DATA2x01: Data Science, Big Data and Data Variety
Practical Assignment Rubric: Greater Sydney Analysis

Data Workflow (Code)

Data Import 0.5 marks	**0.5 - Good** All provided datasets loaded in.		**0.25 - Okay** Only some provided datasets loaded in, or all loaded in with minimal evidence of the process.	**0 - No Marks** No evidence of any provided datasets being loaded in.	
Database Schema 1.5 marks	**1.5 - Very Good** A sensible normalised schema has been implemented to store the data for all datasets used in the scoring function, with suitable data types, consistent naming conventions, primary keys in every table, and foreign keys where applicable.	**1 - Good** One or two tables are missing schema definitions but the rest are well-formed, OR schema definitions exist for all tables, but not all of the aforementioned requirements are met.	**0.5 - Incomplete** Only one or two tables feature schema definitions, OR only a few match the aforementioned requirements.	**0 - No Marks** No evidence that an intentional database schema has been implemented.	
Spatial Join 2 marks	**2 - Very Good** All required spatial datasets have been accurately matched to the ABS SA2 boundaries using PostGIS SQL queries.		**1 - Good** A spatial join was made without PostGIS SQL queries, OR the PostGIS spatial join implemented is not a precise match.	**0.5 - Incomplete** A meaningful attempt using PostGIS is evident, OR non-spatial workarounds were implemented.	**0 - No Marks** No evidence of a spatial approach to join required datasets.
Own Dataset 2 marks	**2 - Very Good** One additional dataset per group member has been sourced, transformed and ingested into the DBMS using a sensible schema. At least one dataset contains spatial data, and at least one involves a type not used so far (e.g. JSON, XML, web scraping).	**1.5 - Close** One additional dataset per group member has been sourced, transformed and ingested into the DBMS, but either none contain spatial data OR none feature a new type or approach as required.	**1 - Good** At least one additional dataset is missing, OR neither of the two conditions (one spatial dataset and one new type) are met.	**0.5 - Incomplete** At least one additional dataset is missing, and the ingestion approach is unclear or misguided.	**0 - No Marks** No additional datasets.

Score Calculation 3 marks	**3 - Very Good** A score has been calculated that is (1) well-implemented (2) using SQL queries, (3) incorporating all required datasets, (4) with correct definitions, (5) for all applicable regions in Greater Sydney, and (6) with integrated extensions for own datasets.	**2.5 - Close** Missing 1/6 requirements.	**2 - Good** Missing 2/6 requirements.	**1.5 - Good** Missing 3/6 requirements.	**1 - Okay** Missing 4/6 requirements.	**0.5 - Okay** Missing 5/6 requirements.	**0 - No Marks** Missing all requirements.

Indexing 1 mark	**1 - Good** More than one useful index created, with at least one spatial index.		**0.5 - Okay** At least one useful index created.	**0 - No Marks** No index beyond primary keys.

Data Analysis (Report)

Dataset Description 1 mark	1 - Good All data sources have been briefly described in a clear but concise manner.		0.5 - Okay A description of data sources exists, but is incomplete or not well-summarised.		0 - No Marks No evidence of a dataset description.	
Database Description 1 mark	**1 - Good** An accurate database schema diagram has been included in the report detailing the datasets involved, as well as primary and foreign keys.		**0.5 - Okay** Some DB schema detail has been provided, but is either not through a diagram, OR through a quite inaccurate diagram.		**0 - No Marks** No evidence of a database description.	

Results Analysis — 2.5 marks

2.5 - Very Good	2 - Close	1.5 - Good	1 - Okay	0.5 - Okay	0 - No Marks
Detailed analysis including (1) a well-explained rationale for the formula computation, (2) the implications of extending the base formula with additional datasets, (3) a summary of the overall distribution, (4) particular trends, regions or scores of interest, and (5) support by relevant tables, figures and graphs.	Detailed analysis, but missing or lacking in 1 of the 5 listed areas.	Lacking in 2 of the 5 listed areas.	Lacking in 3 of the 5 listed areas.	Lacking in 4 of the 5 listed areas.	Lacking in all required areas.

Correlation Analysis — 2 marks

2 - Very Good	1 - Good	0.5 - Incomplete	0 - No Marks
A statistical test for correlation between computed scores and the median income in available SA2 regions has been completed, summarised in the report, and interpreted well for the reader.	A statistical test for correlation has been conducted but the analysis is lacking, OR the summary is intuitive, but based on slightly flawed methodology.	A statistical test for correlation has been conducted but the report contains no meaningful analysis, OR both the summary and test methodology leave much room for improvement.	No evidence of any correlation tests.

Data Visualisations — 2 marks

2 - Very Good	1 - Good	0.5 - Incomplete	0 - No Marks
Great use of appropriate, meaningful data visualisations, including at least one map-overlay visualisation (optional even interactive).	Reasonably good use of data visualisations throughout, but somewhat limited in quantity, creativity or usefulness.	Very few visualisations included, OR those included are inaccurate or wrong.	No data visualisations.

Report Quality — 1.5 marks

1.5 - Very Good	1 - Good	0.5 - Incomplete	0 - No Marks
Great report which is within the length limit, well written, appropriately structured, grammatically correct, free of spelling errors, contains good font size/spacing, and uses figures well to visualise information.	Reasonably good report with at most three things from the previous list that could be improved.	Report contains more than three areas for improvement in the previous list.	No report submitted.

Advanced Tasks (DATA2901 ONLY)

Rank Based Scoring 2.5 marks	**2.5 - Very Good** A rank-based alternative scoring technique with appropriate standardisation techniques is planned, computed, and well-explained, including a comparison to the original scoring results and critical analysis of the differences.	**2 - Close** Minor oversights in the alternative score methodology, OR an analysis that covers the basics but lacks depth or insight.	**1.25 - Good** A reasonable attempt, with notable oversights in the methodology, OR a quite limited analysis.	**0.5 - Incomplete** A meaningful, but incomplete, attempt towards an alternative score.	**0 - No Marks** No attempt evidenced.
Machine Learning 2.5 marks	**2.5 - Very Good** A suitable question for which a machine learning solution is well-implemented. Key findings, implications, and the accuracy of the model are well-explained.	**2 - Close** Minor oversights in the machine learning solution, OR an analysis that covers the basics but lacks depth or insight.	**1.25 - Good** A reasonable attempt, with notable oversights in the methodology, OR a quite limited analysis.	**0.5 - Incomplete** A meaningful, but incomplete, attempt towards a machine learning solution.	**0 - No Marks** No attempt evidenced.