1/19/2020

Personal Development Report

Imran Al-Touqi

Imran Al-Touqi
3475379

Contents

Introduction:

About me:

My name is Imran Al-Touqi and I'm 19 years old currently studying in Fontys UAS. I come from Oman

To study abroad as I feel it's a real privilege to be studying abroad to get a better education and to live a dependable life.

Why I choose Data Science:

I choose Data science out of curiosity, as I've been getting a lot of information from people and social media about the topic. I felt it's interesting to learn data science because I think it would help in my software course and get a broad idea on how machine learning works.

What I Expect from the course:

I read online about Data Science, and expect a lot of machine learning and solving big data problems. I think we would implement machine learning algorithms to solve various problems in companies that help them automate processes. I also expect to take raw data and turn it to a useful and under stable information so others can read it and make use of it. In addition to be able to work with a team together to make projects and exercises to learn more about data science.

Reporting

Goal: *"You report in a methodologically sound way about a data analysis (proposal, process documentation, reports on results and conclusions, etc.)."*

It can be seen from the goal that we must be able to produce effective proposals, documentation and reports. And to be able to find your results and conclusions based on those documents that will be in use.

During the past weeks I made a lot of reports and documentation related to data science. As well I read some chapters from various books to get a clear understanding about the various topics to prepare myself to make the upcoming exercises with my group or individually. Along with that I have made summary document of the various document I read to be able to explain what I understood and use it as a personal note for myself. As going through the whole document can be time consuming, so breaking down the main points in a document is very useful. All of these practices have been very useful to improve my reporting skills and prepare myself of what it yet to come.

I have also read a document regarding Data analysis, and made a report explaining the brief points of what Data analysis is, the types of data analysis and the importance of practicing EDA in an organization. This is proven by the comment I have received by my teacher as it states.

"*Well written summary on EDA; I think you are able to explain the value and key aspects of EDA. However, you didn't* including *anything on metrics/KPIs. Note that for a organization to be truly data driven, key decisions are supported by data, and results of these decisions are monitored via KPIs*." *-Schrijvers, Leon L.*

Although I provided a well written report, I went through some mistakes like not including metrics/KPIs as stated. But we are bound to go through some mistakes which I think is part of the learning process. I have taken into account of the feedback received and will try to improve myself further.

I also made a report on my sub-group challenge, which briefly explains on the company we worked with and how they make use of their data, more in detail is explained in the Data Driven Organization section in the PDR. I feel the report was well I received feedback from the company regarding to make some changes which we considered, learned from it and improved on the document.

I further worked on my personal challenge, where I had to produce multiple reports on my work Like EDA, Prediction Report and business proposal. All of these had to be well structured and defined so the reader can grasp the idea of what I am offering. I have received feedback on my personal challenge as it states.

"Hey Imran, I was assigned to review your report. But before, I want to say that this is only my personal opinion which can be wrong since I am also a student like you. Your report is really good, well written, clearly structured and explained. I love how you give a clear reason of why you use linear regression as your machine learning method. But I am sure that it would be a great idea if you give at a short introduction to your project in the beginning of the report, since usually the clients don't read the business proposal first and directly get their hands on the final

report, or maybe they forgot the business proposal. Moreover, shout out to the clear conclusion and recommendation for the company. It's also good that you put the jupyter notebook file separately, so it wouldn't be confusing for the clients if they do not want to see the code". - Jonathan Christopher Jayakusuma, Jonathan.

Overall, I have received well feedback from the reports I have offered on my personal challenge. There is some advice given like putting an introduction on the final report which I will take into consideration.

I have also worked with my group projects on many reports. Data analysis report, business proposal and many more to come. We have gotten feedback from the reports that we submitted for our project. It stated that we have offered a good report with clear explanation, but we focused too much on the business perspective and not how it's done. So, we learned that it is also good to explain the technical perspective of the project and the steps needed to take to reach a goal. This in result help make our goals that we propose more realistic.

We also provided a final report for our project, which includes explanation of various algorithms and other details which are both in technical and business aspects. We displayed our differences through various graphs explaining what happened before and after a particular event. This in result aided me to understand that it's always a good idea to provide graphs as a proof during explanation to further validate your points. The final report overall provides what we have done, our results and what we recommend to the client as a whole. This helped me look back into the project in one big picture and determine the purpose of data science after reaching the end of our project deliverable.

I have also worked on my Open Program report, where I explained what I have done over the course of the two weeks. I explained the steps, problems, approaches and why I chose this approach as a solution for my problem. As well as what I have learned after the open program work. This in result helped me look back at everything I have done so far into one picture and explaining them through a document aided me to revise my knowledge gained. I did explain what I have learned in the open program within depth detail in each section. I have also provided a graph explaining the difference before and after my solution approach displaying the changes between them.

I did receive feedback for my open programe as it states:

"Hi Imran, This report shows an excellent process of working on the different aspects of data science projects. You have shown that you are able to analyse initial results, investigate your options (including pros and cons) to improve the results, apply them and evaluate the results. Your reasoning in each of these steps is solid. Finally I can agree with your conclusion, that you have learned a lot in your open programme. The notebooks are very thorough. The steps that you took are easy to follow and make sense. Well done!" Schrijvers, Leon L.

Based on the work and feedback I received, I believe I'm in the Proficient level.

Machine Learning

Goal: *"You are able to select and apply modern supervised (machine) learning algorithms for classification and regression to a given data set."*

As we can see from the goal, we must be able to be able to learn and make use of machine learning to be able to choose and apply the machine algorithm for each data set.

I read documents explaining about what is machine learning and the various types of machine learning algorithms and how they work. With this I can choose what is the most appropriate algorithm for each data set whether it can be automated or have some supervision. However, I'm not able yet to apply the machine learning algorithm yet but know the theory behind it. I expect we will be assigned later to apply an algorithm to a given data set later in the course.

I have also worked with tree-based methods, which finds and predicts the outcome of a particular data given and figures out the percentage of each outcome from the trends of the dataset. This is considered to be a type of machine learning as they make us of the data given and train itself to find out the right answer. I have made a tree based on decisions on the current weather with my own made simple dataset and I have received feedback on that preparation.

"Hi Imran Looks good to me, well done. Do you have any idea why you started the tree with Rainfall? It seems to me that Cloudy would have been a better start (higher information gain), but perhaps you have another theory? Also, record 29 is not covered by the tree, perhaps on purpose?" - Michielsen, Bas B.S.H.T.

Based on the feedback the preparation went well, however there was a bit of context the teacher didn't understand which I would like to explain. I have chosen for this dataset that rainfall would be main variable that decides whether the person goes fishing, as I personally don't think it is a good idea to fishing when it's raining too much. Although the teacher stated that cloudy could have been better which I would take into account. Also, it seems that record 29 has not been included which was true as it was a mistake on my behalf.

I worked as well on support vector machines and cross-validation, which both are a form used in machine learning. I have researched online about support vector machines and cross-validation through online documents and videos. For the preparation, we were given a sample with faces from different sort of people in order classify which are wearing classes and which are not. To do this we need to find two features and plot them in a scatter graph. After that, would look into the graph and determine the decision boundary with the help of support vectors. Which in result would display a graph with all the data in their correct set.

The classification of people in the right set is achieved by the cross-validation algorithm. Which simply splits the data set into parts that would also be used to estimate the performance of the algorithm. Although there were some images in the data set that would make the algorithm have difficulty training it. So, we had to specify the images giving or reasonable explanation on why we think so. This helped me realize and understand that many of the datasets will have data that will sort of act like obstacles so preparing ourselves for such situation is essential.

I got a feedback on my preparation which states below:

“Hello Imran I like that you have an image and then in the description say that your image is not really suitable as there are no support vectors needed. It makes me understand that you know the basic theory of SVM even though your image is not so good (and you know it!). :)” - Michielsen, Bas B.S.H.T

Which is true what is stated above, because my graph was just a sample and too simple so no support vectors were needed to determine the decision boundary as this could easily determines without it by just looking as the graph.

I have also applied machine learning on my personal challenge, Where I had to work with several algorithms with the dataset I have. The algorithms would learn from the dataset by training and testing it with the appropriate data that they will need to learn from for the prediction.

When working on my personal challenge, I had to predict a certain column in the dataset which was the sales of the cars. After exploring the and understanding how to gain the value needed. I have decided to use KNN in my first approach. However, I realized that KNN was not the optimal algorithm to be used for this situation. After asking from my colleagues and teachers, I have found out the KNN was used to find categorical values. Which explains why it doesn’t work in my favor as I am trying to predict a numerical value based of correlations within the dataset. After trail and errors, I realized that linear regression was the right algorithm to use. Which did make sense after looking up online on the algorithm it fitted the problem, I was trying to solve which is predicting a column value based on correlations. I have also asked for feedback from colleagues and teachers for my situations and they mostly agreed that linear regression is the right way.

After Working with the personal challenge, I have learned that it a good idea to experiment with some different algorithms for your solution. Using the wrong algorithms helped me understand them more and how to use them in the right way. Perhaps in the future, I can use KNN to prediction a categorical value.

We have also worked on our project with machine learning, we our working on linear and polynomial regression to predict our data. We have chosen these algorithms because we think it’s best to predict our target values based on trends on the dataset. We then displayed and proved those trends by displaying each’s algorithm with its target values on a graph. This helped me understand the difference between the two algorithms as it visualized them clearly and see the overall difference between the two.

I also worked on Evaluation metrics preparation, where I followed an online course and solved multiple questions on the topic. I have learned a lot from this preparation as it gave me ideas I never thought about when working with data, the one that caught my attention was the true and false negatives and positives Which was really interesting.

In addition to preparations, I have also worked with neural networks. Where I can experiment the different situations online and give different outcome for each change. Neural networks are interesting as you can see how the data is being handled in real time as it moves from one node to node. And gives you a general view of how data is being used to give out the appropriate outcome.

However, I haven't received any feedback yet from the preparations which I hope will happen soon.

I have also worked on my open programme, where I had to make a choice on what algorithm needs to be used to get the value needed which was a numerical value. I ended up choosing linear regression algorithm for my datasets. However, in this dataset I had string values that I need to deal with. This in result aided me to understand regression models more and how they compare values with one another. In addition, I have also worked with polynomial regression as well after researching about it online and applied it on my dataset for prediction.

Based on the work, I believe I'm in the Proficient level.

Data Driven Organization

Goal: *"You can explain what a 'data driven organization' is and are able to argue on the applied data science maturity level of such organization including a reflection on the use of machine learning."*

The goal aims that we must be able to understand what is mean to be data-driven, and how companies achieve that by various methods with the use of machine learning. As well as how the take data-driveness further through the explanation of data-driven maturity and what level an organization is placed in

I have read documents regarding data-driveness and the background regarding the topic. I have learned what data-driven is, how and what it means for an organization to be considered data-driven. I have never thought that organizations have many data that they must take into account to work better. But this gave me an insight of how companies can improve as data is considered another key factor to success. Also, I know that companies that are considered data driven lye into different levels based on what they do with their analytics and procedures they take in various situations. It was an eye opening for me that many data-driven companies behave differently, as I thought that every company did the same with their data however this wasn't always the case.

I have also worked in a sub-group challenge with a topic based on data-driven organization. Were we had to contact a specific company and made a meeting with one their employee to interview them and give them a survey to fill out regarding their company. The experience was interesting, as we got a better perspective on how companies handle their data and how much they make use of it.

I have also received feedback from the sub-group challenge as it states: "*Well done! This report is crafted with care and is easy for me to read and understand. The only structural issue I had was matching the paragraphs under Report with the corresponding items in the Maturity Matrix. You used the matrix correctly and added the correct evidence but a few sub headers or pointers would have made it easier to read. Nevertheless, they say that when a teacher complains about small things the work must be rather good, otherwise the teacher would have complained about big things. :) Last year I visited AME myself because I had an internship student there, and they told me that they do not use ML as of yet. The student made a proof of concept on how to use ML for their product quality services. It is nice to read that AME is still trying to do something with this even now. I filled out the criteria matrix, if you still have questions, do not hesitate to contact me via email/Canvas message.*" - *Michielsen, Bas B.S.H.T*

We also have gotten from the feedback from the company where they agree with our thoughts and which level they are on the maturity matrix.

After getting the data we needed, we then had to make a report on what we think about the company and how they handle their data. As well we also classified them on one of the three levels of organizations which in was result lies in the Experienced level. Although some of the methods they take lies in the aspirational level, we chose this level based on the majority of the methods they undergo. We also received feedback from the company to review our document and approved of it including one of the teachers.

Based on the work and feedback I received, I believe I’m in the Proficient level.

Business Requirements:

Goal: *"You can translate business requirements and potential business benefits from applied data science into a structured business proposal, taking into account the compliance with laws on digital data and any ethical considerations that may arise from such a proposal."*

The goal states that we can understand from business requirements provided to us, look in depth on the situation and make a business proposal on that situation. With the help of knowledge from topics from Data Science achieving a well structured and written business proposal can be achieved.

We were given an assignment regarding the business proposal, where we had to criticize the business proposal, we had to work on based on the document we read. The document provided lots of valuable information on how to evaluate a business proposal and read it carefully. However, the preparation didn't go well because I was having trouble finding a source of a business proposal online.

After getting feedback from the business proposal preparation. I have redone the preparation with an article I found online with various business proposals of different companies available. Have applied the business proposal evaluation guide on it as much as possible. Although the proposal that was offered did not specify everything as expected, but I stated it in every different section in the document. It was good that I have redone the preparation to get knowledge of the stuff I missed on the previous time. I have not received any feedback from it yet, but hopefully will soon.

I have also done the business proposal for my individual challenge, after looking into my dataset I have provided a business proposal on what I personally think is appropriate for my situation. I have gotten feedback on it for my firsthand in. I have applied the feedback by changing the document's structure to make it more of a proposal. I have provided each section with the explanation needed like Business understanding, Data preparations etc... I have added more explanations in those sections to make it clearer of what I am proposing. The feedback also mentioned that I haven't specified what Is about to be predicted, which I have mentioned in the proposal and it's the sales of the cars. And to evaluate whether the problem is going to be solved or not is from the prediction section. Based on the prediction, we can then conclude whether the proposed idea is going to be successful or not based on number of different circumstances. I have also Mentioned what steps taken into cleaning the dataset which I applied from the feedback. The amount of data needed is specified in the final report which is something that is found out at the end of the project. The suggestion for the company to bring in cars based on wheelbase is based on trends and statistics for the company. I am well aware that customers won't come in the shop to buy a car with a specific wheelbase. This data is used to see trends for what customers prefer buying even if they are not aware of it themselves. Overall, the change specifies the more technical aspect of the business proposal rather than just focusing on the business side only.

I was able to work with my colleagues on our final report for our group project. We have given recommendations, results and conclusions on what we have found during the project. So, we provided details on what the company should do in terms of business and ethical aspects to succeed further on their business. This made me learn that we must also see the business side of the company and not just the technical parts. As is it essential to be transparent with the

client and explain why to solve a particular problem in a specific way with a logical and ethical explanation.

Based on the work, I believe I’m in the Proficient level.

Predictive Performance

Goal*: "You are able to research improvements on the predictive performance of applied data science models by selecting and applying different techniques for data preparation and normalization as well as model tuning and training."*

Based on the goal, we should be able to come up with improvements of our own by the help predictive performance. Which can be achieved after working on preparing the data and making a business proposal to better understand the whole context of the situation to provide a good prediction.

I have worked with predictive Performance on my personal challenge. After looking at my dataset, I had to decide on what predicting algorithm I am going to use and what are the data am I'm going to use for training at testing the algorithms. I have experimented on two different algorithms for my personal challenge that being KNN and Linear algorithm. After seeing the results of KNN, I decided to try normalizing my data to produce a better accuracy from the algorithm. This did improve slightly but it was still overall inaccurate for what we need to predict. I end up choosing Linear regression to predict the sales of cars using the trends of the dataset. The accuracy was still unfortunately low; however, I did give my reasons on why it was that way on my final report.

After working on the personal challenge, it opened my mind up on what it takes to make an algorithm work better. Like normalizing data helped gave a clear improvement on the algorithm which I kept in mind for future use in case to improve other algorithms as well. I think this is mainly to make the algorithm understand and familiarize the values it's working with after normalizing it. I have also learned that choosing the correct values to train and test an algorithm is the key to good prediction. Which I have chosen after finding correlations between the dataset values in my EDA. It is also a good idea to make a heatmap on correlations on the dataset, so get a clear understanding on what column is related to another to make it easier to choose values for an algorithm.

I have also received feedback from my personal challenge as it states for this section: *"Are you still predicting Sales? Which features correlate with sales? Did you make a heatmap? How did you decide which features to put into your algorithm, and which not?" Michielsen, Bas B.S.H.T*

Yes, I was still predicting sales. And the features correlated with sales was chosen after exploring dataset and getting a brief understanding of it. However, I have not made a heatmap unfortunately, but the knowledge gained from the heatmap is gained, as I already checked correlations for the target variable that I was planning to work with manually in the EDA. So It would have been easier to make a heatmap but I took the longer route essentially which will take into perspective for future use. So, after knowing the correlations of the dataset, and what data effects the others. I have decided to train the algorithm to get the target variable which is sales by testing and training the algorithm with the strongest correlations I could find with sales.

Another comment also states: *"At the end of page 10 I see you picked a small number of features and "magically" got a fair prediction. How did this come to be?" Michielsen, Bas B.S.H.T*

One that part I wasn't predicting the sales, I was predicting the price of the car. To prove essentially why my prediction was low for my sales. Because price stronger correlations with its columns than sales in a small dataset. So, I was giving an example by predicting price to further explain my situation when predicting sales.

I have also worked on final prediction report for my preparation. It specified how to report a prediction report and how to convince your readers on what you are trying to predict and your ideas. But I haven't received feedback on it yet.

I have worked with my colleagues on our group project. We have tried predicting the cleaned data but unfortunately gave out a poor prediction result. However, after applying models on the same algorithms the prediction result increased dramatically. This has taught me that taking extra procedures can be taken to improve an algorithms like using models and normalizing data Which is well worth the time taken for the improvement gained.

We were also given an extra dataset weeks later into the project and merged it with our current dataset. After merging it, it has negatively affected our prediction results from 70% to 50%. We have then decided to exclude this dataset and we all agreed that the new dataset contains incorrect data.

In my open program, I have reasonable amount of work that covers predictive performance. After applying linear regression algorithm on my datasets, I realized that the scores differ every time I run the Jupyter notebook. I then researched online I decided to use cross validation specifically K-folds cross validation. This in result would train and test different parts of the dataset rather than just one section. This gave me different accuracy scores based on the number of times I have provided and took the mean of those scores as my final accuracy. The final accuracy was then fixed with a value of 75% and helped me determine the actual accuracy of my algorithm.

In addition, I have also decided to use polynomial regression algorithm as well on my datasets to improve my accuracy scores. After using polynomial regression my accuracy score increased from 75% to 83% which is a big improvement. This made me realize that it's always a good idea to implement polynomial regression alongside linear regression to improve accuracy scores. I did discuss about my predictive performance in my final open programe report. I have received feedback from it as it states:

"Finally I can agree with your conclusion, that you have learned a lot in your open programme." *Schrijvers, Leon L.,*

As I did discuss about predictive performance in terms of polynomial regression in my conclusion and why I would recommend using it.

Based on the work, I believe I'm in the Proficient level.

Data Quality

Goal: *"You are able to clean data sets according to theories of data quality, in such a way that the process of cleaning and preparing those data is repeatable, transparent to others, and the results are suitable for data analysis.."*

Based on the goal mentioned, we should to improve the quality of the data by cleaning the data set we are working with to make the data consistent and understandable for ourselves and in order to explain it to others well visually which in return will help us for data analysis.

We have started with working with improving on data quality, we are currently cleaning our dataset on the group project and individual challenge as there a lot of data that need to be cleaned.

This helped me look into data into a different perspective, as I thought all data given would be correct and precise with the right data type in each column. But as it turns out, many data sets have wrong data that would interfere with our work so it's important to identify the data types, clean it and then work with it to give us a better look into the data set by giving our correct graphs.

I have worked with my personal challenge on data quality. I have cleaned my datasets during my Exploratory data analysis in various. This includes removing duplicate values, filling in missing data, renaming columns, changing column values and much more. More of data cleaning can be found in my exploratory data analysis document submitted in my personal challenge. After cleaning the data, I was then able to produce informational graphs that was not possible without cleaning.

This in result helped me practice the knowledge learned from preparations, I was then able to apply those knowledges on fixing my dataset. As many dirty data is hidden in the dataset, but I had an easier time cleaning it from the knowledge gained from the preparations that I have done. So, whenever I am handed with dataset, I can expect the dirty data from it and steps needed to take so clean the dataset for each specific problem.

One of the main problems that I have encountered during my personal challenge was the size of the dataset. After cleaning the whole dataset, I have seemed to overlook the actual size of the dataset. As it is very important for training and testing the dataset. However, I was not able to do anything about it as it contained only around 120 records or so. This has taught me that this is another precaution that must be taken before working with a dataset which is the size. So for the next time, I will choose a dataset with enough size to work with around thousands or more records.

I have also worked on handling missing data preparation which I have used this knowledge to apply to my personal challenge. The feedback from the preparation states:

"From the report I can see that you are able to explain the differences between the different types of missing data and that you can describe the different techniques of dealing with missing data. In your analysis on the missing values of the Embark column, you state that the missing data is of type MNAR and that the reason of it can be found by further investigation. However, in your report you don't mention how this can be done. What are your thoughts on this?"
Schrijvers, Leon L.

The reason that further investigation must be taken to find the answer is because unlike missing at random(MAR) and Missing completely at random (MCAR) there is a specific trend on why the data is missing. This could be because of the data of other columns that lead to the data being missing as an example. I haven't look further on the reason but I do think that's the reason as it's missing not at random.

I have also worked on Data Quality on my project. After Exploring our dataset, we found a lot of dirty data that you would normally expect when acquiring a typical dataset. This is result made me work as a team and learn what steps needed to take in order to solve a particular problem in the dataset. Whether its missing values, duplicate rows or unclear data format. All have different steps needed to take in terms of code and logical sense to solve which I learned with my colleagues. This is a must before working with the dataset in order to gain the most out the dataset as possible and to ensure a good prediction result.

We have also been given extra data for our group project, after proceeding to merge the data with the two datasets. It seems to have added extra dirty data so we had to go back and clean the data again.

I did work on my open program as well, I got the opportunity to do more data cleaning as I worked with two datasets during the open program. I have encountered new problems while working on the program.

One caught my attention is that with one of my datasets, it displayed that the datasets contained little to no null values. However, after looking through the dataset manually I can see a lot of missing values. This may be since jupyter had a problem identifying the null values. But I did solve this problem by replacing the null values with an "unknown" string to be able to identify them better. This is in result made me learn that sometimes jupyter is not always right when it comes to identifying null values. As It's a good idea to take an approach to look through the dataset manually if you can identify null values. This may be time consuming but it's always good to make sure.

I also had to work with string values in my dataset, I knew that those columns contained useful information that would be left behind if not dealt with. After researching online, I have made a decision to use One Hot Encoding to deal with the values. As this would create new columns for each unique value in a column with 1 indicating it is present and 0 that it's not. This in result can be used in regression algorithms and was able to make use of that data to get a good accuracy score on my models.

There were some setbacks while dealing with string values. At first, I thought of making each value in the column into a number in terms of enumeration. But then I realized this was not the right approach to take as the algorithms will compare the values incorrectly. Then I used One Hot Encoding which is the correct approach. However, this solution has some setbacks as well as when implementing this on my second dataset was difficult. Because the second dataset contained many unique string values and would create many unique columns which I know is not correct. I tried countering this by narrowing down the unique values or use the top 3 used values but that didn't seem to work as well.

Despite all the setbacks mentioned, I was still able to use this solution in my first dataset. I did learn what problems are expected and how to solve them when dealing with string values in a

dataset. I went through many problems along the way, but I do believe that it is part of the learning process.

I did provide an in depth about my data cleaning in my open program report as well. I did discuss about the overall summary about One Hot Encoding in my conclusion and received feedback as it states:

“Finally I can agree with your conclusion, that you have learned a lot in your open programme.” *Schrijvers, Leon L.,*

Based on the work and feedback, I believe I’m in the Proficient level.

Work Ethos

Goal: *"You are an effective co-worker in project groups, and are able to guide your own study progression by asking for, interpreting and applying feedback by teachers, tutors, coaches and fellow students."*

As mentioned in the goal, we must be able to work efficiently in group works. This includes, projects, exercises, challenges and many others. As well as being able to consistently ask for feedback for our work and applying that to our work.

Meeting my group members on the first day was easy as this course was very crowded in the first day. I am confident that I can work with my colleagues effectively as they are always on time, communicative and always make plans for future work. Sometimes I have issues during a preparation or any other work for that matter, and I always reach out to my colleagues and teachers to help guide me in the right track to be able to keep up with others. We also compare our work with each other and give advice/opinions on our work to be able to improve them.

I have also worked with one of my colleagues on the sub-group challenge. It was a good experience working together to get in contact with and talk to a fellow employee from a company. We discussed together about and shares our opinions between each other on what we think of the company and included it in our report. I really liked working with my sub-group colleague Jonathan as I think he is enthusiastic and added to mood to get the work done. The challenge went well and received good feedback from the teacher which in return motivates me working with in a team even further.

I am also working on the group project with my colleagues, we all discuss with one another during the exercises and group project on what approaches to take and when to work on something during to help us manage our time and work effectively. We also decided to split our group into two groups so increase productivity and get more work done. Rather than six working on one part of the project we don't think its good work management. After we are done with our work, we explain each other how we achieved our goal so each one of us could learn from it. So far, I'm enjoying my time with my team and looking forward to working with them in upcoming work.

After receiving several feedbacks from teachers, I have applied them to my work including preparation, personal challenge, project and PDR. Last PDR evaluation teachers mentioned that I should specify the names of the teachers/ student of the comment I have mentioned in the document which I have changed. Feedback from teachers was useful as it served as an eye opener of where I needed to improve. Especially during the PDR evaluation sessions, it helps me familiarize myself better with the teacher and what is expected next from me and what steps needed to take to further improve. They helped me understand some points I have not understood in some areas like business proposal, and clearly specified what it is by giving examples and tasks recommended to take to be better.

I have also received feedback from my colleagues during my peer reviews. The Results were good which I'm glad that I have been an overall good team member in my group. The peer review is difficult to break down into one document so I can show them to teacher during my next PDR Evaluation session. I have provided an average score of my Peer Review Below from 1 to 4.

Kyle Ritchi: 1.75

Daniel: 1.5

Mike: 1.6

Jonathan: 1.65

I was able to guide myself further by constantly asking for feedback and advice both from my colleagues and my teachers over the course. To be able to get other people's point of view on your work is essential to be able to improve further and see the bigger picture in case I am missing something. I am happy to work with enthusiastic members of the data science course as it helped me to be able to work in collaboration in a team much better. All of the feedback mentioned in this document where looked back on and tried to implement the suggestion from teachers. As well as answering the questions the teachers provided under the comment explaining my situation or why I have done that.

Based on the work and feedback, I believe I'm in the Proficient level.