# ANALYTICS FOR HOSPITAL HEALTH-CARE DATA

**NALAIYA THIRAN PROJECT BASED LEARNING**

# ON

**PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP**

**A PROJECT REPORT**

|  |  |
| --- | --- |
| SABARI RAJA A | 911519205018 |
| SIRAJ ANWARDEEN | 911519205022 |
| VIGNESHWARAN | 911519205027 |
| SHAKKIF IBRAHIM | 911519205019 |
| ` |  |

# BACHELOR OF TECHNOLOGY

**IN**

# INFORMATION TECHNOLOGY

## MOHAMED SATHAK ENGINEERING COLLEGE

# November 2022

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO** | **TITLE** |  |
|  | **ABSTRACT** |  |
| **1** | **INTRODUCTION** |  |
| **2** | **OBJECTIVE** |  |
| **3** | **IDEATION PHASE** |  |
|  | **3.1 Literature Survey** |  |
|  | **3.2 Empathy Map** |  |
|  | **3.3 Ideation** |  |
|  | **3.4 Problem Statement** |  |
| **4** | **PROJECT DESIGN PHASE 1** |  |
|  | **4.1 Proposed Solution** |  |
|  | **4.2 Problem Solution Fit** |  |
|  | **4.3 Solution Architecture** |  |
| **5** | **PROJECT DESIGN PHASE 2** |  |
|  | **5.1 Customer Journey Map** |  |
|  | **5.2 Solution Requirements** |  |
|  | **5.3 Data Flow Diagrams** |  |
|  | **5.4 Technology Stack** |  |
| **6** | **PROJECT PLANNING PHASE** |  |
|  | **6.1 Prepare Milestone and Activity List** |  |
|  | **6.2 Sprint Delivery Plan** |  |
| **7** | **PROJECT DEVELOPMENT PHASE** |  |
| **7.1 Project Development - Delivery of Sprint - 1** | | |
| **7.2 Project Development - Delivery of Sprint - 2** | | |
| **7.3 Project Development - Delivery of Sprint - 3** | | |
| **7.4 Project Development - Delivery of Sprint - 4** | | |
| **8** | **CONCLUSION** |  |
| **9** | **REFERENCES** |  |

# ABSTRACT:

The healthcare sector is dominated by the expertise, skills, and knowledge regarding the ins and outs of the medical field. Since every step is taken and a decision made can alter the course of action and affect human health and life, so alongside the above, there are also factors beyond these that influence the healthcare sector which include accuracy in decision making, and confidence in tasks, and operations. This accuracy is brought through [data analytics](https://www.analyticssteps.com/blogs/what-data-analytics-and-its-types), an aspect that spins the thread on which the healthcare sector operates. With technology constantly evolving, healthcare is on its way to becoming another sector whose future is driven by data.

In any industry, a[data analytics and Business Intelligence (BI)](https://www.analyticssteps.com/blogs/10-analytics-business-intelligence-trends)solution can enhance operational efficiency, curtail costs, and simplify processes by calculating and employing KPIs for detecting opportunities and shaping up decision-making, making every participant from the patients and providers to payers and suppliers, unlocking the value of data benefits for everyone.

While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital.

This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning.

## INTRODUCTION

**1.1Overall description**

**Introduction to Healthcare Analytics** is designed for aspiring healthcare data champions and orients learners to this unwieldy field through the lens of three concepts: organizational mindset, analyst skill set, and technology tool set.  Learners walk through a day in the life of a healthcare analytics manager to illustrate all of the components required to turn raw data into a business success story.  Topics include the analytics life cycle, common use cases for analytics, roles and responsibilities for delivering healthcare analytics, and covers expectations of analytic teams.

* Articulate the role of health analytics in patient care, case management, and business operations
* Summarize the different stages of the analytic product life cycle and the business purpose of each stage
* Convince others why health analytics is an essential competency required if a health organization is to not just be competitive, but thrive
* Compare and contrast the differences between the following terms: analytics and informatics; analysis and analytics; data mining and machine learning; statistics and analytics; analytics and data science; and, business intelligence and visualization
* Summarize the common barriers to healthcare analytics being successful
* Identify the various roles of analytics in an organization

## Problem Statement

* Advancements in telemedicine
* Enhanced patient engagement
* Wearables that provide real-time alerts
* Disease prevention/population health
* Improving/refining treatment standard

## Purpose

**Health Data analytics** refers to analysis of the data using quantitative and qualitative techniques to be able to explore for trends and patterns in the data — to “acquire, manage, analyze, interpret and transform data into accurate, consistent and timely information.”

**Health Informaticq** refers to “a collaborative activity that involves people, processes and technologies” to use the information derived from data analytics to “improve the delivery of health care services and improve patient outcomes.”

## Motivation and scope Product Perspective

It describes the current events or summarizes the past events by generating reports with the help of statistical tools such as tables and graphs. Thus, it helps medical practitioners to study and understand the patient’s behavioral patterns in the past with the help of the patient’s operational data, which can be used to solve problems in cur-rent situations

## Product Features

The healthcare industry generates an enormous amount of data coming from various sources, such as EHRs (Electronic Health Records),LIMS (Laboratory Information

Management System), diagnostic or moni-toring instruments, supply chains, insurance claims/billing, pharmacy,real-time locating systems [3], and social media. The data that are collectedare used in continuous learning by employing various technologies and processes to derive insights from the information, which will help improve the quality of healthcare. In addition, the reduction in storage costs and the development of advanced architectures have led to large volumes of databeing stored, processed, and managed, using the existing traditional sys-tems adopted by the healthcare industry

## Assumptions and Dependencies

* + - Constraints are limitations which are outside the control of the project. The Project must be managed within these constraints.
    - Assumptions are made about events, or facts outside the control of project. External dependencies are activities which need to be completed before an internal activity can proceed.
    - Constraints, assumptions and dependencies can create risks that the project may be delayed because access is not provided to the site (assumption).
    - Assumption will be that the complexity may arise due to large unstructured data set.

## Constraints

* + - Hardware limitation and timing constraints.
    - High feature may not correspond to semantic similarity.
    - System Environment Windows subsystem for Linux with Ubuntu operating system will be required to run the application

## OBJECTIVE

Healthcare prediction is another data analytics method focusing on reducing future medical costs. Predictive technique uses patient medical history to evaluate all the potential health risks and predict a future medical treatment in advance (LexisNexis 2015). (Loginov et al 2012) stated that by retrieving and reviewing past patient details, information and diagnoses from the databases, predictive methods can take a place through forecasting, reducing time and costs. Parkland hospital in Dallas, Texas has launched a predictive system which scans all patients’ details and information to identify potentials risks and outcomes. As a result, the hospital has saved more than half a million dollars, especially in heart failure and disease predictions in terms of performing patients’ monitoring and avoiding future complications . Predictive analytics supports healthcare sectors to achieve a high level of effective overall care and preventive care, as predictive systems’ results allow treatments and actions to be taken when all the risks are recognized in early stages, which aids for minimizing costs. Furthermore, Patients can also work and support medical care by following up and updating their medical status, so they can get the necessary treatment at the right time. The technology era has added significant value to the healthcare decision support system, since decision making systems in healthcare care sectors can be enhanced by focusing on patient diagnoses, behavior, and prevention in order to reach a high level of care and improve healthcare economics . In the healthcare sectors, predictive analytics can be achieved in many ways such as; a medical care delivery success, which can be achieved by using a model that proposes algorithms in order to assist medical treatment for interacting diseases, which can reflect in capturing patient’s behavior and interactions.

1. **IDEATION PHASE**

## Literature Survey

The healthcare industry historically has generated large amounts of data, driven by record keeping, compliance & regulatory requirements, and patient care. While most data stored in hard copy form, the current trend is toward rapid digitization of these large amounts of data. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of data (known as ‘big data’) hold the promise of supporting a wide range of medical and healthcare functions, to derive previously untapped intelligence and insights from data to address many new and important questions

## Data Analytics for Hospital Health Care - Data:

That information technology has improved healthcare industries, but they also highlighted some of the difficulties related to the use of information technology in healthcare sectors, as they noticed that it is hard to implement information technology in small clinics and organizations, with high costs due to reduced efficiencies of scale. Therefore, IT implementation requires long term training and retention of skilled professionals

## Topological Data Analysis For Hospital Health Care-Data

One of the most important elements in dealing with and managing data is to know where and how this data will be stored once when it is collected. The traditional methods of storing and retrieving such data are not efficient anymore, since it was structured and stored in data warehouses and relational databases, after extracting and loading it from different outside sources. However, this data is transformed and classified before being ready to use and function .The numbers of data sources now and that a huge amount of data has become available, so this growth of data will absolutely require an agile database which can deal with the data logically and through data synchronization in order to adapt to the rapid data evolution.

**Patients Role in Healthcare Analytics:**

This section is concerned about how individuals (and patients in specific) can improve healthcare analytics through understanding the small and personal data, as well as educate themselves in how to collaborate with the healthcare data analytics to reach a high level of efficiency and accuracy. It was discussing the same point when he identified the term “citizen science”, where nonprofessional and educated individuals are skilled enough to conduct and support healthcare analytics system. Accordingly, this will require organizations to train individuals how to follow up and track their health information, as well as self-monitoring.

Principally, to perform good data analytics, first of all we should teach individuals how to understand and realize the importance of dealing with such data, for instance how to deal with breast cancer .However, how much our patients are educated and skilled to provide us with the data we expect, medical professionals still highly need to test and clarify this data to consider it and keep it on record. Also, he added that once when the data has been tested and clarified, we then need to find out how to change an individual’s behaviour, starting with parents and

guardians who are responsible for raising their children. However, Understanding of

information are not enough. Furthermore, patients should identify the risks and detect where to change, for instance; some patients know that they have a high level of blood pressure but they don’t know how to deal with it and control it: should they change drugs? Change eating habits and life styles? Do more exercises? information through filling and completing some online forms in order to keep track of their state of health, as well as to provide the suitable advice and treatment when needed. Moreover, patients can share some information with other patients, so they increase their knowledge, background and awareness in the healthcare analytics sectors regarding their conditions. Finally, patient who share their symptoms, diagnoses and results with others can gain benefit from the ability to understand their health conditions by comparing them with other patients.

**Healthcare Predictions and Decision Support System (DSS)**

Healthcare prediction is another data analytics method focusing on reducing future medical costs. Predictive technique uses patient medical history to evaluate all the potential health risks and predict a future medical treatment in advance stated that by retrieving and reviewing past patient details, information and diagnoses from the databases, predictive methods can take a place through forecasting, reducing time and costs. Parkland hospital in Dallas, Texas has launched a predictive system which scans all patients’ details and information to identify potentials risks and outcomes. As a result, the hospital has saved more than half a million dollars, especially in heart failure and disease predictions in terms of performing patients’

monitoring and avoiding future complications a positive influence on operating efficiency, a larger number of full-time employee equivalents hinders efficiency outcomes, which indicates the importance of enhancing labor efficiency

among carriers.

**Role of Predictive Analytics in Medical Healthcare**

Predictive analytics supports healthcare sectors to achieve a high level of effective overall care and preventive care, as predictive systems’ results allow treatments and actions to be taken when all the risks are recognized in early stages, which aids for minimizing costs. (Conley et al 2008). Furthermore, Obenshain (2004) said that patients can also work and support medical care by following up and updating their medical status, so they can get the necessary treatment at the right time. The technology era has added significant value to the healthcare decision support system, since decision making systems in healthcare care sectors can be enhanced by focusing on patient diagnoses, behavior, and prevention in order to reach a high level of care and improve healthcare economics.

**Financial Factors in Healthcare Predictive Analytics**

The most significant and obvious result of using such technology within the healthcare sectors is its results on costs. Because of cost, information is one of the main aspects that have a big effect on the cost of healthcare predictive analytics. Medical care systems have focused on increasing healthcare analytics performance as well as minimizing the cost by simplify unstructured clinical record and reducing irregular information. Consequently, large quantities of information then will be managed and controlled smoothly and efficiently . Predictive analytics can assist to avoid and reduce inaccurate prediction costs plus time for the reason that it makes the data sourcing cost lower by specifying the desired and necessary data only, since the data is simplified, standardized and exists in historical clinical databases.

**Healthcare Analytics & Real Time**

The real time analytics produces more accurate results and information, since it evaluates current patients’ history and conditions, therefore investigating patients’ diagnosis correctly and offering the best treatment. Real time monitoring techniques guarantee to keep data up to date and increase the quality of information, as assumed so by Taylor (2010). She believed that real time matters in healthcare analytics are very significant for the reason that it generates accurate results, such as where diabetes patients can recover if their ailment is discovered and treated correctly in the earlier stages. Walker et al (2012) agreed with that however they also highlighted some of its disadvantages, such as high cost, its high required level of training and long time to complete.The objective of this Report was to conduct a review, which encourages professionals, doctors, medical staff and patients to adopt and utilize technologies in order to assist healthcare analytics and improve decision

making process in our everyday life. This study has found that most relevant keywords technology to support medical information systems.

**Healthcare Data Analytics Platforms and Tools:**

Comparing between traditional analytics and advanced analytics, traditional analytics is focusing on business intelligence, operational research and data mining. However, advanced analytics is focusing on descriptive, predictive and optimization (Raden 2010). The Report has come up with different tools and techniques that would improve healthcare data analytics in order to support descriptive, predictive and prescriptive healthcare data analytics the first tool is:

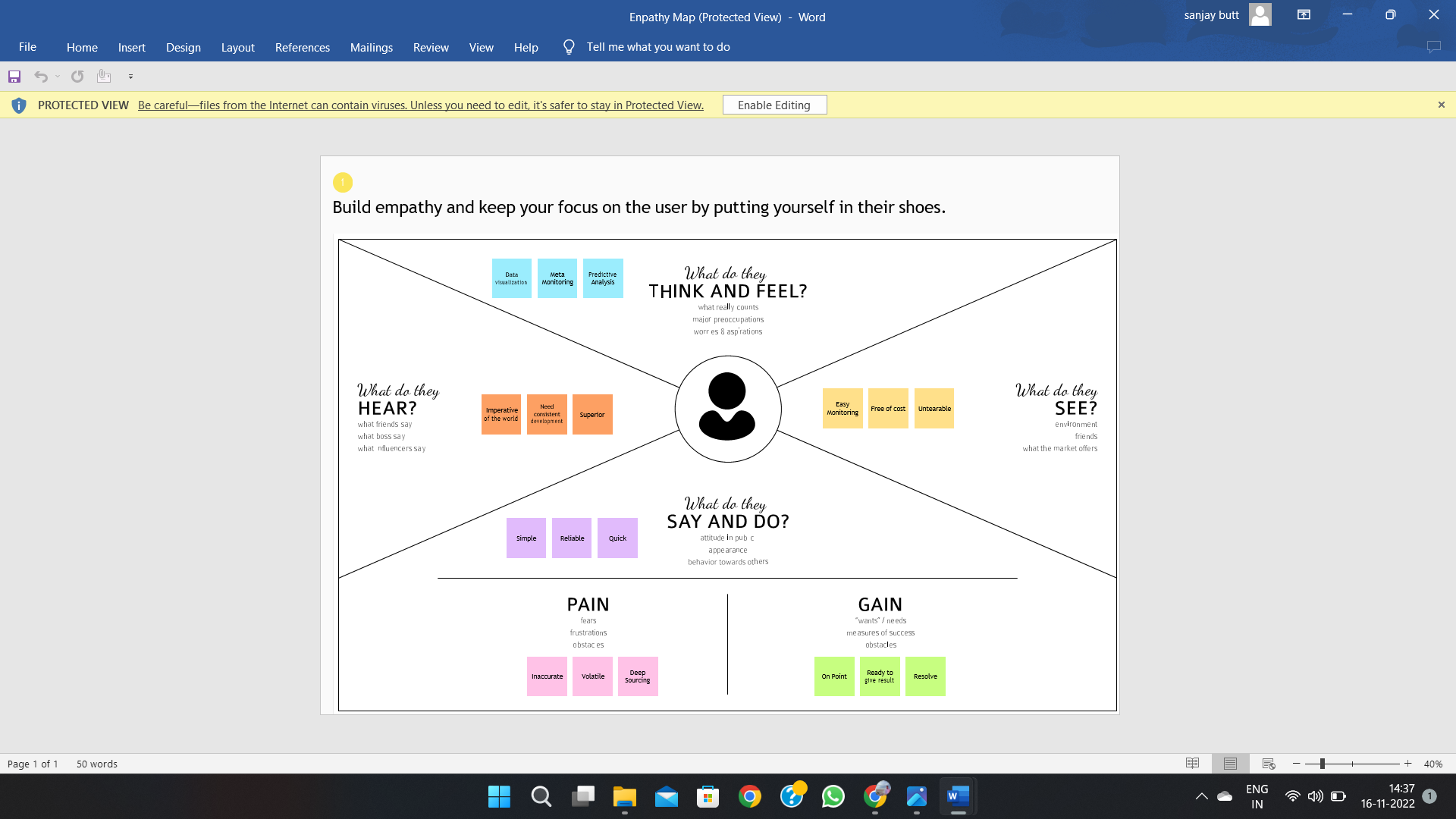
**Advanced Data Visualization (ADV):** ADV is different from other standards bars and line chart, since it can scale its visualization for millions of data points, also can handle different data types.

**Online Analytics Processing (OLAP):** It can improve healthcare system by performing statistical calculation very fast through hierarchal and multidimensional organized data, and can increase data integrity checking, quality control and reporting services. OLAP has the ability to improve healthcare decision making system.

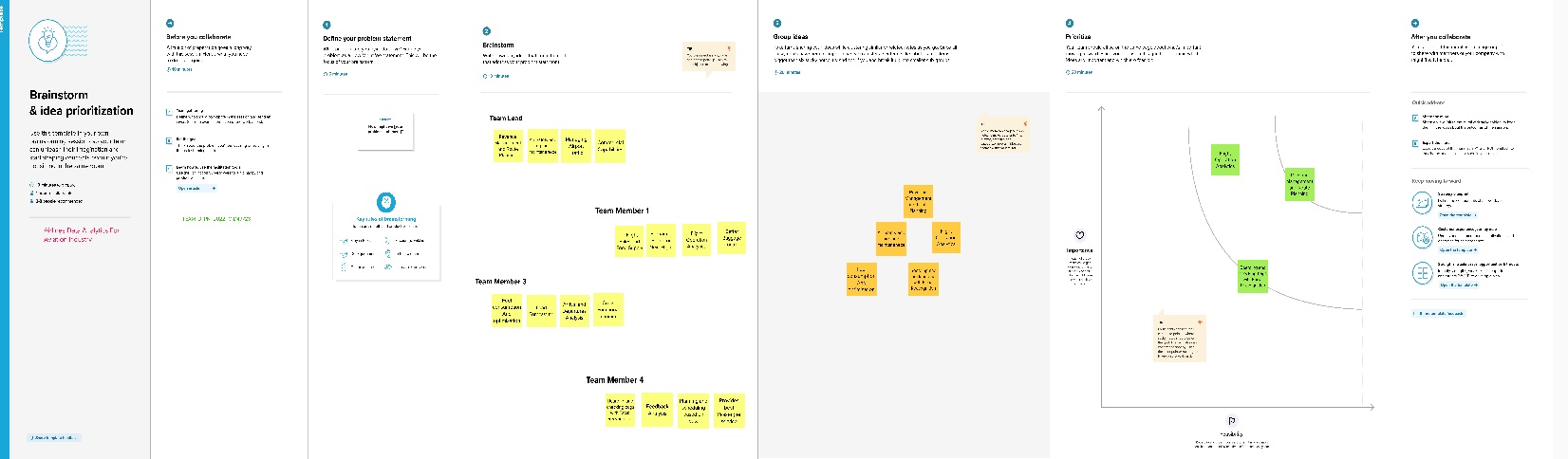
**Casandra File System (CFS**): CFS is also distributed system like HDFS, however CFS is a designated system to perform analytic operation with no single point of failure.

**Online Transaction Processing (OLTP):** It is similar to OLAP, but it is designated to process patient care operations, such as patient registration, hospital documents and results review.

## Empathy Map



* 1. **Brainstroming**

****

## Problem Statement

#### **An Unstable Federal Reimbursement Model.**

#### **Awkward Incentive Model.**

#### **Double Customer Services Standard.**

#### **Lack of Transparency in Assigning Hospital**

#### **Lack of Transparency in Selecting Drugs and Medications Used in Hospitals.**

## PROJECT DESIGN PHASE 1

* 1. **Proposed Solution**

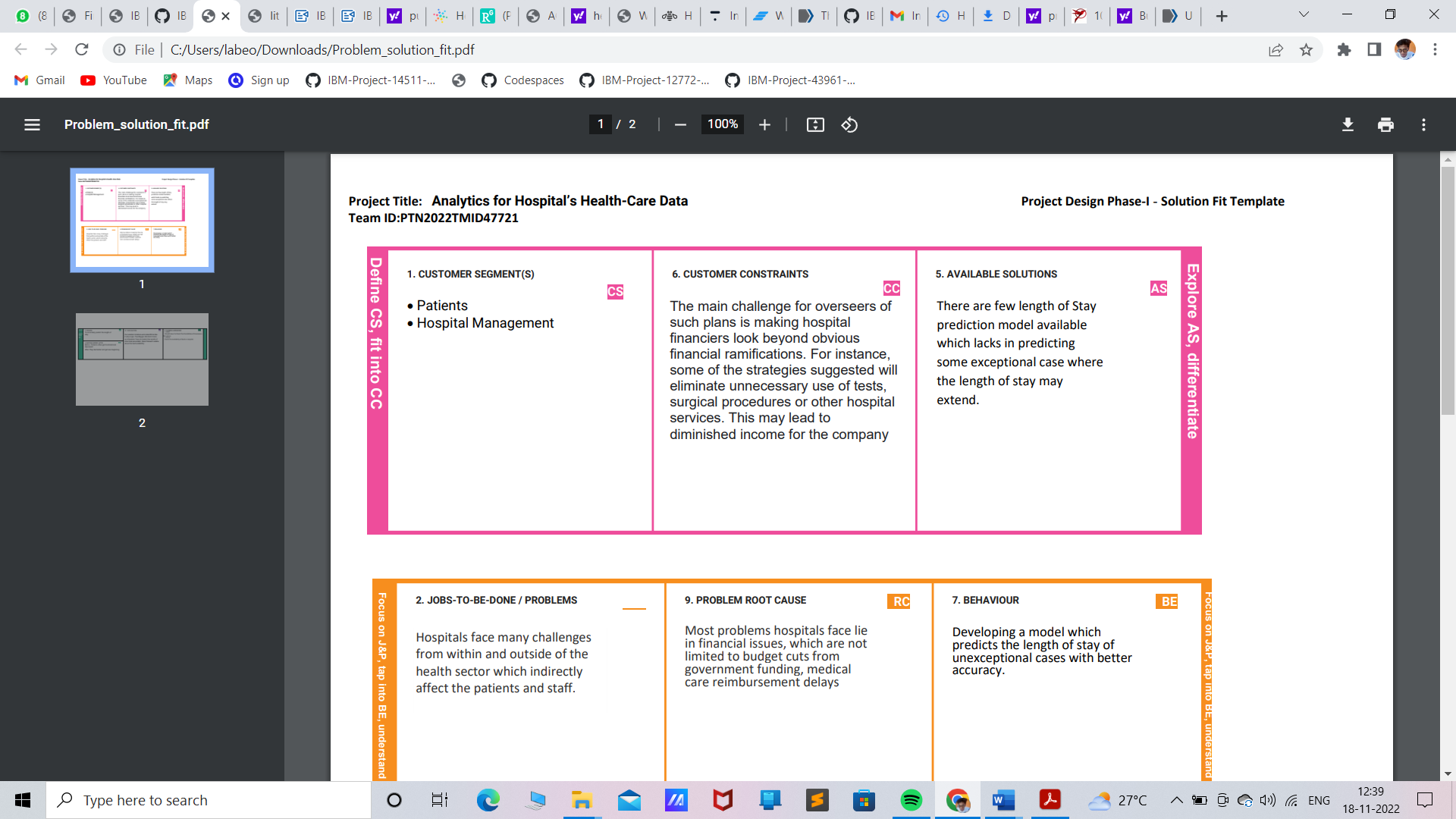
**Proposed Solution Template:**

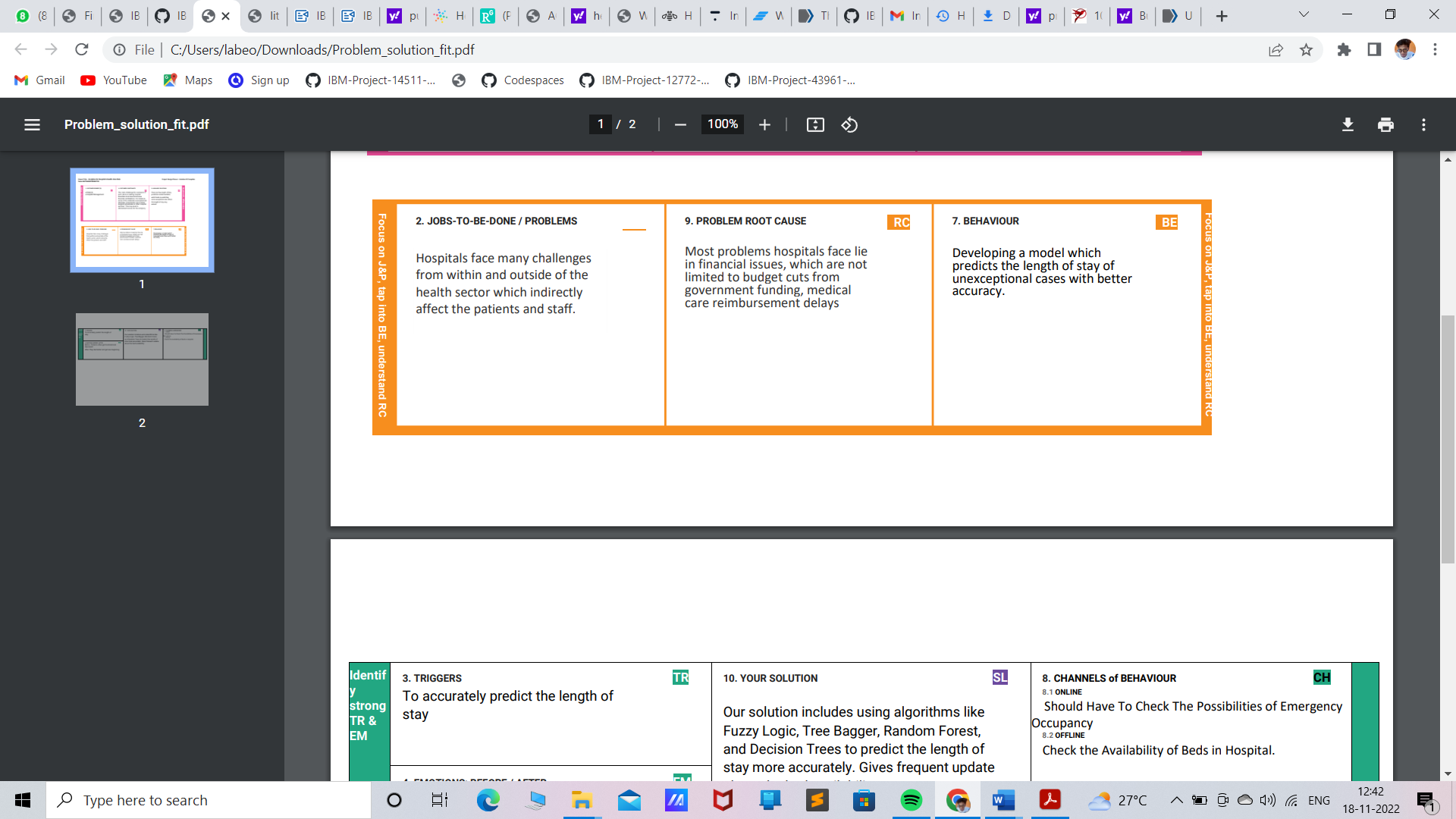
Project team shall fill the following information in proposed solution template.

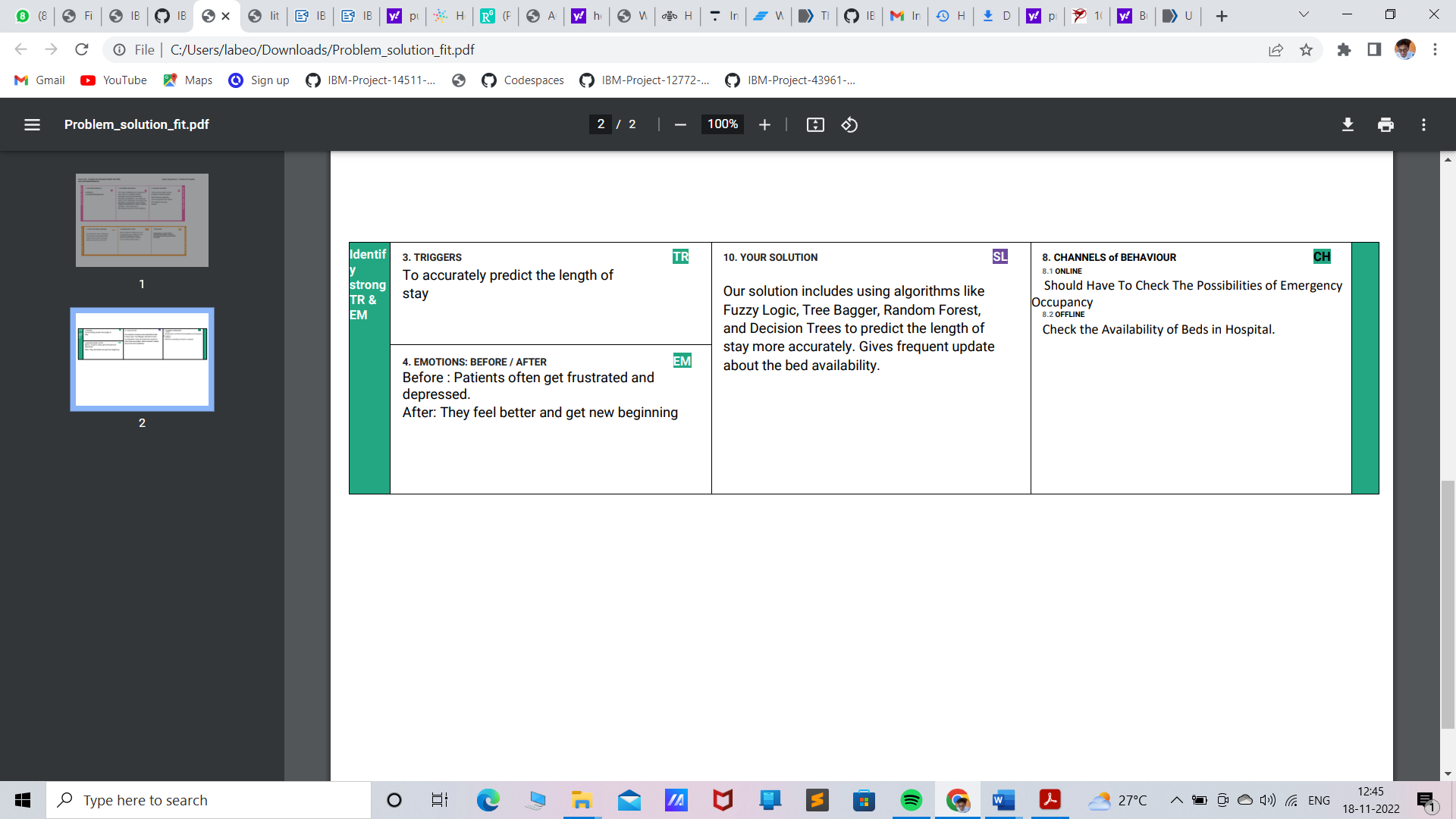
|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Description** |
| 1. | Problem Statement (Problem to be solved) | * To perform good data analytics, first of all we should teach individuals how to understand and realize the importance of dealing with such data, for instance how to deal with breast cancer. Believed that whatever and how much our patients are educated and skilled to provide us with the data we expect, medical professionals still highly need to test and clarify this data to consider it and keep it on record. |
| 2. | Idea / Solution description | * Predictive analytics supports healthcare sectors to achieve a high level of effective overall care and preventive care, as predictive systems’ results allow treatments and actions to be taken when all the risks are recognized in early stages, which aids for minimizing costs. |
| 3. | Novelty / Uniqueness | * 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 Hospital health care-Data. |

|  |  |  |
| --- | --- | --- |
| 5. | Business Model (Revenue Model) | A business model describes the resources, processes, and cost assumptions that an organization makes that will lead to the delivery of a unique value proposition to a customer. As health care organizations are beginning to transform their structure in preparation for a value-based delivery system, understanding business model theory can help in the redesign process. |
| 6. | Scalability of the Solution | * which can reflect in capturing patient’s behaviour and interactions. Another method of using predictive analytics regards how to use applications and software services alongside the electronic clinical records to analyse diagnosis and confirm outcomes in order to provide the correct treatment for the right information from huge databases |

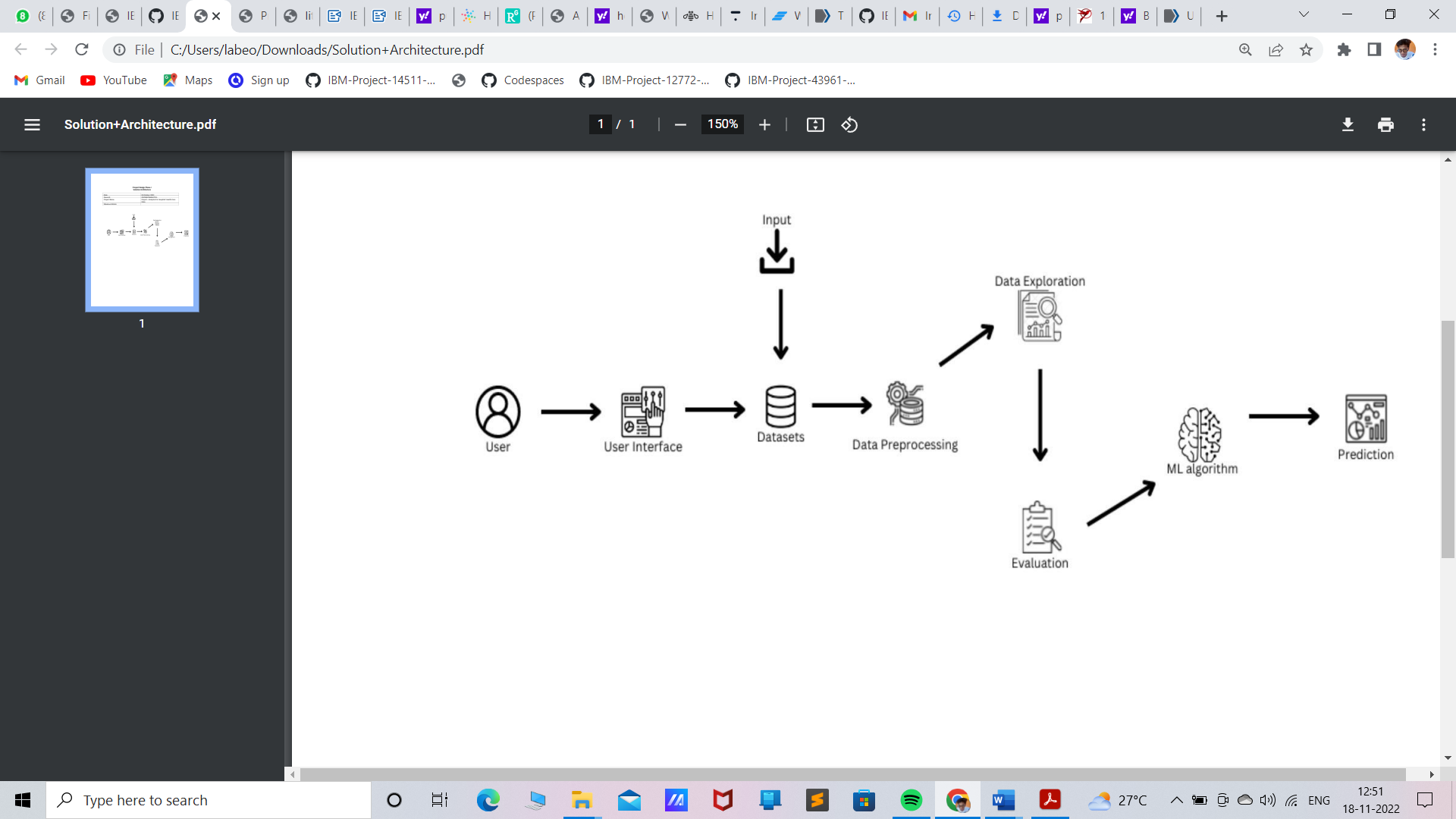
## Problem Solution Fit

****

****

****

* 1. **Solution Architecture**

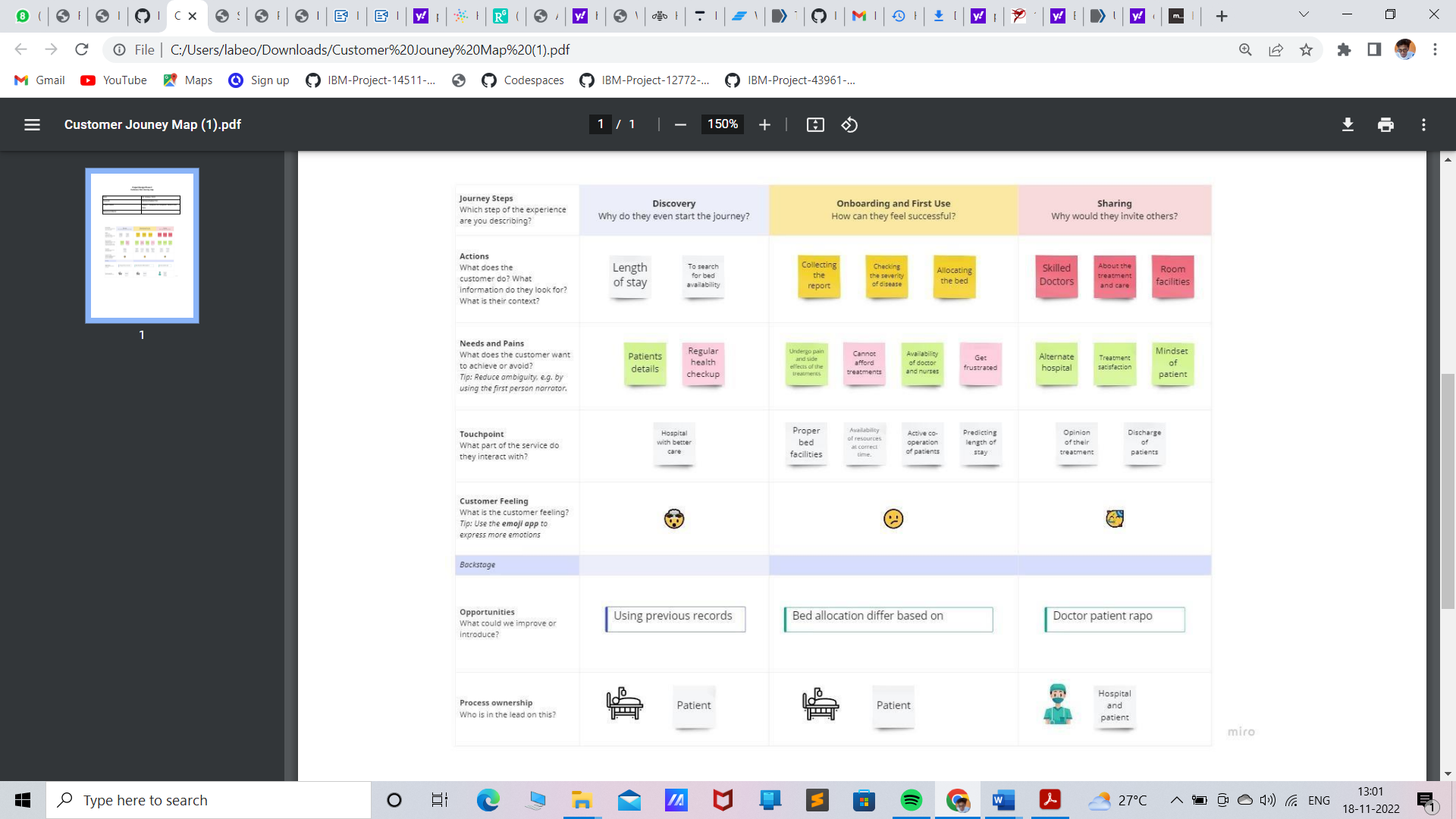
****

## PROJECT DESIGN PHASE 2

**Customer Journey Map**

Customer journey analytics uses machine learning and big data to track and analyze when and through what channels customers interact with an organization, with an aim to influence behavior (e.g., buying behaviors among retail customers). Similarly, healthcare organizations want to influence health-related behaviors, such a taking medication as prescribed and not smoking, to improve outcomes and lower the cost of care. In a partnership with an analytics services provider, a payer organization is leveraging customer journey analytics among healthcare consumers to identify the best opportunities and channels for patient outreach. With this analytics-driven engagement strategy, the payer has found an opportunity to significantly improve patient engagement—a predicted overall increase from 18 percent to 31 percent.

Quantitative research is great for adding in relevant supporting statistics and analytical data, also, it allows you to move past one or two observations/interviews and conduct broader research to get a sense of how the experience of your persona translates on a larger scale. The first step of a customer journey map is to set down the high- level processes a user goes through. These need to be fairly large chunks in the journey, to give you space to dig deeper into each section. Perhaps thinking of them as blocks in a flowchart, can help you break it down into the large sections required, then plot them on a large piece of paper, whiteboard, or use a modeling program. Also, consider at this point the high-level stakeholders that your customer interacts with on their journey. Plot these below and leave space for building on their touch points with the customer.



Typically most designers and developers of applications have some initial understanding of how a user works their way through the application. They’ve either observed the scenario themselves or received documentation of the typical path a user takes from documentation or initial reporting. Adding in additional information from qualitative research, such as from an observational or contextual inquiry study, can give more points along the timeline and highlight some critical events.

## Solution Requirements

**Functional Requirements:**

Following are the functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR**  **No.** | **Functional Requirement (Epic)** | **Sub Requirement (Story / Sub-Task)** |
| FR-1 | Dataset | Upload the dataset to the dashboard |
| FR-2 | Analysis | Data is pre-processed and cleaned. After cleaning the exploration process is carried out |
| FR-3 | Prediction | Machine learning algorithm is used for prediction |
| FR-4 | Visualization | Visualization of the prediction is shown in the dashboard created using IBM Cognos Analytics |

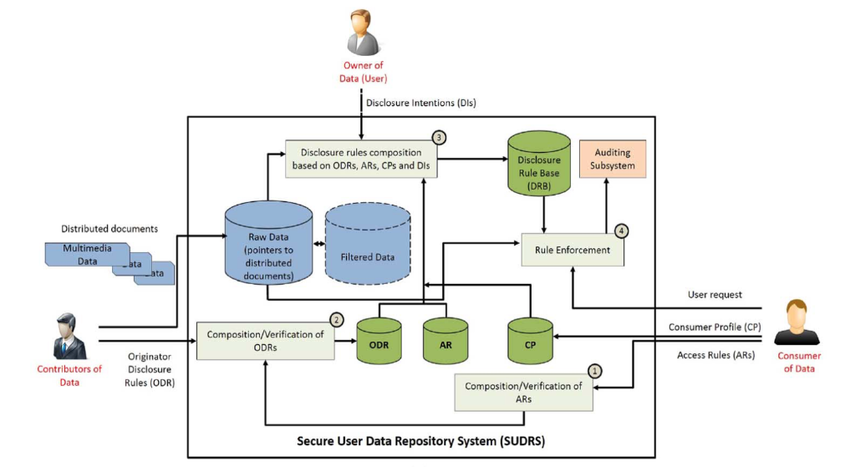
## Non-functional Requirements:

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

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Non-Functional Requirement** | **Description** |
| NFR-1 | **Usability** | Dashboards are created in order to display the length of stay prediction in visual manner. So, the prediction can be easily understood. |
| NFR-2 | **Security** | The dataset uploaded to the dashboard cannot be downloaded or accessed by external sources. |
| NFR-3 | **Reliability** | Dashboard created after the prediction process will be more reliable and shows the result clearly and effectively. |
| NFR-4 | **Performance** | The prediction has more accuracy. |
| NFR-5 | **Availability** | Predicted data will be available for sometime after the prediction. |

# 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.



Use Case:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Functional Requirement (Epic**) | **User Story Number** | **User Story / Task** | **Story Points** | **Priority** | **Team Members** |
| Sprint-1 | API and Database Connectivity | USN-1 | As an Analyzer, I use Kaggle API To gather the dataset | 3 | Medium | SABARI RAJA |
| Sprint-1 |  | USN-2 | As an analyzer, I can create database connectivity using IBM Cloud and IBM Cognos Analytics | 5 | High | VIGNESHWARAN  SIRAJ ANWARDEEN |
| Sprint-1 |  | USN-3 | As an analyzer, I can pre-process the data and Explore it. | 5 | High | SIRAJ  ANWARDEEN  SABARI RAJA |
| Sprint-2 | Dashboard | USN-4 | As an analyzer, I Can add the created visualizations to the dashboard | 8 | High | SHAKKIF IBRAHIM  SIRAJ ANWAERDEEN  VIGNESHWARAN |
| Sprint-3 | Report, Story | USN-5 | As an analyzer, I can add the create the reports for the given dataset(creating report) | 8 | High | VIGNESHWARAN  SABARI RAJA |
|  |  | USN-6 | As an analyzer, I Can create Story for given dataset | 8 | High | SIRAJ ANWAERDEEN  VIGNESHWARAN |
| Sprint-4 | Documentation | USN-7 | We create our entire project documentation | 10 | High | SIRAJ ANWAERDEEN  VIGNESHWARAN  SHAKKIF IBRAHIM |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **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 | 13 | 12 Days | 24 Oct 2022 | 04 Nov 2022 | | | |
| Sprint-2 | 8 | 2 Days | 05 Nov 2022 | 06 Nov 2022 | | | |
| Sprint-3 | 16 | 2 Days | 07 Nov 2022 | 08 Nov 2022 | | | |
| Sprint-4 | 10 | 4 Days | 09 Nov 2022 | 12 Nov 2022 | | | |

**VELOCITY:**



Average velocity for sprint-1:

AV = 13/12 = 1.08

Average velocity for sprint-2:

AV = 8/2 = 4

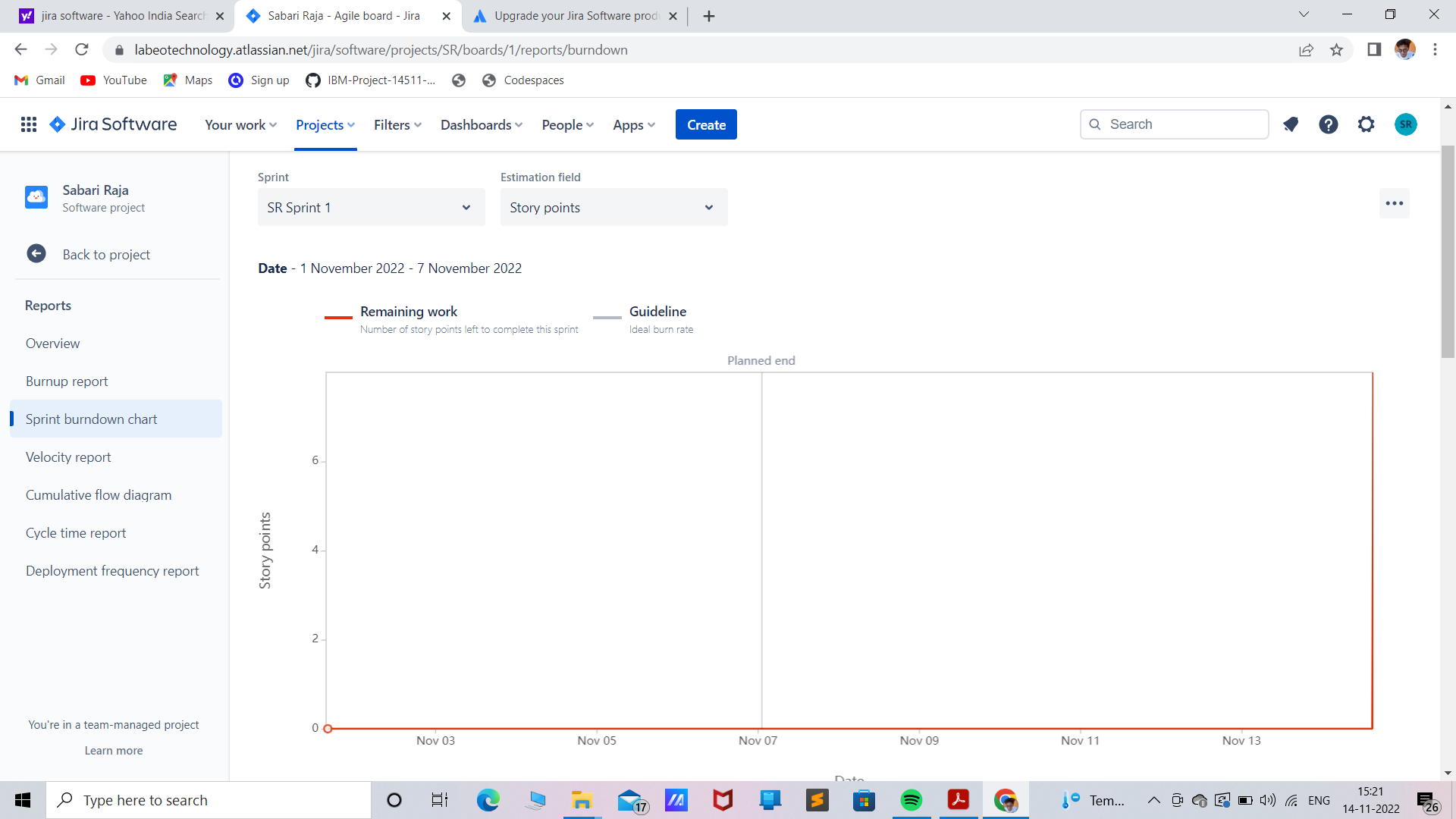
Average velocity for sprint-3:

AV = 16/2 = 8

Average velocity for sprint-4:

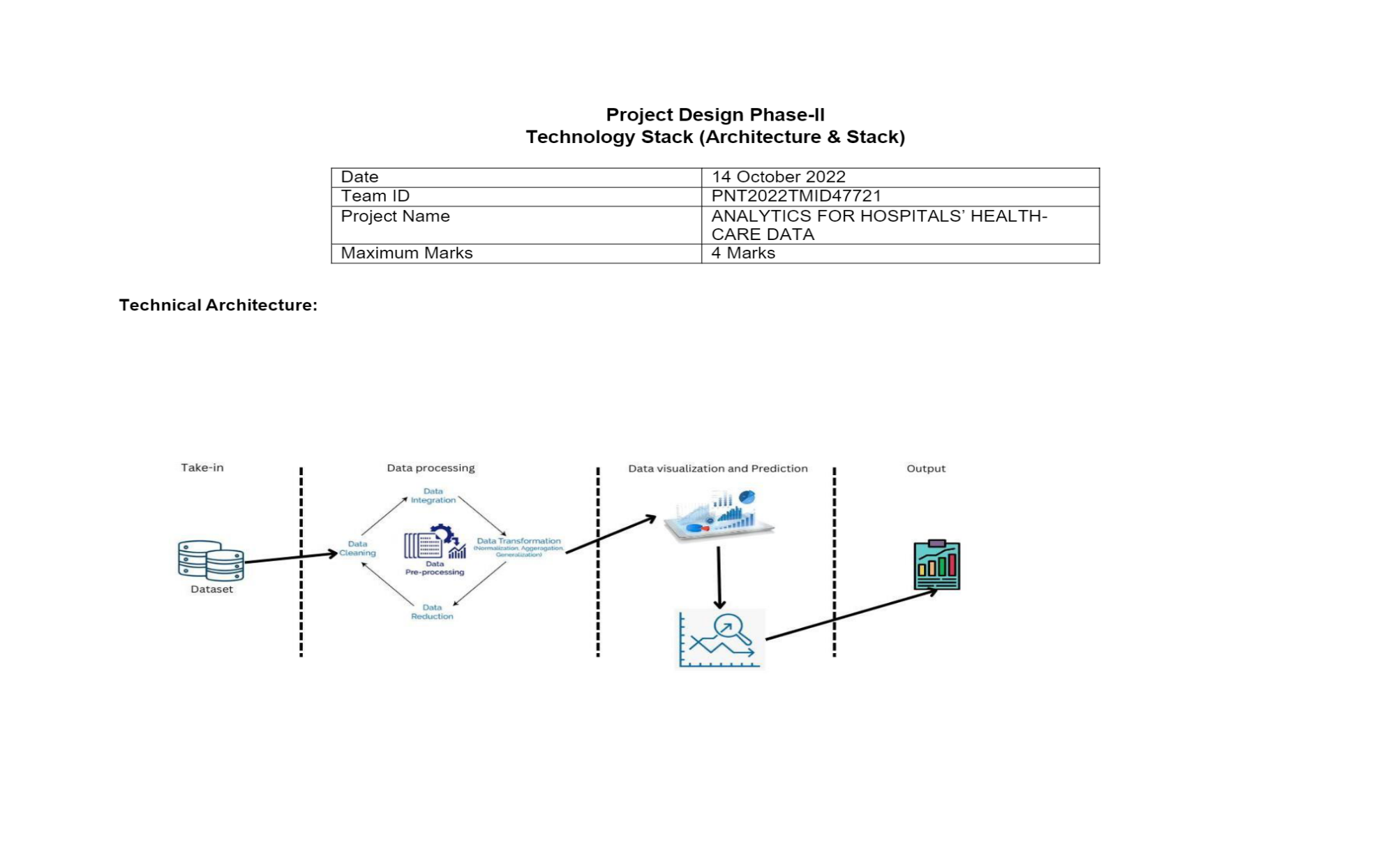
AV = 10/4 = 2.5

**SPRINT BURNDOWN CHART:**



**5.4** **TECHNOLOGY STACK**

As tumultuous as the current healthcare environment is, it’s expected to become even more complex over the next several years. Challenges such as evolving market dynamics, increasing governmental regulation and more demanding consumers will require smarter, more informed decisions from organizations so they can remain competitive and deliver value in their communities. Then, they created a library of custom predictive analytics to identify complex schemes as well as simple billing mistakes. They used both parametric and non-parametric methods for anomaly detection, which included clustering techniques to identify peer groups with normal claims behaviors whose homogeneity has been enhanced not only by specialty designation, but also practice setting, procedure mix fingerprints, geography and other relevant data-driven factors.



**Table-1: Components & Technologies:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Component** | **Description** | **Technology** |
| 1. | User Interface | The user interacts with application using Web UI | HTML, CSS, JavaScript |
| 2. | Data Processing | The data from the dataset is pre-processed | IBM Cognos Analytics |
| 3. | Cloud Database | The clean dataset is stored on IBM Cloud | IBM Cloud |
| 4. | Data visualization | The data is visualized into different forms | IBM Cognos Analytics, python |
| 5. | Prediction | ML algorithms are used for predicting the length of stay | ML algorithms – Fuzzy Logic, Tree Bagger, Random Forest, Decision Tree |

**Table-2: Application Characteristics:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Characteristics** | **Description** | **Technology** |
| 1. | Open-Source Frameworks | Open-source frameworks used | IBM Cognos Analytics, Python |
| 2. | Security Implementations | Request authentication using Encryptions | Encryptions |
| 3. | Scalable Architecture | Scalability consists of 3-tiers | Web Server – HTML, CSS, Javascript Application Server – Python |

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Characteristics** | **Description** | **Technology** |
| Database Server – IBM Cloud | | | |
| 4. | Availability | The application is available for cloud users | IBM Cloud Hosting |
| 5. | Performance | The user can know the prolonged period of stay | ML algorithms |

**PROJECT PLANING PHASE**

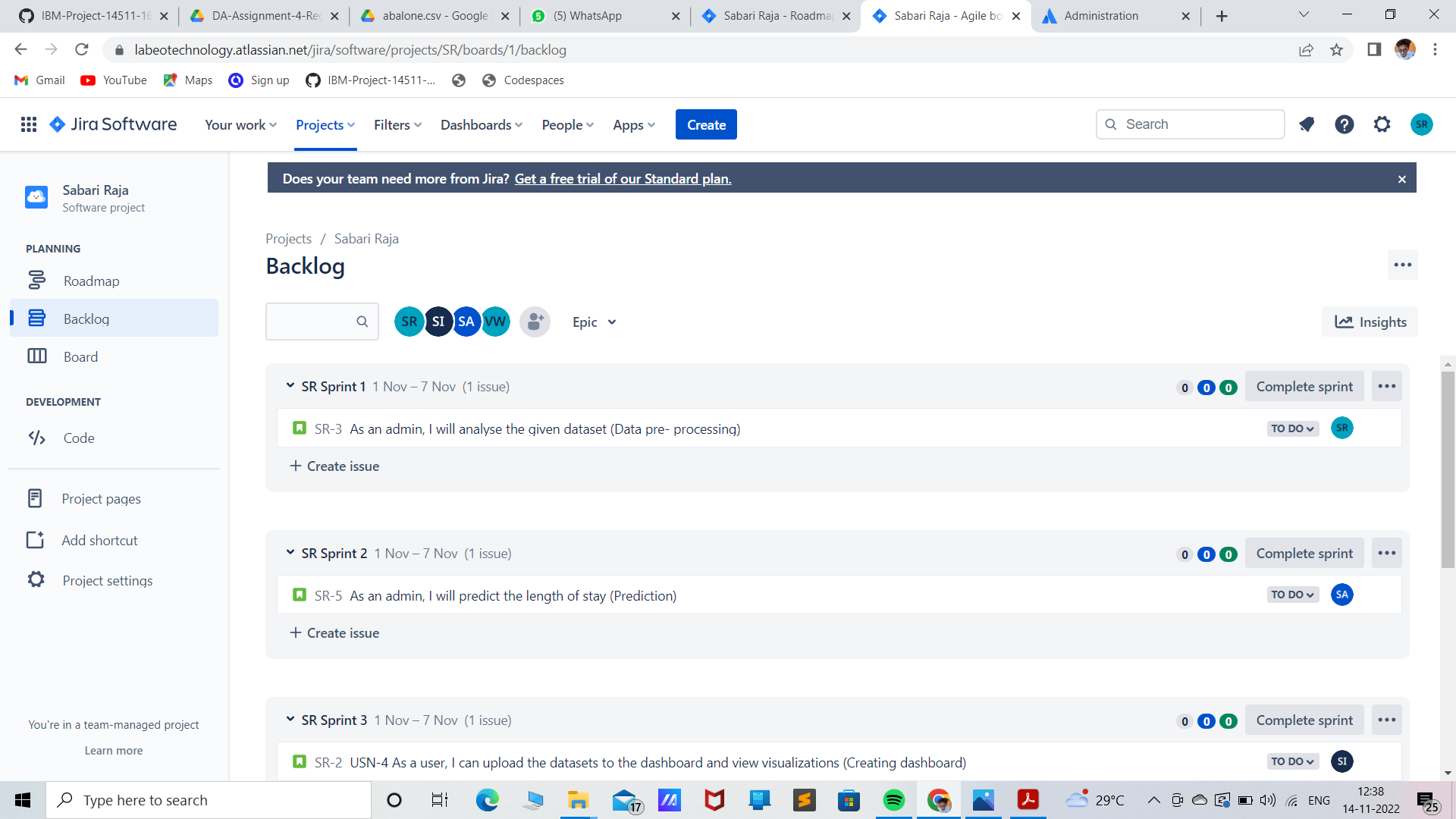
**JIRA SOFTWARE**

Jira Software is part of a product family that helps teams of all kinds to manage their jobs. Jira was originally designed as a tracker for problems and bugs. Today, however, Jira has become a powerful work management tool for all types of applications, from requirements and test case management to agile software. You will learn in this guide what Jira’s features and features can help your team with your unique needs.

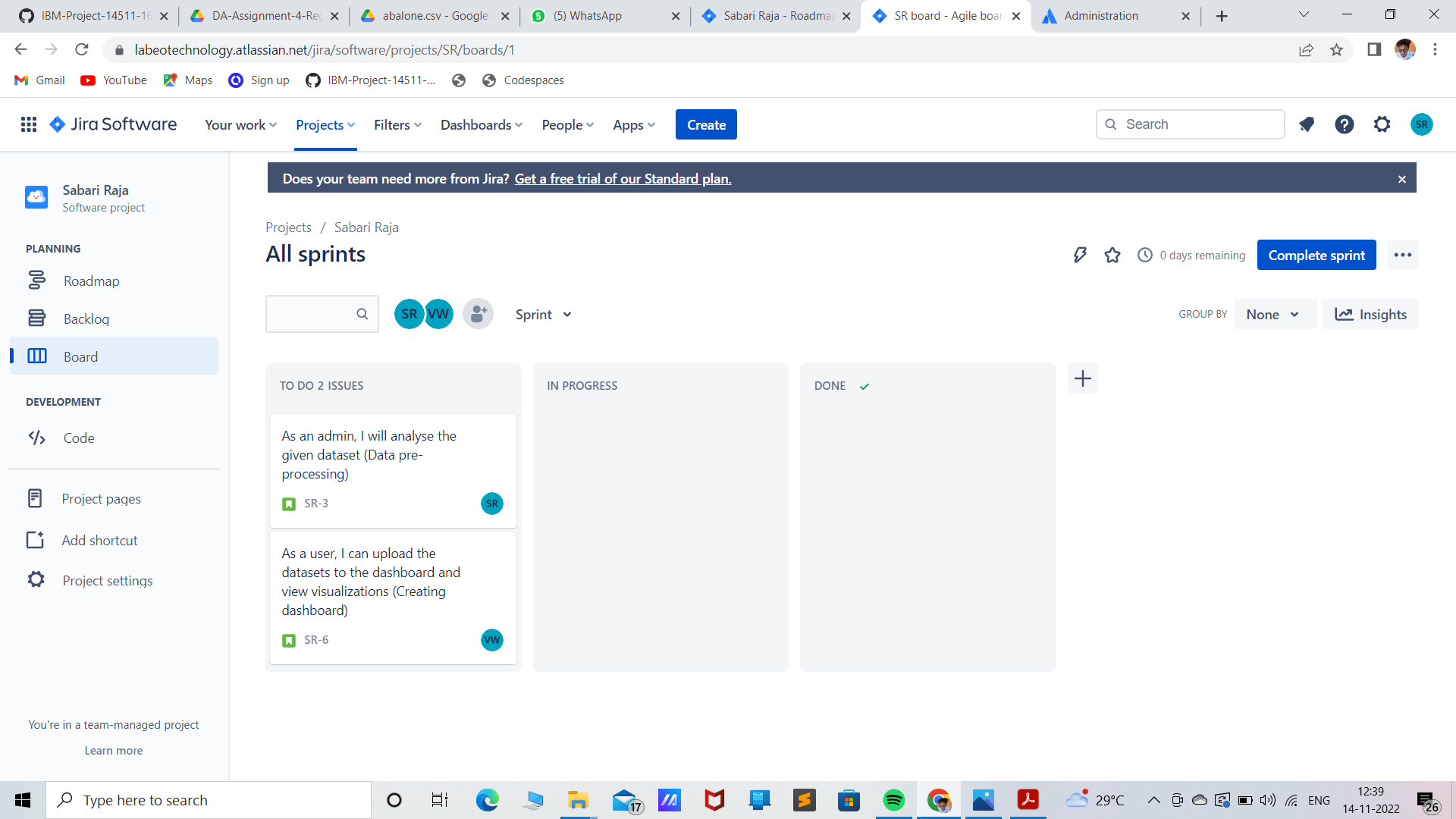
**Key points on Jira:**

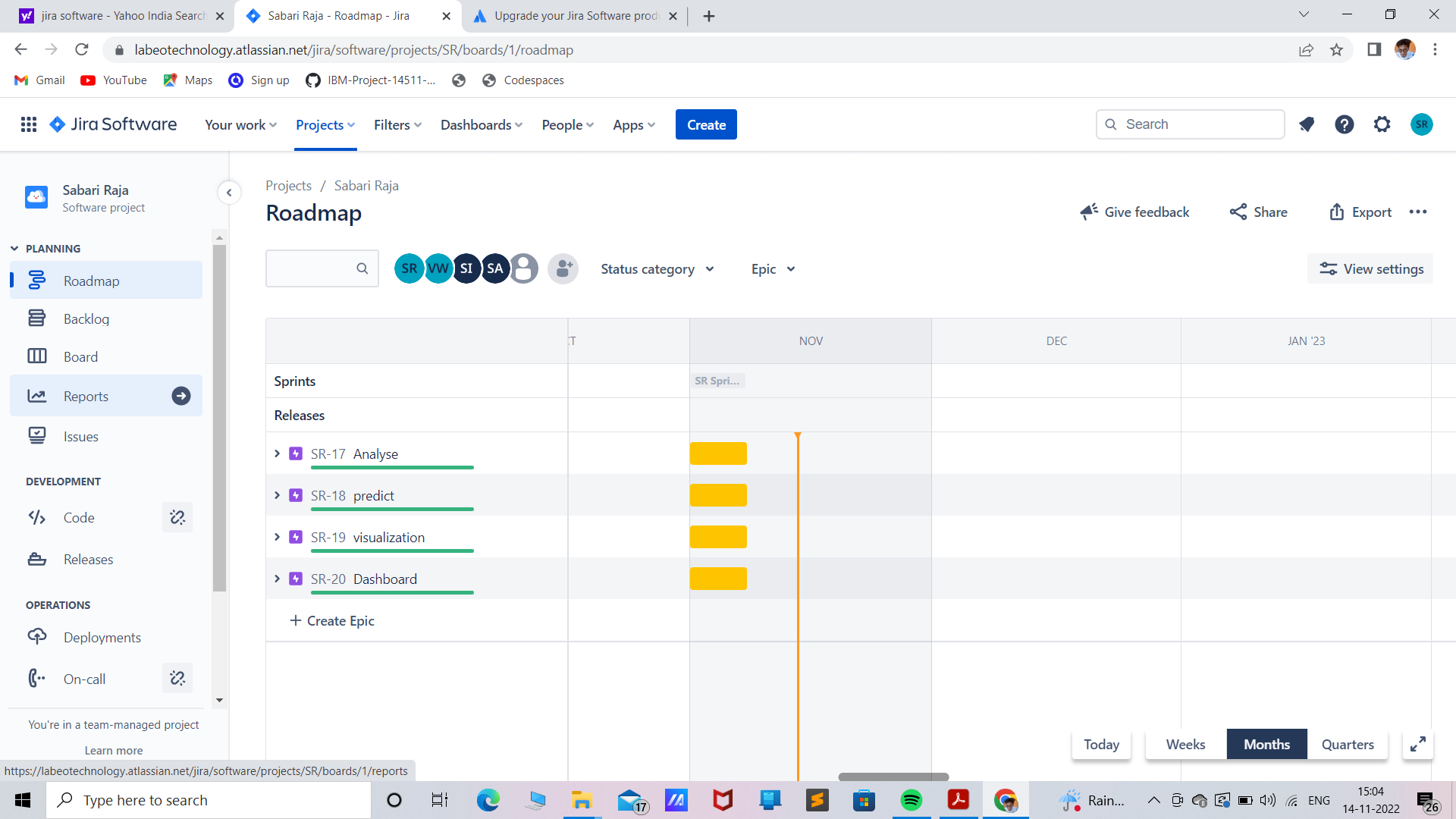
* JIRA name is a truncation of Gojira (the name for Godzilla in Japanese) that can be a reference to Bugzilla (a bug tracker and testing tool of Mozilla).
* It is a platform-independent tool; that can be worked with different types of operating systems.
* It is written in the JAVA programming language.

**BACKLOG:**

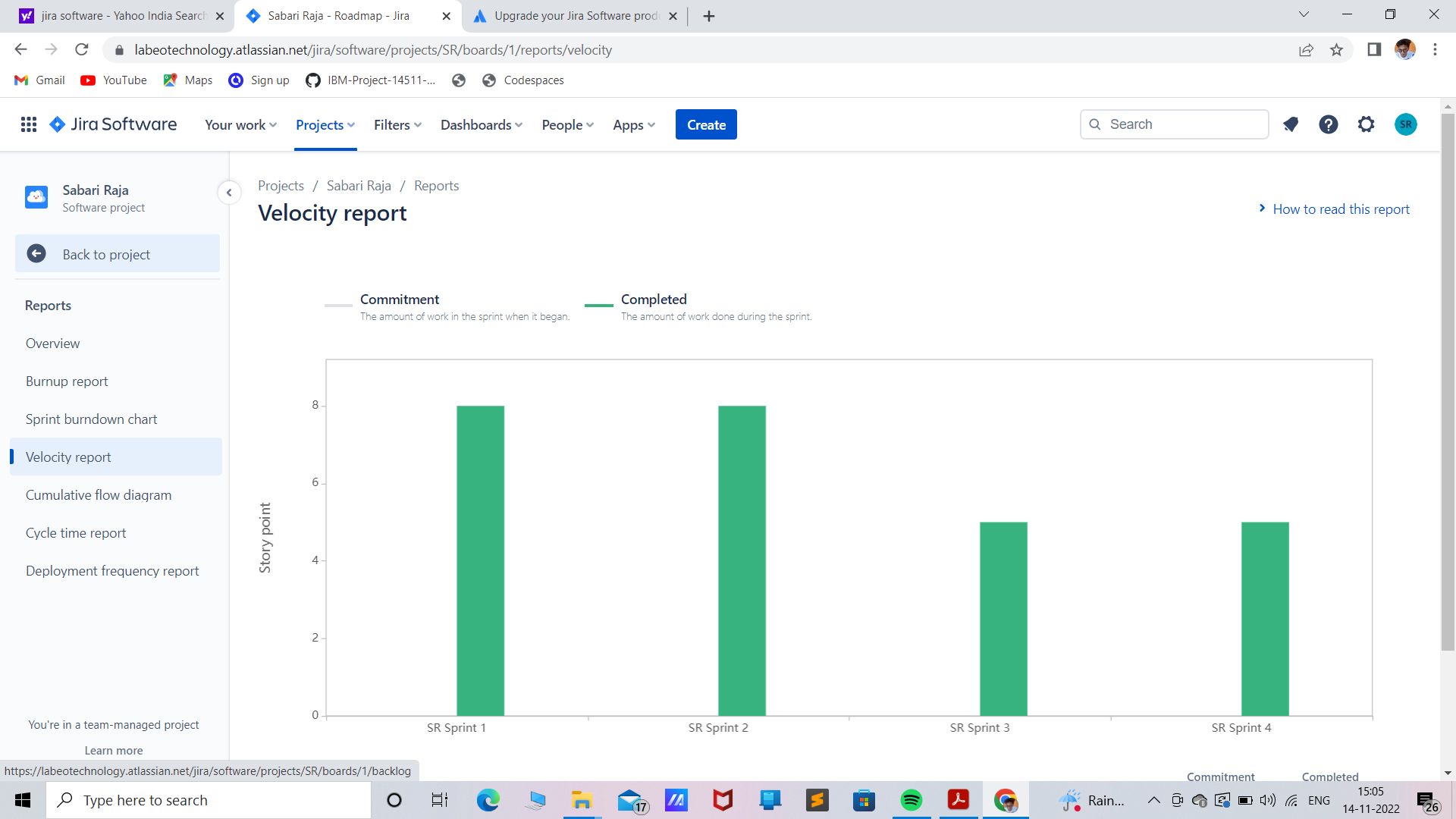


**BOARD:**



****

**VELOCITY REPORT:**

****

|  |  |
| --- | --- |
| **MILESTONES** | **TASKS** |
| MILESTONE-1 | COLLECTION OF DATA |
| MILESTONE-2 | UPLOADING THE REQUIRED DATAS ON THE PLATFORM (IBM COGNOS) |
| MILESTONE-3 | EXPLORATION AND VISUALIZATION OF DATA |
| MILESTONE-4 | CREATING THE INTERACTIVE DASHBOARD |
| MILESTONE-5 | DISPLAY THE INSIGHTS IN THE DASHBOARD |
| MILESTONE-6 | PREPARE A STANDARDIZED DATA SET AND USING THE DATA REQUIRED WITH THE HELP OF PYTHON PROGRAM |

## Technology Stack

**Technical Architecture:**

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2



Table-1: Components & Technologies:

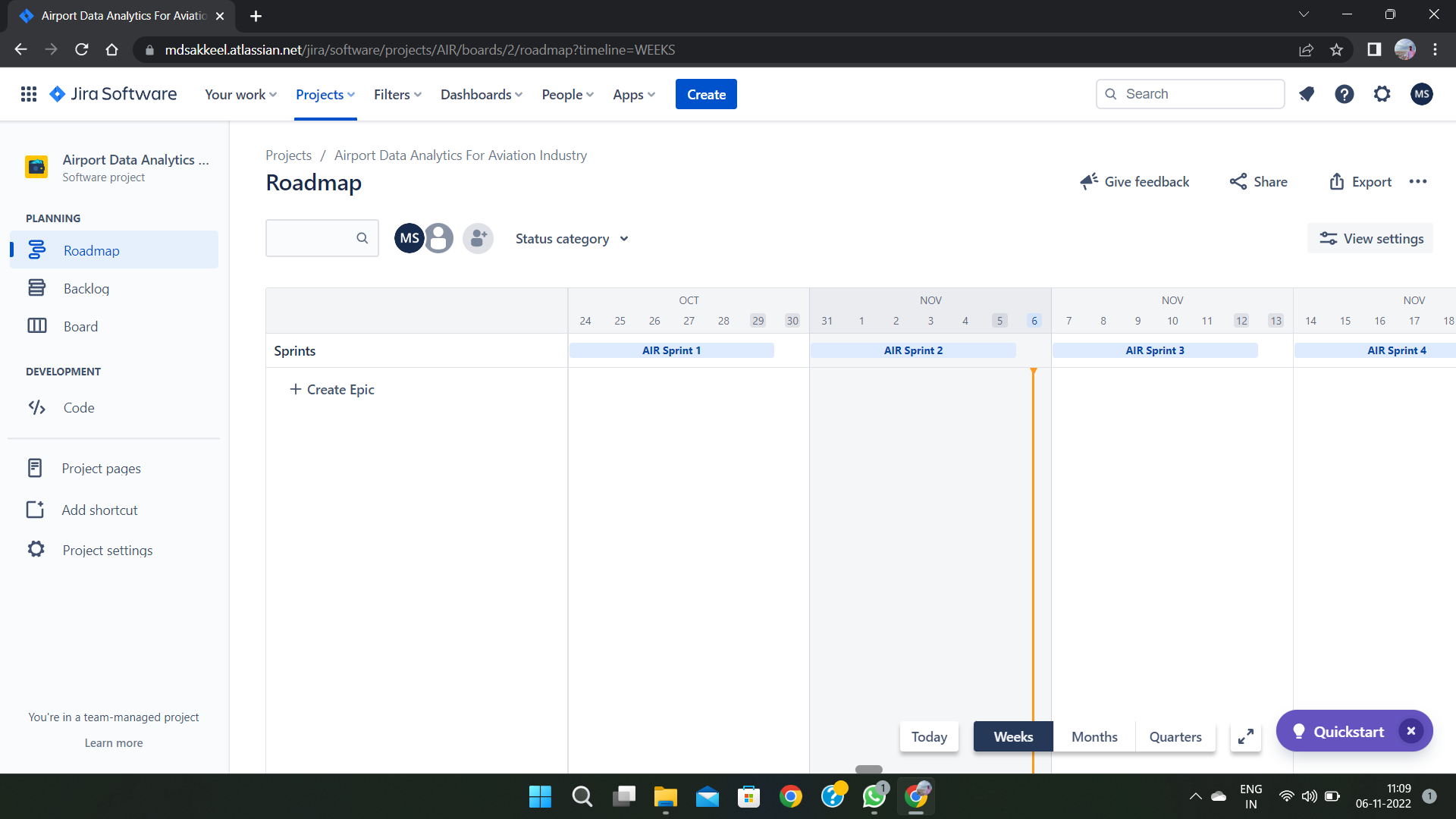
|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Components** | **Description** | **Technology** |
| 1. | User Interface | How user interacts with application.  Example: Mobile App | HTML, CSS, Java Script, Excel |
| 2. | Application Logic-1 | Logic for a process in the application | IBM Watson STT service, Python |
| 3. | Application Logic-2 | Logic for a process in the application | IBM Watson Assistant |
| 4. | Database | Data Type, Configurations | MySQL, NSQL |
| 5. | Cloud Database | Database service on cloud | IBM DB2, IBM  Cloudant |
| 6. | File Storage | File Storage requirements | IBM Blocks Storage or other storage service or Local File system |
| 7. | External API-1 | Purpose of External API used in the application | IBM Weather API |
| 8. | External API-1 | Purpose of External API used in the application | Aadhar API |
| 9. | Infrastructure (Server/Cloud) | Application Deployment on Local System/Cloud Local Server Configuration: Cloud Server Configuration | Local, Cloud Foundry |

Table-2: Application Characteristics:

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Characteristics** | **Description** | **Technology** |
| 1. | Open-Source Frameworks | List the open-source frameworks used | Technology of open- source framework |
| 2. | Security Implementations | List all the security/access controls implemented, use of firewalls. | Example: SHA-256, Encryption, IAM Controls, OWASP |
| 3. | Scalable Architecture | Justify the scalability of architecture | Cognos Used |
| 4. | Availability | Justify the availability of application (e.g: use of load balancers, distributed servers) | AWS Used |
| 5. | Performance | Design consideration for the performance of the application (number of requests per second, use of Cache, use of CDN’s) | Dashboard, Reports, Stories |

## PROJECT PLANNING PHASE

* 1. **Prepare Milestone and Activity List**



## Sprint Delivery Plan

**Product Backlog, Sprint Schedule, and Estimation (4 Marks)**

Use the below template to create product backlog and sprint schedule

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | Team Members |
| Sprint-1 | Registration | USN-1 | As a user, I can register for the application by entering my email, password, and confirming my password. | 3 | High | SHIYAMGANESH M |
| Sprint-1 | Registration | USN-2 | As a user, I will receive confirmation email once I have registered for the application | 3 | High | MOHAMED FARHAN S |
| Sprint-1 | Login | USN-3 | As a user, I can log into the application by entering email & password | 1 | Low | MOHAMED MUSHARAF M |
| Sprint-1 | Accessing the dataset | USN-4 | I can access the dataset and choose the different types of exploration can be done is analyzed as a user. | 5 | Medium | MOHAMED SAKKEEL R |
| Sprint-2 | Exploration | USN-5 | I can explore the given dataset through IBM Cognos Analytics with Watson | 6 | High | MOHAMED FARHAN S |
| Sprint-2 | Visualization | USN-6 | I will use Cognos as a visualization tool for the provided dataset into a dashboard | 6 | High | SHIYAMGANESH M |
| Sprint-3 | Dashboard | USN-7 | I can create the dashboard that is visualized as a user | 6 | High | MOHAMED SAKKEEL R |
| Sprint-3 | Ease of Access | USN-8 | I can simply access and use the dashboard as a user | 5 | Medium | MOHAMED MUSHARAF M |
| Sprint-4 | Generation of Report | USN-9 | I can generate the report with the help of my visualization | 6 | High | MOHAMED SAKKEEL R |
| Sprint-4 | Dashboard Establishment | USN-10 | As a developer I can Established the dashboard into a website and submit the website | 6 | High | SHIYAMGANESH M |

**Project Tracker, Velocity & Burndown Chart: (4 Marks)**

| **Sprint** | **Total Story Points** | **Duration** | **Sprint Start Date** | **Sprint End Date (Planned)** | **Story Points Completed (as on Planned End Date)** | **Sprint Re lease Date (Actual)** |
| --- | --- | --- | --- | --- | --- | --- |
| Sprint-1 | 20 | 6 Days | 24 Oct 2022 | 29 Oct 2022 | 15 | 29 Oct 2022 |
| Sprint-2 | 20 | 6 Days | 31 Oct 2022 | 05 Nov 2022 | 15 | 05 Nov 2022 |
| Sprint-3 | 20 | 6 Days | 07 Nov 2022 | 12 Nov 2022 | 15 | 12 Nov 2022 |
| Sprint-4 | 20 | 6 Days | 14 Nov 2022 | 19 Nov 2022 | 15 | 19 Nov 2022 |

**Velocity:**

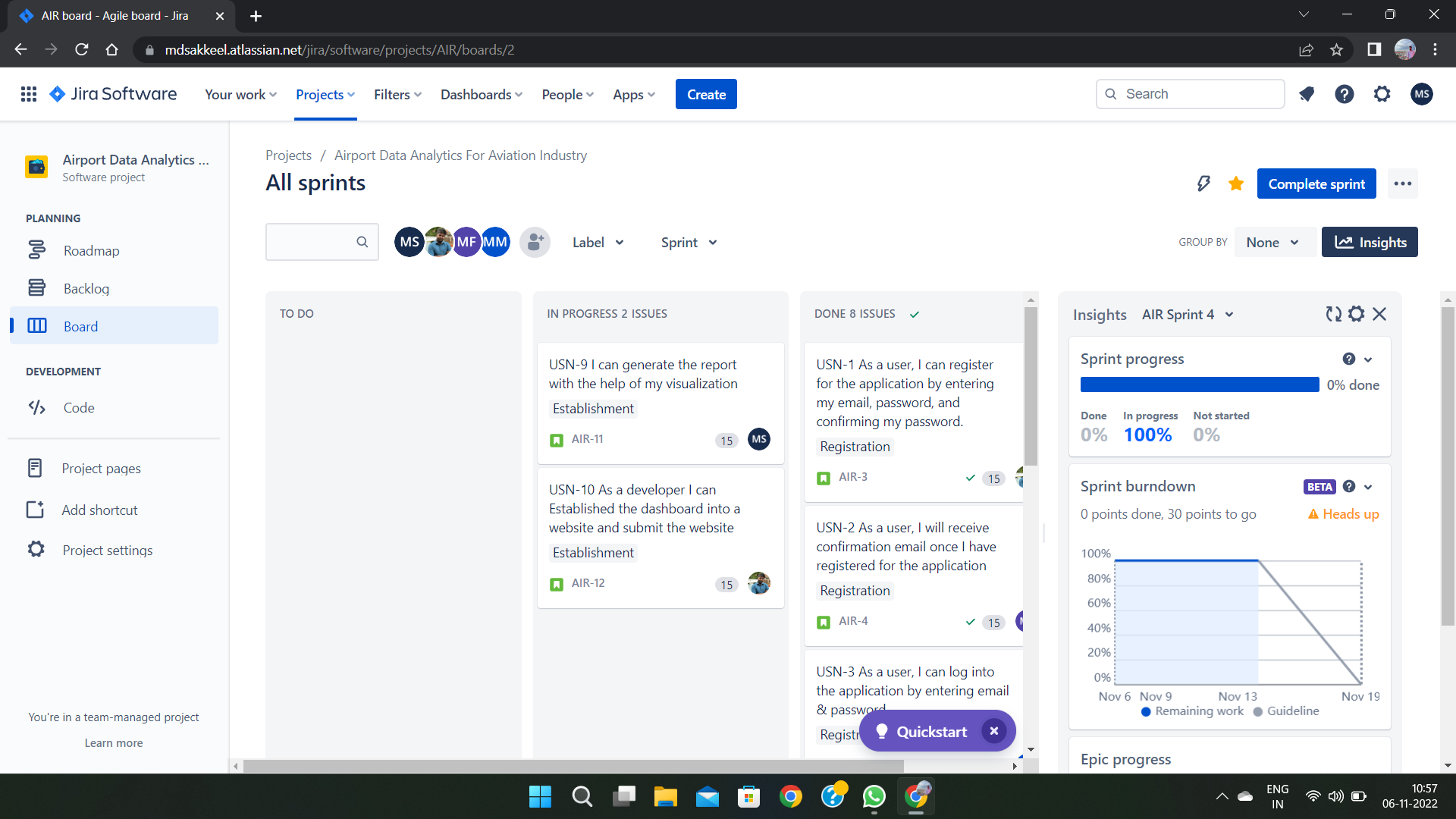
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let’s calculate the team’s average velocity (AV) per iteration unit (story points per day)



Average velocity=Sprint duration / velocity=15/6=2.5

**Burndown Chart:**

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile [software development](https://www.visual-paradigm.com/scrum/what-is-agile-software-development/) methodologies such as [Scrum.](https://www.visual-paradigm.com/scrum/scrum-in-3-minutes/) However, burn down charts can be applied to any project containing measurable progress over time.



## PROJECT DEVELOPMENT PHASE

* 1. **Project Development - Delivery of Sprint – 1**

**Flight count from Top 5 Airlines at Top 5 Airports**

WITH top\_5\_airports AS (

SELECT ORIGIN, COUNT(ORIGIN) AS count FROM

airline-delay-canc.airlines\_data.delay\_canc\_data GROUP BY

1

HAVING

count > 100000 ORDER BY

2 DESC

LIMIT 5

),

top\_5\_airlines AS (

SELECT

OP\_CARRIER, COUNT(OP\_CARRIER) AS count

FROM

airline-delay-canc.airlines\_data.delay\_canc\_data main, top\_5\_airports top5

WHERE

top5.ORIGIN = main.ORIGIN GROUP BY

1

ORDER BY

2 DESC

LIMIT 5

),

airportwise\_carrier\_cnt AS

SELECT

main.ORIGIN AS Airport, main.OP\_CARRIER AS Carrier, COUNT(\*) AS count

FROM

airline-delay-canc.airlines\_data.delay\_canc\_data main, top\_5\_airports top5\_ap,

top\_5\_airlines top\_al

WHERE

top5\_ap. ORIGIN = main. ORIGIN

AND top\_al.OP\_CARRIER = main.OP\_CARRIER GROUP BY

1,

2

),

resut\_cte AS (

SELECT

Airport, Carrier, count,

RANK() OVER (PARTITION BY Airport ORDER BY count) AS rank

FROM

)

Airportwise\_carrier\_cnt

SELECT

Airport, Carrier, count

FROM

resut\_cte

WHERE

rank < 6

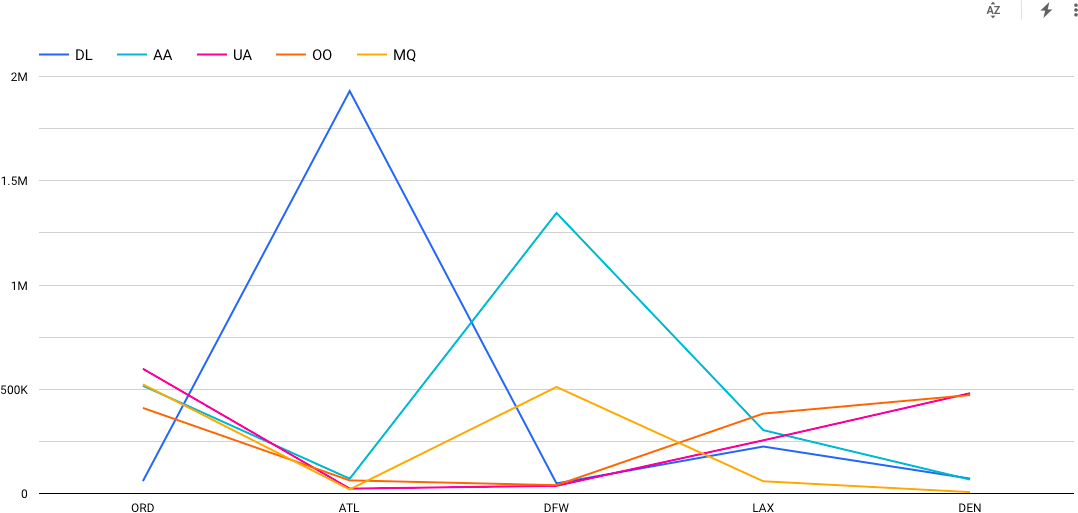
**Results**

*Top 5 Airports with maximum flight count:*

* + 1. **ORD** (O'Hare International Airport)
    2. **ATL** (Hartsfield-Jackson Atlanta International Airport)
    3. **DFW** (Dallas/Fort Worth International Airport)
    4. **LAX** (Los Angeles International Airport)
    5. **DEN** (Denver International Airport)

*Top 5 Airlines with maximum flight count:*

1. **DL** (Delta Air Lines)
2. **AA** (American Airlines)
3. **UA** (United Airlines)
4. **OO** (SkyWest Airlines)
5. **MQ** (American Eagle Airlines)



* From the above, it is realized that on **Delta Airlines** has the highest flight frequence on the **Atlanta** airport.

**Top 5 Airports with Maximum Cancellations (decreasing order)**

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

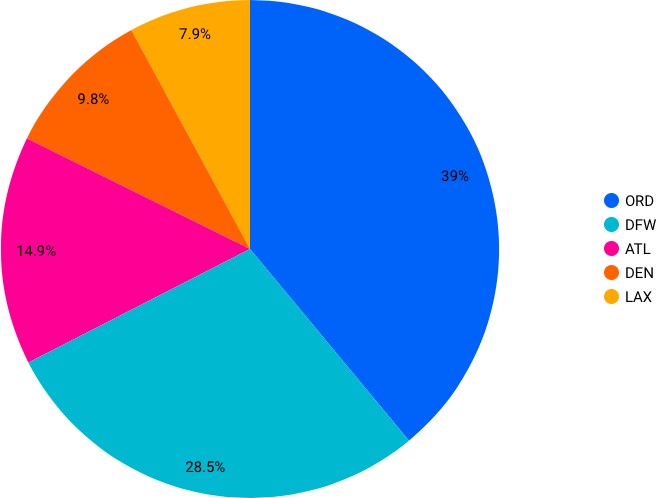
FROM

all\_flights af, cancelled\_flights cf WHERE

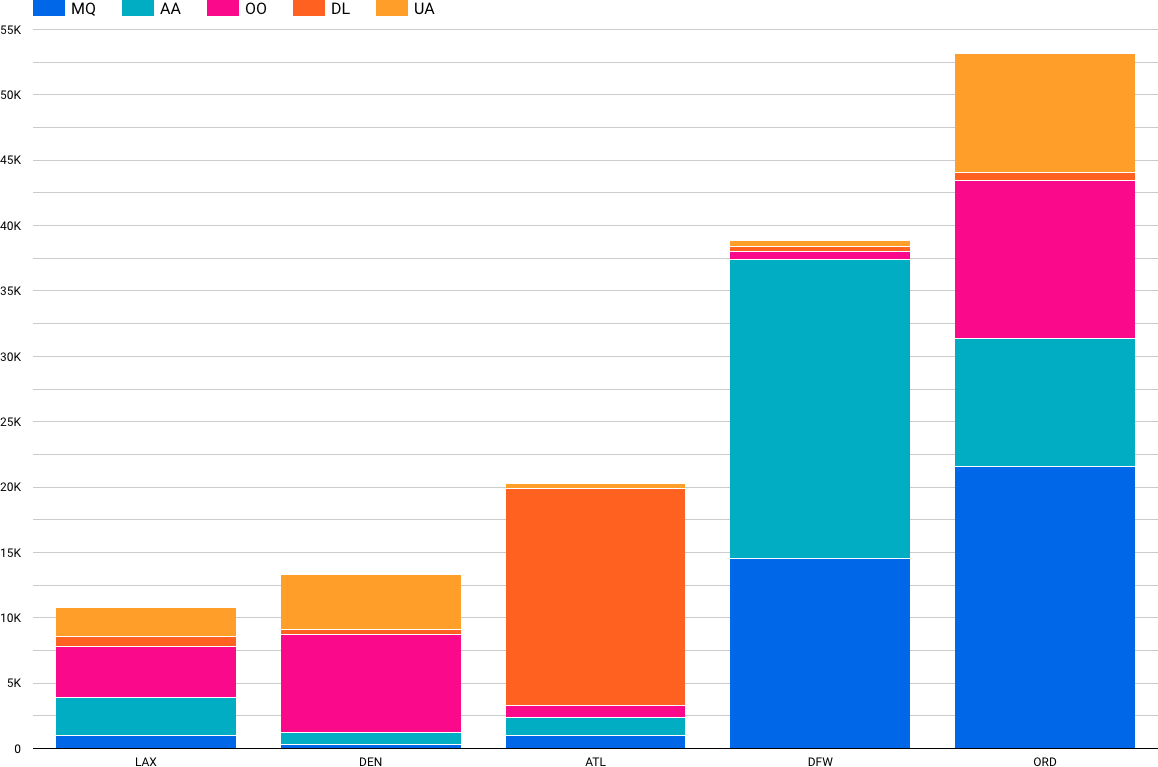
af.Airport = cf.Airport AND af.Carrier = cf.Carrier

**Results**

| S No.| Airport Code | Airport Name | Cancellation (in %) | | - | - | - | - | - | 1. | **ORD** | (O'Hare International Airport) | 39| | 2. | **DFW** | (Dallas/Fort Worth International Airport) | 28.5| | 3. | **ATL** | (Hartsfield-Jackson Atlanta International Airport) | 14.9| | 4. | **DEN** | (Denver International Airport) | 9.8| | 5. | **LAX** | (Los Angeles International Airport) | 7.9|



*Airline-wise Cancellation Bifurcation*



## Project Development - Delivery of Sprint – 2

**Top Cancellation Reasons for Top 5 Busiest Airports**

**Query - JS UDF Function**

## CREATE TEMP FUNCTION

**cancellation\_reason(code string) RETURNS string**

## LANGUAGE js AS """

**switch(code) {**

## case "A":

**return "Airline/Carrier"; break;**

## case "B":

**return "Weather"; break;**

## case "C":

**return "National Air System"; break;**

## case "D":

**return "Security"; break;**

## default:

**return "Others"; break;**

## }

**"""; WITH**

## top\_5\_airports AS ( SELECT

**ORIGIN, COUNT(ORIGIN) AS count FROM**

## `airline-delay-canc.airlines\_data.delay\_canc\_data` GROUP BY

**1**

## HAVING

**count > 100000 ORDER BY**

## 2 DESC LIMIT

**5 ) SELECT**

## top5.ORIGIN,

**cancellation\_reason(main.CANCELLATION\_CODE) AS reason, COUNT(main.CANCELLATION\_CODE) AS count**

## FROM

**`airline-delay-canc.airlines\_data.delay\_canc\_data` main, top\_5\_airports top5**

## WHERE

**CANCELLED = 1 AND EXTRACT(year FROM**

## FL\_DATE) = 2018

**AND top5.ORIGIN = main.ORIGIN GROUP BY**

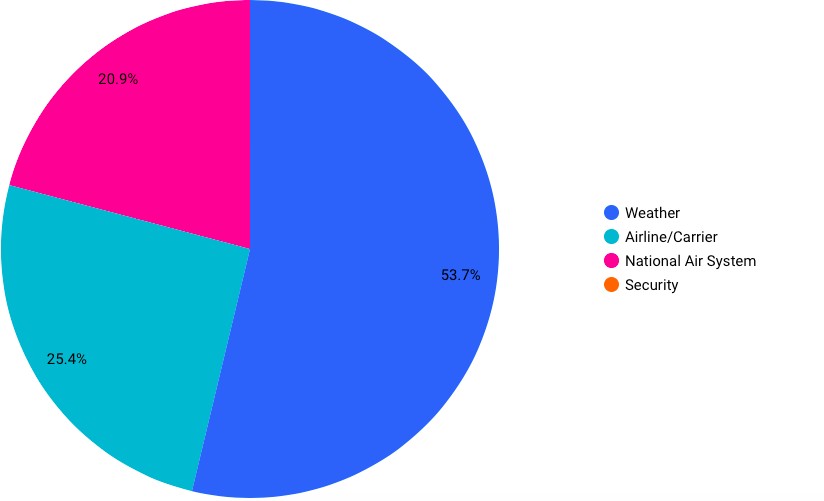
## 1,

**2**

## ORDER BY 1,

**2**

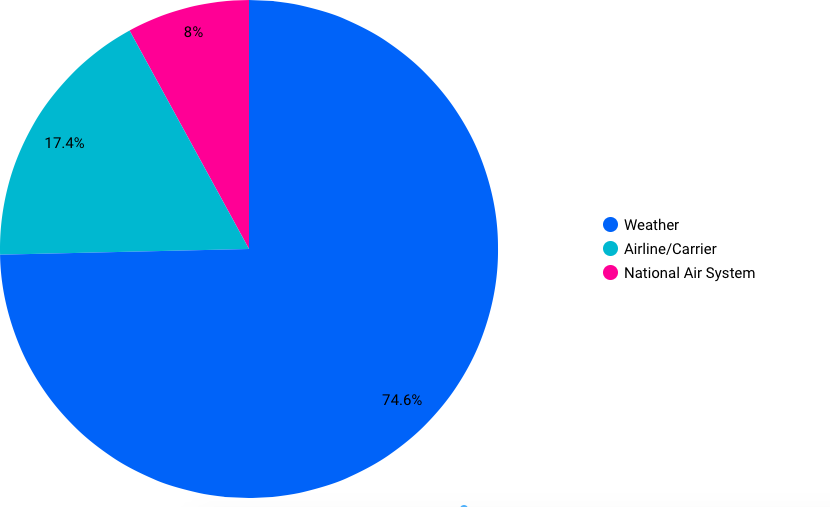
**Result**

| S No.| Reason | Cancellation (in %) | | - | - | - | - | 1. | **Weather** | 53.7| | 2. | **Airline/Carrier Delays** | 25.4| | 3. | **National Air System** | 20.9| | 4. | **Airport Secutiy** | 0.01 (~ 0)|

Top Cancellation Reasons at the Most Busiest Airport in practice (Atlanta)

* Atlanta is one of the largest inter-connect point (airport) for domestic and international flights in USA.

|S No.| Reason | Cancellation (in %) | | - | - | - | - | 1. | **Weather** | 74.6| | 2. | **Airline/Carrier Delays** | 17.4| | 3. | **National Air System** | 8|



## Project Development - Delivery of Sprint – 3

**Overall Delays at Top 5 Airports for top 5 airlines**

**Query**

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

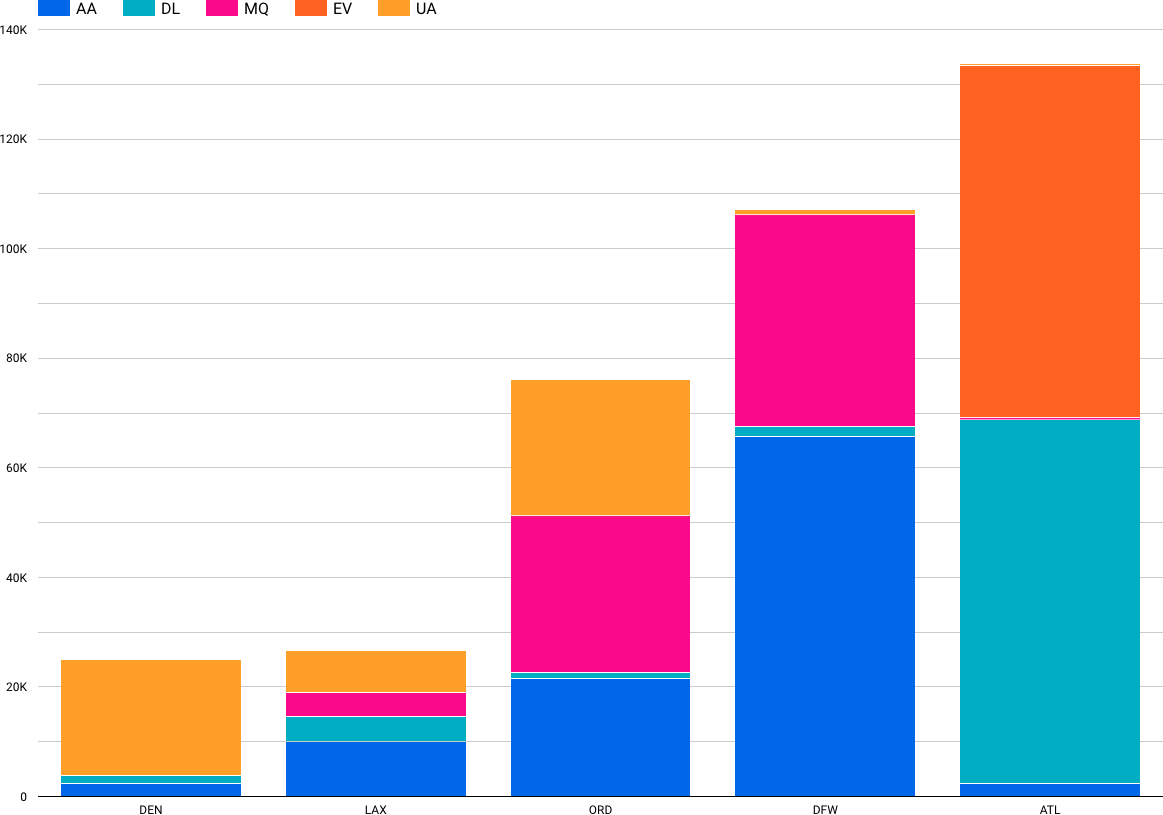
all\_flights af, delayed\_flights df

WHERE

af.Airport = df.Airport

AND af.Carrier = df.Carrier

*Overall Delays at Top 5 Airports with top 5 airlines*



**Overall Delay Time Frequency with Top 5 Airports**

**Query**

CREATE TEMP FUNCTION delay\_bifurcation(slot\_cnt ARRAY<STRUCT<slot int64,count int64>>)

RETURNS STRUCT<cnt\_1\_30 float64, cnt\_30\_2 float64, cnt\_2\_5 float64, cnt\_5\_24 float64, cnt\_24 float64>

LANGUAGE js AS """

let response = {"cnt\_1\_30": 0.0, "cnt\_30\_2": 0.0, "cnt\_2\_5": 0.0, "cnt\_5\_24": 0.0, "cnt\_24":

0.0}

for(let i = 0 ; i < slot\_cnt.length; i++){ let slotCntObj = slot\_cnt[i];

let result = slotCntObj.count; switch(parseInt(slotCntObj.slot)){

case 1:

response["cnt\_1\_30"] = result; break;

case 2:

response["cnt\_30\_2"] = result; break;

case 3:

response["cnt\_2\_5"] = result; break;

case 4:

response["cnt\_5\_24"] = result; break;

case 5:

response["cnt\_24"] = result; break;

default:

response["cnt\_1\_30"] = 0.0;

response["cnt\_30\_2"] = 0.0;

response["cnt\_2\_5"] = 0.0;

response["cnt\_5\_24"] = 0.0;

response["cnt\_24"] = 0.0; break;

}

}

return response

""";

WITH top\_5\_airports as (

SELECT ORIGIN, count(ORIGIN) as count

FROM `airline-delay-canc.airlines\_data.delay\_canc\_data` Group by 1

having count > 100000 order by 2 desc

limit 5

),

delay\_bifurcation as ( select ORIGIN,

(case when ARR\_DELAY > 1440 then 5 when ARR\_DELAY > 300 then 4

when ARR\_DELAY > 240 then 3

when ARR\_DELAY > 30 then 2 else 1 end) as slot

from `airline-delay-canc.airlines\_data.delay\_canc\_data` where ARR\_DELAY is not null and ARR\_DELAY > 0

-- and EXTRACT(year FROM FL\_DATE) = 2018

),

airport\_timeslots as(

select db.ORIGIN, db.slot, count(db.slot) as count from delay\_bifurcation db,top\_5\_airports top5 where top5.ORIGIN = db.ORIGIN

group by 1,2),

airport\_struct as(

select origin, struct(slot,count) as slot\_cnt from airport\_timeslots

),

udf\_result as (select origin, delay\_bifurcation(ARRAY\_AGG(slot\_cnt)) as slot\_struct from airport\_struct

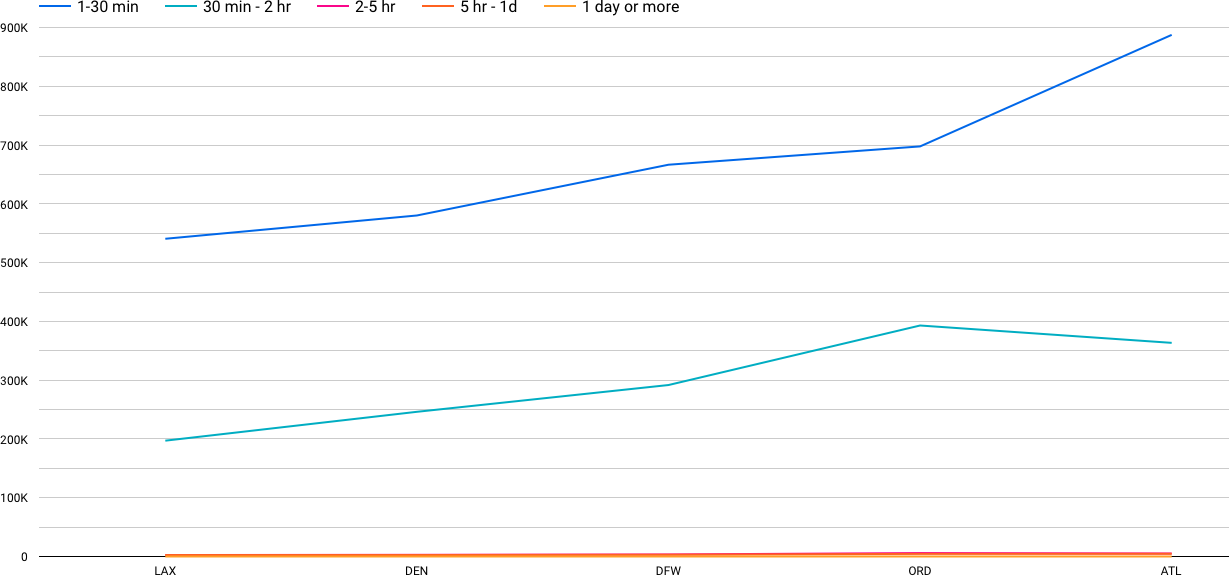
group by 1

)

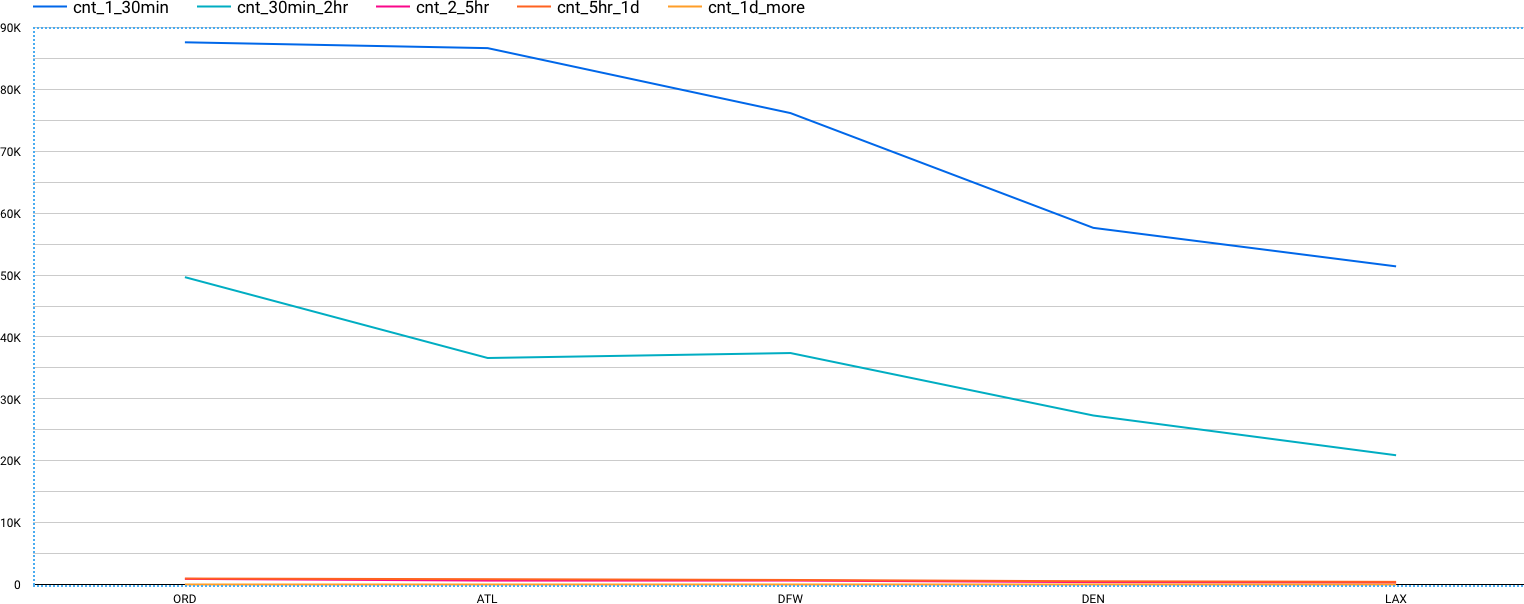
select origin, slot\_struct.cnt\_1\_30 as cnt\_1\_30min, slot\_struct.cnt\_30\_2 as cnt\_30min\_2hr, slot\_struct.cnt\_2\_5 as cnt\_2\_5hr, slot\_struct.cnt\_5\_24 as cnt\_5hr\_1d, slot\_struct.cnt\_24 as cnt\_1d\_more

from udf\_result

*Overall Delay Time Frequency with Top 5 Airports (UDF Function)*



*Overall Delay Frequency (Year with max delays and cancellations)*



## Project Development - Delivery of Sprint – 4

**Delay Percentage for top 5 airports**

**Query**

CREATE TEMP FUNCTION delay\_bifurcation(slot\_cnt ARRAY<STRUCT<slot int64,count int64>>)

RETURNS STRUCT<cnt\_1\_30 float64, cnt\_30\_2 float64, cnt\_2\_5 float64, cnt\_5\_24 float64, cnt\_24 float64>

LANGUAGE js AS """

let response = {"cnt\_1\_30": 0.0, "cnt\_30\_2": 0.0, "cnt\_2\_5": 0.0, "cnt\_5\_24": 0.0, "cnt\_24":

0.0}

let total\_delayed\_flights = 0;

for(let i = 0 ; i < slot\_cnt.length; i++){ total\_delayed\_flights += parseInt(slot\_cnt[i].count);

}

for(let i = 0 ; i < slot\_cnt.length; i++){ let slotCntObj = slot\_cnt[i];

let result = parseFloat(parseInt(slotCntObj.count) / total\_delayed\_flights \* 100).toFixed(2);

switch(parseInt(slotCntObj.slot)){ case 1:

response["cnt\_1\_30"] = result; break;

case 2:

response["cnt\_30\_2"] = result; break;

case 3:

response["cnt\_2\_5"] = result; break;

case 4:

response["cnt\_5\_24"] = result; break;

case 5:

response["cnt\_24"] = result; break;

default:

response["cnt\_1\_30"] = 0.0;

response["cnt\_30\_2"] = 0.0;

response["cnt\_2\_5"] = 0.0;

response["cnt\_5\_24"] = 0.0;

response["cnt\_24"] = 0.0; break;

}

}

return response

""";

WITH top\_5\_airports as (

SELECT ORIGIN, count(ORIGIN) as count

FROM `airline-delay-canc.airlines\_data.delay\_canc\_data` Group by 1

having count > 100000 order by 2 desc

limit 5

),

delay\_bifurcation as ( select ORIGIN,

(case when ARR\_DELAY > 1440 then 5 when ARR\_DELAY > 300 then 4

when ARR\_DELAY > 240 then 3

when ARR\_DELAY > 30 then 2 else 1 end) as slot

from `airline-delay-canc.airlines\_data.delay\_canc\_data` where ARR\_DELAY is not null and ARR\_DELAY > 0

-- and EXTRACT(year FROM FL\_DATE) = 2018 -- used for filtering

),

airport\_timeslots as(

select db.ORIGIN, db.slot, count(db.slot) as count from delay\_bifurcation db,top\_5\_airports top5 where top5.ORIGIN = db.ORIGIN

group by 1,2),

airport\_struct as(

select origin, struct(slot,count) as slot\_cnt from airport\_timeslots

),

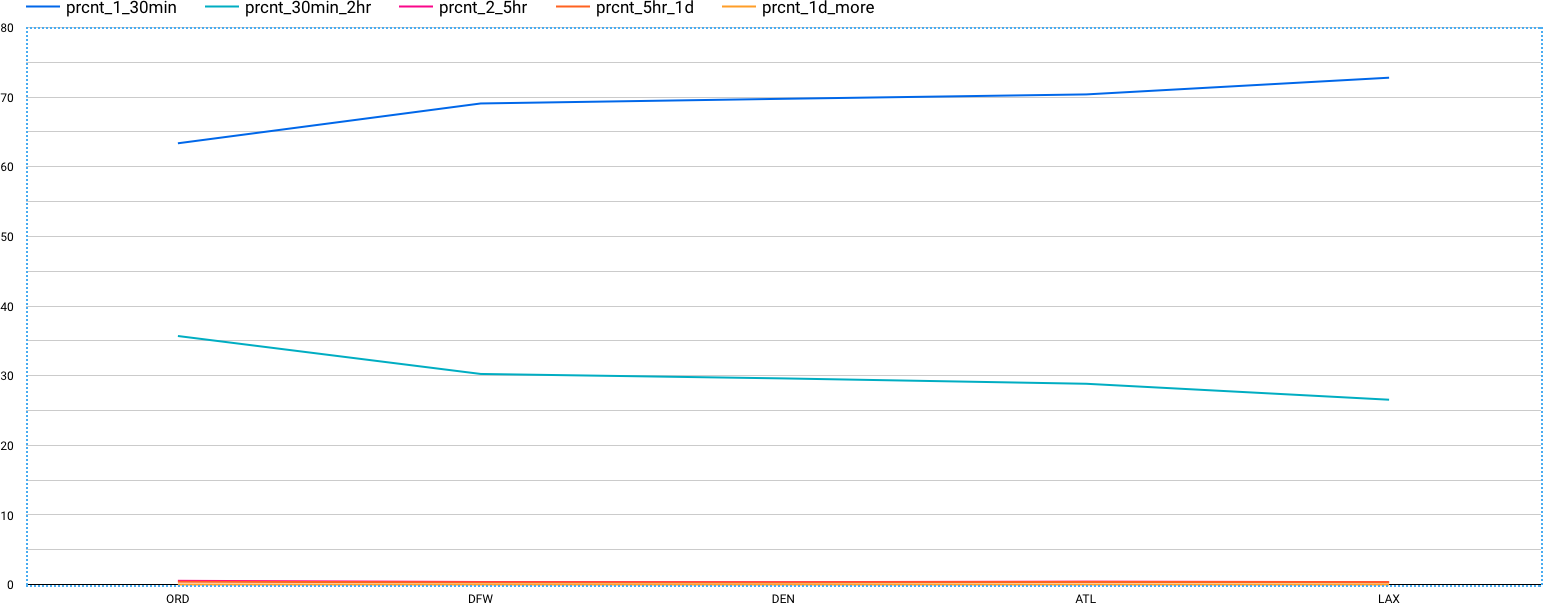
udf\_result as (select origin, delay\_bifurcation(ARRAY\_AGG(slot\_cnt)) as slot\_struct

from airport\_struct group by 1

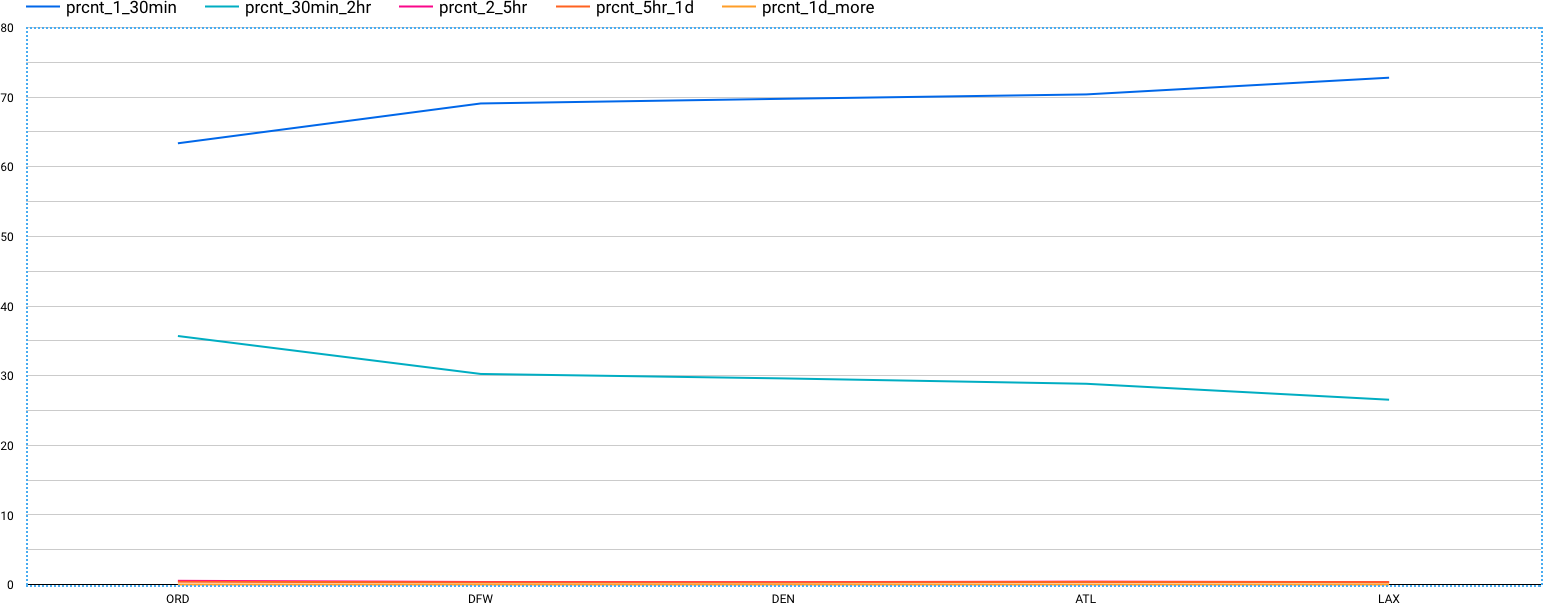
)

select origin, slot\_struct.cnt\_1\_30 as prcnt\_1\_30min, slot\_struct.cnt\_30\_2 as prcnt\_30min\_2hr, slot\_struct.cnt\_2\_5 as prcnt\_2\_5hr, slot\_struct.cnt\_5\_24 as prcnt\_5hr\_1d, slot\_struct.cnt\_24 as prcnt\_1d\_more

from udf\_result

*Delay Percentage for top 5 airports*

Delay Percentage for top 5 airports



Most unreliable month (Cancellations in ascending order)

**Query**

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) )

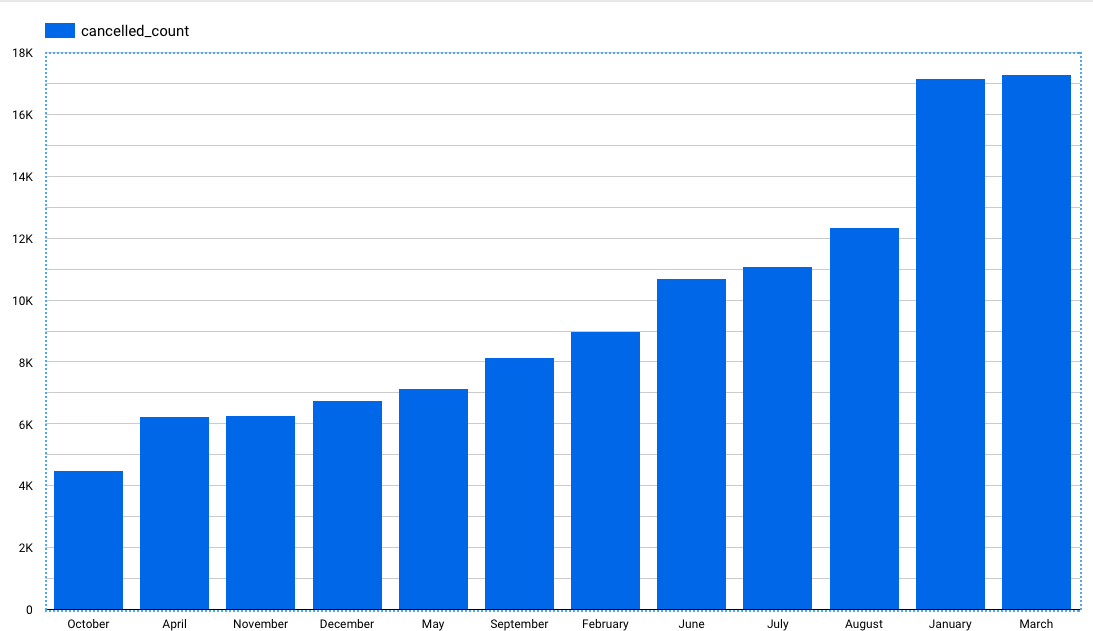
SELECT

month, cancelled\_count

FROM

cancelled\_count\_cte ORDER BY

rank DESC



## CONCLUSION

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

After analyzing the data, a lot of insights have been generated. Most of the delays and cancellations are due to three major reasons:

* + - Weather
    - Airline/Carrier Issues
    - National Air System
  1. **REFERENCES**
     + <https://www.iata.org/en/publications/store/world-air-transport-statistics/>
     + [https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-](https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=data%20analytics%20in%20aviation%20industry&ved=2ahUKEwiW86_T9Zr7AhWkl-YKHSzFC9YQsKwBKAB6BAhDEAE) [wDQ&q=data%20analytics%20in%20aviation%20industry&ved=2ahUKEwiW86\_T9Zr7AhWkl](https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=data%20analytics%20in%20aviation%20industry&ved=2ahUKEwiW86_T9Zr7AhWkl-YKHSzFC9YQsKwBKAB6BAhDEAE)

[-YKHSzFC9YQsKwBKAB6BAhDEAE](https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=data%20analytics%20in%20aviation%20industry&ved=2ahUKEwiW86_T9Zr7AhWkl-YKHSzFC9YQsKwBKAB6BAhDEAE)

* + - [https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-](https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=impact%20of%20covid-19%20on%20aviation%20industry%20research%20paper&ved=2ahUKEwiW86_T9Zr7AhWkl-YKHSzFC9YQsKwBKAJ6BAhDEAM) [wDQ&q=impact%20of%20covid-](https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=impact%20of%20covid-19%20on%20aviation%20industry%20research%20paper&ved=2ahUKEwiW86_T9Zr7AhWkl-YKHSzFC9YQsKwBKAJ6BAhDEAM)

[19%20on%20aviation%20industry%20research%20paper&ved=2ahUKEwiW86\_T9Zr7AhWkl-](https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=impact%20of%20covid-19%20on%20aviation%20industry%20research%20paper&ved=2ahUKEwiW86_T9Zr7AhWkl-YKHSzFC9YQsKwBKAJ6BAhDEAM) [YKHSzFC9YQsKwBKAJ6BAhDEAM](https://www.google.com/search?lei=cl9oY5byKqSvmgesiq-wDQ&q=impact%20of%20covid-19%20on%20aviation%20industry%20research%20paper&ved=2ahUKEwiW86_T9Zr7AhWkl-YKHSzFC9YQsKwBKAJ6BAhDEAM)

* + - <https://dl.acm.org/doi/abs/10.1145/3469028>

<https://www.nap.edu/read/21909/chapter/5>