

DATA VISUALIZATION APPROACH AND ONE WAY ANOVA ANALYSIS: A CASE STUDY OF LI KA SHING LIBRARY ENTRY DATA

REN MENGXI, WANG SIJA, WANG TIANJING

Introduction

Our project focuses on analyzing the entry information from the card reader logs provided by Li Ka Shing Library. The card readers are located at both the main entrance and the linkbridge side entrance. This provides us with the entry log with timestamp and basic demographic information about the student.

Motivation

The management team of Li Ka Shing Library is striving for better user experience. Unlike e-book usage or search request which data can be easily collected, they have little information about physical usage of the library, especially the usage of specific user groups.

Objectives

The objective of our project is to use data analytics tools to understand the general visit pattern of the library users, and compare the patterns between different user groups, in order to draw business insights and give suggestion to our sponsor.

Data Preparation

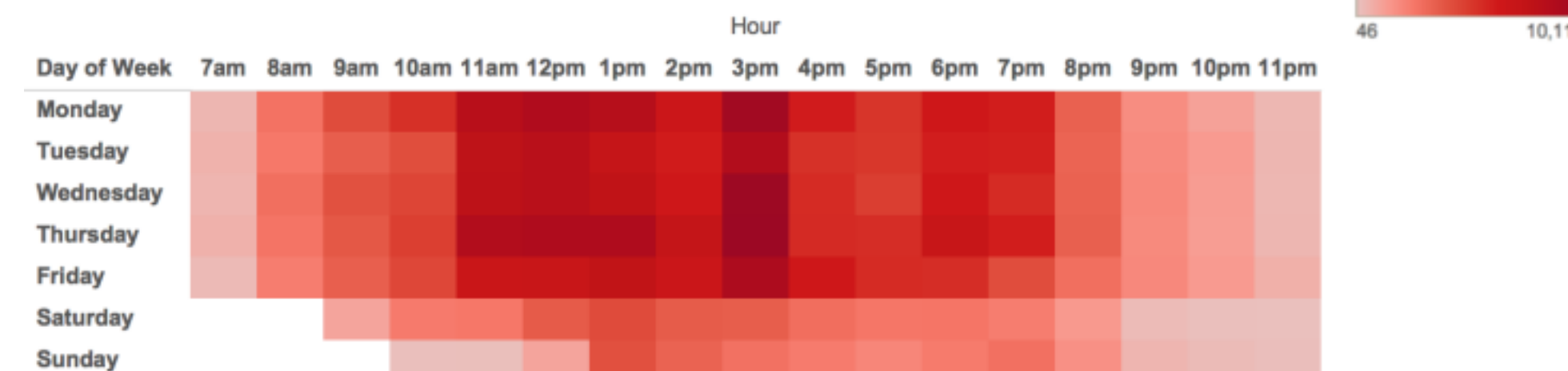
- Data standardization
 - Format 'Date' column
 - Table concatenation
- Rename columns
 - Device Name → Gantry
 - Statistical Category 1 → School
 - Statistical Category 2 → Major
 - Statistical Category 3 → Admission Year
 - Statistical Category 4 → Graduation Year
- Missing value pattern
- Recode column value
 - Gantry: 1, 2, 3, 4, Linkbridge
 - Major: 0 → 'Unknown'
 - Shorten user group and school names
 - Change 'Admission Year' and 'Graduation Year' to numerical value
- Detect outliers
 - Outside main gate opening hours
 - Admission year later than 2017
- Derive new columns
 - IsAlumni
 - YearOfStudy
 - AcademicWeek

After preparation, the clean dataset consists of 481,648 entries generated by 8241 unique users.



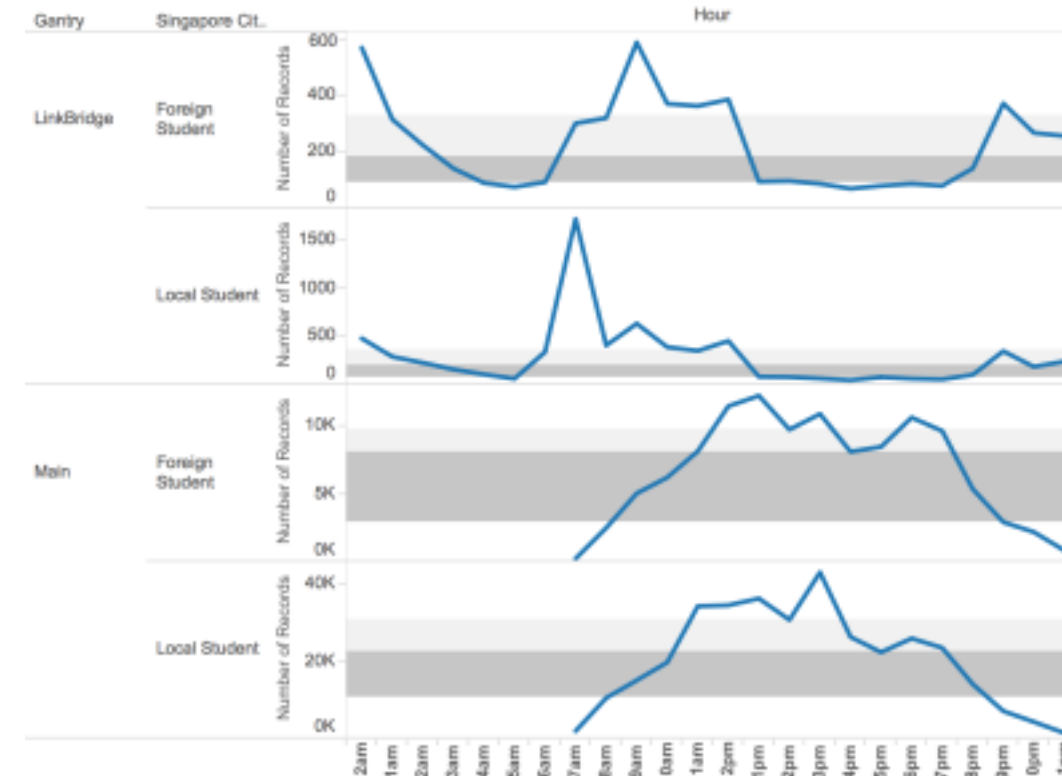
Exploratory Visualization

Main Gantry/Day of Week/Hour



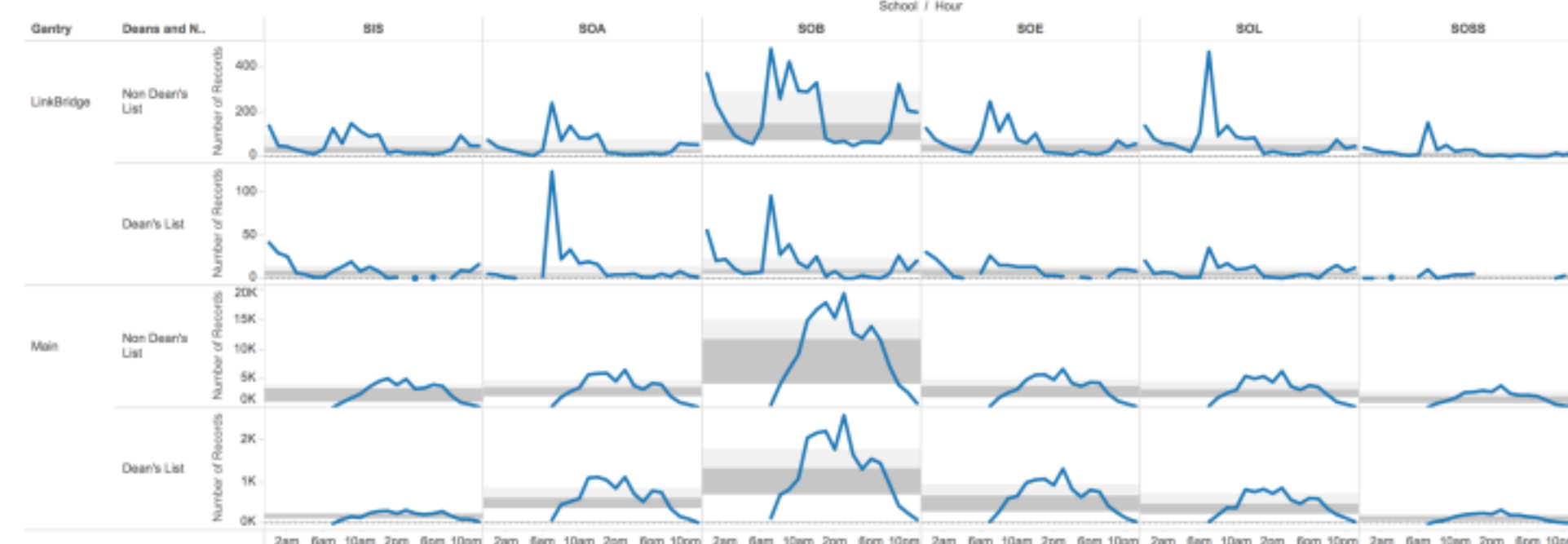
This heatmap shows us the distribution of entries through main gantry at different time and on different day of week. The peak hour lies on Wednesday 3-4pm and Thursday 3-4pm. Also, there are obviously less students coming on weekends.

Hour/Gantry/IsSingaporean



This graph shows the number of entries from different gantry through different hours for local and foreign students. For linkbridge entry, the peak for local students appears at 7am but for international students is at 9am and 12am. For the main entry, there is a clear peak at 3pm but for foreign students, the highest number of entries is at 1pm.

School/Hour/Gantry/IsDeanlister

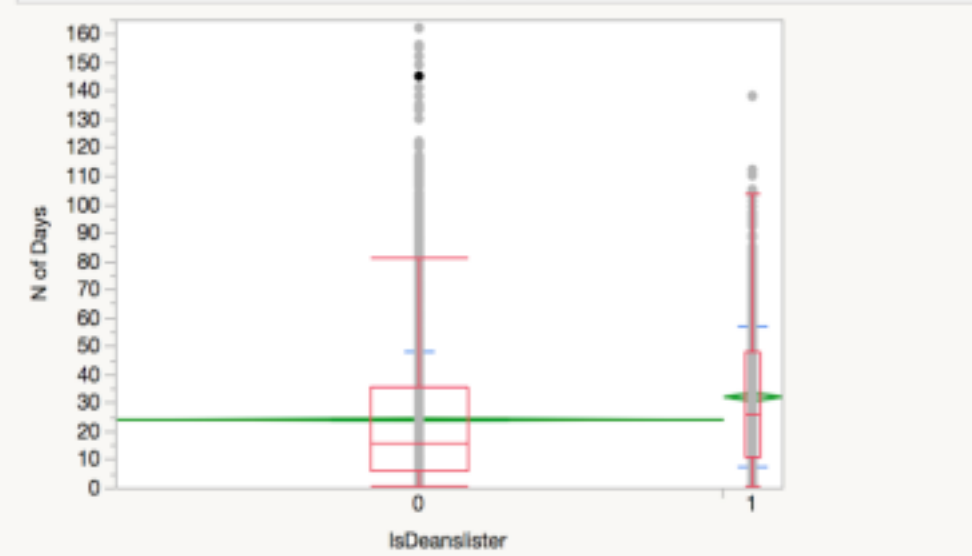


This trellis graph shows the number of entries from main gantry and link bridge through different hours for dean's list and non dean's list students from different schools. We can see that from main entry side, the behaviour of dean's list and non dean's list students and different school students are similar, whereas for linkbridge side, the pattern of dean's list and non dean's list students are different. There is a clear peak at 7am for Business student and Law non dean's list students.



Confirmative Analysis

Oneway Analysis of N of Days By IsDeanslister



Level	Minimum	10%	25%	Median	75%	90%	Maximum
0	1	2	6	16	36	60	162
1	5	11	26	48.25	66.9	138	

Source	DF	Squares	Mean Square	F Ratio	Prob > F
IsDeanslister	1	42053.9	42053.9	74.0175	<.0001*
Error	8239	4681088.7	568.2		
C. Total	8240	4723142.6			

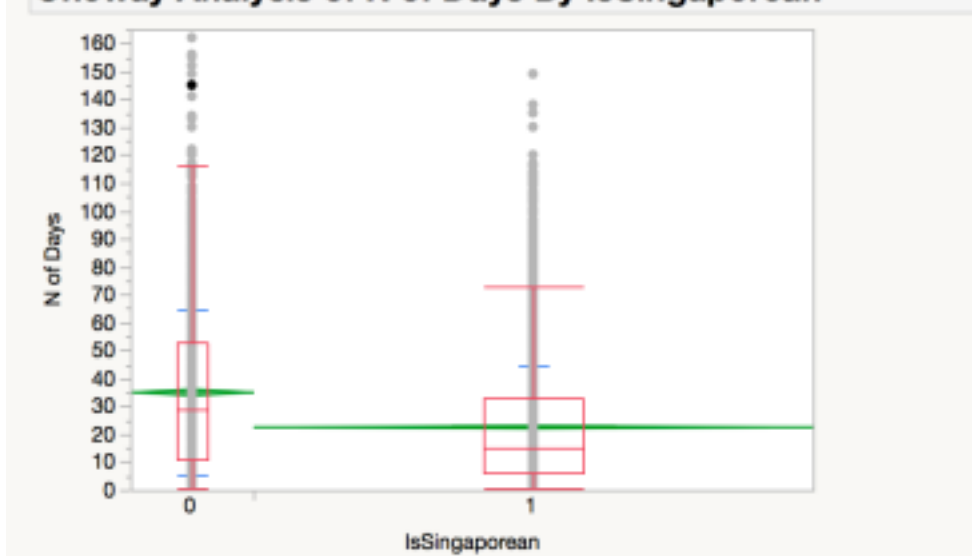
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
0	7511	24.3086	0.27503	23.769	24.848
1	730	32.2589	0.88222	30.530	33.988

Level	Number	Mean	Std Dev	Lower 95%	Upper 95%
0	7511	24.3086	23.7345	0.27386	23.772
1	730	32.2589	24.8592	0.92008	30.453

The line near the center of each diamond represents the group mean. The mean of dean's list is about 32 days. The mean of non-dean's list is about 24. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see dean's list has higher interval than non-dean's list. The median of dean's list is about 26 days. The median of non-dean's list is about 16 days. The median of dean's list is about 10 days higher than the median of non-dean's list.

The Analysis of Variance report shows the standard ANOVA information. The Prob > F (the p-value) is <0.0001, which supports our visual conclusion that there are significant differences between the number of days by dean's list and non-dean's list.

Oneway Analysis of N of Days By IsSingaporean



Level	Minimum	10%	25%	Median	75%	90%	Maximum
0	1	4	11	29	53	77	162
1	2	6	15	33	56	149	

Source	DF	Squares	Mean Square	F Ratio	Prob > F
IsSingaporean	1	188606.6	188607	342.6878	<.0001*
Error	8239	4534536.0	550		
C. Total	8240	4723142.6			

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
0	1477	35.2505	0.61043	34.054	36.447
1	6764	22.7774	0.28525	22.218	23.337

Level	Number	Mean	Std Dev	Lower 95%	Upper 95%
0	1477	35.2505	29.5207	0.76813	33.744
1	6764	22.7774	21.9157	0.26647	22.255

The line near the center of each diamond represents the group mean. The mean of Singaporeans is about 22. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see international student has higher interval than Singaporean student. The mean of international students is about 35 days. The mean of Singaporeans is about 22 days. The median of international student is about 13 days higher than the median of Singaporean student.

The Analysis of Variance report shows the standard ANOVA information. The Prob > F (the p-value) is <0.0001, which supports our visual conclusion that there are significant differences between the number of days by dean's list and non-dean's list.



Conclusion and Recommendation

With exploratory data analysis and confirmative analysis, we analyzed the general patterns of library users and compared the difference between different Singaporean students international students, dean's list and non dean's list students.

General peak hour of library entries lies on Wednesday 3-4pm and Thursday 3-4pm.

For linkbridge usage, international students go to library through linkbridge at midnight while Singaporeans usually go to library through linkbridge in the early morning.

Singaporeans tend to come to library on weekends less.

Dean's listers go to library more often but dean's listers don't go through the linkbridge gate more than non-dean's listers. International students go to library more often than Singaporean students do and international students go through the linkbridge gate more than Singaporean students do.