Indiana University Bloomington

CSCI B 565

Data Mining

Data Analytics for IU Bus System

Group Name : RedMiners

Authors:

Jayasankar,
Siddharth
Nagarajan, Ganesh
Madhavan,
Sarvothaman

Supervisor:
Dr. Dalkilic,
Mehmet

December 17, 2015

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

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 looses speed during the course of the travel.

Thus, primary interest of this study if to understand the time difference between the schedule time and the actual time, $t_{schedule} - t_{actual}$. An model was created in Neo4j depicting the acutal bus system

2 Problem Description

3 Constraints

4 Data Management

4.1 Source Data Model

Text Explanation here

```
DBMAP_StopID
ID : int(11)
Index : double
Stop : varchar(255)
```

```
DBMAP_RouteID
ID : varchar(150)
Index : varchar(150)
Route ID : varchar(150)
Field3 : varchar(150)
```

```
DBMAP_ScheduleData
ID : int(11)
Route : varchar(255)
Time : time
Stop : varchar(255)
```

```
DBMAP_WeatherData
ID : int(11)
EDT : datetime
MinTemp : double
Precipitation : varchar(255)
Events : varchar(255)
```

```
DBMAP_WorkRecord
ID : int(11)
Clock In : time
Clock Out : time
Driver : varchar(150)
Shift Type : varchar(150)
daynum : int(11)
Date : date
Route : varchar(150)
Bus : int(11)
```

```
DBMAP_GPSData
BUS_ID : int(11)
LAT : float
LONGI : float
x : int(11)
y : int(11)
timestamp : varchar(100)
```

```
DBMAP_IntervalData
id : int(10)
from_id : int(10)
to_id : int(10)
time_id : int(10)
bus_id : int(10)
route_id : int(10)
when_time : timestamp
```

Text Explanation here

IU_SPRING_Drivers

Driver : varchar(30)
Driver ID : int(11)

IU_SPRING_Buses

Bus ID : smallint(6)

IU_SPRING_Dates

Date : datetime

IU_SPRING_Schedules

Schedule Key : int(11)
Schedule ID : varchar(10)
Start Times : time

IU_SPRING_ScheduleNames

Schedule Name : varchar(255)

NOAA_Weather

STATION_NAME : varchar(150)
DATE : date
PRCP : int(11)
Measurement Flag : varchar(150)
SNOW : int(11)
TMIN : float
FOG : int(11)
RAIN : int(11)
HEAVY_FOG : int(11)
THUNDER : int(11)

IU_SPRING_Shifts

Shift ID : int(11)
Date : datetime
Schedule Name : varchar(10)
Driver : varchar(30)
Bus ID : smallint(6)

IU_SPRING_Trips

Trip ID : int(11)
Shift ID : int(11)
Schedule ID : varchar(150)
Starting Time : time
Inbound : int(11)
Outbound : int(11)
To Mall : int(11)
From Mall : int(11)
Comments : varchar(500)

4.2 Intermediate Model

INTER_FIRST_STOP

SCHEDULE_ID : mediumtext
SCHEDULE_TIME : time
STOP_ID : int(11)
STOP_ORDER : double
DOW : mediumtext
MIDPOINT : int(11)
UP_DOWN : bigint(20)
TRIP_ID : int(11)

INTER_LAST_STOP

SCHEDULE_ID : mediumtext
SCHEDULE_TIME : time
STOP_ID : int(11)
STOP_ORDER : double
DOW : mediumtext
MIDPOINT : int(11)
UP_DOWN : bigint(20)
TRIP_ID : int(11)

INTER_TRIPS_START_END

SCHEDULE_ID : mediumtext
TRIP_ID : int(11)
Date : date
Inbound : int(11)
outbound : int(11)
trip_start_time : time
trip_end_time : time

Text Explanation here

7

INTER_SPRING_TRIPS

schedule name : varchar(10)
dt : date
Starting Time : time
Inbound : int(11)
outbound : int(11)

INTER_RUN_STOP_TRIPS

Route : varchar(150)
Date : date
TripID : longtext
from_id : int(11)
to_id : int(11)
sec_time : int(11)
when_time : timestamp
stop_order : longtext
bus_id : int(11)
id : mediumtext
integration_key : mediumtext
travel_flag : bigint(20)

INTER_RUN_START_TRIPS

Route : varchar(150)
Date : date
TripID : longtext
from_id : int(11)
to_id : int(11)
sec_time : int(11)
when_time : timestamp
stop_order : longtext
bus_id : int(11)
id : mediumtext
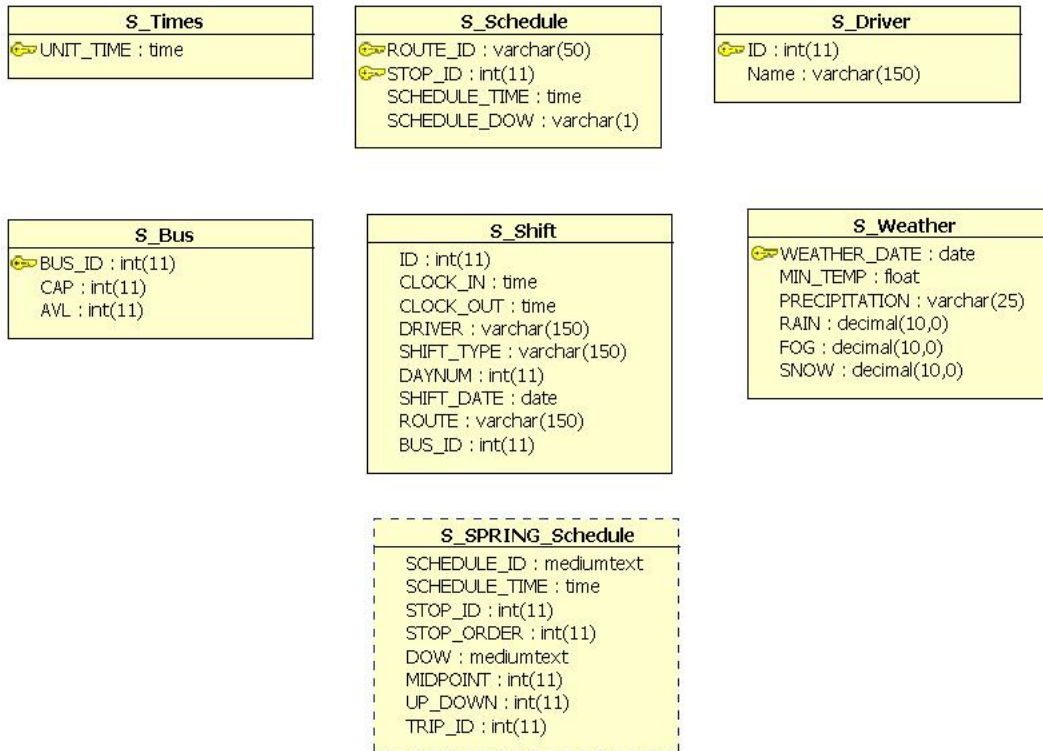integration_key : mediumtext
travel_flag : bigint(20)

INTER_RUNS_START_END

start_time : time
Date : date
Route : varchar(6)
TripID : double
end_time : time

Text Explanation here

8

4.3 Final Data Model

Text Explanation here

S_Times

🔑 UNIT_TIME : time

S_Schedule

🔑 ROUTE_ID : varchar(50)
🔑 STOP_ID : int(11)
SCHEDULE_TIME : time
SCHEDULE_DOW : varchar(1)

S_Driver

🔑 ID : int(11)
Name : varchar(150)

S_Bus

🔑 BUS_ID : int(11)
CAP : int(11)
AVL : int(11)

S_Shift

ID : int(11)
CLOCK_IN : time
CLOCK_OUT : time
DRIVER : varchar(150)
SHIFT_TYPE : varchar(150)
DAYNUM : int(11)
SHIFT_DATE : date
ROUTE : varchar(150)
BUS_ID : int(11)

S_Weather

🔑 WEATHER_DATE : date
MIN_TEMP : float
PRECIPITATION : varchar(25)
RAIN : decimal(10,0)
FOG : decimal(10,0)
SNOW : decimal(10,0)

S_SPRING_Schedule

SCHEDULE_ID : mediumtext
SCHEDULE_TIME : time
STOP_ID : int(11)
STOP_ORDER : int(11)
DOW : mediumtext
MIDPOINT : int(11)
UP_DOWN : int(11)
TRIP_ID : int(11)

Text Explanation here

W_RUNS_F
id : mediumtext
integration_key : varchar(35)
Route : varchar(150)
time_date : date
trip_id : longtext
STOP1_NAME : varchar(255)
STOP2_NAME : varchar(255)
STOP2_TIME : time
TRAVEL2_TIME : int(11)
STOP3_NAME : varchar(255)
STOP3_TIME : time
TRAVEL3_TIME : int(11)
STOP4_NAME : varchar(255)
STOP4_TIME : time
TRAVEL4_TIME : int(11)
STOP5_NAME : varchar(255)
STOP5_TIME : time
TRAVEL5_TIME : int(11)
STOP6_NAME : varchar(255)
STOP6_TIME : time
TRAVEL6_TIME : int(11)
STOP7_NAME : varchar(255)
STOP7_TIME : time
TRAVEL7_TIME : int(11)
STOP8_NAME : varchar(255)
STOP8_TIME : time
TRAVEL8_TIME : int(11)
STOP9_NAME : varchar(255)
STOP9_TIME : time
TRAVEL9_TIME : int(11)
STOP10_NAME : varchar(255)
STOP10_TIME : time
TRAVEL10_TIME : int(11)
▼

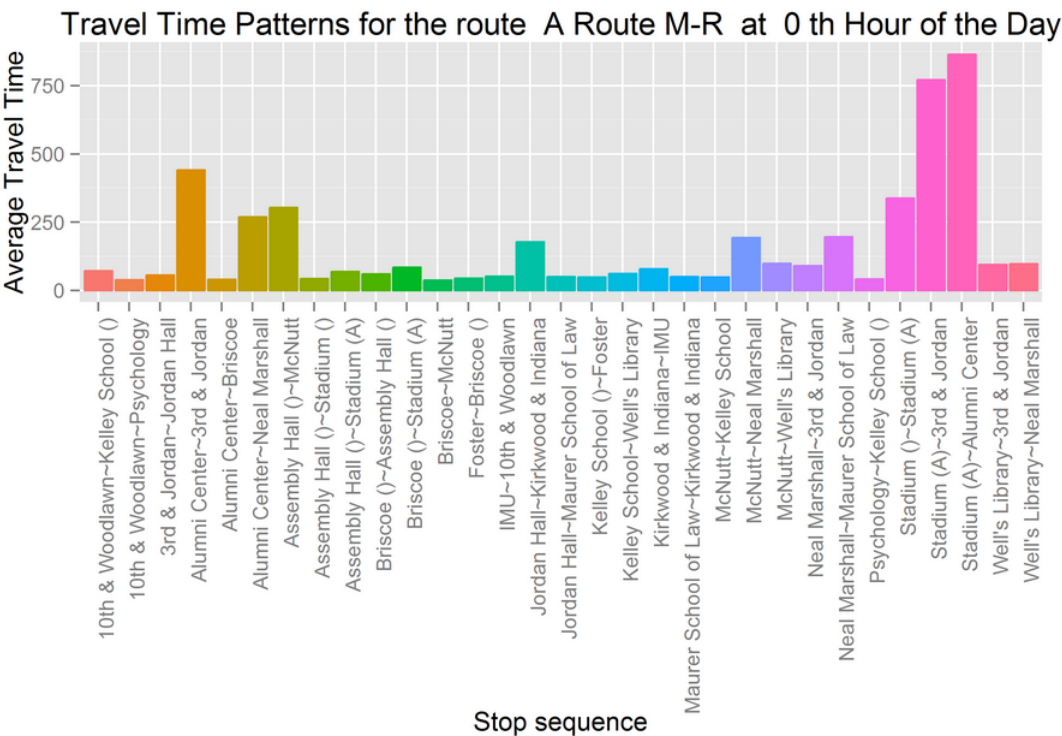| W_SHED_D |
|---|
| integration_key : varchar(35) |
| schedule_id : mediumtext |
| trip_id : int(11) |
| DOW : mediumtext |
| STOP1_NAME : varchar(255) |
| STOP1_TIME : time |
| STOP_1ID : int(11) |
| STOP2_NAME : varchar(255) |
| STOP2_TIME : time |
| STOP2_ID : int(11) |
| STOP3_NAME : varchar(255) |
| STOP3_TIME : time |
| STOP3_ID : int(11) |
| STOP4_NAME : varchar(255) |
| STOP4_TIME : time |
| STOP4_ID : int(11) |
| STOP5_NAME : varchar(255) |
| STOP5_TIME : time |
| STOP5_ID : int(11) |
| STOP6_NAME : varchar(255) |
| STOP6_TIME : time |
| STOP6_ID : int(11) |
| STOP7_NAME : varchar(255) |
| STOP7_TIME : time |
| STOP7_ID : int(11) |
| STOP8_NAME : varchar(255) |
| STOP8_TIME : time |
| STOP8_ID : int(11) |
| STOP9_NAME : varchar(255) |
| STOP9_TIME : time |
| STOP9_ID : int(11) |
| STOP10_NAME : varchar(255) |
| STOP10_TIME : time |
| STOP10_ID : int(11) ▼ |

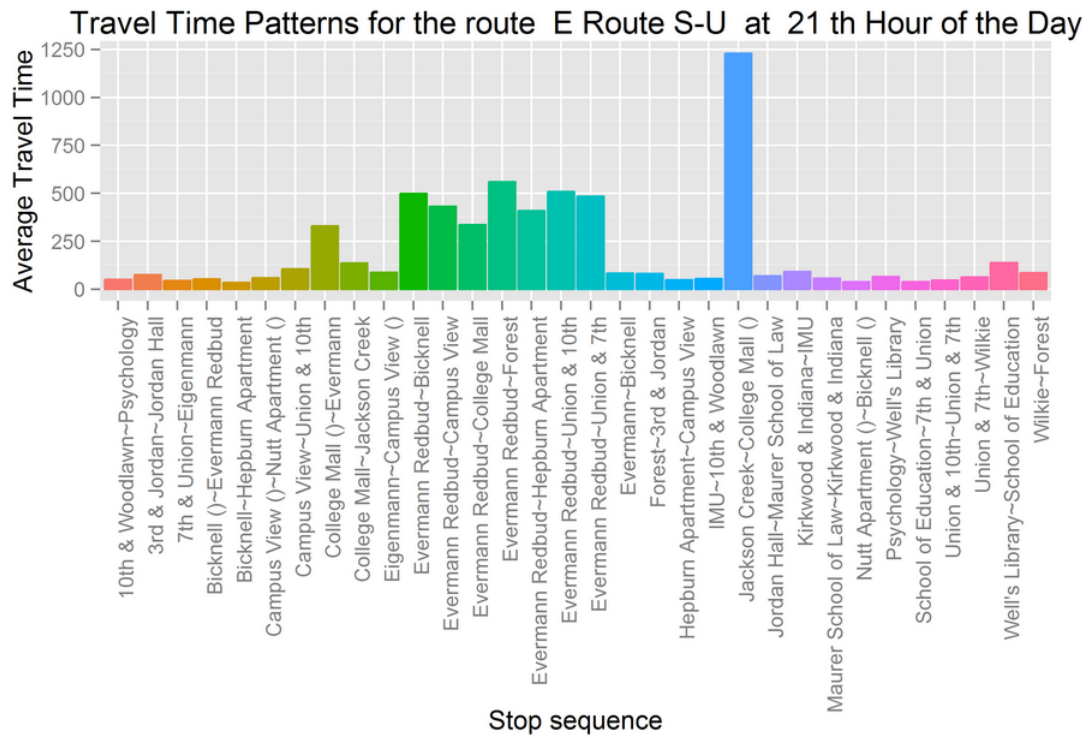| F_TRIPS_COMPARED |
|---|
| SCHEDULE_ID : mediumtext |
| TRIP_ID : int(11) |
| Date : date |
| Inbound : int(11) |
| outbound : int(11) |
| schedule_start : time |
| actual_start : time |
| schedule_end : time |
| actual_end : time |

4.4   Data Preparation
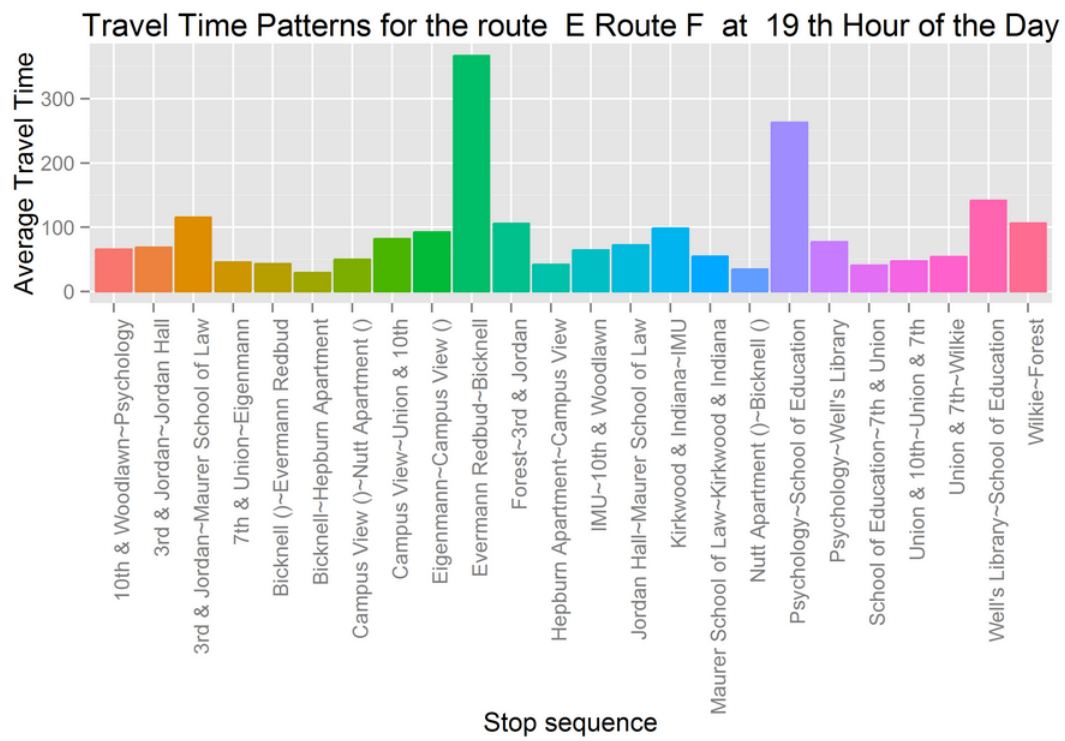
Text Explanation here

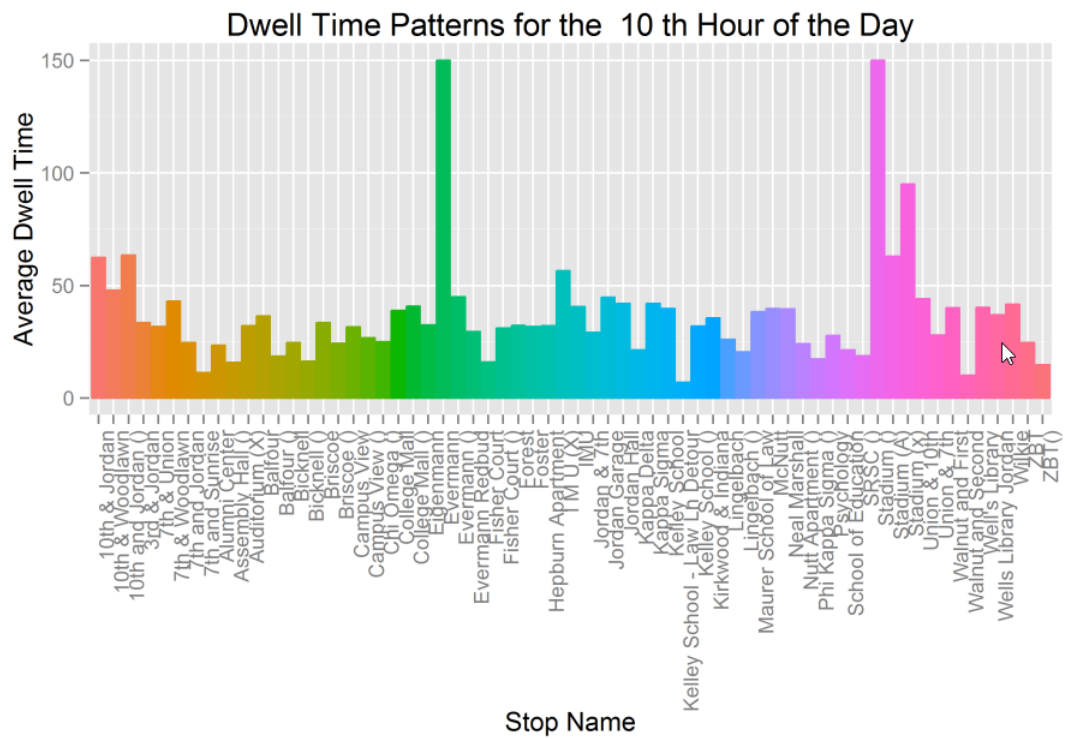# 5   Exploratory Analysis
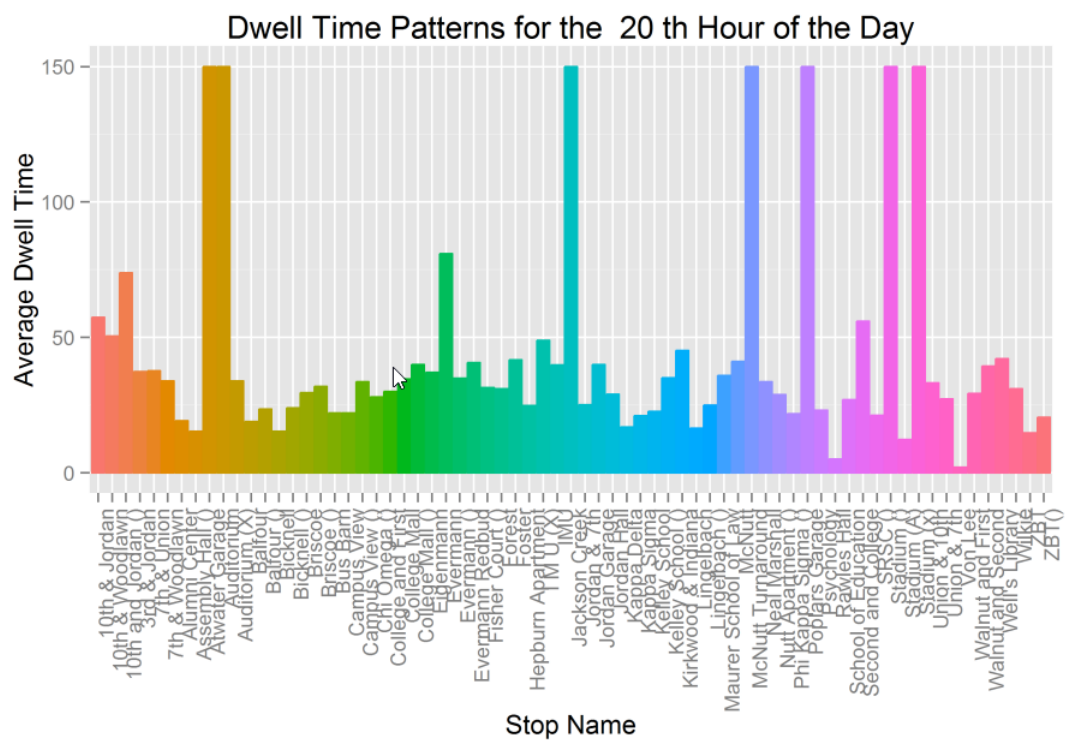
Text Explanation here

Text Explanation here

Travel Time Patterns for the route E Route S-U at 21 th Hour of the Day

Text Explanation here

Travel Time Patterns for the route E Route F at 19 th Hour of the Day

Text Explanation here

Dwell Time Patterns for the 10 th Hour of the Day

Text Explanation here

Dwell Time Patterns for the 20 th Hour of the Day

6   Reporting

Text Explanation here

Text Explanation here

Stadium

Date	Route	Trip #	Stop Name	Scheduled Time	Actual Time	
3/3/2015	T-A1.1	3	Stadium (A)	08:20:00	08:20:47	-47
3/10/2015	T-A1.1	3	Stadium (A)	08:55:00	08:37:42	1,038
3/24/2015	T-A1.1	3	Stadium (A)	08:20:00	08:05:30	870

Wells Library

Date	Route	Trip #	Stop Name	Scheduled Time	Actual Time	
3/3/2015	T-A1.1	3	Well's Library	08:27:00	08:25:25	95
3/10/2015	T-A1.1	3	Well's Library	09:02:00	09:03:48	-108
3/24/2015	T-A1.1	3	Well's Library	08:27:00	08:25:02	118

3rd and Jordan

Date	Route	Trip #	Stop Name	Scheduled Time	Actual Time	
3/3/2015	T-A1.1	3	3rd & Jordan	08:29:00	08:28:39	21
3/10/2015	T-A1.1	3	3rd & Jordan	09:04:00	09:07:18	-198
3/24/2015	T-A1.1	3	3rd & Jordan	08:29:00	08:28:46	14

IMU

Date	Route	Trip #	Stop Name	Scheduled Time	Actual Time	
3/3/2015	T-A1.1	3	IMU	08:39:00	08:34:05	295
3/10/2015	T-A1.1	3	IMU	09:14:00	09:12:58	62
3/24/2015	T-A1.1	3	IMU	08:39:00	08:33:30	330
3/31/2015	T-A1.1	3	IMU	09:14:00	09:12:22	98

Stadium Return

Date	Route	Trip #	Stop Name	Scheduled Time	Actual Time	
3/3/2015	T-A1.1	3	Stadium ()	08:47:00	08:39:47	433
3/10/2015	T-A1.1	3	Stadium ()	09:22:00	09:20:10	110
3/24/2015	T-A1.1	3	Stadium ()	08:47:00	08:39:52	428
3/31/2015	T-A1.1	3	Stadium ()	09:22:00	09:20:30	90

Month
- [ ] (All)
- [ ] January
- [ ] February
- [x] March
- [ ] April

Date
1/13/2015 — 4/30/2015

Route
- [ ] (All)
- [ ] M-A1.1
- [ ] M-A2.1
- [ ] M-A3.1
- [ ] M-A4.1
- [ ] M-A5.1
- [ ] R-A1.1
- [ ] R-A2.1
- [ ] R-A3.1
- [ ] R-A4.1

Trip #
- [ ] (All)
- [ ] 2
- [x] 3
- [ ] 4
- [ ] 5
- [ ] 6
- [ ] 7
- [ ] 8
- [ ] 9

Text Explanation here

Logistic Report

Trip Id	Route	Stop 1 Status	Stop 2 Status	Stop 3 Status	Stop 4 Status	
February 2, 2015, M, 1	M-A3.1	Delayed	On Time	Early	Delayed	9
	M-A4.1	Early	Early	Early	Early	9
	M-A5.1	Early	On Time	Early	Delayed	9
February 2, 2015, M, 2	M-A1.1	Delayed	Delayed	On Time	Delayed	8
	M-A3.1	Delayed	Delayed	On Time	Delayed	10
	M-A4.1	Delayed	On Time	Early	Delayed	9
	M-A5.1	On Time	On Time	Early	Delayed	10
February 2, 2015, M, 3	M-A1.1	On Time	On Time	Early	Delayed	8
	M-A3.1	Delayed	Early	Early	Early	10
	M-A4.1	Delayed	Early	Early	Early	10
	M-A5.1	Delayed	Early	Early	Early	11
February 2, 2015, M, 4	M-A1.1	Delayed	Early	Early	Early	9
	M-A3.1	Delayed	Delayed	Delayed	Delayed	11
	M-A4.1	Delayed	On Time	On Time	Delayed	11
	M-A5.1	Delayed	Early	Early	Early	11
February 2, 2015, M, 5	M-A1.1	Delayed	Delayed	Delayed	Delayed	10
	M-A3.1	Delayed	Delayed	On Time	Delayed	12
	M-A4.1	Early	Early	Early	Early	12
	M-A5.1	Delayed	Early	Early	On Time	12
February 2, 2015, M, 6	M-A1.1	Delayed	On Time	On Time	Delayed	11
	M-A3.1	Delayed	On Time	Early	Early	12
	M-A4.1	Delayed	Early	Early	On Time	12
	M-A5.1	Delayed	On Time	On Time	Delayed	13
February 2, 2015, M, 7	M-A1.1	Data Missing	On Time	Early	Early	11
	M-A3.1	Delayed	Delayed	Delayed	Delayed	13
	M-A4.1	Delayed	Delayed	On Time	Delayed	13
	M-A5.1	Delayed	Early	Early	Early	13

Hr Day
7 — 15

Month Name
- [ ] (All)
- [ ] January
- [x] February
- [x] March
- [x] April

Route
- [x] (All)
- [x] M-A1.1
- [x] M-A2.1
- [x] M-A3.1
- [x] M-A4.1
- [x] M-A5.1
- [x] R-A1.1
- [x] R-A2.1
- [x] R-A3.1
- [x] R-A4.1
- [x] R-A5.1
- [x] T-A1.1
- [x] T-A2.1
- [x] T-A3.1
- [x] T-A4.1
- [x] T-A5.1
- [x] W-A1.1
- [x] W-A2.1
- [x] W-A3.1
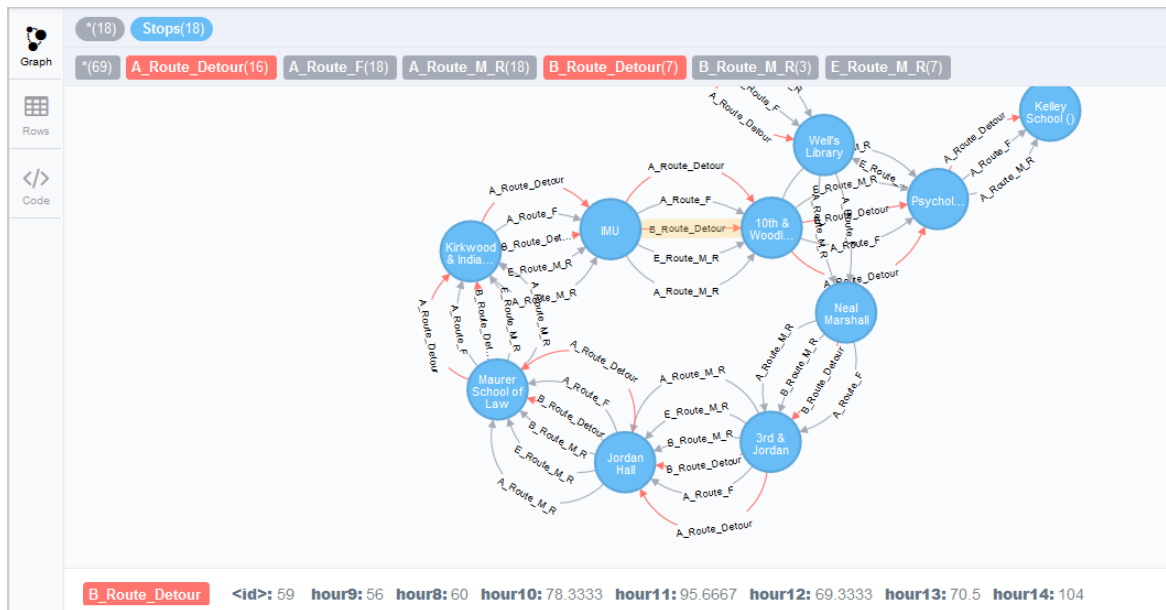- [x] W-A4.1
- [x] W-A5.1
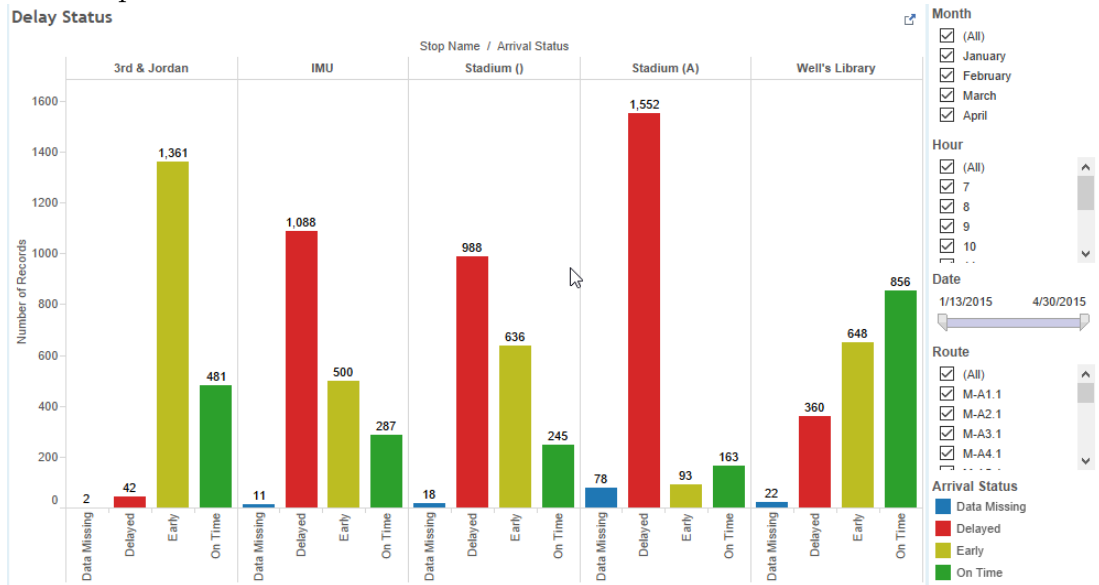
7   Visualization

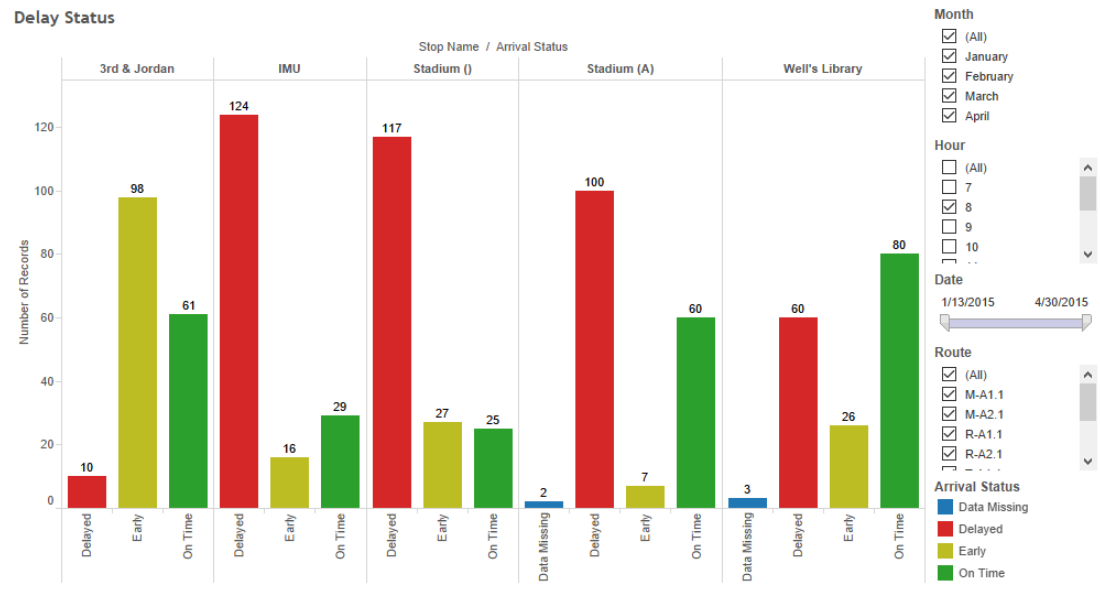Text Explanation here
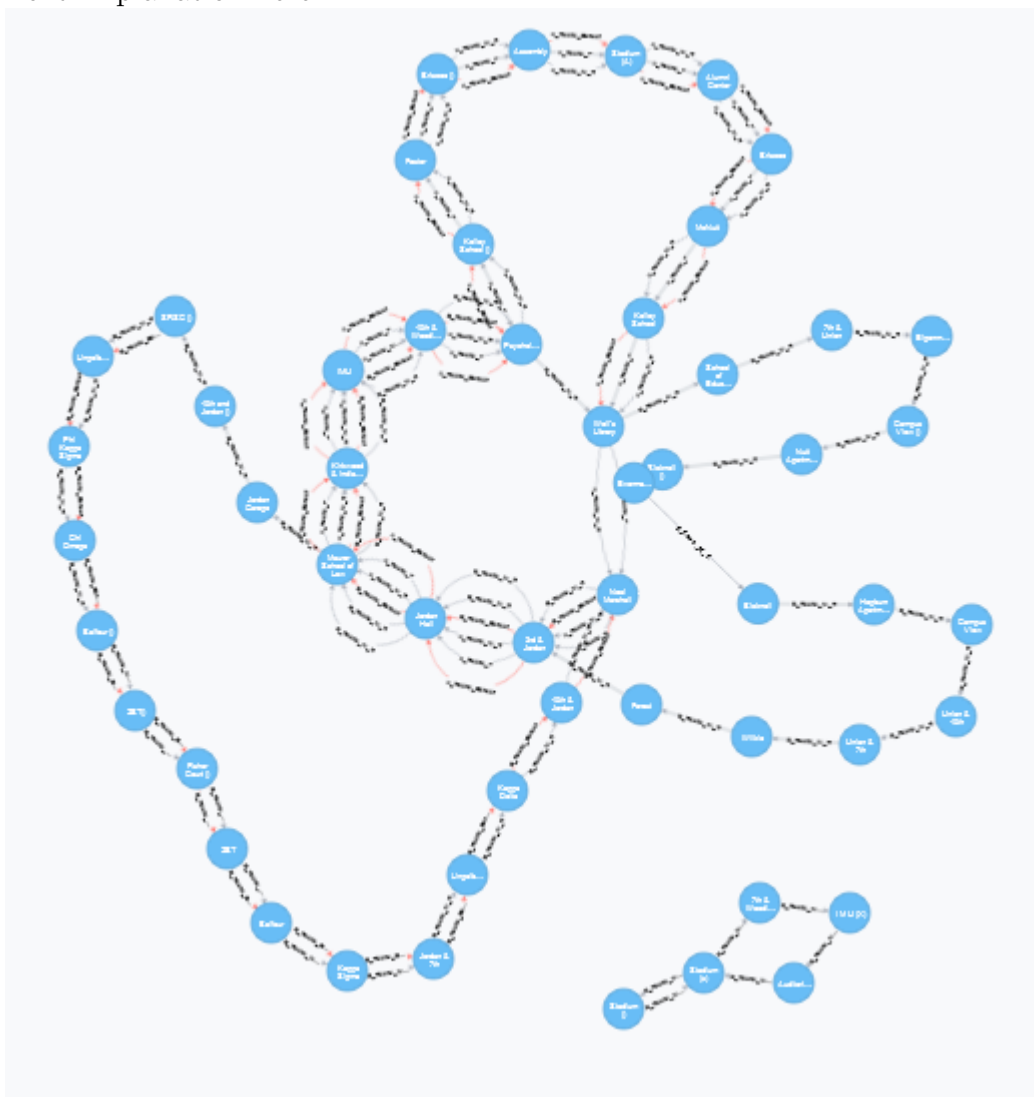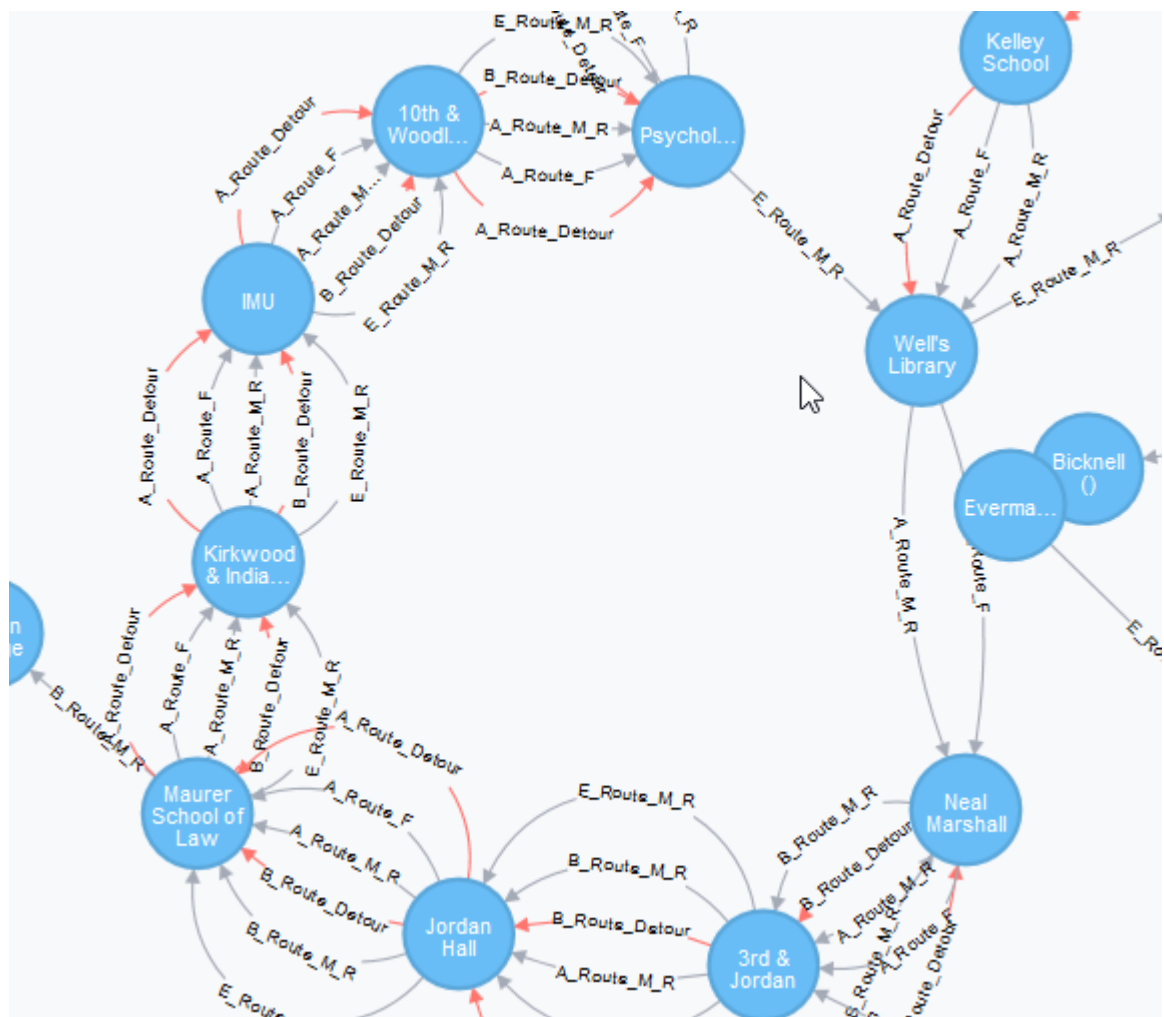
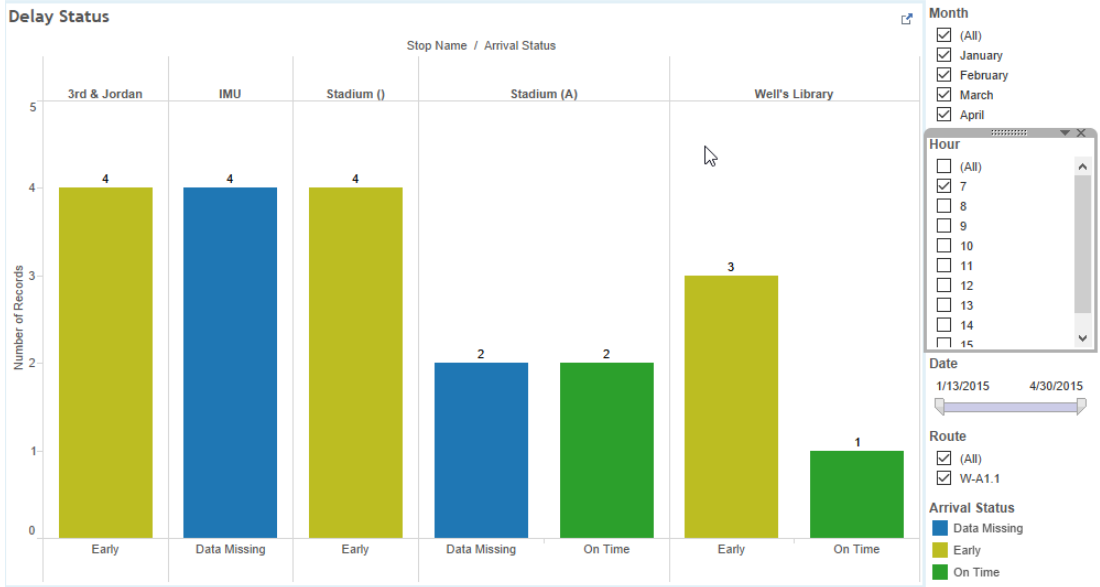
Text Explanation here

Text Explanation here

Text Explanation here

Text Explanation here

Text Explanation here

8   Statistical Model

9   Future Work

10   Concluding remarks

11   References