



SIG731

DATA WRANGLING

Learning Summary Report

YATHARTH DEOLY

yathartdeoly@gmail.com

Self-Assessment Details

The following checklists provide an overview of my self-assessment for this unit.

	Pass (D)	Credit (C)	Distinction (B)	High Distinction (A)
Self-Assessment	✓	✓	✓	✓

Self-Assessment Statement

Declaration

I declare that the portfolio of tasks within the unit is my individual work. I have not copied from any other student's work or from any other source except where due acknowledgment is made explicitly in the text, nor has any part of this submission been written for me by another person.

Signature: **Yatharth Deoly**

Reflections

The most important things I learnt:

This unit story revolves around data checks, pre-processing, removal of errors, analysis, visualization and understanding feature relations where these concepts are achieved using python and different libraries.

With the provided tasks, it made it more interesting and challenging to achieve and understand the core.

I feel I learnt these topics, concepts, and/or tools really well:

After this unit, am confident with below pointers:

- Usage of different python libraries, like pandas, numpy, seaborn, matplotlib.
- Different data visualization with histogram, line plot, pie chart and more.
- Power BI visualization technique and working.

I found the following topics particularly challenging:

Part which I found challenging is how to portray the task. As in this world, story-telling is the most important task, so while solving these tasks, I get to know how to implement the concepts/skills by explaining the task, usage, why it has been done and what we conclude.

I found the following topics particularly interesting:

As I was using python in my day-to-day life, but task-6 was the thing I love the most, as I have used pandas mainly in simple way, but the conversion that has been done from SQL query to python(pandas) way of writing was interesting and made my understanding better towards pandas and usage of it.

I still need to work on the following areas:

Like there are many libraries that can be used for visualization and mainly in tasks also we have come across with different plots and graphs, but when came across via Power BI and checked different plots, it made my more enthusiastic to learn it in python way with using different plotting libraries like plotty, bokeh, pygal and more.

Also, we have come across pre-processing of data, for which we need to dive deeper to get an understanding of tools and techniques that can be used for it.

The things that helped me most were:

Resources which have been used for this unit:

- Olympus platform
- Live sessions
- Surfing over google for more understanding

As there are different tasks that need to be implemented in this unit, which needs research and understanding, these platforms help me with it for easier understanding and implementation.

If I did this unit again, I would do the following things differently:

As I am a professional employee, where I haven't got much time to deliver the tasks on time, if I got the chance to do this unit again, I would have done it in a more systematic way with more descriptive task deliveries and implementation.