

A New Hint to Transportation-Analysis of the NYC Bike Share System

IBM-39037-1662629538

NALAIYA THIRAN PROJECT BASED LEARNING ON PROFESSIONAL READLINESS FOR INNOVATION, EMPLOYNMENT AND ENTREPRENEURSHIP

A PROJECT REPORT

ANEESH AHMED A	722819104007
ANIRUTHRAM K P	722819104008
DEEPAN KUMAR S	722819104026
FAYAZ A	722819104028

**BACHELOR OF ENGINEERING
COMPUTER SCIENCE AND ENGINEERING
SRI ESHWAR COLLEGE OF ENGINEERING
COIMBATORE-641202**

1. INTRODUCTION

1.1 Project Overview

1.2 Purpose

2. LITERATURE SURVEY

2.1 Existing Solution

2.2 References

2.3 Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

3.2 Ideation

3.3 Proposed Solution

3.4 Problem Solution fit

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

5. PROJECT DESIGN

5.1 Data Flow Diagrams

5.2 Solution & Technical Architecture

5.3 User Stories

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

6.2 Sprint Delivery Schedule

7. WORKING WITH THE DATASET&DATAVISUALIZATION

7.1 Understanding the Dataset

7.2 Loading the Dataset

7.3 Visualization Chart

8.CREATING THE DASHBOARD

9. ADVANTAGES & DISADVANTAGES

10. CONCLUSION

11. FUTURE SCOPE

12. SOURCE CODE

13. GITHUB LINK

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 CitiBike 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?”

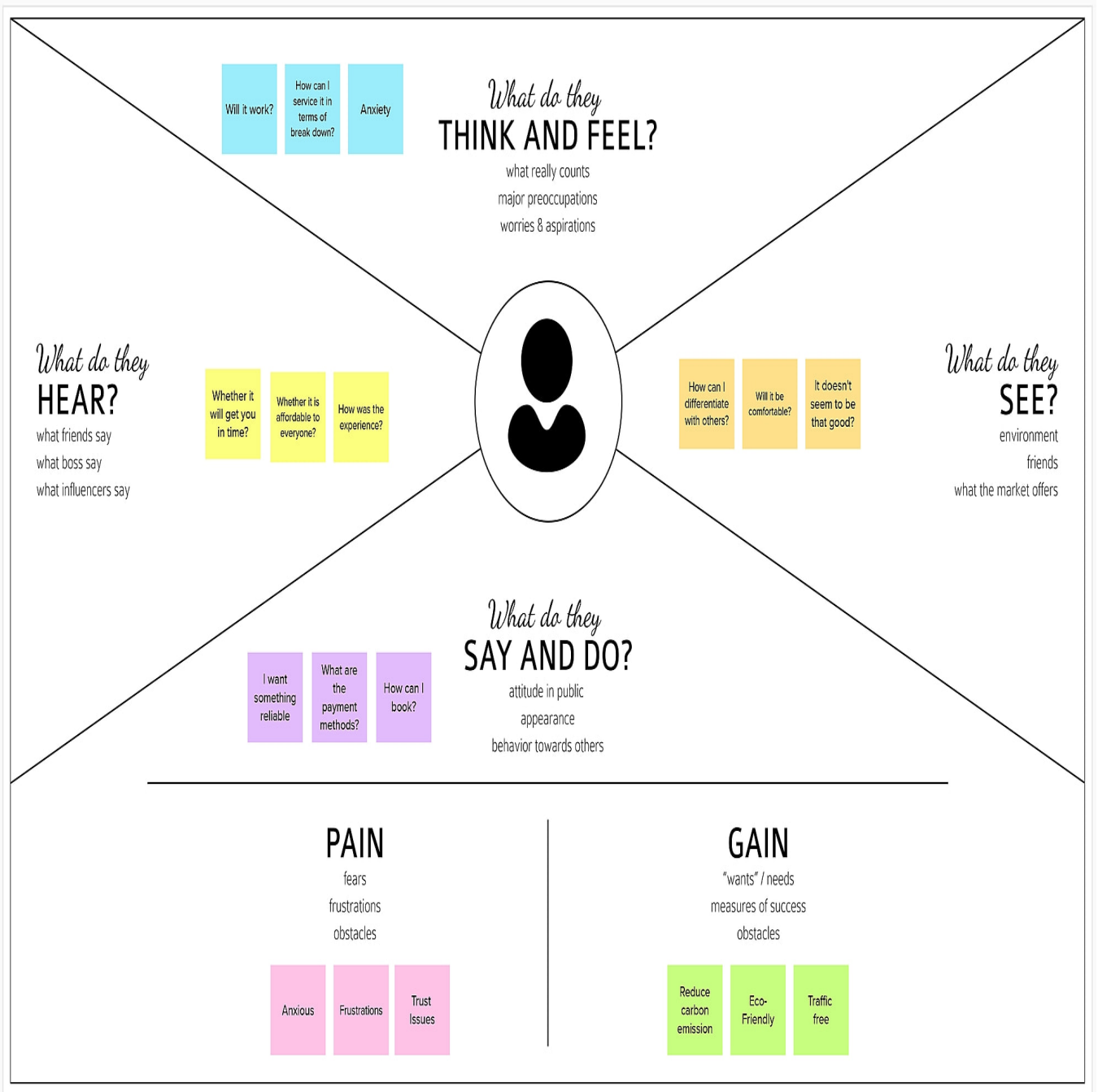
FRANCESCO et al., IEEE Access 2020 “Bike Sharing and Urban Mobility in a Post-Pandemic” “A long-term perspective on the COVID-19: The bike sharing system resilience under the epidemic environment” Journal of Transport & Health, 2021 Nguyen Thi Hoai Thu, Chu Thi Phuong Dung, Vietnam 2017 International 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



3.2 Proposed Solution

Proposed Solution Template:

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

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	<p>The goal of this analysis is to create an operating report of Citi Bike for the year 2018. To create data visualizations to understand</p> <ol style="list-style-type: none">1.Total Number of Trips2. What is Customer and subscriber with gender3.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.	Idea / Solution description	<p>To effectively process and visualize data, we use IBM Cognos because there are number of data and it is difficult to process it with some conventional methods, 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.</p>
3.	Novelty / Uniqueness	<p>Our solution reduces maintenance due to complete report coverage, The conversion rate of customers (who could be non-tourists) to subscribers and also provides fast results.</p>
4.	Social Impact / Customer Satisfaction	<p>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.</p>
5.	Business Model (Revenue Model)	<p>Revenue is generated on renting the bike with the analyzed data we can alter the pricing according to the demand of the customer.</p>
6.	Scalability of the Solution	<p>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</p>

3.3 Problem Solution Fit

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

Project Design Phase-I - Solution Fit

Team ID: **PNT2022TMID19163**

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS Who is your customer? i.e. working parents of 0.5 y.o. kids <ul style="list-style-type: none"> - 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. <ul style="list-style-type: none"> - 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: <ul style="list-style-type: none"> - Easy and simple to implement - Direct interaction with the end users of the bike share system Cons: <ul style="list-style-type: none"> - 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
	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 Citibike 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. <ul style="list-style-type: none"> - 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. <ul style="list-style-type: none"> - 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 	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. <ul style="list-style-type: none"> - 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 	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. 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 through Form

Registration through Gmail

Registration through LinkedIn

FR-2 User Confirmation Confirmation via Email

Confirmation via OTP

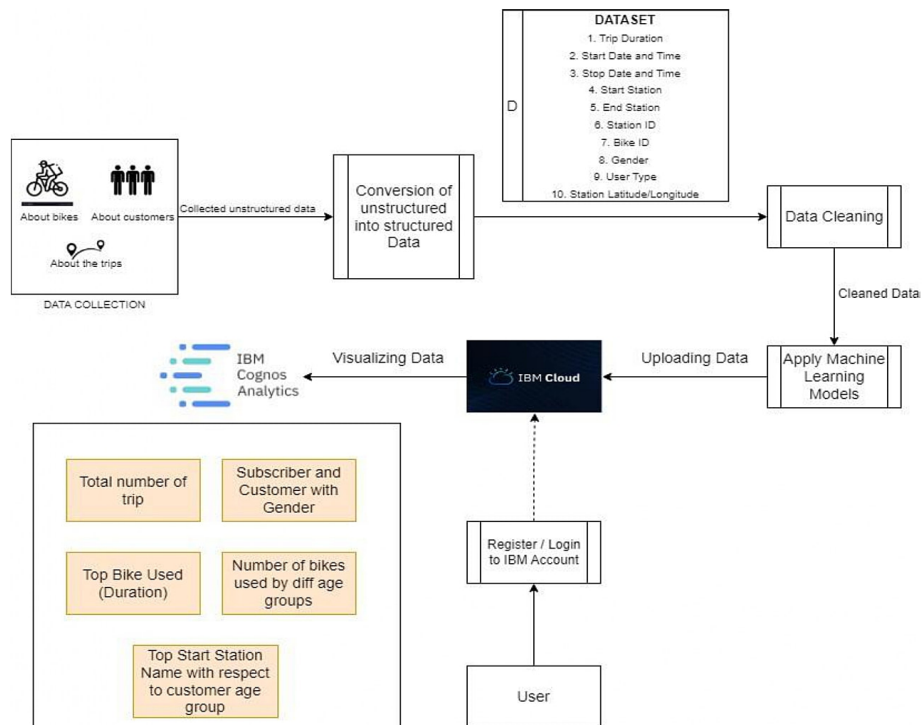
FR-3 Collection of Data Usage of the NYC Citi Bike help generate data regarding the different trips taken by different people using Citi Bike. These data were

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.

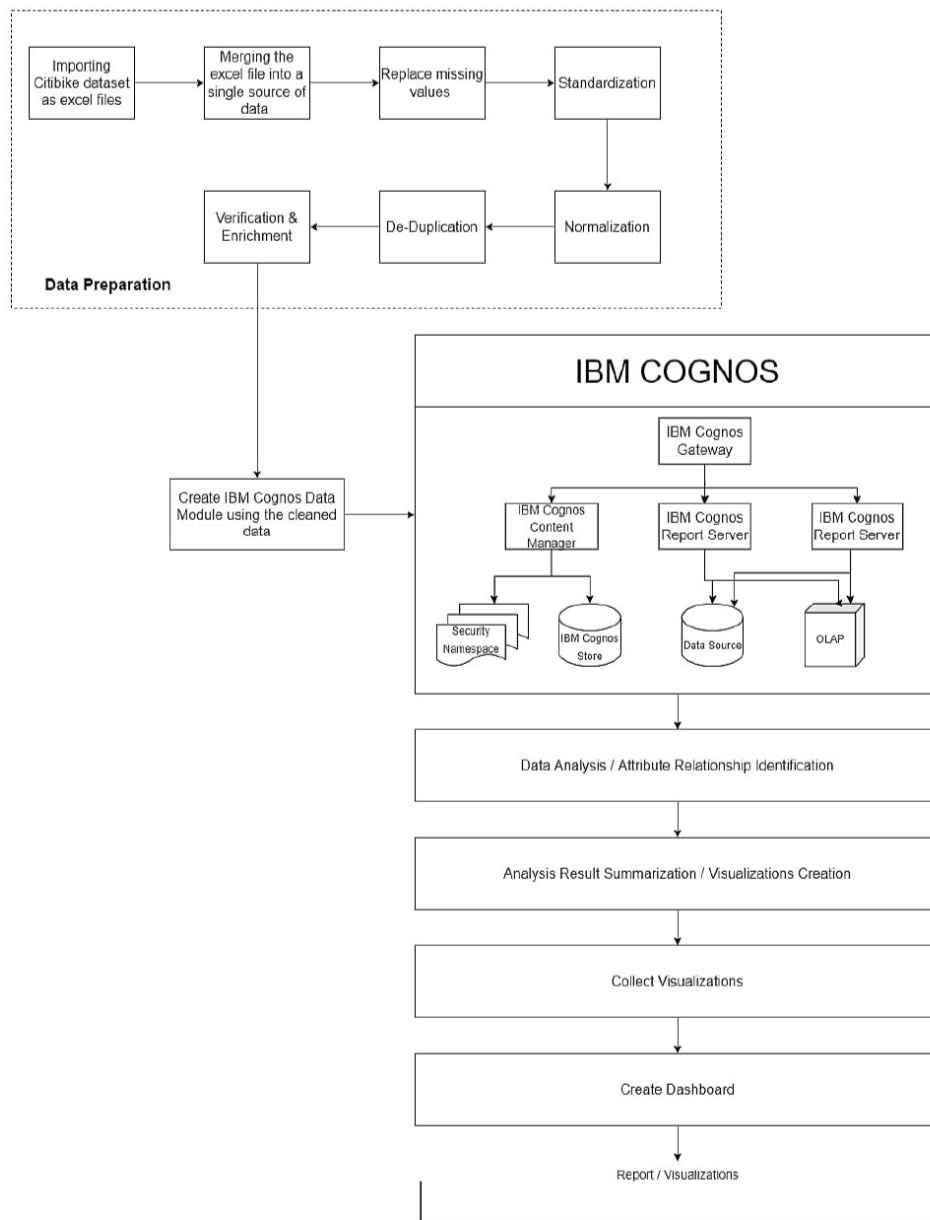
5. PROJECT DESIGN

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:



5.3 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, 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, Government	Login	USN-2	As a user I should be able to login to see the dashboard with the correct credentials	Successful Login with correct credentials	High	Sprint-1
Customer, Analysts, 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 : <ol style="list-style-type: none"> 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, 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
Customer, Analysts, Government	Collection of data	USN-5	Lyft citi bike website provides data for analyse, visualization etc.	Accessing the data in Lyft citi bike	High	Sprint -1
Customer, Analysts, Government	Analysing the data	USN-6	Analysed data is used to create different types visualization and charts.	Can view the analysis	High	Sprint-1

6.PROJECT PLANNING &SCHEDULING

6.1 Sprint Planning & Estimation

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

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	5	High	Aniruthram ,Deepan Kumar,Aneesh,Fayaz
Sprint-1	Registration	USN-2	As a user, I will receive confirmation email once I have registered for the application	2	High	Aneesh,Fayaz
Sprint-1	Registration	USN-3	As a user, I can register for the application through Gmail or phone number	1	Medium	Aniruthram ,Deepan Kumar
Sprint-2	Login	USN-4	As a user, I can log into the application by entering email & password	7	High	Deepan Kumar,Aneesh,Fayaz
Sprint-2	Collection of user data	USN-5	As a user, I can enter my information which includes bank details for easy online payment and my mobile number for OTP confirmation	10	High	Aniruthram ,Deepan Kumar,Aneesh
Sprint -3	Analysis of User data	USN -6	As a analyst, I can access the Lyft Citi bike database to access the bike sharing system data of the user	8	High	Aniruthram ,Deepan Kumar,Aneesh,Fayaz
Sprint-3	Analysis of User data	USN-7	The data can be used as input for creating various types of visualizations and analysis.	5	High	Deepan Kumar,Aneesh,Fayaz
Sprint -4	Visualization	USN-8	As an analyst, I create various visualizations using IBM Cognos based on the knowledge obtained at the end of the EDA process.	7	High	Aniruthram ,Deepan Kumar

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-4	Visualization	USN -8	As an analyst, I create a dashboard with the created visualizations to supplement business insights during the decision-making process at Citi. As an analyst, I apply predictive analytics and additional features to enhance visualizations	8	High	Aniruthram ,Deepan Kumar,Aneesh,Fayaz
Sprint-4	Dashboard	USN - 9	I can create and access the dashboard based on the user data	5	High	Aniruthram ,Deepan Kumar,Aneesh,Fayaz

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	8	6 Days	24 Oct 2022	29 Oct 2022	8	29 Oct 2022
Sprint-2	17	6 Days	31 Oct 2022	05 Nov 2022	25	
Sprint-3	13	6 Days	07 Nov 2022	12 Nov 2022	38	
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	48	

6.2 Sprint Delivery Schedule

MILESTONE AND ACTIVITY LIST:

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the technical papers, research publications, journals etc.	19 SEPTEMBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, prepare list of problem statements that are to be solved by this project.	19 SEPTEMBER 2022
Ideation	List the ideas by organizing a brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	19 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document, which includes novelty, feasibility of idea, revenue model, social impact, scalability of solution, etc.	19 SEPTEMBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	19 SEPTEMBER 2022
Solution Architecture	Prepare solution architecture document.	19 SEPTEMBER 2022

Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application (entry to exit).	3 OCTOBER 2022
Functional Requirement	Prepare the functional requirement document.	3 OCTOBER 2022
Data Flow Diagrams	Draw the data flow diagrams and submit for review.	3 OCTOBER 2022
Technology Architecture	Prepare the technology architecture diagram.	3 OCTOBER 2022
Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	18 OCTOBER 2022
Project Development - Delivery of Sprint-1, 2, 3 & 4	Develop & submit the developed code by testing it.	IN PROGRESS

7. WORKING WITH THE DATASET & DATA VISUALISATION

7.1 Understanding the dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
4	202	#####	#####	3186	Grove St P	40.71959	-74.0431	3213	Van Vorst	40.71849	-74.0477	24689	Subscriber	1962	2								
5	248	#####	#####	3209	Brunswick	40.72418	-74.0507	3203	Hamilton	40.7276	-74.0442	24693	Subscriber	1984	1								
6	903	#####	#####	3195	Sip Ave	40.73074	-74.0638	3210	Pershing	40.74268	-74.0518	24573	Customer		0								
7	883	#####	#####	3195	Sip Ave	40.73074	-74.0638	3210	Pershing	40.74268	-74.0518	24442	Customer		0								
8	445	#####	#####	3186	Grove St P	40.71959	-74.0431	3203	Hamilton	40.7276	-74.0442	24510	Subscriber	1988	2								
9	192	#####	#####	3211	Newark Av	40.72153	-74.0463	3203	Hamilton	40.7276	-74.0442	24625	Subscriber	1980	1								
10	409	#####	#####	3187	Warren St	40.72112	-74.0381	3214	Essex Light	40.71277	-74.0365	24429	Subscriber	1990	1								
11	285	#####	#####	3187	Warren St	40.72112	-74.0381	3214	Essex Light	40.71277	-74.0365	24407	Subscriber	1988	2								
12	206	#####	#####	3183	Exchange	40.71625	-74.0335	3187	Warren St	40.72112	-74.0381	24377	Subscriber	1993	1								
13	251	#####	#####	3213	Van Vorst	40.71849	-74.0477	3209	Brunswick	40.72418	-74.0507	24605	Subscriber	1976	1								
14	292	#####	#####	3186	Grove St P	40.71959	-74.0431	3187	Warren St	40.72112	-74.0381	24721	Subscriber	1992	1								
15	175	#####	#####	3186	Grove St P	40.71959	-74.0431	3213	Van Vorst	40.71849	-74.0477	24601	Subscriber	1991	1								
16	769	#####	#####	3193	Lincoln Pai	40.72461	-74.0784	3211	Newark Av	40.72153	-74.0463	24606	Subscriber	1968	1								
17	384	#####	#####	3209	Brunswick	40.72418	-74.0507	3213	Van Vorst	40.71849	-74.0477	24470	Subscriber	1975	2								
18	364	#####	#####	3209	Brunswick	40.72418	-74.0507	3213	Van Vorst	40.71849	-74.0477	24542	Subscriber	1976	1								
19	326	#####	#####	3186	Grove St P	40.71959	-74.0431	3202	Newport	40.72722	-74.0338	24716	Subscriber	1992	1								
20	341	#####	#####	3186	Grove St P	40.71959	-74.0431	3202	Newport	40.72722	-74.0338	24687	Subscriber	1982	1								
21	634	#####	#####	3194	McGinley	40.72534	-74.0676	3209	Brunswick	40.72418	-74.0507	24675	Subscriber	1981	1								
22	331	#####	#####	3202	Newport	40.72722	-74.0338	3203	Hamilton	40.7276	-74.0442	24687	Subscriber	1982	1								
23	550	#####	#####	3183	Exchange	40.71625	-74.0335	3199	Newport	40.72874	-74.0321	24531	Subscriber	1986	1								
24	988	#####	#####	3196	Riverview	40.74432	-74.044	3209	Brunswick	40.72418	-74.0507	24662	Customer		0								
25	969	#####	#####	3196	Riverview	40.74432	-74.044	3209	Brunswick	40.72418	-74.0507	24563	Subscriber	1984	1								
26	378	#####	#####	3202	Newport	40.72722	-74.0338	3203	Hamilton	40.7276	-74.0442	24450	Subscriber	1981	1								
27	368	#####	#####	3202	Newport	40.72722	-74.0338	3203	Hamilton	40.7276	-74.0442	24627	Subscriber	1982	2								
28	287	#####	#####	3202	Newport	40.72722	-74.0338	3183	Exchange	40.71625	-74.0335	24471	Subscriber	1986	1								
29	286	#####	#####	3214	Essex Light	40.71277	-74.0365	3187	Warren St	40.72112	-74.0381	24407	Subscriber	1988	2								
30	539	#####	#####	3195	Sip Ave	40.73074	-74.0638	3212	Christ Hos	40.73479	-74.0504	24664	Subscriber	1966	2								
31	340	#####	#####	3195	Sip Ave	40.73074	-74.0638	3193	Lincoln Pai	40.72461	-74.0784	24493	Subscriber	1972	1								
32	343	#####	#####	3186	Grove St P	40.71959	-74.0431	3203	Hamilton	40.7276	-74.0442	24526	Subscriber	1986	1								
33	243	#####	#####	3202	Newport	40.72722	-74.0338	3187	Warren St	40.72112	-74.0381	24422	Subscriber	1982	1								
34	1282	#####	#####	3209	Brunswick	40.72418	-74.0507	3194	McGinley	40.72534	-74.0676	24675	Subscriber	1981	1								

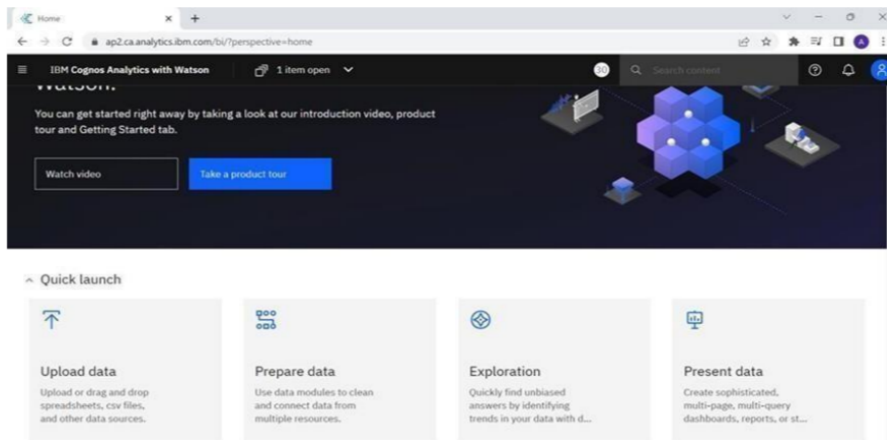
Dataset Link: Dataset

- 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

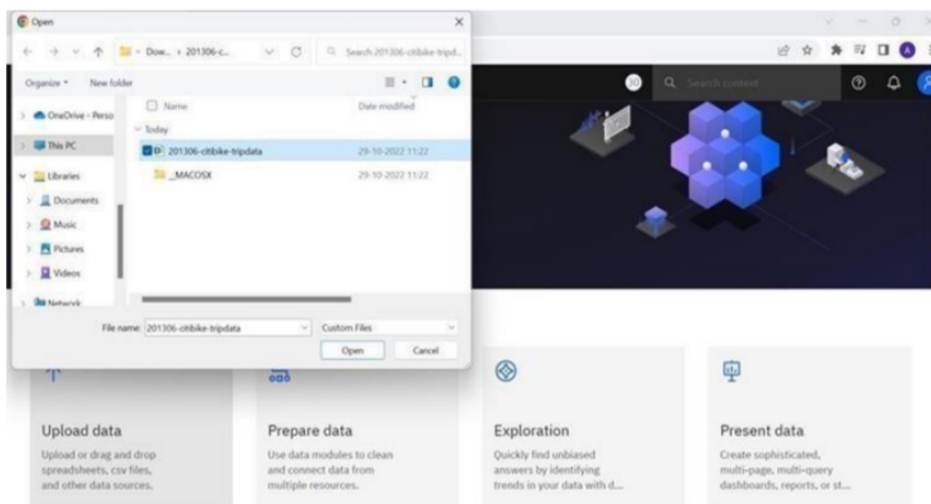
Open Cognos Analytics and click upload data

Open Cognos Analytics and click upload data

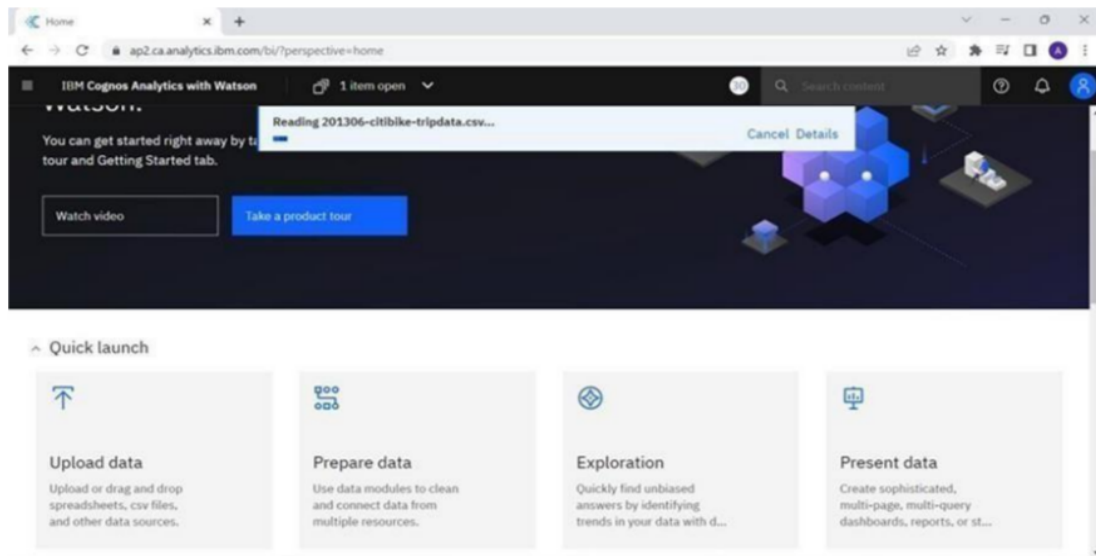


Select the dataset to be uploaded

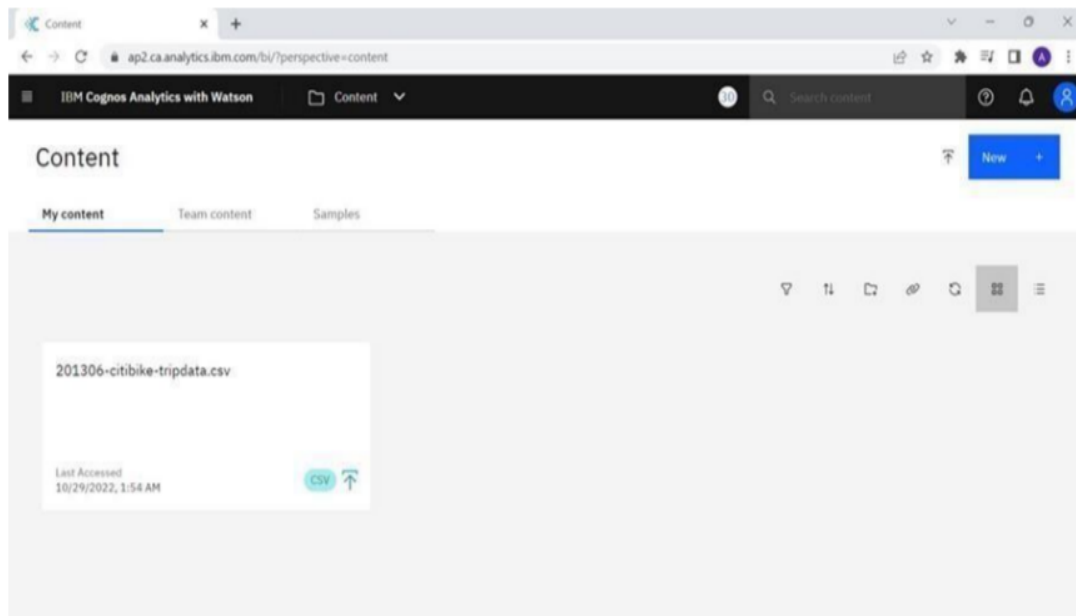
Select the dataset to be uploaded



The excel file is getting uploaded in Cognos Analytics



The dataset can be accessed in My Content in Cognos Analytics



7.3 Data Visualization charts

Number of Trips:

11/2/22, 12:44 PM

* New dashboard

number of trips

bikeid and tripduration

bikeid	tripduration
4,466	649.38

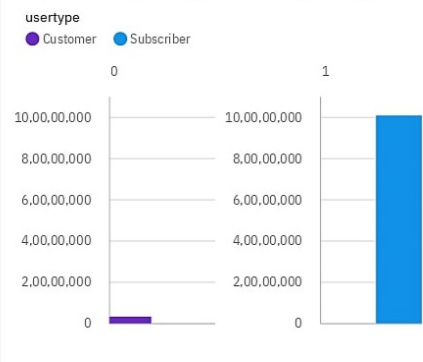
Customer and Subscriber with Gender:

11/7/22, 10:57 AM

* New dashboard

customer and subscriber with gender

tripduration by usertype colored by usertype

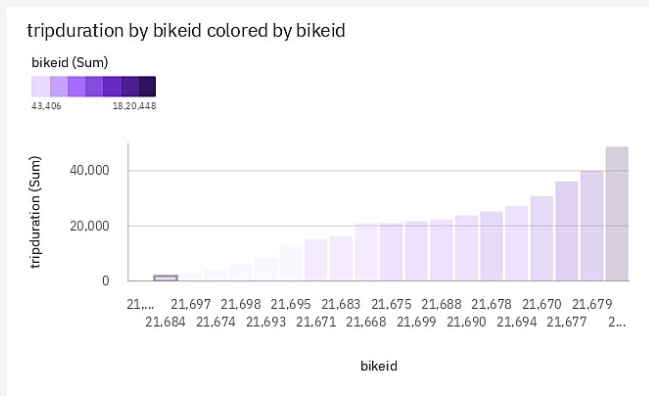


Bike Usage:

11/7/22, 11:05 AM

* New dashboard

bike usage

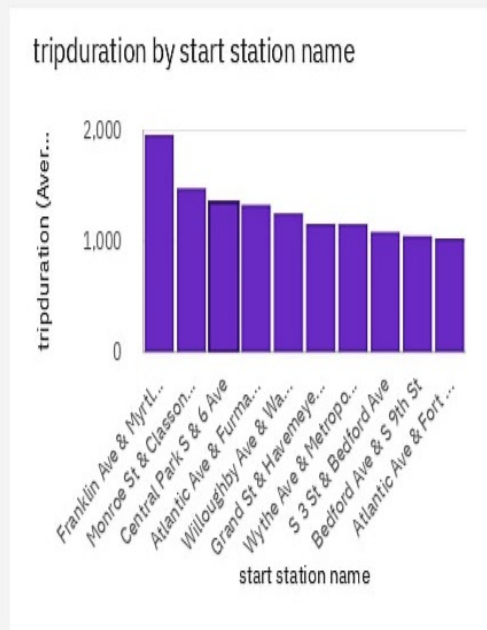


Top 10 Start Station Names with Respect to Customer Age Group:

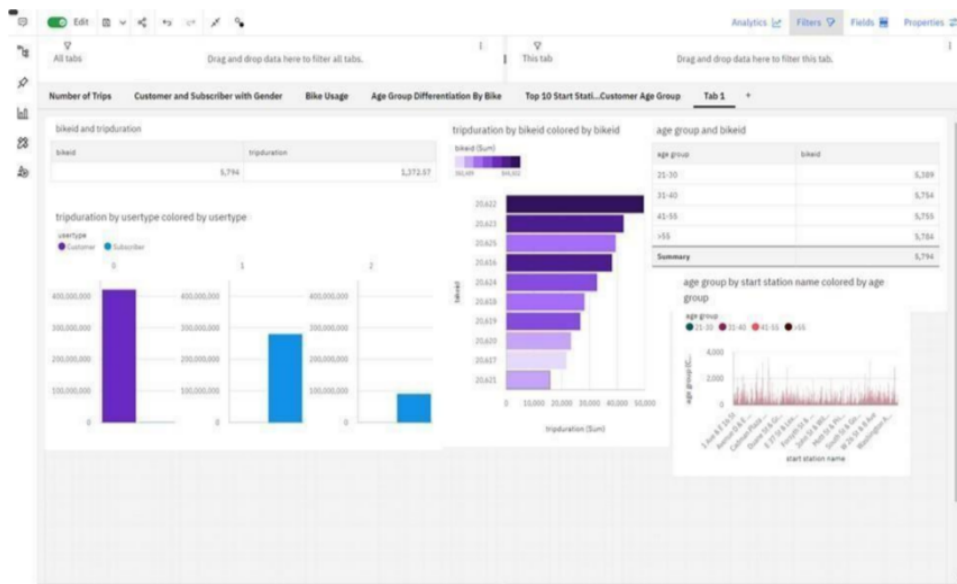
11/9/22, 9:48 AM

* New dashboard

Top 10 Start Station Names With Respect To Customer Age Group



8.CREATING THE DASHBOARD



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 a lot 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 indicate the needs for improved techniques in bike sharing analytics. There exists a lot of scope in this research area.

12. SOURCE CODE

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
from google.colab import files
files.upload()
import pandas as pd
df = pd.read_csv('JC-202102-citibike-tripdata.csv')
df.head()
df.describe()
df.info()
df.isnull().sum()
df[df['started_at'].isnull()]
df[df['ended_at'].isnull()]
df = df[:-1]
print(type(df["start_lat"][0]))
print(df["start_lat"][0])
df['start_lat'].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.info()
df.shape
df.to_csv('New_dataset.csv', index=False)

```

Sprint 4:

```

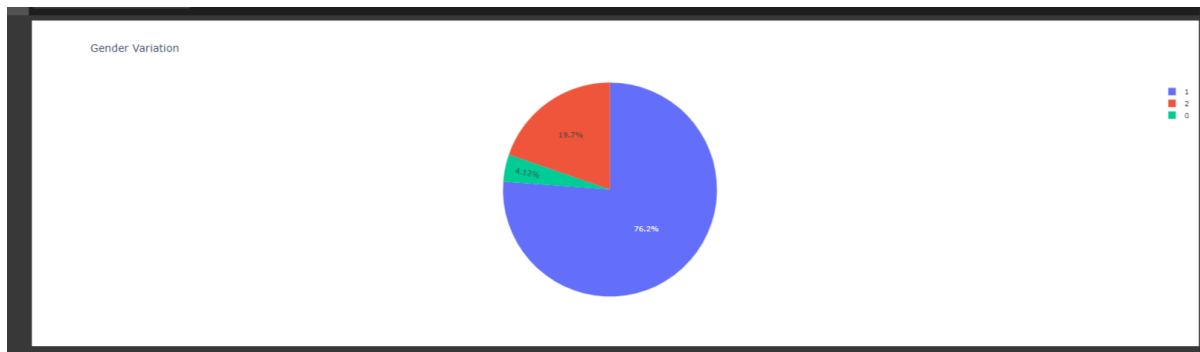
from google.colab import files
files.upload()

```

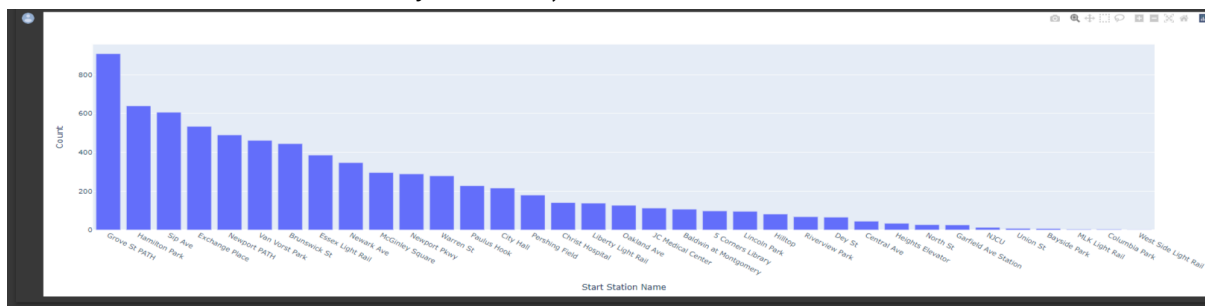
```

import pandas as pd
df = pd.read_csv('JC-20161-citibike-tripdata.csv')
df.head()
temp = df
temp.head()
temp.describe()
temp.info()
temp["Start Time"] = pd.to_datetime(temp["Start Time"])
temp["Stop Time"] = pd.to_datetime(temp["Stop Time"])
temp.info()
temp["Hour"] = temp["Stop Time"].dt.hour - temp["Start Time"].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")

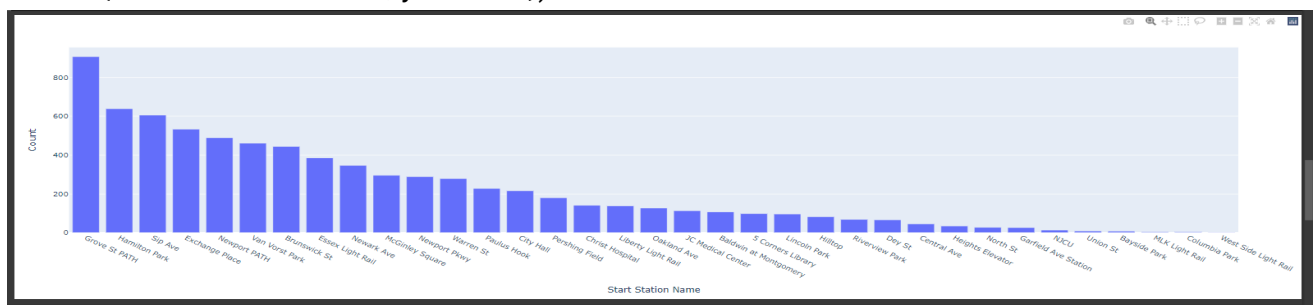
```



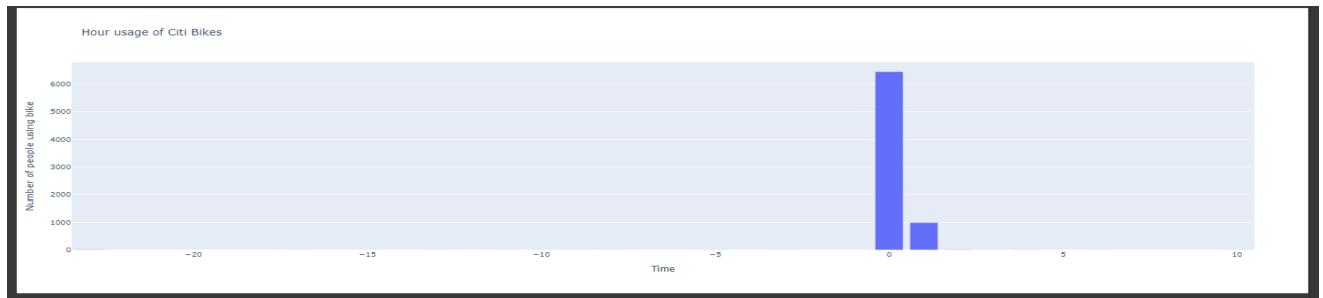
```
px.bar(x=temp["Start Station Name"].value_counts().index,
y=temp["Start Station Name"].value_counts().values,
labels={'x':'Start Station Name','y':"Count"})
```



```
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-39037-1660389703>