# Analysis Case Study and Analytics Project: Heathrow Airport Operations Improvements

## Introduction
The aviation industry stands as a cornerstone of modern transportation, facilitating global connectivity and economic growth. At the heart of this network is London's Heathrow Airport, a critical hub that orchestrates the flow of over 80 million passengers each year to more than 200 destinations around the world. The complexity of operations at Heathrow is unrivaled, with a significant proportion of its passengers relying on the airport for connecting flights. This places immense pressure on the operational capacity of the airport to deliver a seamless transit experience.

Heathrow's commitment to ensuring "happy passengers, traveling on time, with their bags," is not just a service promise but a logistical masterpiece behind the scenes. The linchpin in this endeavor is the Airport Operations Center (APOC), a strategic command center that collates inputs from various stakeholder groups to make real-time decisions. The initial forays into using technology for airport operations have seen substantial improvements in the efficiency of passenger and baggage flows, yet the evolving demands of air travel and passenger expectations call for a continuous enhancement of these systems.

As Heathrow Airport ventures into the next phase of operational excellence, it grapples with the integration of advanced analytics and machine learning to further refine its processes. The objective is clear: to maintain its stature as a premier hub by not only handling the sheer volume of passengers but by elevating the travel experience through smart, data-driven interventions.

More information about the case can be found in the link given below.

https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Real-Time-Analytics-Improve-Airport-Operations

## Problem Statement
Despite the advancements at Heathrow Airport, the challenges of managing transfer passengers remain daunting. The sheer volume of transit passengers, accounting for nearly one-third of the airport's traffic, adds layers of complexity to the operational framework. The goal is to provide these passengers with a frictionless transfer experience, mitigating the risk of misconnections and ensuring a timely departure with their belongings. The current systems, although robust, are not equipped to preempt the variables that can lead to service disruptions.

The problem is twofold: predicting the unpredictable and managing the manifold. Firstly, the prediction of passenger flows within the airport premises—knowing when and where passengers will arrive and identifying potential bottlenecks before they translate into delays. This requires a prognostic approach that can foretell issues and allow preemptive action to be taken. Secondly, the management of resources—whether it's reallocating staff to high-traffic areas, adjusting departure times, or expediting the journey of at-risk passengers through the airport—demands a responsive and adaptable system.

To elevate the level of service to meet the standards expected by passengers and to optimize operational efficiency, Heathrow requires a sophisticated layer of real-time data collection and analytics capabilities. The system should not only assess current data but also apply predictive models to forecast potential outcomes. The integration of such a system into Heathrow's APOC represents a significant step in using technology to harmonize the intricate dance of airport logistics with the overarching goal of passenger satisfaction and operational fluency.

Given this, we can create multiple models to solve these problems. Given our overarching problem, we are faced with goals and their specific problems that can use analytics to improve decision making and outcomes. I will be referencing Goals and their associated problems which we will try to solve (which are numbered). One goal will have at least 1 or more problem associated with it. This will help in a comprehensive analytics solution for the problem at hand.

## Goal 1: Minimize Passenger Transit Time

**Problem 1: Forecasting Passenger Flow Peaks**

**Collecting Data**
In order to develop a robust forecasting model for predicting peak passenger flows, we first need to collect an array of data that reflects the patterns and variables impacting airport traffic. This includes historical data of passenger movements within the airport, encompassing arrival and departure times correlated with specific flight schedules, as well as seasonal variations such as holidays or school breaks. We would also incorporate external factors such as local events or weather conditions that historically have had a significant impact on passenger flow.

Key data points for this purpose include:
- Timestamped entries and exits from terminal gates.

- Flight schedule archives with actual vs. planned departure and arrival times.
- Historical passenger count data from checkpoints and terminal entrances/exits.
- Dates and times of local events that may influence airport traffic.
- Weather records, especially those related to poor conditions affecting travel.
- Operational logs documenting past incidents affecting passenger flow (e.g., security breaches, technical failures).

**ARIMA Model**
Given the collected data, we will develop a ARIMA model. This ARIMA model is an extension of the ARIMA model that adds the capability to model seasonal effects, which is crucial for airport data that has clear patterns of seasonality due to factors like vacation periods and holidays.

To construct the ARIMA model, we will need to specify the following parameters:

1. **Seasonality (S)**: Determined by the periodicity of the data — for instance, weekly, monthly, or yearly cycles.
2. **Autoregression order (p)**: The number of lag observations included in the model.
3. **Differencing order (d)**: The number of times the data have had past values subtracted.
4. **Moving average order (q)**: The size of the moving average window.
5. **Seasonal autoregression order (P)**: The number of seasonal lag observations in the model.
6. **Seasonal differencing order (D)**: The number of times seasonal differences are taken.
7. **Seasonal moving average order (Q)**: The order of the seasonal moving average.

**Rationale for ARIMA Selection**
The choice of ARIMA is based on its proven track record in dealing with seasonal time series data. Airports have distinct peak periods that repeat at regular intervals (e.g., daily rushes, weekend travel spikes, holiday seasons), and these patterns can be captured and predicted using SARIMA. By identifying the model parameters that best fit the historical data (a process known as model identification), we can forecast future passenger flow with greater accuracy.

**Outcomes of Using the ARIMA Model**
The ARIMA model will enable us to:
- Accurately predict the volume of passengers at the airport at different times, which is critical for planning and resource allocation.

- Understand the impact of specific events or seasons on passenger flow, allowing us to adjust operational strategies accordingly.

- Generate detailed forecasts that can inform decision-making at various operational levels, from strategic planning to tactical responses during daily operations. For instance, if the ARIMA model predicts a particularly busy upcoming period, airport management can make decisions such as opening additional security lanes, adjusting staff shifts to cover peak times, or even coordinating with airlines to manage flight schedules more effectively.

Moreover, by continually updating the model with the latest data, we can refine our forecasts over time, thus enhancing the agility of airport operations to meet the demands of passenger flow dynamically. This agility is key to reducing passenger transit times and directly contributes to a smoother and more efficient airport experience.

**Data Dynamics**: Involves continuous, periodic data including seasonal trends and regular fluctuations in passenger volumes.

**Rerun Frequency**:

Ideally, the model should be updated at least quarterly to capture any significant shifts in travel patterns or seasonal effects.

Additionally, after any major event (like a new airline service at the airport or a major local event affecting travel), the model should be rerun to adjust the forecasts accordingly with event-driven updates.

A comprehensive annual review and recalibration of the model should be conducted annually to incorporate the latest trends, historical data, and changes in flight schedules or airport operations.

**Problem 2: Streamlining Security Checkpoints**

**Collecting Data**
The effectiveness of security checkpoints is a major factor in overall passenger transit time. To streamline this process, a comprehensive set of data must be collected that reflects the operations and efficiency of these checkpoints. The data should capture every aspect of the passenger's interaction with the checkpoint, the throughput of the system, and the allocation of staff and technology that contributes to processing speed. The data collection process would include, but not be limited to, the following elements:

- **Passenger Arrival Times**: Continuous data on when passengers arrive at the checkpoint.

- **Processing Times**: Time taken for each passenger to pass through the security process, including scanning and manual checks.

- **Checkpoint Layout**: Details on the number and type of screening lanes (e.g., standard, TSA PreCheck, priority).
- **Staffing Levels**: Number of security personnel available during different times of the day.

- **Incident Reports**: Data on any security incidents that cause delays.

- **Equipment Efficiency**: Performance data of screening equipment like metal detectors and body scanners.

**Discrete Event Simulation Model**
With the collected data, we would construct a Discrete Event Simulation (DES) model. This type of simulation is particularly well-suited to environments like airport security, where the system is influenced by discrete, individual events that occur at particular moments in time.

**Components of the DES Model**

- **Entities**: Passengers, security personnel, and equipment.

- **Resources**: Screening lanes, staff members, and security tools.

- **Queues**: Lines of passengers waiting for screening.

- **Events**: Arrival of a passenger, start and end of a screening, and opening or closing of lanes.

- **Variables**: Number of open lanes, passenger wait times, and throughput rates.

- **Clock**: Simulation time to track the flow and timing of events.

**Simulation Process**

- **Initialization**: Set up initial conditions such as the number of open lanes and available staff.

- **Input Data Integration**: Feed the model with real-time and historical data to simulate typical and peak conditions.

- **Event Handling**: Simulate events where passengers arrive, get processed, and leave the system. Each event will alter the state of the system, such as changing queue lengths and resource availability.

- **Time Advancement**: Move the simulation clock forward to the next scheduled event, recalculating resource states and event outcomes based on new data inputs and system changes.

**Rationale for DES Selection**

The rationale for choosing DES lies in its ability to model complex systems and evaluate the impact of changes without disrupting actual operations. For example, it can simulate the effect of opening an additional security lane during peak times or the potential benefits of implementing new screening technologies.

DES helps answer "what if" questions about operational changes and can be used to forecast the outcomes of different resource allocation strategies without any risk to actual airport operations.

**Outcomes of Using the Discrete Event Simulation Model**

Employing the DES model will yield multiple benefits, such as:

- Providing a detailed understanding of how different factors affect the throughput and efficiency of security checkpoints.

- Allowing management to visualize the potential impact of various changes, such as different staffing levels, checkpoint configurations, or the introduction of new technologies.

- Enabling the airport to experiment with different scenarios to find the optimal setup that minimizes wait times and maximizes passenger satisfaction.

- Offering the ability to stress-test the security system against high-volume scenarios, such as holiday travel peaks, and plan resource allocation accordingly.

In summary, the use of a DES model would offer a virtual environment to refine and optimize checkpoint operations, significantly contributing to reducing passenger transit times and enhancing the overall efficiency of airport security.

**Data Dynamics**
Based on operational data, layout changes, and passenger feedback.

**Rerun Frequency**

Since the setup of security checkpoints does not change daily, rerunning the simulation quarterly or biannually with pre-defined intervals is usually sufficient to capture any significant operational or layout changes.

Before implementing any major changes in operations or checkpoint layout, the simulation should be run to predict the impacts of such changes.

After any major operational changes, the simulation should be rerun to validate the effectiveness of these changes with a post-implementation review to refine the operations further based on actual data.

## Goal 2: Reduce Misconnections

### Problem 1: Identifying Passengers at Risk of Misconnection

To predict the likelihood of passengers missing their connecting flights, a logistic regression model will be used. This model is effective for binary classification tasks, such as determining whether a passenger will miss (1) or make (0) their connection.

**Collecting Data**

Key data required for the logistic regression model includes:

- **Flight Arrival and Departure Times**: Actual and scheduled times to capture potential delays.

- **Historical Misconnection Data**: Past data on misconnections to identify patterns or common factors.

- **Layover Duration**: Time between scheduled arrival of the inbound flight and departure of the outbound flight.

- **Passenger Mobility Information**: Data on whether passengers require special assistance, which might affect their speed through the airport.

**Modeling Variables for Logistic Regression**

Given:

1. **Flight Delay Duration** (continuous variable): The difference in minutes between scheduled and actual arrival times.

2. **Layover Duration** (continuous variable): Scheduled time available for a passenger to make the connection.

3. **Number of Terminal Changes** (categorical variable): Whether the connection requires changing terminals, which can increase the risk of misconnection.

4. **Historical Flight On-time Performance** (continuous variable): Ratio of previous on-time departures for the connecting flight.

5. **Passenger Mobility** (binary variable): 1 if special assistance is needed, 0 otherwise.

6. **Response Variable (Misconnection)**: 1 if the passenger misses the connection, 0 if they make it.

**Choosing the Right P-value for Logistic Regression**
The significance level, or p-value, in the context of logistic regression, is critical for determining which variables significantly impact the likelihood of a passenger missing their connection. This p-value helps in deciding whether to include certain predictors in the model.

**Initial Model Selection**: Start with a full model including all predictors. Utilize backward elimination starting with a higher p-value threshold (e.g., 0.10) to remove non-significant predictors iteratively.

**Refinement and Validation**: Adjust the p-value threshold to a more conventional level (e.g., 0.05) after the initial selection to refine the model. This step helps in ensuring that the model is neither overfitting nor underfitting.

**Validation Against Seasonality and Trends**
Use the output from the SARIMA model regarding expected passenger volumes and seasonality to validate the logistic model's predictions. Adjust the logistic model to ensure it aligns with the broader traffic patterns and seasonal variations identified by the SARIMA model.

**Cross-Validation**
Perform k-fold cross-validation to assess the model's performance across different subsets of data, ensuring its robustness and generalizability.

**Rationale for Logistic Regression Selection**
Logistic regression is chosen for its efficiency in binary classification problems and its ability to provide probabilities that offer more granularity than a simple yes or no prediction. This is essential for operational decision-making where degrees of risk need to be understood to prioritize actions. Moreover, logistic regression models are interpretable, which helps in explaining why certain passengers are flagged as high risk.

**Outcomes of Using the Logistic Regression Model**
Using logistic regression will allow us to:

- Determine the likelihood of each passenger making their connecting flight, which can help prioritize which passengers need assistance first.

- Understand the impact of various factors on the risk of misconnection, allowing for targeted improvements in areas like scheduling, terminal layouts, and passenger communication.
- Enhance the allocation of airport resources, such as rebooking staff, based on predicted misconnection rates, leading to more efficient operations and better passenger experiences.

- Reduce overall misconnection rates, which can lead to higher passenger satisfaction and lower costs associated with rebooking and overnight accommodations for passengers.

The logistic regression model serves as a foundational tool in the proactive management of the passenger experience at Heathrow, reducing the likelihood of misconnections, and smoothing the process for those at risk. It aligns with our larger goal of ensuring that passengers enjoy a seamless transit through Heathrow, regardless of the complexity of their travel itinerary.

**Data Dynamics**
Relies on historical data and real-time inputs about flight delays, passenger behavior, and operational changes.

**Rerun Frequency**

Given the dependency on flight schedules and operational variables that can change frequently, it's crucial to monitor the model's performance continuously.

Model coefficients should be updated quarterly to adjust for any changes in the relationships between predictors and the outcome.

After any significant operational changes or updates in airline schedules that might impact passenger connection risks, the model should be reassessed and recalibrated if necessary.

## Goal 3: Optimize Resource Distribution

**Problem 1: Allocation of Staff and Physical Resources**

**Collecting Data**
Optimal resource allocation requires precise and comprehensive data collection, focusing on the utilization and efficiency of both staff and physical assets. The types of data necessary include:

- **Staff Data**: Number of security personnel, ground handlers, and customer service agents, including their shift schedules, overtime records, and salary rates.

- **Equipment Data**: Information on the availability and operational costs of essential airport equipment such as security scanners, baggage carts, and tarmac buses.

- **Passenger Flow Data**: This includes data from the SARIMA model on expected passenger numbers to align staff and equipment needs.

- **Operational Costs**: Details on utilities costs, maintenance expenses, and other operational expenditures linked to staffing and equipment use.

- **Revenue Data**: Income generated from airport operations, including retail revenues, parking fees, and other ancillary services.

**Linear Programming Model for Cost Optimization**
Given:

1. **Xstaff** - Number of staff to hire.

2. **Xequip** - Number of equipment units to deploy.

3. **Xpass** - Number of passengers (from SARIMA output).

4. **Xfees** - Fees collected per passenger.

5. **Cprofit** - Desired profit for the year.

6. **Csalaries** - Total budget for salaries.

7. **Cequip** - Total cost of equipment use.

8. **Xutility** - Utilities cost (power, electricity, water, heating/AC, maintenance).

9. **XmiscRev** - Miscellaneous revenue (retail, parking, etc.).

Use:

A cost optimization model with the following objective function and constraints:
- **Objective Function**: Minimize total operational costs including salaries, equipment costs, and utilities, while maximizing revenue from fees and miscellaneous sources.

Total Cost= Cstaff*Xstaff+Cequip*Xequip

- **Constraints**:

**Xstaff** ≥ **Xpass**/250 (at least one staff per 250 passengers)

**Xstaff** * Average salary ≤ **Csalaries**

**Xequip** * Operational cost per unit ≤ **Cequip**

(**Xfees** * **Xpass** + **XmiscRev**) - (**Xstaff** * Average salary + **Xequip** * Operational cost per unit + **Xutility**) ≥ **Cprofit**

**Xstaff**, **Xequip**, **XmiscRev**, **Xutility** ≥ 0

Average salary ≤ $60,000 (annual)
**Csalaries** ≤ $50 million

**Cprofit** ≥ $5 million

**Rationale for Using Linear Programming**
Linear programming is ideal for this optimization problem because it allows us to define a clear objective (minimizing costs and maximizing revenues) while adhering to various operational and budgetary constraints. It provides a systematic approach to distribute limited resources effectively across various needs to meet the defined objectives.

**Outcomes of Using the Cost Optimization Model**
The implementation of this linear programming model for resource allocation will result in several beneficial outcomes:

- **Efficient Resource Use**: Optimal scheduling of staff and deployment of equipment will ensure that resources are used efficiently, reducing waste and operational costs.

- **Cost Management**: Ability to tightly control and predict monthly and yearly expenditures, aligning them with revenue streams to meet profitability targets.

- **Enhanced Operational Effectiveness**: By aligning staff and equipment with predicted passenger flows, the model supports smoother airport operations, reducing bottlenecks and improving passenger satisfaction.

- **Financial Optimization**: The model will help to strike a balance between operational costs and revenue generation, ensuring that Heathrow maintains a healthy financial status while improving service levels.

By applying this model, Heathrow Airport can ensure that it not only meets its operational goals but does so in a cost-effective manner that maximizes financial health and service quality, ultimately enhancing the overall airport experience for passengers and staff alike.

**Data Dynamics**

Depends on budgetary constraints, staff availability, and operational demands.

**Rerun Frequency**

Given the fluctuating nature of airport operations and varying passenger volumes, the model should be run at least every one to two months to ensure optimal resource allocation.

Should be rerun following significant changes in operational budgets, staff changes, or shifts in passenger demand patterns(event-driven)

For strategic planning purposes, running the model annually helps align long-term operational strategies with budgetary and resource availability forecasts.

## Conclusion

From a technical standpoint, the use of diverse analytical techniques was instrumental in addressing the complexities of airport operations.

Each of these models was carefully chosen and designed to leverage the data environment of the airport, combining historical data analysis with real-time operational data to produce actionable insights. The flexibility and scalability of these models mean they can be adapted and expanded as new data becomes available or as operational requirements evolve.

The analytical frameworks set up are not static; they are designed for continuous learning and improvement. By incorporating feedback mechanisms and regular updates, the models can evolve in response to changing conditions and new challenges. This adaptive approach ensures that Heathrow Airport remains at the forefront of operational efficiency and passenger service excellence.