# Sky-High Insights: Overview

## Sky-High Insights: The Overview

The British Airways Review Dataset from Kaggle, which spans the years 2012 to 2023, offers a unique opportunity to analyze a wealth of passenger feedback across a diverse range of routes globally. This collection isn't just data; it's a series of personal stories from passengers aboard British Airways, covering routes all over the world. Inside, you'll find:

- Passenger Reviews: At its core, this dataset includes detailed reviews from passengers, offering a narrative insight into their travel experiences.
- Quantitative Ratings: Accompanying the narratives are detailed satisfaction metrics, such as ratings on seat comfort, cabin staff service, food and beverages, ground service, value for money, and entertainment.
- Contextual Information: Enhancing the dataset further is a range of contextual information, including details about the author, date, place, aircraft, traveler type, seat type, route, date flown, whether the trip was recommended, and if the trip was verified.

## What We're Looking To Do

My goal to derive insights from the dataset, with a focus on identifying specific routes where targeted service improvements could significantly improve the passenger experience. Here's how I plan to do it:

- Spot Trends: We want to see how customer satisfaction changes over time and across different routes. Are there any patterns in what makes a flight good or bad?
- Dig Into Details: We're going to look closely at how different parts of the service, like the comfort of the seats or the friendliness of the staff, affect how happy travelers are overall.
- Look at Different Travelers: Not everyone has the same needs or expectations. We'll explore how different factors, like the reason for traveling or the time of year, might change how satisfied people are with their flights.

## Questions We Want to Answer

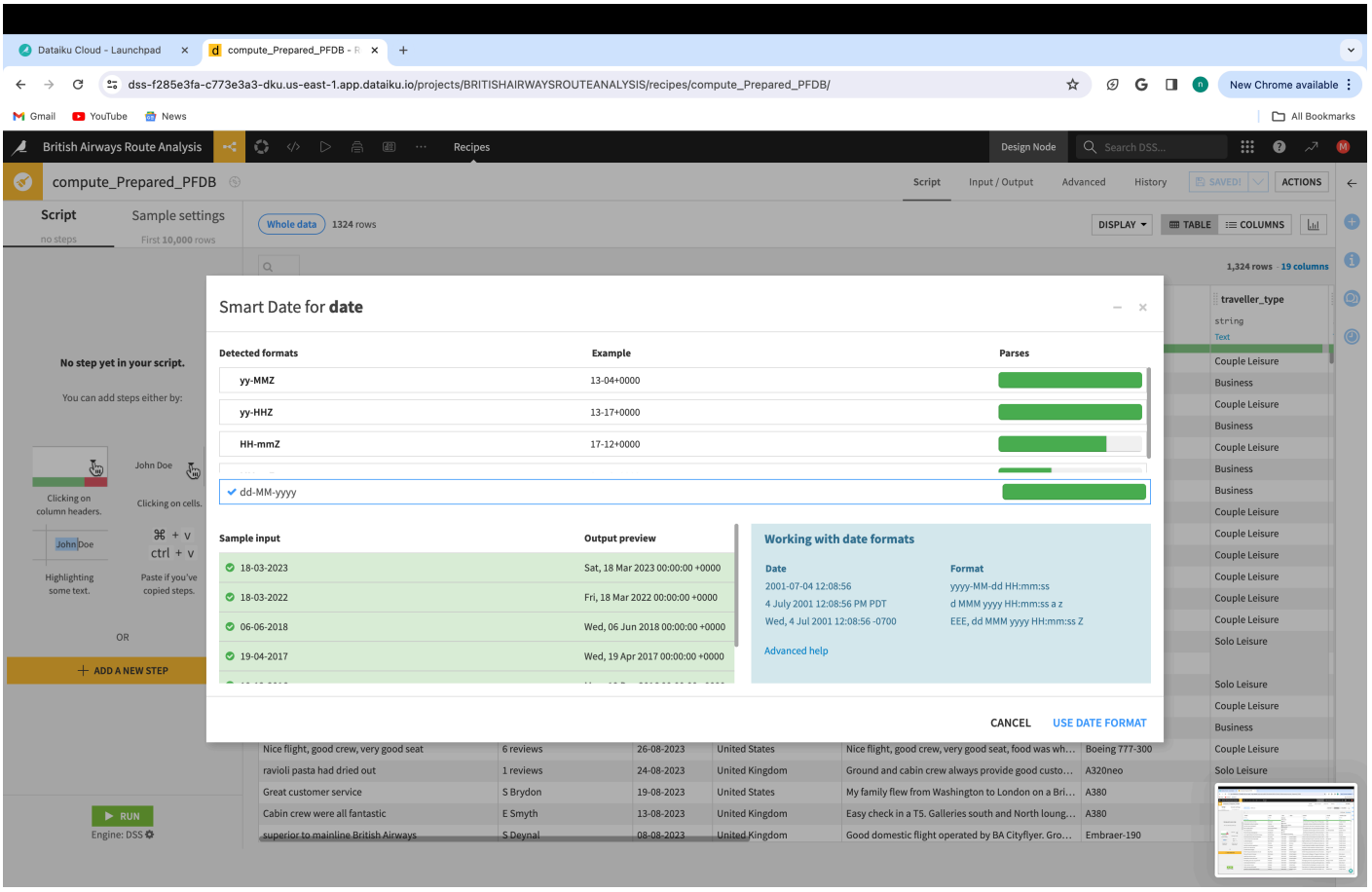
1. How do customer satisfaction and service ratings vary across different routes and over time?
2. Which service aspects are most closely linked to overall satisfaction, and does this vary by route?

# Elevating the Insights: Cleaning and Enriching the British Airways Reviews Dataset

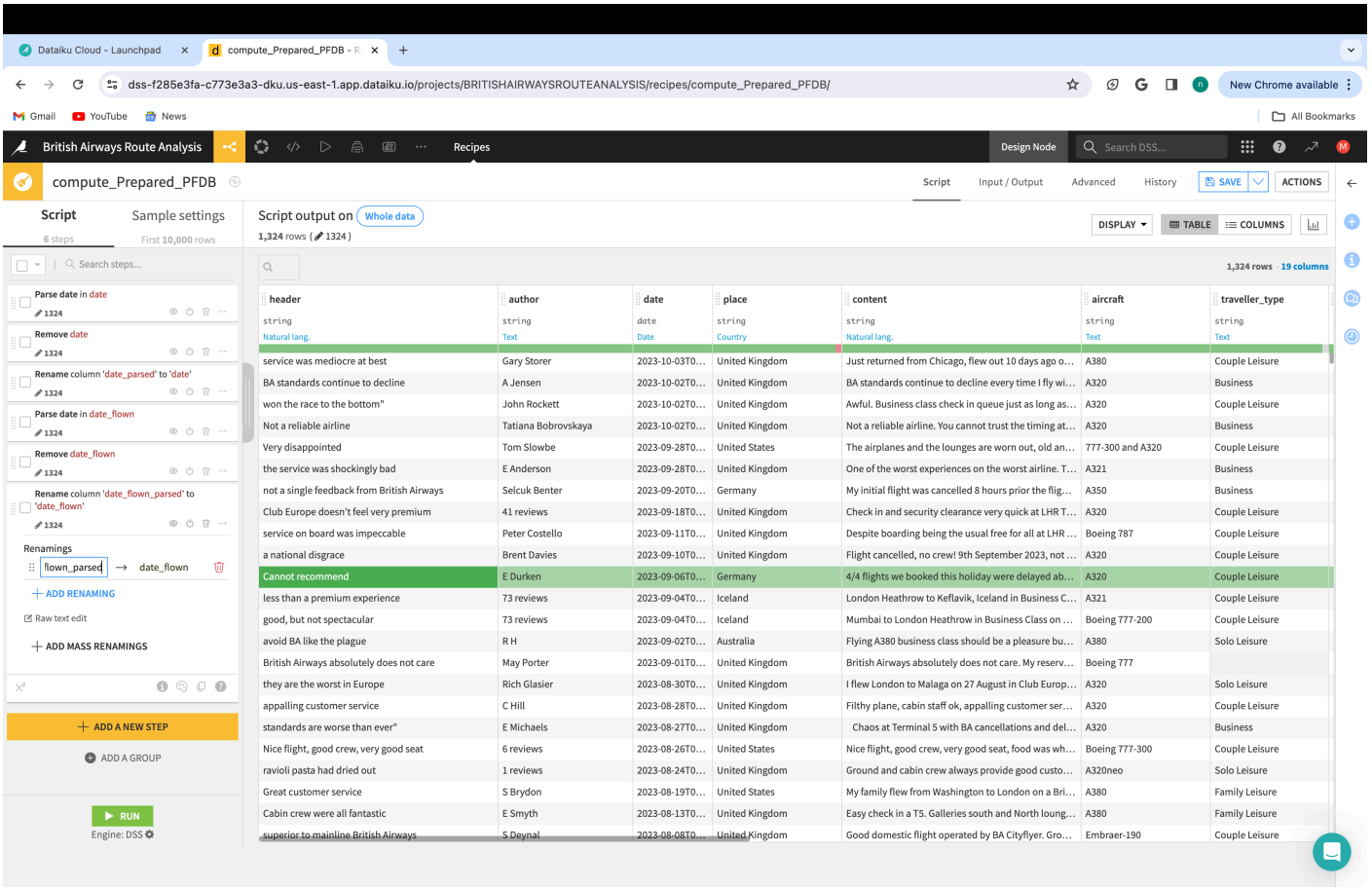
Elevating the Insights: Cleaning and Enriching the British Airways Reviews Dataset

> Note: I've included screenshots of key steps, configurations, and outcomes at appropriate sections within this article to provide visual context to my narrative.

My journey began with uploading the dataset into Dataiku, where I embarked on the initial data preparation. This step involved parsing the date_flown and date columns to standardize the date format, ensuring compatibility for future joins with other datasets. I utilized a Prepare recipe in Dataiku, which allowed me to interactively select and transform these date columns into a consistent format.

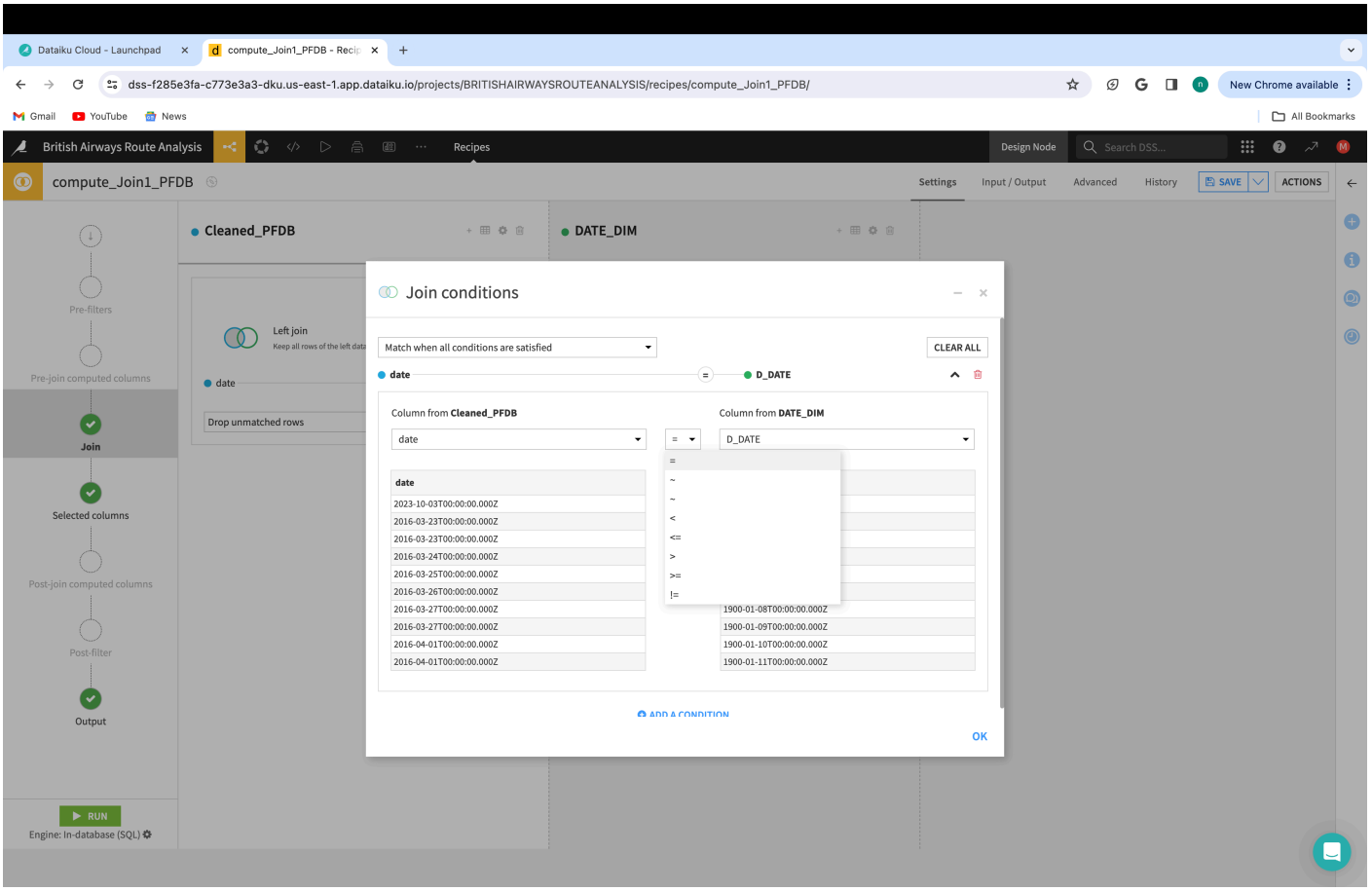



**The result:**

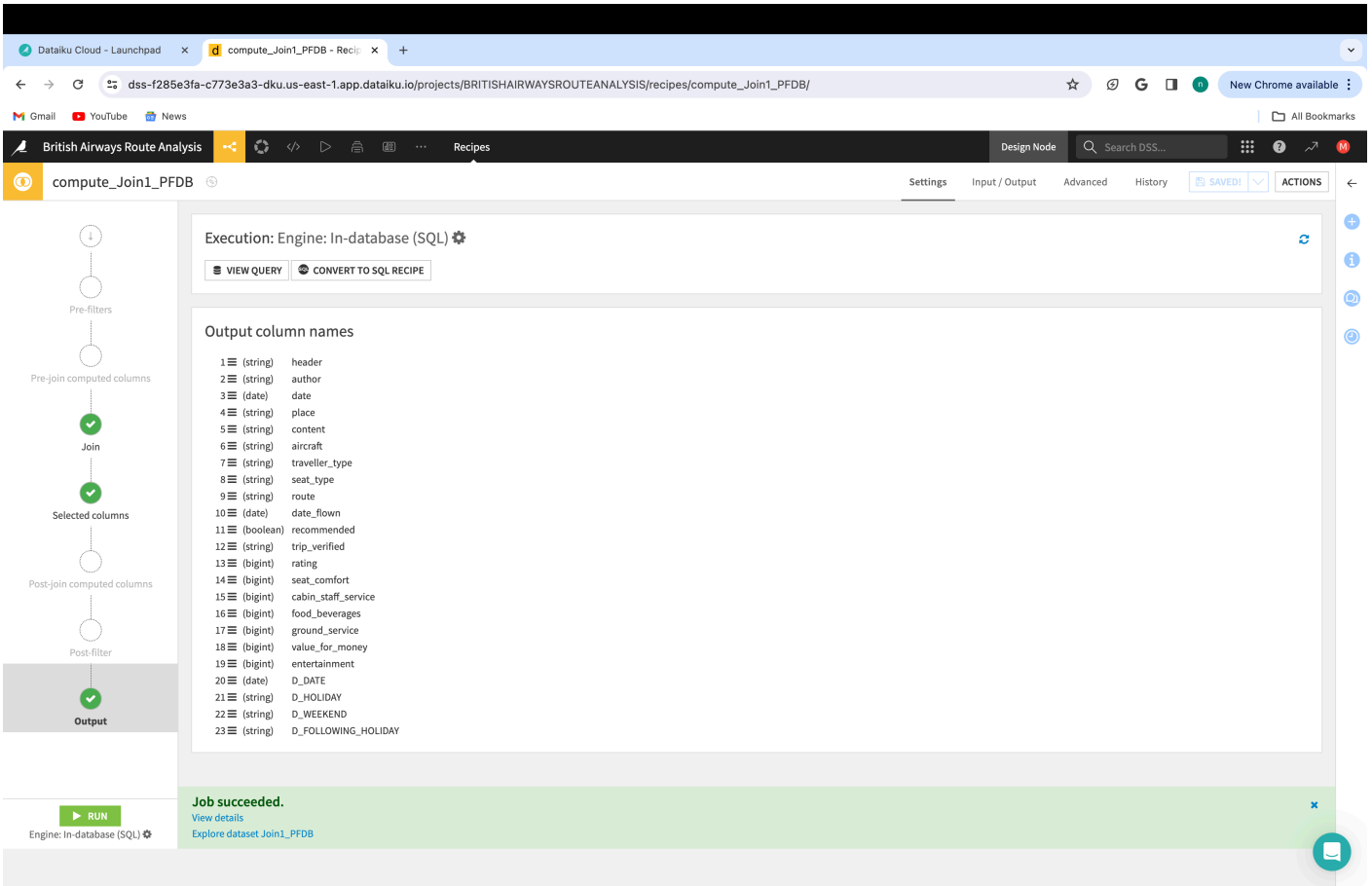



# Joining with DATE_DIM

To add depth and context to my analysis, I enriched my dataset with crucial date-related information through with a left join. Leveraging the Join Recipe within Dataiku, I merged my refined dataset named "Cleaned_PFDB," with the "DATE_DIM" dataset. This join, based on matching 'date' with 'D_DATE', allowed me to incorporate D_HOLIDAY, D_WEEKEND, and D_FOLLOWING_HOLIDAY columns into my dataset.

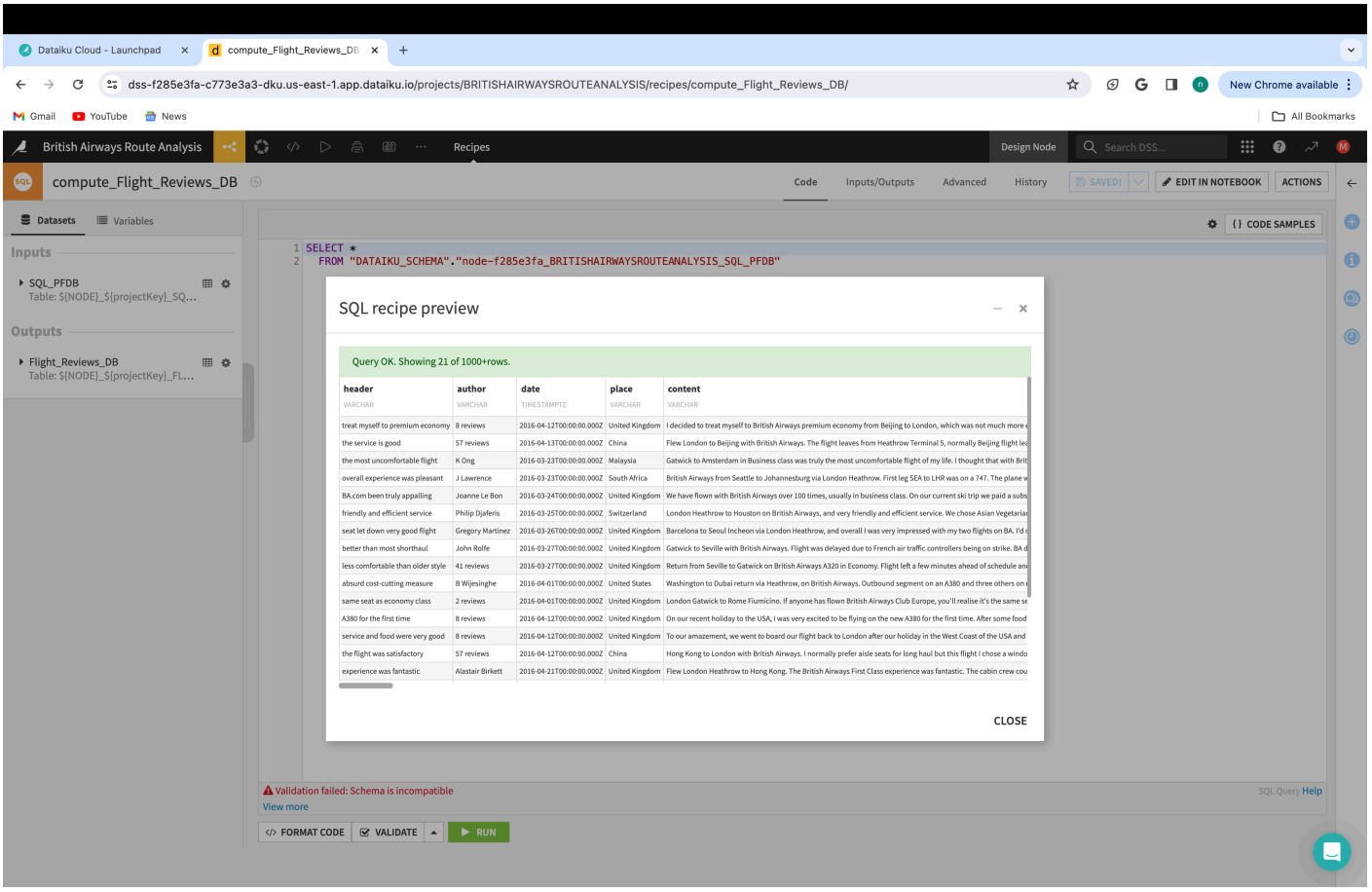



This will be helpful for my later analysis, particularly in understanding the impact of holidays and weekends on British Airways ratings.



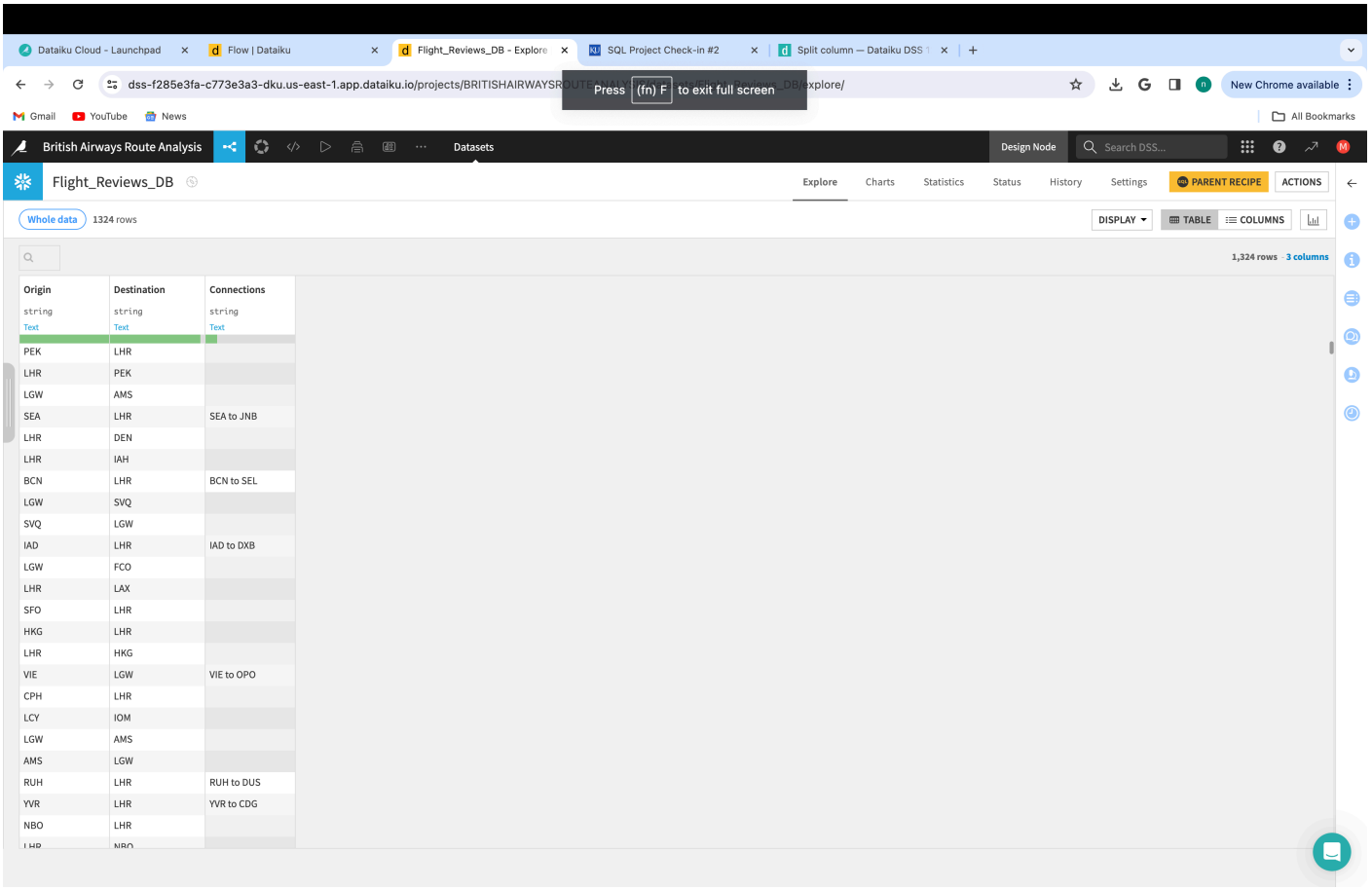

# Synchronization with Snowflake

With my dataset prepared and enriched, I synchronized it with Snowflake using a SQL recipe. This step allowed me to leverage Snowflake's powerful SQL capabilities for further data manipulation and analysis.




# Route Splitting and Data Enrichment

My analysis required dissecting the route information into more granular details, leading me to extract Origin, Destination, and Connections columns. This extraction was performed through a SQL recipe, which split the composite route strings into separate, actionable pieces of information.




**Here's how I did it:**

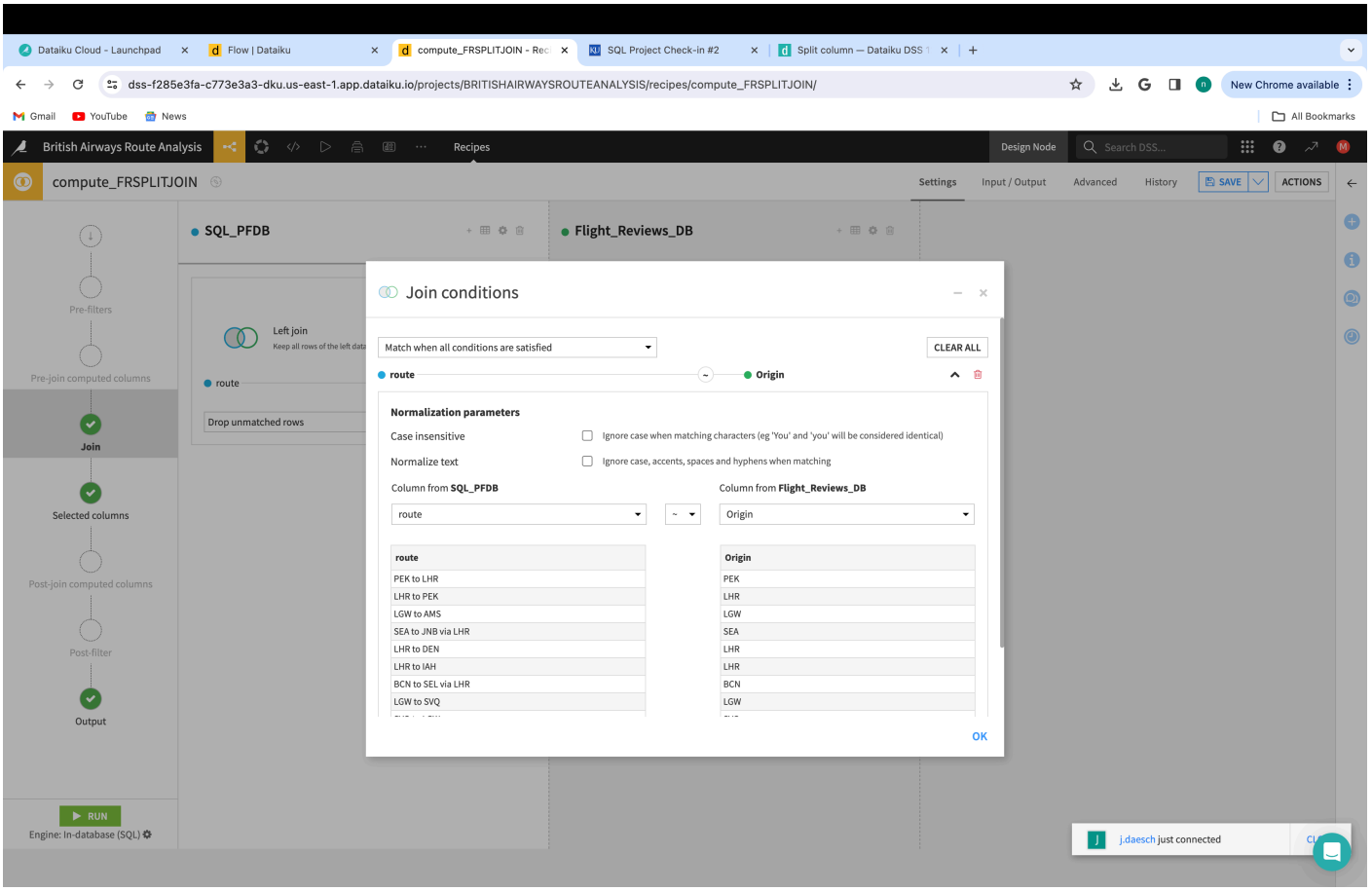
```
SELECT
    SPLIT_PART("route", ' to ', 1) AS "Origin",
    CASE
        WHEN POSITION(' via ' IN "route") > 0 THEN
            SPLIT_PART(SPLIT_PART("route", ' via ', 2), ' to ', 1)
        ELSE
            SPLIT_PART("route", ' to ', 2)
    END AS "Destination",
    CASE
        WHEN POSITION(' via ' IN "route") > 0 THEN
            SPLIT_PART("route", ' via ', 1)
        ELSE
            NULL
    END AS "Connections"
FROM
    "DATAIKU_SCHEMA"."node-f285e3fa_BRITISHAIRWAYSROUTEANALYSIS_SQL_PFDB";
```

This code relies on the use of 'SPLIT_PART' and 'POSITION' functions to navigate through the 'route' strings, identifying the initial departure point as the 'Origin' and the final arrival point as the 'Destination'. For routes that included 'via' in the string, indicating a connection flight, the query extracted this into the 'Connections' column.

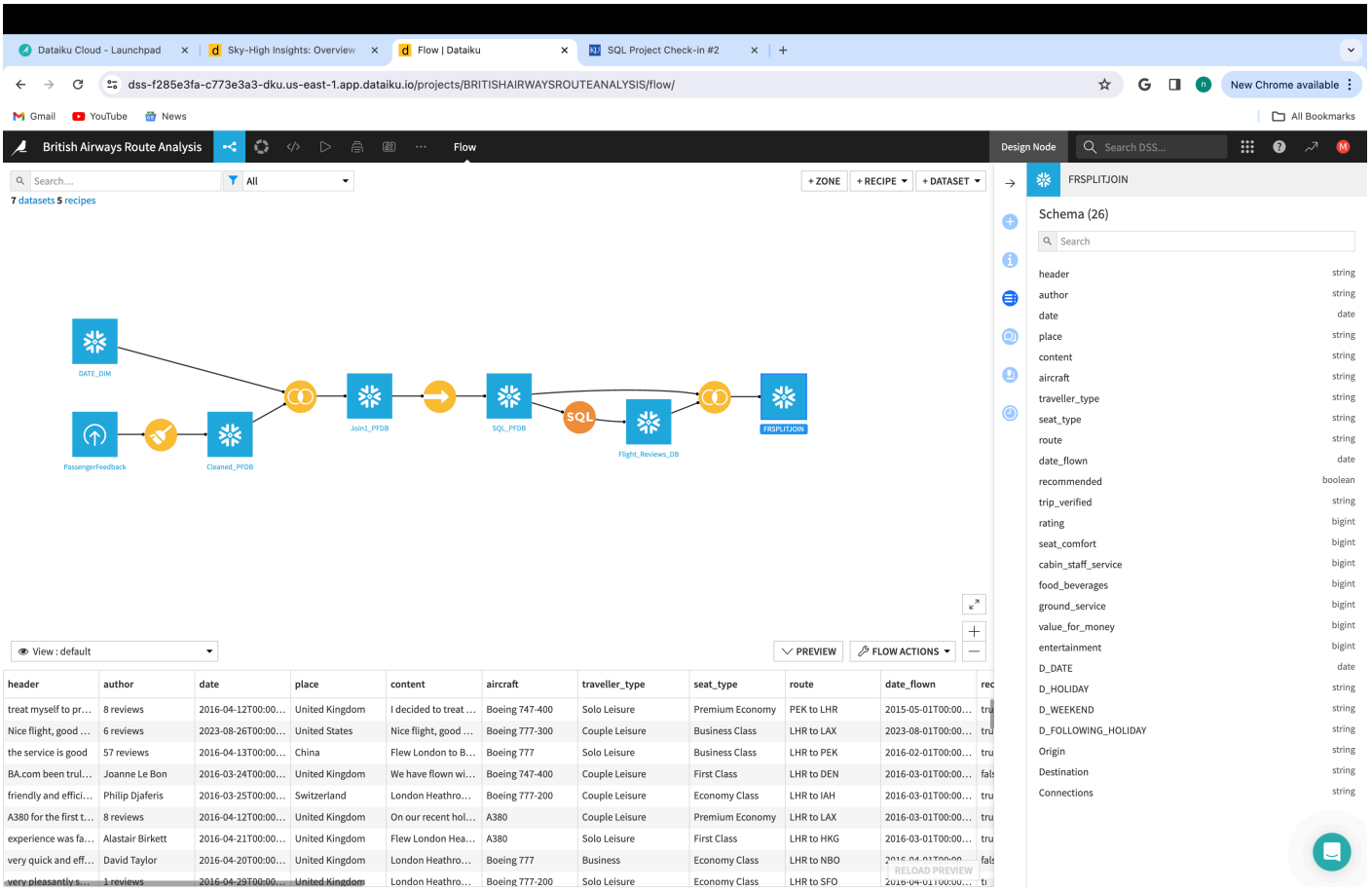
# Addressing Join Challenges

After extracting the Origin, Destination, and Connections columns from the 'route' information through a dedicated SQL recipe, I had to integrate these insights back into the broader, already cleaned and enriched dataset.

To accomplish this, I turned to a left join operation again, which enabled me to merge the detailed route components with the rest of the dataset. I decided to employ the route's starting word (Origin) as the pivot for the join and selected Destination and Connections to be included in the final output.

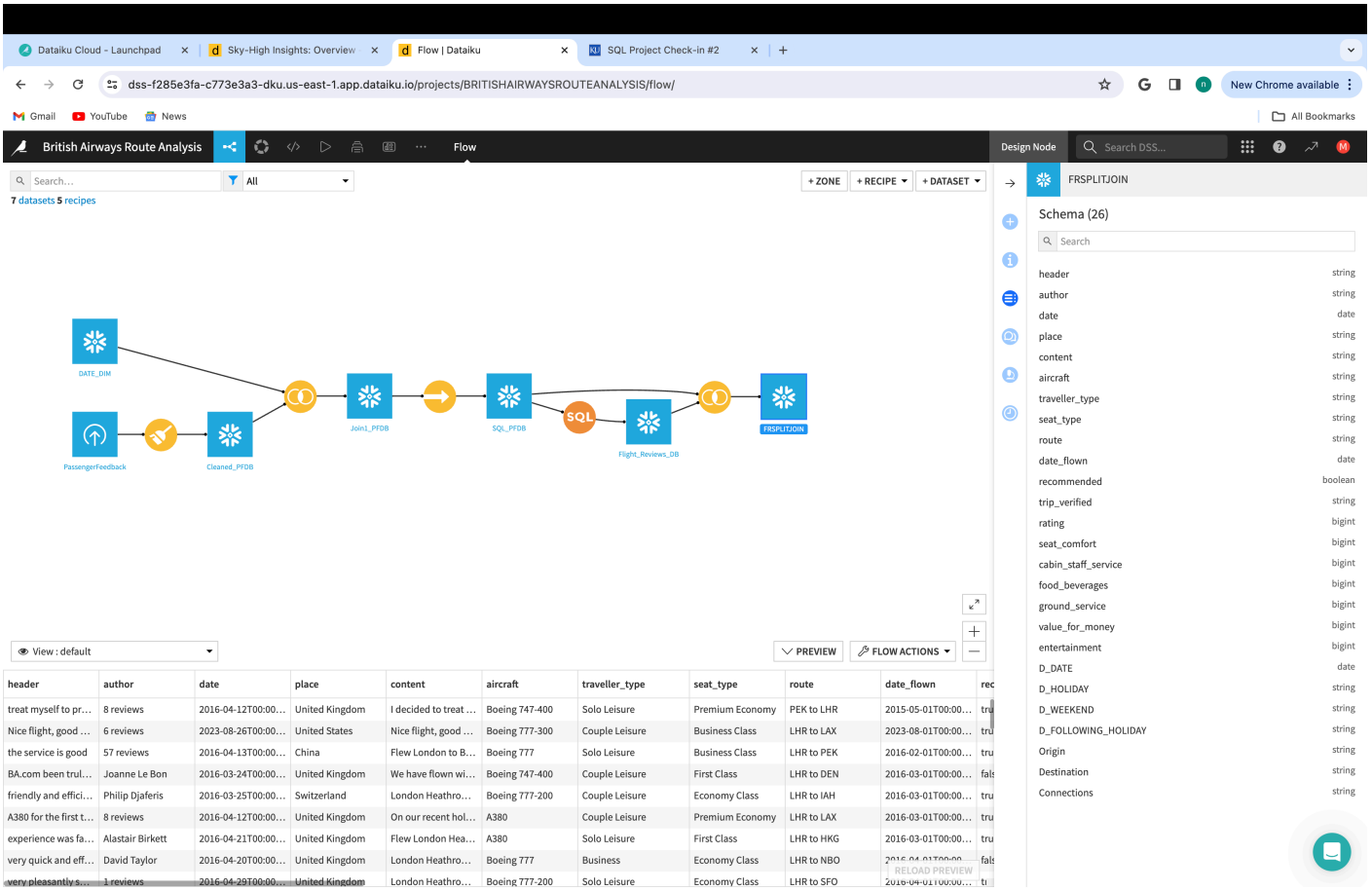



Lastly, I checked columns for potential overlaps and discrepancies, ensuring that the join conditions were precisely defined to avoid data mismatches or exclusions.




| header | author | date | place | content | aircraft | traveller_type | seat_type | route | date_flown |
|---|---|---|---|---|---|---|---|---|---|
| treat myself to pr… | 8 reviews | 2016-04-12T00:00:… | United Kingdom | I decided to treat … | Boeing 747-400 | Solo Leisure | Premium Economy | PEK to LHR | 2015-05-01T00:00:… |
| Nice flight, good … | 6 reviews | 2023-08-26T00:00:… | United States | Nice flight, good … | Boeing 777-300 | Couple Leisure | Business Class | LHR to LAX | 2023-08-01T00:00:… |
| the service is good | 57 reviews | 2016-04-13T00:00:… | China | Flew London to B… | Boeing 777 | Solo Leisure | Business Class | LHR to PEK | 2016-02-01T00:00:… |
| BA.com been trul… | Joanne Le Bon | 2016-03-24T00:00:… | United Kingdom | We have flown wi… | Boeing 747-400 | Couple Leisure | First Class | LHR to DEN | 2016-03-01T00:00:… |
| friendly and effici… | Philip Djaferis | 2016-03-25T00:00:… | Switzerland | London Heathro… | Boeing 777-200 | Couple Leisure | Economy Class | LHR to IAH | 2016-03-01T00:00:… |
| A380 for the first t… | 8 reviews | 2016-04-12T00:00:… | United Kingdom | On our recent hol… | A380 | Couple Leisure | Premium Economy | LHR to LAX | 2016-03-01T00:00:… |
| experience was fa… | Alastair Birkett | 2016-04-21T00:00:… | United Kingdom | Flew London Hea… | A380 | Solo Leisure | First Class | LHR to HKG | 2016-03-01T00:00:… |
| very quick and eff… | David Taylor | 2016-04-20T00:00:… | United Kingdom | London Heathro… | Boeing 777 | Business | Economy Class | LHR to NBO | 2016-04-01T00:00:… |
| very pleasantly s… | 1 reviews | 2016-04-29T00:00:… | United Kingdom | London Heathro… | Boeing 777-200 | Solo Leisure | Economy Class | LHR to SFO | 2016-04-01T00:00:… |

# Some Last Minute Changes

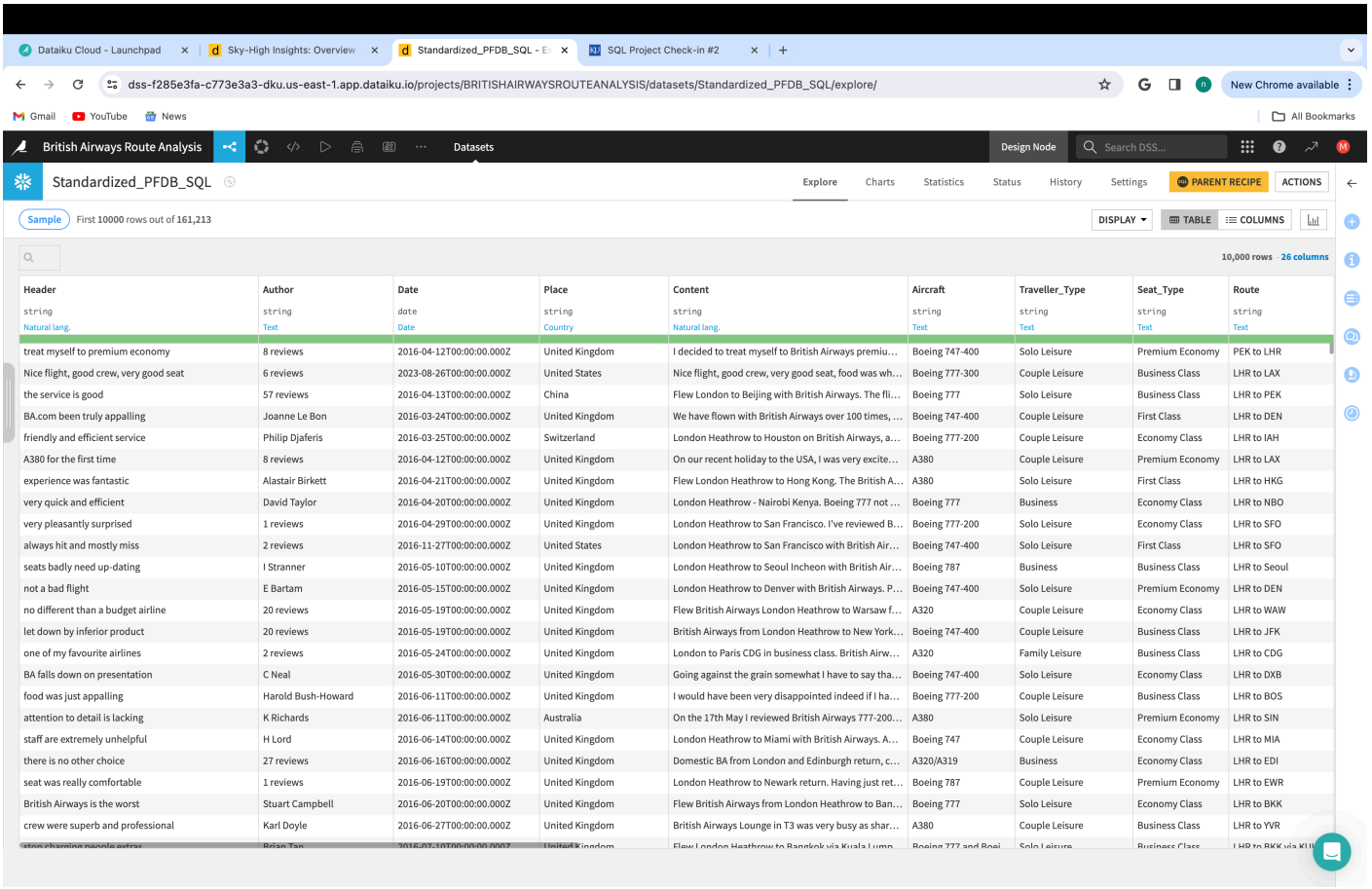



| header | author | date | place | content | aircraft | traveller_type | seat_type | route | date_flown |
|---|---|---|---|---|---|---|---|---|---|
| treat myself to pr… | 8 reviews | 2016-04-12T00:00:… | United Kingdom | I decided to treat … | Boeing 747-400 | Solo Leisure | Premium Economy | PEK to LHR | 2015-05-01T00:00:… |
| Nice flight, good … | 6 reviews | 2023-08-26T00:00:… | United States | Nice flight, good … | Boeing 777-300 | Couple Leisure | Business Class | LHR to LAX | 2023-08-01T00:00:… |
| the service is good | 57 reviews | 2016-04-13T00:00:… | China | Flew London to B… | Boeing 777 | Solo Leisure | Business Class | LHR to PEK | 2016-02-01T00:00:… |
| BA.com been trul… | Joanne Le Bon | 2016-03-24T00:00:… | United Kingdom | We have flown wi… | Boeing 747-400 | Couple Leisure | First Class | LHR to DEN | 2016-03-01T00:00:… |
| friendly and effici… | Philip Djaferis | 2016-03-25T00:00:… | Switzerland | London Heathro… | Boeing 777-200 | Couple Leisure | Economy Class | LHR to IAH | 2016-03-01T00:00:… |
| A380 for the first t… | 8 reviews | 2016-04-12T00:00:… | United Kingdom | On our recent hol… | A380 | Couple Leisure | Premium Economy | LHR to LAX | 2016-03-01T00:00:… |
| experience was fa… | Alastair Birkett | 2016-04-21T00:00:… | United Kingdom | Flew London Hea… | A380 | Solo Leisure | First Class | LHR to HKG | 2016-03-01T00:00:… |
| very quick and eff… | David Taylor | 2016-04-20T00:00:… | United Kingdom | London Heathro… | Boeing 777 | Business | Economy Class | LHR to NBO | 2016-04-01T00:00:… |
| very pleasantly s… | 1 reviews | 2016-04-29T00:00:… | United Kingdom | London Heathro… | Boeing 777-200 | Solo Leisure | Economy Class | LHR to SFO | 2016-04-01T00:00:… |

Upon reviewing the final dataset "FRSPLITJOIN," I noticed inconsistencies in the column names, where some names were appropriately standardized while others were not. To address this issue, I reran the

SQL recipe and applied the following code:

```
SELECT
     "header" AS "Header",
     "author" AS "Author",
     "date" AS "Date",
     "place" AS "Place",
     "content" AS "Content",
     "aircraft" AS "Aircraft",
     "traveller_type" AS "Traveller_Type",
     "seat_type" AS "Seat_Type",
     "route" AS "Route",
     "date_flown" AS "Date_Flown",
     "recommended" AS "Recommended",
     "trip_verified" AS "Trip_Verified",
     "rating" AS "Rating",
     "seat_comfort" AS "Seat_Comfort",
     "cabin_staff_service" AS "Cabin_Staff_Service",
     "food_beverages" AS "Food_Beverages",
     "ground_service" AS "Ground_Service",
     "value_for_money" AS "Value_For_Money",
     "entertainment" AS "Entertainment",
     "D_DATE" AS "D_Date",
     "D_HOLIDAY" AS "D_Holiday",
     "D_WEEKEND" AS "D_Weekend",
     "D_FOLLOWING_HOLIDAY" AS "D_Following_Holiday",
     "Origin" AS "Origin",
     "Destination" AS "Destination",
     "Connections" AS "Connections"
 FROM
     "DATAIKU_SCHEMA"."node-f285e3fa_BRITISHAIRWAYSROUTEANALYSIS_FRSPLITJOIN";
```




| Header | Author | Date | Place | Content | Aircraft | Traveller_Type | Seat_Type | Route |
|---|---|---|---|---|---|---|---|---|
| string<br>Natural lang. | string<br>Text | date<br>Date | string<br>Country | string<br>Natural lang. | string<br>Text | string<br>Text | string<br>Text | string<br>Text |
| treat myself to premium economy | 8 reviews | 2016-04-12T00:00:00.000Z | United Kingdom | I decided to treat myself to British Airways premiu… | Boeing 747-400 | Solo Leisure | Premium Economy | PEK to LHR |
| Nice flight, good crew, very good seat | 6 reviews | 2023-08-26T00:00:00.000Z | United States | Nice flight, good crew, very good seat, food was wh… | Boeing 777-300 | Couple Leisure | Business Class | LHR to LAX |
| the service is good | 57 reviews | 2016-04-13T00:00:00.000Z | China | Flew London to Beijing with British Airways. The fli… | Boeing 777 | Solo Leisure | Business Class | LHR to PEK |
| BA.com been truly appalling | Joanne Le Bon | 2016-03-24T00:00:00.000Z | United Kingdom | We have flown with British Airways over 100 times, … | Boeing 747-400 | Couple Leisure | First Class | LHR to DEN |
| friendly and efficient service | Philip Djaferis | 2016-03-25T00:00:00.000Z | Switzerland | London Heathrow to Houston on British Airways, a… | Boeing 777-200 | Couple Leisure | Economy Class | LHR to IAH |
| A380 for the first time | 8 reviews | 2016-04-12T00:00:00.000Z | United Kingdom | On our recent holiday to the USA, I was very excite… | A380 | Couple Leisure | Premium Economy | LHR to LAX |
| experience was fantastic | Alastair Birkett | 2016-04-21T00:00:00.000Z | United Kingdom | Flew London Heathrow to Hong Kong. The British A… | A380 | Solo Leisure | First Class | LHR to HKG |
| very quick and efficient | David Taylor | 2016-04-20T00:00:00.000Z | United Kingdom | London Heathrow - Nairobi Kenya. Boeing 777 not … | Boeing 777 | Business | Economy Class | LHR to NBO |
| very pleasantly surprised | 1 reviews | 2016-04-29T00:00:00.000Z | United Kingdom | London Heathrow to San Francisco. I've reviewed B… | Boeing 777-200 | Solo Leisure | Economy Class | LHR to SFO |
| always hit and mostly miss | 2 reviews | 2016-11-27T00:00:00.000Z | United States | London Heathrow to San Francisco with British Air… | Boeing 747-400 | Solo Leisure | First Class | LHR to SFO |
| seats badly need up-dating | I Stranner | 2016-05-10T00:00:00.000Z | United Kingdom | London Heathrow to Seoul Incheon with British Air… | Boeing 787 | Business | Business Class | LHR to Seoul |
| not a bad flight | E Bartam | 2016-05-15T00:00:00.000Z | United Kingdom | London Heathrow to Denver with British Airways. P… | Boeing 747-400 | Solo Leisure | Premium Economy | LHR to DEN |
| no different than a budget airline | 20 reviews | 2016-05-19T00:00:00.000Z | United Kingdom | Flew British Airways London Heathrow to Warsaw f… | A320 | Couple Leisure | Economy Class | LHR to WAW |
| let down by inferior product | 20 reviews | 2016-05-19T00:00:00.000Z | United Kingdom | British Airways from London Heathrow to New York… | Boeing 747-400 | Couple Leisure | Business Class | LHR to JFK |
| one of my favourite airlines | 2 reviews | 2016-05-24T00:00:00.000Z | United Kingdom | London to Paris CDG in business class. British Airw… | A320 | Family Leisure | Business Class | LHR to CDG |
| BA falls down on presentation | C Neal | 2016-05-30T00:00:00.000Z | United Kingdom | Going against the grain somewhat I have to say tha… | Boeing 747-400 | Solo Leisure | Economy Class | LHR to DXB |
| food was just appalling | Harold Bush-Howard | 2016-06-11T00:00:00.000Z | United Kingdom | I would have been very disappointed indeed if I ha… | Boeing 777-200 | Couple Leisure | Business Class | LHR to BOS |
| attention to detail is lacking | K Richards | 2016-06-11T00:00:00.000Z | Australia | On the 17th May I reviewed British Airways 777-200… | A380 | Solo Leisure | Premium Economy | LHR to SIN |
| staff are extremely unhelpful | H Lord | 2016-06-14T00:00:00.000Z | United Kingdom | London Heathrow to Miami with British Airways. A… | Boeing 747 | Couple Leisure | Economy Class | LHR to MIA |
| there is no other choice | 27 reviews | 2016-06-16T00:00:00.000Z | United Kingdom | Domestic BA from London and Edinburgh return, c… | A320/A319 | Business | Economy Class | LHR to EDI |
| seat was really comfortable | 1 reviews | 2016-06-19T00:00:00.000Z | United Kingdom | London Heathrow to Newark return. Having just ret… | Boeing 787 | Couple Leisure | Premium Economy | LHR to EWR |
| British Airways is the worst | Stuart Campbell | 2016-06-20T00:00:00.000Z | United Kingdom | Flew British Airways from London Heathrow to Ban… | Boeing 777 | Solo Leisure | Economy Class | LHR to BKK |
| crew were superb and professional | Karl Doyle | 2016-06-27T00:00:00.000Z | United Kingdom | British Airways Lounge in T3 was very busy as shar… | A380 | Couple Leisure | Business Class | LHR to YVR |

This SQL query standardized all column names to ensure consistency and clarity across the dataset. This modification will help later in my analysis as it mitigates potential confusion regarding capitalization discrepancies, minimizing the risk of encountering syntax issues.

# Reflections

Now that I have dedicated "Origin", "Destination" and "Connection" columns, as well as the additional contextual date information from the 'DATE_DIM' dataset, I'm quite happy with the final dataset. I can now effortlessly categorize 'route' information as needed, this enhancement will allow me to analyze how various variables evolve across different routes. Moreover, the inclusion of the date information allows me to establish correlations between satisfaction ratings or airline performance specifically during holidays compared to regular days.

**Here's the data dictionary of my final dataset:**

# Data Dictionary: British Airways Passenger Reviews

| Field Name | Data Type | Data Format | Field Size | Description | Example |
|---|---|---|---|---|---|
| Header | VARCHAR | XNNNN | 100 | Summary of the customer review | "service was mediocre at best" |
| Author | VARCHAR | XNNNN | 50 | Name of the review author | G Storer |
| Date | DATE | DD-mm-yyyy | 20 | Date when the review was written | 2016-03-25 |
| Place | VARCHAR | XNNNN | 20 | Location from where the review is posted | United Kingdom |
| Content | VARCHAR | XNNNN | 300 | Detailed text content of the passenger's review | "Just returned from Chicago, flew out 10 days ago on American Airlines absolutely superb..." |
| Aircraft | VARCHAR | XNNNN | 10 | Model of the aircraft on which reviewer traveled | A381 |
| Traveller_Type | VARCHAR | XNNNN | 20 | Type of traveler as specified by the reviewer | Couple Leisure |
| Seat_Type | VARCHAR | XNNNN | 20 | Class of service used by the reviewer | Economy Class |
| Route | VARCHAR | XNNNN | 50 | The flight route taken by the reviewer | Chicago to Manchester via Heathrow |
| Date_Flown | DATE | DD-mm-yyyy | 20 | Date of the flight | 2016-04-21 |
| Recommended | BOOLEAN | xyz | 5 | Whether the reviewer would recommend the airline | Yes |
| Trip_Verified | VARCHAR | XNNNN | 10 | Indicates if the trip has been verified | Not Verified |
| Rating | NUMBER | # | 5 | Overall rating given by the reviewer | 8 |
| Seat_Comfort | NUMBER | # | 5 | Rating for the comfort of the seating | 5 |
| Cabin_Staff_Service | NUMBER | # | 5 | Rating for the service provided by the cabin staff | 2 |
| Food_Beverages | NUMBER | # | 5 | Rating for the quality of food and beverages | 3 |
| Ground_Service | NUMBER | # | 5 | Rating for the services provided on the ground | 5 |
| Value_For_Money | NUMBER | # | 5 | Rating reflecting the passenger's perception of the value for money | 4 |
| Entertainment | NUMBER | ## | 5 | Rating for the in-flight entertainment options | 4 |
| D_Date | DATE | DD-mm-yyyy | 20 | Date for verification | 2016-04-12 |
| D_Holiday | BOOLEAN | x | 2 | Whether the reviewer travelled during a Holiday | Y |
| D_Weekend | BOOLEAN | x | 2 | Whether the reviewer travelled during the weekend | N |
| D_Following_Holiday | BOOLEAN | x | 2 | Whether the reviewer travelled after a Holiday | N |
| Origin | VARCHAR | XNNNN | 20 | Origin of the flight | PEK |
| Destination | VARCHAR | XNNNN | 20 | Destination of the flight | LHR |
| Connections | VARCHAR | XNNNN | 20 | Flight connections during Trip | London via Singapore |

# Exploring the Insights: Analyzing Customer Satisfaction and Service Ratings

## Exploring the Insights: Analyzing Customer Satisfaction and Service Ratings

In the third phase of my project, I focused on analyzing my dataset to uncover insights related to customer satisfaction and service ratings across different routes and over time. Utilizing the "Standardized_PFDB_SQL" dataset within Dataiku, I aimed to answer two questions;

1. How do customer satisfaction and service ratings vary across different routes and over time?
2. Which service aspects are most closely linked to overall satisfaction, and does this variation depend on the route?

## Addressing Question 1

To analyze how customer satisfaction and service ratings vary across different routes and over time, I wanted to focus on extracting average ratings, recommendation percentages, and considering the impact of holidays, weekends, and different seasons. The fields included in our analysis were:

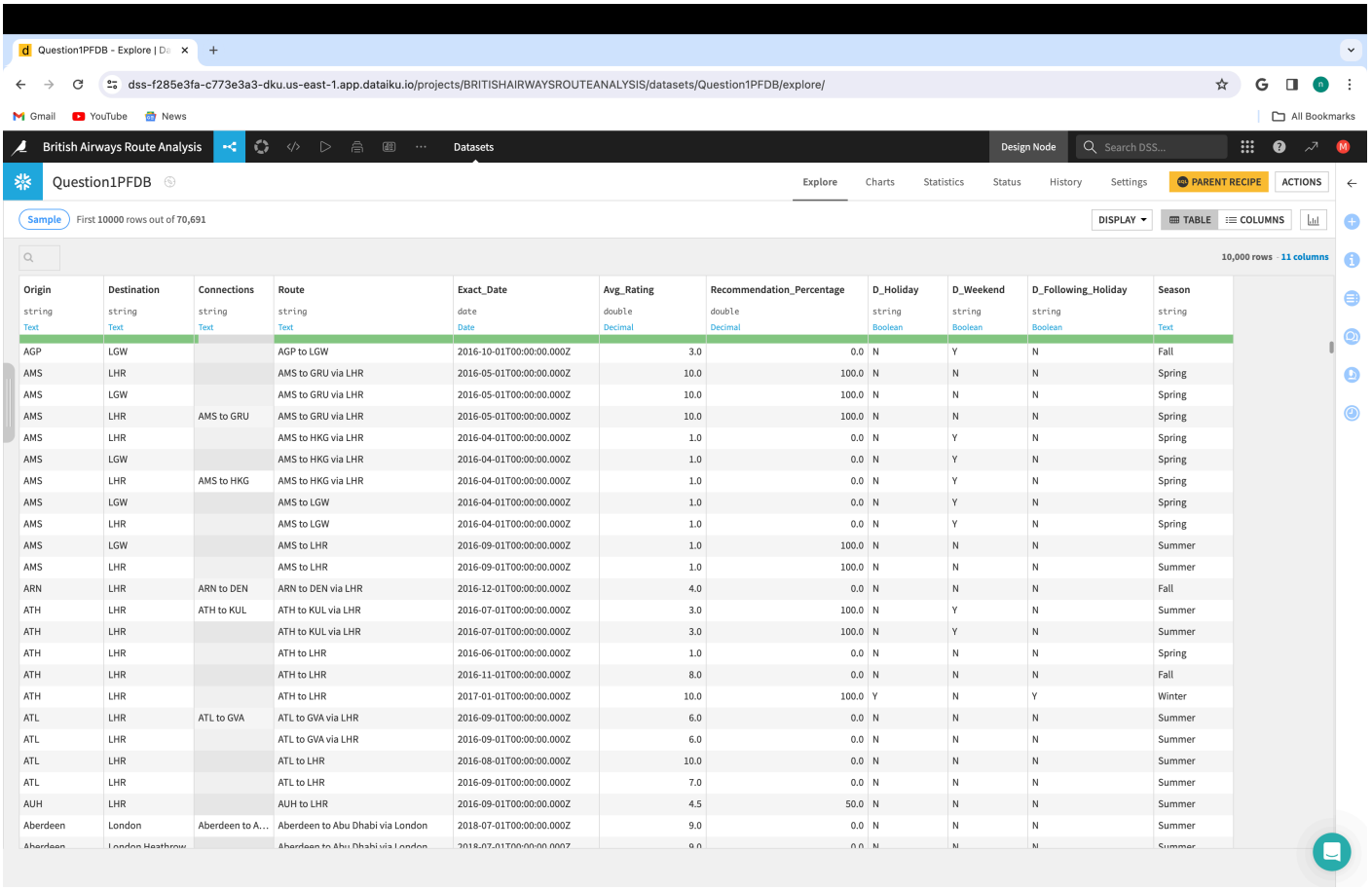
1. Average Rating (AVG("Rating"))
2. Recommendation Percentage (COUNT(CASE WHEN "Recommended" = 'Yes' THEN 1 END) 100.0 / COUNT())
3. Date indicators ("D_Holiday", "D_Weekend", "D_Following_Holiday")
4. Season determination using a CASE statement to categorize the date flown into Spring (March-May), Summer (June-August), Fall (September-November), and Winter (December-February).

   **Here's the query:**

```
SELECT
"Route",
"Date_Flown" AS "Exact_Date",
AVG("Rating") AS "Avg_Rating",
COUNT(CASE WHEN "Recommended" = 'Yes' THEN 1 END) * 100.0 / COUNT(*) AS
"Recommendation_Percentage",
"D_Holiday",
"D_Weekend",
"D_Following_Holiday",
CASE
WHEN MONTH("Date_Flown") IN (3, 4, 5) THEN 'Spring'
```

```
WHEN MONTH("Date_Flown") IN (6, 7, 8) THEN 'Summer'
WHEN MONTH("Date_Flown") IN (9, 10, 11) THEN 'Fall'
WHEN MONTH("Date_Flown") IN (12, 1, 2) THEN 'Winter'
ELSE 'Unknown'
END AS "Season"
FROM "DATAIKU_SCHEMA"."node-f285e3fa_BRITISHAIRWAYSROUTEANALYSIS_STANDARDIZED_PFDB_SQL"
GROUP BY "Route", "Exact_Date", "D_Holiday", "D_Weekend", "D_Following_Holiday", "Season"
ORDER BY "Route", "Exact_Date";
```

**Here's the output:**




| Origin | Destination | Connections | Route | Exact_Date | Avg_Rating | Recommendation_Percentage | D_Holiday | D_Weekend | D_Following_Holiday | Season |
|---|---|---|---|---|---|---|---|---|---|---|
| string | string | string | string | date | double | double | string | string | string | string |
| Text | Text | Text | Text | Date | Decimal | Decimal | Boolean | Boolean | Boolean | Text |
| AGP | LGW | | AGP to LGW | 2016-10-01T00:00:00.000Z | 3.0 | 0.0 | N | Y | N | Fall |
| AMS | LHR | | AMS to GRU via LHR | 2016-05-01T00:00:00.000Z | 10.0 | 100.0 | N | N | N | Spring |
| AMS | LGW | | AMS to GRU via LHR | 2016-05-01T00:00:00.000Z | 10.0 | 100.0 | N | N | N | Spring |
| AMS | LHR | AMS to GRU | AMS to GRU via LHR | 2016-05-01T00:00:00.000Z | 10.0 | 100.0 | N | N | N | Spring |
| AMS | LHR | | AMS to HKG via LHR | 2016-04-01T00:00:00.000Z | 1.0 | 0.0 | N | Y | N | Spring |
| AMS | LGW | | AMS to HKG via LHR | 2016-04-01T00:00:00.000Z | 1.0 | 0.0 | N | Y | N | Spring |
| AMS | LHR | AMS to HKG | AMS to HKG via LHR | 2016-04-01T00:00:00.000Z | 1.0 | 0.0 | N | Y | N | Spring |
| AMS | LGW | | AMS to LGW | 2016-04-01T00:00:00.000Z | 1.0 | 0.0 | N | Y | N | Spring |
| AMS | LHR | | AMS to LGW | 2016-04-01T00:00:00.000Z | 1.0 | 0.0 | N | Y | N | Spring |
| AMS | LGW | | AMS to LHR | 2016-09-01T00:00:00.000Z | 1.0 | 100.0 | N | N | N | Summer |
| AMS | LHR | | AMS to LHR | 2016-09-01T00:00:00.000Z | 1.0 | 100.0 | N | N | N | Summer |
| ARN | LHR | ARN to DEN | ARN to DEN via LHR | 2016-12-01T00:00:00.000Z | 4.0 | 0.0 | N | N | N | Fall |
| ATH | LHR | ATH to KUL | ATH to KUL via LHR | 2016-07-01T00:00:00.000Z | 3.0 | 100.0 | N | Y | N | Summer |
| ATH | LHR | | ATH to KUL via LHR | 2016-07-01T00:00:00.000Z | 3.0 | 100.0 | N | Y | N | Summer |
| ATH | LHR | | ATH to LHR | 2016-06-01T00:00:00.000Z | 1.0 | 0.0 | N | N | N | Spring |
| ATH | LHR | | ATH to LHR | 2016-11-01T00:00:00.000Z | 8.0 | 0.0 | N | N | N | Fall |
| ATH | LHR | | ATH to LHR | 2017-01-01T00:00:00.000Z | 10.0 | 100.0 | Y | N | Y | Winter |
| ATL | LHR | ATL to GVA | ATL to GVA via LHR | 2016-09-01T00:00:00.000Z | 6.0 | 0.0 | N | N | N | Summer |
| ATL | LHR | | ATL to GVA via LHR | 2016-09-01T00:00:00.000Z | 6.0 | 0.0 | N | N | N | Summer |
| ATL | LHR | | ATL to LHR | 2016-08-01T00:00:00.000Z | 10.0 | 0.0 | N | N | N | Summer |
| ATL | LHR | | ATL to LHR | 2016-09-01T00:00:00.000Z | 7.0 | 0.0 | N | N | N | Summer |
| AUH | LHR | | AUH to LHR | 2016-09-01T00:00:00.000Z | 4.5 | 50.0 | N | N | N | Summer |
| Aberdeen | London | Aberdeen to A... | Aberdeen to Abu Dhabi via London | 2018-07-01T00:00:00.000Z | 9.0 | 0.0 | N | N | N | Summer |

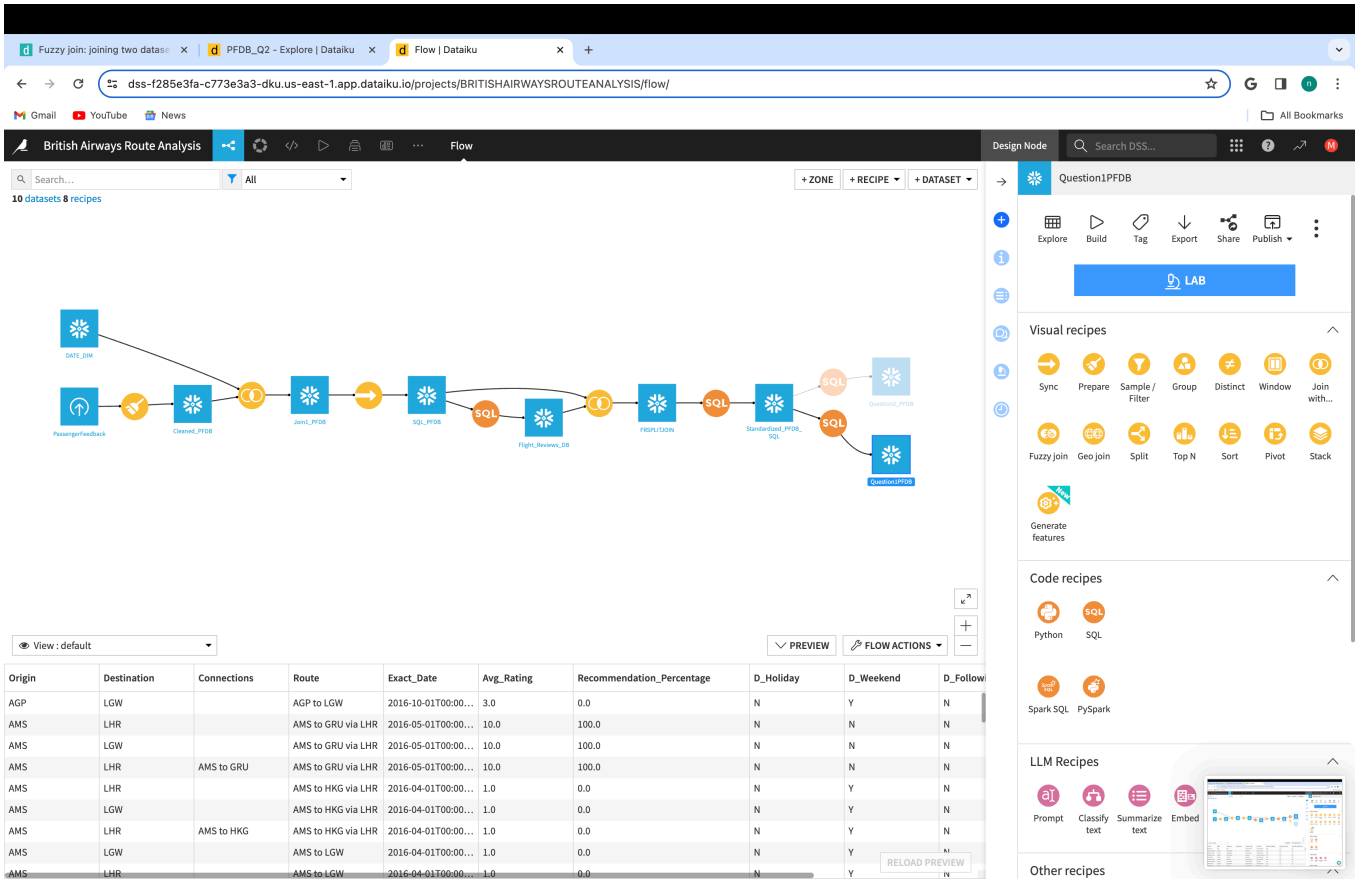
# Findings and Analysis

Upon examining the query output, I observed customer satisfaction and service ratings that did indeed vary by route and date.

- Seasonal Trends: The data reveals a mixed pattern in terms of seasons affecting satisfaction levels, contrary to my initial hypothesis. For instance, the routes AMS to LGW and ATL to LHR showed varying average ratings across different seasons without a clear trend towards higher ratings in any specific season. This suggests that seasonal factors are not the sole determinant of customer satisfaction.

- Holiday and Weekend Influence: My initial analysis suggested a slight increase in satisfaction during holidays and weekends, which is confirmed by specific entries such as the route from AGP to LGW on a holiday showing a lower average rating, which indicates how truly complex customer satisfaction dynamics are during these periods.

- Route-Specific Variations: The data shows significant fluctuations in satisfaction across different routes. For example, the route from AMS to GRU shows consistently high satisfaction ratings regardless of the season, which could indicate a good service on this route. Other routes show a broader range of average ratings, suggesting inconsistency in service quality or varying passenger expectations.
    - Recommendation Percentages: The recommendation percentage appears to be a critical indicator of overall satisfaction. Routes with high recommendation percentages align with higher average ratings, suggesting that customers are more likely to endorse the airline's service when their expectations are met or exceeded.




# Addressing Question 2

> Note: The findings presented are derived from a selection of general samples within the dataset. They should be regarded as preliminary insights rather than conclusions.

To answer the question regarding which service aspects are most closely linked to overall satisfaction, and whether this varies by route, I decided to use an aggregate analysis approach. This involved calculating average ratings for each service aspect along with the overall satisfaction, followed by examining how these averages varied by route.

The specific service aspects I considered for this analysis were:

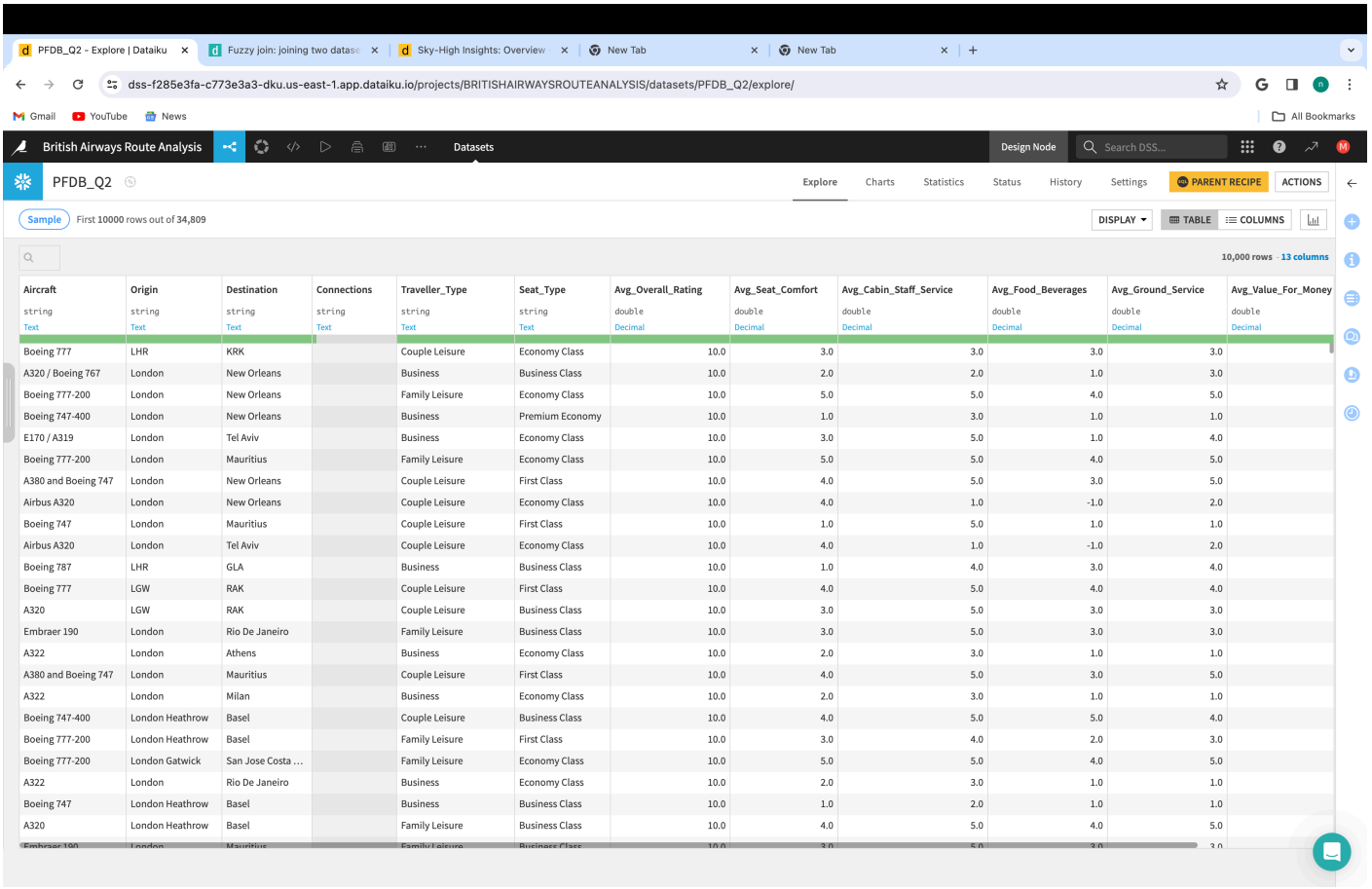
1. Seat Comfort

2. Cabin Staff Service
3. Food and Beverages
4. Ground Service
5. Value for Money
6. Entertainment

**Here's how I structured my query to calculate the average rating for each of these aspects, as well as the overall average rating:**

```
SELECT
  "Aircraft",
  "Origin",
  "Destination",
  "Connections",
  "Traveller_Type",
  "Seat_Type",
  AVG("Rating") AS "Avg_Overall_Rating",
  AVG("Seat_Comfort") AS "Avg_Seat_Comfort",
  AVG("Cabin_Staff_Service") AS "Avg_Cabin_Staff_Service",
  AVG("Food_Beverages") AS "Avg_Food_Beverages",
  AVG("Ground_Service") AS "Avg_Ground_Service",
  AVG("Value_For_Money") AS "Avg_Value_For_Money",
  AVG("Entertainment") AS "Avg_Entertainment"
FROM
  "DATAIKU_SCHEMA"."node-f285e3fa_BRITISHAIRWAYSROUTEANALYSIS_STANDARDIZED_PFDB_SQL"
GROUP BY
  "Aircraft",
  "Origin",
  "Destination",
  "Connections",
  "Traveller_Type",
  "Seat_Type"
ORDER BY
  "Avg_Overall_Rating" DESC;
```

This query was designed to segment the data not only by service aspects but also by aircraft type, origin, destination, connections, traveller type, and seat type, offering a complete view of the factors contributing to overall satisfaction.




| Aircraft | Origin | Destination | Connections | Traveller_Type | Seat_Type | Avg_Overall_Rating | Avg_Seat_Comfort | Avg_Cabin_Staff_Service | Avg_Food_Beverages | Avg_Ground_Service | Avg_Value_For_Money |
|---|---|---|---|---|---|---|---|---|---|---|---|
| string | string | string | string | string | string | double | double | double | double | double | double |
| Text | Text | Text | Text | Text | Text | Decimal | Decimal | Decimal | Decimal | Decimal | Decimal |
| Boeing 777 | LHR | KRK | | Couple Leisure | Economy Class | 10.0 | 3.0 | 3.0 | 3.0 | 3.0 | |
| A320 / Boeing 767 | London | New Orleans | | Business | Business Class | 10.0 | 2.0 | 2.0 | 1.0 | 3.0 | |
| Boeing 777-200 | London | New Orleans | | Family Leisure | Economy Class | 10.0 | 5.0 | 5.0 | 4.0 | 5.0 | |
| Boeing 747-400 | London | New Orleans | | Business | Premium Economy | 10.0 | 1.0 | 3.0 | 1.0 | 1.0 | |
| E170 / A319 | London | Tel Aviv | | Business | Economy Class | 10.0 | 3.0 | 5.0 | 1.0 | 4.0 | |
| Boeing 777-200 | London | Mauritius | | Family Leisure | Economy Class | 10.0 | 5.0 | 5.0 | 4.0 | 5.0 | |
| A380 and Boeing 747 | London | New Orleans | | Couple Leisure | First Class | 10.0 | 4.0 | 5.0 | 3.0 | 5.0 | |
| Airbus A320 | London | New Orleans | | Couple Leisure | Economy Class | 10.0 | 4.0 | 1.0 | -1.0 | 2.0 | |
| Boeing 747 | London | Mauritius | | Couple Leisure | First Class | 10.0 | 1.0 | 5.0 | 1.0 | 1.0 | |
| Airbus A320 | London | Tel Aviv | | Couple Leisure | Economy Class | 10.0 | 4.0 | 1.0 | -1.0 | 2.0 | |
| Boeing 787 | LHR | GLA | | Business | Business Class | 10.0 | 1.0 | 4.0 | 3.0 | 4.0 | |
| Boeing 777 | LGW | RAK | | Couple Leisure | First Class | 10.0 | 4.0 | 5.0 | 4.0 | 4.0 | |
| A320 | LGW | RAK | | Couple Leisure | Business Class | 10.0 | 3.0 | 5.0 | 3.0 | 3.0 | |
| Embraer 190 | London | Rio De Janeiro | | Family Leisure | Business Class | 10.0 | 3.0 | 5.0 | 3.0 | 3.0 | |
| A322 | London | Athens | | Business | Economy Class | 10.0 | 2.0 | 3.0 | 1.0 | 1.0 | |
| A380 and Boeing 747 | London | Mauritius | | Couple Leisure | First Class | 10.0 | 4.0 | 5.0 | 3.0 | 5.0 | |
| A322 | London | Milan | | Business | Economy Class | 10.0 | 2.0 | 3.0 | 1.0 | 1.0 | |
| Boeing 747-400 | London Heathrow | Basel | | Couple Leisure | Business Class | 10.0 | 4.0 | 5.0 | 5.0 | 4.0 | |
| Boeing 777-200 | London Heathrow | Basel | | Family Leisure | First Class | 10.0 | 3.0 | 4.0 | 2.0 | 3.0 | |
| Boeing 777-200 | London Gatwick | San Jose Costa ... | | Family Leisure | Economy Class | 10.0 | 5.0 | 5.0 | 4.0 | 5.0 | |
| A322 | London | Rio De Janeiro | | Business | Economy Class | 10.0 | 2.0 | 3.0 | 1.0 | 1.0 | |
| Boeing 747 | London Heathrow | Basel | | Business | Business Class | 10.0 | 1.0 | 2.0 | 1.0 | 1.0 | |
| A320 | London Heathrow | Basel | | Family Leisure | Business Class | 10.0 | 4.0 | 5.0 | 4.0 | 5.0 | |

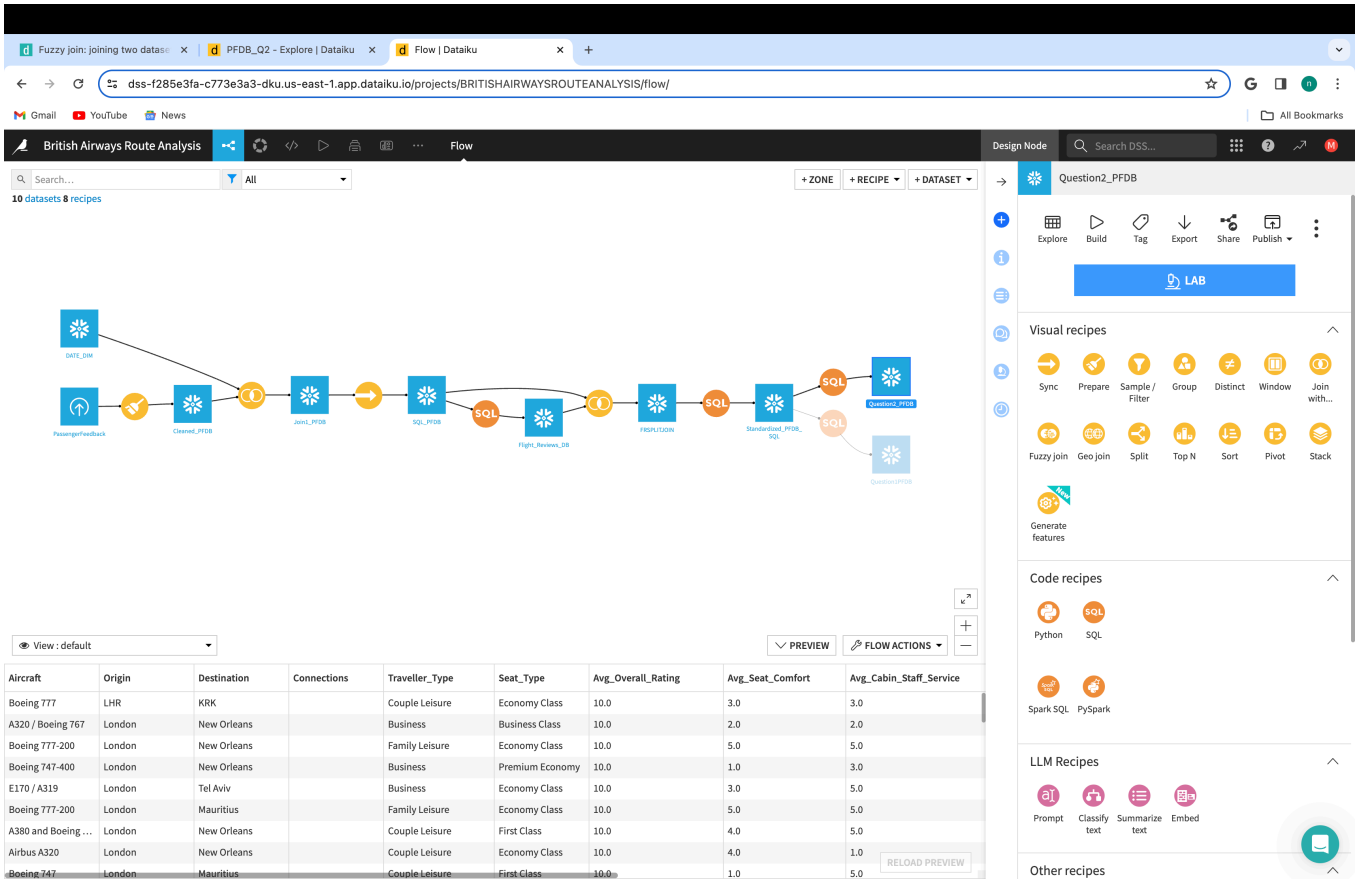
# Findings and Analysis

Upon analyzing the output from the code for Question 2, which aimed to identify service aspects most closely linked to overall customer satisfaction and how they vary by route, here's what I found:

- Seat Comfort: There appears to be a consistent correlation between seat comfort and overall satisfaction across different routes and classes. However, it is interesting to note that the highest overall ratings don't always align with the highest ratings for seat comfort.
- Cabin Staff Service: Cabin staff service generally receives higher ratings, which often aligns with higher overall satisfaction scores. This validates my initial hypothesis that staff interaction plays a critical role in shaping passengers' perceptions of their experience.
- Food and Beverages: The variation in satisfaction with food and beverages is interesting. In some instances, low ratings in this aspect coincide with high overall satisfaction ratings, suggesting that passengers may not prioritize food and beverages as highly as other service aspects, depending on the route or class.
- Ground Service: Ground service shows a diverse range of ratings, but it seems to have less impact on the overall rating compared to in-flight services like seat comfort and cabin staff service.
- Value for Money: Value for money consistently reflects on overall satisfaction, with economic and premium economy classes showing a close relationship between high value for money ratings and overall satisfaction.

- Entertainment: In-flight entertainment ratings are varied, but they do not consistently align with overall satisfaction ratings, indicating that this service aspect may not be significant in overall satisfaction for certain routes and aircraft types.

- Route-Specific Satisfaction: The overall satisfaction ratings appear to vary significantly across different routes. For instance, some routes show perfect overall ratings of 10, while others have lower scores. This suggests that there might be route-specific factors influencing passenger satisfaction.




| Aircraft | Origin | Destination | Connections | Traveller_Type | Seat_Type | Avg_Overall_Rating | Avg_Seat_Comfort | Avg_Cabin_Staff_Service |
|---|---|---|---|---|---|---|---|---|
| Boeing 777 | LHR | KRK | | Couple Leisure | Economy Class | 10.0 | 3.0 | 3.0 |
| A320 / Boeing 767 | London | New Orleans | | Business | Business Class | 10.0 | 2.0 | 2.0 |
| Boeing 777-200 | London | New Orleans | | Family Leisure | Economy Class | 10.0 | 5.0 | 5.0 |
| Boeing 747-400 | London | New Orleans | | Business | Premium Economy | 10.0 | 1.0 | 3.0 |
| E170 / A319 | London | Tel Aviv | | Business | Economy Class | 10.0 | 3.0 | 5.0 |
| Boeing 777-200 | London | Mauritius | | Family Leisure | Economy Class | 10.0 | 5.0 | 5.0 |
| A380 and Boeing ... | London | New Orleans | | Couple Leisure | First Class | 10.0 | 4.0 | 5.0 |
| Airbus A320 | London | New Orleans | | Couple Leisure | Economy Class | 10.0 | 4.0 | 1.0 |
| Boeing 747 | London | Mauritius | | Couple Leisure | First Class | 10.0 | 1.0 | 5.0 |

# Final Observations

In conclusion, my analysis revealed that customer satisfaction and service ratings fluctuate significantly across different routes and time periods. Seasonal changes were initially thought to have a considerable impact on satisfaction; however, the data did not substantiate this across the board. Instead, I observed a more complex pattern, one that wasn't entirely predictable by seasonality. I also explored the effects of holidays and weekends on satisfaction, which revealed nuanced impacts on customer ratings.

While certain service aspects affect passenger satisfaction, the context of the route plays a huge role in shaping the overall experience. The data suggests that we shouldn't only focus on the quality of service delivery but also consider route-specific strategies to enhance passenger satisfaction. This approach to understanding and improving customer experience could lead to increased customer loyalty and a stronger competitive position in the market.

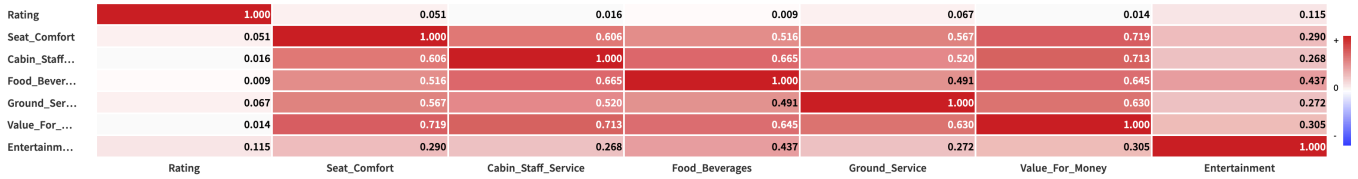
# Navigating the Altitudes of Satisfaction

## Client & Scope of Analysis

For this project I will be working with a British Airways dataset, spanning over a decade from 2012 to 2023. The dataset is a rich repository from Kaggle, includes a range of passenger reviews narrating experiences from different angles – comfort, service, food, and entertainment, to name a few.

The scope of my analysis is focused on uncovering insights that British Airways can leverage to enhance the passenger experience. We will delve into the correlation between various service aspects and overall customer satisfaction, dissect the preferences of traveler types, and assess the impact of aircraft and seat types on the perceived quality of travel.

## Correlation Matrix



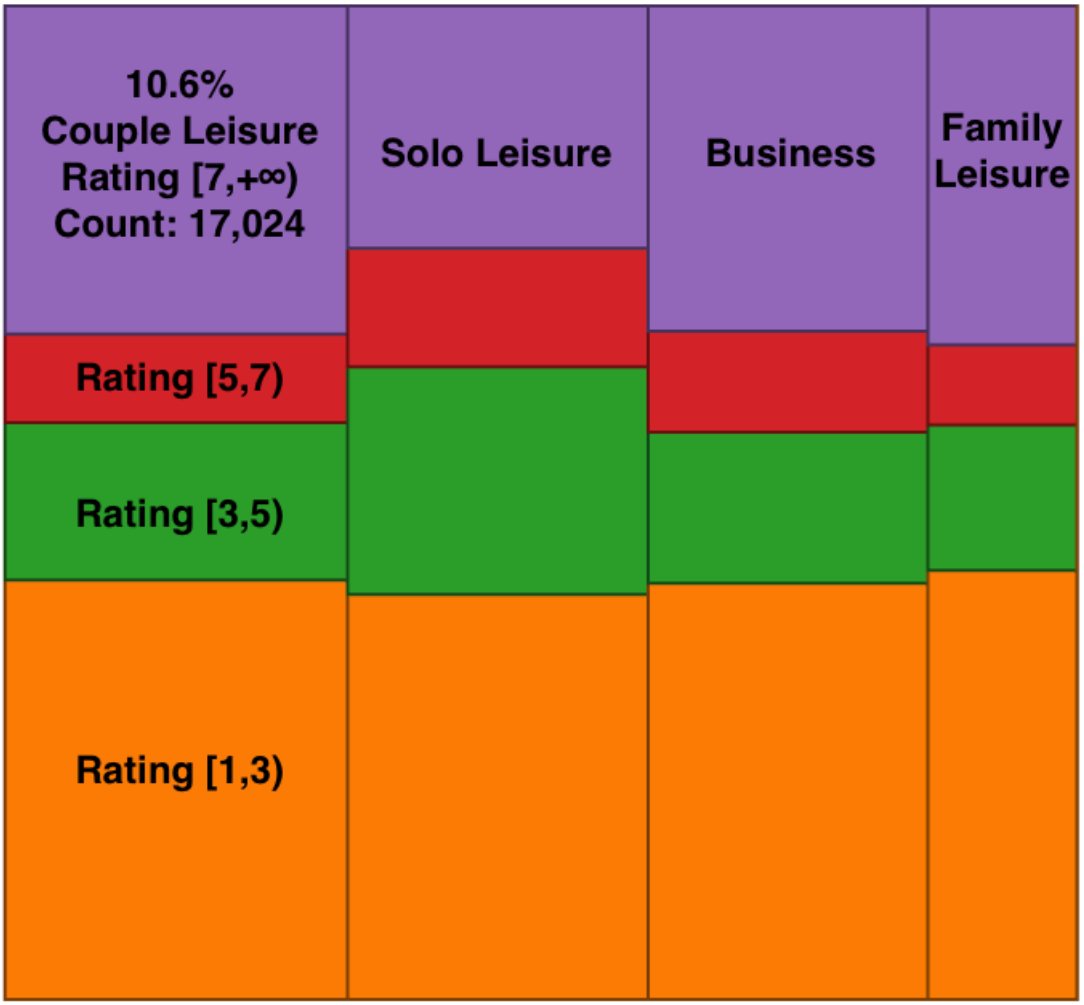

The Correlation Matrix visualization examines the relationship between overall ratings and various service aspects such as **Seat Comfort, Cabin Staff Service, Food & Beverages, and more.**

The values range from -1 to +1, with +1 signifying a perfect positive correlation, 0 meaning no correlation, and -1 indicating a perfect negative correlation. The diagonal, showing a value of 1, is the correlation of each variable with itself (perfect correlation). The values outside the diagonal show the correlation between different variables. In this matrix, we don't observe any negative correlations, which means none of the service aspects has an inverse relationship with the overall rating.

**Here are my findings:**

- No strong correlation between 'Overall Rating' and any single service aspect.
- **'Seat Comfort'** and **'Cabin Staff Service'** correlation: **0.606** – suggesting passenger comfort enhances perception of service.
- High correlation between **'Value For Money'** and both **'Cabin Staff Service'** ( **0.713** ) and 'Food & Beverages' ( **0.645** ) – indicating these factors significantly impact perceived value.
- **'Entertainment'** has moderate to **low** correlations with other factors, indicating it's less pivotal to overall satisfaction.
- **'Ground Service'** shows moderate correlations with **'Cabin Staff Service'** , **'Food & Beverages'** , and **'Value For Money'** , highlighting the importance of the end-to-end travel experience.

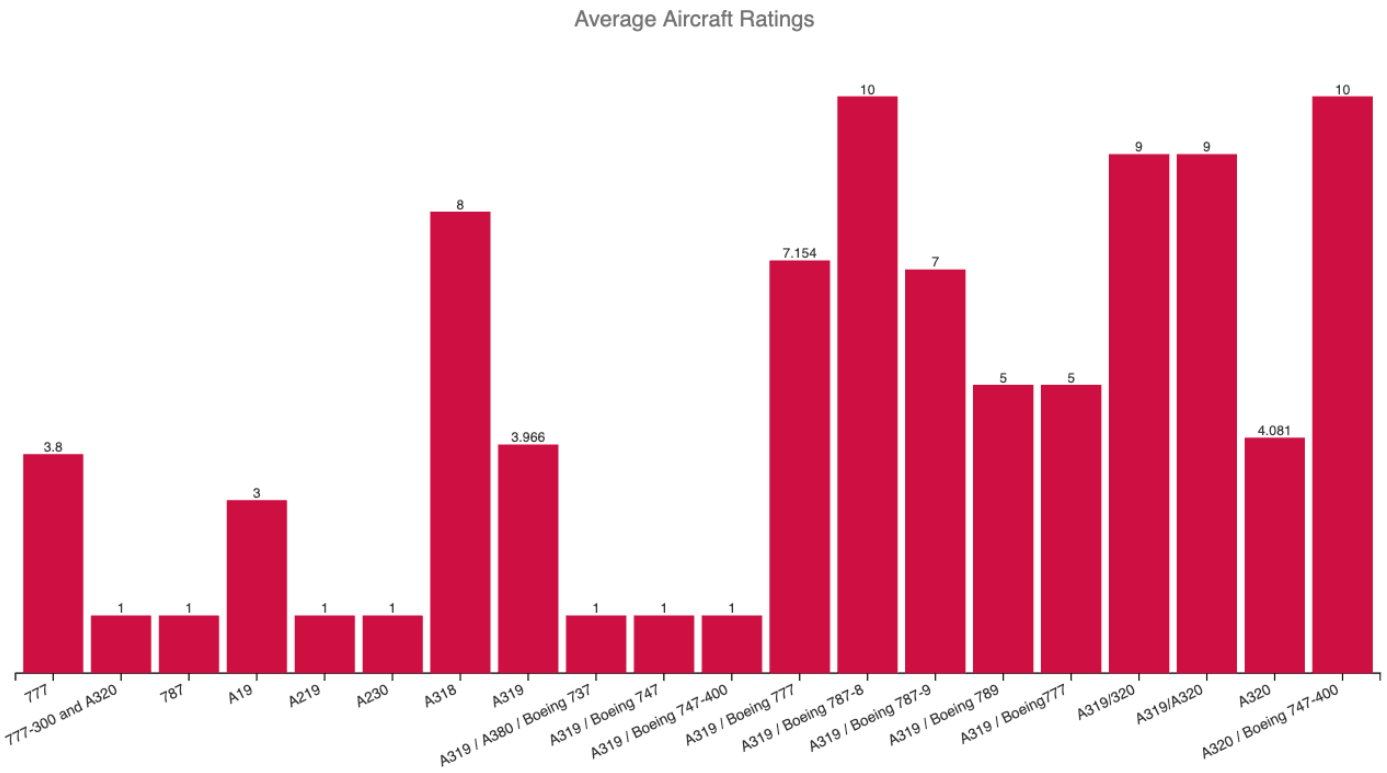
# Mosaic Plot




The mosaic plot offers a visual representation of the distribution of ratings among different types of travelers: **Couple Leisure, Solo Leisure, Business and Family Leisure.**

- **Couple leisure** has the largest representation in the highest rating interval (rating [7,+∞)), comprising 10.6% of all ratings with a count of 17,024. This suggests that couples on leisure trips are more likely to give the highest ratings.
- Each traveller type shows a similar distribution pattern across rating intervals, with the most common ratings falling in the interval [1,3] for all types except for **Solo Leisure** , where the most common rating interval is slightly higher at [3,5].
- **Business travellers** have a significant number of ratings in the highest category as well, which could imply that their satisfaction or their expectations might be well met by the services they use.
- **Family leisure** seems to have a smaller representation across all rating intervals compared to other types of travellers. It's the smallest group in the highest rating interval, which might suggest families are either less likely to give high ratings or there are simply fewer family leisure reviews.

# Bar Chart of Average Aircraft Rating




The "Avg Rating by Aircraft" chart displays how passengers rate their experience on different aircraft types. The **highest satisfaction** is shown for the **A319/Boeing 787-8** and **A320/Boeing 747-400** , indicating these models may offer attributes or services that resonate well with passengers. On the other hand, aircrafts like the **A219** and **A320** have received **lower ratings** , signaling areas where passenger expectations may not be fully met.

# Bar Chart of Average Ratings by Seat Type


The "Avg Rating by Seat_Type" chart compares passenger satisfaction ratings across different classes of airline seats: **First Class, Economy Class, Business Class, and Premium Economy.**

- The horizontal bar chart shows that **First Class** has the highest average rating, surpassing 4.5, which aligns with expectations given the level of service and comfort typically associated with First Class.
- **Economy Class** also has a high average rating, close to 4.5. This suggests that passengers are very satisfied with the Economy Class service, which might reflect well on the value proposition of the offerings in that class.
- **Business Class** follows closely with an average rating of just over 4. This is slightly lower than Economy, which is unexpected since Business Class usually offers a more premium experience.
- **Premium Economy** has the lowest average rating among the four, nearing 4, but it's still a high rating overall, indicating general satisfaction with the services provided.

# Conclusions

In conclusion, the findings underscore the complexity of passenger satisfaction. The Correlation Matrix revealed that perceived value for money has the most substantial albeit moderate relationship with overall satisfaction.

The Mosaic Plot identified Couple Leisure as a particularly satisfied group, hinting at successful service elements that could be replicated or adapted for other traveler types.

In the "Avg Rating by Aircraft" chart, we see a clear demarcation between the best and worst-performing aircraft in terms of passenger ratings. This suggests that certain aircraft features or the service provided

aboard them aligns more closely with passenger expectations and needs.

Lastly, the "Avg Rating by Seat Type" chart puts a spotlight on the varying levels of satisfaction across different seat classes, with First Class leading the pack.

# Analyzing Passenger Satisfaction for British Airways Using Predictive Modeling


# Analyzing Passenger Satisfaction for British Airways Using Predictive Modeling


## Client and Scope of Analysis

**Client:** British Airways Reviews Dataset

**Data Source:** Kaggle

For this project, I used predictive modeling to answer the question:

> "Can we predict which variables are likely to yield higher passenger satisfaction scores based on historical ratings data?"

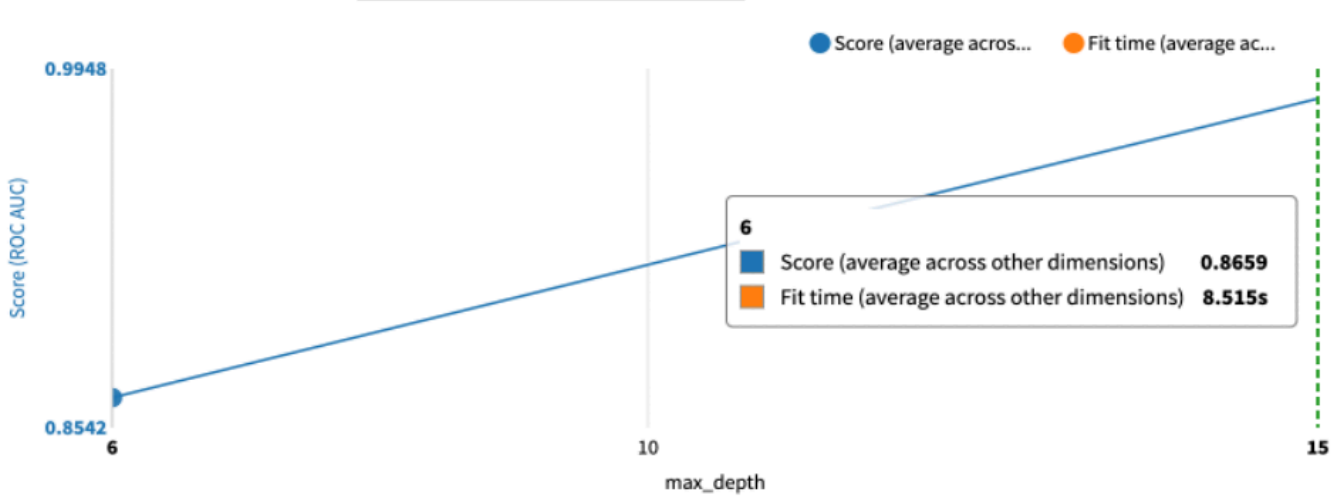
The objective was to uncover insights and patterns that might not be immediately obvious by simply reviewing raw data. By applying predictive modeling techniques such as **Random Forest** , **Decision Tree** , and **XGBoost** , we aim to assist British Airways in enhancing their customer satisfaction strategies.

## Model 1: Random Forest (RF)

The first model I created was Random Forest, which is a popular technique that uses multiple decision trees to make predictions. This type of model is well-suited for classification problems like predicting passenger satisfaction scores, as it can handle large data sets with higher dimensionality.

The following are my observations from the RF model:

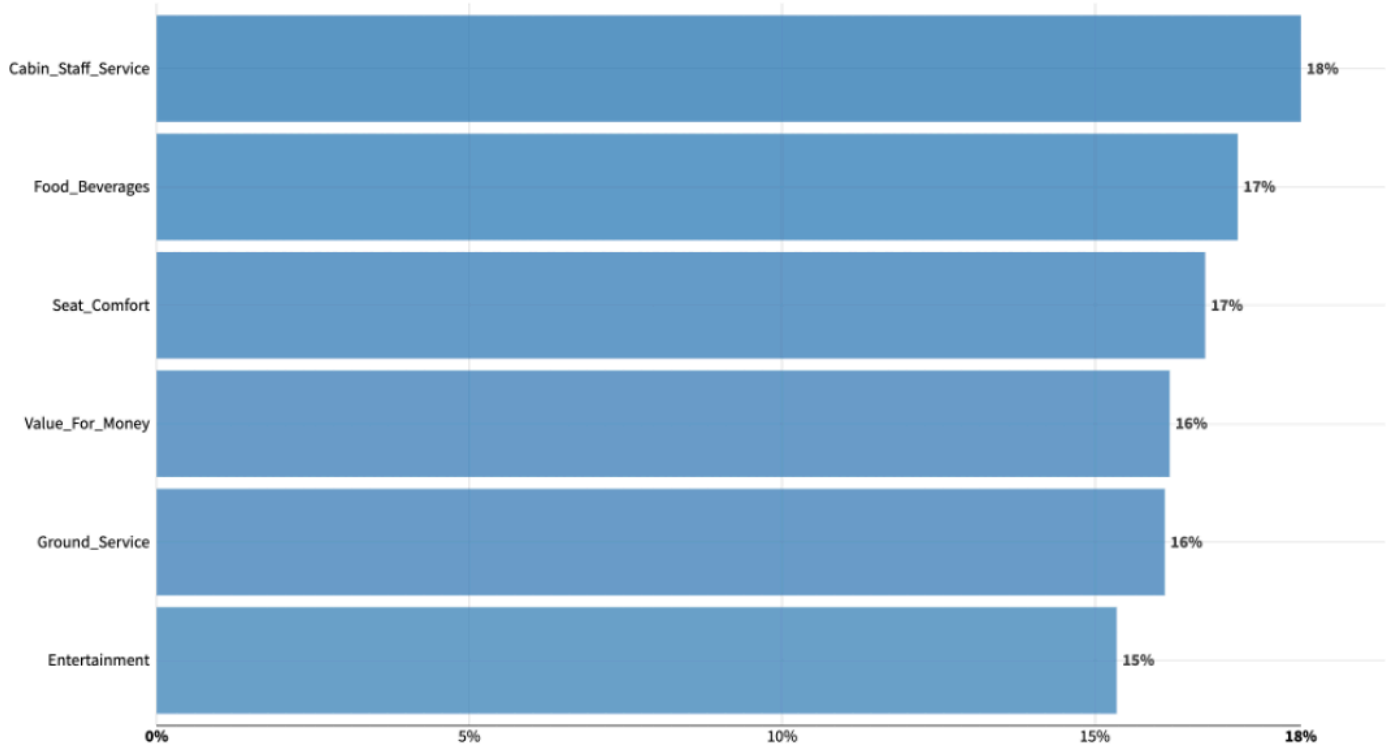
**Hyperparameters**



- The ROC AUC score increases as max_depth increases from 6 to 10, suggesting that allowing the trees to grow more deeply improves the model's ability to differentiate between the classes (satisfied and unsatisfied passengers).
- The model's performance improves with depth, but the rate of improvement slows after max_depth of 10, as indicated by the levelling off of the ROC AUC score.
- The fit time, represented by the orange dot, also increases substantially with depth.

Given the high ROC AUC score and reasonable fit time at max_depth of 10, this appears to be a good balance for my Random Forest model.

**Feature Importance**




The most influential factors affecting passenger satisfaction are predominantly service-related, with 'Cabin Staff Service', 'Food and Beverages', and 'Seat Comfort' being the top predictors. This shows us the importance of service quality in passenger satisfaction. These features collectively shape the passenger's experience, emphasizing areas that might require attention to boost overall ratings.

**Metric Details**

| | |
|---|---|
| Accuracy | **0.9286** |
| Precision | **0.8945** |
| Recall | **0.8719** |
| F1-score | **0.8830** |

- **Accuracy:** 93% of the predictions match the actual ratings, signifying a high level of overall accuracy.
- **Precision:** 89% indicates that when the model predicts a passenger is satisfied, it is correct 89% of the time.
- **Recall:** 87% shows that it correctly identifies 87% of all satisfied passengers.
- **F1-Score:** 88% is a balance between Precision and Recall and indicates a robust model, especially since it was the optimization target.

**Confusion Matrix**

| | Predicted `1` | Predicted `0` | Total |
|---|---|---|---|
| Actually `1` | 5417 | 796 | 6213 |
| Actually `0` | 639 | 13236 | 13875 |
| Total | 6056 | 14032 | 20088 |

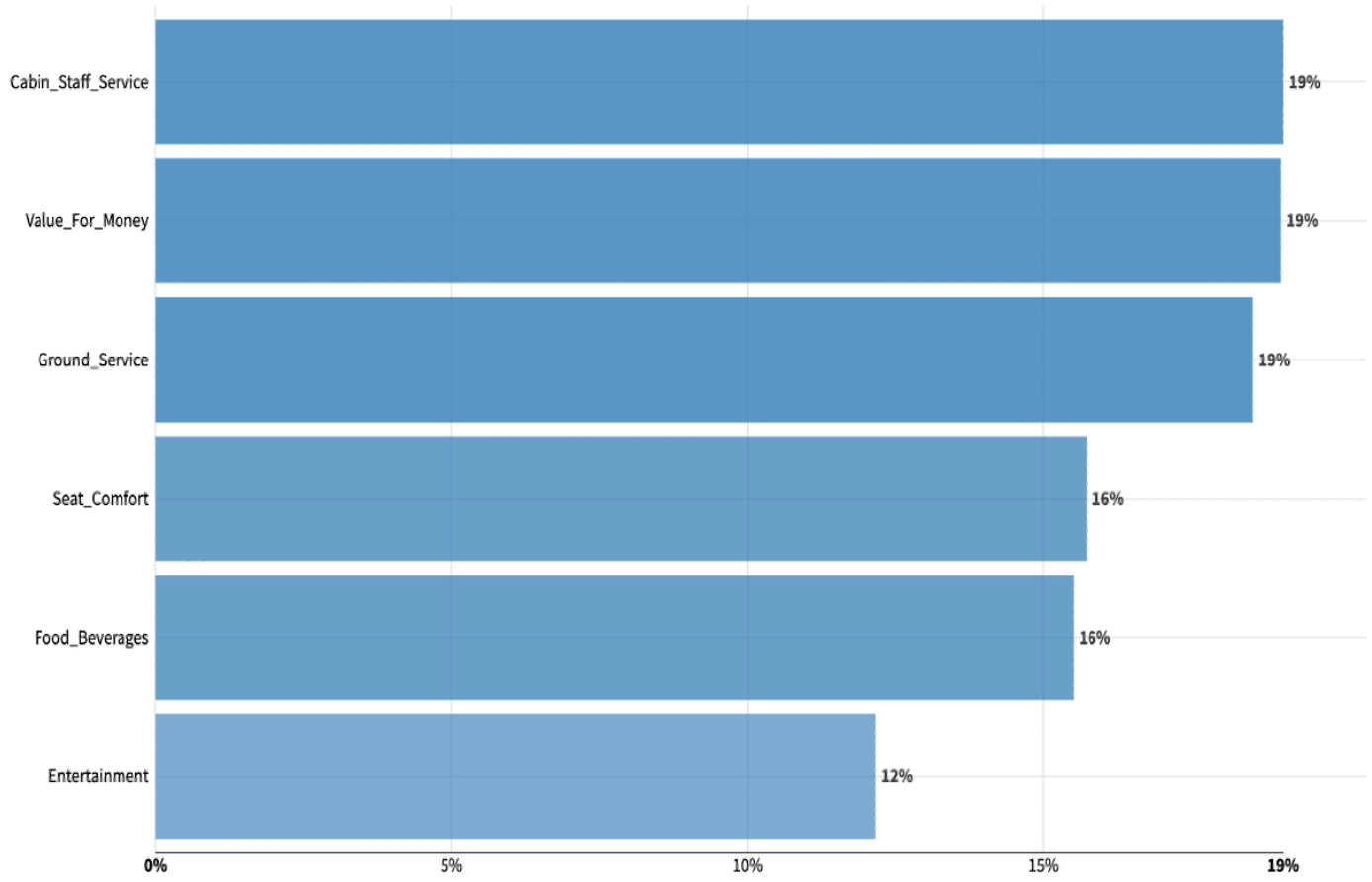
The Confusion Matrix shows:

- **True Positives (TP):** 5417 cases where the model correctly predicted passengers as satisfied.
    - **False Negatives (FN):** 796 cases where the model incorrectly predicted satisfied passengers as unsatisfied.
    - **True Negatives (TN)** : 13236 cases where the model correctly identified unsatisfied passengers.
    - **False Positives (FP):** 639 cases where the model incorrectly labeled unsatisfied passengers as satisfied.

# Model 2: XGBoost (Extreme Gradient Boosting)

The second model implemented in this analysis is the XGBoost. This model is an enhanced version of gradient boosting machines and is highly regarded for its performance in classification tasks through the use of decision trees.

**Feature Importance**




**Standout Features in XGBoost Model**

- Cabin Staff Service (19%)
- Value for Money (19%)
- Ground Service (19%)

These three features equally dominate the importance scale in the XGBoost model, each accounting for 19% of the predictive power. This suggests a very balanced influence between these factors, suggesting that both the perception of value and the quality of service (both on the ground and by the cabin staff) are crucial for passenger satisfaction.

**Features of Lesser Importance**

- In-flight Entertainment (12%)
- Food and Beverages (16%)
- Seat Comfort (16%)

While still significant, these features are slightly less influential in the XGBoost model compared to the leading factors. Notably, in-flight entertainment has the lowest impact among the variables, which suggests that while it contributes to passenger satisfaction, it is not as important as the service quality or perceived value for money.

**Metric Details**

| | |
|---|---|
| Accuracy | **0.8544** |
| Precision | **0.7334** |
| Recall | **0.8318** |
| F1-score | **0.7795** |

- **Accuracy:** 85.44% indicates a strong overall model performance, correctly identifying both satisfied and unsatisfied passengers.
- **Precision:** 73.34% suggests that when the model predicts passenger satisfaction, it is correct about 73% of the time.
- **Recall:** 83.18% indicates the model's capability to correctly identify a high percentage of actually satisfied passengers.
- **F1-Score:** 77.95% offers a balanced measure of precision and recall, vital for scenarios where both types of errors have similar costs.

**Confusion Matrix**

| | Predicted 1 | Predicted 0 | Total |
|---|---|---|---|
| Actually 1 | 5168 | 1045 | 6213 |
| Actually 0 | 1879 | 11996 | 13875 |
| Total | 7047 | 13041 | 20088 |

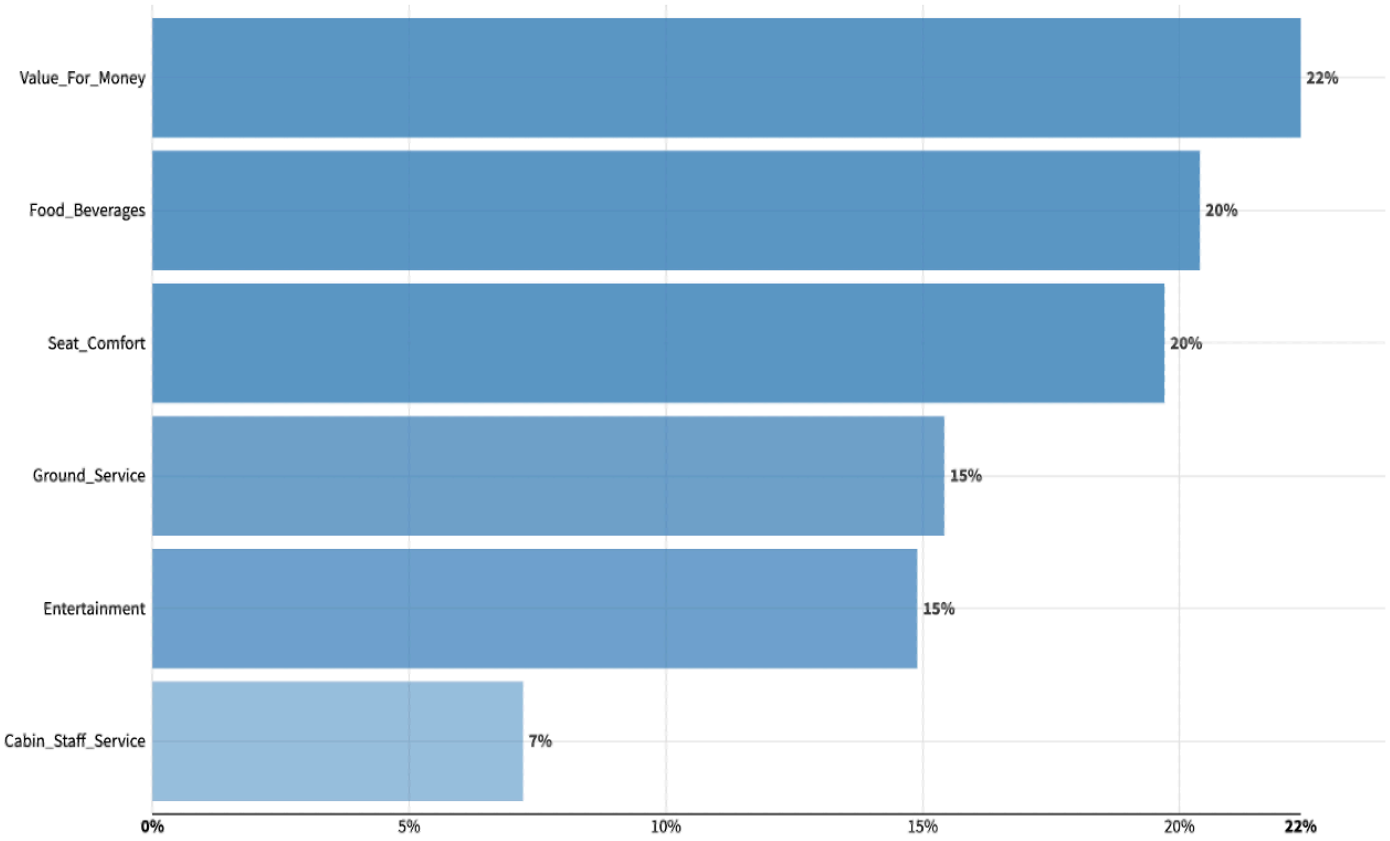
The Confusion Matrix for the XGBoost model reveals:

- **True Positives (TP):** 5,168 - Correct predictions of satisfied passengers.
- **False Positives (FP):** 1,879 - Passengers incorrectly identified as satisfied.
- **True Negatives (TN)** : 11,996 - Correct predictions of unsatisfied passengers.
- **False Negatives (FN):** 1,045 - Satisfied passengers missed by the model.

# Model 3: Decision Tree (DT)

The third model in this analysis is the Decision Tree, which constructs a tree-like graph of decisions and their possible consequences.

**Feature Importance**




- Food and Beverages (27%)
- In-flight Entertainment (20%)

These features stand out as they are deemed significantly more influential in the Decision Tree model than in the previous models. 'Food and Beverages' is particularly notable, as it appears to be the highest-rated factor, suggesting that aspects related to dining services are viewed as important to passenger satisfaction in this model. Interestingly, 'In-flight Entertainment' also receives a notable emphasis, which is a shift from its lesser importance in the Random Forest and XGBoost models.

'Cabin Staff Service' and 'Ground Service', which were more prominent in the previous models, are less emphasized in the Decision Tree. This is particularly striking given its top ranking in both Random Forest and XGBoost models, which suggests that the Decision Tree model may not capture the nuances of service quality's impact as effectively as other models.

**Metric Details**

| | |
|---|---|
| Accuracy | **0.6340** |
| Precision | **0.4438** |
| Recall | **0.7241** |
| F1-score | **0.5503** |

The performance metrics of the Decision Tree model are considerably lower than those of the previous models:

- **Accuracy:** 63.40% - This is significantly lower than other models, indicating that it is less capable of accurately identifying satisfied and dissatisfied passengers.
- **Precision:** 44.38% - Suggests that less than half of the passengers identified by the model as satisfied are actually satisfied, indicating a high rate of false positives.
- **Recall:** 72.41% - Although the model is reasonably effective at identifying who is satisfied, it is not very precise.
- **F1-Score:** 55.03% - Reflects the poor balance between precision and recall, indicating overall inefficacy in the model's predictions.

**Confusion Matrix**

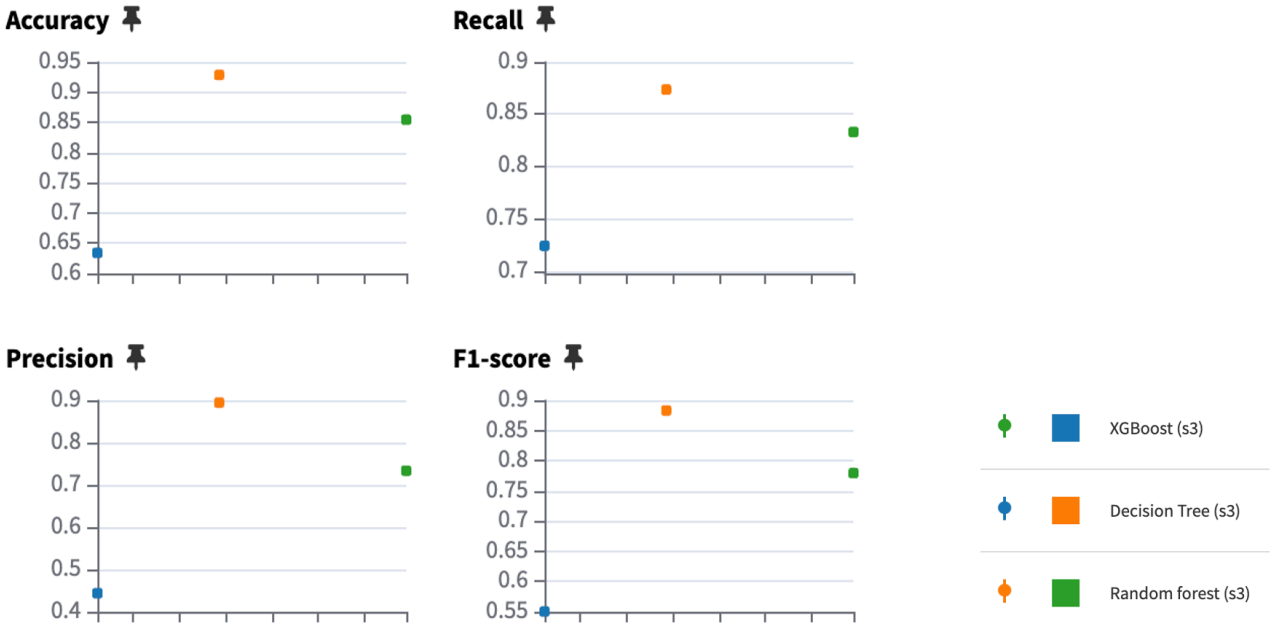
| | Predicted `1` | Predicted `0` | Total |
|---|---|---|---|
| Actually `1` | 4499 | 1714 | 6213 |
| Actually `0` | 5639 | 8236 | 13875 |
| Total | 10138 | 9950 | 20088 |

The confusion matrix for the Decision Tree model reveals:

- **True Positives (TP):** 4,499 - The model correctly predicted these passengers as satisfied.

- **False Positives (FP):** 5,639 - A high number indicating many passengers were incorrectly identified as satisfied.
- **True Negatives (TN):** 8,236 - Correctly identified dissatisfied passengers.
- **False Negatives (FN)** : 1,714 - Satisfied passengers that the model failed to recognize.

# Model Comparisons (RF, XGBoost, DT)




**Feature Importance Ranking Comparison**

- Value for Money: Highly ranked by both XGBoost and Random Forest, less so by the Decision Tree.
- Ground Service: Important in XGBoost and Random Forest; less emphasis in the Decision Tree.
- Seat Comfort: Consistently important across all models, albeit with slight ranking variations.
- Food and Beverages: Most important in the Decision Tree but less so in other models.
- Cabin Staff Service: Lower importance across all models.
- Entertainment: Varies significantly, with higher importance in the Decision Tree and less in XGBoost.

**Selection of the Best Model**

The Random Forest model is the superior choice for predicting passenger satisfaction within the British Airways Reviews Dataset. This model achieves the highest scores in critical metrics like F1-score and ROC AUC.

- **Highest ROC AUC Score (0.984):** Demonstrates its strong discriminative ability between satisfied and unsatisfied passengers.
- **Highest F1-score (88.3%):** Indicates a robust balance between precision and recall, essential for a reliable prediction model.

# Conclusion

After a thorough comparative analysis of the three models—XGBoost, Decision Tree, and Random Forest—conducted to determine the most effective approach for predicting passenger satisfaction at British Airways, it is clear that the Random Forest model stands out as the most suitable choice for future implementation.

British Airways is recommended to adopt the Random Forest model for its future strategies in monitoring and enhancing passenger satisfaction. This model not only provides the most accurate and reliable predictions but also offers detailed insights into the factors that significantly impact passenger satisfaction.