HUMBER INSTITUTE OF TECHNOLOGY

AND ADVANCED LEARNING

(HUMBER COLLEGE)

Group Assignment

COURSE: BIA 5000

TEAM: 6

SUBMITTED BY:

Last Name	First Name	Student Number
Boini	Avinash	N01581336
Panchal	Param	N01579822
Patel	Vrajkumar	N01581006
Prajapati	Pravina	N01579926
Singh	Gurmeher	n01581802
Verma	Neha	N01581181

SUBMITTED TO: Professor Yulia Kosarenko

SUBMISSION DATE: 2023-04-08

Contents

Step 1: Identify a business problem

- **Industry:** Aviation Industry
- **Company name:** Hum-Air
- **Company description:**
  - "Hum-Air" is an airline company that provides scheduled air transport for passengers to destinations inside Canada. Recently, Hum-Air is facing many baggage issues including lost, damaged, and delayed bags. Several passengers are criticizing Hum-Air online on different social media platforms about the airline's baggage problems, which is harming the airline's reputation. Hence, Hum-Air's main focus is to reduce its baggage problems and retain customer loyalty to maintain its brand value.
  - Hum-Air provides its booking services via Hum-Air mobile application. This app can be downloaded by anyone who has a valid Canadian phone number.
- **Business Problem:**
  - Reduction in customer loyalty because of baggage issues including delayed, damaged, and misplaced bags.

Step 2: Ask analytics questions

Analytic Questions:

Descriptive:

1. What percentage of baggage is lost/delayed/damaged during transportation?
   - This descriptive question will tell us which baggage issue is more dominant than the others and based on that solution strategies can be decided.
2. How many days does it take to resolve the baggage incident?
   - This will tell us the total efforts used to resolve incidents. We can analyze which category is taking longer time and effort.
3. What is the average cost of compensating customers for lost/delayed/damaged baggage?
   - This will give us insights into the cost incurred by the company for baggage issue compensations. Based on this, we can analyze how much we can save by reducing the baggage issue as we have to give fewer compensations to passengers in that scenario.
4. Which are the most vulnerable airports for baggage issues?
   - This will provide us with the most vulnerable airport locations. Based on this we can analyze why this particular airport is having more baggage issues. And based on that we can implement our solution strategy in those locations.
5. Are there any specific times of the year (e.g., high travel season, inclement weather) when baggage loss/delay/damage incidents increase?
   - This will give us information about the peak timings of the baggage issue. For example, in December passengers traveling through Hum-Air increased, and hence the baggage issues as well, so we can increase our staff or may provide self-checking to reduce our baggage delay at that particular time of the year.
6. What is the average processing time for each step during baggage processing?
   - This will provide us with the complete process cycle time for the bag to reach from source to destination. The baggage process will include different steps like below.
   - Baggage check-in.
   - Baggage sorting and routing.

- Baggage security screening.
- Baggage loading.
- Baggage unloading and reclaim.
- We can use this information to monitor and improve the efficiency of our baggage handling system and to identify and address bottlenecks or other issues that may be causing delays. Passengers might also be interested in knowing the average baggage processing time so they can plan their travel itinerary and avoid missing their flights.

7. What is the Growth/fall in incidents related to baggage issues in each category?
   - By analyzing data on baggage-related complaints over time, one could identify which categories of complaints are most common, whether there are any seasonal or cyclical trends, and whether the overall number of complaints has increased or decreased over the past five years.

Diagnostic:

1. Which issue is stronger as compared to others and what was the reason?
   - This information can be used to get the dominant issues among the others and we can analyze if any common factors are resulting in those issues. Based on this we can act accordingly.
2. Are there specific routes or airports where these incidents are more frequent and what was the reason?
   - This will tell us the vulnerable routes and airports and will also provide us with the reason for the same. For example, Let's suppose there is a large increase in baggage issues at Montreal Airport. We analyzed and found out that there is a staff shortage for handling passenger luggage hence resulting in baggage delays. So in that scenario, our solution strategy will include hiring more staff for baggage handling or developing an automated process that will require less staff.
3. Are there any common factors that contribute to baggage loss/delay/damage incidents (e.g., weather conditions, baggage handling equipment, human error, security screening, and connecting flights)?

    - This will give us all the common factors which contribute to our baggage issue and based on that solution strategy can be implemented

4. What is the processing time for handling baggage at each stage of the journey (e.g., Baggage check-in, Baggage sorting, and routing, Baggage security screening, Baggage loading, Baggage unloading, and reclaiming)? Which part is taking longer time and the reason for the same?
    - This will give us the process which is taking longer time and what are the common factors that are responsible for that delay. For example, we checked that most of the delay is happening in the check-in section at all the airports. And the reason for this is the slow process of manual check-in. To resolve this, we can implement self-check-in methods where passengers have to check in online through the portal and have to declare all the items in the bags and barcode tags will be provided which they have to stick on their bags. After that, they just have to drop their bags in the mentioned conveyor belts at the airport without any need to go into physical lines.

Predictive:

1. What will be the likelihood of an increase in customer ratings if we provide them with real-time updates on their baggage location/status?
    - This will provide us with the likelihood of customers remaining loyal to Hum-Air by using a communication strategy even after experiencing a problem with their baggage. By analyzing data on the number of customers who experienced baggage issues and whether they continued to use the airline's services, predictive analytics can be used to forecast future retention rates based on specific variables, such as the type of baggage issue, the duration of the issue, and the customer's satisfaction with the airline's response
2. What will be the likelihood of a decreased percentage of baggage issues, if we use RFID or GPS for baggage tracking and monitoring?
    - This information can be used to predict the impact of this technology in reducing or eliminating baggage problems.
3. What will be the likelihood of an increase in customer ratings if we provide good compensation value for baggage issues?

4. What will be the likelihood of a decrease in baggage issues if we use automated systems instead of the physical process?
    - For example, if we are using self-check-in software where users can generate the baggage id at their home one day prior and then can directly put the bag on the belt and it will be automatically checked in for further processing, what will be the reduction rate in the delay issues?

Prescriptive:

1. What value of compensation should be decided so that customers remain loyal even after the baggage issue?
    - For example, if analysis shows that customers who are offered better compensation for baggage issues are more likely to remain loyal to the airline, the airline can develop a compensation strategy to improve customer retention rates.
2. What will be the most effective communication strategies to inform customers about baggage loss/delay/damage incidents?
    - For example, if analysis shows that customers who are offered real-time updates regarding their bag's status are more likely to remain loyal to the airline, the airline can develop a communication strategy to improve customer retention rates.
3. What technologies should be used to automate the baggage process so that baggage issues related to delay can be minimized?
4. What technologies (RFID, GPS, and AIRTAG) should be used which will be in our budget and effective to track the bags which are misplaced or lost?

Step 3: Main data entities and ERD

The main entities and attributes which will be used in our system.

1. **User_Account**: This entity is used to collect user account-related data. It will store user login details, phone number, and last login details. Below are the Attributes of the User Account Entity:
    - User_ID: This will be the primary key in the user_account table and will identify each row uniquely.
    - User_name: This column is used to store user name details. This user name should be unique; hence a unique key constraint is applied to this column. If the username already exists it will not allow another user to enter the same username again.
    - Password: This attribute is used to store the password for the user account and will be encrypted.
    - Phone_Number: This column is used to store phone number details
    - Last_Login_Date: This stores the last login date.
    - Created_Date: Used to store the creation date of the respective account.

2. **Passenger:** This entity is used to store all the passenger-related data like name passenger type, class, age, etc. Data will be inserted into the passenger table when a user from the user's account will book the flight tickets. Users can book multiple tickets (for example for his/her parents and children) hence one user account can have multiple passenger entries based on their booking. If a user only created the user account and hasn't booked any flight yet then there will be no entries in the passenger table for the respective user account.
    - Passenger_ID: This is the primary key for the table passenger; hence each passenger row has its unique passenger id.
    - User_ID: This is a foreign key that will be used to link the user_account table with the passenger table. One user can book multiple tickets hence one user_id can be linked to more than one passenger_id.
    - Name: This attribute is used to store the passenger's full name.

- Passenger_type: This attribute is used to store the type of passenger. Three types of passengers can be selected via the Hum-Air mobile application while booking flight tickets.
    - I. Student: The passenger will be considered a Student if he/she has a valid student ID from any Canadian educational institute.
    - II. Adult: The passenger will be considered an Adult if their age is more than 18 years.
    - III. Senior Citizen: The passenger will be considered a Senior Citizen if their age is more than 65 years.
- Class: This stores class information-Business, Premium Economy, Economy
- Gender: This will be used to store the gender of passengers.
- Created_date: This attribute is used to store the creation date of the record when the passenger data was loaded into the passenger table.

3. **Flight_Ticket**: This entity is used to store flight details and entries will be created in this table when a flight is booked successfully. This will have details like ticket number, source details, destination details, the scheduled and actual time of arrival and departure, etc.
    - Ticket_id: This is the primary key for the table Flight_Ticket and will uniquely identify each row in the table.
    - Passenger_id: This is the foreign key for the table and is used to connect the passenger table with the flight ticket table.
    - Source_location: This is used to store the source location which is the airport from which the flight departed.
    - Destination: This is used to store the destination location which is the airport where the flight arrived.
    - Scheduled_departure: This stores the scheduled departure time and date
    - Scheduled_arrival: This stores the scheduled arrival time and date
    - Actual_departure: This stores the actual departure time and date
    - Actual_arrival: This stores the actual arrival time and date.

4. **Baggage:** This entity is used to store baggage details and one flight ticket can be associated with one baggage id only. The entries will be created on this system when passenger drop their bags at the airport and baggage id is generated and attached to their bags.
    - Baggage_id: This is used to store the primary key values for the baggage table and all ids will be unique.
    - Ticket_id: This is used to link the baggage table with the flight ticket table and this will store the foreign key values. One ticket can have multiple bags hence multiple baggage IDs will be linked to one ticket id.
    - Bag_type-This is used to store the bag type like check-in or carry-on.
    - Bag_colour: This is used to store the color of the bag.
    - Bag_material: This is used to store bag material like hard or soft.
    - Check_status: This is used to store the check status, whether that particular bag cleared the check status or not. If the bag includes any prohibited items then its check status will be 'F' and the bag will not be transported and will be kept at the source airport location.
    - Weight: stores the weight measurement of the particular bag.
    - Extra_luggage_flag: This flag is used to identify bags that are extra other than the provided free luggage by Hum-air. For extra luggage bags, a passenger has to pay an extra amount.
    - Created_date: This is used to store the creation date of the record.

5. **Baggage_Process**: This entity is used to capture the actual time for a bag to reach from source to destination. This table will include separate entries for one bag including details of each process and how much time it took to complete one process and when was a process started and completed.
    - Process_id: This is the primary key of the table.
    - Baggage_id: This is the foreign key of the table which will be used to link the baggage_process table with the Baggage table.
    - Process_type: This is used to define the multiple types of processes for which data is being recorded. Different types of processes included are Baggage check-in,

Baggage sorting and routing, Baggage security screening, Baggage loading, Baggage unloading and reclaim

- Process_start_date_time: stores process starts to date and time.
- Process_end_date_time: stores process end date and time.
- Location: stores location details where the process has been executed.

6. **Baggage_Incident:** This entity is used to store incident details related to baggage issues. For every bag which is delayed, lost, or misplaced an incident is created in the Baggage_Incident table. It can be either created by the user via the Hum-Air app, where the user can report their baggage issue, or it can be created by the Hum-Air to track their baggage issues for future analysis.
    - Incident_id: It stores the primary key.
    - Baggage_id: It stores the foreign key used to link a particular incident with the respective bag for which the incident is reported.
    - Review_id: It stores foreign key values which are used to link the particular incident with the review table where reviews of the customers are captured related to baggage incident resolution.
    - Incident_type: It stores the type of baggage issue: It contains three types delay, damaged, and misplaced.
    - Incident_status: It stores the status of the incident: open, in process, resolved.
    - Description: stores the description of the issue.
    - Reason: stores the reason comments, what was the actual reason for the delay/damage/misplacement of the bag.
    - Compensation_flag: This flag is 'T' when a user is eligible for compensation. This flag is 'F' when a user is not eligible for compensation
    - Created_by: This field is used to store the incident creation details and will have only two values.
        I. One will be 'Created_by =Hum-air'. This means that the incident is created by hum-air for tracking purposes and future analysis related to baggage issues.

II. Other will be ‘Created_by = User’. This means that the incident is created by a user and hence needs to be looked into on high priority.

- Created_date: This is used to store the creation date of the incident.
- Resolution_date: This is used to store the resolution date of the incident.

7. **Compensation:** This entity is used to store the compensation details related to a particular incident. Compensation can be accepted and declined based on the hum-air compensation policies.
    - Compensation_ID: This attribute stores the primary key of the table.
    - Incident_Id: This stores the foreign key of the table which will be used to link the compensation table and Baggage_incident table.
    - Compensation_type: This stores the compensation types like baggage refund, delay expenses, and damage expenses.
    - Compensation_approval_status: This stores the compensation approval status. Compensation requests can be approved and declined based on the compensation policies of the airline.
    - Amount: This stores the compensation amount in Canadian dollars.
    - Created_Date: This stores the creation date of the compensation record.

8. **Feedback:** This table is used to collect user reviews about baggage incident resolution. This will indicate how many customers are satisfied with the current baggage incident resolution.
    - Review_id: This stores the primary key values for the table.
    - Reviewer_name: This stores the reviewer name which will be the passenger name who is providing a review for the respective baggage incident resolution.
    - Rating: This stores the rating from the passenger or in this case we can say that reviewer. 5 represents the best and 1 represents the worst experience.
    - Review_description: This stores the review description provided by the passenger or reviewer.

Relationship Between the entities:

- **User_Account and Passenger:** This is zero to many relationship. One user can book tickets for zero or multiple passengers but one passenger can be linked to only one user account.
- **Passenger and Flight_Ticket**: This is one to many relationship. One passenger can have one or multiple tickets for different locations but one ticket id can have one passenger only.

- **Flight_Tickets and baggage:** This is zero to many relationship. One flight ticket can be associated with zero or multiple baggage IDs based on the number of bags that one passenger is carrying during their travel. But one baggage id cannot be associated with multiple ticket IDs and hence can only be linked with one and only one ticket ID.

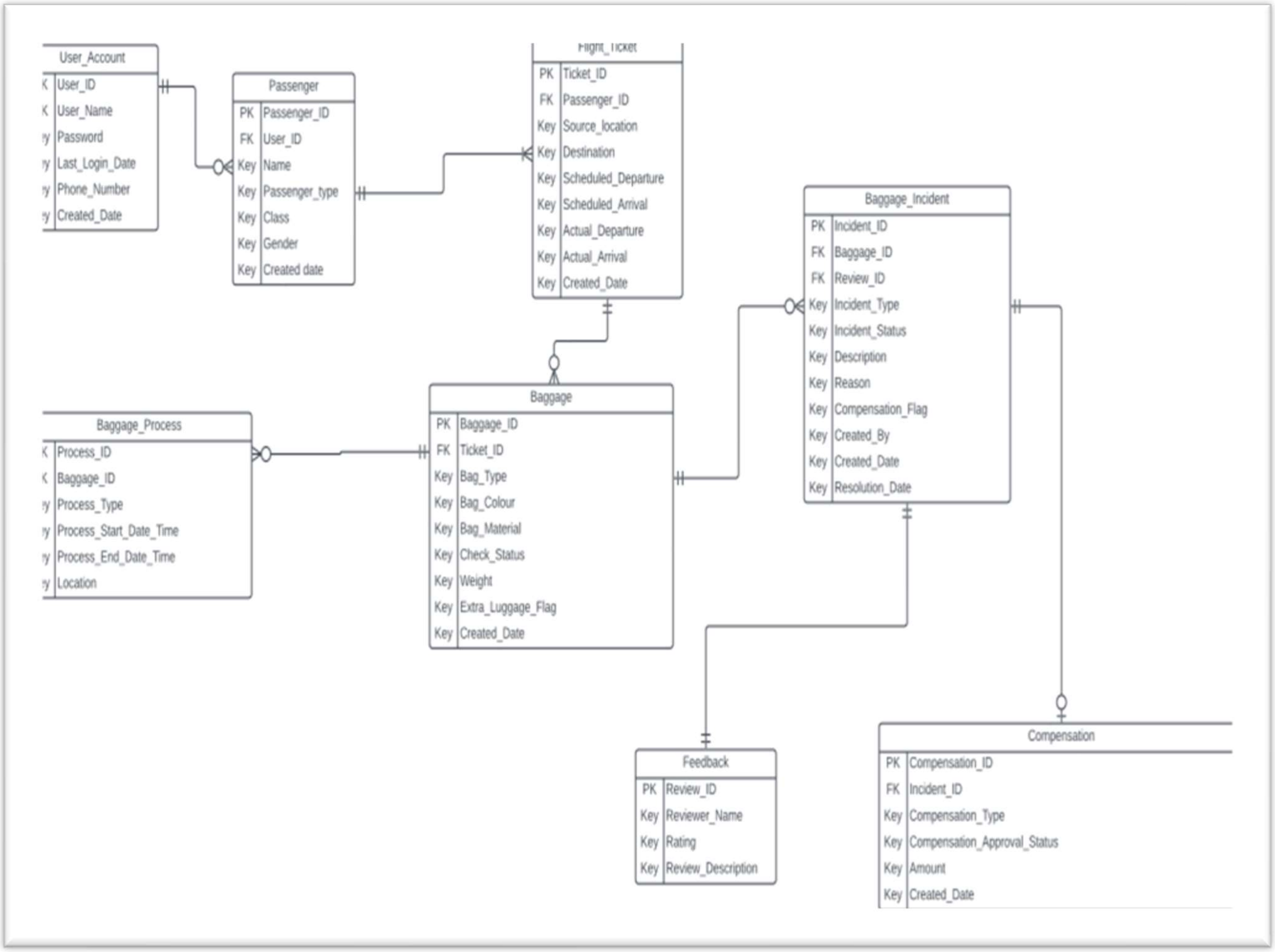
- **Baggage and Baggage_Process:** This is zero to many relationship. If baggage does not clear the check status then it will not be processed hence it will have zero corresponding records in the Baggage_process table. One process id can only be linked to one and only one baggage_id.

- **Baggage and Bagagge_Incident:** This is zero to many relationships. One baggage can have zero or multiple incidents. For example, one bag can be delayed and damaged at the same time and hence for one particular bag two incident entries will be created. One for the delay and the other for damage. One incident can be related to one and only one baggage and cannot be associated with multiple bags.

- **Baggage_Incident and Compensation:** This is zero to one relationship. One incident can be related to zero or one compensation record. And one compensation can be linked to one and only one incident number.

- **Baggage_Incident and Feedback:** This is one to one relationship. One feedback record can be associated with one and only one incident. Entities and Attributes:

ERD Diagram:


User_Account
K User_ID
K User_Name
y Password
y Last_Login_Date
y Phone_Number
y Created_Date
Passenger
PK Passenger_ID
FK User_ID
Key Name
Key Passenger_type
Key Class
Key Gender
Key Created date
Flight_Ticket
PK Ticket_ID
FK Passenger_ID
Key Source_location
Key Destination
Key Scheduled_Departure
Key Scheduled_Arrival
Key Actual_Departure
Key Actual_Arrival
Key Created_Date
Baggage_Incident
PK Incident_ID
FK Baggage_ID
FK Review_ID
Key Incident_Type
Key Incident_Status
Key Description
Key Reason
Key Compensation_Flag
Key Created_By
Key Created_Date
Key Resolution_Date
Baggage_Process
K Process_ID
K Baggage_ID
y Process_Type
y Process_Start_Date_Time
y Process_End_Date_Time
y Location
Baggage
PK Baggage_ID
FK Ticket_ID
Key Bag_Type
Key Bag_Colour
Key Bag_Material
Key Check_Status
Key Weight
Key Extra_Luggage_Flag
Key Created_Date
Feedback
PK Review_ID
Key Reviewer_Name
Key Rating
Key Review_Description
Compensation
PK Compensation_ID
FK Incident_ID
Key Compensation_Type
Key Compensation_Approval_Status
Key Amount
Key Created_Date

Step 4: Identify Systems of Record

The main system of records is explained below:

1. **Customer relationship management system (CRM).**
   - Applicable to: Existing passengers, New customers.
   - Purpose: This SOR will store the data of customers and maintain their credentials, enabling airlines to effectively manage their customer service inquiries. An airline agent can enter a customer's request for a service, such as a seat upgrade, a special meal, or a change in flight schedule, in the SOR.
   - Source of data: customer data from Hum-air Application.
   - Entity: User Account, Passenger details

2. **Flight booking system (FBS):**
   - Applicable to: Passengers, Airline agents.
   - Purpose: The flight booking system is a consolidated software platform that keeps track of all the information about bookings for flights, including customer data, aircraft schedules, seat availability, costs, payment information, and booking history.
   - Source of data: HumAir booking system
   - Entity: Flight Ticket

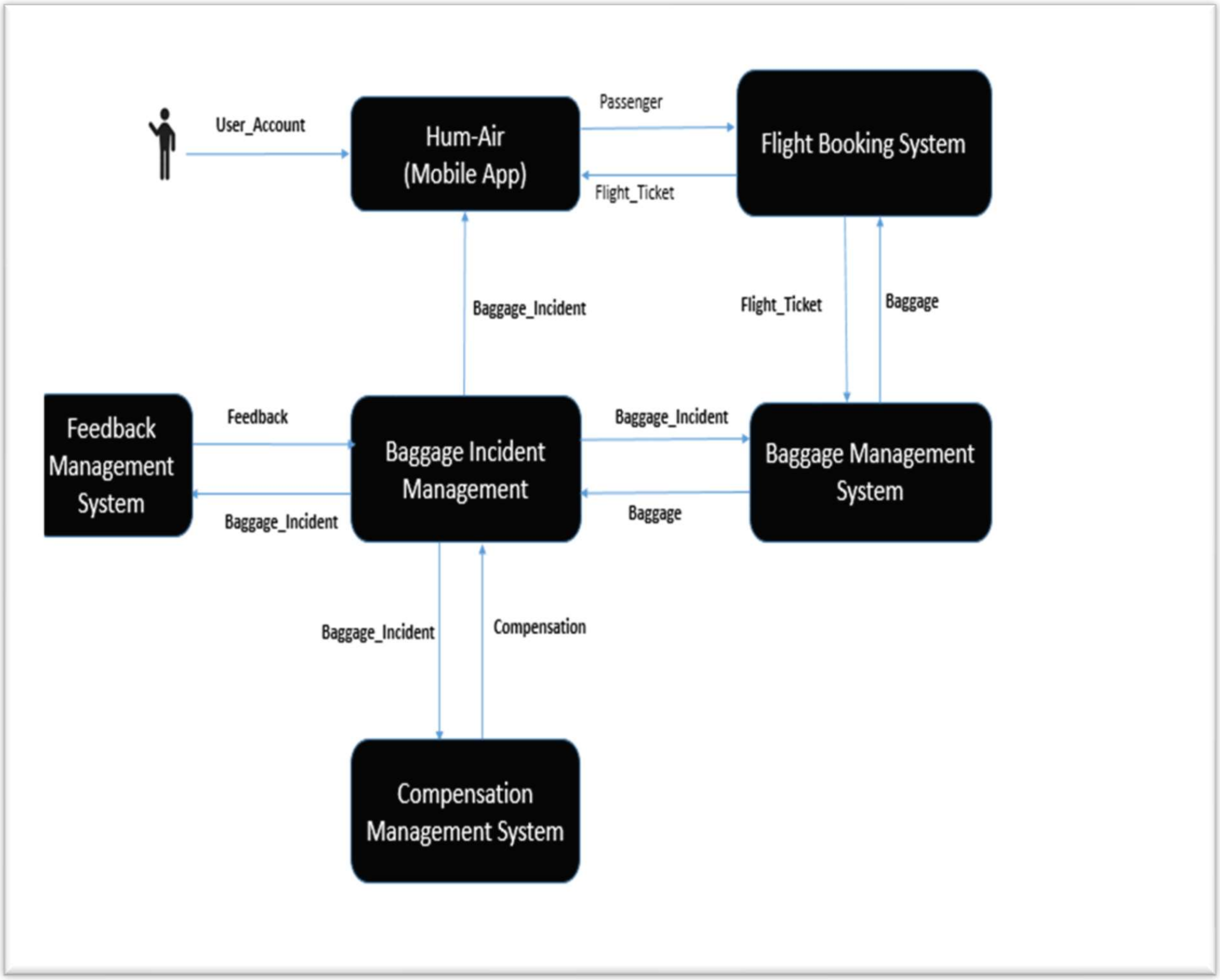
3. **Baggage management system (BMS):**
   - User: Baggage staff, Airline authority, and Users (can get their real-time baggage location).
   - Purpose: the BMS provides accurate, up-to-date, and complete information about each piece of baggage throughout its journey. This information is critical for the airline's operations, as it helps to ensure that baggage is loaded onto the correct flight, transferred between flights as needed, and ultimately delivered to the correct passenger at the final destination. The BMS also helps the airline to comply with various regulations related to baggage handling and security.
   - Source of data: Booking details
   - Entity: Baggage & Baggage Process

4. **Baggage Incident Management System (BIMS)**
    - User: passenger, customer support team, baggage department
    - Purpose: The BIMS provides accurate, up-to-date, and complete information about each baggage incident, including the reason for the incident, the location of the baggage, and any actions taken to resolve the incident. The ability to fix the situation as promptly and effectively as possible, avoiding any inconvenience or disturbance to the passenger's travel plans, depends on this information, which is crucial for the airline's operations.
    - Source of data: Hum-Air app, baggage details
    - Entity: Baggage Incident

5. **Compensation management system (CMS):**
    - User: Management, Accounts
    - Purpose: The CMS provides accurate, up-to-date, and complete information about each compensation claim, including the reason for the claim, the amount of compensation owed, and any actions taken to resolve the claim. This information is critical for the airline's operations, as it helps to ensure that compensation claims are processed quickly and efficiently, minimizing any negative impact on the airline's reputation and customer satisfaction.
    - Source of data: Baggage Incidents
    - Entity: Compensation

Step 5: Systems of Record - context model

Context Model:



Step 6: Data Preparation and Data Wrangling Activities

- Data Preparation and Data Wrangling activities such as data cleansing, data profiling and will be performed to ensure that the data generated is good enough to provide accurate results to perform analytics.
- The data preparation and data wrangling activities mentioned above will impact the data entities that are incomplete, inaccurate, unformatted, and useless.

Data Cleaning

- The goal of this activity is to perform certain checks on the data to ensure that the quality of data is good, data consistency is maintained, and the data is clean.

  The checks included for Data cleaning are:

1. There should be no null entries in the user account table. For example, if there is a record in the ‘User_Account’ table which only includes User ID and no other values are there in that particular record then we will consider that record as a bad record and we will delete that record as a part of our data cleaning process.
   - **Before Data Cleaning:** There is a null record present in the User_Account table for User_ID=’U0000’.

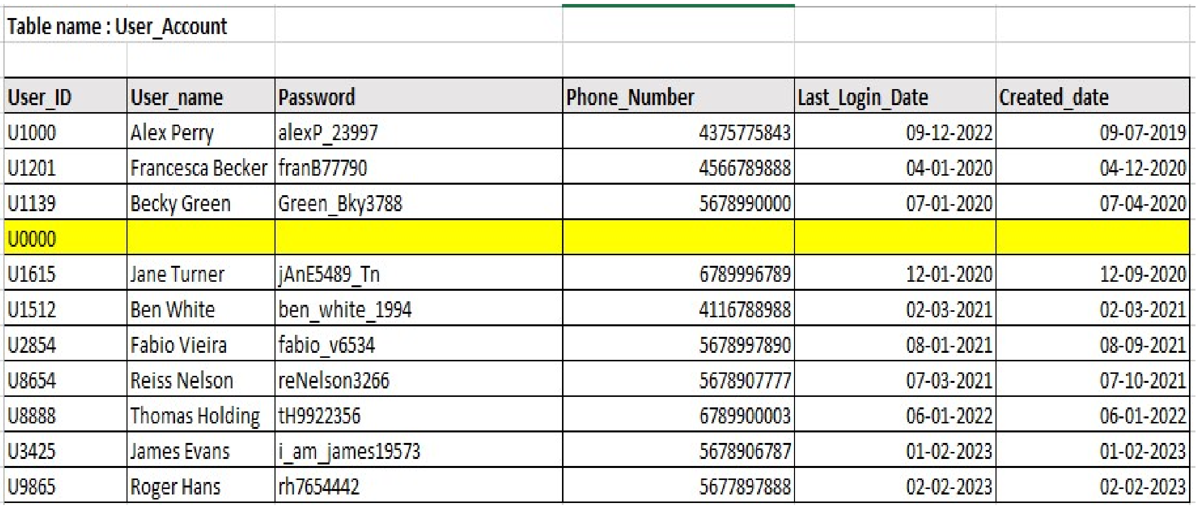


Table name : User_Account

User_ID	User_name	Password	Phone_Number	Last_Login_Date	Created_date
U1000	Alex Perry	alexP_23997	4375775843	09-12-2022	09-07-2019
U1201	Francesca Becker	franB77790	4566789888	04-01-2020	04-12-2020
U1139	Becky Green	Green_Bky3788	5678990000	07-01-2020	07-04-2020
U0000					
U1615	Jane Turner	jAnE5489_Tn	6789996789	12-01-2020	12-09-2020
U1512	Ben White	ben_white_1994	4116788988	02-03-2021	02-03-2021
U2854	Fabio Vieira	fabio_v6534	5678997890	08-01-2021	08-09-2021
U8654	Reiss Nelson	reNelson3266	5678907777	07-03-2021	07-10-2021
U8888	Thomas Holding	tH9922356	6789900003	06-01-2022	06-01-2022
U3425	James Evans	i_am_james19573	5678906787	01-02-2023	01-02-2023
U9865	Roger Hans	rh7654442	5677897888	02-02-2023	02-02-2023

- **After Data Cleaning:** We identified that User_ID =' U0000' has null values for all other attributes. Hence it is a bad record and was deleted as a part of the data cleaning process.

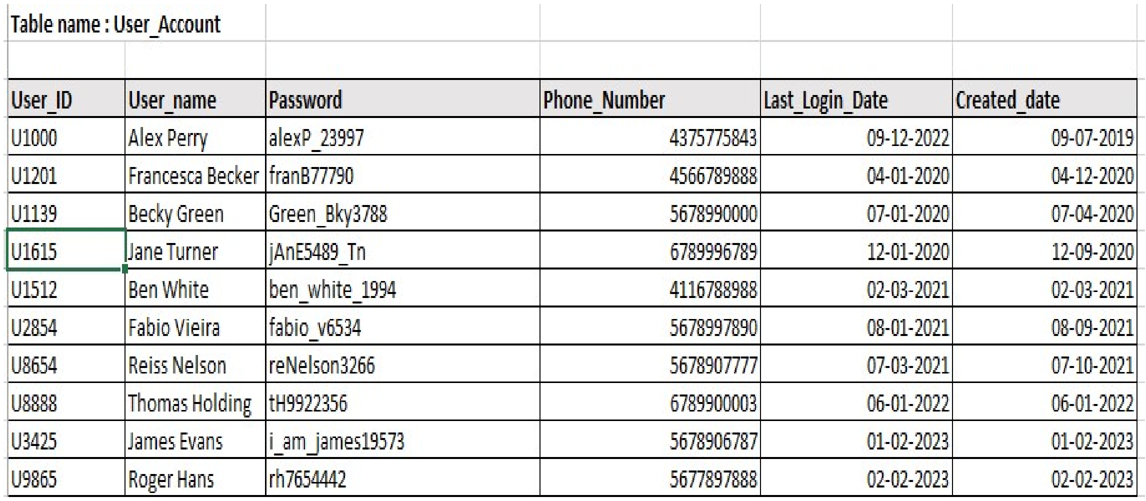


Table name : User_Account

User_ID	User_name	Password	Phone_Number	Last_Login_Date	Created_date
U1000	Alex Perry	alexP_23997	4375775843	09-12-2022	09-07-2019
U1201	Francesca Becker	franB77790	4566789888	04-01-2020	04-12-2020
U1139	Becky Green	Green_Bky3788	5678990000	07-01-2020	07-04-2020
U1615	Jane Turner	jAnE5489_Tn	6789996789	12-01-2020	12-09-2020
U1512	Ben White	ben_white_1994	4116788988	02-03-2021	02-03-2021
U2854	Fabio Vieira	fabio_v6534	5678997890	08-01-2021	08-09-2021
U8654	Reiss Nelson	reNelson3266	5678907777	07-03-2021	07-10-2021
U8888	Thomas Holding	tH9922356	6789900003	06-01-2022	06-01-2022
U3425	James Evans	i_am_james19573	5678906787	01-02-2023	01-02-2023
U9865	Roger Hans	rh7654442	5677897888	02-02-2023	02-02-2023

2. The User_ID in the User_Account table should always start with the 'U' letter. Only then it will be considered valid a record.
   - **Before Data Cleaning:** There is a record in the User_Account table having User_ID =' P1512'

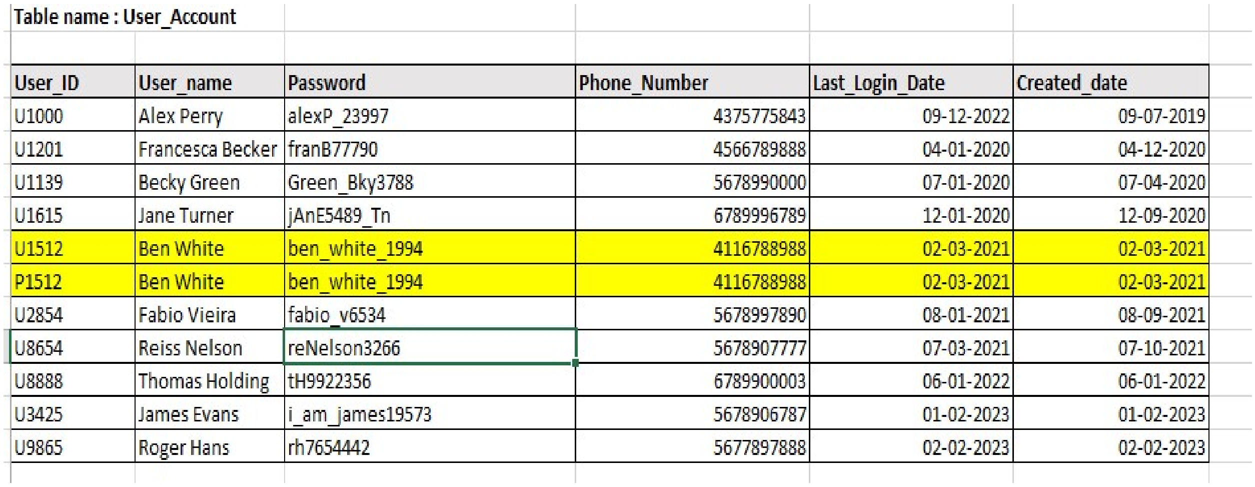


Table name : User_Account

User_ID	User_name	Password	Phone_Number	Last_Login_Date	Created_date
U1000	Alex Perry	alexP_23997	4375775843	09-12-2022	09-07-2019
U1201	Francesca Becker	franB77790	4566789888	04-01-2020	04-12-2020
U1139	Becky Green	Green_Bky3788	5678990000	07-01-2020	07-04-2020
U1615	Jane Turner	jAnE5489_Tn	6789996789	12-01-2020	12-09-2020
U1512	Ben White	ben_white_1994	4116788988	02-03-2021	02-03-2021
P1512	Ben White	ben_white_1994	4116788988	02-03-2021	02-03-2021
U2854	Fabio Vieira	fabio_v6534	5678997890	08-01-2021	08-09-2021
U8654	Reiss Nelson	reNelson3266	5678907777	07-03-2021	07-10-2021
U8888	Thomas Holding	tH9922356	6789900003	06-01-2022	06-01-2022
U3425	James Evans	i_am_james19573	5678906787	01-02-2023	01-02-2023
U9865	Roger Hans	rh7654442	5677897888	02-02-2023	02-02-2023

- **After Data Cleaning:** We identified that User_ID =' P1512' is starting from a letter other than 'U'. Also, the same values are present for record 'U1512. Hence, 'P1512' is not a valid record and was deleted as a part of the data cleaning process.

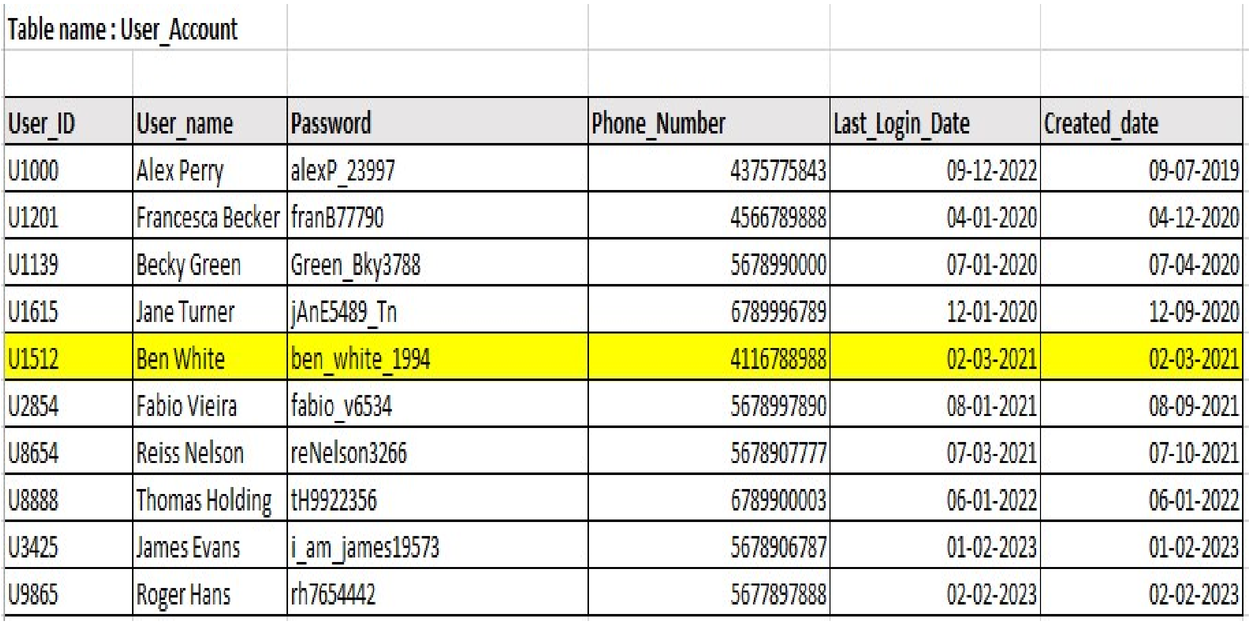


Table name : User_Account

User_ID	User_name	Password	Phone_Number	Last_Login_Date	Created_date
U1000	Alex Perry	alexP_23997	4375775843	09-12-2022	09-07-2019
U1201	Francesca Becker	franB77790	4566789888	04-01-2020	04-12-2020
U1139	Becky Green	Green_Bky3788	5678990000	07-01-2020	07-04-2020
U1615	Jane Turner	jAnE5489_Tn	6789996789	12-01-2020	12-09-2020
U1512	Ben White	ben_white_1994	4116788988	02-03-2021	02-03-2021
U2854	Fabio Vieira	fabio_v6534	5678997890	08-01-2021	08-09-2021
U8654	Reiss Nelson	reNelson3266	5678907777	07-03-2021	07-10-2021
U8888	Thomas Holding	tH9922356	6789900003	06-01-2022	06-01-2022
U3425	James Evans	i_am_james19573	5678906787	01-02-2023	01-02-2023
U9865	Roger Hans	rh7654442	5677897888	02-02-2023	02-02-2023

3. If duplicate records are having the same phone number for two or more user entries and then the latest one will be considered a valid one and the previous ones will be deleted.
   - **Before Data Cleaning:** We have 5 duplicate records in the User_Account table having different User_IDs.

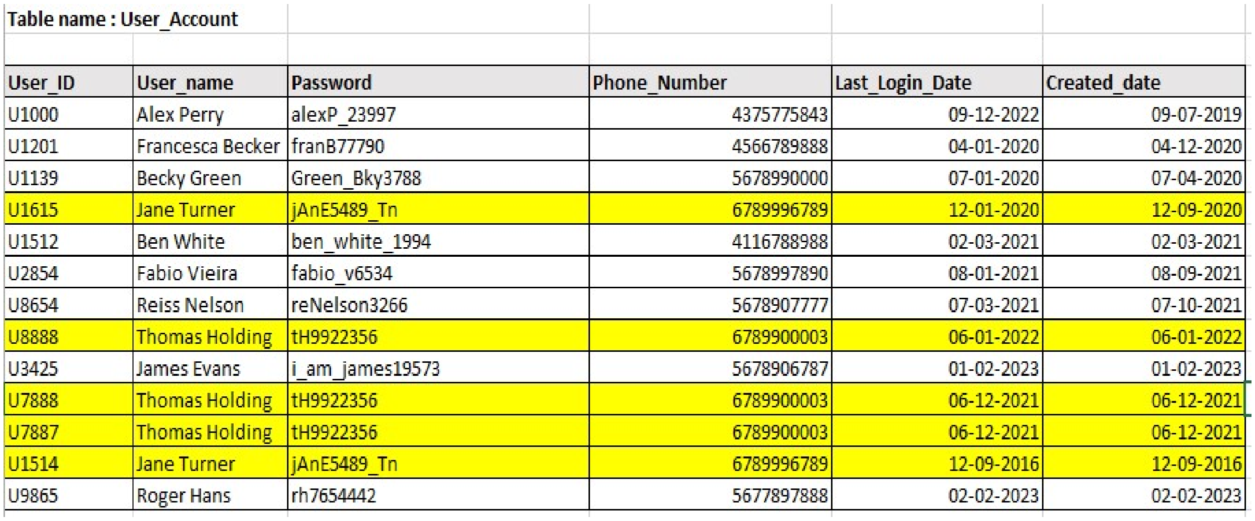


Table name : User_Account

User_ID	User_name	Password	Phone_Number	Last_Login_Date	Created_date
U1000	Alex Perry	alexP_23997	4375775843	09-12-2022	09-07-2019
U1201	Francesca Becker	franB77790	4566789888	04-01-2020	04-12-2020
U1139	Becky Green	Green_Bky3788	5678990000	07-01-2020	07-04-2020
U1615	Jane Turner	jAnE5489_Tn	6789996789	12-01-2020	12-09-2020
U1512	Ben White	ben_white_1994	4116788988	02-03-2021	02-03-2021
U2854	Fabio Vieira	fabio_v6534	5678997890	08-01-2021	08-09-2021
U8654	Reiss Nelson	reNelson3266	5678907777	07-03-2021	07-10-2021
U8888	Thomas Holding	tH9922356	6789900003	06-01-2022	06-01-2022
U3425	James Evans	i_am_james19573	5678906787	01-02-2023	01-02-2023
U7888	Thomas Holding	tH9922356	6789900003	06-12-2021	06-12-2021
U7887	Thomas Holding	tH9922356	6789900003	06-12-2021	06-12-2021
U1514	Jane Turner	jAnE5489_Tn	6789996789	12-09-2016	12-09-2016
U9865	Roger Hans	rh7654442	5677897888	02-02-2023	02-02-2023

- **After Data Cleaning:** We identified that 5 records are duplicates. All records have the same information; hence we kept the records which had the latest created_date and all other records were deleted as a part of data the cleaning process.

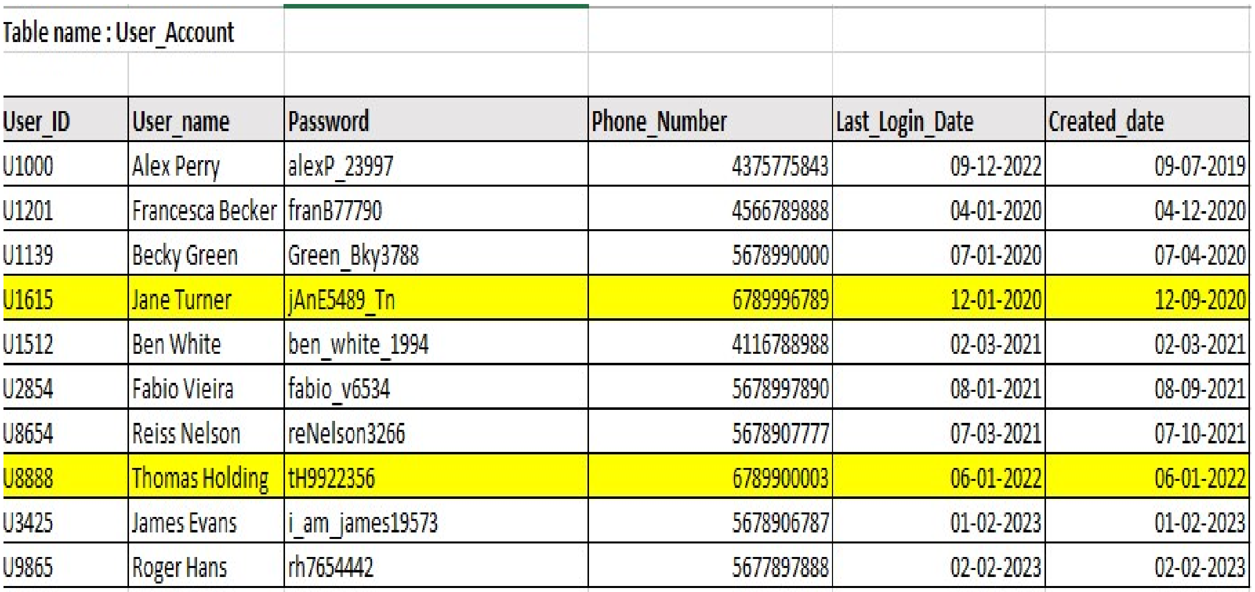


Table name : User_Account

User_ID	User_name	Password	Phone_Number	Last_Login_Date	Created_date
U1000	Alex Perry	alexP_23997	4375775843	09-12-2022	09-07-2019
U1201	Francesca Becker	franB77790	4566789888	04-01-2020	04-12-2020
U1139	Becky Green	Green_Bky3788	5678990000	07-01-2020	07-04-2020
U1615	Jane Turner	jAnE5489_Tn	6789996789	12-01-2020	12-09-2020
U1512	Ben White	ben_white_1994	4116788988	02-03-2021	02-03-2021
U2854	Fabio Vieira	fabio_v6534	5678997890	08-01-2021	08-09-2021
U8654	Reiss Nelson	reNelson3266	5678907777	07-03-2021	07-10-2021
U8888	Thomas Holding	tH9922356	6789900003	06-01-2022	06-01-2022
U3425	James Evans	i_am_james19573	5678906787	01-02-2023	01-02-2023
U9865	Roger Hans	rh7654442	5677897888	02-02-2023	02-02-2023

Data Profiling

- The main purpose of data profiling is to understand the structure, content, relationships, and quality of a dataset. Data profiling is a process of examining and analyzing data from various sources to understand the data better, identify data quality issues, and detect patterns and relationships within the data.
- Before data profiling, issues were there in data in the flight details-related table. In the Scheduled Departure & Scheduled Arrival columns format of the DateTimes not consistent with the rest of the data.

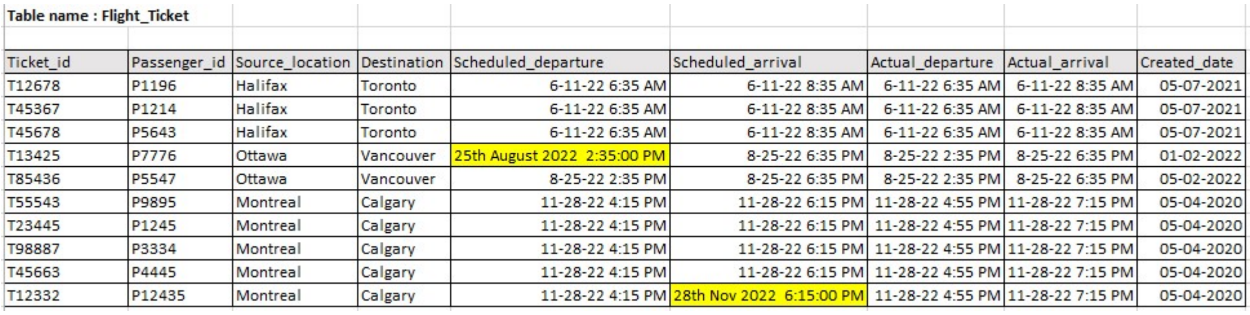


Table name : Flight_Ticket

Ticket_id	Passenger_id	Source_location	Destination	Scheduled_departure	Scheduled_arrival	Actual_departure	Actual_arrival	Created_date
T12678	P1196	Halifax	Toronto	6-11-22 6:35 AM	6-11-22 8:35 AM	6-11-22 6:35 AM	6-11-22 8:35 AM	05-07-2021
T45367	P1214	Halifax	Toronto	6-11-22 6:35 AM	6-11-22 8:35 AM	6-11-22 6:35 AM	6-11-22 8:35 AM	05-07-2021
T45678	P5643	Halifax	Toronto	6-11-22 6:35 AM	6-11-22 8:35 AM	6-11-22 6:35 AM	6-11-22 8:35 AM	05-07-2021
T13425	P7776	Ottawa	Vancouver	25th August 2022 2:35:00 PM	8-25-22 6:35 PM	8-25-22 2:35 PM	8-25-22 6:35 PM	01-02-2022
T85436	P5547	Ottawa	Vancouver	8-25-22 2:35 PM	8-25-22 6:35 PM	8-25-22 2:35 PM	8-25-22 6:35 PM	05-02-2022
T55543	P9895	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T23445	P1245	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T98887	P3334	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T45663	P4445	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T12332	P12435	Montreal	Calgary	11-28-22 4:15 PM	28th Nov 2022 6:15:00 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020

- After data profiling, the data format is consistent with the rest of the data.

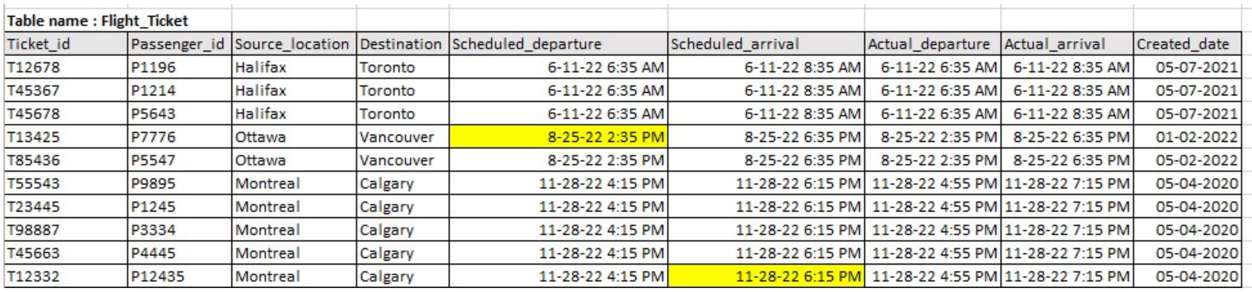


Table name : Flight_Ticket

Ticket_id	Passenger_id	Source_location	Destination	Scheduled_departure	Scheduled_arrival	Actual_departure	Actual_arrival	Created_date
T12678	P1196	Halifax	Toronto	6-11-22 6:35 AM	6-11-22 8:35 AM	6-11-22 6:35 AM	6-11-22 8:35 AM	05-07-2021
T45367	P1214	Halifax	Toronto	6-11-22 6:35 AM	6-11-22 8:35 AM	6-11-22 6:35 AM	6-11-22 8:35 AM	05-07-2021
T45678	P5643	Halifax	Toronto	6-11-22 6:35 AM	6-11-22 8:35 AM	6-11-22 6:35 AM	6-11-22 8:35 AM	05-07-2021
T13425	P7776	Ottawa	Vancouver	8-25-22 2:35 PM	8-25-22 6:35 PM	8-25-22 2:35 PM	8-25-22 6:35 PM	01-02-2022
T85436	P5547	Ottawa	Vancouver	8-25-22 2:35 PM	8-25-22 6:35 PM	8-25-22 2:35 PM	8-25-22 6:35 PM	05-02-2022
T55543	P9895	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T23445	P1245	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T98887	P3334	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T45663	P4445	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020
T12332	P12435	Montreal	Calgary	11-28-22 4:15 PM	11-28-22 6:15 PM	11-28-22 4:55 PM	11-28-22 7:15 PM	05-04-2020

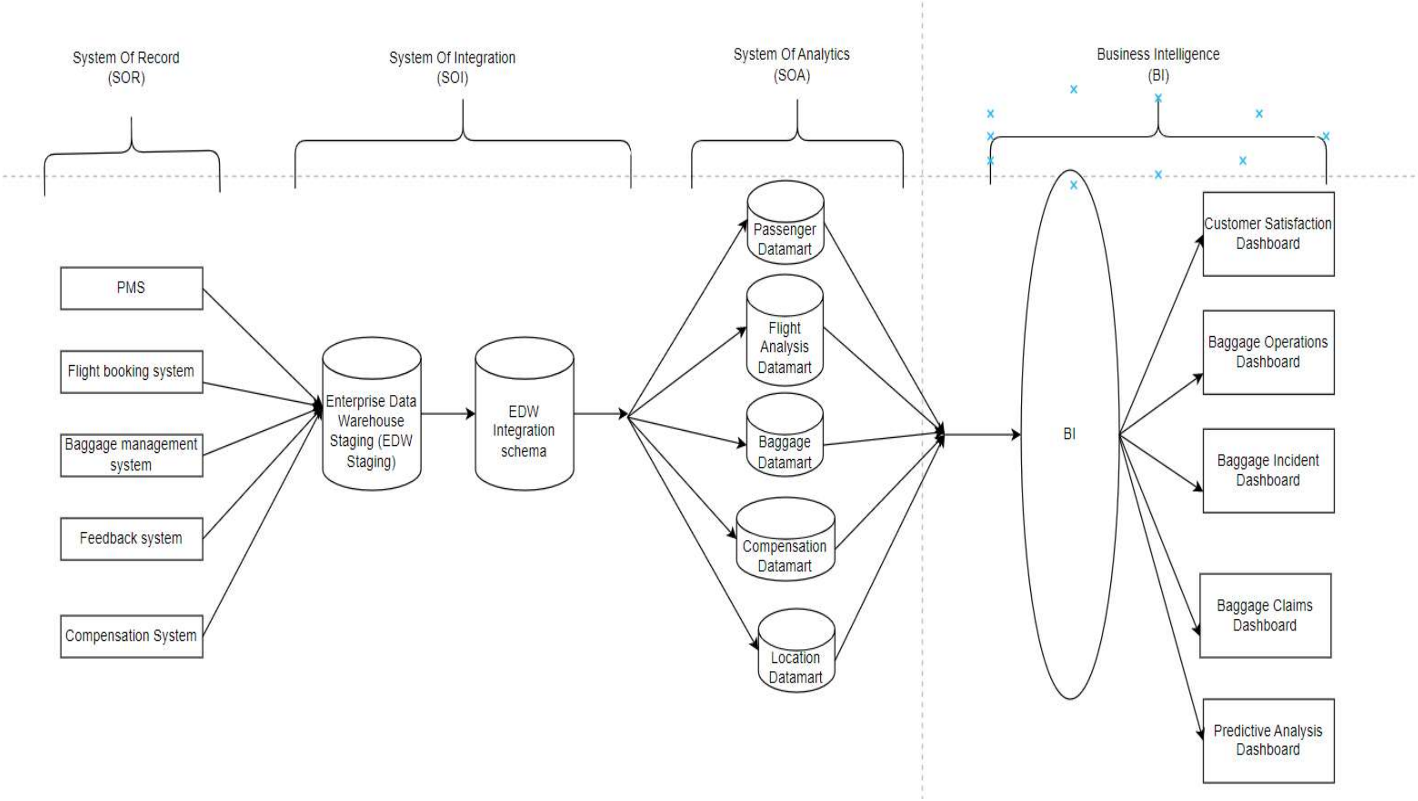
Step 7: Revise data entities and SORs

- After we agreed upon the entities, at a later stage we discovered that the feedback table is very important for our business. Because user feedback is very crucial for Hum Air. Based on the feedback we can be able to make modifications to our system. Hence, we have added a Feedback entity to our data model.

- Moving toward the attributes, in the most important entities, we identified that the created_date column should be there to identify when a particular record is inserted in our system. Also, it is a very important field in the report to show counts by date. Hence, we have added this created_date column in most of the entities.

Step 8. BI architecture diagram

Hub-and-spoke architecture:



Step 9. BI and analytics solutions description

Describe each BI and analytics solution you propose to build:

1. Customer Satisfaction Dashboard:
    a. Style: Dashboards & Scorecards
    b. Analytical Question: This dashboard will provide information about overall customer satisfaction level with the HumAir. It will include customer satisfaction segregated by airport, route, or other factors related to baggage handling operations.
    c. Main Users: Baggage Services Teams, Customer Service Representatives, Operations Managers, and Marketing and Communications Teams will be the key users of this dashboard.

2. Baggage Operations Dashboard:
    a. Style: Dashboards & Scorecards
    b. Analytical Question: Data on the typical processing time for each phase of luggage processing as well as the processing time for handling bags at each stage of the journey may be found in the baggage operation dashboard. The dashboard can assist airlines or airports in implementing targeted solutions to increase efficiency and decrease delays by determining which steps are taking longer than anticipated and why.
    c. Main Users: Baggage Handling Teams, Maintenance and Engineering Teams, Customer Service Representatives, Data analysts

3. Baggage Incident Dashboard:
    a. Style: Data Discovery and Report
    b. Analytical Question: This dashboard provides information on how many bags are lost or delayed over time, which airports or routes have the greatest amount of lost or delayed bags, what the main causes of lost or delayed luggage are, which airports or routes experience the most baggage delays or losses, how many

baggage claims are open or closed, and how long it takes to process a baggage claim, etc.

c. Main User: staff, Baggage handlers, Customer service representatives, Data Analyst

4. Baggage Claims Dashboard:
    a. Style: Data Discovery & Dashboards & Scorecards
    b. Analytical Question: A baggage compensation dashboard can provide insights into the impact of compensation on customer satisfaction and loyalty. By tracking compensation offered to passengers for lost, delayed, or damaged baggage, as well as the number of claims filed and resolution time, HumAir can analyze the correlation between compensation and customer loyalty. This analysis can help HumAir to adjust its compensation strategy to improve customer retention rates, such as increasing the compensation amount for baggage issues.
    c. Main User: Customer service representatives or agents, finance or accounting department
5. Predictive Analysis Dashboard:
    a. Style: Big Data and Predictive Analytics
    b. Analytical Question: This dashboard will give answer to our predictive questions such as How can RFID or GPS be used for baggage tracking and monitoring to decrease the percentage of baggage issues in future, and what is the predicted impact of this technology, How can we use real-time updates on baggage location/status to increase customer ratings and forecast future retention rates, How can automated systems be used to decrease baggage issues, and what is the predicted reduction rate in delay issues if we implement self-check-in software for baggage processing etc.
    c. Main User: Data scientist, Data analyst, management team

Step 10. Plan the analytics project

The team has come up with different ideas on how to create a simple project plan for building the solution to the analytics problem.

1. **Scope Statement**: The project aims to develop an analytical solution to address baggage-related issues in an airline company. The solution will help in minimizing baggage mishandling incidents, improve baggage tracking, and enhance customer satisfaction.
2. **Main stakeholders** involved in this project related to baggage issues in the airline industry.
    - **Airline companies**: Airline companies are the primary stakeholders as they are responsible for baggage handling operations. The success of the project will benefit airline companies by improving their baggage handling operations, reducing costs associated with mishandled baggage, and enhancing the customer experience.
    - **Passengers**: Passengers are another important stakeholder as they are the end-users of the baggage handling system. The project's success will benefit passengers by reducing the likelihood of baggage-related issues, such as lost or damaged baggage and improving their overall travel experience.
    - **Customer service teams:** The customer service teams are responsible for addressing customer complaints related to baggage issues. The success of the project will benefit customer service teams by providing them with insights and data to address customer complaints related to baggage issues, resulting in improved customer satisfaction.
    - **Airport authorities:** Airport authorities are responsible for the overall airport operations, including baggage handling. The success of the project will benefit airport authorities by improving the efficiency of baggage handling operations and enhancing the airport's reputation as a customer-centric airport.

Name of each project phase:

1. Requirement Gathering Phase

2. Data Collection and Preparation Phase

3. Data Analysis Phase

4. Solution Design Phase

5. Solution Implementation Phase

6. Solution Testing Phase

7. Solution Deployment Phase

1. **Requirement Gathering Phase:**
    - Understand the business requirements and the problems faced by the airline's baggage handling system. (Business Analyst)
    - Conduct various meetings with stockholders to clarify doubts (Business Analyst)
    - Define project scope and objectives (Project Manager)
    - Identify project stakeholders (Project Manager)
    - Establish project team roles and responsibilities (Project Manager)
    - Develop project plan and timeline (Project Manager)
2. **Data Collection and Preparation Phase:**
    - Identify relevant data sources (Data Analyst)
    - Collect and compile data (Data Analyst)

- Clean and pre-process data (Data Analyst)

3. **Data Analysis Phase:**
   - Develop statistical models (Data Scientist)
   - Identify trends and patterns (Data Analyst)
   - Conduct root cause analysis (Data Analyst)
4. **Solution Design Phase:**
   - Identify solution requirements (Business Analyst)
   - Develop solution architecture (Solution Architect)
   - Design user interface (UI Designer)
5. **Solution Implementation Phase:**
   - Develop a solution (ETL Developer)
   - Build data pipelines (DevOps Engineer)
   - Implement data visualization tools (Data Visualization Engineer)
6. **Solution Testing Phase:**
   - Test solution components (Quality Assurance Engineer)
   - Conduct user acceptance testing (Business Analyst)
7. **Solution Deployment Phase:**
   - Deploy the solution to the production environment (DevOps Engineer)
   - Provide training to end-users related to dashboards (Training Specialists)
   - Monitor solution performance (Production support engineer)

Roles assigned to team members:

Project Manager: Vraj

Data Analyst: Neha

Data scientist: Neha

Business analyst: Gurmeher

Solution Architect: Gurmeher

UI Designer: Avinash

ETL Developer: Pravina, Vraj

DevOps Engineer: Pravina

Data Visualization: Param, Vraj

QA Engineer: Param

Training Specialist: Avinash

Production Support Engineer: Pravina

Step 11: Legal and ethical concerns

What are the legal and ethical concerns that you must keep in mind when working on this analytics project?

Legal Concerns:

In our aviation industry, Passenger and Baggage information is very crucial. That's why we make sure we follow the privacy rules like Personal Information Protection and Electronic Documents Act (PIPEDA). This involves obtaining people's permission before collecting their personal information, making sure the data is accurate, limiting the amount of data gathered, protecting the data from misuse, and taking responsibility for the data obtained.

Ethical Concern:

In the airline industry, ethical considerations for baggage delay, loss, or damage include timely and transparent communication with affected passengers, accepting responsibility for the problem and offering adequate compensation or reimbursement, safeguarding passengers' private information, and taking steps to prevent or lessen such incidents.

- **What data will you need to protect and how you will do it and what strategy will you use?**

➢ **Passenger data:** This contains individual particulars like name, address, phone number, and itinerary. Only authorized employees should have access to this data, and all data transfers should be done securely using encryption and other security measures.

  - **Data protection Strategy:** - Pseudonymization and Encryption
    - Pseudonymization is a strategy of replacing identifiable data with a reversible, consistent value for operational purposes while limiting access to personal data. This strategy could be used to replace the passenger information with the passanger_id.
    - **Encryption:** To prevent unauthorized access, sensitive data such as password, should be encrypted both during transfer and storage.

- ➢ **Information on baggage:** This includes details on the size, weight, and contents of each passenger's bag as well as statistics on baggage claims and tracking. Only authorized employees should have access to this data, and all data transfers should be done securely using encryption and other security measures.
  - **Data protection Strategy: -** Data minimization & User access controls
    - Data minimization can be used to reducing the amount of baggae data collected to only what is necessary for a specific and compelling purpose, such as identifying and communicating with affected passengers. User access controls can also be implemented to manage access to baggage data through access restriction and monitoring, ensuring that only authorized personnel have access to the information.

- **What legal or ethical risks do you need to take into account?**
  - **Legal risks**: In accordance with national and foreign laws, airlines are required to reimburse passengers for baggage lost or damaged during flight travel. If HumAir doesn’t follow these rules, it becomes legal trouble and social harm.
  - **Ethical risks:** Dealing with confidential data and luggage-related data may cause problems.

Step 12. Lessons learned

7 lessons that we learned while working collaboratively on this project are:

- **Managing Conflicts:** With the diverse educational backgrounds of our team members, it is natural for conflicts to arise. However, we tried to address and resolve conflicts, working together as a cohesive team.
- We learned to stick to the business problem. As there are high chances of deviation from the project objective.
- Making decisions collectively and involving everyone in the process can be difficult, but it is essential for effective teamwork. So, we learned to make **decisions collaboratively**.
- **Communication** is essential. A successful group project depends on the team member's ability to effectively communicate. Goal alignment, task delegation, and conflict resolution are all made easier by consistent and clear communication.
- Good **time management** and adherence to project deadlines are essential in a team environment. We now know how important it is to swiftly finish the duty given to us each week. Setting reasonable deadlines, prioritizing work, and coordinating efforts are required to guarantee the timely delivery of project deliverables.
- **Defining roles and responsibilities**: we learned to Set up roles and responsibilities for each team member helps to avoid confusion and make sure that everyone is aware of their duties. Better coordination and accountability are facilitated by it.
- **Responsibility and dedication:** Each team member needs to be dedicated to carrying out their duties and taking ownership of their tasks. The success of the group project depends on everyone's ability to hold each other and themselves accountable for their efforts.

What was the most difficult?

The most difficult challenge we encountered during our collaborative project was managing conflicts that arose due to the diverse educational backgrounds of our team members. It was sometimes challenging to reconcile different perspectives and opinions, and finding common ground required open communication, active listening, and compromise.

What were your collaboration challenges?

Our collaboration challenges included decision-making, because it was challenging to come to an agreement among team members with different opinions. Also, it took proactive measures to keep everyone informed and on the same page to ensure excellent communication among team members, particularly in a virtual environment. It was difficult to manage time and stick to deadlines since organisation and responsibility were needed to coordinate efforts and guarantee that assignments were completed on time.

Which of these challenges do you expect to encounter in real business projects, and why?

We anticipate comparable difficulties in real business projects because diverse teams with a variety of opinions and communication preferences are typical in business settings. Due to varying interests and points of view, making decisions can be difficult. It can be challenging to communicate effectively, especially in distant or virtual work settings. In order to assure the timely delivery of results and achieve project objectives, time management, and fulfillment deadlines are essential in business initiatives.

By establishing clear communication channels, setting expectations, outlining roles and responsibilities, and encouraging open and friendly communication among team members, it is crucial to foresee and proactively solve these difficulties in real business projects. Encouragement of active decision-making and the development of a cooperative team culture can also help to ease difficulties and guarantee positive project outcomes.