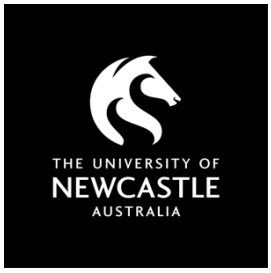
THE UNIVERSITY OF NEWCASTLE

SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTING



WORK INTEGRATED LEARNING

COMP3851B – SEMESTER 2, 2019

Final Report

Authors:

Kepler MANU-LONG

Supervisor:

Kathryn LAMBKIN

November 7, 2019

Contents

Data Engineering Transformation and Business Intelligence Visualisations

1 Background

This year, I completed multiple projects at Komatsu's Smart Solutions Team (SST) [1] in the form of an internship. SST aims to derive insights from data to improve the efficiency of mining operations by "leveraging the Internet of Things (IoT) [2] to rapidly increase onsite insights, powering data-based decision making". [3] Throughout this year, I edited and engineered data that was stored in Komatsu's databases into valuable information to give to end users, usually as a report or in a form where it is easy to understand the end results of the large amounts of data that is being recorded during mining operations.

Komatsu currently uses SQL Server Reporting Service (SSRS) [4] for most of their reporting tasks, but they would like to experiment with Power BI (Figure 1), [5] the tool that I have used this year, due to the demands from customers for a more dynamic business intelligence tool with various end user benefits such as: dynamic reporting, user friendly UI, and ad hoc reporting. [6]

Figure 1: An example of a Power BI dashboard in Power BI Mobile

R [7] can create complicated and statistical graphs that Power BI can't offer. For the purpose of recreating requested graphs from stakeholder into something more useable, I harnessed the power of R and imbedded R files into Power BI to produce the best solution. [8]

Most of the datasets that I worked with have never been looked at before, so there are a multitude of insights that could be used to increase the efficiency of procedures at the mining sites where Komatsu's software has been implemented, which will lead to Komatsu being able to move earth at a faster and more productive rate.

2 Aims

2.1 Overall Goal

The overall goal of the projects undertaken is **to transform data into easy to read and straightforward to understand reports** that will **give the project stakeholders meaningful insights** into the operations that are being undertaken at their mining sites **based on the stakeholders' business questions**. Furthermore, since Komatsu haven't formally used Power BI in the past, this project also serves to **assess the viability of Power BI and suitability of the program** for different styles of datasets.

2.2 Create Visualisations for Payload Data

The data that I was to work with was from a new source that had never been explored; therefore, the first step was to **carry out an exploratory analysis on these datasets**. This included **user engagement and stakeholder liaison to give business context to the data and ascertain requirements**. Further inspection of the datasets made it clear that data quality was a concern, so **data cleansing preparation would be needed to ensure that the data was accurate** and results in the report wouldn't be heavily skewed (Figure 2). [9] **The end product should be data visualised as a user-friendly and interactive report in Power BI** that is easy to understand, even for less-technical users, that **meets stakeholders' needs and answers their business questions, so that they can gain meaningful insights at a glance**.

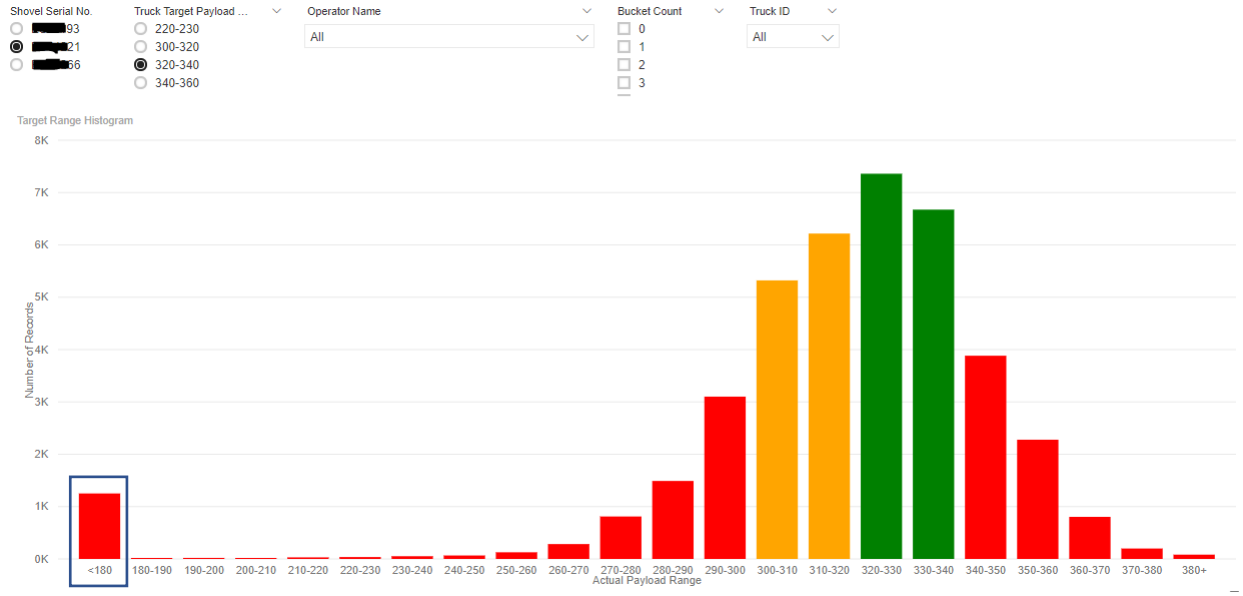


Figure 2: An example of a graph with skewed data due to faulty data

2.3 Create Tools for Ad Hoc Business Questions

I was **provided with four examples of statistical graphs that stakeholders wanted to see replicated within Power BI.** The specialised graphs were outside of the standard Power BI graph library, so I **planned to use various non-default R packages** [10] to be able to deliver these. These graphs include an **XY Plot** (Figure 3) [11], **Scatter Plot** (Figure 4) [12], **Stacked Column Chart** (Figure 5) [13], and a **Heat Map** (Figure 6) [14]. These charts each answer a specific ad hoc business question. **The use of these graphs will ultimately increase reliability of a machine** by; decreasing down-time, anticipating failures and analysing error events when they occur, and finding the optimal combination of results for a machine.

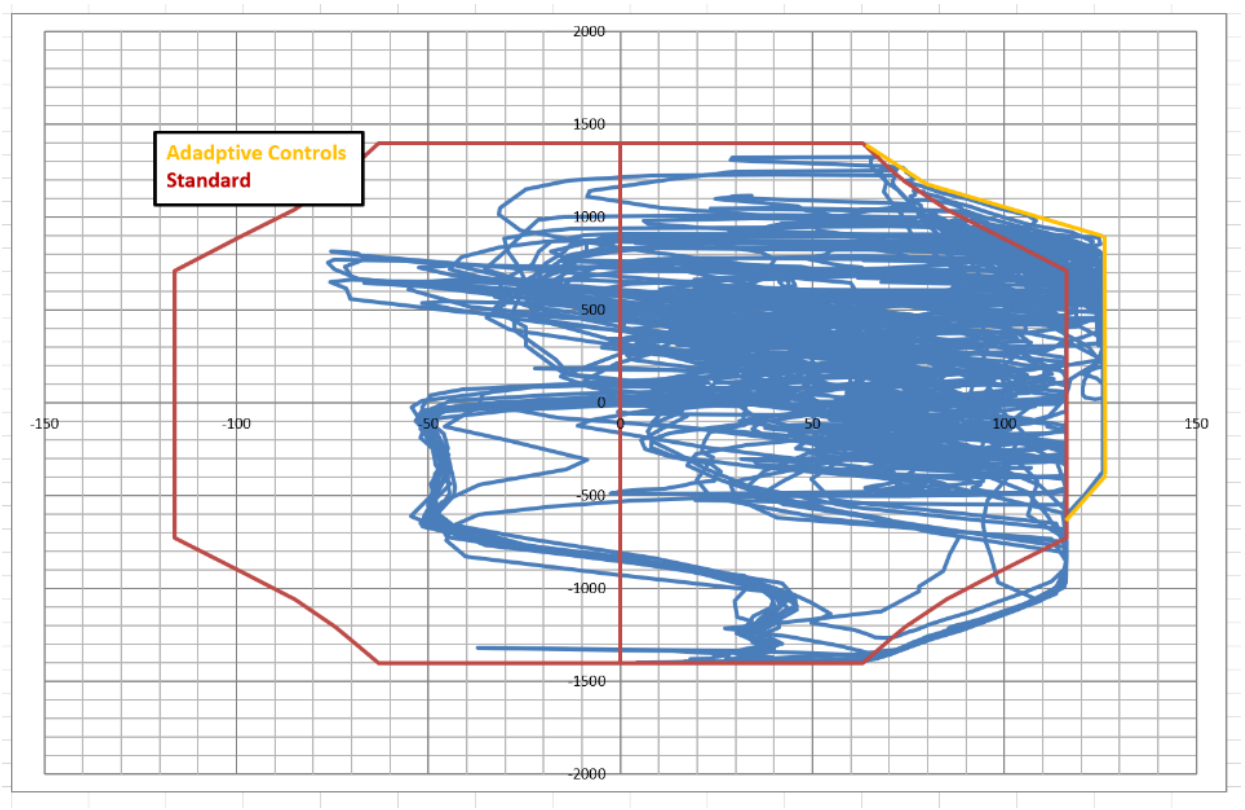


Figure 3: Example XY Plot provided

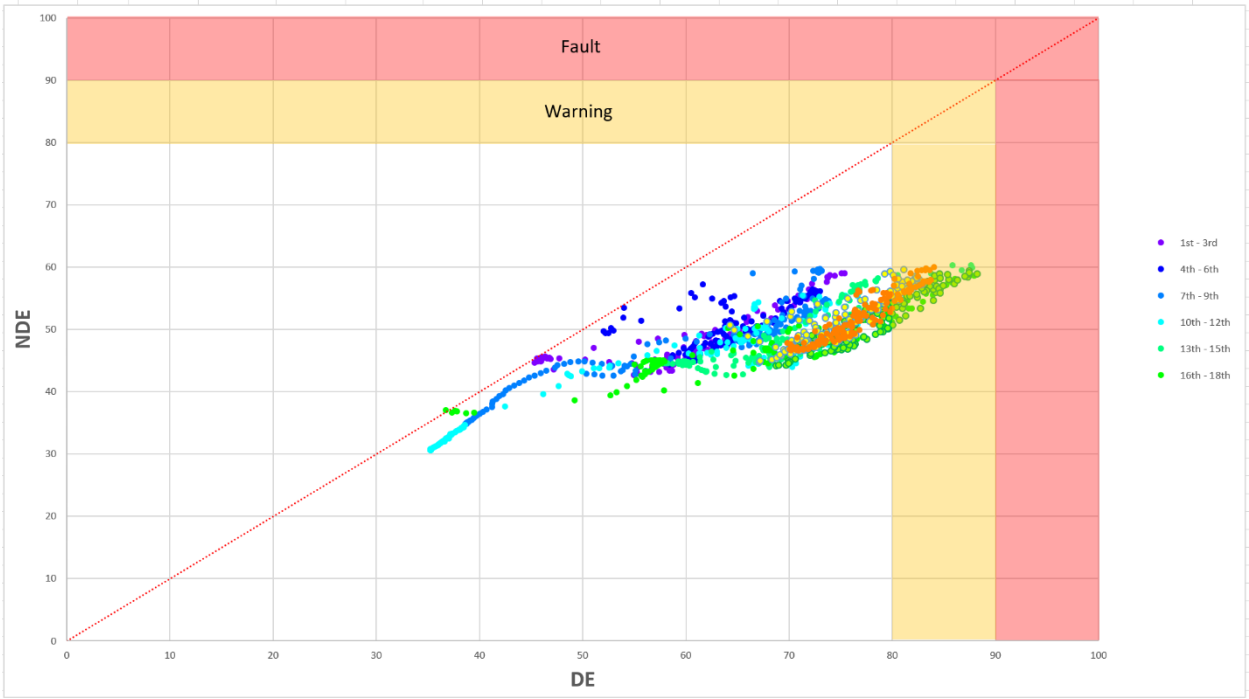


Figure 4: Example Scatter Plot provided

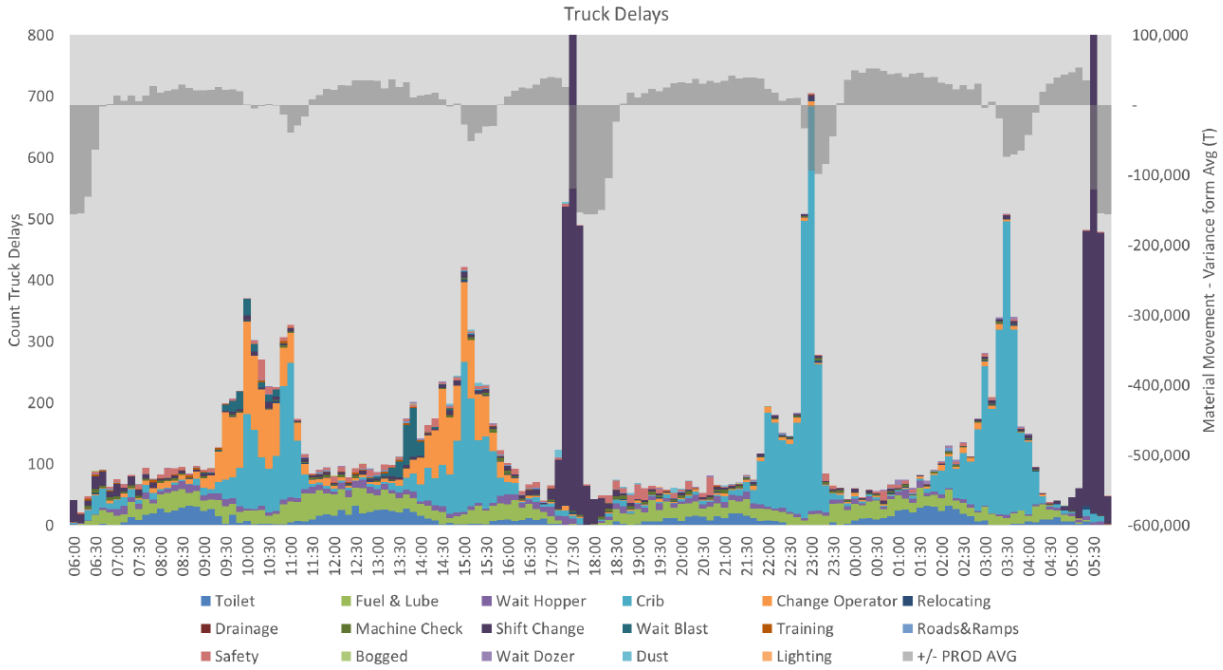


Figure 5: Example Stacked Column Chart provided

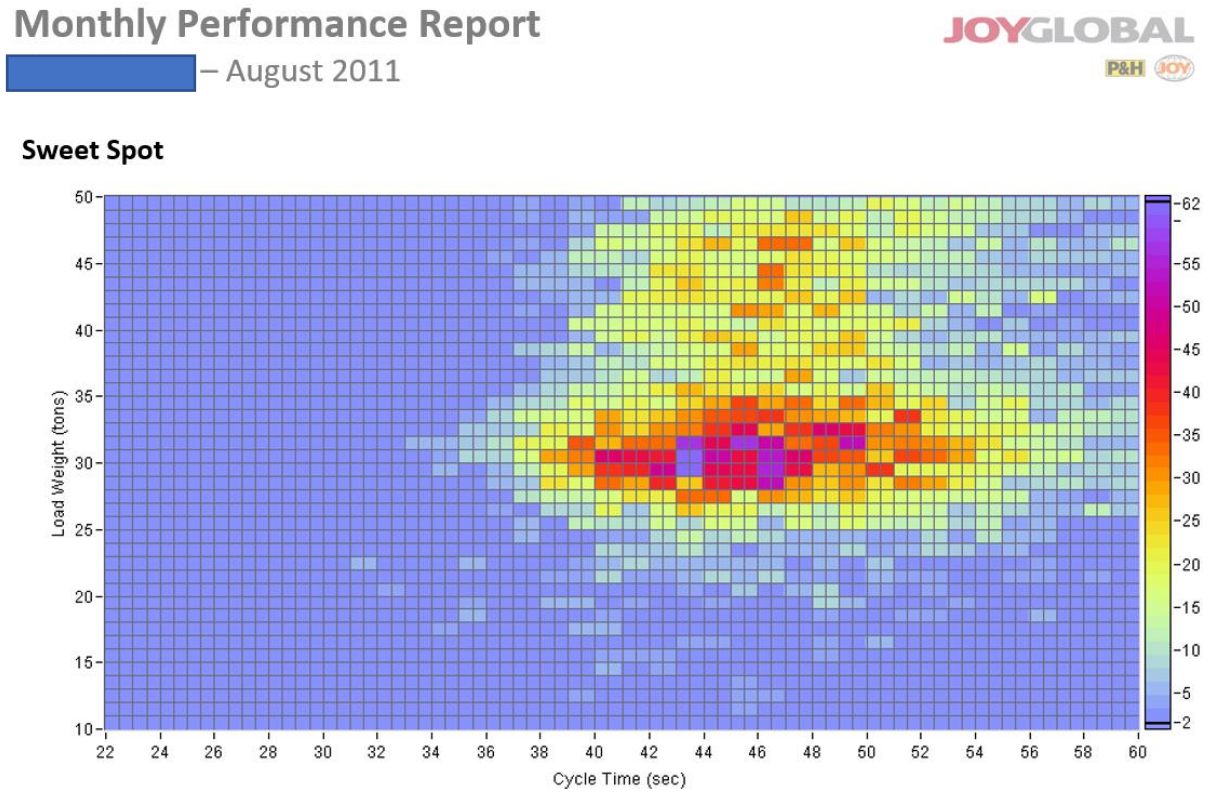


Figure 6: Example Heat Map Sweet Spot example provided

3 Methods

3.1 Creating Reports in Power BI

I facilitated several stakeholder engagement meetings to understand the user requirements and collaborate on the envisioned solution for the report. Upon realisation that some of the data was inaccurate (Figures 7 & 8), **I carried out several data quality investigations**, including crosschecking against other sources for records that were faulty (Figures 9 & 10) **to decide whether proceeding with this report was viable.**

target	loadstart	loadend	dumpend	dip00tons	dip01tons	dip02tons
325	2018-09-06 17:37:34	1536255540	1536256026	113	230	328
326	2018-09-06 19:24:22	1536261921	1536262951	NULL	NULL	NULL
335	2018-09-06 19:53:23	1536263651	1536264815	211	304	NULL
335	2018-09-06 20:19:29	1536265231	1536265714	119	204	266
335	2018-09-06 20:35:49	1536266149	1536266623	NULL	NULL	NULL
323	2018-09-06 20:54:06	1536267251	1536267990	312	NULL	NULL
335	2018-09-06 22:46:23	1536274056	1536274594	94	213	324
323	2018-09-06 23:27:52	1536276547	1536277058	133	226	317
326	2018-09-06 23:31:00	1536276707	1536277389	108	213	291
340	2018-09-07 00:11:25	1536279174	1536280465	NULL	NULL	NULL
355	2018-09-07 01:02:48	1536282281	1536282876	124	216	330
324	2018-09-07 01:11:55	1536282799	1536283323	113	325	NULL
326	2018-09-07 02:57:27	1536289143	1536289874	116	229	320
335	2018-09-07 03:33:17	1536291310	1536292519	115	231	327

Figure 7: dip00tons should always have an integer value

dip00tons	dip01tons	dip02tons	dip03tons	actual_payload
222	203	295	NULL	315
332	177	264	309	334
336	207	309	NULL	329
134	238	314	240	318
118	19	250	364	356
225	314	125	NULL	317
124	241	342	178	353
271	146	NULL	NULL	249
119	191	42	232	231
97	71	201	NULL	191
118	48	193	NULL	196

Figure 8: Implausible for the cumulative number of dips to go down in number

value	day	hours	minutes	seconds
114.500000000	24	0	2	51
108.800000000	24	0	3	35
126.700000000	24	0	4	17

Figure 9: Open TSBD data that should roughly match up with Figure 10(Truck Data dataset)

Crosscheck

loadstart	loadend	dumpend	dip00tons	dip01tons	dip02tons
2019-04-24 00:02:44	1556064261	1556064764	127	234	372

Figure 10: Truck data that should roughly match up with Figure 9(OpenTSBD)

My original plan was to use a simple Impala database connection [15] for the report, but that no longer proved suitable once the data quality issues were discovered. Therefore, **I changed my import method to ODBC, where I could apply filters using SQL to remove faulty records on the fly** (Figure 11). [16]

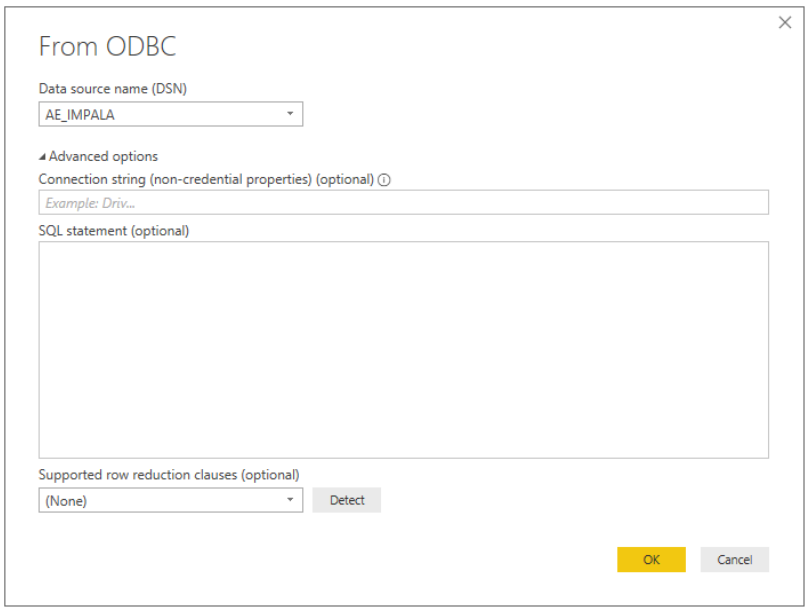


Figure 11: Option to insert SQL statements when importing ODBC data into Power BI

There was a limitation with Power BI where I couldn't customise the order of data within a graph (Figure 12), **so I used Microsoft's Data Analysis Expressions (DAX) language** [17] **to organise the data in the correct order and to code dynamic calculations** corresponding to the filter applied.

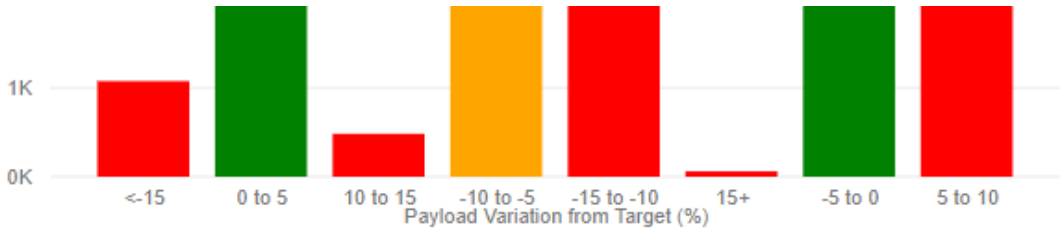


Figure 12: Columns ordered incorrectly, automatically ordered in alphabetical order (pre DAX-use)

3.2 Creating Graphs in R

I had to **rapidly learn R language** in the second semester, using google to find tutorials and by studying the preliminary materials that were given to me. Once I began to understand R's syntax, **I built upon existing research to create statistical graphs. I installed various non-default packages that were necessary transform data** [18] **and draw graphs** [19]. (Figure 13)

```
library(dplyr) #data manipulation, providing a set of verbs that help solve common data manipulation challenges
library(implyr) #SQL back-end to dplyr
library(jsonlite) #JSON parser/generator
library(tidyr) #Organises tabular data
library(stats) #statistical calculations (aggregate function)
library(imputeTS) #Imputates data into time series
library(lubridate) #date time transformation package
library(ggplot2) #draws graphs
```

Figure 13: List of non-default packages used to create the Heat Map graph within R Studio

Variables were crosschecked against the data exploration tools, Apache Hue [20] and Grafana [21], **to validate variable names, check if there were gaps within the time period gaps within the data for a certain machine, and identify if data was complete** and suitable for testing. (Figures 14 & 15)

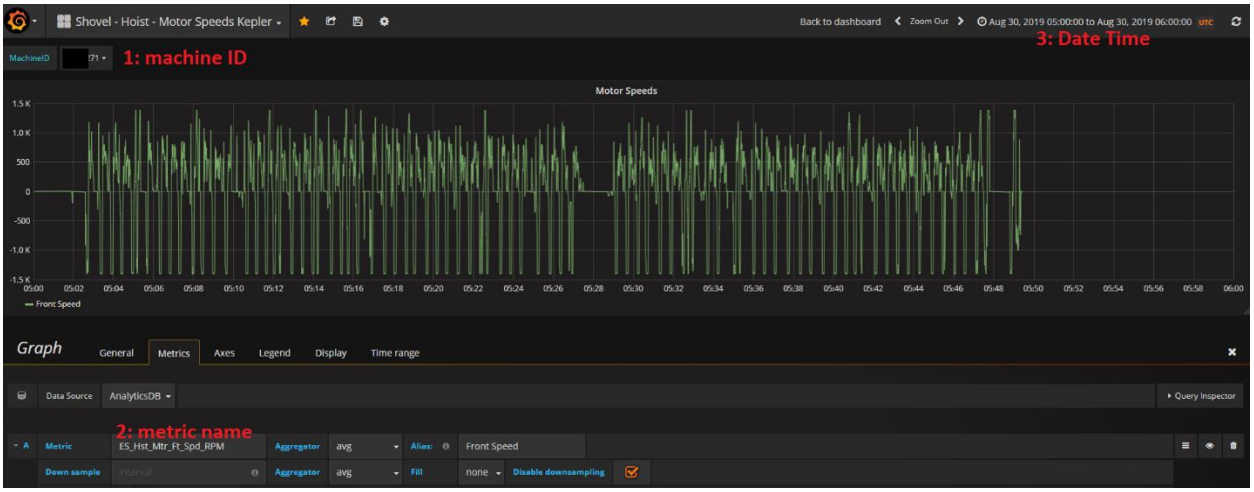


Figure 14: Check Machine ID, Metric Variable name, and Date Time in Grafana to see if the data's accurate and can be used in R

```
#set metrics
machines <- c([redacted]71") 1: machine ID
machineName <- machines[1]

start_ts <- "2019/08/30-05:00:00" 3: Date Time
end_ts   <- "2019/08/30-06:00:00"
metrics <- c("ES_Hst_Mtr_Ft_Spd_RPM","ES_Hst_Mtr_Ft_Torque_Pct")
                2: metric name
```

Figure 15: Insert Machine ID, Metric Variable name, and Date Time that was checked in Grafana

3.3 Combining R and Power BI Functionality

Within Power BI, I imported data from OpenTSDB using an R script (Figure 16). [8]

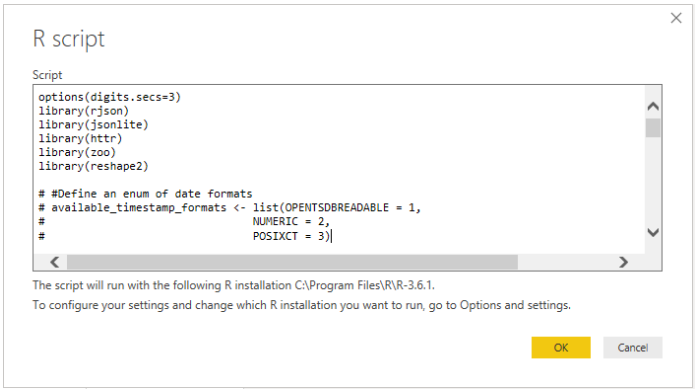


Figure 16: Importing data into Power BI through an R script

When dealing with Big Data [22], there's too much data to import the entirety of the database, so I **needed to facilitate a way that users can import just the subset of data that they need**. There was no native solution for users to easily apply filters to the data that they want to import in Power BI. I didn't want the user to have to go and edit the R import script every time they wanted to change a parameter. Therefore, **to allow a parameter to be changed in simple and easy for the user to understand, when a parameter is changed in the Power BI file, the inserted parameter is passed to the imbedded R script** [23], which is then rerun when the user clicks on a popup prompt. This then **imports only a certain subset of data instead of the whole dataset.**

I wanted to populate dropdown lists of available machines for a user-friendly experience, so I needed to imported datasets, filter for unique values for machines, and then convert that dataset into a list so that they could be configured to populate a parameter option dropdown list. [24]

Back and forth conversation via Outlook emails [25] and Microsoft Teams [26] **were frequent to refine requirements and ensure that the stakeholders' needs were met.**

4 Results

4.1 Payload Data Report

I collated my findings, listing all the problems I found with the dataset's data quality and sent it to the relevant stakeholders so that they could solve these issues at the source of the problem. (Figure 17)

Figure 17: List of faulty record categories from Truck Data Quality document

I built a library of SQL code snippets that can be reused not only on this project, but by other team members on their future work. E.g. I assisted my supervisor in rounding timestamps to the lowest 10-minute period which I then implemented into my own SQL query from this library of SQL code snippets when creating my own report. [27]

As a result of choosing the ODBC connection for Power BI reports which imports the data into the file, the file can be distributed to stakeholders without requiring VPN access, any configuration, or even internet access.

I created a histogram and a data table that showed information related to the histogram directly underneath it (Figure 18). I was able to sort the percentage variance histogram via this use of DAX (Figure 19),[28] as well as insert a percentage of total records column in the Dipper Averages Table based on the selected machine serial number (Figure 20). [29] The functions I developed using DAX codes can be reused for future reference and projects.

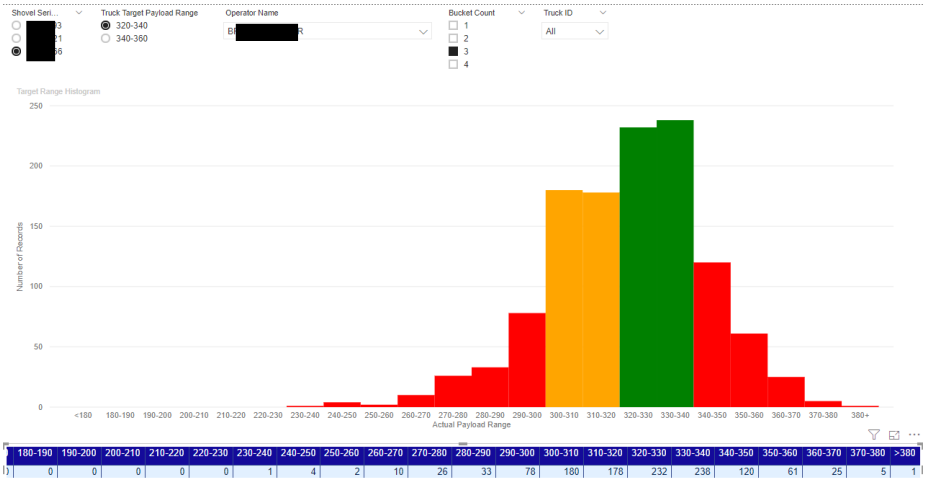


Figure 18: Histogram with data table of exact amount of records for each bin underneath graph

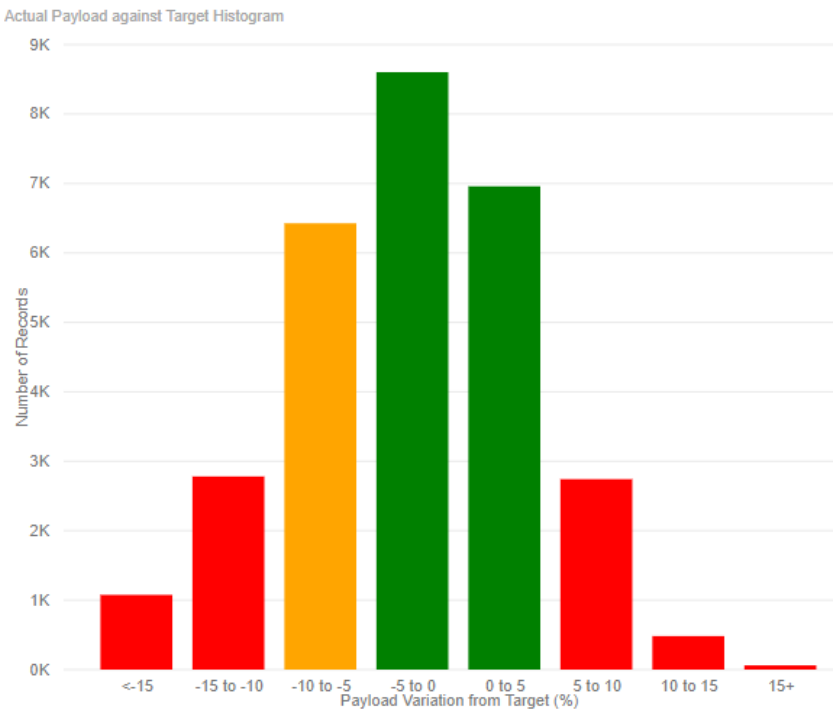


Figure 19: Correct order of figure 12 because of DAX used

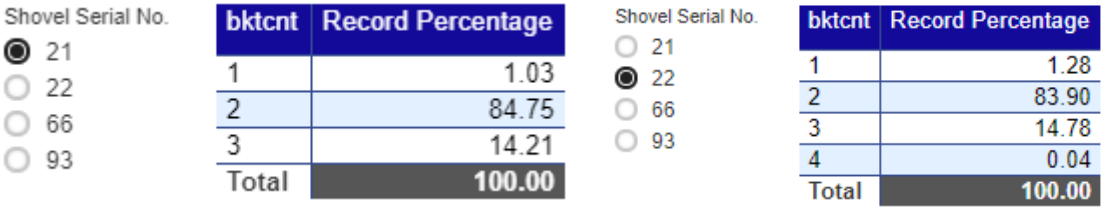


Figure 20: Dynamic Record Percentage column based on selected filter

4.2 Ad Hoc Business Questions Tools

I have stored the completed R scripts in SST's Azure DevOps Boards [30] on the related ticket for the task (Figure 21) so that future developers can see what I have done and continue further development upon my solution. My solution was only constrained to shovels, so a future developer can expand functionality within the R scripts to allow for loaders and other machines if so desired.

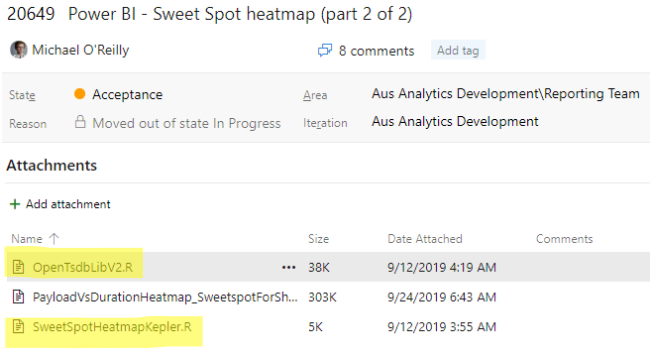
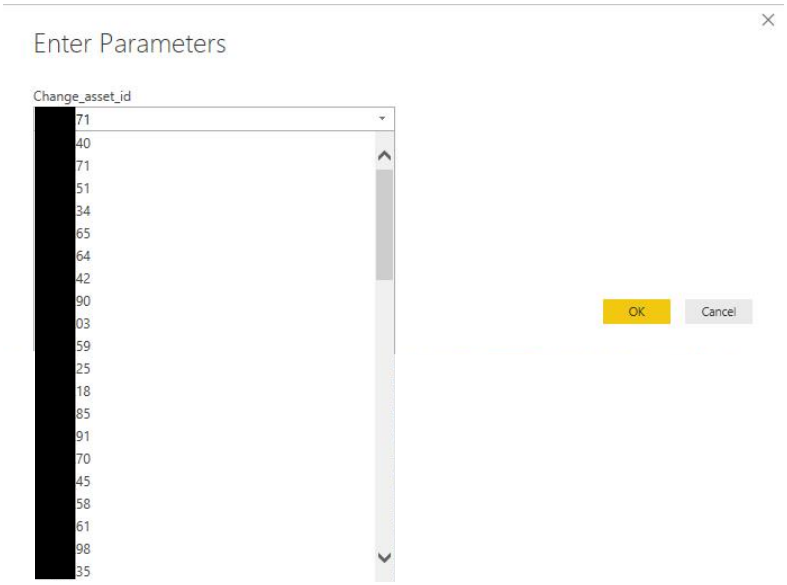


Figure 21: R files uploaded to the appropriate DevOps ticket for future use and reference

One of the constraints of Power BI is that importing large datasets is very slow. I solved this by only importing the data for a certain machine and time period through parameters which leads to faster importing because of the lesser amount of data.

I have also implemented a user-friendly drop-down list with preloaded selections for only AC shovels (Figure 22) since the ad hoc business question tools were only to be applied for AC shovels. These preloaded dropdown lists are based off a separate dataset within the same file that imports with ODBC to populate the dropdown lists (Figure 23). This is setup in such a way that developers can easily expand functionality with DC shovels and loaders if required.



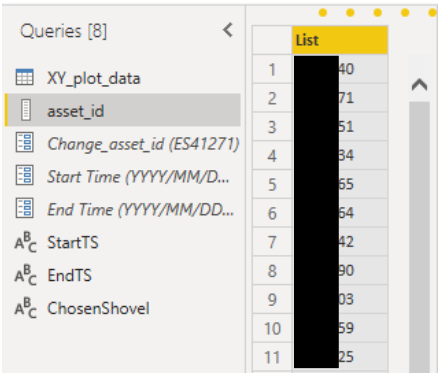


Figure 23: Asset ID (AC Shovels only) dataset converted to list

Figure 22: Drop down list with preloaded selections based off list in Figure 23

All the files that I created were uploaded to Power BI's online application [31] SST workspace (Figure 24). However, since Power BI is a relatively new tool that is still under development by Microsoft, parameters aren't functional yet on the Online Workspace for Power BI. [32] Therefore, employees would need to take the extra step to download the report to edit the parameters and view information for other machines and timestamps that aren't the default settings.

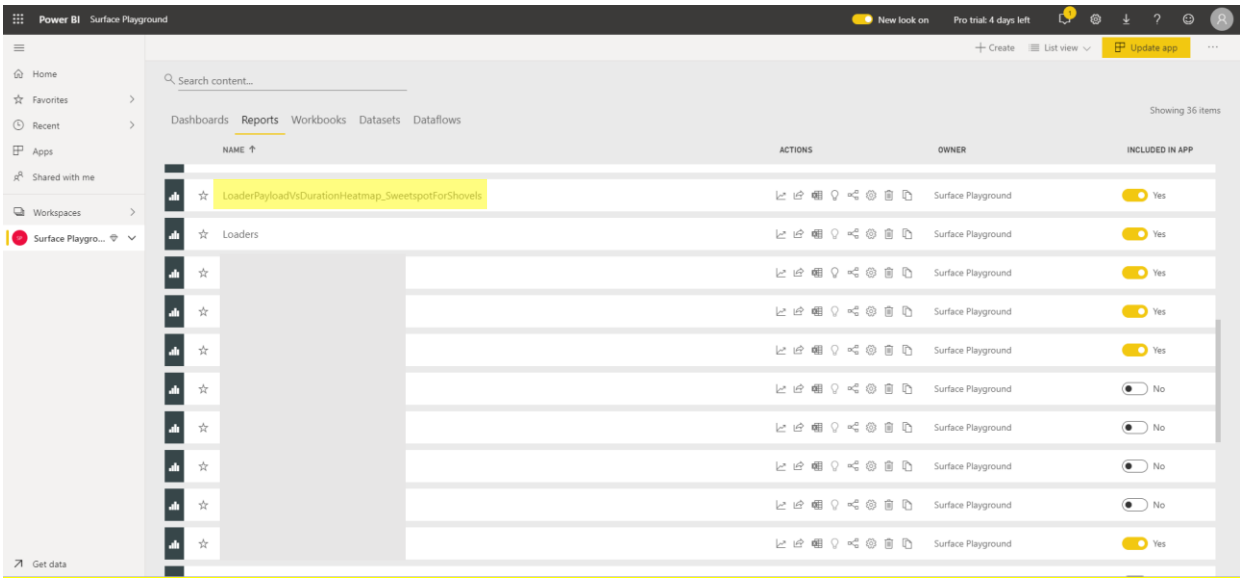


Figure 24: Power BI online workspace for surface mining production reports. Highlighted is a report that I uploaded.

There are some errors that occur when trying to run these Power BI files with Power BI's default settings, such as when trying to import data from OpenTSDB through an R script, and importing data from OpenTSDB with an ODBC connection at the same time that's being passed through to the R script, an error occurred, which I figured to be a privacy issue which was solved when changing the Privacy Levels setting of Power BI to "Always ignore Privacy Level Settings". [33] I included an instructions page in the report file with a step-by-step tutorial on how to solve this issue with developer's comments (Figure 25), as well as instruction on how to change parameters [34] in Power BI (Figure 26). I also commented on all the R files I created so developers can continue expanding on my work if required.

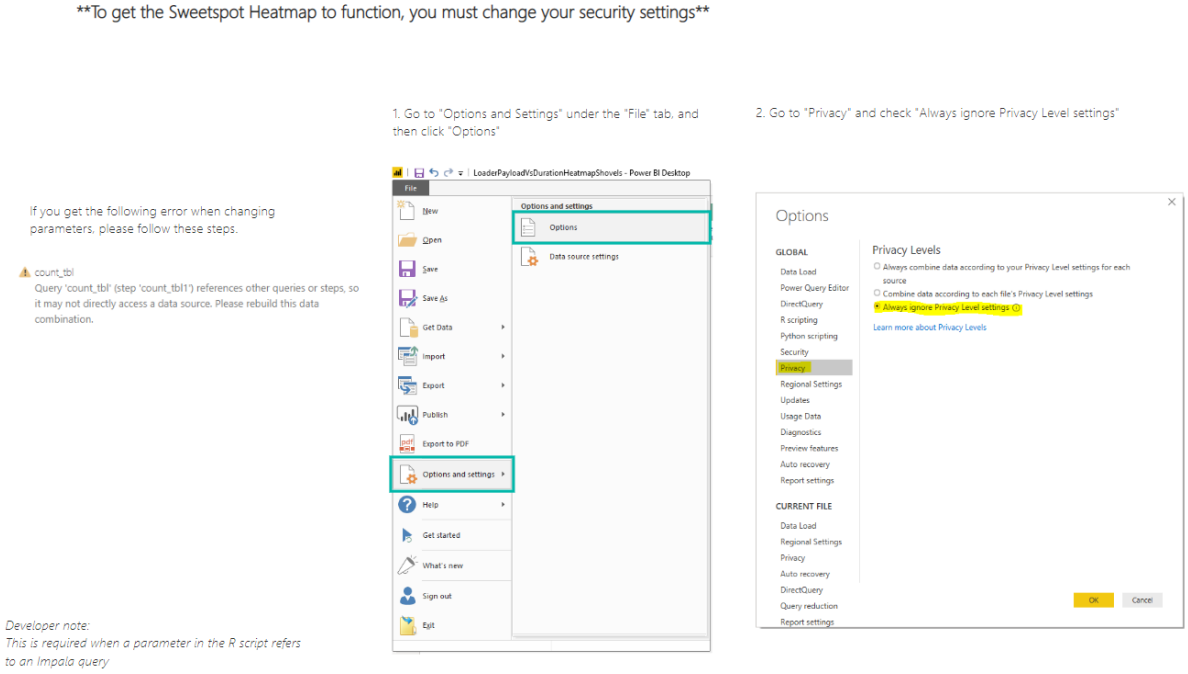


Figure 25: Instructions page on how to solve an error that appears when trying to run the file with default Power BI settings + developer's note

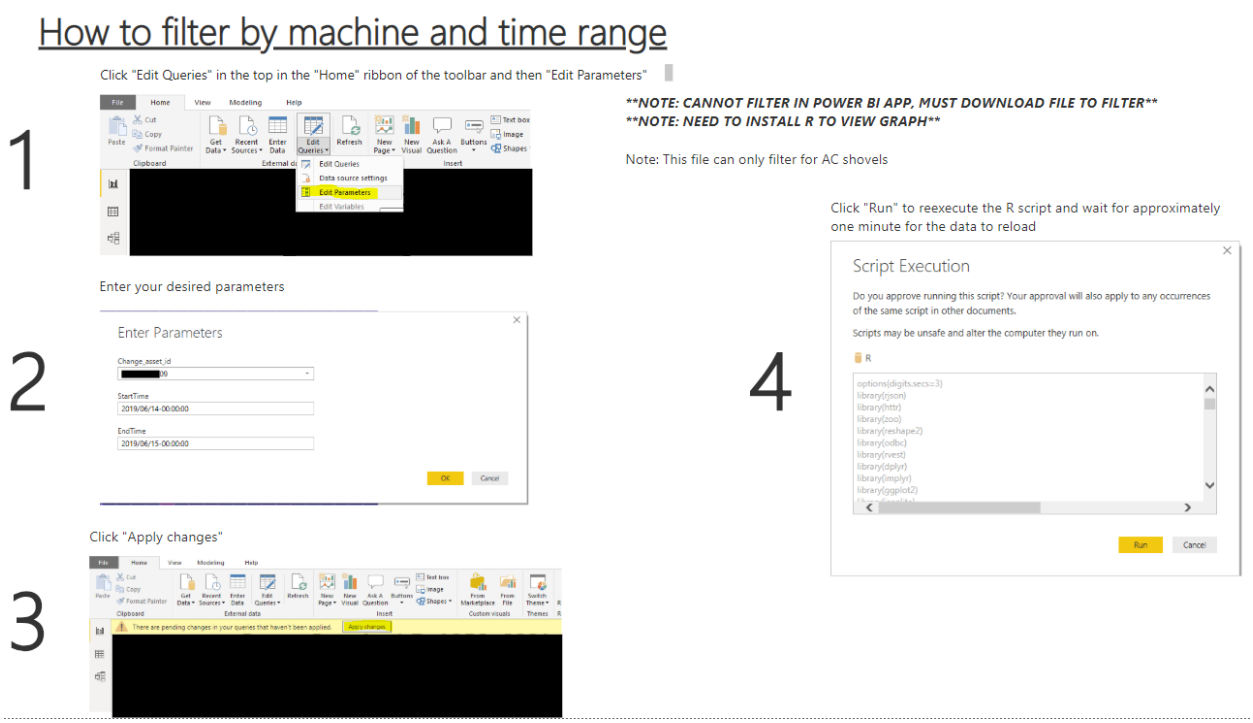


Figure 26: Instructions on how to change parameters within Power BI so even low-technical users can use report file

4.3 Power BI Viability Assessment

Power BI has many merits and limitations, but in the end, I have concluded that Power BI isn't a suitable product for the projects that Komatsu wants to use it for. There are many limitations that prevent Power BI from being the ultimate reporting service that Microsoft wants it to be. Power BI gives end users access to the self-serve, easy to use tools that allow them more control over reports [35] than SSRS, which is why Komatsu is pushing forward the notion of using Power BI over SSRS. However, development within Power BI is much more limited than what is possible within SSRS. [36]

Ideally, we would want to use Power BI's Direct Query functionality and link the report directly to the data source so that it's possible to build visualisations over larger datasets, and always uses current data without having to import the data every time. [37] However, there are too many limitations with using Direct Query, such as; not being able to filter the data that will be imported with SQL and a limitation of 1 million rows which will be insufficient for projects larger than the small-scale ones that I have undertaken over the year. Therefore, I resorted in using the Import option over Direct Query, yet on larger projects, the load time would be far too long for end users when loading a report. Import also has a 1GB limit, [38] which is insufficient for the Big Data projects.

A good data model is also required for efficient use of Power BI. Missing data and data inaccuracy were much higher than anticipated resulting in difficulties when trying to create an accurate report. I recommend they use SQL Server Integration Services (SSIS) [39] to transform their data first before importing it into Power BI.

Power BI receives regular monthly updates [40] and is a good effective tool to use for small to medium size businesses to analyse their data, but for large companies with a tremendous amount of data inflow every second like Komatsu, developer-wise, it is simply lacking basic functionality that SSRS has.

5 Ethics

One of the biggest ethics considerations when handling data is the European Union General Data Protection Regulation (GDPR). This regulation states that data related to a subject should not be able to be related back to that subject from merely observing said data without the use of additionally stored information. [41] i.e. Looking at data based on an Employee, just by looking at this data, it should not be able to be traced back to the employee, so names should be either pseudonymised of fully anonymised. The data should at all means not be publicly available without explicit and informed consent from said subject or owner of said subject.

All measures possible are taken, such as only keeping company data on the company's computers and making sure all USBs are locked with a password, so that the company doesn't leak confidential information to the public. Sensitive data should not ever be publicly revealed to people besides the management of a mining site, but non-sensitive data, such as data about the performance of an electrical shovel, may be viewed by even the employees of that mining site because it doesn't involve information about other personnel. 2-factor authentication [42] was used when accessing work emails and Microsoft teams outside of the workplace.

Considering these ethics considerations, I ensured all employee names and shovel ids were pseudonymised in my presentation and this report so they can't be identified and traced back to their allocated mining site.

6 Self-reflection, project management and ethics

6.1 Personal Experience

SST has a wide and well set-up workspace with a relaxed workplace environment and friendly staff. I mainly interacted with my supervisor and the main project stakeholders, sometimes emailing other personnel when reporting problems about data quality and machines or when in need of administration help. These stakeholders included people from differing technical backgrounds, such as engineers and marketing assistants, and not just IT specialists. I met with my supervisor at the end of each day to show her my progress and get feedback and suggestions, which I proceed to work on the next day I was at the office.

Any time that I encountered a problem that I was uncapable of doing with my own knowledge, I would first search for the solution through Google and implement it, or slightly alter a solution that I found online. Most problems and challenges were overcome through this method, but If I couldn't find the solution online, I would contact and discuss the problem with my supervisor. For example, when I first initially started studying R, my supervisor helped me understand the syntax by sitting down with me and discussing the research that was supplied. They would advise me what they thought the best solution to the problem was, but it was ultimately up to me about how I wanted to approach each problem. I was advised that if I took too long to solve a problem and my supervisor wasn't available, just skip it and move onto the next thing.

e.g. 1) The connection to Apache Impala wasn't working in R. After Googling the problem, I realised that only the 32-bit version of the Cloudera ODBC Driver for Impala was installed and the proceeded in installing the 64-bit version which solved my problem. [43]

e.g. 2) The sweet spot heat map didn't appear as originally intended (Figure 27), so I utilised a solution from a website to fix the problem. [44] The heatmap still didn't appear as intended (Figure 28), therefore, looking closely at the graph, I realised I should enter a value of 0 into the columns that are blank so that the fill function could fill up the rest of the heatmap automatically. (Figure 29)

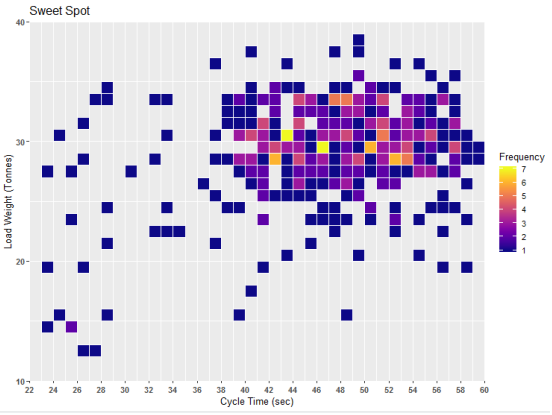


Figure 27: Sweet Spot Heatmap 1st iteration – only tiles that have more than a frequency of 0 are filled

Use solution found online [44]

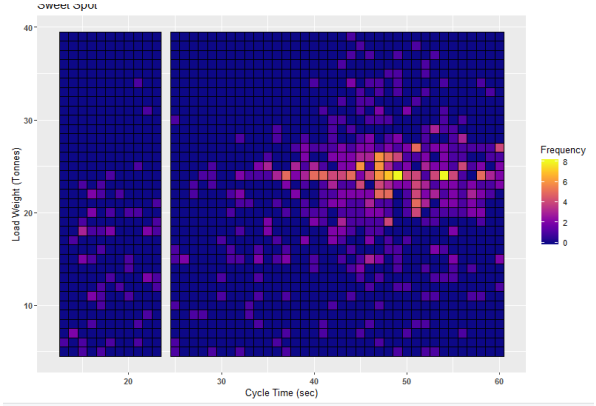


Figure 28: Sweet Spot Heatmap 2nd iteration - columns in which there are no values for frequency aren't filled properly, leaving gaps in the graph

Use observation skills and gained knowledge to solve issue

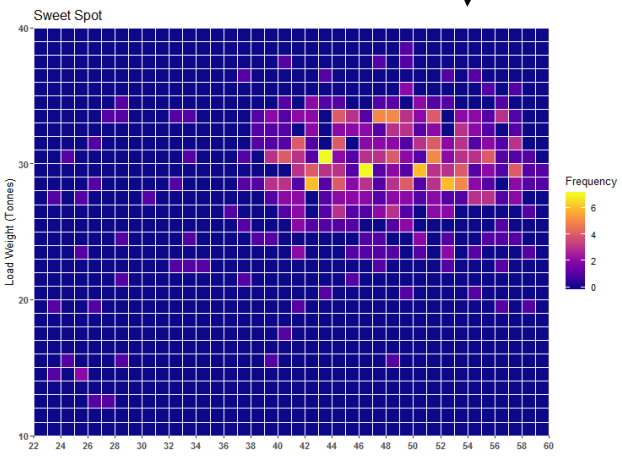


Figure 29: Sweet Spot Heatmap 3rd iteration - all tiles filled properly with no unnecessary grey border around outside of graph

6.2 Project Management Tools and Procedures

Throughout the year, I commuted to Komatsu approximately 3 days a week and worked 5-hour days. Komatsu uses Azure DevOps Boards for their project management tool of choice. They have sprints named alphabetically running on 2-week periods. My tasks were managed by myself on a Kanban Board (Figure 30). [45] I planned the tasks in tickets at the start of the second semester and wrote what I needed to achieve by the end of the task within each ticket. These tickets were linked to other relevant tickets with information related to the tasks that I had to undergo. I also assigned the Reviewer, who would be my supervisor, the Acceptor, who would be the main stakeholder of the task/project, and the Implementer, who would be myself. (Figure 31) [46]

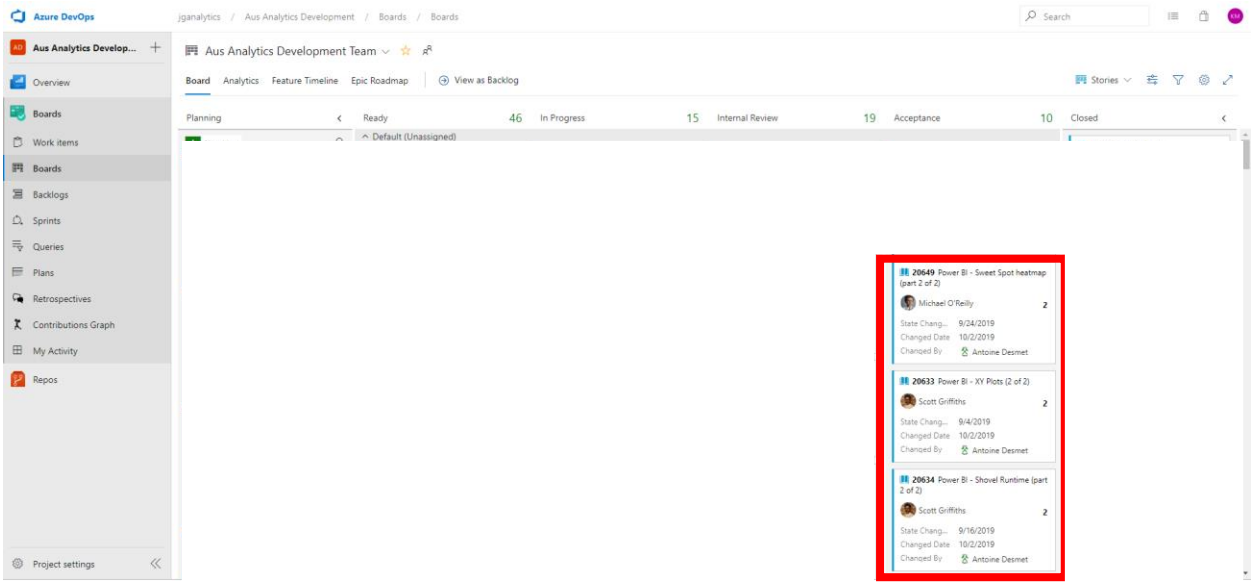


Figure 30: Komatsu's Smart Solutions Team's Azure DevOps Boards Kanban Board. My tasks are enclosed in the red box under acceptance as I have finished all tasks and am waiting for the main stakeholder to approve of the results.

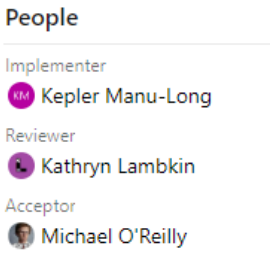


Figure 31: Example of assigned Implementer (myself), Reviewer (supervisor), and Acceptor (main stakeholder for task)

Files that were completed that day were attached to the ticket at the end of every shift on the off chance that a stakeholder would want to view progress on the task. I would commonly comment on these tasks with what I had completed within that day and what I had lined up next for the task. I would also comment and tag stakeholders when I needed more information or resources for a certain task, then change the status of that task to “Internal Approval” while I continue to work on another task (Figure 32). In this sense, Azure DevOps Boards made it simple for me to multitask many projects at once because it was easy to understand which tasks still needed to be reviewed by a stakeholder, and which tasks were safe to continue to work on.

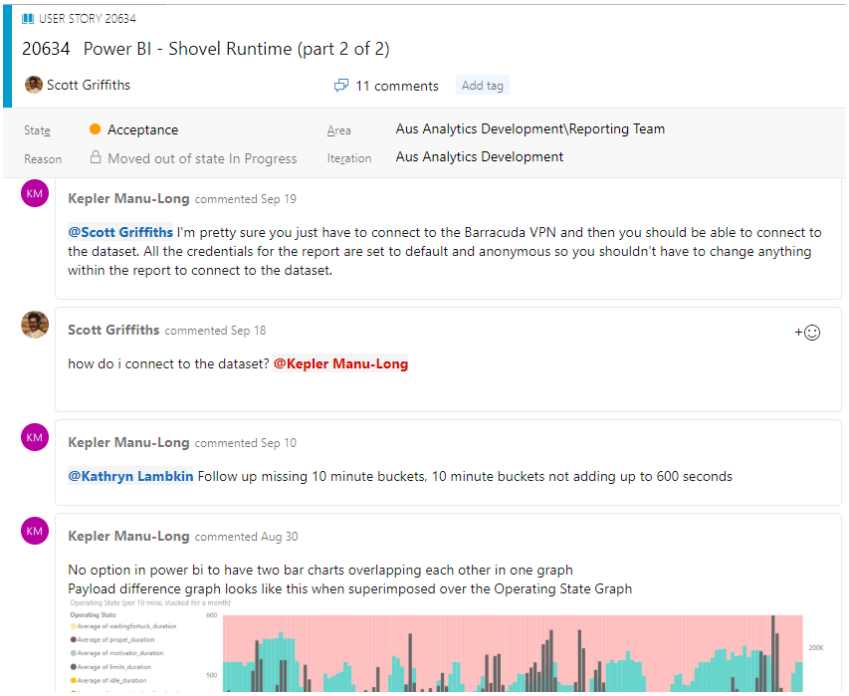


Figure 32: Example of communication with main stakeholder about project on a DevOps ticket. State would be in "Internal Approval" but picture was taken after task was completed.

Microsoft Teams and Outlook is used for communication within the Smart Solution's Team. WebEx video conference calls [47] are used for large meetings with many participants. Any conversations/task threads that needed to be officially kept track of were sent via email and meetings were scheduled in Outlook. For immediate problems and when asking for general advice, communication was undergone via Microsoft Teams.

It wasn't necessary to make a project requirements document and all notes were taken down either in my laptop or my notebook. Notes were taken frequently throughout each meeting, which frequently widened the scope of the project and I also had four tasks assigned to me within the second semester, so I always had something to work on. I always made sure to comment my code as a method of documentation to make sure my code is maintainable and reusable.

I approached this project with agile project development practices. [48] I would complete a prototype of the report or one of the ad hoc business question tools, show it to my supervisor or a stakeholder, then wait to receive feedback while I was working on another project, to see what aspects needed to be pivoted in another direction, or to receive approval.

There were no time constraints on this project, provided I finished the project within the time allowed for COMP3851A/B. My supervisor and project sponsor were satisfied with my working pace, provided I regularly produced deliverables.

6.3 Workplace and Project Reflection

Looking back at the project and my experience at Komatsu, one of the things that I would do differently if restarting the project would be to be less hesitant when contacting stakeholders. At the beginning of my internship, I was too cautious of contacting higher-ups and stakeholders of the project because it

was intimidating, considering it was my first time at a large-scale company like Komatsu. I thought I would be interrupting them or taking their time away from other important responsibilities. However, I learned throughout my internship that it was important to be a bit persistent with contacting these personnel as they are in fact interested in my work, even if they are busy. At the end, I was able to contact them more carefreely on Microsoft Teams as I needed to get feedback for my Ad Hoc Business Question Tools at a faster rate. Things could have been done much faster if, instead of waiting for my supervisor to come check my work every time and ask her if I should send a file off for it be checked or not, I just contacted a stakeholder via email to ask for feedback instead.

Another thing I slightly regret is not meeting more of the team. I would have like to chat with more of the team informally and see what projects they were doing and discuss my own project with them. This way I could have created more personal relations that would be useful for my future career. Even within the workspace that I was in, I kept more to myself instead of spreading out and having conversations with people at other desks, which I recognise now as wasted potential.

However, overall, my work integrated learning experience has been very successful, and I gained a multitude of practical and communicational skills that will assist me on my future endeavours.

7 References

[1] Komatsu.com.au. (2019). *Komatsu Australia*. [online] Available at: https://www.komatsu.com.au/ [Accessed 4 Jun. 2019].

[2] Minerals, W. (2019). *How Industrial Internet of Things is changing the mining industry*. [online] MINING.com. Available at: http://www.mining.com/web/industrial-internet-things-changing-mining-industry/ [Accessed 4 Jun. 2019].

[3] Mining.komatsu. (2019). *Smart Solutions*. [online] Available at: https://mining.komatsu/docs/default-source/non-product-documents/services/joysmart-solutions/smart-solutions-brochure.pdf?sfvrsn=4aa50c6b_65 [Accessed 5 Apr. 2019].

[4] Sparkman, M., Guyer, C., Schmidtke, R., Saxton, A., Hassler, M., Kumar, S., Milener, G., Ghanayem, M. and Howell, J. (2019). *What is SQL Server Reporting Services (SSRS)?*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/sql/reporting-services/create-deploy-and-manage-mobile-and-paginated-reports?view=sql-server-2017 [Accessed 4 Jun. 2019].

[5] Powerbi.microsoft.com. (2019). *Power BI | Microsoft Power Platform*. [online] Available at: https://powerbi.microsoft.com/en-us [Accessed 5 Nov. 2019].

[6] Johnson, S. (2018). *So Why Use Microsoft BI? Here Are Its Key Benefits | Inside Info*. [online] Insideinfo.com.au. Available at: https://www.insideinfo.com.au/microsoft-business-intelligence-and-analytics/so-why-use-microsoft-bi-here-are-its-key-benefits [Accessed 5 Nov. 2019].

[7] R-project.org. (2019). *R: What is R?*. [online] Available at: https://www.r-project.org/about.html [Accessed 5 Nov. 2019].

[8] Iseminger, D., Sebolt, M., Blythe, M., Saxton, A., Hu, J. and Peterson, T. (2019). *Running R Scripts in Power BI Desktop - Power BI*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/desktop-r-scripts [Accessed 5 Nov. 2019].

[9] Downs, E. (2018). *The Staggering Impact of Dirty Data | MarkLogic*. [online] MarkLogic. Available at: https://www.marklogic.com/blog/the-staggering-impact-of-dirty-data/ [Accessed 5 Nov. 2019].

[10] Wickham, H. (n.d.). *Introduction · R packages*. [online] R-pkgs.had.co.nz. Available at: http://r-pkgs.had.co.nz/intro.html [Accessed 5 Nov. 2019].

[11] Exploreanalytics.com. (2015). *XY Chart - Explore Analytics: The Wiki*. [online] Available at: https://www.exploreanalytics.com/wiki/index.php?title=XY_Chart [Accessed 5 Nov. 2019].

[12] Mathsisfun.com. (2017). *Scatter (XY) Plots*. [online] Available at: https://www.mathsisfun.com/data/scatter-xy-plots.html [Accessed 5 Nov. 2019].

[13] Exceljet.net. (n.d.). *Stacked Column Chart | Exceljet*. [online] Available at: https://exceljet.net/chart-type/stacked-column-chart [Accessed 5 Nov. 2019].

[14] Hurley Halls, S. (2019). *What is a Heat Map, How to Generate One, Example and Case Studies*. [online] The Daily Egg. Available at: https://www.crazyegg.com/blog/understanding-using-heatmaps-studies/ [Accessed 5 Nov. 2019].

[15] Saxton, A., Iseminger, D., Blythe, M., Agiewich, R., Peterson, T., Hu, J., Sparkman, M. and Brown, S. (2019). *Connect to an Impala database in Power BI Desktop*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/desktop-connect-impala [Accessed 4 Jun. 2019].

[16] Iseminger, D., Harvey, B., Schonning, N., Blythe, M., Sparkman, M., Saxton, A., Petersen, T., Hu, J. and Sebolt, M. (2019). *Connect to data using generic interfaces in Power BI Desktop*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/desktop-connect-using-generic-interfaces#data-sources-accessible-through-odbc [Accessed 4 Jun. 2019].

[17] Iseminger, D., Petersen, T., Blythe, M., Duncan, O., Saxton, A., Sebolt, M., Hu, J., Schonning, N. and Koudelka, M. (2019). *DAX basics in Power BI Desktop*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/desktop-quickstart-learn-dax-basics [Accessed 4 Jun. 2019].

[18] Dplyr.tidyverse.org. (n.d.). *A Grammar of Data Manipulation*. [online] Available at: https://dplyr.tidyverse.org/ [Accessed 5 Nov. 2019].

[19] Ggplot2.tidyverse.org. (n.d.). *Create Elegant Data Visualisations Using the Grammar of Graphics*. [online] Available at: https://ggplot2.tidyverse.org/ [Accessed 5 Nov. 2019].

[20] Gethue.com. (2019). *Hue*. [online] Available at: http://gethue.com/ [Accessed 4 Jun. 2019].

[21] Grafana Labs. (2019). *Grafana*. [online] Available at: https://grafana.com/ [Accessed 4 Jun. 2019].

[22] Oracle.com. (2019). *What is Big Data?*. [online] Available at: https://www.oracle.com/au/big-data/guide/what-is-big-data.html [Accessed 4 Jun. 2019].

[23] Sheldon, R. (2018). *Power BI Introduction: Working with Parameters in Power BI Desktop —Part 4 - Simple Talk*. [online] Simple Talk. Available at: https://www.red-gate.com/simple-talk/sql/bi/power-bi-introduction-working-with-parameters-in-power-bi-desktop-part-4/ [Accessed 5 Nov. 2019].

[24] Guy in a Cube (2019). *Populate a Power BI parameter list using a query*. [video] Available at: https://www.youtube.com/watch?v=7TfV8qTUyfU [Accessed 5 Nov. 2019].

[25] Outlook.live.com. (2019). *Microsoft free personal email*. [online] Available at: https://outlook.live.com/owa/ [Accessed 4 Jun. 2019].

[26] Products.office.com. (2019). *Microsoft Teams*. [online] Available at: https://products.office.com/en-us/microsoft-teams/group-chat-software [Accessed 4 Jun. 2019].

[27] Lang, P. (2010). *Round date to 10 minutes interval*. [online] Stack Overflow. Available at: https://stackoverflow.com/questions/2192424/round-date-to-10-minutes-interval [Accessed 4 Jun. 2019].

[28] Community.powerbi.com. (2018). *change column based on filter*. [online] Available at: https://community.powerbi.com/t5/Desktop/change-column-based-on-filter/m-p/498138#M232332 [Accessed 4 Jun. 2019].

[29] Iseminger, D., Petersen, T., Blythe, M., Duncan, O., Saxton, A., Sebolt, M., Hu, J., Schonning, N. and Koudelka, M. (2019). *DAX basics in Power BI Desktop*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/desktop-quickstart-learn-dax-basics [Accessed 4 Jun. 2019].

[30] Thurman, P., Danielson, S., Lewis, A. and Comley (2019). *Understand what you get with Azure Boards - Azure Boards*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-au/azure/devops/boards/get-started/what-is-azure-boards?view=azure-devops&tabs=basic-process [Accessed 5 Nov. 2019].

[31] Sparkman, M., Simpson, D., Saxton, A., Blythe, M., Rigney, R., Petersen, T., Lee, D., Hu, J. and Hart, M. (2019). *Collaborate in a classic workspace - Power BI*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/service-collaborate-power-bi-workspace [Accessed 5 Nov. 2019].

[32] Community.powerbi.com. (2017). *change parameter values on PowerBI Online*. [online] Available at: https://community.powerbi.com/t5/Service/change-parameter-values-on-PowerBI-Online/td-p/238281 [Accessed 5 Nov. 2019].

[33] Social.technet.microsoft.com. (2015). *Formula.Firewall: Query 'x' (step 'Source') references other queries or steps and so may not directly access a data source. Please rebuild this data combination.*. [online] Available at: https://social.technet.microsoft.com/Forums/en-US/4e409d5b-260d-4b6c-baca-10cc94d61df5/formulafirewall-query-x-step-source-references-other-queries-or-steps-and-so-may-not?forum=powerquery [Accessed 5 Nov. 2019].

[34] Guy in a Cube (2017). *Dynamic Power BI reports using Parameters*. [image] Available at: https://www.youtube.com/watch?v=iiNDq2VrZPY [Accessed 5 Nov. 2019].

[35] Ulag, A. (2018). *Power BI expands self-service prep for big data, unifies modern and enterprise BI*. [online] Powerbi.microsoft.com. Available at: https://powerbi.microsoft.com/en-us/blog/power-bi-expands-self-service-prep-for-big-data-unifies-modern-and-enterprise-bi/ [Accessed 4 Jun. 2019].

[36] Team, D. (2018). *SSRS Vs Power BI - 11 Major Difference Between Power BI Vs SSRS - DataFlair*. [online] DataFlair. Available at: https://data-flair.training/blogs/ssrs-vs-power-bi/ [Accessed 4 Jun. 2019].

[37] Iseminger, D., Blythe, M., Saxton, A., Petersen, T., Sebolt, M., Hamilton, B., Hu, J., Caserio, C. and Cummin (2019). *Use DirectQuery in Power BI Desktop - Power BI*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/desktop-use-directquery [Accessed 5 Nov. 2019].

[38] Mabee, D., Iseminger, D., Saxton, A., Thompson, W., Petersen, T., Hu, J., Blythe, M., Clem, E. and Sparkman, M. (2019). *Data sources for Power BI - Power BI*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/power-bi/service-get-data [Accessed 5 Nov. 2019].

[39] Rouse, M. (2014). *What is Microsoft SSIS (SQL Server Integration Services)? - Definition from WhatIs.com*. [online] SearchSQLServer. Available at: https://searchsqlserver.techtarget.com/definition/Microsoft-SSIS-SQL-Server-Integration-Services [Accessed 5 Nov. 2019].

[40] Community.powerbi.com. (2016). *Power BI Service update frequency*. [online] Available at: https://community.powerbi.com/t5/Service/Power-BI-Service-update-frequency/td-p/60755 [Accessed 5 Nov. 2019].

[41] Oaic.gov.au. (2019). *Australian businesses and the EU General Data Protection Regulation*. [online] Available at: https://www.oaic.gov.au/resources/agencies-andorganisations/business-resources/privacy-business-resource-21-australianbusinesses-and-the-eu-general-data-protection-regulation.pdf [Accessed 5 Apr. 2019].

[42] Rouse, M. (2015). *What is two-factor authentication (2FA)? - Definition*. [online] SearchSecurity. Available at: https://searchsecurity.techtarget.com/definition/two-factor-authentication [Accessed 5 Nov. 2019].

[43] Stack Overflow. (2013). *[Microsoft][ODBC Driver Manager] Data source name not found and no default driver specified*. [online] Available at: https://stackoverflow.com/questions/17115632/microsoftodbc-driver-manager-data-source-name-not-found-and-no-default-drive [Accessed 5 Nov. 2019].

[44] Rpubs.com. (2016). *RPubs - Faceted Heatmap*. [online] Available at: https://rpubs.com/omicsdata/faceted_heatmap [Accessed 5 Nov. 2019].

[45] Comley (2019). *About boards and Kanban - Azure Boards*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/azure/devops/boards/boards/kanban-overview?view=azure-devops [Accessed 5 Nov. 2019].

[46] Danielson, S., Sebold, M. and Comley (2019). *Understand how work items are used to track user stories, issues, bugs, features, and epics - Azure Boards and TFS*. [online] Docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/azure/devops/boards/work-items/about-work-items?view=azure-devops&tabs=agile-process [Accessed 5 Nov. 2019].

[47] Webex. (n.d.). *Crystal-clear conference calls with Cisco Webex Meetings*. [online] Available at: https://www.webex.com/conference-call.html [Accessed 5 Nov. 2019].

[48] Linchpin SEO Team (2019). *A Beginners Guide To The Agile Method & Scrums*. [online] Linchpin SEO. Available at: https://linchpinseo.com/the-agile-method/ [Accessed 4 Jun. 2019].