HSLU Hochschule Luzern

The Impact of Adverse Weather Conditions on Flight Delays and Cancellations

Course and Semester: W.MSCIDS_DWL03.H24

Group Name: SOA

Students: Stefan Durrer, 16-539-231, stefan.durrer@stud.hslu.ch
Oliver Heisel, 22-888-135, oliver.heisel@stud.hslu.ch
André Kuhn, 11-725-256, andre.kuhn@stud.hslu.ch

Examiners: Dr. Luis Terán
José Mancera
Jhonny Pincay

GitHub: https://github.com/WLdurr/Data-Warehouse-and-Data-Lake-Systems

Lucerne, September 28, 2024

Contents

1 Background and Motivation

The aviation industry is highly sensitive to weather conditions, which frequently lead to flight delays and cancellations, as well as heightened safety concerns. As global air traffic continues to rise, the challenge for airlines and air traffic controllers to maintain operational efficiency intensifies (Burbidge, 2016). Factors such as storms, fog, and extreme temperatures can disrupt flight schedules and affect overall performance, creating a pressing need for accurate, real-time data analysis to mitigate these effects (Borsky & Unterberger, 2019; Coffel *et al.*, 2017; Koetse & Rietveld, 2009; Schultz *et al.*, 2021). While climate change has a long term effect on the aviation industry, extreme weather events already occur occasionally. A predictive, data-driven approach could help to mitigate the adverse effects such disruptions have on transportation networks (Stamos *et al.*, 2015)

Our project aims to address this issue by leveraging the power of Amazon Web Services (AWS) to analyze vast amounts of aviation and weather data. AWS provides a robust, scalable infrastructure that allows for the processing of real-time data streams in a cost-effective and reliable manner. Using services such as Amazon S3 for data storage and AWS Lambda for serverless computing, we can effectively integrate various data sources to predict flight delays and cancellations based on weather conditions. By employing Amazon's cloud solutions, we ensure that our system is not only fast and reliable but also accessible to stakeholders from any given location.

The motivation for this project lies in the potential to improve decision-making processes within the aviation industry, helping airlines, air traffic controllers, and passengers alike. By providing more accurate predictions, airlines can better manage their resources, reduce operational costs, and enhance passenger satisfaction. For travelers, this could translate to more timely and reliable updates on flight statuses, potentially minimizing the frustration and inconvenience often associated with delays and cancellations.

In essence, this project not only demonstrates the capabilities of AWS in handling complex, large-scale data analysis but also highlights the practical applications of data science in improving both safety and efficiency in the aviation sector. The use of cloud computing enables us to provide scalable, real-time solutions that can adapt to the needs of a growing industry, making this an impactful and timely endeavor.

2 Research Questions

2.1 Problem Statement

Flight delays and cancellations due to adverse weather conditions present significant economic and operational challenges for the aviation industry. Airlines, airports, and passengers are affected, leading to increased costs, logistical difficulties, and reduced customer satisfaction. This highlights the urgent need for a predictive model that integrates both weather and aviation data to forecast flight delays. Such a model could help mitigate risks, enhance operational efficiency, and improve the overall passenger experience. Our project aims to use AWS to develop this predictive system, leveraging real-time data from both aviation and weather sources. By analyzing large datasets, we seek to provide accurate, actionable insights that airlines and airports can use to anticipate delays and cancellations, thereby optimizing resource allocation and enhancing decision-making.

2.2 Research Questions and Objectives

1. **What are the effects of different weather conditions on the frequency and duration of flight delays in commercial aviation?**

   Objective: Classify weather conditions and quantify their impact on flight delays using a case study analysis.

   Key Results:

   (a) Identify and categorize specific weather conditions that contribute to flight delays.

   (b) Predict future flight delay patterns based on a 16-day weather forecast, considering conditions such as wind speed, pressure, and precipitation.

2. **How do specific weather conditions contribute to flight cancellations and delays, depending on the climate region?**

Objective: Assess how various weather phenomena influence flight cancellations and delays in different climate regions.

Key Results:

(a) Identify weather conditions responsible for flight disruptions across different climate zones.

(b) Predict how these conditions will impact cancellation and delay rates in the future.

3. **Which airports experience the highest frequency of weather-induced flight delays and cancellations, and what specific weather conditions cause these disruptions at each airport?**

   Objective: Determine which airports are most affected by weather-related disruptions and analyze the specific conditions responsible for these events.

   Key Results:

   (a) Rank the top 10 airports with the most weather-related delays and cancellations.

   (b) Predict future disruptions at these airports based on seasonal weather forecasts.

4. **How do weather-related delays and cancellations vary among airlines, and which airlines are most affected?**

   Objective: Analyze how weather-related delays and cancellations differ among major airlines.

   Key Results:

   (a) Compare weather-induced delays and cancellations across five major airlines.

   (b) Identify airlines that deviate significantly from average disruption rates.

3 Data Sources

Our project utilizes a combination of historical and real-time datasets to assess the impact of weather on flight delays and cancellations. These data sources are carefully chosen to ensure comprehensive coverage of both flight operations and meteorological conditions. The four primary data sources are as follows:

3.1 Flight Data

Flight operations data, including schedules, delays, and cancellations for both domestic and international flights, are provided by AirLabs and AeroDataBox APIs. These platforms offer detailed flight tracking information, such as departure and arrival times, flight statuses, and any deviations from scheduled operations. This granularity is essential for correlating flight delays and cancellations with specific weather conditions (AeroDataBox, n.d.; AirLabs, n.d.). AirLabs: https://airlabs.co AeroDataBox: https://aerodatabox.com

3.2 Weather Data

Weather data is obtained via the Open-Meteo API, which provides comprehensive meteorological records. This includes real-time and historical data on temperature, wind speed, precipitation, and severe weather events. Such data is critical for understanding the specific atmospheric conditions that contribute to flight disruptions. The API allows for querying based on location and timeframe, making it suitable for the temporal analysis required in this study (Open-Meteo, n.d.). Open-Meteo API: https://open-meteo.com/en/docs/historical-weather-api

3.3 Geolocation of Airports

Accurate geolocation data for each airport is necessary for making precise weather API calls. This information is sourced from a global airport database, which contains detailed geographic coordinates for over 4,200 airports worldwide. The dataset can be exported as a CSV file or

a SQLite database, facilitating easy integration with other data sources (Partow, n.d.). Global Airport Database: https://www.partow.net/miscellaneous/airportdatabase

4 Scope

The primary objective of this project is to develop a robust architecture for data wrangling and processing using AWS as well as training a predictive model to forecast flight delays and cancellations under specific weather conditions. The scope of this project covers several key areas:

- **Data Integration**: The project will involve the collection, preprocessing, and integration of large-scale datasets, including historical and real-time weather and flight data. Weather data will be sourced from platforms such as Open-Meteo, while flight operations data will come from APIs like AirLabs and AeroDataBox. The geolocation of airports will also be included to ensure accurate weather predictions for specific locations. AWS services such as Amazon S3 for data storage and AWS Lambda for serverless computing will be leveraged to process these datasets efficiently.

- **Predictive Modeling**: Using machine learning techniques, the project will develop models that analyze the relationship between various weather conditions, such as wind speed, precipitation, and temperature, and flight disruptions. These models will focus on forecasting both delays and cancellations, offering predictions based on short-term (e.g., 16-day forecasts) and seasonal weather data. This can be achieved through the use of time series analysis which has already been done in this field (Markovic *et al.*, 2008).

- **Regional Analysis**: The scope also includes the analysis of how weather-induced delays and cancellations vary depending on the climate region and airport location. By evaluating the impact of weather on different airports and airlines, the project aims to identify which locations and carriers are most affected by specific weather conditions.

- **Scalability and Cloud Computing**: The project will make extensive use of Amazon's scalable infrastructure to handle the real-time processing of data. The solution will be designed to accommodate large amounts of data and provide timely predictions that can be accessed by aviation stakeholders.

5 AUDIENCE

The solution developed in this project is intended to benefit a wide range of end users within the aviation industry. Key stakeholders include:

- **Airline Operators and Airport Management**: These professionals can leverage the predictive model to implement proactive strategies aimed at minimizing weather-related disruptions. By forecasting delays and cancellations, airlines can improve resource allocation, optimize scheduling, and enhance overall passenger satisfaction.
- **Travelers and Passengers**: With more accurate information regarding potential delays and cancellations, passengers can make informed decisions about their travel plans. This leads to reduced frustration, better planning, and the possibility of rescheduling or rerouting in advance, ultimately improving the customer experience.
- **Aviation Authorities and Regulatory Bodies**: Organizations responsible for maintaining the safety and efficiency of air travel can use these predictions to enforce guidelines and allocate resources more effectively, ensuring smoother operations during adverse weather conditions.
- **Tourism and Business Industries**: The ability to predict flight delays and cancellations will also benefit industries that rely heavily on air travel, such as tourism and corporate travel, by enabling better planning and reducing the economic impact of flight disruptions.

6 METHODS

The project employs a comprehensive set of methods to ensure the development of an accurate and scalable predictive model that integrates weather and flight data. The key methods used are as follows:

- **Data Collection**: Historical and real-time data will be gathered from APIs such as AirLabs, AeroDataBox, and Open-Meteo, which provide detailed flight and weather information. This includes data on flight schedules, delays, cancellations, and meteorological conditions such as wind speed, precipitation, and temperature.

- **Data Preprocessing**: After collecting the data, it will be cleaned and preprocessed to remove any inconsistencies, such as missing values or outliers. Datasets from different sources will be merged into a unified database, ensuring all necessary variables—flight details, weather conditions, and airport locations—are aligned for analysis.

- **Exploratory Data Analysis (EDA)**: Before building predictive models, EDA will be performed to identify key patterns and correlations between weather conditions and flight disruptions. Visualization techniques and statistical analyses will be used to explore relationships and trends that could affect flight delays and cancellations.

- **Prediction Models**: The focus will be on building models that not only predict delays and cancellations but also factor in different climate regions and seasonal variations. These models will be tested on a validation dataset to ensure they perform well in forecasting delays and cancellations under various weather scenarios. This can be achieved with time series analysis. In the literature, especially autoregressive (AR) models are widely used in this field (Markovic *et al.*, 2008).

- **AWS Infrastructure Implementation**: The project will utilize AWS to process, store, and deploy the data and models. Services like Amazon S3 will be used for scalable storage, Amazon EC2 for computational processing, and Amazon SageMaker for model training and deployment. AWS Lambda will enable serverless computing for handling data streams and API requests, ensuring real-time predictions are possible.

- **API Integration**: The real-time aspect of the project will be enabled through the integration of weather and flight APIs. These APIs will provide continuous updates on weather conditions and flight schedules, which will be used to feed the predictive models, ensuring up-to-date and reliable forecasts.

7 TIMELINE

Week 1: Introduction and project kickoff, defining objectives, familiarizing with AWS resources, and setting up the basic project environment.

Week 2: Initial data collection begins, gathering historical and real-time data from APIs (flight and weather). Introduction to AWS services and identifying the tools necessary for data storage and processing.

Week 3: Data preprocessing starts, including cleaning and merging datasets. Initial exploration of how AWS storage solutions (e.g., Amazon S3) can be utilized. Basic testing of data ingestion workflows.

Week 4: Exploratory Data Analysis (EDA) begins. This week will also focus on identifying the services needed for processing, such as AWS Lambda or EC2, and gaining familiarity with deploying small-scale data processes in the cloud.

Week 5: Model selection for predictive modeling starts. Research into AWS tools like SageMaker for machine learning model development, though initial testing will still be local as AWS skills develop.

Week 6: Refining machine learning models with local testing and optimization. Focus will also be on continuing to learn about deploying models on AWS. Initial exploration of cloud-based computing with AWS services.

Week 7: Midterm presentation to present project progress, initial data findings, and the planned approach for integrating AWS infrastructure.

Week 8: Begin serious work on integrating predictive models into AWS, exploring scalable deployment options. Implement data storage on Amazon S3 and basic data processing using EC2 or Lambda for small-scale workflows.

Week 9: Focus on learning more advanced AWS services like SageMaker for model training and deployment, while ensuring that the model can handle real-time data inputs via APIs.

Week 10: Model validation and performance testing using cloud infrastructure. Begin working on automation pipelines for collecting, processing, and feeding real-time weather and flight data into the model.

Week 11: Integration of real-time data APIs with the cloud infrastructure. Focus on optimizing workflows and understanding AWS services for data streaming and real-time processing.

Week 12: Refining predictive models in AWS, ensuring that model scalability and performance are optimized for large-scale data processing. Continued learning about AWS best practices for deployment.

Week 13: Final preparations, focusing on testing the entire system with real-time weather and flight data inputs. Ensure everything is set up in AWS for a seamless final presentation.

Week 14: Final presentation, showcasing the predictive model, the AWS infrastructure setup, and how the system handles real-time data. Documentation and submission of the final project report, including a detailed description of the AWS infrastructure and lessons learned during implementation.

8 Table of Contents of the Final Report

1. **Introduction**
    - 1.1 Project Idea
    - 1.2 Use Case
    - 1.3 Motivation
    - 1.4 Problem Definition and Goals
    - 1.5 Business and Research Questions
2. **System Architecture and Infrastructure**
    - 2.1 Data Lake Architecture
    - 2.2 Data Sources and Justification
    - 2.3 Data Source Limitations
3. **Data Ingestion**
    - 3.1 Data Source Description
    - 3.2 ETL (Extract, Transform, Load) Process Overview
    - 3.3 API Rate Limits and Constraints
4. **Data Storage**
    - 4.1 Data Storage Formats (Parquet, CSV, etc.)
    - 4.2 Relational Database Setup
    - 4.3 NoSQL/Vector Database Setup
    - 4.4 Indexing and Data Retrieval
5. **Data Transformation**
    - 5.1 Data Transformation Steps

References

AeroDataBox. (n.d.). Aerodatabox api [Accessed: 2024-09-26]. https://aerodatabox.com

AirLabs. (n.d.). Airlabs api [Accessed: 2024-09-26]. https://airlabs.co

Borsky, S., & Unterberger, C. (2019). Bad weather and flight delays: The impact of sudden and slow onset weather events. *Economics of Transportation*, *18*, 10–26. https://doi.org/10.1016/J.ECOTRA.2019.02.002

Burbidge, R. (2016). Adapting european airports to a changing climate. *Transportation Research Procedia*, *14*, 14–23. https://doi.org/10.1016/j.trpro.2016.05.036

Coffel, E. D., Thompson, T. R., & Horton, R. M. (2017). The impacts of rising temperatures on aircraft takeoff performance. *Climatic Change*, *144*, 381–388. https://doi.org/10.1007/s10584-017-2018-9

Koetse, M. J., & Rietveld, P. (2009). The impact of climate change and weather on transport: An overview of empirical findings. *Transportation Research Part D: Transport and Environment*, *14*, 205–221. https://doi.org/10.1016/j.trd.2008.12.004

Markovic, D., Hauf, T., Röhner, P., & Spehr, U. (2008). A statistical study of the weather impact on punctuality at frankfurt airport. *Meteorological Applications*, *15*, 293–303. https://doi.org/10.1002/MET.74

Open-Meteo. (n.d.). Open-meteo historical weather api [Accessed: 2024-09-26]. https://open-meteo.com/en/docs/historical-weather-api

Partow, A. (n.d.). Global airport database [Accessed: 2024-09-26]. https://www.partow.net/miscellaneous/airportdatabase

Schultz, M., Reitmann, S., & Alam, S. (2021). Predictive classification and understanding of weather impact on airport performance through machine learning. *Transportation Research Part C: Emerging Technologies*, *131*. https://doi.org/10.1016/j.trc.2021.103119

Stamos, I., Mitsakis, E., Salanova, J. M., & Aifadopoulou, G. (2015). Impact assessment of extreme weather events on transport networks: A data-driven approach. *Transportation Research Part D: Transport and Environment*, *34*, 168–178. https://doi.org/10.1016/j.trd.2014.11.002