

**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 aspects 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.

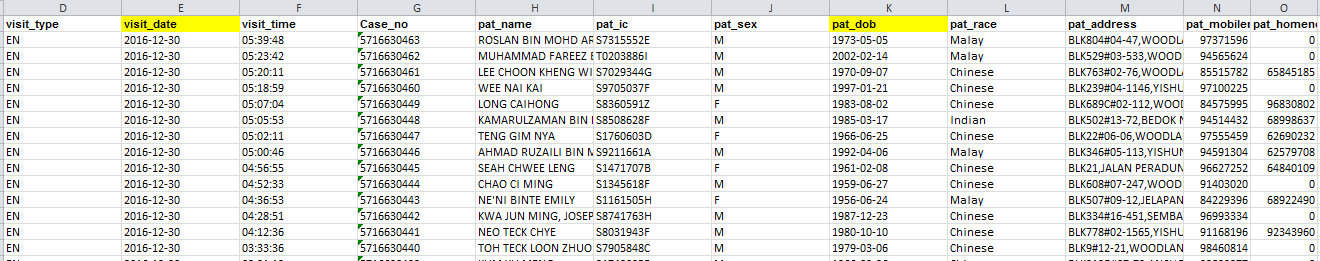
Analytics Practicum project not only provides us with a hands-on experience on solving a real-life business problem with analytics methodologies. Most importantly, it draws linkage between skills developed through multiple courses so as to construct an integrated analytics solution. We hope to go beyond ‘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**

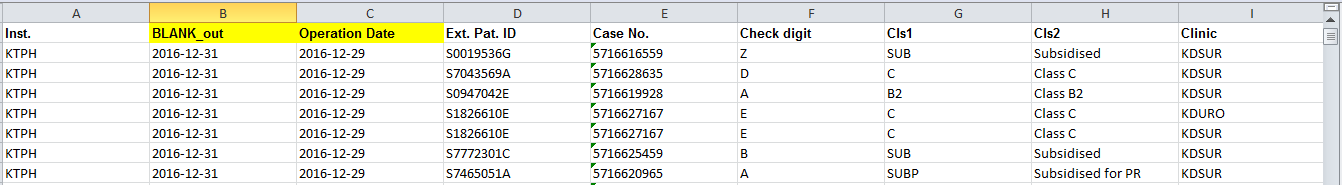
1. To automate data cleaning and report generation process
2. To provide visualization for data changes so as to identify the trend

**Data**

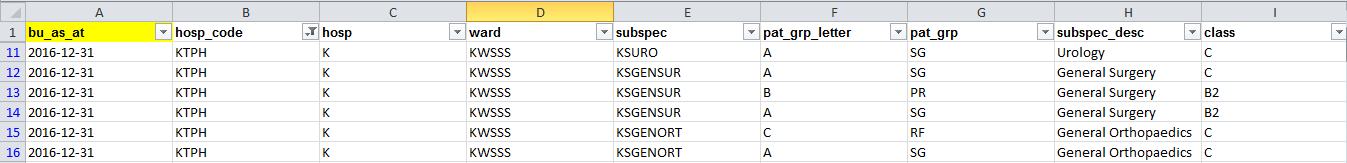
We have gathered 4 data files in Oct, Nov and Dec of 2016, 25 report templates and verification procedures from our sponsor.

1. VTAE (Acute & Emergency)  


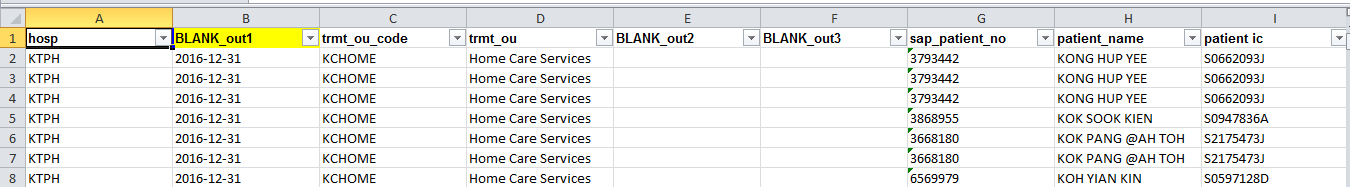
2. VTOP



3. VTPD



4. VTSOC (Specialized Outpatient Clinic)



**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** |
| *Gather Requirements from Sponsor* | Learn about business background, operating model, and data cleaning procedures | WD(Wu Di),  XY(Zheng Xiye),  XX(Wei Xiaoxin) |
| *Collect Data and Report format from spnsor* | Conduct research to know more about the SMU Li Ka Shing Library | WD(Wu Di),  XY(Zheng Xiye),  XX(Wei Xiaoxin) |
| *Research R functions for data cleaning and transformation* | Explore different types of R functions for data cleaning according to the procedures | WD |
| *Research R functions for formatting and exporting report* | Explore different types of R functions and library for generating report and .xlsx file formatting | XX |
| *Research R functions for UI* | Explore different type of UI library to bulid a user interface with document uploading and dynamic report generation | XY |
| *Research R functions for data visualization* | Explore different type of library to perform data visualization in R | WD,XY |
| *Create Wiki page (proposal)* | Create wiki page for proposal submission with relative content | WD,XX,XY |
| *Report (proposal)* | Prepare proposal report | WD,XX,XY |
| *Create test data sets* | Create a subset data of the raw data set, to test the logic applied for the data cleaning procedures | WD,XY |
| *Explore the report cells relationships with existing data set* | Learn the required report format, and figer out the derived cells’ relationship with raw data set | XX |
| *Data cleaning for SOC, A&E, Inpt and PD* | Create R code to perform data cleaning for the raw data sets according to the procedures correspondingly | WD |
| *Data transformation for SOC, A&E, Inpt and PD files according to report format* | Create R code to perform data/metadata transformation and mapping according to the relative report format correspondingly | XY |
| *Testing for SOC ,A&E, Inpt and PD file with standard data file* | *Testing the R code of data cleaninga and transformation for SOC ,A&E, Inpt and PD file with original raw data set correspondingly* | WD |
| *Report format design* | Create R code to generate .xlsx file with the same design as corresponding report | XX |
| *Report export* | Create R code to generate and export report | XX |
| *Testing with clean data set* | Test the R code for report formatting and exporting with clean data set(cleaned from raw data) input | XX |
| *Prepare Interim presentation* | Prepare content and slides for interim presentation | WD,XX(finalize),XY |
| *Interim Report* | Write interim report and revise according to interim presentation | XY |
| *Update Wiki(Interim)* | Update Wiki page content with interim progress | WD,XY |
| *Visualize data trend and comparsion in R* | Create R code to perform data visualization for some cetain trend and comparsion according to sponsor needs | WD,XX |
| *Build visualization dashboard* | Collaborate the visualization R code into a dashborad with existing UI | XY |
| *Integrate data cleaning report code* | Integrate R code for data cleaning and report exporting and ensure the application perform properly | XX |
| *Integrate UI and report code* | Integrate R code for UI and report reporting part and ensure the application perform properly | XY |
| *Integrate UI and data cleaning code* | Integrate R code for UI and data cleaning part and ensure the application perform properly | XY |
| *Integrate final automatic process with data visualization* | 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 |
| *Update final wiki page* | Ensure the wiki page is updated to final stage with final outcomes | WD,XX,XY |
| *Update final report* | Integrate final outcomes to write final report (need to collaborate the interim report feedback from client and supervisor ) | WD,XX,XY(finalize) |
| *Prepare Poster* | Prepare poster for final presentation | WD,XX,XY |
| *Presentation to Client* | Present outcome, report and poster to client and gather feedback | WD,XX,XY |
| *Final Adjustment-Application* | Adjust final application based on presentation feedbacks | WD,XX,XY |
| *Final Adjustment-Report* | Adjust final report based on presentation feedbacks | WD,XX,XY |
| *Buffer* | 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 |