Reflective Diary Template

Group Number: 149

Full Name	Sachit Anil Kumar (29392624)
Role in group:	Team Member
Role responsibilities:	● Coordinating meetings: Keeping communication open; arranging meetings. ● Dirty Data – Exploratory Analysis ● Dirty Data – Detecting and Fixing Errors ● Dirty Data – Testing ● Dirty Data – Documentation
Contribution to Group:	50% - Worked on the dirty data file to find errors and correct them
Learnings from Project	● Working with various different packages – numpy.linalg, network, sklearn, seaborn etc. ● Overview of data cleansing processes ● Linear Regression modelling, Implementing Dijkstra's algorithm, finding patterns from exploratory data analysis ● Working with dataframes extensively – using Pandas to avoid more expensive looping ● Using CSS stylesheets with Jupyter Notebook
Learning Techniques	● Applying new ideas or algorithms to small parts and checking validity or feasibility before testing on broader scope. E.g. Checking validity of ideas on one file and checking before applying to all ● Reading documentation of packages for syntax and functionality ● Running through tutorial notebooks and lecture demo notebooks ● Using online forums such as stackoverflow
What went well	● Starting the assignment early and thus being able to clear doubts early ● Using consultation times to our advantage ● Was generally happy with how things went this time around
What went wrong?	● A lot of the methods were repeated as this was an iterative process. Had to rework many things as I found better and faster methods at a later point – had to make these improvements to the existing functions as well.
Resolution to solve issues:	● Use more generic coding practices such that it is easy to adapt new solutions to all the iterative methods
Overall Conclusion: How would you do it, if asked to do it again? Would want to start early and have more generic methods and structure to the code.	

Full Name	Xinming Huang (26989166)
Role in group:	Team member
Role responsibilities:	● Missing Data - Preliminary analysis ● Missing Data - Detect and impute missing data ● Missing Data - Documentation ● Outliers Data - Preliminary analysis ● Outliers Data - Detect and Remove outlier data ● Outliers Data - Documentation
Contribution to Group:	50% - Worked on the missing data file to impute missing values; detect and delete outliers on the outliers data file.
Learnings from Project	● Practically go through the entire process of data cleansing ● Use linear regression model to help imputing missing data ● Use boxplot and residual to detect outliers ● Working with dataframes extensively – using Pandas to avoid more expensive looping
Learning Techniques	● Read the documentation of packages thoroughly and write some small demos before using them in the assignment code. ● Run through and understand tutorial materials. ● Attend consultations to get clear clarification of assignment requirements. ● Seek help from online forums like stackoverflow
What went well	We start the assignment early so we have sufficient time to work on it. Tutorial materials and packages documentations are very helpful resources. My tasks are less tricky compared to Sachit's and for some tasks we are actually sharing the same methodology and thus we save a lot of time.
What went wrong?	The task of detecting missing data and outliers is more straightforward than the dirty data task but it is still time-consuming . There are a lot of exploration and detection tasks and it is easy to get lost.
Resolution to solve issues:	Spend more time on the assignment. Always keep track with the high level task while working on small tasks.
Overall Conclusion: **How would you do it, if asked to do it again?** I will start the assignment earlier and have a better code structure.	