



# AIRPORT TRAVELER REVIEW ANALYSIS

Using Sentiment, Bigram, and Trigram analysis  
with Python.

Fierdy Pandu

2023 May 04





# Objective

# Objective

**Scraping data from Skytrax**

**Get insights about which services or sections that need improvement concerned by travelers**



# Objective

**Scraping data from Skytrax**

**Get insights about which services or sections that need improvement concerned by travelers**





# Preparation

9/10

"Everything went smooth"

Roman Pawlyk (Poland) 2nd April 2023

Not Verified

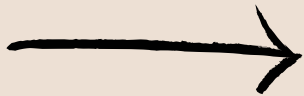
 | I was pleasantly surprised by level of service and experience on this flight. Everything went smooth, food was very tasty, great choices of dishes and beverages. Overall, staff was very helpful always trying to accommodate the customer needs

Aircraft	Boeing 777
Type Of Traveller	Couple Leisure
Seat Type	Economy Class
Route	Warsaw to Phuket via Dubai
Date Flown	November 2022
Seat Comfort	★★★★★
Cabin Staff Service	★★★★★
Food & Beverages	★★★★★
Inflight Entertainment	★★★★★
Ground Service	★★★★★
Wifi & Connectivity	★★★★★
Value For Money	★★★★☆
Recommended	✓



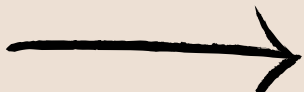
# Data to use

Determine what and the type of data that I will use for the analysis



# Scraping data

Scrap the data with Beautiful Soup python package.



# Prepare data

Storing the data and make it ready to use for data cleaning, manipulation, and analysis with pandas.



**CLICK HERE FOR**  
**ALL THE CODE**

---

reviews	
0	✓ Trip Verified   Jakarta's Soekarno-Hatta has...
1	Not Verified   I arrived in the airport at ni...
2	✓ Trip Verified   You can't transfer between ...
3	✓ Trip Verified   International terminal 3 is...
4	✓ Trip Verified   Most of the airport was hot...
...	...
166	Soekarno-Hatta has a good architecture inside ...
167	Very negative experience. Arrived around midni...
168	CGK can't be compared to Changi or KLIA howeve...
169	Transferred from a Garuda domestic to a Garuda...
170	As a resident in Indonesia have frequently use...
171 rows × 1 columns	

Got 171 data from  
the scraping

Here is the output of the data from scraping with BeautifulSoup. Upon exploring the data, I found a text pattern that can hinder the data to be manipulated and analyzed further.

## Major text patterns problems to solve

There are four major text patterns that will interfere with the analysis. This can lead to ambiguity when doing word counting. The text patterns are as follows:

✓ Verified Review
✓ Trip Verified |
✗ Unverified
Not Verified

Using custom functions in Python can remove these anomalies in our raw data.





## Creating a function to clean the data

```
# Removing unimportant part of the review text with a function
def process_reviews(reviews):
    if reviews.startswith('✅ Trip Verified'):
        return reviews.split('|', 1)[1]
    elif reviews.startswith('Not Verified'):
        return reviews.split('|', 1)[1]
    elif reviews.startswith('✅ Verified Review'):
        return reviews.split('|', 1)[1]
    elif reviews.startswith('❌ Unverified'):
        return reviews.split('|', 1)[1]
    elif reviews.startswith('❌ Not Verified'):
        return reviews.split('|', 1)[1]
    else:
        return reviews
```

The goal here is to make the data ready to proceed to our next step of our project (Manipulating, Exploring, Analysis)

## Result

	reviews
0	Jakarta's Soekarno-Hatta has been for a coupl...
1	I arrived in the airport at night. The airpo...
2	You can't transfer between domestic and inte...
3	International terminal 3 is a boiling pot. ...
4	Most of the airport was hot. Only felt moder...
...	...
59	Unbelievable stupidity. The whole process is ...
60	This is the old terminal (1 & 2), but after r...
61	I had a very good experience traveling from ...
62	I have been using Terminal 3 since its day of...
63	Jakarta is a total disaster of an internatio...



	reviews	clean_reviews	POS tagged	Lemma	Sentiment	Analysis
0	Jakarta's Soekarno-Hatta has been for a coupl...	Jakarta s Soekarno Hatta has been for a coupl...	[(Jakarta, n), (Soekarno, n), (Hatta, n), (cou...	Jakarta Soekarno Hatta couple year home base...	0.4019	Neutral
1	I arrived in the airport at night. The airpo...	I arrived in the airport at night The airport...	[(arrived, v), (airport, n), (night, n), (airp...	arrive airport night airport surprisingly em...	0.5267	Positive
2	You can't transfer between domestic and inte...	You can t transfer between domestic and inter...	[(transfer, n), (domestic, a), (international,...	transfer domestic international section with...	-0.8309	Negative
3	International terminal 3 is a boiling pot. ...	International terminal is a boiling pot Air c...	[(International, n), (terminal, n), (boiling, ...	International terminal boil pot Air conditio...	-0.6808	Negative
4	Most of the airport was hot. Only felt moder...	Most of the airport was hot Only felt moderat...	[(airport, n), (hot, a), (felt, v), (moderatel...	airport hot felt moderately cool quiet area ...	0.6769	Positive

Regex

Using the Regex package in Python to clean the text data from any special characters like: "" !@#\$%^&\*()

NLTK

The nltk.corpus package defines a collection of corpus reader classes, which can be used to access the contents of a the diverse set of corpora.

Using Natural Language Toolkit (NLTK), It will be removed unimportant words like "was", "were", "I", "you" etc.

Lemmatization

Lemmatization is the process of transforming a word to its base form (lemmatization) by removing word inflections such as suffixes and prefixes, resulting in a base word called a lemma.  
The purpose of lemmatization is to obtain a base word that has the same meaning as other forms of the word.

Vader Sentiment

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a rule-based sentiment analysis tool that is specifically designed for analyzing the the sentiment of the text.  
  
I used this package to better define whether the reviews are negative or positive with a score.



By using the data from the "Lemma" columns, I've created another three separate data frames to get better insights for my analysis

## Word Counting

	Word	Frequency
0	airport	301
1	terminal	258
2	immigration	110
3	time	90
4	check	87
...	...	...
1927	yell	1
1928	yellow	1
1929	yesterday	1
1930	youth	1
1931	zero	1

Counting the total occurrences of words in the reviews.

## Bigram Counting

	Bigram	Frequency
0	new terminal	24
1	jakarta airport	24
2	baggage claim	19
3	security check	18
4	soekarno hatta	16
...	...	...
8071	friendly enviromnment	1
8072	enviromnment reasonably	1
8073	reasonably price	1
8074	price good	1
8075	quality food	1
8076 rows × 2 columns		

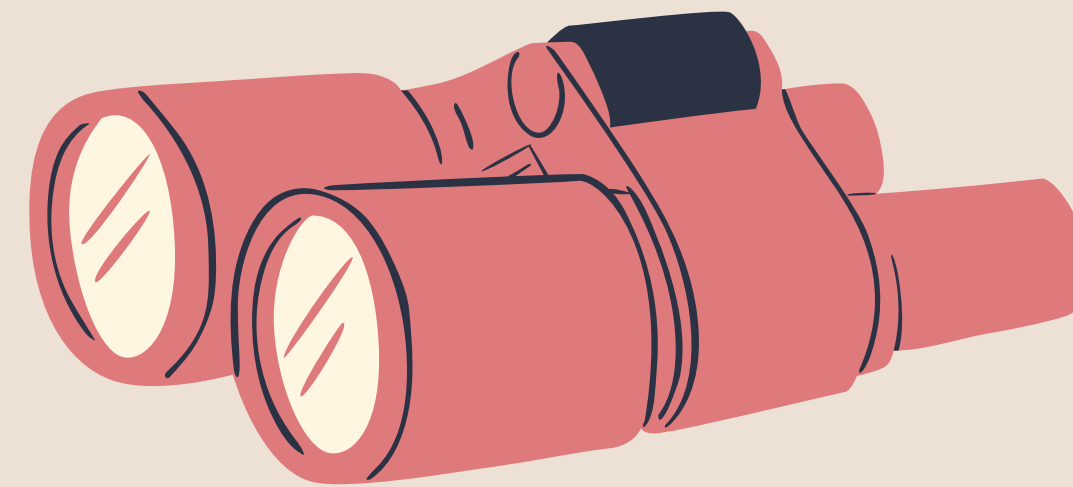
Bigram is an n-gram in natural language processing and text analysis that consists of two words that appear consecutively in a text. In a bigram, the text is divided into a sequence of consecutive word pairs without overlapping.

## Trigram Counting

	Bigram	Frequency
0	new terminal	24
1	jakarta airport	24
2	baggage claim	19
3	security check	18
4	soekarno hatta	16
...	...	...
8071	friendly enviromnment	1
8072	enviromnment reasonably	1
8073	reasonably price	1
8074	price good	1
8075	quality food	1
8076 rows × 2 columns		

Trigram is an n-gram in natural language processing and text analysis that consists of three words that appear consecutively in a text. In a trigram, the text is divided into a sequence of consecutive three-word units without overlapping.

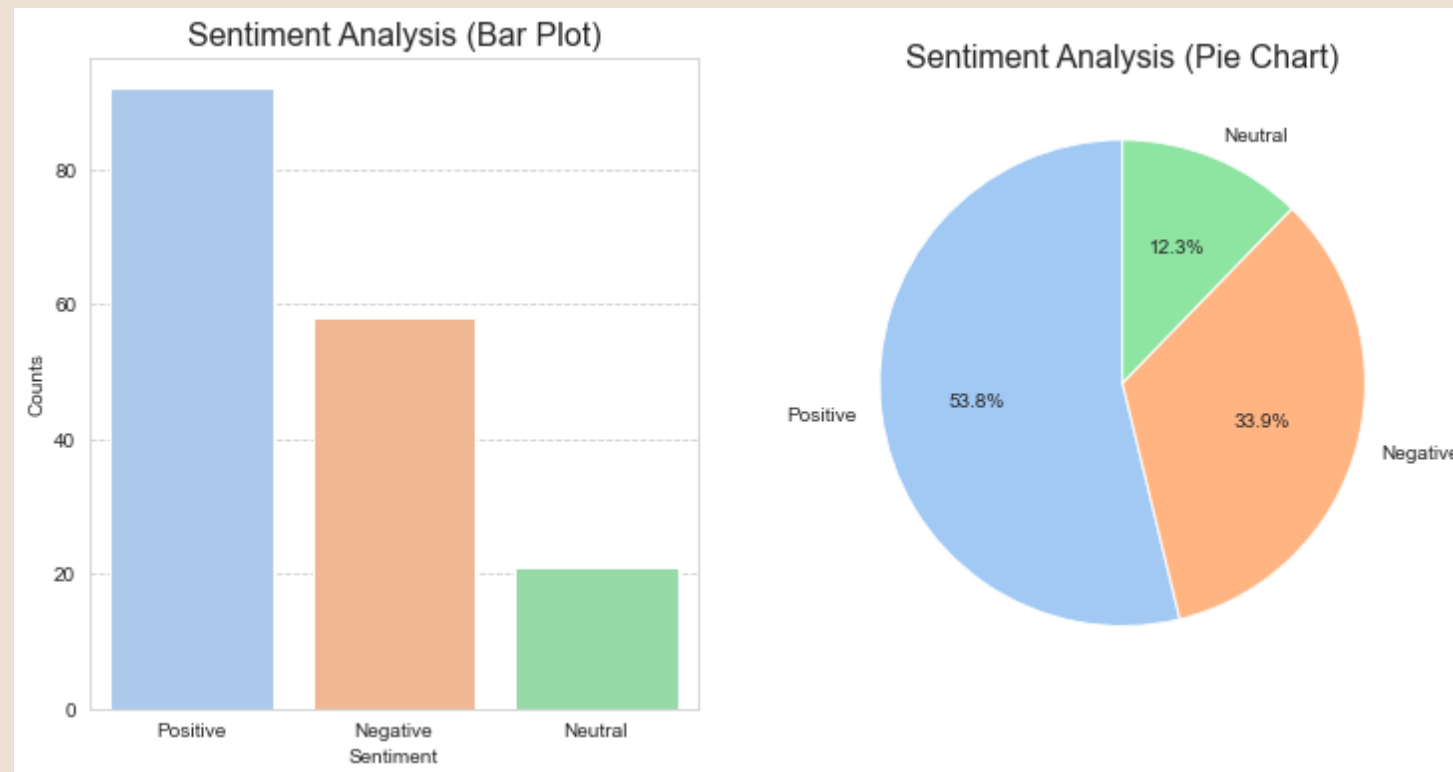




# Exploration



# Sentiment Analysis



The given data represents the results of sentiment analysis of text reviews on Soekarno-Hatta International Airport services. The text reviews were classified into three sentiment categories, positive, negative, and neutral.

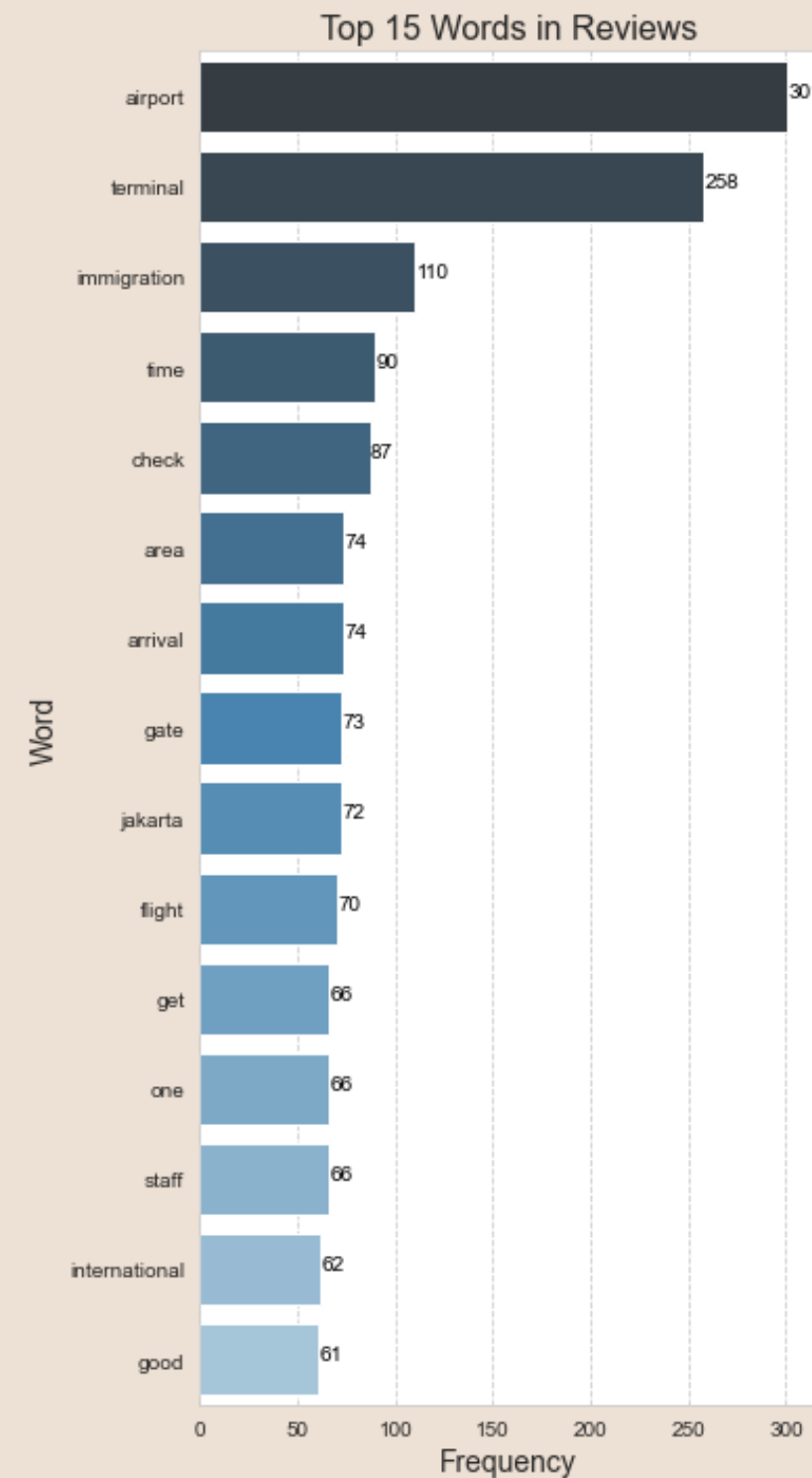
The data shows that out of all the reviews, 92 were categorized as positive, 58 were categorized as negative, and 21 were categorized as neutral. This means that the majority of reviews about Soekarno-Hatta International Airport services are positive.

This suggests that many passengers appreciate the facilities, services, and overall environment provided by the airport. However, the negative reviews, accounting for 33.9% of the total, indicate that there is still a considerable number of passengers who have encountered issues, challenges, or inconveniences at the airport.

Overall, while Soekarno-Hatta International Airport seems to receive predominantly positive feedback, it is crucial to address the negative and neutral reviews to identify areas for improvement and enhance the airport experience for all passengers. Airport management can use this information to target specific issues, make necessary changes, and work towards improving customer satisfaction.



# Word Counting Analysis

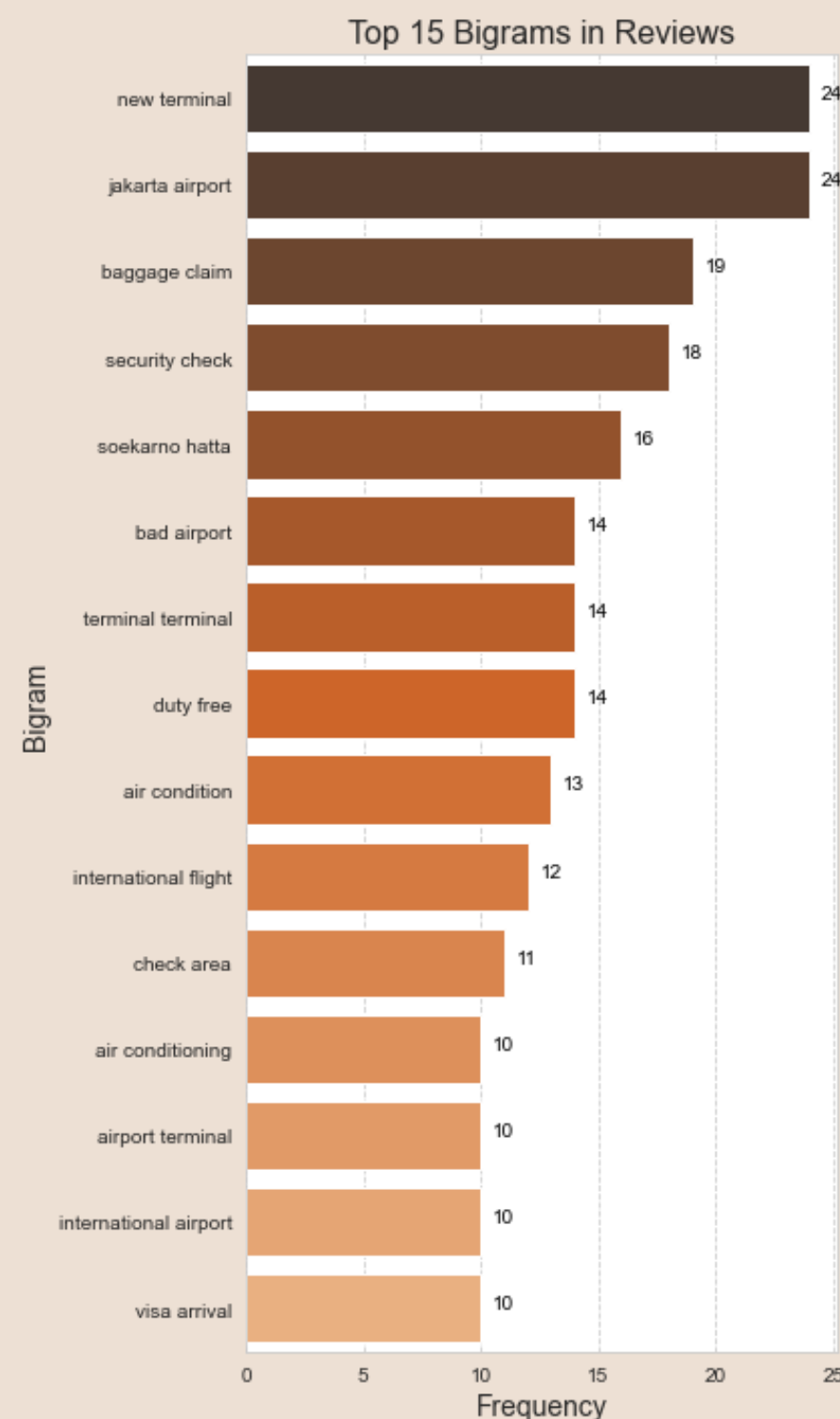


This data frame shows the most frequently occurring words in the analyzed text. From this data, we can see that the word "airport" appears the most, with a count of 301, followed by "terminal", "immigration", "time", and "check". Other common words include "area", "arrival", and "gate".

This indicates that passengers' primary experiences and concerns at Soekarno-Hatta International Airport involve the terminals, immigration process, time management, security checks, specific airport areas, and airport gates. Airport management should focus on these aspects, analyze the sentiment associated with these words, and address the factors causing any negative experiences to improve overall customer satisfaction.



# Bigram Analysis

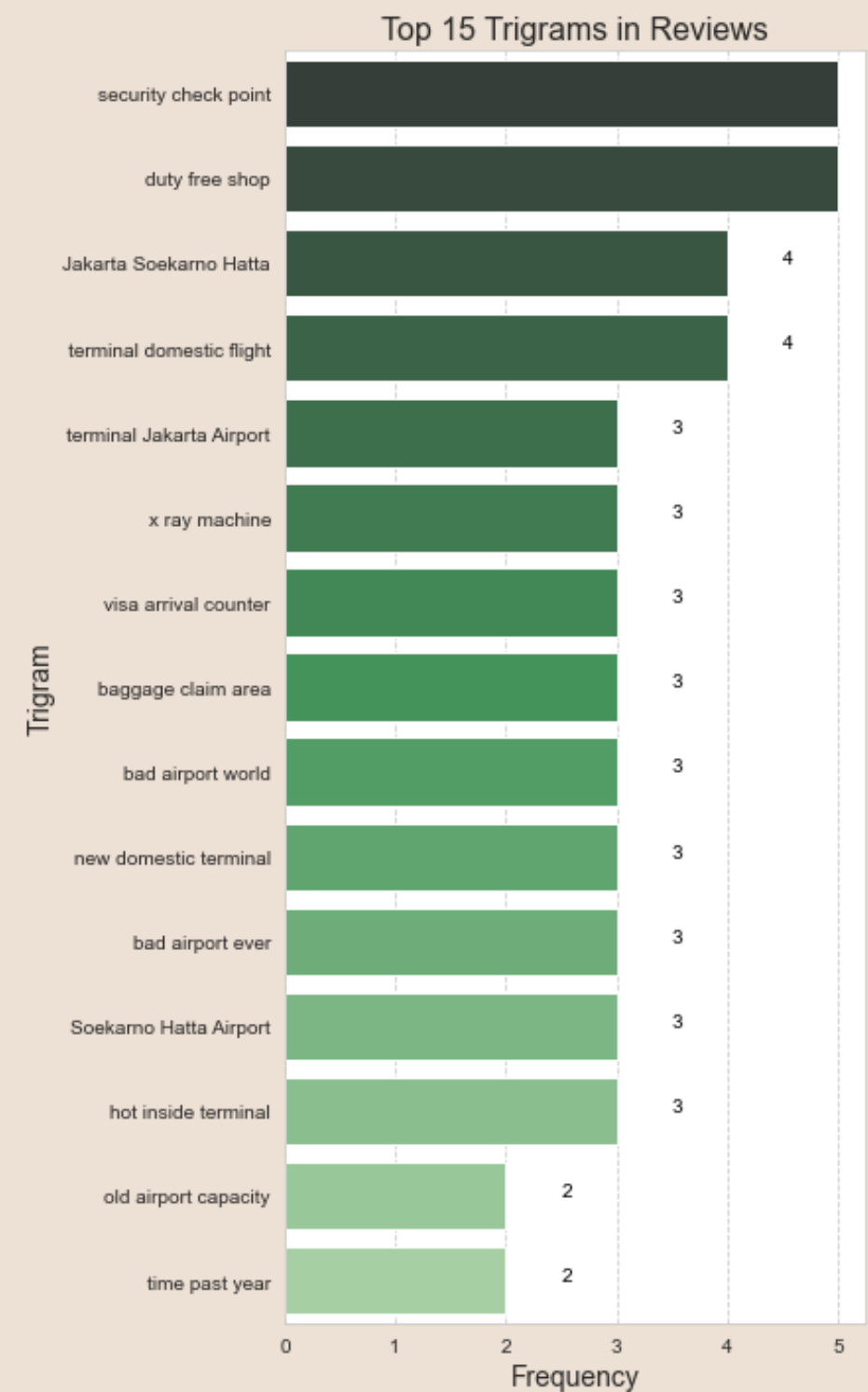


This data frame shows the most frequently occurring pairs of words in the analyzed text. From this data, the most common bigram are "new terminal" and "jakarta airport", followed by "baggage claim", and "security check" and "soekarno hatta". Other common bigrams include "bad airport", "duty free", "terminal terminal" and "air condition".

The top bigrams reveal that passengers' primary experiences and concerns at Soekarno-Hatta International Airport involve the new terminal, baggage claim, security checks, duty-free shopping, and air conditioning. Airport management should focus on these areas, analyze the sentiment associated with these bigrams, and address the factors causing any negative experiences to improve overall customer satisfaction.



# Trigram Analysis



This data frame shows the most frequently occurring pairs of words in the analyzed text.

The top trigrams reveal that passengers' primary experiences and concerns at Soekarno-Hatta International Airport involve security checks, duty-free shopping, terminal experiences, visa-on-arrival services, baggage claims, negative perceptions, and temperature control. Airport management should focus on these areas, analyze the sentiment associated with these trigrams, and address the factors causing any negative experiences to improve overall customer satisfaction.







# Conclusion

# Conclusion

53,8%

53.8% of reviews describe the service as positive for Soekarno–Hatta International Airport. While 33.9% are negative and neutral for the rest.



Address things involving the terminals, immigration process, time management, security checks, specific airport areas, and airport gates. These words are occurring frequently in the reviews for Soekarno–Hatta International Airport and can be investigated further for better airport experiences.



# Thank you!