DAV 5400 Module 10 Assignment (30 Points)

Tidying and Reshaping Data

	Month	Category	Caltex	Gulf	Mobil
2	Open	Engine Oil	140 : 000	199 : 000	141 : 000
3		GearBox Oil	198 : 000	132 : 000	121 : 000
4	Jan	Engine Oil	170 : 103	194 : 132	109 : 127
5		GearBox Oil	132 : 106	125 : 105	191 : 100
6	Feb	Engine Oil	112 : 133	138 : 113	171 : 101
7		GearBox Oil	193 : 148	199 : 119	134 : 127
8	Mar	Engine Oil	184 : 100	141 : 141	114 : 108
9		GearBox Oil	138 : 121	172 : 133	193 : 115
10	Apr	Engine Oil	149 : 150	117 : 118	117 : 118
11		GearBox Oil	185 : 125	191 : 133	119 : 121
12	May	Engine Oil	170 : 139	104 : 119	200 : 117
13		GearBox Oil	168 : 117	138 : 102	121 : 146
14	Jun	Engine Oil	159 : 129	170 : 138	169 : 105
15		GearBox Oil	107 : 129	195 : 141	141 : 112

The chart above describes purchases and use of marine oil available at a major seaport. There are two types of oil available for use at the seaport: **engine oil** and **gearbox oil**. Each type of oil is provided by three distinct oil manufacturers/suppliers: **Caltex, Gulf**, and **Mobil**. The contents of the 'Caltex', 'Gulf', and 'Mobil' columns contain the number of gallons of oil purchased and consumed (e.g., **purchased : consumed**) for each month, with the '**Open**' indicator shown at the top of the chart telling us how much of each type of oil was on hand at the start of the chronological period (i.e., the 'purchased' amounts are the starting inventories for each type/brand of oil). The content of the chart has been re-created within the provided **M10_Data.csv** file. Get started as follows:

- Upload the provided **M10_Data.csv** file to your online AIM 5001 GitHub repository.
- Using the **pd.read_csv()** function, read the **M10_Data.csv** file from your GitHub repository into a Jupyter Notebook.

1.1 (**12 Points**): Use your knowledge of combining and reshaping data in Pandas to tidy and transform/reshape the data contained within the dataframe. To get started, think about how you would want the data to appear if it were converted to "long" format, e.g., how would you define a "single observation" for the data shown in the graphic?; How many key values are associated with each data value?; How many

columns should your long format structure contain based on the information provided in the graphic shown above?; What would the column headings for the long structure be?; etc. Use your answers to these questions to guide your reshaping/transformational work on the data. **Your reshaping/transformational steps must include converting the above table to a "tidy" long format.** Additional transformational steps (e.g., filling in missing data values, renaming columns, etc.) should be performed as needed to ensure that your data is, in fact, "tidy".

1.2 (10 Points) Using your reshaped/transformed data, perform analysis to answer the following questions:

- What was the amount of oil remaining **at the end of the chronological period** for each pair of oil types / brands?
- At the end of the chronological period, what was the most consumed brand of oil for each of the two separate types of oil?

1.3 (8 Points) Finally, given your "tidy" long format structure, describe what, if any, changes you would make to the visual presentation of the data if you were then asked to transform your "long" data back into a "wide" format: would you mimic the structure of the graphic shown above? If not, how might you transform your "long" data to "wide" format to make its "wide" presentation easier to understand and work with? Provide an example of your recommendation and explain your rationale for preferring your specific structure.

Save all of your work for this assignment within **a single Jupyter Notebook** and upload / submit it within the provided M10 Assignment Canvas submission portal. Be sure to save your Notebook using the following nomenclature: **first initial_last name_M10_assn**" (e.g., J_Smith_M10_assn).