SKYHACK
HACKATHON

2K20/MC/049 Gaurav Jain

TABLE OF CONTENTS

EXECUTIVE SUMMARY

- In this project, we conducted a comprehensive analysis of customer feedback data related to United Airlines' food and beverage (F&B) services. Our approach involved dataset segmentation, thorough data cleaning, and sentiment analysis to understand customer sentiments. We also examined potential factors impacting F&B satisfaction, categorizing data by satisfaction types, and creating unique value frequency summaries. In addition to these steps, we generated multiple bar charts to visualize the relationship between customer satisfaction and various influencing factors, providing a well-rounded view of F&B service dynamics.

UNDERSTANDING THE PROBLEM

- Root Cause Analysis for F&B Satisfaction During Summer Months:
  - Identify the underlying reasons behind fluctuations in food and beverage (F&B) satisfaction scores specifically during the summer months and analyze data to pinpoint the key drivers influencing F&B satisfaction during this period.
- Analysis of Survey Comments:
  - Gain insights from customer feedback by analyzing survey comments and understand the major themes and issues related to F&B that customers are complaining about through natural language processing (NLP) or text analysis
- Coding Skills Showcase:
  - Demonstrate technical proficiency in SQL, Python, or R to manipulate and summarize data effectively and write queries or scripts to extract, transform, and summarize data at various levels for meaningful conclusions.
- Storytelling and Initial Recommendations:
  - Present the analysis findings cohesively in a narrative format and tell a story using data to convey the factors impacting F&B satisfaction during summer and provide initial recommendations based on insights.

APPROACHING THE DATASETS

Step 1: Data Understanding

- Importing essential libraries such as 'numpy' and 'pandas.'
- Beginning by comprehensively understanding the dataset using the dataset description to gain insights into its context and purpose and also examination of the types of values and frequency of such values present in each column.

Step 2: Data Cleaning

- Interpreting and addressing missing values using appropriate methods.
- Identifying and dealing with outliers to ensure data integrity.
- Creating more descriptive categories to enhance data interpretability.
- Removing redundant or unnecessary columns that do not contribute to the analysis.

APPROACHING THE DATASETS

Step 3: Data Segmentation

- Split the cleaned dataset into two distinct tables based on the "question_type" column.
- This separation helps us focus on each question separately since the “score” column now contains answers from either question in different tables

Step 4: Further Segmentation

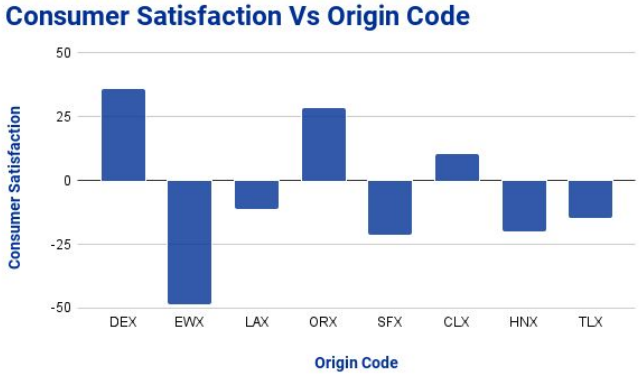
- Within one of the segmented tables, further split the data based on "satisfaction_type."
- This results in two additional tables—one for satisfied customers and one for dissatisfied customers.
- This is useful for fine data manipulation and data extraction

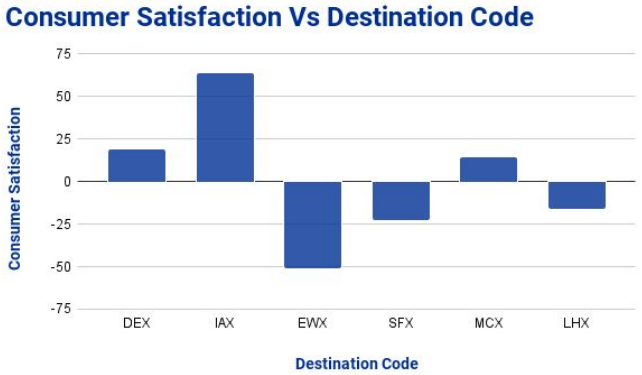
APPROACHING THE DATASETS

Step 5: Creating Factors Affect DataFrame

- Create a new dataframe to store unique values and their frequencies from each column of the previously segmented dataset.
- For each column in the original dataset:
  - Calculate the frequency of each unique value across the entire dataset.
  - Calculate the frequency of each unique value separately for satisfied and dissatisfied customers.
  - Add these frequencies to the new dataframe.
  - Also add the percent of each unique value with respect to the number of entries in that column
- Using all values entered in this dataframe, calculate the effect of the occurrence of each unique value and its positive or negative effects on the customer satisfaction
- Lastly, remove all factors that have very less effect on the consumer satisfaction

ANALYSING THE RESULTS



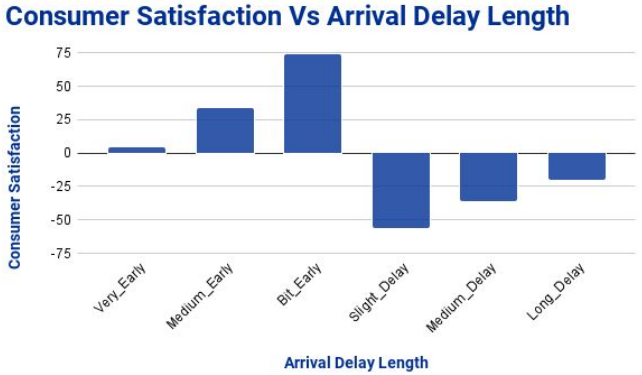


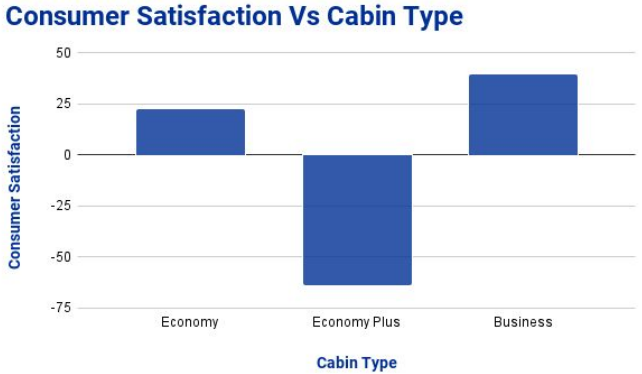
Using the previously generated table, I extracted the impact of each factor and represented it visually through bar charts that showcase the relationship between customer satisfaction with Food and Beverages (F&B) and select relevant factors from categories.

In these particular charts:

- There are some airports that appear both times, for example:
  - EWX: this airport seems to decreases the F&B satisfaction relatively a lot
  - SFX: similar to EWX, this also decreases F&B satisfaction but not as much
  - DEX: this airport generally increases the F&B satisfaction by a bit

ANALYSING THE RESULTS





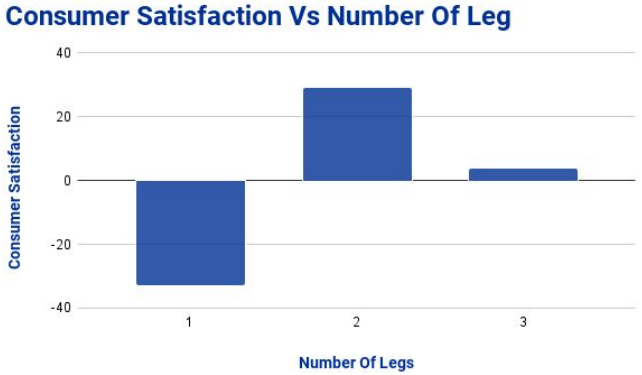
In the first chart:

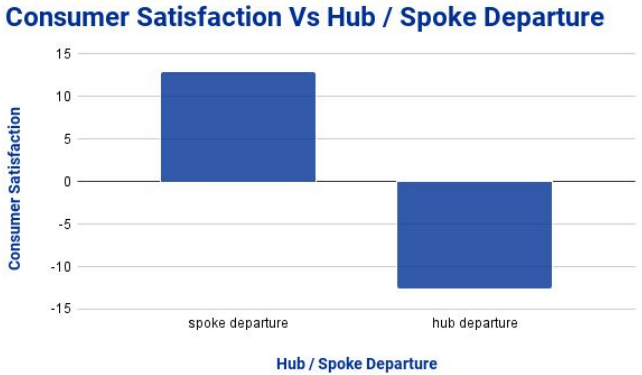
- We can visually see a very reasonable relationship:
  - If delay, then dissatisfaction and vice versa
- However, there is a point that a long-delay would cause more dissatisfaction than slight delay. The reason why this is not the case is because the calculation also takes into account the frequency of such cases. Since the algorithm focuses more on more frequent factors of consumer satisfaction, the rare cases are somewhat less considered

In the second case:

- It is interesting to see that those who pay for economy plus are less satisfied than those who get economy or business class. Other cabin types are near 0, hence not considered

ANALYSING THE RESULTS





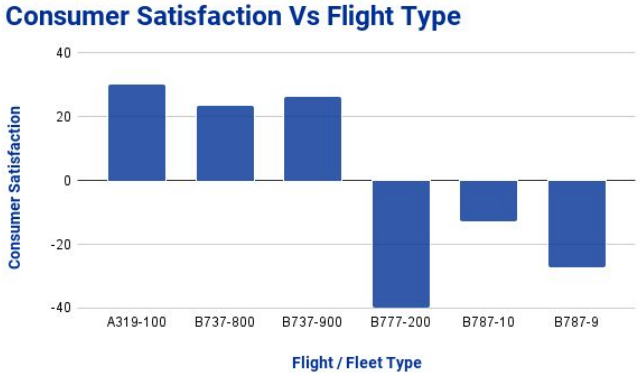
In the first chart:

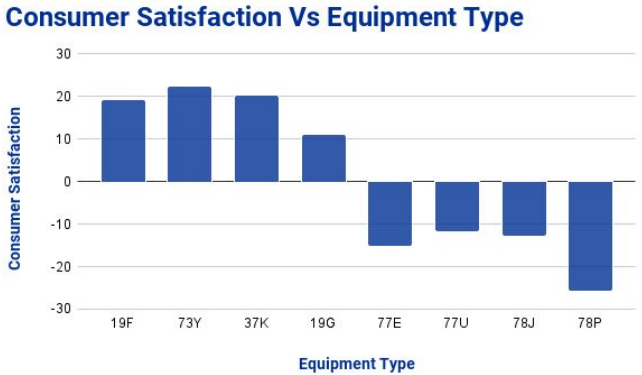
- It is interesting to see that those without a stop in between are the most dissatisfied whereas those who have stop(s) are more satisfied, it can be assumed that this is due to the fact that many 1 leg flights are short in duration and some don't have meal time resulting in dissatisfaction

In the second chart:

- The hub departures are those leaving from a larger airport having many facilities, hence one would assume the F&B services here would be satisfactory but the data seem to show otherwise. It is also worth noting the scale that each chart is showing as they are all ranging differently

ANALYSING THE RESULTS





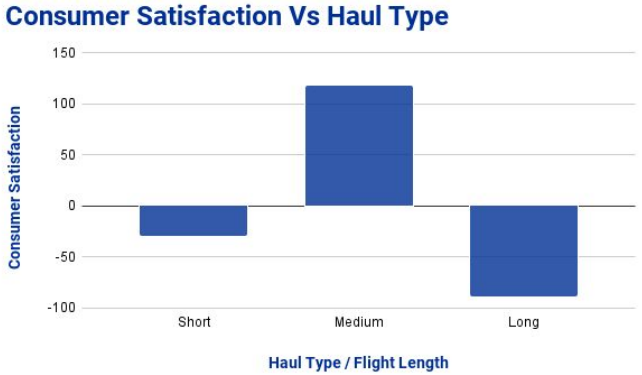
In the first chart:

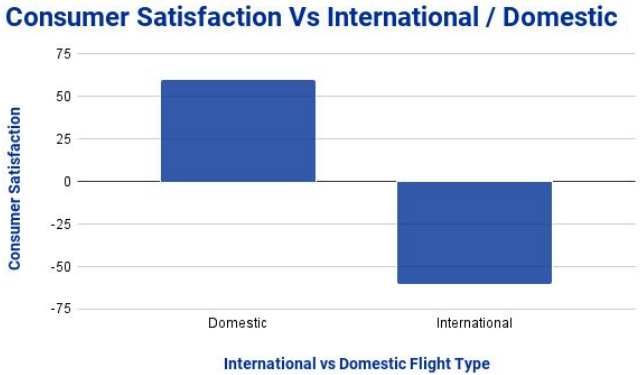
- We can clearly see that some of the fleet / airplane types are preferred over the others, namely:
  - A319-100, B737-800, B737-900
  - It would be smarter to try to acquire more of these fleet types to increase the customer satisfaction

Similarly in the second chart:

- We can that even for equipment type, there are some that are better as they increase consumer satisfaction, like:
  - 19F, 73Y, 37K, 19G

ANALYSING THE RESULTS

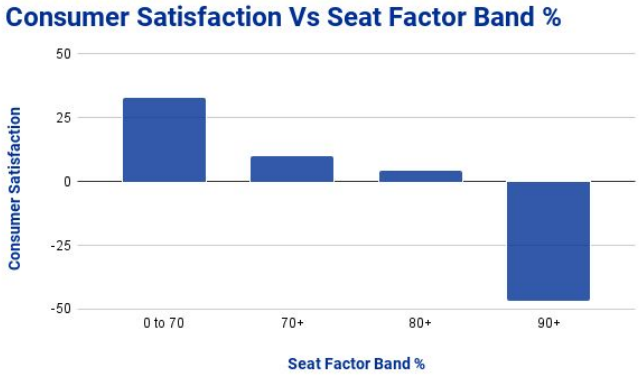


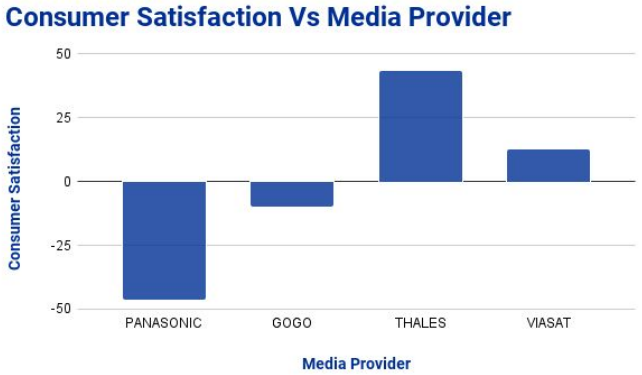


In these charts:

- We can see an understandable dissatisfaction when it comes to long flights (which is also reflected in the second chart as most international flight tent to be longer) as at this point we can assume that the journey itself is creating more dissatisfaction than the F&B service
- The short journeys are also slightly in dissatisfaction due to the possibility that they may not serve food at all
- However we can see that the medium and domestic flights are really satisfied with their services, paying attention the first chart, the medium length has even broken the 100 unit barrier

ANALYSING THE RESULTS





In the first chart:

- Flight that are filled to the brim also cause customer dissatisfaction as even your personal space gets cramped
- We can assume that the flight attendants are busy hence your food also may get delay leading to high dissatisfaction
- And we can see this hypothesis work when we consider the converse of this situation

In the second chart:

- It is interesting that even a media provider can change a customers F&B experience for example Parasonic causes a lot of dissatisfaction while Thales is causing satisfaction.

ANALYSING THE RESULTS

In this chart:

- It is fascinating to see that those who are not a member of united are way more satisfied than those who are a member of the loyalty program.
- It can be assumed that those who are non-member are not frequent-flyers meaning that the high satisfaction is coming not from F&B but the flying experience itself.

PLANNING FOR THE FUTURE

With more time available, I would have proceeded by taking the partially segmented table from first dataset and applying various machine learning algorithms to assess which one best predicts customer satisfaction or dissatisfaction.

I would then select the most effective algorithm and apply it to the second half of the first dataset that I had segmented in the 3rd step , which contains responses to the second question and details about the types of food and beverages (F&B). This approach would have allowed us to discern which specific types of F&B contribute to customer satisfaction and which do not.

PLANNING FOR THE FUTURE

Subsequently, I would have used natural language processing on the second dataset, which contains customer reviews of F&B. This analysis would involve sentiment analysis to identify which F&B items are well-received and which are criticized. If my hypothesis holds true, the ratio of positive to negative sentiment for each F&B category would align with the results obtained from the machine learning model. Throughout this process, I would have also noted common themes recurring in customer reviews for further drawing conclusions.

Finally, I would have conducted data analysis on the planned versus consumed F&B counts available in the inventory dataset to gain insights into consumption patterns and potential areas for optimization.

PROJECT LINK

Hackathon Link:

https://skyhack.hackerearth.com/

GitHub Link:

https://github.com/Gaurav7877/SkyHack_Hackathon

SkyHack Hackation Excel Sheet Link (G-Drive)

THANK YOU