A NEW HINT TO TRANSPORTATION - ANALYSIS OF THE NYC BIKE SHARE SYSTEM

DOMAIN: DATA ANALYTICS

TEAM MEMBERS:

1.SHRUTHI.S

2.NITHYA.B

3.TANUSH.V

4.SANTHOSH .M

1. INTRODUCTION

1.1 OVERVIEW:

Bike-sharing systems have been deployed in many major cities around the world today. Bike sharing systems provide great advantages as a means of urban public transportation facilitating a green solution for daily commuters and tourists. Users tend to use this type of transportation for their daily needs. The key to success for such systems is the efficient distribution of bikes among the bike stations in order to satisfy high user demands. Existing schemes in the literature focus either on predicting the bike station demand and modeling user mobility mainly focusing on making cycling more accessible to people, or on minimizing the costly and time-consuming movement of bikes among the stations while the system is in use. In this work our objective is to gain insights into the usage of bike sharing systems and in particular the pick-up and drop-off operations. Our goal is to get a better understanding of the bike mobility patterns and identify the key factors that lead to imbalances in the distribution of the bikes at the stations, towards creating effective and sustainable bike sharing systems.

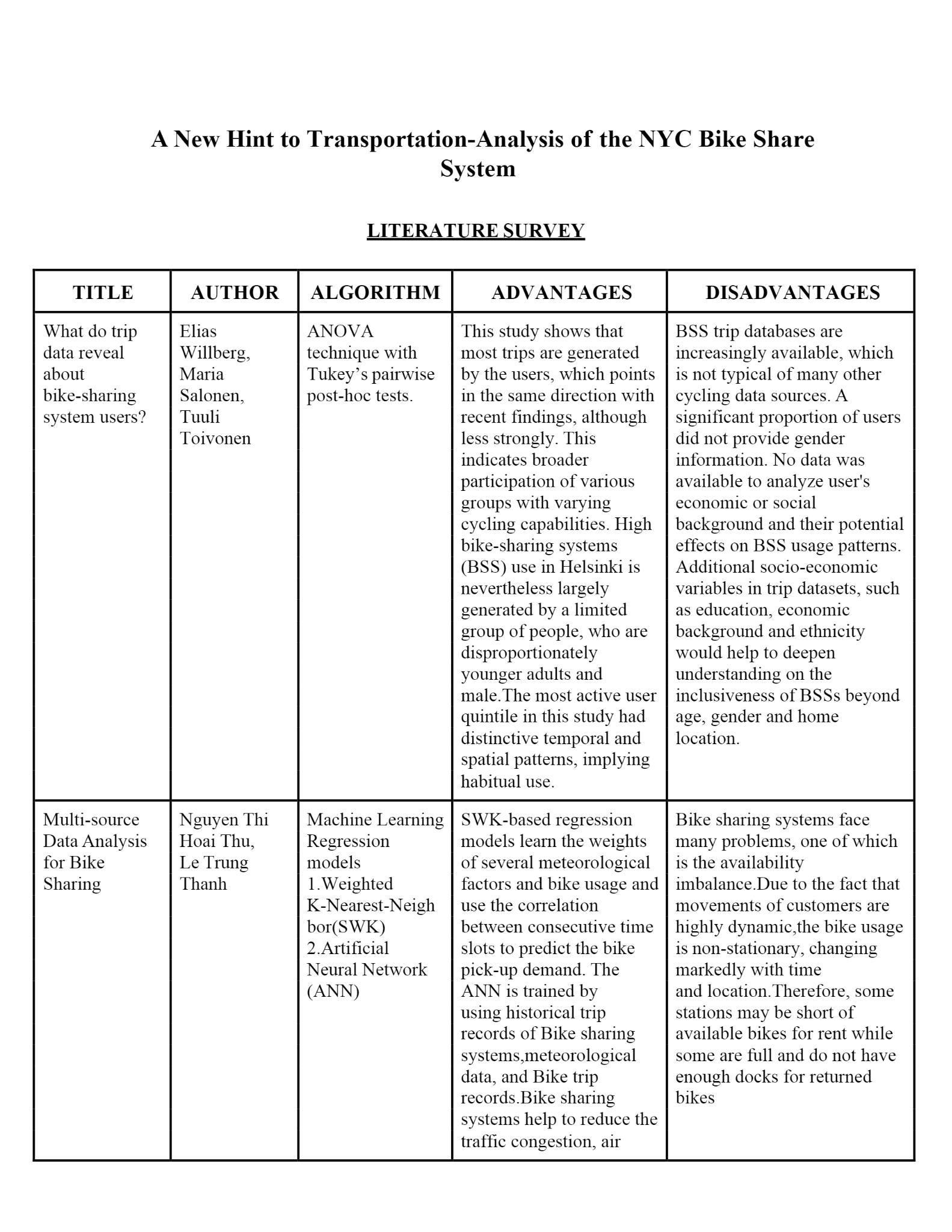
Any Citi Bike client has come up against two frustrating scenarios: the empty dock at the start and full dock at the end of the trip. Researchers call this a "rebalancing" problem as part of "fleet optimization" questions. This problem has attracted the attention of data scientists to develop complex methodologies to optimize the available bikes and open docks.

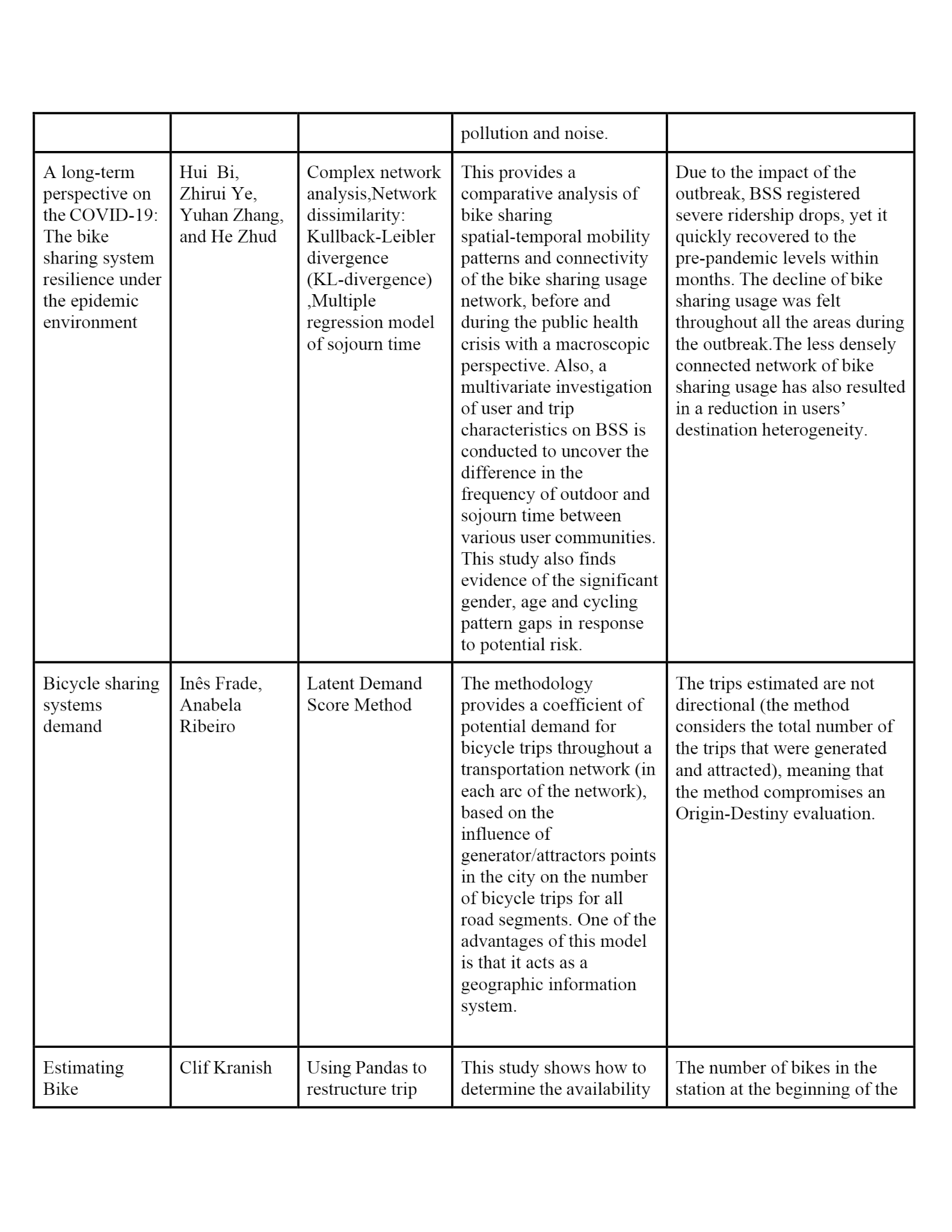
1.2. PURPOSE:

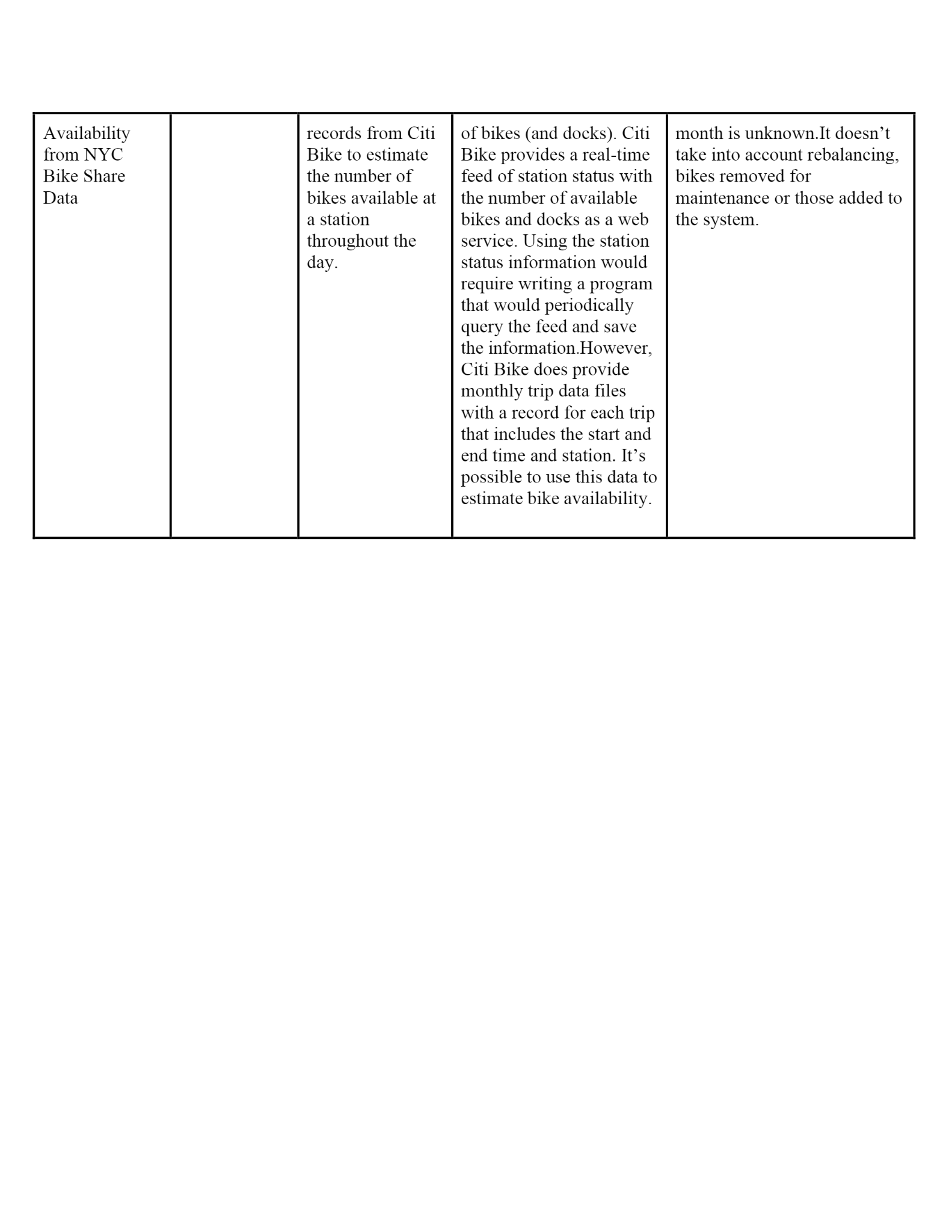
We attempted to utilize the IBM cognos to provide a hint for the 3 questions:

1. Fleet Routing Pattern Detection: what are the most popular routes during peak hours and off-peak? What is the direction of the flow?
2. Station Balance Prediction: what is the average volume of imbalance in the distributed system? What is the station-level inflow and outflow? Is it sensitive to the time? What does it look like in a time series?
3. Reducing rebalancing demand: What are the riders' activities like? Is it possible to rebalance through pricing schemes?

IBM cognos visualization is intended to provide a way to explore different comparative measures at the route, station and system levels with spatial attributes and time series.

2. LITERATURE SURVEY:





2.2 Reference Links:

1. NYC Bike Share https://www.nyc.gov/html/dot/downloads/pdf/bike-share-outreach-r eport.pdf

2. Exploring NYC Bike Share Data https://towardsdatascience.com/exploring-bike-share-data-3e3b2f287 60c

3. Bike Share Opportunities in NYC https://www1.nyc.gov/assets/planning/download/pdf/plans/transportat ion/bike\_share\_complete.pdf

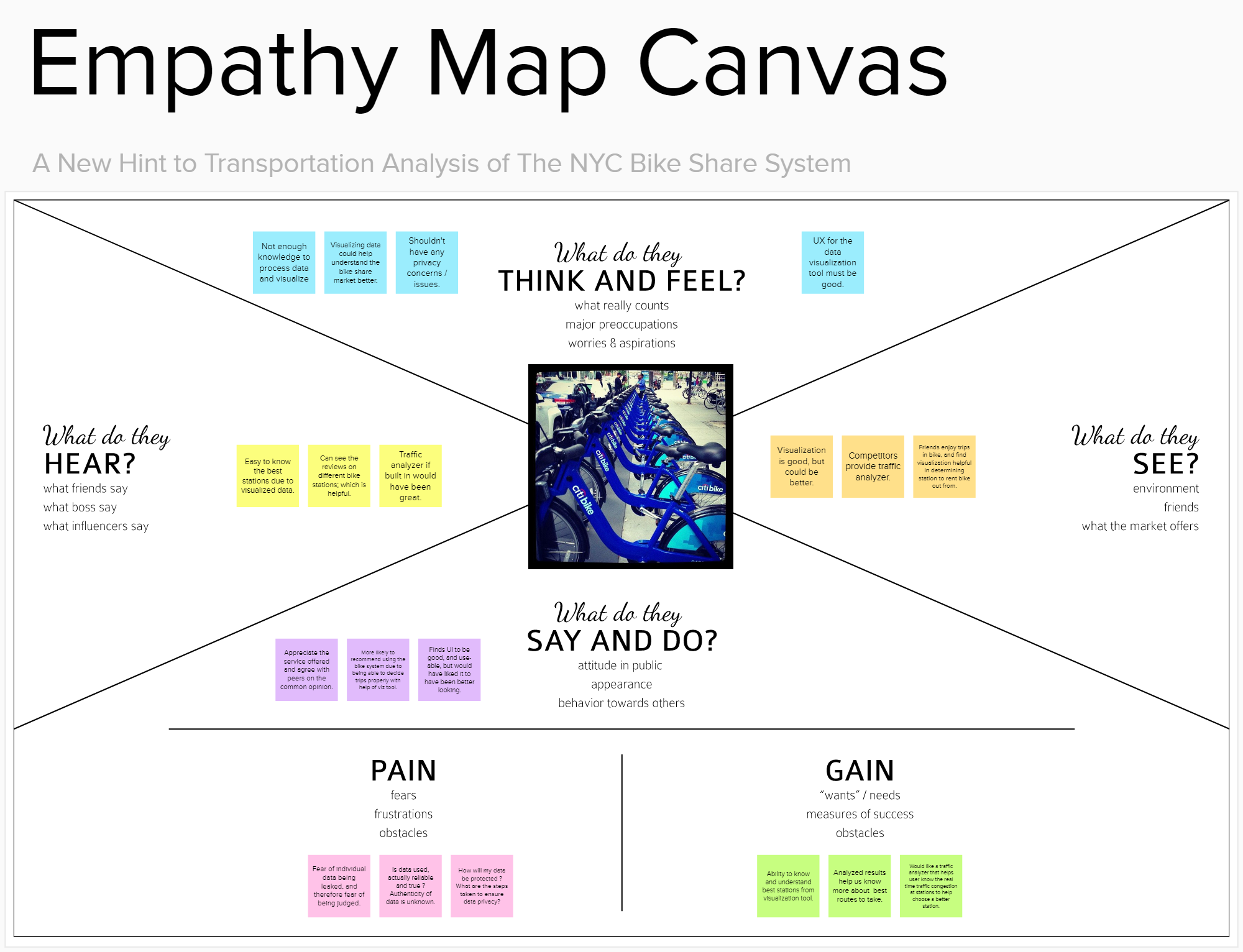
4. Data Visualization on NYC Citi Bike https://nycdatascience.com/blog/r/data-visulization-on-nyc-citi-bike/

5. Impact of Bike Sharing in New York City

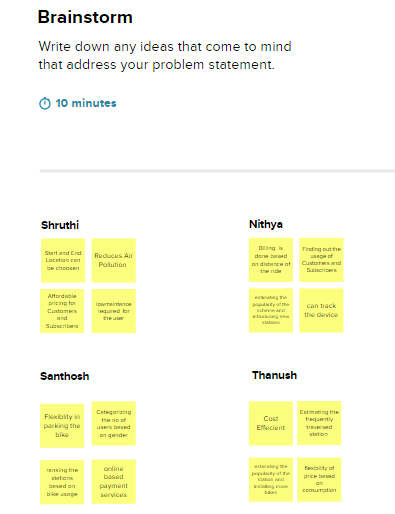
<https://arxiv.org/pdf/1808.06606.pdf>

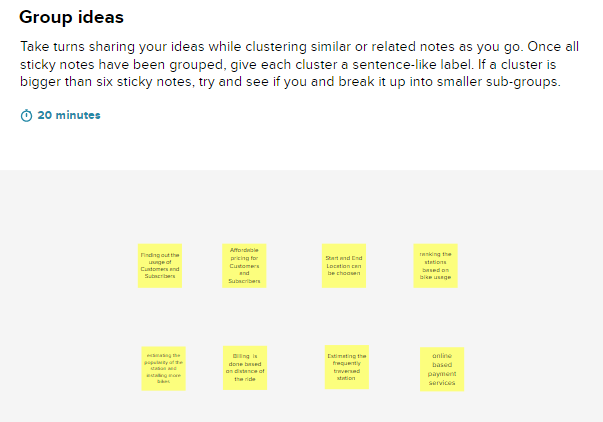
3. IDEATION & PROPOSED SOLUTION:

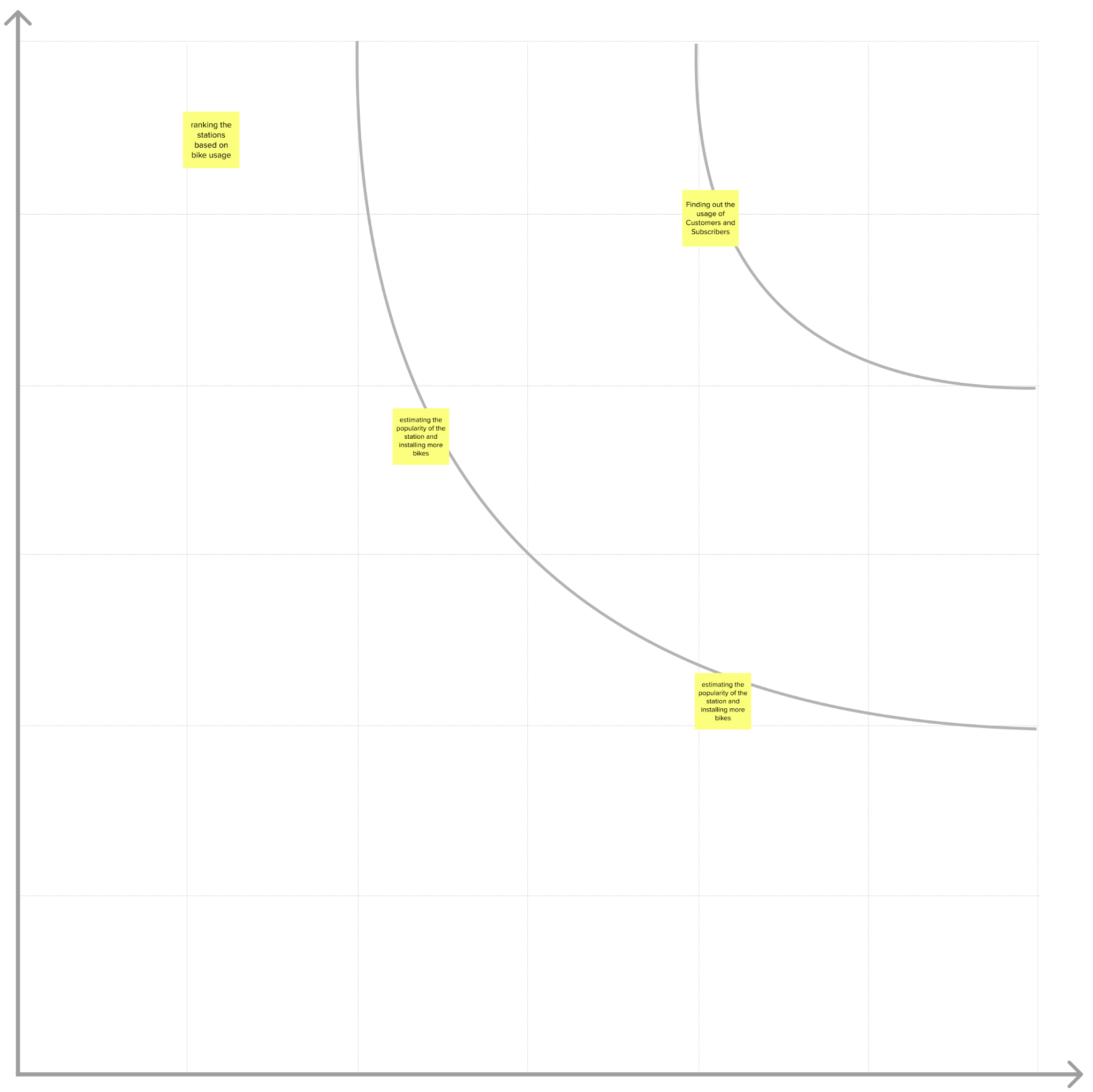
3.1EMPATHY MAP:



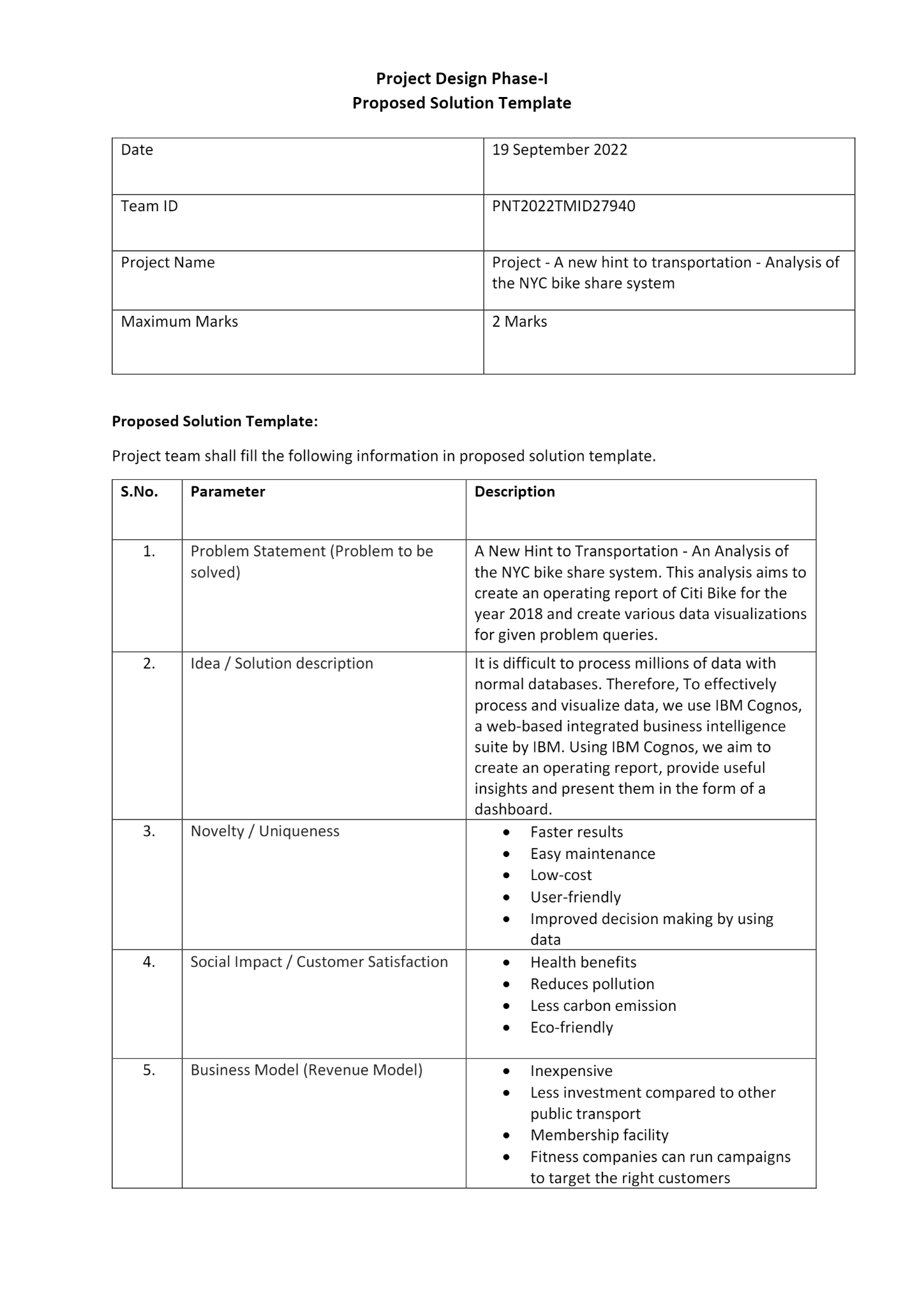
3.2IDEATION & BRAIN STORMING:



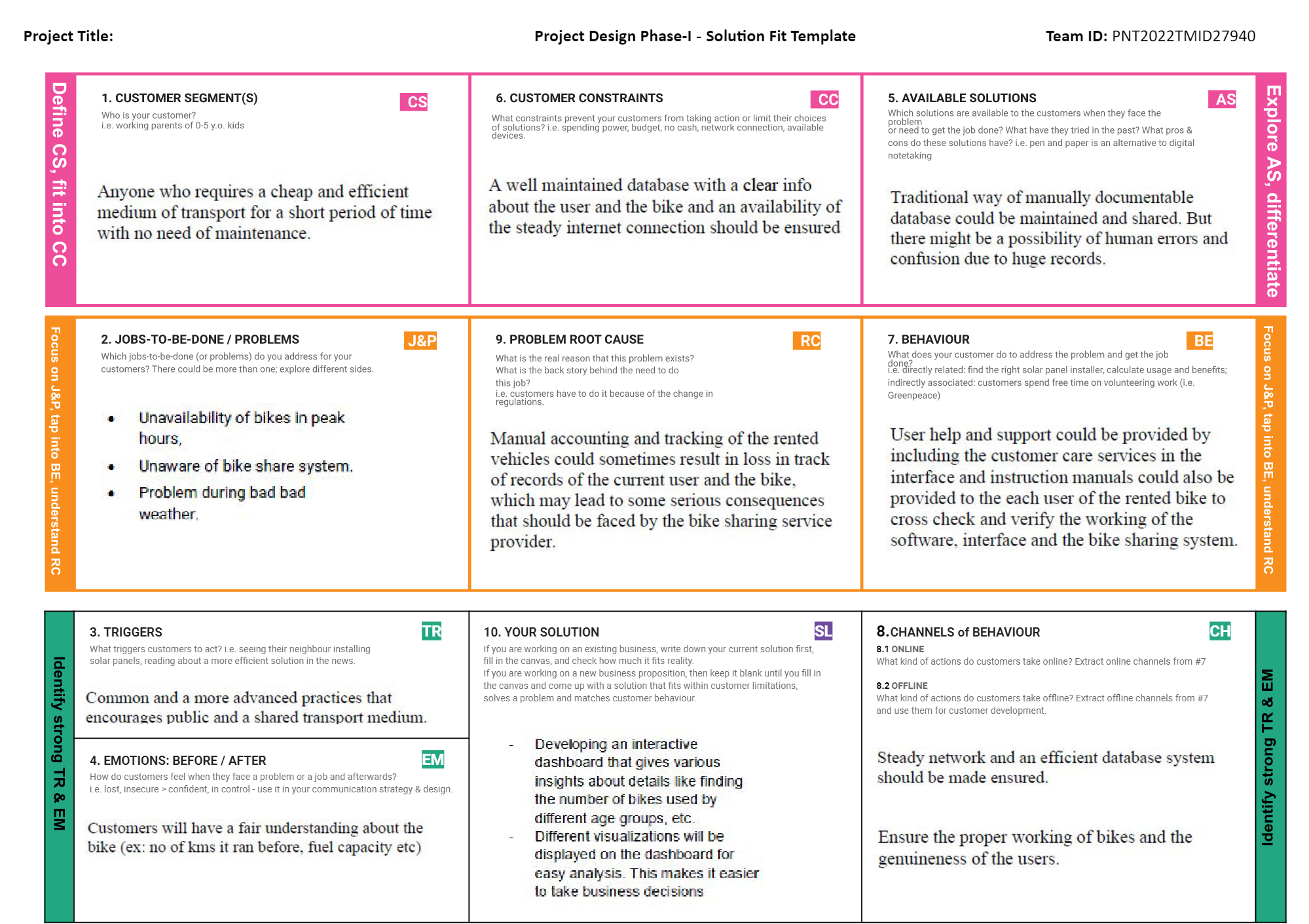




3.3PROPOSED SOLUTION:







**4.REQUIREMENT ANALYSIS:**

**Project Design Phase-II**

**Solution Requirements (Functional & Non-functional)**

| Date | 03 October 2022 |
| --- | --- |
| Team ID | PNT2022TMID27940 |
| Project Name | A New Hint To Transportation - Analysis of the NYC Bike Share System |
| Maximum Marks | 4 Marks |

**Functional Requirements:**

Following are the functional requirements of the proposed solution.

| **FR No.** | **Functional Requirement (Epic)** | **Sub Requirement (Story / Sub-Task)** |
| --- | --- | --- |
| FR-1 | User Registration | Registration through Form  Registration through Gmail  Registration through LinkedIN |
| FR-2 | User Confirmation | Confirmation via Email  Confirmation via OTP |
| FR-3 | Collection of Data | Usage of the NYC Citi Bike helps generate data regarding the different trips taken by different people using Citi Bike. These data were then categorized and provided as datasets, on which further analysis and visualization are to be carried out |
| FR-4 | Analysis of Data | Usage of the NYC Citi Bike helps generate data regarding the different trips taken by different people using Citi Bike. These data were then categorized and provided as datasets, on which further analysis and visualization are to be carried out |
| FR-5 | Display (Visualization) of Data | Different visualizations are carried out depending on the sub-task dealt with. These visualizations are then pooled and displayed on a dashboard - which serves as a tool to provide business insights to customers. Some of the different sub-tasks involved in this requirement include finding the top 10 Start station names with respect to customer age group, displaying the top bikes used with respect to trip duration etc. |

**Non-functional Requirements:**

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

| **FR No.** | **Non-Functional Requirement** | **Description** |
| --- | --- | --- |
| NFR-1 | **Usability** | The dashboard provides an easily understandable and usable operational report which facilitates easy-to-grasp business insights and trends for the customers. Using an interactive dashboard helps to drill down and filter operational information so data can be viewed from different perspectives & in more detail |
| NFR-2 | **Security** | Several crucial business decisions will be made based on the Citi Bike usage data and its analysis, which will be secured appropriately. Data and visualization reports are restricted to a limited number of customers/users |
| NFR-3 | **Reliability** | This analysis provides a reliable and efficient way to grasp on the performance of this bike-sharing system in the year 2018. Usage of the IBM Cognos Platform offers standard reliability to the creation, maintenance, and access of the operational report (dashboard). |
| NFR-4 | **Performance** | A bike-sharing system's performance is determined by its operational efficiency and spatial effectiveness. It is important to evaluate the conditions of bike lanes from the perspective of public bike users in order to improve the operational efficiency of the bike-sharing system. The bike-sharing system dashboard analyzes the  characteristics of bike stations and the accessibility between bike stations and other facilities. It is possible to improve the public bike-sharing program using the evaluation results. |
| NFR-5 | **Availability** | The bicycle-sharing system is a shared transport service where bicycles are available for short-term sharing by individuals at a low cost. There are two types of docking systems offered by CitiBike: docking systems, where users can borrow a bike from a dock and return it at another dock within the system; and dockless systems, which are node-free and rely on smart technology. Both formats can use smartphone web mapping to locate available bikes and docks. |
| NFR-6 | **Scalability** | Bike sharing system can provide alternate transport options to urban residents quickly, which may contribute to a more resilient transport system. The application becomes more scalable as more data is fetched involving areas that are currently inaccessible through this transport methodology, as well as expanding it to other cities apart from New York City provided the required data is available and obtained. Eventually, as more data becomes available, especially in other areas with similar comprehensive bike-sharing systems, this analysis will be able to provide a more detailed picture of the role of bike-sharing in emergency situations. |

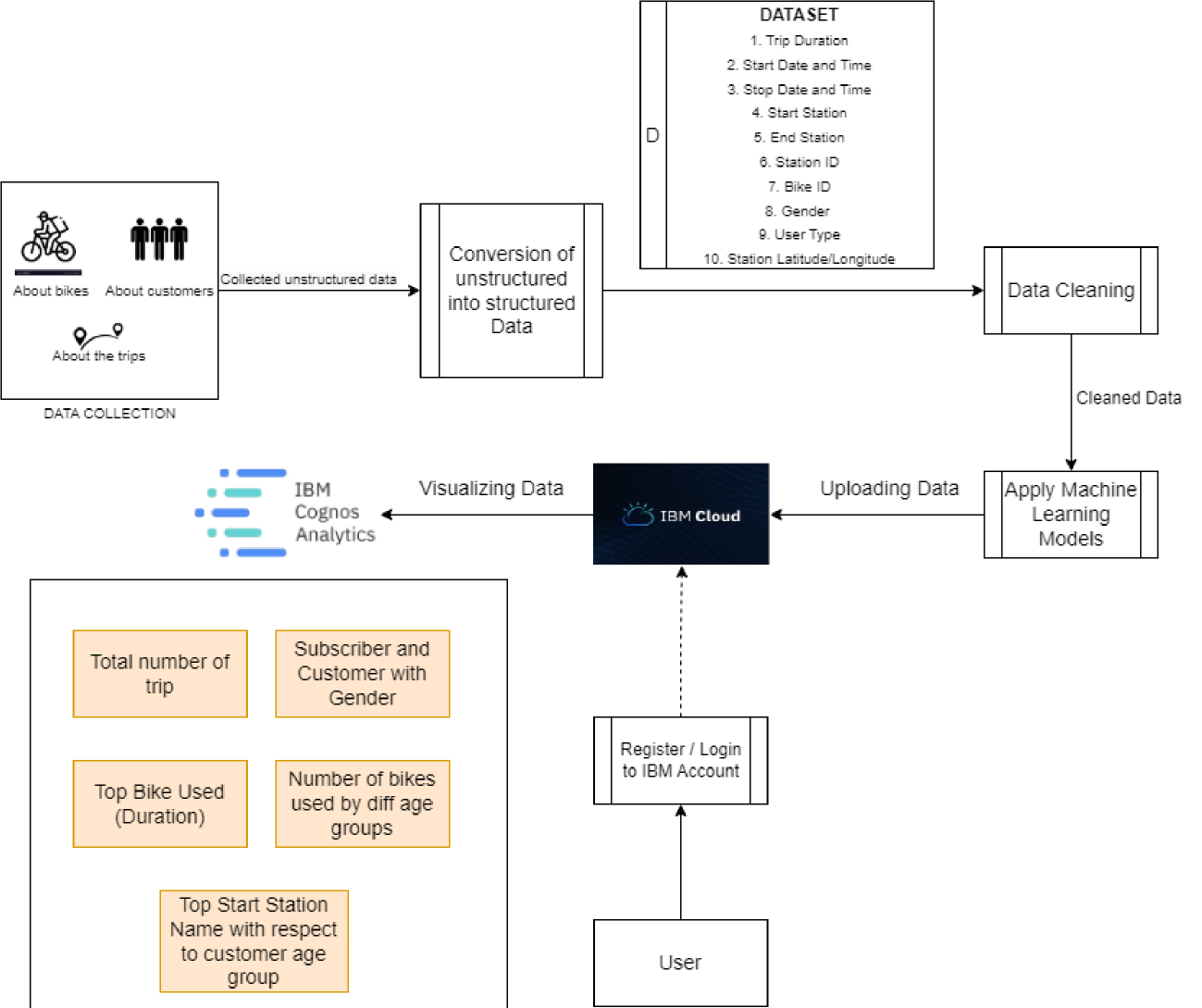
5.PROJECT DESIGN:

5.1 DATA FLOW ARCHITECTURE:

**Project design phase-II**

**data flow diagram and user stories**

| Date | 03 October 2022 |
| --- | --- |
| Team ID | PNT2022TMID27940 |
| Project Name | A New Hint To Transportation - Analysis of the NYC Bike Share System |
| Maximum Marks | 4 Marks |



**User Stories**

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

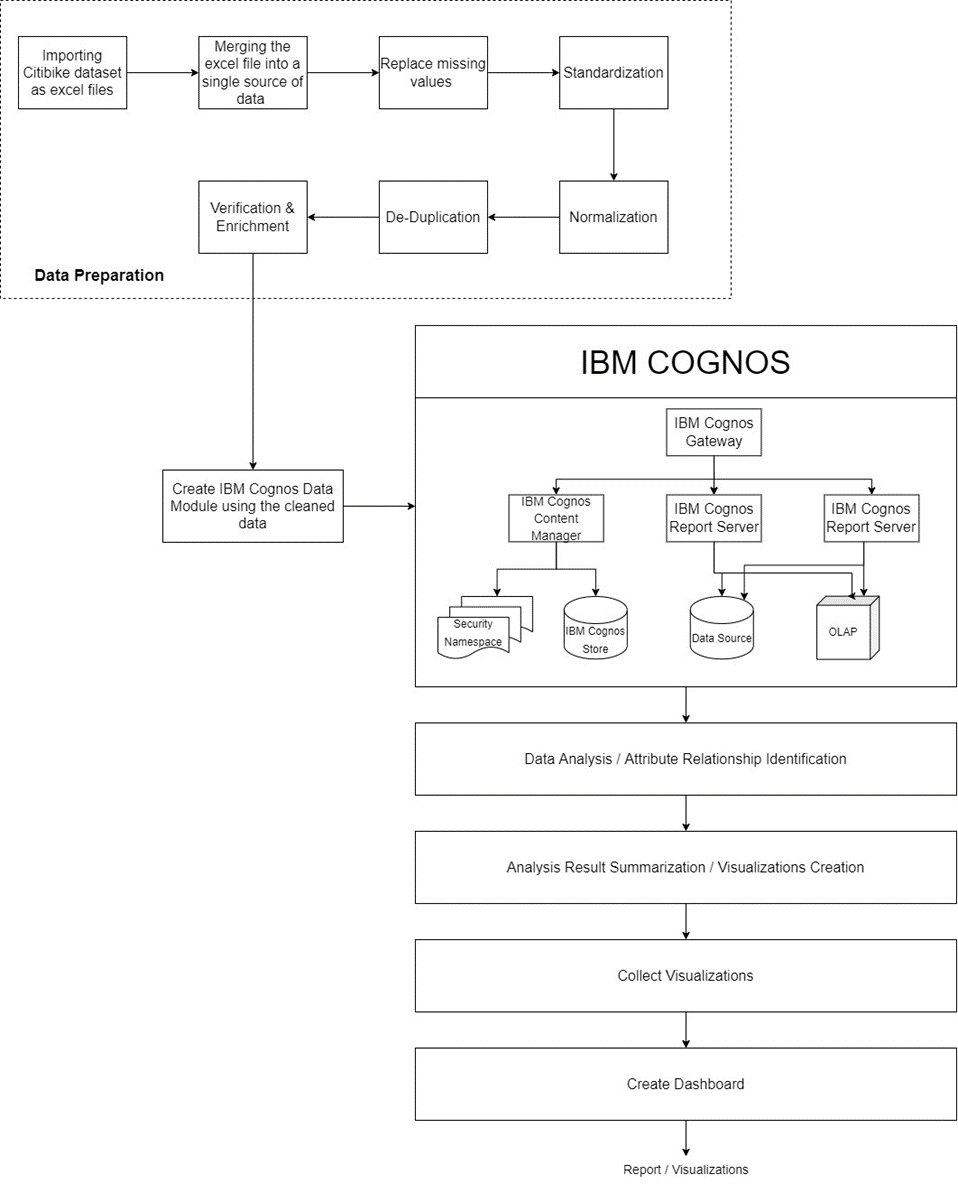
| **User Type** | **Functional**  **Requirement**  **(Epic)** | **User Story**  **Number** | **User Story / Task** | **Acceptance criteria** | **Priority** | **Release** |
| --- | --- | --- | --- | --- | --- | --- |
| Customer(Analysts at Citi, Government) | Registration | USN-1 | As a user, I should be able to register to see the dashboard as a new user | Successful Registration | High | Sprint-1 |
| Customer(Analysts at Citi, Government) | Login | USN-2 | As a user I should be able to login to see the dashboard with the correct credentials | Succesful Login with correct credentials | High | Sprint-1 |
| Customer(Analysts at Citi, Government) | Accessing the dashboard | USN-3 | As a user, I should be able to view the visualizations displayed | Should be able to view the following analysis among others :  1. Total number of  trips  2. Subscriber and Customer with gender  3. Top Bike used  with respect to  duration  4. Number of bikes used by different age groups  5. Top start station name with respect to customer age  group | High | Sprint-1 |
| Customer(Analysts at Citi, Government) | Manipulating the data | USN-4 | As a user I should be able to apply some modifications to the data to see how the resultant visualizations change | I should have the permission to manipulate the data | High | Sprint-2 |

**Project Design Phase-I**

**Solution Architecture**

| Date | 19 September 2022 |
| --- | --- |
| Team ID | PNT2022TMID27940 |
| Project Name | Project - A new hint to transportation - Analysis of the NYC bike share system |
| Maximum Marks | 4 Marks |

**5.2 SOLUTION AND TECHNICAL ARCHITECTURE:**

****

6. PROJECT PLANNING AND SCHEDULING

SPRINT PLANNING AND ESTIMATION:

**Project Planning Phase**

**Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)**

| Date | 18 October 2022 |
| --- | --- |
| Team ID | PNT2022TMID27940 |
| Project Name | A New Hint To Transportation - Analysis of the NYC Bike Share System |
| Maximum Marks | 8 Marks |

**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 | Data Preparation | USN-1 | As an analyst, I filter and extract the Citi-bike data for the year 2018 from the given bucket of datasets. | 4 | Medium | Shruthi, Santhosh |
| Sprint-1 | Data Preparation | USN-2 | As an analyst, I upload the filtered dataset to IBM Cognos. | 1 | Medium | Thanush, Nithya |
| Sprint-2 | Data Preparation | USN-3 | As an analyst, I can prepare the data for analysis by handling missing values and outliers | 7 | Medium | Thanush, Nithya |
| Sprint-2 | Analysis | USN-4 | As an analyst, I perform Exploratory Data Analysis on the filtered dataset to identify patterns and relationships between various features present. | 8 | High | Shruthi, Santhosh |
| Sprint-3 | Visualization | USN-5 | As an analyst, I create various visualizations using IBM Cognos based on the knowledge obtained at the end of the EDA process. | 10 | High | Nithya, Shruthi |
| Sprint-3 | Visualization | USN-6 | As an analyst, I create a dashboard with the created visualizations to supplement business insights during the decision-making process at Citi. | 10 | High | Thanush, Santhosh |
| Sprint-4 | Visualization | USN-7 | As an analyst, I apply predictive analytics and additional features to enhance visualizations | 5 | Medium | Shruthi, Thanush |
| Sprint-4 | Registration | USN-8 | As a user, I can register for the application by entering my email and password, and confirming my password. | 5 | Low | Santhosh, Nithya |

**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 Release Date (Actual)** |
| --- | --- | --- | --- | --- | --- | --- |
| Sprint-1 | 5 | 6 Days | 25 Oct 2022 | 30 Oct 2022 | 5 | 30 Oct 2022 |
| Sprint-2 | 15 | 6 Days | 31 Oct 2022 | 05 Nov 2022 | 20 |  |
| Sprint-3 | 20 | 6 Days | 07 Nov 2022 | 12 Nov 2022 | 40 |  |
| Sprint-4 | 10 | 6 Days | 14 Nov 2022 | 19 Nov 2022 | 50 |  |

| **Sprint** | **Average Velocity** |
| --- | --- |
| Sprint-1 | 0.833 |
| Sprint-2 | 2.500 |
| Sprint-3 | 3.333 |
| Sprint-4 | 1.666 |

**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)

**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.

