

**Baggage Handling System**

**Optimization for American Airlines**

**Vishal Vallapu -11618660**

**Vamsi Krishna Bellam - 11593718**

**Aakanksha Reddy Poreddy- 11608373**

**Parvataneni Naga Satya Sai - 11599516**

**University of North Texas Toulouse Graduate School of Business**

**ADTA 5910-Capstone**

**Dr. Denise Philpot**

**Fall 2023**

# TABLE OF CONTENTS

**[CHAPTER 1 - INTRODUCTION 4](#_19c6y18)**

[Background 4](#_28h4qwu)

[Research Questions](#_2lwamvv) 9

**[CHAPTER 2 – LITERATURE REVIEW 5](#_3ygebqi)**

American Airlines History…………………………………………………………………………………………………..6

Historical Evolution of Baggage Handling…………………………………………………………………………..6

Challenges and Issues in Bag Handling…………………………………………………………………………………………..6

[Impact of Baggage Handling on Customer Satisfaction 7](#_1664s55)

Best Practices in Baggage Handling [9](#_2iq8gzs)

[Future Trends and Recommendations 9](#_3hv69ve)

[Conclusion 12](#_2250f4o)

**CHAPTER 3 – METHODOLOGY: DATA PREPARATION 14**

[Software 14](#_40ew0vw)

[Data Collection 14](#_3ep43zb)

[Data Wrangling 17](#_3bj1y38)

[Data Compiling and Cleaning 19](#_2hio093)

**CHAPTER 4 – METHODOLOGY: EXPLORATORY DATA ANALYSIS 21**

[Data Description 21](#_1f7o1he)

[Missing Data 21](#_thw4kt)

[Outliers 21](#_4cmhg48)

[Response Variable 23](#_44bvf6o)

[Categorical Variables 24](#_1xrdshw)

[Numeric Variables 28](#_vgdtq7)

**CHAPTER 5 – METHODOLOGY: MODELING 41**

[Test of Means 41](#_4ddeoix)

Multiple Linear Regression 42

[K-Fold Cross Validation 48](#_2pcmsun)

[Decision Tree 50](#_3g6yksp)

Principal Component Analysis and Principal Component Regression 56

**[CHAPTER 6 – MODEL EVALUATION 59](#_1p04j8c)**

**[CHAPTER 7 – CONCLUSION 60](#_hkkpf6)**

[Discussion 60](#_1gpiias)

[Applications 62](#_27jua8u)

[Limitations and Future Research 62](#_45tpw02)

**[APPENDIX 64](#_4j8vrz3)**

**CHAPTER 1: INTRODUCTION**

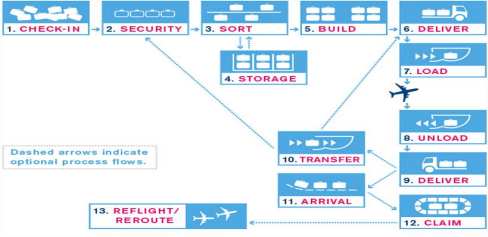
**Background**

American Airlines is driven by the motto” to care for people on life’s journey**”.** They offer thousands of flights to over 350 destinations over 60 countries. There are many departments in the Airlines namely Maintenance, Operations, Customer Experience, Revenue Management , Network Planning, Crew Planning , Flight Operations , Marketing & Sales, Airport Operations, Loyalty, Network Planning.

In our capstone project, we address the BHS (Baggage Handling System) problem of analyzing incoming baggage and aim to address the challenges faced by an airline company with a shortage of personnel and equipment in its BHS with central objective to predict and analyze the number of baggage items that will arrive for any given outbound flight.

The Baggage Handling System (BHS) is a critical component of airline's operations, ensuring that passengers' luggage is efficiently transported from departure airports to their final destinations and avoid mishandling of the baggage. It is a crucial part of airlines operation as it affects both passenger happiness and operational expenses.

There are several stages in the BHS namely :



**Source:** https://xwiki.avinor.no/display/BARIN/Baggage+process+and+data+elements+- +airport+perspective

1. **Check-in:** At the counter, passengers check their bags.

2. **Security:** Bags are inspected for the sake of security reasons.

3. **Sort:** Bags are arranged according to flight number and airport of destination.

4. **Build:** Bags are placed inside luggage carts or other holding spaces.

5. **Deliver:** The bags are brought to the plane.

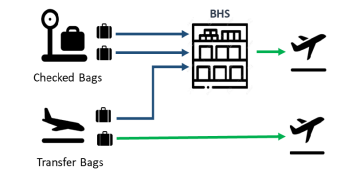
6. **Load:** Bags are loaded into the cargo hold of the aircraft.

7. **Unload:** Bags are unloaded from the cargo hold of the aircraft.

8. **Storage:** If there is any waiting time then baggage will be transferred into storage during this period.

9. **Transfer:** Based on the barcode tag attached to bags, will be delivered to claim area, re-route to connecting flight or storage.

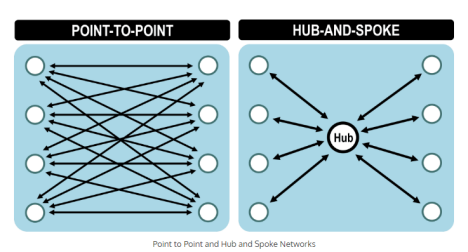
10.**Baggage claim:** At the baggage claim area, travelers claim their



**Source:** American Airlines Data Challenge pdf

Checked bags and Transfer bags (for connecting flights) are passed through the BHS and reaches the passenger at their destination.

**Scheduling system**: American Airlines follows a hub and spoke system. Hub-and spoke system became norm for most of the major airline’s post 1978 after US federal government deregulated the airlines whereas few airlines follow point to point transportation.



**Source:**https://transportgeography.org/contents/chapter2/geography-of-transportation networks/point-to-point-versus-hub-and-spoke-network/

American Airlines have 6 hubs namely: DFW, CLT, MIA, LAX, PHX, ORD

**DFW:** Dallas/Fort Worth International Airport is in Texas, Primary and largest hub.

**CLT:** Charlotte Douglas International Airport is in North Carolina, Major hub on East coast.

**MIA:** Miami International Airport is in Florida, American Airlines uses MIA as a significant hub, primarily for flights to the Caribbean and Latin America.

**LAX:** Los Angeles International Airport is in California, American Airlines uses LAX as a major hub, primarily for flights to and from the West Coast and overseas locations.

**PHX:** Phoenix Sky Harbor International Airport is in Arizona, American Airlines uses PHX as a hub, notably for flights to the southern United States and Mexico.

**ORD:** Chicago O'Hare International Airport is in Illinois; American Airlines relies heavily on ORD for both local and international flight.

## **Research Questions**

## Can we predict the number of baggage items that will arrive for any given outbound flight with a 2-3 days prediction horizon based on historical data and flight information? To explore patterns and trends which can lead to better operational measures.

**Hypothesis:**

**Null Hypothesis (H0):** There is no significant relationship between historical flight data and the number of baggage items arriving for outbound flights with a 2-3 days prediction horizon.

**Alternative Hypothesis (H1):** Historical flight data, including factors such as airline code, origin, destination, scheduled departure time, and actual departure time, can be used to predict the number of baggage items arriving for outbound flights with a 2-3 days prediction horizon.

**Target Variable and Dependent Variable**

**Target Variable:** The number of baggage items arriving is the primary target variable in this analysis. It represents the outcome we want to predict.

**Dependent Variables:** The Airline Code, Flight Number, Origin, Destination, Scheduled Departure Time, Actual Departure Time, and Arrival Timestamp are the dependent variables used as input features to predict the number of baggage items arriving.

## **CHAPTER 2: LITERATURE REVIEW**

**American Airlines History**

American Airlines, established in 1930, has a storied history in the aviation industry. Initially focusing on air mail transport, the airline expanded rapidly and became a major domestic carrier in the 1930s and 1940s, introducing coast-to-coast passenger service and pioneering the concept of flight attendants. Post-World War II, the company continued to grow, modernizing its fleet by incorporating jet aircraft and expanding its domestic and international routes.

Through mergers and acquisitions, such as acquiring Air California, AirCal, and TWA, American Airlines significantly extended its network. Notable for its industry innovations, American Airlines introduced the first frequent flyer program, AAdvantage, revolutionizing loyalty programs. However, financial challenges led to a bankruptcy reorganization in 2011, culminating in a merger with US Airways in 2013, forming the world's largest airline at that time. Continuing to enhance its services and customer experience, the company remains a key player in shaping commercial aviation. American Airlines is driven by the motto” to care for people on life’s journey”. They offer thousands of flights to over 350 destinations over 60 countries. There are many departments in the Airlines namely Maintenance, Operations, Customer Experience, Revenue Management, Network Planning, Crew Planning, Flight Operations, Marketing & Sales, Airport Operations, Loyalty, Network Planning. Certainly! American Airlines has a compelling history filled with milestones, innovations, and significant developments in the aviation industry. Here's a more detailed look at the history of American Airlines. American Airlines has a history of introducing significant industry innovations.

In 1981, it introduced the world's first frequent flyer loyalty program, AAdvantage, revolutionizing customer loyalty practices in the airline industry. The company's focus has remained on improving customer experiences, improving services, and investing in technological improvements.

**Historical Evolution of Baggage Handling**

In air travel, ensuring the safe and timely arrival of passengers’ luggage is a complex and well-executed behind the scenes as tourists take the opportunity to easily check-in and collect their bags, staff dedicated custodians work tirelessly to ensure this smooth experience. The luggage is then transported through conveyor belts to specialized personnel who load it onto wagons or baggage dollies. Upon the aircraft's arrival, the luggage is unloaded, transported to a collection point near the terminal, and then delivered to passengers via conveyor belts or carousels. Historical baggage handling systems initially used simple conveyor belts like those in open mining systems. Over time, as airports and the volume of luggage grew, more advanced facilities such as round conveyor belts, sorting lanes, and storage facilities were introduced. In 1971, the first automated baggage handling system was developed by BNP Associates. This automation marked a significant advancement in the efficiency of baggage handling. Automation brought about a central control system using specialized baggage handling software. Today, mobile baggage handling software allows managers to monitor and address issues remotely via mobile phones. Modern baggage handling systems can be broadly categorized into three types which are Conveyor Technology which involves low to medium-speed transportation of loose luggage on a conveyor belt. A vast network of conveyor belts and junctions ensures automated transport to various destinations. Tilt-Tray Loop Technology with high-speed sortation system uses trays to handle baggage items of different shapes and sizes. Each item is placed on an individual tray, allowing for efficient tracking and tracing. The Individual Carrier System (ICS) Technology is done by each piece of luggage is placed inside a container, cart, or tray and is tracked throughout its journey. This system is suitable for regional airports and large airport hubs, offering higher speeds and efficient sorting. Security Screening Integration Post-9/11, airports worldwide integrated baggage screening directly into their handling systems.

**Technological Advancements of American Airlines**

In covid-19, not just American Airlines, but the entire aviation business, collapsed owing to global upheavals. Bad weather and a lack of pilots resulted in an excessive number of flight cancellations. According to Reuters, more than 100,000 flights in the United States have been cancelled because of pandemic. In that same time, American Airlines shifted to the cloud and set up data centers for both customers and operations. This system allows us not only to understand how events have affected american airlines in the past but also improve customers and operational outcomes as they happen, so that we can use real-time data from many major moving parts of the world's largest airline. American Airlines has collaborated with Microsoft to make Azure its preferred cloud platform for airline apps and critical workloads. AI, machine learning, and data analytics are being applied to every aspect of the company's operations, from reducing taxi time (saving thousands of gallons of jet fuel per year and giving connecting customers more time to make their next flight) to putting real-time information at the fingertips of maintenance personnel, ground crews, pilots, flight attendants, and gate agents. Airlines' IT departments are risk averse, but it doesn't mean the industry is anti-innovation. Consider American Airlines, which, in collaboration with IBM, pioneered the first reservation system in 1960. Some technological changes, such as program updates or real-time notifications, exist on a small scale, with improvements woven into the bigger system. According to Pilla, the purpose of AI is not to create a solution that fits within the present technological framework. AA, on the other hand, desired to create a modest prototype and web application that could exist outside of the traditional technology stack. In order to alleviate congestion in baggage handling, AA developed an application for station managers that helped determine which bags to treat first. American Airlines tried to collect data about a bag from various points throughout the luggage process in order to document what actually happened to a bag, including locating it if it became misplaced. It is feasible to forecast in real time

whether a bag is going to become a problem bag, allowing it to be avoided. Examine the luggage

processes to identify trouble areas and document compliance with any service level agreement. To make it easier to interpret the acquired data, data items must be strictly specified, and the same set of data definitions must be used across all data sources. The data sources may send data in a variety of formats. Then, all formats should be converted to the common format.

**Challenges and Issues in Baggage Handling**

There are various reasons for missing bags when it comes to baggage handling. One of the most common causes of lost baggage is because the connecting flight departs from a different gate than the arriving one. When a plane lands, the most crucial thing is to unload the bags for the passengers on the journey. Where all of luggage on flight is placed on the conveyor for pickup, transfer luggage is set aside to be sorted. If there is a big volume of luggage, luggage may be accidentally placed with baggage for another connecting aircraft when the transfer luggage is sorted by flight number and destination. Baggage handlers are working incredibly hard to load and unload aircraft due to the high volume of arriving and departing flights.

Despite being careful with the project, there are a few human errors due to severe weather and other factors such as heavy aircraft capacity, which forces airlines to halt a few bags and send them on the following flight. There is a possibility of an electronic device appearing in a suitcase during a security check, thus airlines must cease scanning the luggage for several reasons. Most airports have mastered the baggage procedure through a succession of security checks, conveyor belts, and hand-offs — some of the larger ones even have sophisticated robots — but there are still some hubs that have little to no baggage technology. The main issue is a labor shortage, and it might be difficult to forecast aircraft capacity when there is no appropriate analysis of how many bag handlers should be present for each flight that arrives.

**Impact of Baggage Handling on Customer Satisfaction**

Customer satisfaction is related to baggage handling systems. It is probable that luggage handling will be a major issue at busy airports. The level of satisfaction is determined by different aspects of baggage handling, such as checkin, waiting time, and luggage collecting. Baggage handling is critical in the airline sector for transporting bags from origin to destination. With the increase of air passenger travelling nowadays, it is critical to report the bags securely to the destination. Because of the large number of passengers, many airports have been automated. It is critical in the future to increase the use of automated machines, such as artificial intelligence, robots, and blockchain technology, to improve the speed, accuracy, and security of baggage handling processes. If a customer's luggage experience is negative, it will have an impact on the airline. There was less awareness in the past, but now everything is on the internet. Every customer review is critical to the airline's success. Bag Handling System must work on each customer bag because there could be a mistake that causes a bag delay or a lost bag. The way they compensate will also play a role in the favourable perception of a reputable airline.

**Best Practices in Baggage Handling**

The baggage handling system is critical for the transfer of consumers' bags. A well-trained workforce is essential for appropriate luggage handling. A high level of service should be provided when bags arrive at their destination. Maintain constant and clear communication with passengers about the status of their luggage. When there are delays, airlines should provide sufficient explanations as well as an expected time for the bag to arrive. When there is a connecting flight, airlines should build effective ground handling teams with enough staff to execute quickly and transfer all luggage in the shortest amount of time. Regular reconciliation checks aid in the prevention of baggage-related errors and improve operational dependability. Quality control must be performed on luggage handling equipment such as conveyors, belts, and machinery on a regular basis. To prevent security threats, effective security checks should be implemented. To have a positive perception of airlines, staff should be highly trained to handle all queries, handle delayed bags, and treat customers with empathy and provide them with appropriate solutions. When luggage goes missing, suitable compensation should be provided so that the consumer is satisfied at the end of the day. Several recommended practices must be followed while constructing a predictive model to estimate the quantity of luggage items for outgoing flights over a 2-3-day prediction horizon. The first step is to acquire a comprehensive dataset that contains various scenarios and pertinent past flight data. As a result, careful treatment of missing data and outliers is required during data pre-processing, as is adequate encoding of categorical variables such as airline codes.

**Future Trends and Recommendations**

There are numerous remedies to American Airlines' shortcomings and issues. Automation can eliminate manual errors and improve overall baggage handling efficiency. Large numbers of bags can be handled using predictive analytics and machine learning. Baggage handling system (BHS) should provide proper resource management. Prior planning is required for departure flights to arrive on time and to supply workers and resources. Staff will be reduced by forecasting how many flights will arrive at a certain moment. If a customer's bag is delayed, they should be able to follow it in real time and receive notifications about the delay. Partnership with Airport Authorities Encourage engagement with airport authorities and other airlines to build industry-wide standardized processes and technologies. This has the potential to result in a more seamless and linked baggage handling network. To protect passenger information throughout the baggage handling process, data security measures must be used. These steps must be made to improve American Airlines' baggage handling system, assuring a smoother, more efficient.

**Conclusion**

Airline departments provide a variety of functions, including customer service, revenue management, and so on. baggage handling system, one of the operations that works on arriving baggage to report baggage to the destination is considered. Baggage can go missing or be misplaced in a variety of ways. There are various phases from origin to destination, even while checking in. Even very professional workers can make blunders. Bag must be subjected to a security check. There are numerous reasons why the bag could have been halted by security for security reasons, or the bag could have been misplaced. otherwise, the bag must be moved to another lane. Because of the specified time, it is possible that it will be misplaced when sorting. Others are Build, Deliver, Load, Unload, Storage, Transfer, and Baggage Claim. American Airlines has several hubs, and the system involves connecting flights via these hubs. Based on historical data and travel information, we infer that if we anticipate the quantity of baggage items for departing flights within a 2-3-day prediction window. Airlines should increase their resources and staffing levels during peak periods to reduce the frequency of missing or delayed luggage.

The analysis aims to explore patterns and trends in historical flight data to improve operational measures related to baggage handling. This comprehensive overview highlights the intricate nature of American Airlines' operations, emphasizing the significance of the Baggage Handling System in ensuring a positive passenger experience. The study seeks to uncover patterns and trends in past flight data to improve operational baggage handling techniques. This in-depth look into American Airlines' operations emphasizes the importance of the Baggage Handling System in delivering a great passenger experience.

# CHAPTER 3: METHODOLOGY

## **Software**

## **Snowflake** Snowflake is the cloud-based warehouse that offers data storage and analytics services. It does not have their own infrastructure that offers pay per use model. Snowflake can be used for creating pipelines and visual dashboards as well. It can also process semi structured data without need for ETL tools. Snowflake has in built performance optimization with its micro partitions and cluster keys. It can allows for data backup and recovery as well as data capture and sharing. With its architecture it perform query processing in effective way. Snowflake provides us with dashboards and tables by default and also worksheets that creates folders that stores queries to run. Snowflake has web based UI for connecting and can also connect with CLI as well as with JDBC drivers, through native connectors available in ETL tools. Snowflake has a feature that allows for autoscaling of the virtual warehouses that processes queries and works for DML operations.

## **Pyspark**-

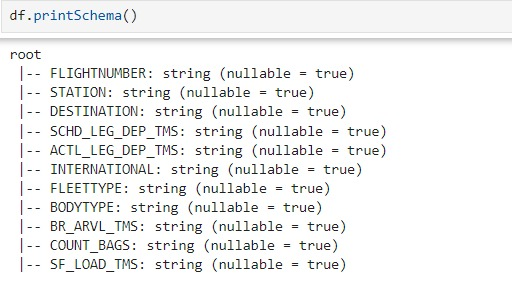
## Pyspark is an Python API for spark. PySpark allows for large data processing in real time with distributed environment. Pyspark supports features of spark like dataframes, spark SQL as well as for ML and streaming. With pyspark it allows to work with data using data frames and SQL queries to analyze large structured and unstructured data that allows for processing. Pyspark allows you to interact with Resilient distributed datasets(RDD) in spark and python programming. Pyspark allows for interaction with libraries such as Py4J that can interface with JVM objects. Spark jobs can be run that takes computations for preforming tasks that are generated. Here in cluster mode driver runs one of the worker nodes. cluster mode is used to run jobs

## **Python-**

## Python is the interpreted programming language that can be easily understood and way to work. It can have its lot of applications with web applications, ml and image processing as well. There are many libraries that can be worked with python. Pandas is one of the library that can be used for data manipulation, data analysis in faster and efficient way. Pandas allows for creating data frames that can be make many computations easier. Numpy is the another such library that performs mathematical computations well and allows for creating matrices, arrays etc. Pandas works well with tabular data and performs better data cleaning tasks like handling missing values, removing duplicates etc. it can also supports for aggregations on data as well as alignment. Numpy operates well with multi dimensional data by performing various operations on the data with element wise as well.

## **Data Collection**

## Several variables are required to answer the research questions “Can we predict the number of baggage items that will arrive for any given outbound flight with a 2-3 days prediction horizon based on historical data and flight information? “.These variables can broadly be broken into the following categories:



**Station:** Name of the station.

**Flight Number:** A unique identifier for each flight.

**Destination:** The airport where the flight is scheduled to arrive.

**SCHD\_LEG\_DEP\_TMS:** The planned departure time for the flight.

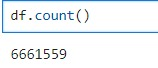
**ACTL\_LEG\_DEP\_TMS:** The actual departure time of the flight.

**FLEETTYPE:** Describes the aircraft type.

**COUNT\_BAGS:** Number of bags arriving with the flight.

**INTERNATIONAL:** 1 for International and 0 for non-international.

**BODYTYPE:** Tells if the aircraft is narrow type or wide type.



*FIG: TOTAL NO.OF RECORDS*

## This dataset provides a comprehensive set of features that can be used to analyze and predict the number of baggage items for outbound flights. It includes information about flight schedules, actual departure times, aircraft types, baggage counts, and other relevant details. Depending on your specific objectives, you can employ various data analysis and machine learning techniques to extract insights and build predictive models. The total number of records in the dataset are 6661559.

## **Data Wrangling**

The process of converting unprocessed data into comprehensible formats and assembling sets into a single structure for additional processing is known as data wrangling. It includes gathering raw data, cleaning it, organizing it, and putting it in a manner.

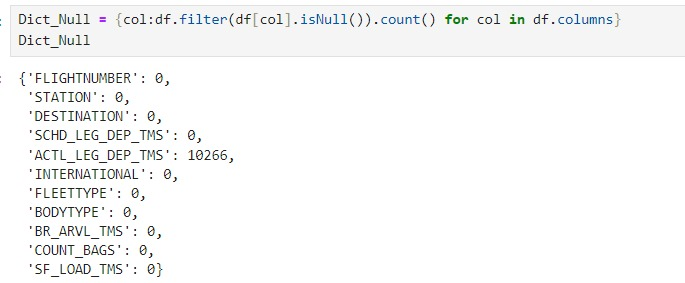
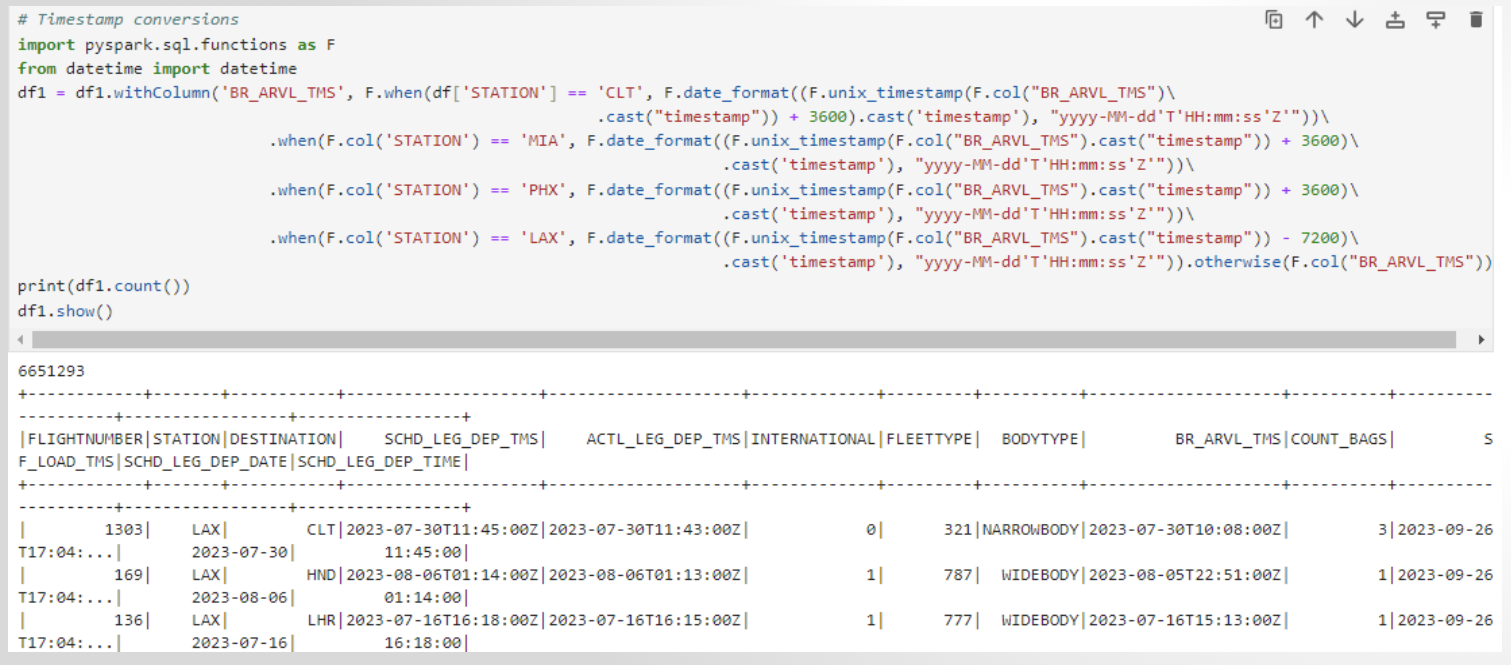


Figure: Finding null values in the data set.

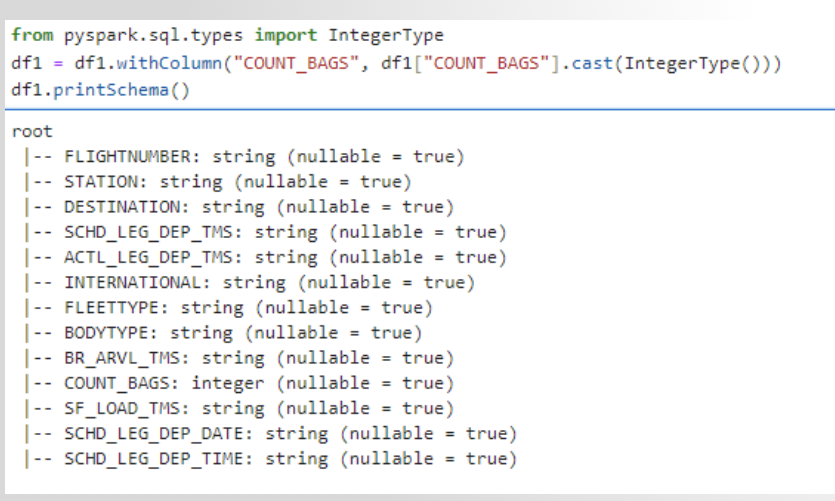
We can see that the null values are present in the ACTL\_LEG\_DEP\_TMS variable. There are 10266 null values by defining a new variable Dict\_Null to display all the null values.

**Timestamp conversions**

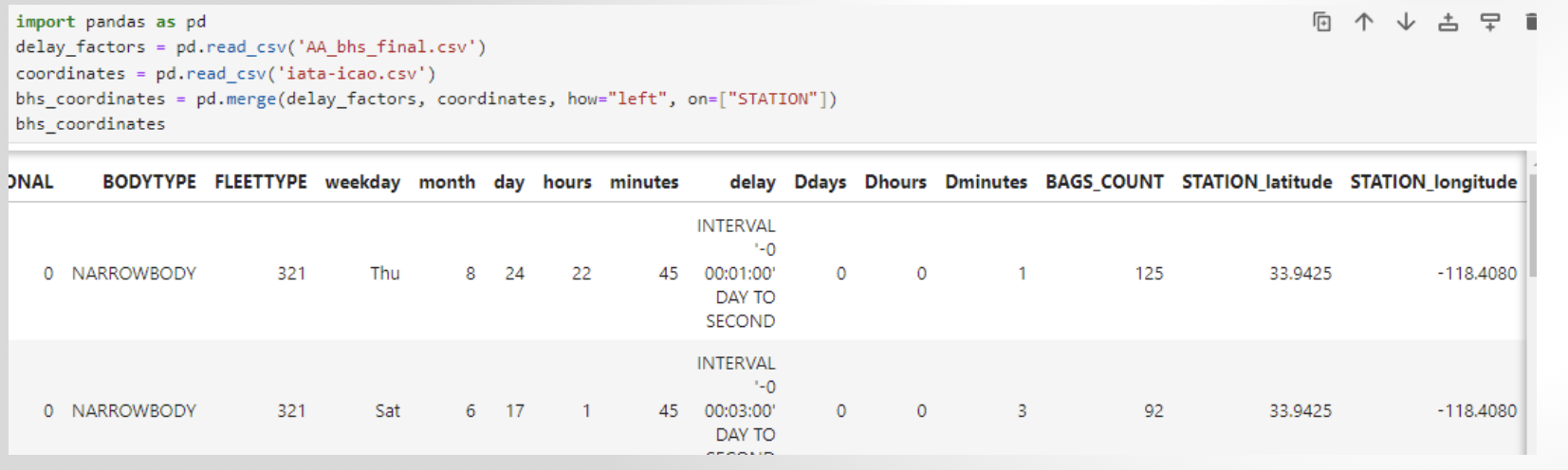
Here for each arrival timestamp entered into the system with est i.e Dallas timestamp. We convereted the station to local timestamp zone.



**Converting the bags count variable into integer**

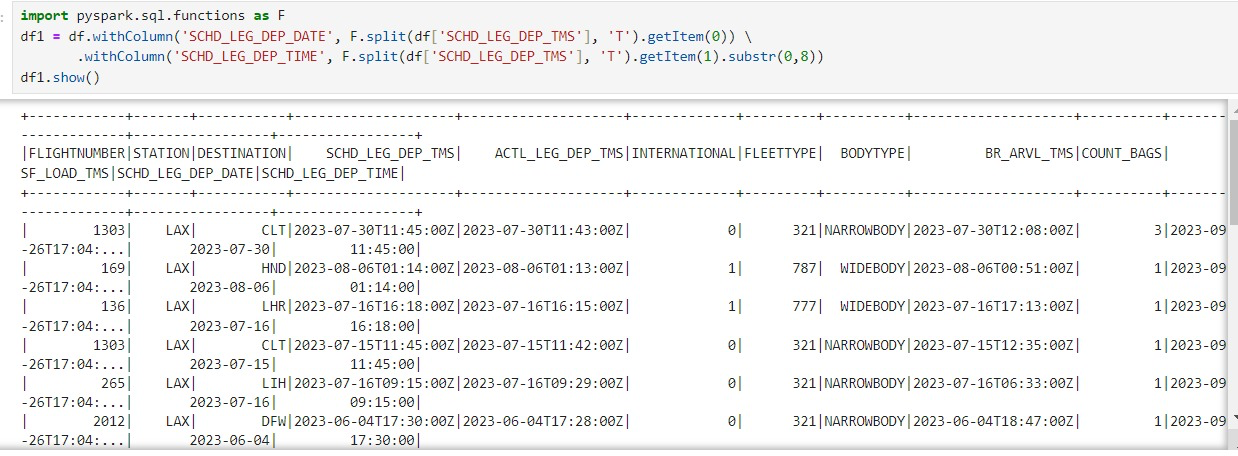


**Feature creation**

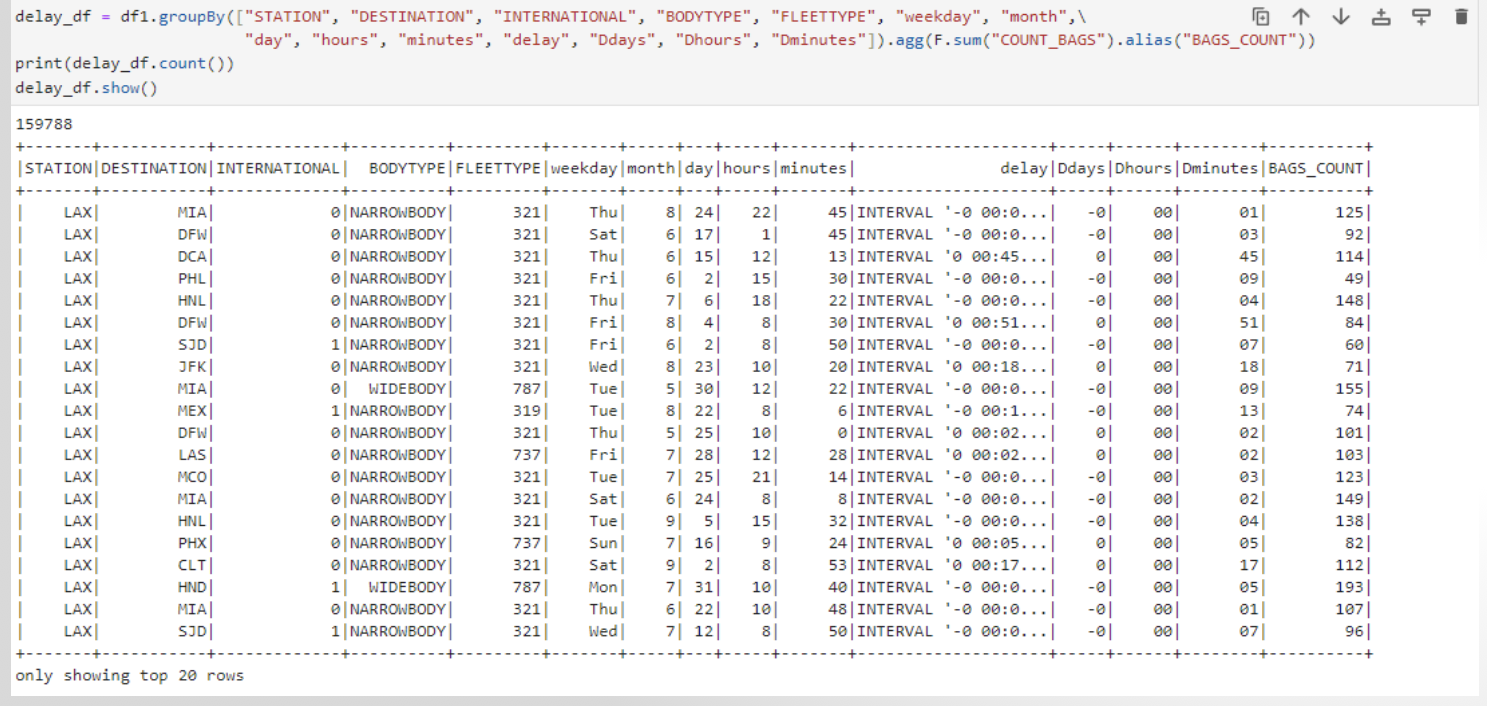


**DATA PREPROCESSING**

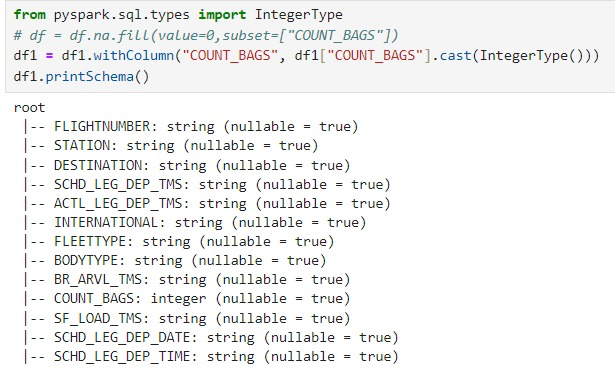
For the given SCHD\_LEG\_DEP\_TMS we split the column into two columns as date and time using withColumn function and getitem( method).



**Data Aggregration**



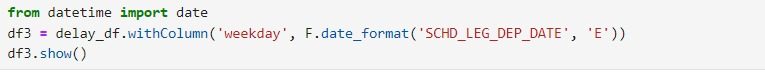
We converted the target variable from string type to intger type to predict target variable for given outbound flight.

****

Splitting the SCHD\_LEG\_DEP\_TMS into minutes and hours involves extracting the hours and minute components from a timestamp or time . The imports are required modules from PySpark. `SparkSession` is needed to create a Spark session, and `split` and `col` functions are used for string splitting and working with Data Frame columns, respectively. This assumes that you already have a Spark session named 'spark'.` SparkSession.builder`. `split(col('SCHD\_LEG\_DEP\_TIME'). This uses the `split` function to split the values in the 'SCHD\_LEG\_DEP\_TIME' column based on the ':' delimiter. This results in a column of arrays. getItem(0)`: This is used to get the first element of the array, which corresponds to the hours part.cast('int')`: This is used to cast the extracted hours as integers. The same logic is applied for the minutes part, where `.getItem(1)` is used to get the second element of the array, corresponding to the minutes. Finally, the results are stored in two new columns, 'hours' and 'minutes', in the DataFrame `df3`.This code assumes that the 'SCHD\_LEG\_DEP\_TIME' column contains values in the format 'HH:MM'. If your actual data format is different, you may need to adjust the code accordingly.

****

For the scheduled departure date we converted date into weekday column.

****

We converted scheduled date column into month and day columns

****

Here we considered the following columns "STATION", "DESTINATION", "SCHD\_LEG\_DEP\_TMS", "SCHD\_LEG\_DEP\_DATE", "SCHD\_LEG\_DEP\_TIME", "INTERNATIONAL", "BODYTYPE", "FLEETTYPE" as unique parameters to find the particular flight in a particular date. Based on these columns we changed the equation of bags arrived every minute to bags arrived per flight.



Afterthe data preprocessing webought down the data from 6661559 to 1597788 records.

****

# CHAPTER 4: EXPLORATORY DATA ANALYSIS (EDA)

## **Data Description**

Station - Name of the airport or station where the flight information is recorded.

Flight Number - Unique identifier assigned to each flight for tracking and reference.

Destination - The airport where the flight is scheduled to arrive, indicating the final destination of the flight. ACTL\_LEG\_DEP\_TMS - Actual departure time of the flight, indicating the real-time when the aircraft departs.

FLEETTYPE -Descriptive attribute indicating the type of aircraft used for the flight (e.g., narrow type or wide type).

COUNT\_BAGS - Number of bags associated with the flight, providing information about the baggage load.

BODYTYPE -Describes the body type of the aircraft, specifying whether it is a narrow type or wide type.

INTERNATIONAL: 1 for International and 0 for non-international.

Weekday – It defines the which day of the week

Month – This column states which month is it

Minute – minute column is used for finding out how many flights per minute.

## **Response Variable**

## BAGS\_COUNT is the considered as the independent variable as it is the prime thing to be figured out for every flight. Our main aim is to find out how many number of bags per flight.

## **Categorical Variables**

categorical or qualitative variables were considered for inclusion in the model

*Based on the information provided, here are the assumed data types for each variable:*

*1. Station: String*

*2. Flight Number: String or Numeric (depending on the format of the flight number)*

*3. Destination: String*

*4. SCHD\_LEG\_DEP\_TMS Timestamp or String (if not converted to a timestamp format)*

*5. ACTL\_LEG\_DEP\_TMS:Timestamp or String (if not converted to a timestamp format)*

*6. FLEETTYPE: String (assuming it contains categories like "narrow" or "wide")*

*7..BODYTYPE: String (assuming it contains categories like "narrow" or "wide")*

**‘**

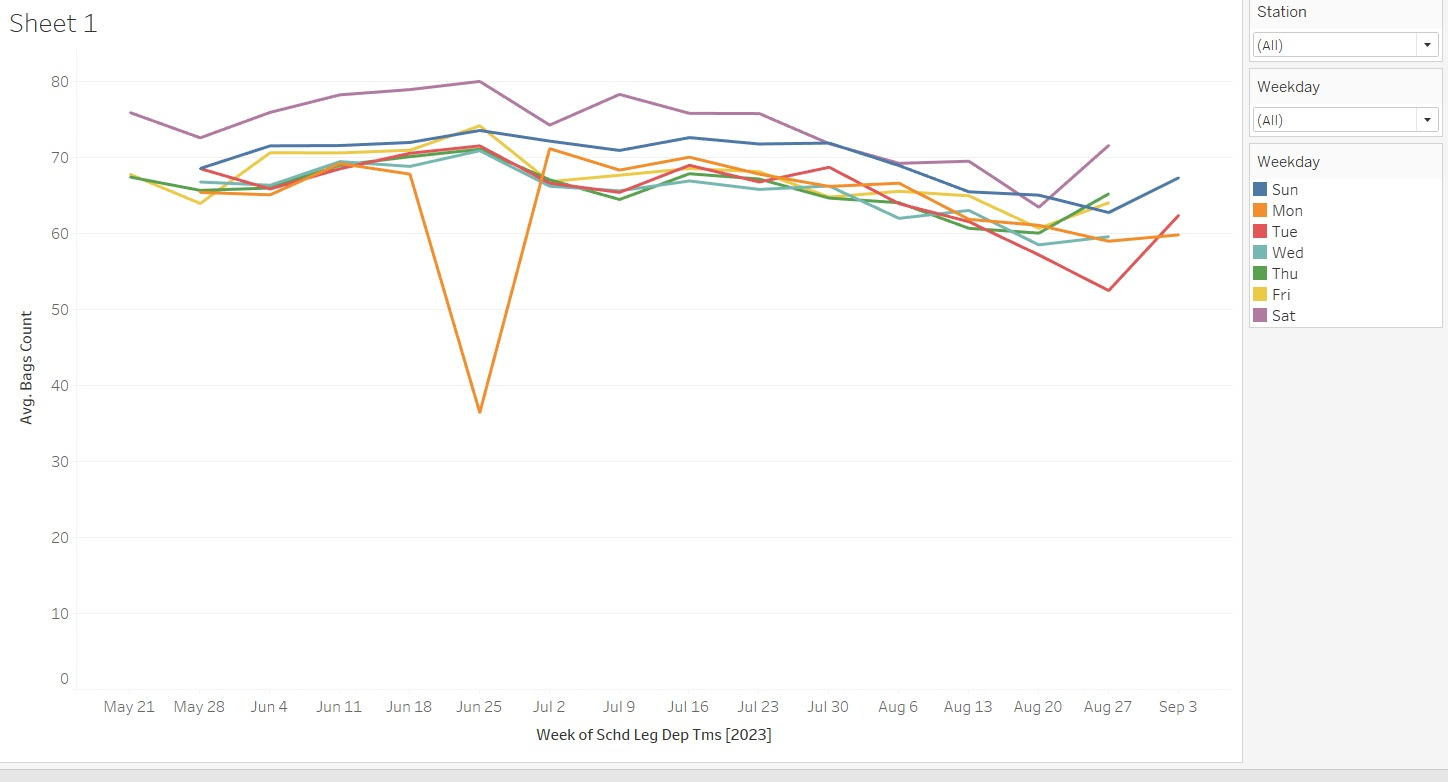
**Numeric Variables**

COUNT\_BAGS: Integer

INTERNATIONAL: Integer (Binary: 0 or 1)

Minute – Integer

Hours – Integer



The above figure illustrates about the week of Schd Leg Dep Tms on x-axis and Avg Bags count on y-axis with a relation between Avg bag count and the week scheduled departed flights. There is no particular pattern in the figure.

**Exploratory Data Analysis:**

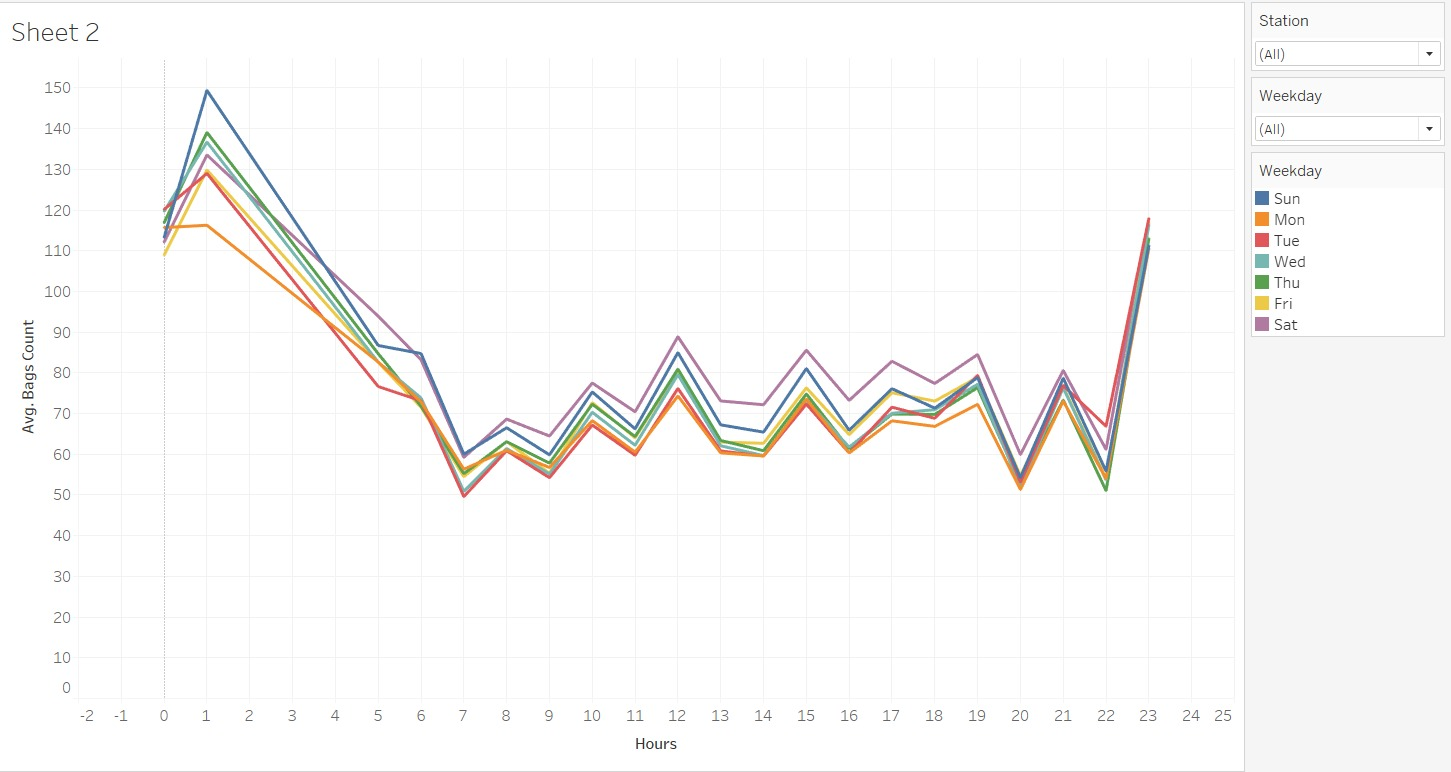
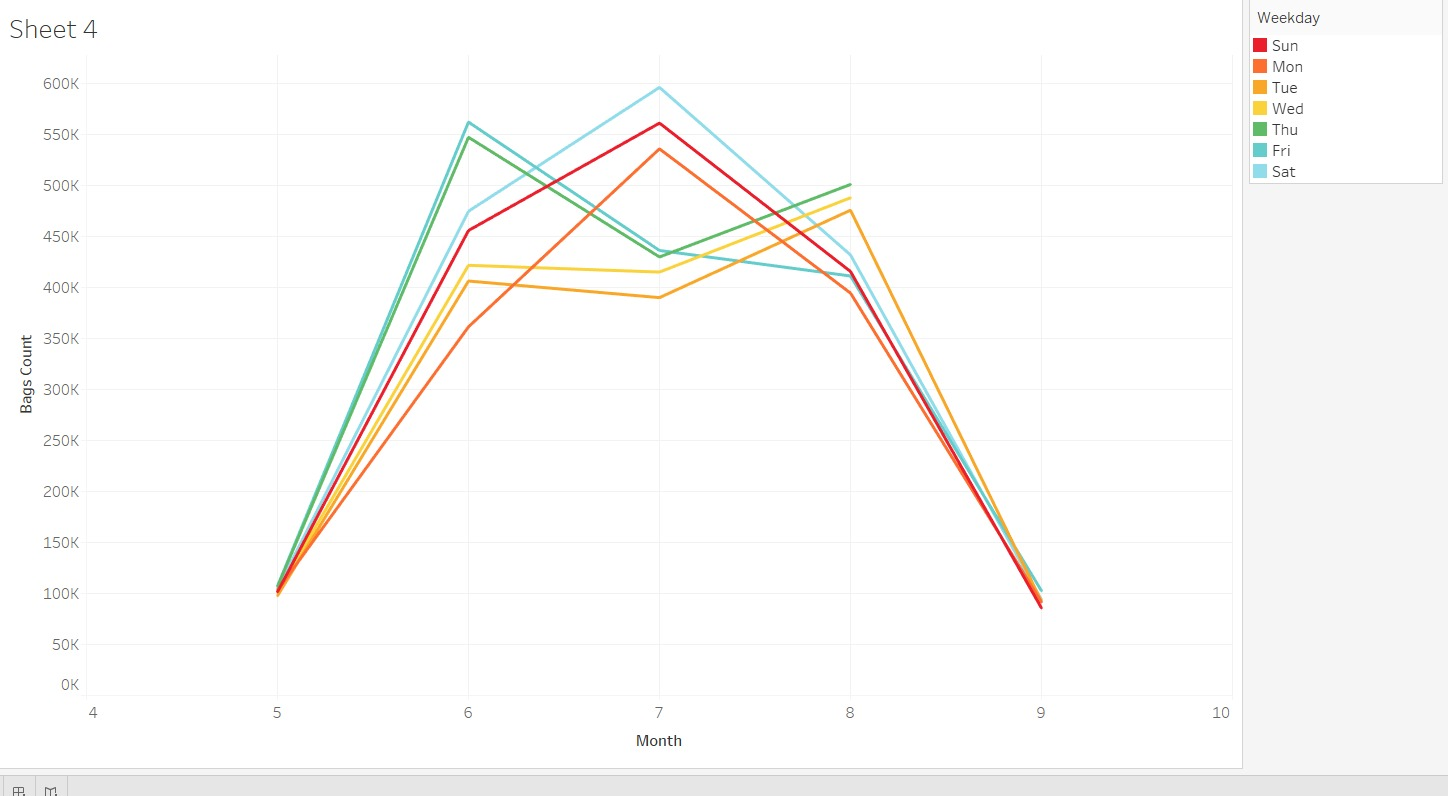
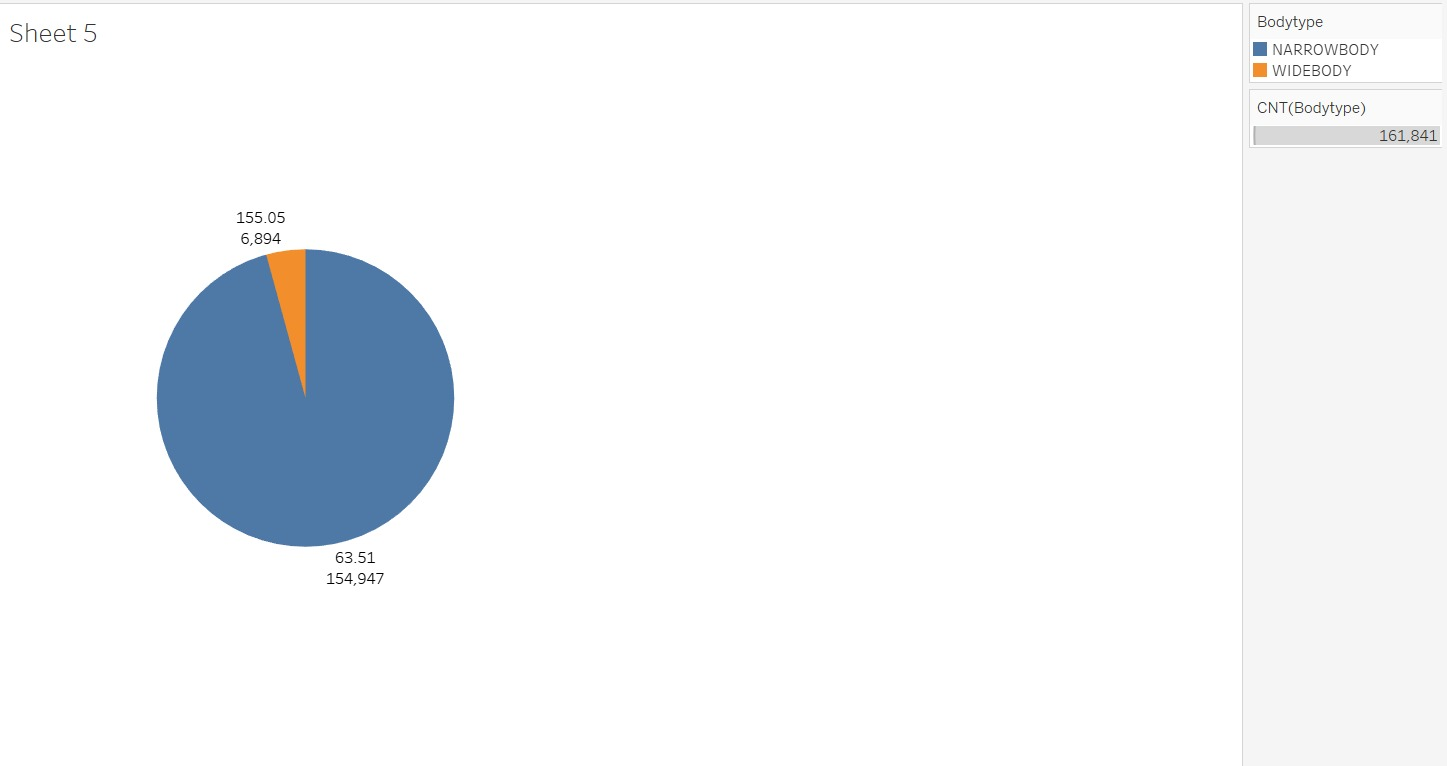


Fig: Relation between Avg Bags Count and Hours.

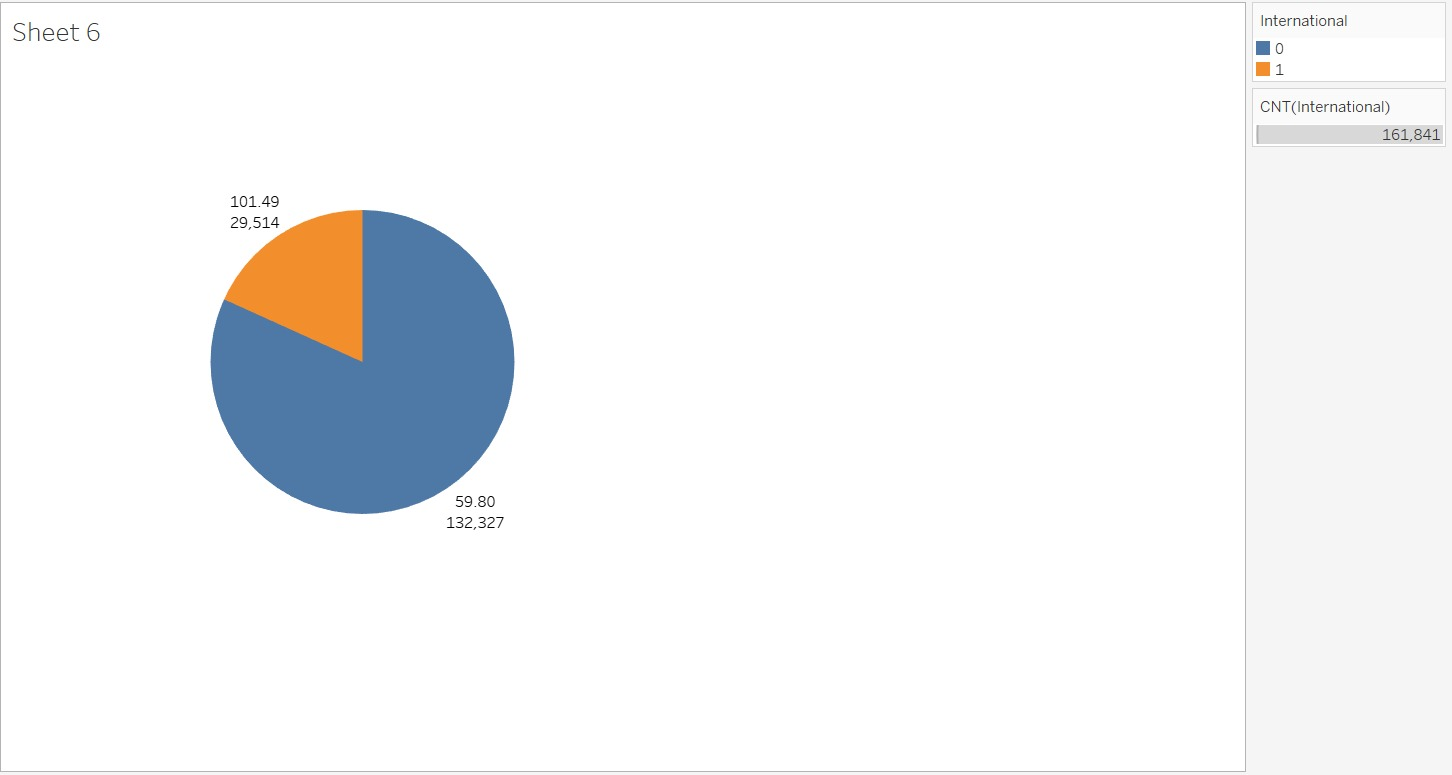
The above figure illustrates by comparing both the graphs, we can see that the Hourly analysis of number of bags count is almost the same for all the days. The relationship has same trend with bag count and hours for each weekday. We can see the average bag count on Sundays the bag count is high and Mondays and Tuesday average bag count is less.



The above figure illustrates that It’s been observed that sum of the bag count is high for the month of June on Fridays. In the month of July the sum of bag count is high on Saturdays. There is a high bag count on Thursdays for august month. We can see seasonality for these three months.



Observation: We can see that flights with Narrow body type are more in number compared to Widebody type. There are 6,894 widebody flights with 155 bags per flight and 154,947 narrow body type with 64 bags per flight.

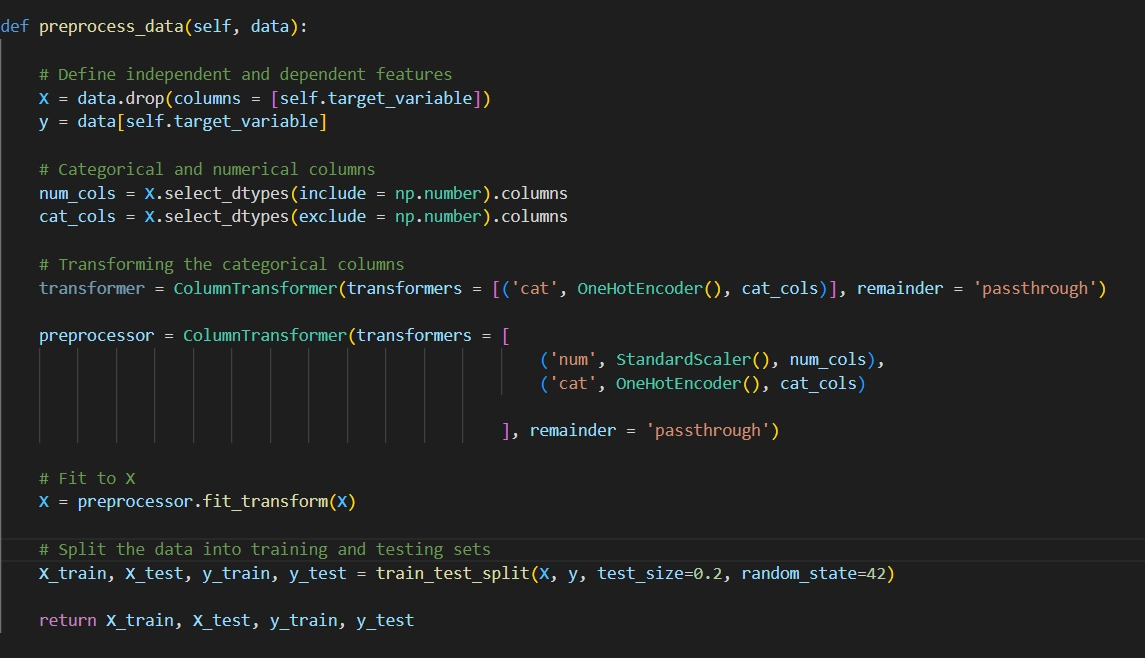


Observation: The above pie chart illustrates that the international flights are 29514 which carries 101 bags. Domestic flights are 132,327 with 60 bags per flight.

# CHAPTER 5: MODELING

We preprocessed the data by defining independent and dependent variables, Categorical and numerical columns and then transformed the categorical columns using OneHotEncoder, which is used for converting

categorical information to a format which can be fitted into a machine learning algorithm in order to improve the accuracy of prediction model. We used StandarScaler to remove the mean and scale the data to unit variance and split the data into training and testing sets.



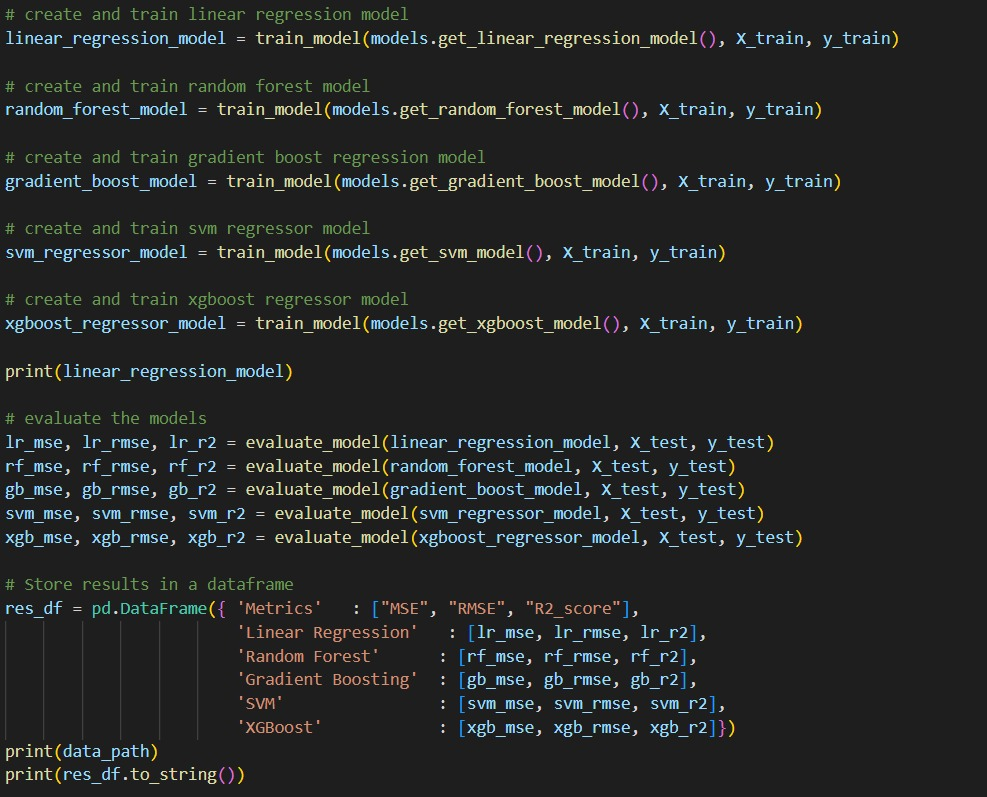
**Models used : Linear Regression Model:** Modeling the relationship between a scalar answer and one or more explanatory variables—also referred to as dependent and independent variables—is done using a linear technique called linear regression. Simple linear regression refers to the situation where there is just one explanatory variable. Mean Square Error of the Linear Regression Model obtained is 579.6173, RMSE of 24.075 and r2 value of 0.6477.

**Random Forest Model**: Random forests work by building a large number of decision trees during the training phase. They are an ensemble learning technique for classification, regression, and other problems. The class that the majority of the trees choose is the random forest's output for classification problems. The mean or average prediction made by each individual tree is returned for regression tasks. The tendency of decision trees to overfit to their training set is compensated for by random decision forests. We obtained the MSE of 336.494,RMSE of 18.343 and r2 value of 0.79.

**Gradient Boost regression model:** Machine learning techniques like gradient boosting are applied to applications like regression and classification. An ensemble of weak prediction models, or models with little assumptions about the data—typically straightforward decision trees—is provided as a prediction model. We obtained the MSE,RMSE and r2 values 570.95,23.89 and 0.65 respectively.

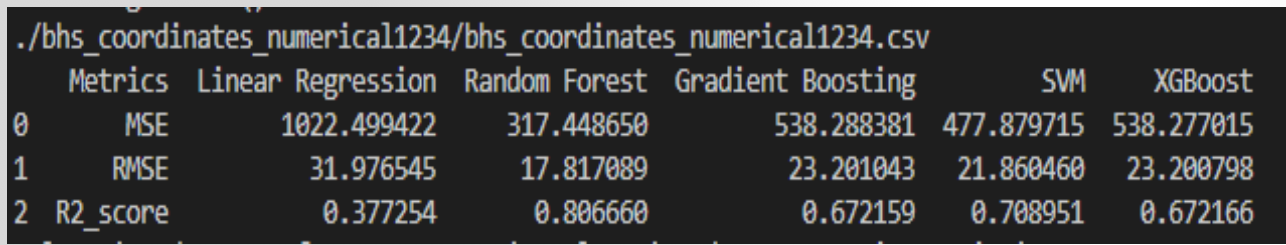
**SVM Regressor model:** A machine learning approach used for regression analysis is called Support Vector Regression (SVR) or SVM regression. Rather than fitting a line to the data points, as is done with typical linear regression methods, it determines a hyperplane that best matches the data points in a continuous space. We obtained the MSE,RMSE and r2 values 524,37,22.903, 0.681 respectively.

**Xgboost model:** A machine learning library called XGBoost (eXtreme Gradient Boosting) uses the Gradient Boosting framework to implement supervised machine learning models. We obtained the MSE,RMSE and r2 values 571.36, 23.903,0.652 respectively.

**Discussion/Interpretation:**

We can see the r2 value for Random Forest is near to 1 indicating that it is the best possible machine learning model among others.

**CONCLUSION**



In conclusion, analysis involved a comprehensive preprocessing of the data, encompassing the definition of independent and dependent variables, the identification of categorical and numerical columns, and the transformation of categorical columns using OneHotEncoder. After all the analysis mean squared and root mean squared are randomforest with 80 percentage. So, we consider randomforest model to predict the target variable from various variables.

# REFERENCES

1-

<https://www.researchgate.net/profile/Rene-Sorensen-2/publication/338403526_Deep_Reinforcement_Learning_for_Route_Optimization_in_Baggage_Handling_Systems/links/5ef5a88fa6fdcc4ca4311ad8/Deep-Reinforcement-Learning-for-Route-Optimization-in-Baggage-Handling-Systems.pdf>

This paper explores the application of Deep Reinforcement Learning (Deep RL), specifically the Double Deep Q-Network (Double DQN), to optimize routing in Baggage Handling Systems (BHS) at airports. The study compares the performance of Double DQN with traditional routing methods, specifically Dijkstra's shortest path algorithm calculated statically and dynamically. The paper focuses on route planning for Baggage Handling Systems in airports, emphasizing the complexity and dynamic nature of BHS, which can lead to congestion and deadlocks.

APA-

Sørensen, R. A., Nielsen, M., Karstoft, H., & Yurish, S. Y. (2019, March). Deep reinforcement learning for route optimization in baggage handling systems. In *Proceedings of the 1st International Conference on Advances in*.

2- [A simulation approach to modelling baggage handling systems at an international airport - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1569190X15303129)

The paper presents a microscopic simulation model for a baggage handling system (BHS) that integrates all baggage-related subsystems at Santiago International Airport in Chile. The BHS model covers passenger arrival, check-in queues, baggage check-in, security screening, sorting, transport to the aircraft, and loading. The simulation platform developed allows for the analysis of individual subsystems and their interactions. The study aims to address the challenges faced by the airport due to the growing demand, exceeding the BHS's operating capacity.

[Sci-Hub | Intelligent Component-Based Automation of Baggage Handling Systems With IEC 61499 | 10.1109/tase.2008.2007216 (hkvisa.net)](https://sci-hub.hkvisa.net/https:/ieeexplore.ieee.org/abstract/document/4745827)

APA-

Zhang, T., Ouyang, Y., & He, Y. (2008). Traceable air baggage handling system based on RFID tags in the airport. *Journal of Theoretical and Applied Electronic Commerce Research*, *3*(1), 106-115.

3- [Internet-of-Things-augmented dynamic route planning approach to the airport baggage handling system - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0360835222007902)

This paper proposes an IoT-augmented dynamic route planning approach for airport baggage handling systems (BHS). The key contributions of the study include incorporating real-time baggage tracking information (BTI) and device disruption information (DDI) into the baggage route planning algorithm based on Internet-of-Things (IoT) technologies. In summary, the paper introduces an innovative approach to dynamic route planning for airport baggage handling systems by integrating IoT technologies. The proposed method addresses the limitations of traditional static routing approaches and considers real-time conditions, disruptions, and control errors, demonstrating its efficiency through computational results and a real airport case study

APA-

Yang, X., Feng, R., Xu, P., Wang, X., & Qi, M. (2023). Internet-of-Things-augmented dynamic route planning approach to the airport baggage handling system. *Computers & Industrial Engineering*, *175*, 108802.

4.

[Sci-Hub | Traceable Air Baggage Handling System Based on RFID Tags in the Airport | 10.3390/jtaer3010011 (hkvisa.net)](https://sci-hub.hkvisa.net/10.3390/jtaer3010011)

[JTAER | Free Full-Text | Traceable Air Baggage Handling System Based on RFID Tags in the Airport (mdpi.com)](https://www.mdpi.com/0718-1876/3/1/11)

In summary, the paper underscores the utility of RFID in aviation, particularly in air baggage handling. It introduces the IATA RP1740c protocol, describes a distributed traceable application, and presents results from an RFID-based baggage tracking experiment in Beijing Capital International Airport. The advantages of RFID adoption, including improved accuracy and real-time monitoring, are discussed, along with the associated economic benefits.

APA-

Zhang, T., Ouyang, Y., & He, Y. (2008). Traceable air baggage handling system based on RFID tags in the airport. *Journal of Theoretical and Applied Electronic Commerce Research*, *3*(1), 106-115.

5.

[Airport baggage handling system simulation modeling using SysML | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/7093764)

[Sci-Hub | Airport baggage handling system simulation modeling using SysML. 2015 International Conference on Industrial Engineering and Operations Management (IEOM) | 10.1109/ieom.2015.7093764 (hkvisa.net)](https://sci-hub.hkvisa.net/https:/ieeexplore.ieee.org/abstract/document/7093764)

In response to a substantial increase in passenger numbers at Taiwan Taoyuan International Airport (TPE), surpassing its designed capacity, this study addresses the crucial challenge of efficiently operating the airport's baggage handling system. APA-

Lin, J. T., Shih, P. H., Huang, E., & Chiu, C. C. (2015, March). Airport baggage handling system simulation modeling using SysML. In *2015 International Conference on Industrial Engineering and Operations Management (IEOM)* (pp. 1-10). IEEE.

6. [A generalised data analysis approach for baggage handling systems simulation | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/6377979)

[Sci-Hub | [IEEE 2012 IEEE International Conference on Systems, Man and Cybernetics - SMC - Seoul, Korea (South) (2012.10.14-2012.10.17)] 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC) - A generalised data analysis approach for baggage handling systems simulation | 10.1109/icsmc.2012.6377979 (hkvisa.net)](https://sci-hub.hkvisa.net/10.1109/icsmc.2012.6377979)

In summary, the paper focuses on proposing a standardized approach for assessing the performance of baggage handling systems in airports. It employs discrete event simulation techniques and introduces a set of measures that can be applied not only to baggage handling systems but also to general network systems. The evaluation results provide insights into operational characteristics, including metrics like peak throughput, in-system time, and system recovery time. The study contributes to the understanding and optimization of airport operations.

APA-

Le, V. T., Zhang, J., Johnstone, M., Nahavandi, S., & Creighton, D. (2012, October). A generalised data analysis approach for baggage handling systems simulation. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1681-1687). IEEE.

7.

[Fuzzy control in the simulation model of airport baggage handling systems - IOPscience](https://iopscience.iop.org/article/10.1088/1757-899X/919/4/042017/meta)

[Sci-Hub | Fuzzy control in the simulation model of airport baggage handling systems. IOP Conference Series: Materials Science and Engineering, 919, 042017 | 10.1088/1757-899x/919/4/042017 (hkvisa.net)](https://sci-hub.hkvisa.net/https:/iopscience.iop.org/article/10.1088/1757-899X/919/4/042017/meta)

It is proposed to include into the composition of simulation models used in solving problems of efficiency analysis and design of baggage handling systems, a fuzzy algorithm that simulates the strategy of a human operator who controls the process of pre-flight passenger service and baggage handling in the airport terminal.

Romanenko, V. A., Skorokhod, M. A., & Guzha, E. D. (2020, September). Fuzzy control in the simulation model of airport baggage handling systems. In *Iop conference series: Materials science and engineering* (Vol. 919, No. 4, p. 042017). IOP Publishing.

8.

[Routing in congested baggage handling systems using deep reinforcement learning - IOS Press](https://content.iospress.com/articles/integrated-computer-aided-engineering/ica190613)

[Sci-Hub | Routing in congested baggage handling systems using deep reinforcement learning | 10.3233/ICA-190613 (hkvisa.net)](https://sci-hub.hkvisa.net/10.3233/ICA-190613)

This paper proposes to use a single global deep reinforcement learning agent to route a fleet of baggage-totes to continuously pick up and deliver baggage in simple yet functionally realistic simulations of baggage handling systems.

APA-Sørensen, R. A., Nielsen, M., & Karstoft, H. (2020). Routing in congested baggage handling systems using deep reinforcement learning. *Integrated Computer-Aided Engineering*, *27*(2), 139-152.

9.

[Techniques for Detecting and Tracking of Baggages in Airports | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/document/8081953)

[Sci-Hub | [IEEE 2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT) - Bangalore, India (2017.3.16-2017.3.17)] 2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT) - Techniques for Detecting and Tracking of Baggages in Airports | 10.1109/ICRAECT.2017.78 (hkvisa.net)](https://sci-hub.hkvisa.net/https:/ieeexplore.ieee.org/document/8081953)

In this paper the different, effective and advanced technologies for detecting and baggages in airports being adopted in the past one and half decade (2000 to 2015) have been studied and being compared.

APA-

Kishan, K. K., & Prashanth, K. M. (2017, March). Techniques for detecting and tracking of baggages in airports. In *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)* (pp. 333-338).