COMP60711 - DATA ENGINEERING

Formative Coursework

Marking Scheme and Model Answer

MARKING SCHEME

Rubric Detail

Levels of Achievement

Criteria	Novice	Competent	Proficient
Column 'Flags' Update  **Weight 11.00%**	**0.00 %** The updated column does not appear on a screenshot OR the performed update is incorrect.	**50.00 %** The updated column appears on a screenshot but the level of correctness of its update is only partial.	**100.00 %** The updated column appears on a screenshot of the relevant file and the performed update is totally and completely correct.
Column 'Flag Text' Update **Weight 11.00%** 	**0.00 %** The updated column does not appear on a screenshot OR the performed update is incorrect.	**50.00 %** The updated column appears on a screenshot but the level of correctness of its update is only partial.	**100.00 %** The updated column appears on a screenshot of the relevant file and the performed update is totally and completely correct.

Levels of Achievement

Criteria	Novice	Competent	Proficient
Code/Recipe for Column Updates  **Weight 11.00%**	**0.00 %** Code/recipe associated with updates on columns 'Flags' and 'Flag Text' is incorrect or missing.	**50.00 %** Code/recipe associated with updates on columns 'Flags' and 'Flag Text' is only partially correct and/or partially complete.	**100.00 %** Code/recipe associated with columns 'Flags' and 'Flag Text' updates is totally and completely correct.
Code Explanation **Weight 11.00%** 	**0.00 %** Explanation associated with code/recipe is not sensible/correct and/or clear, or is missing.	**50.00 %** Explanation associated with code/recipe is only partially sensible/correct and/or partially clear/complete. It also represents a good attempt at providing comments about the tool/language features used to develop the task.	**100.00 %** Explanation associated with code/recipe is not only correct, sensible and clear but also brings interesting/non-obvious comments about the tool/language used in the development of the task.
Traffic Volume Calculation (TUE) **Weight 11.00%**	**0.00 %** Tuesday's total traffic volume is incorrect.	**0.00 %** NOT APPLICABLE.	**100.00 %** Tuesday's total traffic volume is correct.

Levels of Achievement

Criteria	Novice	Competent	Proficient
Traffic Volume Calculation (FRI) **Weight 11.00%**	**0.00 %** Friday's total traffic volume is incorrect.	**0.00 %** NOT APPLICABLE.	**100.00 %** Friday's total traffic volume is correct.
Code/Recipe for Traffic Volume Calculation **Weight 11.00%**	**0.00 %** The provided code/recipe associated with traffic volumes calculation is incorrect or is missing. 5	**50.00 %** The provided code/recipe associated with traffic volumes calculation is only partially correct and/or partially complete.	**100.00 %** The provided code/recipe associated with traffic volumes calculation is totally correct and complete.
Question as to whether any Data Preparation was required to perform the analysis. **Weight 11.00%**	**0.00 %** No (yes/no) answer is provided.	**0.00 %** NOT APPLICABLE.	**100.00 %** An (yes/no) answer is provided.

Levels of Achievement

Criteria	Novice	Competent	Proficient
Identification of Data Preparation steps and explanation **Weight 12.00%**	**0.00 %** According to the answer that was provided, the given explanation is incorrect and/or missing.	**50.00 %** According to the answer that was provided, the given explanation is only partially correct/sensible and/or complete/clear, as the student was able to correctly identify at least a subset of data preparation steps in the code and explain why the identified steps represent data preparation procedures reasonably well; or explain that none of the steps in the code represents data preparation procedures reasonably well/correctly; either way, linking the explanation to what has been learned in lectures.	**100.00 %** According to the answer that was provided, the given explanation is totally correct/sensible and complete/clear, as the student was able to correctly identify all data preparation steps in the code and perfectly explain why the identified steps represent data preparation procedures; or perfectly explain that none of the steps in the code represents data preparation procedures; either way, linking the explanation to what has been learned in the lectures.

Undetailed Model Answers

Note that we avoid providing you with code and explanations to avoid the re-use of these in later coursework.

Python

The tasks can be fully completed using a programming language and as such, Python is one of the best solutions for this coursework. In this set of solutions each task is solved independently of the others, however this is not required, as for example the dataset resulting from solving Task 1 can be saved and used for ulterior tasks.

The solutions below have been designed using Python 2.7.6, Pandas 0.13.1 and NumPy 1.8.0 so as to be run on the Kilburn computers. Using later versions for Python and the libraries would enable using new built-in functions which simplify the code.

Task 1

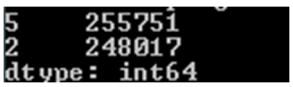
Update the 'Flag' and 'Flag Text' columns (for the entire file) by creating an index for each day of the week in the dataset. Rows where the date falls on a Tuesday should have the 'Flags' entry as 2 and 'Flag Text' as ‘Tuesday’, while Fridays should have 'Flags' 5 and ‘Flag Text’ ‘Friday’. Also, calculate the total traffic volume for each day of the week. Note that the total traffic volume should not be placed anywhere in the data file, but merely described in the pdf file to be submitted.

Required Output: Two values corresponding to the total traffic volume of each day of the week + a screenshot of the updated dataset

The following is a possible solution in Python with clear, correct/sensible and detailed explanations:

(…)

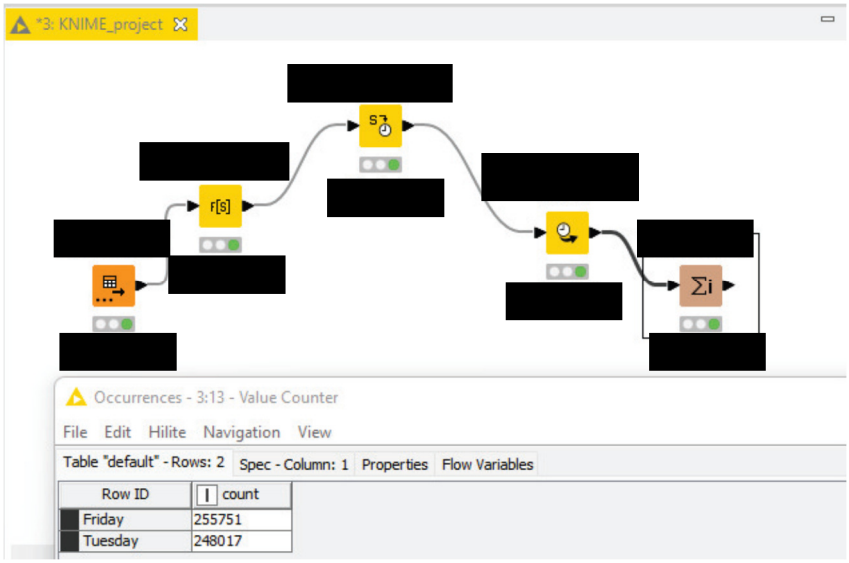
Resulting output: A Series object as observed in the command line interface screenshot below:



The resulting Data Frame would look as such:

	A	B	C	D	E	F	G	H	I	J	K
1		Date	Lane	Lane Nam	Direction	Direction	Speed (m	Headway	Gap (s)	Flags	Flag Text
2	0	02/02/2018 00:00:03	6	SB_NS	2	South	38.525			5	Friday
3	1	02/02/2018 00:00:22	5	SB_MID	2	South	32.31			5	Friday
4	2	02/02/2018 00:00:22	4	SB_OS	2	South	44.739			5	Friday
5	3	02/02/2018 00:00:36	6	SB_NS	2	South	33.554			5	Friday
6	4	02/02/2018 00:00:49	6	SB_NS	2	South	39.768	12.3	11.847	5	Friday
7	5	02/02/2018 00:00:52	2	NB_MID	1	North	64.623			5	Friday
8	6	02/02/2018 00:00:55	1	NB_NS	1	North	29.205	6.319		5	Friday
9	7	02/02/2018 00:00:58	2	NB_MID	1	North	37.283	6.2	6.089	5	Friday
10	8	02/02/2018 00:01:03	6	SB_NS	2	South	44.739	14.8	14.575	5	Friday
11	9	02/02/2018 00:01:04	2	NB_MID	1	North	41.01	5.155	5.242	5	Friday
12	10	02/02/2018 00:01:05	2	NB_MID	1	North	37.283	1.47	0.949	5	Friday

A solution in Knime would look like this (with clear, correct/sensible and detailed text-based explanations):



(….)