A NEW HINT TO TRANSPORTATION-ANALYSIS OF THE NYCBIKE SHARE SYSTEM

A project report submitted in partial fulfillment of
the requirements of the award of the degree of

Bachelor of Engineering
in
Electronics and Communication Engineering
By

SANTHAMURTHY R	(312419106115)
SURIYA KUMAR A	(312419106140)
SIVA SURYA V	(312419106130)
SUBIKSHAN M	(312419106134)

TEAM ID :PNT2022TMID28308

TABLE OF CONTENTS

1. INTRODUCTION

1.1 PROJECT OVERVIEW

Bike share programs have risen in popularity in recent years and have been promoted as a lower carbon alternative to other forms of transit. Interest in bicycle sharing has been growing exponentially over the past decade, resulting in a proliferation of bike share systems in 712 cities across the world, encompassing 806,000 bicycles and 37,500 stations (Shaheen et al., 2014).

This can be largely attributed to the successful incorporation of information technology in docking stations and mobile devices as well as improved logistics such as bicycle rebalancing to ensure responsive supply management. Cities often hope bike sharing will bring many benefits such as extending the reach of transit, substituting motorized trips, and encourage non-cyclists to try cycling (Shahen et al., 2014).

The premise of bicycle sharing is that it is a short-term bike rental system, based on varying timed memberships. Members of the bike share network have access to stations, comprised of a pay-station and multiple bike docks, across the system where bikes can be checked out from one station and returned to another ne nearest to their destination.

The appeal of membership is 24/7 access to an automated bike rental network and utility of bikes in completing “last-kilometer connections” (Shaheen et al., 2014) without the worry of storage or maintenance. The price system is set to encourage shorter trips (less than 30 minutes in time), with additional fees for any time used over that maximum.

There is evidence that bike share users switch to bike share from motorized transport, such as bus and auto (Shaheen et al., 2014), creating the potential for significant reductions in transportation related greenhouse gas or CO2e emissions.

However, there is significant heterogeneity between different cities (Shaheen et al., 2014), showing that there is not a guaranteed CO2e reduction benefit from instituting bike share, especially if the trips would not have been made otherwise or are substituting walking and private bicycle trips.

1.2 PURPOSE

Many bike share systems make available their trip data for those who want to understand how their systems are used.

The bike share system in New York City, Citi Bike, is one of them, but they don't provide much more than the data. I've got some experience in obtaining and preparing their data for visualization, so in this article I will show you how to get started with this rich data source.

In the Before Times I commuted from suburban New Jersey to my job as a Product Manager in New York City at an office, now shuttered, ab';ove Penn Station. To get around in the City at lunch or after work I often relied on Citi Bike, New York's bike share system.

I found I could get to destinations in midtown and even further afield faster than walking and cheaper than the bus or subway. When I discovered that Citi Bike made trip data publicly available I thought that it might provide an interesting use case for the data preparation product that I managed.

The trip data files contain one record for each ride, around two million records per month, depending on the season. It's a traditional bike share system with fixed stations where a user picks up a bike at one dock, using a key fob or a code, and returns it at another.

Using the large table graphic, the moderator was able to show participants images of the kiosk, membership key, smart phone apps, and more.

The features section allowed DOT to solicit participant options on the usefulness of these features, and collect suggestions for additional features. Sites must have unrestricted, 24/7 public access.

• Sites should ensure maximum visibility and access.

• Sites must not impede the use of any existing .

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

Bike sharing is an emerging industry and it is very popular in western countries, while people have tried to start the same in India, we will look into some of the stats regarding how many people use bike sharing systems. According to Wikipedia by August 2014 only 600 cities in the world had bike sharing systems and most of them were in western countries with a fleet of about 500000 bicycles with them. There is a sharp
increase in Next Bike Cog Bike Share are some of the leading Bike Sharing systems that are currently in operation in the world.

While considering Indian perspective in the Bike Share industry, India has not yet adapted the application of this emerging industry. Currently there are a few bike share systems.

2.2 REFERENCE

Burrows; et al. (1999). Gotham. Oxford Press. ISBN 0-19-511634-8.
Koeppel, Gerard (2015). City on a Grid: How New York Became New York. Boston: Da Capo Press. ISBN 978-0-306-82284-1.
"Glorification! The Cities Celebrate the Work That Makes Them One". Brookly n Daily Eagle. May 24, 1883. p. 12. Retrieved June 26, 2019 – via Brooklyn Public Library; newspapers.com open access.
Walker, James Blaine (1918). Fifty Years of Rapid Transit, 1864–1917. Retrieved January 7, 2015.
Caro, Robert (1974). The Power Broker: Robert Moses and the Fall of New York. New York: Knopf. ISBN 978-0-394-48076-3. OCLC 834874.
U.S. Census Bureau, American Community Survey 2006, Table S0802.
Metropolitan Transportation Authority. "The MTA Network". Retrieved July 15, 2016.
Bureau of Transportation Statistics, U.S. Department of Transportation (2001). "Highlights of the 2001 National Household Travel Survey ". Archived from the original on October 2, 2006. Retrieved May 21, 2006.
Bellafante, Ginia (July 12, 2019). "New York Was Supposedly Getting Better for Cyclists. What Happened?". The New York Times. ISSN 0362-4331. Retrieved November 5, 2019.

2.3 PROBLEM STATEMENT DEFINITION

1) The government needs a way to analyze the NYC bike share system so that they can enhance the system and give residents and visitors a fun, safe, affordable and convenient alternative to walking, taxis, buses etc.

2) The goal of this analysis is to create an operating report of Citi Bike for the year 2018.

3) Citi Bike officials are pushing to make the program more robust and to broaden its reach. Financial viability increases with such larger bike-sharing programs. This could be seen by analyzing the total number of trips.

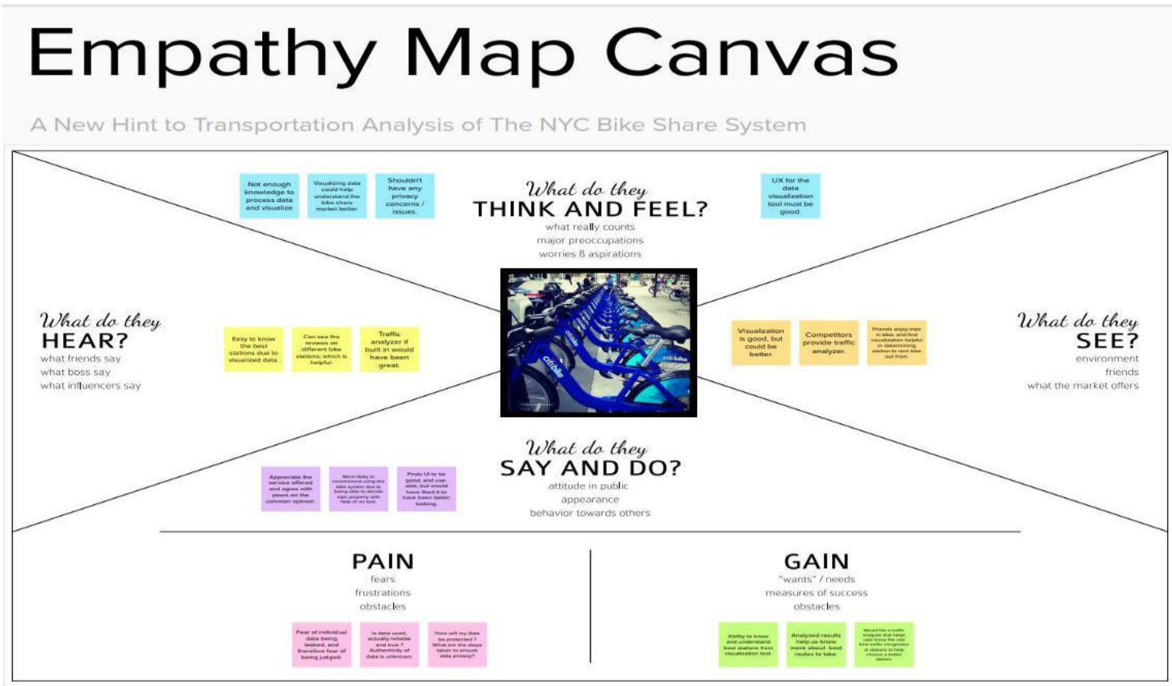
4) The top bikes used with respect to trip duration could be found by this analysis so that more of these bikes can be produced and more users can be attracted.

5) With the help of this analysis, the top 10 Start station names with respect to customer age group could be found so that the government can broaden the bike sharing system by increasing the number of bikes in those stations to make them readily available to all the potential users.

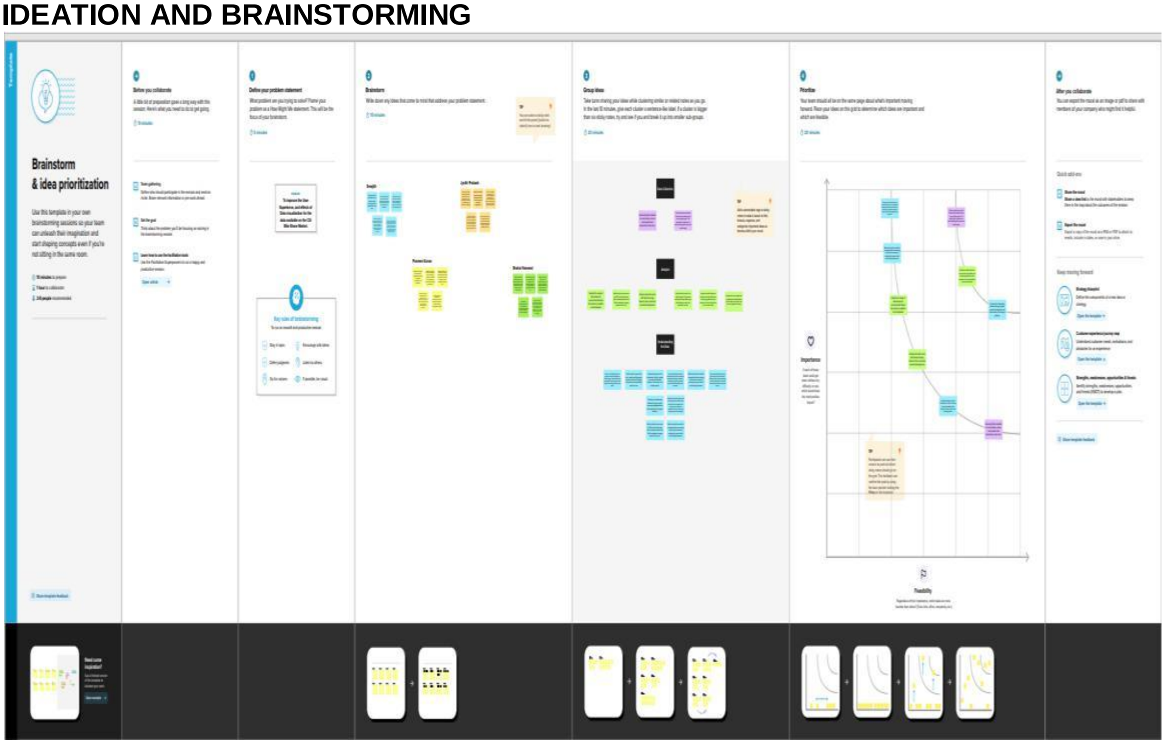
6) The gender of the customer as well as the subscriber could be assessed and the number of bikes used by respective age groups could also be computed.

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAINSTORMING



3.3 PROPOSED SOLUTION

Proposed Solution:

S No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The government needs a way to analyze the NYC bike share system so that they can enhance the system and give residents and visitors a fun, safe, affordable and convenient alternative to walking, taxis, buses etc.
2.	Idea / Solution description	The goal of this analysis is to create an operating report of Citi Bike for the year 2018. We are going to create different types of data visualizations using the various features of IBM Cognos Analytics so that the user can better understand the results of the analysis. It integrates reporting, modeling,analysis, dashboards etc. so that the users can understand the available data, and make effective decisions. It includes predictive, descriptive, and exploratory techniques and provides an intuitive and straightforward interface that is easy to understand. Python's analytical functions can also be used for generating descriptive statistics and visualizations can also be created using Python's visualization libraries.
3.	Novelty / Uniqueness	Our solution gives faster results, reduces maintenance due to complete report coverage, and improved decision making – our reports and dashboards present the data in easily-understood formats.
4.	Social Impact / Customer Satisfaction	Bike share engages riders in physical activity, beneficial to health. In addition, it promotes green mobility and contributes to carbon neutrality. This analysis will help in understanding the association between bike share usage and the environment which is essential for system management and urban transportation planning.
5.	Business Model (Revenue Model)	This analysis might show that bike share is a relatively inexpensive and quick-to-implement urban transportation option compared to other transportation modes. The relative cost of launching a bikeshare system is less than investments in other transportation infrastructure, such as public transit and highways.
6.	Scalability of the Solution	This analysis presents evidence of the possible contribution of bike sharing systems to a more resilient transport system, as it can quickly provide alternative transport options to urban residents. As moredata becomes available, particularly in otherareas with identically comprehensive bike sharing systems, a clearer picture of the role of this transport mode in these emergency situations can be better evaluated by this analysis and provide results with an increased accuracy.

3.4 PROBLE SOLUTION FIT

PROJECT TITLE: A NEW HINT TO TRANSPORTATION-ANALYSIS OF THE NYC BIKE SHARE SYSTEM *PROJECT DESIGN PHASE-I - SOLUTION FIT* **TEAM ID:** *PNT2022TMID23006*

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

Who is your customer?
i.e. working parents of 0-5 y.o. kids

- Sales team of Citi
- Marketing team of Citi
- Firms looking to start a new bike sharing system

6. CUSTOMER CONSTRAINTS CC

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

- Lack of availability of data obtained through detailed data analysis of available information pertaining to the bike sharing system
- Limited access to statistical information

5. AVAILABLE SOLUTIONS AS

Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking

Surveys and studies to understand the active user age groups, frequently visited locations, riding patterns, peak hours etc.

Pros:
- Easy and simple to implement
- Direct interaction with the end users of the bike share system

Cons:
- Limited sample audience - might lead to inadequate understanding.
- Lack of utilization of all available data
- Information collected is hard to extend when needed in the future

Explore AS, differentiate

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

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

Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.

We create an operating report with various forms of visualisations using huge volumes of Citibike user data.
The existing data is filtered to extract the essential information.
For eg Finding the number of bikes used by different age groups

9. PROBLEM ROOT CAUSE RC

What is the real reason that this problem exists? What is the back story behind the need to do this job?
i.e. customers have to do it because of the change in regulations.

Data Analytics can help find patterns and useful insights using data which is necessary for the Ctibike team to analyze their product delivery system and find areas with scope for improvement

7. BEHAVIOUR BE

What does your customer do to address the problem and get the job done?
i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace)

They do not have any insights about gained from user data. Therefore they are unable to promote their product (Citibike) in the best possible way.

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

Identify strong TR & EM

3. TRIGGERS TR

What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.

- Realizing how unhealthy they are becoming and finding out using bikes can be healthy - this makes the users use the bikes more often which gives the Citi teams more sales
- Realizing how much pollution they are causing by making use of vehicles that give out CO2

10. YOUR SOLUTION SL

If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.
If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.

- Developing an interactive dashboard that gives various insights about details like finding the number of bikes used by different age groups, etc.
- Different visualizations will be displayed on the dashboard for easy analysis. This makes it easier to take business decisions

8. CHANNELS of BEHAVIOUR CH

8.1 ONLINE
What kind of actions do customers take online? Extract online channels from 7

8.2 OFFLINE
What kind of actions do customers take offline? Extract offline channels from 7 and use them for customer development.

ONLINE:

The teams at Citi will be able to keep track of the statistics of the usage of Citi bikes online by looking at the dashboards and visualizations.

Identify strong TR & EM

4. EMOTIONS: BEFORE / AFTER EM

How do customers feel when they face a problem or a job and afterwards?
i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

- Users of the bikes will feel extremely satisfied after a good ride which in turn will give the teams at Citi satisfaction
- Customers will feel good about giving back to the community by reducing carbon footprint

OFFLINE:

The teams at Citi will be involved in offline work like installing new bike hubs and trying to work off site to find the problems faced by users of the Citi bike. They also try to keep new bikes in stock in all hubs.

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

Following are the functional requirement of the proposed solution

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Collection of user data	Citi bike's official website provides the data to help with analysis , development, visualization etc. Data is collected from these published files.
FR-2	Analysing the user data	This data is used as input for creating various types of visualizations and analysis is done and a dashboard is created.
FR-3	Display the data	The dashboard is used to display the top bike used with respect to trip duration ,top 10 Start Station Names with respect to customer age group, to find the customer and subscriber with gender, to find total number of trips & calculating the number of bikes used by respective age groups.

4.2 NONFUNCTIONAL REQUIREMENT

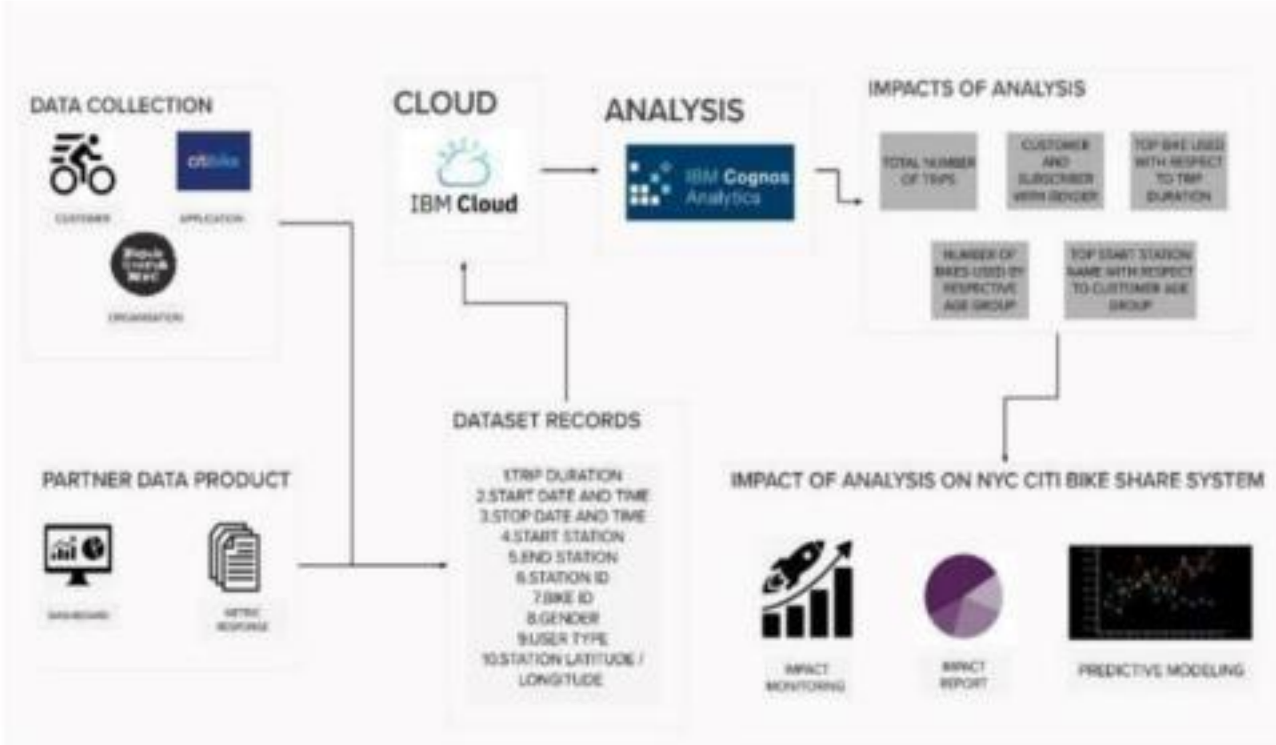
Following are the nonfunctional requirement of the proposed solution

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	This dashboard provides an easily understandable report which facilitates many people and tourists who use bicycles to complete their work and enjoy themselves. It provides many benefits such as measures data like distance, and help with tasks such as route planning, expansion of the bicycle sharing system, manufacturing of desired bikes etc . The benefits of Bicycle sharing systems could be reduced vehicle emissions, reduces energy consumption, improve health benefits, financial savings for individuals, reduced congestion and fuel consumption.
NFR-2	**Security**	The citi bike usage data is secured with appropriate caution as crucial decisions will be made based on this data. We can restrict access to this data and the visualization reports.
NFR-3	**Reliability**	This analysis provides a reliable and an efficient way to grasp on the performance of the citi bike sharing system in the year 2018. It makes use of the available dataset precisely and gives accurate data visualizations that can be used to improve the citi bike sharing system.
NFR-4	**Performance**	Performance of bike sharing system is defined as operational efficiency and spatial effectiveness of bike sharing system. The operational efficiency of bike sharing system aims at understanding the characteristics of public bike users, and evaluating the conditions of bike lanes from the perspective of public bike users .The effectiveness of bike sharing system dashboard aims at analyzing the characteristics of bike stations, and accessibility between bike stations and other facilities. The evaluation results can be used to improve the public bicycle sharing program.
NFR-5	**Availability**	A bicycle-sharing system is a shared transport service where bicycles are available for shared use by individuals for a short-term at low or zero Cost .The programs themselves include both docking and dockless systems, where docking systems allow users to borrow a bike from a dock and return at another node or dock within the system — and dockless systems, which offer a node-free system relying on smart technology. In either format, systems may incorporate smartphone web mapping to locate available bikes and docks.
NFR-6	**Scalability**	This analysis presents evidence of the possible contribution of bike sharing systems to a more resilient transport system,as it can quickly provide alternative transport options to urban residents.As more data becomes available, particularly in other areas with identically comprehensive bike sharing systems, a clearer picture of the role of this transport mode in these emergency situations can be better evaluated by this analysis and provide results with an increased accuracy.

5. PROJECT DESIGN

A data flow diagram is traditional visual representation of the information flow with in a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how enters and leaves the system, what changes the information and when data is stored.

5.1 DATA FLOW DIAGRAM



5.2 USER STORIES

User Stories

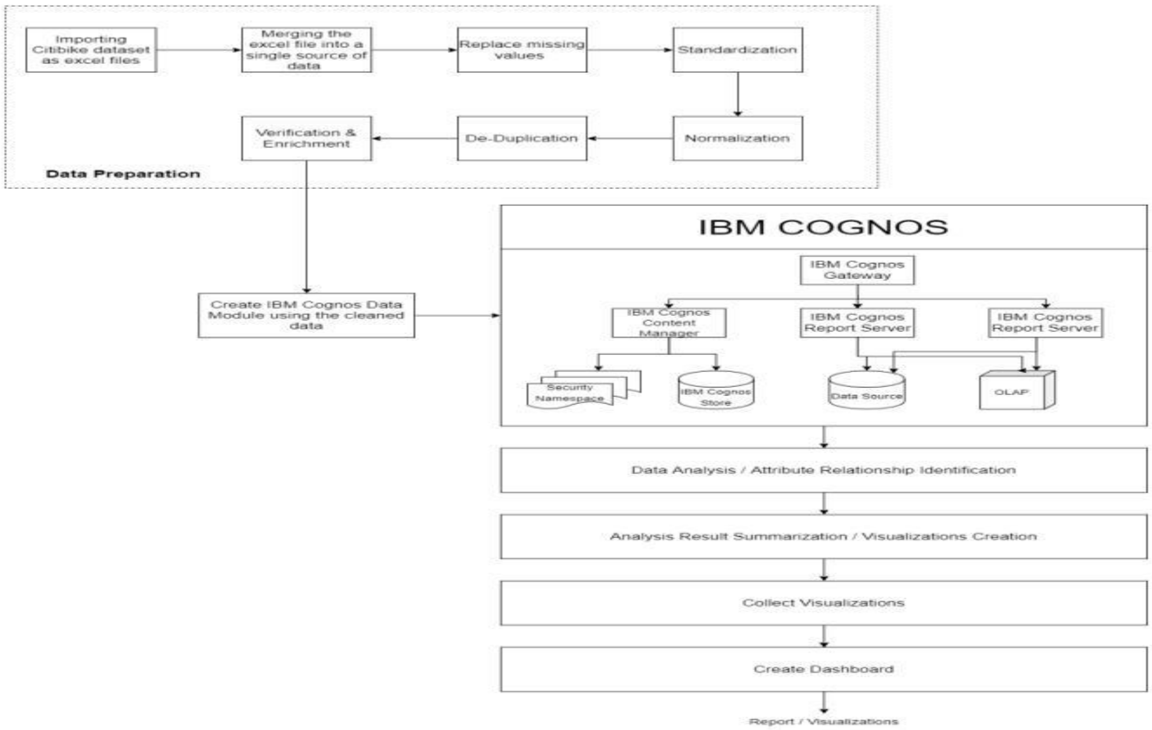
Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer, Analysts, Organizations, Government	Collection of user data	USN-1	Lyft citi bike's official website provides the data to help with analysis, development, visualization etc. Data is collected from these published files.	I can access the data on Lyft citi bike's official website	High	Sprint-1
Customer, Analysts, Organizations, Government	Analysing the user data	USN-2	This data is used as input for creating various types of visualizations and analysis	I can view the analysis of the citi bike	High	Sprint-1

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
			is done and a dashboard is created			
Customer, Analysts, Organizations, Government	Dashboard	USN-3	The dashboard is used to display the top bike used with respect to trip duration, top 10 Start Station Names with respect to customer age group, to find the customer and subscriber with gender, to find total number of trips & calculating the number of bikes used by respective age groups.	I can register & access the dashboard with login	Low	Sprint-2

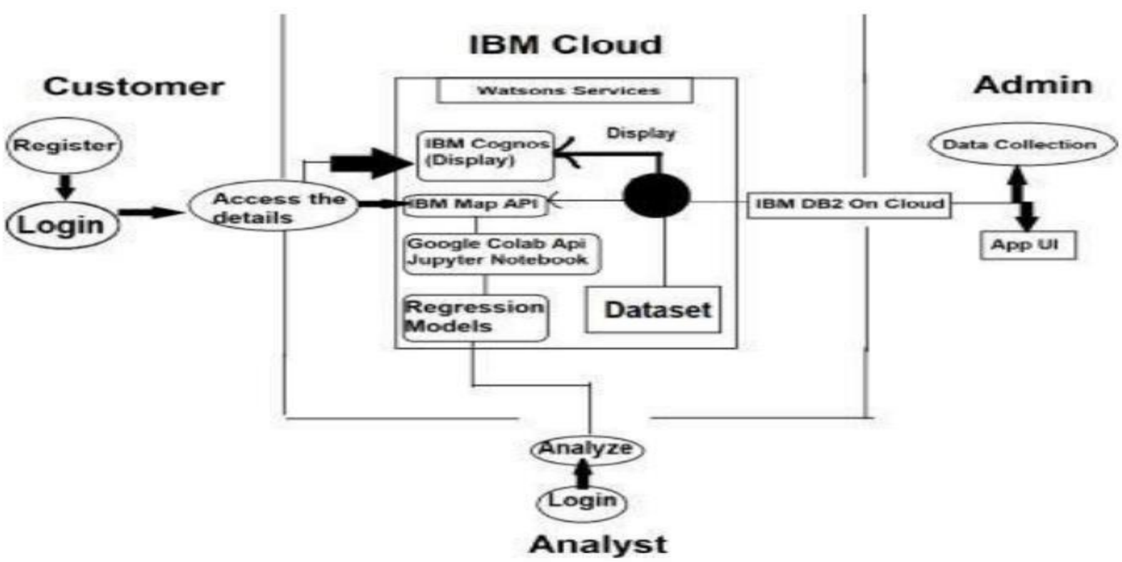
5.3 SOLUTION AND TECHNICAL ARCHITECTURE

SOLUTION ARCHITECTURE



TECHNICAL ARCHITECTURE

The Deliverable shall include the architectural diagram as below

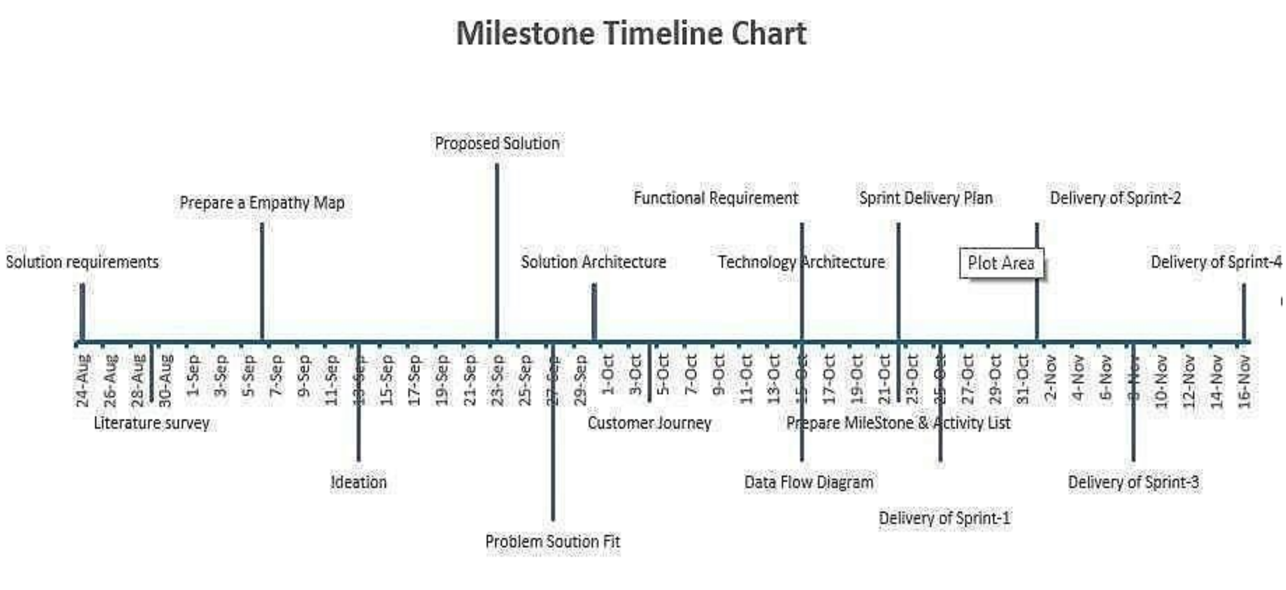


6 . PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Milestone Timeline Chart:

A milestone schedule, or milestone chart, is a timeline that uses milestones to divide a project schedule into major phases. Due to its simplicity, it's used when project managers or sponsors need to share an overview of the project schedule with stakeholders or team members without going over every detail.



6.2 SPRINT DELIVERY SCHEDULE

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	Jyothi Prakash, Sreejith
Sprint-1	Data Preparation	USN-2	As an analyst, I upload the filtered dataset to IBM Cognos.	1	Medium	Praveen,Shahul
Sprint-2	Data Preparation	USN-3	As an analyst, I can prepare the data for analysis byhandling missing values and outliers	7	Medium	Jyothi Prakash,Sreejith
Sprint-2	Analysis	USN-4	As an analyst, I perform Exploratory Data Analysison the filtered dataset to identify patterns and relationships between various features present.	8	High	Praveen,Jyothi
Sprint-3	Visualization	USN-5	As an analyst, I create various visualizations using IBM Cognos based on the knowledge obtained at theend of the EDA process.	10	High	Shahul,Jyothi
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	Sreejith,Praveen
Sprint-4	Visualization	USN-7	As an analyst, I apply predictive analytics and additional features to enhance visualizations	5	Medium	Shahul,Sreejith
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	Jyothi,Sreejith

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

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 FEATURE

People use bike-share for various reasons. Some who would otherwise use their own bicycle have concerns about theft or vandalism, parking or storage, and maintenance.

The Citi Bike System Data page describes the information provided. The specific information for each ride is:

- Trip Duration (seconds)
- Start Time and Date
- Stop Time and Date
- Start Station Name
- End Station Name
- Station ID
- Station Lat/Long
- Bike ID
- User Type (Customer = 24-hour pass or single ride user; Subscriber = Annual Member)
- Gender (Zero=unknown; 1=male; 2=female)
- Year of Birth

7.2 FEATURE

EASY INSTALLATION

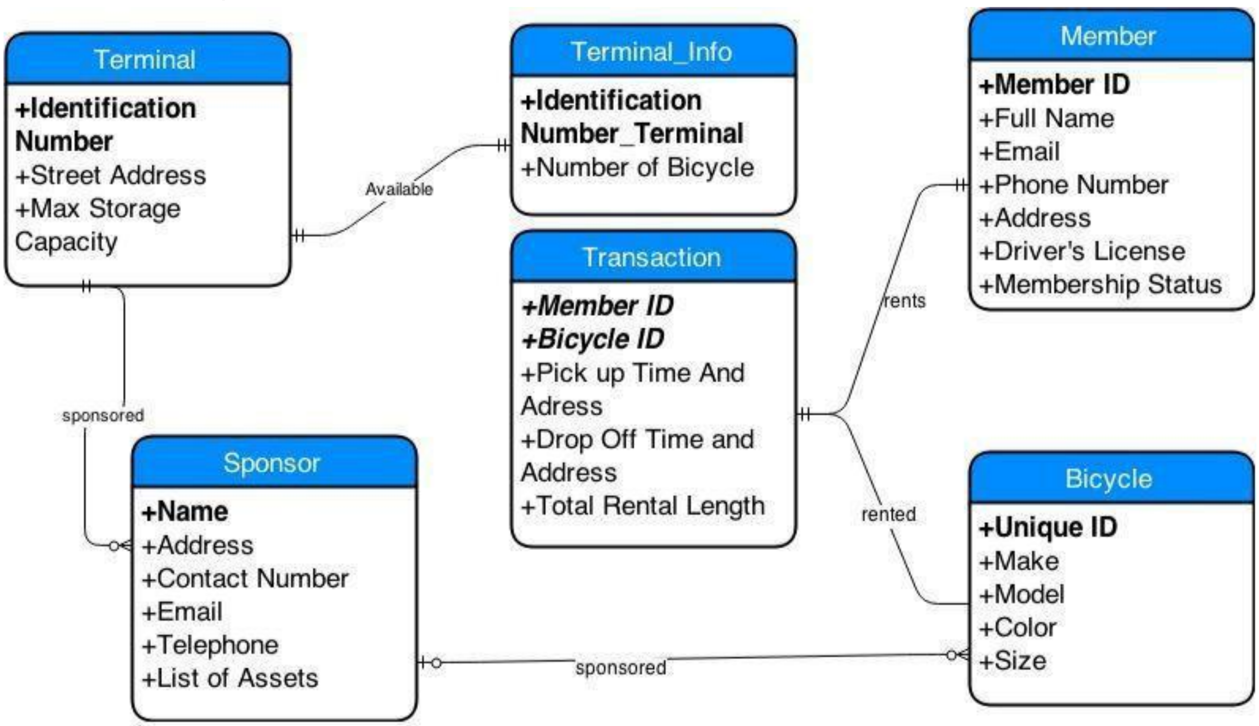
- Stations install in under an hour
- Solar powered and wireless
- No digging or roadwork required

BUSINESS PLAN

- ◊ NYC Bike share pays for all system costs- revenues from users and sponsorship
- ◊ Sponsorship in NYC is highly valued
- ◊ Profit split 50/50 between NYC Bike Share and the City
- ◊ Bike Share in NYC will help spread the word about safe, respectful cycling

7.3 DATABASE SCHEMA

The database schema is the structure of a database described in a formal language supported by the database management system. The term "schema" refers to the organization of data as a blueprint of how the database is constructed.



8.TESTING

Testing is the process of evaluating and verifying that a software product or application does what it is supposed to do. The benefits of testing include preventing bugs, reducing development costs and improving performance.

8.1 TEST CASES

Test case includes information such as test steps, expected results and data while a test scenario only includes the functionality to be tested.

- ◊ UI Test Cases for Bike
- ◊ Positive Test Cases for Bike
- ◊ Negative Test Cases for Bike

UI Test Cases for Bike

- ◊ Verify that design and dimension of the application are as per the specifications.
- ◊ Verify that the different colors used in the bike are of the correct shades as per the specifications.
- ◊ Verify that the weight of the bike is as per the specifications.
- ◊ Check the material used in different parts of the bike – outer body, tires, seat, etc.

Positive Test Cases for Bike

- ◊ Check if the bike is of type electric start, manual start or both.
- ◊ Verify that the bike starts smoothly using the available options.
- ◊ Check the amount of force to kick-start the bike.
- ◊ Verify that bike runs smoothly and attain desired speed when accelerated.
- ◊ Verify that the maximum speed attained by bike is as per the specification.

Negative Test Cases for Bike

- ◊ Check if the bike starts when fuel other than prescribed fuel is filled in the bike.
- ◊ Check the condition of the bike when tires are filled with pressure less or more than specified.

- ◊ Check the condition of the bike when both the tires have different air pressure.
- ◊ Check the bike's condition when it is ridden at high speed on first gear only.

8.2 USER ACCEPTANCE TESTING

User Acceptance Testing (UAT), which is performed on most UIT projects, sometimes called beta testing or end-user testing, is a phase of software development in which the software is tested in the "real world" by the intended audience or business representative.

- ◊ Before product goes live.
- ◊ Done by end users.
- ◊ Fix usability issue.
- ◊ Ensures viable product.

UAT test cases will look like this:

- ◊ Are testers filling out the correct information in bike?
- ◊ Do they understand what's happening when being redirected to the in bike system?

9.RESULTS

9.1 PERFORMANCE METRICS

The main metrics are used to judge the performance of bikeshare systems: average number of daily uses per bike and average daily trips per resident (of the coverage area). These two metrics tend to have an inverse relationship.

A system with a low number of bikes could have high per-bike usage because demand is high, but fail to meet that demand and therefore have a lower number of trips per resident. On the other hand, a system could have a high number of trips per resident but also a very high number of bikes, and therefore a low number of trips per bike.

Both of these extremes are inefficient; a sustainable system should find a balance of having just enough bikes to satisfy demand with around 4 daily trips per bike share system.

Average daily trips per bike

Target: 4-8 daily uses per bike

Turnover is critical to a successful bikeshare system, and this metric gets at how efficiently the bikes are being used. Fewer than four daily uses per bike can result in financial unsustainability for the operator (i.e., user fees not able to cover cost to operate each bike), while more than eight daily uses can indicate limited bike availability, especially during peak hours. New York City (6.4), Barcelona (6.4), Mexico City (5.4), and Guangzhou (5.0) showed solid daily usage numbers in 2017.

Average daily trips per 1,000 residents (in service area)

Target: city-generated, improvement over time

This is a metric of market penetration, that is, how many people in the service area are using the system. A high number of uses spread across residents in the service area is key the increasing bicycle mode share, decreasing vehicle and transit network congestion, and promoting safe, clean, healthy modes of transport.

Trips per 1,000 residents should be monitored as the system matures, with the goal of increasing market penetration over time (a more prescriptive target for annual improvement in market penetration could be created from baseline trip numbers). An increase in trips per 1,000 residents indicates more trips being taken by bike, and can help to evaluate progress toward citywide mode shift goals.

10. ADVANTAGES AND DISADVANTAGES

ADVANTAGES

1. Convenient Mode of Transportation

The most common benefit of this program is its accessibility. If you are fond of cycling, you will find this method helpful as there are numerous systems just about everywhere for you to use. Bigger cities are supporting the use of bike-sharing. This is why it is common to find bike-sharing systems in downtown areas. Driving a car through congested avenues can be frustrating and irritating. This makes bike sharing very convenient.

2. Healthy Method for Traveling

Wellness, fitness, and health are an essential part of your life. You should care about what you consume and what your daily activity is. Bike-sharing systems can help and encourage you to live a better and healthier life. They help you stay in shape even when you are away from home vacationing.

3. Environmental Benefits

Maintaining a clean environment is as important as maintaining your health. Living in a heavily polluted environment can cause various health issues. Bikes do not release greenhouse gases, unlike buses and cars. So, if you are renting a bike from a bike-sharing system, you reduce the carbon footprint and take measures to keep your environment safe.

DISADVANTAGES

1. Congestion in the Users of Bike Sharing

As bike-sharing systems can help you travel throughout the city, they do not exist in infinite numbers. Waiting can be annoying if the bike is not available. You may face this problem during peak hours.

2. A Helmet is a Requirement

When you are planning to ride a bike, you need to keep safety measures in your mind. Therefore, wearing a helmet is essential. Some bike sharing systems may require you to use a helmet but might not provide them for you. Sometimes you will need to bring your own which can be a hassle.

3. Bikes Are Not Clean

Shared bikes can be unhygienic as many people have probably used it before you. The seat and handlebars are a particularly high-traffic area for germs, so cleanliness is always a concern.

11. CONCLUSION

Bicycle sharing systems can be the new boom in India, with use of various prediction models the ease of operations will be increased. The four algorithms are applied on the bikeshare dataset for predicting the count of bicycles that will be rented per hour. We got some good results and accuracy with random forest and by using Tune RF function with the original random forest algorithm. The accuracy and performance has been compared between the models using Root Mean Squared Logarithmic Error (RMSLE).
If these systems include the use of analytics the probability of building a successful system will increase.

12. FUTURE SCOPE

One aspect of the data that I did not explore in great detail is the intra-day variation in usage of the system. This is also a key aspect that bike share system operators are interested in because knowing the variation in demand on an hourly basis is another very useful metric for identifying the times of the day when the need for artificial rebalancing is maximum.

Additionally, this work will feed into a larger study calculating the life cycle environmental impacts of a bikeshare system and its ability to substitute other modes of transit with the aim of reducing the overall Greenhouse gas (GHG) emissions due to transportation.

13. APPENDIX

Research Methodology and Framework Methodology and Data Sources

Using spatial data primarily from the United States Census54 and New York City's Department of Information Technology & Telecommunications,55 ridership and station activity data from New York City Department of Transportation and NYC Bike Share, LLC,56 and station location data from NYC Bike Share, LLC, this study used ArcGIS software57 to analyze and show connections between Citi Bike and public transit. Data from Divvy Bikes,58 Chicago Open Data Portal,59 Capital Bike Share,60 the District Department of Transportation,61 Hubway,62 the Massachusetts Bay Transportation Authority,63 Nice Ride,64 and MetroGIS65 allowed for comparisons in station coverage area and station density in New York City, Chicago, Washington, DC, and Minneapolis/St. Paul. Researchers conducted interviews with New York City Department Transportation and NYC Bikeshare LLC staff.

Framework

We examine connections between New York City's bike share program, Citi Bike, and the previously existing transportation options in New York City. After observing the system's success in its first year of operation, this study analyzes connections between bike share stations and from stations to transit options. New York City's bike share system offers a solution to the "last mile"66 problem, the problem of getting riders short distances, under a mile, to and from transit stations. A key component of this "last mile" analysis came through calculating the number of Citi Bike stations with 100, 200, 500, and 1320 feet67 of subway station entrances and comparing the proximity and density of bike share stations in New York City, Washington, DC, Chicago, Boston, and Minneapolis/St. Paul.

Source Code

```
#cleaning_and_understanding_the_data.ipynb

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
from datetime import datetime
from pprint import pprint

from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
```

```
from google.colab import auth
from oauth2client.client import GoogleCredentials

path = "/content/dataset.csv"
df = pd.read_csv(path)
print(df)

df.head()

df.describe()

df.info()

df.isnull().sum()

trips_df = pd.DataFrame()
trips_df = temp.groupby(['start station name','end station name']).size().reset_index(name
= 'Number of Trips')
trips_df = trips_df.sort_values('Number of Trips',ascending = False)
trips_df["start station name"] = trips_df["start station name"].astype(str)
trips_df["end station name"] = trips_df["end station name"].astype(str)
trips_df["Routes"] = trips_df["start station name"] + " to " + trips_df["end station
name"] trips_df = trips_df[:50]
trips_df = trips_df.reset_index()
trips_df

px.pie(values = temp['gender'].value_counts(),
     names =temp['gender'].value_counts().index,
     title ="Gender Variation")

px.bar(x=temp["start station name"].value_counts().index,
     y=temp["start station name"].value_counts().values,
     labels={'x':'Start Station Name',"y":"Count"})

px.bar(x=temp["end station name"].value_counts().index,
     y=temp["end station name"].value_counts().values,
     labels={'x':'End Station Name',"y":"Count"})

px.bar(x=temp["Hour"].value_counts().index,
     y=temp["Hour"].value_counts().values, title = "Hour
     usage of Citi Bikes", labels={'x':'Time',"y":"Number
     of people using bike"})
```

```
# index.js

const mysql = require('mysql');
const express = require('express');
const session = require('express-session');
const path = require('path');
const { dirname } = require('path');
const { request } = require('express');

// Requiring dotenv and Creating variables to store the Env values to maintain a bit of
secrecy require("dotenv").config()
const  DB_HOST  =  process.env.DB_HOST
const DB_USER = process.env.DB_USER
const DB_PASSWORD = process.env.DB_PASSWORD
const DB_DATABASE = process.env.DB_DATABASE

const connection = mysql.createConnection({
   host: DB_HOST,
   user: DB_USER,
   password: DB_PASSWORD,
   database: DB_DATABASE
});

// Creating an express application
const app = express();

// storing the value of username for further displaying the respective user's
info let uname = '';

// setting the view engine to ejs -> as I wanted to render the user info dynamically, and have
it viewed through an engine [This was the only possible way to render / pass values to
HTML] app.set('view engine', 'ejs');

app.use(session({
    secret: 'secret',
    resave: true,
    saveUninitialized: true
}));
app.use(express.json());
app.use(express.urlencoded({ extended: true }));
app.use(express.static(path.join(__dirname, 'static')));
```

```
// http://localhost:3000/
app.get('/', function(request, response) {
    // Render login template
    response.sendFile(path.join(__dirname + '/login.html'))
})

// http://localhost:3000/auth
app.post('/auth', function(request, response) {
    // Capture the input fields
    let username = request.body.username;
    uname = username;
    let password = request.body.password;
    // Ensure the input fields exist and are not
    empty if (username && password) {
        // Execute SQL query that'll select the account from the database based on the
specified username and password
        connection.query('SELECT * FROM accounts WHERE username = ? AND password
= ?',[username, password], function(error,results,fields){
            // If there is an issue with the query, output the error
            if (error) throw error;
            // If the account exists
            if (results.length > 0) {
                // Authenticate the user
                request.session.loggedin = true;
                request.session.username = username;
                  // Redirect to home page
                 response.redirect('/home');
            } else {
                response.send('Incorrect Username and/or Password!');
                response.end();
            }
        });
    } else {
        response.send('Please enter Username and
        Password!'); response.end();
    }
});

// http://localhost:3000/home
app.get('/home', function(request, response) {
    // If the user is loggedin
    if (request.session.loggedin) {
```

```
        // Creating a home page for the users to be welcomed.
        response.sendFile(path.join(__dirname + '/home.html'));
        // // Output username
        // response.send('Welcome back, ' + request.session.username +
    '!'); } else {
        // Not logged in
        response.send('Please login to view this page!');
    }
    //response.end();
    // Uncommenting this usually led me to have errors regarding 'Can't set headers after they
are sent' -> so it meant that responses were closed, and communication wasn't possible after
that ? });

// http://localhost:3000/dashboard
app.post('/dashboard', function(request, response) {
    // Return the dashboard for the user if he is logged
    in. if (request.session.loggedin) {
        // // Return the dashboard to be viewed for the user.
        // response.sendFile(path.join(__dirname + '/dashboard.html'));
        // Instead of sending the path and hence thereby opening the html file, we can instead
render an ejs file, using the ejs view engine.
        response.render('dashboard');
    } else {
        // Not logged in
        response.send('Please login to view this page!');
    }
    //response.end();
});

// http://localhost:3000/info
app.post('/info', function(request, response) {
    // Return the dashboard for the user if he is logged
    in. if (request.session.loggedin) {
        // // Return the dashboard to be viewed for the user.
        // response.sendFile(path.join(__dirname + '/info.html'));
        // Instead of sending the file and it's path, to be displayed when accessing this route, we
can actually pass the sql values as parameters to the view engine, to print them out, dynamically
corresponding to the data available from the sql database.
        // response.render('info', { rinzler: 'Hello Tron!'});
        // console.log(uname);
        connection.query('SELECT id,username,email FROM accounts WHERE username
= ?',[uname], function(error,results,fields){
            // Output error if there is some kind of problem with the
```

query 25

```
        if (error) throw error;
        // If the account exists then perform the operations associated with
        it! else {
            // console.log(results);
            // // This basically stores all the resulting row's fields in it.
            // // In the following steps, we have to seperate the different fields (4 of them) and
just show only the three fields to the user.
            // // Also try to change the query so that only the three non-private fields are sent back as
a result
            // console.log("Id is : ",results[0].id);
            // console.log("Username is : ",results[0].username);
            // console.log("Email is : ",results[0].email);
            response.render('info', { id: results[0].id, name: results[0].username, mail:
results[0].email})
        }
      })
   } else {
      // Not logged in
      response.send('Please login to view this page!');
   }
   //response.end();
});

// Add a port to enable the nodejs server to listen to incoming connections
// Ideally when we want to deploy our login system to a production server, we want to listen on
port `80` so that we don't have to specify the port number in the URL.
app.listen(3000);
```

----

home.html

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <meta http-equiv="X-UA-Compatible" content="IE=edge">
   <meta name="viewport" content="width=device-width, minimal-
   scale=1.0"> <title>Home</title>
   <!-- form awesome library for adding icons to the form -->
   <link rel="stylesheet" href="https://use.fontawesome.com/releases/v5.7.1/css/all.css">
   <link rel="stylesheet" href="style.css" type="text/css">
</head>
```

```
<body>
    <div class="login">
        <!--
        <h2><a
        href="/dashboard">Dashboard</a></h2>
        <h2><a href="/info">User Info</a></h2> -->
        <form action="/dashboard" method="post">
            <label for="dashboard">
                <h2 class="fas fa-chart-line"></h2>
            </label>
            <!--
            -->
            <input type="submit" value="analytics">
        </form>
    </div>
    <div class="login">
        <form action="/info" method="post">
            <label for="info">
                <h2 class="fas fa-info"></h2>
            </label>
            <!--
            -->
            <input type="submit" value="info">
        </form>
    </div>
</body>
</html>
```
