Optimizing American Airlines Bag Room Operations Nationwide: A Data-Enabled Approach

Aamani Ramanathan Parthasarthy

Mohammed Zubairuddin

Naga Praveena Kollipara

University of North Texas, Denton, Texas

Dr. Denise Philbot

ADTA 5910: Capstone Project

Table of Contents

[CHAPTER 1: CONCEPT PAPER 3](#_Toc152880084)

[Introduction 4](#_Toc152880085)

[Problem Statement 5](#_Toc152880086)

[Research questions 6](#_Toc152880087)

[CHAPTER 2: LITERATURE REVIEW 8](#_Toc152880088)

[Previous Theories and Research in the Field 9](#_Toc152880089)

[CHAPTER 3: METHODOLOGY 16](#_Toc152880090)

[Data Description 16](#_Toc152880091)

[Data Collection 17](#_Toc152880092)

[Data Wrangling 18](#_Toc152880093)

[Tools 19](#_Toc152880094)

[Exploratory Data Analysis (EDA) 19](#_Toc152880095)

[Total Bags 20](#_Toc152880096)

[Total Bags count at each station. 21](#_Toc152880097)

[Total Bags count at each station – month wise. 23](#_Toc152880098)

[Total Bags count in each station – date wise. 24](#_Toc152880099)

[Total Bags count in each station - week wise. 25](#_Toc152880100)

[Average number of flights scheduled for departure – Hourly wise. 28](#_Toc152880101)

[Average baggage arrival time. 30](#_Toc152880102)

[Average baggage arrival time. 31](#_Toc152880103)

[ANALYSIS 32](#_Toc152880104)

[CONCLUSION 40](#_Toc152880105)

[REFERENCES 42](#_Toc152880106)

# CHAPTER 1: CONCEPT PAPER

## Introduction

American Airlines is a major US-based airline headquartered in Fort Worth, Texas, within the Dallas–Fort Worth metroplex. It is the largest airline in the world when measured by scheduled passengers carried and revenue passenger mile (Wikipedia contributors, 2023).

**American Airlines Flight Operations:**

American Airlines, as one of major carriers in the United States, has successfully implemented the hub-and-spoke system to connect passengers across the country and beyond. This system allows them to efficiently manage their flight operations, provide a broad range of travel options, and compete effectively in the highly competitive airline industry. The hub-and-spoke system is an operational model to efficiently manage the flight routes and passenger traffic.

***Hub***: A hub is a central airport that serves as a major connecting point for an airline's flights. American Airlines operates 10 hubs – Charlotte, Chicago- O’Hare, Dallas / Fort Worth, Los Angeles, Miami, New York (JFK), New York (LaGuardia), Philadelphia, Phoenix (Sky Harbor), Washington (Regan) (Wikipedia contributors, 2023).

These hubs are strategically located across the United States to facilitate connections between flights from various origin and destination cities.

***Spokes***: Spokes are airports that are not hubs but serve as destinations for flights originating from the hubs. These airports are connected to one or more hub airports and typically have a mix of regional and mainline flights arriving and departing.

**Bag-room Operations:**

Airlines face a huge challenge when it comes to bag-room throughput, particularly with the staggering number of travelers they serve daily. In American Airlines, with 500,000 travelers passing through their gates every day, airlines need to efficiently handle their luggage as well. On average, each traveler carries approximately 2.5 bags. To manage this massive volume of luggage, airlines typically operate bag-rooms and Baggage Handling Systems at the first level of their airport facilities. Bag-room handles the checked bags and transfer bags except shorter tail-to-tail connections transfer bags. Gate bags do not enter bag-room. The efficiency of the bag-room's operations is paramount to prevent delays, lost luggage, and ultimately, to provide travelers with a smooth and hassle-free experience from check-in to baggage claim.

# Problem Statement

Every airline carrier reports baggage mishandling rate to Office of Airline Information (OAI) each month. These data are used to monitor each carrier's baggage, wheelchairs, and scooters handling, and to provide information to consumers (Office of Airline Information (OAI), 2018). In turn, this information on public portal adds to consumer trust and passenger attraction.

Additionally, the personnel and equipment required for managing an airline's bag-room represent a substantial financial investment, contributing significantly to the operational costs of the airline. Hence, optimizing bag-room's operations is crucial to minimize the risk of flight delays, mitigate the occurrence of mishandled luggage.

A diagram of a flight

Description automatically generated

## Research questions

* What are the key performance indicators (KPIs) for bag-room operations, and how can bag-room data be leveraged to improve these metrics?
* How can historical bag-room data be analyzed to identify trends and patterns in baggage handling, and how can this analysis inform operational improvements?
* What are the most common causes of mishandled baggage, and how can bag-room data help in developing proactive strategies to reduce such incidents and reduce baggage Mishandling Rate? Does arrival of baggage have a pattern / distribution and what are the factors influencing the pattern? Model arrival curves to facilitate precise resource allocation and workflow scheduling within the bag-room to enhance efficiency and minimize congestion?
* How many carts would be required to carry baggage from the check-in point to the baggage storage area?
* How many bags would pass through baggage room per flight?

# CHAPTER 2: LITERATURE REVIEW

The challenges to control the high number of footfalls have always been inevitable. It could be a shopping mall, airport or any venue where there could be a greater number of footfalls. There are many other challenges associated with operational workflows when the population footfall is not controlled. Let us take us example of Hospital or a shopping mall. If the footfalls are not streamlined or controlled, it would be difficult to manage the operational process and generate the profits in the respective businesses.

The airport industry is a complex and dynamic field where customer satisfaction is of paramount importance. A sub-domain in this industry is baggage handling system (BHS), a critical operation in terms of customer satisfaction and cost. It is important to set a solid foot in understanding baggage handling procedures, methods currently in the industry to study mishandling rates and further methods to optimize prediction models. This literature review embarks on a comprehensive exploration of existing studies and research within the intersection of aviation and baggage handling data analysis. This literature review begins to identify the flight and bag-room operations, the discusses baggage mishandling rates and theories in this field and finally conclude with takeaway pointers.

**Baggage mishandling rates:**

Baggage systems are an important and key performance indicator of the airport’s service quality. Inefficient baggage handling systems lead to passenger and airline dissatisfaction. This reduces the preferability of the airport operator. In addition, the competition between airports, which is becoming fiercer day-by-day, for passengers and airlines also makes service quality an important factor. Using fast transfer baggage systems for transfer, outbound or inbound baggage attracts new airlines to the airport as a destination or hub and motivates passengers. It offers an attraction to use the same airport in the future (International Airport Review, 2022). Nearly 220,000 bags were “mishandled” by U.S. airlines in April 2022, meaning they were lost, damaged, delayed or stolen, according to the most recent data published by the U.S. Department of Transportation (Iacurci, 2022). 0.72 bags per hundred are mis handled in American Airlines in April 2022 (Air Travel Consumer Reports for 2022, n.d.).

According to SITA (Passenger IT Insights 2023, n.d.), mishandled baggage rate of 7.6 bags per thousand passengers in 2022. This can be traced back in large part to the resumption of international and long-haul lights throughout 2022, meaning more transfers, where bags are most susceptible to being mishandled. In other words, mishandling rates for international bags are more likely to be transferred from one flight to another – are eight times higher than for domestic flights.

# Previous Theories and Research in the Field

As airlines continuously strive to enhance their service quality, safety, and efficiency, understanding the existing data analysis studies and theories is indispensable for driving improvements and innovation in baggage handling services. First and foremost, to set the basics, let’s understand the current practices / theories / algorithms / research to identify baggage mishandling events. Followed by methods to assess the quality of airline BHS. Further, discussion focuses on loading and unloading baggage zones and role of BHS.

**Predicting Mishandling events**

In the airline industry, several methods have been adopted to understand the significant factors influencing customer satisfaction. Research commonly involved on-time performance, denied boardings, baggage mishandling incidents and customer complaints (Airline Quality Rating 2022). Baggage mishandling has been a consistent factor affecting Airline Quality Rating. Cluster analysis is a statistical method used to classify groups. It has been identified that cluster analysis has been effective method in grouping similar flights in terms of several dimensions and sub dimensions (Wang & Pham, 2020). This provides an idea to concentrate on a particular area of interest or improvement in this industry. Different Machine Learning (ML) models like Feed-Forward Neural Networks (FFNNs);K-Nearest Neighbors (KNN);Gradient Boosting Machine (GBM);Random Forests (RF); ExtraTrees (ET) are developed for streamlining this process (Z. Wang et al., 2020).

A Similar approach to streamline the inflow and outflow of taxi routes has been implemented at Beijing International Airport by identifying bottle necks and reduced the average taxi-in time by 5.1 minutes. Controlling pushback reduces average taxi-out time by 3.7 minutes. It is a novel integration of data analysis and optimization tailored to the studied airport to evaluate and improve surface operations. The data-driven customization and optimization results showcase the benefits of this approach (Ma et al., n.d.).

To avoid the high turnaround time, Copenhagen Airport and The Technical University Denmark, developed the algorithm for handling outbound luggage at the airport. This has eliminated the need to increase the real estate area for baggage handling at the airport (Pisinger, D., & Rude, S. Í. H. 2020).

A new algorithm called the oppositional krill herd algorithm (OKHA) was developed for global numerical optimization. OKHA is based on the opposition-based learning and krill herd algorithms. It introduces an opposition-based population initialization method and new movement equations for krill individuals. Performance of OKHA is tested on a set of 23 numerical benchmark functions and compared to other optimization algorithms like genetic algorithm, particle swarm optimization, and original krill herd algorithm. Results show OKHA converges faster and finds better optimal solutions than the other algorithms for most test functions. The opposition-based initialization increases population diversity. The new movement equations improve exploration and exploitation. The authors conclude that OKHA is an effective new algorithm for solving global optimization problems. The opposition-based learning and new krill movement mechanisms enhance its global search ability and convergence speed. In summary, the article proposes a new oppositional krill herd optimization algorithm and demonstrates its performance improvements over other algorithms for numerical test functions. The results highlight the benefits of opposition-based learning and new krill movement equations for global optimization problems (Wang, X. 2022).

**Quality assessment of airline baggage handling systems**

The SERVQUAL model was suggested by the study to evaluate how well the baggage handling system is judged to be performing. The most crucial qualities are dependability, responsiveness, certainty, tangibles, and empathy, according to literature evaluations and data from the best worst method (BWM).

An array of methods is used in this study to assess the service quality aspects of the airline baggage handling system. It recognizes a hierarchy of criteria and establishes their relative weights. The best worst method (BWM) and the SERVQUAL framework are utilized to generate quantitative findings. The study, which is based on passenger interviews and surveys, identifies 13 fundamental service characteristics. The findings imply that designs focusing on the underlying consumer clusters may be more successful than those focusing on the aggregate (Rezaei, J., Kothadiya, O., Tavasszy, L., & Kroesen, M. 2018, June 1).

**On load balancing strategies for baggage screening at airports**

The efficiency of the screening subsystem is the focus of this paper's examination of load balancing policies for an airport baggage handling system. The paper offers a join-shortest-queue (JSQ) policy that may be used in conjunction with round-robin (RR) and first-available (FA) policies. It does this using discrete-event simulations. The outcomes of the simulation demonstrate that RR-JSQ can enhance the efficiency of the system.

An airport's baggage handling system (BHS) is a logistical device that automatically moves passenger bags from one location to another. It entails activities including registration, transportation, screening, locating, classifying, and early storage. Conveyors, trays, and carts are just a few of the logistical tools that BHS uses. Individual bag speed control is made possible by destination-coded vehicles (DCVs), a recent technological innovation. BHS is composed of baggage processing flow-based subsystems, such as screening equipment and loading robots. Systems with cascading queues may result in delays problems.

Prior to loading into the aircraft, security screening of all passenger luggage entering a BHS is required. The effectiveness of the screening subsystem is significantly influenced by the load balancing policy. The RR, FA, and JSQ load balancing strategies, as well as a potential combination of the RR-FA (current widespread practice) and RR-JSQ, were all examined in this study (Wu, X., & Xie, L. 2017, July 1).

**Optimal Assignment of Airport Baggage Unloading Zones to Outgoing Flights:**

A stochastic optimization model for assigning baggage unloading zones (chutes) to outgoing flights at an airport. Uncertainty in flight departure times, baggage volumes, and handling times are captured through scenarios. Objective is to minimize expected total assignment costs across scenarios. Costs include handling workload, airline preferences, overlap if multiple flights assigned to one chute. Model incorporates practical constraints like consistent assignment of flights on different days and overlap penalties. Implemented case study on major Asian airport with around 850 flights per week. Compared optimal solution to heuristic rules like LIFO. Optimal stochastic solution reduces costs by 23-27% versus heuristic policies and deterministic solution. Just 2-3 priority levels extract most revenue benefits. Overall, this is a strong paper that provides an optimization model for a complex real-world airport operations problem. The stochastic modeling and computational results demonstrate the value of using optimization tools versus heuristics for this application. The model provides a valuable framework that could be extended to additional operational considerations in future work (Huang et al., 2016).

**Baggage arrival theories and models**

A queuing model encompasses the entire system, including both the arrival process and the service process, while an arrival model specifically focuses on modeling how entities arrive at the system. Arrival models are a fundamental component of queuing models and play a crucial role in understanding and optimizing queuing systems. According to field of operations research and queuing theory, arrival models describe the probability distribution and characteristics of how entities arrive at the queue over time. Common arrival processes include Poisson arrivals (constant arrival rate over time) or non-Poisson arrivals (varying arrival rates) (Wrediningsih et al., 2019). Steady state is a condition when the properties of a system do not change within time (constant). The queue process generally is assumed as the time between arrivals and service times following the exponential distribution, or equal to the numbers of arrivals and the number of services following the poisson distribution. Some distributions deviate from the strict assumptions of the poisson distribution, which assumes a constant event rate and independence such as weibull distribution, exponential distribution (for constant-rate arrivals), negative binomial distribution (for over-dispersed arrivals), and more complex time series models like the Autoregressive Integrated Moving Average (ARIMA) for modeling time-varying arrival rates.

Queueing theory (Haviv & Ravner, 2021) provides tools for the analysis of the waiting times and associated costs. If customers have the option of deciding when to join the queue, they will face a decision dilemma of when to arrive. The level of congestion one suffers from depends on others, behavior and not only that of the individual under consideration. This fact leads customers to make strategic decisions regarding their time of arrival. In addition, multiple decision makers that affect each other’s expected congestion call for non-cooperative game-theoretical analysis of this strategic interaction. Customers choose both – their arrival time and priority level to join the queue, trading off the cost of waiting before service begins vs. cost of obtaining higher priority services. Strategic customer choices lead to self-organization into priority classes. The numerical results suggest that a few priority levels extract most of the revenue benefits. Limitations are the stylized model with simplified cost functions. Extensions could consider more general cost functions and incorporate psychology and bounded rationality in decision making. Overall, it's an interesting model that yields economic and operational insights into designing priority queueing systems (Talak et al., 2019b).

**Case Study: An Application of Cluster Analysis Method to Determine Vietnam Airlines Ground Handling Service Quality Benchmarks**

Here, use of cluster analysis to analyze ground handling service quality data from Vietnam Airlines at Noi Bai International Airport. Data was collected through questionnaires completed by 315 international passengers in 2019. The questionnaire had 28 service quality attributes across 5 dimensions (reliability, assurance, tangibles, empathy, responsiveness). Cluster analysis identified 3 service quality clusters: high, medium, and low. The high cluster had high scores across all dimensions. The medium cluster had moderate scores. The low cluster had low scores on reliability, assurance, and empathy. Based on the cluster analysis, benchmarks were proposed for ground handling service quality at Noi Bai Airport. For the high-quality benchmark, scores should be above 4.1 across all dimensions. For the minimum quality benchmark, scores should be above 3.4 across all dimensions. The research provides data-driven quality benchmarks that Vietnam Airlines can use to evaluate and improve their ground handling services. The methodology can also be applied to other airlines and airports (Wang, T. W., & Pham, Y. T. H. 2020). In conclusion, this literature review has provided an overview of current studies and theories and practices in the domain of baggage handling system.

# CHAPTER 3: METHODOLOGY

## Data Description

**Dataset Name:** American Airlines baggage operations.

**Dataset Source:** The dataset is provided by Operations Research and Advanced Analytics team in American Airlines Headquarters after signing a non-disclosure agreement. The data is accessed through an ETL (Extract, Transform, Load) tool snowflake.

Figure 1 – Dataset in Snowflake.

A screenshot of a computer

Description automatically generated

**Dataset Period:** The dataset covers bag room operations in Summer 2023 which ranges approximately between May 27, 2023 (Memorial Day weekend) to September 3, 2023 (Labor Day Weekend)

**Dataset shape**: The dataset contains 11 variables and more than 6 million records.

**Dataset format**: The data is structured containing values in string, integer, date and time stamp formats. Accessible using SQL queries.

**Data variables**:

1. FLIGHTNUMBER – Integer variable

This is the unique number temporarily assigned to flight.

1. STATION – Departure Airport/location.

The airport from which the flight departs. Here, we have values from six primary airports, Charlotte Douglas International Airport, Dallas Fort Worth International Airport, Chicago O’Hare International Airport, Los Angeles International Airport, Miami International Airport and Phoenix Sky Harbor international airport.

1. DESTINATION – Arrival Airport/location.

The airport to which the flight arrives.

1. SCHD\_LEG\_DEP\_TMS – Scheduled time of flight departure

The time scheduled for flight departure.

1. ACTL\_LEG\_DEP\_TMS – Actual time of flight departure

The actual time when the flight departed.

1. INTERNATIONAL – International or Domestic flight.

If the destination is international, it is an international flight, similarly domestic flight.

1. FLEETTYPE

The model number of the flight. Example: ‘777’ in Boeing.

1. BODYTYPE – Narrow or wide body of flight.
2. BR\_ARVL\_TMS – Baggage arrival time
3. COUNT\_BAGS – Baggage count for each
4. SF\_LOAD\_TMS – Baggage load time

## Data Collection

American Airlines team has provided the data of baggage operations in primary airports, Charlotte Douglas Airport, Dallas Fortworth Airport, Chicago O’Hare International Airport, Los Angeles International Airport, Miami International Airport and Phoenix Sky Harbor international airport.

The data was integrated by connecting snowflake to python using the queries below.

!pip install snowflake-connector-python

!pip install "snowflake-connector-python[pandas]”

## Data Wrangling

After a first glance at data, we identified there are integer, string and date time stamp data types for all the variables. There are approximately 6million observations. The data set is significantly more challenging than assumed. Before we analyzed the data, we observed the dataset and made a high-level outline about the insights. We have separated the time and date for all the date and time variables for better analysis. Also, the time zone in the time stamp are not uniform, hence they are changed to CST to get the right duration. Also, we have integrated the dataset with globe latitude and longitude data to understand the flight route trends.

Figure 1.1 – Data Transformation python query

A screenshot of a computer program

Description automatically generated

Figure 2 – Data Transformation result

A screenshot of a computer

Description automatically generated

## Tools

* Jupyter notebook – Anaconda – used for data wrangling and EDA.
* Snowflake – To access and download the data. SQL and Python workspaces to analyze the data.
* Tableau – for EDA and data visualization.

## Exploratory Data Analysis (EDA)

The objective of the EDA was to understand the relationship between the variables. As the core problem is operational flaws and the objective is to re-structure the baggage operational flow it is important to perform EDA in-depth as it would give many insights from the façade area. Also, all the variables are in integer, string, date and time stamp format, hence the usual EDA methods cannot be applied where correlation is analyzed between variables. We made a different approach in this step. The values of the variable SF\_LOAD\_TMS are the same for all the flights which made it insignificant. Using the other variables, the below visualizations are created to understand the relationship between bags count, destination, date and time.

### Total Bags

This visualization in the figure 3, below depicts the count of bags loaded in all flight from each airport. Most number of bags are flown to DFW international airport and from DFW international airport to various destinations primarily to Miami, New York, Chicago, Los Angeles and Pheonix cities.

Figure 3 – Total bags count for each destination from primary airports.

A screenshot of a computer

Description automatically generated

### Total Bags count at each station.

In figure 4, total bag count for both international and domestic flights at major airports. DFW has the highest bag count followed by Miami International Airport. The total number of bags per station is depicted in the bar chart, figure 5, and the findings indicate that, when compared to other airports, DFW airport has the largest number, with Miami airport coming in second. As a result, the amount of luggage at the airports in Charlotte, Chicago, and Lux differs very little. Additionally, out of all airports, Phoenix Airport has the lowest amount of luggage.

Figure 4 – Total bags count at each destination.

A graph of different colored squares

Description automatically generated with medium confidence

In the below figure 6 it depicts the density of bags count across the major airports in USA. This visualization depicts the east coast has more flights and baggage count compared to west.

Figure 5 – Total bags count at each airport.

A graph of orange squares

Description automatically generated with medium confidence

Figure 6 – Total bags count at each airport across the country.

A map of the united states

Description automatically generated

### Total Bags count at each station – month wise.

The total number of bags flown at each station is depicted in the heatmap in figure 7, from May to September. To determine the total number of bags, a filter is applied to the stations. The findings indicate that the months of June, July, and August have the most luggage among all airports. Moreover, DFW airport has the highest bag count in these 3 months. From June to August, the airports in Miami and Charlotte had the second and third highest counts, respectively. Consequently, the remaining airports have some similar counts but have a smaller number of bags compared with DFW, MIA and CLT airports.

Figure 7 – Total bags count every month.

A screenshot of a computer

Description automatically generated

### Total Bags count in each station – date wise.

This line graph in figure 8, consists of six colored lines, where the y-axis indicates the number of bags, and the x-axis shows the dates in the order of time. Stations and months from May to September are filtered. An airport is represented by each colored line. Throughout the whole-time frame displayed, the DFW airport stays level and steady. With occasional fluctuations, the airports with the fewest luggage are LAX and PHX. Similar ascending trends are seen at the Chicago and Charlotte stations, which begin with lower values on the left and rise to better values on the right. However, between June 25 and June 27, count bags decreased at every location.

Figure 8 – Total bags count date wise.

A graph of colored lines

Description automatically generated

Figure 9 – Total bags count date wise – Miami International Airport.

A graph with blue lines and text

Description automatically generated

The y-axis in the line graph above in figure 9, depicts the number of bags, while the x-axis displays the dates in chronological order. The single-colored blue line on this graph illustrates the count of bags on all days at Miami Airport from May to September. Filtered are stations and months ranging from May to September. The total number of luggage checked into Miami airport for the selected months is displayed in the above line graph. It is accurate to say that Saturday is the day with the most luggage checked at Miami airport. The number of bags is lowest during the week and highest during the weekend.

### Total Bags count in each station - week wise.

The y-axis in the box plot above, figure 10, indicates the number of bags, while the x-axis shows the number of days in a week. station and month columns are filtered. It is not difficult to see that DFW airport handles the most bags each day. The airports in Miami and Charlotte have the second and third most luggage, respectively. Subsequently, the least amount of luggage is checked each day at the airports of LAX and PHX. In addition, Monday, Tuesday, and Wednesday have the minimum bags at every airport.

Figure 10 – Total bags count day wise.

A graph of a graph

Description automatically generated with medium confidence

In the below figure 11, the density of number of bags is depicted separately for the international flights and domestic flights. DFW has highest density for International flights and Miami International Airport has highest density for domestic flights.

Figure 11 – Total bags count at each airport for international flights

A screenshot of a computer

Description automatically generated

Figure 12 – Total bags count week wise.

A graph of a graph

Description automatically generated with medium confidence

The bar graph above in figure 12, illustrates how each station's weekly total bag count is determined. Days in a week are represented on the x-axis, and count bags are represented on the y-axis. The station is colored differently and filtered. We can see that DFW airport, which is shown in orange on the y-axis, takes up greater space every day. Charlotte comes in third, and Miami comes in second, as indicated by the purple and blue color. The remaining stations take considerably less area. Lastly, on weekdays, all stations deal with some of the lowest bags.

### Average number of flights scheduled for departure – Hourly wise.

Figure 13 – Total flights count.



We created a heatmap as depicted in the above, figure 13, graphic to show the average number of flights planned for departure hourly. The station is filtered in this graphic, which shows that the morning hours of 8 to 10 in May had the largest daily average of flights. later in the evening, after which it decreased slightly. The average number of flights increased again from 4 p.m. to 10 p.m. When it comes to June, we can state that the morning hours of 7 to 10 am have a high average number of flights, with a minor reduction later in the day.

Figure 14 – Total flights count.

A screenshot of a computer

Description automatically generated

### Average baggage arrival time.

In the below chart, figure 14, we can see the heatmap that shows the average baggage arrival time which is highest in the flights flying from DFW and Chicago to various destinations.

Figure 15 – Average bags count.

A screenshot of a computer

Description automatically generated

## Average baggage arrival time.

The below infographic visualization depicts the flight route map from DFW airport. Most of the domestic flights from DFW fly to east coast cities and most of the international flights fly towards Europe. The same trends are observed in the flights flying from Chicago O’Hare international airport.

Figure 16 – Flights from DFW airport

A map of the world with blue lines

Description automatically generated

Figure 17 – Flights maps.

A map of the world

Description automatically generated

## ANALYSIS

The below info graph, figure 18, depicts the average baggage arrival time which is highest in Miami International Airport and least in the Phoenix International Airport.

Figure 18 – Average Scheduled baggage arrival time.

A graph of blue bars

Description automatically generated with medium confidence

The below info graph, figure 19, depicts the flight maps. Most of the flights from USA fly towards Mexico and Europe. The average baggage arrival time difference is high in Miami and Loas Angeles airports as depicted in the figure 20.

Figure 19 – Flights map

A map of the world with many points

Description automatically generated

Figure 20 – Average time difference between baggage arrival and scheduled departure..

A chart with text and numbers

Description automatically generated with medium confidence

From 7th to 19th of a month, bag count number is very high and lowest in the first week of the month (as depicted in figure 21). The figure 22 depicts the average flight count in a month, where similar trends are observed.

Figure 21 – Count of bags in a month.

A graph of blue bars

Description automatically generated with medium confidence

Figure 22 – Flight count in a month.

A graph of orange bars

Description automatically generated with medium confidence

Figure 23 – Count of bags from May to September.

A blue and black text on a white background

Description automatically generated

The biggest challenge in our analysis is applying any models to this dataset. This type of data set where target variables are date and time variables and problem statement is operational structure re-engineering, the EDA part is of high emphasis. This part itself would provide insights and problem causing factors.

Figure 24 – LAX Random Forest Regression.

A graph with blue dots

Description automatically generated

The Random Forest model for Los Angeles Airport, has generated the R-Squared value 0.37 and Mean Squared Error 1475.9. The plot, figure 24, plotted against predicted values and true values has depicts few outliers which cannot be eliminated. The cluster is observed between 0-150.

Figure 26 – PHX Random Forest model.

A graph with blue dots

Description automatically generated

The Random Forest model for Pheonix Airport, has generated the R-Squared value 0.72 and Mean Squared Error 199.9. The plot, figure 24, plotted against predicted values and true values has depicts few outliers which cannot be eliminated. The cluster is observed between 0-150.

Figure 26 – Ishikawa diagram.

A black background with white squares

Description automatically generated

We did an extensive exploratory data analysis (EDA) on a dataset associated with flight data. The analysis is conducted with the help of Python, SQL, and Pandas, demonstrating a flexible and effective method for interpreting, and drawing conclusions.

Further dataset information is extracted by the code through SQL queries. To get an immediate estimation of the size of the dataset, a SQL query is run to count the total number of records in the designated table. To figure out the dataset's date range, an additional SQL query is utilized for obtaining the lowest and maximum dates from a designated column. Interpreting the duration span of the dataset requires knowledge of this time-related data. We used Pandas to recognize and manage null values in the dataset. A Data Frame that indicates whether null values are present is Boolean and is created using the isna() function. Next, 'df\_null,' a new Data Frame with at least one null value in each row, is constructed. This stage is crucial for determining the amount of missing information and developing processing or estimation procedures. We used an advanced SQL query to find planes that connect inside the dataset. 'UNIQUE\_COMBI' is the Common Table Expression defined by the code to extract a distinctive set of aircraft numbers, locations, and routes. By comparing the final location for a single flight section with the corresponding station of the succeeding flight for an identical flight number, the following query determines connecting flights. 'df\_connectingflights,' a Pandas DataFrame containing the outcomes, is enlarged so it displays all rows and columns.

## CONCLUSION

This paper aimed to provide the operational flow issues of American Airlines flight and baggage handling operations, with a focus on analyzing baggage handling data to identify factors influencing baggage mishandling rates. The exploratory data analysis has provided valuable insights into the operational flow of baggage across various airports, below are the key insights:

1. Dallas Fort Worth International Airport handles the highest volume of baggage across all airports studied, followed by Miami International Airport and Charlotte Douglas International Airport.
2. Baggage volumes tend to peak on weekends, especially Saturdays. Weekdays see lower baggage volumes.
3. Morning hours from 8-10am tend to have the highest number of scheduled flight departures on average.
4. Dallas Fort Worth and Chicago O'Hare airports have the highest average baggage arrival times.

To further analyze the root causes of baggage mishandling and optimize baggage handling operations, additional data such as weather conditions, baggage handling staff schedules, and time stamps of baggage at each step of process is required.

The significant business insights from the analysis include:

**1**. **Bag Count variability across destinations and airports**: The DFW International Airport witnessed high baggage rate and flights count followed by Miami, Los Angeles and Pheonix.

**2. Temporal Patterns in Baggage Flow**: The busiest months witnessed in most of the airports are June, July, and August, experiencing the highest number of bag count. This is probably due to international students arriving. In the weekends most of the airports observed a steep increase in the number of bags. This is a clear depiction that airport must prepare for the vacation period.

**3. Operational Challenges and Opportunities:** The airports must have a different approach during these busy days to handle the bags. There is an opportunity to keep extra manpower and equipment to save the time.

**4. Flight Patterns and Baggage Arrival Times:** The trends in the average baggage arrival times at different airports like Miami and Los Angeles indicate areas that require more efficient handling procedures.

**6. Operational Structure Re-engineering:** EDA has provided a strong foundation for operational process structure re-engineering. Insights into problem-causing factors, such as high bag counts and varying arrival times, will be foundation in improving the overall efficiency and effectiveness of baggage operations.

The Random forest regression model for Los Angeles further strengthened out hypothesis with R-Squared value 0.37 and Mean Square error value 1475.9. Similarly, for Pheonix R-Squared value 0.72 and Mean Square error value 199.19.

In conclusion, the exploratory data analysis has successfully uncovered patterns and trends in baggage handling across major airports, offering a comprehensive view of the operational challenges and opportunities. This analysis lays a solid foundation for further investigations and potential re-engineering of baggage operational flows to enhance efficiency and effectiveness.

# REFERENCES

Air Travel Consumer Reports for 2022. (n.d.). US Department of Transportation. <https://www.transportation.gov/individuals/aviation-consumer-protection/air-travel-consumer-reports-2022>

British Airways and Philips projects in world’s first trials of “smart label” technology to revolutionise airline baggage handling. (2000). *Facilities*, *18*(3/4). <https://doi.org/10.1108/f.2000.06918cab.011>

Haviv, M., & Ravner, L. (2021). A survey of queueing systems with strategic timing of arrivals. *Queueing Systems*, *99*(1–2), 163–198. <https://doi.org/10.1007/s11134-021-09717-8>

Huang, E., Mital, P., Goetschalckx, M., & Wu, K. (2016). Optimal assignment of airport baggage unloading zones to outgoing flights. *Transportation Research Part E: Logistics and Transportation Review*, *94*, 110–122. <https://doi.org/10.1016/j.tre.2016.07.012>

Iacurci, G. (2022, July 19). Airlines are struggling with lost and delayed bags: What to know and how to pack if you’re traveling this summer. *CNBC*. <https://www.cnbc.com/2022/07/19/what-travelers-need-to-know-as-airlines-deal-with-lost-delayed-bags.html>

International Airport Review. (2022, March 1). The importance of baggage handling systems in winning passengers and airlines. <https://www.internationalairportreview.com/article/164476/the-importance-of-baggage-handling-systems-in-winning-passengers-and-airlines/>

Lu, M., Dufour, J. S., Weston, E. B., & Marras, W. S. (2018). Effectiveness of a vacuum lifting system in reducing spinal load during airline baggage handling. *Applied Ergonomics*, *70*, 247–252. <https://doi.org/10.1016/j.apergo.2018.03.006>

Ma, J., Zhou, J., Liang, M., & Delahaye, D. (n.d.). Data-driven trajectory-based analysis and optimization of airport surface movement. *Transportation Research Part C: Emerging Technologies*, *145*, 103902. <https://doi.org/10.1016/j.trc.2022.103902>

Passenger IT Insights 2023. (n.d.). <https://www.sita.aero/resources/surveys-reports/passenger-it-insights-2023/>

Pisinger, D., & Rude, S. Í. H. (2020). Advanced algorithms for improved baggage handling. Journal of Airport Management, 14(3), 280-290. <https://trid.trb.org/view/1753585>

Rezaei, J., Kothadiya, O., Tavasszy, L., & Kroesen, M. (2018, June 1). *Quality assessment of airline baggage handling systems using SERVQUAL and BWM*. Tourism Management; Elsevier BV. <https://doi.org/10.1016/j.tourman.2017.11.009>

Talak, R., Manjunath, D., & Proutière, A. (2019b). Strategic arrivals to queues offering priority service. *Queueing Systems*, *92*(1–2), 103–130. <https://doi.org/10.1007/s11134-019-09604-3>

The Bureau of Transportation Statistics (BTS), Office of Airline Information (OAI). (2018). Technical reporting directive #30 - mishandled baggage and wheelchairs and scooters.

<https://www.bts.gov/topics/airlines-and-airports/number-30-%E2%80%93-technical-directive-mishandled-baggage-effective-jan-1-2019>

Wrediningsih, A. P., Sugito, S., Prahutama, A., & Hakim, A. R. (2019). Non-Poisson queueing model’s identification (Case study: AKAP and AKDP bus on the West Lines bus service of Tirtonadi Surakarta). *Journal of Physics*, *1217*(1), 012102. <https://doi.org/10.1088/1742-6596/1217/1/012102>

Wang, T. W., & Pham, Y. T. H. (2020). An application of cluster analysis method to determine Vietnam Airlines’ ground handling service quality benchmarks. *Journal of Advanced Transportation*, *2020*, 1–13. <https://doi.org/10.1155/2020/4156298>

Wang, X. (2022). Measures for airlines to reduce airport congestion fees: Scheme design and performance assessment. *Mathematical Problems in Engineering*, *2022*, 1–12. <https://doi.org/10.1155/2022/5235813>

Wang, Z., Liang, M., & Delahaye, D. (2020). Automated data-driven prediction on aircraft Estimated Time of Arrival. *Journal of Air Transport Management*, *88*, 101840. <https://doi.org/10.1016/j.jairtraman.2020.101840>

Wikipedia contributors. (2023). American Airlines. Wikipedia. [https://en.wikipedia.org/wiki/American\_Airlines](https://en.wikipedia.org/wiki/American_Airlines%20)

Wu, X., & Xie, L. (2017, July 1). *On load balancing strategies for baggage screening at airports*. Journal of Air Transport Management; Elsevier BV. <https://doi.org/10.1016/j.jairtraman.2017.02.009>