AIRLINES DATA ANALYTICS FOR AVIATION INDUSTRY

THE PROJECT REPORT
Submitted by

TEAM ID

PNT2022TMID22723

TEAM MEMBERS
Sathyasri J (732919CSR095)
Ranjana P(732919CSR083)
Yukesh S S (732919CSL003)
Surya M (732919CSR112)

	6.1 SPRINT PLANNING & ESTIMATION
	6.2 SPRINT DELIVERY SCHEDULE
7	**RESULTS**
8	**ADVANTAGES & DISADVANTAGES**
9	**CONCLUSION**
10	**FUTURE SCOPE**
11	**APPENDIX**

1. INTRODUCTION :

An Airport has a huge amount of data related to number of flights, data and time of arrival and dispatch, flight routes, No. of airports operating in each country, list of active airlines in each country. The problem they faced till now it's, they have the ability to analyze limited data from databases. The Proposed model intention is to develop a model for the airline data to provide a platform for new analytics based on the following queries.Data analyst can be used to predict future glitches, prevent them from happening, and make the maintenance procedures more accurate and thorough. As a result, it is possible to lower costs related to maintaining an aircraft. The airport codes may refer to either the IATA airport code, a three-letter code that is used in passenger reservation, ticketing and baggage handling systems, or the ICAO airport code which is a four-letter code used by ATC systems and for airports that do not have an IATA airport code.In this project based on the customer reviews and flight arrival timing and cost the best flight is determined.

1.1 PROJECT OVERVIEW :

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

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

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

1.2 PURPOSE :

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

- The aim is to provide airports, airlines, and the traveling 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.

- Based on the third party review of the customer ,the best flight which covers the destination in a short time will be decided.

2. LITERATURE SURVEY

2.1 EXISTING SOLUTION :

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

AUTHORS : Leonardo Carvalho, Alice Stenberg, Leandro maia gonçalves,
Ana Beatriz cruz, Jorge A,soares

YEAR : 2018

DESCRIPTION :

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

PROS :

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

CONS :

- Due to weather predicting this is difficult.

- A few factors responsible for the flight delays like runway construction and excessive traffic are rare, but bad weather seems to be a common cause.

TITLE : Aviation management

AUTHORS : Shi Qiang Liu, Andrea D'Ariano, Erhan Kozan, Mahmoud Masoud CARRS-Q, SaiHo Chung.

YEAR : 2019

DESCRIPTION :

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

PROS :

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

CONS :

- Air traffic control operations and related issues are still scheduled by human controllers.

- Ignore any military/defense use of drones.

TITLE : Predictive Analytics Platform for Airline Industry.

AUTHORS : P. H. K Tissera, A.M.R.S.P. llwana, K.T. Waduge, M.A.l.

Perera, D.P. Nawinna, D. Kasthurirathna.

YEAR : 2020

DESCRIPTION :

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

PROS :

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

CONS :

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

2.2 PROBLEM STATEMENT DEFINITION :

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

3 . IDEATION & PROPOSED SOLUTION :

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

Nowadays, the aviation industry plays a crucial role in the world's transportation sector, and a lot of businesses rely on various airlines to connect them with other parts of the world. But, extreme weather conditions may directly affect the airline services by means of flight delays. Ultimate benefits of big data analytics include timely responses to current and future market demands, improved planning and strategically aligned decision making, as well as crystal clear comprehension and monitoring of all main performance drivers relevant to the airline industry. Data mining produces insights around the decisions for adding or subtracting the flights

to the routes where more or less passenger movement is found.The purpose of this project is to look at the approaches used to build models for predicting flight delays that occur due to bad weather conditions. Based on the customer review and other data the delay of the flight is calculated then compared with other flights the best flight with shortest time delay will be delivered.

3.1 EMPATHY MAP CANVAS :

An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment. The empathy map was originally created by Dave Gray and has gained much popularity within the agile community.

3.2 BRAINSTORM :

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

Brainstorm

Write down any ideas that come to mind that address your problem statement.

🕐 10 minutes

TIP
You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

Sathyasri

Summary Cards	Airports and no of muncipality	Waterfall Chart
no of airports by continent	flight count by categories	data visualization
pie chart continent vise	packed buble cart	continent filter

Ranjana

understanding data set	latitude_deg	airstats data in airport
airport performance report	resources coverage data	load the data set
perform join of data set table	IBM	multiple analytical graphic

Yukesh

flight type filter	coloumn chart	exploration of data
understand the descriptive statistics	data preparation	calculated field
data set	IBM cognos	data stored in spreadsheet

Surya

fundamental concepts of IBM cognos	create meaning full dashboard	third party view
create meaning full virtualization	planning phase	promise the passenger from point to point on time
development	provide airline airport	traveling public with a metal

Job Recommentations

Based on user search criteria	based on experience or skills
Based on users salary requirements	suitable for experience and age

Resume

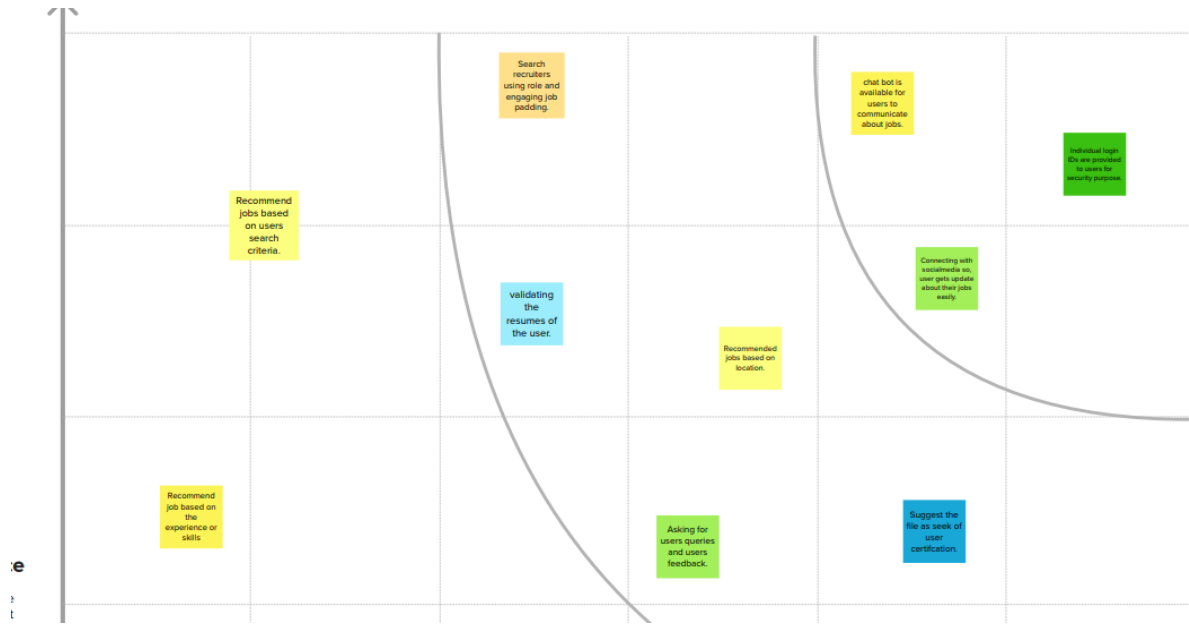|	
Validating the resumes of the user	provide companys specifc resume builder
	automatic discards of application if criteria doesn't meet

Social

Notifications of the job vacancies will reach quickly	Suggestions of job will be given to the users up to-date through social connectives
	Connected with social media so that users will get update easily

TIP
Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.

Security

indiviucidual login id's are provided to user for security purpose	checking the false id information
can track who viewed our profle	

Feedback

Asking users for queries and feedback	Suggest the skills which user need to improve
Displaying company's rating	

Other Information

discents of application if doesn't need	Candidates can take mock test as test their ability
User can apply for multi jobs by single registration	User can apply for multi jobs by single registration
Connected with recruitment companies as they can announce vacancies	

Chatbot

Chatbot is avaliable for users to communicate about jobs

Refinement

Suggest the fle as seek of user certifcation	Search the recruiters using stats like engaging jobs and visitings

3.3 PROPOSED SOLUTION :

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To provide better airlines solutions and avoid flight delays during air travel across different regions.
2.	Idea / Solution description	Displaying the current flight location in the dashboard
3.	Novelty / Uniqueness	Unique Visualization of data from different datasets and graphical representation
4.	Social Impact / Customer Satisfaction	Data analytics helps the industry to understand customers' preferences and other maintenance issues.
5.	Business Model (Revenue Model)	Creating a application in a subscription based model
6.	Scalability of the Solution	Size and number of the data on the datasets can be large and sometimes very hard to visualize.

3.4 PROBLEM SOLUTION FIT :

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

<table>
<tr>
<td rowspan="2">Define CS, fit into CC</td>
<td>

1. CUSTOMER SEGMENT(S) `CS`

Who is your customer?
It is difficult to keep track of forecasting data and planes' arrivals and departures for airline and airport customers. Airlines bear significant costs as a result of delays and cancellations, which include maintenance expenses and compensation to passengers stranded in airports. Predictive analytics applied to fleet technical support is a reasonable solution to nearly 30 percent of total delay time caused by unplanned maintenance.

</td>
<td>

6. CUSTOMER CONSTRAINTS `CC`

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.Since the consumer experience in the airline business is frequently described as a customer's perceptions and responses as he or she travels through the various departure stages and arrives at an airport, it is crucial to connect with customers mid-flight and understand their in-flight requirements. The post-landing phase is a great chance to interact with passengers and listen to their opinions. In addition to seating comfort and crew decorum, start with the basics, such as seating comfort and crew etiquette. That's a terrific way to boost your online reputation, post-flight.

</td>
<td>

5. AVAILABLE SOLUTIONS `AS`

Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital note taking Flight Turnaround Analytics: Using video monitoring for ground activities, the video annotation service helps to capture process inefficiencies in a flight turnover. Using video monitoring for ground activities, process inefficiencies in a flight turnover are captured. Planning and Schedule Analytics: It provides information on how much revenue an airline makes on a specific route and the amount of money spent on fuel and personnel. It is used to rebalance aircraft fleets, estimate fuel needs, and plan crew rosters.

</td>
<td rowspan="2">Explore AS, differentiate</td>
</tr>
<tr>
<td>

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

Which jobs-to-be-done (or problems) do you address for your customers?
There could be more than one; explore different sides.Using proprietary software like Airmax, or simple tools like Microsoft Excel, you will collect information about important performance indicators (KPIs) such as flight operations and inventory. As an example, you will use statistics to optimise flight operations. You will use quantitative data analysis to identify trends and bottlenecks, and then advise your management on them so they can take the necessary action.

</td>
<td>

9. PROBLEM ROOT CAUSE `RC`

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

The purpose of conducting a root cause analysis is to identify the causal factors that trigger substandard safety performance within an event, whether it be an accident, a minor incident, or a close call. Your aviation SMS manual defines risk management processes.

</td>
<td>

7. BEHAVIOUR `BE`

What does your customer do to address the problem and get the job done?
Airport data analysts can gather information about passengers as they go through various checkpoints, such as whether they are male or female, when they arrived, and if they checked their baggage, in order to better understand passenger behaviour. This understanding can be used to improve the service.

</td>
</tr>
</table>

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

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

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

| FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Customer Registration	Customers can register using their Gmail account.
FR-2	User Confirmation	The consumer will receive mail confirmation following registration.
FR-3	Visualizing data	Using IBM cognos Analytics, a user can see the regular trends in flight delay.
FR-4	Generating Report	Viewing the flight delay report is possible.

4.2 NON - FUNCTIONAL REQUIREMENT

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

| FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The programme will have an easy-to-use graphical user interface. All the elements of the application would be simple for users to comprehend and utilize. Any activity must be carried out in a matter of clicks.
NFR-2	**Security**	Since user accounts are the main target of security concerns, adequate login procedures should be followed to prevent hacking. The system should not make public user personal information or other organization information.
NFR-3	**Reliability**	The system should save all user processes made up to the point of abnormal occurrences when it disconnects or freezes as a result of excessive simultaneous access.
NFR-4	**Performance**	The system need to require some speed, especially when navigating the catalogue.
NFR-5	**Availability**	The system must be accessible every day of the week, 24 hours a day. Access is available at any time.

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

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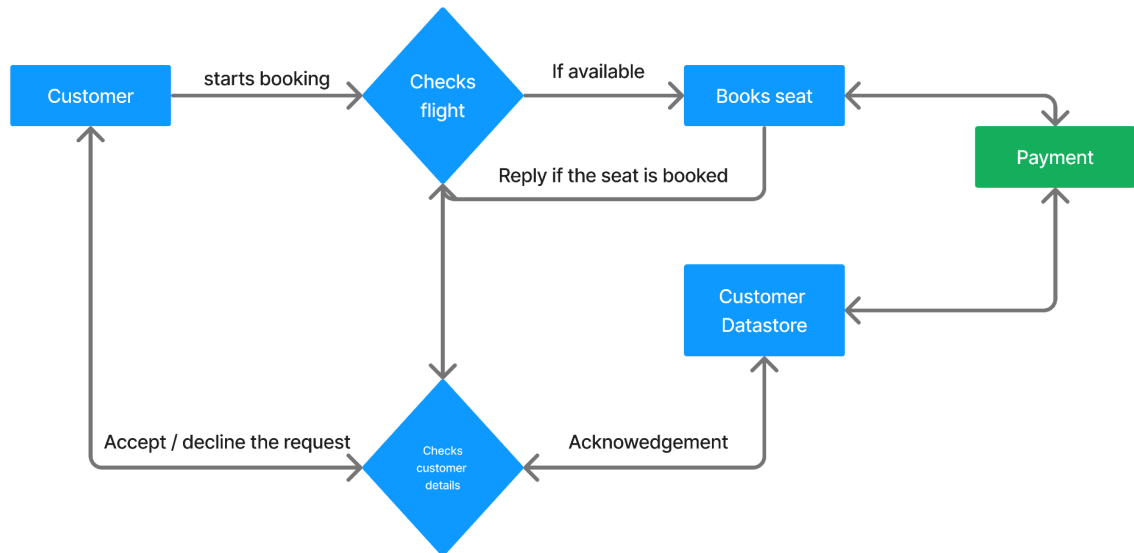
5.2 SOLUTION & TECHNICAL ARCHITECTURE

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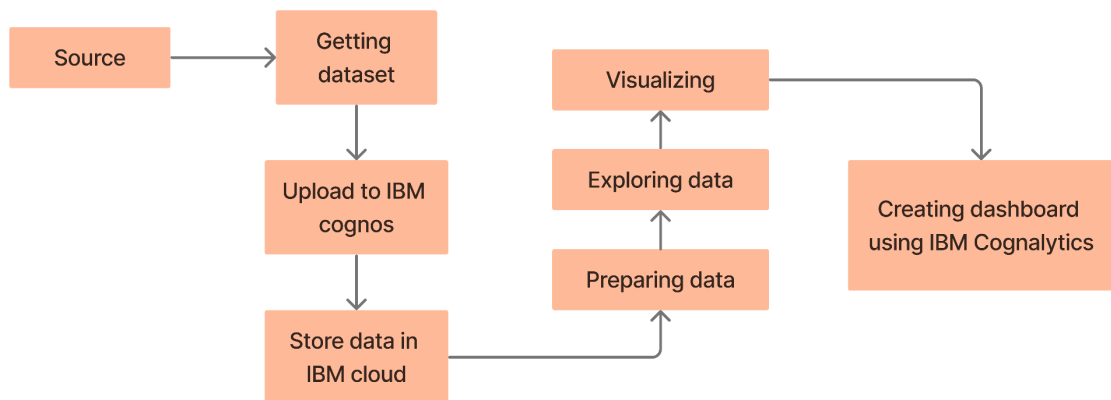
5.3 USER STORIES

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 (Web 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	User will receive confirmation email once he have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	User can register for the application through Gmail.		Medium	Sprint-1
	Login	USN-4	User can log into the application by entering email & password.	I can get to access my web portal	High	Sprint-1
	Dashboard	USN-5	User can get to know what my dashboard consists of.	I can my details of my registration.	Low	Sprint-2
Customer Care Executive	Organization	USN-6	• Consumers will have the ability to contact the company that owns this aeroplane analysis system if they have any issues with the system for interacting with customers or if there are any problems with the aeroplane itself, such as delays or landing in an unexpected place.	The customer care workers will help out the customers in trouble.	High	Sprint-1
Administrator	Administration	USN-7	Organization takes in-charge of the administrative policies of different departments like: • registration • flight booking • delay visualization • generation of delay report	As an administrator, confirmation of user while registration is done.	High	Sprint-1

6. PROJECT PLANNING & SCHEDULING
6.1 SPRINT PLANNING & ESTIMATION

Product Backlog, Sprint Schedule, and Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint1	Data Preprocessing and Exploratory Data Analysis(EDA)	USN-1	Data cleaning is implemented to check whether, there are any null values or any outliers are found	10	Medium	SATHYASRI J RANJANA P YUKESH S SURYA M
		USN-2	Testing and Training the data model is implemented using Jupyter notebook	10	High	SATHYASRI J RANJANA P YUKESH S SURYA M

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint2	Working with dataset	USN-3	Working with the Dataset. Understanding the Dataset Loading the Dataset Exploring the dataset Visualize the Data.	20	Low	SATHYASRI J RANJANA P YUKESH S SURYA M
Sprint3	Data Visualization	USN-4	We plan to create various graphs and charts to highlight the insights and visualizations with the given attributes	20	Medium	SATHYASRI J RANJANA P YUKESH S SURYA M
Sprint4	Dashboard	USN-5	Dashboard Showing Different Types Of Visuals	15	High	SATHYASRI J RANJANA P YUKESH S SURYA M

Project Tracker, Velocity & Burndown Chart:

| 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	02 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	09 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	16 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)
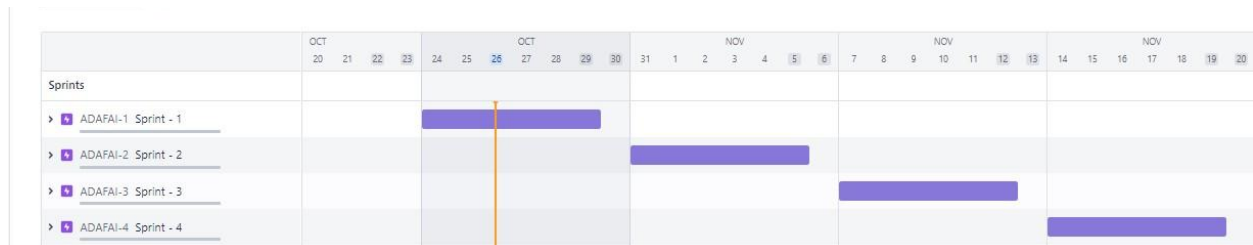
$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Burndown Chart:

A burndown 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.
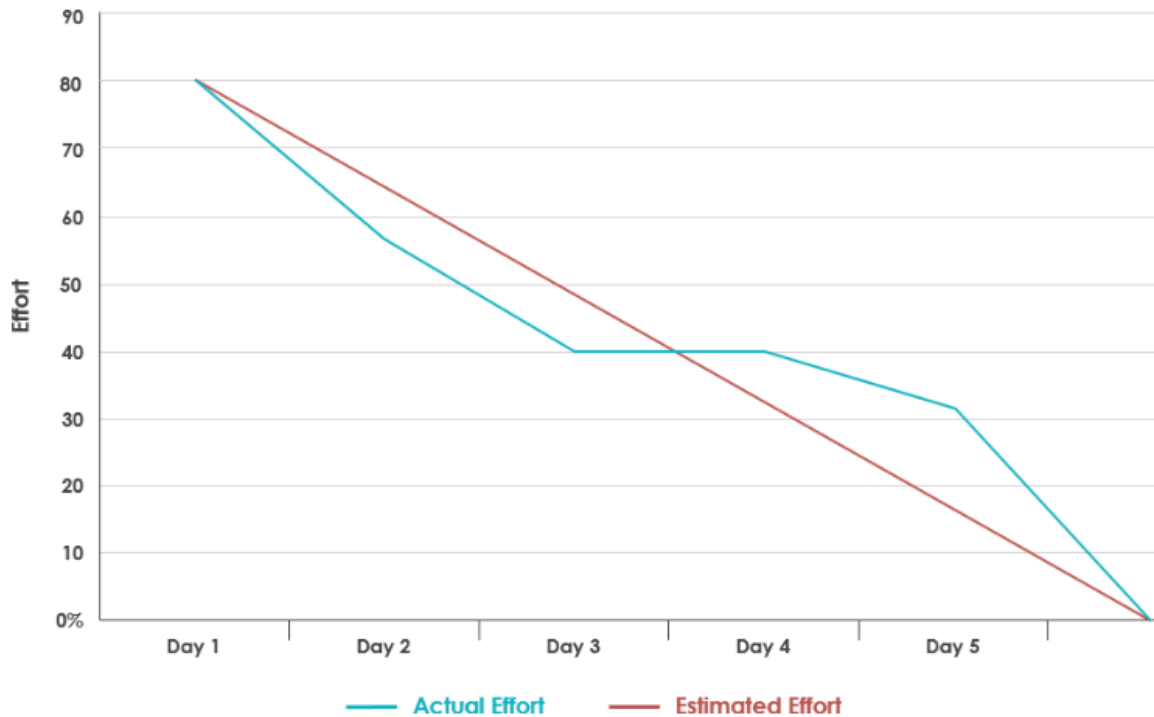
6.2 SPRINT DELIVERY SCHEDULE

7. RESULTS

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

destination in short time when compared to the other flight is calculated easily with more accuracy.The accuracy rate is higher by using data analytics.

8. ADVANTAGES & DISADVANTAGES
ADVANTAGES
- The advantages include being able to fly to almost any destination in the world and having a variety of different aircraft for different purposes, and cut down on travel time.
- There is less need for heavy packaging - Air exports, in general, entail less hard packaging than ocean shipments. This ensures you save both time and money by not having to provide extra packaging services.
- Fast Service - Air transportation offers convenient, reliable and fast services of transport. It is considered the cheapest way to ship peregrinate goods. It offers a standard, convenient, reliable and fast service.

DISADVANTAGES
- Cost - Air travel is considered to be the most expensive means of transportation. The cost of maintaining aircraft is higher and the costs for the building of aerodromes and avions are much higher. That's why air travel is so expensive that it gets beyond ordinary people's grasp.
- Risky - Air travel is the riskiest mode of transport, since there can be considerable losses to goods, customers and crews as a result of a minor crash. Compared to other means of travel, the risks of collisions are higher.
- Accident-prone - Compared to other modes air travel is always at high risk of accidents. There are more accidents on count while traveling by air transport. The reason can be bad weather, signal issues or machine parts failure which causes loss of people, crew or goods.

9. CONCLUSION

Customer experience is always at the top of the priority list for airlines. Customers that are dissatisfied or disengaged inevitably result in fewer passengers and less money. It is critical that clients have a positive experience every time they travel.Looking at the bright prospects of the aviation industry, it makes sense to invest in airline stocks as they are likely to benefit from the government's push to make the aviation industry a bulwark of the transportation industry in India. From this project we conclude that ,The usage of big data analytics is booming today, with its ability to be used to draw useful insights from past data research. Its uses in the aviation industry have a wide array of applications ranging from predicting flight delays to detecting faults in airplane parts. In this paper, we conducted exploratory data analysis on flight dataset to draw inferences on arrival and departure delays and to identify relationships between flight timings and delays. Using the flight delay data, we identified which flight is mostly prone to delays. The arrived upon conclusions are useful for selecting flights in the future.from the review of the customer and the flight which covers the destination in correct time and in shortest time that airline flight will be selected as a best airline service.

10. FUTURE SCOPE

With the growth of data, the use of analytics in the airline industry is the next big wave. The ultimate benefits of big data analytics include timely responses to current and future market demands, improved planning and strategically aligned decision making, as well as crystal clear comprehension and monitoring of all main performance drivers relevant to the airline industry. In future this project has been developed with some extra features.The customer can give query for any dissatisfaction that query will be solved review of the customer will be collected.Then if a customer want to change the destination in a midway they can give one alert message to the service and that nearby destination will be given for the customer.

11. APPENDIX

GITHUB LINK
https://github.com/IBM-EPBL/IBM-Project-33022-1660213837
PROJECT DEMO LINK
https://drive.google.com/file/d/1OOWQUCnXlM7FDUpIBDhOEqFR_Z65-sJR/view?usp=share_link