IU Bus Route Optimization
Group Name: Data Driven Unicorns

Authors:
Amruta Chavan
Chirag Chheda
Maria Soledad Elli
Tejashree Khot

Supervisor:
Mehmet M. Dalkilic

December 18, 2015
Contents

1 Introduction 3

2 Problem Description 4
  2.1 Case study ........................................ 4
  2.2 Mathematical formulation ......................... 5

3 Data Description 6
  3.1 Existing model .................................... 6
  3.2 Proposed model .................................... 6

4 Computational Implementation 7
  4.1 Schedule time fluctuations ........................ 7
  4.2 Average travel time between two stops and Average arrival time 8

5 Visualization 8
  5.1 Delay at stops throughout the day .................. 8
  5.2 Delay by weather .................................. 10
  5.3 Delay by passenger count .......................... 11
    5.3.1 By hour ...................................... 11
    5.3.2 By day ....................................... 12
    5.3.3 At major stops ................................. 13
    5.3.4 By weather .................................... 16

6 Proposed Changes 16

7 Concluding remarks 19
Abstract

Whenever an organization providing mobility is entrusted with aiding people, a question that comes to mind is how effectively can such services be made available. This basic requirement of efficient mobility of passengers gives rise to, among many other things, the subject areas of optimal routing and scheduling. In this project, we work on improving the existing transit system made available to the students of Indiana University, Bloomington.

We work on a year long dataset that is provided by the IU Student Transportation Board. We begin with preprocessing the data by cleaning it and making it consistent. Later we analyze this information and find different possible ways that will assist in improving the efficiency of the IU Bus transit. We depict the same by providing some insights into the problem and the solution by presenting some captivating visualizations.

By taking into consideration the observations made so far, we put forward few suggestions that can possibly help increase the efficiency of the transportation system at Indiana University and meet the desired goals.
1 Introduction

Indiana University Bloomington is a public research university located in Bloomington, Indiana, United States. With over 40,000 students, IU Bloomington is the flagship institution of the Indiana University system and its largest university. With an incessant flow of students that join IU every year, it becomes a requisite for the university to meet the basic requirements of good transportation for them. This enables the students to travel daily between the university and their place of residence with greater ease and comfort.

While meeting the needs of the students it also becomes necessary to keep a track of how efficiently is the University able to do this. Increasing the efficiency of the transport system leads to an increase in the availability of time for the students. It is always desired that our time should be better utilized and in this project we avail this opportunity to the students of Indiana University, by making their commute fast and hassle free.

In order to do so we make use of various software and tools that assist us in data crunching as we make sense of the data provided. We make use of MS Access as we work on the databases that are made available by Indiana University. The database technologies used are MySQL and MongoDB while the programming languages which we have made use of consists of R and Python.
2 Problem Description

2.1 Case study

In real life scenario, we come across some hurdles which prevent us from achieving an ideal result always. However, one can thrive to minimize the difference between the expected and observed answer. In this project, we focus on increasing the efficiency of the IU Bus transit by providing insights on some situations that lead to anomaly.

The IU Campus Bus helps thousands of students daily by allowing them to travel different parts of the campus throughout the day. There are a total of 4 routes followed by the IU Campus Bus (figures 1, 2), with a total of 27 buses that can serve them. Although the buses follow a fixed schedule for a given semester, there is a slight deviation observed in meeting this assumption. Increase in the passengers riding the bus, change in the climate, the day and time when the bus is running are some of the factors that are thought to have an impact on the schedule of the bus.

![IU Routes A - B](image)

Figure 1: IU Routes A - B

With the help of the dataset at hand, we look into all the possible factors causing such delay. An analysis of this data enables us to gain valuable insights to the problem. With the aid of this dataset, we provide suggestions that can help in improving the existing efficiency of IU Campus Bus.
2.2 Mathematical formulation

The main goals are listed as:

- We wish to meet the expected schedule. If $t_{exp}$ represents expected arrival and $t_{obs}$ represents observed arrival of the bus then we wish to minimize:

  $$t_{exp} - t_{obs}$$

- We also observe the average travel time taken by a bus to traverse between any two pickup stops. We do this using the formula:

  $$\sum_{i=1}^{n} t_{ab_{iw}}/n$$

  Where $t_{ab_{iw}}$ indicates the time taken for the bus to cover the distance between stop $a$ and $b$ during $i$th trip of the bus on week day $w$.

- In addition, we also want to study the effects of various factors on dwell timing, namely, weather and passenger count such as:
  find if there exists a function $f$ such that $f : weather \rightarrow dwell\text{timing}$
  find if there exists a function $g$ such that $g : passenger \rightarrow dwell\text{timing}$
3 Data Description

3.1 Existing model

The existing model is based upon several data collected by the University. It gives us information about various content that have been captured over time. Below is a brief description of the tables that we have:

- We have a table named Stop ID that can be referenced to see which bus stops are assigned what ID in the tables.
- The information present in the Schedule Data table gives information about the expected bus location at a particular time, running on a particular route.
- The Route ID table gives a unique identifier that is associated with all the buses servicing on a given route.
- We have data in the Work Record table which is used to link a bus assignment on doublemap to a schedule number.
- The data present in RidershipSchedule2015 provides information about the passengers that have traveled on a given route over a certain duration. Given a day, it also tells us about the total number of passengers that have traveled on a particular route.
- The Servicebook Draft speaks about the expected schedule of all the buses for a given route.

3.2 Proposed model

In order to achieve the goals described before in section problems, the data analysis was made with certain constrains.

The first one is that we are considering the data for the Spring 2015 semester. This is because some data for the Fall semester 2014 is missing, like Work Record. Secondly, we are considering weather data in degrees such as 0 for good weather (with no precipitations) or on a scale of 1 to 3 for bad
weather (like rain, snow, etc).

Finally, we have been able to arrange a table that aggregates all the necessary information. This table contains information regarding:

- The bus stop ID and it’s respective name.
- Dwell time, Scheduled time and Real time.
- The time delay in seconds and absolute time difference in minutes.
- Sum of inbound and outbound at that particular sleep.
- Daily information regarding the weather conditions and the day of the week.

4 Computational Implementation

Building the model proposed before required a lot of data processing and we chose to do it outside the DB, using Python and R, as we thought this was more convenient. In the following sections we will describe the processes carried out for merging all the data together in order to build the model described above.

4.1 Schedule time fluctuations

In contemplation of the difference between the scheduled times and the real times each bus is reaching a designated stop, we matched the Intervaldata2014-2015 table and the Work Record table by "bus id" taking into account the time in column "when" from the first table must be between the respective values of "Date Clock In" and "Date Clock Out" columns. After this, we combined this new joined table by Route and by time with the Schedule Data table and calculate the difference between the matched scheduled and real times.

Moreover, for simplicity, we match every row depicting behavior of bus at stops with more information about it’s trip. Now we also know for which trip of which day the instance belongs. We also match each row with passenger counts of the trip and weather during the day.
Now from this table, we can have data about for each time any bus was at major. From this we can relations between delays, dwell times, stops, trips, routes, days, passenger counts and weather.

4.2 Average travel time between two stops and Average arrival time

We have calculated average travel time between any two major stops. We made a decision to restricted ourselves to only major stops mostly because they are most important for the scheduling purposes. Moreover, we also derived a table that averages the arrival time of buses for particular trips on particular day of the week.

5 Visualization

We are presenting the analysis of route A in order to determine the factors that affect the schedule time of that route. This analysis is scalable to the remaining routes as well. The plot tool we used throughout this project is plotly for R (https://plot.ly/plot/)

5.1 Delay at stops throughout the day

The graphs in figures 3, 4, and 5 reflects the delays of route A for different trips on different days. We found that the delay is increasing from Well’s Library to 3rd & Jordan and to IMU. This shows that the delay is steadily increasing through the trip. Hence, for most trips the actual travel time is longer than schedule. A less optimistic schedule that does not consistently underestimate delay should be created.
Figure 3: 3rd & Jordan

Figure 4: Indiana Memorial

Figure 5: Well’s Library
5.2 Delay by weather

The graph in figure 6 is a boxplot of delay distribution for different weather conditions. The median of delay falls almost always close to 0. The 2nd, 98th, 25th percentile and 75th percentile for all distributions of delay with respect to weather are also almost same. This goes to prove that delays are independent of weather.

![Figure 6: Weather implications in delay](image-url)

Figure 6: Weather implications in delay
5.3 Delay by passenger count

5.3.1 By hour

A visualization of the amount of passengers using the bus during the day is shown in figure 8. This indicates the behavior of students going to and from classes during day. As the graph shows, certain times of the day, like Thursday around 10:45 am, are more critical than others. We can see from this graph how the class schedule is reflected and modifies the number of passengers taking the bus.

Figure 7: Delay and passenger count by days

Figure below shows the k-means algorithm run on the number of passengers vs the dwell time shows that there has been an exponential rise in the dwell time as the number of passengers increase. You can see how the dwell time shoots up in cluster number 4 as compared to the other 3 clusters. An increase in dwell time increases the travel time required to reach the destination, thus affecting the expected schedule. This adds up to the point made by figure 7, that the time required to cover the distance between two given stops is indeed influenced by the increase in passenger count.
5.3.2 By day

In this subsection the graphs to represent Passenger Count and Time Difference over interval of 24Hrs on any day of the week will be presented:

![Figure 9: Delay and passenger count on Monday](image)

Inferences:
As per the daily Time Delay and Passenger count figures 9, 10, 11, 12, we can draw following conclusions:

- The Average Time delay fluctuations tend to reduce as the week progresses.
- High peaks for passenger count and successively Time delay can be seen in all the plots depicting that passenger count increases during
scheduled class timings which results into increasing dwelling time at the stops. This contributes to increased delay time during the further trip.

- The Inbound Passenger count is higher in the morning while Outbound is higher in the evening.

### 5.3.3 At major stops

The following figures 13, 14, 15, and 16 shows the delay and passenger count at major stops of route A.

Inferences:

- The time delay fluctuations are maximum at the 1st and the last major stop where maximum passengers aboard the bus.
Figure 12: Delay on Thursday

Figure 13: Delay and passenger count at Stadium

- Also Well’s Library being centrally located major stop maximum passenger strength
Figure 14: Delay and passenger count at Indiana Memorial Union

Figure 15: Delay and passenger count at Wells Library

Figure 16: Delay and passenger count at 3rd & Jordan
5.3.4 By weather

The diagram in figure 17 reflects the implication with respect to Weather, where 0 is very good weather and 3 is very bad weather. As per this the passenger counts increase slightly when the weather gets slightly bad (only snow or rain). This may display that people prefer to use buses under these circumstances (Snow or rain) and buses needs to cater to small amount of extra people. The buses don’t get these extra crowds in more extreme weathers.

This is disappointing because we have previously established that buses work well in bad weather conditions. We expect people to rely more on buses in these bad weather conditions. However, this does not happen indicating a possible image issue.

Figure 17: Delay and passenger count driven by weather

6 Proposed Changes

As we can see in the figure 18, the four bus routes are distributed through IU campus to provide students who live in the main residential halls with a transportation system that allows them to go to class easily. Since this is the main goal of the bus system, we do not think we should delete any route
from the current schedules. Although, a rescheduling is necessary in order to improve the travel time of the buses and, as result, increase the student’s satisfaction towards the bus system.

As seen from the visualization above, in the graph of students traveling in the bus during different time of the day, we notice sharp peaks during some particular time frames. Such an increase in the number of students traveling via the IU Campus bus during some peculiar timings is observed consistently throughout the week. We propose to increase the frequency of buses during this time as it might lead to a decrease in the dwell time and help the students commute with greater comfort.

Another picture, which is obtained after performing k-means, on the dwell time vs the number of passengers confirms the supposition that with an increase in the number of passengers in the bus, there is an exponential rise in the dwell time. An increase in the number of buses will eventually even help in improving this situation. An alternative way to achieve this is to find a faster way to allow the passengers to board and alight the bus, thus reducing the dwell time.
In order to meet the expected schedule we put forward the following suggestions:

- Work along with the University to obtain the class enrollment data in order to predict the time zones where there is a comparatively higher demand for transportation by the students to reach a particular destination on the campus. Schedule the frequency of buses based on this factor.

- Schedule the arrival time of a bus based on the observed average arrival time. This will lead to a decrease in the difference between the expected and observed schedule of the bus.

We also derived two additional tables from our main table for quick usability. One of these tables contains average travel times for between two major stops in a route for specific trip on a specific day of a week. Then we also derived another table that holds average arrival time of bus at major stops of a route running a for specific trip on a specific day of a week. These table could aid for planning schedules.
7 Concluding remarks

From data provided to us, we modeled it into a structure that holds information about buses when they reach major stops and its dependencies on passenger count and weather and derived two tables for schedule planning.

We have analyzed this model and found many interesting relations. From these observations, we proposed that while changing entire route is not advisable but we have suggestions for change in schedules. We propose to add an extra bus in peak passenger traffic times which is very consistent enough to be added to a static schedule. Moreover, actual scheduling process can be improved by collaborating with university and getting class schedule from university. Also, we have derived average travel time between any two major stops and average arrival time at any major stop for each trip on every day of the week to help with the planning process.