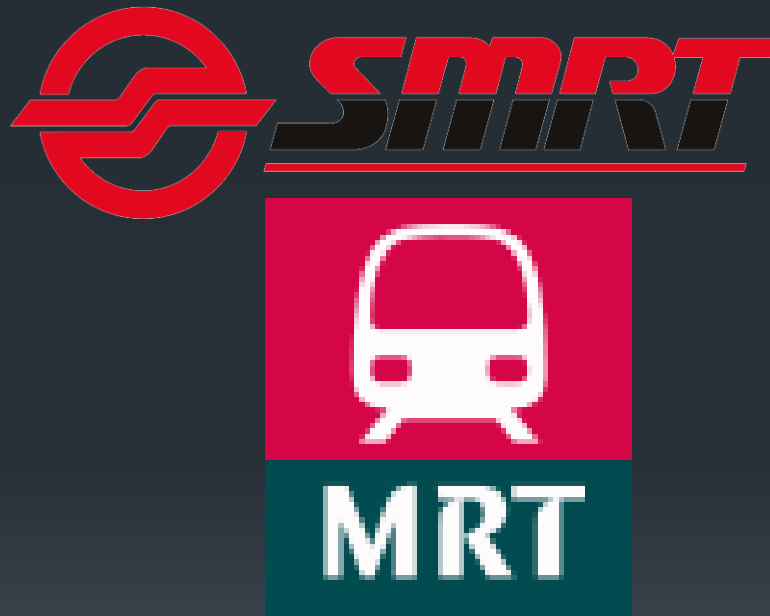


ANLY482 - Analytics Practicum

Computational Transportation Science



Group 6

Koh Ying Ying Trecia

Luqman Haqim Bin Ab Rahman

Project Sponsor / Supervisor



- Associate Professor of Information Systems
- Senior Advisor, SIS Programmes in Analytics



Data Source

- Land Transport Authority (LTA) provides the dataset which is a weeks' worth of smart card (Ezlink) transactions used in Singapore's public transport.
- It is made available for LARC research initiative
- Data consist of the different tables:
 - Bus_service_mapping
 - Location_gis_mapping
 - Location_mapping
 - Lta_ride
 - LTA ride data subset (1 November 2011 – 6 November 2011)

Roles and Responsibility



Luqman



Trecia Koh

- Data Preparation
- Data Analysis & Reporting
 - Wiki
- Presentation slides
 - Final Report

Introduction

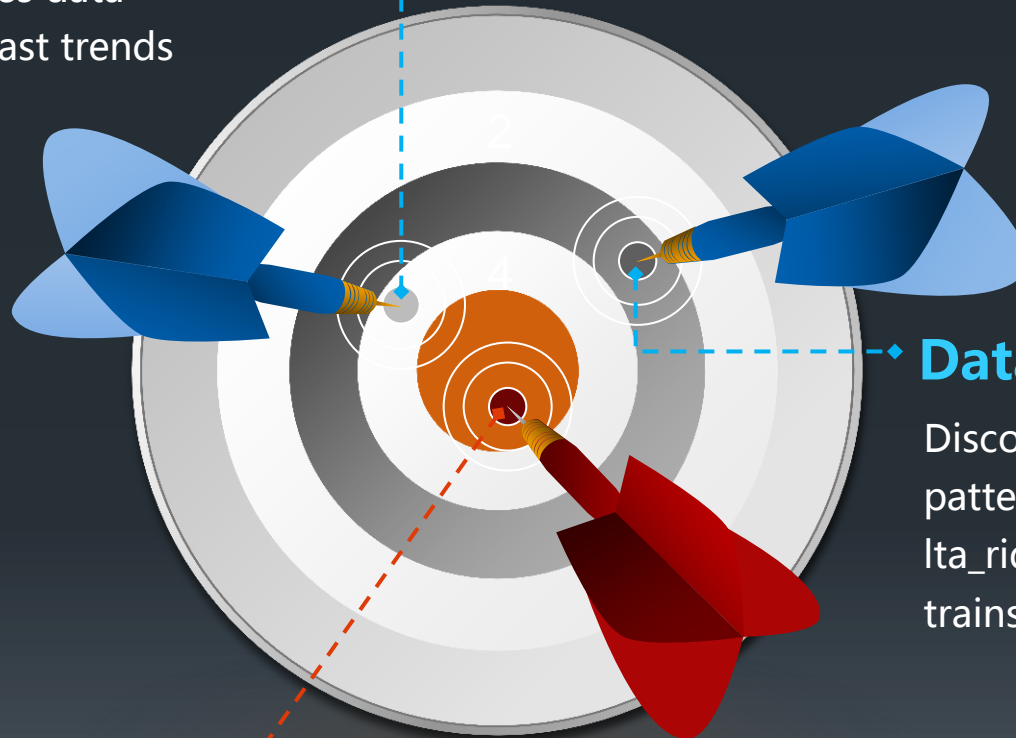


The Ezlink cards have, since its inception become an essential part of our life. It contains rich contextual data ranging from fare prices to entry and exit date and time as well as geographical areas. It provide new window to allow people to discover travelling patterns and insightful knowledge on commuter's behavior.

Objective

Forecasting

Using time-series data mining to forecast trends



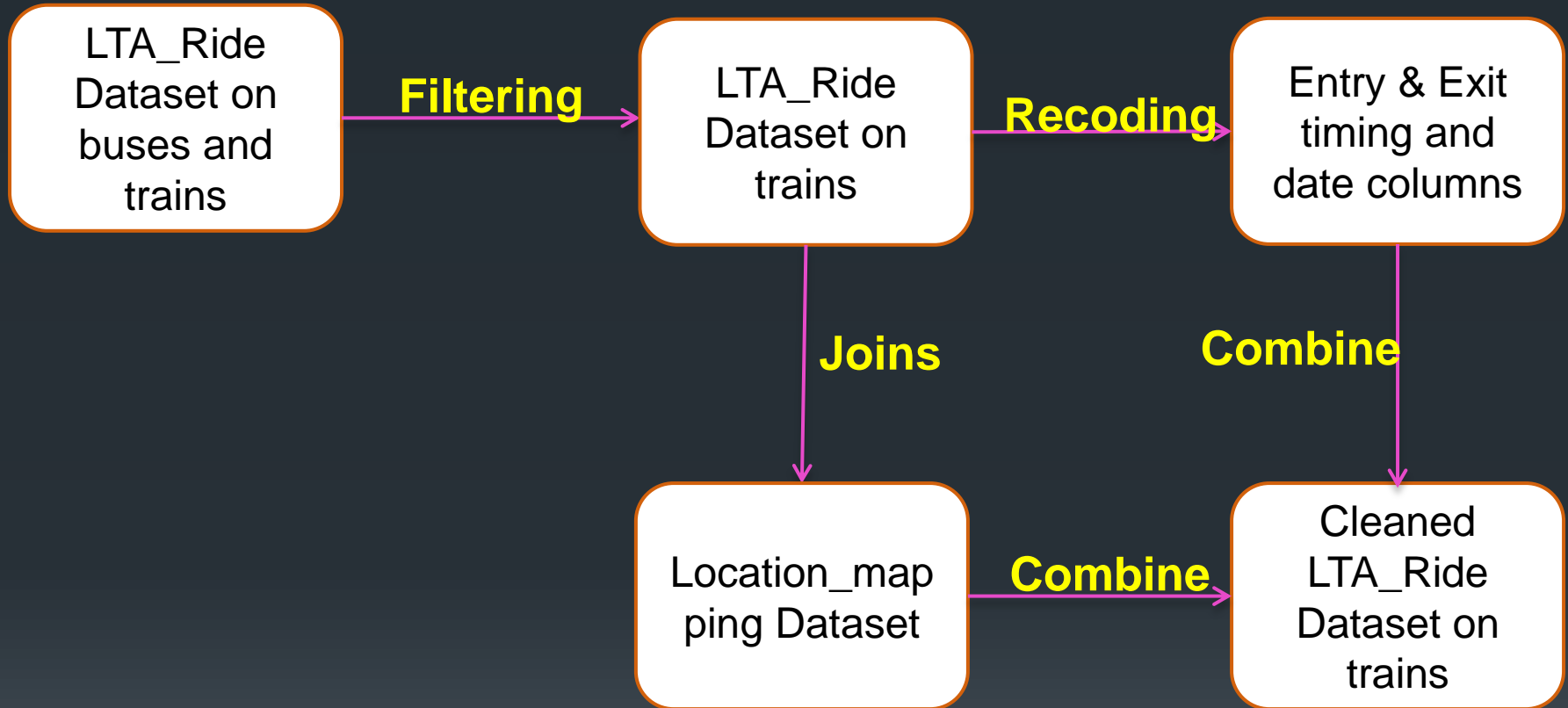
Data discovery

Discover the unusual pattern from the Ita_ride dataset on trains

Pattern detection

Do descriptive analysis on patterns so can classify by commuter's type or different areas or payment mode

Data Collection & Preparation Process



Data Collection & Preparation Process

LTA_Ride Dataset on buses and trains

card_num	transport	entry_dat	entry_tim	exit_date	exit_time	payment	commute	origin_loc	destination	bus_servi	entry_bus	exit_bus	distance	travelling_time
SioLOS	BUS	01/11/201	20:21.0	01/11/201	29:07.0	Pass	Student	1926	2351	246	1	1	1.9	8.767
Gms8+EL	BUS	01/11/201	52:56.0	?	?	Pass	Student	2874	-99	190	1?	?	20.1?	?
LJ/ovW+	BUS	01/11/201	12:32.0	?	?	Pass	Adult	4137	-99	161	1?	?	21.3?	?
EYUJlx+	BUS	01/11/201	07:15.0	01/11/201	20:24.0	Pass	Student	4150	6095	109	1	1	2.6	13.15
U06Vsa+	BUS	01/11/201	50:48.0	?	?	Pass	Adult	6906	-99	902	1?	?	1.1?	?
YhlPgLe+	RTS	01/11/201	10:52.0	01/11/201	52:43.0	Pass	Student	26	3?	?	?	?	19.5	41.85
XI8m9/d	BUS	01/11/201	40:17.0	01/11/201	56:41.0	Pass	Student	3563	3651	76	1	1	4	16.4
Ts3mkXj	BUS	01/11/201	48:07.0	?	?	Pass	Adult	4391	-99	70	1?	?	9.5?	?
NZ8K/Ba	RTS	01/11/201	49:05.0	01/11/201	58:03.0	Pass	Adult	22	43?	?	?	?	3.2	8.967
PVciS9fc	RTS	01/11/201	21:11.0	01/11/201	24:16.0	Pass	Adult	111	110?	?	?	?	0.8	3.083
bJaKWIE+	RTS	01/11/201	08:24.0	01/11/201	45:04.0	Pass	Adult	1	64?	?	?	?	15.2	36.667
B3mlvKv+	BUS	01/11/201	42:19.0	?	?	Pass	Student	6824	-99	81	1?	?	34.5?	?
NYKM7+	BUS	01/11/201	23:50.0	01/11/201	34:54.0	Pass	Student	6878	2785	945	1	1	2.7	11.067
RC+6nRf	RTS	01/11/201	30:27.0	01/11/201	44:45.0	Pass	Adult	43	1?	?	?	?	9.4	14.3
iF57ir5G	RTS	01/11/201	03:51.0	01/11/201	40:06.0	Pass	Adult	45	66?	?	?	?	20.9	36.25
QZAHjxc	BUS	01/11/201	26:49.0	01/11/201	30:52.0	Pass	Adult	6837	4943	229	1	1	0.8	4.05

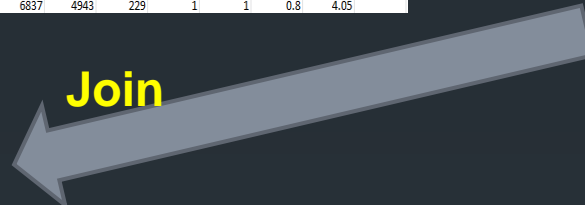
LTA_Ride Dataset on trains

card_num	transport	entry_dat	entry_tim	exit_date	exit_time	payment	commute	origin_loc	destination	distance	travelling_time
YhLpGle+	RTS	01/11/201	10:52.0	01/11/201	52:43.0	Pass	Student	26	3	19.5	41.85
NZ8K/Ba	RTS	01/11/201	49:05.0	01/11/201	58:03.0	Pass	Adult	22	43	3.2	8.967
PVciS9fc	RTS	01/11/201	21:11.0	01/11/201	24:16.0	Pass	Adult	111	110	0.8	3.083
bJaKWIE+	RTS	01/11/201	08:24.0	01/11/201	45:04.0	Pass	Adult	1	64	15.2	36.667
RC+6nRf	RTS	01/11/201	30:27.0	01/11/201	44:45.0	Pass	Adult	43	1	9.4	14.3
iF57ir5G	RTS	01/11/201	03:51.0	01/11/201	40:06.0	Pass	Adult	45	66	20.9	36.25
VDPAP9dm	RTS	01/11/201	37:03.0	01/11/201	09:49.0	Pass	Adult	215	101	14.5	32.767
fiLVpxxik	RTS	01/11/201	21:58.0	01/11/201	07:20.0	Pass	Adult	5	48	26.7	45.367
T1L+JO3X	RTS	01/11/201	49:00.0	01/11/201	22:43.0	Pass	Adult	34	21	17.3	33.717
SzK3csEQ	RTS	01/11/201	22:35.0	01/11/201	05:42.0	Pass	Adult	64	39	22.8	43.117
KIGOMslg	RTS	01/11/201	29:27.0	01/11/201	16:09.0	Pass	Adult	115	46	26.8	46.7
MwLlZjjH	RTS	01/11/201	17:18.0	01/11/201	41:31.0	Pass	Adult	111	37	10	24.217
C50v3Aqj	RTS	01/11/201	51:33.0	01/11/201	19:55.0	Pass	Adult	7	16	8	28.367
JmWVY0G	RTS	01/11/201	11:27.0	01/11/201	58:49.0	Pass	Adult	1	61	25.7	47.367
e659Qq+	RTS	01/11/201	46:57.0	01/11/201	09:14.0	Pass	Adult	111	102	11.3	22.283
ZT1oaRqT	RTS	01/11/201	53:47.0	01/11/201	04:41.0	Pass	Adult	64	17	5.1	10.9
TRDCVJrg	RTS	01/11/201	59:01.0	01/11/201	05:25.0	Pass	Adult	25	27	3.2	6.4
Wo9WCF	RTS	01/11/201	21:18.0	01/11/201	29:43.0	Pass	Adult	9	8	1	8.417

Filtering



Join



origin_location_id	origin_location_name	destination_location_id	destination_location_name
26	Lakeside	3	Bishan
22	Yishun	43	Sembawang
111	Little India	110	Farrer Park
1	Yio Chu Kang	64	Dover
43	Sembawang	1	Yio Chu Kang
45	Woodlands	66	Pioneer
215	Bras Basah	101	Punggol
5	Toa Payoh	48	Yew Tee
34	Aljunied	21	Clementi
64	Dover	39	Tanah Merah
115	Outram Park NEL	46	Marsiling

Combine



Recode columns



entry_tim	hourEnter	exit_time	hourExit
:0:06:10:52	6	:0:06:52:43	6
:0:08:49:05	8	:0:08:58:03	8
:0:12:21:11	12	:0:12:24:16	12
:0:07:08:24	7	:0:07:45:04	7
:0:08:30:27	8	:0:08:44:45	8
:0:16:03:51	16	:0:16:40:06	16
:0:18:37:03	18	:0:19:09:49	19
:0:19:21:58	19	:0:20:07:20	20
:0:07:49:00	7	:0:08:22:43	8
:0:21:22:35	21	:0:22:05:42	22
:0:22:29:27	22	:0:23:16:09	23

Data Collection & Preparation Process

card_num	transport	entry_date	Weekday	entry_time	hourEnter	minuteEnter	exit_date	Weekday	exit_time	hourExit	minuteExit	payment	commute	origin_id	origin_loc	destination	dest_loc	distance	travelling	diffInTim	travel_time_min
YhLpGleR	RTS	01-11-201	3	:06:10:5	6	1	01-11-201	3	:06:52:4	6	4	Pass	Student	26	Lakeside	3	Bishan	19.5	41.85	2511	41
NZ8K/BaN	RTS	01-11-201	3	:08:49:0	8	4	01-11-201	3	:08:58:0	8	4	Pass	Adult	22	Yishun	43	Sembawa	3.2	8.967	538	8
PVciS9fcg	RTS	01-11-201	3	:12:21:1	12	2	01-11-201	3	:12:24:1	12	2	Pass	Adult	111	Little Indi	110	Farrer Par	0.8	3.083	185	3
bJaKWIE7	RTS	01-11-201	3	:07:08:2	7	1	01-11-201	3	:07:45:0	7	3	Pass	Adult	1	Yio Chu Ki	64	Dover	15.2	36.667	2200	36
RC+6nRfD	RTS	01-11-201	3	:08:30:2	8	2	01-11-201	3	:08:44:4	8	3	Pass	Adult	43	Sembawa	1	Yio Chu Ki	9.4	14.3	858	14
iF57lr5G8	RTS	01-11-201	3	:16:03:5	16	1	01-11-201	3	:16:40:0	16	3	Pass	Adult	45	Woodlan	66	Pioneer	20.9	36.25	2175	36
VDPA9dn	RTS	01-11-201	3	:18:37:0	18	3	01-11-201	3	:19:09:4	19	1	Pass	Adult	215	Bras Bas	101	Punggol	14.5	32.767	1966	32
fILVpxxiK	RTS	01-11-201	3	:19:21:5	19	2	01-11-201	3	:20:07:2	20	1	Pass	Adult	5	Toa Payol	48	Yew Tee	26.7	45.367	2722	45
T1L+jO3X+	RTS	01-11-201	3	:07:49:0	7	4	01-11-201	3	:08:22:4	8	2	Pass	Adult	34	Aljunied	21	Clementi	17.3	33.717	2023	33
SzK3csEQf	RTS	01-11-201	3	:21:22:3	21	2	01-11-201	3	:22:05:4	22	1	Pass	Adult	64	Dover	39	Tanah Me	22.8	43.117	2587	43
KIGOMslp	RTS	01-11-201	3	:22:29:2	22	2	01-11-201	3	:23:16:0	23	2	Pass	Adult	115	Outram Pi	46	Marsiling	26.8	46.7	2802	46
MwLIZjjH	RTS	01-11-201	3	:18:17:1	18	2	01-11-201	3	:18:41:3	18	3	Pass	Adult	111	Little Indi	37	Kembang	10	24.217	1453	24
C50v3AqN	RTS	01-11-201	3	:11:51:3	11	4	01-11-201	3	:12:19:5	12	2	Pass	Adult	7	Newton	16	Tiong Bah	8	28.367	1702	28
JmWY0Gs	RTS	01-11-201	3	:19:11:2	19	1	01-11-201	3	:19:58:4	19	4	Pass	Adult	1	Yio Chu Ki	61	Senja	25.7	47.367	2842	47
e659Qq+4	RTS	01-11-201	3	:19:46:5	19	4	01-11-201	3	:20:09:1	20	1	Pass	Adult	111	Little Indi	102	Sengkang	11.3	22.283	1337	22
ZT1oaRqT	RTS	01-11-201	3	:20:53:4	20	4	01-11-201	3	:21:04:4	21	1	Pass	Adult	64	Dover	17	Redhill	5.1	10.9	654	10

Methodology

1

Descriptive Analysis

Tool Used:

- JMP Pro 11 / SAS Enterprise Guide

2

Time-series Data Mining

Tool Used:

- SAS Enterprise Miner

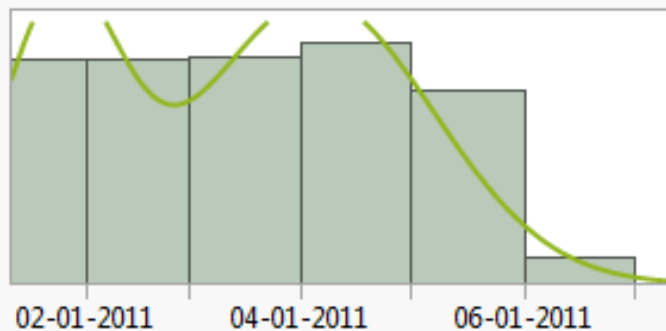
Findings



- Summary Statistics
 - Individual variable
 - Commuter's pattern by commuter type
 - Commuter's pattern by different zones

Commuter's Pattern by Date

entry_date



Normal 2 Mixture

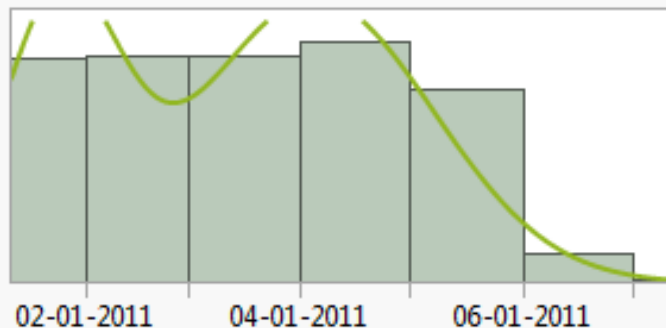
Quantiles

100.0%	maximum	3.39e+9
99.5%		3.39e+9
97.5%		3.39e+9
90.0%		3.39e+9
75.0%	quartile	3.39e+9
50.0%	median	3.38e+9
25.0%	quartile	3.38e+9
10.0%		3.38e+9
2.5%		3.38e+9
0.5%		3.38e+9
0.0%	minimum	3.38e+9

Summary Statistics

Mean	3.3828e+9
Std Dev	3729964.8
Std Err Mean	1123.3722
Upper 95% Mean	3.3828e+9
Lower 95% Mean	3.3828e+9
N	11024582
Skewness	0.0897487

exit_date



Normal 2 Mixture

Quantiles

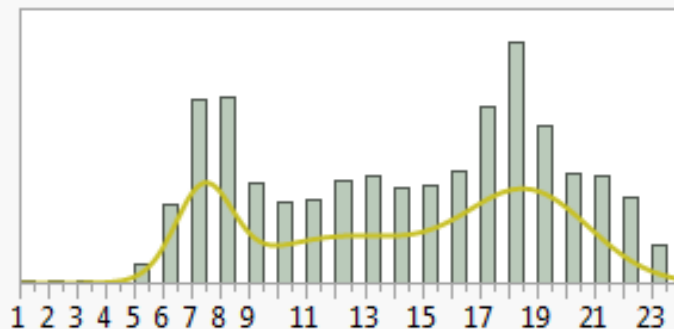
100.0%	maximum	3.39e+9
99.5%		3.39e+9
97.5%		3.39e+9
90.0%		3.39e+9
75.0%	quartile	3.39e+9
50.0%	median	3.38e+9
25.0%	quartile	3.38e+9
10.0%		3.38e+9
2.5%		3.38e+9

Summary Statistics

Mean	3.3828e+9
Std Dev	3739508
Std Err Mean	1126.2464
Upper 95% Mean	3.3828e+9
Lower 95% Mean	3.3828e+9
N	11024582
Skewness	0.0952385

Commuter's Pattern by Hour

hourEnter



Normal 3 Mixture

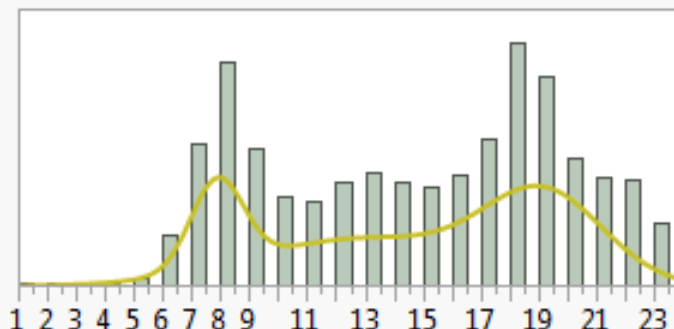
Quantiles

100.0%	maximum	23
99.5%		23
97.5%		22
90.0%		21
75.0%	quartile	18
50.0%	median	15
25.0%	quartile	9
10.0%		7
2.5%		6
0.5%		5
0.0%	minimum	0

Summary Statistics

Mean	14.198121
Std Dev	5.0479003
Std Err Mean	0.0015203
Upper 95% Mean	14.201101
Lower 95% Mean	14.195142
N	11024582
Skewness	-0.158553

hourExit



Normal 3 Mixture

Quantiles

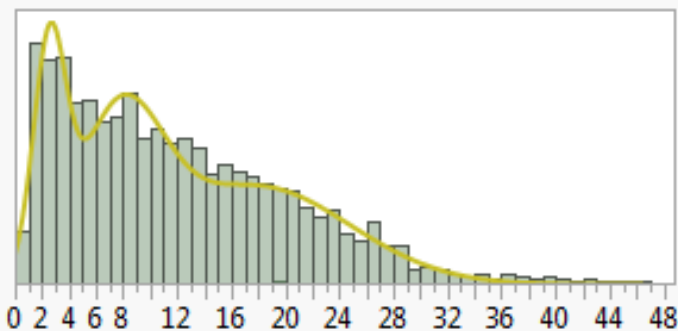
100.0%	maximum	23
99.5%		23
97.5%		23
90.0%		21
75.0%	quartile	19
50.0%	median	15
25.0%	quartile	9
10.0%		8
2.5%		6
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	14.492401
Std Dev	5.1044312
Std Err Mean	0.0015373
Upper 95% Mean	14.495414
Lower 95% Mean	14.489388
N	11024582
Skewness	-0.221068

Commuter's Pattern by Travel time and distance

distance_travelled



Normal 3 Mixture

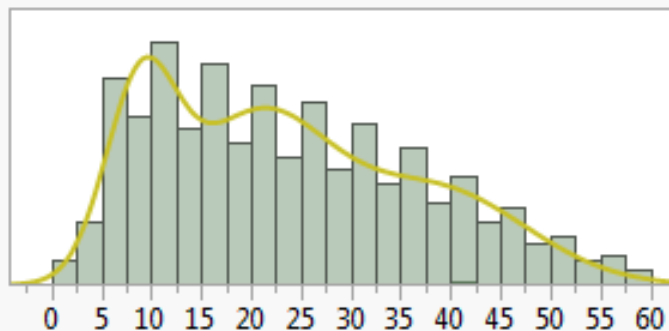
Quantiles

100.0%	maximum	46.6
99.5%		36.6
97.5%		28.8
90.0%		23.2
75.0%	quartile	17.1
50.0%	median	10
25.0%	quartile	4.9
10.0%		2.2
2.5%		1
0.5%		0.5
0.0%	minimum	0

Summary Statistics

Mean	11.611661
Std Dev	8.0827817
Std Err Mean	0.0024343
Upper 95% Mean	11.616432
Lower 95% Mean	11.60689
N	11024582
Skewness	0.7458813

travel_time_min



Normal 3 Mixture

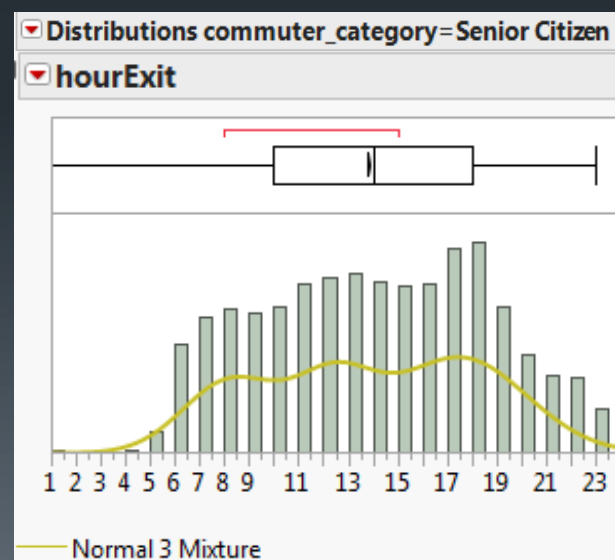
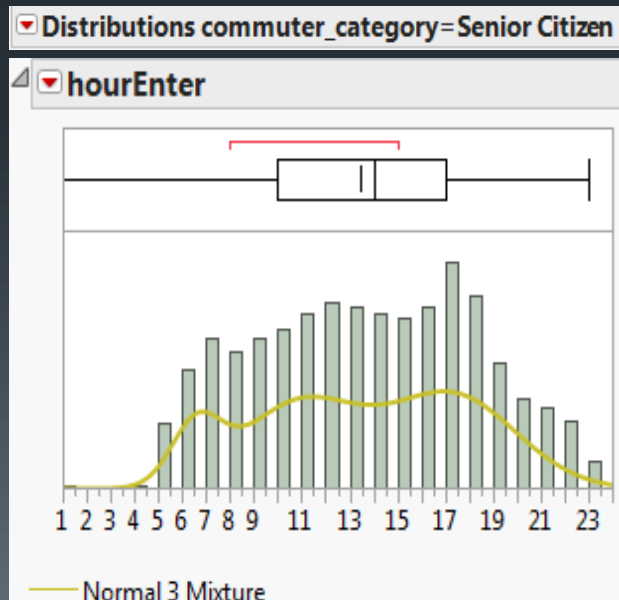
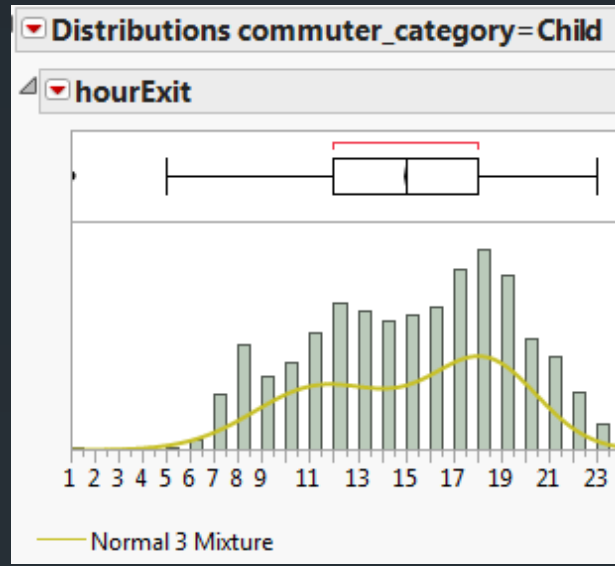
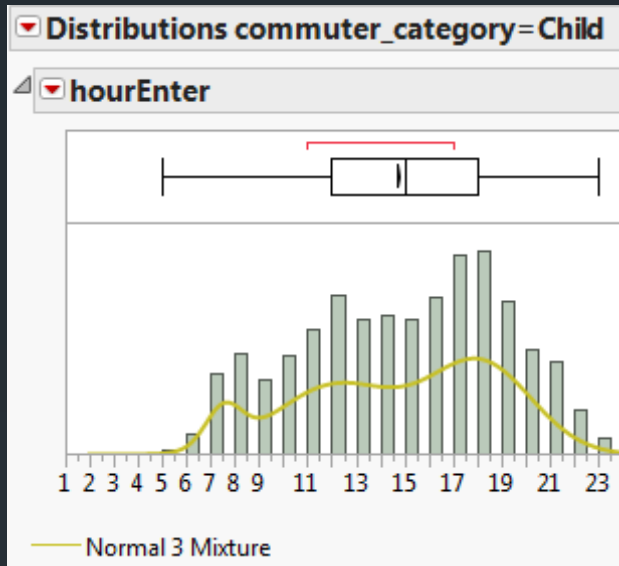
Quantiles

100.0%	maximum	59
99.5%		57
97.5%		52
90.0%		43
75.0%	quartile	33
50.0%	median	22
25.0%	quartile	12
10.0%		7
2.5%		4
0.5%		1
0.0%	minimum	0

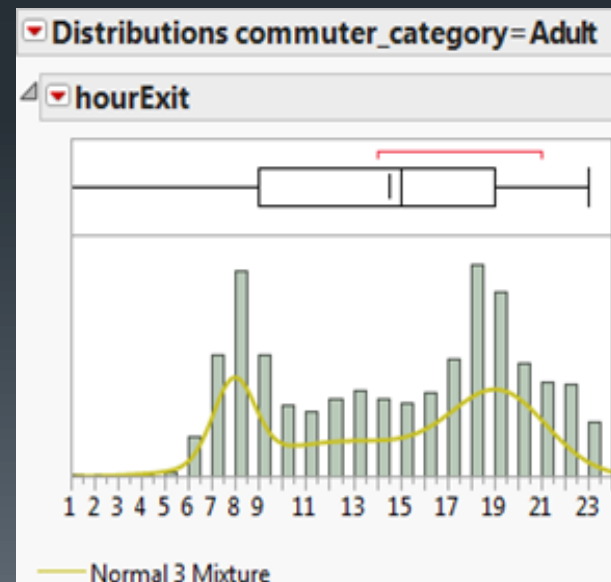
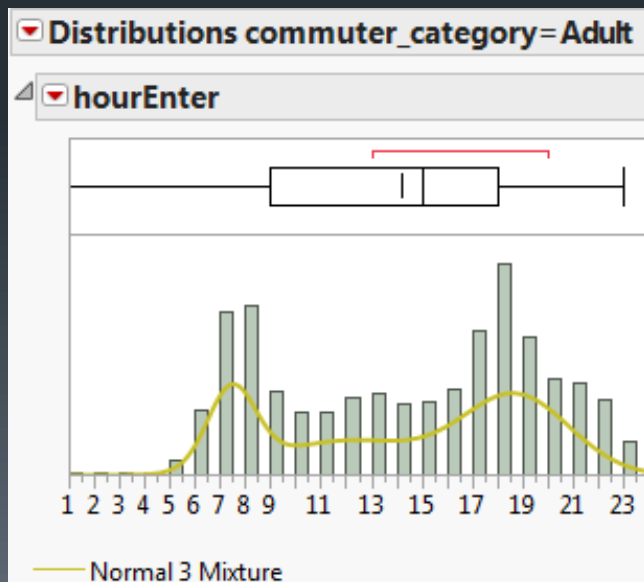
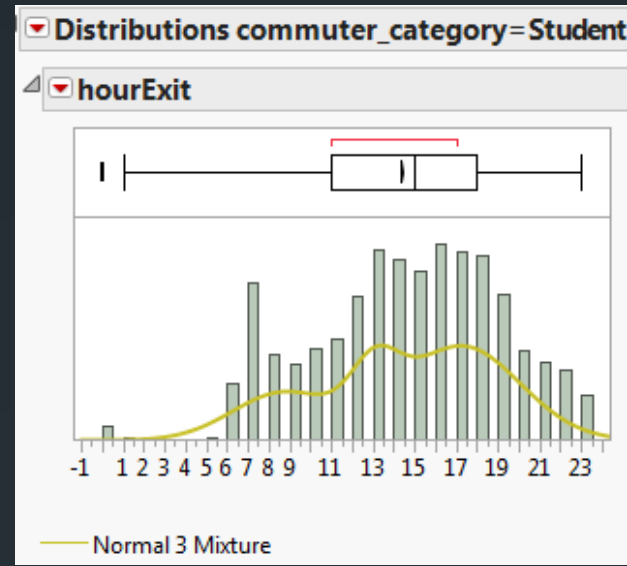
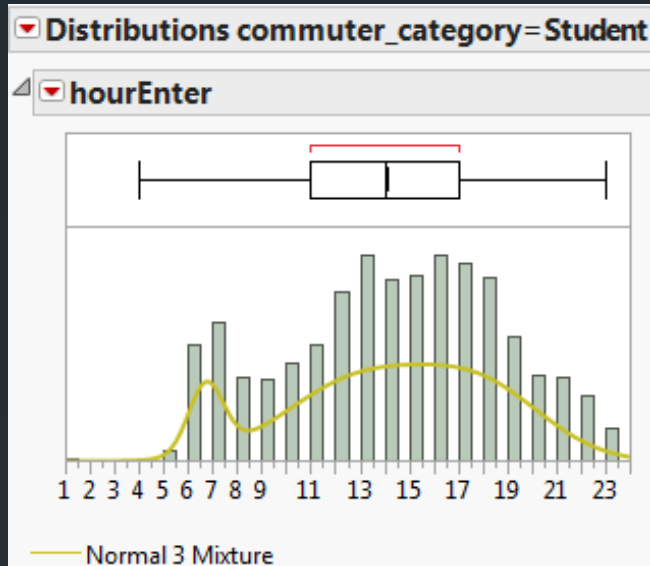
Summary Statistics

Mean	23.460301
Std Dev	13.427014
Std Err Mean	0.0040439
Upper 95% Mean	23.468227
Lower 95% Mean	23.452375
N	11024582
Skewness	0.4879247

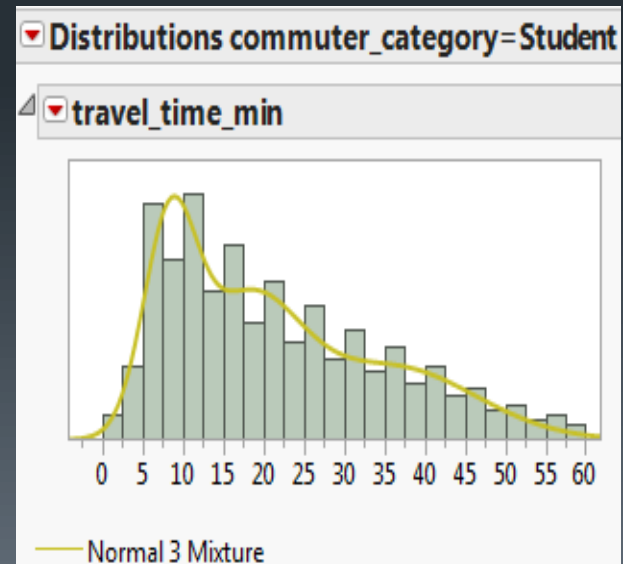
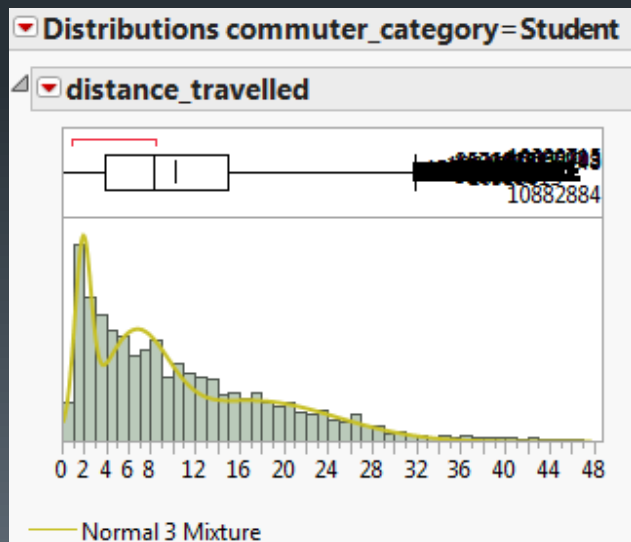
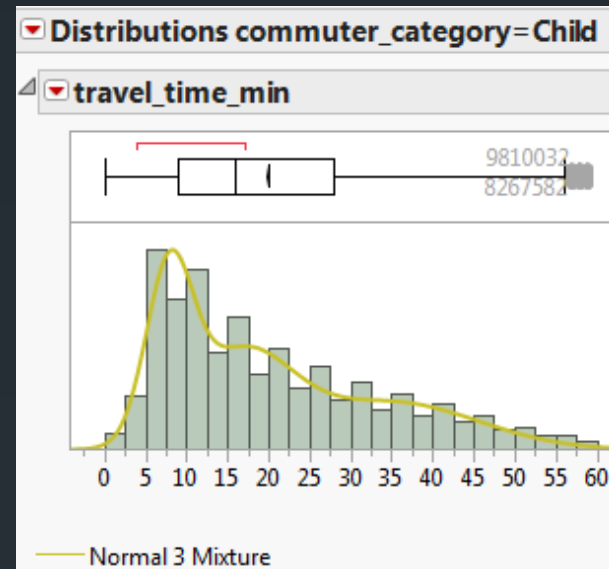
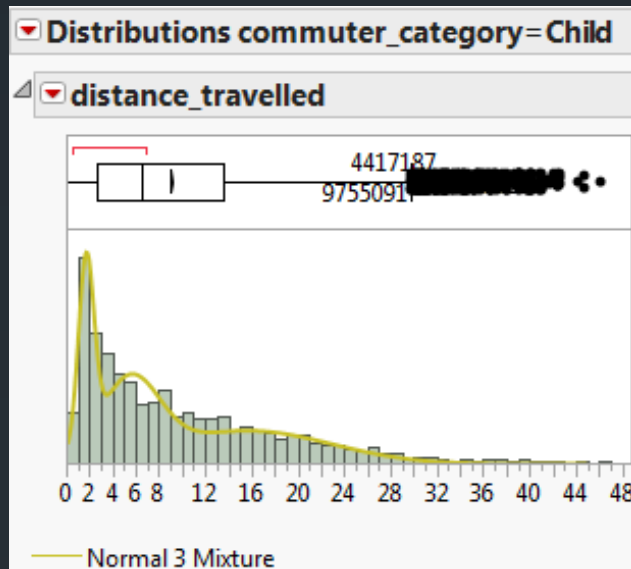
Entry & Exit time by commuter's type



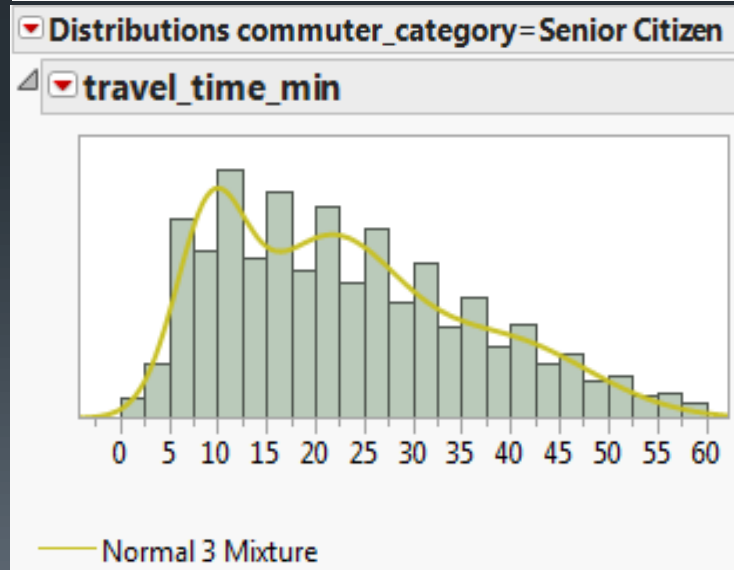
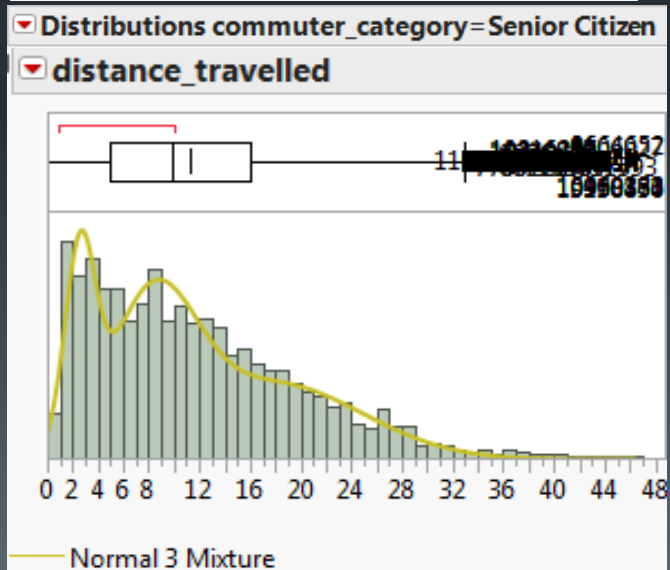
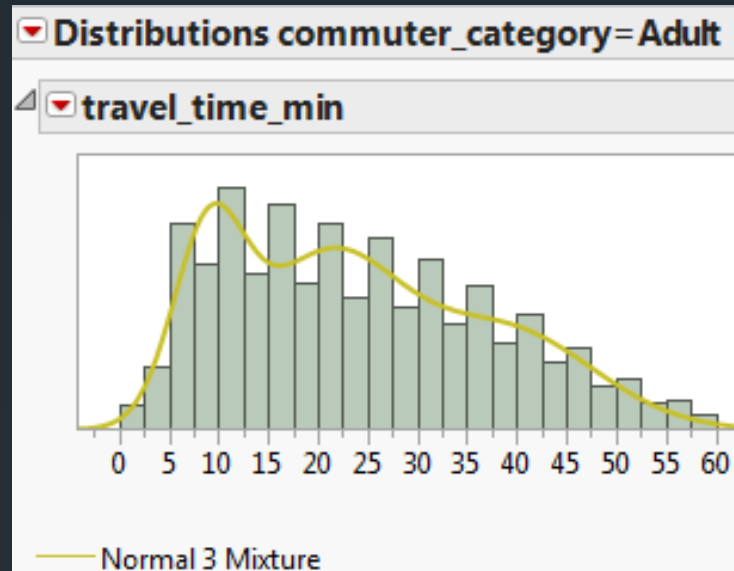
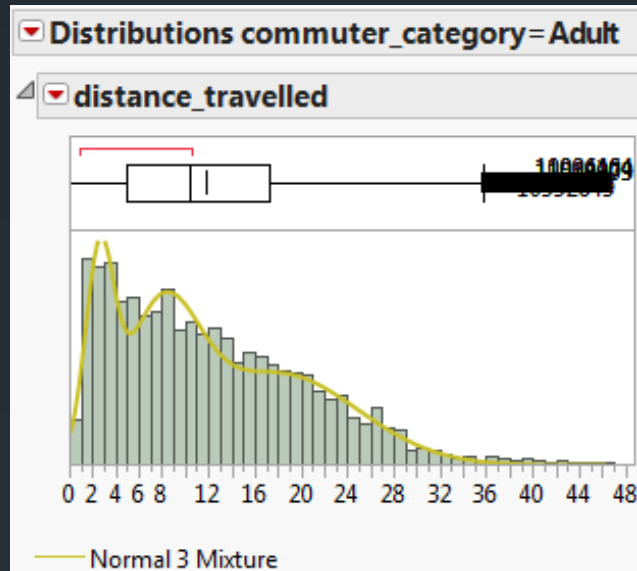
Entry & Exit time by commuter's type



Distance travelled and time by commuter's type



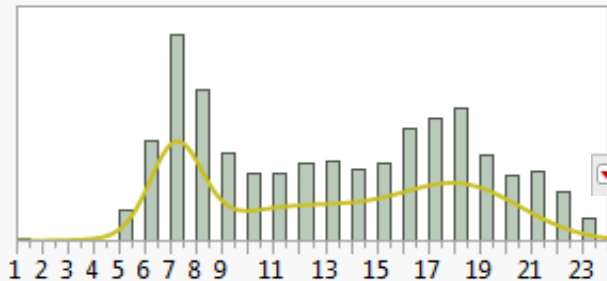
Distance travelled and time by commuter's type



Entry time of different zone

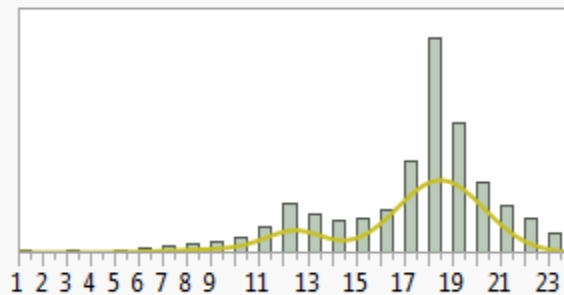
Distributions origin_location_name=Woodlands

hourEnter



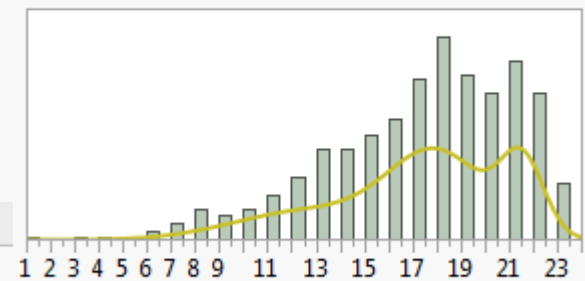
Distributions origin_location_name=Raffles Place

hourEnter



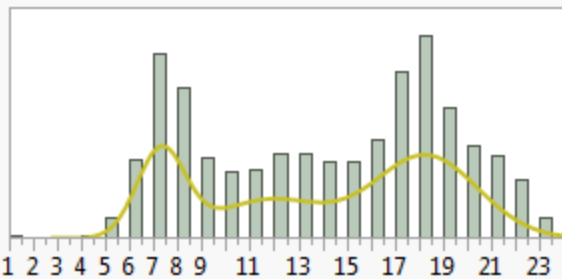
Distributions origin_location_name=Orchard

hourEnter



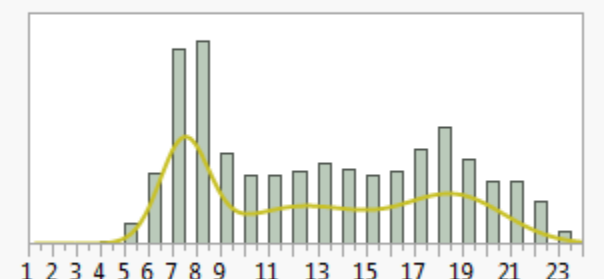
Distributions origin_location_name=Boon Lay

hourEnter



Distributions origin_location_name=Tampines

hourEnter



Normal 3 Mixture

Normal 3 Mixture

Normal 3 Mixture

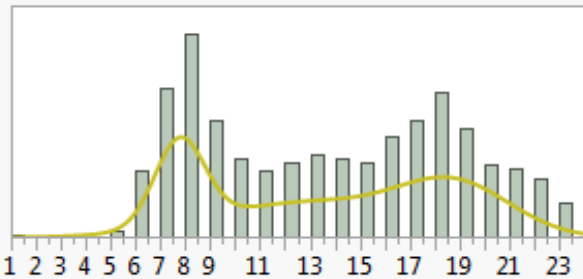
Normal 3 Mixture

Normal 3 Mixture

Exit time of different zone

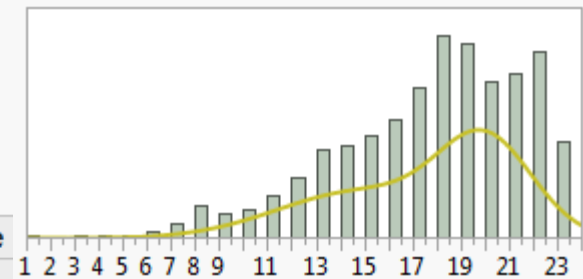
Distributions origin_location_name=Woodlands

hourExit



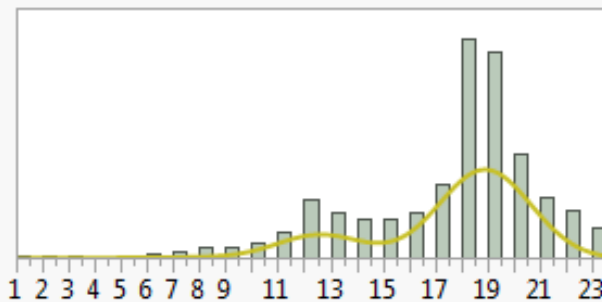
Distributions origin_location_name=Orchard

hourExit



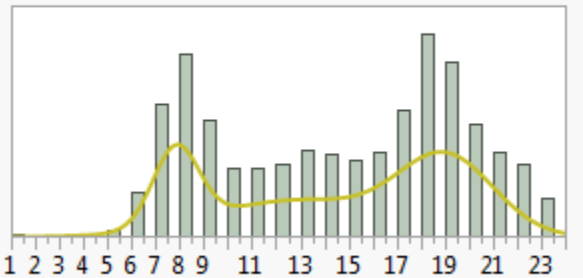
Distributions origin_location_name=Raffles Place

hourExit



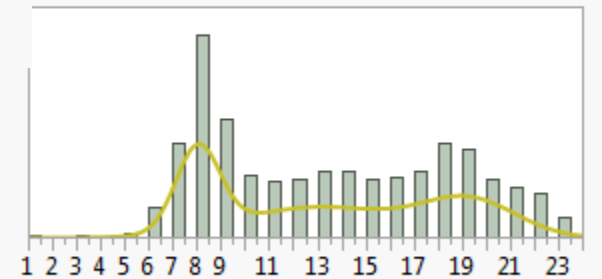
Distributions origin_location_name=Boon Lay

hourExit



Distributions origin_location_name=Tampines

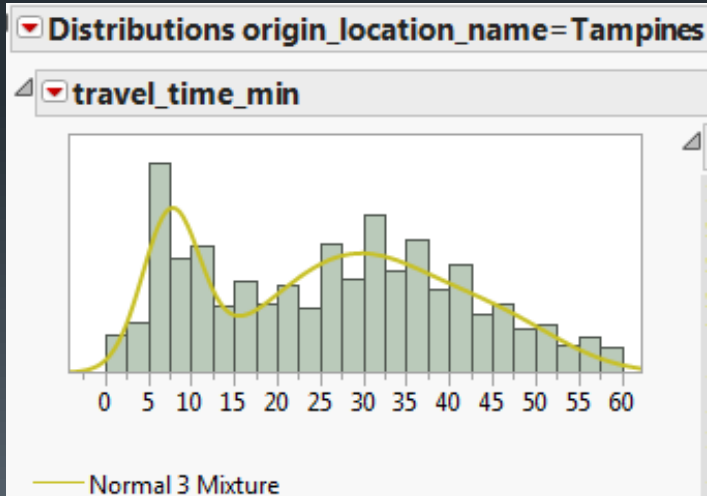
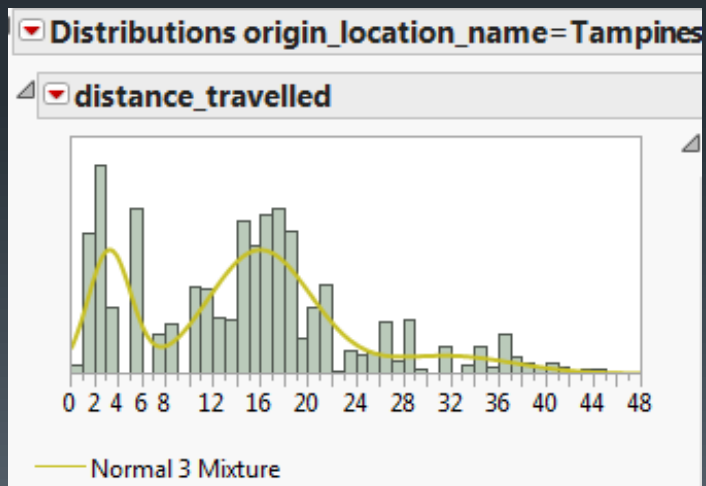
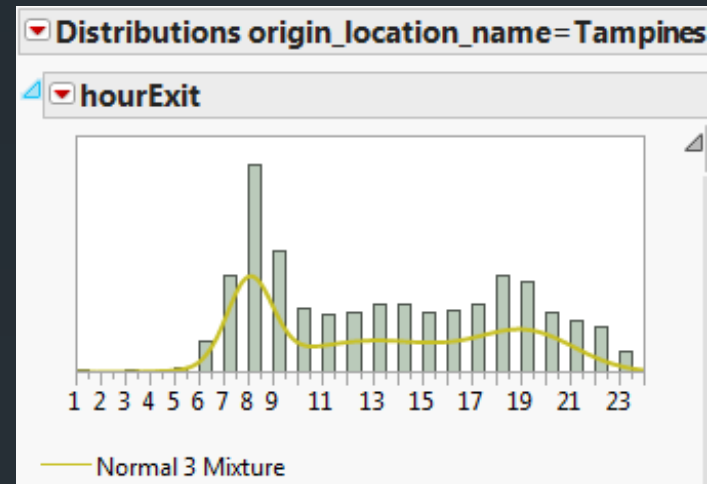
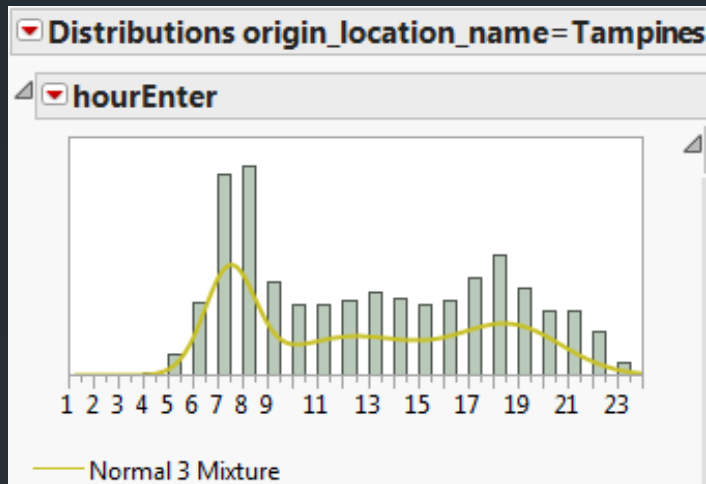
hourExit



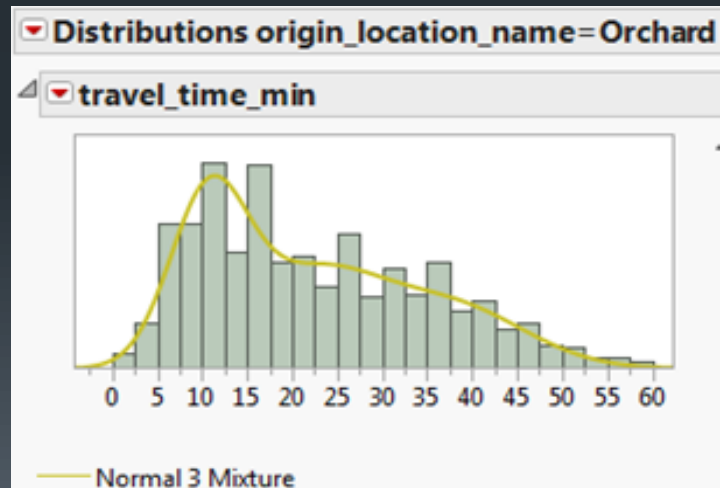
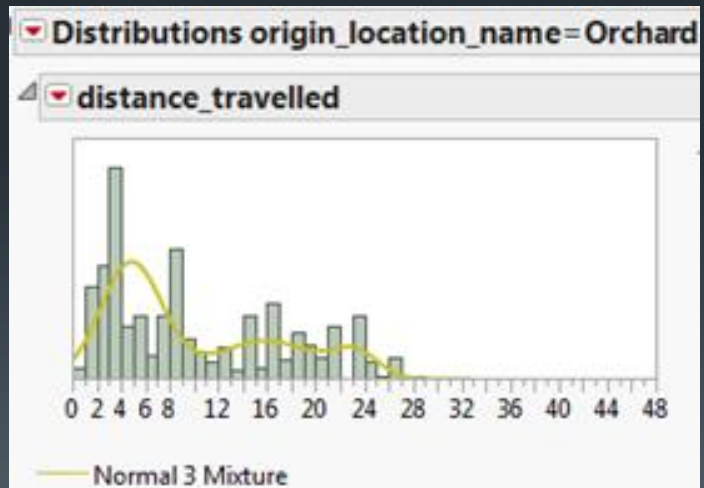
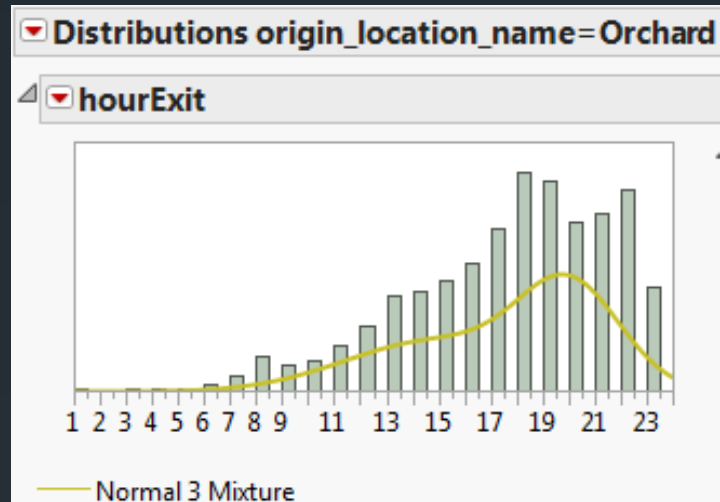
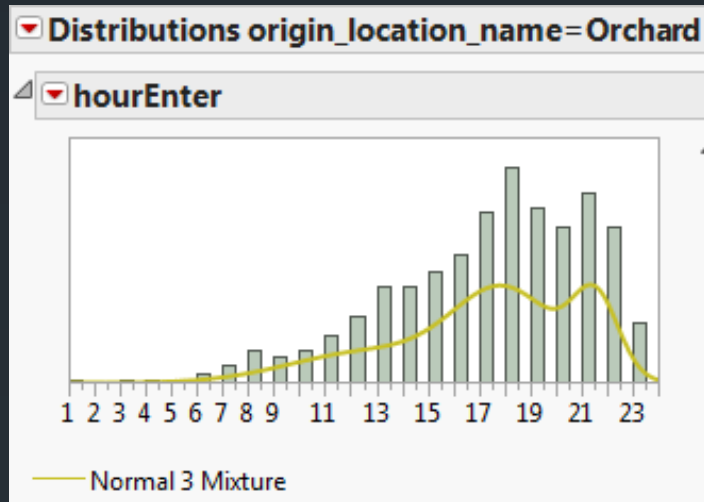
Normal 3 Mixture

Normal 3 Mixture

Commuter's Pattern in Tampines



Commuter's Pattern in Orchard



Review of Project Schedule

ALOS Project

- Source and finalize on project
- Readings on ALOS
- Research on various models

Week 1-3

Cancelation of Current Project

- Consider taking up other project
- Taking up LTA project
- Exploring and readings on LTA

Week 4-5

- Proposal Due

Data Preparation Process / Midterm phase

- Collect dataset
- Data exploration and cleaning
- Analysis & Reporting
- Prepare report & power point slides
- Update wiki

Week 6-8

- Mid-term due

Analysis Process

- Perform further analysis on SAS Enterprise Miner
- Analysis & Reporting
- Research and Report on supporting results
- Draft Research Paper

Week 9-12

Final Phase

- Final Research Paper write up & power point slides
- Poster
- Update wiki

Week 13-14

- Final Artifacts due

**THANK YOU &
WE WISH YOU A HAPPY
CHINESE NEW YEAR!!**

▪ Questions?

