IBM REPORT

| | |
|---|---|
| Date | 19 November 2022 |
| Team Id | PNT2022MID12865 |
| Project Name | Airline Data Analytics for Aviation Industry |
| Team Members | Narain Muthiah,Poornima,Prawin Kumar,Swethadarshni |

# 1.INTRODUCTION

## 1.1 Project Overview

To know fundamental concepts and can work on IBM Cognos Analytics. Gain a broad understanding of plotting different visualizations to provide suitable solution. Able to create meaningful Visualizations and Dashboard(s).

## 1.2 Purpose

Technology is drastically changing the way businesses connect with their customers, take business decisions, and build workflows. No doubt, the world of aviation has been affected too: data is transforming airlines from pre-flight to post-flight operations, including ticket purchase, seat selection, luggage, boarding, ground transportation, etc.

# 2.LITERATURE SURVEY

## 2.1 Existing Problem

## 2.2 References

1. Paper : Data Analytics for Air Travel Data: A Survey and New Perspectives.

Author : Haiman Tian, Maria Presa-Reyes, Yudong Tao, Tianyi Wang, Samira Pouyanfar, Alonso Miguel, Steven Luis, Mei-Ling Shyu, Shu-Ching Chen, Sundaraja Sitharama Iyengar

Description: The airline industry has remarkably connected countries all over the world through rapid long-distance transportation, helping people overcome geographic barriers. Consequently, this has ushered in substantial economic growth, both nationally and internationally.

2. Paper : Airline Route Profitability Analysis and Optimization Using BIG DATA Analytics on Aviation Data Sets under Heuristic Techniques

Author : E.Kasturia, PrasannaDevi, VinuKiran, Manivannan

Description : Applying vital decisions for new airline routes and aircraft utilization are important factors for airline decision-making. For data driven analysis key points such as airliners route distance, availability on seats/freight/mails and fuel are considered. The airline route profitability optimization model is proposed based on performing Big data analytics over large scale aviation data under multiple heuristic methods, based on which practical problemsareanalysed.

3. Paper : Towards a maturity model for big data analytics in airline network planning

Author : Iris Hausladen, Maximilian Schosser

Description: The evaluation, acquisition and use of newly available big data sources has become a major strategic and organizational challenge for airline network planners. We address this challenge by developing a maturity model for big data readiness for airline network planning.

4. Paper : Big Data Analytics in Airlines: Opportunities and Challenges

Author : Hamida Abd El Samie Mohamedorcid; Mahmoud Ramadan Al-Azab

Description: Big data refers to the huge amounts of information in the structured and unstructured form that cannot be processed using traditional data systems. Big data technology facilitates the utilization of high volumes of external and internal data to create new products, services and improve business operations.

5. Paper : A new approach of social media analytics to predict service quality: evidence from the airline industry

Author: Xin Tian, Wu He, Chuanyi Tang, Ling Li, Hangjun Xu, David Selover

Description: Research on how to use social media data to measure and evaluate service quality is still limited. To fill the research gap in the literature, the purpose of this paper is to open a new avenue for future work to measure the service quality in the service industry by developing a new analytical approach of using social media analytics to evaluate service.

6. Paper : Big data analytics platform for flight safety monitoring

Author: Bo Li; Xinguo Ming, Guoming Li

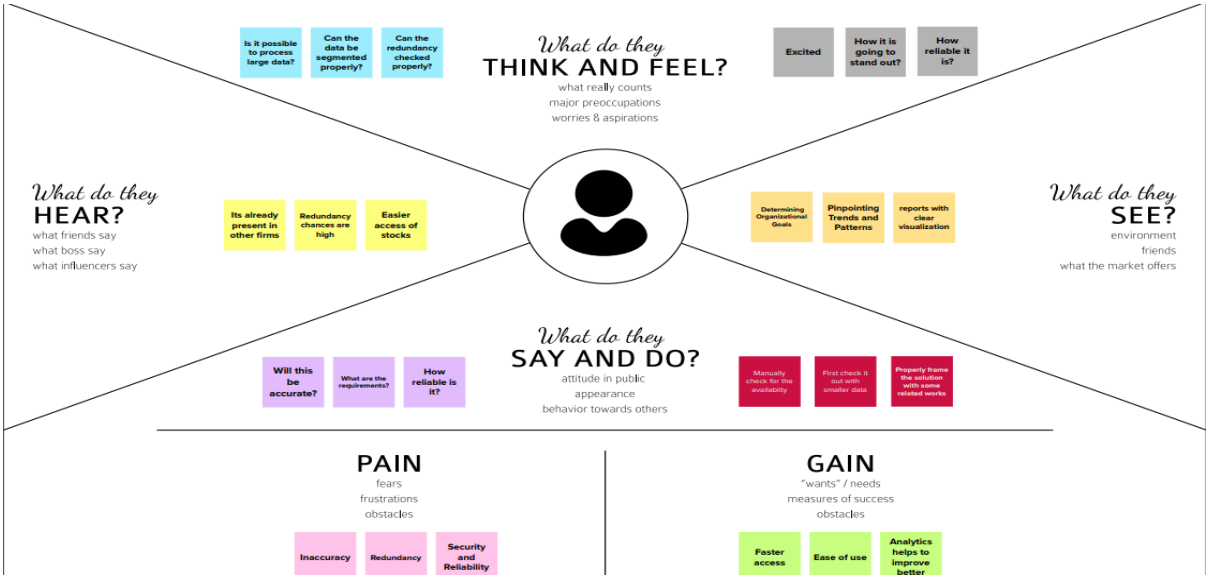
Description: The conventional methods of data analytics for flight safety monitoring have met many bottlenecks. This paper analyzes the insufficiencies of the preliminary business process of an airline. For the purpose of meeting requirements of efficiency and accuracy and avoiding the drawbacks encountered before, the architectural framework of the flight safety monitoring platform utilizing big data technology is proposed and demonstrated by the function module structure and logical structure.

2.3 Problem Statement Defenition

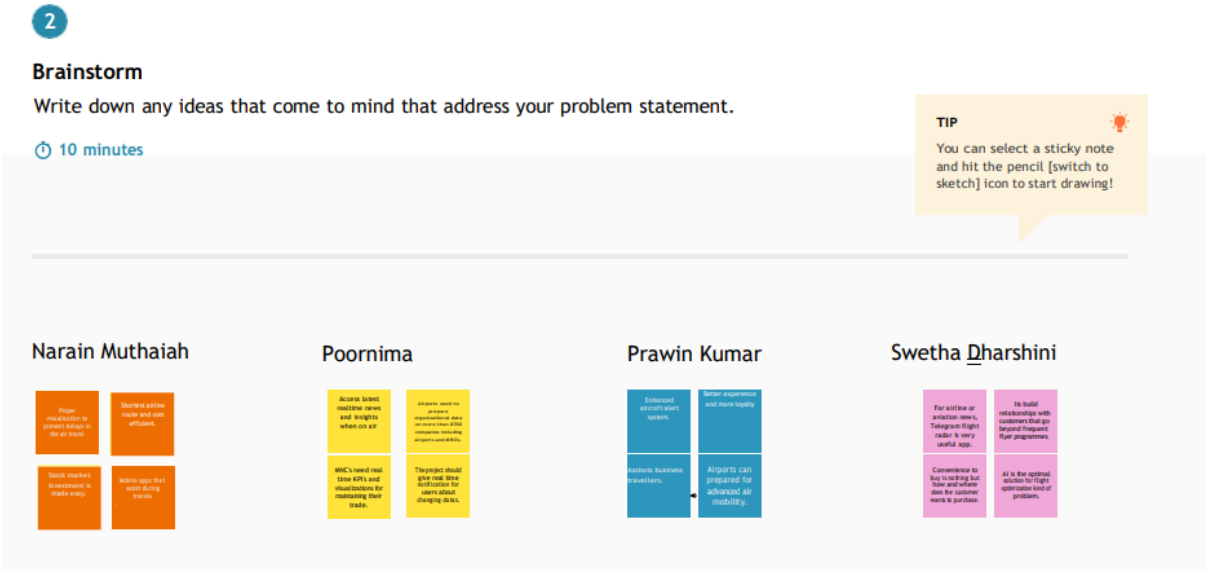
Technology is drastically changing the way businesses connect with their customers, take business decisions, and build workflows. Data is transforming airlines from pre-flight to post-flight operations, including ticket purchase, seat selection, luggage, boarding, ground transportation, etc. The ultimate 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.

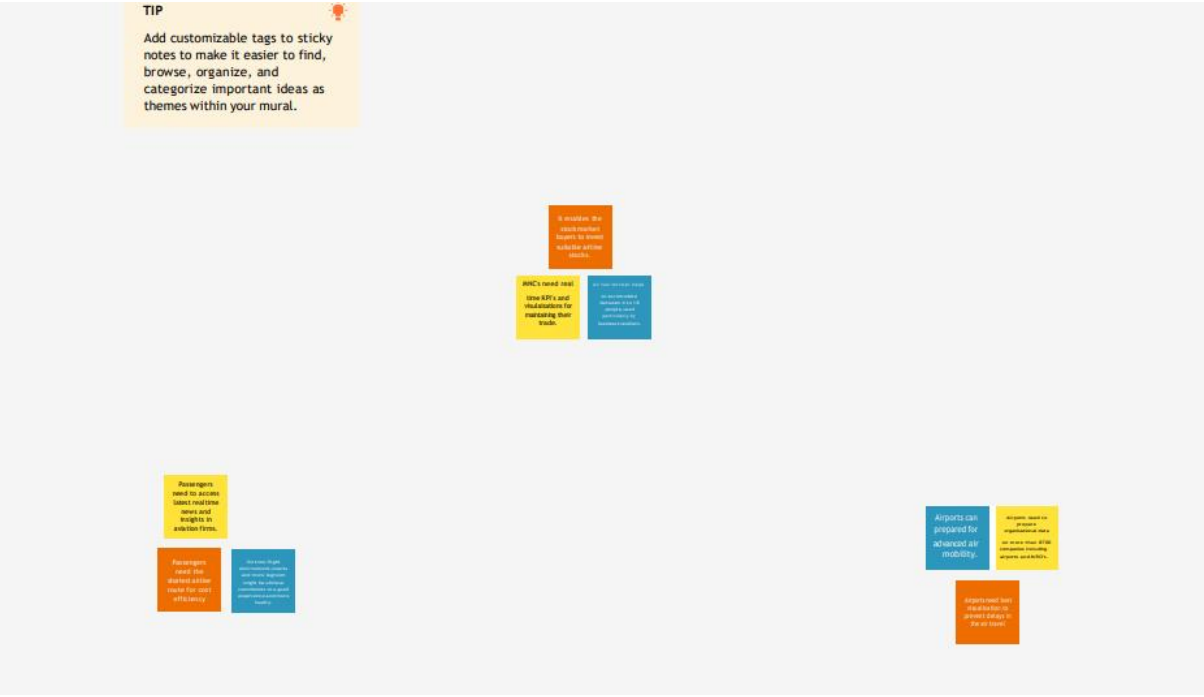
3.Ideation and Proposed Solution

3.1 Empathy Map Canvas




## 3.2 Ideation and Brainstorming







## 3.3 Proposed Solution

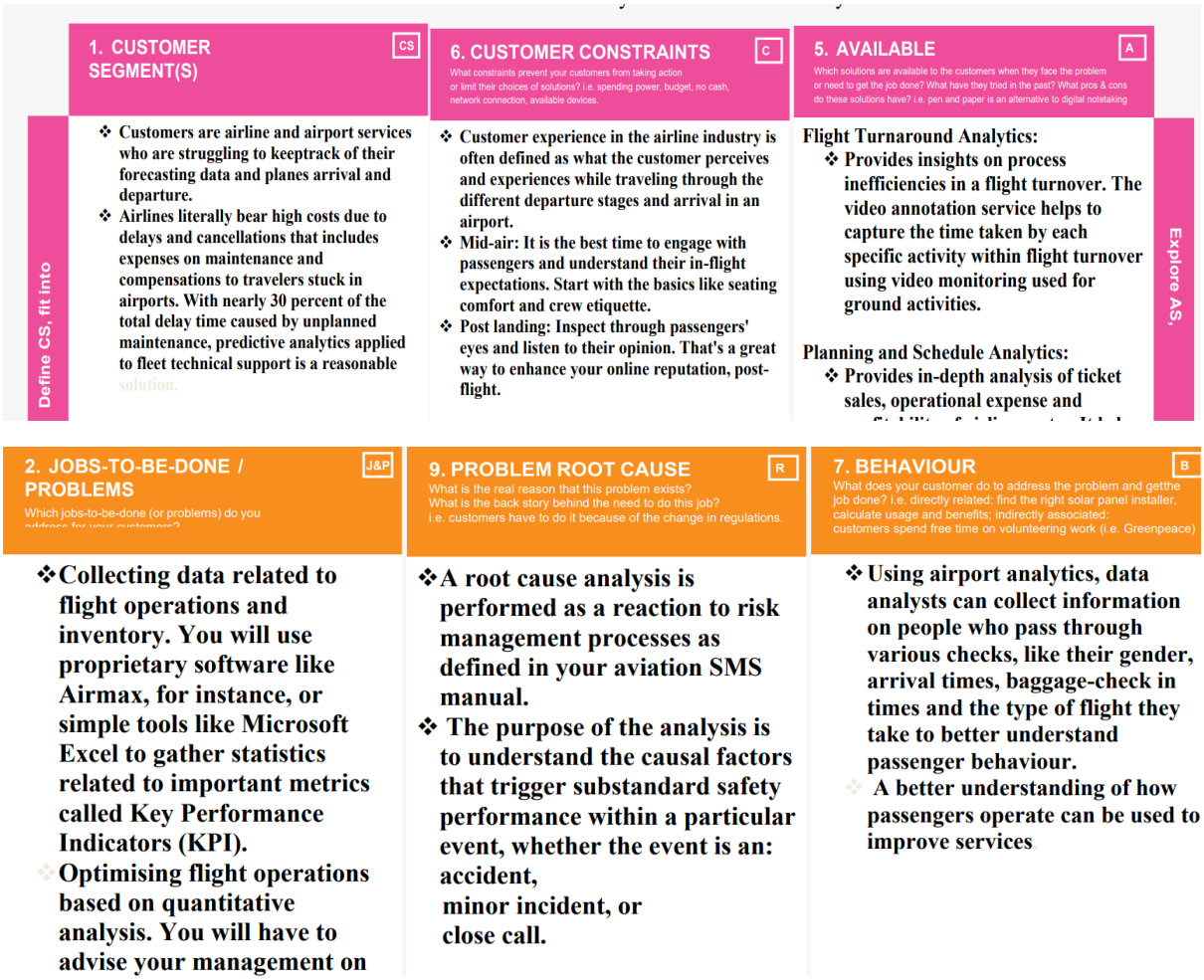
| S.no | Parameter | Description |
|---|---|---|
| 1. | Problem Statement (Problem to be solved) | Air travel is becoming increasingly popular among travelers, owing to its convenience and, in some cases, comfort. This has resulted in |

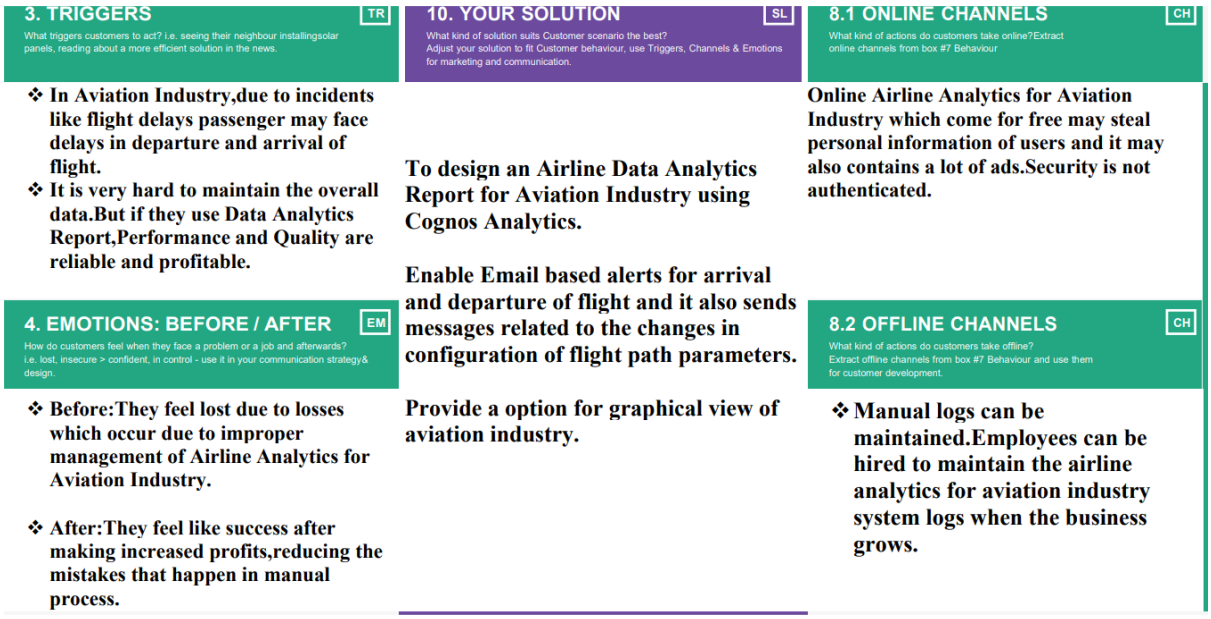
| | | |
|---|---|---|
| | | phenomenal growth in both air traffic and ground traffic. • Increased air traffic has resulted in massive levels of aircraft.There are delays both on the ground and in the air. These delays are to blame for significant monetary losses • It is critical to provide better airline and airport services while avoiding delays. in Air Travel across various locations and promise to transport passengers from On time from Location A to Location B. |
| 2. | Idea / Solution description | Understanding traveler demand for specific city pairs and pricing flights can be done using data analytics project. •Airlines use this biometric technology as a boarding option. The equipment scans travelers' faces and matches them with photos stored in border control agency databases. These can be handled with the aforementioned project. |
| 3. | Novelty/Uniqueness | The priority tags are recognized first using the priority selection technique, and then all other tags within its range are identified. As a result, the priority tag can receive service as soon as it is identified, while the other tags must wait in a line and receive service later. |

| | | |
|---|---|---|
| | | Keep connections with passengers and the cabin staff, if you can, and share information on the specifics and status of the trip so that you can interact with the right employees. |
| 4. | Social Impact/Customer Satisfaction | Data analytics helps the industry to understand customers' preferences and other maintenance issues. • For instance, analysis of ticket booking helps the industry to target the customers with personalised offers while optimising the price in real-time using predictive analysis techniques. As a result, by gathering meaningful data, airlines can fetch more bookings in the given timeframe. |
| 5. | Business Model(Revenue Model) | There are numerous apps available in this regard right now.However, once properly developed, our solution has the potential to ensure that passengers travel without delay due to air trafficking and that they are aware of any flight delays. |
| 6. | Scalability of Solution | The Cloud Cognos Analytics is not only for particular organization/governments. • Aviation industry acting under international, domestic or private are also getting satisfied with the aviation data |

| | | |
|---|---|---|
| | | analysing process provided as per their needs. |

## 3.4 Problem Solution Fit






# 4.Requirement Analysis

## 4.1 Functional Requirement

| FR No. | Non-Functional Requirement | Description |
|---|---|---|
| NFR-1 | **Usability** | User can easily understand and make use of the features of the application effectively. The application has a simple and user friendly graphical interface. Any action has to be performed with just a few clicks. |
| NFR-2 | **Security** | A proper login mechanism is used to avoid hacking. Personal information of users and other organization details to public will not be disclosed by the organization system. |
| NFR-3 | **Reliability** | Even when the system is disconnected or frozen due to over access at the same time, it should save all the process of the users made up to the point of abnormal happenings. |
| NFR-4 | **Performance** | The system should require an appropriate amount of speed especially while browsing through the catalogue. |

| FR No. | Functional Requirement (Epic) | Sub Requirement (Story / Sub-Task) |
|---|---|---|
| FR-1 | Customer Registration | Registrations can be made by the customers through Gmail. |
| FR-2 | User Confirmation | Once the registration is completed, the customer will get confirmation through mail. |
| FR-3 | Visualizing data | The Regular trends of delay of flights<br>Using IBM cognos Analytics can be visualized by the users. |
| FR-4 | Generating Report | The flight delay report can be viewed by the user. |

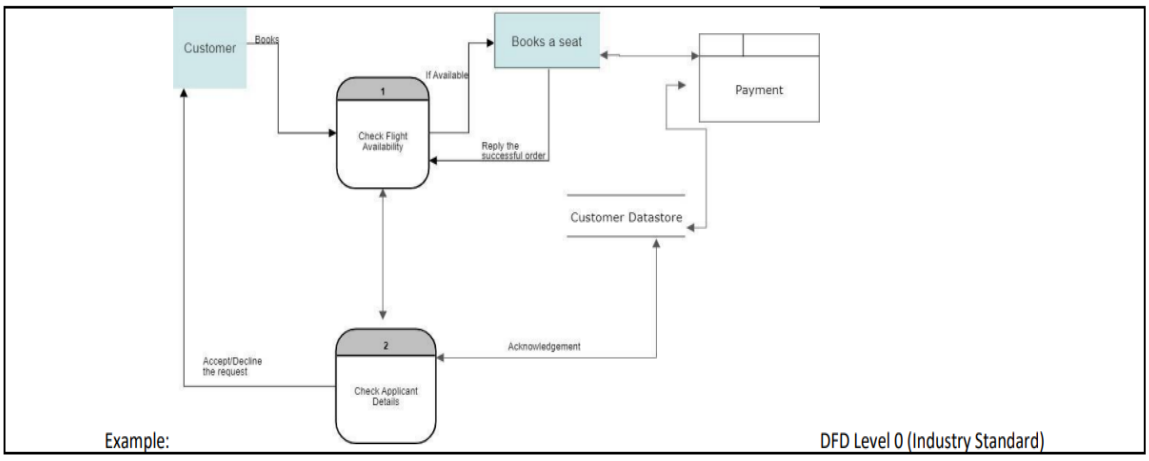
## 4.2 Non - Functional Requirements

| | | |
|---|---|---|
| NFR-6 | **Scalability** | The website can be accessed by a large number of users. |

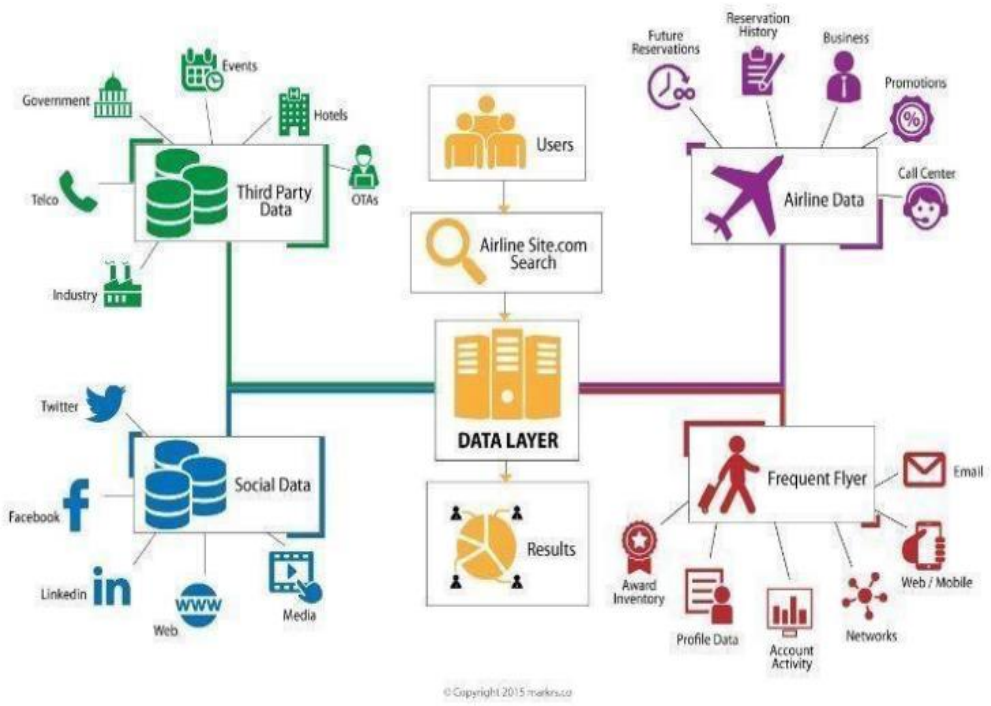
# 5.Project Design

## 5.1 Data Flow Diagrams

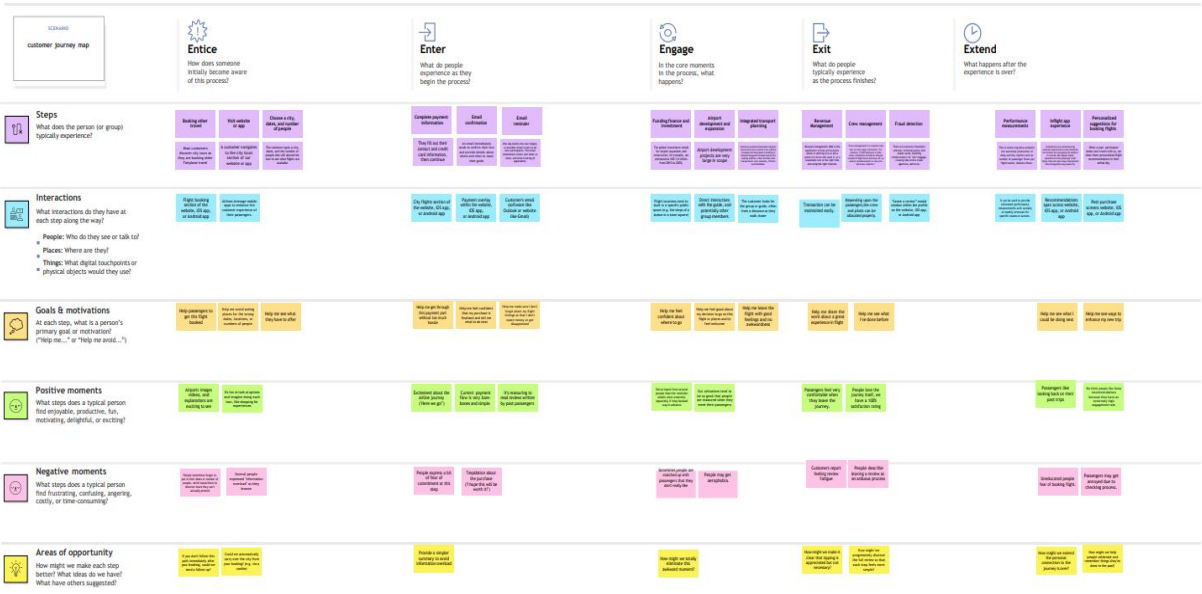
A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. It is a traditional visual representation of 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 and Technical Architecture




## 5.3 User Stories



# 6. Project Planning and Scheduling

## 6.1 Project Planning and Estimation

| Activity Name | Activity Number | Activity Description | Tasks Assigned | Status |
|---|---|---|---|---|
| Preparation Phase | 1 | a) Access the resources in project dashboard.<br>b) Explore the dataset provided in workspace.<br>c) Create GitHub account & collaborate with Project Repository in project workspace.<br>d) Set-up the prerequisites for the project. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| Ideation Phase | 2 | a) Literature survey relevant to the selected project.<br>b) Preparation of Empathy Map to identify the user pros and cons.<br>c) List the ideas by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| **Project Design Phase-I** | 3 | | | |
| Proposed Solution | 3.1 | Preparation of proposed solution document, which includes the Problem statement , Idea description, novelty, feasibility of idea, business model, social impact and scalability of the solution. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |

| | | | | |
|---|---|---|---|---|
| Problem Solution Fit | 3.2 | Prepared problem solution fit document which have designed a value proposition that addresses the customers' job, pros and cons to the particular application. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| Solution Architecture | 3.3 | Develop effective architecture for the proposed solution which provides ground for application development projects. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| **Project Design Phase-II** | 4 | | | |
| Solution Reqirements | 4.1 | Identify the Functional and Non-Functional requirements of the proposed solution. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| Customer Journey | 4.2 | Preparation of customer journey map to understand the user interactions which describes the stages that the customer experiences over time. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| Data Flow Diagram and User stories | 4.3 | Generate Data flow diagram for the Project which maps out the flow of information for the application. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| Technology Architecture | 4.4 | Develop effective technical architecture for the proposed solution which describes the logical software and hardware capabilities that are required to support the development of the application. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| **Project Planning Phase** | 5 | | | |
| Milestones & Activity List | 5.1 | Prepare Milestone and Activity list of the project. | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| Sprint Plan | 5.2 | Prepare Sprint Delivery plan of the project | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | Completed |
| **Project Development** | 6 | | | |
| Delivery of Sprint-1 | 6.1 | Implement the coding phase of Sprint- 1 | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | In Progress |
| Delivery of Sprint-2 | 6.2 | Implement the coding phase of Sprint-2 | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | In Progress |
| Delivery of Sprint-3 | 6.3 | Implement the coding phase of Sprint- 3 | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | In Progress |
| Delivery of Sprint-4 | 6.4 | Implement the coding phase of Sprint- 4 | Narain Muthiah V<br>Poornima V<br>Prawin Kumar P<br>Swetha Dharshini M | In Progress |

## 6.2 Sprint Delivery and Schedule

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-1 | Registration | USN-1 | I can sign up for the application as a user by providing my email address, password, and confirming that. | 2 | High | Narain Muthaiah |
| Sprint-1 | Registration | USN-2 | When I register for the application as a user, I will get a confirmation email. | 3 | High | Prawin Kumar P |
| Sprint-1 | Login | USN-3 | I've grown accustomed to using credentials to access the system as a user. | 2 | Low | Poornima V |
| Sprint-1 | Collection of dataset | USN-4 | I can collect the dataset and choose the area of interest to be tracked and analysed as a user. | 5 | Medium | Swetha Dharshini M |
| Sprint-2 | Dataset Exploration | USN-5 | I can explore the given dataset through IBM cognos | 6 | High | Prawin Kumar P |
| Sprint-2 | Dataset Visualization | USN-6 | I will use cognos as a developer to visualise the provided dataset into a dashboard. | 6 | High | Narain Muthiah |
| Sprint-3 | Dashboard Customization | USN-7 | I can personalise the dashboard that is visualised as a user. | 6 | Medium | Poornima V |
| Sprint-3 | Ease of Access | USN-8 | I can simply access and use the dashboard as a user. | 6 | Medium | Swetha Dharshini M |

| 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 | 12 | 29 Oct 2022 |
| Sprint-2 | 20 | 6 Days | 31 Oct 2022 | 05 Nov 2022 | 12 | 11 Nov 2022 |
| Sprint-3 | 20 | 6 Days | 07 Nov 2022 | 12 Nov 2022 | 12 | 15 Nov 2022 |
| Sprint-4 | 20 | 6 Days | 14 Nov 2022 | 19 Nov 2022 | 12 | 19 Nov 2022 |

# 7 CODING AND SOLUTIONING:

## 7.1 Feature 1

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

sns.set_theme(style="darkgrid",color_codes=True)

sns.catplot(x='continent', y='elevation_m',data=data)

## 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')

# subplot of the number of airports per type

counts = data['type'].value_counts().sort_index()

labels =[counts.index[i] for i in range(len(counts))]

sizes = [counts.values[i] for i in range(len(counts))]


counts_US = US_airports['type'].value_counts().sort_index()

labels_US =[counts_US.index[i] for i in range(len(counts))]

sizes_US = [counts_US.values[i] for i in range(len(counts))]



plt.figure(figsize=(14,8))


plt.subplot(1,2,1)

plt.pie(sizes_US,labels=labels_US,autopct='%1.1f%%',startangle=90,textprops={'fontsize':10})

plt.title('types of airports in the US',fontsize=20)


plt.subplot(1,2,2)

plt.pie(sizes,labels=labels,autopct='%1.1f%%',startangle=90,textprops={'fontsize':15})

plt.title('types of airports in the rest of the world',fontsize=20)

plt.figure(figsize=(25,16))


plt.show()

print("In the US there is:\n",counts_US.sort_values())

print("\nwhile there is in the rest of the world :\n",counts.sort_values())
```

7.2 Feature 2

```
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
All cancellations
-- 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
Analytics function
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
```

Front End Screenshot:


8. Testing:

8.1 UAT

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 | 7 | 5 | 1 | 2 | 17 |
| Duplicate | 1 | 0 | 4 | 0 | 5 |
| External | 2 | 5 | 0 | 1 | 8 |
| Fixed | 11 | 2 | 4 | 15 | 32 |
| Not Reproduced | 0 | 0 | 1 | 0 | 1 |
| Skipped | 0 | 0 | 1 | 3 | 4 |
| Won't Fix | 0 | 5 | 4 | 1 | 10 |
| Totals | 21 | 17 | 15 | 22 | 77 |

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 | 8 | 0 | 0 | 8 |
| Client Application | 14 | 0 | 0 | 14 |
| Security | 3 | 0 | 0 | 3 |
| Outsource Shipping | 4 | 0 | 0 | 4 |
| Exception Reporting | 8 | 0 | 0 | 8 |
| Final Report Output | 3 | 0 | 0 | 3 |
| Version Control | 2 | 0 | 0 | 2 |

9. Results

9.1 Performance Analysis:

**Model Performance Testing:**

Project team shall fill the following information in model performance testing template.

| S.No. | Parameter | Screenshot / Values |
|---|---|---|
| 1. | Dashboard design | No of Visulizations / Graphs - 8 |
| 2. | Data Responsiveness | High |
| 3. | Amount Data to Rendered (DB2 Metrics) | 240 entries |
| 4. | Utilization of Data Filters | Null values are removed |
| 5. | Effective User Story | No of Scene Added - 2 |
| 6. | Descriptive Reports | No of Visulizations / Graphs - 2 |

10. Advantages and Disadvantages:

ADVANTAGES

* Lower operating costs

* Better customer service

*Helps in the customer in choosing the better fit

* Market leading competitiveness

* Increased profit margin

DISADVANTAGES

*The data that has been collected from the persons are prone to network threat and must be safeguarded with authenticated network

*Proprietary softwares must be used to handle such big data and the management must be advised on trends and bottlenecks that they observe from data analytics so they can take necessary action.

11.Conclusion:

We have designed an Airline Data Analytics Report for Aviation Industry using Cognos analytics. This enable email based alerts for arriv and departure of flight and it also sends messages to the changes in configuration of flight path parameters. Apart from this an option for graphical view of aviation industry is also given. This better understanding of how the passengers operate can be used to improve services. This way we could help connect the business with their customers with both benefitting from each other and build workflows.

12.Future Scope:

The work as in is a helpful entity but with further development on the amount of data acquired and the technology stack used for them we could implement it  with a real time access. With more of strategies built around it and cleasmr comprehension and monitoring ,it would encompass more meaningful use.

13.Appendix

Source Code

Sprint 1

```
@import
url('https://fonts.googleapis.com/css2?family=Poppins:wght@400;500;600&display=swap');

* {
    margin: 0;
    padding: 0;
    box-sizing: border-box;
    font-family: "Gill Sans Extrabold",sans-serif;
}


body {
```

```
    margin: 50px;
    display: flex;
    align-items: center;
    justify-content: start;
    background-image:
url("https://wallpaperaccess.com/full/1307395.jpg");
    background-repeat: no-repeat;
    background-position: center;
    background-size: cover;
}
::selection {
    color: #fff;
    background: #53f0e3;
}

.wrapper {
    width: 380px;
    padding: 40px 30px 50px 30px;
    background: linear-gradient(rgba(224, 123, 64,.7),rgba(230, 86, 86,
0.7),rgb(90, 87, 87));
    border-radius: 5px;
    text-align: center;
    box-shadow: 10px 10px 15px rgba(0, 0, 0, 0.1);
}

.wrapper header {
    font-size: 35px;
    font-weight: 600;
```

```
}

.wrapper form {
    margin: 40px 0;
}

form .field {
    width: 100%;
    margin-bottom: 20px;
}

form .field.shake {
    animation: shake 0.2s ease-in-out;
}

@keyframes shake {
    0%,
    100% {
        margin-left: 0px;
    }
    20%,
    80% {
        margin-left: -12px;
    }
    40%,
    60% {
```

```
        margin-left: 12px;
    }
}

form .field .input-area {
    height: 50px;
    width: 100%;
    position: relative;
}

form input {
    width: 100%;
    height: 100%;
    outline: none;
    padding: 0 45px;
    font-size: 20px;
    background: none;
    caret-color: #eb6122;
    border-radius: 5px;
    border: 1px solid #bfbfbf;
    border-bottom-width: 2px;
    transition: all 0.2s ease;
}

form .field input:focus,
form .field.valid input {
    border-color: #f06b53;
```

```
}

form .field.shake input,
form .field.error input {
    border-color: #242121;
}

.field .input-area i {
    position: absolute;
    top: 50%;
    font-size: 20px;
    pointer-events: none;
    transform: translateY(-50%);
}

.input-area .icon {
    left: 15px;
    color: #bfbfbf;
    transition: color 0.2s ease;
}

.input-area .error-icon {
    right: 15px;
    color: #d8b9bc;
}

form input:focus~.icon,
```

```
form .field.valid .icon {
    color: #3d3e42;
}

form .field.shake input:focus~.icon,
form .field.error input:focus~.icon {
    color: #bfbfbf;
}

form input::placeholder {
    color: #bfbfbf;
    font-size: 18px;
}

form .field .error-txt {
    color: #130608;
    text-align: left;
    margin-top: 5px;
}

form .field .error {
    display: none;
}

form .field.shake .error,
form .field.error .error {
    display: block;
```

```
}

form .pass-txt {
    text-align: left;
    margin-top: -10px;
}

.wrapper a {
    color: #110804;
    text-decoration: none;
}

.wrapper a:hover {
    text-decoration: underline;
}

form input[type="submit"] {
    height: 50px;
    margin-top: 30px;
    color: #fff;
    padding: 0;
    border: none;
    background: #d49278;
    cursor: pointer;
    border-bottom: 2px solid rgba(0, 0, 0, 0.1);
    transition: all 0.3s ease;
}
```

```
form input[type="submit"]:hover {
    background: #0d4c5f;
}
```

Script.js

```
const form = document.querySelector("form");
eField = form.querySelector(".email"),
eInput = eField.querySelector("input"),
pField = form.querySelector(".password"),
pInput = pField.querySelector("input");

form.onsubmit = (e) => {
    e.preventDefault();

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

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

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

    function checkEmail() {
```

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

            (eInput.value != "") ? errorTxt.innerText = "Enter a valid email address": errorTxt.innerText = "Email is required";
        } else {
            eField.classList.remove("error");
            eField.classList.add("valid");
        }
    }

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

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

```
}
```

Style.css

Style.css

```
@import
url('https://fonts.googleapis.com/css2?family=Poppins:wght@400;500;600&display=swap');
* {
    margin: 0;
    padding: 0;
    box-sizing: border-box;
    font-family: "Gill Sans Extrabold",sans-serif;
}


body {
    margin: 50px;
    display: flex;
    align-items: center;
    justify-content: start;
    background-image:
url("https://wallpaperaccess.com/full/1307395.jpg");
    background-repeat: no-repeat;
    background-position: center;
    background-size: cover;
}
::selection {
    color: #fff;
    background: #53f0e3;
}
```

```
.wrapper {
    width: 380px;
    padding: 40px 30px 50px 30px;
    background: linear-gradient(rgba(224, 123, 64,.7),rgba(230, 86, 86, 0.7),rgb(90, 87, 87));
    border-radius: 5px;
    text-align: center;
    box-shadow: 10px 10px 15px rgba(0, 0, 0, 0.1);
}

.wrapper header {
    font-size: 35px;
    font-weight: 600;

}

.wrapper form {
    margin: 40px 0;
}

form .field {
    width: 100%;
    margin-bottom: 20px;
}

form .field.shake {
    animation: shake 0.2s ease-in-out;
```

```
}

@keyframes shake {
    0%,
    100% {
        margin-left: 0px;
    }
    20%,
    80% {
        margin-left: -12px;
    }
    40%,
    60% {
        margin-left: 12px;
    }
}

form .field .input-area {
    height: 50px;
    width: 100%;
    position: relative;
}

form input {
    width: 100%;
    height: 100%;
    outline: none;
```

```
    padding: 0 45px;
    font-size: 20px;
    background: none;
    caret-color: #eb6122;
    border-radius: 5px;
    border: 1px solid #bfbfbf;
    border-bottom-width: 2px;
    transition: all 0.2s ease;
}

form .field input:focus,
form .field.valid input {
    border-color: #f06b53;
}

form .field.shake input,
form .field.error input {
    border-color: #242121;
}

.field .input-area i {
    position: absolute;
    top: 50%;
    font-size: 20px;
    pointer-events: none;
    transform: translateY(-50%);
}
```

```
.input-area .icon {
    left: 15px;
    color: #bfbfbf;
    transition: color 0.2s ease;
}

.input-area .error-icon {
    right: 15px;
    color: #d8b9bc;
}

form input:focus~.icon,
form .field.valid .icon {
    color: #3d3e42;
}

form .field.shake input:focus~.icon,
form .field.error input:focus~.icon {
    color: #bfbfbf;
}

form input::placeholder {
    color: #bfbfbf;
    font-size: 18px;
}
```

```
form .field .error-txt {
    color: #130608;
    text-align: left;
    margin-top: 5px;
}

form .field .error {
    display: none;
}

form .field.shake .error,
form .field.error .error {
    display: block;
}

form .pass-txt {
    text-align: left;
    margin-top: -10px;
}

.wrapper a {
    color: #110804;
    text-decoration: none;
}

.wrapper a:hover {
    text-decoration: underline;
```

```
}

form input[type="submit"] {
    height: 50px;
    margin-top: 30px;
    color: #fff;
    padding: 0;
    border: none;
    background: #d49278;
    cursor: pointer;
    border-bottom: 2px solid rgba(0, 0, 0, 0.1);
    transition: all 0.3s ease;
}

form input[type="submit"]:hover {
    background: #0d4c5f;
}
```

Sprint 2:

Scale_fix.js

```
var metas = document.getElementsByTagName('meta');
var i;
if (navigator.userAgent.match(/iPhone/i)) {
  for (i=0; i<metas.length; i++) {
    if (metas[i].name == "viewport") {
      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 == "viewport") {
      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;
  }

  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

{{ site.description | default: site.github.project_tagline }}

{% for nav in site.navigation %} {% if nav.url contains "://" %} {% else %} {% endif %} {% endfor %}

{{ nav.title }}

{{ nav.title }}

{% if site.show_downloads %}

Download ZIP File

Download TAR Ball

View On GitHub

{% endif %}

{{ content }}

{% if site.github.is_project_page %}

This project is maintained by {{ site.github.owner_name }}

{% endif %}

Sprint 4

3,4 delays

```
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
```

```
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
```

Data Visualisation using Python

```
import numpy as np
import pandas as pd
```

```
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.graph_objects as go
import plotly.express as px
import os
for dirname, _, filenames in os.walk('C:/Users/DELL/Downloads/archive (1)'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
data=pd.read_csv("C:/Users/DELL/Downloads/archive (1)/airports.csv")


print("Lines: ",data.shape[0])
print("columns: ",data.shape[1])
data.head()
countries=pd.read_csv("C:/Users/DELL/Downloads/archive (1)/countries.csv")
print("Lines: ",countries.shape[0])
print("columns: ",countries.shape[1])
countries.head()
regions=pd.read_csv("C:/Users/DELL/Downloads/archive (1)/regions.csv")
print("Lines: ",regions.shape[0])
print("columns: ",regions.shape[1])
regions.head()
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

data['continent']=data['continent'].fillna('NoA') # NA for north america is a little bit confusing so we will fill it with NoA

countries['continent']=countries['continent'].fillna('NoA')

data['iso_country']=data['iso_country'].fillna('NAM') # well change NA to NAM for Namibia

countries['code']=countries['code'].fillna('NAM')

countries.loc[countries['name']=='Perú','name']='Peru'

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()

data.set_index('ident',inplace=True) ## set index to ident

print("index is unique ") if data.index.is_unique else print("index is not (duplicated index )");

a=data.groupby('continent').size().sort_values(ascending=True).index # we will get the continents with the most airports

data=data[data['continent'].isin(a)] # we will keep only the continents with the most airports

number_per_country=data['iso_country'].value_counts()

medium_airports=data[data['type']=='medium_airport']["iso_country"].value_counts()

small_airports=data[data['type']=='small_airport']["iso_country"].value_counts()

number_per_country

# 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()

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

sns.set_theme(style="darkgrid",color_codes=True)

sns.catplot(x='continent', y='elevation_m',data=data)

## 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')

# subplot of the number of airports per type

counts = data['type'].value_counts().sort_index()

labels =[counts.index[i] for i in range(len(counts))]

sizes = [counts.values[i] for i in range(len(counts))]


counts_US = US_airports['type'].value_counts().sort_index()

labels_US =[counts_US.index[i] for i in range(len(counts))]

sizes_US = [counts_US.values[i] for i in range(len(counts))]



plt.figure(figsize=(14,8))


plt.subplot(1,2,1)

plt.pie(sizes_US,labels=labels_US,autopct='%1.1f%%',startangle=90,textprops={'fontsize':10})

plt.title('types of airports in the US',fontsize=20)
```

```
plt.subplot(1,2,2)
plt.pie(sizes,labels=labels,autopct='%1.1f%%',startangle=90,textprops={'
fontsize':15})
plt.title('types of airports in the rest of the world',fontsize=20)
plt.figure(figsize=(25,16))

plt.show()
print("In the US there is:\n",counts_US.sort_values())
print("\nwhile there is in the rest of the world :\n",counts.sort_values())
for i in countries.index:
    #add a new column with the number of airports per country
    code=countries.loc[i,'code']

    if code in number_per_country.index:
        total=number_per_country[code]
        countries.loc[i,'number_of_airports']=total
    else:
        countries.loc[i,'number_of_airports']=0



    if code in medium_airports.index:
        total_medium=medium_airports[code]
        countries.loc[i,'number_of_medium_airports']=total_medium
    else:
        countries.loc[i,'number_of_medium_airports']=0

    if code in small_airports.index:
```

```
        total_small=small_airports[code]
        countries.loc[i,'number_of_small_airports']=total_small
    else:
        countries.loc[i,'number_of_small_airports']=0



countries.head()
# drop the US because it has too many airports
px.choropleth(countries, locations="name",
                locationmode='country names', color="number_of_airports",
                hover_name="number_of_airports",
                color_continuous_scale=px.colors.sequential.Plasma,
                title='number of airports per country'
)
```

Github link:

https://github.com/IBM-EPBL/IBM-Project-37392-1660306922/blob/main/Project%20Development%20Phase/Sprint%201/index%20(4).html

Project Demo link:

https://drive.google.com/file/d/1IhSrKxyM3K2wjzoD5WA3sxknxEkMg9JI/view?usp=sharing