

Data Analysis and Visualization of Hangzhou Metro Traffic

Project for Principles and Practice of Problem Solving

1 Introduction

There is a surge in urban metro traffic in China in the past couple years, which causes overcrowding in metro systems of many cities. In this project, you will be asked to analyze the traffic data obtained from Hangzhou metro system and visualize it to gain a better understanding of the underlying passenger flow patterns. We hope your work could potentially help addressing the aforementioned issue.

2 Dataset Description

The dataset is a real-world automated fare collection (AFC) dataset of Hangzhou metro system downloaded from [Tianchi Competition](#). We excerpted the data from January 7, 2019 to January 13, 2019. Each record contains fields shown in Table 1.

Table 1. Samples of Hangzhou AFC dataset.

fields	type	description	example
time	String	In/out time	2019-01-07 19:58:27
lineID	String	ID of subway lines	C
stationID	int	43	ID of stations
deviceID	int	2002	ID of card reader
status	int	Out=0, In=1	1
userID	String	ID of users	B2ecb7b23922956b70bed08426b45d9a6
payType	int	Card type of users	1

The dataset available contains 3 subway lines with a total of 81 stations. Take the first row of Table 2 as an example, a passenger entered station 43 of line C and swiped his card at 19:58:27 on January 7th, 2019. More detailed description of the dataset can be found in [Tianchi Competition](#).

Table 2. Samples of Hangzhou AFC dataset.

time	lineID	stationID	deviceID	status	userID	payType
2019-01-07 19:58:27	C	43	2002	1	B2ecb7b23922956b70bed08426b45d9a6	1
2019-01-07 09:14:43	A	73	3330	0	Aa28ed7df81394a01313302b8b883b817	0
2019-01-07 09:15:54	B	15	782	1	D32a644707664c5290bd7bb5d3c101013	3
2019-01-07 16:52:43	B	8	387	1	C1f3ff2039d69f8f121d50b5017639fd1	2

3 Tasks

3.1 Mandatory Part

Design a graphic user interface, which

- 1) allows the user to **load the dataset**. Implement a **filter** to import the selected fields of user's interest.
- 2) allows the user to select a station and plot the trend of traffic inflow and outflow over time. The user should be able to **tune the parameters** including starting time, ending time, time step, station ID, and etc. graphically. **Interpolation** is required to smooth the line when only a few data points are available.
- 3) allows the user to input the entering station ID and exiting station ID, and then plan one or more **possible routes** for him/her. Hint: With the adjacency matrix of stations, we can infer the possible travel routes of specific passengers.

3.2 Elective Tasks

Implement additional analysis by yourself. Please NOTE that we won't limit views from which data is counted and analyzed. Use your knowledge and freely exert your creativity.

4 Submission Requirements

The project will graded on the basis of your **project report**, **video demo** and the **source codes**. You need to zip and upload these files to the course website. It is due by **11:59pm on December 29th, 2019. No late submission will be accepted.**

4.1 Report Requirements

1. No more than 5 pages. No formatting requirements.
2. Should at least contain
 - a) **Introduction.** Describe your main ideas and what you do in this project.
 - b) **Implementation Details.**
 - c) **Results.**
 - d) **Discussions.** You could discuss the performance of your application or discuss the interesting results you have revealed from the data.
3. Report is in English only.

4.2 Demo Requirements

1. Make a video to present your work by screen recording or camera shooting.
2. No more than 5 minutes.
3. Presentation is in English only.

Examples of screen recording:

- [YouTube Demo](#)
- [Bilibili Demo](#)

Examples of camera shooting:

- [Presentation Demo](#)

4.3 Source Code Requirements

1. Add a README file, which includes instructions to run your codes.
2. Make sure it runs smoothly on your own computer.