

INDIANA UNIVERSITY BLOOMINGTON

CSCI B 565

DATA MINING

Data Analytics for IU Bus System

Group Name : RedMiners

Authors:

NAGARAJAN, GANESH
MADHAVAN,
SARVOTHAMAN
PAGADALA,
SIDDHARTHA

Supervisor:

DR. MEHMET,
DALKILIC

December 17, 2015



INDIANA UNIVERSITY
BLOOMINGTON

Abstract

University Bus systems are widely used for shuttling students, faculty and patrons in and out of University campuses and still remains the easiest way to reach the University Campuses. Indiana University operates its bus system under the name IU Bus and the details are available at http://www.iubus.indiana.edu/campus_bus/index.html

This paper attempts to develop an uniform framework and processes for analyzing IU Bus data and thus along the process hopes to establish a general framework. Any such framework should be able to transform practical nuances into presentable data analytics. The areas focused in this paper consider various aspects of discussion with the key stake holders, developing glossary, developing metrics for measuring performance, data pipe lining, data visualization, Model creation and model verification.

This project uses MySql as its preferred database for storing operational data, Neo4j graphical database for holding aggregated data, Tableau Public for visualizing the data and R for model creation and verification.

Contents

3	1	Introduction	
4	2	Problem Description	
	2.1	Presentation of case study	4
	2.2	Mathematical formulation	4
4	3	Data Description	
	3.1	Existing model	4
	3.2	Proposed model	4
4	4	Computational Implementation	
4	5	Visualizations	
4	6	Proposed Changes	
4	7	Concluding remarks	
4	8	References	

1 Introduction

University Bus systems are critical for shuttling students in and out of university campuses. These bus systems are not only used at Indiana University, but also at other universities like University Transit Service, University of Virginia <http://www.virginia.edu/parking/uts/>, The Bus, Boston University system <http://www.bu.edu/thebus/>. All these universities have similar kind of schedules for different terms like Spring, Fall and special trips during late nights, called *NightOwl* in the IU System. All these services have an live tracking system which records the bus statuses, Geo-locations of buses and their stops and is presented real-time to the users.

IU Bus system has four routes A,B,E and X and has about 52 stops including the minor and major stops. These busses are scheduled about one bus in every 15 minutes for use of its passengers. IU Bus System commits itself to keep up the time as per the schedule at http://www.iubus.indiana.edu/campus_bus/bus_schedule.html, however as with any transportation system, the bus gains or loses speed during the course of the travel.

Thus, primary interest of this study is to understand the time difference between the schedule time and the actual time, $t_{schedule} - t_{actual}$. Also, it is imperative for any bus system administration to

2 Problem Description

2.1 Presentation of case study

2.2 Mathematical formulation

3 Data Description

3.1 Existing model

3.2 Proposed model

4 Computational Implementation

5 Visualizations

6 Proposed Changes

7 Concluding remarks

8 References