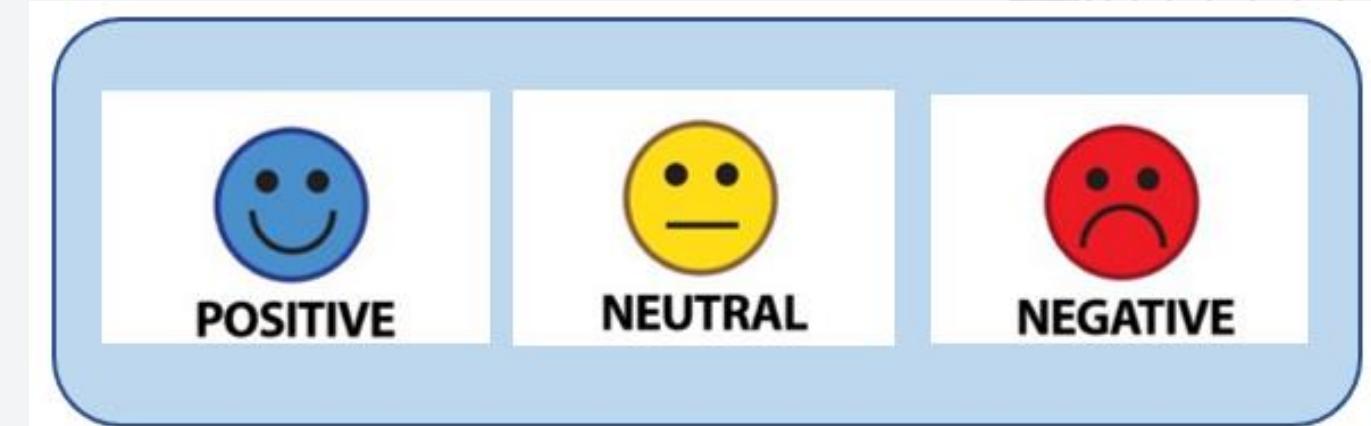
Remove Non-alphanumeric magnificent very well placed very high end service

Remove Stopwords magnificent well placed high service

Processed Text magnificent well placed high service

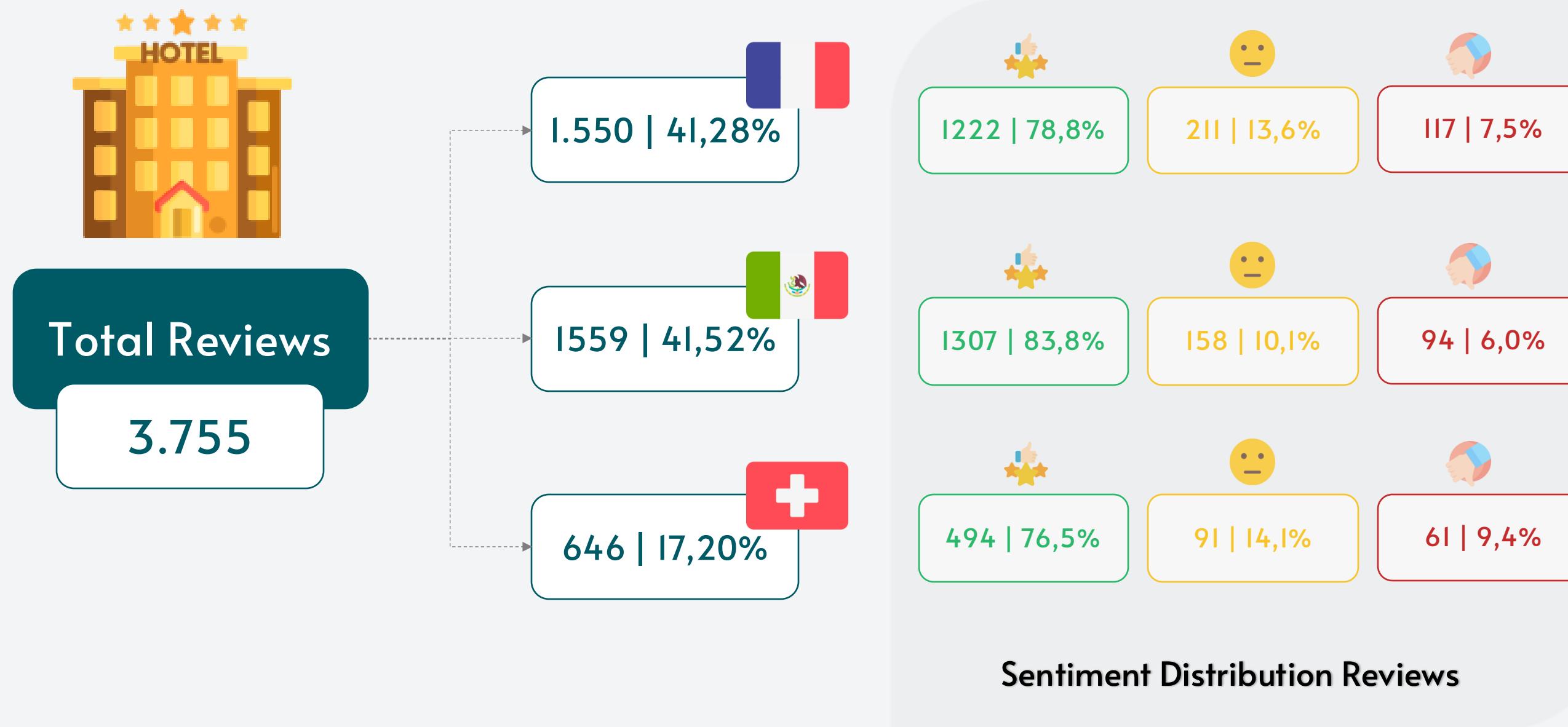
```
# Step 10: Text preprocessing function
def preprocess_text(text):
    # Tokenize the text
    tokens = word_tokenize(text)
    # Convert tokens to lowercase
    tokens = [word.lower() for word in tokens]
    # Remove non-alphanumeric tokens and stopwords
    tokens = [word for word in tokens if word.isalnum()]
    tokens = [word for word in tokens if word not in stopwords.words('english')]
    # Join tokens back into a single string
    return ' '.join(tokens)
```

Sentiment Analysis



→ Raw text: Magnificent, very well placed very high -end service
Sentiment Polarity: 0.4693333333333333

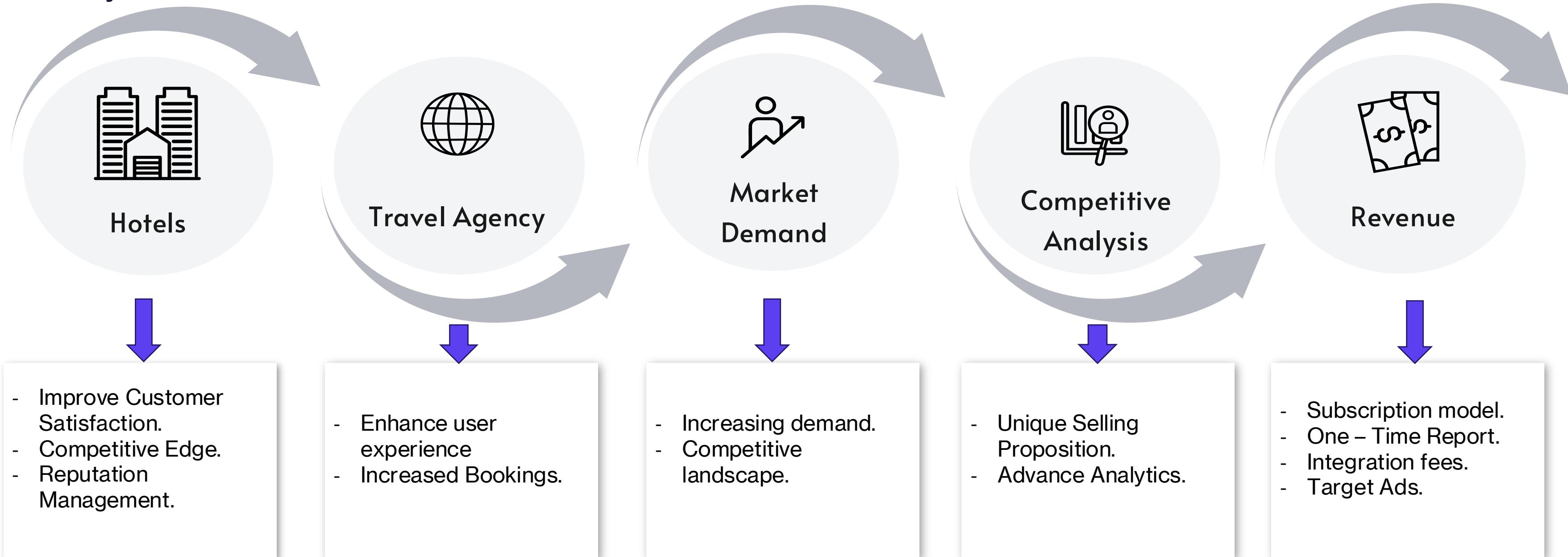
Exploratory Analysis



Financial

The financial viability of this project is rooted in its strong value proposition, multiple revenue streams, and the growing demand for data-driven decision-making in the luxury hotel industry.

This positions the project as a profitable and sustainable in the competitive landscape of luxury Hotel.



Cost Breakdown

Human Resource	Cost Per Month	Total Months	Total Cost
Business Analyst	\$ 5,969.00	2	\$ 11,938.00
Data Analyst	\$ 4,418.00	2	\$ 8,836.00
Data Analyst	\$ 4,418.00	2	\$ 8,836.00
Data Scientist	\$ 6,178.00	2	\$ 12,356.00
QA Specialist	\$ 4,418.00	2	\$ 8,836.00
Business intelligence	\$ 6,178.00	2	\$ 12,356.00
Total Human Resources	\$ 31,579.00		\$63,158.00
IT Resources	Cost Per Month	Total Months	Total Cost
Software Tools (Cloud - Google)	\$ 500.00	12	\$ 6,000.00
Data storage	\$ 100.00	12	\$ 1,200.00
Monitoring and Maintenance	\$ 700.00	12	\$ 8,400.00
Total IT Resources	\$ 1,300.00		\$15,600.00
Miscellaneous Cost	Cost Per Month	Total Months	Total Cost
Legal And Administrative	\$ 300.00	12	\$ 3,600.00
Total Project	\$ 33,179.00		\$82,358.00

Future Work

The sentiment analysis project for Expedia luxury hotel reviews provides valuable insights into customer feedback. However, there are areas for future work and potential enhancements to improve the project's utility.

REAL-TIME DATA PROCESSING

Develop capabilities for real-time sentiment analysis

PERSONALIZATION

Create user profiles based on review history and sentiment analysis

INTEGRATION WITH OTHER DATA SOURCES

Incorporate additional data sources, for a more holistic view.



Thank You

Gerente General