AIRLINES DATA ANALYTICS IN AVIATION INDUSTRY

Team ID: PNT2022TMID15316

NALAIYA THIRAN PROJECT BASED LEARNING

On

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

A PROJECT REPORT

TABLE OF CONTENTS

ABSTRACT

In the contemporary world, Data analysis is a challenge in the era of varied intersdisciplines though there is a specialization in the respective disciplines. In other words, effective data analytics helps in analyzing the data of any business system. But it is the big data which helps and axial rates the process of analysis of data paving way for a success of any business intelligence system. With the expansion of the industry, the data of the industry also expands. Then, it is increasingly difficult to handle huge amount of data that gets generated no matter what‘s the business is like, range of fields from social media to finance, flight data, environment and health.

An Airport has huge amount of data related to number of flights, data and time of arrival and dispatch, flight routes, No. of airports operating in each country, list of active airlines in each country. The problem they faced till now it‘s, they have ability to analyze limited data from databases.

How can it be gathered, stored, processed and analyzed it to turn the raw data information to support decision making. In this paper Big Data is depicted in a form of case study for Airline data.

1. INTRODUCTION

1.1Overall description

Researchers working in the structured data face many challenges in analyzing the data. For in_stance the data created through social media, in blogs, in Facebook posts or Snap chat. These types of data have different structures and formats and are more difficult to store in a traditional business data base. The data in big data comes in all shapes and formats including structured. Working with big data means handling a variety of data formats and structures. Big data can be a data created from sensors which track the movement of objects or changes in the environment such as temperature fluctuations or astronomy data. In the world of the internet of things, where devices are connected and these wearable create huge volume of data. Thus big data approaches are used to manage and analyze this kind of data. Big Data include data from a whole range of fields such as flight data, population data, financial and health data such data brings as to another V, value which has been proposed by a number of researcher i.e., Veracity.

Most of the time social media is analyzed by advertisers and used to promote produces and events but big data has many other uses. It can also been used to assess risk in the insurance industry and to track reaction to products in real time. Big Data is also used to monitor things as diverse as wave movements, flight data, traffic data, financial transactions, health and crime. The challenge of Big Data is how to use it to create something that is value to the user. How to gather it, store it, process it and analyze it to turn the raw data information to support decision making.

An Airport has huge amount of data related to number of flights, data and time of arrival and dispatch, flight routes, No. of airports operating in each country, list of active airlines in each country. The problem they faced till now it‘s, they have ability to analyze limited data from databases. The Proposed model intention is to develop a model for the airline data to provide platform for new analytics based on the following queries.

1.1 Problem Statement

- Big amount of data generated on hourly basis.
- A single twin engine aircraft with an average 12 hour flight time can produce up to 844 TB of data
- There are many active users of flights
- Many flights are scheduled everyday
- User varies from common man to celebrities

1.2 Purpose

The main purpose of the project to explore detailed analysis on airline data sets such as listing airports operating in the India, list of airlines having zero stops, list of airlines operating with code share which country has highest airports and list of active airlines in united states. The main objective of project is the processing the big data sets using map reduce component of Hadoop ecosystem in distributed environment.

1.3 Motivation and scope

Product Perspective

The main purpose of the project to explore detailed analysis on airline data sets such as listing airports operating in the India, list of airlines having zero stops, list of airlines operating with code share which country has highest airports and list of active airlines in united states. The main objective of project is the processing the big data sets using map reduce component of Hadoop ecosystem in distributed environment.

Product Features

Airline data analysis can provide a solution for businesses to collect and optimize large datasets, improve performance, improve their competitive advantage, and make faster and better decisions.

- By using airline data analysis, we can save time of users.
- The data could even be structured, semi-structured or unstructured.
- Cost savings
- Implementing new strategies
- Fraud can be detected the moment it happens

1.4 Assumptions and Dependencies

- Constraints are limitations which are outside the control of the project. The Project must be managed within these constraints.
- Assumptions are made about events, or facts outside the control of project. External dependencies are activities which need to be completed before an internal activity can proceed.
- Constraints, assumptions and dependencies can create risks that the project may be delayed because access is not provided to the site (assumption).
- Assumption will be that the complexity may arise due to large unstructured data set.

1.5 Constraints

- Hardware limitation and timing constraints.
- High feature may not correspond to semantic similarity.
- System Environment Windows subsystem for Linux with Ubuntu operating system will be required to run the application

2. OBJECTIVE

Safety
Air navigation Capacity
Efficiency
Security
Facilitation

The security seals **ensure the traceability of the food at all times duringtransport and avoid unauthorized handling**.

3. IDEATION PHASE

3.1 Literature Survey

Authors :Sai-Ho-Chung,Hoi-Lam-ma

The researcher in this article cited that,Due to the rapid development of advanced technologies nowadays, a massive amount of real time data regarding flight information, flight performance, airport conditions, air traffic conditions, weather, ticket prices, passengers comments, crew comments, etc., are all available from a diverse set of sources, including flight performance monitoring systems, operational systems of airlines and airports, and social media platforms. Development of data analytics in aviation and related applications is also growing rapidly. This paper concisely examines data science and analytics in aviation studies in several critical areas, namely big data analysis, air transport network management, forecasting, and machine learning. The papers featured in this special issue are also introduced and reviewed, and future directions for data science and analytics in aviation are discussed.

Data Analytics for Air Travel Data:
Authors:Haiman Tian,Yudong Tao

The researcher in this article cited that,From the start, the airline industry has remarkably connected countries all over the world through rapid long-distance transportation, helping people overcome geographic barriers. Consequently, this has ushered in substantial economic growth, both nationally and internationally. The airline industry produces vast amounts of data, capturing a diverse set of information about their operations, including data related to passengers, freight, flights, and much more. Analyzing air travel data can advance the understanding of airline market dynamics, allowing companies to provide customized, efficient, and safe transportation services. Due to big data challenges in such a complex environment, the benefits of drawing insights from the air travel data in the airline industry have not yet been fully explored. They introduce existing data sources commonly used in the papers surveyed and summarize their availability. Finally, we discuss several potential research directions to better harness airline data in the future. They anticipate this study to be used as a comprehensive reference for both members of the airline industry and academic scholars with an interest in airline research.

Topological Data Analysis For Aviation Applications:

Authors: Max Z. Li,Megan S. Ryerson and Hamsa Balakrishnan

Aviation data sets are increasingly high-dimensional and sparse. Consequently, the underlying features and interactions are not easily uncovered by traditional data analysis methods. Recent advancements in applied mathematics introduce topological methods, offering a new approach to obtain these features. This paper applies the fundamental notions underlying topological data analysis and persistent homology (TDA/PH) to aviation data analytics. We review past aviation research that leverage topological methods, and present a new computational case study exploring the topology of airport surface connectivity. In each case, we connect abstract topological features with real-world processes in aviation, and highlight potential operational and managerial insights.

The Global Airline Industry:

Author:Hoi-Lam-ma

The researcher in this article cited that,The events of September 11th, 2001 precipitated an almost unprecedented financial crisis for the world airline industry. However, it is not clear that these events represent a discrete, industry disruption or whether, in fact, airlines were already entering a period of economic challenges that would demand new strategic orientations on their part. This study investigates the structural drivers of operational efficiency as well as the financial posture of airlines on the eve of September 11th. A sample of 38 airlines from North America, Europe, Asia and the Middle East was utilized to investigate whether relative operational efficiency implied superior financial mobility (as defined by Donaldson). Data envelopment analysis was utilized to derive efficiency scores for individual airlines. The underlying structural drivers of efficiency were then investigated. It was found that the traditional framework developed in the literature still provided reasonable explanatory power for realized relative operational efficiency. However, the second stage of the analysis found that relative operational efficiency did not inherently imply superior financial mobility. As such, airlines that had chosen relatively efficient operational strategies found themselves in positions of vulnerability with regard to financial mobility and thus suffered the consequences in the post-September 11th environment.

An Evaluation Of The Operational Performance And Profitability Of The U.S. Airlines:

Author:Emilio Collar

The researcher in this article cited that,Since 2008, a series of megamergers has dramatically changed the U.S. airline industry. Despite the presence of fewer airlines in the market, the competition remains intense, which forces airlines to continually search for ways to increase their efficiency to maintain survival and financial sustainability. To evaluate airline performance and disentangle the causes of inefficiency, this paper applied a two-stage network data envelopment analysis approach and a truncated regression to investigate the performance of nine U.S.-based airlines from 2015 to 2019. Our empirical results reveal that during the sample period, airlines 'operating efficiency steadily improved, but the efficiency in the profitability stage stagnated. Therefore, strategic resource allocations are needed for airlines to see further advances in their overall efficiency. On average, airlines operating in the low-cost business model yielded higher efficiency scores than their peers operating in the full-service framework. While an airline's size, measured in terms of total assets, has a positive influence on operating efficiency, a larger number of full-time employee equivalents hinders efficiency outcomes, which indicates the importance of enhancing labor efficiency among carriers.

The Relationship Between On-Time Performance And Airline Market Share:
Author:Yoshinori Suzuki

The researcher in this article cited that,We propose a new method of modeling the relationship between on-time performance and market share in the airline industry. The idea behind the method is that the passengers 'decision to remain (use same airline) or switch (use other airlines) at time t depends on whether they have experienced flight delays at time t−1 or not. More specifically, we posit that the passengers who experienced flight delays are more likely to switch airlines for the subsequent flight than those passengers who did not experience delays. To capture such effect, we develop an aggregate level Markovian type model that estimates the transition probability matrices separately for the passengers who experienced flight delays at time t−1 and for those who did not experience delays. The model was calibrated with the US DOT data. The study results imply that, once experiencing flight delays, passengers are more likely to switch airlines. The results also imply that on-time performance affects a carrier's market share primarily through the passengers 'experience, and not though the "advertisement" of performance.

Airline Finance:
Author:Peter.S.Morel

The researcher in this article cited that, It is supported at each stage by practical airline examples and recent data, Airline Finance examines the financial trends and longer term prospects for the airline industry as a whole, contrasting the developments for the major regions and airlines together with critical discussion of key issues that affect the industry as a whole. Important techniques in financial analysis are applied to the airlines as well as their investors such as banks and other financial institutions.This book is written for employees of airlines, airports and their suppliers, and investment bank and other analysts. It is also popular for use by universities and in-house courses on air transport management, within both academia and industry.

Airline Route Profitability Analysis And Optimization Using Big Data Analytics On Aviation Data Sets Under Heuristic Techniques:
Authors: Kasturi E, Prasanna Devi Sb, Vinu Kiran Sb, Manivannan Sc

Researchers in this article cited that ,applying vital decisions for new airline routes and aircraft utilization are important factors for airline decision making.For data driven analysis key points such as airliners route distance, availability on seats/freight/mails and fuel are considered. The airline route profitability optimization model is proposed based on performing Big data analytics over large scale aviation data under multiple heuristic methods, based on which practical problems are analyzed. Analysis should be done based on key criteria, identified by operational needs and load revenues from operational systems e.g. passenger, cargo, freights, airport, country, aircraft, seat class etc. The result shows that the analysis is simple and convenient with concrete decision. Analysis Of Flight Data Using Clustering

Techniques For Detecting Abnormal Operations:
Author:Lishaui Li,Santanu Das

The researcher in this article cited that,the airline industry is moving toward proactive risk arrangement, which aims to identify and mitigate risks before accidents occur. However, existing methods for such efforts are limited. They rely on predefined criteria to identify risks, leaving emergent issues undetected. This paper presents a new method, cluster-based anomaly detection to detect abnormal flights, which can support domain experts in detecting anomalies and associated risks from routine airline operations. The new method, enabled by data from the flight data recorder, applies clustering techniques to detect abnormal flights of unique data patterns. Compared with existing methods, the new method no longer requires predefined criteria or domain knowledge. Tests were conducted using two sets of operational data consisting of 365 B777 flights and 25,519 A320 flights. The performance of clusterbased anomaly detection to detect abnormal flights was compared with those of multiple kernel anomaly detection, which is another data-driven anomaly detection algorithm in recent years, as well as with exceedance detection, which is the current method employed by the airline industry. Results showed that both cluster-based anomaly detection to detect abnormal flights and multiple kernel anomaly detection were able to identify operationally significant anomalies surpassing the capability of exceedance detection. Cluster-based anomaly detection to detect abnormal flights performed better with continuous parameters, whereas multiple kernel anomaly detection was more sensitive toward discrete parameters.

Data Analytics Of Skytrax's Airport Review And Ratings:
Author:KrityaBunchongchit

The researcher in this article cited that,This study investigates the perception of passengers of airport service attributes, using data from the Skytrax Airport Review websites. Overall, a total of 7358 reviews were collected from the website, together with other related passenger data, namely review headers, passenger types, rating scores of airport attributes and the overall rating. This study focused on investigating each group of passenger types to identify underlying differences amongst airport's passenger segmentation, particularly on the leisure travellers. The study performed different techniques of data analysis including sentiment analysis, lemmatization and partial least square – structural equation modelling (PLS-SEM) to reveal key patterns derived from the available data, which the normal survey data or the interview data may not have revealed. The research contributes to airport passenger segmentation by highlighting the differences found in the travellers segmented by Skytrax. The study also provides practical implications to airport managers.

Post Pandemic Aviation Market Recovery:
Experience and lessons from China
Author:Achim.I.Czemy

The researcher in this article cited that,China was the first aviation market in the world hit hard by COVID-19 and has been recovering gradually as the pandemic became largely under control within mainland China. This study reviews the recovery pattern influenced by the Chinese government's aviation policy choices, in the hope that our discussions and findings will help improve aviation policy responses elsewhere. While the domestic market in mainland China has enjoyed a quick recovery to about 80% of the pre-crisis level by July 2020, the recovery of international services has been much slower, due to the bilateral route and flight frequency/capacity control and strict requirements for health check and quarantine. China's domestic aviation market was recovered by about 80% in two months after the pandemic became under good control. Most other countries with a “curve flattening” strategy, instead of full pandemic control, may not expect the fast recovery path China has achieved. A British “travel corridor” approach may be more practical for Western countries to follow, albeit more likely to be subject to serious setbacks and disruptions. The aviation fee reductions and cost support China and many other countries have been using are helpful by reducing airlines' marginal costs, but not sufficient for carriers to return to profitability or sustainable operations. Capital injection and/or credit guarantee may be needed for many airlines to survive. With various, often uncoordinated, regulations imposed in international markets, airlines based in open economies that have small domestic markets will face particularly serious challenges during the recovery process.

Sustainability Reporting In The Airline Industry:
Current Literature And Future Research Avenues:
Authors: Malgorzata Zieba and Eljas Johansson

Researchers in this article cited that,sustainability reporting (SR) allows organizations to communicate their non-financial impacts to stakeholders. It has also become a widespread business practice in aviation, a transport sector that contributes significantly to global warming. Academia has begun to examine SR in the context of airlines surprisingly late, and no comprehensive reviews of its respective developments have been made so far. Consequently, a systematic literature review was performed with an exclusive focus on airline SR to synthesis its associated scholarly research and distinguish the common concerns and gaps that have emerged from it. The analyzed publications indicate that the industry has lacked a unified policy and common understanding of how to define and measure sustainability, which has led to inconsistent SR practices. This causes ambiguity between the real actions and promotional communication through which airlines may legitimize their operations. Academia and various airline stakeholders would benefit from more in-depth studies examining the stakeholder views and quality of disclosures, helping the industry improve its SR.

Changes In Air Passenger Demand As A Result Of
The COVID-19 Crisis:Using Big Data To Inform
Tourism Policy:
Author:Galle Immaculadago

The researcher in this article cited that,this paper develops a methodology for the early detection of reactivation of tourist markets to help mitigate the effects of the COVID-19 crisis, using Skyscanner data on air passenger searches (>5,000 million) and picks (>600 million), for flights between November 2018 and December 2020, through Forward Keys. For future travel during the May to September 2020 period, the desire to travel (based on the number of flight searches) has dropped by about 30% in Europe and the Americas, and by about 50% in Asia, while intention to travel (the number of flight picks, the final selections amongst flight searches) has dropped a further 10–20%. Most source markets remain optimistic about air travel during the last quarter of 2020, suggesting a U shape recovery. However, optimism has dwindled as time passes, suggesting a flatline L shape. A traffic light dashboard for domestic and inbound air travel demand to Spain shows how destination managers might use Big Data relating to the early recovery of key source markets to develop targeted marketing strategies. We show how Big Data provides timely granular data essential in highly volatile situations, and we argue that destination management organizations must improve their Big Data analytical and evidence-based, decision-making skills.

The Impact Of Social Media And Offline

Influences On Consumer Behaviour: An Analysis Of The Low-Cost Airline Industry:
Author:Carla Ruiz

The researcher in this article cited that,this study analyses the impact of social media as well as offline environments upon tourist online purchase and recommendation behaviour of low cost airline services. Drawing on the Theory of Reasoned Action (TRA), this research considers the effect of offline social influences (interpersonal and external influences) and analyses online Consumer-to-Consumer (C2C) information exchanges as a driver of customer attitude towards online purchases. We propose that these factors improve online repurchase intentions and positive word-of-mouth communication (WOM and e-WOM) in low-cost settings. Using structural equation modelling, the conceptual model is tested with a sample of 441 Spanish Internet buyers of low-cost airline services. Interpersonal offline influences (e.g. friends, relatives, and family) have a significant effect on online repurchase intentions and WOM but do not affect e-WOM. External offline influences (e.g. media and experts), however, only affect consumer intentions to recommend future purchases of lowcost airline services on social networking travel sites and have no effect on online repurchase intentions or WOM.

Assessing Quality Of Air Transport Service: A Comparative Analysis Of Two Evaluation Models:
Author:Denise Dumiko De Medeiros

The researcher in this article cited that,this paper aims to analyze the opinion of tourists about airlines ’service in a developing country. For this, the study proposes to make a comparative analysis of two evaluation models (SERVQUAL and SERVPERF) to investigate the factors that influence the formation of perceived quality in airline services, using statistical techniques such as Cluster Analysis and Structural Equation Modeling. Although the results were not the same, the result of both analyzes indicated two common dimensions (tangibles and empathy) that influence the customer’s perception of the airline service quality. The main conclusion of this study is that the two analyzes are convergent for the study sample. The SERVQUAL and cluster analysis allow airline managers to identify and prioritize gaps in service delivery according to criticality, aiming at the allocation of efficient resources by the airline. The SERVPERF and SEM provide statistical evidence of the impact of different dimensions of service quality on customer satisfaction, highlighting the direct relationship between satisfaction and dimensions. Considering how customers evaluate the service provided by airlines, particularly regarding the service they receive from airport employees, this study has relevance for decisions taken by airline managers to develop quality services, and provide guidelines for improvements in airline services.

Domestic Code Sharing, Alliances, And Airfares In The U.S. Airline Industry:
Author:Harumi Ito,Darmin Lee

The researcher in this article cited that,this paper examines the impact of domestic codesharing alliances on airfares. Our analysis yields two novel and somewhat surprising findings that have yet to be documented in the literature. First, unlike with international code sharing, we find that the overwhelming majority of domestic codeshare itineraries involve a single operating carrier, a phenomenon that we refer to as virtual code sharing. Second, we find that these virtual code-sharing itineraries are priced lower than itineraries operated and marketed by a single carrier in the same market. We suggest that carriers may be using virtual code sharing—in large part—as a generic product to compete for the most price-sensitive passengers.

Customer segmentation revisited: The case of the airline industry:
Author:Iwanvon Wartburg

The researcher in this article cited that,although the application of segmentation is a topic of central importance in marketing literature and practice, managers tend to rely on intuition and on traditional segmentation techniques based on socio-demographic variables. In the airline industry, it is regarded as common sense to separate between business and economy passengers. However, the simplicity of this segmentation logic no longer matches the ever more complex and heterogeneous choices made by customers. Airline companies relying solely on flight class as the segmentation criterion may not be able to customize their product offerings and marketing policies to an appropriate degree in order to respond to the shifting importance and growing complexity of customer choice drivers, e.g. flexibility and price as a result of liberalization in the airline industry. Thus, there is a need to re-evaluate the traditional market segmentation criterion.
By analyzing the stated preference data of more than 5800 airline passengers, we show that segmenting into business and leisure (a) does not sufficiently capture the preference heterogeneity among customers and (b) leads to a misunderstanding of consumer preferences. We apply latent class modeling to our data and propose an alternative segmentation approach: we profile the identified segments along behavioral and socio-demographic variables. We combine our findings with observable consumer characteristics to derive pronounced fencing mechanisms for isolating and addressing customer segments receptive for tailored product packages.

A Hierarchical Model Of Service Quality In The Airline Industry:

Author:Ching Chang-Cheng

The researcher in this article cited that,the purpose of this study is to enhance understanding of service quality in the airline industry by developing a conceptual framework and measurement scale. Based on an extensive literature review, qualitative and empirical research, a hierarchical model of service quality for the airline industry is proposed. Analysis of data from 544 passengers indicates that the proposed model fits the data well. Reliability and validity of the measurement scale are established using a pilot test and the substantive survey. This study extends the literature on service quality in the fields of transportation management by providing a comprehensive framework and measurement scale. Theoretical and managerial implications are discussed.

Applied Cognitive Task Analysis in Aviation:

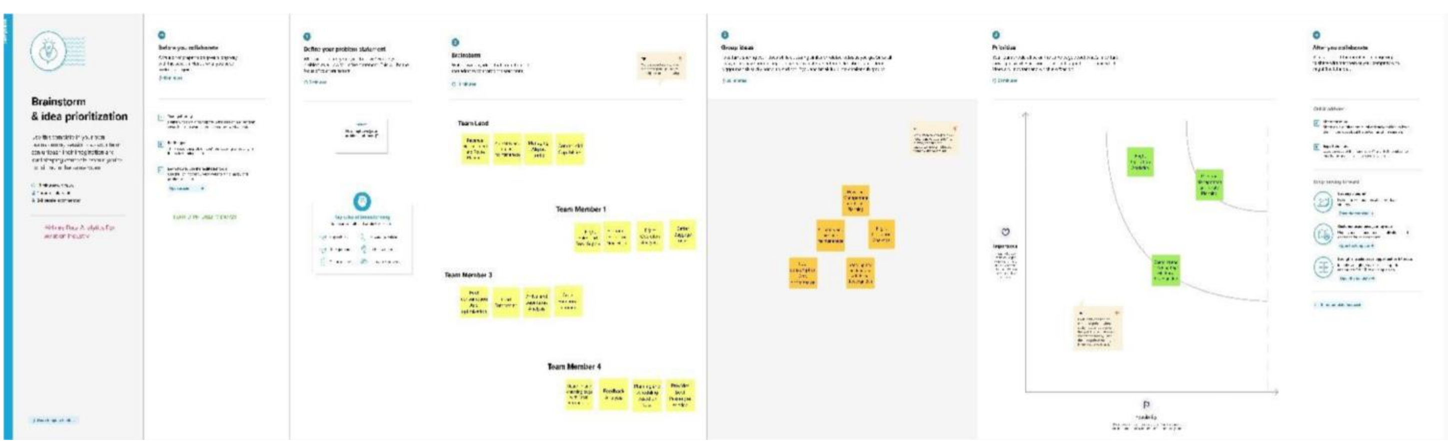
Authors: Thomas L. Seamster and Richard E. Redding

Researchers in this article cited that, due to the requirements of automatic system design, and new needs for the training of complex tasks, Cognitive Task Analysis (CTA) has been used with increasing frequency in recent years by the airline industry and air traffic control community. Its power is reflected in the literature on professional training and systems design, where CTA is often cited as one of the most promising new technologies, especially for the complex cognitive tasks now confronting those working in aviation. The objective of this book is to bridge the gap between research and practice, to make what we know about CTA available to practitioners in the field. The book focuses on cognitive psychology and artificial intelligence analyses of aviation tasks. It is designed to help readers identify and solve specific design and training problems, in the flight deck, air traffic control and operations contexts. Distilling experience and guidelines from the best aviation cognitive analyses in accessible form, it is the first comprehensive volume on CTA, and is written for practitioners of cognitive analysis in aviation. It provides an overview of analyses to date; methods of data collection; and recommendations for designing and conducting CTA for use in instructional design, systems development, and evaluation. The first part of the book provides the principles and foundations of CTA, describing traditional approaches to task analysis and ways that cognitive analyses can be integrated with the analysis and development processes. The next part details how to: select the appropriate method or methods; determine job tasks that can be trained for automatic performance; extract knowledge structures; analyse mental models; and identify the decision-making and problemsolving strategies associated with experienced job .performance. The authors also describe when to use and how to design and conduct a cognitive task analysis; how to use CTA along with traditional task analysis and ISD; and how to use CTA in training program development and systems design, as well as in personnel selection and evaluation. The current demand for cognitive analyses makes this a timely volume for those in aviation and, more generally, the industrial development and training communities. Readers will find this a thorough presentation of cognitive analyses in aviation and a highly usable guide in the design, implementation and interpretation of

CTA. The book will be useful to instructional developers, aviation equipment and systems designers, researchers, government regulatory personnel, human resource managers, instructors, pilots, air traffic controllers, and operations staff.

3.2 Empathy Map

3.3 Brainstroming



3.4 Problem Statement

 Analyze passsenger traffic and analyze their travelling

 Analyze and help in maintaining the services of the aeroplane

 Provides broad opportunities for airspace management, enhancing flexibility in dealing with each passenger, boostingproblem solving, supporting decision, providing safe flights.

 Flight delay for a specific period of time caused due to climate, security, carrier, NAS, Arrival and Departure can beovercomed

1. PROJECT DESIGN PHASE 1

4.1 Proposed Solution

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	**Parameter**	**Description**
1.	Problem Statement (Problem to besolved)	With the growing demand forair transportation and the limited ability toincrease capacity at some key points inthe air transportation system, there are concerns that in the future the system will not scale to meet demand. This situation will result in the generation and the propagation of delays throughout the system, impacting passengers' quality of travel and more broadly the economy.

2.	Idea / Solution description	▯ Understanding traveler demand for specific city pairs and pricing flights can be done using data analytics project. ▯ Airlines use this biometric technology as a boarding option. The equipmentscans travelers' faces and matches with photos stored in border control agency databases. These can behandled with the aforementioned project.
3.	Novelty / Uniqueness	▯ The ultimate benefits of big data analytics include timely responses to current and future market demands, improved planning and strategically aligned decision making, as well as crystal clear comprehension and monitoring of all main performance drivers relevant to the airline industry. ▯ Due to the use of smartdata analytics, passengers will avoid many issues with baggage tracking. While radio- frequency identification prevents mishandling the baggage, predictive analysis assists in improving the predictability of fleet reliability

5.	Business Model (Revenue Model)	Business models innovation in airlines can contribute to the creation of value, competitive advantage and profitabilitywith new possibilities of action. A revenue model is a blueprint that shows how a startup business will earn revenue or gross income from its standard business operations, and how it will pay for operating costs and expenses.
6.	Scalability of the Solution	The Cloud Cognos Analytics is not only for particular organization/governments. Aviation industryacting under international, domestic or private are also getting satisfied with the aviation

4.2 Problem Solution Fit

Problem-Solution fit canvas 2.0 Airline Data Analytics For Aviation Industry AMALTAMA

Define CS, fit into

1. CUSTOMER SEGMENT(S) CS

- Customers are airline and airport services who are struggling to keeptrack of their forecasting data and planes arrival and departure.
- Airlines literally bear high costs due to delays and cancellations that includes expenses on maintenance and compensations to travelers stuck in airports. With nearly 30 percent of the total delay time caused by unplanned maintenance, predictive analytics applied to fleet technical support is a reasonable solution.

6. CUSTOMER CONSTRAINTS C

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

- Customer experience in the airline industry is often defined as what the customer perceives and experiences while traveling through the different departure stages and arrival in an airport.
- Mid-air: It is the best time to engage with passengers and understand their in-flight expectations. Start with the basics like seating comfort and crew etiquette.
- Post landing: Inspect through passengers' eyes and listen to their opinion. That's a great way to enhance your online reputation, post-flight.

5. AVAILABLE A

Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking

Flight Turnaround Analytics:

- Provides insights on process inefficiencies in a flight turnover. The video annotation service helps to capture the time taken by each specific activity within flight turnover using video monitoring used for ground activities.

Planning and Schedule Analytics:

- Provides in-depth analysis of ticket sales, operational expense and profitability of airline routes. It helps in fleet rebalancing, fuel needs and crew planning for a flight.

Explore AS,

Focus on J&P, tap into BE, understand

2. JOBS-TO-BE-DONE / PROBLEMS J&P

Which jobs-to-be-done (or problems) do you address for your customers?

- Collecting data related to flight operations and inventory. You will use proprietary software like Airmax, for instance, or simple tools like Microsoft Excel to gather statistics related to important metrics called Key Performance Indicators (KPI).
- Optimising flight operations based on quantitative analysis. You will have to advise your management on trends and bottlenecks that you observe from data analysis so they can take the necessary action

9. PROBLEM ROOT CAUSE R

What is the real reason that this problem exists? What is the back story behind the need to do this job? i.e. customers have to do it because of the change in regulations.

- A root cause analysis is performed as a reaction to risk management processes as defined in your aviation SMS manual.
- The purpose of the analysis is to understand the causal factors that trigger substandard safety performance within a particular event, whether the event is an: accident, minor incident, or close call.

7. BEHAVIOUR B

What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)

- Using airport analytics, data analysts can collect information on people who pass through various checks, like their gender, arrival times, baggage-check in times and the type of flight they take to better understand passenger behaviour.
- A better understanding of how passengers operate can be used to improve services

Focus on J&P, tap into BE, understand

Define CS, fit into CL

3. TRIGGERS TR

What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news

- In Aviation Industry,due to incidents like flight delays passenger may face delays in departure and arrival of flight.
- It is very hard to maintain the overall data.But if they use Data Analytics Report,Performance and Quality are reliable and profitable.

4. EMOTIONS: BEFORE / AFTER EM

How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

- Before:They feel lost due to losses which occur due to improper management of Airline Analytics for Aviation Industry.
- After:They feel like success after making increased profits,reducing the mistakes that happen in manual process.

10. YOUR SOLUTION SL

What kind of solution suits Customer scenario the best? Adjust your solution to fit Customer behaviour, use Triggers, Channels & Emotions for marketing and communication

To design an Airline Data Analytics Report for Aviation Industry using Cognos Analytics.

Enable Email based alerts for arrival and departure of flight and it also sends messages related to the changes in configuration of flight path parameters.

Provide a option for graphical view of aviation industry.

8.1 ONLINE CHANNELS CH

What kind of actions do customers take online? Extract online channels from box #7 Behaviour

Online Airline Analytics for Aviation Industry which come for free may steal personal information of users and it may also contains a lot of ads.Security is not authenticated.

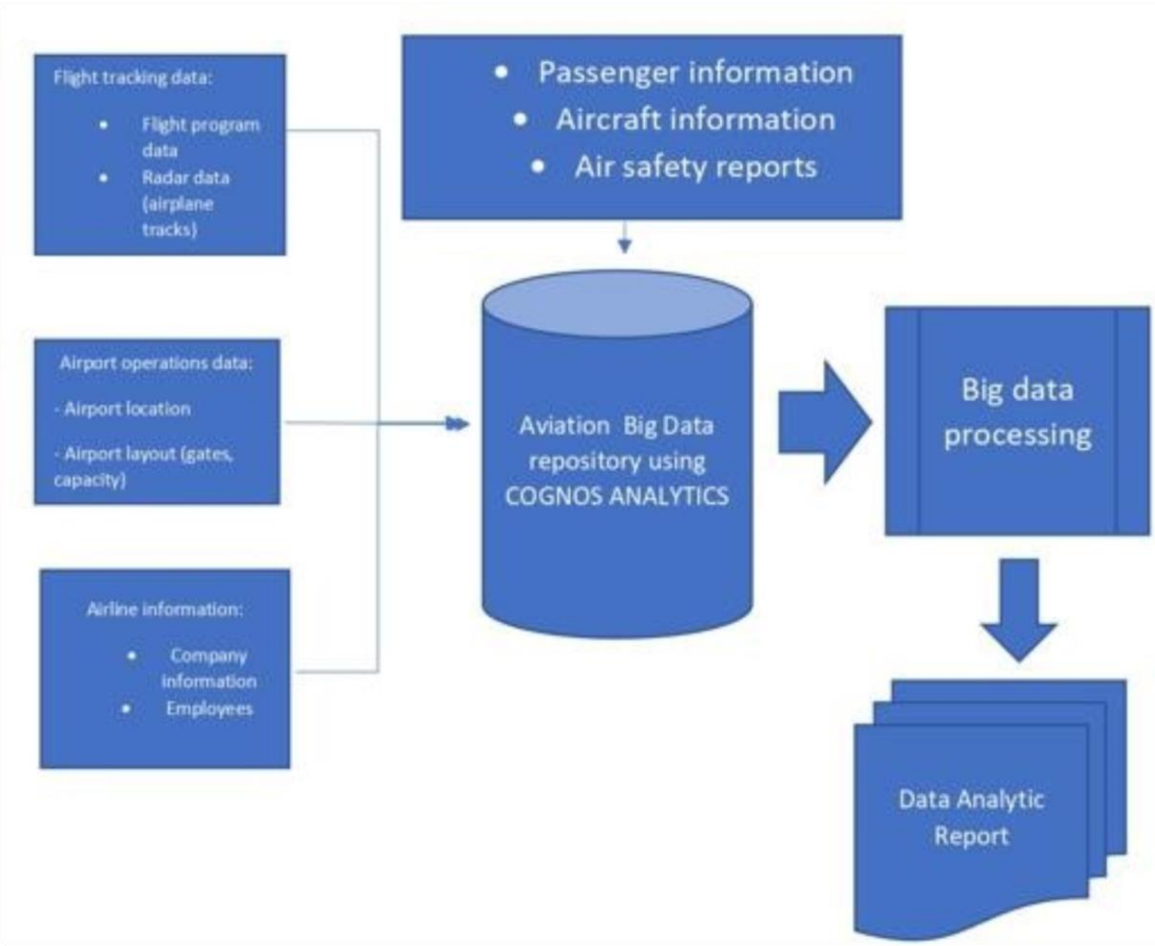
8.2 OFFLINE CHANNELS CH

What kind of actions do customers take offline? Extract offline channels from box #7 Behaviour and use them for customer development

- Manual logs can be maintained.Employees can be hired to maintain the airline analytics for aviation industry system logs when the business grows.

Explore AS, differentiate

4.3 Solution Architecture



2. PROJECT DESIGN PHASE 2

5.1 Customer Journey Map

Customer journey mapping is ‘’the process of tracking and describing all the experiences customers have as they encounter a service or set of services, taking into account not only what happens to them, but also their responses to these experiences’’ (Dent, 2015). As stated in Eva Manrique’s blog: (2016) ‘’when airlines adopt the customer journey mapping strategy, it helps them to clarify what each individual customer expects at each touchpoint and which fitting service or product the airline could provide in order to fulfil these expectations.’’

The customer journey map is divided into various phases as the customer has various different options in regards to approaching the airline, these phases include the three communication arena’s: physical, digital and social.

• Phase 1: attract: the customer recognizes the airline for (potentially) the first time, via toolssuch as social media but also through discount offers posted on the website or social media.

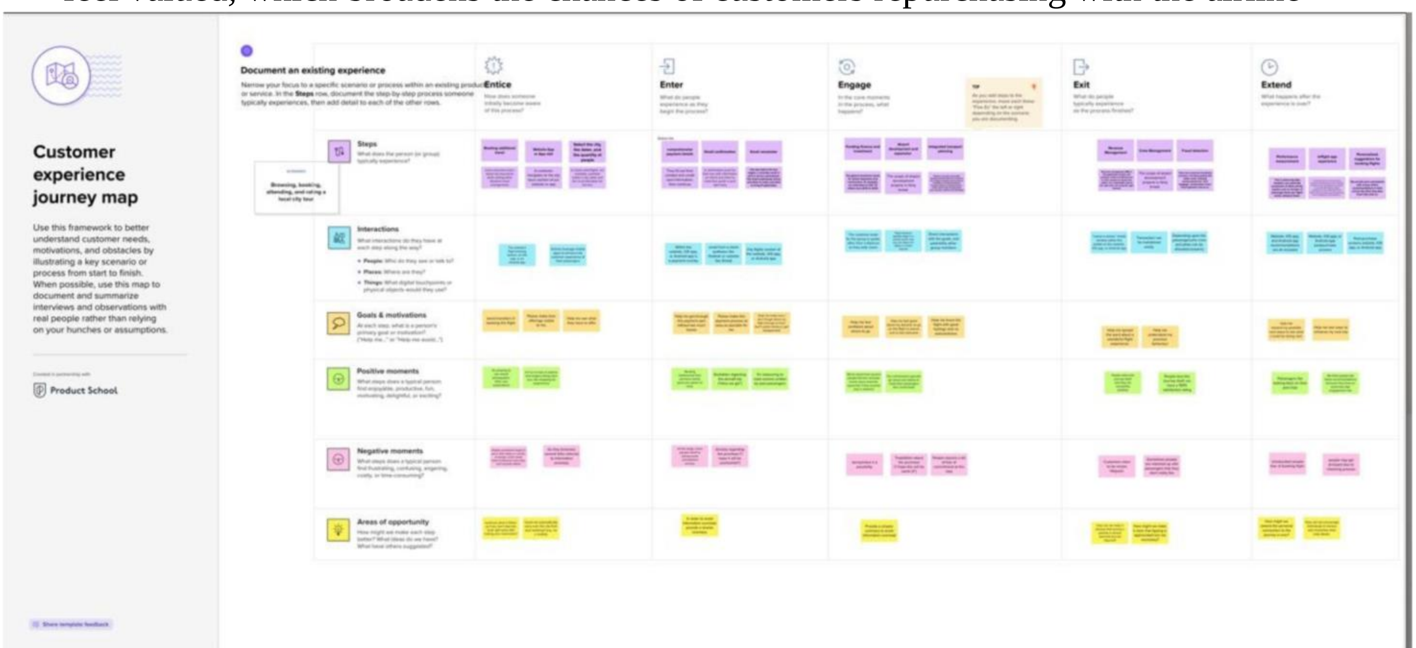
So, this stage takes place in both the digital as the social arena.

• Phase 2: decide: the customer decides whether or not to purchase the flight ticket on digitally or via the phone with a call-centre employee / through a physical store that offers the airline’s tickets. Furthermore, the customer is getting familiar with the airline by purchasing the ticket, which leads to an increase of flows to the social media accounts of the airline.

• Phase 3: use: happens when the customer will actually experience the flight that waspurchased with the airline. As the customer receives the boarding pass through email communication, the flight information as well as maps will be provided to the customer in order to prepare for the trip, which can be saved in a digital wallet. Other additionalresources, such as gate information and travel guides are provided to the customer either through email, social media or actual mail.

• Phase 4: support, as it is crucial for an airline to maintain contact with the customer before,during as well as after the flight. It is inevitable that things could go wrong during the flight which is why it is important that airlines show their support to the customer through either the digital and/or social arena.

• Phase 5: retain: airlines need to be able to retain customers, especially frequent flyers.Airlines can implement this by letting customers manage their own bookings on the website as well as offering discounts to returning customers as this makes the customer feel valued, which broadens the chances of customers repurchasing with the airline



5.2 Solution Requirements

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Customers can register using their Gmail account
FR-2	User Confirmation	The consumer will receive mail confirmation followingregistration
FR-3	Visualization of data	Using IBM cognos Analytics, a user can see the regulartrends in flight delay.
FR-4	Generation of report	Viewing the flight delay report is possible

Non-functional Requirements:

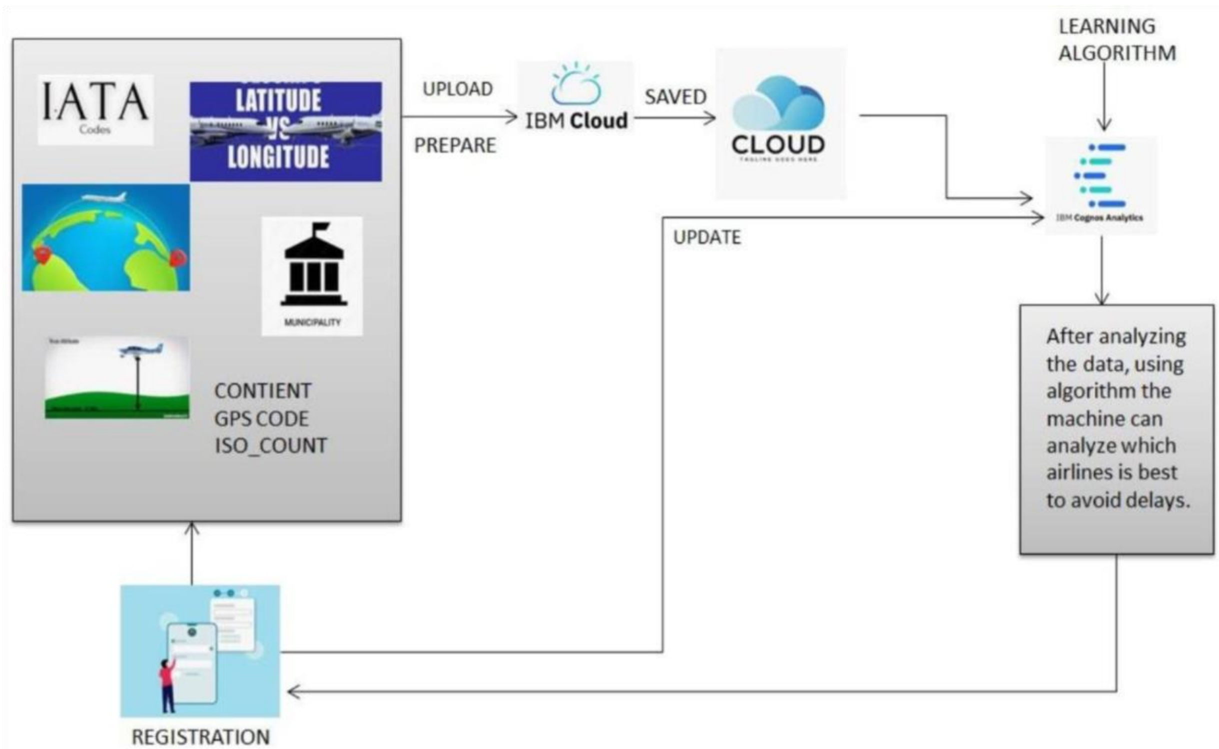
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The programme will have an easy-to-use graphical user interface. All the elements of the application would be simple for users to comprehend and utilize. Any activity must be carried out in a matter of clicks.
NFR-2	**Security**	Since user accounts are the main target of securityconcerns, adequate login procedures should be followed to prevent hacking. The system should not make public user personal information or other organization information.

NFR-3	**Reliability**	The system should save all user processes made upto the point of abnormal occurrences when it disconnects or freezes as a result of excessive simultaneous access.
NFR-4	**Performance**	The system need to require some speed, especially when navigating the catalogue.
NFR-5	**Availability**	The system must be accessible every day of the week,24 hours a day. Access is available at any time.

5.3 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



Use Case:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint1	Registration	USN-1	As a user, I can register for the application by	3	High	SURYAPRAK ASH S

			entering my email, password, and confirming my password.			
Sprint1	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	3	High	SANTHOS H P
Sprint1	Login	USN-3	As a user, I can log into the application by entering email & password	1	MEDIU M	SATHIYAMO ORTHI M
Sprint1	Accessing the dataset	USN-4	I can access the dataset and choose the different types of exploration can be done is analyzed as a user.	5	Medium	SIVACHAN DIRAN S
Sprint2	Exploration	USN-5	I can explore the given dataset through IBM Cognos Analytics with Watson	6	High	SURYAPRA KASH S
Sprint2	Visualization	USN-6	I will use Cognos as a visualization tool for the provided dataset into a dashboard	6	High	SANTHOSH P
Sprint3	Dashboard	USN-7	I can create the dashboard that is visualized as a user	6	High	SURYAPRA KASH S

Sprint3	Ease of Access	USN-8	I can simply access and use the dashboard as a user	5	Medium	SANTHOSH P
Sprint4	Generation of Report	USN-9	I can generate the report with the help of my visualization	6	High	SURYAPRAKASH S
Sprint4	Dashboard Establishment	USN-10	As a developer I can Established the dashboard into a website and submit the website	6	High	SIVACHANDIRAN S

5.4 Technology Stack

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

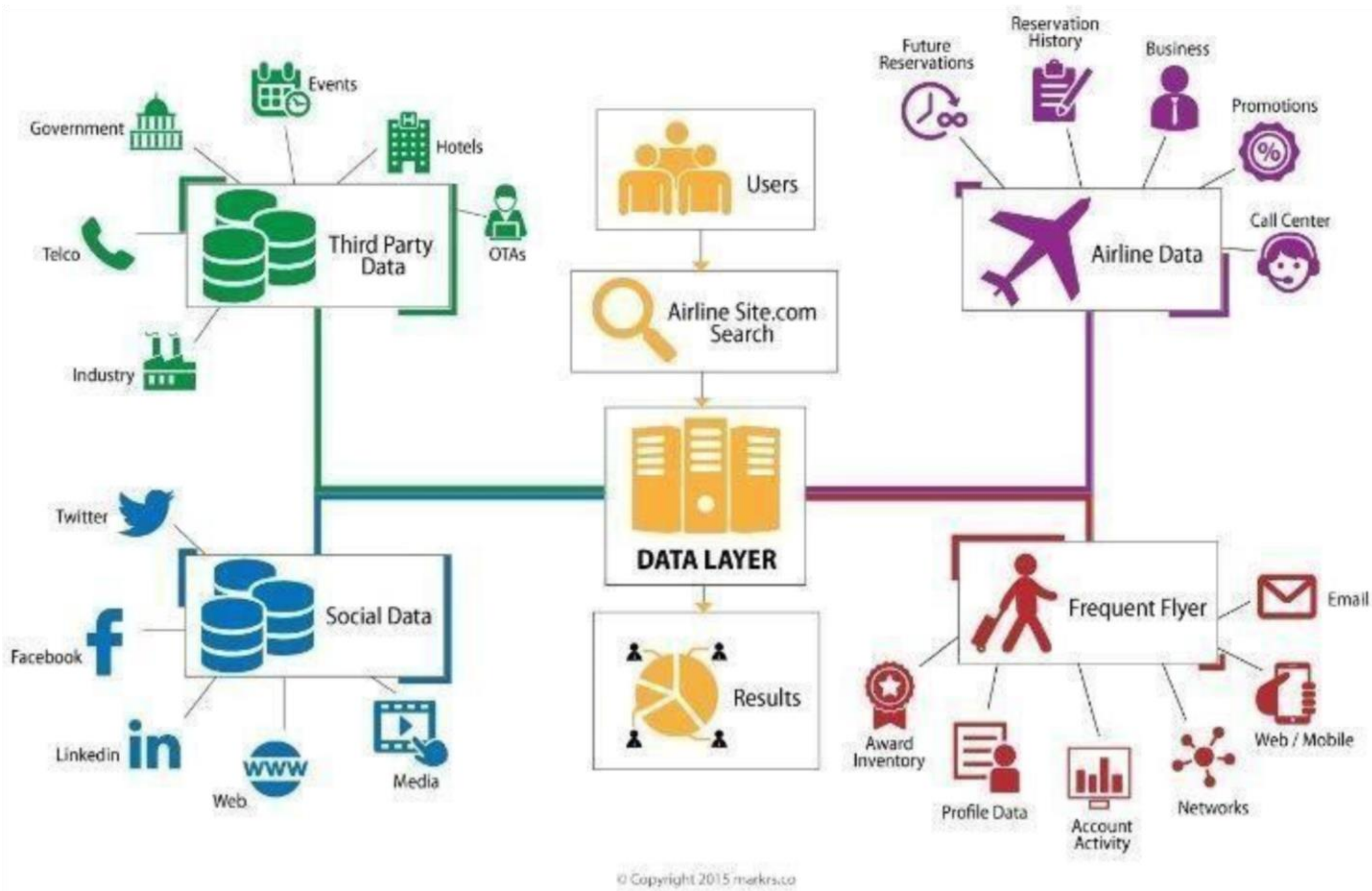


Table-1: Components & Technologies:

S.No	Components	Description	Technology
1.	User Interface	How user interacts with application. Example: Mobile App	HTML, CSS, Java Script,Excel
2.	Application Logic1	Logic for a process in theapplication	IBM Watson STTservice, Python

3.	Application Logic2	Logic for a process in theapplication	IBM Watson Assistant
4.	Database	Data Type, Configurations	MySQL, NSQL
5.	Cloud Database	Database service on cloud	IBM DB2, IBM Cloudant
6.	File Storage	File Storage requirements	IBM Blocks Storage orother storage serviceor Local File system
7.	External API-1	Purpose of External APIused in the application	IBM Weather API
8.	External API-1	Purpose of External APIused in the application	Aadhar API
9.	Infrastructure (Server/Cloud)	Application Deploymenton Local System/Cloud Local Server Configuration: Cloud Server Configuration	Local, Cloud Foundry

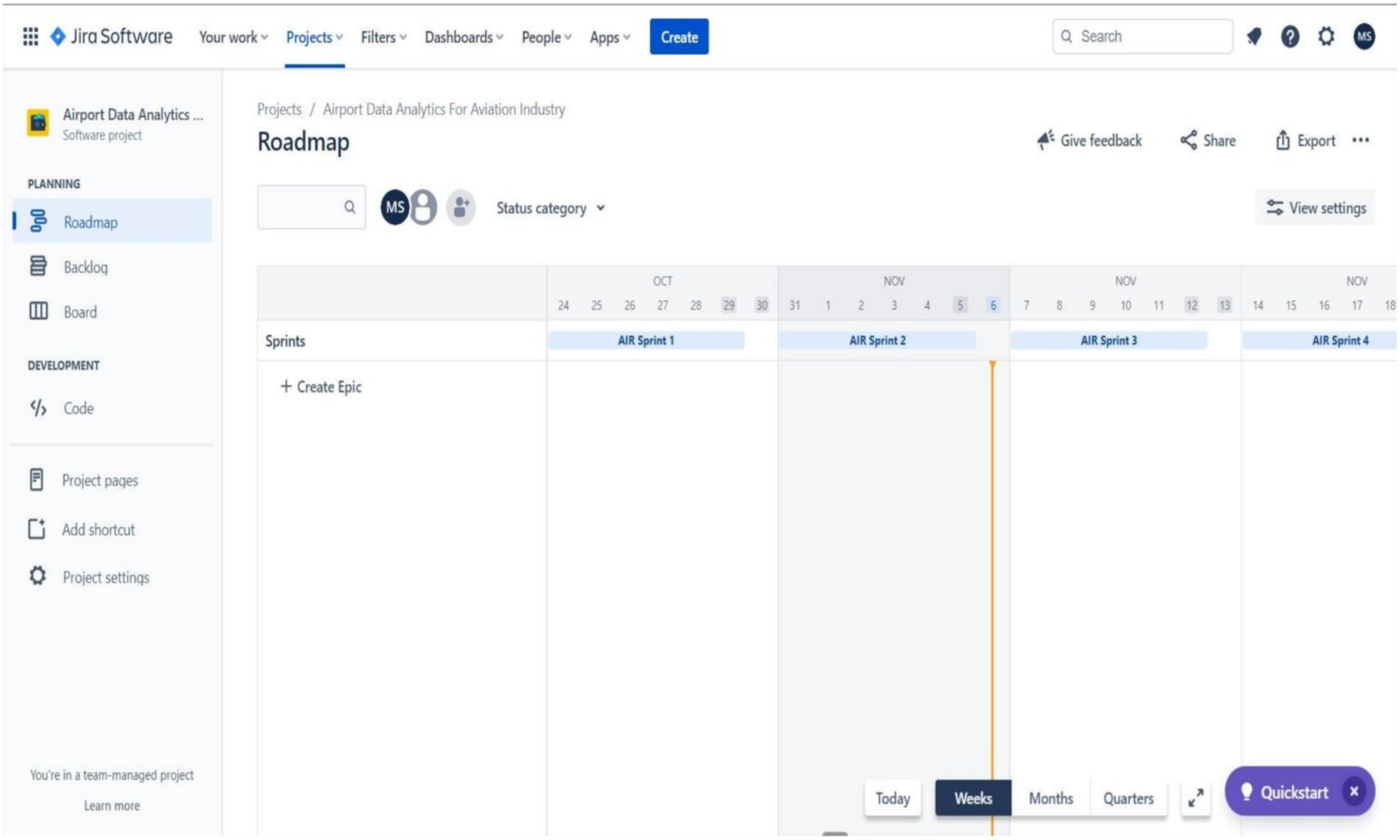
Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Technology of opensource framework
2.	Security Implementations	List all the security/access controls implemented,use of firewalls.	Example: SHA-256, Encryption, IAM Controls,OWASP

3.	Scalable Architecture	Justify the scalability ofarchitecture	Cognos Used
4.	Availability	Justify the availability of application (e.g: use of load balancers, distributed servers)	AWS Used
5.	Performance	Design consideration for the performance of the application (number of requests per second, use of Cache, use of CDN's)	Dashboard, Reports, Stories

6. PROJECT PLANNING PHASE

6.1 PrepareMilestoneandActivityList



6.2 Sprint Delivery Plan

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	3	High	SURYAPRAKASH S

Sprint1	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	3	High	SANTHOSH P
Sprint1	Login	USN-3	As a user, I can log into the application by entering email & password	1	Low	SATHIYAMOORTHI M
Sprint1	Accessing the dataset	USN-4	I can access the dataset and choose the different types of exploration can be done is analyzed as a user.	5	Medium	SIVACHANDHIRAN S
Sprint2	Exploration	USN-5	I can explore the given dataset through IBM Cognos Analytics with Watson	6	High	SURYAPRAKASH S
Sprint2	Visualization	USN-6	I will use Cognos as a visualization tool for the provided dataset into a dashboard	6	High	SANTHOSH P
Sprint3	Dashboard	USN-7	I can create the dashboard that is visualized as a user	6	High	SURYAPRAKASH S
Sprint3	Ease of Access	USN-8	I can simply access and use the dashboard as a user	5	Medium	SIVACHANDHIRAN S

Sprint4	Generation of Report	USN-9	I can generate the report with the help of my visualization	6	High	SURYAPRAKASH S
Sprint4	Dashboard Establishment	USN-10	As a developer I can Established the dashboard into a website and submit the website	6	High	SANTHOSH P

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Re lease Date (Actual)
Sprint1	20	6 Days	24 Oct 2022	29 Oct 2022	15	29 Oct 2022
Sprint2	20	6 Days	31 Oct 2022	05 Nov 2022	15	05 Nov 2022
Sprint3	20	6 Days	07 Nov 2022	12 Nov 2022	15	12 Nov 2022
Sprint4	20	6 Days	14 Nov 2022	19 Nov 2022	15	19 Nov 2022

Velocity:

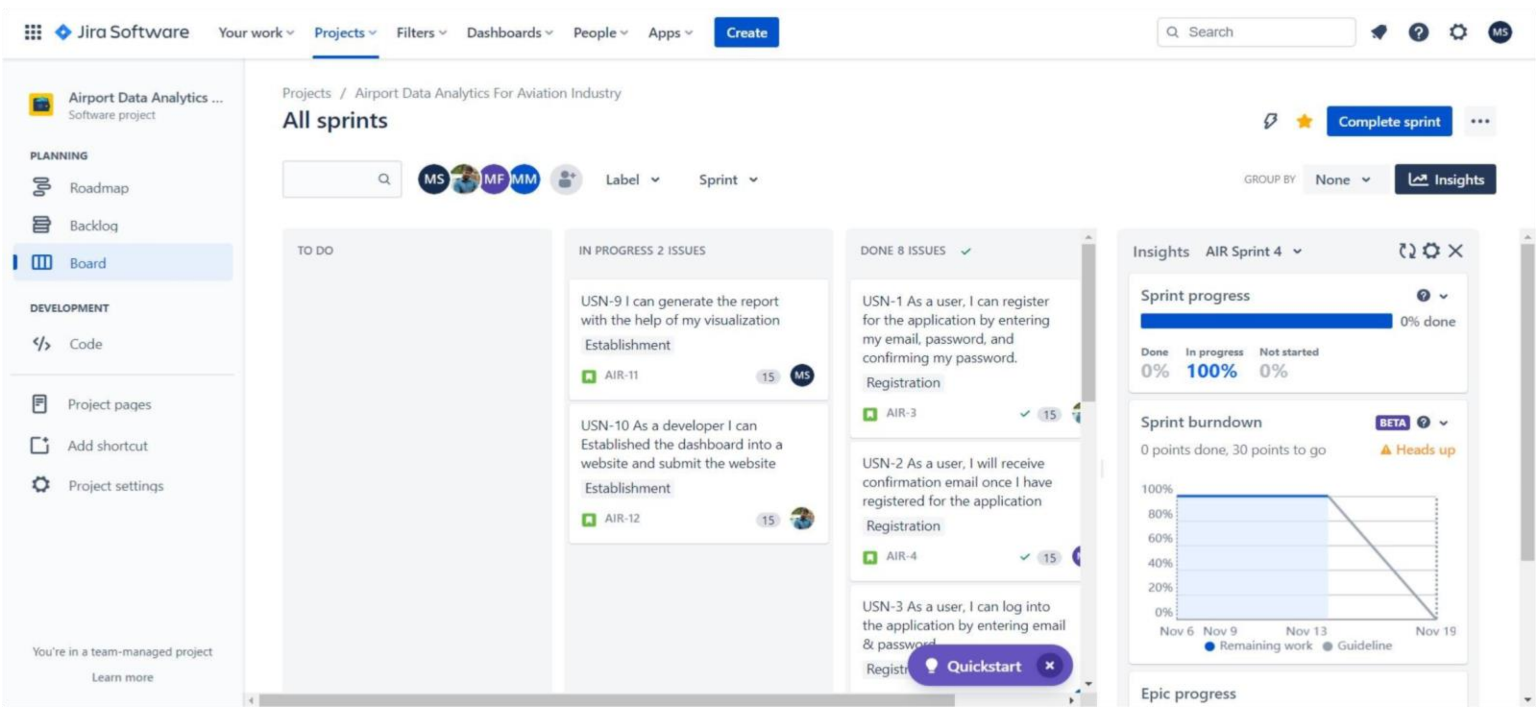
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let'scalculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Average velocity=Sprint duration / velocity=15/6=2.5

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to anyproject containing measurable progress over time.



7. PROJECT DEVELOPMENT PHASE

7.1 Project Development - Delivery of Sprint – 1

Flight count from Top 5 Airlines at Top 5 Airports

```
WITH top_5_airports AS (
    SELECT ORIGIN, COUNT(ORIGIN)
    AS countFROM
        airline-delay-
    canc.airlines_data.delay_canc_dataGROUP
    BY
        1
    HAVING
        count >
    100000ORDER
    BY
        2 DESC
    LIMIT 5
),
top_5_airlines AS (
    SELECT
        OP_CARRIER,
        COUNT(OP_CARRIER)
        AS count
    FRO
    M   airline-delay-
        canc.airlines_data.delay_canc_data main,
        top_5_airports top5
    WHERE
        top5.ORIGIN =
    main.ORIGINGROUP BY
        1
    ORDER BY
        2 DESC
    LIMIT 5
```

```
),
airportwise_carrier_c
        nt AS
        (SELECT
                main.ORIGIN AS
                Airport,
                main.OP_CARRIER AS
                Carrier,COUNT(*)   AS
                count
        FRO
        M       airline-delay-
                canc.airlines_data.delay_canc_data
                main,       top_5_airports       top5_ap,
                top_5_airlines top_al
        WHERE
                top5_ap.ORIGIN = main.ORIGIN
                AND top_al.OP_CARRIER =
        main.OP_CARRIERGROUP BY
                1,
                2
),
resut_cte AS (
        SELECT
                Airpo
                rt,
                Carrie
                r,
                count,
                RANK() OVER (PARTITION BY Airport ORDER BY count) AS rank
        FRO
        M       airportwise_carrier_cnt

)
SELECT
        Airpo
        rt,
        Carrie
        r,
```

```
    count
FROM
    resut_cte
WHERE
    rank < 6
```

Results

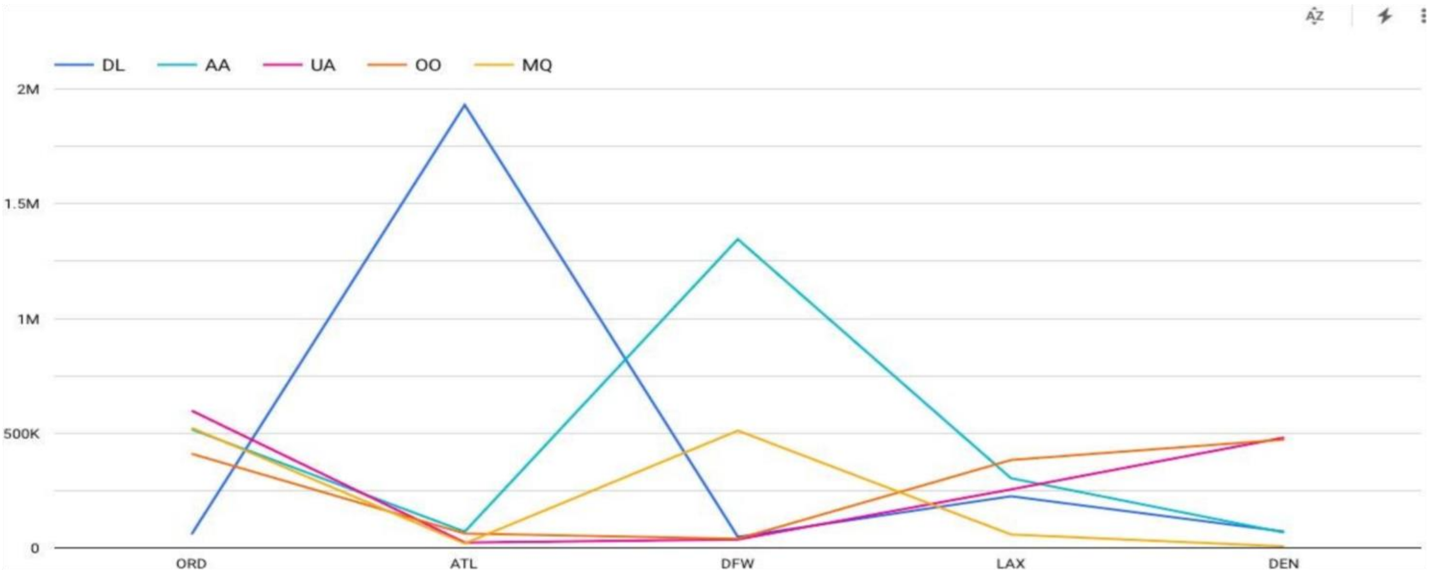
Top 5 Airports with maximum flight count:

1. **ORD** (O'Hare International Airport)
2. **ATL** (Hartsfield-Jackson Atlanta International Airport)
3. **DFW** (Dallas/Fort Worth International Airport)
4. **LAX** (Los Angeles International Airport)
5. **DEN** (Denver International Airport)

Top 5 Airlines with maximum flight count:

1. **DL** (Delta Air Lines)
2. **AA** (American Airlines)
3. **UA** (United Airlines)
4. **OO** (SkyWest Airlines)

5. **MQ** (AmericanEagleAirlines)



 From the above, it is realized that on **Delta Airlines** has the highest flight frequence onthe **Atlanta** airport.

Top 5 Airports with Maximum Cancellations (decreasing order)

```
WITH
 top_5_airports
 AS  (SELECT
 ORIGIN,
  COUNT(ORIGIN) AS
 countFROM
  `airline-delaycanc.airlines_data.delay_canc_data`
 GROUP BY
  1
 ORDER
 BY2
 DESC
 LIMIT
  5 ),
 top_5_airlines
 AS (SELECT
 OP_CARRIER,
  COUNT(OP_CARRIER) AS count
```

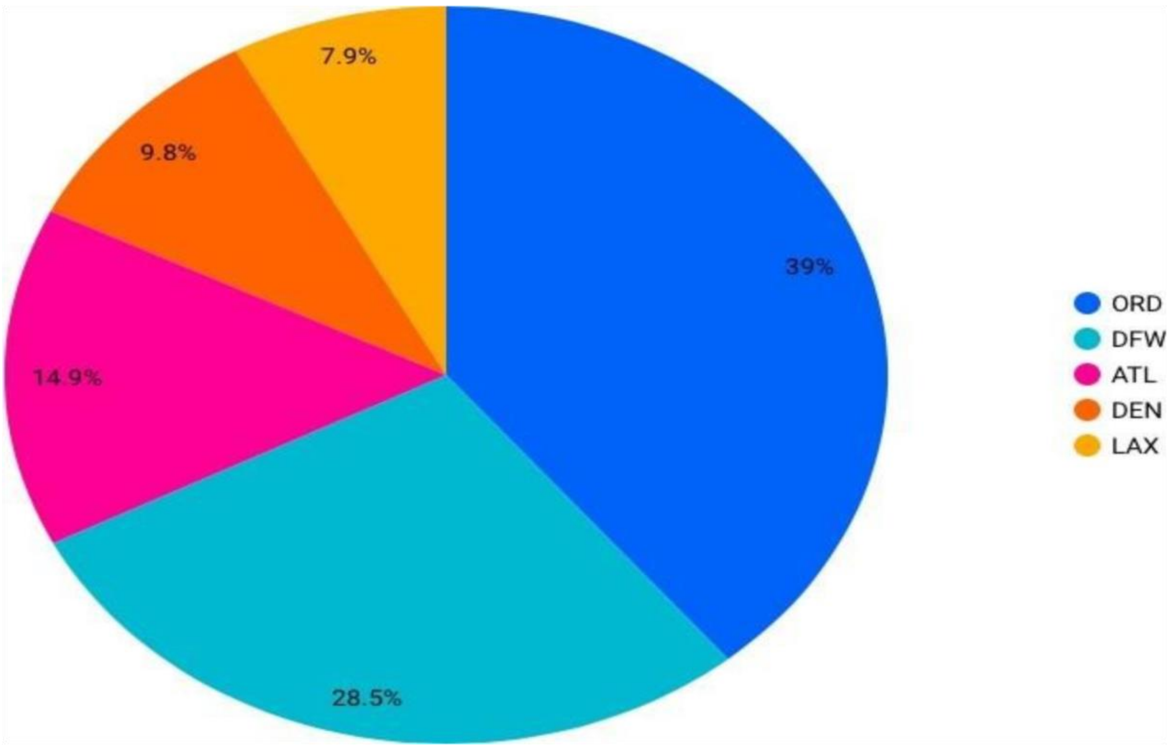
```
FROM
 `airline-delay-canc.airlines_data.delay_canc_data`
 main,top_5_airports top5
WHERE
top5.ORIGIN =
main.ORIGINGROUP
BY
 1
ORDER
BY2
DESC
LIMIT
 5),
all_flights
AS
(SELECT
 main.ORIGIN AS
 Airport,
 main.OP_CARRIER AS
 Carrier,COUNT(*) AS
 all_cnt
FROM
 `airline-delay-canc.airlines_data.delay_canc_data`
 main,top_5_airports
top5_ap, top_5_airlines
top_alWHERE
 top5_ap.ORIGIN = main.ORIGIN
AND top_al.OP_CARRIER =
main.OP_CARRIERGROUP BY
1,
 2 ),
cancelled_flights
AS (SELECT
 main.ORIGIN AS
Airport,
main.OP_CARRIER AS
Carrier, COUNT(*) AS
cancelled_cnt FROM
```

```
  `airline-delay-canc.airlines_data.delay_canc_data`
  main,top_5_airports
 top5_ap, top_5_airlines
 top_alWHERE
  top5_ap.ORIGIN = main.ORIGIN
 AND top_al.OP_CARRIER =
 main.OP_CARRIERAND cancelled = 1
 GROUP
  BY1,
  2 )
SELE
CT
 af.Airpo
 rt,
 af.Carri
 er,
 af.all_cnt
 -
 cf.cancell
 ed_cnt
 AS
     all_cnt
 ,cf.cancel
 led_cnt
FROM
 all_flights af,
cancelled_flight
s cf WHERE

 af.Airport            =
 cf.Airport       AND
 af.Carrier = cf.Carrier
```

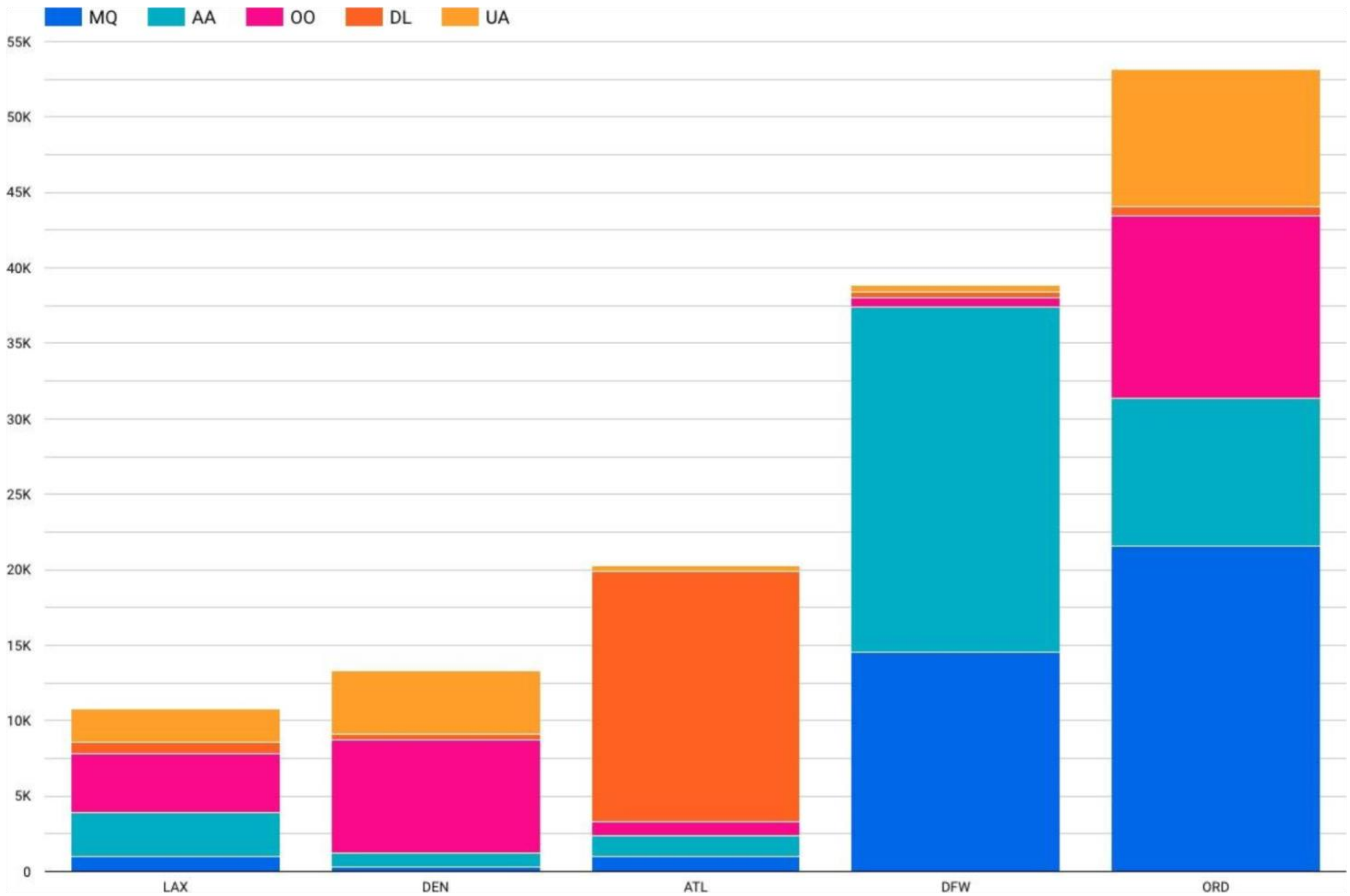
Results

| S No.| Airport Code | Airport Name | Cancellation (in %) | | - | - | - | - | - | 1. | **ORD** |

(O'Hare International Airport) | 39| | 2. | **DFW** | (Dallas/Fort Worth International Airport) | 28.5| | 3. | **ATL** | (Hartsfield-Jackson Atlanta International Airport) | 14.9| | 4. | **DEN** | (Denver International Airport) |9.8| | 5. | **LAX** | (Los Angeles International Airport) | 7.9|



Airline-wise Cancellation Bifurcation



7.2 Project Development - Delivery of Sprint – 2

Top Cancellation Reasons for Top 5 Busiest Airports

Query - JS UDF Function

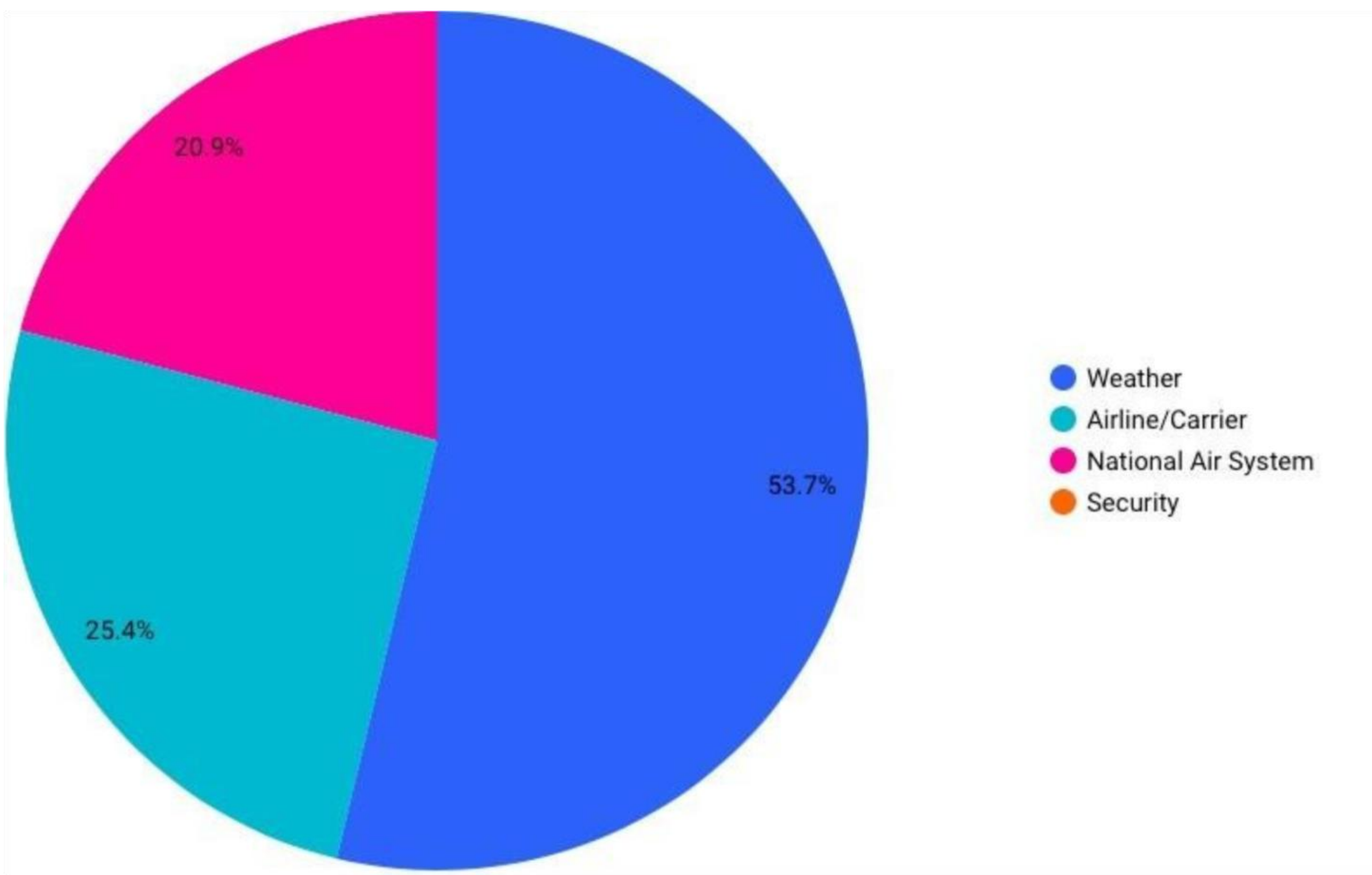
```
CREATE TEMP FUNCTION
  cancellation_reason(code string)
  RETURNS string
  LANGUAGE js AS """
    switch(code) {
        case "A":
          return "Airline/Carrier";
```

```
      break;
      case "B":
       return "Weather";
      break;
      case "C":
       return "National Air System";
      break;
      case "D":
       return "Security";
      break;
      default:
       return "Others";
      break;
 }
""";
WITH
 top_5_airports AS
 (SELECT
 ORIGIN,
 COUNT(ORIGIN) AS count
 FROM
   `airline-delay-canc.airlines_data.delay_canc_data`
 GROUP BY
   1
 HAVING
   count > 100000
 ORDER BY
   2 DESC
 LIMIT
   5 )
SELECT
 top5.ORIGIN,
 cancellation_reason(main.CANCELLATION_CODE) AS reason,
 COUNT(main.CANCELLATION_CODE) AS count
FROM
 `airline-delay-canc.airlines_data.delay_canc_data` main,
 top_5_airports top5
```

```
WHERE
  CANCELLED = 1
  AND EXTRACT(year
  FROM
    FL_DATE) = 2018
  AND top5.ORIGIN = main.ORIGIN
GROUP BY
  1,
  2
ORDER BY
  1,
  2
```

Result

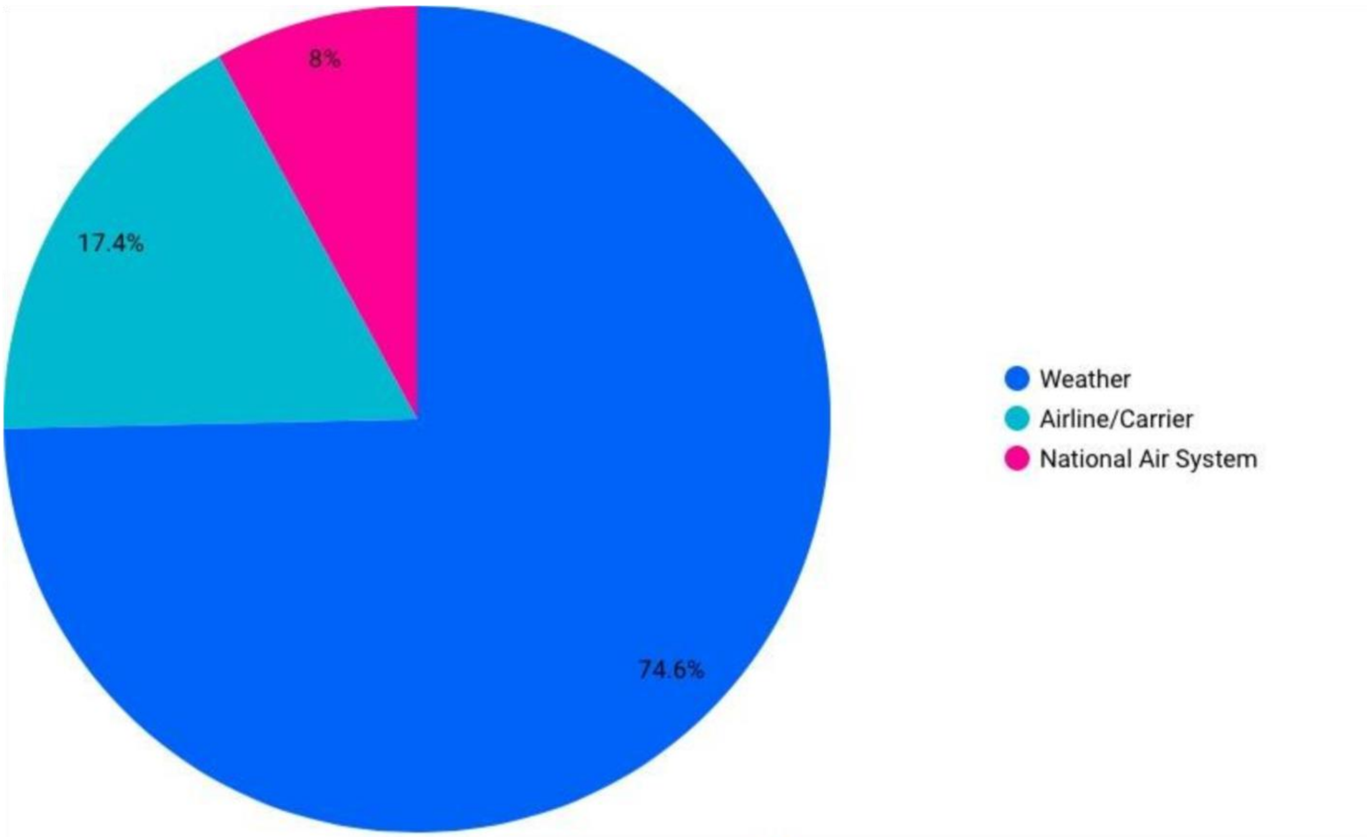
| S No.| Reason | Cancellation (in %) | | - | - | - | - | 1. | **Weather** | 53.7| | 2. | **Airline/Carrier Delays** |25.4| | 3. | **National Air System** | 20.9| | 4. | **Airport Secutiy** | 0.01 (~ 0)|



Top Cancellation Reasons at the Most Busiest Airport in practice (Atlanta)

- Atlanta is one of the largest inter-connect point (airport) for domestic and international flights inUSA.

|S No.| Reason | Cancellation (in %) | | - | - | - | - | 1. | **Weather** | 74.6| | 2. | **Airline/Carrier Delays** |17.4| | 3. | **National Air System** | 8|



7.3 Project Development - Delivery of Sprint – 3

Overall Delays at Top 5 Airports for top 5 airlines

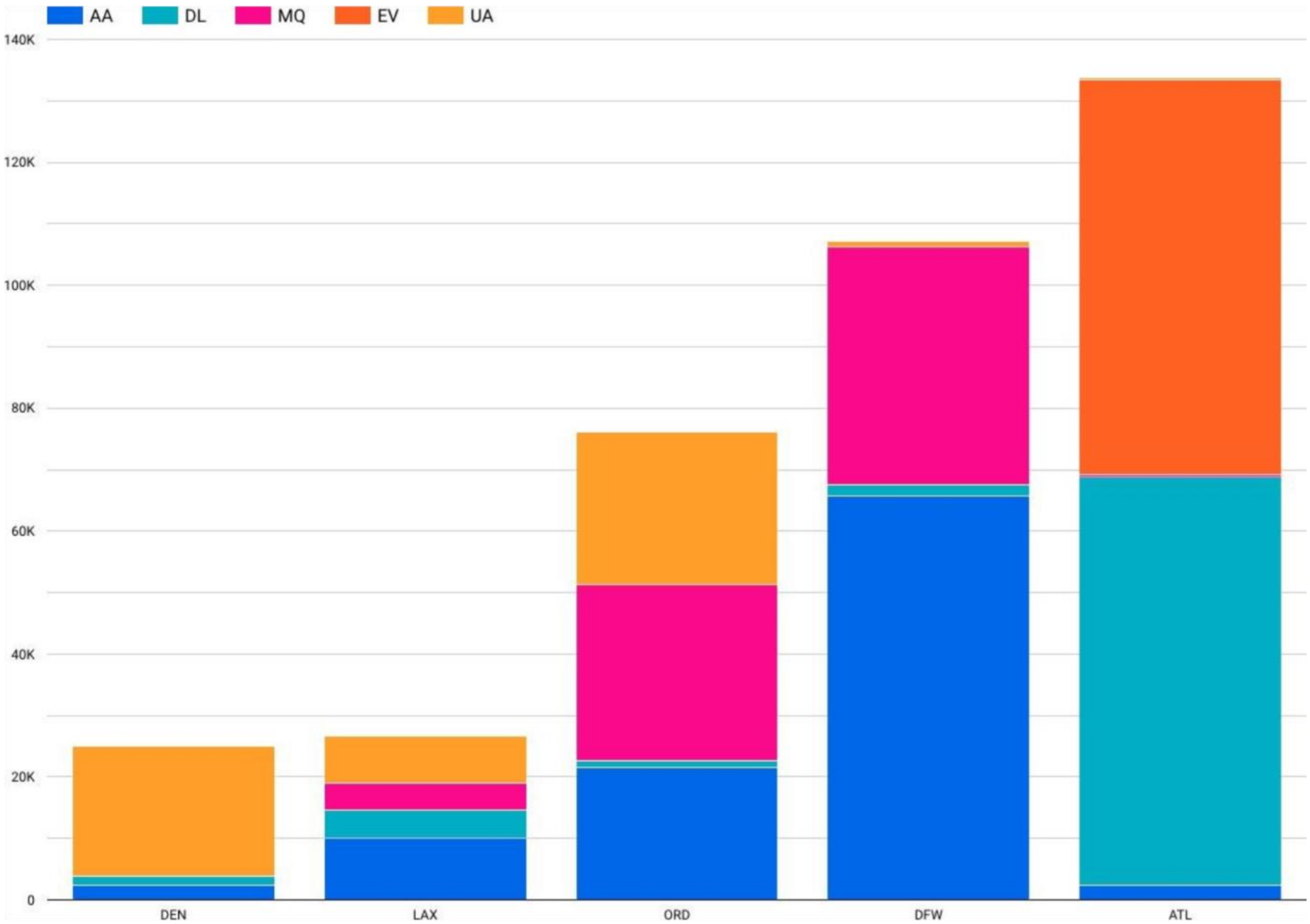
Query

```
WITH
  top_5_airports AS (
```

```
SELECT
   ORIGIN,
   COUNT(ORIGIN)
AS countFROM
   `airline-
delaycanc.airlines_data.delay_canc_data`GROU
P BY 1
ORDER BY
   2
DESC
LIMIT
   5                    ),
top_5_airlines AS
(SELECT
   OP_CARRIER,
   COUNT(OP_CARRIER)
   AS count
FROM
   `airline-delay-canc.airlines_data.delay_canc_data`
   main,top_5_airports top5
WHERE
   top5.ORIGIN =
main.ORIGINGROUP
BY
   1
ORDER BY
   2
DESC
LIMIT
   5),
all_flights AS
(SELECT main.ORIGIN AS
   Airport,
   main.OP_CARRIER    AS
   Carrier,COUNT(*) AS
   all_cnt
FROM
   `airline-delay-canc.airlines_data.delay_canc_data`
```

```
main,top_5_airports top5_ap,
top_5_airlines top_alWHERE
top5_ap.ORIGIN = main.ORIGIN
AND top_al.OP_CARRIER =
main.OP_CARRIERGROUP BY
1,
2 ),
delayed_flights AS
(SELECT main.ORIGIN AS
Airport,
main.OP_CARRIER AS
Carrier,COUNT(*) AS
delayed_cnt
FROM
`airline-delay-canc.airlines_data.delay_canc_data`
main,top_5_airports top5_ap,
top_5_airlines top_alWHERE
top5_ap.ORIGIN = main.ORIGIN
AND top_al.OP_CARRIER =
main.OP_CARRIERAND
(CARRIER_DELAY IS NOT NULL
AND CARRIER_DELAY > 0
OR ARR_DELAY IS NOT NULL

AND
ARR_DELAY > 0)
GROUP BY
1,
2 )
SELE
CT
af.Airport, af.Carrier, af.all_cnt all_with_del,
df.delayed_cnt, af.all_cnt - df.delayed_cnt AS
all_without_del
FROM
all_flights af,
delayed_flights
df
WHERE
```

af.Airport = df.Airport
AND af.Carrier =
df.Carrier *Overall Delays at Top 5 Airports with top 5 airlines*



Overall Delay Time Frequency with Top 5 Airports

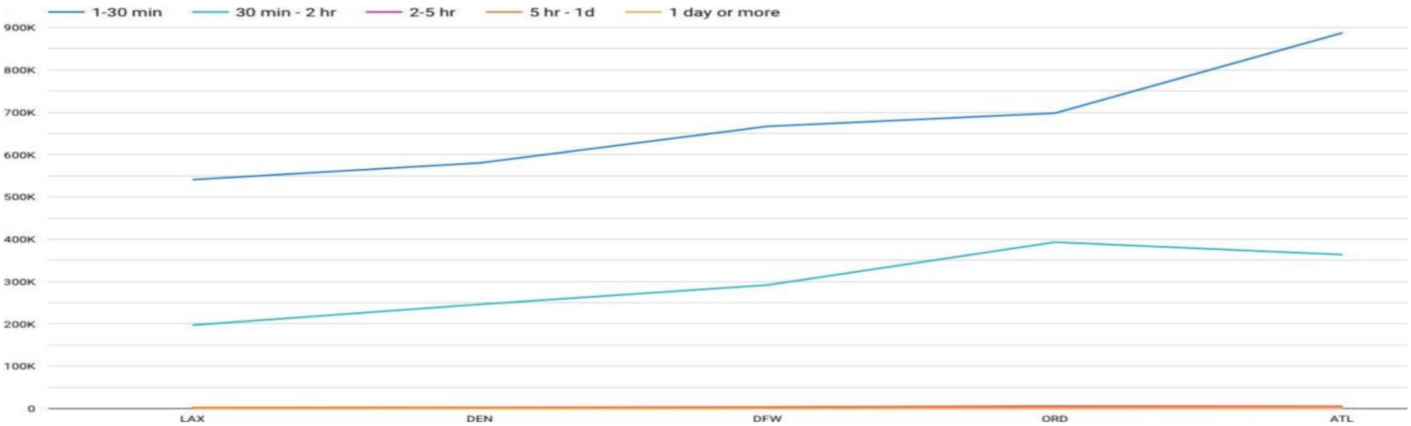
Query

CREATE TEMP FUNCTION delay_bifurcation(slot_cnt ARRAY<STRUCT<slot int64,count int64>>)
RETURNS STRUCT<cnt_1_30 float64, cnt_30_2 float64, cnt_2_5 float64, cnt_5_24 float64, cnt_24float64>
LANGUAGE js AS """
let response = {"cnt_1_30": 0.0, "cnt_30_2": 0.0, "cnt_2_5": 0.0, "cnt_5_24": 0.0, "cnt_24": 0.0} for(let i = 0 ; i < slot_cnt.length;
i++){let slotCntObj = slot_cnt[i];

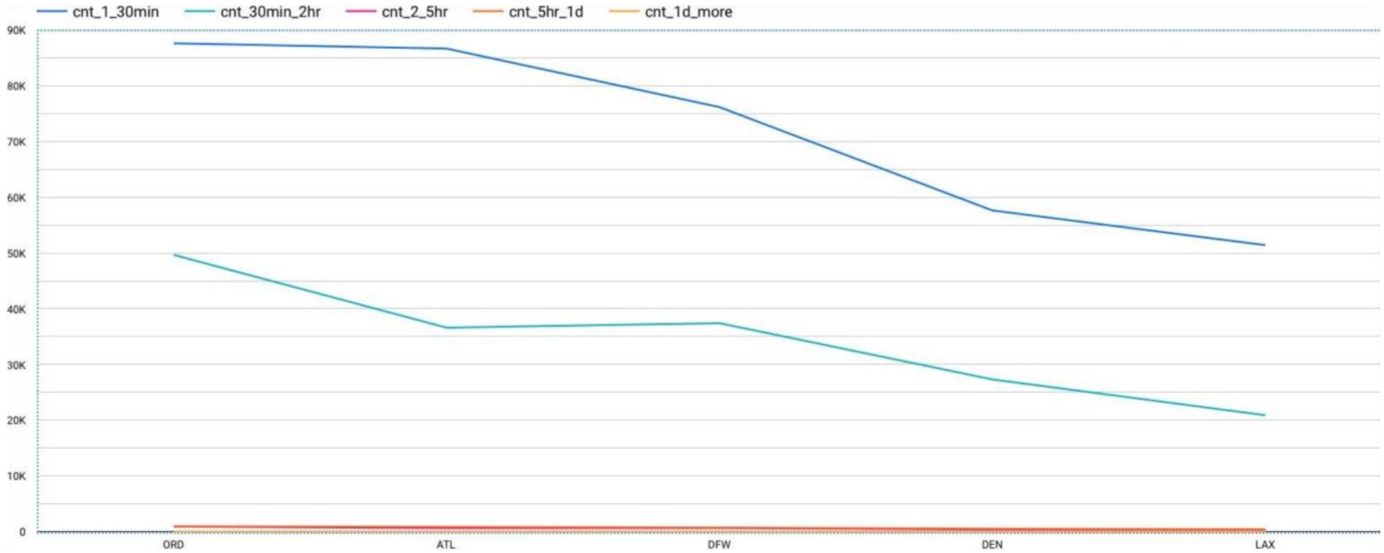
```
        let result = slotCntObj.count; switch(parseInt(slotCntObj.slot))
        {
          case  1:  response["cnt_1_30"]  =
            result; break;
          case                          2:
            response["cnt_30_2"] =
            result;break;
          case                        3:
            response["cnt_2_5"] =
            result;break;
          case                          4:
            response["cnt_5_24"] =
            result;break;
          case 5:

            response["cnt_24"]      =      result;
            break;
          default:
            response["cnt_1_30"] = 0.0;
            response["cnt_30_2"] = 0.0;
            response["cnt_2_5"]  =  0.0;
            response["cnt_5_24"] = 0.0;
            response["cnt_24"] =
            0.0;break;
        }
      }            return
      response
""";

WITH top_5_airports as (
        SELECT ORIGIN, count(ORIGIN) as count
        FROM `airline-delay-canc.airlines_data.delay_canc_data`
        Group by 1
        having count >
        100000order by
        2 desc limit 5
        ),
      delay_bifurcation as
        (select ORIGIN,
```

```
            (case when ARR_DELAY >
              1440 then 5when
              ARR_DELAY > 300 then 4 when
              ARR_DELAY > 240 then 3
              when ARR_DELAY >
          30 then 2else 1 end) as slot
    from `airline-delaycanc.airlines_data.delay_canc_data`where
    ARR_DELAY is not null and ARR_DELAY > 0
--      and EXTRACT(year FROM FL_DATE) = 2018
    ),

    airport_timeslots as( select db.ORIGIN,
    db.slot, count(db.slot) as countfrom
    delay_bifurcation db,top_5_airports top5
    where top5.ORIGIN = db.ORIGIN group by
    1,2),

    airport_struct as( select origin, struct(slot,count) as slot_cnt from
       airport_timeslots
    ),
    udf_result as (select origin, delay_bifurcation(ARRAY_AGG(slot_cnt)) as
    slot_structfrom airport_struct group by 1
    )
    select origin, slot_struct.cnt_1_30 as
       cnt_1_30min,slot_struct.cnt_30_2 as
       cnt_30min_2hr, slot_struct.cnt_2_5 as
       cnt_2_5hr, slot_struct.cnt_5_24 as
       cnt_5hr_1d, slot_struct.cnt_24 as
       cnt_1d_more
    from udf_result
```

Overall Delay Time Frequency with Top 5 Airports (UDF Function)



Overall Delay Frequency (Year with max delays and cancellations)



7.4. Project Development - Delivery of Sprint – 4

Delay Percentage for top 5 airports

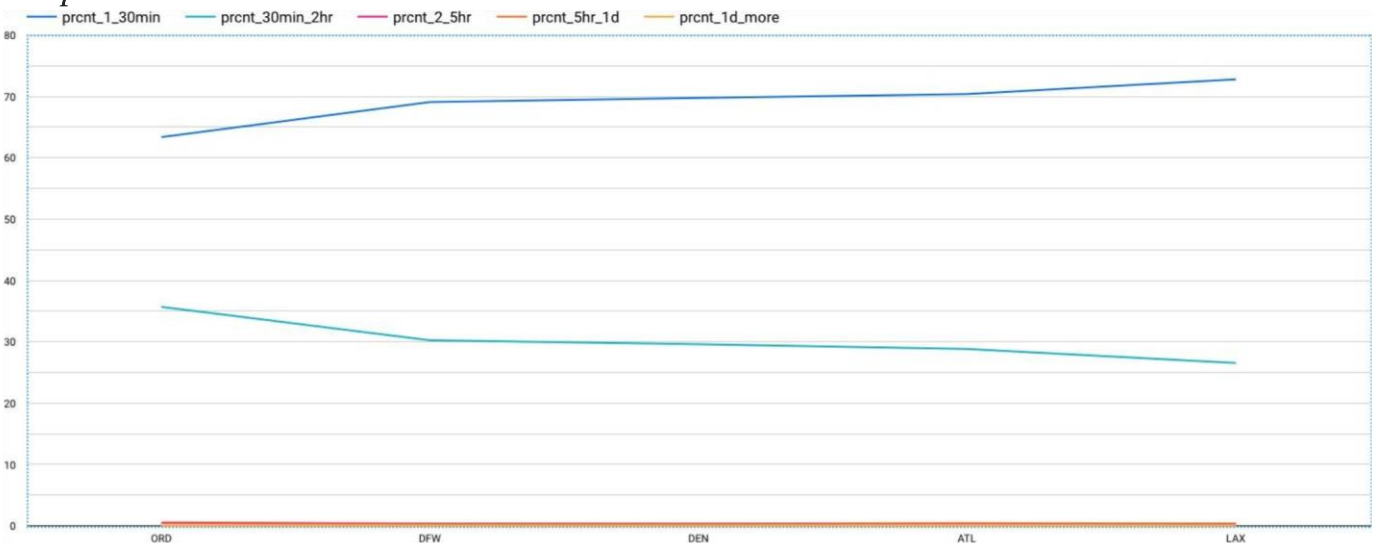
Query

```
CREATE TEMP FUNCTION delay_bifurcation(slot_cnt ARRAY<STRUCT<slot int64,count
int64>>)
    RETURNS STRUCT<cnt_1_30 float64, cnt_30_2 float64, cnt_2_5 float64, cnt_5_24 float64,
cnt_24float64>
  LANGUAGE js AS """
```

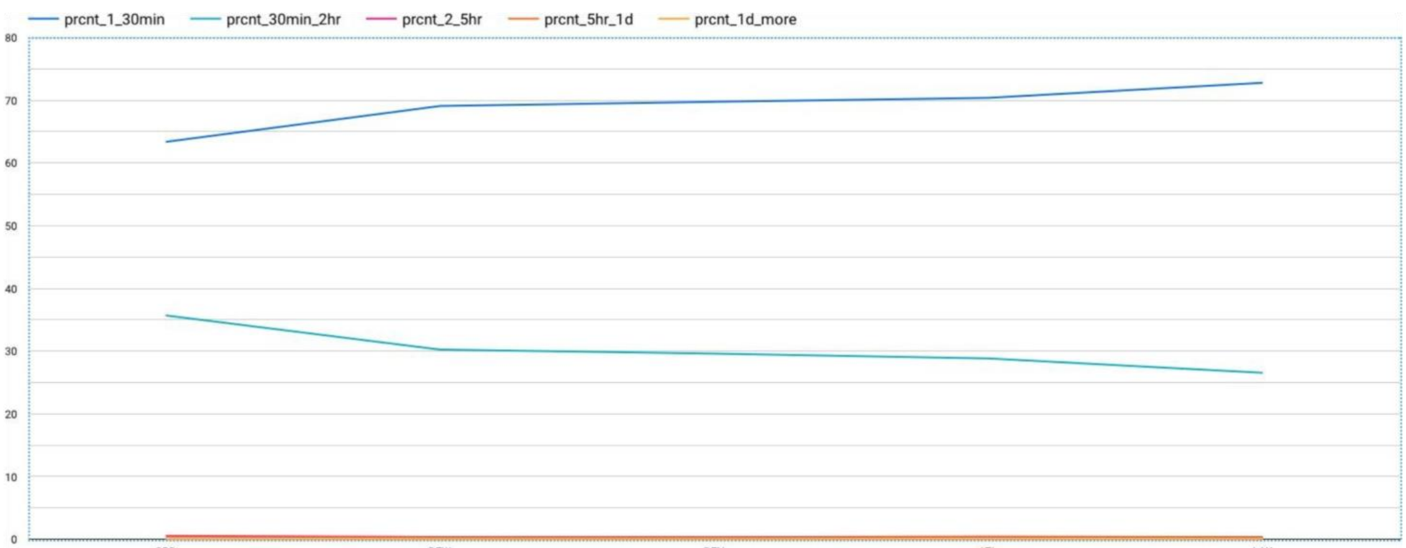
```
  let response = {"cnt_1_30": 0.0, "cnt_30_2": 0.0, "cnt_2_5": 0.0, "cnt_5_24": 0.0, "cnt_24":
0.0} let total_delayed_flights = 0;
  for(let i = 0 ; i < slot_cnt.length;
    i++){ total_delayed_flights +=
    parseInt(slot_cnt[i].count);
  }
  for(let i = 0 ; i < slot_cnt.length; i++){let slotCntObj
      = slot_cnt[i];
      let result =      parseFloat(parseInt(slotCntObj.count) / total_delayed_flights
*100).toFixed(2);
      switch(parseInt(slotCntObj.slot)){ca
      s e 1: response["cnt_1_30"] = result;
      break;

        case 2:
          response["cnt_30_2"] =
          result;break;
        case                            3:
          response["cnt_2_5"] =
          result;break;
        case                              4:
          response["cnt_5_24"] =
          result;break;
        case  5:  response["cnt_24"]  =
          result; break;
        default:
          response["cnt_1_30"] = 0.0;
          response["cnt_30_2"] = 0.0;
          response["cnt_2_5"]  =  0.0;
          response["cnt_5_24"] = 0.0;
          response["cnt_24"] =
          0.0;break;
      }
    }            return
    response
""";

WITH top_5_airports as (
      SELECT ORIGIN, count(ORIGIN) as count
```

```
FROM `airline-delay-canc.airlines_data.delay_canc_data`
Group by 1
having count >
100000order by 2
desc
limit 5
),
delay_bifurcation as
(select ORIGIN,
(case when ARR_DELAY > 1440
then 5when
ARR_DELAY > 300 then 4
when ARR_DELAY > 240 then 3 when
ARR_DELAY >
30 then 2else 1 end) as slot

from `airline-delaycanc.airlines_data.delay_canc_data`where
ARR_DELAY is not null and ARR_DELAY > 0

--    and EXTRACT(year FROM FL_DATE) = 2018 -- used for filtering
),

airport_timeslots as( select db.ORIGIN,
db.slot, count(db.slot) as countfrom
delay_bifurcation db,top_5_airports top5
where top5.ORIGIN = db.ORIGIN group by
1,2),

airport_struct as( select origin, struct(slot,count) as slot_cnt from
airport_timeslots
),
udf_result as (select origin, delay_bifurcation(ARRAY_AGG(slot_cnt)) as slot_struct

from
airport_struct
group by 1
)
select origin, slot_struct.cnt_1_30 as
prcnt_1_30min,slot_struct.cnt_30_2 as
prcnt_30min_2hr, slot_struct.cnt_2_5 as
```

prcnt_2_5hr, slot_struct.cnt_5_24 as
prcnt_5hr_1d, slot_struct.cnt_24 as
prcnt_1d_more
from udf_result *Delay Percentage for top 5 airports*



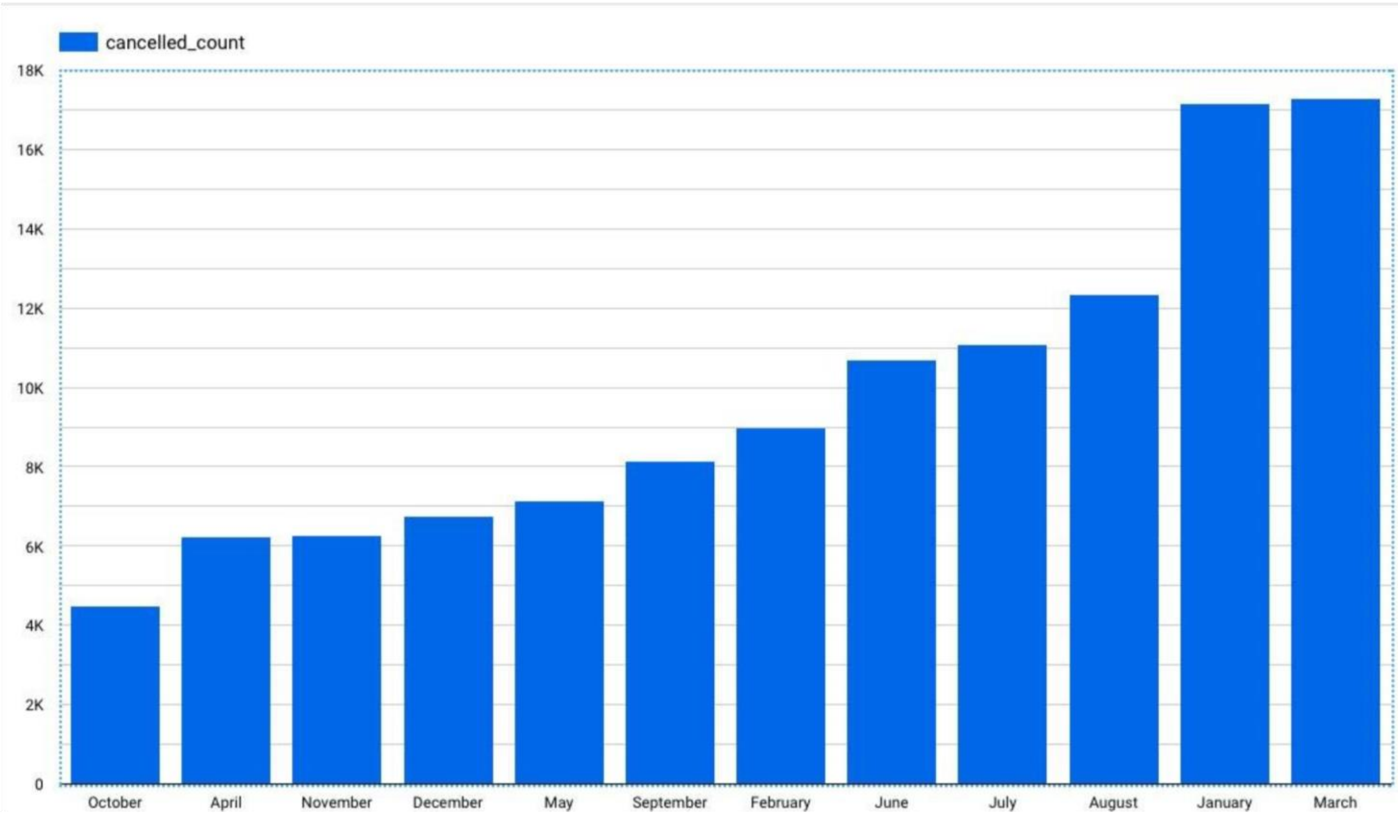
Delay Percentage for top 5 airports



Most unreliable month (Cancellations in ascending order)

Query

```
WITH
  cancelled_count_cte
  AS (SELECT
    *,
    ROW_NUMBER()         OVER         (ORDER         BY
  cancelled_count) AS RANKFROM
    (SELECT
      FORMAT_DATE('%B', FL_DATE) AS month,
      SUM(CANCELLED) AS
    cancelled_countFROM
      `airline-delay-canc.airlines_data.delay_canc_data`
    WHERE
      EXTRACT
      (year
      FROM
        FL_DATE) =
    2018GROUP BY
      1) )
SELECT
  month,
  cancelled_co
  u nt
FROM
  cancelled_count_c
teORDER       BY
  rank DESC
```


cancelled_count
18K
16K
14K
12K
10K
8K
6K
4K
2K
0
October
April
November
December
May
September
February
June
July
August
January
March

CONCLUSION

It can be used **to predict future glitches, prevent them from happening, andmake the maintenance procedures more accurate and thorough**.

After analyzing the data, a lot of insights have been generated. Most of the delaysand cancellations are due to three major reasons:

Weather

Airline/Carrier Issues

National Air System

9. REFERENCES

https://www.iata.org/en/publications/store/world-air-transport-statistics/

https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=data%20analytics%20in%20aviation%20industry&ved=2ahUKEwiW86_T9Zr7AhWkl
-YKHSzFC9YQsKwBKAB6BAhDEAE

https://www.google.com/search?lei=cl9oY5byKqSvmg
esiq-wDQ&q=impact%20of%20covid-
19%20on%20aviation%20industry%20research%20paper&ved=2ahUKEwiW86_T9Zr7Ah
W kl- YKHSzFC9YQsKwBKAJ6BAhDEAM https://dl.acm.org/doi/abs/10.1145/3469028
https://www.nap.edu/read/21909/chapter/5