DATA MODELLING PROJECT

2024

By:

- Ahmed Jumaa
- Mohamed El-Feky
- Malak Sherif

DATA MODELLING

Contents

DATA MODELLING

Project Requirements

Objective: The objective of this project is to assist the executive management of a major airline company in analyzing their current business processes and identifying new opportunities for business expansion. The first deliverable will focus on analyzing flight activity to ensure ongoing business process efficiency.

Stakeholders:

- Executive Management: Requires insights to make informed decisions about business processes and expansion.
- Marketing Department: Interested in analyzing frequent flyer behavior for targeted marketing strategies.
- Finance Team: Interested in analyzing reservation processes for profit analysis.
- Customer Care Department: Requires analysis of customer interactions for service improvement.

Requirements:

1. **Flight Activity Analysis:**
    - Analyze flights taken by frequent flyers.
    - Determine fare basis paid by frequent flyers.
    - Track frequency of upgrades by frequent flyers.
    - Analyze earning and redemption of frequent flyer miles.
    - Evaluate response of frequent flyers to special fare promotions.
    - Determine duration of overnight stays by frequent flyers.
    - Track proportion of frequent flyers with gold, platinum, or titanium status.
2. **Reservation Process Analysis:**
    - Analyze reservation processes through multiple channels.
    - Determine company profit from reservation processes.
3. **Customer Care Interaction Analysis:**
    - Analyze customer care interactions before, during, and after trips.
    - Categorize interactions based on type (inquiry, complaint, feedback).
    - Determine problem severity (if applicable).

DATA MODELLING

Data Modeling Approach

The data modeling process is driven by the specific business needs and requirements of the organization. These requirements are translated into logical and conceptual designs that form the foundation of the data model. In this section, we will outline the key business processes that need to be defined and the associated dimensions and facts that will be used to capture and analyze the relevant data.

Each business process has its level of granularity, which is determined by the specific business requirements it aims to address. This granularity ensures that the data model captures the necessary details to support the business processes effectively. By defining the appropriate level of granularity for each business process, we can ensure that the data model meets the organization's needs and enables insightful analysis and decision-making. We will outline the key business processes that need to be defined and the associated dimensions and facts that will be used to capture and analyze the relevant data.

1. **Flight Activity**

    - Summary: This process involves analyzing the flights taken by frequent flyers, including fare basis, upgrades, frequent flyer miles, response to promotions, overnight stays, and status levels (gold, platinum, titanium).
    - Insights: It can provide insights into customer travel patterns, preferences, and loyalty levels, helping the airline tailor its services and promotions.

2. **Reservation**

    - Summary: This process includes analyzing the reservation processes through multiple channels to understand customer behavior and profitability.
    - Insights: It can help the finance team analyze company profits and identify areas for revenue optimization.

3. **Customer Care**

    - Summary: This process involves analyzing interactions with customers before, during, and after their trip to handle inquiries, complaints, and feedback.
    - Insights: It can provide insights into customer satisfaction levels, common issues, and opportunities for service improvement.

DATA MODELLING

4. **Frequent Flyer**

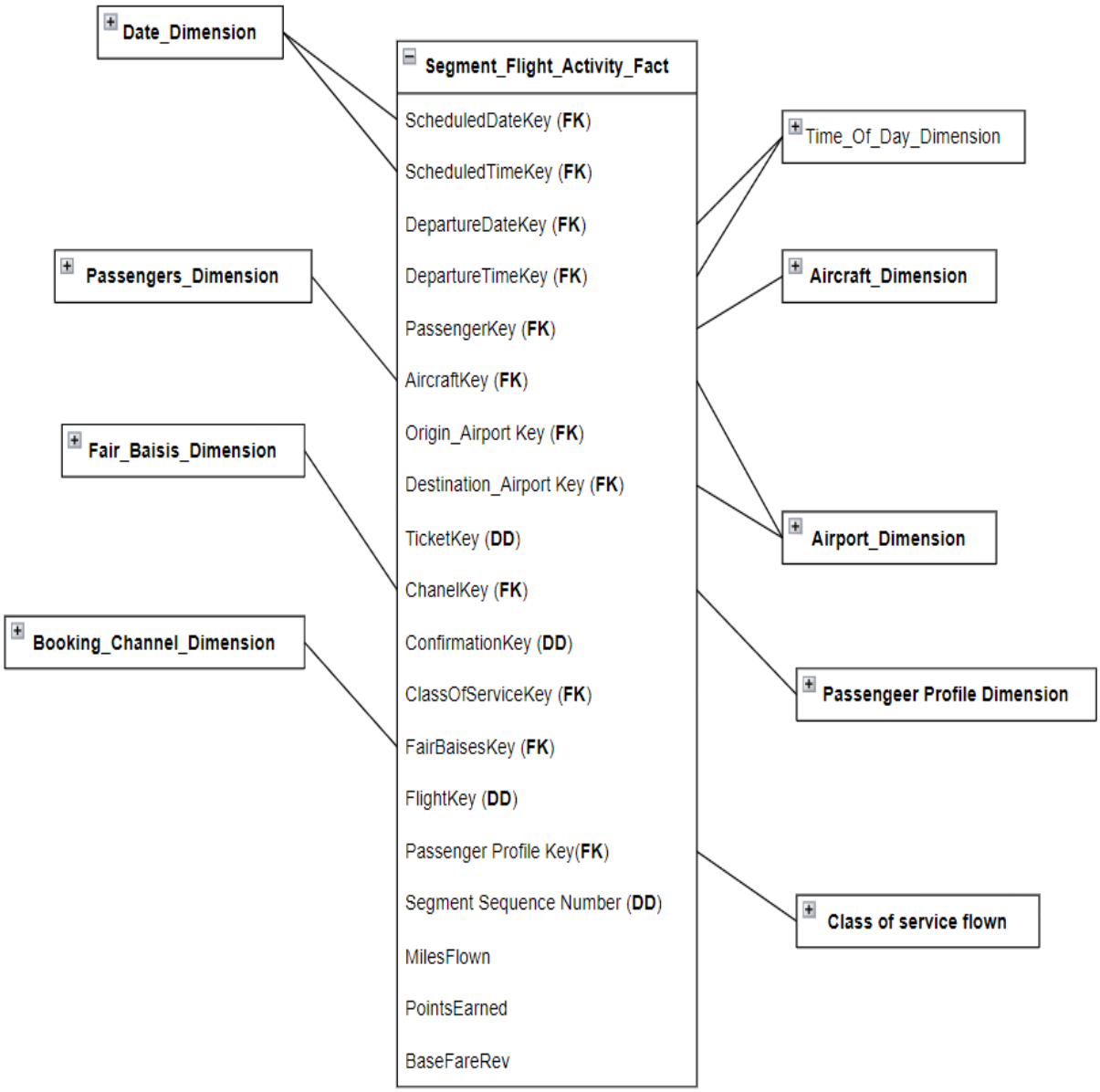
    - Summary: This process focuses on analyzing the behavior of frequent flyers, including their tier status, home airports, club membership, and lifetime mileage tier.
    - Insights: It can help the marketing department analyze activity by frequent flyer tier, understand customer preferences, and tailor loyalty programs and promotions accordingly.

5. **Travel Services Hotel**
    - This process involves analyzing the hotel booking process and its impact on overall travel experience and customer satisfaction. It helps in understanding the preferences of customers when it comes to accommodation and can lead to partnerships with hotel chains for mutual benefit.

By defining and modeling these business processes, the airline company can gain a comprehensive understanding of its operations, customer behavior, and revenue streams, enabling data-driven decision-making and business growth.

DATA MODELLING

Flight Activity Business Process:



In the flight activity analysis, the granularity refers to the level of detail captured in the data. In this case, the airline captures data at multiple levels of granularity:

DATA MODELLING

1. **Leg Level:** This represents the most granular level where data is captured for each individual leg of a flight, from takeoff to landing without any stops. Capacity planning and flight scheduling analysts are interested in this level as they can calculate load factors and analyze operational metrics such as flight duration and on-time arrivals.
2. **Segment Level:** A segment represents a single flight number flown by a single aircraft. A segment may consist of one or more legs. For example, a flight from San Francisco to Minneapolis with a stop in Denver would be two legs but one segment (SFO-MSP). Passenger revenue and mileage credit are determined at the segment level, making it important for marketing and revenue analysis.
3. **Trip Level:** A trip encompasses all the segments a passenger takes from origin to destination, even if there are layovers or aircraft changes. Understanding trip-level data helps in analyzing overall customer demand and travel patterns.
4. **Itinerary Level:** This represents the entire airline ticket or reservation, including all the segments and trips associated with it.

The decision to start the analysis at the segment level is driven by the meaningful revenue metrics available at this level. Although the data warehouse will eventually need to handle more granular leg-level data, the initial focus on segments allows for a detailed analysis of customer behavior and revenue generation patterns.

The fact table will have one row for each boarding pass collected from passengers. To support this analysis, the data model will include extensive dimensionality, utilizing role-playing dimensions for date, time, and airport details. This approach allows for flexibility in analyzing segments from different perspectives without duplicating dimension tables.

Data Collection Strategy: The segment level is chosen as the starting point for data collection due to its granularity and meaningful revenue metrics. This allows for a detailed analysis of customer behavior on individual flights.

Decision-Making Considerations:

- **Leg vs. Segment Level:** While some departments may focus on leg-level operations for capacity planning and flight scheduling, the marketing and revenue groups primarily focus on segment-level metrics due to their relevance to passenger revenue and mileage credit determination.
- **Trip Level Data:** While analyzing segments, it's important to consider the entire trip a passenger takes, as the trip provides an accurate picture of customer demand. This holistic view helps sales and marketing analysts understand the overall travel patterns and preferences of customers.

DATA MODELLING

Data Collection and Analysis:

- **Data Sources:** Data for each segment is collected from various sources, including booking systems, flight operations data, and customer loyalty programs.
- **Data Integration:** Integrating data from these sources allows for a comprehensive view of each segment, including passenger information, flight details, and fare information.
- **Data Cleansing and Transformation:** Before analysis, the data undergoes cleansing and transformation to ensure accuracy and consistency. This involves removing duplicates, correcting errors, and standardizing formats.
- **Data Analysis Techniques:** Analysis of segment-level data involves various techniques, including descriptive analytics to understand patterns and trends, and predictive analytics to forecast future behavior based on historical data.
- **Key Performance Indicators (KPIs):** KPIs for segment-level analysis include load factors, revenue per available seat mile (RASM), and passenger yield. These metrics help evaluate the performance of each segment and identify areas for improvement.

Metrics of Interest:

- **Flights Taken:** Number of flights taken by frequent flyers on specific segments.
- **Fare Basis:** The fare basis for each segment, providing insights into pricing strategies and customer preferences.
- **Upgrades:** Frequency of upgrades on specific segments, indicating customer loyalty and satisfaction.
- **Frequent Flyer Miles:** Analysis of how customers earn and redeem their frequent flyer miles on each segment.
- **Response to Promotions:** Whether frequent flyers respond to special fare promotions on particular segments.
- **Overnight Stays:** Duration of overnight stays associated with each segment.
- **Status Levels:** Proportion of frequent flyers with gold, platinum, or titanium status on specific segments.

DATA MODELLING

Fact Table: Each row in the fact table represents a boarding pass collected from passengers, providing a detailed record of each segment.

1. **Scheduled Date Key:** Represents the date of the scheduled flight. This key is included to track the scheduled dates of flights, which is important for analyzing trends in flight scheduling and capacity planning.

2. **Scheduled Time Key:** Represents the scheduled departure time of the flight. This key is included to analyze trends in flight departure times and to identify any patterns or anomalies in scheduling.

3. **Departure Date Key:** Represents the actual departure date of the flight. This key is included to track the actual departure dates of flights, which is important for analyzing delays and operational performance.

4. **Departure Time Key:** Represents the actual departure time of the flight. This key is included to track the actual departure times of flights, which is important for analyzing delays and operational performance.

5. **Passenger Key:** Represents a unique identifier for each passenger. This key is included to track individual passenger behavior and preferences, which is important for personalized marketing and customer satisfaction.

6. **Aircraft Key:** Represents a unique identifier for each aircraft. This key is included to track aircraft performance and usage, which is important for maintenance and operational planning.

7. **Origin Airport Key:** Represents the airport code for the origin airport of the flight. This key is included to track the origin of each flight segment, which is important for analyzing route performance and demand.

8. **Destination Airport Key:** Represents the airport code for the destination airport of the flight. This key is included to track the destination of each flight segment, which is important for analyzing route performance and demand.

DATA MODELLING

9. **Ticket Key:** Represents a unique identifier for each ticket. This key is included to track ticket sales and revenue, which is important for revenue analysis and forecasting.

10. **Channel Key:** Represents the sales channel through which the ticket was purchased. This key is included to track ticket sales channels, which is important for analyzing marketing effectiveness and distribution strategies.

11. **Fare Basis Key:** Represents the fare basis code for the ticket. This key is included to track fare types and pricing structures, which is important for revenue management and pricing strategies.

12. **Flight Key:** Represents a unique identifier for each flight. This key is included to track individual flight segments, which is important for analyzing flight performance and scheduling.

13. **Class Key:** The "class key" in the fact table documentation refers to a foreign key (FK) that represents the class of service for a particular flight segment or reservation. This key is used to link the fact table to the class of service dimension table, which contains detailed information about different classes of service offered by the airline.

14. **Passenger Profile Key:** Represents a unique identifier for each passenger profile. This key is included to track passenger preferences and behavior, which is important for personalized marketing and customer satisfaction.

15. **Segment Sequence Number:** Represents the sequence number of the flight segment within the trip. This key is included to track the order of flight segments, which is important for analyzing trip itineraries and customer travel patterns.

16. **Confirmation Key:** A confirmation key is a unique identifier assigned to a booking or reservation to distinguish it from other bookings. It is typically used in the travel industry, such as airlines or hotels, to track and manage reservations. The confirmation key is generated when a booking is made and is often included in the confirmation email or document provided to the customer. This key allows the customer to easily reference and manage their booking, and it is also used internally by the company to identify and process the reservation.

DATA MODELLING

17. **Miles Flown**: This metric represents the actual distance traveled by passengers on the flight segment, measured in miles. It is a critical factor for revenue calculation, as ticket prices often vary based on the distance flown. Additionally, miles flown is important for operational planning, as it helps airlines assess fuel consumption, flight times, and overall route efficiency.

18. **Points Earned**: refers to the number of frequent flyer points earned by passengers for a specific flight segment or reservation. Frequent flyer points are typically awarded based on factors such as the distance traveled, fare class, and membership tier.

19. **Base Fare Revenue**: Base fare revenue is the revenue generated from the base fare of the ticket for the flight segment. It excludes additional fees and charges, such as taxes, surcharges, and ancillary fees. Base fare revenue is a key metric for revenue analysis and forecasting, as it reflects the core revenue generated from ticket sales.

Dimensionality:

Dimension tables for segment-level analysis include date, time, airport, flight, and passenger dimensions. These dimensions provide context to the segment data and facilitate analysis from different perspectives. Utilization of role-playing dimensions for date, time, and airport details, ensuring efficient linkage to the underlying physical dimension tables. This allows for flexibility in analyzing segments from different perspectives without duplicating dimension tables.

DATA MODELLING

Date Dimension:

The Date Dimension in the context of the airline's flight activity business process is a fundamental component that provides a comprehensive view of time-related information associated with flight scheduling and operations. This dimension contains various attributes that facilitate the analysis of scheduled and actual flight dates, times, and other temporal aspects. Here's a detailed explanation of the attributes in the Date Dimension:

Date Key: This is a unique identifier for each date record in the dimension table, allowing for easy reference and integration with other data.

Full Date: Represents the complete date in the format YYYY-MM-DD, providing a precise timestamp for each flight segment.

Day of Week: Indicates the day of the week for the date, represented either as a numeric value (1-7) or a textual representation (e.g., Monday-Sunday), helping to analyze trends based on weekdays.

Day Number in Calendar Month: Specifies the sequential day number within the calendar month (1-31), useful for understanding the progression of days within a month.

Day Number in Calendar Year: Indicates the sequential day number within the calendar year (1-365 or 1-366 for leap years), providing a chronological reference within the year.

Day Number in Fiscal Month: Represents the sequential day number within the fiscal month, which may differ from the calendar month in organizations with a non-standard fiscal calendar.

Day Number in Fiscal Year: Specifies the sequential day number within the fiscal year, which may vary from the calendar year in organizations with a non-standard fiscal calendar.

Last Day in Month Indicator: A flag (1 or 0) indicating whether the date is the last day of the calendar month, useful for identifying month-end dates.

DATA MODELLING

Calendar Month Name: The name of the calendar month (e.g., January, February) corresponding to the date, providing a descriptive representation of the month.

Calendar Quarter: Indicates the quarter of the calendar year (1-4) to which the date belongs, helping to analyze performance and trends by quarter.

Calendar Year: Specifies the calendar year corresponding to the date, providing a reference point for temporal analysis.

Role-Playing Dimension:

The Date Dimension serves as a role-playing dimension in the airline's data model, providing two meanings within one fact table. It serves as both the scheduled date and the departure date for each segment flight. This dual role allows for more detailed analysis of late departure flights and helps in monitoring and optimizing flight schedules.

Significance and Analysis:

The Date Dimension plays a crucial role in analyzing flight schedules, departure times, and operational efficiency. By leveraging the attributes in this dimension, airlines can gain valuable insights into scheduling patterns, identify trends in late departures, and optimize operational processes to enhance overall efficiency and customer satisfaction.

DATA MODELLING

Time of Day Dimension:

The Time of Day Dimension in the airline's data model provides a granular view of time, allowing for analysis and reporting based on different times of the day. This dimension is essential for understanding booking patterns, flight schedules, and passenger behavior at different times.

Attributes:

1. **TimeKey** (PK): Primary key uniquely identifying each record in the dimension table.
2. **TimeID** (NK): Natural key representing the unique identifier of the time slot. It is a non-sequential, business-recognized identifier used for data integration and reference.
3. **TimeOfDay**: Categorizes the time slot into different segments of the day, such as morning, afternoon, evening, and night.
4. **HourOfDay**: Represents the hour component of the time slot, providing a detailed view of the time.
5. **MinuteOfDay**: Represents the minute component of the time slot, allowing for precise time-based analysis.
6. **TimeDescription**: Describes the time slot in human-readable format, providing context and clarity.
7. **DayOfWeek**: Indicates the day of the week for the time slot, enabling analysis of weekly patterns and trends.
8. **WeekendIndicator**: Flags whether the time slot falls on a weekend (Saturday or Sunday), providing insights into weekend travel behavior.
9. **HolidayIndicator**: Flags whether the time slot falls on a holiday, enabling analysis of holiday travel patterns and trends.
10. **MonthOfYear**: Indicates the month of the year for the time slot, facilitating seasonal analysis and planning.

Significance and Analysis:

The Time of Day Dimension is critical for analyzing time-related patterns and trends in airline operations and passenger behavior. By leveraging the attributes in this dimension, airlines can optimize flight schedules, and enhance customer service.

DATA MODELLING

Aircraft Dimension:

The Aircraft Dimension is a confirmed dimension in the airline's data model, focusing on tracking the status of each aircraft used in the operation. It includes the following attributes:

1. **AircraftKey (PK):** Primary key uniquely identifying each record in the dimension table.
2. **AircraftID (NK):** Natural key representing the unique identifier of the aircraft. This key is non-sequential and can be used for lookups.
3. **Status:** The current status of the aircraft, indicating whether it is active, in maintenance, retired, etc. This attribute is subject to change over time.
4. **StartDate:** The date when the aircraft status was first recorded or became active.
5. **EndDate:** The date when the aircraft status changed or became inactive. This attribute is used for historical tracking in the context of SCD type 2.

Slowly Changing Dimension (SCD) Type 2:

The Aircraft Dimension implements SCD Type 2 to track historical changes to the status of each aircraft. When a change occurs in the status attribute, a new record is inserted into the dimension table with a new surrogate key (AircraftKey). The original record remains in the table but is marked as inactive (using the EndDate attribute) to maintain historical accuracy.

Significance and Analysis:

The Aircraft Dimension is crucial for analyzing fleet status, monitoring changes over time, and identifying patterns in aircraft utilization and maintenance. By tracking changes to aircraft status, airlines can optimize maintenance schedules, identify potential issues early, and ensure a reliable fleet operation.

Airport Dimension:

The Airport Dimension in the airline's data model serves as a role-playing dimension, providing meaningful information for both origin and destination airports in flight segments. It includes the following attributes:

DATA MODELLING

1. **AirportKey (PK):** Primary key uniquely identifying each record in the dimension table.
2. **AirportID (NK):** Natural key representing the unique identifier of the airport. This key is non-sequential and can be used for lookups.
3. **AirportName:** The name of the airport, providing a descriptive identifier for the airport.
4. **Country:** The country where the airport is located, providing geographical context.
5. **State:** The state or region where the airport is located, providing further geographical context.
6. **City:** The city where the airport is located, providing additional geographical information.

Role-Playing Dimension:

The Airport Dimension serves as a role-playing dimension in the airline's data model, providing two meanings within one fact table. It serves as both the origin and destination airport for each flight segment. This dual role allows for more detailed analysis of flight routes, passenger traffic, and regional performance.

Significance and Analysis:

The Airport Dimension plays a crucial role in analyzing flight routes, passenger traffic patterns, and regional performance. By leveraging the attributes in this dimension, airlines can gain valuable insights into airport utilization, identify trends in passenger demand, and optimize route planning to enhance overall operational efficiency and customer satisfaction.

Fare Basis Dimension:

The Fare Basis Dimension in the airline's data model is a confirmed dimension that provides essential information about fare basis codes used in pricing airline tickets. It includes the following attributes:

1. **FareBasisKey (PK):** Primary key uniquely identifying each record in the dimension table.
2. **FareCode:** The fare basis code, which is a unique identifier for a specific fare class or type of ticket.

3. **Description:** A description of the fare basis code, providing additional information about the fare class or ticket type.
4. **Conditions:** Any conditions or restrictions associated with the fare basis code, such as minimum stay requirements or advance purchase rules.
5. **Price Range:** The price range associated with the fare basis code, indicating the typical price range for tickets with this fare basis.

The Fare Basis Dimension is confirmed because it is a reliable and accurate source of information about fare basis codes, which are essential for pricing airline tickets. It provides a consistent framework for understanding fare structures, pricing strategies, and revenue management practices in the airline industry.

Passenger Dimension:

The Passenger Dimension in the airline's data model is designed to capture comprehensive information about passengers, providing valuable insights into passenger demographics, preferences, and behavior. This dimension plays a crucial role in customer segmentation, loyalty programs, and travel pattern analysis.

Attributes:

1. **PassengerKey** (PK): Primary key uniquely identifying each record in the dimension table.
2. **PassengerID** (NK): Natural key representing the unique identifier of the passenger. It is a non-sequential, business-recognized identifier used for data integration and reference.
3. **FirstName**: The first name of the passenger, providing personal identification.
4. **LastName**: The last name of the passenger, providing personal identification.
5. **Gender**: The gender of the passenger, providing demographic information.
6. **DateOfBirth** (DOB): The date of birth of the passenger, enabling age-related analysis and targeted marketing.
7. **PhoneNumber**: The phone number of the passenger, facilitating communication and customer service.
8. **Nationality**: The nationality of the passenger, providing citizenship information.
9. **Country**: The country of residence of the passenger, providing geographical context.

10. **State**: The state or region of residence of the passenger, providing further geographical context.
11. **City**: The city of residence of the passenger, providing additional geographical information.

Significance and Analysis: The Passenger Dimension is a critical component for understanding and analyzing customer behavior in the airline industry. By analyzing the attributes in this dimension, airlines can segment customers effectively, tailor marketing campaigns, and enhance customer service to drive customer satisfaction and loyalty.

Passenger Profile Dimension

The Passenger Profile Dimension is a supplemental dimension table designed to address specific analytical requirements and manage the complexity of the main Passenger Dimension, which can contain tens to hundreds of millions of rows for a large airline. This dimension contains one record for each unique combination of attributes related to frequent flyer elite tier, home airport, club membership status, and lifetime mileage tier.

Attributes:

- **Passenger Profile Key (PK):** Primary key uniquely identifying each record in the dimension table.
- **Frequent Flyer Tier:** The tier or level of the frequent flyer program to which the passenger belongs, indicating their status and benefits.
- **Home Airport:** The airport code or identifier for the airport nearest to the passenger's residence, providing insights into passenger demographics and travel patterns.
- **Club Membership Status:** Indicates whether the passenger belongs to the airline's airport club at the time of each flight, helping in targeted marketing and customer service.
- **Lifetime Mileage Tier:** The tier or level based on the passenger's lifetime mileage, reflecting their loyalty and long-term relationship with the airline.

Slowly Changing Dimension (SCD) Type 2 Implementation:

Effective Date: Indicates the date when the record became effective or valid.

End Date: Indicates the date when the record expired or became inactive.

DATA MODELLING

Current Flag: A flag indicating whether the record is the current record for the passenger profile.

Version Number: A unique identifier for each version of the passenger profile record, helpful for auditing and tracking changes.

Purpose:

- The Passenger Profile Dimension is generated to manage the complexity and size of the main Passenger Dimension, facilitating easier and faster analysis and reporting.
- It provides a more focused dataset for analysis, which is particularly useful for marketing analysts who frequently leverage this dimension for their analysis needs.
- By isolating specific attributes into a separate dimension, it simplifies the process of tracking changes in frequent flyer tier, home airport, club membership status, and lifetime mileage tier over time.
- The dimension enhances performance for analysis and reporting tasks by creating a more manageable dataset with a reduced number of records.

Significance:

- Enables the airline to track changes in passenger attributes over time, providing valuable insights into passenger behavior and preferences.
- Simplifies the analysis process for marketing analysts and other users who frequently access passenger data for their analytical needs.
- Enhances the airline's ability to target marketing efforts and tailor services to specific passenger segments, such as frequent flyers, who are valuable customers for the airline.

Business Impact:

- **Marketing Strategies:** Insights from segment-level analysis help the marketing department tailor promotional offers and loyalty programs to specific customer segments, improving customer retention and revenue.
- **Revenue Optimization:** By analyzing fare basis and passenger behavior, the revenue management team can optimize pricing strategies for each segment, maximizing revenue potential.
- **Operational Efficiency:** Understanding flight activity at the segment level allows for better capacity planning and scheduling, leading to improved operational efficiency and customer satisfaction.

DATA MODELLING

Booking Channel Dimension:

The Booking Channel Dimension in the airline's data model serves as a critical component for analyzing reservation data and understanding customer interaction behavior across various booking channels. This dimension is indispensable for multiple teams within the airline company, primarily benefiting the Marketing, Sales, and Customer Care departments.

Attributes:

ChannelKey (PK): This is the surrogate key uniquely identifying each record in the dimension table. It facilitates efficient data retrieval and joins with fact tables.

ChannelID (NK): The natural key representing the identifier of the booking channel in the business domain. It is a non-surrogate, business-recognized identifier used for data integration and reference.

ChannelName: Descriptive name or label for the booking channel, providing clarity and context for users. It helps in identifying and distinguishing between different channels.

URL: Provides the Uniform Resource Locator (URL) associated with the booking channel, facilitating direct access to the channel's online platform or portal.

StartDate: Specifies the date when the booking channel becomes active or valid. It allows for temporal analysis and tracking of changes in channel usage patterns.

EndDate: Indicates the date when the booking channel is no longer active or valid. This attribute enables historical tracking of changes to the booking channels over time.

ChannelType: Categorizes the booking channel based on its nature, such as online travel agency (OTA), direct booking, call center, or mobile app.

Significance and Analysis:

The Booking Channel Dimension plays a crucial role in analyzing booking patterns, channel effectiveness, and customer behavior. By leveraging the attributes in this dimension, airlines can gain valuable insights into channel performance, customer preferences, and booking trends. This information enables the airline to optimize marketing strategies, improve sales efforts, and enhance customer service, ultimately leading to increased revenue and customer satisfaction.

DATA MODELLING

Class of Service Dimension:

The Class of Service Dimension in the airline's data model is designed to cater to the business community's feedback, focusing on tracking both the "Class of Service Flown" and the "Booking Class Purchased." Additionally, users expressed a need to easily filter and report on activity based on whether an upgrade or downgrade occurred.

Initial Reaction: Initially, the instinct was to include a second role-playing dimension and foreign key in the fact table to support both the purchased and flown class of service, along with a third foreign key for the upgrade indicator. This approach would have required complex logic in the BI application to identify various scenarios as upgrades, including different class transitions such as economy to premium economy or economy to business.

Combined Dimension Approach: However, due to the small number of distinct values in both class of service dimensions, a more efficient approach was adopted. The decision was made to combine them into a single "Class of Service" dimension. This combined dimension effectively represents the Cartesian product of the separate class dimensions, resulting in a 16-row dimension table (4 class purchased rows times 4 class flown rows).

Attributes and Relationships:

The combined dimension provides an opportunity to describe the relationship between the purchased and flown classes, such as a class change indicator. This combined class of service dimension can be compared to a type of "junk dimension," where the attributes within the dimension are tightly correlated.

Benefits:

1. Simplicity: Consolidating the separate dimensions into a single dimension simplifies the schema and makes it more intuitive for users to navigate and analyze.
2. Efficiency: The combined dimension reduces the number of foreign keys in the fact table, leading to improved query performance and reduced storage requirements.
3. Contextual Insights: The combined dimension allows for additional attributes that provide context and meaning, enhancing the depth of analysis and insights derived from the data.

By combining the separate class dimensions into a single "Class of Service" dimension, the data model remains scalable, maintainable, and optimized for analytical purposes, meeting the business requirements effectively.

DATA MODELLING

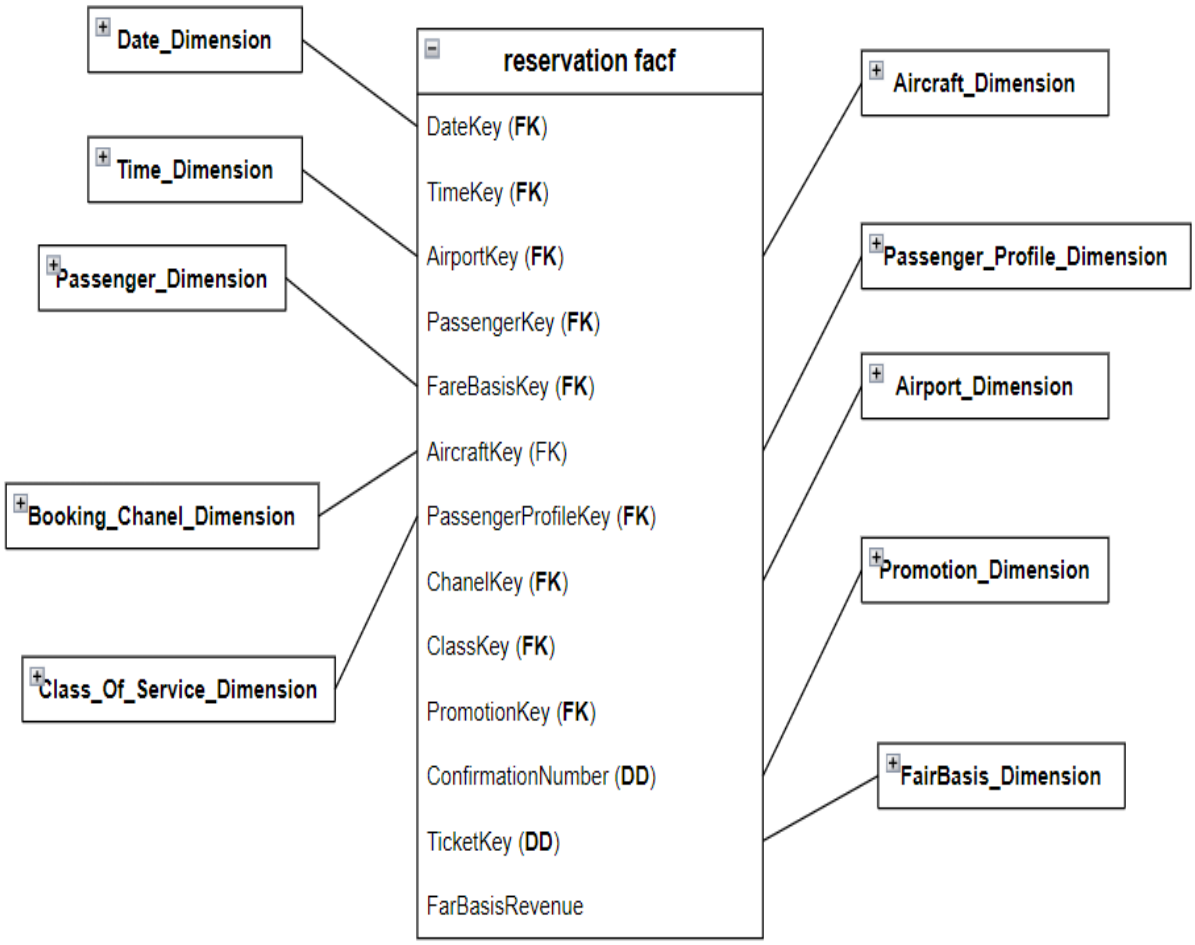
Reporting and Visualization:

- **Dashboards:** Dashboards are created to visualize segment-level data, providing executives and managers with a clear overview of performance metrics.
- **Ad Hoc Analysis:** Users can perform ad hoc analysis on segment data using tools like Tableau or Power BI, allowing for deeper insights and exploration of data trends.

Conclusion: Segment-level analysis is essential for understanding the behavior of frequent flyers and optimizing business processes in the airline industry. By focusing on this level of granularity and leveraging advanced analytics techniques, airlines can gain valuable insights that drive business growth and customer loyalty.

DATA MODELLING

Reservation Business Process:



Business Process:

The reservation process is a critical aspect of airline operations, involving the booking and allocation of seats for passengers on flights. It encompasses various steps, including ticket reservation, seat allocation, and revenue generation. Analyzing this process provides insights into passenger demand, revenue management, and flight occupancy, which are crucial for optimizing operations and enhancing customer satisfaction.

DATA MODELLING

Grain Level:

The decision to capture the reservation process at the ticket level was driven by several factors. While other grain levels, such as capturing each seat reservation or each passenger reservation, are viable options, the ticket level grain was chosen for its alignment with the business aspects and analytical needs of the airline company.

Ticket-Level Grain vs. Other Grain Levels:

1. **Seat-Level Grain:** Capturing each seat reservation would provide detailed information about seat occupancy and availability. However, this grain level could result in a significantly larger dataset and may not be as relevant for analyzing overall revenue and passenger booking patterns, which are more efficiently captured at the ticket level.
2. **Passenger-Level Grain:** Capturing each passenger reservation would allow for detailed analysis of individual passenger behavior and preferences. However, this grain level might not provide a comprehensive view of overall flight occupancy and revenue generation, which are key metrics for optimizing flight schedules and revenue management.

Key Differences:

- **Granularity:** The ticket-level grain provides a balance between granularity and relevance. It captures enough detail to analyze revenue and booking patterns effectively without overwhelming the analysis with unnecessary details.
- **Analytical Focus:** By focusing on the ticket level, the data mart can provide insights into overall revenue generation, passenger demand, and flight occupancy, which are essential for optimizing operations and marketing strategies.
- **Scalability:** The ticket-level grain is more scalable than seat-level or passenger-level grains, as it reduces the complexity of the data model while still providing valuable insights for decision-making.

In summary, the choice of the ticket-level grain for the reservation data mart allows the airline company to strike a balance between granularity and relevance, enabling effective analysis of key metrics while ensuring scalability and manageability of the data model.

DATA MODELLING

Data Mart Purpose: The data mart is designed to facilitate in-depth analysis of the reservation process, providing a comprehensive view of ticket reservations and associated metrics. By storing this process in a data mart, the airline company can:

- Analyze passenger booking patterns and preferences.
- Optimize seat allocation and flight scheduling.
- Evaluate the effectiveness of marketing promotions and pricing strategies.
- Monitor revenue generation and profitability.

Fact Table: The fact table, **Fact_Reservation**, captures key metrics and attributes related to each ticket reservation. It includes the following fields:

- **DateKey (FK)**: Foreign key referencing the date dimension.
- **TimeKey (FK)**: Foreign key referencing the time dimension.
- **PassengerKey (FK)**: Foreign key referencing the passenger dimension.
- **AircraftKey (FK)**: Foreign key referencing the aircraft dimension.
- **ClassKey (FK)**: Foreign key referencing the class of service dimension.
- **AirportKey (FK)**: Foreign key referencing the airport dimension.
- **TicketKey (DD)**: Degenerate dimension representing the unique identifier for each ticket reservation.
- **ChannelKey (FK)**: Foreign key referencing the booking channel dimension.
- **PassengerProfileKey (FK)**: Foreign key referencing the passenger profile dimension.
- **PromoKey (FK)**: Foreign key referencing the promotion dimension.
- **FairBasisKey (FK)**: Foreign key referencing the fare basis dimension.
- **FlightKey (DD)**: Degenerate dimension representing the unique identifier for each flight.
- **FairBasisRevenue**: Amount of revenue generated from the ticket reservation.

Related Dimensions:

- **Date Dimension**: Provides temporal context for the reservation process, allowing analysis over time.
- **Time Dimension**: Enhances the temporal analysis with detailed time-based attributes.

DATA MODELLING

- **Passenger Dimension**: Captures detailed information about passengers, enabling passenger-centric analysis.
- **Aircraft Dimension**: Provides insights into the aircraft used for each flight, facilitating analysis of seat availability and utilization.
- **Class of Service Dimension**: Describes the class of service for each reservation, allowing for analysis based on service level.
- **Airport Dimension**: Contains information about airports, enabling analysis based on departure and destination airports.
- **Booking Channel Dimension**: Describes the booking channels used by passengers, aiding in channel-specific analysis.
- **Passenger Profile Dimension**: Provides additional details about passengers, such as frequent flyer status, to enhance passenger-centric analysis.
- **Promotion Dimension**: Describes promotional offers used in ticket reservations, aiding in promotional analysis.
- **Fare Basis Dimension**: Contains details about fare basis codes, facilitating analysis based on fare types.

Reporting and Visualization:

- The reservation data mart provides a wealth of information that can be leveraged for insightful reporting and visualization. Here are some key aspects of analyzing and reporting on the data:
-
- Revenue Analysis: Utilize the reservation data mart to analyze revenue trends over time, by route, and by passenger segment. Visualizations such as line charts, bar charts, and pie charts can be used to depict revenue distributions and trends.
- Flight Occupancy: Analyze flight occupancy rates to optimize flight schedules and resource allocation. Use visualizations like heat maps or histograms to visualize seat occupancy on different flights.
- Booking Patterns: Identify booking patterns based on factors like booking channel, class of service, and time of booking. Visualizations such as scatter plots or trend lines can help visualize these patterns.
- Customer Segmentation: Segment customers based on their booking behavior, such as frequent flyers, last-minute bookers, or premium class travelers. Use

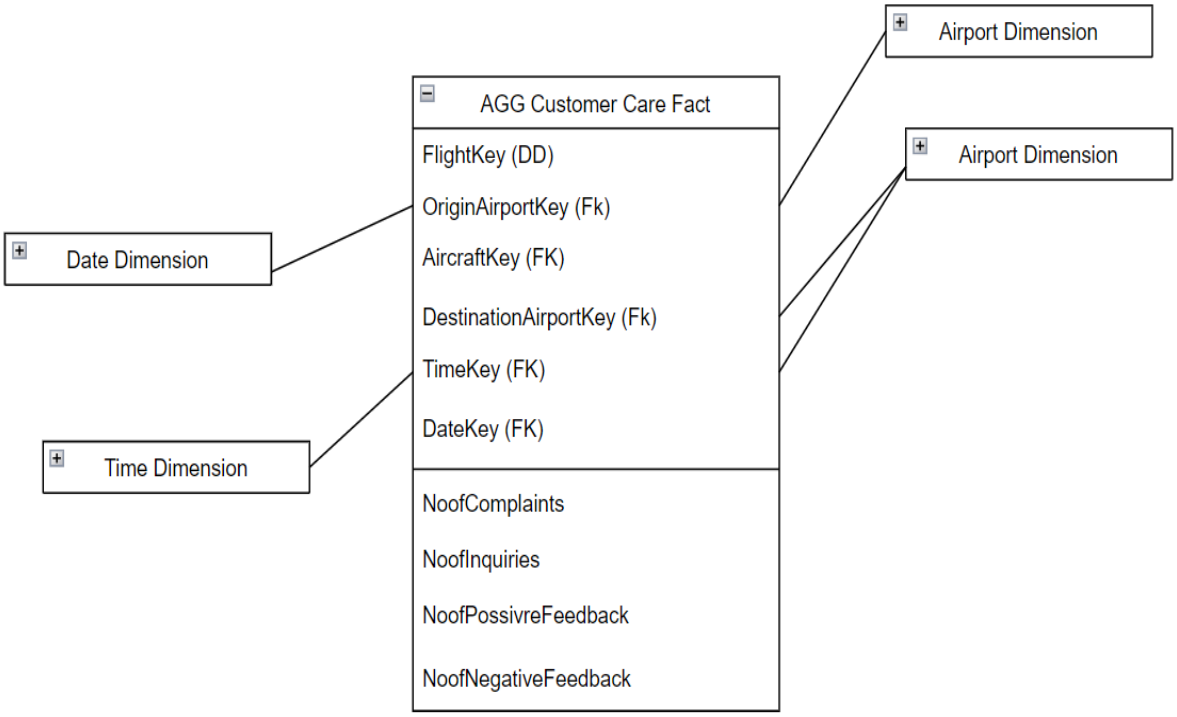
segmentation techniques to tailor marketing strategies and enhance customer experience.

- Operational Efficiency: Analyze operational metrics such as flight delays, cancellations, and customer complaints to identify areas for improvement. Visualizations like dashboards can provide a comprehensive view of operational performance.
- Predictive Analytics: Use historical reservation data to predict future booking trends and revenue projections. Machine learning models can be applied to forecast demand and optimize pricing strategies.

DATA MODELLING

Customer Care Business Process:

Agg_Customer_Care_Fact:



The Customer Care business process in the airline industry is a crucial component of ensuring customer satisfaction and loyalty. It involves handling customer inquiries, complaints, and feedback before, during, and after their trip. This process aims to address customer concerns promptly, resolve issues effectively, and enhance the overall customer experience.

Grain for Segment-Level Ratings: In the Customer Care business process, passengers have the opportunity to rate various aspects of their flight experience at the segment level. This includes rating aspects such as the flight crew, comfort level, onboard amenities, and overall satisfaction for each flight segment within a trip. Segment-level ratings provide

detailed insights into passenger feedback, allowing airlines to identify areas for improvement and enhance customer satisfaction.

Aggregated Fact Table for Flight Ratings: To consolidate and analyze passenger ratings across different aspects of their flight experience, an aggregated fact table can be used. The **"agg_fact_customer_care"** table serves this purpose by aggregating ratings and feedback from passengers for each flight. This table includes the following attributes:

- **FlightKey** (DD): A unique identifier for each flight.
- **OriginAirportKey** (FK): Foreign key referencing the origin airport.
- **AircraftKey** (FK): Foreign key referencing the aircraft used for the flight.
- **DestinationAirportKey** (FK): Foreign key referencing the destination airport.
- **TimeKey** (FK): Foreign key referencing the time dimension.
- **DateKey** (FK): Foreign key referencing the date dimension.
- **NoofComplaints**: Number of complaints received for the flight.
- **NoofInquiries**: Number of inquiries made by passengers.
- **NoofPositiveFeedback**: Number of positive feedback received.
- **NoofNegativeFeedback**: Number of negative feedback received.

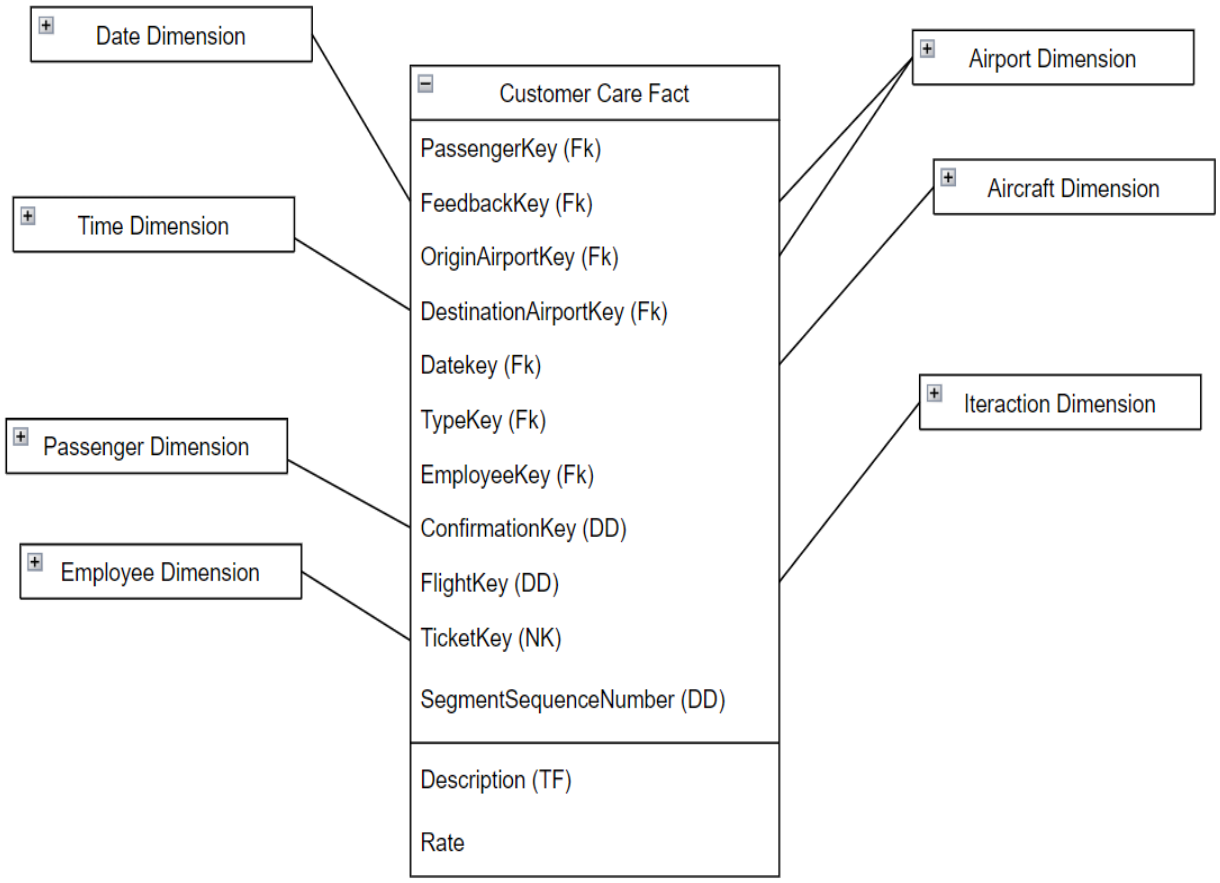
This aggregated fact table provides a comprehensive view of passenger feedback and ratings for each flight, allowing airlines to analyze trends, track performance, and make informed decisions to improve the overall customer experience.

The "**agg_fact_customer_care**" fact table is linked to the following dimensions:

1. **Date Dimension**: Provides temporal context for the flight ratings, allowing analysis based on dates such as day, month, and year.
2. **Time Dimension**: Offers more granular time-related attributes for analyzing ratings based on specific times of the day.
3. **Airport Dimension** (Origin and Destination): Contains information about the airports, including their names, countries, states, and cities, providing context for the flight ratings.
4. **Aircraft Dimension**: Includes details about the aircraft used for the flights, such as their statuses, start dates, and end dates, enabling analysis of ratings based on the aircraft's operational status.

DATA MODELLING

Customer_Care_Fact:



Another fact for the same business process is the customer care fact that provide more details about the specific customer rating on a specific flight with a description.

The customer care fact table provides detailed information about specific customer ratings and feedback for a particular flight, including descriptions of the feedback. This table is essential for analyzing and improving customer satisfaction and service quality.

Attributes:

- **PassengerKey (FK):** Foreign key linking to the Passenger Dimension, identifying the passenger giving the feedback.
- **FeedbackKey (FK):** Foreign key linking to the Feedback Dimension, providing details about the feedback given by the passenger.

DATA MODELLING

- **OriginAirportKey (FK):** Foreign key linking to the Airport Dimension, representing the origin airport of the flight.
- **DestinationAirportKey (FK):** Foreign key linking to the Airport Dimension, representing the destination airport of the flight.
- **DateKey (FK):** Foreign key linking to the Date Dimension, indicating the date of the flight.
- **TypeKey (FK):** Foreign key linking to the Type Dimension, specifying the type of feedback (e.g., complaint, inquiry, positive feedback, negative feedback).
- **EmployeeKey (FK):** Foreign key linking to the Employee Dimension, identifying the employee handling the feedback.
- **ConfirmationKey (DD):** Degenerate dimension representing the confirmation key associated with the flight.
- **FlightKey (DD):** Degenerate dimension representing the flight key associated with the flight.
- **TicketKey (NK):** Natural key identifying the ticket associated with the flight.
- **SegmentSequenceNumber (DD):** Degenerate dimension representing the segment sequence number of the flight.
- **Description (TF):** Text field describing the feedback provided by the passenger.
- **Rate:** Numeric field indicating the rating given by the passenger (e.g., satisfaction level).

Dimension Links:

- **Passenger Dimension:** Provides details about the passenger giving the feedback.
- **Airport Dimension:** Provides context about the origin and destination airports of the flight.
- **Date Dimension:** Offers temporal context for the flight, including the date of the flight.
- **Interaction Dimension:** Specifies the type of feedback given by the passenger (e.g., complaint, inquiry, positive feedback, negative feedback).
- **Employee Dimension:** Contains details about the employee handling the feedback.

DATA MODELLING

This customer care fact table enables airlines to analyze and understand specific customer feedback and ratings for each flight, helping them identify areas for improvement and enhance overall customer satisfaction.

Employee Dimension

providing essential information about employees working within the organization. This dimension enables comprehensive analysis of employee-related data, facilitating various HR-related analytics and reporting.

Attributes:

1. **Employee SID** (Surrogate Key): This is a unique identifier for each employee record in the dimension table, facilitating easy reference and integration with other data.
2. **Name:** Represents the name of the employee, providing identification and personalization in analytical reports.
3. **Age**: Specifies the age of the employee, enabling analysis of workforce demographics and age-related trends.
4. **Language**: Indicates the primary language(s) spoken by the employee, facilitating language-specific communication and support.
5. **Country**: Represents the country of residence or nationality of the employee, providing geographical context for workforce distribution and diversity analysis.
6. **Gender**: Indicates the gender of the employee, allowing for gender-based analysis and diversity reporting.
7. **Nationality**: Specifies the nationality of the employee, providing additional context for workforce diversity analysis.
8. **Job Title**: Represents the title or position held by the employee within the organization, facilitating hierarchical analysis and organizational structure mapping.
9. **Employment Status**: Indicates the employment status of the employee (e.g., full-time, part-time, contractor), enabling workforce planning and analysis of employment types.
10. **Department**: Specifies the department or functional area to which the employee belongs, facilitating analysis of workforce distribution and departmental performance.

DATA MODELLING

Interaction Dimension

capturing various types of customer interactions and feedback. This dimension facilitates detailed analysis of customer sentiment, feedback sources, and resolution statuses, providing insights to improve customer service and satisfaction.

Attributes:

1. **TYPE_ID** (Surrogate Key): This serves as a unique identifier for each interaction record in the dimension table, facilitating easy reference and integration with other data.
2. **TYPE**: Represents the type or category of interaction (e.g., inquiry, complaint, feedback), providing insights into the nature of the customer's communication.
3. **SENTIMENT**: Indicates the sentiment expressed in the interaction (e.g., positive, negative, neutral), enabling sentiment analysis and customer satisfaction monitoring.
4. **FEEDBACK_SOURCE**: Specifies the source or channel through which the interaction was received (e.g., website, email, phone), providing insights into customer communication preferences.
5. **RESOLUTION_STATUS**: Indicates the status of the interaction resolution (e.g., pending, resolved), enabling tracking of issue resolution and customer follow-up.

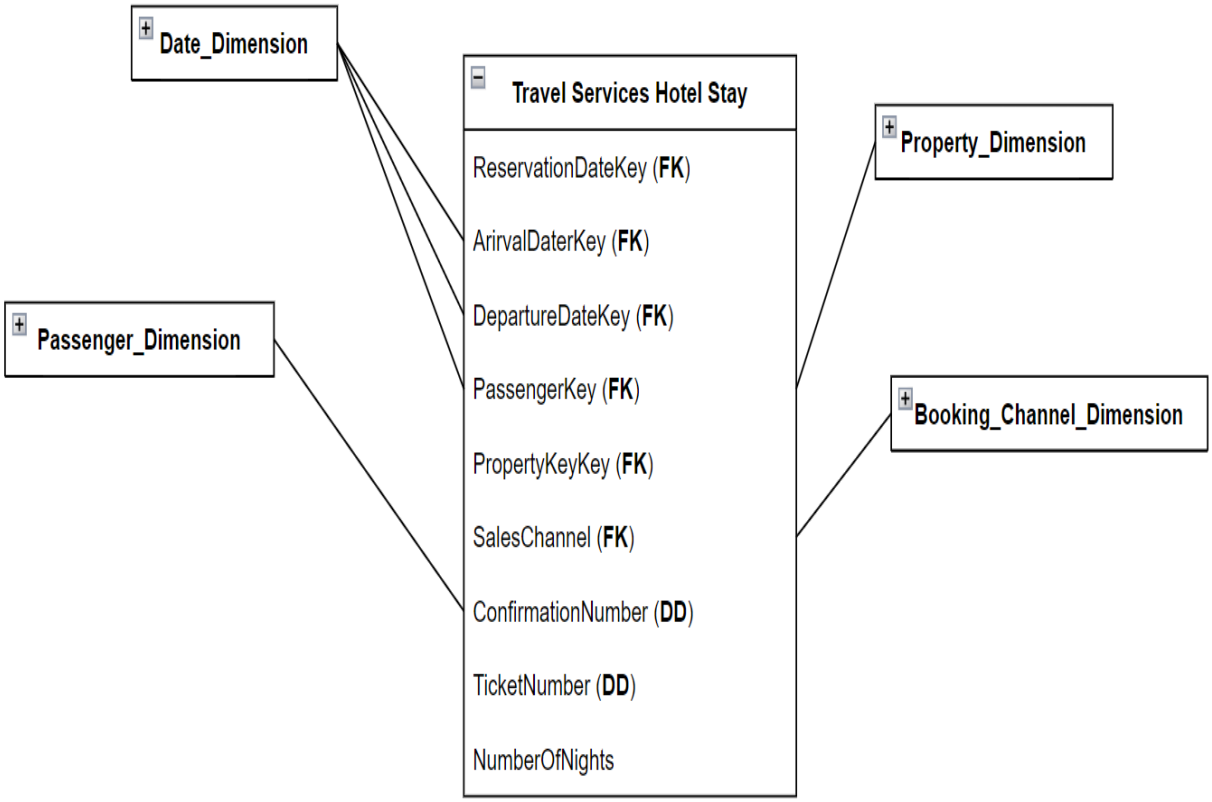
Significance and Analysis:

The Interaction Dimension plays a crucial role in analyzing customer feedback and understanding customer sentiment across different communication channels. By leveraging the attributes in this dimension, customer service teams and organizational stakeholders can gain valuable insights into customer satisfaction levels, service performance, and areas for improvement.

For example, analysis of interaction types and sentiment enables airlines to prioritize and address customer issues effectively, enhancing overall service quality and customer experience. Understanding feedback sources helps in optimizing communication channels and ensuring a seamless customer experience across all touchpoints. Tracking resolution statuses enables airlines to monitor issue resolution timelines and ensure timely response and follow-up with customers.

DATA MODELLING

Travel Services Hotel Stay:



The **travel_service_hotel_stay_fact** fact table captures detailed information about hotel stay reservations made by passengers. Each record in the fact table represents a single hotel reservation and contains key information such as reservation dates, passenger details, hotel property details, and sales channel information. The fact table is linked to various dimension tables, allowing for detailed analysis and reporting on hotel stay reservations.

Grain Level: The chosen grain level for the Travel Service Hotel Stay data mart is at the reservation level. This means that each record in the fact table represents a single hotel reservation made by a passenger. The decision to use this grain level was based on the

need to capture detailed information about each hotel stay reservation, including reservation dates, passenger details, property details, and sales channel information.

Business Process: The Travel Service Hotel Stay data mart focuses on capturing and analyzing hotel stay reservations made by passengers. This process involves passengers making hotel reservations through various sales channels for a specified number of nights. The data mart stores information about each reservation, including the reservation date, arrival date, departure date, passenger details, hotel property details, sales channel information, confirmation number, ticket number, and the number of nights booked.

Fact Table: The fact table in the Travel Service Hotel Stay data mart is named **travel_service_hotel_stay_fact** and contains the following columns:

1. ReservationDateKey (FK): Foreign key referencing the Date dimension table, representing the date the hotel reservation was made.
2. ArrivalDateKey (FK): Foreign key referencing the Date dimension table, representing the arrival date of the hotel stay.
3. DepartureDateKey (FK): Foreign key referencing the Date dimension table, representing the departure date of the hotel stay.
4. PassengerKey (FK): Foreign key referencing the Passenger dimension table, representing the passenger making the hotel reservation.
5. PropertyKey (FK): Foreign key referencing the Hotel Property dimension table, representing the hotel property where the reservation was made.
6. SalesChannel (FK): Foreign key referencing the Sales Channel dimension table, representing the sales channel through which the reservation was made.
7. ConfirmationNumber (DD): Degenerate dimension representing the confirmation number of the hotel reservation.
8. TicketNumber (DD): Degenerate dimension representing the ticket number associated with the hotel reservation.
9. NumberOfNights: The number of nights booked for the hotel stay.

Dimensions: The Travel Service Hotel Stay data mart includes the following dimensions:

1. Date Dimension: Contains detailed date information for the reservation date, arrival date, and departure date.
2. Passenger Dimension: Contains information about the passengers making the hotel reservations, including passenger details.

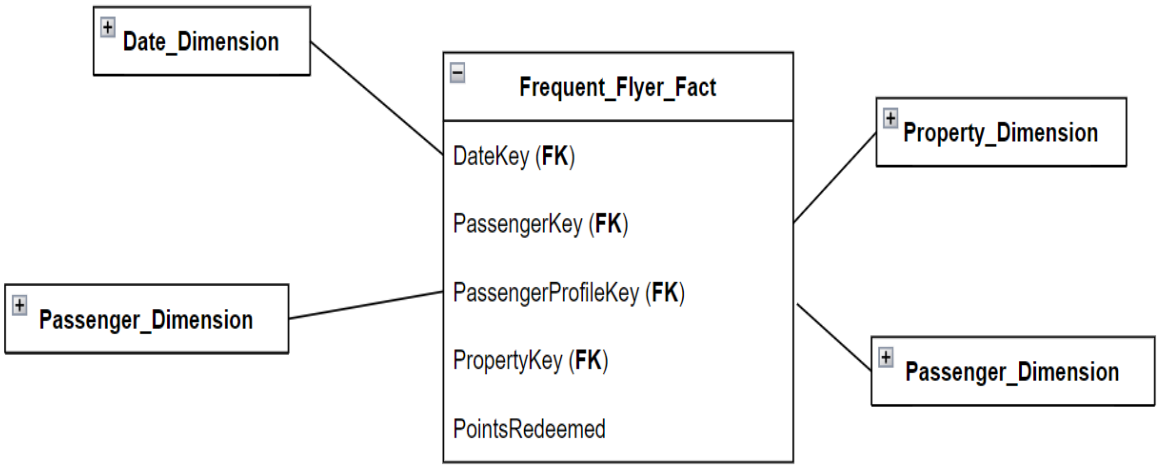
3. Hotel Property Dimension: Contains information about the hotel properties, including property details and location information.
4. Sales Channel Dimension: Contains information about the sales channels through which hotel reservations are made, including channel details.

Reporting and Visualization: Reporting and visualization for the Travel Service Hotel Stay data mart can include analysis of hotel stay trends, revenue generation, booking patterns, and customer segmentation. Tools like Tableau, Power BI, or QlikView can be used to create interactive dashboards and visualizations to analyze and report on the hotel stay reservations data.

1. **Revenue Analysis:** The fact table provides crucial data for analyzing revenue generated from hotel stay reservations. By aggregating revenue data at different levels (e.g., daily, weekly, monthly), the airline company can identify revenue trends, peak booking periods, and optimize pricing strategies.
2. **Customer Behavior Analysis:** Analyzing the fact table data allows the airline company to understand customer behavior regarding hotel reservations. By identifying patterns such as preferred booking channels, popular hotel properties, and booking lead times, the company can tailor marketing strategies to target specific customer segments more effectively.
3. **Operational Efficiency:** The fact table helps in monitoring and managing operational aspects of hotel reservations, such as booking volume, occupancy rates, and stay durations. This information is valuable for optimizing hotel inventory, staffing levels, and overall operational efficiency.
4. **Marketing and Sales Strategies:** Insights from the fact table can be used to develop targeted marketing campaigns and sales strategies. By understanding customer preferences and booking patterns, the airline company can offer personalized promotions and incentives to drive hotel booking conversions.
5. **Customer Satisfaction:** The fact table can be used to track customer satisfaction levels related to hotel stays. By analyzing feedback and complaints associated with hotel reservations, the company can identify areas for improvement and enhance overall customer satisfaction.

DATA MODELLING

Frequent Flyer Redeem Fact:



The Frequent_Flyer_Redemption_Fact table records the redemption of frequent flyer points by passengers for various rewards or benefits. Each record in the fact table represents a single redemption transaction and includes key information such as the date of redemption, the passenger involved, the passenger's profile, the property where the redemption occurred, and the number of points redeemed.

Grain Level Selection:

- The selected grain level for the Frequent_Flyer_Redemption_Fact table is at the redemption transaction level.
- This grain captures each instance of a passenger redeeming frequent flyer points for rewards or benefits.
- Other potential grain levels considered were:
    - Redemption at the passenger level, aggregating all redemptions for each passenger.

    - Redemption at the property level, aggregating all redemptions for each property.
- The chosen grain level aligns with the business need to analyze individual redemption transactions, providing detailed insights into the redemption behavior of passengers.

2. Business Process Description:

- The Frequent_Flyer_Redemption_Fact table captures the redemption of frequent flyer points by passengers for various rewards or benefits.
- The process involves passengers selecting and redeeming their accrued points for rewards such as flight upgrades, hotel stays, car rentals, or other promotional items.
- Each redemption transaction is recorded with key details including the date of redemption, the passenger involved, the passenger's profile, the property where the redemption occurred, and the number of points redeemed.

3. Importance of the Fact:

- **Reward Analysis:** Enables analysis of popular rewards, redemption trends, and helps tailor rewards to meet customer preferences.
- **Customer Loyalty:** Provides insights into customer loyalty and engagement with the frequent flyer program.
- **Operational Insights:** Aids in monitoring and managing the operational aspects of frequent flyer rewards.
- **Marketing Strategies:** Helps develop targeted marketing strategies based on redemption behavior.
- **Customer Satisfaction:** Tracks customer satisfaction levels related to the frequent flyer program.

4. Fact Table Details:

- **Fact Table Name:** Frequent_Flyer_Redemption_Fact
- **Columns:**
    - DateKey (FK): Date dimension foreign key for the date of redemption.
    - PassengerKey (FK): Passenger dimension foreign key for the passenger redeeming points.

- PassengerProfileKey (FK): Passenger profile dimension foreign key for additional passenger details.
- PropertyKey (FK): Property dimension foreign key for the property where redemption occurred.
- PointsRedeemed: Number of frequent flyer points redeemed in the transaction.

5. Associated Dimensions:

- **Date Dimension:** Provides details about the date of redemption.
- **Passenger Dimension:** Contains information about the passenger redeeming points.
- **Passenger Profile Dimension:** Includes additional details about the passenger.
- **Property Dimension:** Describes the property where the redemption occurred.

6. Reporting and Visualization:

- Reporting on the Frequent_Flyer_Redemption_Fact table can include:
  - Redemption trends over time.
  - Most popular rewards.
  - Analysis of points redeemed by passenger segments.
- Visualizations can include charts and graphs showing redemption patterns, reward popularity, and passenger engagement.