AIRLINES-DATA ANALYTICS IN AVIATION INDUSTRY

19CSP14 - PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

Submitted by

TEAM ID: PNT2022TMID22185
VIKAS MAHESH

SATHISH KUMAR.S

RAHUL.R. A

MOHAMMAD SHAFIQ HUSSAIN.S

BACHELOR OF ENGINEERING

in

ELECTRONICS AND COMMUNICATION ENGINEERING

VEL TECH HIGH TECH

Dr. RANGARAJAN Dr. SAKUNTHALA ENGINEERING COLLEGE

An Autonomous Institution

Approved by AICTE-New Delhi, Affiliated to Anna University, Chennai

Accredited by NBA, New Delhi & Accredited by NAAC with “A” Grade & CGPA of 3.27

NOVEMBER 2022

INDEX

1. INTRODUCTION

Data analytics is widely used in every sector in the 21st century. A career in the field of data analytics is highly lucrative in today's times, with its career potential increasing by the day. Out of the many job roles in this field, a data analyst's job role is widely popular globally. A data analyst collects and processes data; he/she analyzes large datasets to derive meaningful insights from raw data.

Most companies are collecting loads of data all the time—but, in its raw form, this data doesn't really mean anything. This is where data analytics comes in. Data analytics is **the process of analyzing raw data in order to draw out meaningful, actionable insights**, which are then used to inform and drive smart business decisions.

A data analyst will extract raw data, organize it, and then analyze it, transforming it from incomprehensible numbers into coherent, intelligible information. Having interpreted the data, the data analyst will then pass on their findings in the form of suggestions or recommendations about what the company's next steps should be.

You can think of data analytics as a form of business intelligence, used to solve specific problems and challenges within an organization. It's all about finding patterns in a dataset which can tell you something useful and relevant about a particular area of the business—how certain customer groups behave, for example, or how employees engage with a particular tool.

Data analytics helps you to make sense of the past and to predict future trends and behaviors; rather than basing your decisions and strategies on guesswork, you're making informed choices based on what the data is telling you. Armed with the insights drawn from the data, businesses and organizations are able to develop a much deeper understanding of their audience, their industry, and their company as a whole—and, as a result, are much better equipped to make decisions and plan ahead.

1.1 PROJECT OVERVIEW

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.

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.

1.2 PURPOSE OF THE PROJECT

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.

Social Impact

• Air transports provide significant economic and social benefits Business Model/Impact

• Ease of trade for businesses having business-to-consumer (B2C) model such as e-commerce

• Air transport is a driver of global trade and e-commerce, allowing globalization of production.

The aviation/airline industry is well-known for being capital-intensive. The acquisition of aircraft itself is cost-intensive, the fleet requires heavy maintenance, and everything about the airline operations – from fuel to airport parking facilities to passengers to staff – all incur heavy expenditure. Clearly, any minute lost while an aircraft is out for maintenance stops it from being a revenue-generating asset.

2. LITERTURE SURVEY

ABSTRACT:

Flight delays in air transportation are a major concern that has adverse effects on the economy, the passengers, and the aviation industry. This matter critically requires an accurate estimation for future flight delays that can be implemented to improve airport operations and customer satisfaction. Thus, we propose an interactive dashboard in which user can register his/her details to book the airplane tickets and predict the delays, if occurs.

1. **Life Data Analysis with Applications for the Airline Industry (Julio Pulido, NortekDana Moore, William Hill -2020)**

• In this study **, Julio Pulido, NortekDana Moore, William Hill** proposed the analysis of non-repairable systems.

• The three techniques, namely the time to failure, stress-strength or condition-based approach, is generally adopted.

• The mixed Weibull distribution (also known as a multimodal Weibull) is used to model data that do not fall on a straight line on a Weibull probability plot.

2. Exploratory Data Analysis on Aviation Dataset
(Saba Firdous, Haseeba Fathiya, Lipsa Sadath -2021)

• In this work , **Saba Firdous, Haseeba Fathiya, Lipsa Sadath** performed Aviation informational collection and performed analytics.

• The first step was to organize all the events into categories depending on their risk level.

• The next step was to use an SVM to learn the relationships between the events.

• The third step was to combine the results from both the models to improve the accuracy of the predictions made.

3. Applying Machine Learning to Aviation Big Data for Flight Delay Predictio (Yushan Jiang,Yushan Jiang -2020) :

• In this study**, Yushan Jiang,Yushan Jiang** developed several machine learning models to predict flight arrival delays.

• Firstly, data pre-processing is needed including the data merging and cleansing.

• Next, data visualization can be performed to extract and visualize the graphic representation of data clearly and efficiently.

• Lastly, prediction models are built and trained using different machine learning methods, then evaluated.

4.Transportation Research Part E IN AVIATION (Xu et al- 2020) :

• In this work, Xu **et al** proposed a hybrid model to forecast statistical indicators in the aviation industry, which employs the seasonal autoregressive integrated moving average (SARIMA) and support vector regression (SVR) methods.

• SARIMA is employed to analyze the raw time series. Gaussian White Noise is then used
for calculation according to the SARIMA's results.

5.Towards a maturity model for big data analytics in airline network planning (Iris Hausladen, Maximilian Schosser -2020)

• In this study, **Iris Hausladen, Maximilian Schosser** address this challenge by developing a maturity model for big data readiness for airline network planning.

The transfer steps have been combined with the model evaluation. In the second stage, the maturity levels are conceptualized and formulated, before the complete model is evaluated by the practitioner group.

2.1 EXISTING PROBLEM

ACCURATE AVIATION DATA

Access critical information regarding aircraft orders, fleet figures, route information and finances of every global commercial airlines. Review complete profiles on airlines, airports and MROs.

PLANNING AND FORECASTING

Analyze and respond to key trends and markets developments affecting the aviation industry today, including pre-built reports and analysis with the most current data.

2.2 REFERRENCES

1. Iris Hausladen, Maximilian **Schosser " Towards a maturity model for big data analytics in airline network planning",** ELSEVIER-Journal of Air Transport Management,Volume 82,(2020).

2. Julio Pulido, NortekDana Moore, William Hill **"Life Data Analysis with Applications for the Airline Industry "**, IEEE XPLORE- **Published in:** 2016 Annual Reliability and Maintainability Symposium (RAMS)

3. Saba Firdous, Haseeba Fathiya, Lipsa Sadath **"Exploratory Data Analysis on Aviation Dataset "**,IEEE XPLORE,Conferences-2021

4. Yushan Jiang,Yushan Jiang **"Applying Machine Learning to Aviation Big Data for Flight Delay Prediction"** IEEE XPLORE, Conferences -2021

5. Xu et al **"Transportation Research Part E IN AVIATION"**, SCIENCE DIRECT Journals and books, Volume 167,(2022).

2.3 PROBLEM STATEMENT DEFINITION

To build a user interface application to analyze the delays so airports organization can adjust and allocate the resources(airports) nearby quickly.

3 IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

The above Empathy Map talks about the Mindset of our team members in completing this project.

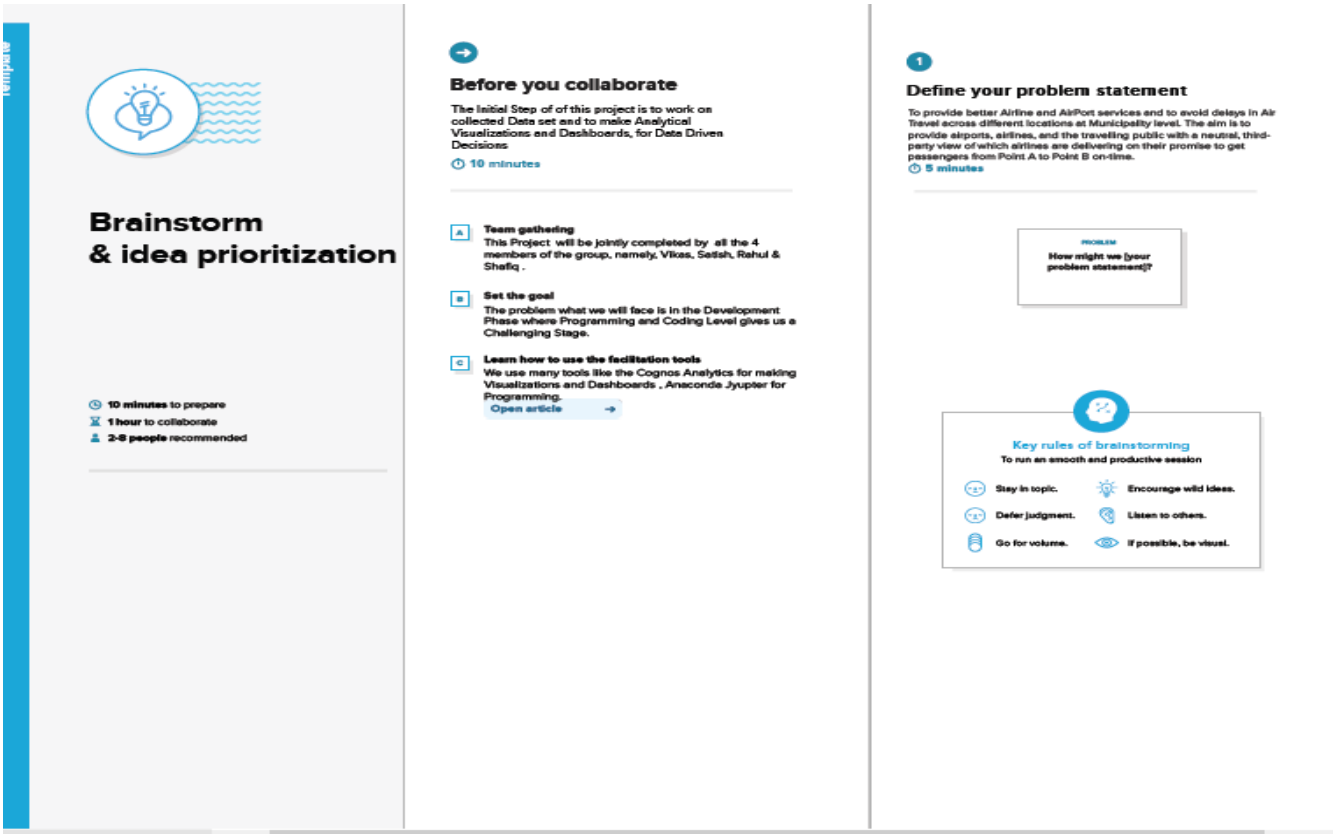
Fig 1 EMPATHY MAP

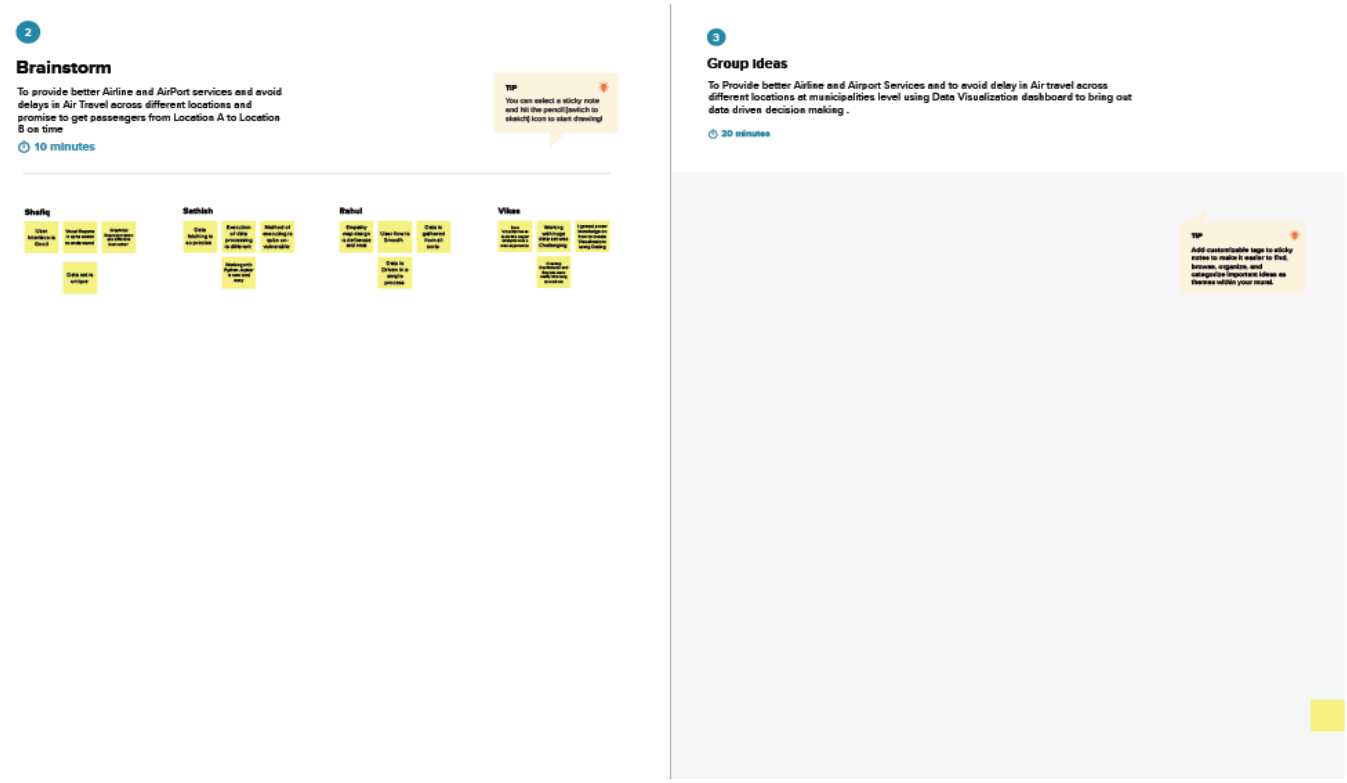
- THE WAY WE HEAR
    - This talks about the importance of travelling in flight and its after effects and pros and cons.
    - This helped to know more how customers hear about Travelling by Air
- WHAT DO THEY THINK AND FEEL
    - This talks about how customers will feel after completing of the object. Also, it add-ons the way we think and our thought process
- WHAT DO THEY SEE
    - It talks about the way we showcase our project and how the customers will create an impact on it
    - It also speaks the way customers will handle the project once its been implemented in the industry

- WHAT DO THEY SAY AND DO
    - This talks about the work wee do and how we deliver it in front of the panel
    - It also talks about the performance of the team

- PAIN
    - This talks of the hardships and challenges that the team had faced during the development of the project.
    - The team faced lots of hardships specially to work programming and coding level
- GAIN
    - The team gain a lot knowledge in the field of Data Analytics
    - Also, the team got a huge exposure new tools like Cognos Analytics and Jyupter Python

3.2 IDEATION AND BRAINSTORMING





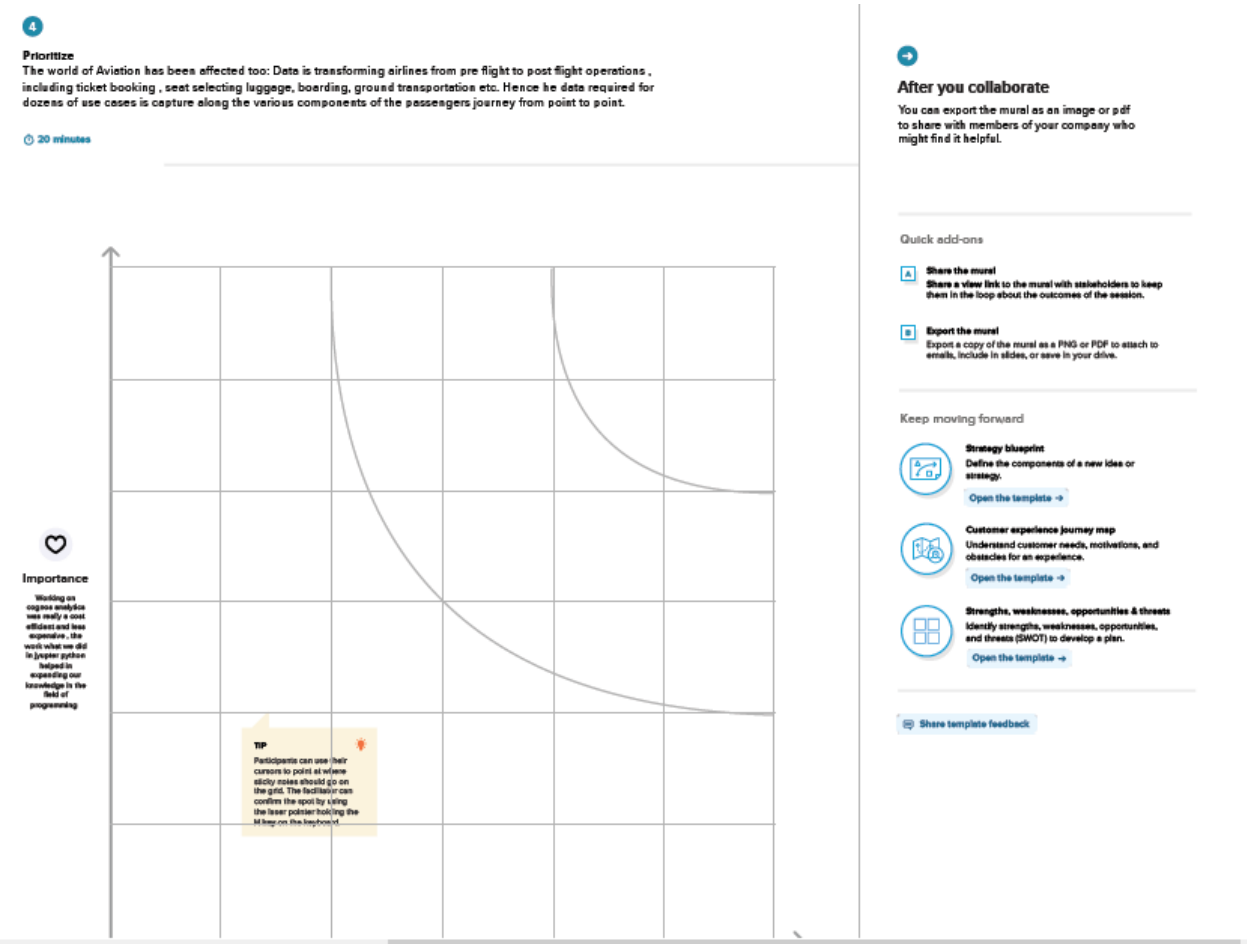


Fig 2 IDEATION BRAINSTORM

- Brainstorming stimulates the building of ideas by a process of association. Embrace the most out-of-the-box notions and build, build, build. Be positive and build on the ideas of others. Brainstorming works well when participants use each other's ideas to trigger their own thinking.
- Ideation is often closely related to the practice of brainstorming, a specific technique that is utilized to generate new ideas. A principal difference between ideation and brainstorming is that ideation is commonly more thought of as being an individual pursuit, while brainstorming is almost always a group activity.

3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	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.	Idea / Solution description	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.	Novelty / Uniqueness	Unique data is given for each chart and dashboards that has been create
4.	Social Impact / Customer Satisfaction	Concerning airlines services is often to help airlines to better understand how the customer views their services compared to their competitors
5.	Business Model (Revenue Model)	From the studies selected for the proposed literature review, the most common data analysis models mainly like the Regression Models and the other Visualization Dashboards and Reports
6.	Scalability of the Solution	As for airport-related studies, some differences also emerge in airlines literature in terms of types of collected data and evaluation scale. In this case, it is even more complex to organize the data collection and choose the time and place for collecting the data, because the interviews for investigating airport services are generally realized at the departure gates of the airport, where passengers are waiting for the flight.

3.4 PROBLEM SOLUTION FIT

CUSTOMER SEGMENTS

Various People who travel by flight, Airlines and Airport Service Authorities are our customers, where we stay in touch with ATC as well.

CUSTOMER CONSTRAINTS

To Understand the given data and try to find the actual issue in it and clean and mine the given dataset and create Attractive and impressive visualizations through charts reports and dashboards. So that the Airline Services can analyze their performance and chart the flights on time without any delay.

TRIGGERS

It makes more attractive to the customers to find the solution easily just by checking the visualizations charts as One Picture talks a lot than Thousand words.

EMOTIONS BEFORE/AFTER

The customers feel very easy to get the idea on which flight to choose for their comfortable journey and according to analysis they can even rate the airlines as well according to their departure and arrival timings.

AVAILABLE SOLUTONS

The current available solution in the industry is from the ATC and the reviews in Flight Booking apps which may fake and false used for marketing. According to Data Analytics the Airlines can make sure that they are providing proper service on time.

CUSTOMER CONSTRAINTS

Main constraint of this project is the Data, without which the project is impossible, so if the customers are providing the actual data without any failure, it makes the project run smoothly

BEHAVIOUR

Customers approach a data analyst with correct amount of data so that the analyst can help the customer by predicting the solution to the problem by Data Driven Decision Making.

CHANNELS OF BEHAVIOUR

Customers look the ratings by which other customers have given and the progress of the performance and feedback of other customers, which all counts in data by which it helps the customers the beet Airline.

Customers can view the bar charts which will be provided in the airport through which the customers can chose their best Airlines according to the performance

PROBLEM ROOT CAUSE

The real problem exists in the past executions of departure and arrivals of the Flights where they don't take off and land on the stipulated given time due to continues follow up by the airlines, as they won't be having a data to act on how to make flights land on perfect timings

YOUR SOLUTION

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.

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration and Creating IBM Cloud Account Registration and Creating IBM Cognos Analytics Account Registration and Creating GitHub Account and Repository
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Downloading Data Set	Data Collection was done in Kaggle.com website
FR-4	Working On IBM Cognos Analytics	Creating Charts, Reports and Dashboards using IBM Cognos Analytics
FR-5	Downloading and Installing Jyupter Python	Jyupter Python was Downloaded and installed from Anaconda for performing programming and codes for data analysis
FR-6	User Registration	Assignments were done in Google Collab; hence an account was created

4.2 NON-FUNCTIONAL REQUIREMENTS

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Usage of IBM Cognos Analytics were really beneficial in creating reports and dashboards. When compared to other tools like Power Bi, this has less features
NFR-2	Security	Quite Securable but not efficient enough and not known in large amounts
NFR-3	Reliability	Moderate Reliable, data exploration has less features, and colors cannot be changed as per the wish of the user
NFR-4	Performance	Performance wise, it’s not so bad for the use of Students, consumes lot of Data and has less features when compared to Power Bi
NFR-5	Availability	Less available for outside users, but one-year free subscription for course enrolled students
NFR-6	Scalability	7/10

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

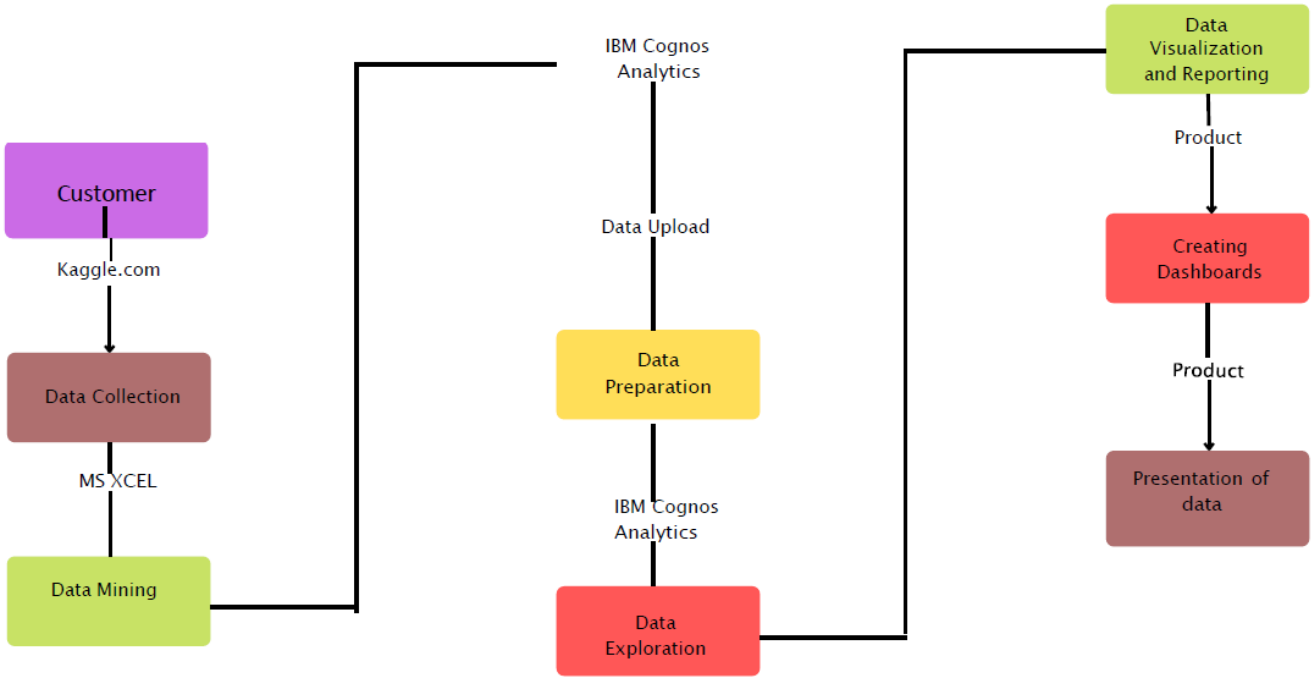


Fig 3 DATA FLOW DIAGRAM

The above figure shows how is the flow of this project in a precise manner. We basically collect the data from the customer through Kaggle.com. From there, we upload those data in Cognos Analytics tool, where we Mine the data by taking the useful data and removing the unwanted data. Later on, we prepare the data in a proper tabular form and with which we create attractive Explorations. These explorations together we create further reports. Keeping those reports on mind we create dashboards for better understanding and thereby we export and present it for the same.

5.2 SOLUTION & TECHNICAL ARCHITECTURE

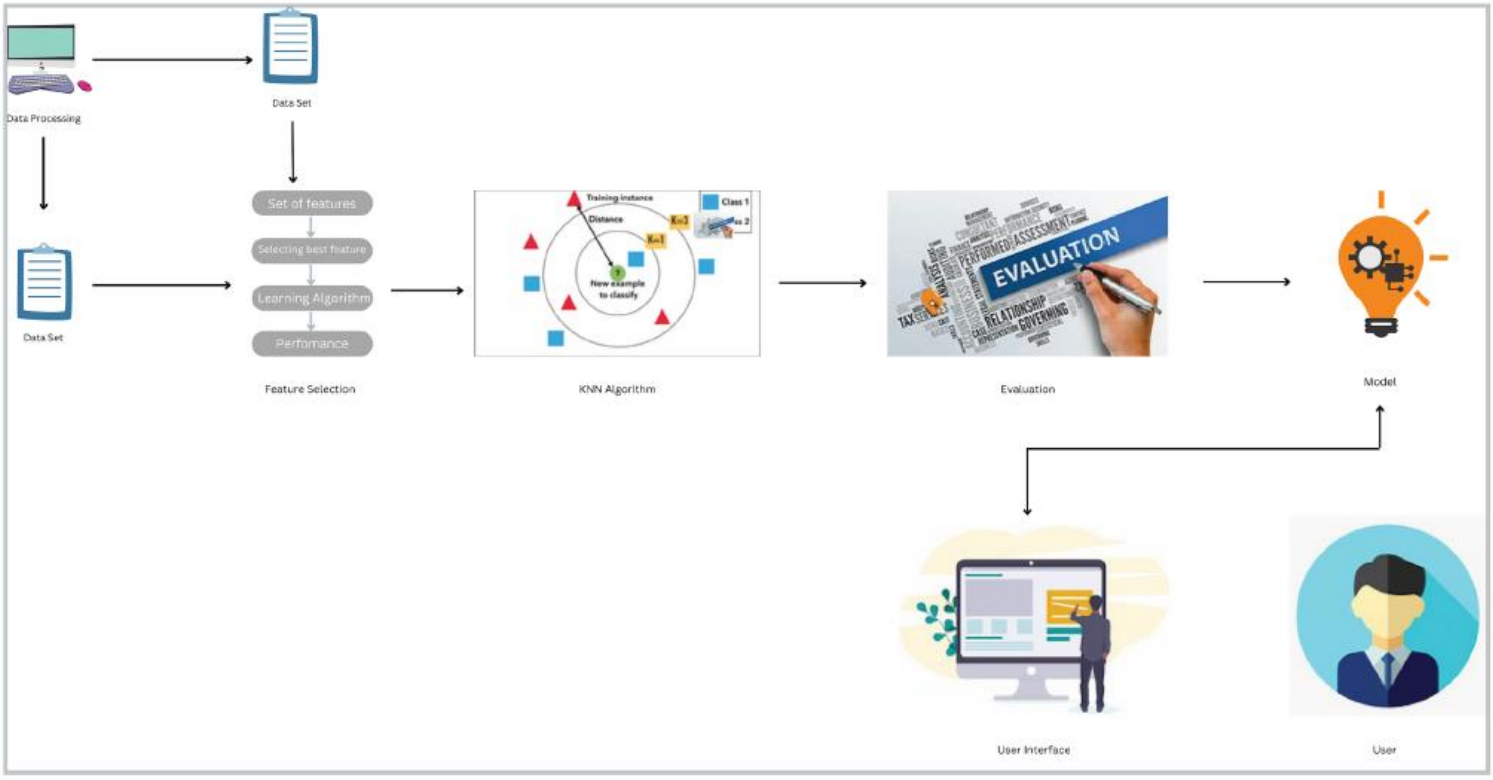


Fig 4 SOLUTION ARCHITECTURE

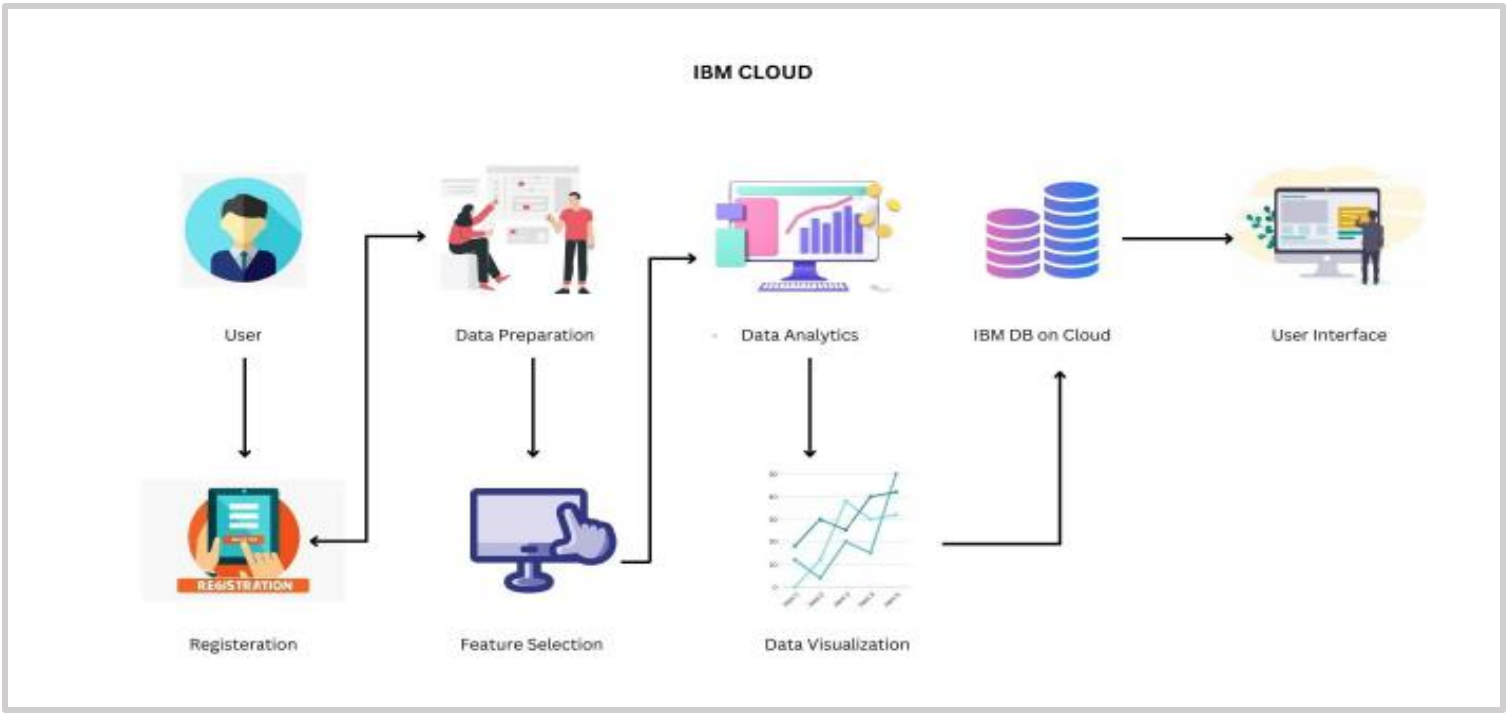


Fig 5 TECHNOLOGY ARCHITECTURE

5.3 USER STORIES

How Airlines Use Data Analytics? | Use Cases

Here are some big data use cases and how they could — and already are, in some contexts — be leveraged to boost revenue.

- **Predictive Maintenance**

Aviation industries can reduce expenses and generate new revenue streams by utilizing Big Data, IoT, and predictive analytics (or any other machine learning technique). Some aircraft manufacturers use this data to perform predictive maintenance or repair the aircraft before it breaks down. About 10% of airline operational expenses go to maintenance, which causes many delays. Therefore, the manufacturer and the carrier will benefit financially by knowing when a part requires refurbishing.

- **Real-time data analytics to boost trip revenue**

Almost 24% of airline expenses constitute fuel prices, and airlines fear the price's impact on their profits yearly. Airlines can better control their fuel expenses due to the Satcom systems' increasing throughput and the development of 5G mobile communications. And here is how it works: the aircraft sends all the data to ground-based flight operations, where an analytics engine process it using its algorithms and sends back to the aircraft a recalculated optimal route.

- **Augmented reality for indoor airport navigation**

Finding your way around an airport's confusing environment is one of the most unpleasant experiences for passengers. Airports are exploring other technologies due to GPS's extreme inaccuracy for indoor navigation. For instance, Bluetooth-based data beacons placed throughout the building or augmented reality (AR).

- **Airline Data Analysis with Python**

This big data analytics project uses Python's SK learn libraries to predict customer satisfaction and aggregate customers. You will start working on this project by importing the Python libraries Pandas, Numpy, Matplotlib, Seaborn, and SK learn. This project involves applying feature selection and one hot encoding to perform dissatisfaction prediction analysis. Additionally, you will use principal component analysis (PCA) for feature transformation and customer clustering.

6. PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING AND ESTMATION

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 by entering my details to work on Cognos Analytics	2	High	Vikas Mahesh Sathish Kumar.S Rahul R. A Mohammad Shafiq Hussain
		USN-3	As a user, I can register for the application through Gmail	2	Medium	Vikas Mahesh Mohammad Shafiq Hussain
	Login	USN-4	As a user, I can Call and request or Approach for dataset	4	High	Rahul.R.A Sathish Kumar.S
	Working with the Dataset	USN-5	To work on the given dataset, Understand the Dataset.	2	High	Vikas Mahesh Sathish Kumar.S Rahul R. A Mohammad Shafiq Hussain
		USN-6	Load the dataset to Cloud platform then Build the required Visualizations.	10	High	Vikas Mahesh Mohammad Shafiq Hussain
Sprint-2	Completion of Tasks for Project Design and Planning Phase	USN-7	Completion of Empathy Map, Literature Survey, Ideation Brain Storm, Proposed Solution, Problem Solution, Solution Architecture, Customer Journey, Functional Requirements, Data Flow Diagrams and Technology Architecture	10	High	Vikas Mahesh Sathish Kumar.S Rahul R. A Mohammad Shafiq Hussain
			Creation of Empathy Map, Brainstorm Ideation	2	High	Rahul.R.A Mohammad Shafiq Hussain
		USN-8	Completion of Literature Survey	2	High	Sathish Kumar.S
		USN-9	Completion of Project Phase 1	5	High	Vikas Mahesh Sathish Kumar.S Mohammad Shafiq Hussain
		USN-10	Completion of Project Phase 2	5	High	Rahul.R.A Vikas Mahesh Mohammad Shafiq Hussain
Sprint-3	Completion of Visualization Charts	USN-11	Using the Airlines Dataset, create various graphs and charts to high light the insights and visualizations. Build a Visualizations and Dashboards to showcase Different Types of Airports in Various Countries and Continents.	4	Medium	Vikas Mahesh
		USN-12	Representation of Flight Count by Categories	4	Medium	Vikas Mahesh
		USN-13	Continent wise Count of Airports using Geo Map	4	Medium	Mohammad Shafiq Hussain
		USN-14	Country wise Airports with Types	4	Medium	Rahul.R.A
	Creating The dashboard and Exporting	USN-15	Create the Dashboard by using the created visualizations.	20	High	Vikas Mahesh
Sprint-4	Developing Programming Codes and Visualizing it through Programming Analysis	USN-16	Creation of Visuals from the data set through Programming	20	High	Vikas Mahesh Satish Kumar.S

6.2 SPRINT DELIVERY SCHEDULE

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

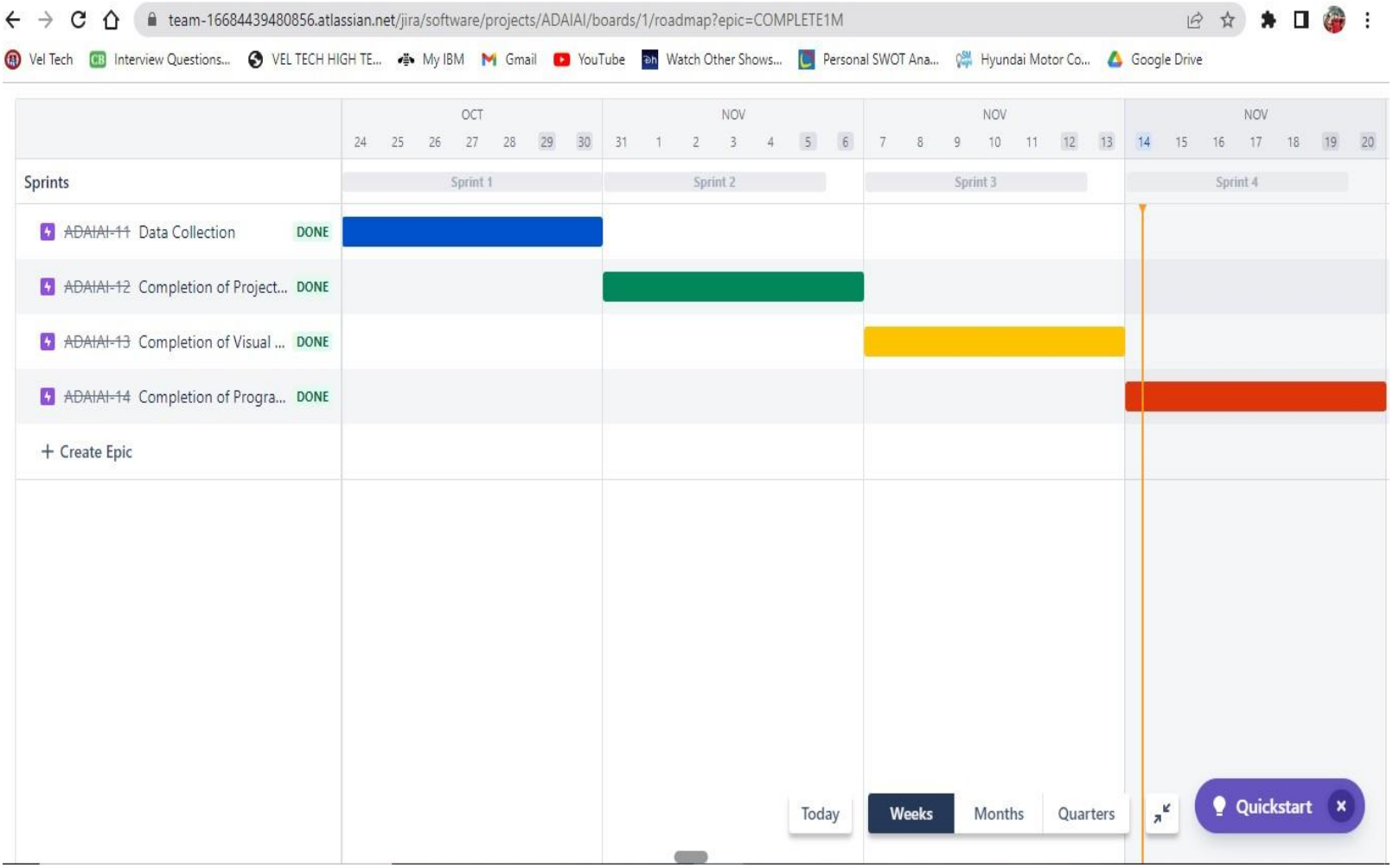


Fig 6 SPRINT DELIVERY SCHEDULE

6.3 REPORTS FROM JIRA

[SPRINT-1] Data Collection Created: 14/Nov/22 Updated: 15/Nov/22 Due: 30/Oct/22 Resolved: 14/Nov/22	
Status:	Done
Project:	AIRLINES DATA ANALYYTICS IN AVIATION INDUSTRY
Components:	KAGGLE.COM
Affects versions:	None
Fix versions:	None

Type:	Epic	**Priority:**	Medium
Reporter:	VIKAS MAHESH	**Assignee:**	rahulink29
Resolution:	Done	**Votes:**	20
Labels:	None		
Remaining Estimate:	Not Specified		
Time Spent:	Not Specified		
Original estimate:	Not Specified		

Attachments:	airports.csv countries.csv regions.csv
Rank:	0|i0001b:
Start date:	24/Oct/22
Issue color:	dark blue

Description

Data has to be collected from Kaggle.com and we had to work on this data to bring out attractive visualizations.

Generated at Tue Nov 15 08:54:17 UTC 2022 by VIKAS MAHESH using Jira 1001.0.0-SNAPSHOT#100210-sha1:28a36363a81be3fec088cc03de57ea0d3b868a26.

Fig 7 REPORT-SPRINT 1

SPRINT-2] Completion of Project Design and Planning Phase Created: 14/Nov/22 Updated: 15/Nov/22 Due: 06/Nov/22 Resolved: 14/Nov/22	
Status:	Done
Project:	AIRLINES DATA ANALYYTICS IN AVIATION INDUSTRY
Components:	MURAL
Affects versions:	None
Fix versions:	None

Type:	Epic	**Priority:**	Medium
Reporter:	VIKAS MAHESH	**Assignee:**	s.mohammadshafiqhussain
Resolution:	Done	**Votes:**	20
Labels:	None		
Remaining Estimate:	Not Specified		
Time Spent:	Not Specified		
Original estimate:	Not Specified		

Attachments:	CUSTOMER JOURNEY.pdf DATA FLOW DIAGRAMS.pdf EMPATHY MAP.pdf IDEATION-BRAINSTORM.pdf LITERATURE SURVEY-AIRLINES DATA ANALYTICS.pdf PROBLEM SOLUTION FIT.pdf PROPOSED SOLUTION.pdf SOLUTION ARCHITECTURE.pdf SOLUTION REQUIREMENTS.pdf TECHNOLOGY ARCHITECTURE.pdf
Rank:	0|i0001j:
Start date:	31/Oct/22
Issue color:	dark_green

Description

Many task were assigned as part of the project that has to be completed on stipulated given time. The task was varying from creating empathy maps to customer journey in different phases. Hence all the given task was able to completed successfully.

Generated at Tue Nov 15 08:54:33 UTC 2022 by VIKAS MAHESH using Jira 1001.0.0-SNAPSHOT#100210-sha1:28a36363a81be3fec088cc03de57ea0d3b868a26.

Fig 8 REPORT SPRINT-2

[SPRINT-3] Completion of Visual Charts Created: 14/Nov/22 Updated: 14/Nov/22 Due: 13/Nov/22 Resolved: 14/Nov/22	
Status:	Done
Project:	AIRLINES DATA ANALYYTICS IN AVIATION INDUSTRY
Components:	IBM COGNOS ANALYTICS
Affects versions:	None
Fix versions:	None

Type:	Epic	**Priority:**	Medium
Reporter:	VIKAS MAHESH	**Assignee:**	VIKAS MAHESH
Resolution:	Done	**Votes:**	20
Labels:	None		
Remaining Estimate:	Not Specified		
Time Spent:	Not Specified		
Original estimate:	Not Specified		

Attachments:	CONTINENT WISE AIRPORT COUNT.pdf COUNTRY WISE AIRPORTS.pdf DASHBOARD.pdf FLIGHT BY CATEGORIES.pdf FLIGHTS BY COUNTRIES, REGIONS AND AIRPORTS.pdf
Rank:	0|i0001r:
Start date:	07/Nov/22
Issue color:	yellow

Description

To complete different types of explorations and dashboards from given data and to bring out data driven decision making from the data and to come out with a conclusion.

Generated at Mon Nov 14 18:55:06 UTC 2022 by VIKAS MAHESH using Jira 1001.0.0-SNAPSHOT#100210-sha1:583150f45e96fe66b2cb2898eb1e9ae5719d8732.

Fig 9 REPORT SPRINT-3

[SPRINT-4] Completion of Programming Codes and Bring out Visualizations Created: 14/Nov/22 Updated: 15/Nov/22 Due: 20/Nov/22 Resolved: 14/Nov/22	
Status:	Done
Project:	AIRLINES DATA ANALYYTICS IN AVIATION INDUSTRY
Components:	JYUPTER PYTHON
Affects versions:	None
Fix versions:	None

Type:	Epic	**Priority:**	Medium
Reporter:	VIKAS MAHESH	**Assignee:**	Sathish kumar
Resolution:	Done	**Votes:**	20
Labels:	None		
Remaining Estimate:	Not Specified		
Time Spent:	Not Specified		
Original estimate:	Not Specified		

Attachments:	FINAL CODE AIRLINES DATA ANALYTICS.ipynb
Rank:	0|i0001z:
Start date:	14/Nov/22
Issue color:	dark_orange

Description

The task was to develop code and python programs to bring out visualization through pure analysis of data and thereby creating attractive visuals for the same.

Generated at Tue Nov 15 08:54:44 UTC 2022 by VIKAS MAHESH using Jira 1001.0.0-SNAPSHOT#100210-sha1:28a36363a81be3fec088cc03de57ea0d3b868a26.

Fig 10 SPRINT- 4

7. CODING AND SOLUTIONING

7.1 FEATURE 1

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In this project, these modules has been imported so that we are able to deliver following features such as Data Importing, Data Cleaning, Data Converting etc…

Data importing

```
[ ] data=pd.read_csv("../input/airstats-data-on-airports-around-the-world/airports.csv")

    print("Lines: ",data.shape[0])
    print("columns: ",data.shape[1])
    data.head()
```

```
Lines:  67312
columns:  18
```

	id	ident	type	name	latitude_deg	longitude_deg	elevation_ft	continent	iso_country	iso_region	municipality	scheduled_service	gps_code	ia
0	6523	00A	heliport	Total Rf Heliport	40.070801	-74.933601	11.0	NaN	US	US-PA	Bensalem	no	00A	
1	323361	00AA	small_airport	Aero B Ranch Airport	38.704022	-101.473911	3435.0	NaN	US	US-KS	Leoti	no	00AA	
2	6524	00AK	small_airport	Lowell Field	59.947733	-151.692524	450.0	NaN	US	US-AK	Anchor Point	no	00AK	
3	6525	00AL	small_airport	Epps Airpark	34.864799	-86.770302	820.0	NaN	US	US-AL	Harvest	no	00AL	

Data cleaning

Drop useless columns and values

```
[ ] data.drop(['wikipedia_link','home_link','keywords'],axis=1,inplace=True)
    countries.drop(['wikipedia_link','keywords'],axis=1,inplace=True)
    regions.drop(['wikipedia_link','keywords'],axis=1,inplace=True)
    data=data[data['continent']!='AN'] # we will remove Antarctica because it has few airports
    countries=countries[countries['continent']!='AN']
    data=data[data['type']!='closed'] # drop closed airports
    regions=regions[regions['local_code']!='U-A'] # drop unassigned airports in regions
```

convert elevation to meters

```
data.rename(columns={'elevation_ft':'elevation_m'},inplace=True) #change attribut name
data['elevation_m']=data['elevation_m']/3.2808 ## convert to meters
data.head()
```

	id	ident	type	name	latitude_deg	longitude_deg	elevation_m	continent	iso_country	iso_region	municipality	scheduled_service	gps_code	iata
0	6523	00A	heliport	Total Rf Heliport	40.070801	-74.933601	3.352841	NoA	US	US-PA	Bensalem	no	00A	
1	323361	00AA	small_airport	Aero B Ranch Airport	38.704022	-101.473911	1047.000732	NoA	US	US-KS	Leoti	no	00AA	
2	6524	00AK	small_airport	Lowell Field	59.947733	-151.692524	137.161668	NoA	US	US-AK	Anchor Point	no	00AK	
3	6525	00AL	small_airport	Epps Airpark	34.864799	-86.770302	249.939039	NoA	US	US-AL	Harvest	no	00AL	
5	322127	00AS	small_airport	Fulton Airport	34.942803	-97.818019	335.284077	NoA	US	US-OK	Alex	no	00AS	

Fig 11 FEATURE 1

7.2 FEATURE 2

```
import plotly.graph_objects as go
import plotly.express as px
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

In this project, we have included the above given modules, which makes the data visualization much easier and effective.

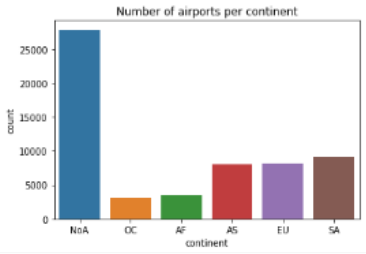
Data visualisation

```
[ ] # as the united states has too many airports, we will separate it in US_airports
    US_airports=data[data['iso_country']=='US']
    US_airports.head()
```

ident	id	type	name	latitude_deg	longitude_deg	elevation_m	continent	iso_country	iso_region	municipality	scheduled_service	gps_code	iata_code	local_code
00A	6523	heliport	Total Rf Heliport	40.070801	-74.933601	3.352841	NoA	US	US-PA	Bensalem	no	00A	NaN	00A
00AA	323361	small_airport	Aero B Ranch Airport	38.704022	-101.473911	1047.000732	NoA	US	US-KS	Leoti	no	00AA	NaN	00AA
00AK	6524	small_airport	Lowell Field	59.947733	-151.692524	137.161668	NoA	US	US-AK	Anchor Point	no	00AK	NaN	00AK
00AL	6525	small_airport	Epps Airpark	34.864799	-86.770302	249.939039	NoA	US	US-AL	Harvest	no	00AL	NaN	00AL
00AS	322127	small_airport	Fulton Airport	34.942803	-97.818019	335.284077	NoA	US	US-OK	Alex	no	00AS	NaN	00AS

```
sns.countplot(x='continent',data=data).set_title('Number of airports per continent')
```

```
Text(0.5, 1.0, 'Number of airports per continent')
```



For US airports

```
## plot the number of airports per region in the US
US_airports.loc[:,"iso_region"]=US_airports.loc[:,"iso_region"].replace({'US-':''}, regex=True)
## set figure size
sns.set(rc={'figure.figsize':(30,8)})
## seaborn title
sns.countplot(x='iso_region',data=US_airports).set_title('Number of airports per region in the USA')
```

```
/opt/conda/lib/python3.7/site-packages/pandas/core/indexing.py:1773: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  self._setitem_single_column(ilocs[0], value, pi)
Text(0.5, 1.0, 'Number of airports per region in the USA')
```

Fig 12 FEATURE 2

8. TESTING

8.1 TEST CASES

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visulizations / Graphs - 7
2.	Data Responsiveness	Highly Responsive
3.	Amount Data to Rendered (DB2 Metrics)	66000 was rendered from which Maximum of 120 data was filtered.
4.	Utilization of Data Filters	NIL
5.	Effective User Story	No of Scene Added - NIL
6.	Descriptive Reports	No of Visulizations / Graphs – 12

8.2 USER ACCEPTENCE TESTING

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	3	3	20
Duplicate	0	2	0	3	5
External	2	3	0	1	6
Fixed	0	2	4	10	16
Not Reproduced	0	0	0	2	1
Skipped	0	0	0	2	2
Won't Fix	0	0	0	1	1
Totals	12	11	7	14	55

3. Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	20	0	0	20
Client Application	40	0	0	40
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS

9.1 PERFORMANCE METRICS

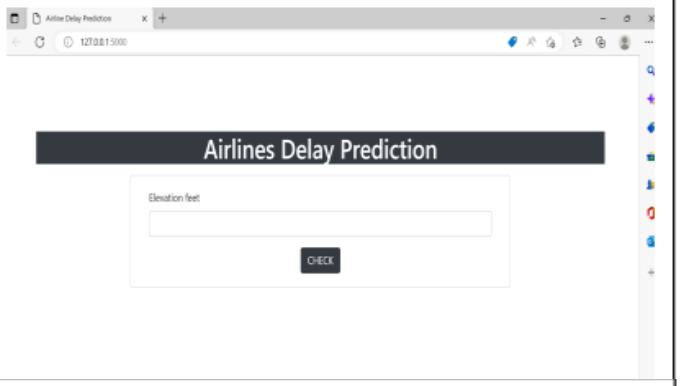
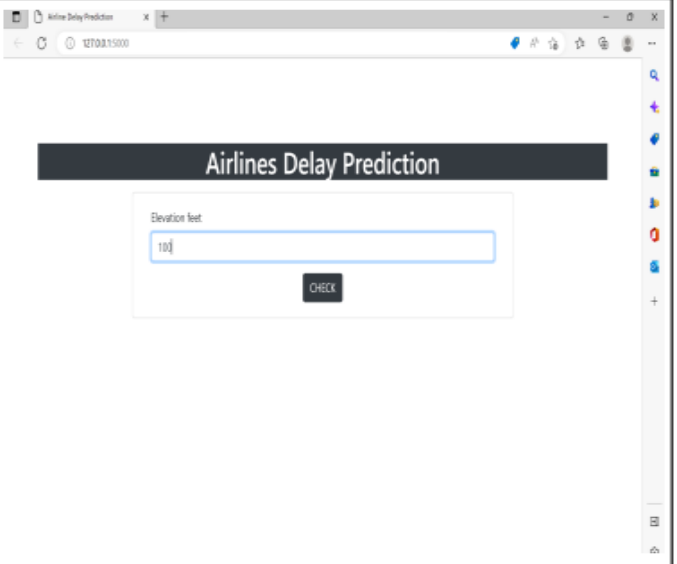
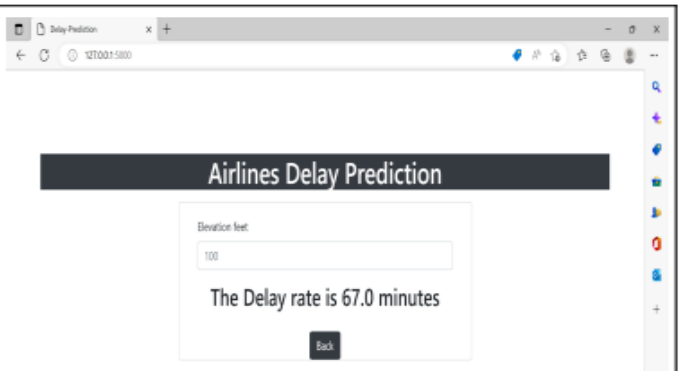
S.No	Parameter	Screenshot
1	**Dashboard design**	
2	**Data entry**	
3	**Data responsiveness**	
4	**Descriptive report of delay analysis**	

Fig 13 PERFORMANCE METRICS

10. ADVANTAGES AND DISADVANTAGES

ADVANTAGES	DISADVANTAGES
Analytics can help aviation players save significantly on costs by allowing them to address failure mechanisms proactively.	There is a lack of alignment between different teams or departments within an organization.
With the right information, MROs can minimize the risks associated with overstocking or stock outs by planning their inventory wisely.	Analytics solutions are not difficult to implement, however, they are costly, and the ROI is not immediate.
With reduced AOG (aircraft on ground) events, the airworthiness of the aircraft increases.	One of the biggest limitations of data analytics is lack of access to quality data.
Analytics makes it possible to enhance fleet reliability, ultimately trickling down into reduced delays and cancellations for passengers.	Sometimes, data collection might breach the privacy of the customers as their information such as purchases, online transactions, and subscriptions are available to companies whose services they are using.
It ensures higher safety for passengers by reducing the risk of safety incidents.	Some of the analytics tools developed by companies are more like a black box model. What is inside the black box is not clear or the logic the system uses to learn from data and create a model is not readily evident.
Analytics helps businesses achieve true collaboration by bringing together different stakeholders onto a single platform allowing them to “talk” to each other in real-time. Analytics is instrumental in ensuring greater airworthiness.	Organizations need to be cautious of what sort of data they are collecting from customers and ensure the security and confidentiality of the data. Only the data required for the analysis needs to be captured and if there is sensitive data, it needs to be anonymized so that sensitive data is protected. Data breaches can cause customers to lose trust in the organizations which may result in a negative impact on the organization.

11. CONCLUSION

The aim of this review was to provide an exhaustive summary of the most relevant studies published in the last ten years in the field of the evaluation of air transport service quality on the basis of passengers' perceptions. We selected a series of papers published in the most important journals of the transportation sector, divided between studies investigating services managed by the airport companies and studies analyzing services managed by airlines. The literature review was structured by three main criteria: the service attributes analyzed in the various studies; the methods adopted for collecting the data; the methods used for analyzing the data. The reason why we selected these criteria is linked to the aim to create a picture of the studies by providing the most important information for researchers and practitioners, which are just the analyzed service aspects and the methodologies adopted for discovering the most relevant ones. From our literature review study, it can be concluded that there is a large variety of methods both for collecting and analyzing data, even if some of them were adopted by several researchers, whereas others were adopted only in few cases (e.g., SEM as data analysis method). Anyway, the suitability of each method depends on the objectives of the study, as well as other practical aspects, linked to the types of available data or the opportunity to collect certain data rather than others.

Definitively, we retain that it is important to investigate much more on the issue of air transport service quality, which is an emerging sector in the public transport service quality analysis. In fact, although in the last 20 years the research works in the field of air transport service quality have become increasingly numerous, this topic is still largely unexplored and requires a thorough investigation and further developments.

12. FUTURE SCOPE

Living in the 21st century, you might have often come across the word ‘data analytics. Currently, it is one of the most buzzing terminologies. Companies around the globe generate vast volumes of data daily, in the form of log files, web servers, transactional data, and various customer-related data. In addition to this, social media websites also generate enormous amounts of data.

Companies ideally need to use all of their generated data to derive value out of it and make impactful business decisions. Data analytics is used to drive this purpose. Data analytics is the process of exploring and analyzing large datasets to find hidden patterns, unseen trends, discover correlations, and derive valuable insights to make business predictions.

It improves the speed and efficiency of your business. Businesses use many modern tools and technologies to perform data analytics. This is data analytics for beginners, in a nutshell.

According to a survey by Oliver Wyman, the global fleet of commercial aircraft could generate 98 million terabytes of data per year by 2026. In-flight recorders, operation systems, and staff all generate a vast amount of data. However, such data is no good if users cannot access it in a timely manner or are unable to use it to derive meaningful insights.

Apart from the data from the Flight Data Recorder, Engine Health Management data or Airplane Health Management data, maintenance-related data of the components/ aircraft recorded in their MRO platforms provides another data stream for predictive maintenance – for instance, non-routines, removals, no fault found and minimum equipment list occurrences, operator maintenance program and part reliability. Analyzing both the operational data from the sensors and the MRO data with the right statistical tools will be the key to a high accuracy prediction.

Three types of analytics can help MROs deal with the problem of unpredictability in the area of maintenance and repair of their fleet –

1. Descriptive and diagnostic analytics

Descriptive and diagnostic analytics helps users analyze information based on past events. They answer questions like “What happened?” Alerts, notifications, and standard reports are some key features of descriptive analytics.

2. Predictive analytics

Predictive models/analytics is more sophisticated and lets users know why something happened and what is likely to happen next. Using predictive analytics and modelling, users can study emerging trends to predict where they are headed. The effectiveness of predictive maintenance is in its ability to leverage the historical data alongside the live operational data to make predictions. This is purely aided by the latest developments on the processing of a high volume of dynamic data feeds and analyzing with sophisticated statistical tools.

3. Prescriptive analytics

Prescriptive analytics lets users know what will happen in the future as well as the possible impact it could have on the business. It also prescribes the best next actions to take in order to minimize adverse impacts.

13. APPENDIX

SOURCE CODE

https://colab.research.google.com/drive/1RHjqkWnU2vUYQa_qRjpjVw_YxBFsbs2f?usp=sharing

GIT HUB LINK

https://github.com/IBM-EPBL/IBM-Project-1747-1658411332

DEMO LINK

https://youtu.be/OxtWjzDDIqM