AIRLINES DATA ANALYTICS FOR AVIATION INDUSTRY

Team-ID : PNT2022TMID49991

Team Members: Benilton Azwalt B, Jenish Immauel Raj J, Leeban Raj J, Thanga Selvakumar D

CHAPTER 1

INTRODUCTION

1.1 PROJECT OVERVIEW

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.For that purpose we use the method of data analytics and internet of things to store the data which is used for the needed. Data set of various countries , Number of flights based on continent , weather , region ,Air station , Number of airports that are suitable for the running of the flights are being collected in the form of the data sets and penned down in charts . In IBM Cognos cloud area the collected information is stored for further process. On following the collection of the data set those processed information are dropped in the spreadsheets.Using the given dataset, we plan to create various graphs and charts to highlight the insights and visualizations. On doing this a perfect information of the flights are obtained and many overhead delay of the arrival , departure of the planes can be avoided.

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.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING PROBLEM

Air travel has been increasingly preferred among travelers, mainly because of its speed and in some cases comfort. This has led to phenomenal growth in air traffic and on the ground. • An increase in air traffic growth has also resulted in massive levels of aircraft delays on the ground and in the air. These delays are responsible for large economic losses. • It's important to provide better Airline and AirPort services and avoid delays in Air Travel across different locations and promise to get passengers from Location A to Location B on time.

2.2 REFERENCES

Akerkar, R. A. (2013) Big Data Computing. CRC Press.

Akerkar, R. A.; Lingras, P. (2008). An Intelligent Web: Theory and Practice, 1st edn. Johns and Bartlett, Boston.

Armour, F. (2012) Introduction to big data, presentation at the symposium Big Data and Business Analytics: Defining a Framework, Center for IT and Global Economy, Kogod School of Business, American University, Washington, DC.

Bell, G., Hey, T., Szalay, A. (2009) Beyond the data deluge, Science 323 (5919), 1297–1298.

Boland, D., Morrison, D., and O'Neill, S. The future of CRM in the airline industry: A new paradigm for customer management. IBM Institute for Business Value. http://www05.ibm.com/innovation/nl/pdf/highlights/integration/crm_airline.pdf

Diebold, F.X. (2000). Big data dynamic factor models for macroeconomic measurement and forecasting. Discussion read to the 8th World Congress of the Econometric Society, Seattle, August. http://www.upenn.edu/~fdiebold/papers107/ABCD_HOED.pdf.

Dumbill, E. (2012) Making sense of big data. DOI: 10.1089/big.2012.1503.

Gualtieri, M., The pragmatic definition of big data. Forrester Research Blog. http://blogs.forrester.com/mike_gualtieri/12-12-05-the_pragmatic_definition_of_big_data John-Paul Clarke. (2013) Enabling Pro-Active Decisions via Predictive Analytics. Air Transport IT Summit.

L. Neumeyer, B. Robbins, A. Nair, and A. Kesari. (2010) S4: Distributed Stream Computing Platform, in Proceedings of IEEE International Conference on Data Mining Workshops (ICDMW), pp. 170–177.

Laney, D. (2001) 3d Data management: controlling data volume, velocity and variety, Apply. Delivery Strategies Meta Group (949). http://refhub.elsevier.com/S0020-0255(14)00034- 6/h0650

Lynch, C. (2008) Big data: how do your data grow?. Nature 455 (7209), 28–29. Mark van Rijmenam. (2014) Southwest Airlines Uses Big Data To Deliver Excellent Customer Service. Datafloq blog. https://datafloq.com/read/southwest-airlines-uses-big-data-deliverexcellent/371

McKinsey Global Institute. (May 2011) Big data: The next frontier for innovation, competition, and productivity.

Nathan Marz and James Warren. (2012) Big Data- Principles and best practices of scalable real-time data systems. Manning publishing

Savitz, E. (2012) Gartner: 10 Critical Tech Trends for the Next Five Years. http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-thenext-five-years/

Savitz, E. (2012) Gartner: Top 10 Strategic Technology Trends for 2013. http://www.forbes.com/sites/ericsavitz/2012/10/23/gartner-top-10-strategic-technologytrends-for-2013/.

Smith, J.A. (2012) Field Note: What Makes Big Data Big – Some Mathematics Behind Its Quantification, Data Scientist Insights. http://datascientistinsights.com/2012/12/19/field-notewhat-makes-big-data-big-some-mathematics-behind-its-quantification/ Szalay, A., Gray, J. (2006) Science in an exponential world, Nature 440, 23–24.

TechAmerica Foundation. (2012) Demystifying Big Data. Washington, DC. Zikopoulos, P. Chris, E. (2011) Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Professional.

2.3 PROBLEM STATEMENT DEFINITION

Customer ProblemStatement Template: Create a problem statement to understand your customer's point of view. The Customer Problem Statement template helps you focus on what matters to create experiences people will love.

A well-articulated customer problem statement allows you and your team to find the ideal solution for the challenges your customers face. Throughout the process,you'll also be able to empathize with your customers, which helps you better understand how they perceive your product or service.

Airlines Data Analytics for Aviation Industry:

Problem Statement (PS)	**I am (Customer)**	**I'm trying to**	**But**	**Because**	**Which makes me feel**
PS-1	Passenger who expects to	Find best solutions and different	Facing issues in searching	Details and information provided in	To enjoy my journey in the safest way

	be comfortable during my airtime travel	ways to meet my expectation	for some unique resources	public platform was not confidential	

CHAPTER 3

IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

THINK AND FEEL

1. Have to take precautions
2. Feeling uncomfortable
3. Stressful

HEAR

1. Inner atmosphere hygienic level is normal.
2. Customer service is average.
3. Food and beverage quality is medium.

SAY AND DO

1. Its causes economic loses.
2. Hard to prepare for the next flight.
3. Poor work on reducing flight delays.

SEE

1. Passengers are dissatisfied.
2. Our Flights are not on time.
3. Decrement in passengers quantity.

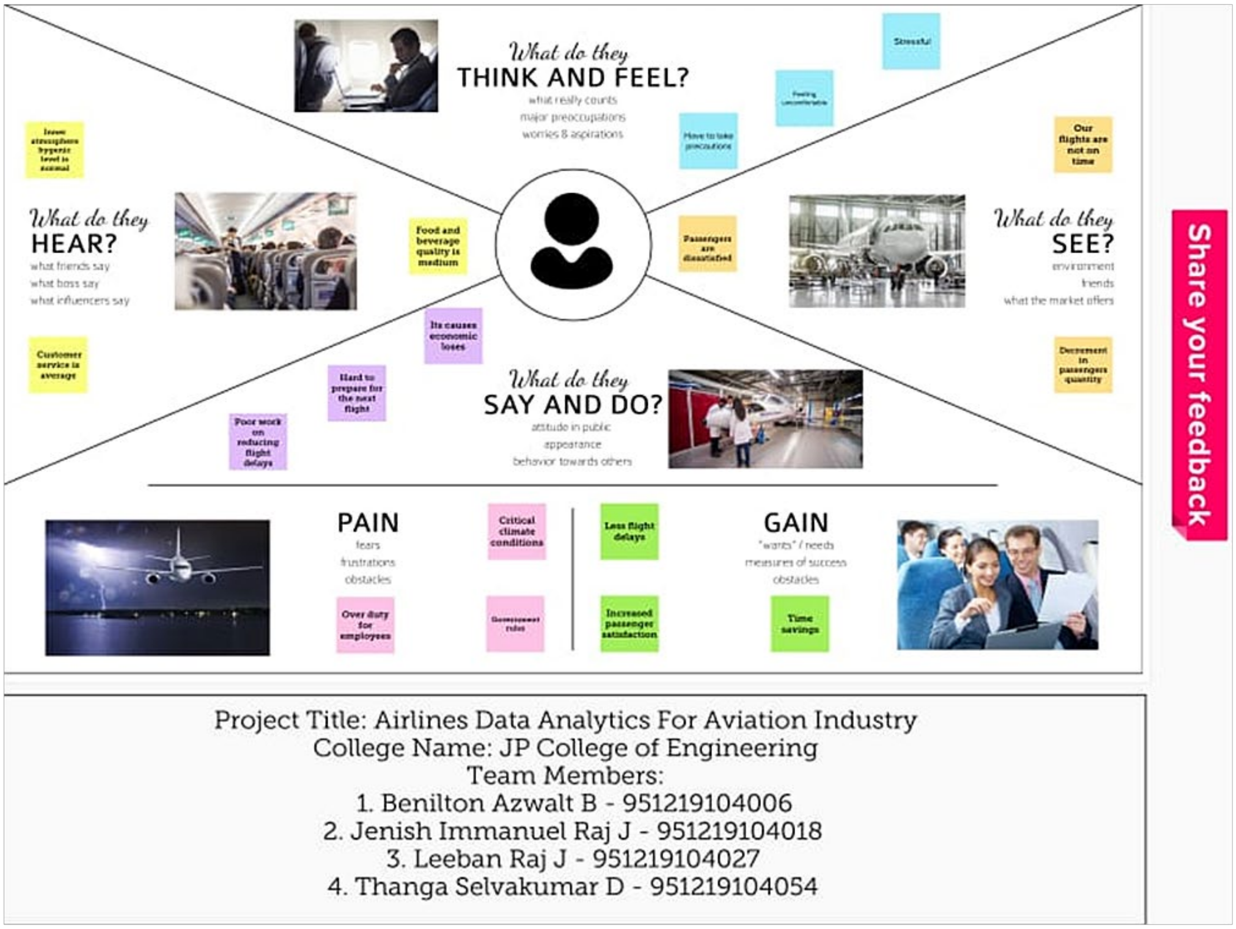
GAIN

1. Less flights.

2. Increased passenger satisfaction.
3. Time savings

PAIN

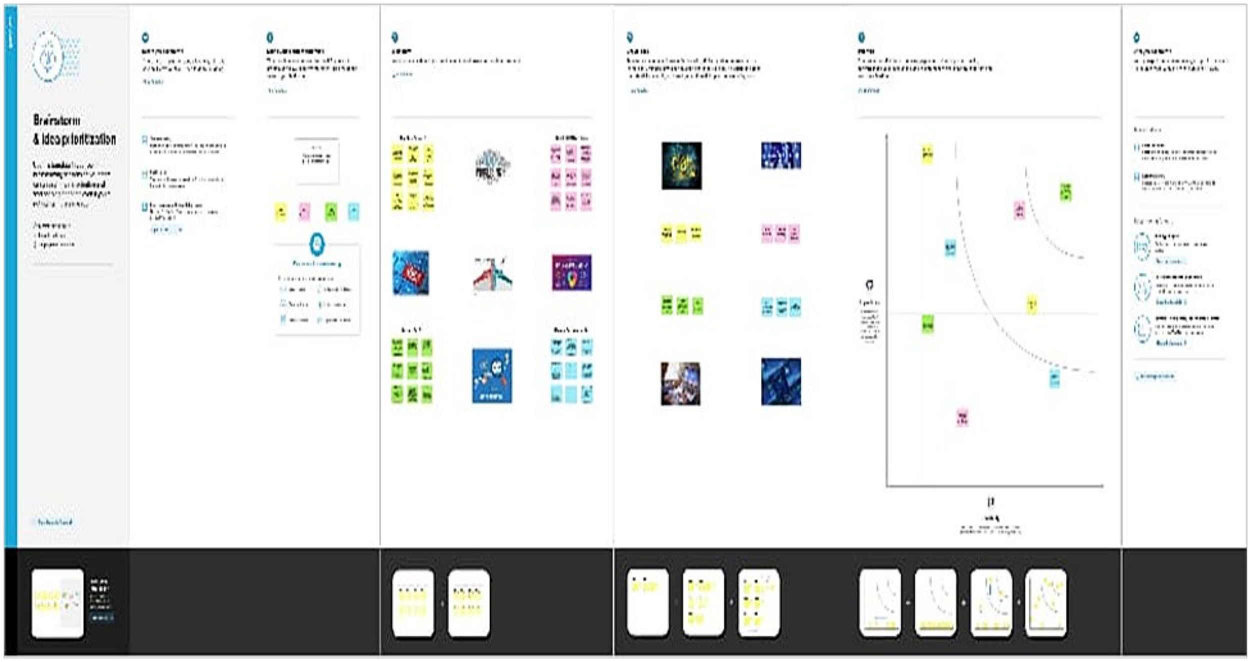
1. Critical climate conditions.
2. Government rules.
3. Over Duty for employees.



Project Title: Airlines Data Analytics For Aviation Industry
College Name: JP College of Engineering
Team Members:
1. Benilton Azwalt B - 951219104006
2. Jenish Immanuel Raj J - 951219104018
3. Leeban Raj J - 951219104027
4. Thanga Selvakumar D - 951219104054

3.2 IDEATION & BRAINSTORMING

1. Data collected from complaints.
2. Create high-quality data for the best business intelligence.
3. Translating data and information in a visual context.
4. Irrelevant information discovered and removed.
5. Hierarchical methods for information clusters.
6. Collecting passengers feedback.
7. Valuable insights to extract.



3.3 PROPOSED SOLUTION

S.No.	Parameter	Description

1.	ProblemStatement(Problemtobesolved)	To provide better Airline and AirPort servicesandtoavoiddelaysinAirTravelacrossdifferentlocations at M aim is to provide airports, airlines, and the travelling public with a view of whichairlinesaredelivering ontheirpromisetoget passengersfromPoint Ato Point Bon-time.
2.	Idea/Solutiondescription	Able to create meaningful Visualizations andDashboardofAirlinesdatafromtheCSVDataset files.
3.	Novelty/Uniqueness	Uniquevisualizationofdata fromdatasetsinsimpleandeasytounderst
4.	SocialImpact/CustomerSatisfaction	Reducesflightdelays.Improvespassengersatisfaction.
5.	BusinessModel(RevenueModel)	Wecanprovidetheapplicationinasubscriptionbased model.
6.	ScalabilityoftheSolution	Size and number of data can be large and notlimitedtosmallscale.Visualizationofdatacanbedoneinsimplegrap t files.

3.4 PROBLEM SOLUTION FIT

CUSTOMER SEGMENTS

- Who is your customer?

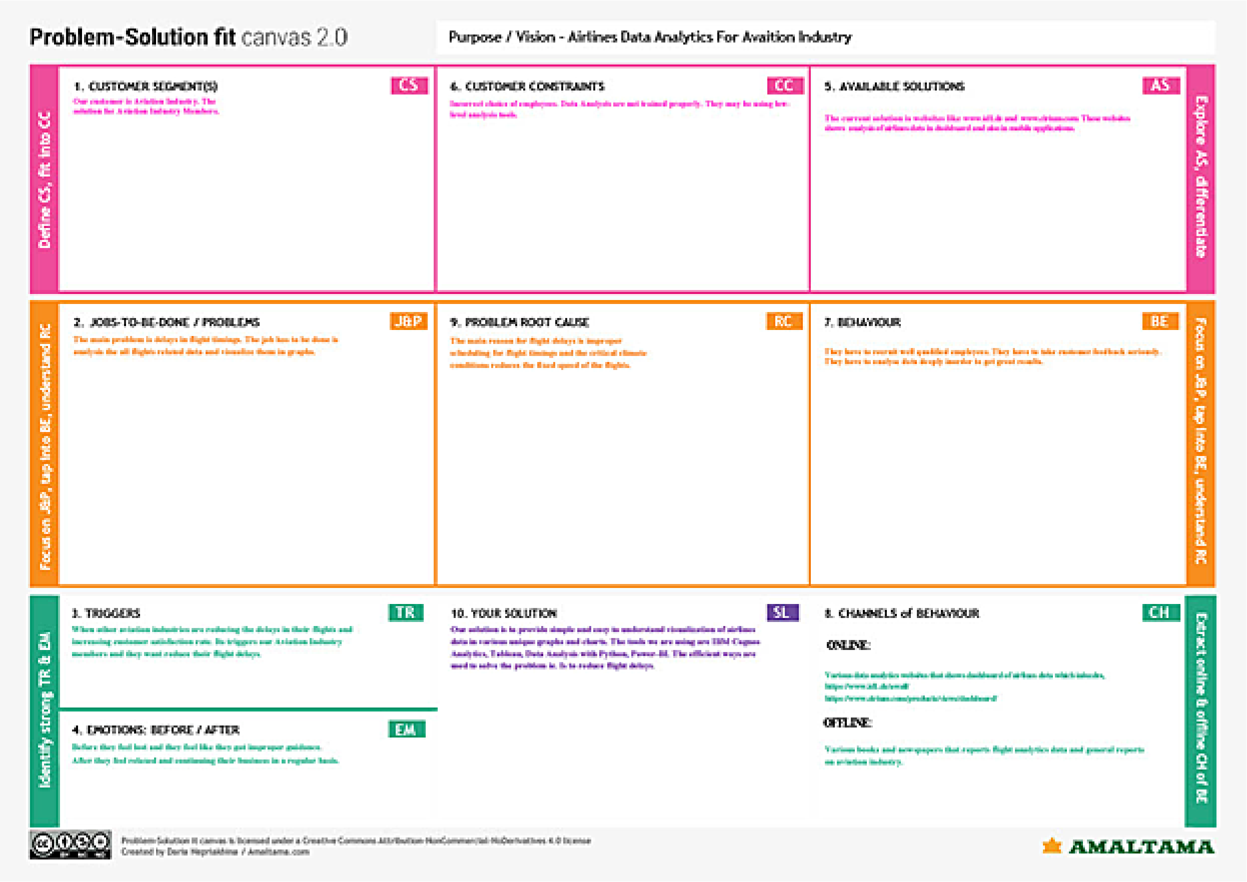
PROBLEM ROOT CAUSE

- What is the real reason that this problem exists?
- What is the back story behind the need to do this job?

YOUR SOLUTION

- 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?



Problem-Solution fit canvas 2.0

Purpose / Vision - Airlines Data Analytics For Avaition Industry

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

Our customer is Aviation Industry. The solution for Aviation Industry Members.

6. CUSTOMER CONSTRAINTS CC

Incorrect choice of employees. Data Analysts are not trained properly. They may be using low-level analysis tools.

5. AVAILABLE SOLUTIONS AS

The current solution is websites like www.[illegible].de and www.cirium.com These websites shows analysis of airlines data in dashboard and also in mobile applications.

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

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

The main problem is delays in flight timings. The job has to be done is analysis the all flights related data and visualize them in graphs.

9. PROBLEM ROOT CAUSE RC

The main reason for flight delays is improper scheduling for flight timings and the critical climatic conditions reduces the fixed speed of the flights.

7. BEHAVIOUR BE

They have to recruit well qualified employees. They have to take customer feedback seriously. They have to analyse data deeply inorder to get great results.

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS TR

When other aviation industries are reducing the delays in their flights and increasing customer satisfaction rate. Its triggers our Aviation Industry members and they want reduce their flight delays.

4. EMOTIONS: BEFORE / AFTER EM

Before they feel lost and they feel like they got improper guidance.
After they feel relaxed and continuing their business in a regular basis.

10. YOUR SOLUTION SL

Our solution is to provide simple and easy to understand visualization of airlines data in various unique graphs and charts. The tools we are using are IBM Cognos Analytics, Tableau, Data Analysis with Python, Power-BI. The efficient ways are used to solve the problem ie. Is to reduce flight delays.

8. CHANNELS of BEHAVIOUR CH

ONLINE:

Various data analytics websites that shows dashboard of airlines data which inlcude,
https://www.[illegible].de/[illegible]/
https://www.cirium.com/products/views/dashboard/

OFFLINE:

Various books and newspapers that reports flight analytics data and general reports on aviation industry.

Extract online & offline CH of BE

Problem-Solution fit canvas is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 license
Created by Daria Nepriakhina / Amaltama.com

AMALTAMA

CHAPTER 4

REQUIREMENT ANALYSIS

4.1 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	Registration through email.
FR-2	User Confirmation	Confirmation via email.
FR-3	User Interaction	Viewing and Interacting with the dashboard that visualize the flights details in graphs and charts.
FR-5	Report Generation	Generating flight reports regarding flight delays and other flight details.
FR-6	User Feedback	Getting feedback and customer satisfaction through survey.

4.2 NON-FUNCTIONAL REQUIREMENTS:

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

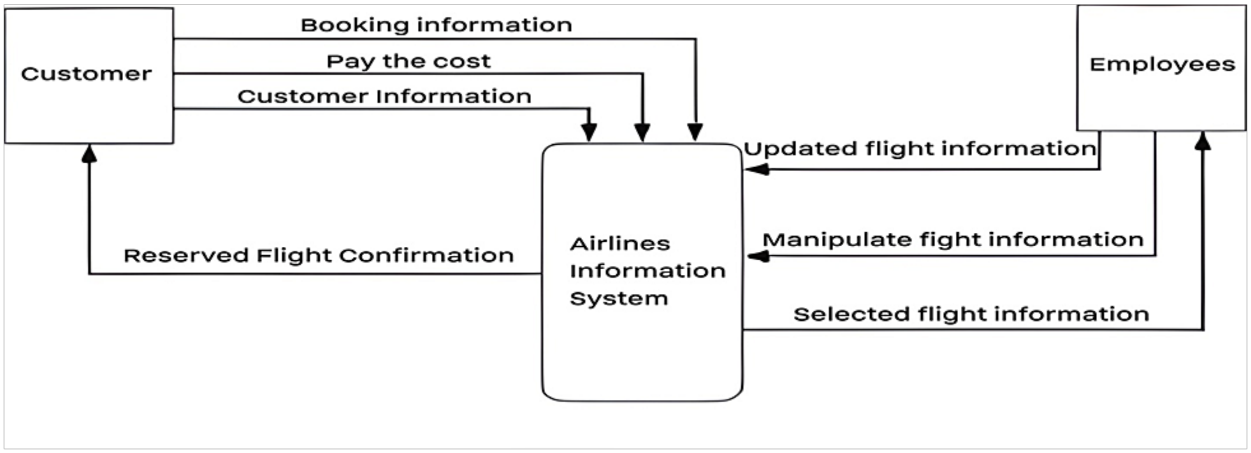
FR No.	**Non-Functional Requirement**	**Description**
NFR-1	**Usability**	The applications graphical user interface is user friendly and easy to use. Viewing and Interacting with the dashboard is very easy. All the features are easily accessible.
NFR-2	**Security**	All the user data from the application is stored securely. User’s privacy and security are maintained strictly.
NFR-3	**Reliability**	When errors happen in the server or client the application is reliable. The user will not face any critical issues with errors.
NFR-4	**Performance**	The application will run very efficiently on the less memory and less performing system.
NFR-5	**Availability**	The application is available 24x7 so user can access it anytime.
NFR-6	**Scalability**	The application is large so large amount of user can access it at a time.

CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

A Data flow Diagrams(DFD) is a traditional visual representation of the information flows within a system .A neat and clear DFDcan depict the right amount of the system requirements graphically . it shows how data enters and leave the system, what changes the information, and where data is stored.


Customer
Booking information
Pay the cost
Customer Information
Employees
Updated flight information
Airlines Information System
Reserved Flight Confirmation
Manipulate fight information
Selected flight information

5.3 USER STORIES:

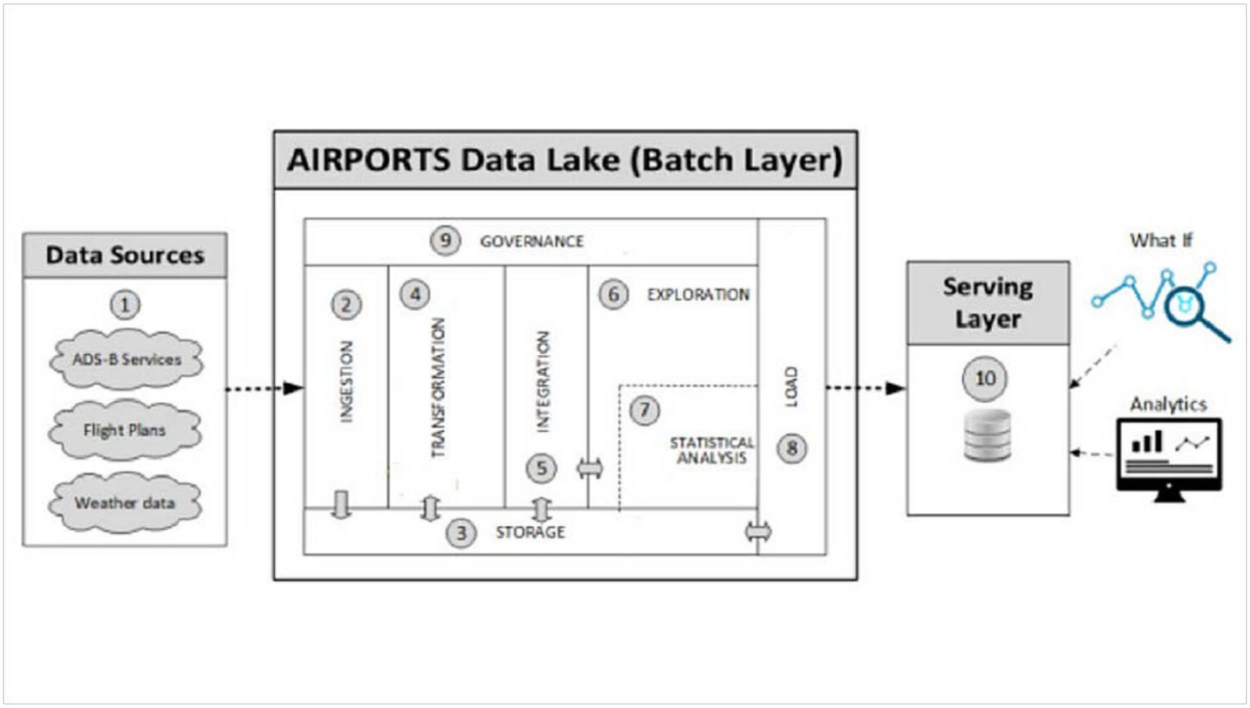
Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Gmail	I can register & access the dashboard with Gmail Login	Medium	Sprint-1

	Login	USN-4	As a user, I can log into the application by entering email & password	I can register & access the dashboard with email & password	High	Sprint-1
	Dashboard	USN-5	User can customize the dashboard visualization with filters and interactive actions.	I can get complete details regarding all the queries with the flight.	Low	Sprint
Customer Care Executive	Organization	USN-6	User can contact the airline industry if they had any issues or any unconfirmed information.	The customer care team will solve these issues.	High	Sprint
Administrator	Administration	USN-7	Uses must read and agree to the administrative policies regarding the analytics process and application or websites privacy and security policies.	The administration checks the agreement and verify my registration.	High	Sprint

5.2 SOLUTION & TECHNICAL ARCHITECTURE

SOLUTION ARCHITECTURE:



TECHNICAL ARCHITECTURE:

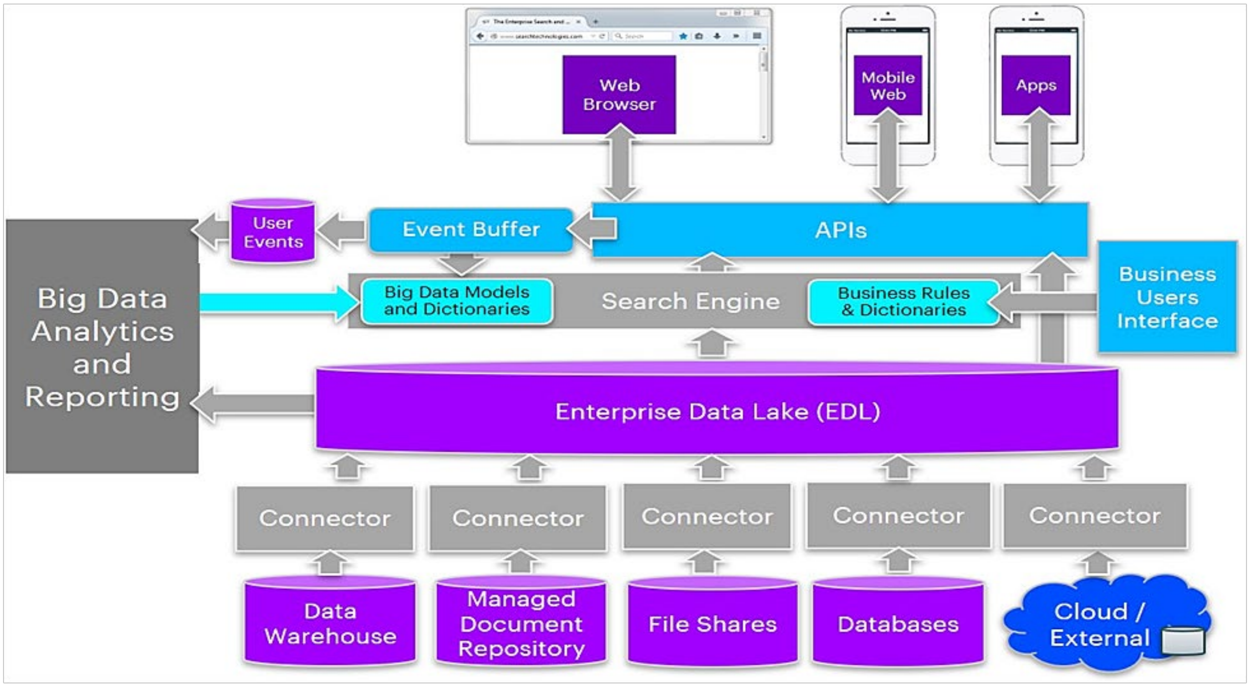


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI.	HTML, CSS, JavaScript / Angular.js / Node.js/ React.js
2.	Application Logic-1	Logic for a process in the application	Python
3.	Application Logic-2	Logic for a process in the application	IBM Watson STT service
4.	Application Logic-3	Logic for a process in the application	IBM Watson Assistant
5.	Database	Data Type, Configurations etc.	MySQL, NoSQL.
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloudant.
7.	File Storage	File storage requirements	IBM Block Storage.
8.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :	Local, Cloud Foundry, Kubernetes

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Python	Technology of Open source framework
2.	Security Implementations	Security / access controls implemented, use of firewalls.	SHA-256, Encryption, IAM Controls, OWASP
3.	Scalable Architecture	The application is large so large amount of user can access it at a time.	IBM Cognos, IBM Cloud
4.	Availability	The application is available 24x7 so user can access it anytime.	Load balancers, Distributed servers
5.	Performance	The application will run very efficiently on the less memory and less performing system.	Cache memory, CDN

CHAPTER 6

PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	5	High	Benilton Azwalt B
Sprint-1	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	8	High	Jenish Immanel Raj J
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password	7	Low	Leeban Raj J
Sprint-2	Data Collection	USN-4	As a developer, I can collect the dataset and analyze it before visualization	8	Medium	Thanga Selvakumar D
Sprint-2	Data Exploration	USN-5	As a developer, I can explore the uploaded dataset through IBM Cognos	12	High	Jenish Immanuel Raj J
Sprint-3	Data Visualization	USN-6	As a developer, I can visualize the uploaded dataset into dashboard	6	Medium	Leeban Raj J
Sprint-3	Dashboard	USN-7	As a user, I can personalize the dashboard with my desired choice.	6	High	Benilton Azwalt B
Sprint-3	Dashboard	USN-8	As a user, I can interact with the dashboard	8	High	Thanga Selvakumar D
Sprint-4	Report Generation	USN-9	As a user, I can view the reports of my visualization	10	Medium	Jenish Immanel Raj J
Sprint-4	Dashboard Establishment	USN-10	As a developer, I can establish the dashboard into a website.	10	High	Benilton Azwalt B

Sprint	Total Story Points	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	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

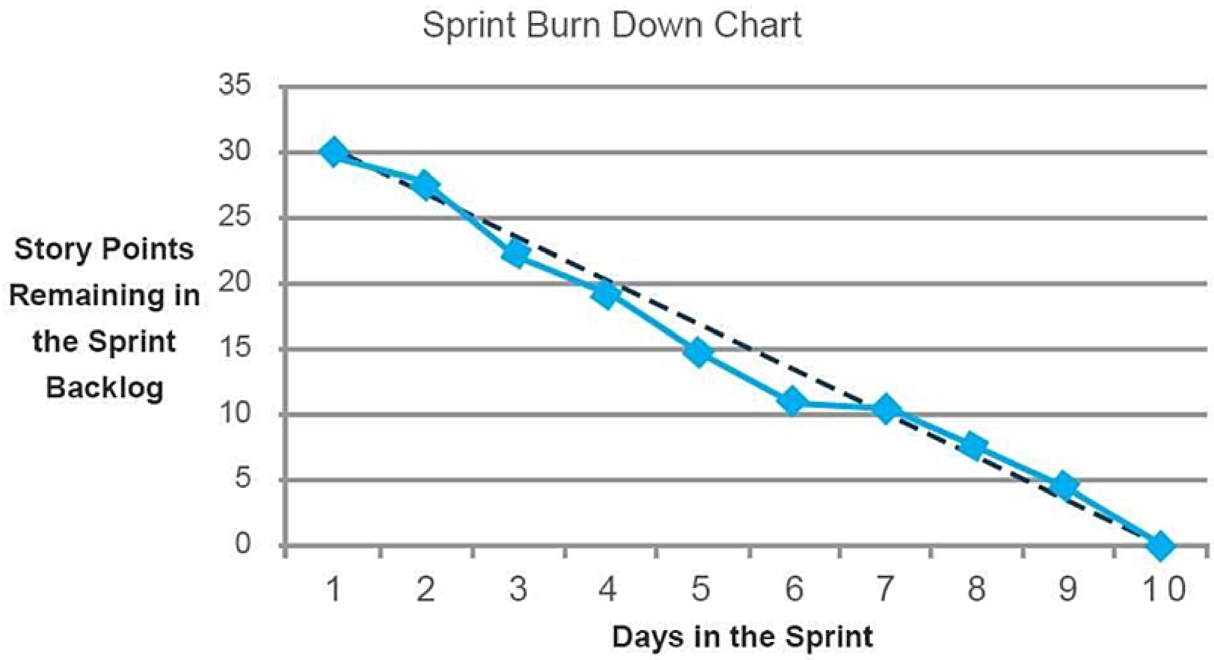
Velocity:

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

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

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 any project containing measurable progress over time.



6.2 SPRINT DELIVERY SCHEDULE:

TITLE	DESCRIPTION	DATE
Ideation Phase: Literature Survey On The Selected On The Selected Project & Information	Collect the relevant information on project use case, refer the existing solutions, technical papers, research publications etc.	3 September 2022
Prepare Empathy Map	Submit the Empathy Map Canvas and List of problem statements as per the template in GitHub	10 September 2022
Ideation	Participate in Brainstorming & Ideation, list the ideas and shortlist the top 3 ideas as per the template in GitHub	17 September 2022
Project Design Phase – I: Proposed Solution	Submit the proposed solution in the prescribed template in GitHub	24 September 2022
Problem Solution Fit	Submit the Problem-Solution fit Template in GitHub	01 October 2022
Solution Architecture	Submit the Solution Architecture in GitHub	01 October 2022
Project Design Phase – I:	Submit the Customer / User Journey Maps in GitHub	8 October 2022

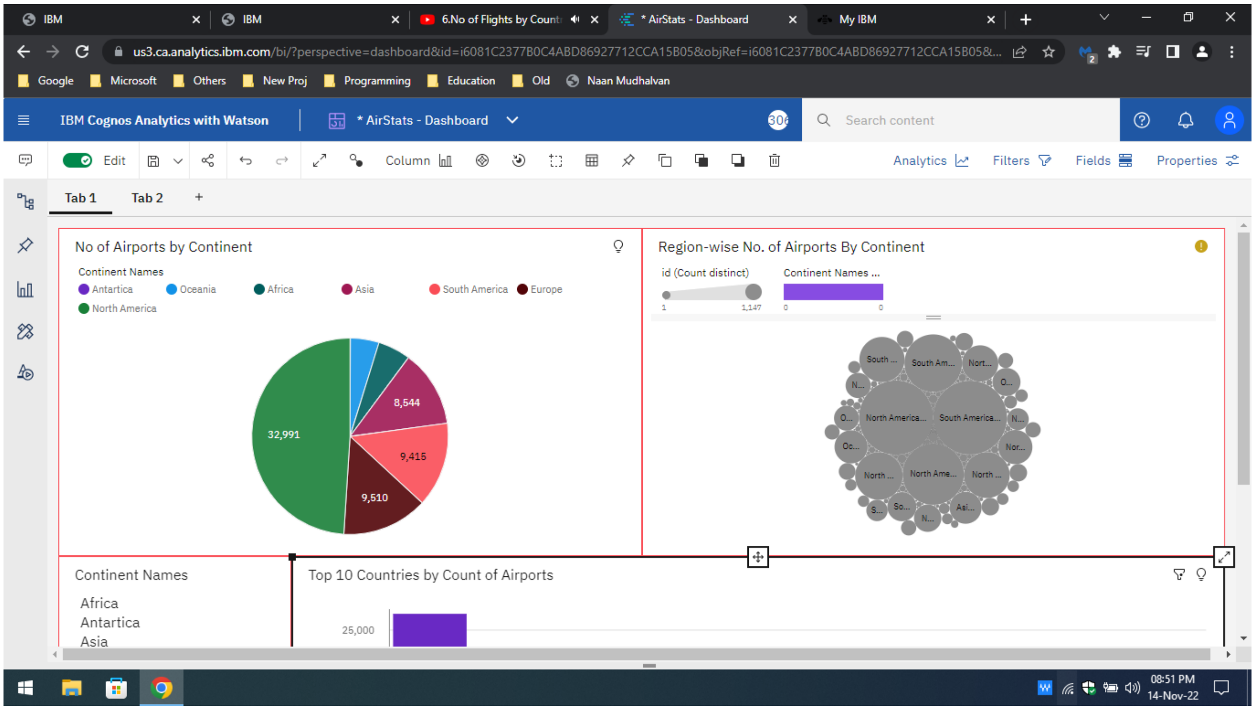
Customer Journey		
Functional Requirement	Submit the Functional Requirements in GitHub	8 October 2022
Data Flow Diagrams	Submit the Data flow dailgrams in GitHub	15 October 2022
Technology Architecture	Submit the Technology Architecture in GitHub	
Project Planning Phase: Prepare Milestone & Activity List	Submit the Milestone & Activity List	22 October 2022
Sprint Delivery Plan	Submit the Project Delivery Schedule as per the standard template in GitHub	22 October 2022
Project Development Phase: Sprint-1, Sprint-2, Sprint-3 and Sprint-4	Develop the Code, Test and push it to GitHub.	19 November 2022

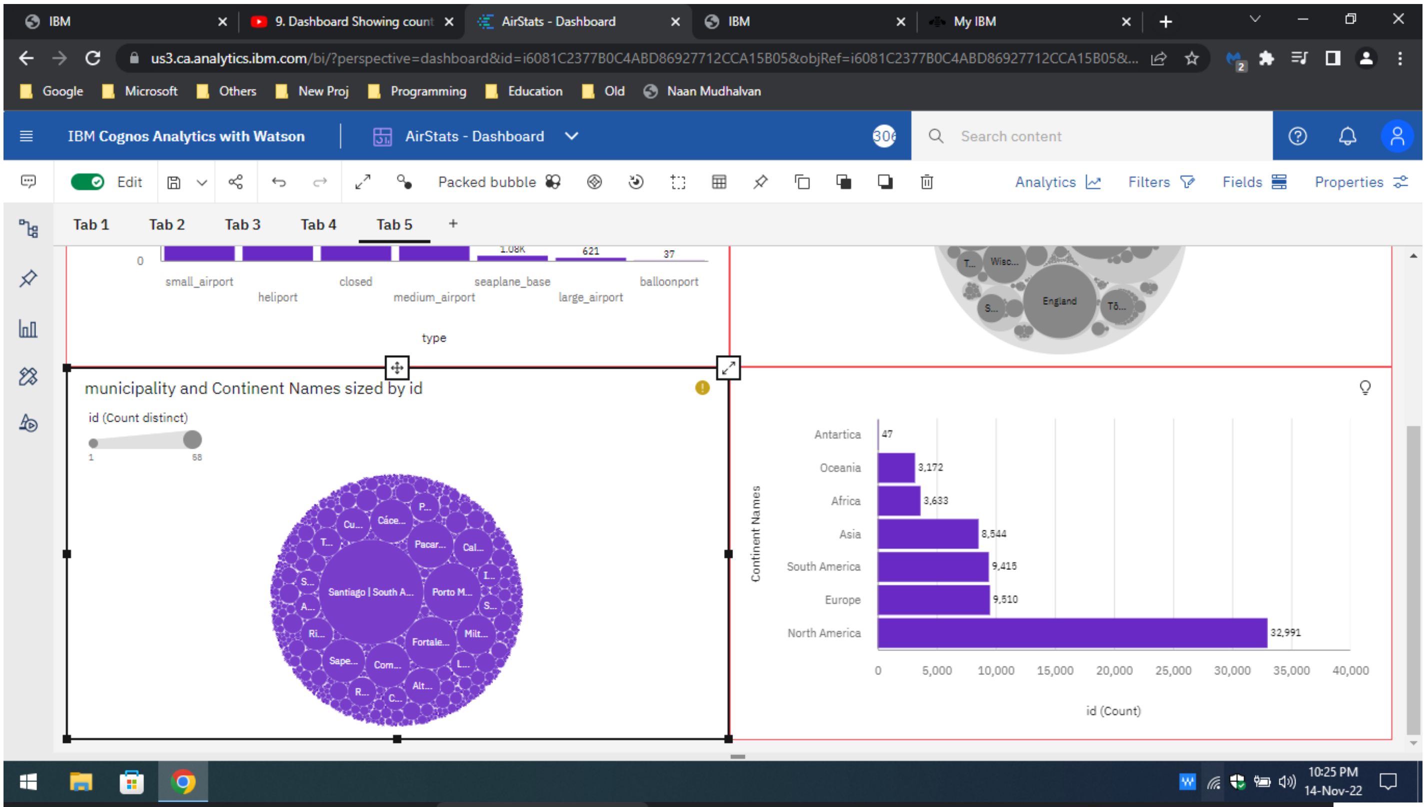
CHAPTER -7

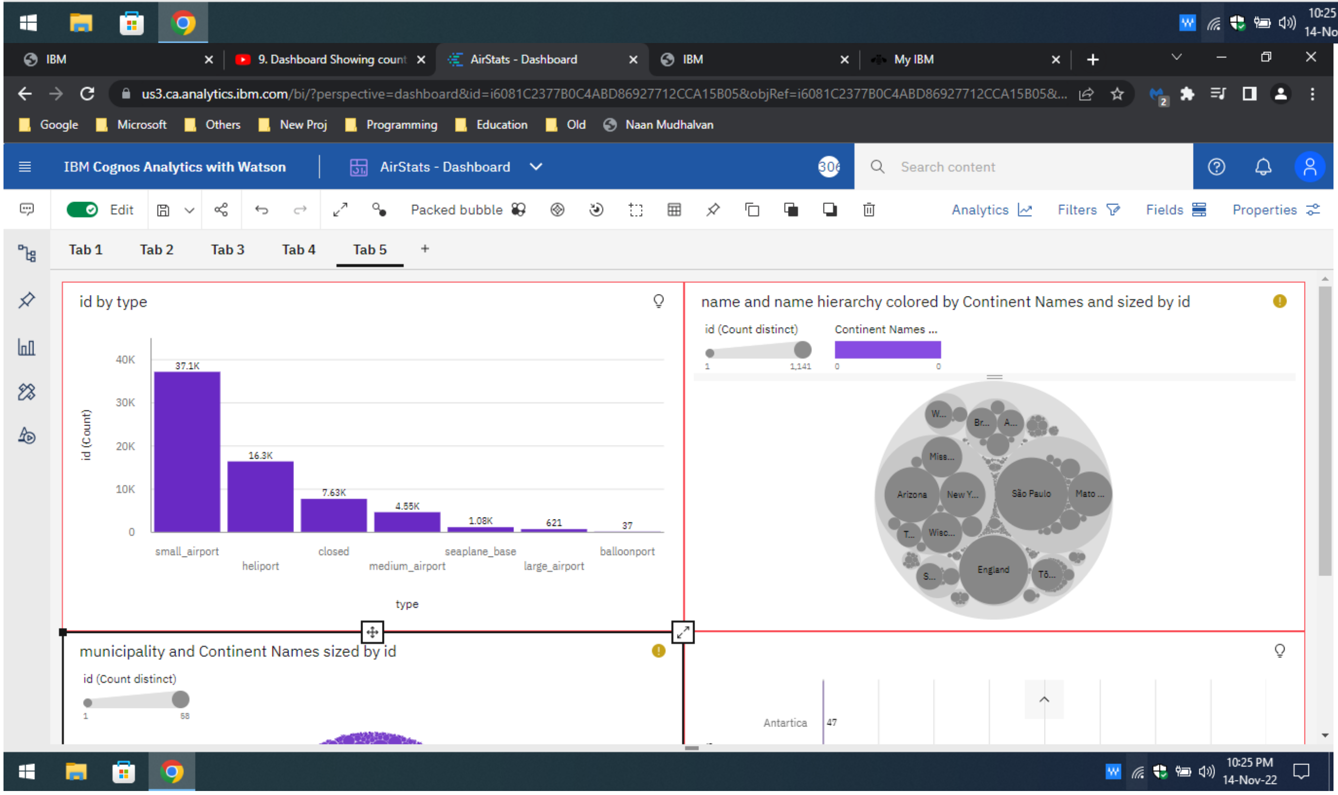
CODING & SOLUTIONING

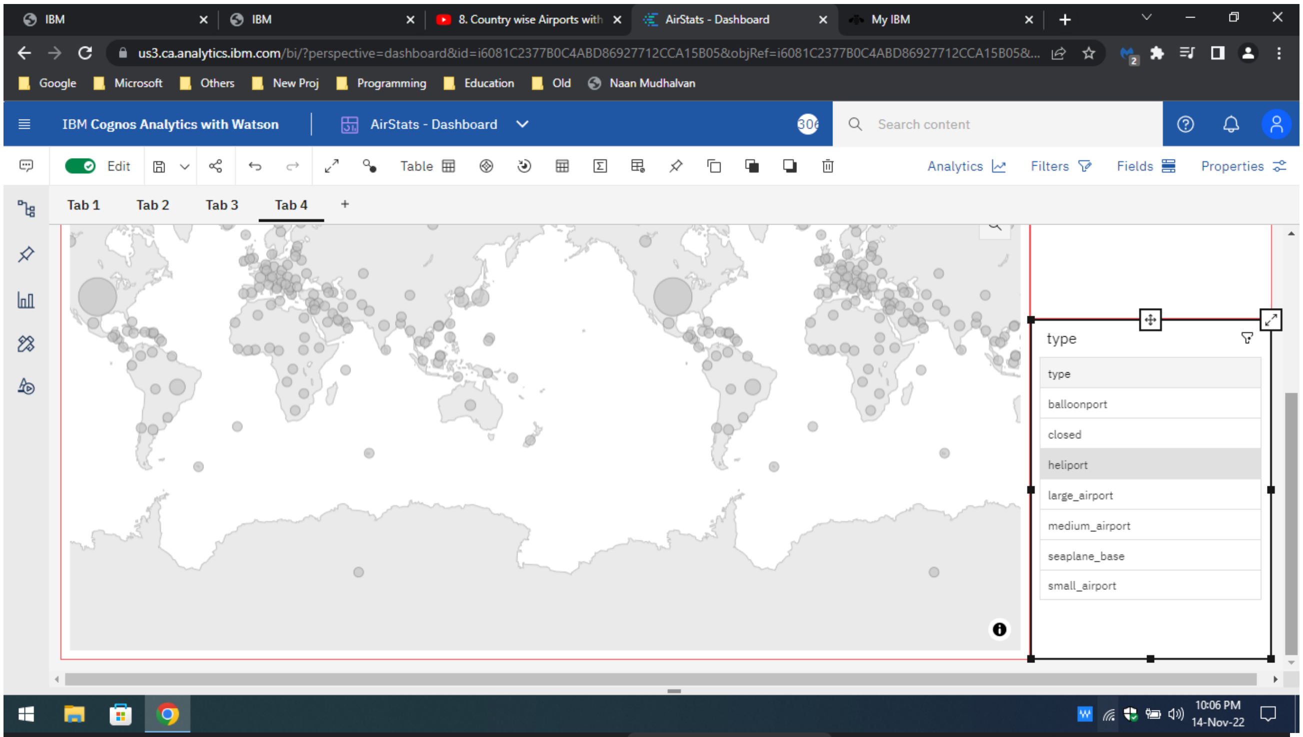
FEATURE 1

Interactive dashboard allows user to interact with dashboard visualizing various details of flights including no of airports in different continent and in different countires. And accurate information on the flight arrival and departure timing that reducing delays on time.




IBM
9. Dashboard Showing count
AirStats - Dashboard
IBM
My IBM
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...
Google
Microsoft
Others
New Proj
Programming
Education
Old
Naan Mudhalvan
IBM Cognos Analytics with Watson
AirStats - Dashboard
Search content
Edit
Packed bubble
Analytics
Filters
Fields
Properties
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
1.08K
621
37
0
small_airport
heliport
closed
medium_airport
seaplane_base
large_airport
balloonport
type
T...
Wisc...
England
S...
Tõ...
municipality and Continent Names sized by id
id (Count distinct)
1
58
P...
Cu...
Cáce...
T...
Pacar...
Cal...
S...
Santiago | South A...
Porto M...
I...
A...
S...
Ri...
Milt...
Fortale...
Sape...
Com...
L...
R...
C...
Alt...
Antartica
Oceania
Africa
Asia
South America
Europe
North America
47
3,172
3,633
8,544
9,415
9,510
32,991
Continent Names
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
id (Count)
10:25 PM
14-Nov-22


IBM
9. Dashboard Showing count
AirStats - Dashboard
IBM
My IBM
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...
Google
Microsoft
Others
New Proj
Programming
Education
Old
Naan Mudhalvan
IBM Cognos Analytics with Watson
AirStats - Dashboard
Search content
Edit
Packed bubble
Analytics
Filters
Fields
Properties
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
id by type
40K
30K
20K
10K
0
id (Count)
37.1K
16.3K
7.63K
4.55K
1.08K
621
37
small_airport
heliport
closed
medium_airport
seaplane_base
large_airport
balloonport
type
name and name hierarchy colored by Continent Names and sized by id
id (Count distinct)
1
1,141
Continent Names ...
0
0
W...
Br...
A...
Miss...
Arizona
New Y...
São Paulo
Mato ...
T...
Wisc...
S...
England
Tõ...
municipality and Continent Names sized by id
id (Count distinct)
1
58
Antartica
47
10:25 PM
14-Nov-22



IBM | IBM | 8. Country wise Airports with | AirStats - Dashboard | My IBM

us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...

Google | Microsoft | Others | New Proj | Programming | Education | Old | Naan Mudhalvan

IBM Cognos Analytics with Watson | AirStats - Dashboard

Search content

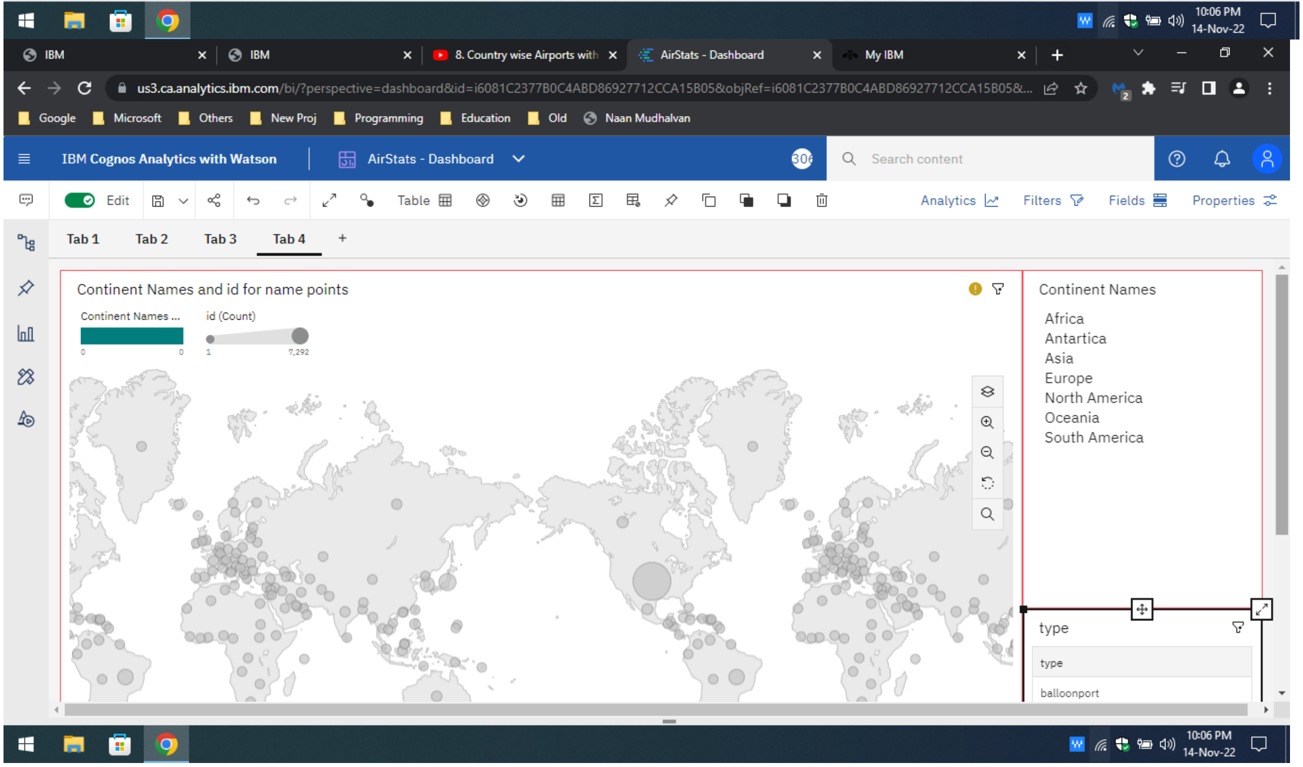
Edit | Table | Analytics | Filters | Fields | Properties

Tab 1 | Tab 2 | Tab 3 | Tab 4 | +

type

- type
- balloonport
- closed
- heliport
- large_airport
- medium_airport
- seaplane_base
- small_airport

10:06 PM
14-Nov-22



IBM | IBM | 8. Country wise Airports with | AirStats - Dashboard | My IBM

us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...

Google | Microsoft | Others | New Proj | Programming | Education | Old | Naan Mudhalvan

IBM Cognos Analytics with Watson | AirStats - Dashboard

Search content

Edit | Table | Analytics | Filters | Fields | Properties

Tab 1 | Tab 2 | Tab 3 | Tab 4 | +

Continent Names and id for name points

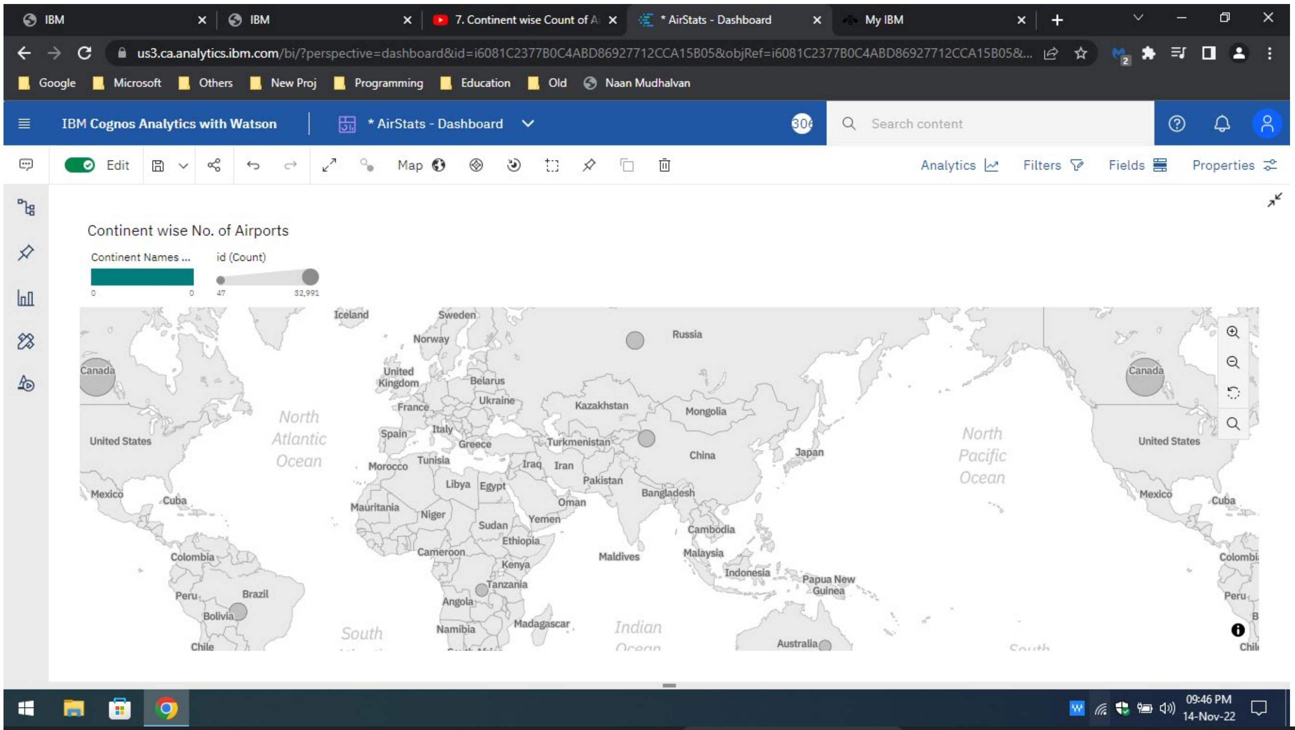
Continent Names

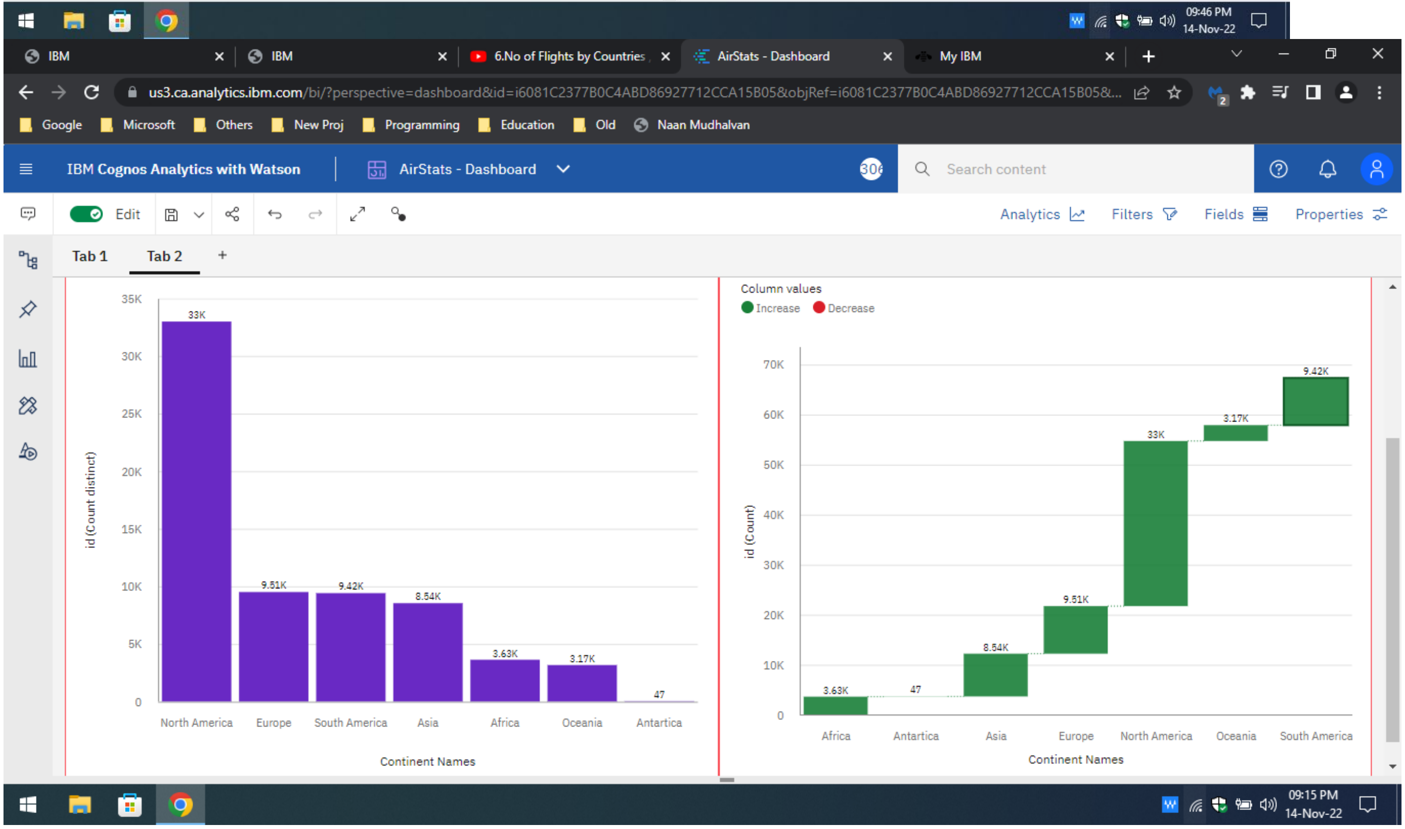
- Africa
- Antartica
- Asia
- Europe
- North America
- Oceania
- South America

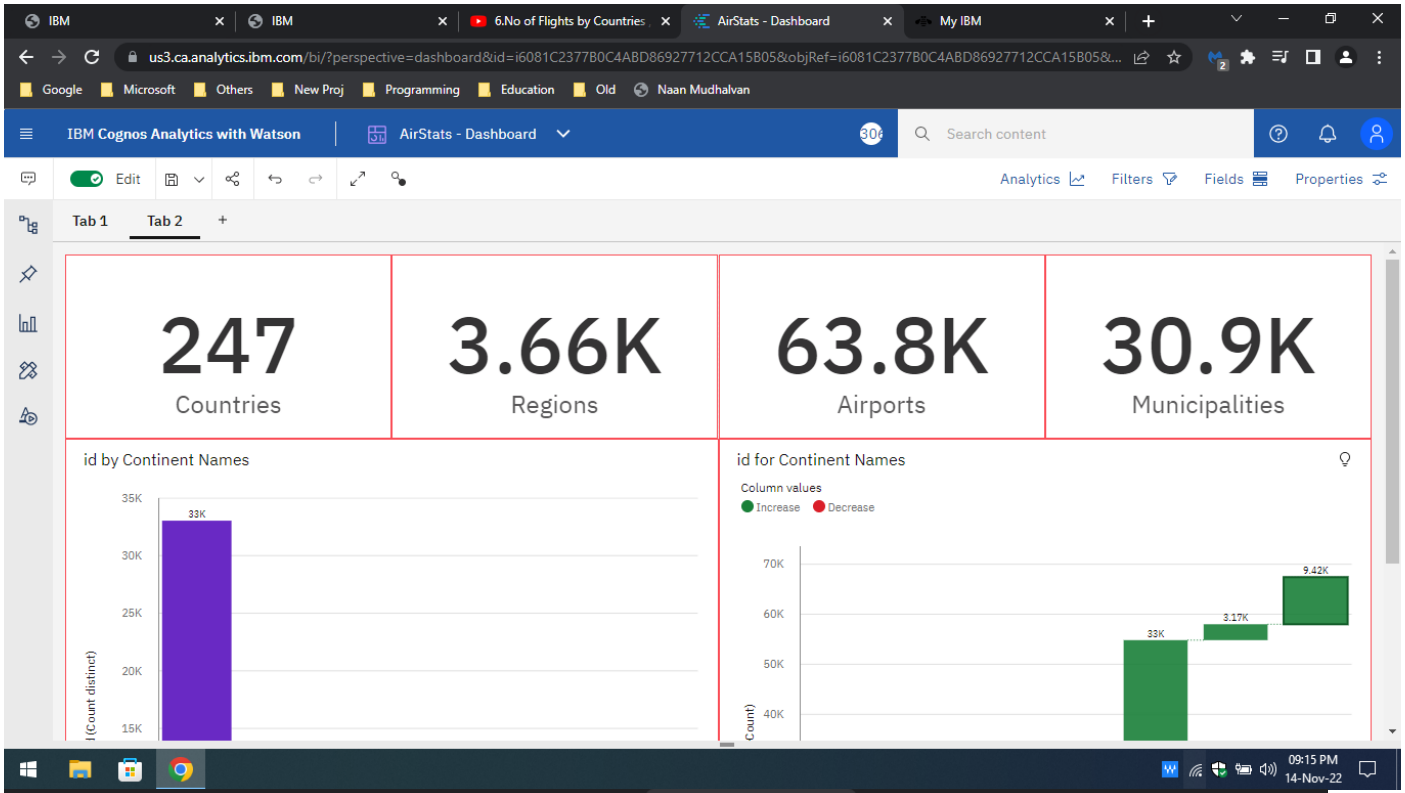
type

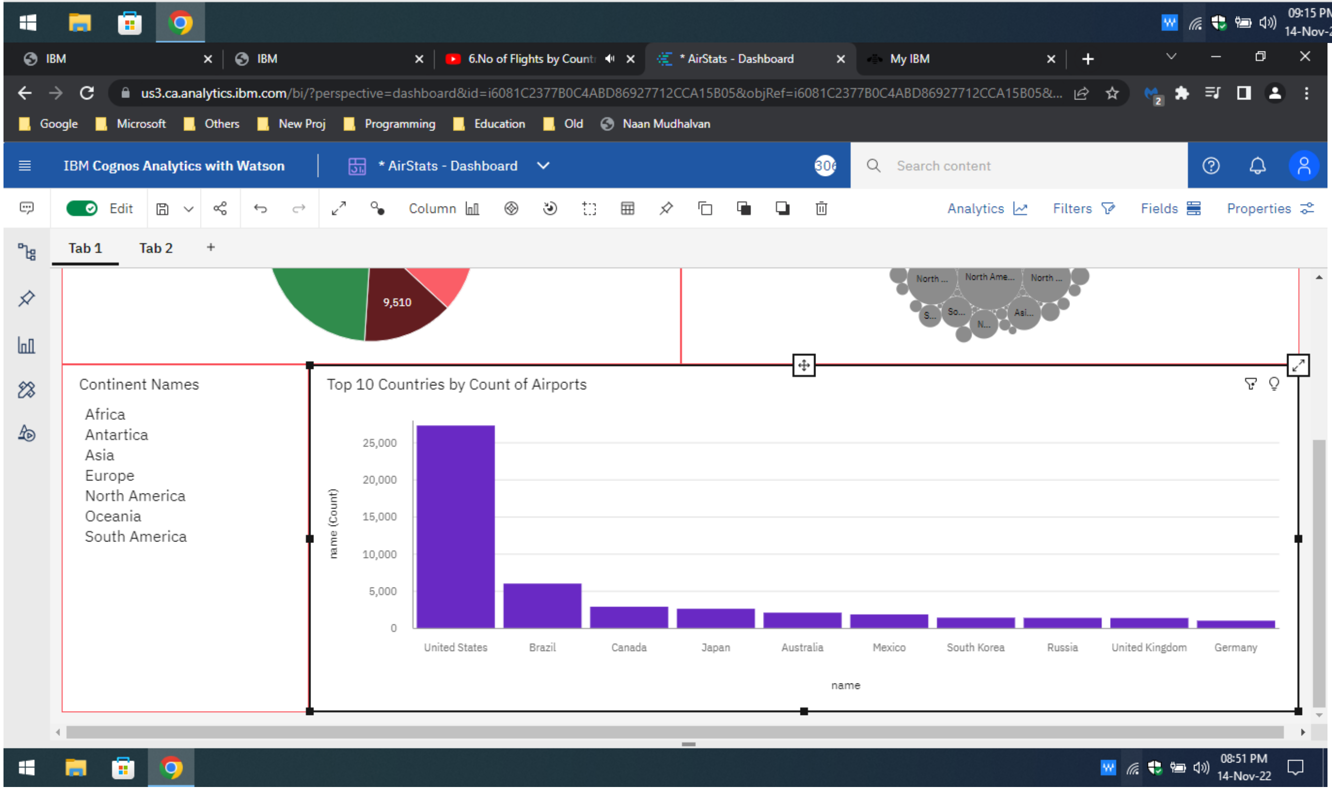
- type
- balloonport

10:06 PM
14-Nov-22


IBM
IBM
7. Continent wise Count of A
* AirStats - Dashboard
My IBM
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...
Google
Microsoft
Others
New Proj
Programming
Education
Old
Naan Mudhalvan
IBM Cognos Analytics with Watson
* AirStats - Dashboard
Search content
Edit
Map
Analytics
Filters
Fields
Properties
Continent wise No. of Airports
Continent Names ...
id (Count)
0
0
47
32,991
Iceland
Sweden
Norway
Russia
Canada
United Kingdom
Belarus
Ukraine
France
Kazakhstan
Mongolia
North Atlantic Ocean
Spain
Italy
Greece
Turkmenistan
China
Japan
North Pacific Ocean
United States
Tunisia
Morocco
Iraq
Iran
Libya
Egypt
Pakistan
Bangladesh
Mexico
Cuba
Mauritania
Oman
Niger
Yemen
Sudan
Cambodia
Ethiopia
Cameroon
Malaysia
Colombia
Kenya
Maldives
Indonesia
Tanzania
Papua New Guinea
Peru
Brazil
Angola
Bolivia
Madagascar
Namibia
Indian Ocean
South
Chile
Australia
09:46 PM
14-Nov-22


IBM
IBM
6.No of Flights by Countries
AirStats - Dashboard
My IBM
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...
Google
Microsoft
Others
New Proj
Programming
Education
Old
Naan Mudhalvan
IBM Cognos Analytics with Watson
AirStats - Dashboard
Search content
Edit
Analytics
Filters
Fields
Properties
Tab 1
Tab 2
+
id (Count distinct)
35K
30K
25K
20K
15K
10K
5K
0
33K
9.51K
9.42K
8.54K
3.63K
3.17K
47
North America
Europe
South America
Asia
Africa
Oceania
Antartica
Continent Names
Column values
● Increase
● Decrease
id (Count)
70K
60K
50K
40K
30K
20K
10K
0
3.63K
47
8.54K
9.51K
33K
3.17K
9.42K
Africa
Antartica
Asia
Europe
North America
Oceania
South America
Continent Names
09:15 PM
14-Nov-22


IBM
IBM
6.No of Flights by Countries
AirStats - Dashboard
My IBM
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...
Google
Microsoft
Others
New Proj
Programming
Education
Old
Naan Mudhalvan
IBM Cognos Analytics with Watson
AirStats - Dashboard
Search content
Edit
Analytics
Filters
Fields
Properties
Tab 1
Tab 2
+
247
Countries
3.66K
Regions
63.8K
Airports
30.9K
Municipalities
id by Continent Names
35K
30K
25K
20K
15K
33K
id (Count distinct)
id for Continent Names
Column values
Increase
Decrease
70K
60K
50K
40K
Count)
33K
3.17K
9.42K
09:15 PM
14-Nov-22


IBM
IBM
6.No of Flights by Countr
* AirStats - Dashboard
My IBM
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=i6081C2377B0C4ABD86927712CCA15B05&objRef=i6081C2377B0C4ABD86927712CCA15B05&...
Google
Microsoft
Others
New Proj
Programming
Education
Old
Naan Mudhalvan
IBM Cognos Analytics with Watson
* AirStats - Dashboard
Search content
Edit
Column
Analytics
Filters
Fields
Properties
Tab 1
Tab 2
+
9,510
North ...
North Ame...
North ...
S...
So...
N...
Asi...
Continent Names
Africa
Antartica
Asia
Europe
North America
Oceania
South America
Top 10 Countries by Count of Airports
25,000
20,000
15,000
10,000
5,000
0
name (Count)
United States
Brazil
Canada
Japan
Australia
Mexico
South Korea
Russia
United Kingdom
Germany
name
08:51 PM
14-Nov-22

CHAPTER – 8

TESTING

8.1 TEST CASES

Section	TotalCases	Not Tested	Fail	Pass
PrintEngine	9	0	0	9
ClientApplication	45	0	0	45
Security	2	0	0	2
OutsourceShipping	3	0	0	3
ExceptionReporting	9	0	0	9
FinalReportOutput	4	0	0	4
VersionControl	2	0	0	2

8.2 USER ACCEPTANCE TESTING

DefectAnalysis

Thisreportshowsthenumberofresolvedor closed bugs at each severity level, and how they were resolved

Resolution	Severity1	Severity2	Severity3	Severity4	Subtotal
By Design	5	4	2	3	14
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	9	2	4	15	30
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won'tFix	0	5	2	1	8
Totals	17	14	13	21	65

CHAPTER – 9

RESULT

9.1 PERFORMANCE METRICES

Learning Algorithm	**Test Data**	**Training Data**	**Training Time**
Machine Learning	50	50	5 mins
Artificial Intelligence	100	100	10 mins
Neural Networking	70	70	7mins
Quantum Computing	40	40	4mins

CHAPTER – 10

ADVANTAGES & DISADVANTAGES

ADVANTAGES:

The advantages of this project are reducing delays in flight and hence improving passenger satisfaction. Using advanced software technologies like Data Analytics further reduces the solution cost and improves future scaling of the project. The interactive dashboard dynamically adjusts the details of flight with user generated input and hence user only gets the desired information, Ease of use with easy login and signin options in a secured environment

DISADVANTAGES:

The disadvantages of this project is it requires and stable internet connection to access the highly interactive dashboard and reports generation. The interactive dashboard also includes advanced and complex options. The accuracy of decreasing flight delays is still not perfect. The aviation industry have to regularly update the datasets which causes time and space complexity. Flight details from other aviation industry also needed for better visualization of the details in the dashboard.

CHAPTER -11

CONCLUSION

To date, several studies have discussed the technical aspects of big data, but fewer focused on the organizational outcomes of big data technology especially in the tourism industry. It is also significant to understand the mechanisms and processes through which big data can add business value to tourism enterprises. Future studies need to be addressed for the previous domains. Besides, further studies can address the modern technology techniques and their impacts on the tourism and aviation industry. It is suggested that further research explores more opportunities and challenges in big data technology, and its relations with block-chain, cloud computing, artificial intelligence can be used to predict future glitches, prevent them from happening, and make the maintenance procedures more accurate and thorough. After analyzing the data, a lot of insights have been generated. Most of the delays and cancellations are due to three major reasons weather, Airline/Carrier Issues and National Air System

CHAPTER – 12

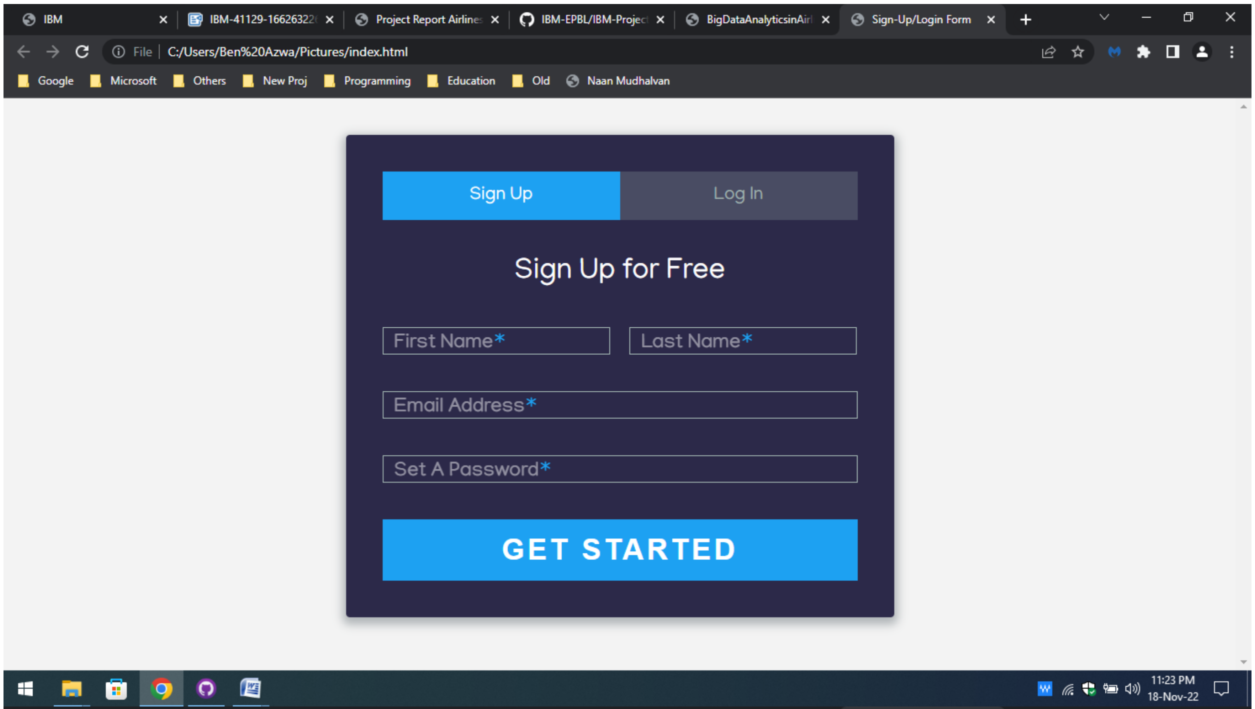
FUTURE SCOPE

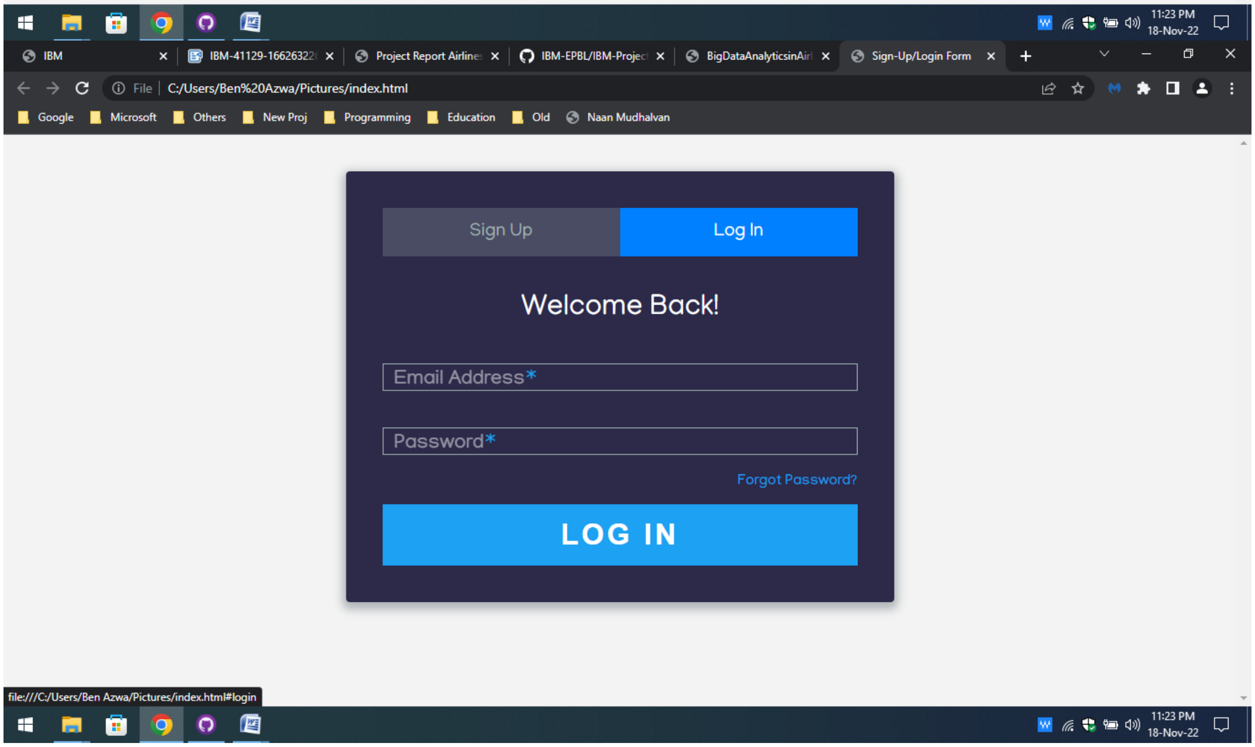
The future scope to add more sign-in option for easy convenient methods and easy to access dashboard design and add features like multiple user login and other security and privacy updates by reducing user information collection and improving User experience and user interface.

CHAPTER – 13

APPENDIX

13.1 SCREENSHOTS





13.2 SOURCE CODE

Sprint-1

index.html

```
            <!DOCTYPE html>
            <html lang="en" >
            <head>
              <meta charset="UTF-8">
              <title>Sign-Up/Login Form</title>
              <link href="https://fonts.googleapis.com/css?family=Manjari&display=swap"
      rel="stylesheet">
              <link rel="stylesheet"
      href="https://cdnjs.cloudflare.com/ajax/libs/normalize/5.0.0/normalize.min.css">
              <link rel="stylesheet" href="style.css">
            </head>
            <body>

<div class="form">

  <ul class="top-area">
    <li class="tab active"><a href="#signup">Sign Up</a></li>
    <li class="tab"><a href="#login">Log In</a></li>
  </ul>

  <div class="tab-content">
    <div id="signup">
      <h1>Sign Up for Free</h1>
```

```
<form action="/" method="post">

<div class="top-row">
  <div class="label-field">
    <label>
      First Name<span class="req">*</span>
    </label>
    <input type="text" required autocomplete="off" />
  </div>

  <div class="label-field">
    <label>
      Last Name<span class="req">*</span>
    </label>
    <input type="text"required autocomplete="off"/>
  </div>
</div>

<div class="label-field">
  <label>
    Email Address<span class="req">*</span>
  </label>
  <input type="email"required autocomplete="off"/>
</div>

<div class="label-field">
```

```
    <label>
      Set A Password<span class="req">*</span>
    </label>
    <input type="password"required autocomplete="off"/>
  </div>

  <button type="submit" class="button button-block"/>Get Started</button>

  </form>

</div>

<div id="login">
  <h1>Welcome Back!</h1>

  <form action="/" method="post">

    <div class="label-field">
    <label>
      Email Address<span class="req">*</span>
    </label>
    <input type="email"required autocomplete="off"/>
  </div>

  <div class="label-field">
    <label>
      Password<span class="req">*</span>
    </label>
```

```
            <input type="password"required autocomplete="off"/>
          </div>

          <p class="forgot"><a href="#">Forgot Password?</a></p>

          <button class="button button-block">Log In</button>

          </form>

        </div>

      </div>
</div>

  <script src='http://cdnjs.cloudflare.com/ajax/libs/jquery/2.1.3/jquery.min.js'></script>
<script  src="function.js"></script>

</body>
</html>
```

function.js

```
$('.form').find('input, textarea').on('keyup blur focus', function (e) {

  var $this = $(this),
      label = $this.prev('label');

        if (e.type === 'keyup') {
```

```
      if ($this.val() === '') {
          label.removeClass('active highlight');
        } else {
          label.addClass('active highlight');
        }
    } else if (e.type === 'blur') {
      if( $this.val() === '' ) {
      label.removeClass('active highlight');
      } else {
          label.removeClass('highlight');
      }
    } else if (e.type === 'focus') {

      if( $this.val() === '' ) {
      label.removeClass('highlight');
      }
      else if( $this.val() !== '' ) {
          label.addClass('highlight');
      }
    }

});

$('.tab a').on('click', function (e) {

  e.preventDefault();
  $(this).parent().addClass('active');
  $(this).parent().siblings().removeClass('active');

  target = $(this).attr('href');

  $('.tab-content > div').not(target).hide();

  $(target).fadeIn(600);

});
```

style.css

```
*, *:before, *:after {
  box-sizing: border-box;
}
html {
  overflow-y: scroll;
}
body {
  background: #f3f3f3;
  font-family: 'Manjari', sans-serif;
}
a {
  text-decoration: none;
  color: #1da1f2;
  transition: .5s ease;
```

```
}
a:hover {
  color: #0080ff;
}
.form {
  background: rgb(22,19,54, 0.9);
  padding: 40px;
  max-width: 600px;
  margin: 40px auto;
  border-radius: 4px;
  box-shadow: 0 4px 10px 4px rgba(19, 35, 47, 0.3);
}
.top-area {
  list-style: none;
  padding: 0;
  margin: 0 0 40px 0;
}
.top-area:after {
  content: "";
  display: table;
  clear: both;
}
.top-area li a {
  display: block;
  text-decoration: none;
  padding: 15px;
  background: rgba(160, 179, 176, 0.25);
```

```
}

  color: #a0b3b0;
  font-size: 20px;
  float: left;
  width: 50%;
  text-align: center;
  cursor: pointer;
  transition: .5s ease;
}
.top-area li a:hover {
  background: #0080ff;
  color: #ffffff;
}
.top-area .active a {
  background: #1da1f2;
  color: #ffffff;
}

.tab-content > div:last-child {
  display: none;
}

h1 {
  text-align: center;
  color: #ffffff;
  font-weight: 300;
  margin: 0 0 40px;
}
```

```
label {
  position: absolute;
  -webkit-transform: translateY(6px);
          transform: translateY(6px);
  left: 13px;
  color: rgba(255, 255, 255, 0.5);
  transition: all 0.25s ease;
  -webkit-back face-visibility: hidden;
  pointer-events: none;
  font-size: 22px;
}
label .req {
  margin: 2px;
  color: #1da1f2;
}
label.active {
  -webkit-transform: translateY(50px);
          transform: translateY(50px);
  left: 2px;
  font-size: 14px;
}
label.active .req {
  opacity: 0;
}
label.highlight {
  color: #ffffff;
}
```

```
input, textarea {
  font-size: 22px;
  display: block;
  width: 100%;
  height: 100%;
  background: none;
  background-image: none;
  border: 1px solid #a0b3b0;
  color: #ffffff;
  border-radius: 0;
  transition: border-color .25s ease, box-shadow .25s ease;
}
input:focus, textarea:focus {
  outline: 0;
  border-color: #1da1f2;
}
textarea {
  border: 2px solid #a0b3b0;
  resize: vertical;
}
.label-field {
  position: relative;
  margin-bottom: 40px;
}
.top-row:after {
  content: "";
  display: table;
  clear: both;
}
.top-row > div {
  float: left;
  width: 48%;
  margin-right: 4%;
}
.top-row > div:last-child {
  margin: 0;
}
.button {
  border: 0;
  outline: none;
  border-radius: 0;
  padding: 15px 0;
  font-size: 2rem;
  font-weight: 600;
  text-transform: uppercase;
  letter-spacing: .1em;
  background: #1da1f2;
  color: #ffffff;
  transition: all 0.5s ease;
  -webkit-appearance: none;
}
.button:hover, .button:focus {
  background: #0080ff;
}
```

```
.button-block {
  display: block;
  width: 100%;
}
.forgot {
  margin-top: -20px;
  text-align: right;
}
@media (max-width: 765px) {
    label {
      left: 0;
    }    }
```

Sprint-2

scale_fix.js

```
var met as = document.getElementsByTagName('meta');
var i;
if (navigator.userAgent.match(/iPhone/i)) {
  for (i=0; i<metas.length; i++) {
    if (metas[i].name == "view port") {
      metas[i].content = "width=device-width, minimum-scale=1.0, maximum-scale=1.0";
    }
  }
  document.addEventListener("gesturestart", gestureStart, false);
}
function gestureStart() {
 for (i=0; i<metas.length; i++) {
  if (metas[i].name == "view port") {
   metas[i].content = "width=device-width, minimum-scale=0.25, maximum-scale=1.6";
  }
 }
}
```

style.css

```
body {
  background-color: #fff;
  padding:50px;
  font: 15px/1.5 -apple-system, BlinkMacSystemFont, "Segoe UI", Roboto, Oxygen, Ubuntu, Cantarell, "Fira Sans", "Droid Sans", "Helvetica Neue", Arial, sans-serif, "Apple Color Emoji", "Segoe UI Emoji", "Segoe UI Symbol";
  color:#595959;
  font-weight:400;
 }

 h1, h2, h3, h4, h5, h6 {
  color:#222;
  margin:0 0 20px;
 }

 p, ul, ol, table, pre, dl {
  margin:0 0 20px;
 }
```

```
}

h1, h2, h3 {
  line-height:1.1;
}

h1 {
  font-size:28px;
  font-weight: 500;
}

h2 {
  color:#393939;
  font-weight: 500;
}

h3, h4, h5, h6 {
  color:#494949;
  font-weight: 500;
}

a {
  color:#39c;
  text-decoration:none;
}

a:hover {
  color:#069;
}
```

```
a small {
  font-size:11px;
  color:#777;
  margin-top:-0.3em;
  display:block;
}

a:hover small {
  color:#777;
}

.wrapper {
  width:1080px;
  margin:0 auto;
}

blockquote {
  border-left:1px solid #e5e5e5;
  margin:0;
  padding:0 0 0 20px;
  font-style:italic;
}

code, pre {
  font-family:Monaco, Bitstream Vera Sans Mono, Lucida Console, Terminal, Consolas, Liberation Mono, DejaVu Sans Mono, Courier New, monospace;
  color:#333;
}
```

```css
pre {
  padding:8px 15px;
  background: #f8f8f8;
  border-radius:5px;
  border:1px solid #e5e5e5;
  overflow-x: auto;
}

table {
  width:100%;
  border-collapse:collapse;
}

th, td {
  text-align:left;
  padding:5px 10px;
  border-bottom:1px solid #e5e5e5;
}

dt {
  color:#444;
  font-weight:500;
}

th {
  color:#444;
```

```
}

img {
  max-width:100%;
}

header {
  width:270px;
  float:left;
  position:fixed;
  -webkit-font-smoothing:subpixel-antialiased;
}

header ul {
  list-style:none;
  height:40px;
  padding:0;
  background: #f4f4f4;
  border-radius:5px;
  border:1px solid #e0e0e0;
  width:270px;
}

header li {
  width:89px;
  float:left;
  border-right:1px solid #e0e0e0;
  height:40px;
}

header li:first-child a {
  border-radius:5px 0 0 5px;
}

header li:last-child a {
  border-radius:0 5px 5px 0;
}

header ul a {
  line-height:1;
  font-size:11px;
  color:#999;
  display:block;
  text-align:center;
  padding-top:6px;
  height:34px;
}

header ul a:hover {
  color:#999;
}

header ul a:active {
  background-color:#f0f0f0;
```

```
}

strong {
  color:#222;
  font-weight:500;
}

header ul li + li + li {
  border-right:none;
  width:89px;
}

header ul a strong {
  font-size:14px;
  display:block;
  color:#222;
}

section {
  width:750px;
  float:right;
  padding-bottom:50px;
}

small {
  font-size:11px;
}
```

```
hr {
  border:0;
  background:#e5e5e5;
  height:1px;
  margin:0 0 20px;
}

footer {
  width:270px;
  float:left;
  position:fixed;
  bottom:50px;
  -webkit-font-smoothing:subpixel-antialiased;
}

@media print, screen and (max-width: 960px) {

  div.wrapper {
    width:auto;
    margin:0;
  }

  header, section, footer {
    float:none;
    position:static;
    width:auto;
  }
  header {
    padding-right:320px;
  }

  section {
    border:1px solid #e5e5e5;
    border-width:1px 0;
    padding:20px 0;
    margin:0 0 20px;
  }

  header a small {
    display:inline;
  }

  header ul {
    position:absolute;
    right:50px;
    top:52px;
  }
}

@media print, screen and (max-width: 720px) {
  body {
    word-wrap:break-word;
  }
  header {
    padding:0;
  }

  header ul, header p.view {
    position:static;
  }

  pre, code {
    word-wrap:normal;
  }
}

@media print, screen and (max-width: 480px) {
  body {
    padding:15px;
  }

  header ul {
    width:99%;
  }

  header li, header ul li + li + li {
    width:33%;
  }
}

@media print {
    body {
      padding:0.4in;
      font-size:12pt;
      color:#444;
    }
  }
```

Sprint-3

default.html

```
<!doctype html>
<html>
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="chrome=1">
    <link rel='shortcut icon' type='image/x-icon' href='assets/img/airplane_icon.ico' />
    <title>{{ site.title | default: site.github.repository_name }} by {{ site.github.owner_name }}</title>

    <link rel="stylesheet" href="{{ '/assets/css/styles.css?v=' | append: site.github.build_revision | relative_url }}">
    <meta name="viewport" content="width=device-width">
  </head>
  <body>
    <div class="wrapper">
      <header>
        <h1>{{ site.title | default: site.github.repository_name }}</h1>
        <p>{{ site.description | default: site.github.project_tagline }}</p>
  <table>
    {% for nav in site.navigation %}
      {% if nav.url contains "://" %}
        <tr><th><a href="{{ nav.url }}">{{ nav.title }}</a></th></tr>
      {% else %}
        <tr><th><a href="{{ nav.url | relative_url }}">{{ nav.title }}</a></th></tr>
      {% endif %}
    {% endfor %}
  </table>

  {% if site.show_downloads %}
    <ul>
      <li><a href="{{ site.github.zip_url }}">Download <strong>ZIP File</strong></a></li>
      <li><a href="{{ site.github.tar_url }}">Download <strong>TAR Ball</strong></a></li>
      <li><a href="{{ site.github.repository_url }}">View On <strong>GitHub</strong></a></li>
    </ul>
  {% endif %}
</header>
<section>

{{ content }}

</section>
<footer>
  {% if site.github.is_project_page %}
```

```
        {% endif %}
      </footer>
    </div>
    <script src="{{ '/assets/js/scale.fix.js' | relative_url }}"></script>
  </body>
</html>
```

full.html

```
<!doctype html>
<html>
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="chrome=1">
    <link rel='shortcut icon' type='image/x-icon' href='assets/img/airplane_icon.ico' />
    <title>{{ site.title | default: site.github.repository_name }} by {{ site.github.owner_name }}</title>
    <link rel="stylesheet" href="{{ '/assets/css/styles.css?v=' | append: site.github.build_revision | relative_url }}">
    <meta name="viewport" content="width=device-width">
  </head>
  <body>
    <div class="wrapper">

      {{ content }}

    </div>
    <script src="{{ '/assets/js/scale.fix.js' | relative_url }}"></script>
  </body>
```

```
</html>
```

Sprint-4

1_analytical_func_busiest_airport_airline.sql

```
create database analytical_func_busiest_airport_airline
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 ),
  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.ORIGIN
  GROUP BY
    1
  ORDER BY
    2 DESC
  LIMIT
    5),
  airportwise_carrier_cnt AS (
  SELECT
    main.ORIGIN AS Airport,
    main.OP_CARRIER AS Carrier,
    COUNT(*) AS count
  FROM
    `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_CARRIER
  GROUP BY
    1,
    2 ),
  resut_cte AS (
  SELECT
    Airport,
      Carrier,
      count,
      RANK() OVER(PARTITION BY Airport ORDER BY count) AS rank
    FROM
      airportwise_carrier_cnt)
  SELECT
    Airport,
    Carrier,
    count
  FROM
    resut_cte
  WHERE
    rank < 6
```

2_all-cancellations.sql

```
-- 3 Cancellation Bifurcation
WITH
  top_5_airports AS (
  SELECT
    ORIGIN,
    COUNT(ORIGIN) AS count
  FROM
    `airline-delay-canc.airlines_data.delay_canc_data`
  GROUP 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.ORIGIN
GROUP 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_al
  WHERE
    top5_ap.ORIGIN = main.ORIGIN
    AND top_al.OP_CARRIER = main.OP_CARRIER
  GROUP 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_al
  WHERE
    top5_ap.ORIGIN = main.ORIGIN
    AND top_al.OP_CARRIER = main.OP_CARRIER
    AND cancelled = 1
  GROUP BY
    1,
    2 )
SELECT
  af.Airport,
  af.Carrier,
  af.all_cnt - cf.cancelled_cnt AS all_cnt,
  cf.cancelled_cnt
FROM
  all_flights af,
  cancelled_flights cf
WHERE
  af.Airport = cf.Airport
  AND af.Carrier = cf.Carrier
```

3,4_all_delays.sql

```
WITH
  top_5_airports AS (
  SELECT
    ORIGIN,
    COUNT(ORIGIN) AS count
  FROM
    `airline-delay-canc.airlines_data.delay_canc_data`
  GROUP 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.ORIGIN
GROUP 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_al
WHERE
  top5_ap.ORIGIN = main.ORIGIN
  AND top_al.OP_CARRIER = main.OP_CARRIER
GROUP 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_al
  WHERE
    top5_ap.ORIGIN = main.ORIGIN
    AND top_al.OP_CARRIER = main.OP_CARRIER
    AND (CARRIER_DELAY IS NOT NULL
      AND CARRIER_DELAY > 0
      OR ARR_DELAY IS NOT NULL
      AND ARR_DELAY > 0)
  GROUP BY
    1,
    2 )
SELECT
  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
```

5,6_year wise_canc_delay.sql

```
WITH
  cancellation_data AS (
  SELECT
    EXTRACT(year
    FROM
      FL_DATE) AS year,
    COUNT(*) AS cancellation_cnt
  FROM
    `airline-delay-canc.airlines_data.delay_canc_data`
  WHERE
    CANCELLED = 1
  GROUP BY
    year
  ORDER BY
    year ),
  delayed_data AS (
  SELECT
    EXTRACT(year
    FROM
      FL_DATE) AS year,
    COUNT(*) AS delay_cnt
  FROM
    `airline-delay-canc.airlines_data.delay_canc_data`
  WHERE
    (CARRIER_DELAY IS NOT NULL
      AND CARRIER_DELAY > 0
      OR ARR_DELAY IS NOT NULL
      AND ARR_DELAY > 0)
  GROUP BY
    year
  ORDER BY
    year )
SELECT
  c.year,
  c.cancellation_cnt,
  d.delay_cnt
FROM
  cancellation_data c,
  delayed_data d
WHERE
  c.year = d.year
ORDER BY
  c.year
```

7_js_udf.sql

```
-- Cancellation Reason bifurcation in top 5 airports
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

8_js_udf_struct.sql

```
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_24 float64>
  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 > 100000
      order by 2 desc
      limit 5
      ),
    delay_bifurcation as (
      select ORIGIN,
          (case when ARR_DELAY > 1440 then 5
             when ARR_DELAY > 300 then 4
             when ARR_DELAY > 240 then 3
          when ARR_DELAY > 30 then 2
      else 1 end) as slot

  from `airline-delay-canc.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 count
  from 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 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
```

9_js_udf_struct.sql

```
-- Finding delay frequency Overall
-- Delay time bifurcation in 2018
-- Order | Descriptiopn
-- 1 | 1 - 30 min
-- 2 | 30 - 120 min
-- 3 | 2 hr - 5 hr
-- 4 | 5 hr - 24 hr
-- 5 | 24 hr +

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_24 float64>
  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)){
        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 > 100000
      order by 2 desc
      limit 5
      ),
    delay_bifurcation as (
      select ORIGIN,
          (case when ARR_DELAY > 1440 then 5
             when ARR_DELAY > 300 then 4
             when ARR_DELAY > 240 then 3
             when ARR_DELAY > 30 then 2
        else 1 end) as slot

  from `airline-delay-canc.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 count
  from 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
```

10_analytical_func.sql

```
-- Most unreliable month
WITH
 cancelled_count_cte AS (
 SELECT
  *,
  ROW_NUMBER() OVER (ORDER BY cancelled_count) AS RANK
 FROM (
  SELECT
   FORMAT_DATE('%B', FL_DATE) AS month,
   SUM(CANCELLED) AS cancelled_count
  FROM
```

```
      `airline-delay-canc.airlines_data.delay_canc_data`
    WHERE
      EXTRACT(year
      FROM
        FL_DATE) = 2018
    GROUP BY
      1) )
SELECT
  month,
  cancelled_count
FROM
  cancelled_count_cte
ORDER BY
  rank DESC
```

13.2 GITHUB REPOSITORY LINK:

https://github.com/IBM-EPBL/IBM-Project-41129-1660639597

13.3 PROJECT DEMO LINK :

https://drive.google.com/file/d/1ne86Y7fvX38Jt_Wt4scR5SXQFGhoLcU8/view?usp=share_link