PROJECT REPORT

BALA PRIYA.M	19104030
ABINAYA.A	19104002
ARUN PRAKASH.H	19104024
ARUNBALAJI.S	19104025

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. 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 encouraging non-cyclists to try cycling.

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, consisting of a pay-station and multiple bike docks, across the system where bikes can be checked out from one station and returned to another 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" 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, creating the potential for significant reductions in transportation related greenhouse gas or CO2e emissions. However, there is significant heterogeneity between different cities, 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

The purpose of this analysis is to create an operating report of Citi Bike for the year 2018. From this analysis, the following data visualizations will be created.

1. Total Number of Trips

2. What is Customer and subscriber with gender

3.Find the top bike used with respect to trip duration?

4.Calculating the number of bikes used by respective age groups.

5.Top 10 Start Station Names with respect to Customer age group

2.LITERATURE SURVEY

2.1 Existing Problem

Spinlister -Spinlister is an online hub for renting bikes from individuals or bike rental shops.

Zagster - Life is better on a bike! They are bringing bike share to communities across the USA.

Motivate International - Motivate is a global full-service bike share operator and technology innovator.

Spin - Spin is a stationless bike and electric scooter sharing service.

2.2 References

https://craft.co/citi-bike/competitors

Ines et al.,ScienceDirect-Social and Behavioral Sciences 111 (2014) 518 – 527 “ Bicycle sharing systems demand”

Elias et al.,ScienceDirect Journal of Transport Geography 91 (2021) 102971”What do trip data reveal about bike-sharing system users? “

FRANCESCO et al.,IEEE Access 2020"Bike Sharing and Urban Mobility in a Post-Pandemic"

"A long-term perspective on the COVID-19: The bike sharing system resilience under the epidemic environment"Journal of Transport & Health ,2021

Nguyen ThiHoai Thu, Chu Thi Phuong Dung, Vietnam 2017 International Conference on Advanced Technologies for Communications - Multi-source Data Analysis for Bike Sharing Systems

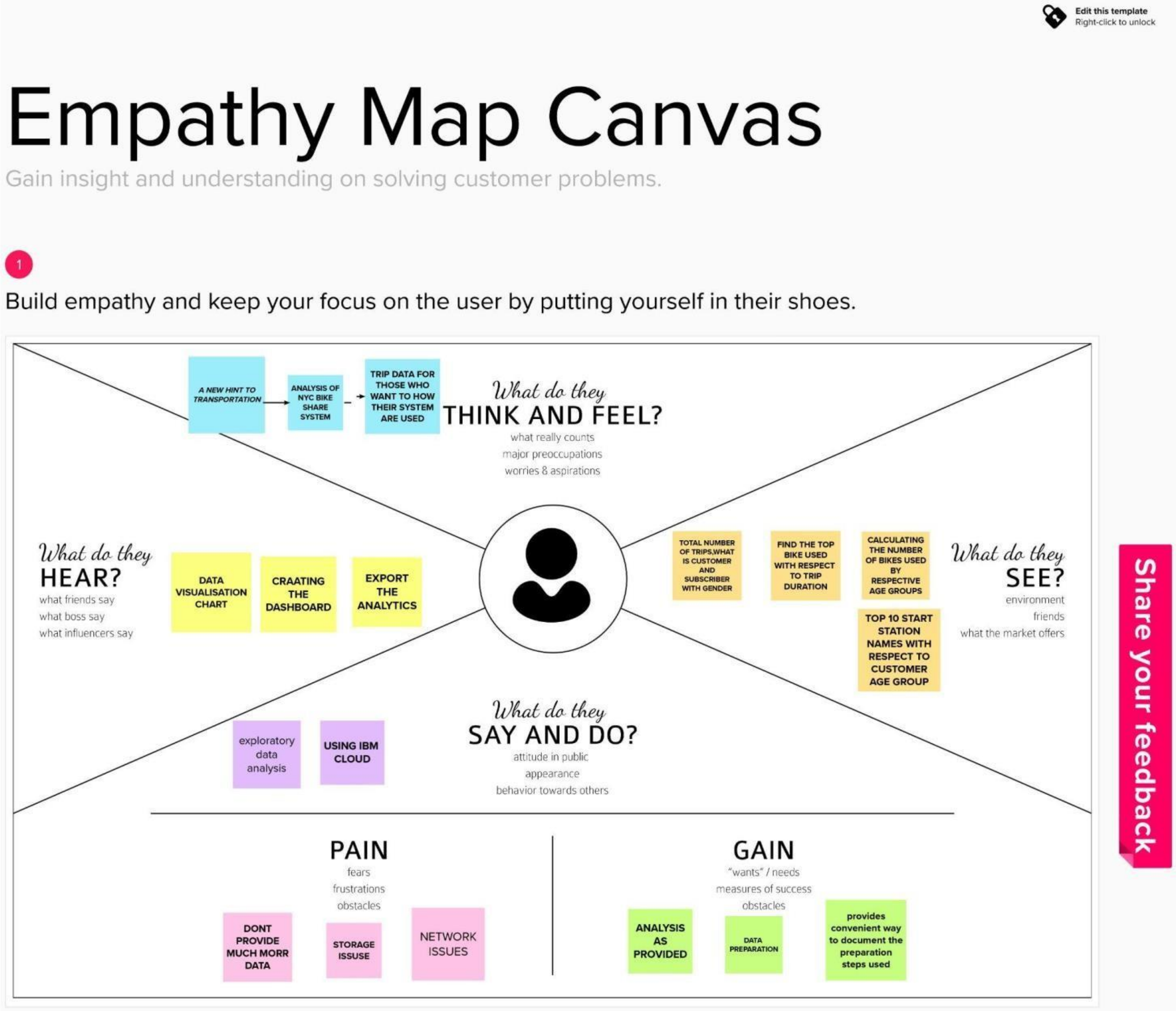
2.3 Problem statement Definition

In busy cities like New York the people are facing difficulties in analyzing the demand for bikes during peak hours.

The main objective of this project is to predict bike patterns that will be extremely helpful for people to plan their travel.

3.IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP



Empathy Map Canvas

Gain insight and understanding on solving customer problems.

Build empathy and keep your focus on the user by putting yourself in their shoes.

What do they THINK AND FEEL?

what really counts
major preoccupations
worries & aspirations

A NEW HINT TO TRANSPORTATION → ANALYSIS OF NYC BIKE SHARE SYSTEM → TRIP DATA FOR THOSE WHO WANT TO HOW THEIR SYSTEM ARE USED

What do they HEAR?

what friends say
what boss say
what influencers say

DATA VISUALISATION CHART

CRAATING THE DASHBOARD

EXPORT THE ANALYTICS

What do they SEE?

environment
friends
what the market offers

TOTAL NUMBER OF TRIPS,WHAT IS CUSTOMER AND SUBSCRIBER WITH GENDER

FIND THE TOP BIKE USED WITH RESPECT TO TRIP DURATION

CALCULATING THE NUMBER OF BIKES USED BY RESPECTIVE AGE GROUPS

TOP 10 START STATION NAMES WITH RESPECT TO CUSTOMER AGE GROUP

What do they SAY AND DO?

attitude in public
appearance
behavior towards others

exploratory data analysis

USING IBM CLOUD

PAIN

fears
frustrations
obstacles

DONT PROVIDE MUCH MORR DATA

STORAGE ISSUSE

NETWORK ISSUES

GAIN

"wants" / needs
measures of success
obstacles

ANALYSIS AS PROVIDED

DATA PREPARATION

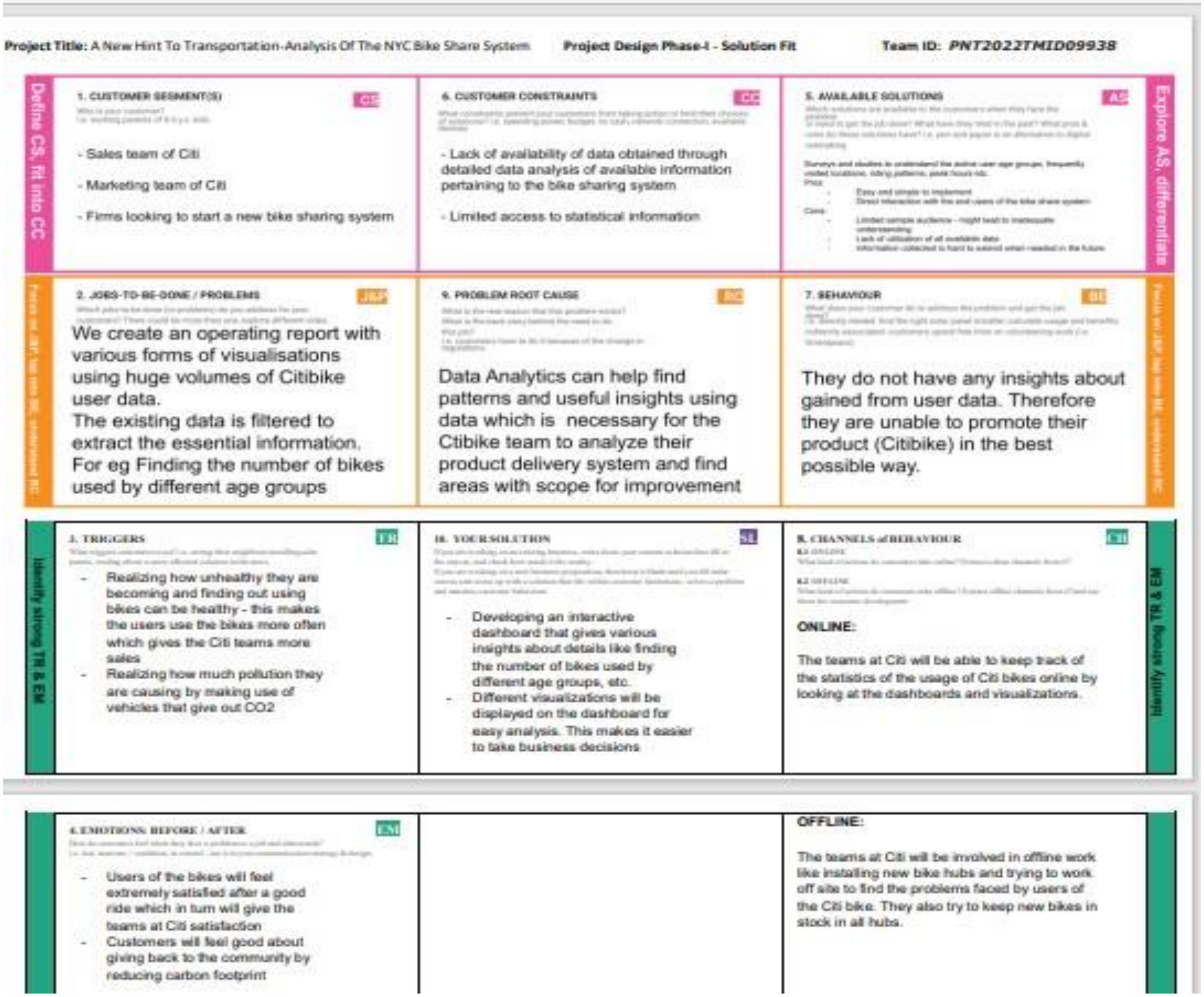
provides convenient way to document the preparation steps used

3.2 Proposed Solution

s.no	Parameter	Description
1	Problem Statement (Problem to be solved)	A New Hint to Transportation - An Analysis of the NYC bike share system. This analysis aimsto create an operating report of Citi Bike for the year 2018 and create various data visualizations for given problem queries.
2	Idea / Solution Description	We have with us millions of recorded data. It is difficult to process the data with normal databases. To effectively process and visualize data, we use IBM Cognos, a web-based integrated business intelligence suite by IBM. Using IBM Cognos, we aim to create an operating report, provide useful insights and present them in the form of a dashboard.
3	Novelty / Uniqueness	-> Apart from creating just the data visualizationsthat have been asked, we aim to create an interactive dashboard that can take certain user inputs and display visualizations that are specific to the input. Example: Take as input the start date and end date - Display the total number of trips between the given start and end date -> We plan to find other statistics that can

		be of use to the manufacturers: ● Coordinates, where any given bike is parked the longest in between the start and end hub - these coordinates, could be those of a potential new hub. ● The conversion rate of customers(who could be non-tourists) to subscribers. -> Sending the final resultant statistics as an email to a specific set of people who might benefit greatly from it - they can store it for future use.
4	Social Impact/ Customer Satisfaction	Identifying the age groups and frequently visited locations can help us come up with a more targeted business approach. Thisis highly helpful when the current system is expanded and new stations are introduced. Availability of an improved, eco-friendly, alternative transport solution that provides health benefits encourages customers to opt for it instead of currently existing transport options.
5	Business Model (Revenue Model	Government can promote environment-friendly bicycles. Fitness companies can run campaigns to target the right customers. Citi Bike has five different sources of revenue, with an annual membership, sponsorship, and casual membership being the three most important. Together these three categories made up over 85% of Citi Bike's total revenue (in 2019).
6	Scalability of the Solution	The creation of the operating report (solution) involves an extended analysis of data presented for the year 2018. Our solution offers high scalability, as not only can it be extended for any number of years provided the right data if offered, but it can also be made use of by other companies as a scalable template that wants to make similar reports on different solutions they offer.

3.3Problem Solution Fit



Project Title: A New Hint To Transportation-Analysis Of The NYC Bike Share System **Project Design Phase-I - Solution Fit** **Team ID:** ***PNT2022TMID09938***

Define CS, fit into CC
Explore AS, differentiate
Focus on J&P, tap into BE, understand RC
Identify strong TR & EM
CS
CC
AS
J&P
RC
BE
TR
SL
CH
EM

1. CUSTOMER SEGMENT(S)

Who is your customer?

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

6. CUSTOMER CONSTRAINTS

- 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

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 utilisation of all available data
- Information collected is hard to extend when needed in the future

2. JOBS-TO-BE-DONE / PROBLEMS

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

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

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.

3. TRIGGERS

- 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

- 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

8.1 ONLINE

8.2 OFFLINE

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.

4. EMOTIONS: BEFORE / AFTER

- 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:

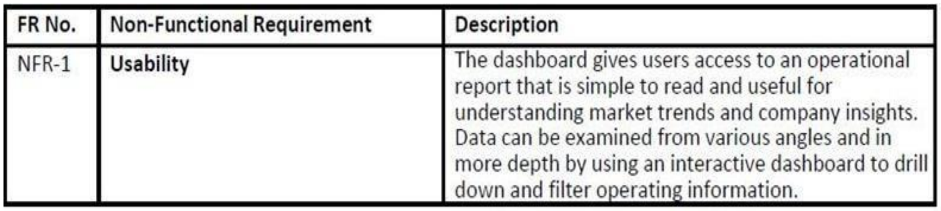
The teams at Citi will be involved in offline work like installing new bike hubs and trying to work off site to find the problems faced by users of the Citi bike. They also try to keep new bikes in stock in all hubs.

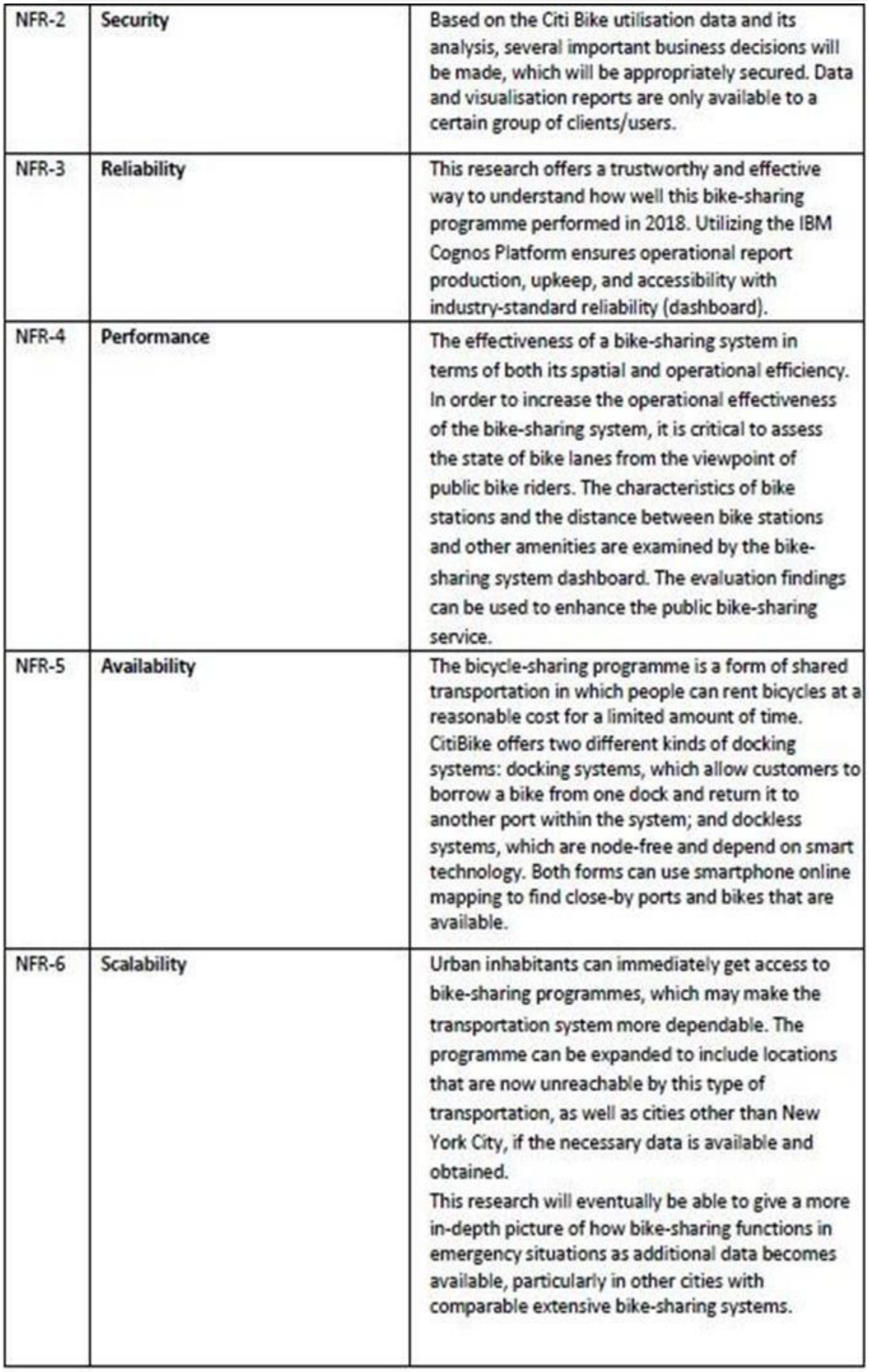
4.REQUIREMENT ANALYSIS

4.1 Functional Requirement

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	Analysis of the given data includes carrying out preprocessing & filtering the data as per the requirement posed by the sub-task. The usage of Machine Learning techniques to gain further insights into the data also contributes to the analysis, and as a result visualization of data.
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.

4.2 NON-Functional Requirement



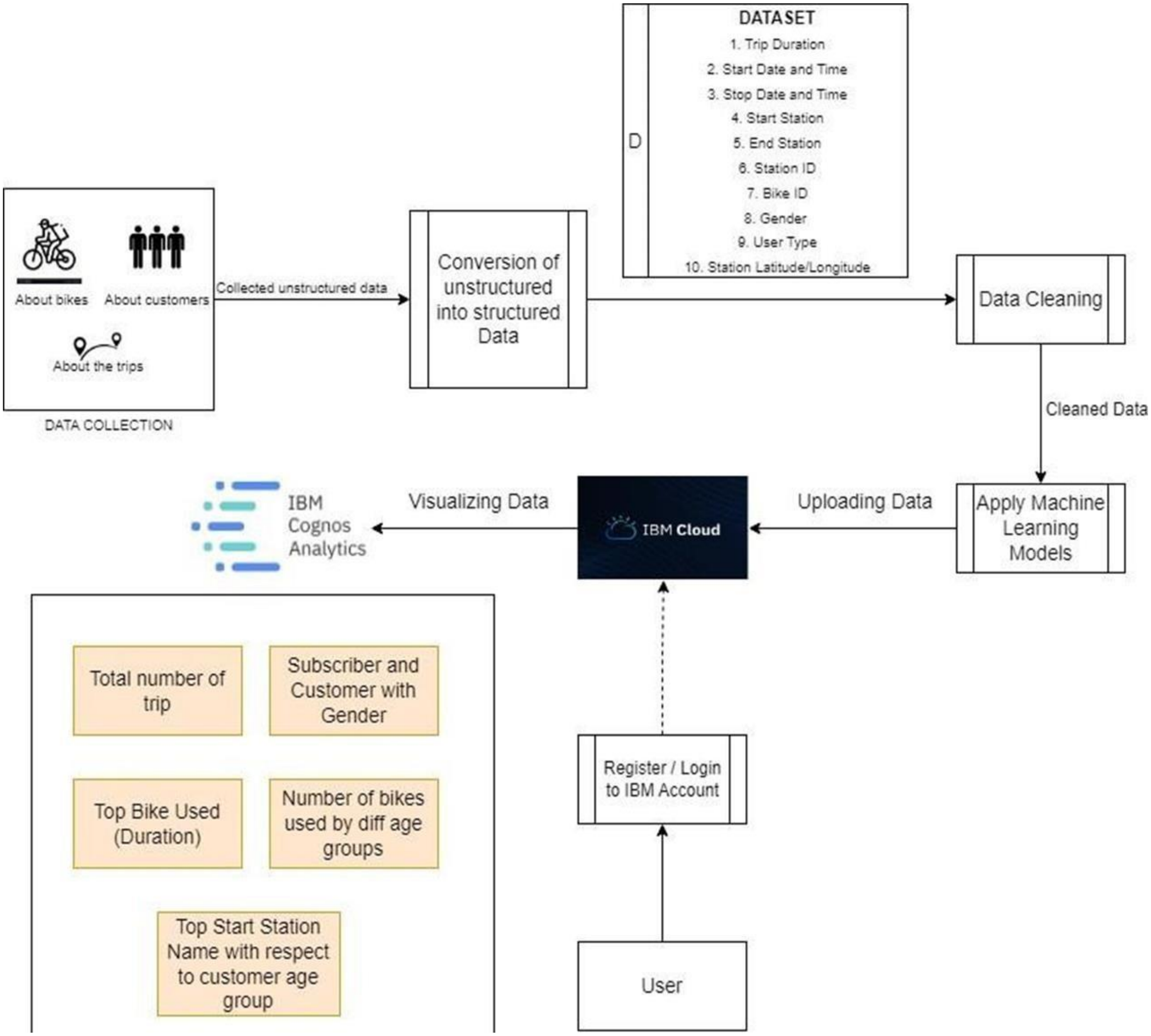


FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The dashboard gives users access to an operational report that is simple to read and useful for understanding market trends and company insights. Data can be examined from various angles and in more depth by using an interactive dashboard to drill down and filter operating information.
NFR-2	**Security**	Based on the Citi Bike utilisation data and its analysis, several important business decisions will be made, which will be appropriately secured. Data and visualisation reports are only available to a certain group of clients/users.
NFR-3	**Reliability**	This research offers a trustworthy and effective way to understand how well this bike-sharing programme performed in 2018. Utilizing the IBM Cognos Platform ensures operational report production, upkeep, and accessibility with industry-standard reliability (dashboard).
NFR-4	**Performance**	The effectiveness of a bike-sharing system in terms of both its spatial and operational efficiency. In order to increase the operational effectiveness of the bike-sharing system, it is critical to assess the state of bike lanes from the viewpoint of public bike riders. The characteristics of bike stations and the distance between bike stations and other amenities are examined by the bike-sharing system dashboard. The evaluation findings can be used to enhance the public bike-sharing service.
NFR-5	**Availability**	The bicycle-sharing programme is a form of shared transportation in which people can rent bicycles at a reasonable cost for a limited amount of time. CitiBike offers two different kinds of docking systems: docking systems, which allow customers to borrow a bike from one dock and return it to another port within the system; and dockless systems, which are node-free and depend on smart technology. Both forms can use smartphone online mapping to find close-by ports and bikes that are available.
NFR-6	**Scalability**	Urban inhabitants can immediately get access to bike-sharing programmes, which may make the transportation system more dependable. The programme can be expanded to include locations that are now unreachable by this type of transportation, as well as cities other than New York City, if the necessary data is available and obtained. This research will eventually be able to give a more in-depth picture of how bike-sharing functions in emergency situations as additional data becomes available, particularly in other cities with comparable extensive bike-sharing systems.

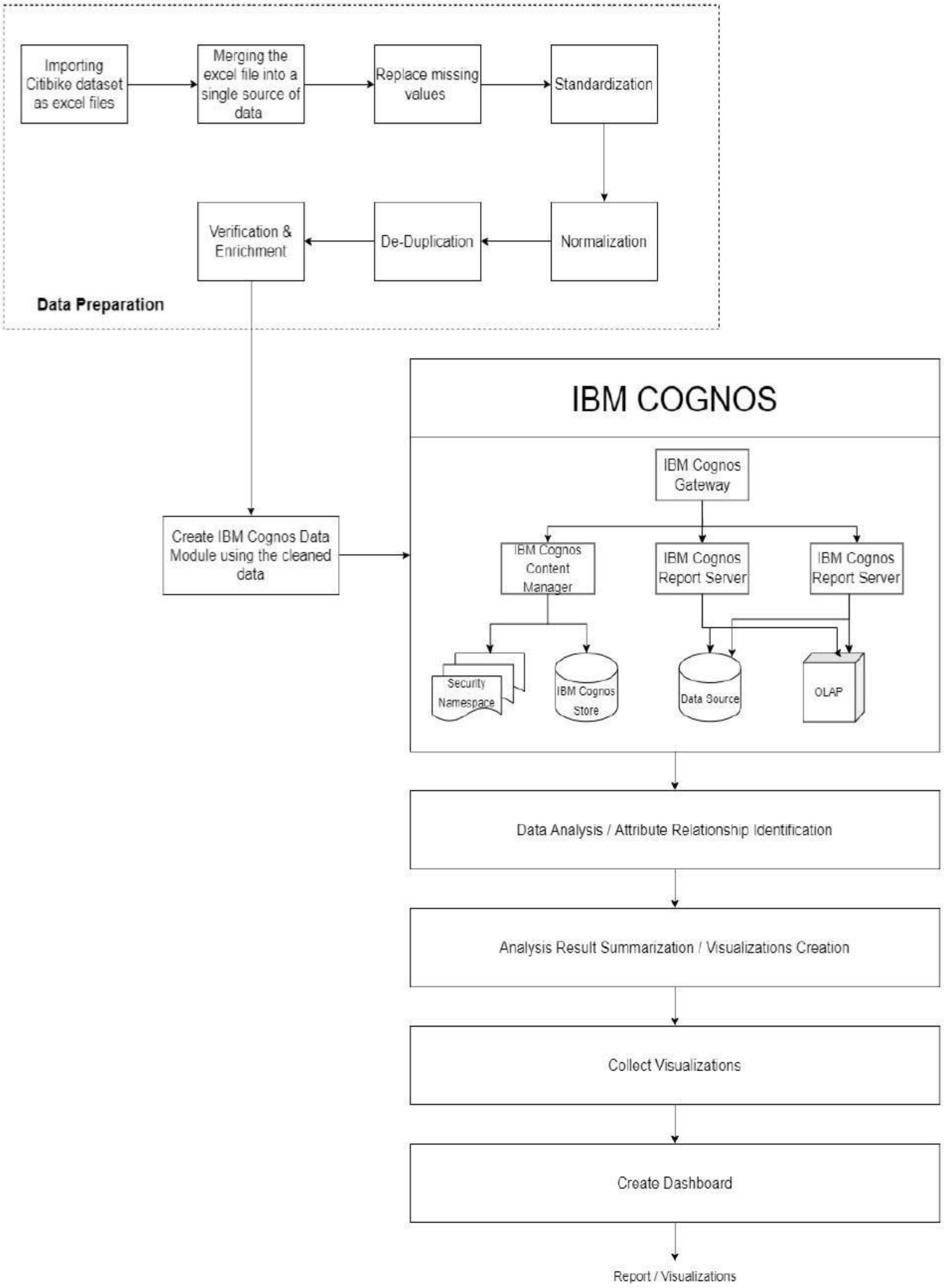
5.PROJECT DESIGN

5.1 Data Flow Diagram

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



5.2 Solution Architecture Diagram:


Importing Citibike dataset as excel files
Merging the excel file into a single source of data
Replace missing values
Standardization
Normalization
De-Duplication
Verification & Enrichment
Data Preparation
Create IBM Cognos Data Module using the cleaned data
IBM COGNOS
IBM Cognos Gateway
IBM Cognos Content Manager
IBM Cognos Report Server
IBM Cognos Report Server
Security Namespace
IBM Cognos Store
Data Source
OLAP
Data Analysis / Attribute Relationship Identification
Analysis Result Summarization / Visualizations Creation
Collect Visualizations
Create Dashboard
Report / Visualizations

\

5.3 User Stories:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Releae
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

6.PROJECT PLANNING & SCHEDULING

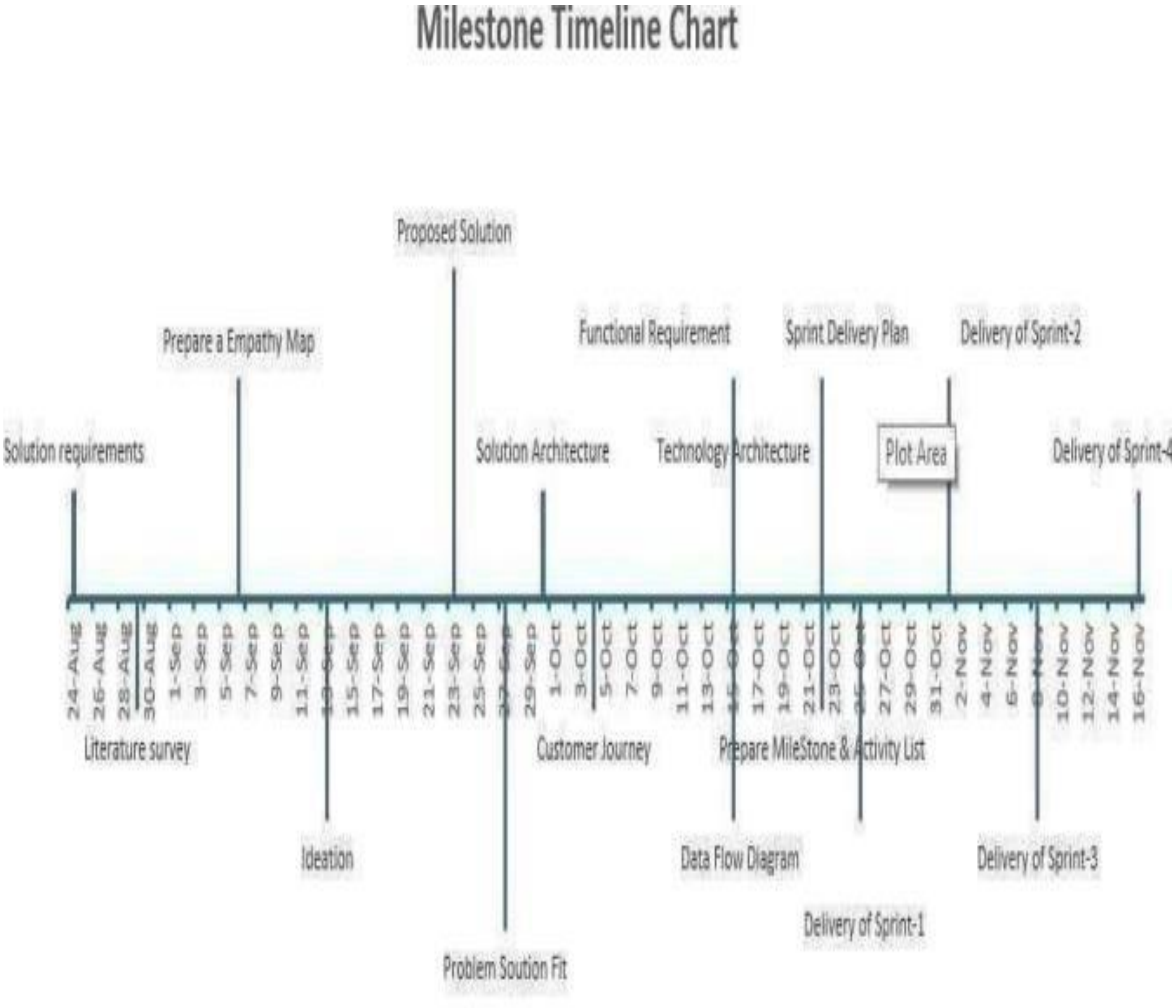
6.1 Sprint Planning & Estimation

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.	2	High	Balapriya M, Abinaya A
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	2	High	Arunbalaji S, Arunprakash H
Sprint-1		USN-3	As a user, I can register for the application through Gmail	2	Medium	Balapriya M, Arunprakash H.
Sprint-2	Login	USN-4	As a user, I can log into the application by entering email & password	2	High	Balapriya M, Abinaya A, Arunbalaji S.
Sprint-2	Collection of user data	USN-5	I can access and collect the citi bike share system data from Lyft citi bike's official website that has the published files.	2	Medium	Balapriya M, Arunprakash H.
Sprint-2		USN-6	I can use the citi bike share system data for analysis purposes	5	High	Balapriya M, Abinaya A.
Sprint-3	Analysing the user data	USN-7	The data is used as input for creating various types of visualizations and analysis is done. I can view the analysis of the citi bike	8	High	Balapriya M, Abinaya A, Arunbalaji S.

Sprint-3	Dashboard	USN-8	I can register & access the dashboard created based on the analysis by logging in	3	Medium	Abinaya A, Arunprakash H, Arunbalaji S.
Sprint-3		USN-9	As a user I can view the dashboard that displays the top bike used with respect to trip duration	5	High	Balapriya M.
Sprint-4		USN-10	As a user I can view the dashboard that displays the top 10 Start Station Names with respect to customer age group	5	High	Abinaya A.
Sprint-4		USN-11	As a user I can view the dashboard that displays the customer and subscriber with respect to gender	5	High	Arunbalaji S.
Sprint-4		USN-12	As a user I can view the dashboard that displays the total number of trips	5	High	Arunprakash H.

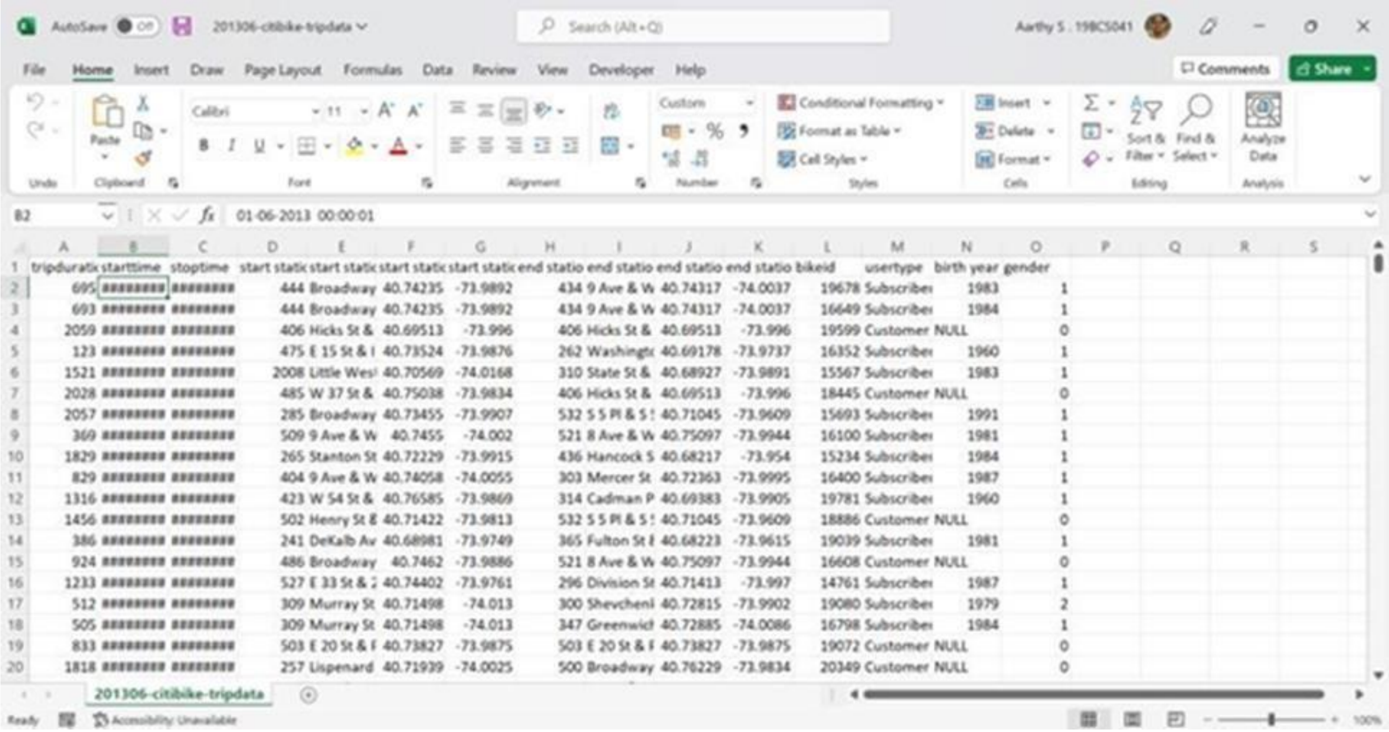
Sprint	**Total Story Points**	**Duration**	**Sprint Start Date**	**Sprint End Date (Planned)**
Sprint-1	6	6 Days	24 Oct 2022	29 Oct 2022
Sprint-2	9	6 Days	31 Oct 2022	05 Nov 2022
Sprint-3	16	6 Days	07 Nov 2022	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022

6.2 Sprint Delivery Schedule



7.WORKING WITH THE DATASET & DATA VISUALISATION

7.1 Understanding the dataset



B2 01-06-2013 00:00:01

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	tripduratic	starttime	stoptime	start static	start static	start static	start static	end statio	end statio	end statio	end statio	bikeid	usertype	birth year	gender
2	695	########	########	444	Broadway	40.74235	-73.9892	434	9 Ave & W	40.74317	-74.0037	19678	Subscriber	1983	1
3	693	########	########	444	Broadway	40.74235	-73.9892	434	9 Ave & W	40.74317	-74.0037	16649	Subscriber	1984	1
4	2059	########	########	406	Hicks St &	40.69513	-73.996	406	Hicks St &	40.69513	-73.996	19599	Customer	NULL	0
5	123	########	########	475	E 15 St & I	40.73524	-73.9876	262	Washingtc	40.69178	-73.9737	16352	Subscriber	1960	1
6	1521	########	########	2008	Little Wes	40.70569	-74.0168	310	State St &	40.68927	-73.9891	15567	Subscriber	1983	1
7	2028	########	########	485	W 37 St &	40.75038	-73.9834	406	Hicks St &	40.69513	-73.996	18445	Customer	NULL	0
8	2057	########	########	285	Broadway	40.73455	-73.9907	532	S 5 Pl & S !	40.71045	-73.9609	15693	Subscriber	1991	1
9	369	########	########	509	9 Ave & W	40.7455	-74.002	521	8 Ave & W	40.75097	-73.9944	16100	Subscriber	1981	1
10	1829	########	########	265	Stanton St	40.72229	-73.9915	436	Hancock S	40.68217	-73.954	15234	Subscriber	1984	1
11	829	########	########	404	9 Ave & W	40.74058	-74.0055	303	Mercer St	40.72363	-73.9995	16400	Subscriber	1987	1
12	1316	########	########	423	W 54 St &	40.76585	-73.9869	314	Cadman P	40.69383	-73.9905	19781	Subscriber	1960	1
13	1456	########	########	502	Henry St &	40.71422	-73.9813	532	S 5 Pl & S !	40.71045	-73.9609	18886	Customer	NULL	0
14	386	########	########	241	DeKalb Av	40.68981	-73.9749	365	Fulton St &	40.68223	-73.9615	19039	Subscriber	1981	1
15	924	########	########	486	Broadway	40.7462	-73.9886	521	8 Ave & W	40.75097	-73.9944	16608	Customer	NULL	0
16	1233	########	########	527	E 33 St & 2	40.74402	-73.9761	296	Division St	40.71413	-73.997	14761	Subscriber	1987	1
17	512	########	########	309	Murray St	40.71498	-74.013	300	Shevchenl	40.72815	-73.9902	19080	Subscriber	1979	2
18	505	########	########	309	Murray St	40.71498	-74.013	347	Greenwicl	40.72885	-74.0086	16798	Subscriber	1984	1
19	833	########	########	503	E 20 St & F	40.73827	-73.9875	503	E 20 St & F	40.73827	-73.9875	19072	Customer	NULL	0
20	1818	########	########	257	Lispenard	40.71939	-74.0025	500	Broadway	40.76229	-73.9834	20349	Customer	NULL	0

Dataset Link: Dataset

1. Trip Duration: How long a trip lasted in seconds

2. Start Date and Time: EX->01-06-2013 00:00:01

3. Stop Date and Time: EX->01-06-2013 00:11:36

4. Start Station ID: Unique identifier for each station

5. Start Station Name

6. Start Station Latitude: Coordinates

7. Start Station Longitude: Coordinates

8. End Station ID: Unique identifier for each station

9. End Station Name

10. End Station Latitude

11. End Station Longitude

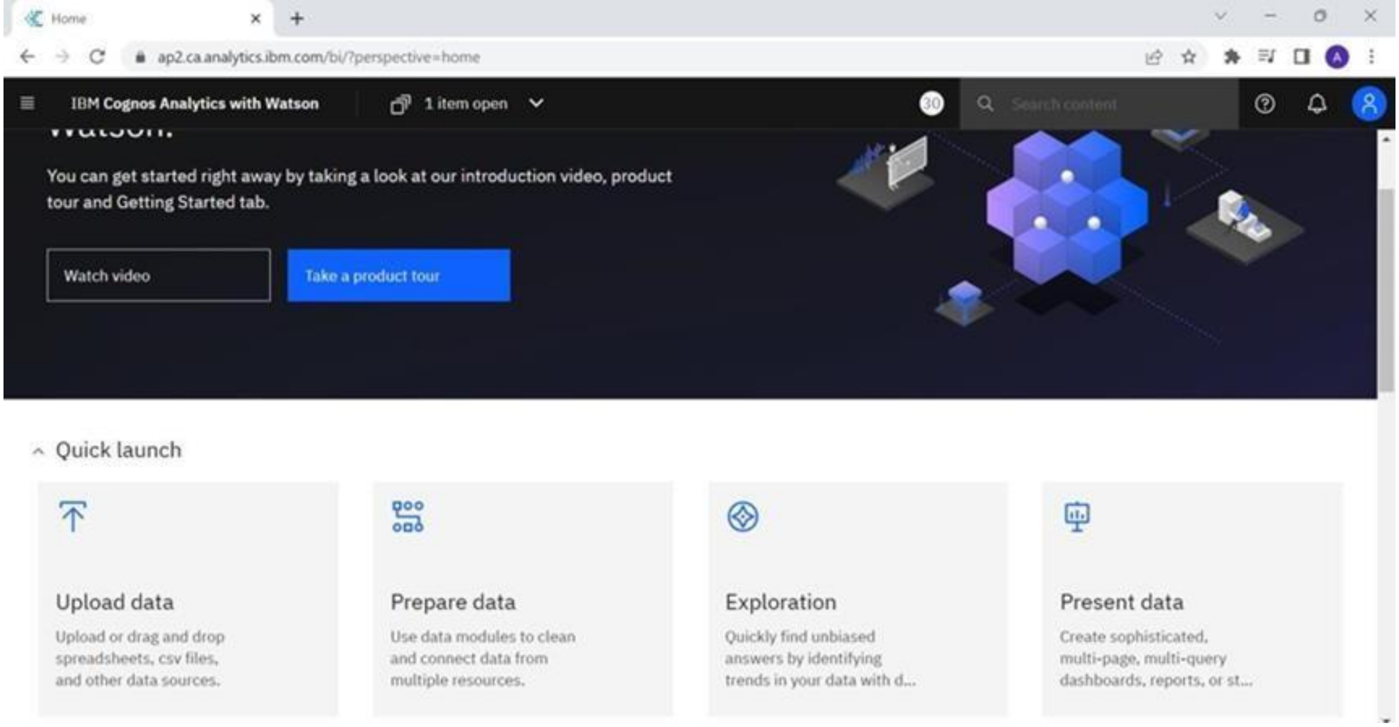
12. Bike ID: Unique identifier for each bike

13. User Type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member): Customers are usually tourists, subscribers are usually NYC residents

14. Year of Birth: Self-entered, not validated by an ID Gender (Zero=unknown; 1=male; 2=female): Usually unknown for customers since they often sign up at a kiosk

7.2 Loading the dataset

Open Cognos Analytics and click upload data


Home
ap2.ca.analytics.ibm.com/bi/?perspective=home
IBM Cognos Analytics with Watson
1 item open
Search content
Watch video
Take a product tour
Quick launch

You can get started right away by taking a look at our introduction video, product tour and Getting Started tab.

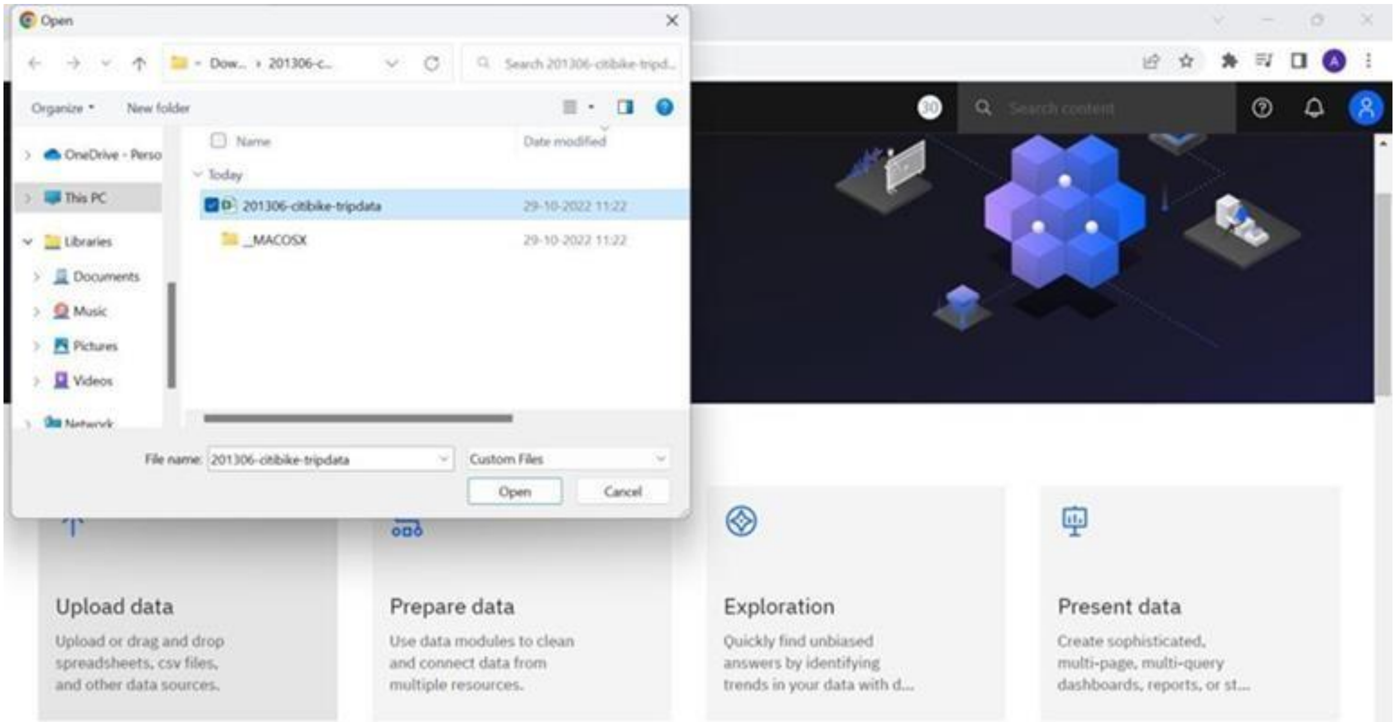
Upload data
Upload or drag and drop spreadsheets, csv files, and other data sources.

Prepare data
Use data modules to clean and connect data from multiple resources.

Exploration
Quickly find unbiased answers by identifying trends in your data with d...

Present data
Create sophisticated, multi-page, multi-query dashboards, reports, or st...

Select the dataset to be uploaded



Name	Date modified
201306-citibike-tripdata	29-10-2022 11:22
_MACOSX	29-10-2022 11:22

File name: 201306-citibike-tripdata
Custom Files

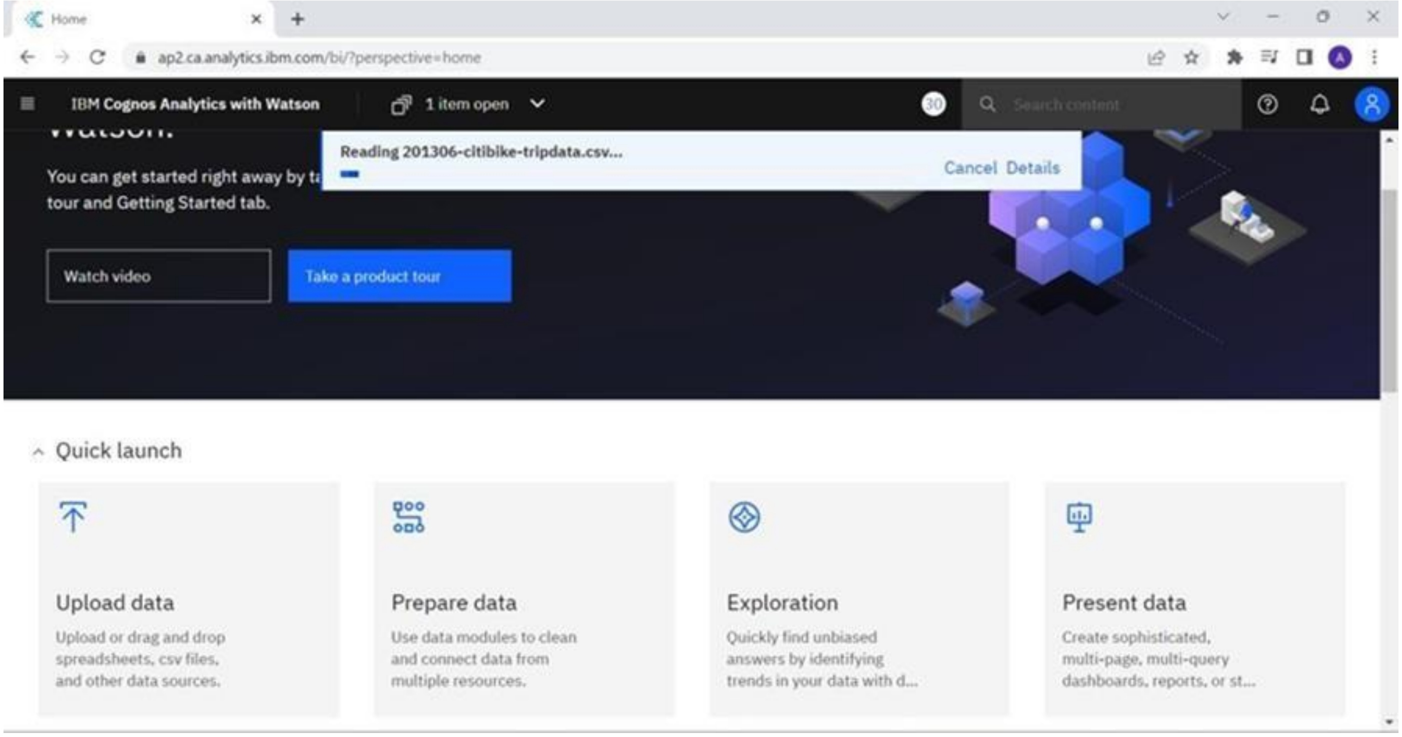
Upload data
Upload or drag and drop spreadsheets, csv files, and other data sources.

Prepare data
Use data modules to clean and connect data from multiple resources.

Exploration
Quickly find unbiased answers by identifying trends in your data with d...

Present data
Create sophisticated, multi-page, multi-query dashboards, reports, or st...

The excel file is getting uploaded in Cognos Analytics


Home
ap2.ca.analytics.ibm.com/bi/?perspective=home
IBM Cognos Analytics with Watson
1 item open
30
Search content
Watch video
Take a product tour
Cancel
Details

You can get started right away by ta… tour and Getting Started tab.

Reading 201306-citibike-tripdata.csv...

Quick launch

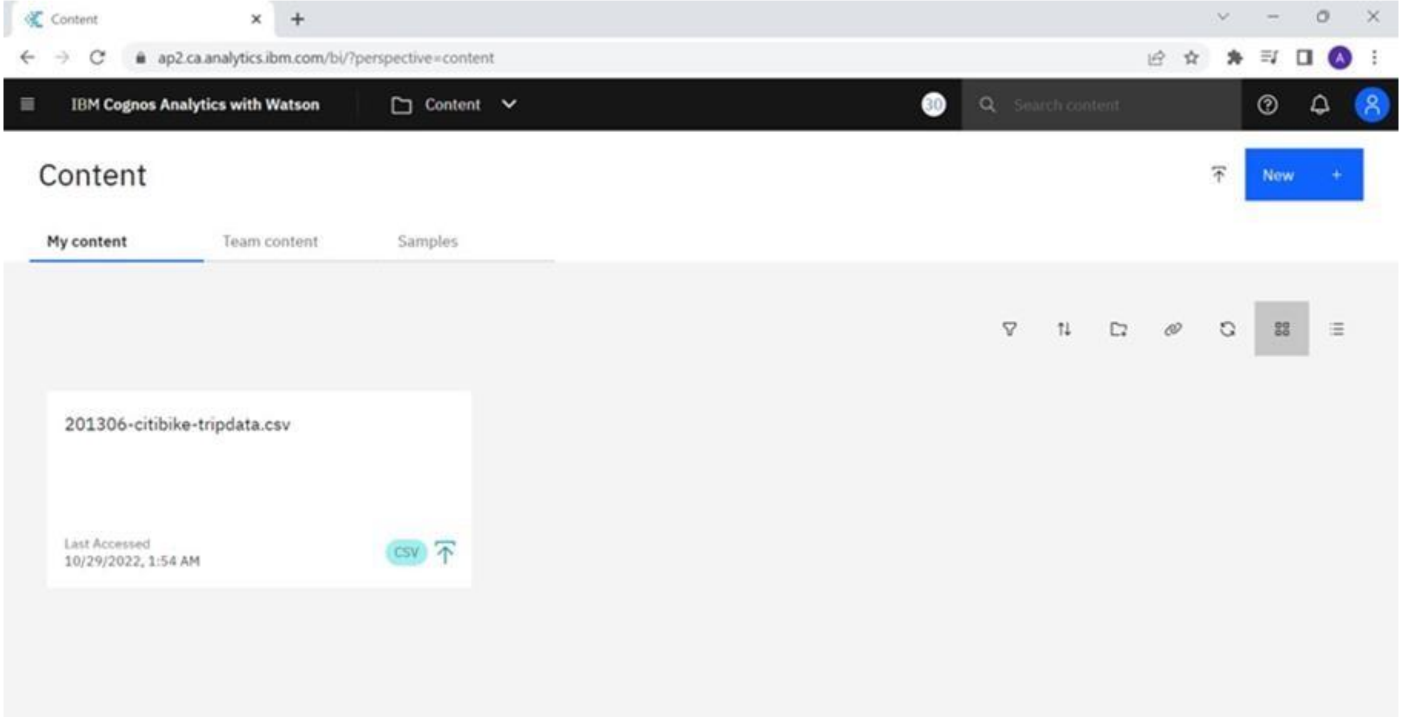
Upload data
Upload or drag and drop spreadsheets, csv files, and other data sources.

Prepare data
Use data modules to clean and connect data from multiple resources.

Exploration
Quickly find unbiased answers by identifying trends in your data with d...

Present data
Create sophisticated, multi-page, multi-query dashboards, reports, or st...

The dataset can be accessed in My Content in Cognos Analytics


Content
ap2.ca.analytics.ibm.com/bi/?perspective=content
IBM Cognos Analytics with Watson
Content
30
Search content
New +
My content
Team content
Samples
CSV

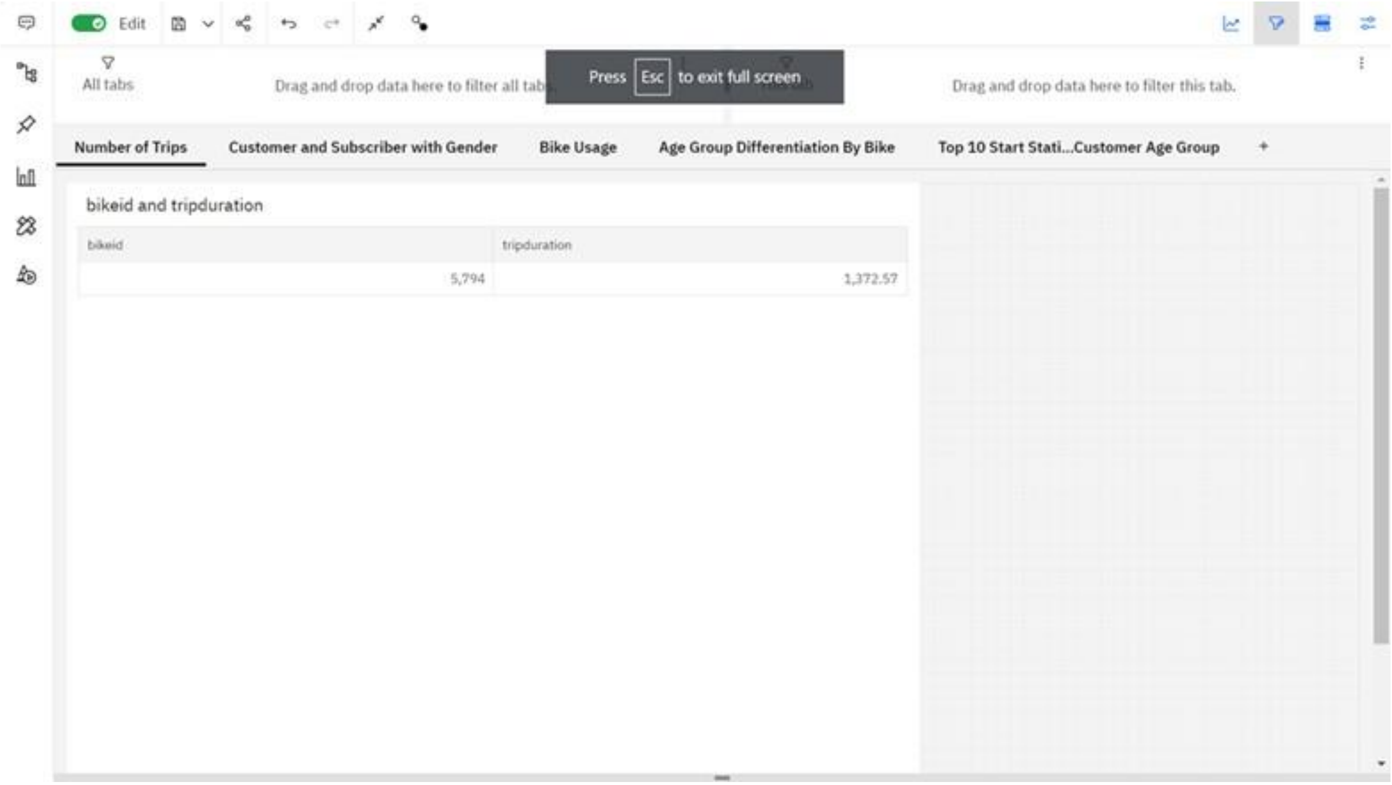
Content

201306-citibike-tripdata.csv

Last Accessed
10/29/2022, 1:54 AM

7.3 Visualization charts

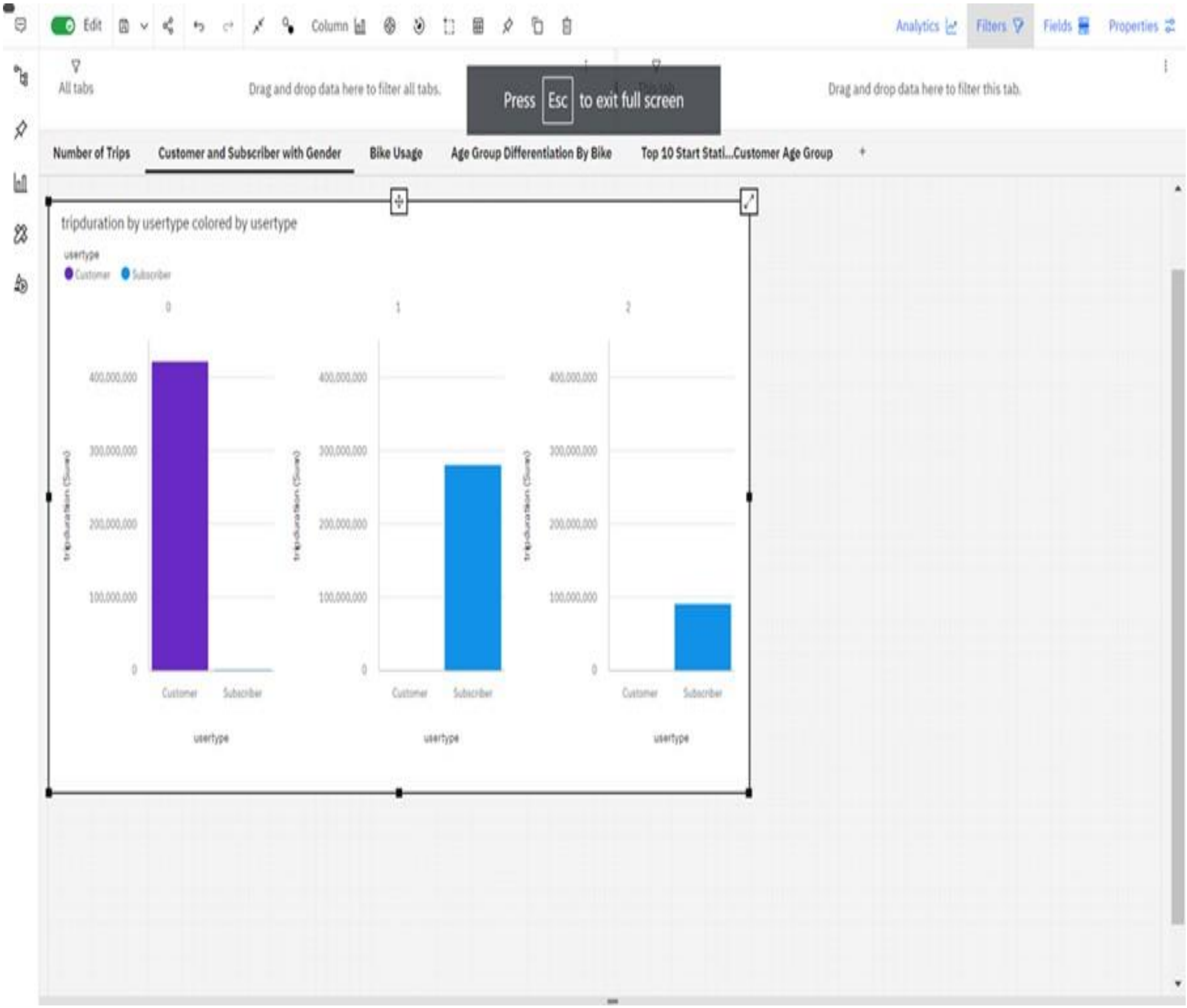
Number of Trips:


Edit
All tabs
Drag and drop data here to filter all tabs.
Press Esc to exit full screen
Drag and drop data here to filter this tab.
Number of Trips
Customer and Subscriber with Gender
Bike Usage
Age Group Differentiation By Bike
Top 10 Start Stati...Customer Age Group
+

bikeid and tripduration

bikeid	tripduration
5,794	1,372.57

Customer and Subscriber with Gender:



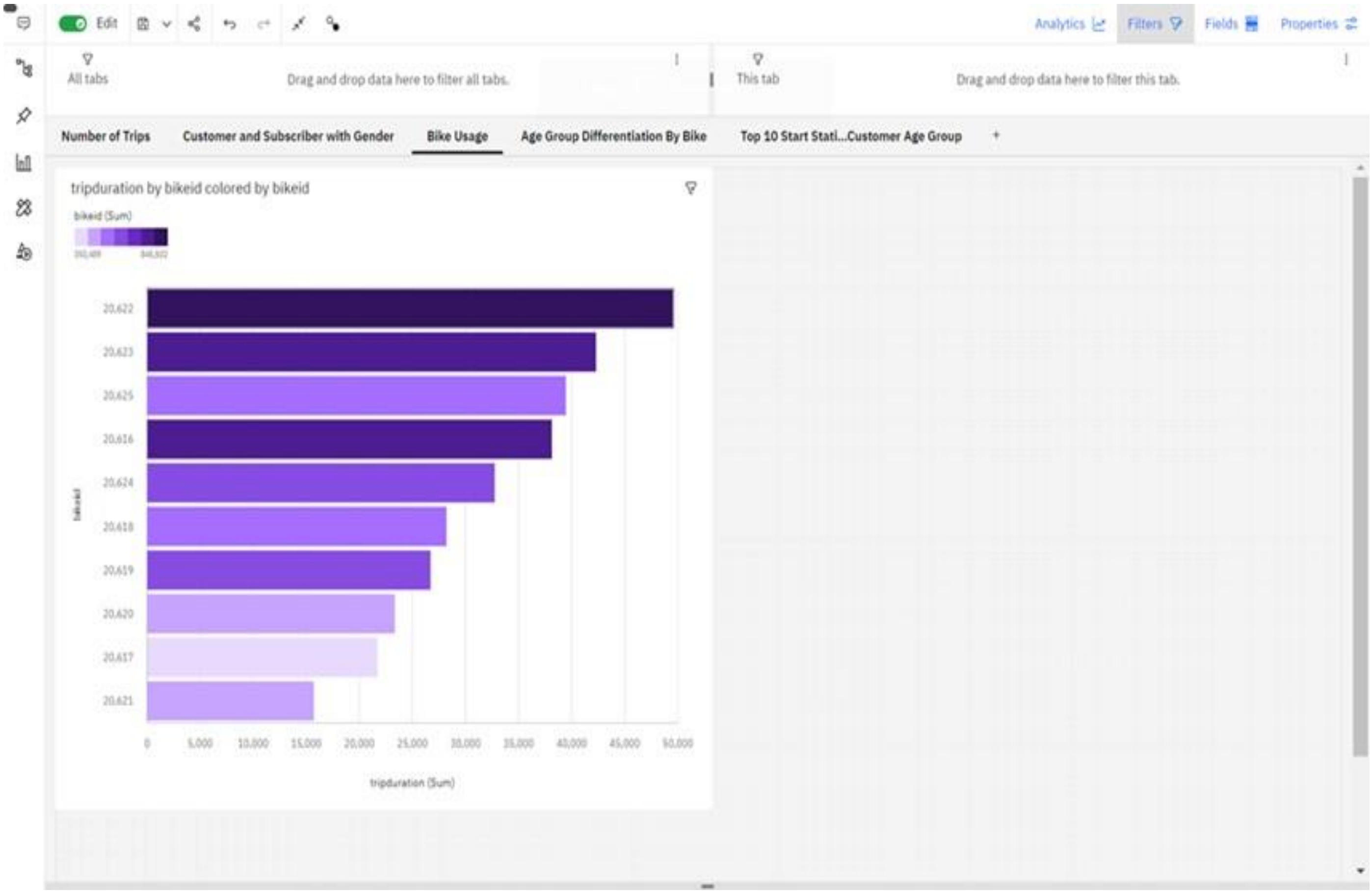
Drag and drop data here to filter all tabs.

Press Esc to exit full screen

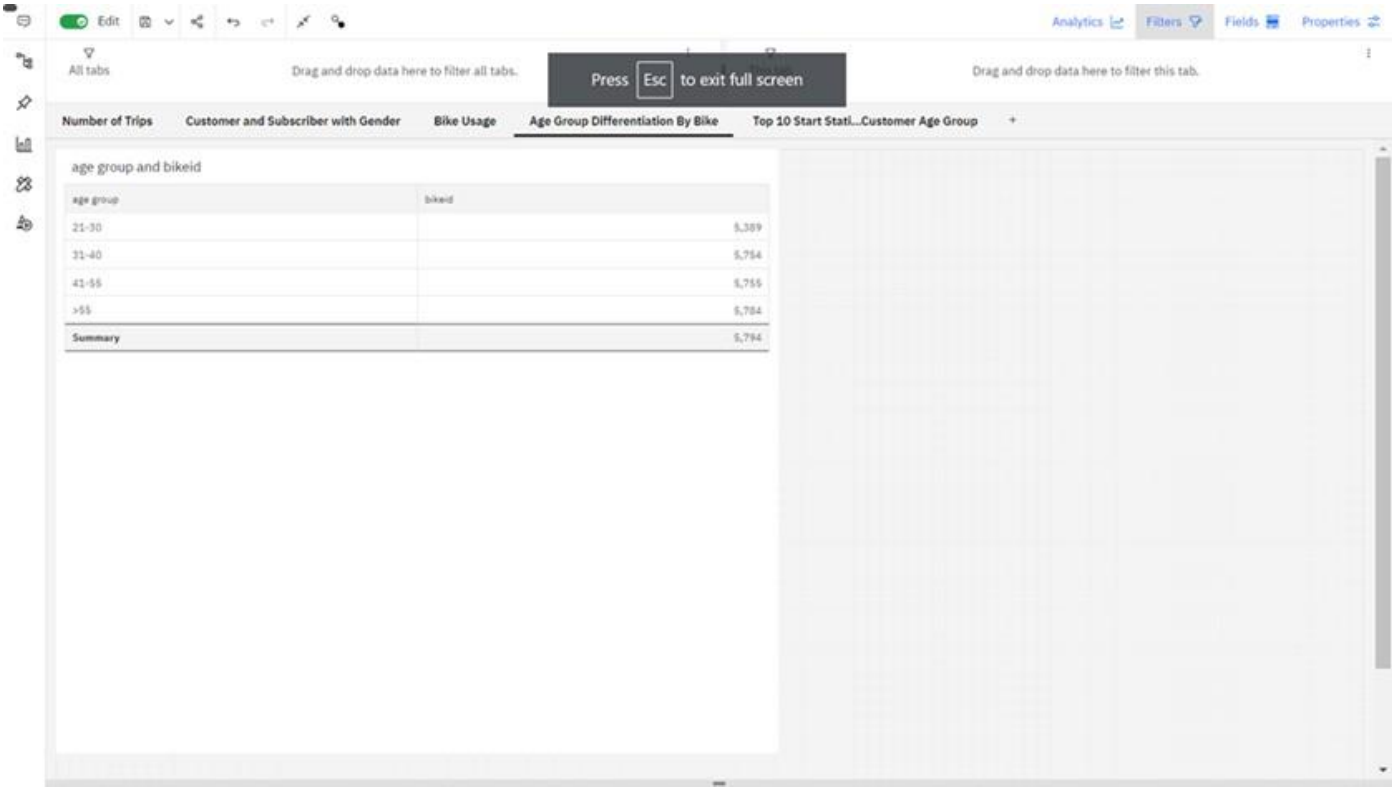
Drag and drop data here to filter this tab.

tripduration by usertype colored by usertype

Bike Usage:


Edit
Analytics
Filters
Fields
Properties
All tabs
Drag and drop data here to filter all tabs.
This tab
Drag and drop data here to filter this tab.
Number of Trips
Customer and Subscriber with Gender
Bike Usage
Age Group Differentiation By Bike
Top 10 Start Stati...Customer Age Group
+
tripduration by bikeid colored by bikeid
bikeid (Sum)
20,622
20,623
20,625
20,616
20,624
20,618
20,619
20,620
20,617
20,621
bikeid
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
50,000
tripduration (Sum)

Age group differentiation by bike

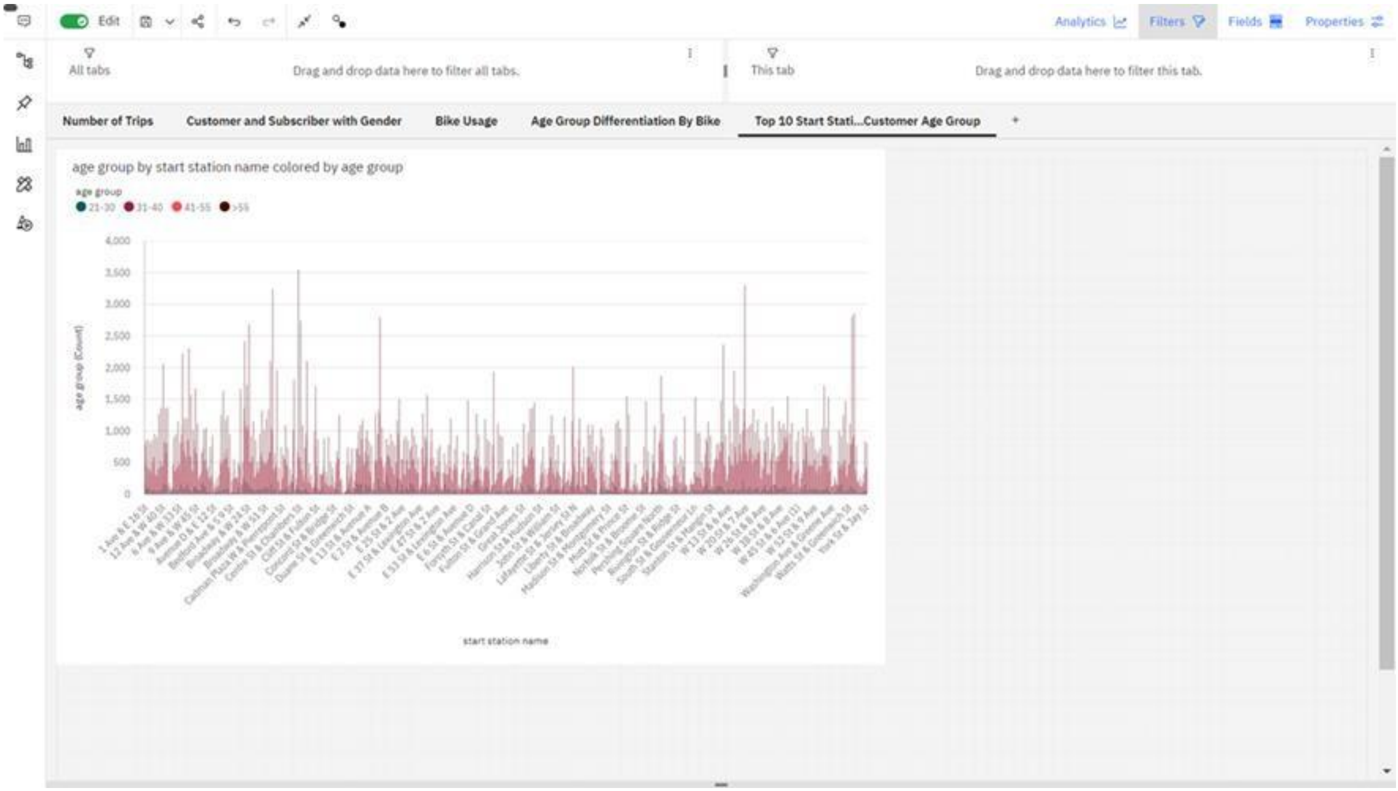


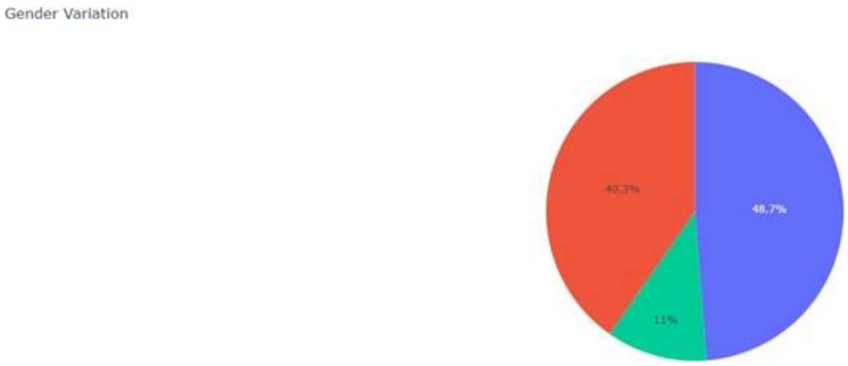
Edit
Analytics
Filters
Fields
Properties
All tabs
Drag and drop data here to filter all tabs.
Press Esc to exit full screen
Drag and drop data here to filter this tab.
Number of Trips
Customer and Subscriber with Gender
Bike Usage
Age Group Differentiation By Bike
Top 10 Start Stati...Customer Age Group
+

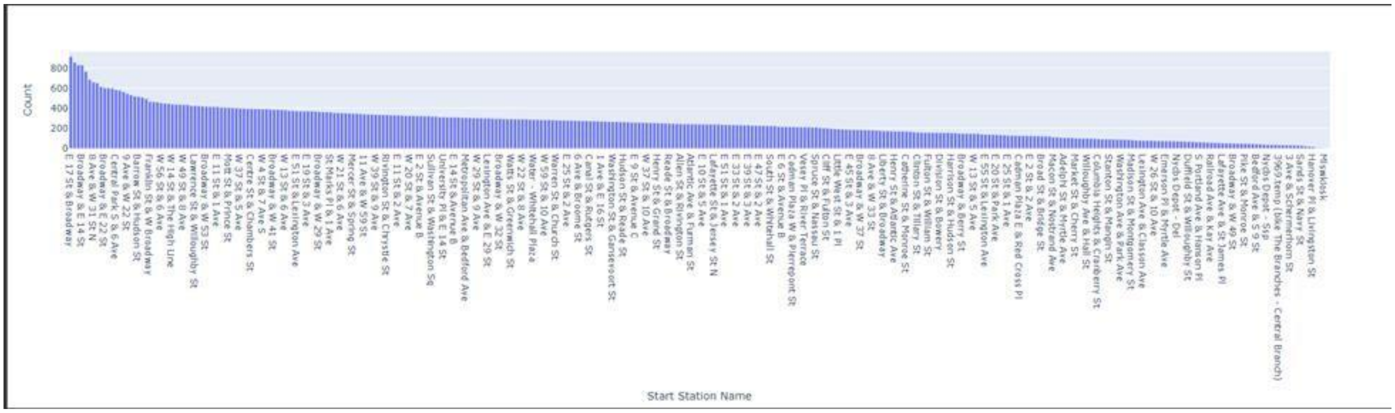
age group and bikeid

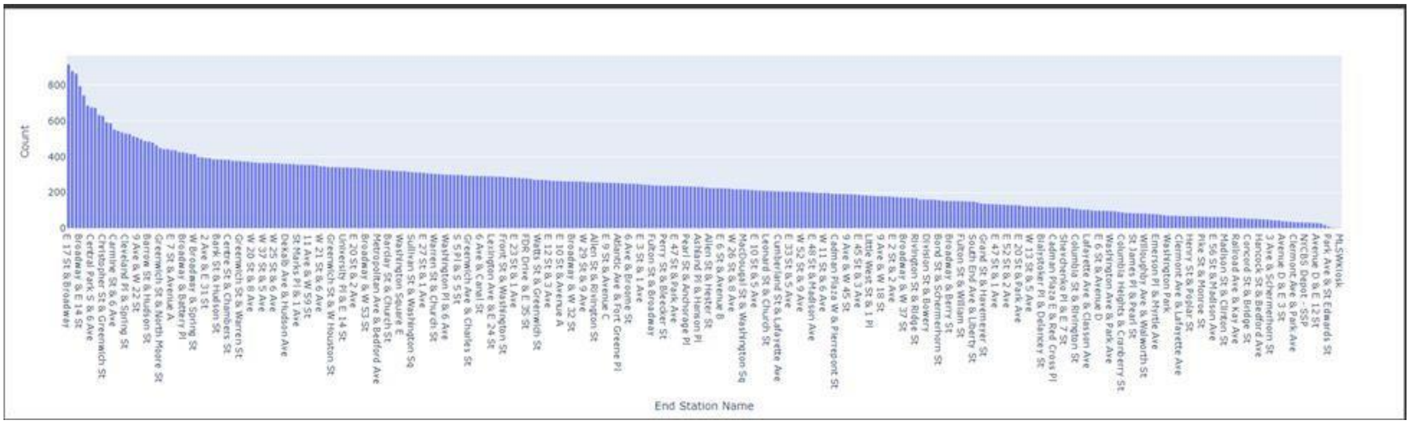
age group	bikeid
21-30	5,389
31-40	5,754
41-55	5,755
>55	5,784
Summary	5,794

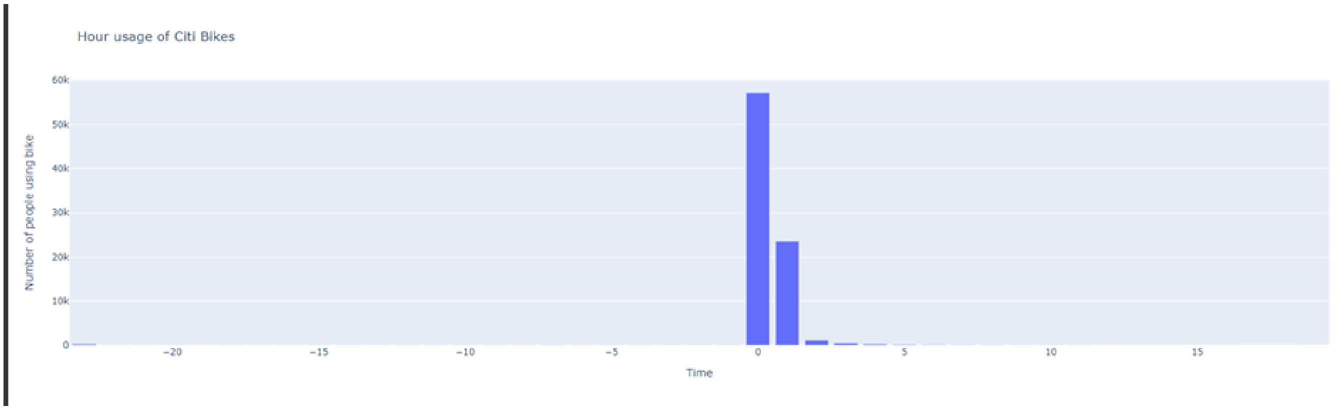
Top 10 Start Station Names with Respect to Customer Age Group:



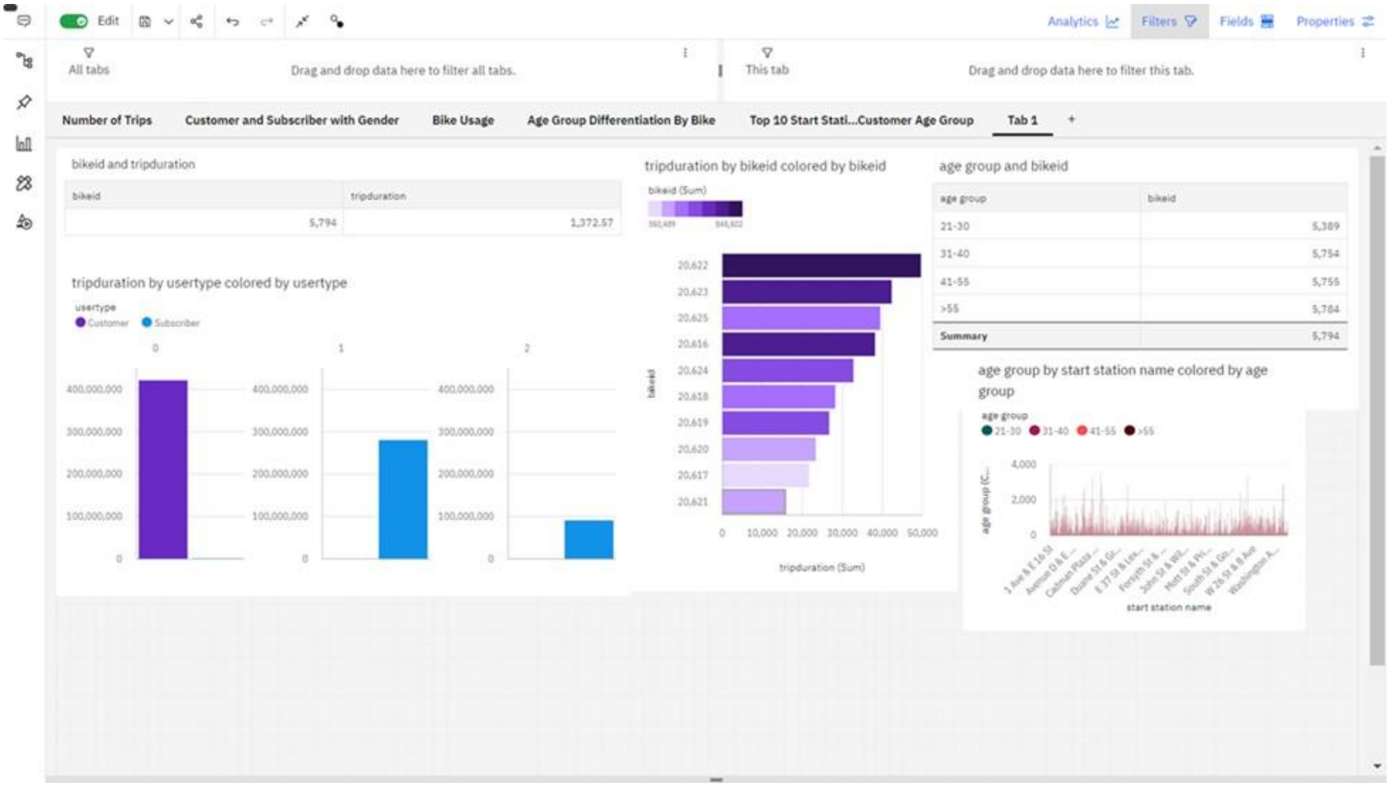



Count
800
600
400
200
0
Start Station Name


Count
800
600
400
200
0
End Station Name


Hour usage of Citi Bikes
Number of people using bike
60k
50k
40k
30k
20k
10k
0
−20
−15
−10
−5
0
5
10
15
Time

8.CREATING THE DASHBOARD



bikeid and tripduration

bikeid	tripduration
5,794	1,372.57

age group and bikeid

age group	bikeid
21-30	5,389
31-40	5,754
41-55	5,755
>55	5,784
Summary	5,794

9. ADVANTAGES AND DISADVANTAGES

The benefits of bike sharing schemes include transport flexibility, reductions to vehicle emissions, health benefits, reduced congestion and fuel consumption, and financial savings for individuals.

One can easily analyze and understand trends in bike sharing patterns with the created dashboard. With no prior skills and knowledge about the tools that we use for analysis, anyone (literate or illiterate) can easily infer the knowledge that we represent in various charts or graphs or maps. So that it would be helpful to users and companies to make appropriate decisions in the future.

10. CONCLUSION

Based on the quantitative as well as visual analysis of the New York bike share system, a number of interesting insights were gained.

One obvious conclusion was that there is a strong seasonal variation in the system usage with maximum usage in summer and minimum usage in winter. This was initially hypothesized because of the harshness of New York's harsh winters and the treacherous riding conditions that exist during that time. However, despite the adverse weather conditions, there is a strong core demographic that consistently uses the system. This conclusion is based on that fact that even during the months of January and February which are the peak winter months, there are more than two hundred thousand trips in the system.

New York has a strong public transit system, and the bike share system seems to complement it quite well with a majority of the highest used stations located either close to subway lines or the commuter rail stations in the city.

Based on the locations of the stations and the duration of trips, it can be hypothesized that bike shares are replacing last mile trips that would otherwise be done either on foot or on public transit. This is particularly true in case of New York where a combination of dense public transit network, the road congestion during peak hours and the average trip distance as calculated create a situation where the only potential trips that the bike share system is replacing currently are those that would otherwise have been undertaken either on foot or on public bus.

11.FUTURE SCOPE

NYC is a very crowded and happening place which leads to lots of pollution. And in this busy world people are always worried about transportation this bike sharing system reduces that stress. With increase in population pollution also increases. So it is in our hands to reduce pollution and to make a better future for our younger generations. We can analyze which station needs more bikes and any area needs new station to be installed. The survey outcomes indicates the needs for improved techniques in bike sharing analytics. There exists a lot of scope in this research area.

12.SOURCE CODE

```
#%% md

# SPRINT **3**

#%%

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()

#%%

df[df['starttime'].isnull()]

#%%

df[df['stoptime'].isnull()]

#%%

df = df[:-1]

#%%

df.isnull().sum()
```

```
#%%

print(type(df["start station latitude"][0]))
print(df["start station latitude"][0])

#%%

df['start station name'].unique()

#%%

def camel_case(city):
   try:
      city = city.split(' ')
      city = ' '.join([x.lower().capitalize() for x in city])
      if city == 'Unknown':
            return np.nan
      else:
            return city
   except:
      return np.nan

# Apply camel case function to City column
df['start station name'] = df['start station name'].apply(camel_case)
df['start station name'].value_counts()

#%%

df.count()

#%%

df["tripduration"] = pd.to_numeric(df["tripduration"])
res = df.iloc[52323]
print(res["tripduration"])

#%%

df_filtered = df[df['tripduration'] != "tripduration"]
df_filtered["tripduration"] = pd.to_numeric(df_filtered["tripduration"])

df = df_filtered
type(df["tripduration"][0])

#%%

type(df["start station latitude"][0])

#%%

type(df["end station longitude"][0])

#%%

type(df["bikeid"][0])

#%%

type(df["birth year"][0])

#%%
```

```
type(df["gender"][0])

#%%

type(df["starttime"][0])

#%%

df["starttime"] = pd.to_datetime(df["starttime"])
df["stoptime"] = pd.to_datetime(df["stoptime"])
type(df["starttime"][0])

#%%

df["starttime"][0] <df["stoptime"][0]

#%%

df.info()

#%%

def find_outliers_IQR(df):
  q1=df.quantile(0.25)
  q3=df.quantile(0.75)
  IQR=q3-q1
  outliers = df[((df<(q1-1.5*IQR)) | (df>(q3+1.5*IQR)))]
  return outliers
outliers = find_outliers_IQR(df["birth year"])
print("number of outliers: " + str(len(outliers)))
print("max outlier value: " + str(outliers.max()))
print("min outlier value: " + str(outliers.min()))

#%%

df["gender"].value_counts()

#%%

temp_df = df[df["birth year"] <= 1957]
temp_df["gender"].value_counts()

#%%

df.shape

#%%

df.to_csv('cleaned_dataset.csv', index=False)

#%% md

# **SPRINT 4**

#%%

path = "/content/cleaned_dataset.csv"
edadf = pd.read_csv(path)
print(edadf)
```

```
#%%

temp = edadf

#%%

temp.head()

#%%

temp.describe()

#%%

temp.info()

#%%

temp["starttime"] = pd.to_datetime(temp["starttime"])
temp["stoptime"] = pd.to_datetime(temp["stoptime"])
temp.info()
temp["Hour"] = temp["stoptime"].dt.hour - temp["starttime"].dt.hour
temp.head()

#%%

temp.shape

#%%

temp['Age'] = 2022 - temp['birth year']
temp.head()

#%%

Age_Groups = ["<20", "20-29", "30-39", "40-49", "50-59", "60+"]
Age_Groups_Limits = [0, 20, 30, 40, 50, 60, np.inf]
Age_Min = 0
Age_Max = 100
temp["Age_group"] = pd.cut(temp["Age"], Age_Groups_Limits, labels=Age_Groups)
temp.head()

#%%

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"})
```

13.GITHUB LINK

https://github.com/IBM-EPBL/IBM-Project-34851660278076

14.OUTPUT DEMO LINK

https://drive.google.com/file/d/1wSmFrMOOsu5Mfzgo1a09MPjxr6DUHvHE/view?usp=share_link