

# **Analysis of the Tartu public city bike sharing service as well influence of Metallica concert on public transportation service.**

**By:Group M15:**

**Niyi Solomon Adebayo,**

**Michael Joseph,**

**Abdul-Rasheed Ottun.**

## **Task 2. Business understanding**

### **Background**

Bike-sharing systems in Tartu is an innovative program of providing rental bicycles in inner urban areas, which differ from traditional mostly leisure-oriented bicycle rental services. This systems provide residents (or tourists) with a fast and easy access to public bicycles that can be used for both daily mobility and one way use. The Tartu bike-sharing system consists of public bicycles that can be picked up and dropped off at a number of fixed bike stations across the urban area. On the public bicycles in their respective docking stations it has a service terminal for system access and user registration. Public bicycles are available to the general public with a small fee, and can be used in a unimodal fashion or as feeder mode to other transportation systems as it is in Tartu city, where the bike system is complementary to the city bus services.

### **Goal(s)**

As aforementioned , the increasing interest of the public in public transportation services, necessitates the need for constant improvement of the services so as to maintain its current appeal and strategically seduce new users to the bike-sharing services. This is background upon which the analysis of the data is premise upon, that is, to inquire the data in such way to be able give recommendations of strategic planning in the future in the event any public in such a magnitude as the Metallica concert. The inherent goal will be to help the city government in strategic planning of public transportation especially bike-sharing services especially during public gatherings in such as to reduce stress on critical infrastructures, roads, minimize risk associated automobile in

crowded environments, reduce environmental pollution and generally increase bike usage thereby promoting healthy lifestyle.

**Success criteria:** This is relatively subjective criteria, this study provides a complete insight into how the bike-sharing system is used (at least for the duration under investigation), in terms of both trips, riding trends and activities. Therefore, there's really no wrong answer.

**Inventory of resources:** The resources available for these project includes: Human resources; 3 data science student with backgrounds in robotics, bio-engineering and business management respectively. There also access to 3 data science instructors (of which one is an associate professor of data science) for troubleshooting and idea discussion if need be.

Hardwares and softwares: 3 PCs and access to a high-end computer(GPU:2 \*NVIDIA RTX 2080) in the case of perform machine learning tasks. We have access to almost all needed data science softwares and APIs: Anaconda, Tableau, Data warehouse, TensorFlow, Pandas, Sci-Learn etc.

**Requirements, assumptions, and constraints:** The data to used for this project is already being acquired from the city government through the computer science institute and as such it is fairly large and robust data set which quite clean. The legal obligation is for no part of data to be exposed in any way or form. The deadline is barely 3 weeks and it is expected each student expends at least 30 hours of work on the project after which the will a poster session to display our findings and result as te case maybe.

**Risks and contingencies:** one of the issues already surfacing the time management and work delegation. It is a busy time for student in the university of Tartu especially as the semester comes to and end. The plan is to work 4-6 hours daily on the project. And contingency plan would be to work long hours on weekends for 10-14 hours.

**Costs and benefits:** the apparent cost of this project will be time and mental investment and the benefits will be perfecting and applying all we've lean rt in the classroom setting to real world problems and recommending a practicable solution to the appropriate people thereby contributing our quote to development.

## Task 3. Data understanding

### Data Gathering:

As hitherto mentioned, the data for the analysis was provided by the Tartu city government and hence we need not conduct any data gathering.

The Selection criteria for the particular data was whole data set of the activities of rental bike service and bus card verification for a particular number of days.

### Describing data

The project data are Tartu City bicycles, buses and bicycle stations for the period of the Metallica Concert that held in 2019. The data was sourced from the companies managing the bike sharing service and the public bus transportation service of the city and gathered in CSV format.

The bike data is a combination of two datasets name location and routes. Location dataset has a total of 759,278 observations and 6 features.

SN	Features	Description
1.	route_code	Route code for routes traveled by the bicycle
2.	cyclenumber	Bicycle unique identification number
3.	latitude	Latitude coordinate
4.	longitude	Longitude coordinate
5.	coord_date	Datestamp for the coordinate
6.	coord_time	Timestamp for the coordinate

The route dataset has a total of 236642 observations over four days for the period and 14 features.

SN	Features	Description
1.	route_code	Route code for routes traveled by the bicycle
2.	unlockedat	Datestamp for unlock of bicycle for use
3.	unlockedattime	Timestamp for unlock of bicycle for use
4.	lockedat	Datestamp for lock of bicycle after use
5.	lockedattime	Timestamp for lock time of the bicycle after use
6.	coord_time	Timestamp for the coordinate
7.	startstationname	Station where the bicycle was unlocked for use
8.	endstationname	Station where the bicycle was locked after use
9.	rfidnumber	Means of accessing the bicycle
10.	length	Distance covered with the bicycle for single use
11.	DurationMinutes	Duration for which the bicycle was used
12.	CycleType	Category of bicycle
13.	costs	Cost of the usage
14.	Membership	Membership class of user

### Verifying data quality

The data is of good quality with less than 10% missing values in some features.

## Task 4. Planning project

To achieve a proper result for the project, we have derived a few number tasks which needs to be done. Since we are already being provided with the data, we do not have to gather data. The first of few tasks are:

1. Cleaning the data: We must clean the data if there is any noise or NaN. So, for that we must analyze the data in order to decide what to keep and what not to. The data is divided into 4 respective dates, so each member of the project will have to analyze and clean at least one dataset of the respective date.
2. Data inspection: There are two types of dataset available one is with respect to locations of the bikes in every 5 seconds and the other with respect to routes the bikes have taken each day. So, we will do pattern mining and try to derive what was the normal and common features among the data and what are some interesting facts that could help us come with a hypothesis.
3. Frequent pattern mining: Trying to find the frequent routes and docks used on normal days versus the concert day. For that we will divide each day among ourselves.
4. Data visualization: Visualizing the data using pandas and tableau. After analysis we must visualize the data in some form, maybe even create a heat-map or route-map for the bikes dock stations and locations of bikes on the day of the concert and other days.
5. Exploratory analysis: After all the tasks we will try to form questions from all the information and analysis. Divide those questions among ourselves and come up with answers inferring from all the tasks. And select to keep those questions which would be more useful for the future.