A NEW HINT TO TRANSPORTATION

ANALYSIS OF THE NYC BIKE SHARE SYSTEM

NALAIYA THIRAN PROJECT BASED LEARNING

ON

PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

A PROJECT REPORT

N.LAVANYA	*410119106028*
R KOWSHIKA	*410119106026*
K.MAHESHWARI	*410119106029*
R.NATHASHA	*410119106040*

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

ADHI COLLEGE OF ENGINEERING AND TECHNOLOGY

Approved by AICT, New Delhi, Permanent affiliation status
ANNA UNIVERSITY, Chennai:
Accredited By NAAC, New Delhi:Recognized
U/S12(B)&2(F) of UGC Act 1956
An ISO Certified Institution
Munu Adhi Nagar, Sankarapuram, Near Wallajabad

KANCHIPURAM-631 605

NOVEMBER 2022

INTERNAL MENTOR:

Dr.K.DINESH BABU

Associate professor

Department of Electronics and Communication Engineering

Adhi College of Enginnering and Technology,

Kanchipuram-631 60

INDUSTRY MENTOR

HARI PRABU

IBM

ABSTRACT

Like all other sharing systems, Airbnb the housing sharing system, Uber the car sharing system, Citi Bike is the network of bicycle rental stations intended for point-to-point transportation.Data shows Citi Bike is NewYork City's largest bike sharing system. It's a convenient solution for trips that are too far to walk but too short for a taxi or the subway. Thebike sharing system is combined with all other transportation methods available in the area for commuters. So in order to help users better understand data used to find out the best stations for renting bikes, and thebike stations that have the best service available for users, there is a needto visualise the data to come up with proper conclusions.

The Process would be :
1. Collecting data regarding user's who have used the bike sharing system.
2. Visualising the data collected. 3. Understanding the data, in order to come up with conclusions regarding the bike system.

TABLE OF CONTENTS

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

1.INTRODUCTION

74% of New Yorkers support bike share (August 2012 Quinnipiac poll) Janette Sadik-Khan - Former commissioner of the New York City Department of Transportation (2007–2013).
In just the last five years, New York City has made huge strides in creating modern, safer streets. Drawing from Mayor Michael Bloomberg's PlaNYC sustainability agenda, we've established more than 300 miles of bike lanes, 30 plazas and made expansive street safety redesigns to accommodate all street users citywide— all while recording the five safest years in city history and logging remarkable economic gains in corridors where projects were implemented.

Citi Bike presents a new way for New Yorkers to get around that takes advantage of these changes to our streets, and it also marks a new standard for publicparticipation in planning.

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.

2.OBJECTIVE

The trip purpose influences the probability of using the bicycle (Marleau, Larsen, & Geneidy, 2011). For instance, the probability of using a bicycle for leisure trips is greater than for shopping purposes, because it can be difficult carrying shopping bags on a bicycle (Mcneil, 2011)(PROBICI team, 2010).

The bike sharing demand is also affected by the trip purpose, as referred, and there are three typical user groups: commuters, recreational/errand riders and tourists. Thus to each one it must be considered different initial rates of bicycle trips (Rn) per purpose (n) based on other study cases.

For short distances (between 2 and 8 minutes) in urban areas, bicycle can be the most efficient transportation mode. However, while travel distance increases the competitiveness will be negatively affected, and Inês Frade and Anabela Ribeiro / Procedia - Social and Behavioral Sciences 111 (2014) 518 – 527 523 consequently the potential demand of this mode will decrease:

In other words, the potential demand is affected by an elasticity which causes a fall in the percentage of bicycle trips when the distance travelled increases.

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.

3.IDEATION PHASE

3.1 LITERATURE SURVEY:

Bike-share programs represent a unique opportunity for the City ofNew York to re-envision transportation within the urban sphere. As a transportation system, bike-shares are ideally designed for densely populated cities like New York. Distances between many major destinations are small and almost 50% of New York's workforce lives within a reasonable bicycling distance (less than 5 miles) of their place of work. Importantly, bike-shares offer immediate transportation solutions as they can be built, installed and open for business in months rather than years. Bike-share programs offer options for economic growth and job creation, as well as providing considerable health benefits.

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 as "rebalancing" problem as part of "fleet optimization" questions.

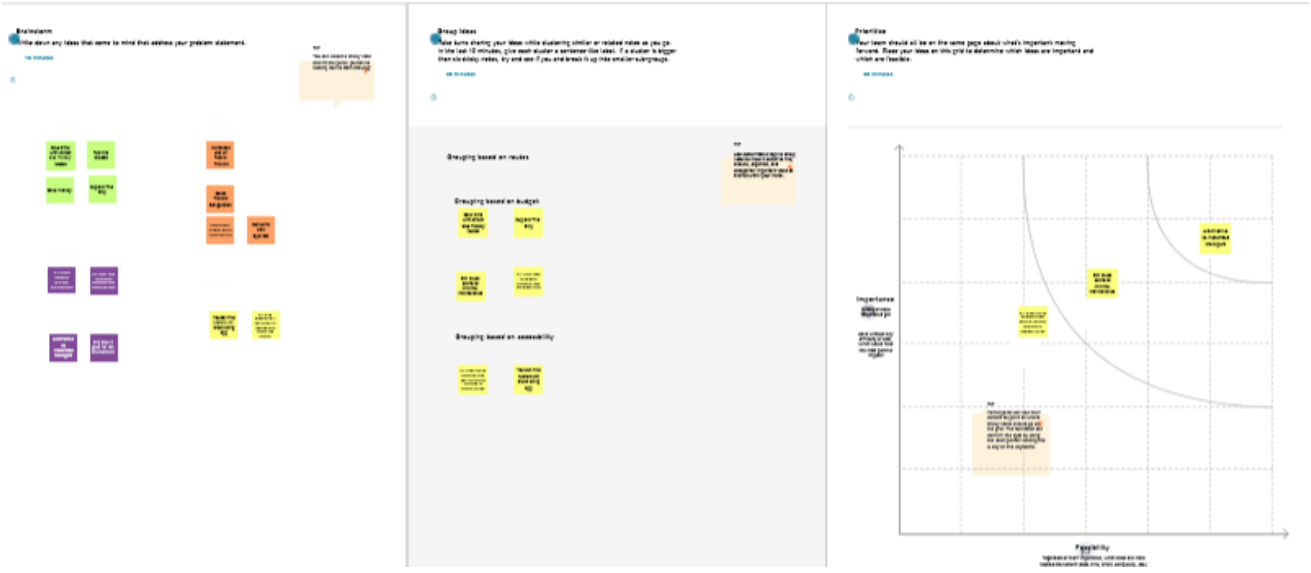
Hubway Bikeshare (Boston, MA) started to pilot programs of subsidized memberships while implementing stations in low-revenue areas in order to increase access and equity of ridership Cities stand to gain $2.6 billion annually in indirect savings based on lower road construction costs, reduced accidents, and lower carbon dioxide emission.

3.2 Empathy Map:

3.3 Ideation:

Who does the Problem Affect?	**This will affect the people who suffers to buy diesel for vehicle and mainly affects the poor people.**
What is the issue	**The main issue is the unavailability of the public transportation at the desired timings, In addition with the environmental factors like pollution, resource extinction etc.**
When does the issue occur	**It occurs due to over-exploiting of Natural resources and making it extinct.**
Where is the issue coming	**Fluctuations of fuel rates and hazard to the environment.**
Why is it important that we fix the problem	**It is important that we fix the problem to reduce environmental effects.**

3.4 Brainstroming:



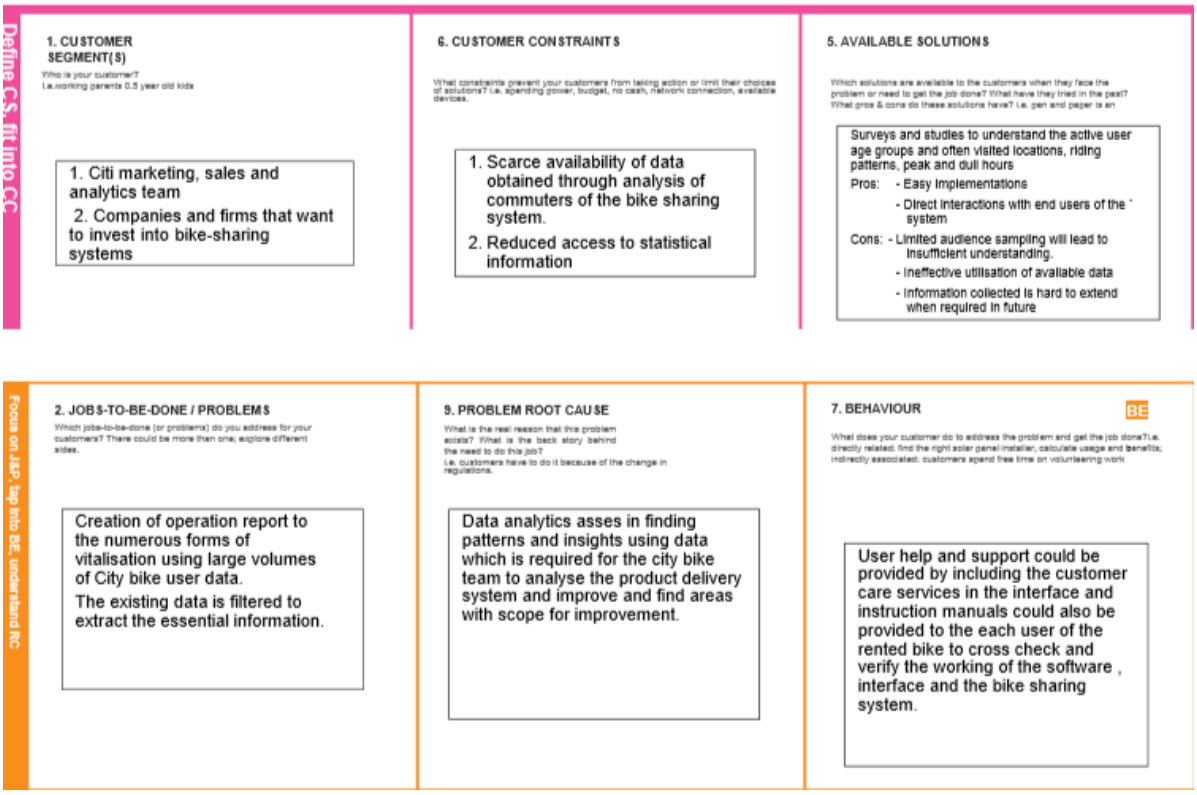
4.PROJECT DESIGN PHASE 1

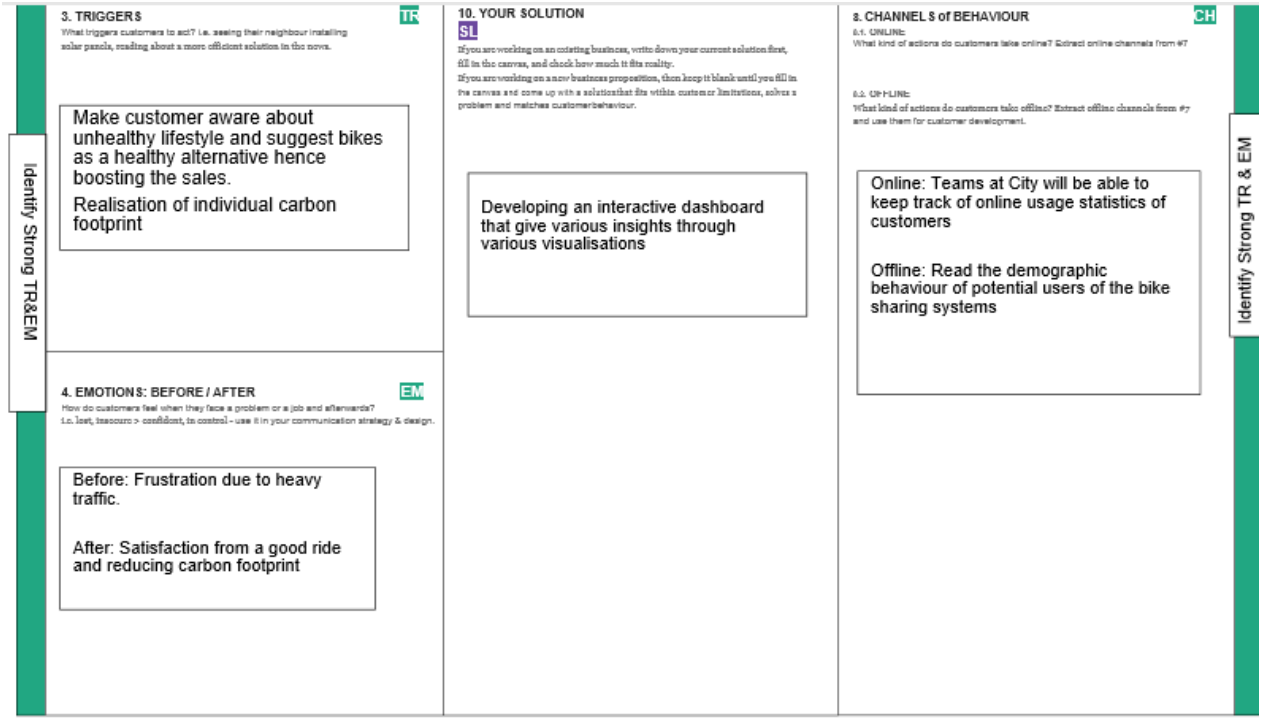
4.1 Proposed Solution:

S.No	***Parameter***	***Description***
1	Problem statement(Problem to be solved)	1. Bike share programs have risen in popularity in recent years and have been promoted as a lower carbon alternative to other forms of transit. 2. The premise of bicycle sharing is that itis a short-term bike rental system, based on varying timed memberships 3. The trips would not have been made otherwise or are substituting walking and private bicycle trips .
2	Idea / Solution description	1. The planning process for the Citi Bike program established an open door policy, encouraging input early and often from the citizens of New York City. 2. Low-income people are less likely than middle- and upper-income people to have a credit card. 3. Sites should ensure maximum visibility and access. 4. Sites must not impede the use of any existing facilities, such as bus stops or fire hydrants.
3	Novelty / Uniqueness	1. Transport flexibility 2. Reductions to vehicle emissions 3. Health benefits 4. Reduced congestion and fuel consumption 5. Financial savings for individuals.
4	Social Impact / Customer Satisfaction	1. Transportation

		2. Recreation of cycling 3. Enjoyable sport 4. Low cost
5	Business Model (Revenue Model)	1. Zero deaths since it launched. 2. Lack of public subsidies 3. Battery-powered bikes 4. The model is trained using open
6	Scalability of the Solution	1. Improved customer engagement 2. Reduce customer acquisition cost 3. Economical development 4. Immediate response for customer queries

4.2 Problem Solution Fit:





4.3 Solution Architecture:

Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

• Find the best tech solution to solve existing business problems.

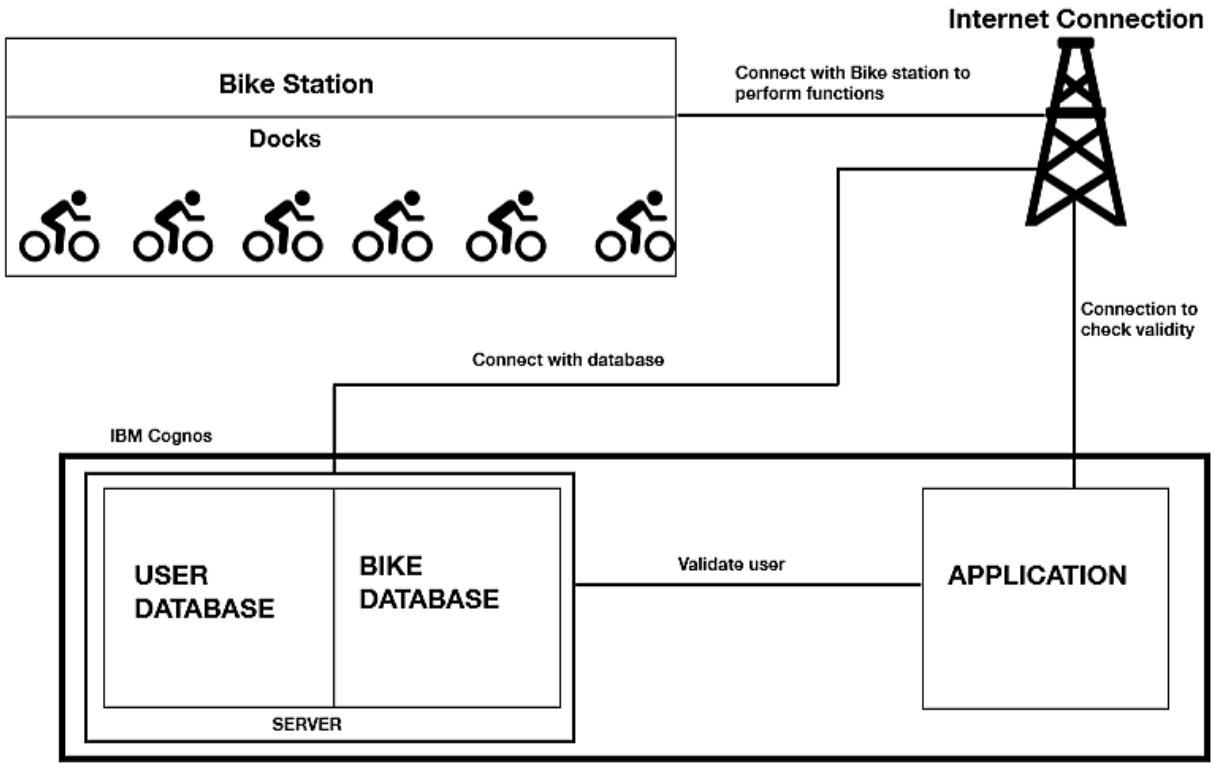
• Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.

• Define features, development phases, and solution requirements.

• Provide specifications according to which the solution is defined, managed, and delivered.

Example-Solution Architecture Diagram:

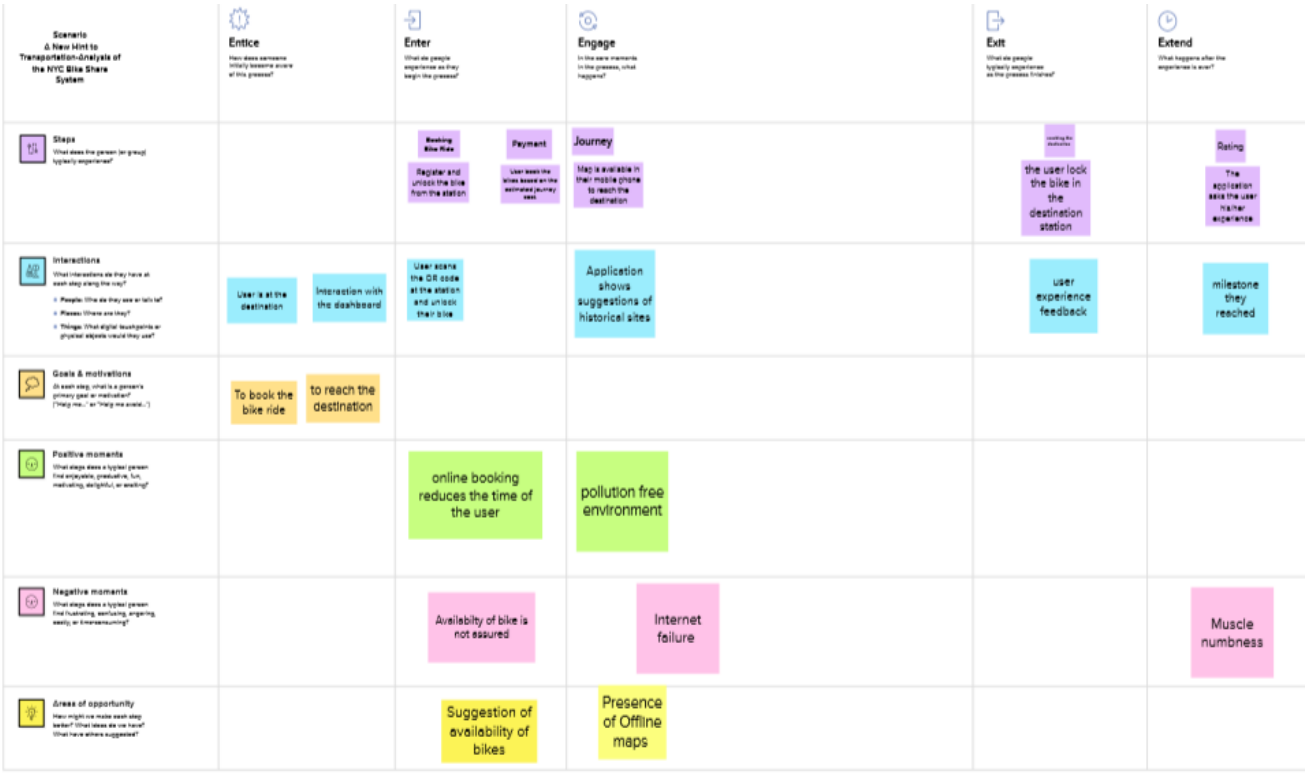
A New Hint To Transportation-Analysis Of The NYC Bike Share System



Team ID: PNT2022TMID53270

5.PROJECT DESIGN PHASE 2

5.1 Customer Journey Map:



5.2 Solution Requirement:

Functional Requirements:

Following are the functional requirements 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.

Non-funtional Requirement:

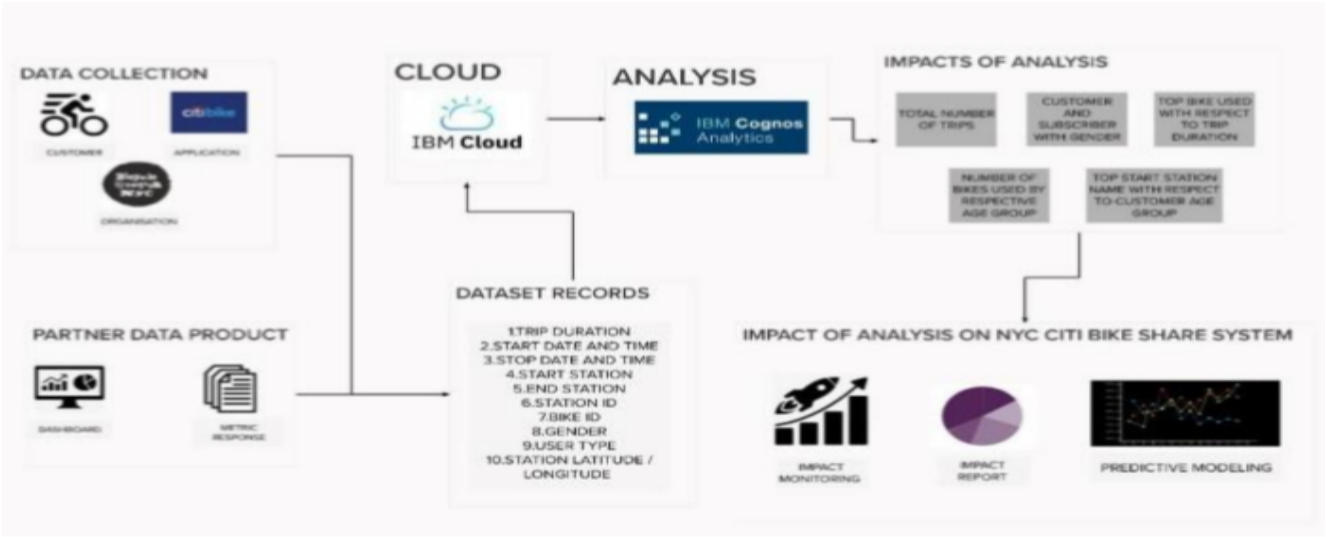
Following are the non-functional requirement of the proposed solution.

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	***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-3	***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-4		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

	Availability of bikes	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-5	***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.

5.3 Data Flow Diagrams:

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



lser Stories

lse 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

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	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
Analysts	Analysing the user data	USN-2	This data is used as input for creating various types of visualizations and analysis is	I can view the analysis of the citi bike	High	Sprint-2

			done and a dashboard is created			
Organisation and 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 and access the dashboard with login	High	Sprint-3

5.4 Technology Stack:

Technical Architecture:

The Deliverable shall include the architectural diagram as below:
Table-1:Components &Technologies:

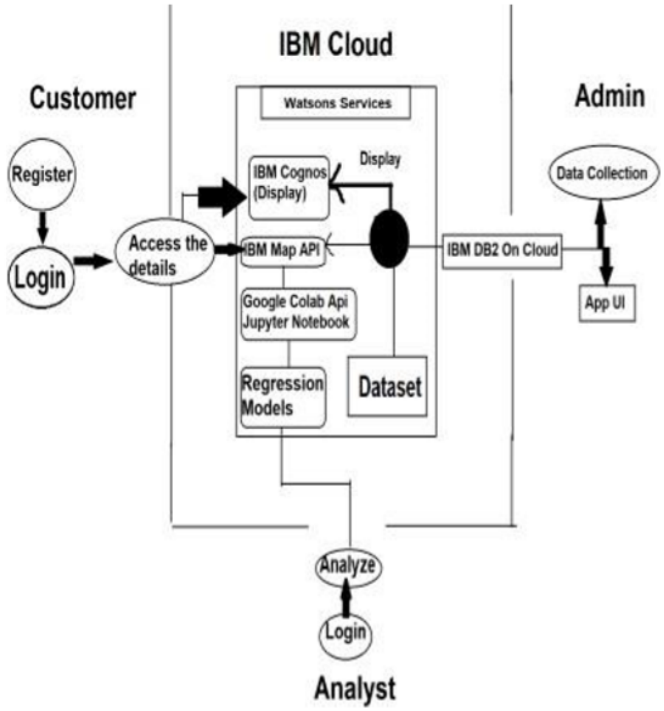


Table-1:Components &Technologies:

S.No	***Component***	***Description***	***Technology***
	User interface	1.Display this visualization of the analysed data 2.Display the interferences from the analysed data	HTML,CSS,Java Script and IBM Cognos
	Application logic-1	display details	HTML
	Database	Data Type,Configurations etc.	MySQL
	Cloud Database	Data base service on cloud	IBM DB2,IBM Cloudant etc.
	File Storage	File stroge requirements	IBM Block Stroge or Other Storage Service or Local Filesystem
	External API-1	To map the Citibike ride in NYC	IBM map API,etc.
	External API-2	Analysis of the data	Google Cloab,Jupyter Notebook
	Machine Learning Model	To plot the graph and predict values	Regression models

	Infrastructure (Server/Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Local Syst Cloud Server Configuration : IBM Cloud	Local,Cloud Foundary,Kubernets,etc.

6.PROJECT PLANNING PHASE

6.1 Prepare Milestone and Activity list:

Activity Name	Activity Number	Activity Description	Tasks Assigned	Status
Preparation Phase	1	a) Access the resources in projectdashboard b) Explore the dataset provided inworkspace c) Create GitHub account & collaborate with Project Repository inproject workspace d) Set-up the prerequisites for theproject	Mageshwari Natasha Kowshika Lavanya	Completed
Ideation Phase	2	a) Literature survey relevant to theselected project b) Preparation of Empathy Map toidentify the user pros and cons c) List the ideas by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility& importance	Mageshwari Lavanya Mageshwari	Completed
Project Design Phase-	3			

I				
Problem Solution Fit	3.2	Prepared problem solution fit which provides effective solutions for the problem	Natasha	Completed
Solution Architecture	3.3	Develop effective architecture for the proposed solution	Lavanya Kowshika	Completed
Project Design Phase-II	4			
Requirement Analysis	4.1	Identify the Functional and Non-Functional requirements	Kowshika	Completed
Customer Journey	4.2	Preparation of customer journey map to understand the user interactions & experiences with the application from the entry level to exit level	Mageshwari	Completed
Data Flow Diagram and User stories	4.3	Generate Data flow diagram of the project	Natasha	Completed
Technical Architecture	4.4	Develop effective technical architecture for the proposed solution	Lavanya	Completed
Project Planning Phase	5			
Milestones & Activity List	5.1	Prepare Milestone and Activity list of the project	Mageshwari	Completed
Sprint Plan	5.2	Prepare Sprint Delivery plan of the project	Lavanya	Completed
Project Development	6			

Phase				
Delivery of Sprint-1	6.1	Implement the coding phase of Sprint-1	Kowshika	In Progress
Delivery of Sprint-2	6.2	Implement the coding phase of Sprint-2	Natasha	In Progress
Delivery of Sprint-3	6.3	Implement the coding phase of Sprint-3	Lavanya	In Progress
Delivery of Sprint-4	6.4	Implement the coding phase of Sprint-4	Natasha	In Progress

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.

Milestone Timeline Chart

6.2 Sprint Delivery Plan:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
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
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	2	High
Sprint-1		USN-3	As a user, I can register for the application through Gmail	2	Medium
Sprint-2	Login	USN-4	As a user, I can log into the application by entering email & password	2	High

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
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
Sprint-2		USN-6	I can use the citi bike share system data for analysis purposes	5	High
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
Sprint-3	Dashboard	USN-8	I can register & access the dashboard created based on the analysis by logging in	3	Medium
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
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
Sprint-4		USN-11	As a user I can view the dashboard that displays the customer and subscriber with respect to gender	5	High
Sprint-4		USN-12	As a user I can view the dashboard that displays the total number of trips	5	High

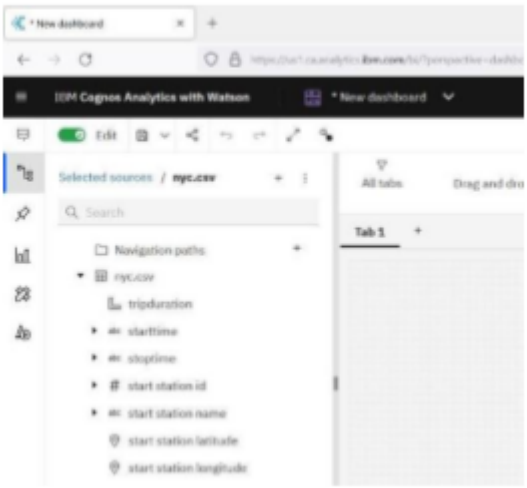
Project Tracker, Velocity & Burndown Chart:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)
Sprint-1	6	6 Days	24 Oct 2022	03 nov 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

7.PROJECT DEVELOPMENT PHASE

7.1 Project Development - Delivery of Sprint-1:

Uploading the dataset



Data cleaning:

Finding the duplicates:

```
df = data[data.birthyear.notnull()]
df.duplicated().sum()
    0 # There are no null values present in birthyear
```

Checking null values:

d=df.isna().sum()

d.head()

```
tripduration            0
starttime               0
stoptime                0
start station id        47
start station name      47
dtype: int64
```

Replace null values:

df.replace(np.nan,'-',inplace = True)

df.isnull().sum()

```
tripduration               0
starttime                  0
stoptime                   0
start station id           0
start station name         0
start station latitude     0
start station longitude    0
end station id             0
end station name           0
end station latitude       0
end station longitude      0
bikeid                     0
usertype                   0
birthyear                  0
gender                     0
tripduration_bins          0
Age                        0
dtype: int64
```

Feature Engineering:

calculating Age from birth year

```
from datetime import datetime, date
age=2018-df['birth_year']
df['Age']=age
df.head()
```

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid	usertype	birth_year	gender	tripduration_bins	Age
0	11.583333	2013-06-01 00:00:01	2013-06-01 00:11:36	444	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	19679	Subscriber	1983.0	1	(0.0, 30.0]	35.0
1	11.550000	2013-06-01 00:00:08	2013-06-01 00:11:41	444	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	16649	Subscriber	1984.0	1	(0.0, 30.0]	34.0
3	2.050000	2013-06-01 00:01:04	2013-06-01 00:03:07	475	E 15 St & Irving Pl	40.735243	-73.987586	262.0	Washington Park	40.691782	-73.973730	16352	Subscriber	1960.0	1	(0.0, 30.0]	58.0
4	25.350000	2013-06-01 00:01:22	2013-06-01 00:26:43	2008	Little West St & 1 Pl	40.705693	-74.016777	310.0	State St & Smith St	40.689269	-73.989129	15567	Subscriber	1983.0	1	(0.0, 30.0]	35.0
6	34.283333	2013-06-01 00:02:33	2013-06-01 00:36:50	285	Broadway & E 14 St	40.734546	-73.990741	532.0	S 5 Pl & S 5 St	40.710451	-73.960876	15693	Subscriber	1991.0	1	(30.0, 60.0]	27.0

calculating age group from age

```
max_limit = df['Age'].max()
max_limit
bins = [0,20,40,60,max_limit]
agegroup = pd.cut(df['Age'], bins=bins).value_counts()
Agegroup
```

```
(20.0, 40.0]     161563
(40.0, 60.0]     148805
(60.0, 119.0]     27014
(0.0, 20.0]           0
Name: Age, dtype: int64
```

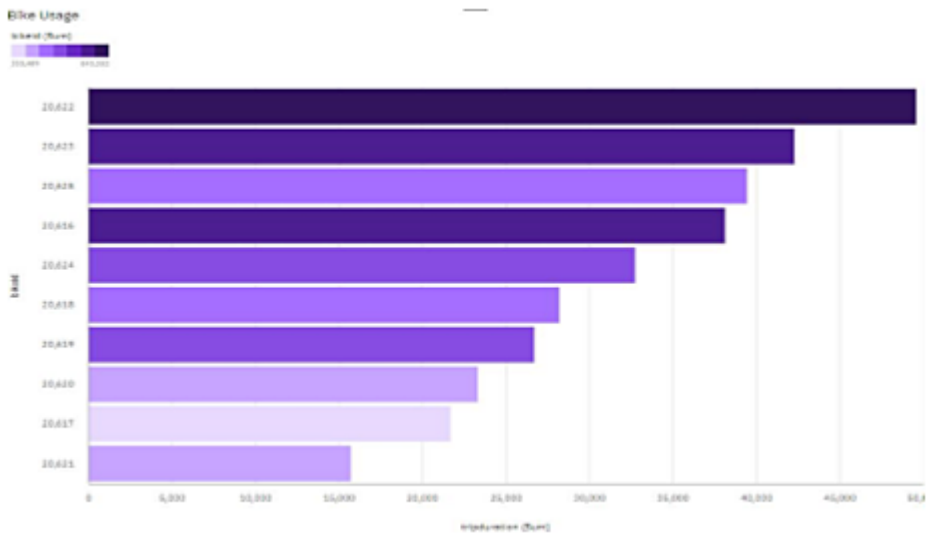
Calculating Hour

```
peak_hour['Start Date'] = pd.to_datetime(df['starttime'])
peak_hour['Stop Date'] = pd.to_datetime(df['stoptime'])
peak_hour['year'] =peak_hour["Start Date"].dt.year
peak_hour["Hour"] = peak_hour["Start Date"].dt.hour
```

	Start Date	Stop Date	year	Hour	bikeid
0	2013-06-01 00:00:01	2013-06-01 00:11:36	2013	0	19678
1	2013-06-01 00:00:08	2013-06-01 00:11:41	2013	0	16649
3	2013-06-01 00:01:04	2013-06-01 00:03:07	2013	0	16352
4	2013-06-01 00:01:22	2013-06-01 00:26:43	2013	0	15567
6	2013-06-01 00:02:33	2013-06-01 00:36:50	2013	0	15693
...	...	...	...	...	...
577687	2013-06-30 23:58:09	2013-07-01 00:05:25	2013	23	19454
577689	2013-06-30 23:57:52	2013-07-01 00:00:57	2013	23	16746
577690	2013-06-30 23:58:39	2013-07-01 00:08:34	2013	23	19290
577698	2013-06-30 23:59:27	2013-07-01 00:14:52	2013	23	15250
577700	2013-06-30 23:59:33	2013-07-01 00:02:14	2013	23	18910

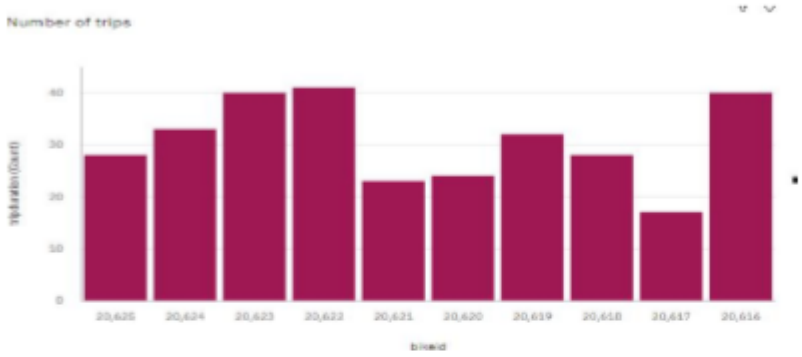
337382 rows × 5 columns

Bike Usage -Bike Id Vs Trip Duration:

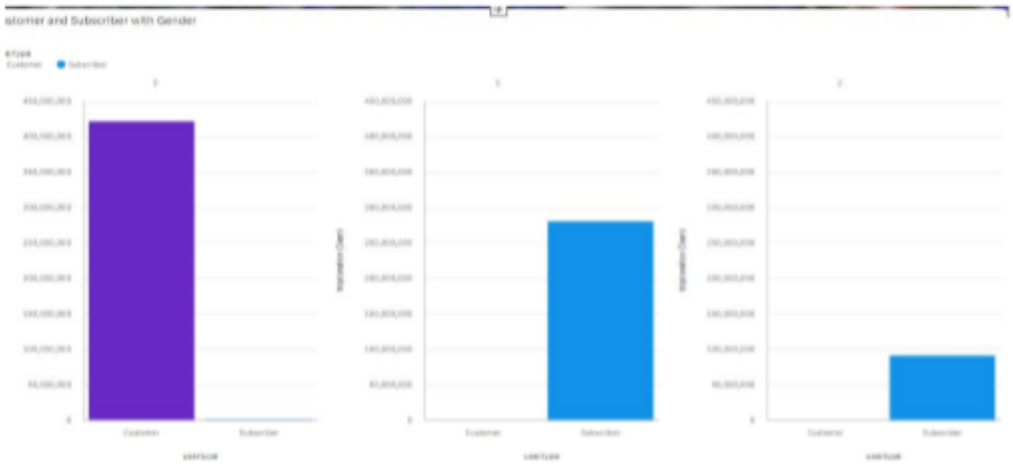


Visualization of the dataset in COGNOS Platform:

Finding the number of trips per each bike:



Finding the percentage of customers and subscribers



Age Group Differentiation by BikeId:

Calculation:

if(age<=20) then

('<20')

else if(age>=21 and age<=30) then

('21-30')

else if(age>=31 and age<=40) then

('31-40')

else if(age>=41 and age<=55) then

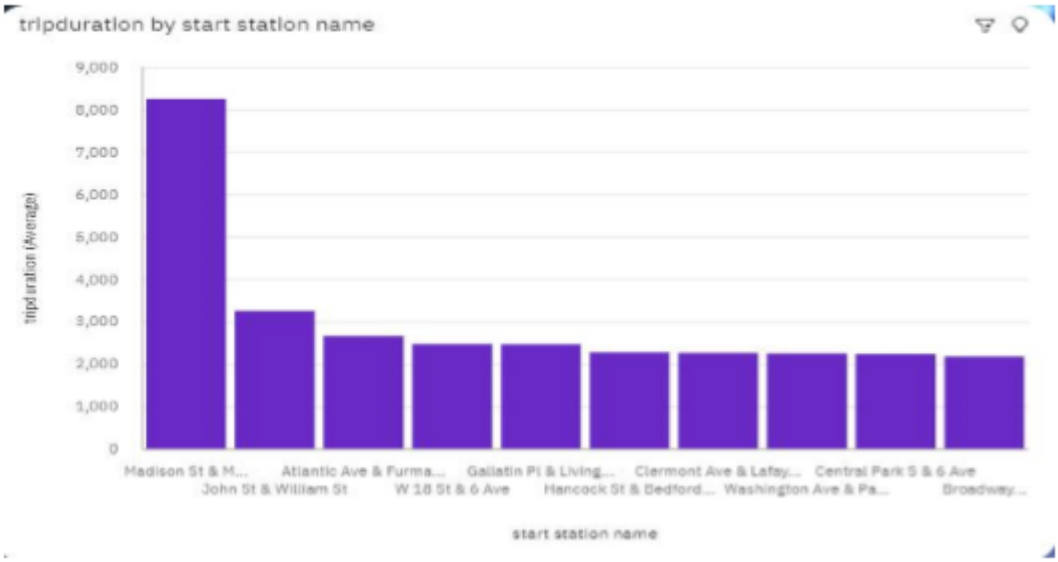
('41-55')

else('>55')

bikeid and Age_Group

Age_Group	bikeid
21-30	5,721
31-40	5,749
41=55	5,741
<20	1,625
>55	5,781
Summary	5,794

Finding the top 10 start stations with customer age group:



7.3 Project Development-Delivery of 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)
      tripduration          starttime           stoptime  \
0              695  2013-06-01 00:00:01  2013-06-01 00:11:36
1              693  2013-06-01 00:00:08  2013-06-01 00:11:41
2             2059  2013-06-01 00:00:44  2013-06-01 00:35:03
2             2059  2013-06-01 00:00:44  2013-06-01 00:35:03
3              123  2013-06-01 00:01:04  2013-06-01 00:03:07
4             1521  2013-06-01 00:01:22  2013-06-01 00:26:43
...            ...                  ...                  ...
84087         1478  2013-06-08 14:35:22  2013-06-08 15:00:00
84088          873  2013-06-08 14:35:31  2013-06-08 14:50:04
```

```
84089         2054  2013-06-08 14:34:51  2013-06-08 15:09:05
84090         1179  2013-06-08 14:35:28  2013-06-08 14:55:07
84091          804  2013-06-08 14:35:33                  NaN

      start station id       start station name  start station latitude  \
0                444.0      Broadway & W 24 St             40.742354
1                444.0      …
start station longitude  end station id          end station name  \
0               -73.989151         434.0          9 Ave & W 18 St
1               -73.989151         434.0          9 Ave & W 18 St
2               -73.995951         406.0     Hicks St & Montague St
3               -73.987586         262.0          Washington Park
4               -74.016777         310.0       State St & Smith St
...                    ...         ...                     ...
84087           -74.008387         342.0  Columbia St & Rivington St
84088           -73.985162         404.0           9 Ave & W 14 St
84089           -73.993790        2012.0           E 27 St & 1 Ave
84090           -74.012342         383.0  Greenwich Ave & Charles St
84091                  NaN           NaN                     NaN

      end station latitude  end station longitude   bikeid    usertype  \
0               40.743174          -74.003664  19678.0  Subscriber
1               40.743174          -74.003664  16649.0  Subscriber
2               40.695128          -73.995951  19599.0    Customer
3               40.691782          -73.973730  16352.0  Subscriber
4               40.689269          -73.989129  15567.0  Subscriber
...                   ...                ...      ...         ...
84087           40.717400          -73.980166  19730.0    Customer
84088           40.740583          -74.005509  15606.0    Customer
84089           40.739445          -73.976806  18597.0  Subscriber
84090           40.735238          -74.000271  14665.0  Subscriber
84091                  NaN                NaN      NaN         NaN


birth year  gender
0        1983.0     1.0
1        1984.0     1.0
2           NaN     0.0
3        1960.0     1.0
4        1983.0     1.0
...         ...     ...
84087       NaN     0.0
84088       NaN     0.0
84089    1977.0     1.0
84090    1968.0     2.0
84091       NaN     NaN

[84092 rows x 15 columns]
```

df.head()

OUTPUT:

```
In [6]: df.head()
```

Out[6]:

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid
0	695	2013-06-01 00:00:01	2013-06-01 00:11:36	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	19678.0
1	693	2013-06-01 00:00:08	2013-06-01 00:11:41	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	16649.0
2	2059	2013-06-01 00:00:44	2013-06-01 00:35:03	406.0	Hicks St & Montague St	40.695128	-73.995951	406.0	Hicks St & Montague St	40.695128	-73.995951	19599.0
3	123	2013-06-01 00:01:04	2013-06-01 00:03:07	475.0	E 15 St & Irving Pl	40.735243	-73.987586	262.0	Washington Park	40.691782	-73.973730	16352.0
4	1521	2013-06-01 00:01:22	2013-06-01 00:26:43	2008.0	Little West St & 1 Pl	40.705693	-74.016777	310.0	State St & Smith St	40.689269	-73.989129	15567.0

7.4 Project Development-Delivery of sprint-4:

```
path = "/content/cleaned_dataset.csv"
edadf = pd.read_csv(path)
print(edadf)
      tripduration          starttime           stoptime  \
0              695  2013-06-01 00:00:01  2013-06-01 00:11:36
1              693  2013-06-01 00:00:08  2013-06-01 00:11:41
2             2059  2013-06-01 00:00:44  2013-06-01 00:35:03
3              123  2013-06-01 00:01:04  2013-06-01 00:03:07
4             1521  2013-06-01 00:01:22  2013-06-01 00:26:43
...            ...                  ...                  ...
84086         1506  2013-06-08 14:35:22  2013-06-08 15:00:28
84087         1478  2013-06-08 14:35:22  2013-06-08 15:00:00
84088          873  2013-06-08 14:35:31  2013-06-08 14:50:04
84089         2054  2013-06-08 14:34:51  2013-06-08 15:09:05
84090         1179  2013-06-08 14:35:28  2013-06-08 14:55:07

      start station id      start station name  start station latitude  \
0                444.0      Broadway & W 24 St               40.742354
1                444.0      Broadway & W 24 St               40.742354
```

```
2             406.0   Hicks St & Montague St           40.695128
3             475.0      E 15 St & Irving Pl           40.735243
4            2008.0    Little West St & 1 Pl           40.705693
...             ...                 ...                ...
84086         422.0        W 59 St & 10 Ave            40.770513
84087         337.0     Old Slip & Front St            40.703799
84088         447.0         8 Ave & W 52 St            40.763707
84089         229.0          Great Jones St            40.727434
84090         259.0  South St & Whitehall St            40.701221

      start station longitude  end station id          end station name  \
0              -73.989151         434.0           9 Ave & W 18 St
1              -73.989151         434.0           9 Ave & W 18 St
2              -73.995951         406.0      Hicks St & Montague St
3              -73.987586         262.0           Washington Park
4              -74.016777         310.0        State St & Smith St
...                   ...         ...                      ...
84086          -73.988038         212.0      W 16 St & The High Line
84087          -74.008387         342.0  Columbia St & Rivington St
84088          -73.985162         404.0           9 Ave & W 14 St
84089          -73.993790        2012.0           E 27 St & 1 Ave
84090          -74.012342         383.0  Greenwich Ave & Charles St

      end station latitude  end station longitude   bikeid    usertype  \
0             40.743174          -74.003664  19678.0  Subscriber
1             40.743174          -74.003664  16649.0  Subscriber
2             40.695128          -73.995951  19599.0    Customer
3             40.691782          -73.973730  16352.0  Subscriber
4             40.689269          -73.989129  15567.0  Subscriber
...                 ...                ...     ...        ...
84086         40.743349          -74.006818  19225.0    Customer
84087         40.717400          -73.980166  19730.0    Customer
84088         40.740583          -74.005509  15606.0    Customer
84089          -73.993790        2012.0           E 27 St & 1 Ave
84090          -74.012342         383.0  Greenwich Ave & Charles St

      end station latitude  end station longitude   bikeid    usertype  \
0             40.743174          -74.003664  19678.0  Subscriber
1             40.743174          -74.003664  16649.0  Subscriber
2             40.695128          -73.995951  19599.0    Customer
3             40.691782          -73.973730  16352.0  Subscriber
4             40.689269          -73.989129  15567.0  Subscriber
...                 ...                ...     ...        ...
84086         40.743349          -74.006818  19225.0    Customer
84087         40.717400          -73.980166  19730.0    Customer
84088         40.740583          -74.005509  15606.0    Customer
84089         40.739445          -73.976806  18597.0  Subscriber
84090         40.735238          -74.000271  14665.0  Subscriber
```

	birth year	gender
0	1983.0	1.0
1	1984.0	1.0
2	NaN	0.0
3	1960.0	1.0
4	1983.0	1.0
...	...	...
84086	NaN	0.0
84087	NaN	0.0
84088	NaN	0.0
84089	1977.0	1.0
84090	1968.0	2.0

OUTPUT:

In [43]:
```
temp = edadf
```

In [44]:
```
temp.head()
```

Out[44]:

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid	usertype	b
0	695	2013-06-01 00:00:01	2013-06-01 00:11:36	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	19678.0	Subscriber	19
1	693	2013-06-01 00:00:08	2013-06-01 00:11:41	444.0	Broadway & W 24 St	40.742354	-73.989151	434.0	9 Ave & W 18 St	40.743174	-74.003664	16649.0	Subscriber	19
2	2059	2013-06-01 00:00:44	2013-06-01 00:35:03	406.0	Hicks St & Montague St	40.695128	-73.995951	406.0	Hicks St & Montague St	40.695128	-73.995951	19599.0	Customer	
3	123	2013-06-01 00:01:04	2013-06-01 00:03:07	475.0	E 15 St & Irving Pl	40.735243	-73.987586	262.0	Washington Park	40.691782	-73.973730	16352.0	Subscriber	19
4	1521	2013-06-01 00:01:22	2013-06-01 00:26:43	2008.0	Little West St & 1 Pl	40.705693	-74.016777	310.0	State St & Smith St	40.689269	-73.989129	15567.0	Subscriber	19

8.CONCLUSION:

Based on the quantitative as well as visual analysis of the new york bike share system was that there is a strong seasonal variation in the system usage with maximum usage in summer and minimum usage in winter. In this conclusion is based on that fact that even during the month of January and February which are the peak winter months , there are more than two hundred thousand group in this system.

Based on the hotspot analysis, it is clear that the most used bike stations are the ones closest to public transit. 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.

9.REFERENCES

1. http://scholarworks.sjsu.edu/mti_all/1
2. https://www.nyc.gov/html/dot/downloads/pdf/bike-share-outreach-report.pdf
3. https://towardsdatascience.com/exploring-bike-share-data-3e3b2f28760c
4. https://nycdatascience.com/blog/r/data-visulization-on-nyc-citibike/
5.https://www1.nyc.gov/assets/planning/download/pdf/plans/transportation/bike_share_complete.pdf