# Table of Contents

# Introduction

**Problem Statement:**

Analyse passenger feedback data to identify key drivers of dissatisfaction related to food and beverage services during flights, and provide actionable recommendations for improvement.

**Summary of Analysis**:

In this Report, we delve into a comprehensive analysis of passenger feedback data to uncover insights that contribute to enhancing in-flight F&B experiences. Employing a blend of techniques including Exploratory Data Analysis (EDA), Data Visualisation, and Natural Language Processing (NLP), we dissected the data from various angles. Our analysis involved:

- **Performing EDA** to understand overall trends and patterns in the dataset.
- **Joining and Preprocessing** multiple datasets to create a comprehensive data source.
- **Utilizing NLP** to extract keywords, uncover pain points, and recommend improvements.
- **Keyword Extraction** We used a technique called "Keyword Extraction" to identify keywords and phrases in passengers' comments.

Our findings provide a valuable resource for identifying pain points and offering practical solutions, thus boosting passenger satisfaction and loyalty related to food and beverage services during flights, and providing actionable recommendations for improvement.

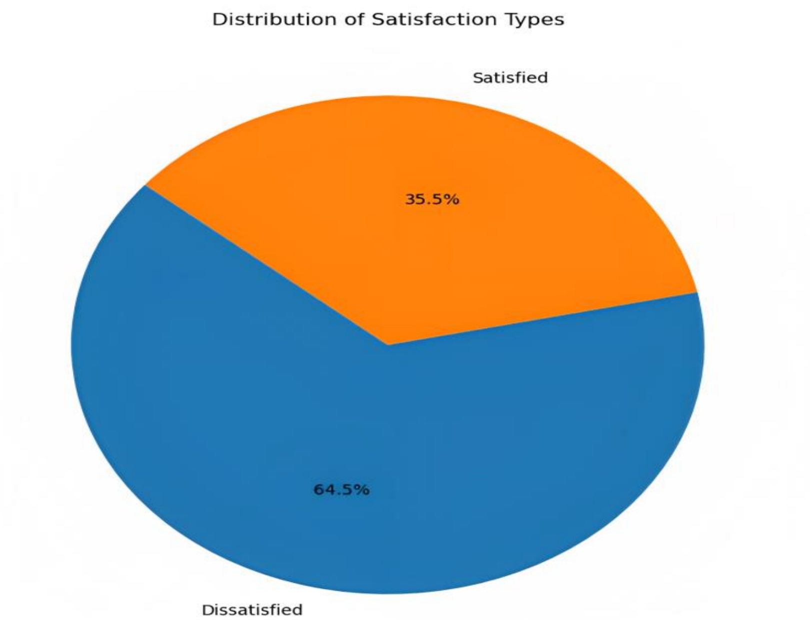
# Exploratory Data Analysis

- **Data Cleaning**
    - **Inventory Dataset (Table 1)** - Key Columns Selection
        - We narrowed down the focus by selecting key columns from Table 1 (inventory dataset), including flight details, entree information, and consumption statistics, while ensuring that the dataset is free of any null or missing values.
    - **Inflight Service_Pre order data (Table 2)** - Missing Values Handling
        - From the "Inflight Service_Pre order data" table, we carefully excluded 20,311 rows out of the total 212,039 rows due to missing values in columns such as "meal_group," and "meal_category".
        - This curation process was undertaken to ensure data integrity and accuracy for our analysis.
    - **Survey data _Customer comments (Table 3)** - Relevant Columns Selection
        - From the "Survey data _Customer comments" table, we selected pertinent columns including 'flight_number,' 'origin_station_code,' 'destination_station_code,' 'arrival_delay_group,' 'departure_delay_group,' 'entity,' 'verbatim_text,' 'seat_factor_band,' 'ques_verbatim_text,' and 'loyalty_program_level.'
        - While there were 2,504 missing values in 'loyalty_program_level' out of a total of 9,424 values, they were supplemented with "NA" for completeness.

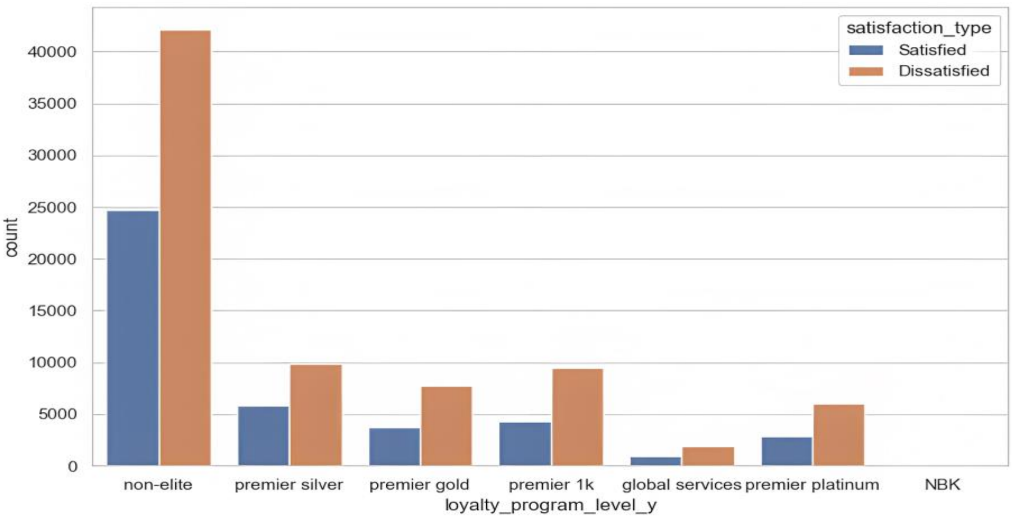
- It is noteworthy to emphasize that our primary focus revolves around 'verbatim_text' as it is pivotal for the conducted Natural Language Processing (NLP) analysis.

- **Survey data_Inflight Satisfaction Score (Table 4)** - Passenger Satisfaction Calculation
    - In the 'Survey data_Inflight Satisfaction Score' table, comprising essential attributes like flight details, passenger feedback scores, and contextual information, we derived the passenger satisfaction score by aggregating the count of 'satisfied' responses and dividing it by the total count of responses for each flight.
    - This satisfaction score reflects the percentage of satisfied passengers in the dataset.
    - We also managed missing values by imputing 12,111 instances of 'satisfaction_type' with the mode and replacing 11,616 missing values in 'loyalty_program_level' with 'NA', ensuring a comprehensive analysis.

## Data Visualization

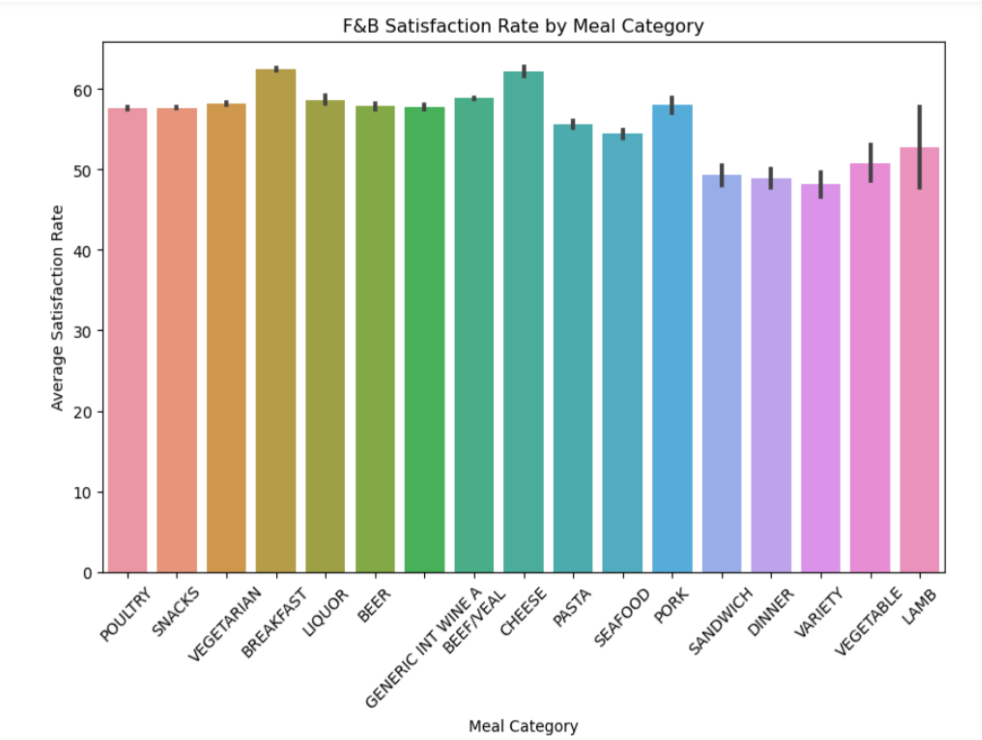



Indicates that a significant portion of passengers, approximately **64.5%,** expressed dissatisfaction with the in-flight dining experience based on the survey responses.



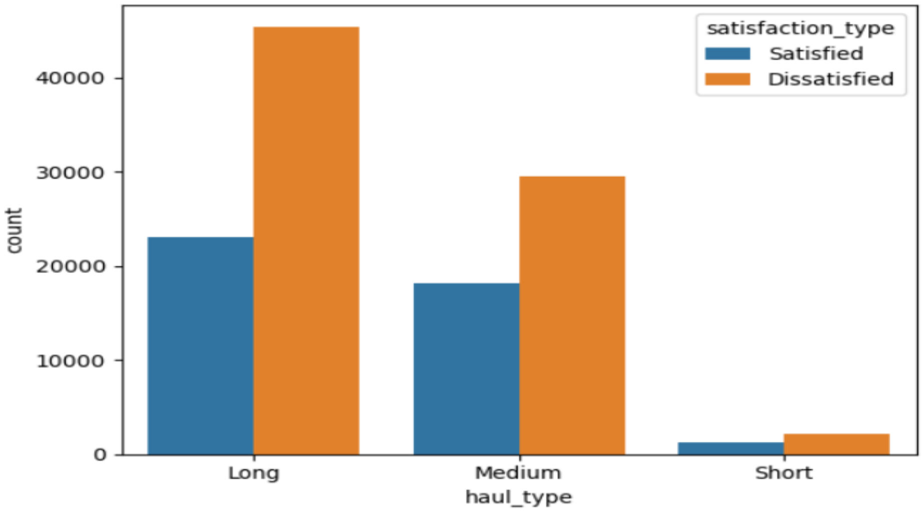

```
satisfaction_type        Dissatisfied Percentage
loyalty_program_level_y
NBK                                    54.545455
global services                        66.363954
non-elite                              63.040000
premier 1k                             68.640558
premier gold                           67.371996
premier platinum                       68.074534
premier silver                         62.680767
```

Dissatisfied percentage is highest among "**premier 1k**" (68.6%), "**premier platinum**" (68.0%7), "**premium gold**" (67.3%) for Loyalty Program levels



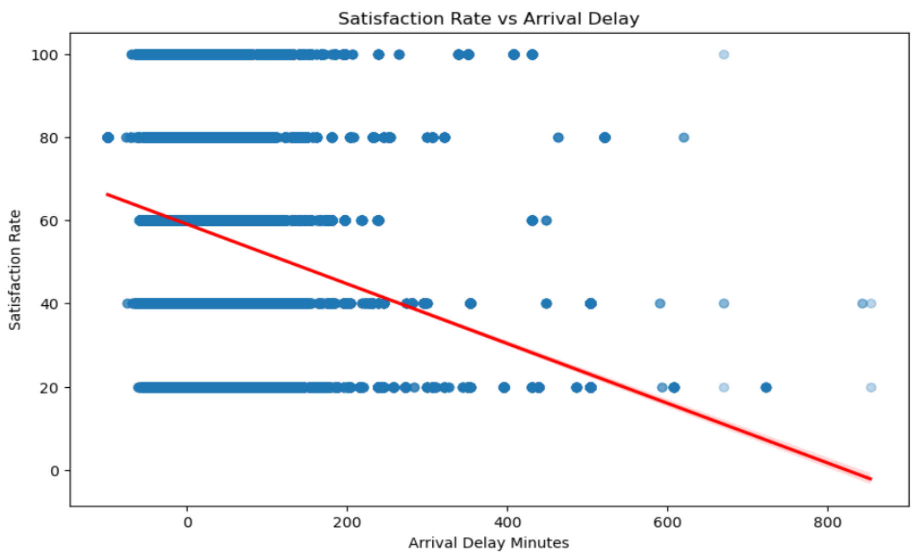

```
Meal Category: VARIETY
Average Satisfaction Rate: 48.16%
------------------------------
Meal Category: DINNER
Average Satisfaction Rate: 48.92%
------------------------------
Meal Category: SANDWICH
Average Satisfaction Rate: 49.29%
------------------------------
Meal Category: VEGETABLE
Average Satisfaction Rate: 50.83%
------------------------------
Meal Category: LAMB
Average Satisfaction Rate: 52.73%
------------------------------
```

**Variety** (48.16%), **sandwich**(49.29%) , **dinner** (48.92%) shows the least satisfaction rate among meal categories.

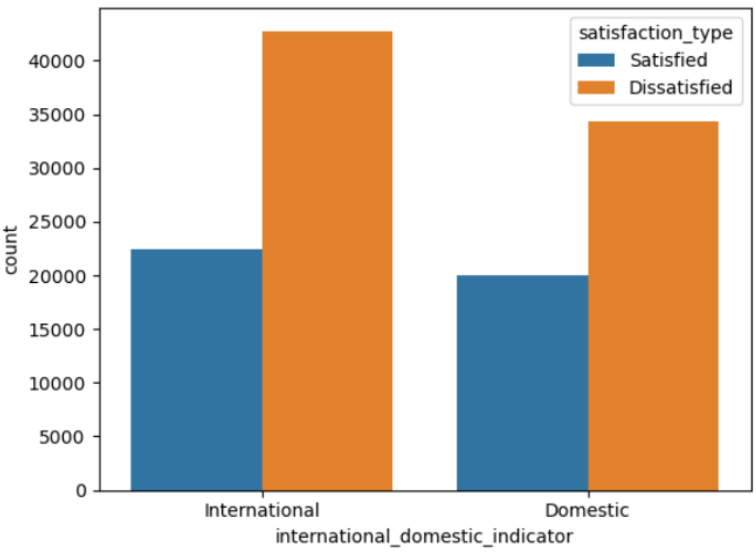



| satisfaction_type | Dissatisfied Percentage |
|---|---|
| haul_type | |
| Long | 66.352917 |
| Medium | 61.960488 |
| Short | 62.934132 |

Dissatisfied percentage is highest for **Long haul_type (66.35%)**

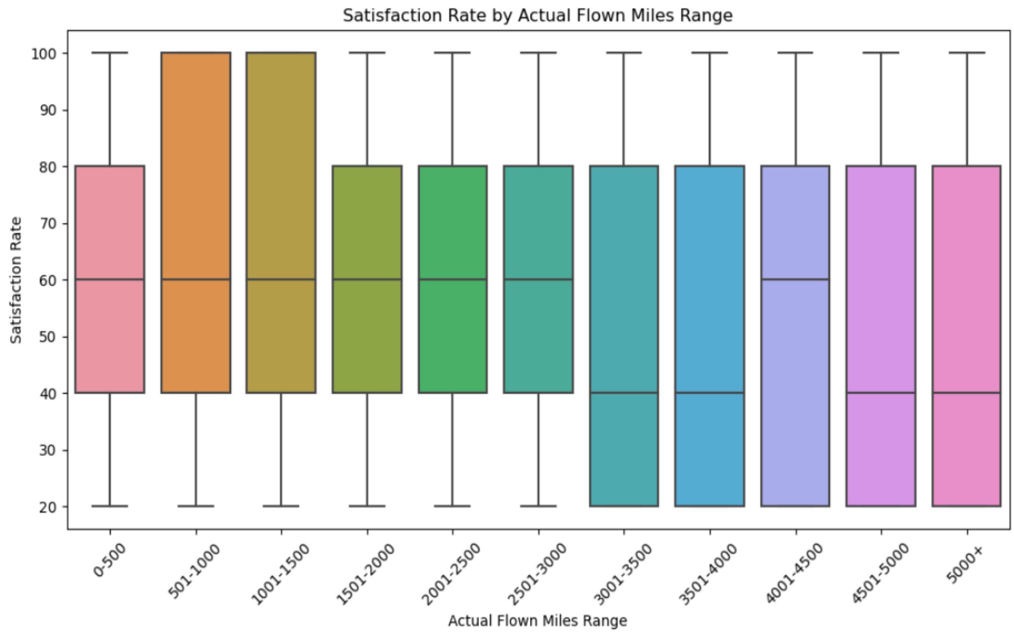



The satisfaction rate tends to decrease as the arrival delay minutes increase.

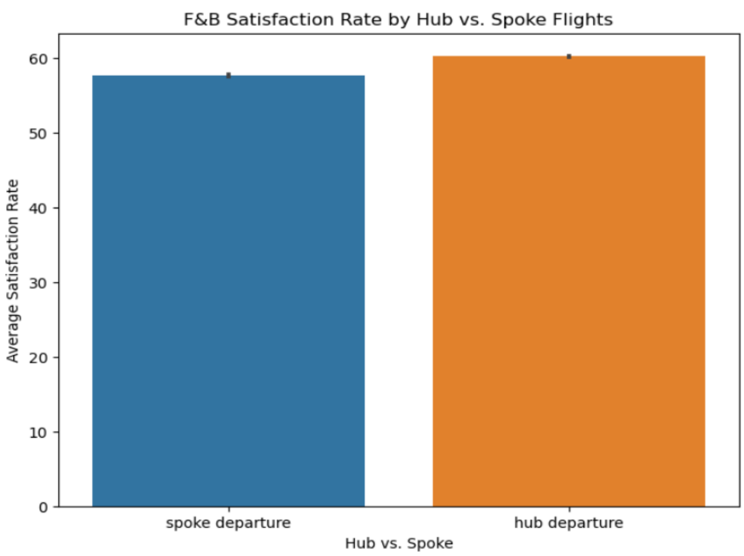



| satisfaction_type | Dissatisfied Percentage |
| --- | --- |
| international_domestic_indicator | |
| Domestic | 63.145410 |
| International | 65.640538 |

Dissatisfied percentage is more for **International** Flights. (65.64%)

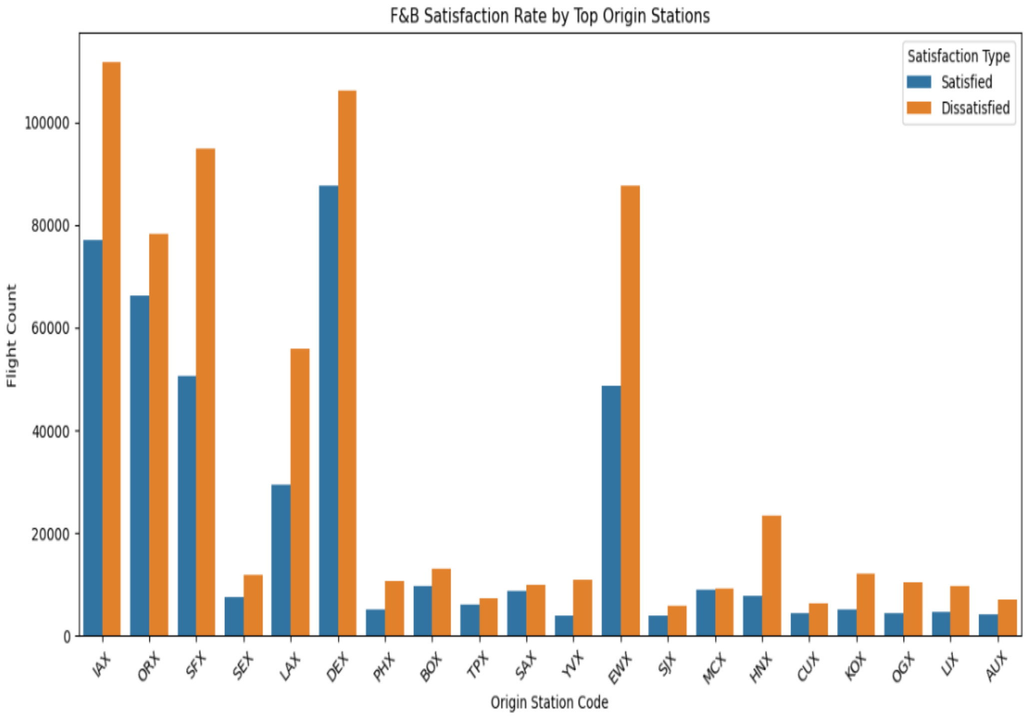



From the Graph we can see that Satisfaction Rate is less for flights that have flown miles greater than **3000+** miles.



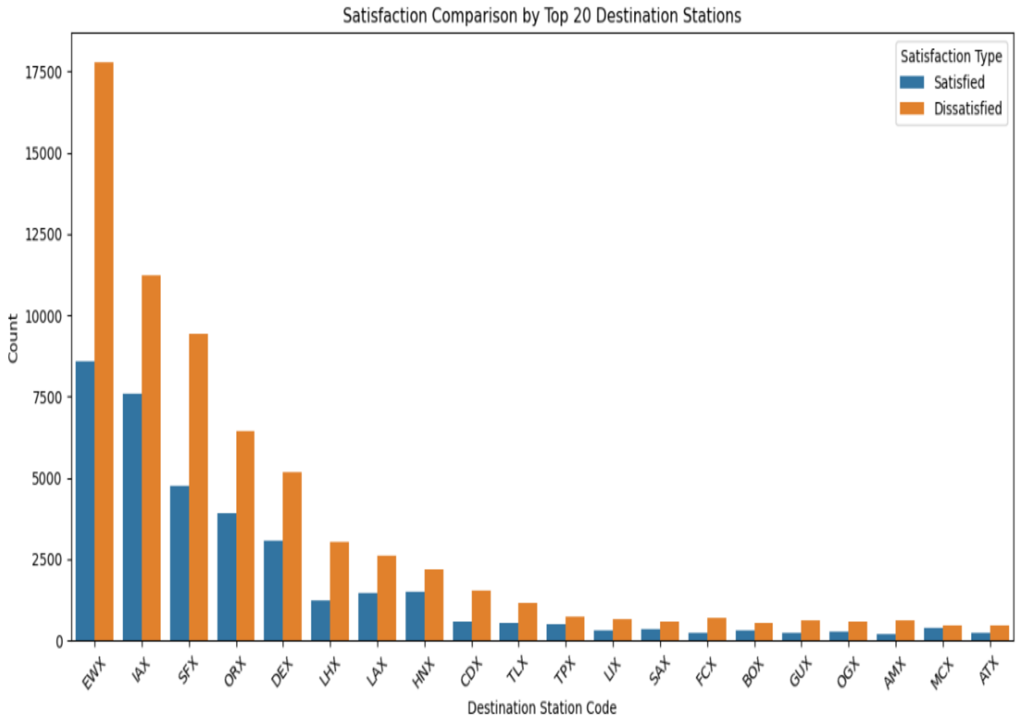

Satisfaction rate for **spoke departures** is slightly less than **hub departures**.

A clear pattern emerges: Factors such as flight type, distance, and departure location significantly influence passenger satisfaction. International flights exhibit a higher dissatisfaction rate, while longer distances and spoke departures are associated with lower satisfaction.




| satisfaction_type | origin_station_code | Dissatisfied | Satisfied |
|---|---|---|---|
| 5 | HNX | 0.751231 | 0.248769 |
| 7 | KOX | 0.701699 | 0.298301 |
| 8 | LAX | 0.654547 | 0.345453 |
| 9 | LIX | 0.669492 | 0.330508 |
| 11 | OGX | 0.708353 | 0.291647 |
| 13 | PHX | 0.683461 | 0.316539 |
| 16 | SFX | 0.652787 | 0.347213 |
| 19 | YVX | 0.740647 | 0.259353 |

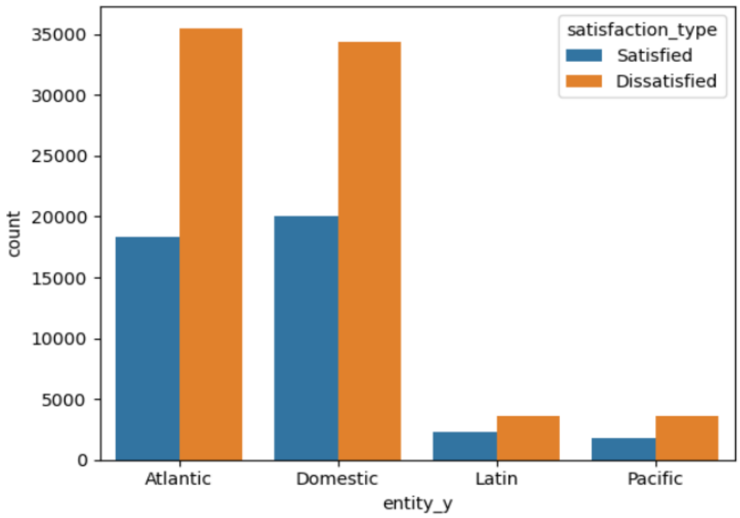
Origin Stations with Dissatisfied percentage more than 65%. Some origin_station_codes with highest dissatisfied percentage are HNX (75%) , YVX (74%) , OGX (70.8%)




| station_code_x | Dissatisfied | Satisfied |
|---|---|---|
| AMX | 0.748810 | 0.251190 |
| CDX | 0.728933 | 0.271067 |
| EWX | 0.674446 | 0.325554 |
| FCX | 0.752391 | 0.247609 |
| GUX | 0.726857 | 0.273143 |
| LHX | 0.708460 | 0.291540 |
| LIX | 0.670020 | 0.329980 |
| OGX | 0.674584 | 0.325416 |
| SFX | 0.664925 | 0.335075 |
| TLX | 0.678819 | 0.321181 |

Destination Stations with Dissatisfied percentage more than 65%. Some destination_station_codes with highest dissatisfied percentage are FCX (75.23%) , AMX (74.8%) , CDX (72.8%)

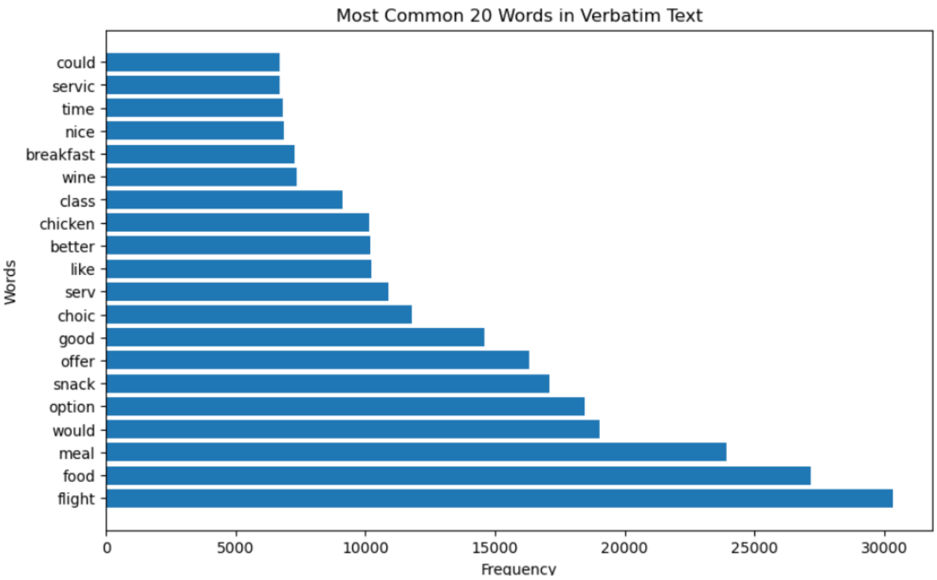
Key Note :- Station Codes [“**LIX**” , “**DGX**” , “**SFX**”] are common in both destination stations and origin stations with more than 65% Dissatisfaction rate .



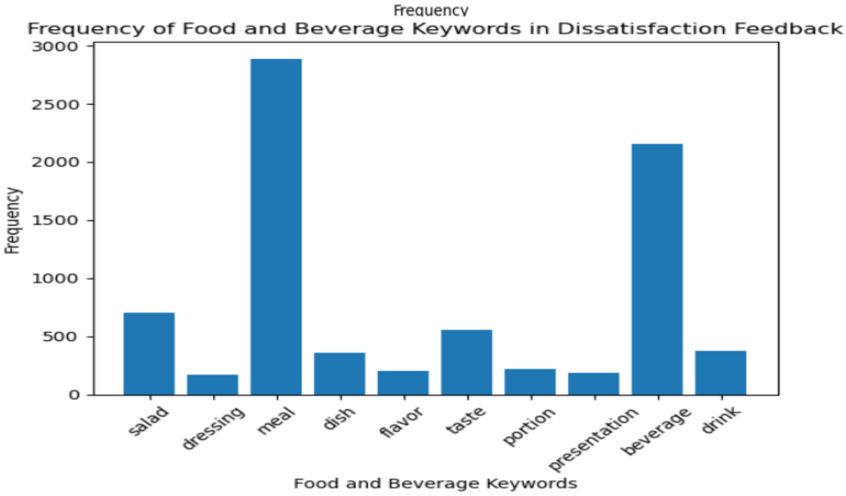

```
satisfaction_type  Dissatisfied Percentage
entity_y
Atlantic                         66.022592
Domestic                         63.145410
Latin                            61.105459
Pacific                          66.776920
```

Dissatisfied Percentage is highest for **Atlantic** and **Pacific entities** with 66.02% and 66.77% respectively.

# NLP




```
Pain Point Keywords and Their Frequencies:
delay: 480
cancel: 66
uncomfortable: 0
issue: 0
problem: 539
bad: 1229
poor: 1425
disappoint: 2658
```




Based on the analysis of these pain point keywords, it seems that flight delays, meals ,cancellations, Salad, dressing , problems, dissatisfaction, beverage ,and disappointment are prominent pain points mentioned in the customer feedback.

```
# Create a DataFrame to store the sentences
food_beverage_sentences = pd.DataFrame({'sentence': unique_sentences_list})

# Define a function to classify sentiment
def classify_sentiment(text):
    analysis = TextBlob(text)
    polarity = analysis.sentiment.polarity
    return "negative" if polarity < 0 else "positive" if polarity > 0 else "neutral"

# Apply sentiment classification to each sentence
food_beverage_sentences['sentiment'] = food_beverage_sentences['sentence'].apply(classify_sentiment)

# Filter out only the negative feedback sentences
negative_feedbacks = food_beverage_sentences[food_beverage_sentences['sentiment'] == 'negative']

print("Negative Feedback Sentences:")
for index, row in negative_feedbacks.iterrows():
    print(f"Sentence {index + 1}: {row['sentence']}")
```

This code extracts **only negative feedbacks** from all sentences given in verbatim_text (as food_beverages_sentences) using **NLP sentiment analysis .**

```
# Load stopwords
stop_words = set(stopwords.words('english'))

# Preprocess sentences by removing stopwords and tokenizing
def preprocess_sentence(sentence):
    tokens = word_tokenize(sentence)
    tokens = [word for word in tokens if word.lower() not in stop_words]
    return ' '.join(tokens)

# Preprocess all sentences in the DataFrame
negative_df['preprocessed_sentence'] = negative_df['sentence'].apply(preprocess_sentence)

# Create TF-IDF vectorizer
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(negative_df['preprocessed_sentence'])

# Calculate cosine similarity matrix
cosine_sim = cosine_similarity(X)

# Group similar sentences based on similarity threshold
similar_sentence_groups = {}

# Iterate through the cosine similarity matrix
for i, row in enumerate(cosine_sim):
    sentence = negative_df.iloc[i]['sentence']
    cluster = negative_df.iloc[i]['cluster']
    similar_sentences = [negative_df.iloc[j]['sentence'] for j, score in enumerate(row) if score > 0.2 and j != i]

    if len(similar_sentences) > 1:  # Adjust this threshold as needed
        if cluster not in similar_sentence_groups:
            similar_sentence_groups[cluster] = {sentence: similar_sentences}
        else:
            similar_sentence_groups[cluster][sentence] = similar_sentences
```

Considering the substantial number of sentences present within the 'negative_feedbacks' dataset, extracting meaningful insights from them presented a considerable challenge. To effectively address this complexity, we adopted a **sentence clustering** methodology utilizing the **TF-IDF vectorization and cosine similarity to identify similar sentences based on their content.**

2 Examples Given below :-

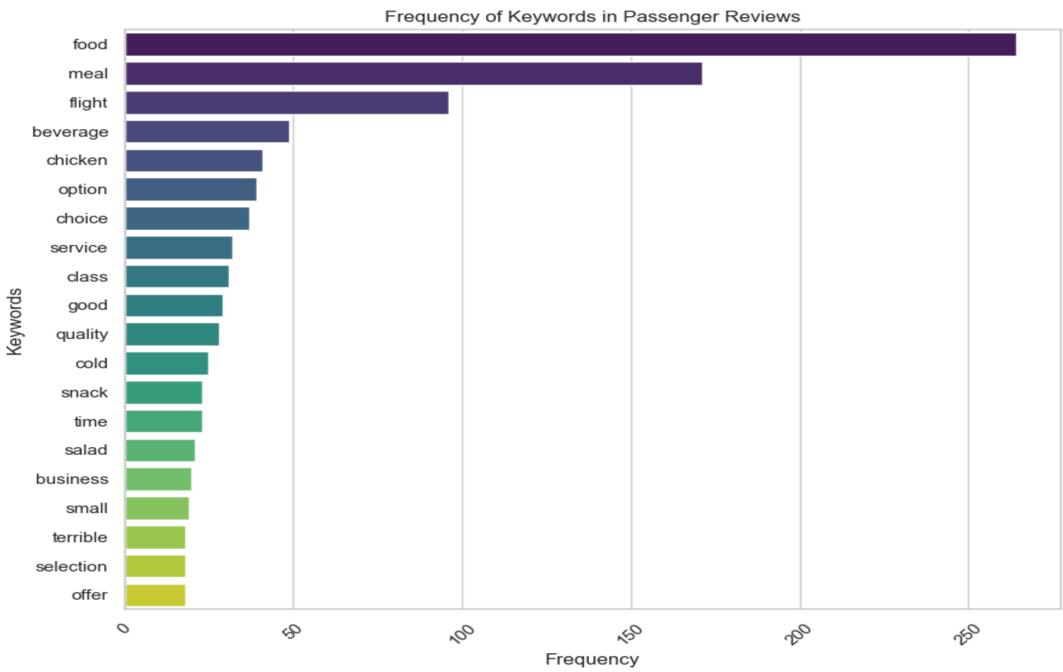
```
Original Sentence: I assumed there was not meal for my late night flight but was pleasantly surprised when I was on the airpl
ane and received a meal.
Similar Sentences: ['The meal I received was very salty, Chicken Risotto.', 'I felt that the meal service came way too late i
n this flight.', 'Surprised that there was no snack or any other food offered for an 8:40 a.m. flight.', 'Now this is an inte
rnational flight, the least you can do is serve a hot meal especially being that the flight was so late in the evening.', 'In
past seemed quality was going down so I was pleasantly surprised that food service seemed to be improved.', 'The crew meal lo
oked better than the chicken meal I received.']

Original Sentence: Very little interaction with United flight attendants on the long haul after meal was served.
Similar Sentences: ['Meals should be served on long flights if it is meal time.', 'The breakfast was very little and the food
too.', 'It is been a long time since United had my GF meal on the flight.', 'Strange change to meal service: Flight attendant
s removed wine glasses from meal trays as they were served and did not offer any beverages with dinner.', "Why should United
dictate what I must have, when it is United's fault that the meal United offered me during preordering is not available?.",
'Flight attendants great- but the chicken orzo meal tasted awful.']
```

Furthermore, following the original sentence and similar sentence clustering process, we employed the provided code to extract and present the **top 30 most recurrent sentences** from the clustered data. This step enabled us to spotlight the most frequently voiced concerns and sentiments within the passenger feedback, serving as a valuable resource for pinpointing significant pain points and crafting tailored improvements.

```python
import spacy
from collections import Counter

# Load the spaCy model
nlp = spacy.load("en_core_web_sm")

# Combine all preprocessed sentences into a single text
all_preprocessed_text = ' '.join(negative_df['preprocessed_sentence'])

# Process the combined text using spaCy
doc = nlp(all_preprocessed_text)

# Extract nouns and adjectives as keywords
keywords = []
for token in doc:
    if token.pos_ in ['NOUN', 'ADJ']:
        keywords.append(token.lemma_)

# Count the frequency of each keyword
keyword_counts = Counter(keywords)

# Print the top keywords and their counts
num_keywords_to_display = 20  # You can adjust this value
for keyword, count in keyword_counts.most_common(num_keywords_to_display):
    print(f"Keyword: {keyword} | Count: {count}")

# Print example sentences for the top keywords
num_sentences_per_keyword = 5  # You can adjust this value
for keyword, _ in keyword_counts.most_common(num_keywords_to_display):
    print(f"\nExamples for Keyword: {keyword}")
    sentences_with_keyword = [sentence for sentence in negative_df['sentence'] if keyword in sentence.lower()]
    for sentence in sentences_with_keyword[:num_sentences_per_keyword]:
        print(f"- {sentence}")
```

Subsequently, we used a provided code snippet that leveraged spaCy's language processing capabilities. By combining preprocessed sentences, spaCy identified important nouns and adjectives, which we considered as **key themes**. We then counted the frequency of these themes to understand what came up most often. Additionally, for the top 20 themes, we showed how often they appeared and provided the **five most common sentences** related to each theme. This process gave us a clear view of the main recurring concerns and feedback, helping us uncover actionable insights and focused recommendations.

```
Examples for Keyword: food
- In general I think the food choices on United flights is not great.
- This is in regards to the food in general (Houston to cun), you have to get rid of the zatar chicken.
- The food is strangely odd.
- I was confused about the food/ snacks on board.
- In general food provided across the airline industry is poor to terrible.

Examples for Keyword: salad
- Chicken meal was served with couscous - basically the same taste/ingredients as the salad.
- The salad was very strange - I was not able to eat it.
- Cold wheat berry salad has been on the menu for a number of years and it is inedible and most of the passengers Avoid eatin
g it.
- They provided me a curry dish with Chicken in it and a salad with some white creamy dressing.
- The dinner was served hot but salad as it should be was cold.
```




```
Keyword: food | Count: 227
Keyword: meal | Count: 161
Keyword: flight | Count: 74
Keyword: chicken | Count: 42
Keyword: salad | Count: 42
Keyword: Food | Count: 39
Keyword: beverage | Count: 38
Keyword: service | Count: 30
Keyword: class | Count: 30
Keyword: options | Count: 29
Keyword: quality | Count: 28
Keyword: good | Count: 28
Keyword: cold | Count: 27
Keyword: other | Count: 23
Keyword: choices | Count: 18
Keyword: flights | Count: 18
Keyword: terrible | Count: 18
Keyword: time | Count: 18
Keyword: choice | Count: 18
Keyword: business | Count: 18
```

# Major Pain Points & Improvements

| | |
|---|---|
| Food | **Problems** :<br>• Overall dissatisfaction with food quality across different meal types.<br>• Inconsistencies in taste and flavor profiles.<br>• Expectations for a broader and more diverse range of food options.<br>• Concerns about freshness and ingredients used.<br>• Lack of diversity in menu options, Negative perception of the presentation and visual appeal of dishes.<br>**Improvements**:<br>• Invest in sourcing higher quality ingredients.<br>• Implement quality control measures to ensure consistent taste.<br>• Prioritize using fresh and locally sourced ingredients.<br>• Collaborate with chefs to create rotating seasonal menus that cater to different tastes.<br>• Expand the menu to include a variety of international and local dishes. |
| Meals | **Problems** :<br>• Complaints about portion sizes being insufficient.<br>• Limited options for different dietary preferences.<br>• Concerns about the variety of side dishes and accompaniments.<br>• Passengers looking for more appealing and innovative meal offerings.<br>• Concerns about meal choices during long-haul flights.<br>**Improvements**:<br>• Increase portion sizes to satisfy hunger for passengers.<br>• Focus on improving the taste and flavor of meals.<br>• Include more options for vegetarian, vegan, and gluten-free diets.<br>• Introduce special menu items that showcase local cuisine and cultural flavors.<br>• Rethink the menu offerings to provide more appealing choices. |
| Options & Choices | **Problems** :<br>• . Limited food options for passengers with dietary restrictions.<br>• Inadequate options for vegetarian, vegan, and gluten-free diets.<br>• Passenger dissatisfaction with the available meal choices.<br>**Improvements**:<br>• Develop a comprehensive menu that includes options for various dietary needs.<br>• Clearly label menu items to indicate dietary preferences and allergen information.<br>• Provide more customizable options that allow passengers to create their own meals.<br>• . Conduct surveys to understand passengers' preferred meal options. |

| | |
|---|---|
| Salad & Dressing | **Problems** :<br>• Many Complaints about Salad Dressing.<br>• Mixed opinions about the quality and taste of salad options.<br>• Passengers looking for more flavorful and appealing salad choices.<br>• Dissatisfaction with the overall freshness and presentation of salads.<br><br>**Improvements**:<br>• Source high-quality and fresh ingredients for salads.<br>• Experiment with diverse ingredients and dressings to enhance salad flavors.<br>• Prioritize proper storage and preparation to maintain salad freshness. |
| Beverages | **Problems** :<br>• Limited selection of beverages, especially non-alcoholic options.<br>• Dissatisfaction with the taste and quality of available beverages.<br>• Concerns about beverage choices offered with meals.<br>• Inadequate beverage options for passengers with dietary restrictions.<br>**Improvements**:<br>• Expand the range of beverage choices, including non-alcoholic drinks.<br>• Improve the taste and quality of both alcoholic and non-alcoholic beverages.<br>• Pair appropriate beverages with different meal options for a better pairing.<br>• Ensure availability of popular beverages preferred by passengers. |
| Snacks | **Problems** :<br>• Dissatisfaction with the quality and variety of offered snacks.<br>• Passengers looking for more satisfying and substantial snack options<br>**Improvements**:<br>• Offer a range of high-quality snacks that cater to different preferences.<br>• Introduce heartier and healthier snack choices to provide more satisfaction.<br>• Ensure availability of appealing snacks throughout the duration of the flight. |

| | |
|---|---|
| Chicken | **Problems** :<br>• Mixed opinions about the taste and flavor of chicken dishes.<br>• Dissatisfaction with the quality of chicken used in meals.<br>• Inconsistencies in the tenderness and juiciness of chicken.<br>• Negative feedback about the seasoning and sauces used with chicken.<br>**Improvements**:<br>• Conduct taste tests to refine the flavor profile of chicken dishes.<br>• Implement cooking techniques that maintain chicken's tenderness.<br>• Train kitchen staff to consistently prepare chicken to the desired quality.<br>• Experiment with different seasonings and sauces to enhance chicken dishes. |

| | |
|---|---|
| Cold Meals | **Problems** :<br>• Complaints about meals being served cold.<br>• Passengers dissatisfied with improper heating of food.<br>**Improvements**:<br>• Implement strict quality checks to ensure meals are served at the correct temperature.<br>• Use appropriate packaging and serving methods to retain heat during meal service.<br>• Train cabin crew to prioritize proper heating and presentation of meals. |

| | |
|---|---|
| Timing of meals | **Problems** :<br>• Passengers expecting timely delivery of meals and snacks.<br>• Frustration when meals are not served during expected mealtime hours.<br>**Improvements**:<br>• Coordinate meal service with flight schedules to align with mealtime expectations.<br>• Train cabin crew to prioritize prompt and efficient meal delivery.<br>• Communicate clearly with passengers about meal service timings and changes. |

| | |
|---|---|
| Inconsistency in food items in different classes | **Problems** :<br>• Discrepancies in meal quality between different flight classes.<br>• Inconsistency in meal presentation and options based on class.<br>**Improvements**:<br>• Maintain a consistent standard of meal quality across all flight classes.<br>• Regularly assess and improve meal options in all classes, not just premium ones. |

## Conclusion

In our comprehensive analysis, we have distilled key pain points that consistently emerge from passenger feedback. These include concerns related to **food quality**, **meal portions**, **menu diversity**, and **presentation**. Additionally, issues such as **inconsistencies in taste**, **freshness**, and **ingredient sourcing**, as well as **variations** in dining experience across flight classes and **long-haul flights**, have surfaced as significant challenges. The culmination of these insights underscores the imperative to address these core pain points to enhance the overall in-flight dining experience, making air travel a more satisfying journey for passengers across all flown miles.

## Next Steps

In our journey of uncovering insights from passenger comments, we've identified significant pain points related to various aspects like food quality, meal portions, and in-flight experience. Moving forward, to enhance our analysis, a larger dataset encompassing a broader range of summer months would provide more comprehensive results. Additionally, refining the TF-IDF vectorizer by exploring n-gram ranges and adjusting stop words could yield improved keyword extraction. For cosine similarity, incorporating more advanced techniques like Word2Vec or Doc2Vec might enhance sentence clustering. Furthermore, augmenting our analysis with contextual data, such as flight routes, passenger demographics, and airline policies, could unlock deeper insights into passenger sentiments and preferences, enriching our understanding of their experiences.

GitHub Link for this Project