

DATA MANAGEMENT(DATA 604)

Project Paper

Maryland Transit Administration

Team:

Bheemishetty Sai Baba (US53017)

Jashwanth Kadem (JT23739)

Anil Kumar Mayala (IN00574)

Shankar Chandra Pothinindi (MS77377)

Charan Ganemaneni (WW88694)

M.P.S DATA SCIENCE, FALL 2022

Instructor: Sarika Rajeev

University of Maryland Baltimore county

December 2022

Contents:

1. Abstract
2. Introduction
3. Literature review
4. Data Specification
5. Process and Tools Used
6. Data Processing
 - 6.1. Tables of the Data set
7. Entity Relationship Diagram of the Data set
8. Queries and the Results of the project
 - 8.1. Queries
 - 8.2. Visualization
 - 8.2.1. Power Bi
 - 8.2.2. Jupyter
9. Future Scope of the project
10. Analysis and Conclusion
11. References

1. Abstract:

Maryland Transit Administration is a state-operated mass transit administration in Maryland state in the United States, This operates a comprehensive transit system in between Baltimore-Washington metropolitan area^[1]. In this project, we are specifically taking the data of the Baltimore city transit and visualizing the transit routes according to their journey time and distance in every way. In the 4 major routes that we considered for visualizing the Baltimore transit, we found that the Longest transit route Purple has a travel duration of 71 minutes, the busiest Orange route has 328 bus trips per day and on average each bus trip (including each route) is about 46:17 minutes.

2. Introduction:

Baltimore is the largest city in the state of Maryland. The MTA took over the operations of the ancient Baltimore Transit Company on April 30, 1970. It was once initially recognized as the Baltimore Metropolitan Transit Authority, then the Mass Transit Administration earlier and it modified to its contemporary title in October 2001. Many of the routes of most of the agency's contemporary bus traces are based totally on the unique streetcars operated via the Baltimore Transit Company and its dad or mum businesses from the Eighteen Nineties to the 1960s^[2]. All these routes have been subsequently transformed into rubber tire bus operations, and many of them had been consolidated, prolonged into newly developed areas, or in any other case reconfigured to hold up with the ridership needs of the times. Additional routes and extensions have been introduced in later years to serve communities that had been later developed and to feed into Metro and Light Rail stations that have been later built.

Table 1: The fleet of the Maryland Transit Administration

SERVICE	NUMBER
Heavy Rail	100
Light Rail	53
Commuter Rail	175
Bus	842

Mobility Vans	303
---------------	-----

A Majority portion of the residents in Baltimore lacks personal transportation such as cars. So, the people of Baltimore use the Baltimore city link(transit service) to commute for their daily needs. Baltimore Link is a complete overhaul and rebranding of the core transit system operating within the city and throughout the greater Baltimore region.



Picture 1: MTA Ford E-Series van used for Mobility services.



Picture 2: MTA Bus services.



Picture 3: MTA Light rail services.

The Annual Ridership of the Maryland Transit administration is approximately 44,612,100 (Four Crores Forty-Six Lakhs Twelve Thousand One Hundred). The average daily ridership of the Maryland Transit administration is about 1,49,200 (One Lakh Forty-Nine Twelve Thousand Two Hundred). So, this shows how the people of Maryland are dependent on Public transportation. Since last few years, the On-time performance has been gradually increasing. In 2019, the On-time performance of the MTA has raised from 59.5% to 74% till the end of December.

3. Literature Survey

Case Study 1:

<https://nap.nationalacademies.org/read/24972/chapter/9>

In this particular case study, the creator has described the primary threats facing MTA'S infrastructure and services including flooding, storm surge and sea level rise. It also shows the operations and vehicles operated. It concludes that involving operations, maintenance, and police directly in safety and emergency management decision-making, as well as project prioritizations, helps embed a safety orientation as well as risk awareness into everyday operations.

Case Study 2:

<https://www.foursquareitp.com/project/purple-line-lrt-metrobus-efficiency-study/>

The corridor-wide study seeks to create a bus transit feeder plan that responds to service commencement, associated growth and changing land use patterns, and shifting demographics and travel expectations. This project included reviewing existing conditions and developing cost-neutral and visionary recommendations for Metrobus service along the corridor.

4. Data Specification:

Source of the Data:

The data has been taken from the Baltimore City Department of Transportation Transit Bureau which operates the public transport system in the state of Maryland.

This data set contains 4 tables which have approximately 30,000 rows. This dataset contains information about the scheduled arrival and departure times, bus stops, and GPS coordinates of all the bus routes, in Baltimore city.

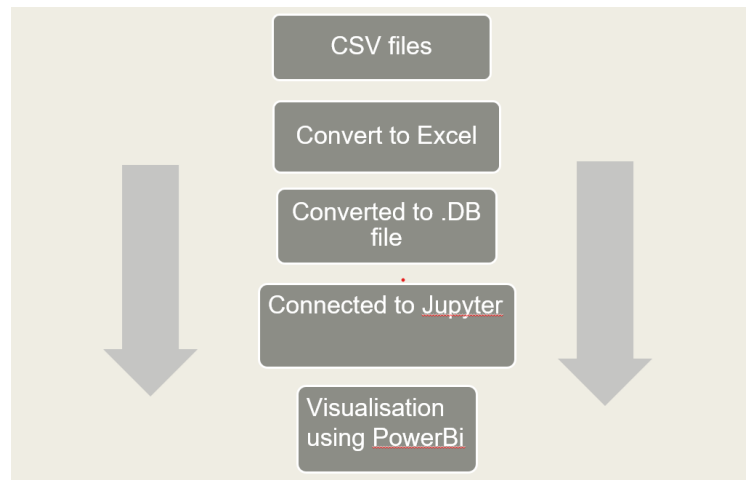
The columns include route_id, trip_id, stop_id, arrival_time, departure_time, shape_pt_lat, shape_pt_lon.

5. Process and Tools Used :

1. **SQL Lite:** For this project, SQL lite is used to create the Maryland Transit Administration (MTA) database (.db) from the Comma Separated Values (.csv) files.
2. **Microsoft SQL Server Management Studio:** SQL server Management studio is used to execute the queries for the Database.
3. **Jupyter:** Jupyter Notebook is used to create a python notebook, data cleaning and visualizing the data.
4. **Power Bi:** Power Bi is used to create the Visualization for the data set.

6. Data Processing:

There are a few steps that we followed to process our data that would lead to the visualization.



Picture 4: MTA Light rail services.

We got our data from data.gov website in the form of .csv files. We have used Microsoft Excel to convert them into .xlsx format. Then we used online tools to convert our .xlsx data into a query that could be run on SQLite software to create tables. Later we used SQLite software to create the MTA database(.db file) which was later used to connect to Jupyter notebook. Queries for the problem statements were executed on Microsoft SQL Server Management Studio 18 and on Jupyter. The data and the query results were visualized using Microsoft Power BI and Jupyter.

6.1 Tables of the Data Set :

From processing the Data by using the above described method and in the SQL Lite we can have the Tables of the data set.

There are 4 tables that are present in the Dataset. They are

1. **Route map table:** The route map table describes the number of routes and types of routes for the bus transit.

route_id	route_long_name	route_color	route_type
10	Orange Route	f2800e	3
12	Banner Route	377ed4	3
13	Purple Route	d51bdb	3
21	Green Route	84c74d	3

Picture 5: Route map table

2. **Stop_Times table:** This table gives the information about the stop timings of the busses in each route according to their trip id.

trip_id	arrival_time	departure_time	stop_id	stop_sequence	timepoint
8074	06:38:00	06:38:00	163	0	1
8074	06:53:00	06:53:00	24	3	1
8074	06:55:00	06:55:00	166	4	1
8074	06:56:00	06:56:00	167	5	1
8074	06:58:00	06:58:00	168	6	1
8074	06:59:00	06:59:00	169	7	1
8074	07:00:00	07:00:00	170	8	1
8074	07:01:00	07:01:00	171	9	1
8074	07:03:00	07:03:00	172	10	1
8074	07:04:00	07:04:00	173	11	1
8074	07:06:00	07:06:00	174	12	1
8074	07:07:00	07:07:00	175	13	1
8074	07:16:00	07:16:00	176	14	1
8074	07:18:00	07:18:00	177	15	1
8074	07:20:00	07:20:00	178	16	1
8074	07:22:00	07:22:00	179	17	1
8074	07:23:00	07:23:00	180	18	1
8074	07:25:00	07:25:00	181	19	1
8074	07:26:00	07:26:00	182	20	1
8074	07:27:00	07:27:00	183	21	1
8074	07:28:00	07:28:00	184	22	1
8074	07:29:00	07:29:00	185	23	1
8074	07:30:00	07:30:00	187	24	1
8074	07:32:00	07:32:00	188	25	1
8074	07:33:00	07:33:00	189	26	1
8074	07:35:00	07:35:00	190	27	1

Picture 6: Stop_Times table

3. **Shape table:** The Shape table gives the data of the Latitude and longitude of the bus stops present across the Baltimore.

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence
13	39.28321	-76.6129	0
13	39.2833	-76.61291	1
13	39.28335	-76.61291	2
13	39.28379	-76.61297	3
13	39.28414	-76.61302	4
13	39.28422	-76.61303	5
13	39.28436	-76.61305	6
13	39.28481	-76.61311	7
13	39.28498	-76.61313	8
13	39.28503	-76.61313	9
13	39.2851	-76.61313	10
13	39.28516	-76.61313	11
13	39.28524	-76.61311	12
13	39.28531	-76.61309	13
13	39.2854	-76.61305	14
13	39.28547	-76.61302	15
13	39.28554	-76.61298	16
13	39.2856	-76.61294	17
13	39.28567	-76.61287	18
13	39.28571	-76.61283	19
13	39.28573	-76.61281	20
13	39.28576	-76.61278	21
13	39.2858	-76.61271	22
13	39.28586	-76.61265	23
13	39.28601	-76.61248	24
13	39.28608	-76.6124	25

Picture 7: Shape table

4. **Trips table:** The Trips table gives the information about the number of trips that each route makes.

trip	route	service	shape	block
8074	13	75	13	P5
8027	13	75	13	P1
8050	13	75	13	P3
8062	13	75	13	P4
8086	13	75	13	P6
8028	13	75	13	P1
8039	13	75	13	P2
8051	13	75	13	P3
8063	13	75	13	P4
8075	13	75	13	P5
8087	13	75	13	P6
8029	13	75	13	P1
8040	13	75	13	P2
8052	13	75	13	P3
8064	13	75	13	P4
8076	13	75	13	P5
8088	13	75	13	P6
8030	13	75	13	P1
8041	13	75	13	P2
8053	13	75	13	P3
8065	13	75	13	P4
8077	13	75	13	P5
8089	13	75	13	P6
8031	13	75	13	P1
8042	13	75	13	P2
8054	13	75	13	P3

Picture 8: Trips table

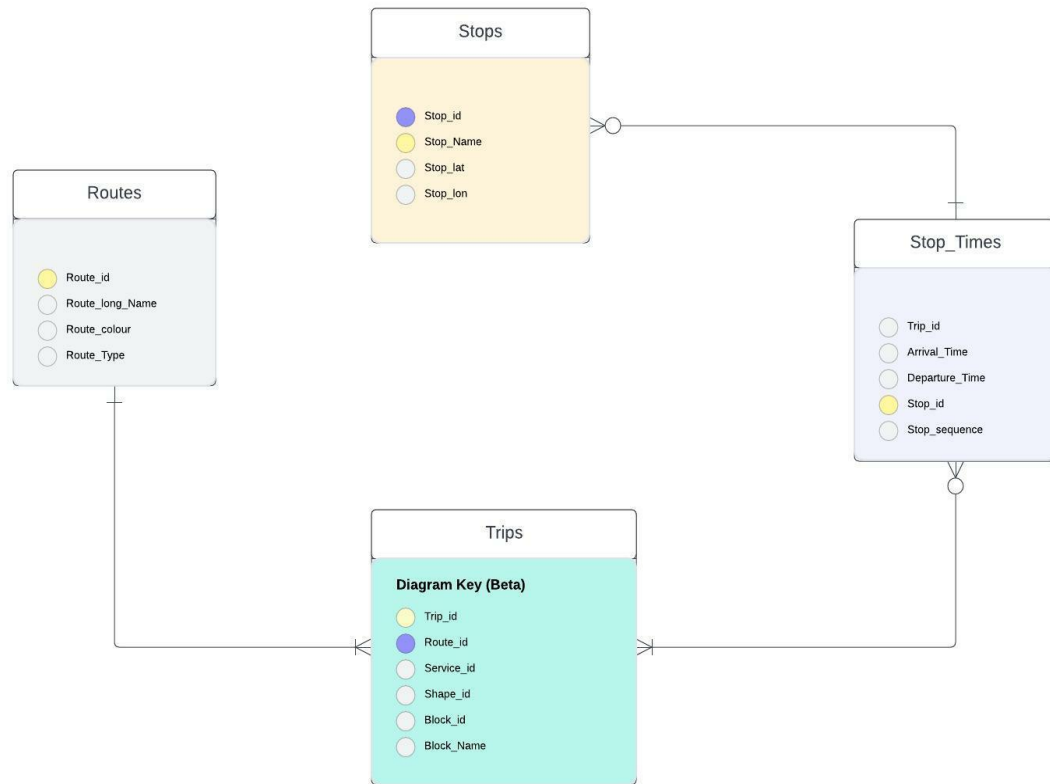
7. Entity Relationship Diagram of the Data set:

An Entity Relationship diagram gives an overview of the data in what way each

table is dependent on the other and gives information about the primary and foreign keys of the tables.

So, In our MTA dataset, there are 4 major tables containing about 30,000 rows and different types of columns. So, now the ER diagram gives the relation between the 4 different tables that are in the MTA Dataset.

Here, the yellow spotted column is the primary key and the blue spotted column is the foreign key of the table.



Picture 9: Entity Relationship Diagram for MTA Dataset

8. Queries and the Results of project:

The following were the problem statements to which queries have been written.

1. To analyze the busiest bus service route in Baltimore.
2. To find out the number of bus trips per route.
3. To find out the longest and shortest bus routes.

8.1 Queries:

1. a) To analyze the busiest bus service route in Baltimore.

Query:

```
SELECT TOP 1 trips.route_id AS ROUTE_ID, route_long_name AS
ROUTE_NAME, COUNT(service_id) AS [NUMBER of TRIPS]
FROM routes INNER JOIN trips ON routes.route_id= trips.route_id
GROUP BY trips.route_id, routes.route_long_name
ORDER BY COUNT(service_id) DESC
```

Result:

Results		Messages	
	ROUTE_ID	ROUTE_NAME	NUMBER of TRIPS
1	10	Orange Route	328

Picture 10: Query Result

- b) To find the less busy bus service route in Baltimore.

Query:

```
SELECT TOP 1 trips.route_id AS ROUTE_ID, route_long_name AS
ROUTE_NAME, COUNT(service_id) AS [NUMBER of TRIPS]
FROM routes INNER JOIN trips ON routes.route_id= trips.route_id
GROUP BY trips.route_id, routes.route_long_name
ORDER BY COUNT(service_id)
```

Result:

Results		Messages	
	ROUTE_ID	ROUTE_NAME	NUMBER of TRIPS
1	12	Banner Route	164

Picture 11: Query Result

2. To find out the number of bus trips per route.

Query:

```
SELECT trips.route_id AS ROUTE_ID, route_long_name AS
ROUTE_NAME,COUNT(service_id) AS [NUMBER of TRIPS]
FROM routes INNER JOIN trips ON routes.route_id= trips.route_id
GROUP BY trips.route_id, routes.route_long_name
```

Result:

	ROUTE_ID	ROUTE_NAME	NUMBER OF TRIPS
0	10	Orange Route	328
1	12	Banner Route	164
2	13	Purple Route	284
3	21	Green Route	182

Picture 12: Query Result

3. To find out the longest and shortest bus routes.

Query:

```
SELECT route_id, trips.trip_id AS Trip_Id,
MIN(arrival_time) AS Start_Time ,MAX(arrival_time) AS End_Time
FROM stop_times, trips WHERE stop_times.trip_id=trips.trip_id
GROUP BY trips.trip_id, trips.route_id
ORDER BY trip_id
```

Results:

Longest Bus Route:

route_id	Trip_Id	Start_Time	End_Time	Duration
530	13	9430	06:38:00 07:49:00	0 days 01:11:00

Picture 13: Query Result

Shortest Bus Route:

route_id	Trip_Id	Start_Time	End_Time	Duration
883	12	9828	09:10:00 09:40:00	0 days 00:30:00

Picture 14: Query Result

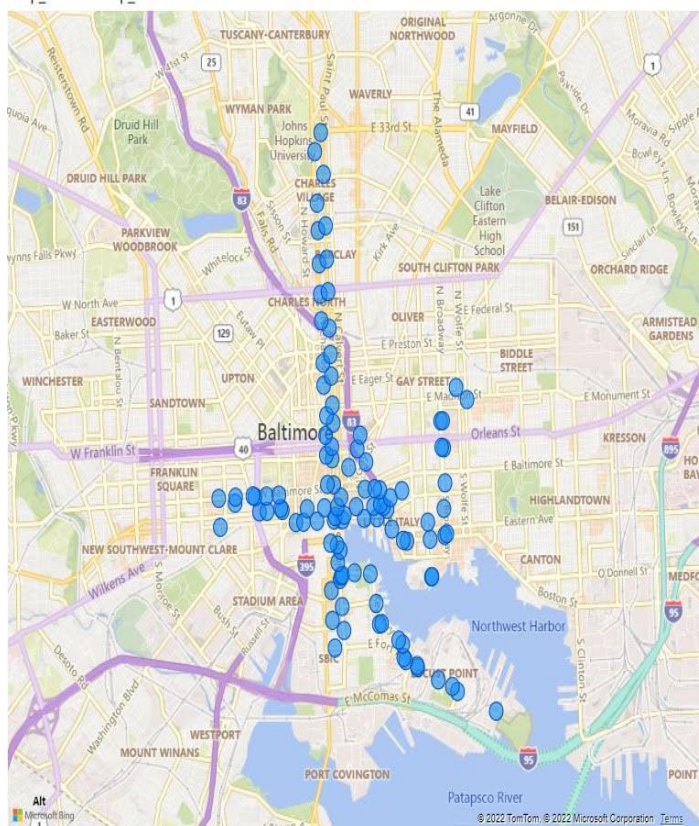
8.2 Visualization:

Visualization in Data science is used to convey our idea to the audience so that they can understand the data, make it simpler and interact with them. Whether the data is simpler or complex by visualizing the dataset we can bring everyone to the same page.

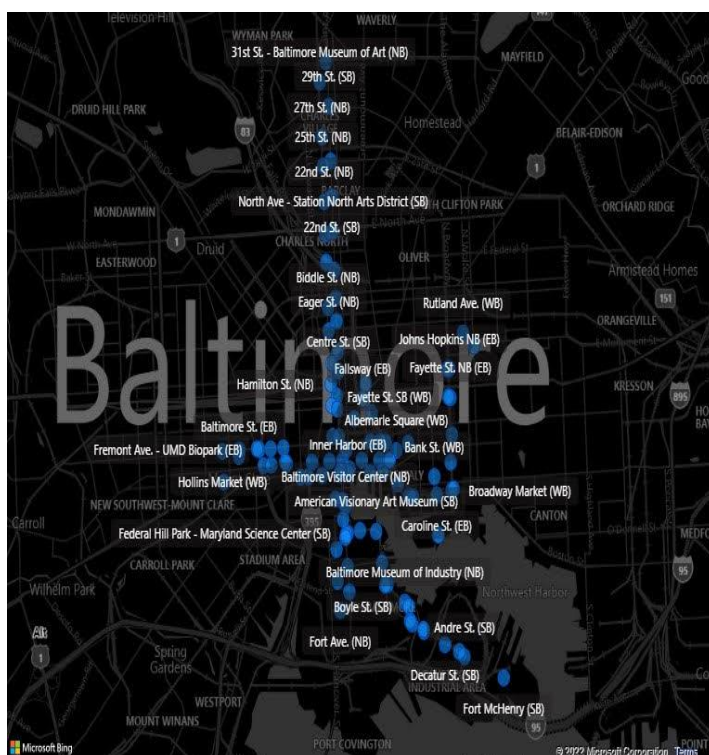
8.2.1. Visualization in Power Bi

Power Bi is a Microsoft technology-driven business intelligence product, Power BI, that analyzes and visualizes raw data to provide usable knowledge. It brings together business analytics, data visualization, and best practices to support an organization's decision-making using data.

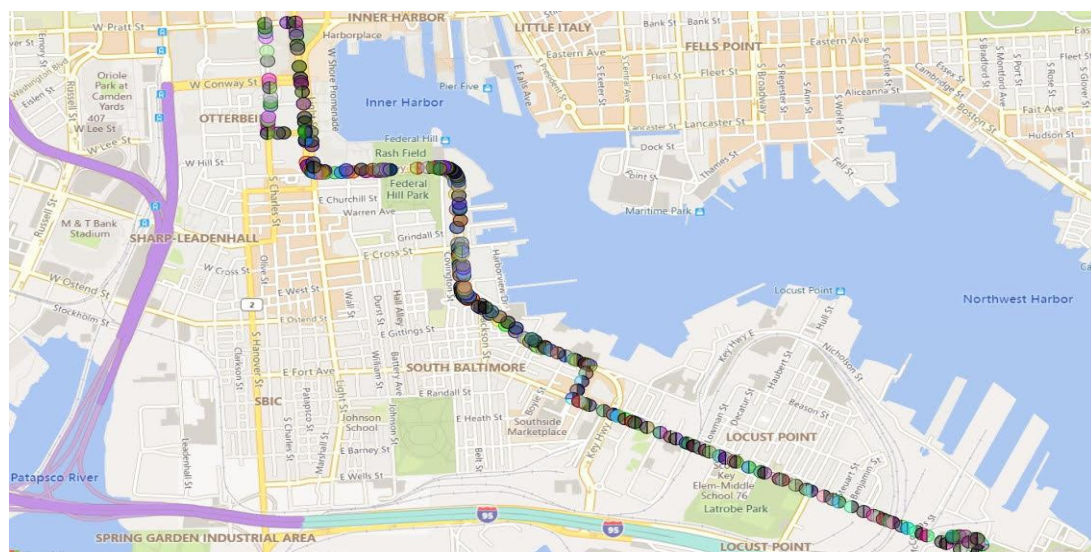
We have used Power Bi software to visualize our data and the query results, which helps us analyze our data clearly.



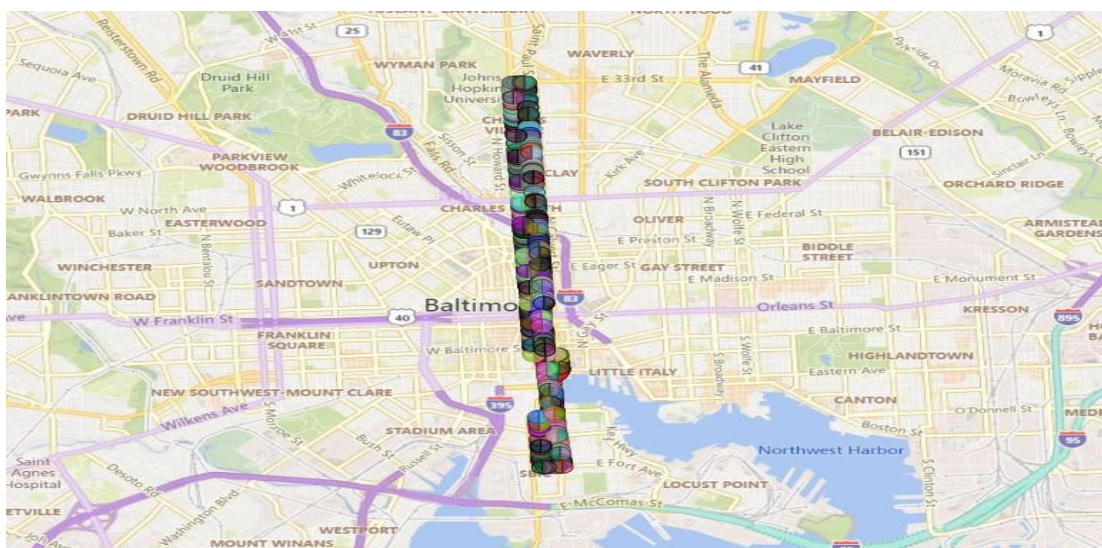
Picture 15: Map of bus stops in Baltimore city



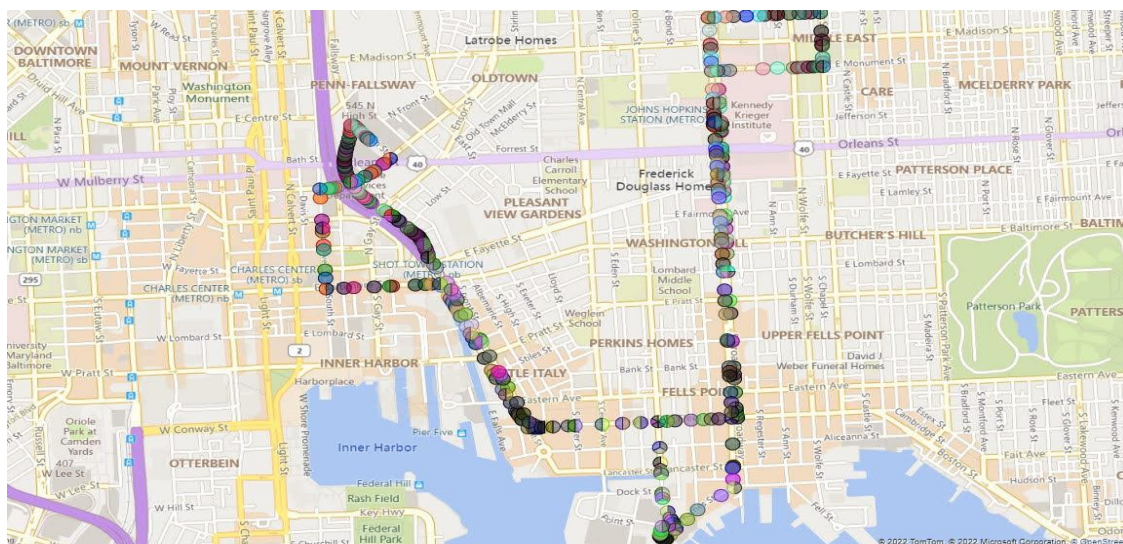
Picture 16: Map of bus stops in Baltimore city in dark view



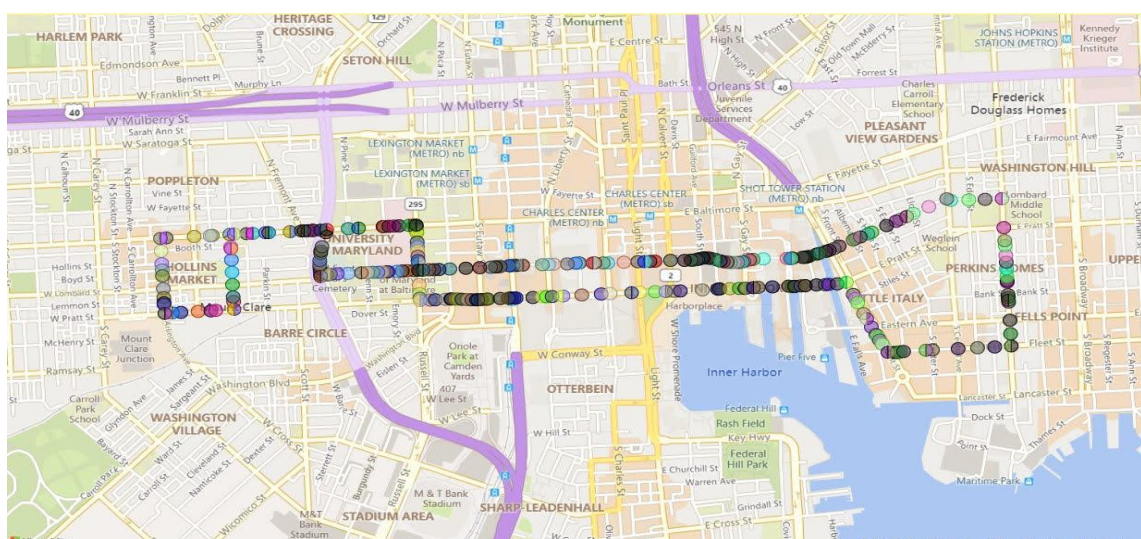
Picture 17: Map of route number 12



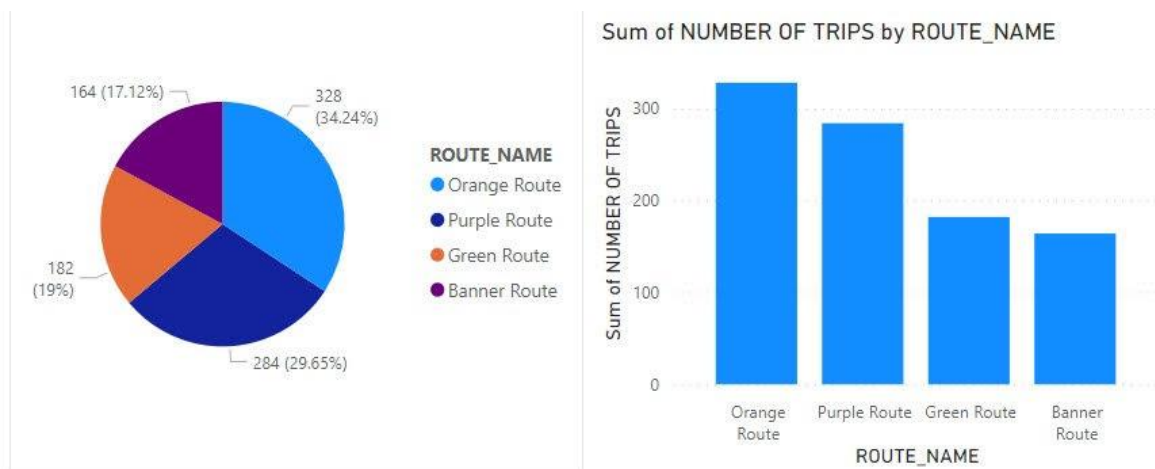
Picture 18: Map of route number 13



Picture 19: Map of route number 21

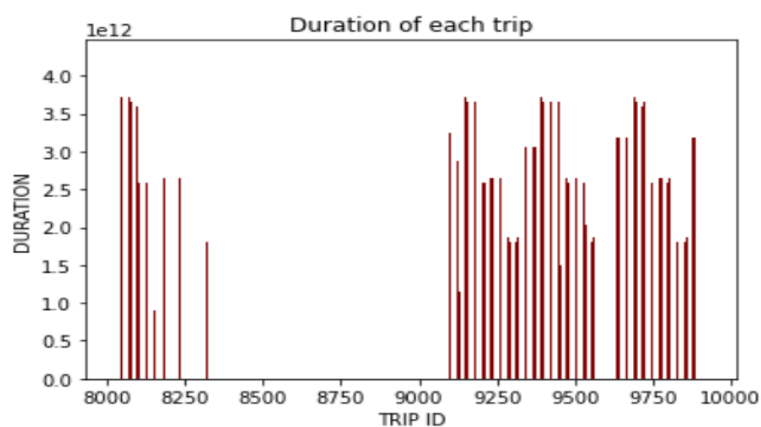


Picture 20: Map of route number 10

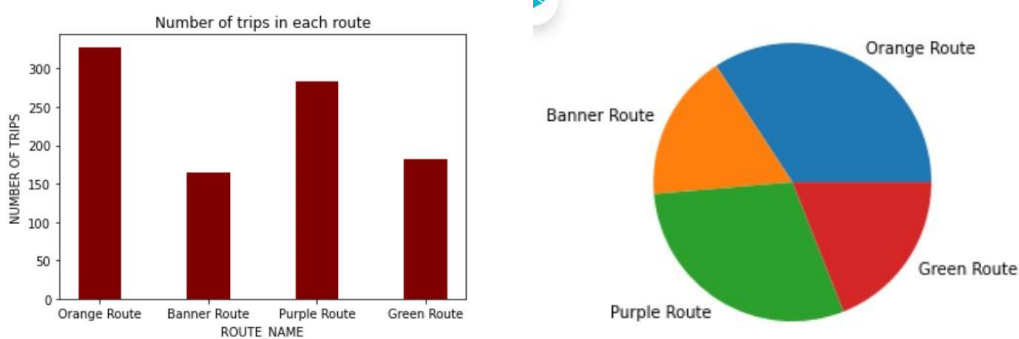


Picture 21: Query result of the number of bus trips in each bus route

8.2.1. Visualization in Jupyter



Picture 22: Query result of the duration of each bus trip



Picture 23: Query result of the number of bus trips in each bus route

9. Future Scope of the project

In future, we could focus on analyzing more bus services including other modes of public transport like light rail, subway link and MARC in the city of Baltimore. We would also analyze the busiest, longest, and shortest transport services across different modes. Visualizing these insights will help MTA in the betterment of services provided to the public.

10. Analysis and Conclusion

After analyzing and visualizing the data, we were able to conclude the following.

1. On average each bus trip is about 46:17 Mins.
2. Orange route is the busiest route with 328 bus trips.
3. Banner route is the least busy route with 168 bus trips.
4. Purple route is the longest route with a journey duration of 01:11 Hrs.
5. Banner route is the shortest route with a journey duration of 00:30 Hrs.

11. References

1. https://en.wikipedia.org/wiki/Maryland_Transit_Administration
2. <https://www.mta.maryland.gov/about>
3. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/Baltimore_FR_LM_Final_Repo
4. <https://catalog.data.gov/dataset/ntm-baltimore-city-department-of-transportation-transit-bureau/resource/6da92db2-40c1-4042-8007-1d5d04aadd4e>