PROJECT REPORT

Airlines Data Analytics for Aviation Industry

TEAM ID: PNT2022TMID35420

TEAM MEMBERS:

1. RAHIL JAMAL (2019105070)
2. SRUTHI RAJESH (2019105059)
3. PRASHANNAA VINAYAK RAJAN (2019105039)
4. INDRAJEET S (2019105015)

INDEX

7. RESULTS

Performance Metrics

8. ADVANTAGES & DISADVANTAGES

9. CONCLUSION

10. FUTURE SCOPE

11. APPENDIX Source Code GitHub & Project Demo Link

I. INTRODUCTION

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 intension is to develop a model for the airline data to provide platform for new analytics based on the following queries. Data analyst can be used to predict future glitches, prevent them from happening, and make the maintenance procedures more accurate and thorough. As a result, it is possible to lower costs related to maintaining an aircraft.

The airport codes may refer to either the IATA airport code, a three-letter code that is used in passenger reservation, ticketing and baggage handling systems, or the ICAO airport code which is a four-letter code used by ATC systems and for airports that do not have an IATA airport code. In this project based on the customer reviews and flight arrival timing and cost the best flight is determined.

PROJECT OVERVIEW

- The main aim is to provide better Airline and AirPort services and to avoid delays in Air Travel across different locations at Municipality level.

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

- Data analysis on flight dataset to draw inferences on arrival and departure delays and to identify relationships between flight timings and delays. Using the flight delay data, we identified which flight is mostly prone to delays. The arrived upon conclusions are useful for selecting flights in the future. From the review of the customer and the flight which covers the destination in correct time and in shortest time that airline flight will be selected as a best airline service.

PURPOSE

- To provide better Airline and AirPort services and to avoid delays in Air Travel across different locations at Municipality level.

- The aim is to provide airports, airlines, and the travelling public with a neutral, third-party view of which airlines are delivering on their promise to get passengers from Point A to Point B on-time.

2. LITERATURE SURVEY

TITLE: On the relevance of data science for flight delay research.

AUTHORS: Leonardo Carvalho, Alice Stenberg, Leandro maia goncalves,

Ana Beatriz cruz, Jorge A,soares.

YEAR: 2018.

DESCRIPTION:

Flight delays are a significant problem for society as they evenly impair airlines, transport companies, facility managers, and passengers. Studying prior flight data is an essential activity for every player involved in the air transportation system. Besides, developing accurate prediction models for flight delays is a crucial component of the decision-making process. Prescribing actions to solve on-going delays is an even challenging task due to the air transportation system complexity. In this regard, this paper presents a thorough literature review of data science techniques used for investigating flight delays. This work proposes a taxonomy and compiles the initiatives used to address the flight delay studies.

PROS:

- Accurately predicting these flight delays allows passengers to be well prepared for the deterrent caused to their journey.
- Enables airlines to respond to the potential causes of the flight delays in advance to diminish the negative impact.

CONS:

- Late due to weather predicting this is difficult.
- A few factors responsible for the flight delays like runway construction to excessive traffic are rare, but bad weather seems to be a common cause.

TITLE: Aviation management.

AUTHORS: Shi Qiang Liu, Andrea D'Ariano, Erhan Kozan, Mahmoud

Masoud CARRS-Q, SaiHo Chung**.**

YEAR: 2019.

DESCRIPTION:

Aviation or air transportation refers to the activities surrounding mechanical flights in the airlines and the aircraft industries. In this paper, we present a recent literature survey on aviation management. The literature review is classified into the following main categories: Airline Capacity Analysis; Air Traffic Flow Management; Airline Fleet Assignment;
Tail Assignment with Aircraft Maintenance Routing; Airline Crew Pairing; Airline Recovery and Rescheduling; Airline Revenue Management; Collaborative Decision Making; Aircraft Scheduling. This classification aims to motivate the researchers and practitioners in aviation management to develop more applicable, realistic and wide-ranging optimization methodologies for meeting the current needs of aviation industry.

PROS:

- Advanced scheduling optimization tools for the better management of the available infrastructure and resources.
- Accurate timing information so that conflicts between aircraft are resolved.

CONS:

- Air traffic control operations and related issues are still scheduled by human controllers.
- Ignore any military/defiance use of drones.

TITLE: Predictive Analytics Platform for Airline Industry.

AUTHORS: P. H. K Tissera, A.N.M.R.S.P. llwana, K.T. Waduge, M.A.l. Perera, D.P. Nawinna, D. Kasthurirathna.

YEAR: 2020.

DESCRIPTION:

The research is to develop accurate demand forecasting model to control the availability in Airline industry. The primary outcome of the model is that the Airline organization can maximize the revenue by controlling the availability. The product in airline industry is the seat, which is an expensive, unstock able product. The demand for the seats is almost uncertain, the capacity is constraint and difficult to increase and the variable costs are very high. The revenue is derived by the number of passengers and the fares they pay which vary for each flight. Hence, it is challenging to develop an accurate method to project the revenue for each route. We have the current ticketed revenue plus we have the current booked passengers. We also have the ticketed passenger details of previous flights. Hence most of the information is available, however changing market conditions is an unknown variable which can have a significant impact on passenger travel patterns.

PROS:

- Focus on the passenger demand forecasting, average fare forecasting, no show forecasting and visualizing the passenger demand and annual revenue prediction for od level point of sales.
- Reliability is improved.

CONS:

- With limitation of predictors because of sensitivity of the data and limited access to the data it may have impacted the models and the accuracy.

TITLE: Exploratory data analysis on aviation dataset.

AUTHORS: Saba Firdous; Haseeba Fathiya; Lipsa Sadath.

YEAR: 2021.

DESCRIPTION:

The usage of big data analytics is booming today, with its ability to be used to draw useful insights from past data research. Its uses in the aviation industry have a wide array of applications ranging from predicting flight delays to detecting faults in airplane parts. In this paper, we conducted exploratory data analysis on flight dataset to draw inferences on arrival and departure delays and to identify relationships between flight timings and delays. Using the flight delay data, we identified which flight is mostly prone to delays. The arrived upon conclusions are useful for selecting flights in the future.

PROS:

- Data collected from customer profiles, social behavior, etc. can be efficiently used by airlines to provide personalized services to customers.
- They can also be used to analyze passenger flow, cost reduction and to enhance revenue.

CONS:

- When the number of flight arrivals and departures is very high, there can be and disparity between the capacity of the flight to handle the demands and its capacity, leading to many delays.
- Bad weather such as floods, hurricanes could also be the cause

REFERENCE

[1]. Leonardo Carvalho, Alice Stenberg, Leandro maia goncalves, Ana Beatriz cruz, Jorge A,soares, On the relevance of data science for flight delay research, 2018.

[2]. Shi Qiang Liu, Andrea D'Ariano, Erhan Kozan, Mahmoud Masoud CARRS-Q, SaiHo Chung**,** Aviation management, 2019.

[3]. P. H. K Tissera, A.N.M.R.S.P. llwana, K.T. Waduge, M.A.l. Perera, D.P. Nawinna, D. Kasthurirathna, Predictive Analytics Platform for Airline Industry, 2020.

PROBLEM STATEMENT DEFINITION

The airport codes may refer to either the IATA airport code, a three-letter code that is used in passenger reservation, ticketing and baggage handling systems, or the ICAO airport code which is a four-letter code used by ATC systems and for airports that do not have an IATA airport code. To provide better Airline and AirPort services and to avoid delays in Air Travel across different locations at Municipality level. The aim is to provide airports, airlines, and the travelling public with a neutral, third-party view of which airlines are delivering on their promise to get passengers from Point A to Point B on-time.

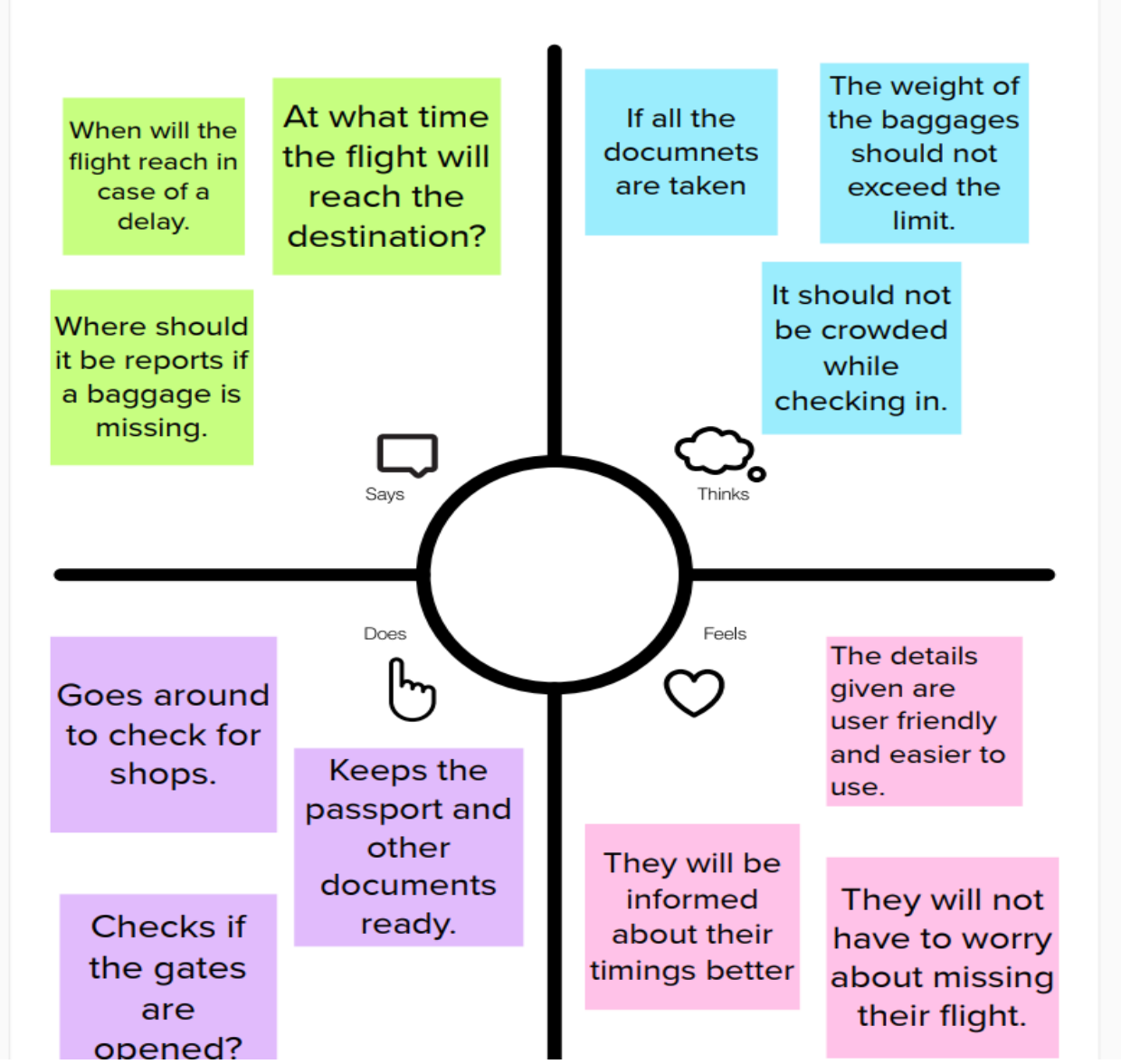
3. IDEATION & PROPOSED SOLUTION

Average aircraft delay is regularly referred to as an indication of airport capacity. Flight delay is a prevailing problem in this world. It's very tough to explain the reason for a delay. A few factors responsible for the flight delays like runway construction to excessive traffic are rare, but bad weather seems to be a common cause. Some flights are delayed because of the reactionary delays, due to the late arrival of the previous flight. It hurts airports, airlines, and affects a company's marketing strategies as companies rely on customer loyalty to support their frequent flying programs.

Nowadays, the aviation industry plays a crucial role in the world's transportation sector, and a lot of businesses rely on various airlines to connect them with other parts of the world. But, extreme weather conditions may directly affect the airline services by means of flight delays**.** 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. Data mining produces insights around the decisions for adding or subtracting the flights to the routes where more or lesser passenger movement is found. The purpose of this project is to look at the approaches used to build models for predicting flight delays that occur due to bad weather conditions. In this based on the customer review and other data the delay of the flight is calculated then comparing with other flight the best flight with shortest time delay will be delivered.

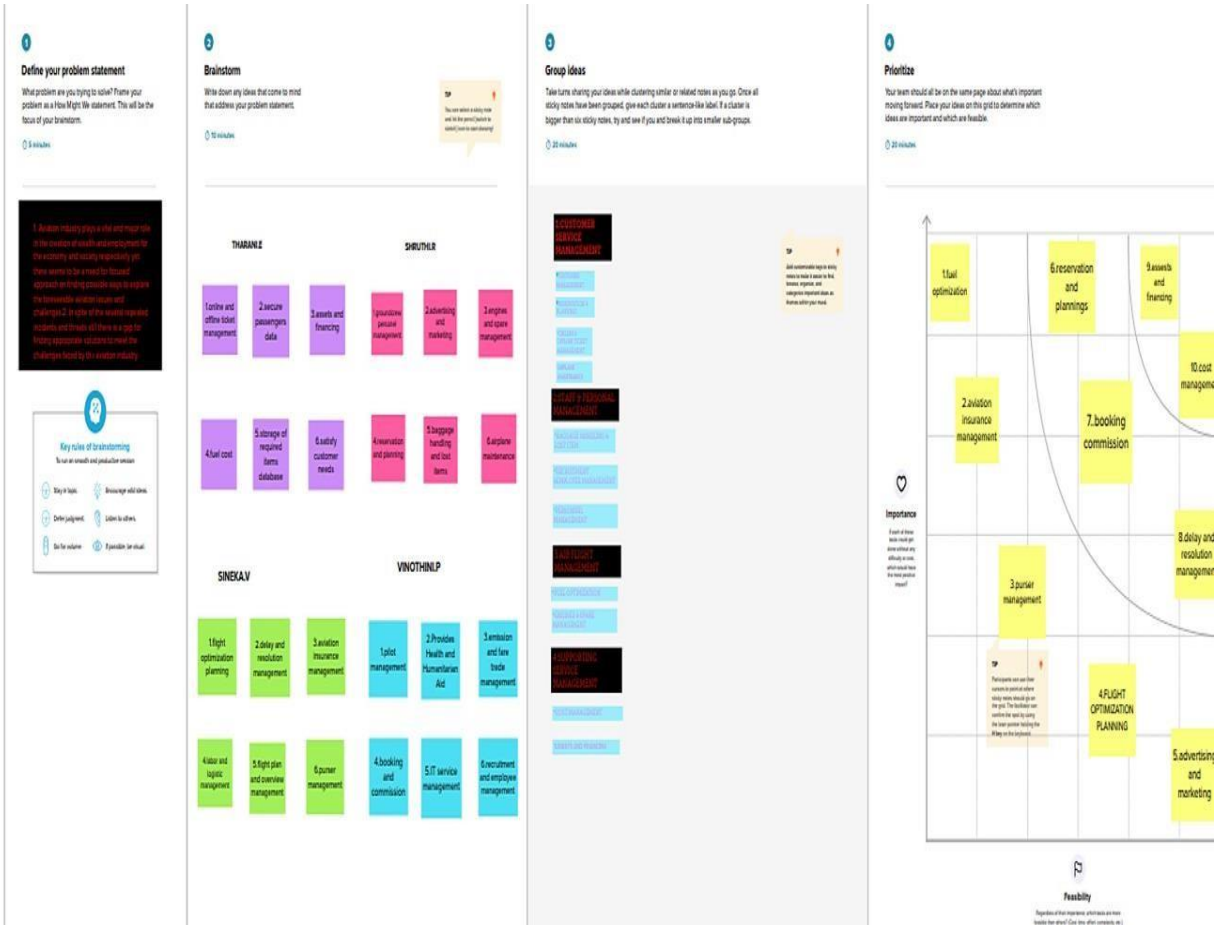
EMPATHY MAP CANVAS

An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment.



BRAIN STORM

A mind map is a diagram used to visually organize information into a hierarchy, showing relationships among pieces of the whole. It is often created around a single concept, drawn as an image in the center of a blank page, to which associated representations of ideas such as images, words and parts of words are added.



PROPOSED SOLUTION

DOMAIN :
DATA
ANALYTICS

- **PROBLEM STATEMENT**
- Infrastructure deficit: unavailability of land to expand airports at their current sites, particularly in major cities
- High Fuel Costs: Fuel cost as a percentage of operating charges amounts to 45% in India as compared to the global average of 30 per cent.
- Procedural Complexities: Taxes and approvals by both Central and States governments increases the transaction costs of the aviation sector
- Inadequate Human Resource Capabilities: Shortage and gaps in the availability of industry-recognised skills – from airline pilots and crew to maintenance and ground handling personnel – constrains the growth of different segments of the sector.

THE CHALLENGES

S.no	Parameter	Description
1.	Problem Statement (problem to be solved)	The airport codes may refer to either the IATA airport code, a three letter code that is used in passenger reservation, ticketing and baggage handling systems, or the ICAO airport code which is a four-letter code used by ATC systems and for airports that do not have an IATA airport code.
2.	Idea/Solution Description	Machine learning and analytics have touched almost all the fields around the globe including the aviation industry. With the growth of data, the use of analytics in the airline industry is the next big wave. The purpose of data analytics in aviation is to examine the vast amount of data generated daily and provide useful information to airlines,

		airports and other aviation stakeholders so that they can improve their operational planning and execution, as well as any related products and services. Airlines use AI systems with built-in machine learning algorithms to collect and analyze flight data regarding each route distance and altitudes, aircraft type and weight, weather, etc. Based on findings from data, systems estimate the optimal amount of fuel needed for a flight
3.	Novelty/Uniqueness	**1.Cost Reduction** Airlines are very concerned about baggage handling metrics like lost-bag tally, SLAs. They rely on real-time baggage tracking data to avoid losing damaging or delaying

		bags and face compliance issues. **2.Fuel Management-** Airlines track real-time fuel consumption data on Dashboards from take-off to landing. This monitoring is crucial to be ultra-efficient in reducing fuel costs and airline emissions. **3.Revenue Maximization-**Airlines segment customers, target with personalized offers, optimize pricing in real time using predictive analytics techniques such as modelling and forecasting.

4.	Social Impact/Customer Satisfaction	Trajectory Optimization • Predictive Maintenance • Delay Estimation • Targeted Advertising • Crew Performance Assessment • Sentiment Analysis • Prediction of Customer Behavior.

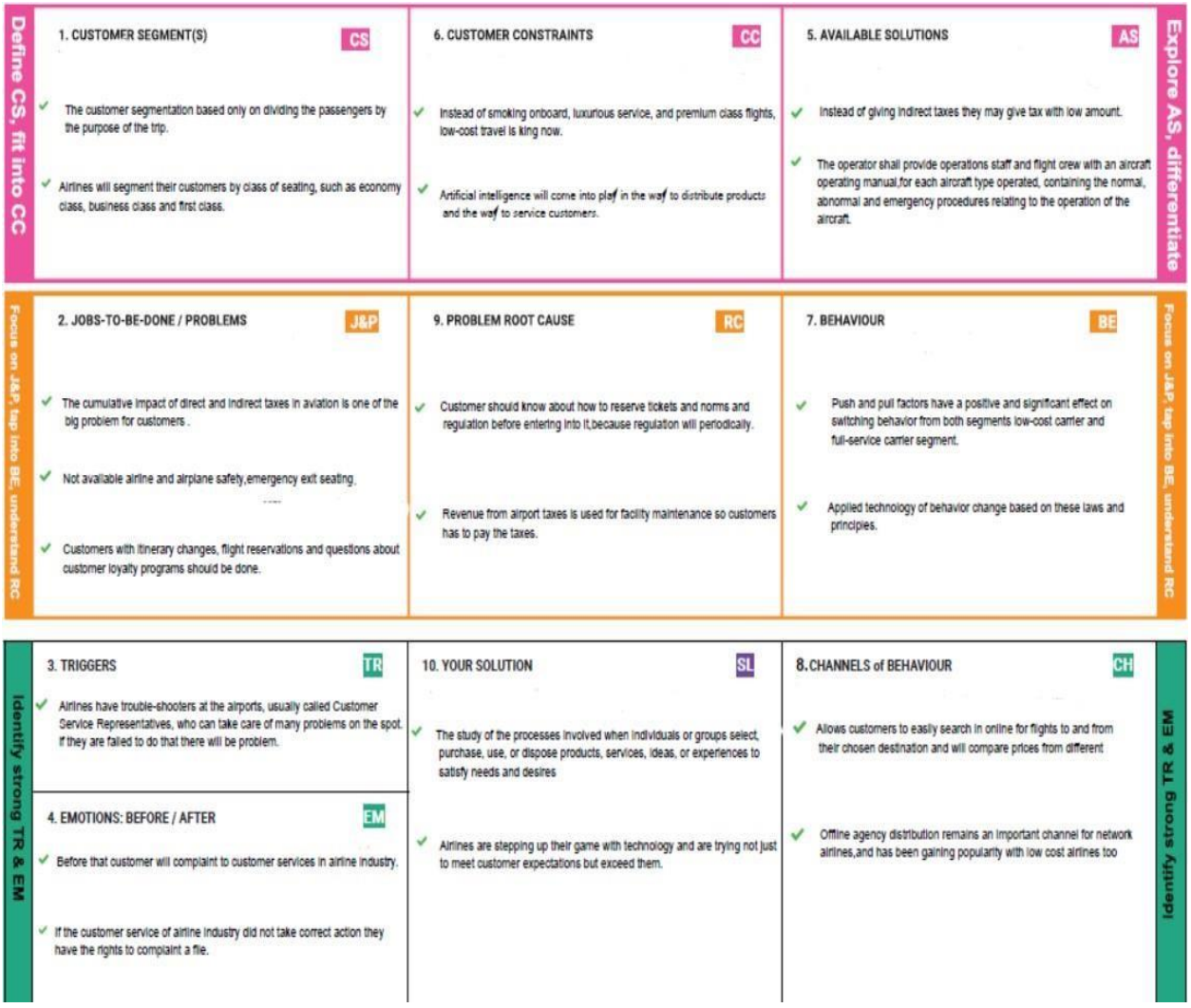
5.	Business Model (Revenue Model)	The 4 Most Important Business Models for Airlines 1. Full-Service Carriers. Full-service carriers are airlines that operate with a business model that includes offering a range of preflight and onboard services with the price of the ticket. 2. Low-Cost Carriers 3. Charter Airlines. 4. Cargo Airlines.

6.	Scalability of the solution	Data analytics has revolved around every industry, including aviation. Technology has changed how business is conducted and helps to make better decisions. As a result, data analytics plays a vital role in the aviation industry. It assists in collecting data and planning a powerful strategy that helps to grow business overall. According to a report, after adopting

		Big Data and Data Analytics in the airline industry, the sector has witnessed 57% more growth. From maintaining flights to unplanned maintenance, Data Analytics in the airline industry unfolds everything. Big Data tailors the flight experience better and uses data to improve performance. There are plenty of advantages, but most of all, it's how Data Analytics transforms the airline industry. It gains insights and enhances operations to make it successful. According to a report, Data Analytics in the airline industry is expected to reach $7 million by 2023.

PROBLEM SOLUTION FIT

Problem-solution fit is a term used to describe the point validating that the base problem resulting in a business idea really exists and the proposed solution actually solves that problem. Validate that the problem exists: When you validate your problem hypothesis using real-world data and feedback.



REQUIREMENT ANALYSIS

FUNCTIONAL REQUIREMENT

Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish.

Following are the functional requirements of the proposed solution.

FR No	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Gmail.
FR-2	User Confirmation	Confirmation via Email
FR-3	Search for flights	The registered user can search one way, round trip and multiple destination flights by choosing specific dates and destination.
FR-4	Specify passenger	Customer select the number of passengers and their category either adults, infant or child.
FR-5	Sorting flight	Customer will sort the flight either by price or duration of the flight and will register.

FR-6	Better airline service	Provide better airline service by analyzing time consuming, comfort of passenger.

PROJECT DESIGN

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.

SOLUTION AND TECHNICAL ARCHITECTURE

Technical Architecture (TA) is a form of IT architecture that is used to design computer systems. It involves the development of a technical blueprint with regard to the arrangement, interaction, and interdependence of all elements so that system-relevant requirements are met.

The Deliverable shall include the architectural diagram as below and the information as per the Table1 & Table2.

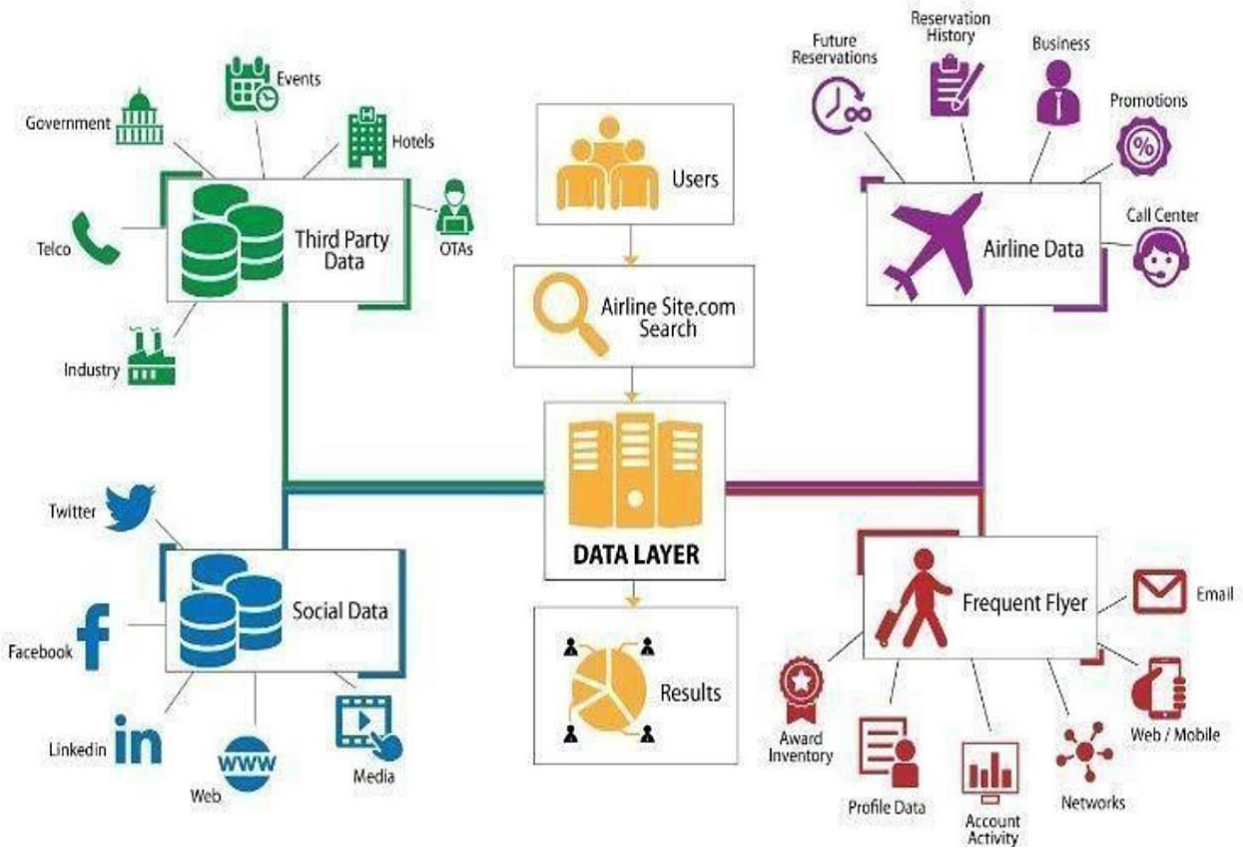


Table-1: Components & Technologies:

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

3.	Application Logic-2	Logic for a process in the application	IBM Watson Assistant
4.	Database	Data Type, Configurations	MySQL, NSQL
5.	Cloud Database	Database service on cloud	IBM DB2, IBM Cloud ant
6.	File Storage	File Storage requirements	IBM Blocks Storage or other storage service or Local File system
7.	External API-1	Purpose of External API used in the application	IBM Weather API
8.	External API-1	Purpose of External API used in the application	Aadhar API
9.	Infrastructure (Server/Cloud)	Application Deployment on 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 open source framework.

2.	Security Implementations	List all the security/access controls implemented, use of firewalls.	Example: SHA256, Encryption, IAM Controls, OWASP.
3.	Scalable Architecture	Justify the scalability of architecture.	Cognos Used.
4.	Availability	Justify the availability of application (e.g. Use of load balancers, distributed servers).	AWS Used.
5.	Availability	Design consideration for the performance of the application (number of requests per second, use of Cache, use of CDN's).	Dashboard, Report stories.

PROJECT PLANNING & SCHEDULING

Sprint Planning & Estimation

Product Backlog, Sprint Schedule, and Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members

Sprint-1	Retrieve the Data	USN-1	Retrieving the data from the passengers those who are traveling in flight and the data of flight	2	High	INDRAJEET PRASHANNAA
Sprint-1	Visualize the data	USN-2	After retrieving the data, we have to visualize the data for better understanding	1	High	INDRAJEET PRASHANNAA SRUTHI.R
Sprint-2	Track the flight timing and airline names	USN-3	Tracking the delays which are made by the flights and in other situations	2	High	PRASHANNAA RAHIL JAMAL
Sprint-2	Create interactive graph	USN-4	At each scenario, we have to create a graph for better visualization	2	High	PRASHANNAA RAHIL JAMAL
Sprint-3	Create dashboard	USN-5	Creating interactive dashboard with the given dataset and information	1	High	SRUTHI.R RAHIL JAMAL

Sprint-3	Creation of story	USN-6	Creating the story for each respective phase		High	SRUTHI. R RAHIL JAMAL
Sprint-4	Predict the delays	USN-7	Finally, this project delivers the airlines which made most of the delays in airport and flight	1	High	INDRAJEET SRUTHI.R PRASHANNAA

Project Tracker, Velocity & Burndown Chart:

Sprint	Total Story Point	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	18 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	18 Nov 2022

Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

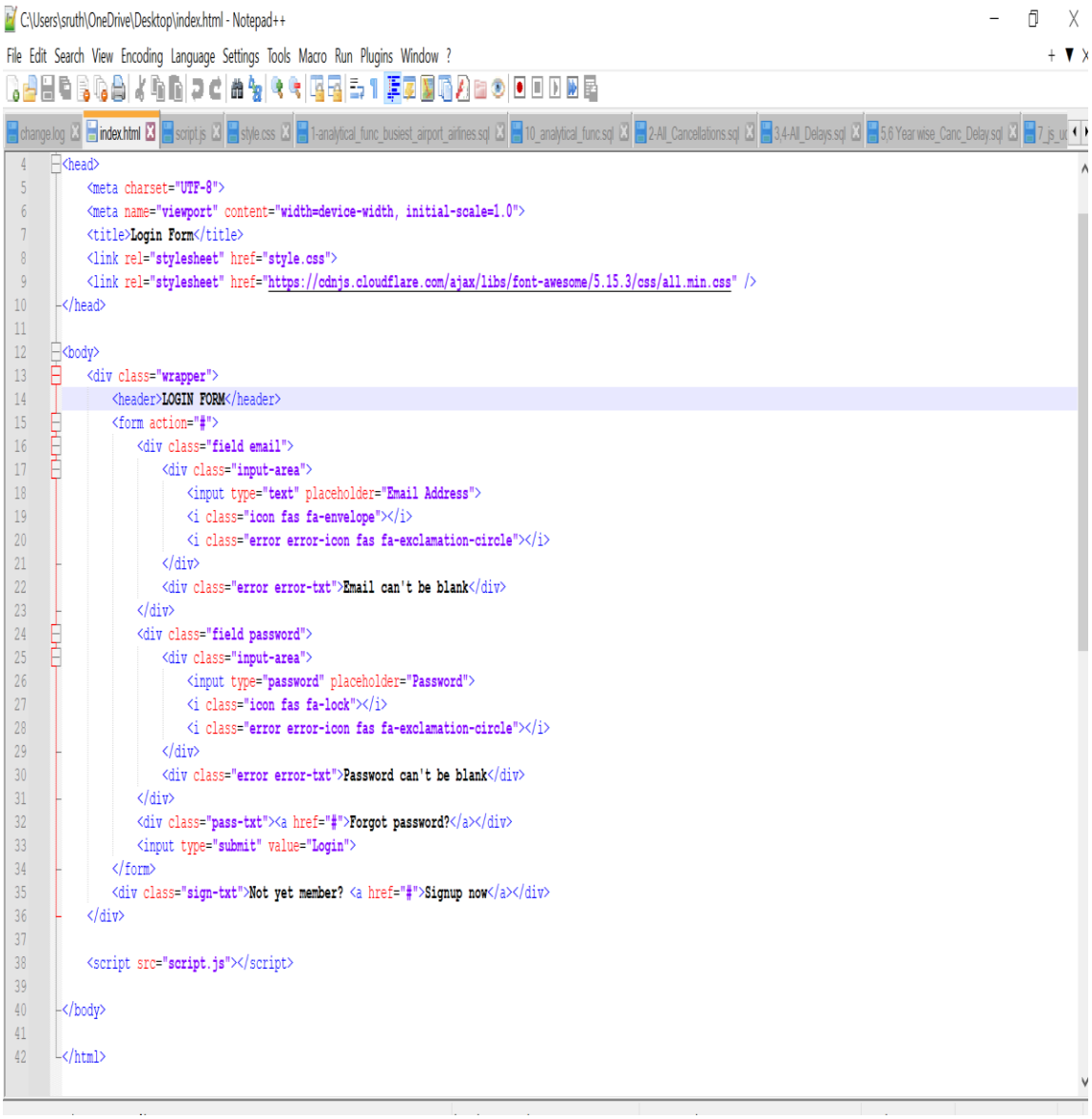
SPRINT DELIVERY SCHEDULE

MILESTONE&TASKS

Milestone	**Task**	**Duration**
Milestone-1	Collection of Data	October-24/10/22
Milestone-2	Uploading the required data on the platform	October-31/10/22
Milestone-3	Visualizing of data	October-31/10/22
Milestone-4	Creating a dashboard	November-2/11/22
Milestone-5	Display the data's in the dashboard	November-5/11/22
Milestone-6	Prepare a standardized data set and using the data's required with the help of python program	November-18/11/22
Milestone-7	Usage of various algorithm to obtain the desired result	November-18/11/22

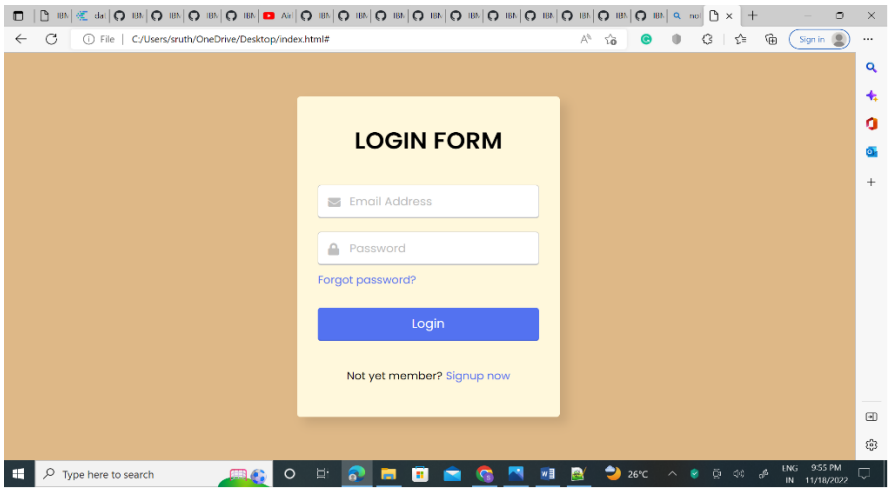
Milestone-8	Display them in the required format	November-18/11/22
Milestone-9	Deployed in the GitHub	November-18/11/22

7. CODES AND SOLUTIONING

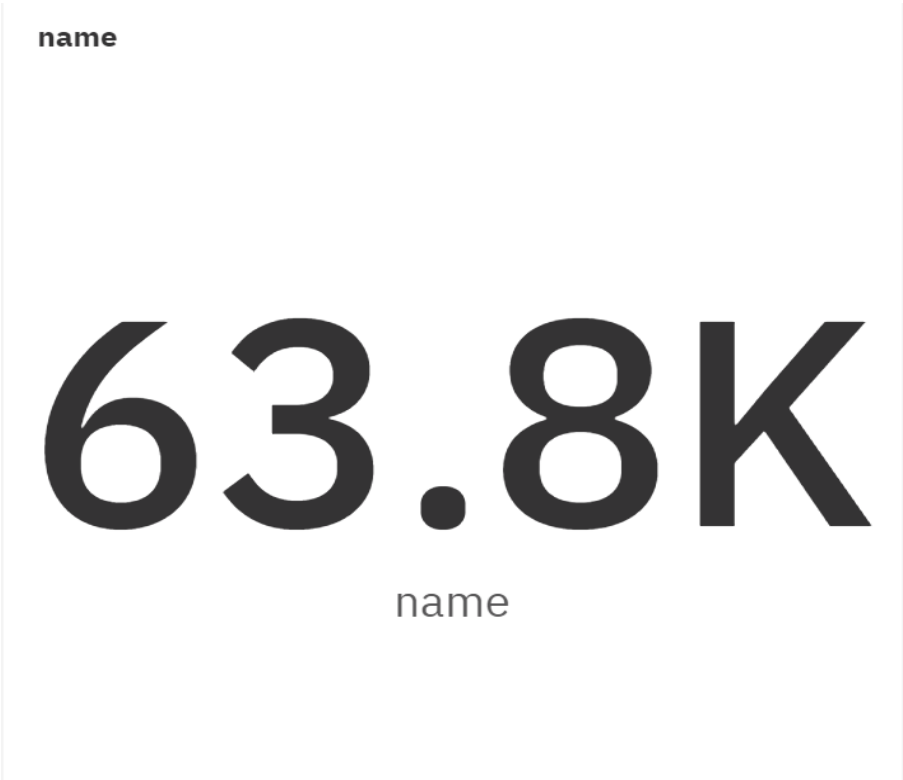


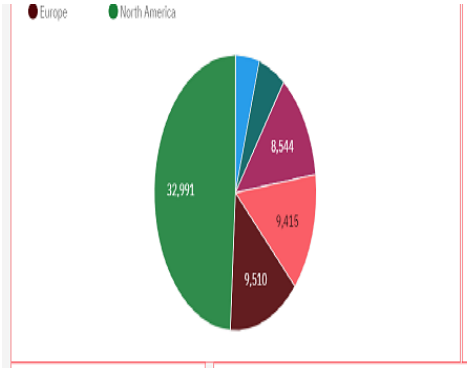
```
(pInput.value == "") ? pField.classList.add("shake", "error"): checkPass();

setTimeout(() => {
    eField.classList.remove("shake");
    pField.classList.remove("shake");
}, 500);

eInput.onkeyup = () => { checkEmail(); }
pInput.onkeyup = () => { checkPass(); }

function checkEmail() {
    let pattern = /^[^ ]+@[^ ]+\.[a-z]{2,3}$/;
    if (!eInput.value.match(pattern)) {
        eField.classList.add("error");
        eField.classList.remove("valid");
        let errorTxt = eField.querySelector(".error-txt");

        (eInput.value != "") ? errorTxt.innerText = "Enter a valid email address": errorTxt.innerText = "Email can't be blank";
    } else {
        eField.classList.remove("error");
        eField.classList.add("valid");
    }
}

function checkPass() {
    if (pInput.value == "") {
        pField.classList.add("error");
        pField.classList.remove("valid");
    } else {
        pField.classList.remove("error");
        pField.classList.add("valid");
    }
}

if (!eField.classList.contains("error") && !pField.classList.contains("error")) {
    window.location.href = form.getAttribute("action");
}
}
```

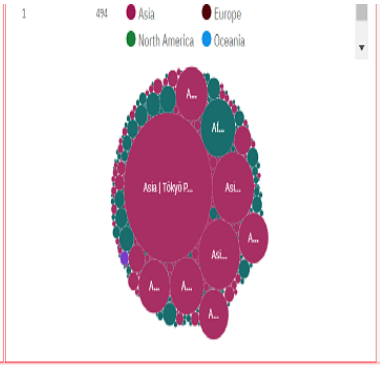
cript file | length : 1,646 lines : 48 | Ln : 48 Col : 2 Pos : 1,647 | Windows (CR LF) | UTF-8 | INS



8. RESULTS

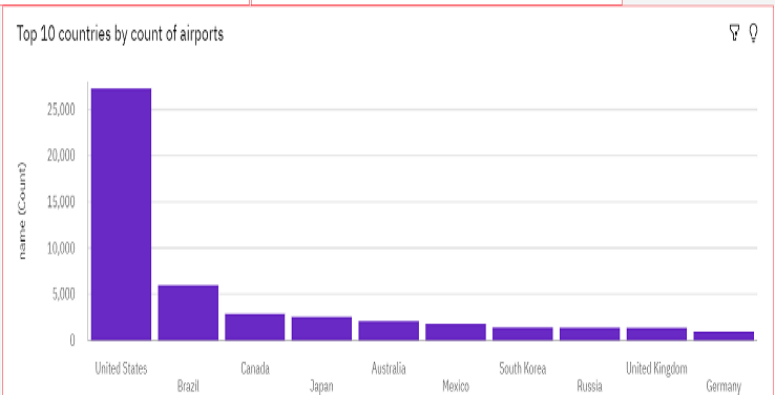


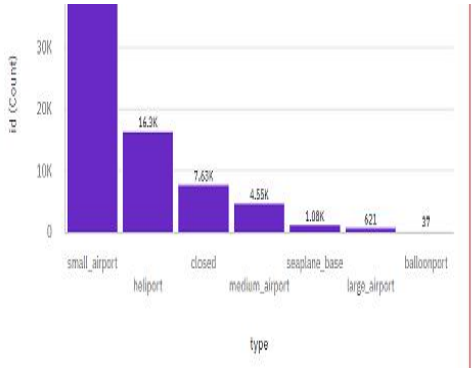


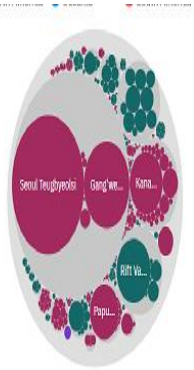


Continent Name

- Africa
- Antartica
- Asia
- Europe
- North America
- Oceania
- South America



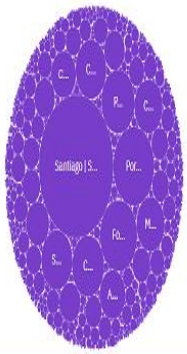




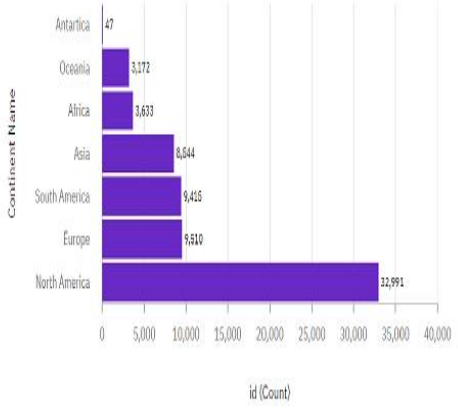
municipality and Continent Name sized by id

id (Count distinct)

1 88



id by Continent Name



PERFORMANCE MATRICES

There are various metrics to calculate the efficiency of the data models itself. Performance of a data model developed by data scientists is a direct way to measure their efficiency. Methods include confusion matrix, F1 score, Precision Recall Curve, Receiver Operating Characteristics, among others. The idea is to see if the performance is better than the baseline models. It is important to consider that a model takes time to improve and that models are not foolproof.

In this project with the help of the data analytics the flight which covers the destination in short time when comparing to the another flight is calculated easily with more accuracy. The accuracy rate is more high by using a data analytics.

8. ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

- The advantages include being able to fly to almost any destination in the world and having a variety of different aircraft for different purposes, and cut down on travel time.
- High Speed - It makes this model an optimum choice if the client has an urgent need to ship a product. It is the quickest transport mode and is therefore ideal for long-distance transport of goods.
- There is less need for heavy packaging - Air exports, in general, entail less hard packaging than ocean shipments. This ensures you save both time and money by not having to provide extra packaging services.
- Fast Service - Air transportation offers convenient, reliable and fast services of transport. It is considered the cheapest way to ship peregrinated goods.

It offers a standard, convenient, reliable and fast service.

- Natural Route - An aircraft can fly to any location without seeing any natural obstacles or barriers. Since customs formalities are easily compiled. It eliminates the need for more time to seek clearance. Air travel is used for relief operations during earthquakes, floods, accidents, and famines.

DISADVANTAGES:

- Risky - Air travel is the riskiest mode of transport, since there can be considerable losses to goods, customer and crews as a result of a minor crash. Compared to other means of travel, the risks of collisions are higher.
- Cost - Air travel is considered to be the most expensive means of transportation. The cost of maintaining aircraft is higher and the costs for the building of aerodromes and avian are much higher. That's why air travel is so expensive that it gets beyond ordinary people's grasp.
- Capacity for Small Carriage - The aircraft have no room and therefore are not ideal for carriage of voluminous and cheaper materials. As is seen for rails, the load volume cannot be raised.
- Accident-prone - Compared to other modes air travel is always at high risk of accidents. There are more accidents on count while travelling by air transport. The reason can be bad weather, signal issues or machine parts failure which causes loss of people.

9. CONCLUSION

Customer experience is always at the top of the priority list for airlines. Customers that are dissatisfied or disengaged inevitably result in fewer passengers and less money. It is critical that clients have a positive experience every time they travel. Looking at the bright prospects of the aviation industry, it makes sense to invest in airline stocks as they are likely to benefit from the government's push to make the aviation industry a bulwark of the transportation industry in India.

From this project we conclude that, the usage of big data analytics is booming today, with its ability to be used to draw useful insights from past data research. Its uses in the aviation industry have a wide array of applications ranging from predicting flight delays to detecting faults in airplane parts. In this paper, we conducted exploratory data analysis on flight dataset to draw inferences on arrival and departure delays and to identify relationships between flight timings and delays. Using the flight delay data, we identified which flight is mostly prone to delays. The arrived upon conclusions are useful for selecting flights in the future. From the review of the customer and the flight which covers the destination in correct time and in shortest time that airline flight will be selected as a best airline service.

10. FUTURE SCOPE

With the growth of data, the use of analytics in the airline industry is the next big wave. 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.
In future this project has been developed with some extra features. The customer can give query for any dissatisfaction that query will be solved review of the

customer will be collected. Then if a customer wants to change the destination in a midway they can give one alert message to the service and that nearby destination will be given for the customer.

11. APPENDIX

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-9469-1659009988.git

PROJECT DEMO LINK:

https://youtu.be/2NaHq8rIBSU