



ANLY482 - ANALYTICS PRACTICUM

PROJECT PROPOSAL

TEAM Enigma

Wei Xiaoxin

Wu Di

Zheng Xiye

Report Automation and Data Analytics for KTPH

Overview

Khoo Teck Puat Hospital (KTPH) is a 590-bed general and acute care hospital, managed by Alexandra Health System. Opened in June 2010, KTPH offers a comprehensive range of medical services and specialist care to the community in the north.

Every month, KTPH needs to manually prepare 25 reports based on 4 data files, each of which contains around 60k rows, and submit the reports to Ministry of Health (MOH). This brings in 3 major problems for the hospital.

Firstly, the process is very time-consuming. The staff needs to manually go through all the data records so as to identify any suspicious data records, and then contact relevant departments for verification. Only when all the suspicious data have been verified, the staff could then proceed to prepare the 25 reports based on different templates. It takes the hospital about 2-week time to verify the data and prepare the reports each month.

Secondly, the process is error-prone as it involves intensive manual works. Human errors are hardly avoidable.

Thirdly, the process is fairly tedious. The staff repeats the same procedures every month for data cleaning and report preparation. However, he could actually spends this amount of time for more intelligent works.

As such, our project will focus on automate the data cleaning and report generation process for KTPH, so as to improve the efficiency, accuracy and allow for better time-spending. Moreover, as currently there is a lack of data visualization for KTPH to view the changes in data records documented in reports; our project would also implement a dashboard for KTPH to view the important changes in data records.

Motivation

In recent years, digital disruption in every aspect of the commercial world has been prevailing. The ever-increasing trend of digital transformation can be attributed to the widespread adoption of business intelligence and data analytics approaches. Both of which are essential in facilitating organization's decision-making process and effectively bridging the gap between IT capabilities and business functions. In Singapore's context, Smart Nation initiative effectively drives adoption of data analytics among industry partners by harnessing the power of data technologies to create substantial business benefits.

In a sense, the rising industry needs drive employers to look out for individuals capable of drawing insights from numerous amount of data in seek of restructuring and optimizing ongoing business processes. As a group of graduating analytics students looking for hiring opportunities, our skills developed so far are mostly in line with market needs, which sparked off our desire to embark on a career in analytics.

Specifically for this project, we could have a good opportunity to learn and practice R language, which we haven't systematically learnt in other modules. R is definitely

important in data science because of its versatility in the field of statistics. It can be used for various data analysis works and it supports data visualization. It has many associated tools (i.e. Rstudio, R Markdown) and powerful libraries. Building up competence in R would greatly help us pursue a career in data analysis.

Analytics Practicum project not only provides us with a hands-on experience on solving a real-life business problem with analytics methodologies. Most importantly, it draws linkage between skills developed through multiple courses so as to construct an integrated analytics solution. We hope to go beyond 'low-hanging fruits' like, simple model construction with numerous assumptions. Rather, we believe that the 'touch-and-feel' experience on analytics project lifecycle, will get us further down our career path in the near future.

Objectives

1. To automate data cleaning and report generation process
2. To provide visualization for data changes so as to identify the trend

Data

We have gathered 4 data files in Oct, Nov and Dec of 2016, 25 report templates and verification procedures from our sponsor.

1. VTAE (Acute & Emergency)

D	E	F	G	H	I	J	K	L	M	N	O
visit_type	visit_date	visit_time	Case_no	pat_address	pat_ic	pat_sex	pat_dob	pat_race	pat_address	pat_mobiler	pat_homeno
EN	2016-12-30	05:39:48	5716630463	ROSLAN BIN MOHD AF	S7315532E	M	1973-05-05	Malay	BLK804#04-47,WOODL	97371596	0
EN	2016-12-30	05:23:42	5716630462	MUHAMMAD FAREEZ I	T70203886I	M	2002-02-14	Malay	BLK529#03-533,WOOD	94565624	0
EN	2016-12-30	05:20:11	5716630461	LEE CHHOON KHENG WI	S7020344G	M	1970-09-07	Chinese	BLK783#02-76,WOODL	85515782	65845185
EN	2016-12-30	05:18:59	5716630460	WEE NAI KAI	S9705037F	M	1997-01-21	Chinese	BLK239#04-1145,YISHL	97100225	0
EN	2016-12-30	05:07:04	5716630449	LONG CAIHONG	S8360528F	F	1983-08-02	Chinese	BLK689#02-112,WOOD	84575995	96830802
EN	2016-12-30	05:05:53	5716630448	KAMARULZAMAN BIN	S8508628F	M	1985-03-17	Indian	BLK502#13-72,BEDOK	94514432	68998637
EN	2016-12-30	05:02:11	5716630447	TENG GIM NYA	S1760603D	F	1966-06-25	Chinese	BLK228#06-06,WOODLA	97555459	62690232
EN	2016-12-30	05:05:46	5716630446	AHMAD RUZALI BIN	S9211661A	M	1992-04-06	Malay	BLK346#05-113,YISHUP	94591304	62579708
EN	2016-12-30	04:56:55	5716630445	SEAH CHWEE LENG	S1471707B	F	1961-02-08	Chinese	BLK21,JALAN PERADUN	96627252	64840109
EN	2016-12-30	04:52:33	5716630444	CHAO CI MING	S1345618F	M	1959-06-27	Chinese	BLK608#07-247,WOOD	91403020	0
EN	2016-12-30	04:38:53	5716630443	NENI BINTE EMILY	S1161505H	F	1956-06-24	Malay	BLK507#09-12,JELAPAN	84229396	68922490
EN	2016-12-30	04:28:51	5716630442	KWA JUN MING, JOSEF	S8741763H	M	1987-12-23	Chinese	BLK334#16-451,SEMBA	69993334	0
EN	2016-12-30	04:12:36	5716630441	NEO TECK CHYE	S8031943F	M	1980-10-10	Chinese	BLK778#02-1565,YISHL	91168196	92343960
EN	2016-12-30	03:33:36	5716630440	TOH TECK LOON ZHUO	S7905848C	M	1979-03-06	Chinese	BLK9812-21,WOODLAN	98460814	0

2. VTOP

A	B	C	D	E	F	G	H	I
Inst.	BLANK_out	Operation Date	Ext. Pat. ID	Case No.	Check digit	Cls1	Cls2	Clinic
KTPH		2016-12-29	S0019536G	5716616559	Z	SUB	Subsidised	KDSUR
KTPH		2016-12-31	S7043569A	5716628655	D	C	Class C	KDSUR
KTPH		2016-12-31	S0947042E	5716619928	A	B2	Class B2	KDSUR
KTPH		2016-12-31	S1826610E	5716627167	E	C	Class C	KDURO
KTPH		2016-12-31	S1826610E	5716627167	E	C	Class C	KDSUR
KTPH		2016-12-31	S7772301C	5716625459	B	SUB	Subsidised	KDSUR
KTPH		2016-12-31	S7465051A	5716620965	A	SUBP	Subsidised for PR	KDSUR

3. VTPD

A	B	C	D	E	F	G	H	I	
bu_as_at	hosp_code	hosp	ward	subspec	pat_grp_letter	pat_grp	subspec_desc	class	
11	2016-12-31	KTPH	K	KWSSS	KSURO	A	SG	Urology	C
12	2016-12-31	KTPH	K	KWSSS	KSGENSUR	A	SG	General Surgery	C
13	2016-12-31	KTPH	K	KWSSS	KSGENSUR	B	PR	General Surgery	B2
14	2016-12-31	KTPH	K	KWSSS	KSGENSUR	A	SG	General Surgery	B2
15	2016-12-31	KTPH	K	KWSSS	KSGENORT	C	RF	General Orthopaedics	C
16	2016-12-31	KTPH	K	KWSSS	KSGENORT	A	SG	General Orthopaedics	C

4. VTSOC (Specialized Outpatient Clinic)

A	B	C	D	E	F	G	H	I
hosp	BLANK_out1	trmt_ou_code	trmt_ou	BLANK_out2	BLANK_out3	sap_patient_no	patient_name	patient ic
1	KTPH	2016-12-31	KCHOME	Home Care Services		3793442	KONG HUP YEE	S0662093J
2	KTPH	2016-12-31	KCHOME	Home Care Services		3793442	KONG HUP YEE	S0662093J
3	KTPH	2016-12-31	KCHOME	Home Care Services		3793442	KONG HUP YEE	S0662093J
4	KTPH	2016-12-31	KCHOME	Home Care Services		3793442	KONG HUP YEE	S0662093J
5	KTPH	2016-12-31	KCHOME	Home Care Services		3668955	KOK SOOK KIEN	S0947836A
6	KTPH	2016-12-31	KCHOME	Home Care Services		3668180	KOK PANG @AH TOH	S2175478J
7	KTPH	2016-12-31	KCHOME	Home Care Services		3668180	KOK PANG @AH TOH	S2175478J
8	KTPH	2016-12-31	KCHOME	Home Care Services		6569979	KOH YIAN KIN	S0597128D

Methodology

1. Understanding the data and procedures

We gain understanding of the data and procedures by communicating with our sponsor. We further explore the procedures by manually verifying the data set according to the rules given.

Scope of Work

Task	Task Description	Job Allocation
<i>Gather Requirements from Sponsor</i>	Learn about business background, operating model, and data cleaning procedures	WD(Wu Di), XY(Zheng Xiye), XX(Wei Xiaoxin)
<i>Collect Data and Report format from sponsor</i>	Conduct research to know more about the SMU Li Ka Shing Library	WD(Wu Di), XY(Zheng Xiye), XX(Wei Xiaoxin)
<i>Research R functions for data cleaning and transformation</i>	Explore different types of R functions for data cleaning according to the procedures	WD
<i>Research R functions for formatting and exporting report</i>	Explore different types of R functions and library for generating report and .xlsx file formatting	XX
<i>Research R functions for UI</i>	Explore different type of UI library to build a user interface with document uploading and dynamic report generation	XY
<i>Research R functions for data visualization</i>	Explore different type of library to perform data visualization in R	WD,XY
<i>Create Wiki page (proposal)</i>	Create wiki page for proposal submission with relative content	WD,XX,XY
<i>Report (proposal)</i>	Prepare proposal report	WD,XX,XY
<i>Create test data sets</i>	Create a subset data of the raw data set, to test the logic applied for the data cleaning procedures	WD,XY
<i>Explore the report cells relationships with existing data set</i>	Learn the required report format, and figure out the derived cells' relationship with raw data set	XX
<i>Data cleaning for SOC, A&E, Inpt and PD</i>	Create R code to perform data cleaning for the raw data sets according to the	WD

	procedures correspondingly	
<i>Data transformation for SOC, A&E, Inpt and PD files according to report format</i>	Create R code to perform data/metadata transformation and mapping according to the relative report format correspondingly	XY
<i>Testing for SOC ,A&E, Inpt and PD file with standard data file</i>	<i>Testing the R code of data cleaning and transformation for SOC ,A&E, Inpt and PD file with original raw data set correspondingly</i>	WD
<i>Report format design</i>	Create R code to generate .xlsx file with the same design as corresponding report	XX
<i>Report export</i>	Create R code to generate and export report	XX
<i>Testing with clean data set</i>	Test the R code for report formatting and exporting with clean data set(cleaned from raw data) input	XX
<i>Prepare Interim presentation</i>	Prepare content and slides for interim presentation	WD,XX(finalize),XY
<i>Interim Report</i>	Write interim report and revise according to interim presentation	XY
<i>Update Wiki(Interim)</i>	Update Wiki page content with interim progress	WD,XY
<i>Visualize data trend and comparsion in R</i>	Create R code to perform data visualization for some cetain trend and comparsion according to sponsor needs	WD,XX
<i>Build visualization dashboard</i>	Collaborate the visualization R code into a dashborad with existing UI	XY
<i>Integrate data cleaning report code</i>	Integrate R code for data cleaning and report exporting and ensure the application perform properly	XX
<i>Integrate UI and report code</i>	Integrate R code for UI and report reporting part and ensure the application perform properly	XY
<i>Integrate UI and data cleaning code</i>	Integrate R code for UI and data cleaning part and ensure the application perform properly	XY

<i>Integrate final automatic process with data visualization</i>	Integrate R code with UI for both data cleaning and report exporting part with data visualization dashboard, and ensure the application perform properly	WD,XY
<i>Update final wiki page</i>	Ensure the wiki page is updated to final stage with final outcomes	WD,XX,XY
<i>Update final report</i>	Integrate final outcomes to write final report (need to collaborate the interim report feedback from client and supervisor)	WD,XX,XY(finalize)
<i>Prepare Poster</i>	Prepare poster for final presentation	WD,XX,XY
<i>Presentation to Client</i>	Present outcome, report and poster to client and gather feedback	WD,XX,XY
<i>Final Adjustment-Application</i>	Adjust final application based on presentation feedbacks	WD,XX,XY
<i>Final Adjustment-Report</i>	Adjust final report based on presentation feedbacks	WD,XX,XY
<i>Buffer</i>	For us to catch up or use due to any changes of our schedule.	

Work Plan

* For weekly sponsor meeting, supervisor consultation, internal meetings:

Weekly Meeting	Task	Job Allocation
Sponsor meeting	Meeting minutes, Presentation slides	XY(minutes), WD,XX,XY(take turn to prepare slides)
Supervisor consultation	Meeting minutes	WD
Internal meetings	Meeting minutes	XX