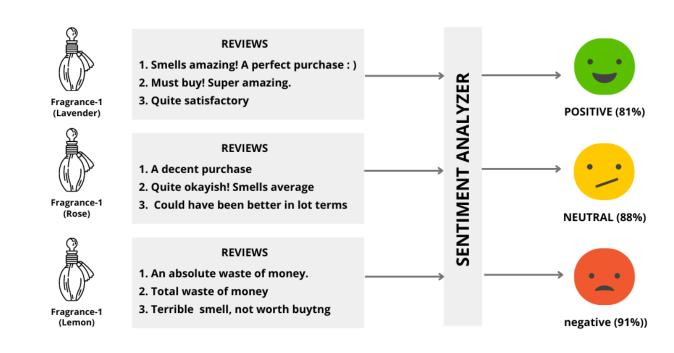# Marketing Analytics Business Case

Harsh Wardhan Singh

Day-3
Sentiment Analysis

Sentiment analysis is a natural language processing (NLP) technique used to identify and quantify emotions in text. It helps businesses understand customer opinions, categorize feedback, and make data-driven decisions based on public sentiment and tone.

Fragrance-1
(Lavender)

REVIEWS
1. Smells amazing! A perfect purchase :)
2. Must buy! Super amazing.
3. Quite satisfactory

Fragrance-1
(Rose)

REVIEWS
1. A decent purchase
2. Quite okayish! Smells average
3. Could have been better in lot terms

Fragrance-1
(Lemon)

REVIEWS
1. An absolute waste of money.
2. Total waste of money
3. Terrible smell, not worth buytng

SENTIMENT ANALYZER

POSITIVE (81%)

NEUTRAL (88%)

negative (91%))

Steps Followed in Sentiment Analysis.

1. Imported required libraries(e.g., pandas, pyodbc, nltk, matplotlib)

2. Connected to SQL Server databaseUsed pyodbc to connect and fetch customer reviews from a SQL table.

3. Loaded data into a DataFrameUsed pandas.read_sql() to retrieve and explore the dataset.

4. Downloaded and initialized VADER sentiment analyzernltk.download('vader_lexicon') and initialized SentimentIntensityAnalyzer().

```python
import pandas as pd
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import nltk

# Download VADER lexicon
nltk.download('vader_lexicon')

# Initialize VADER
vad = SentimentIntensityAnalyzer()

# Define sentiment scoring function
def sentiment_score(review):
    sentiment = vad.polarity_scores(review)
    return sentiment['compound']

# Define sentiment categorization using rating and score
def categorize_sentiment(score, rating):
    if score > 0.05:
        if rating >= 4:
            return 'Positive'
        elif rating == 3:
            return 'Mixed Positive'
        else:
            return 'Mixed Negative'
    elif score < -0.05:
        if rating <= 2:
            return 'Negative'
        elif rating == 3:
            return 'Mixed Negative'
        else:
            return 'Mixed Positive'
    else:
        if rating >= 4:
            return 'Positive'
        elif rating <= 2:
            return 'Negative'
        else:
            return 'Neutral'
```
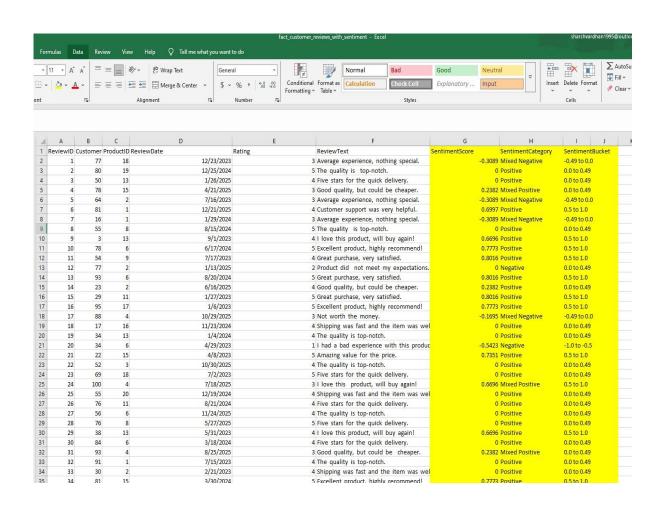
Steps Followed in Sentiment Analysis.

5. Calculated sentiment scoresApplied sia.polarity_scores() to extract compound sentiment scores for each review.

6. Classified sentiment categoriesUsed a custom function combining sentiment score and star rating to assign sentiment labels like Positive, Negative, Mixed, etc.

7. Bucketed sentiment scoresGrouped scores into ranges (e.g., 0.5 to 1.0) for visualization and aggregation.

8. Exported final datasetSaved the enriched DataFrame to a CSV for further use in visualization tools like Power BI.

```python
# Define sentiment bucket
def sentiment_bucket(score):
    if score >= 0.5:
        return '0.5 to 1.0'
    elif 0.0 <= score < 0.5:
        return '0.0 to 0.49'
    elif -0.5 <= score < 0.0:
        return '-0.49 to 0.0'
    else:
        return '-1.0 to -0.5'

# Apply functions to your DataFrame (assuming `df` already contains your data)
df['SentimentScore'] = df['ReviewText'].apply(sentiment_score)

df['SentimentCategory'] = df.apply(
    lambda row: categorize_sentiment(row['SentimentScore'], row['Rating']), axis=1)

df['SentimentBucket'] = df['SentimentScore'].apply(sentiment_bucket)

# Show results
print(df.head())

# Save to CSV
df.to_csv('fact_customer_reviews_with_sentiment.csv', index=False)

# Test example
print(vad.polarity_scores("I'm so happy"))
print(vad.polarity_scores("It is the worst product i have"))
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\Lenovo\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
   ReviewID  CustomerID  ProductID  ReviewDate  Rating  \
0         1          77         18  2023-12-23       3
1         2          80         19  2024-12-25       5
2         3          50         13  2025-01-26       4
3         4          78         15  2025-04-21       3
4         5          64          2  2023-07-16       3

                              ReviewText  SentimentScore SentimentCategory  \
0       Average experience, nothing special.          -0.3089    Mixed Negative
1              The quality is   top-notch.           0.0000          Positive
2   Five stars for the quick delivery.           0.0000          Positive
3  Good quality, but could be cheaper.           0.2382    Mixed Positive
```

Result

Thank You

- [Sentiment Analysis Excel](Sentiment%20Analysis%20Excel)