

2023 Sem2

Group Project Stage 2

Due: 11:59pm on Sunday (end of week 9)

Value: 10% of the unit

Note: these instructions are long and somewhat complicated, but the work you need to do is not actually very much. It should be easy to fit into the provided two weeks of your time, if you interact frequently and apply any feedback from the tutors. Don't wait till near the due date to start! If anything in the instructions is unclear or confusing, please ask about it on Edstem, using the category "Group Report", and sub-category "Stage 2".

GROUPS

Rules

This assignment is done in groups. You must be part of the **same lab session**. Under **VERY** exceptional circumstances, you may be allocated to an already formed group or you may be allocated to a group by the coordinator when dealing with group disputes and/or when dealing with group changes due to member discontinuing this unit. The group will get the same mark for the assignment.

THE PROJECT WORK

Task	Description	individual	Details
1	Identify topic and datasets	Group	The analysis done in this Stage must all be relevant to a single topic or question, which you are investigating because it matters to some stakeholders. You need to then have one or more datasets that you will analyse, to produce results that are relevant to this topic/question. You are allowed to use the same topic as in Stage 1, but you are equally free to change topic. The members of the group are allowed to all work with the same dataset, or some (or all) may choose to work

Task	Description	Group/ individual	Details
			<p>with different datasets. These datasets are allowed to be cleaned data from Stage 1, or integrated data from Stage 1, or you may choose to obtain new/extra data. There are no requirements for particular origin or volume in the datasets for this Stage. We will make available a dataset (on a topic of our choice) and any group can use that data instead, if they prefer. Note that all members of the group must be working on the same topic/question as each other, even if they use different datasets that deal with e.</p> <p>you produce from etely resolve the t, but each result lly able to provide le, if your topic is ge level of wealth ysis may calculate mmunities having g density, and a wealth relates to ng alone in each hat your question actual matter, but ips where insights some stakeholder not a good choice hich country has ?").</p>
2	Choose summaries and charts to produce	Group	<p>Each member needs to calculate one or more grouped-aggregate summaries from the dataset they are using, and they also must produce one or more charts from that dataset. The number of summaries and charts, and some constraints on what sort of attributes are used, depends on the level of score you are seeking. Details are in the marking scheme below. It is required that all the summaries be distinct from one another, and similarly each chart must be distinct. So you need to coordinate among the</p>

Task	Description	Group/ individual	Details
			members, in case two members want to do the same calculation, one at least will need to change!
3a	Use Python to produce a few tables from parts of the data	Individual	Each member then needs to work with their chosen dataset, to produce the material for their section in Part A of the report. This will involve writing Python code to calculate one or more summaries, and running that code to get the output, this can then be formatted as a Table in the report.
3b	Produce a few charts from parts of the data and evaluate their effectiveness	Individual	Producing charts can be done either by spreadsheet such as Excel or by using a programming language like Python. You need to think about how each chart will be used to evaluate its effectiveness, and the approach you take in tasks 7A and 8B, which is to evaluate how the data attributes get used in your chart.
3c	Write your section of the report		Write not more than 1000 words on what they have learned.
4	Write Part B of the report. Communicate your findings to interested readers.		In a group, you need to think about what your analysis has shown (or suggested), and back up these statements with some of the tables or charts taken from what members produced in their individual work. We realise that your work is likely to be limited, and indeed it may be that your analysis suggests that some attributes are not related in any simple way (for example, it may be that wealth and housing density seem fairly independent of one another, or at least, that your data doesn't show any connection!) – that's ok, just be honest in saying what you

Task	Description	Group/ individual	Details
			<p>expected, and what you found.</p> <p>Working together as a group, you need to produce a report. The structure of the report is described below in detail, as the report is the main basic for grading in this project. The report has sections for each member's separate work, as well as a brief combined introduction that explains the topic or issue, and a combined presentation of conclusions.</p>
5	Produce a PDF of the whole report	Group	<p>From the combined document, you need to produce a PDF. As well, there needs to be a folder, within which each member, the dataset the member made or spreadsheet (both summaries and data). Each member submits both PDF and data. Submission links on the whole group. Every member will get the marks for their submission.</p>



SUBMISSION FOR

1. There are **two deliverables** for this task.
2. All two deliverables should be submitted to the group.
3. The overall mark for this task will be shown in the gradebook.

Deliverable	Description
Report	<p>The report should have a front page, that gives the group name, and lists the members involved (giving their SID and unikey, not their name), and then the body of the report has two parts as follows:</p> <p>Part A should be targeted at a tutor or lecturer whose goal is to see what you achieved, so they can allocate a mark.</p> <ol style="list-style-type: none"> 1. There is an initial section which briefly states the topic of interest, and the stakeholders who care about this. This is not marked as such, it is just so the marker can understand the setting for the rest of the report. (max 1 page) 2. There should be one section for each member (the section should state the

	<p>SID/unikey of the group member who did the work reported in this section). In this section, there should be some subsections. (max 4 pages per member)</p> <ol style="list-style-type: none"> A brief description of the dataset being used by this member; showing at least the schema of the dataset. You do not need here to describe the provenance or give a detailed data dictionary. This is not marked as such, it is just so the marker can understand the tables and charts that follow. One or more subsections, each giving a grouped-aggregate summary. In any subsection, you should show the Python code that calculates a summary, followed by a table that presents the output of that summary. One or more subsections, each giving a chart. In any subsection, you should describe how you produced the chart, followed by a display of the chart. Following each chart, you should describe the effectiveness of the description of how you produced the chart; if you used a chart, you should describe in words the chart. <p>3. If you have a dataset with a large number of relationships, there will be an extra subsection for each relationship. This subsection should be no more than 1 page. (if produced)</p>
<p>Part B</p> <p>invest</p>	<p>or issue you are</p>
<p>4. The</p> <p>in</p> <p>the</p> <p>co</p> <p>sh</p> <p>fr</p> <p>ju</p>	<p>readers who are</p> <p>in it, you describe</p> <p>you present some</p> <p>your analysis. You</p> <p>data and chosen</p> <p>ted in Part A), to</p> <p>er group)</p>
	<p>Write whatever is needed to show the reader that you have earned the marks, and don't say more than that! In most cases, the code to produce a summary or chart will be fairly short (a few dozen lines at most), and the evaluation of a chart should not take more than a half- page.</p>
<p>Data and Code</p>	<p>This should be submitted through the Canvas system, as a single zip or tar.gz file. You should put them in a single folder, with <u>subfolders for each member</u>. The subfolder for a member should contain the dataset used, the Python code to calculate some summaries, and either Python code or a spreadsheet for producing the charts. You should also include</p> <ul style="list-style-type: none"> the provenance of the data,

	<ul style="list-style-type: none"> any licence or other restrictions on use of the data, description of all the changes you did between the original datasets and the final dataset; and the meaning of each attribute, what format or units are used, etc. <p>Compress the top folder (with all these subfolders and their contents), then submit the single compressed file.</p>
--	---

MARKING

The score (out of 10) is the sum of separate scores for each of the five components. A student's overall Stage 2 mark will be the average of the scores for the five components (20%). However, if all members agree to share their report, the score for the report will be 100% (20%). However, if all members agree to share their report, the score for the report will be 100% (20%). However, if all members agree to share their report, the score for the report will be 100% (20%).

- M1, 2 & 3 are assessed as part of Section 2 of the report.
- M4 & M5 are assessed as part of Section 2 of the report.

M1: GROUP AGGREGATE CALCULATION

This component is assessed either based on the section prepared by the specific member or the corresponding subsections of all the separate member sections in Part A of the report; the uploaded data and code may be checked by the marker as supporting evidence for claims made in the report.

Full marks	The data is presented in a clear and concise manner, and each group is clearly identified.
Distinction	Every member has written Python code that correctly calculates a grouped-aggregate summary of some data, and each group is clearly identified. All the code pieces are distinct from one another.
Pass	Every member has written Python code that correctly calculates a grouped-aggregate summary of some data. All the code pieces are distinct from one another.
Flawed	There is a correct calculation of some grouped-aggregate summary of some data.

M2: CHART PRODUCTION [2 POINTS]

This component is assessed either based on the section prepared by the specific member or the corresponding subsections of all the separate member sections in Part A of the report, as well as in the final 4-aspect chart section if that is present; the uploaded data and code may be checked by the marker as supporting evidence for claims made in the report.

Full marks	The Distinction criteria holds, and also there is at least one chart which illuminates connections that involve at least four aspects or attributes of the data that are relevant to the question, and where there is a reasonable expectation of a relationship where all four attributes interact together [not just that each pair are related, but that the way in which any two relate, is impacted by the values of the other two!]. This chart must be compelling in communicating the information to the reader (e.g. it draws the reader to easily gain a deep awareness of the patterns, especially how the relationship of some are impacted by the other attributes) and makes them keen to learn more.
Distinction	Each member produces at least two charts that accurately convey the relationship between aspects or attributes from their data that are relevant to the topic. For each member, at least one chart must show information about at least three aspects or attributes. For each member, at least one of the aspects shown must be distinct from one another. (possible starting points in the report such as distortion or missing data).
Pass	Each member produces at least one chart that accurately conveys the relationship between aspects or attributes from their data that are relevant to the topic. [The chart must show a trend or relationship between at least two attributes or attributes of one aspect and another. The chart must be distinct from one another and not be misleading or missing data.]
Flawed	Some members produce charts that do not accurately convey the relationship between aspects or attributes from their data that are relevant to the topic.

M3: CHART EVALUATION [20 marks]

This component is assessed on the corresponding section of the report; the uploaded data and code for each member or based on the corresponding section of the report; the uploaded data and code for each member or based on the corresponding section of the report; the uploaded data and code for each member or based on the corresponding section of the report.

Full marks	The Distinction criteria hold, and, for each chart, there are good reflections on how well (or not) the chart design would work if much more data is obtained.
Distinction	Every member has written an evaluation for each chart in their section, which correctly documents the encoding between data attributes and visual attributes, and documents other decisions (such as style of chart, scale etc.), and it sensibly justifies the decisions in view of the effectiveness of communication. All the charts in the report must be distinct from one another.
Pass	Every member has written an evaluation for each chart in their section, which correctly documents the encoding between data attributes and visual attributes. All the charts in the report must be distinct from one another.

Flawed	Some reasonable attempts to evaluate the effectiveness of some of the charts.
--------	---

M4: CONCLUSION – CONTENT [2 POINTS]

This component is assessed based on Part B of the report. Material in Part A, or the submitted data and code, may be checked by the marker as supporting evidence for claims made in the report.

Full marks	The Conclusion section has all the Distinction criteria, and discusses honestly and with insight, the limitations, and uncertainties about the results.
Distinction	The Conclusion section provides some accurate information which provides insight into important issues in the topic, supported by at least four relevant tables and at least four relevant charts; the tables and charts must include something produced by each member of the group in their sections of the report.
Pass	The Conclusion section provides some accurate information which provides insight into important issues in the topic, supported by at least four relevant tables and at least four relevant charts which were produced by each member of the group in their sections of the report.
Flawed	The Conclusion section provides some accurate information which provides insight into important issues in the topic, supported by at least one relevant table and at least one relevant chart.

CONCLUSION – COMMUNICATION

This component is assessed based on Part B of the report.

Full marks	The Conclusion section draws the reader in and provides a clear summary of the findings.
Distinction	The Conclusion section provides a clear summary of the findings, and includes a clear explanation of the limitations and uncertainties about the results. The summary includes at least four relevant tables and at least four relevant charts which were produced by each member of the group in their sections of the report.
Pass	The Conclusion section provides a clear summary of the findings, and includes a clear explanation of the limitations and uncertainties about the results. The summary includes at least one relevant table and at least one relevant chart.
Flawed	A reasonable attempt to provide a clear summary of the findings, and includes a clear explanation of the limitations and uncertainties about the results. The summary includes at least one relevant table and at least one relevant chart.

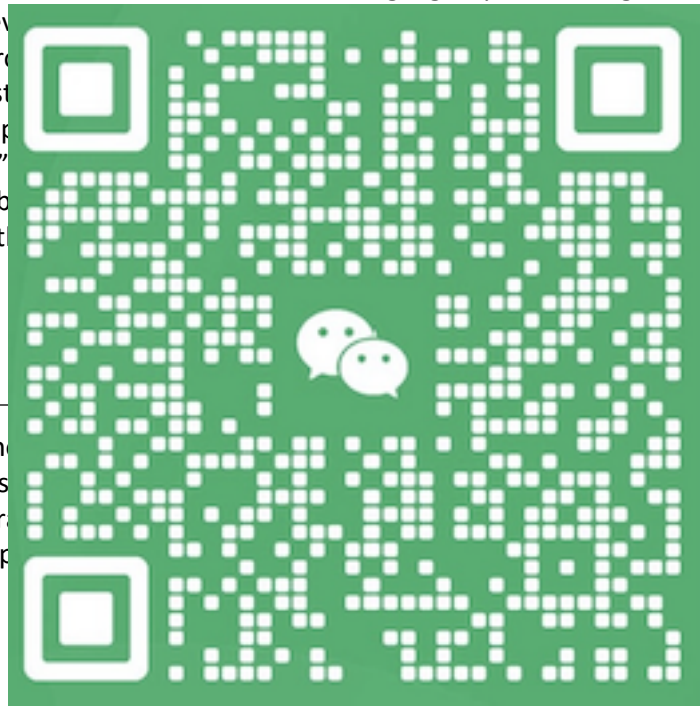
GROUP PROCESS

During the project, you need to manage the work among the group members. *We insist that every person do each activity and describe what they did and found in the appropriate section of the report and in the appropriate subfolder of the compressed folder that gets submitted.* We intend for the members to compare regularly and learn from one another (as well as from tutor feedback during lab sessions). Because any member's poor work will reduce everyone's score, make sure to quickly report any difficulty in working together to the unit coordinator as described above.

DISPUTE RESOLUTION

If, during the course of the assignment work, there is a dispute among group members that you can't resolve, or that will impact your group's capacity to complete the task well, you need to inform the unit coordinator, joshiah.poon@sydney.edu.au. Make sure that your email names the group, and is explicit about the difficulty; also make sure this email is copied to *all* the members of the group (including anyone you are complaining about). We need to know about problems in time to help fix them, so set early deadlines for group members, and deal with non-performance promptly (don't wait till a few days before the work is due, to complain that someone is not delivering on their tasks). If necessary, the coordinator will split a group, and leave anyone who didn't participate effectively, in a group by themselves (they will need to achieve all the outcomes on their own). **This option is only available up until Friday of week 8**, which is the last day with time to resolve the issue before the due date. **For any group issues that arise after this time, you will need to try to resolve the problem on your own**, and you will continue to be treated as a single group which all get the same mark for this Stage, based on whatever you can tell us about them).

If someone doesn't participate, or doesn't meet the standard, you should still put a "Note to marker" in the report. The report may be empty if they don't participate. In such cases, please put a "Note to marker" in the report. In such instances, that way, we can consider how best to proceed, if needed or sensible, for other members to do the work.



LATE WORK

As announced in the (and any other arrangements) suffers from the same problem. That is, we subtract the cost of the late work from the cost of the early work. No consideration or day after the due date for the work. No