



UNSW
SYDNEY

ASSESSMENT GUIDE

COMM5000

Data Literacy

**Sandbox PwC Distribution Project
Case Study Business Report**

Term 1, 2023

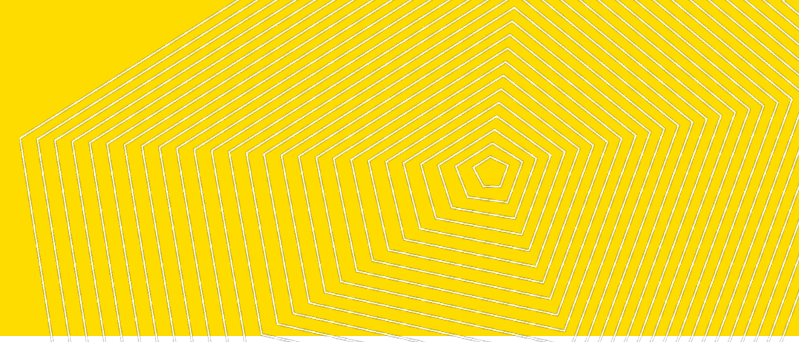


Table of Contents

Assessment Summary	2
Assessment Administrative Details	3
Turnitin	3
Late Submissions	3
Extensions.....	3
Special Consideration	3
CASE STUDY INFORMATION-- PricewaterhouseCoopers (PwC) Distribution Project Statement	4
COMM5000 Context	4
PwC schedule of engagement	4
Sandbox Project Showcase at PwC	4
Business Report	5
.....	5
Description of assessment task.....	5
DATA Considerations	5
Statistical Analysis	6
Executive Report	9

Assessment Summary

Assessment Task	Weighting	Due Date*	Course Learning Outcomes
Milestone 1: Case Study Preliminary Insight Development (due in Week 4 20%)	20%	Week 4 (Friday 5PM)	1, 2
Milestone 2: Case study project proposal	20%	Week 7 (Friday 5PM)	1, 2, 3, 4
Smarthinking feedback (formative task)	0%	Week 9 (Thursday 5PM)	
Case Study business report	60%	Week 11 (Monday 5PM)	2, 3, 4, 5

* Due dates are set at [Australian Eastern Standard/Daylight Time](#) (AEST/AEDT). If you are located in a different time-zone, you can use the [time and date converter](#).

Assessment Administrative Details

Turnitin

Turnitin is an originality checking and plagiarism prevention tool that enables checking of submitted written work for improper citation or misappropriated content. Each Turnitin assignment is checked against other students' work, the Internet and key resources selected by your Course Coordinator.

If you are instructed to submit your assessment via Turnitin, you will find the link to the Turnitin submission in your Moodle course site. You can submit your assessment well before the deadline and use the Similarity Report to improve your academic writing skills before submitting your final version.

You can find out more information on the [Turnitin information site for students](#).

Late Submissions

The parameters for late submissions are outlined in the [UNSW Assessment Implementation Procedure](#). For COMM5000, if you submit your assessments after the due date, you will incur penalties for late submission unless you have Special Consideration (see below). Late submission is 5% per day (including weekends), calculated from the marks allocated to that assessment (not your grade). Assessments will not be accepted more than 5 days late.

Extensions

You are expected to manage your time to meet assessment due dates. If you do require an extension to your assessment, please make a request as early as possible before the due date via the [special consideration portal](#) on [myUNSW](#) (My Student profile > Special Consideration). You can find more information on Special Consideration and the application process below. Lecturers and tutors do not have the ability to grant extensions.

Special Consideration

Special consideration is the process for assessing the impact of short-term events beyond your control (exceptional circumstances), on your performance in a specific assessment task.

What are circumstances beyond my control?

These are exceptional circumstances or situations that may:

- Prevent you from completing a course requirement,
- Keep you from attending an assessment,
- Stop you from submitting an assessment,
- Significantly affect your assessment performance.

Available here is a [list of circumstances](#) that may be beyond your control. This is only a list of examples, and your exact circumstances may not be listed.

You can find more detail and the application form on the [Special Consideration](#) site, or in the [UNSW Special Consideration Application and Assessment Information for Students](#).

CASE STUDY INFORMATION-- PricewaterhouseCoopers (PwC) Distribution Project Statement

Wholesale distribution companies typically purchase products from manufacturers/suppliers and then sell them to retail stores, making them available for consumers. Typically, wholesale distributors deal in large quantities of goods and are set up to have warehouses, distribution centres and logistic functions to manage and deliver inventory to retail stores. We are interested in better understanding the profitability of wholesale distribution companies.

Looking at the profitability of wholesale distribution companies globally over the past five years (PwC to provide excel containing raw data), is there a correlation (positive or negative) between their profitability and their local jurisdiction's GDP and other key economic metrics or events (e.g., the COVID-19 pandemic). If so, what may be the reasons for the correlation? Please provide both quantitative and qualitative analysis supporting any findings.

In addition, with a straightforward business model, wholesale distributors aren't involved in other key business functions such as manufacturing, R&D, retail trade etc. Are researchers able to review the publicly available information of key global distribution companies and corroborate their key functions, assets, and risks across various jurisdictions (e.g., comparing the activities performed, assets held, and risks borne by wholesale distributors based in the US vs China) to determine the other drivers of profitability that may exist? Please also provide any supporting analysis for these additional considerations.

The key jurisdictions we are interested in are the US, UK, China, Japan, South Korea, Australia, and New Zealand.

COMM5000 Context

This is a real business question that PwC is investigating. The role you are to play is one of a consultant contracted by PwC to assist with the analysis of data using the COMM5000 data analysis tools, which include descriptive and inferential statistics.

The work will be scaffolded into two milestones M1 (20%) & M2 (20%) and a final project report (60%). Every milestone will require you to use what you have learned to address specific aspects of the data. Generally, M1 consists of an exploratory data analysis, whereas M2 is concerned with identifying hypotheses and formulating key inferential questions. In the final project report, all the insights gathered from M1 and M2 are used to model the data to answer the project questions.

PwC schedule of engagement

It is very important that you attend these sessions where delegates from PwC will hold live synchronous sessions to provide information about the importance of analysing the factors affecting the profitability of distribution companies to their operations. During these sessions, you are free to ask questions and discuss any aspects of the project.

- 1) **Week 2, Tuesday 21st February**, from 12:00-13:00 (30-minute presentation followed by Q&A)
- 2) **Week 7, Tuesday 28th March**, from 12:00-12:30 (Q&A and mentoring)

Sandbox Project Showcase at PwC

A selection of the **five** best reports in Week 10. Those selected are invited to present their analysis of the PwC distribution project problem in person.

Note that this presentation is not part of COMM5000 assessments. This is an important career opportunity to be invited to showcase your work to an industry leader. However, anyone selected in the top 5 can accept or decline the invitation.

Business Report

Report details



Week 11, Monday 24th April 5 PM (Extension from Friday 21st)



60%



Case Study business report. Reports will be checked for plagiarism.



3000 Words

(PwC Team: You can prepare a separate report for PwC with more words. But keep it below within 4000 words)



Via Moodle course site

Description of assessment task

In this final report, you will use the information you have collected in your analysis in M1 and M2 to build a linear regression model. This model will capture the key objectives we have been pursuing since M1.

- (1) We want to investigate what company characteristics are determinant factors of its profitability.
- (2) We want to know if the country of jurisdiction has a premium or effect on a company's profitability.
- (3) We want to address the question of whether Covid19 has affected the profitability of wholesale companies.

In M2, you have performed a series of hypothesis testing to see whether there is evidence that profits differ between countries (at least for some industries) and whether there was a significant change from 2019 to post Covid19 years. Hypothesis testing is very informative but does not provide us with a means to control for confounding effects. For example, suppose you conclude that you reject a null of equal population means of operating revenues growth rates between the UK and Australia. In that case, you can't conclude that this difference is due solely to the country's effect. Other factors may affect revenues even if you control for the country of jurisdiction. Meaning that rejecting the null hypothesis here is driven by factors other than differences in the country.

To provide a ceteris paribus analysis, we need to use a regression model to identify and estimate the country effect (if any) while controlling for other factors. Other factors are defined by the data we are given. We will use the information on the company characteristics to control for the factors that may drive profitability and estimate the effect of the country of jurisdiction.

DATA Considerations

(i) Profitability Variables

For the dependent variable in your regression model, choose one of the variables you have already analysed in M2.

- i. Operating Revenue (\$'000)
- ii. EBITDA: earnings before interest, taxes, depreciation, and amortisation

<https://www.investopedia.com/terms/e/ebitda.asp>

Team PwC: Those who want to target PwC (whom I call TeamPwC hereinafter 😊) must consider analysing multiple profit variables. In the Q&A session, they listed the following profits variables as targets: *EBIT/Revenue*, *EBIT/Total Cost*, *EBIT/Operating Costs*, *Gross Profit/Revenue*, *Gross Profit/Operating Expenses*, *Gross Profit/COGS*, *EBIT/Total Assets*. You don't have to have all of them, but my understanding is if you consider 2-3 variables, that should hit their target.

(ii) Countries for analysis

You need to have **three countries** to analyse the country effect. You can use the countries allocated to you in M1 and add to that to make three. **You need to have Australia on your list.** If you want to change countries, you can choose any combination of three from the list Australia, the UK, NZ, and the US.

Note that changing the list of countries for the regression analysis will not necessarily create a significant workload compared to using the same countries you had at the start. We use the insights from M2 to motivate using dummy variables to capture the country effect. But none of the computations done in M2 will be used in the regression model.

(iii) Company Classification

For the report rubric, there is no requirement to use any classification. You may keep the pool of companies as is to estimate the regression model. But some of you may find it easier to focus on samples of a small group of companies belonging to a class/ industry and do the regression analysis.

Team PwC: PwC seems to like the idea of using some classification of the companies in each country. The wording they have used is as follows: '...part of the exercise where looking across a few regions; the students find industries with a meaningful number of companies in each jurisdiction (i.e., pharmaceutical goods distribution is common while tobacco distribution is very uncommon) SIC codes can be grouped for ease of analysis - students should go by the broad category, i.e., automobile vehicles industry and automobile parts industry can be grouped.' From the Q&A, a classification along the broad industry groups (like in the Korea Workfile), company size or any classification that you find meaningful is encouraged.

Statistical Analysis

Let Y_i be (your selected) profit variable for the wholesale company i . The aim is to build and estimate a linear regression model in the form:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

Task (A): Within-Country Analysis for Australia

- (1) We investigate what company characteristics determine factors for the selected profits variable Y_i in Australia. First, consider the year 2019 and estimate a regression model for Y .

How do you select what X variables go into the regression? One simple way is to start by putting all the variables in the Australian dataset in the regression. Then perform a process of eliminating those that are not statistically significant.

Model Australia2019: Consider all the 2019 company variables and the past years' variables as regressors, including the Y variable's past values.

- (i) Estimate the model and use the estimation results to remove all the variables that are not statistically significant.
- (ii) Repeat the regression with these significant variables and check if all are statistically significant.
- (iii) Repeat this process until you get a model where all regressors are significant.

- (2) **Model Australia2020:** Now, you can use your selected set of regressors from the 2019 model as your regressors for the 2020 model. Estimate this model and check if your regressors are all significant.
- (3) **Covid19 Effect on Australian Wholesale companies:**
To estimate the Covid19 effect in the case of Australia, we have to pool both 2019 and 2020 observations and estimate one full model.

In the Excel spreadsheet for Australia:

- (i) stack the column for 2019 on top of the 2020 columns for both the Y and X variables.
- (ii) create a dummy variable $Covid19_i$ that takes the value '0' for each 'i' observation from the year 2019 and takes the value '1' for each observation 'i' from the year 2020.
- (iii) Regress the stacked Y_i on the stacked X_i and the $Covid19_i$

Table 1: Example of regression table

	Operating Revenues Growth Rate (Y) Australia		
	Model (1)	Model (2)	Pooled Model (3)
X1	b_1 *** (std)	b_1 (std)	b_1 (std)
X2	b_2 (std)	b_2 (std)	b_2 (std)
...
Covid19	-	-	$b_{Covid19}$ (std)
Constant	b_0 (std)	b_0 (std)	b_0 (std)
Observations	n_{2019}	n_{2020}	$n_{2019} + n_{2020}$
SE of regression
R-squared

You can use a table similar to the one above for summarising the results of the three regression models in (1)-(3).

- (4) **Analysis of the within-country results for Australia:** using the results from the three estimated regression models, what are the determinant factors for profits?
- (i) Consider two of the determinant factors in 2019 and interpret the meaning of their estimated coefficients. What is the effect of a unit change of these factors on the profit variable Y?
 - (ii) Is there a change in the factors determining Y when you compare it to the 2020 results? Explain.
 - (iii) Lastly, what is the estimated effect of Covid19 as captured by the dummy variable? Interpret the coefficient on the dummy variable $Covid19_i$

Team PwC: You can analyse the within-country models by company classification. You may find that what affects one class of companies doesn't affect another, and you may find some groups of companies are more affected by Covid19. PwC didn't seem to require an analysis of all industries but a selection of interesting classes or industries. Those more ambitious can keep the data together and add a class dummy. If you have, say 3 classes or industries, you can use a dummy $class_i$ that takes 0 if the company 'i' belongs to class 1, 1 if it belongs to class 2, 3 if it belongs to class 3.

Task (B): Between-Country Analysis

Consider now the three countries' data. We will use the determinants (regressors) we have selected in Task(A) for Australia and use them as a starting set of regressors for the between-country models. For illustration in this document, assume that the other countries are NZ and the UK.

- (1) Consider the year 2019, estimate a regression model of Y on the selected regressors in (A) for both the UK and NZ
- (2) **Country effect:** To estimate the country effect, we need to stack the data from the three countries in one file. We use a dummy variable to capture the regional effect or effect of the country of jurisdiction. There are two ways to add a dummy variable with multiple categories.

- (i) Pooled Model (1): $Country_i$ that takes say, 0 if company 'i' is in Australia, takes the value 1 if the company is in NZ and takes the value 2 if company 'i' is in the UK. The coding of the countries is arbitrary.
- (ii) Pooled Model (2): create three dummies, one for each country: $Australia_i = 1$ if company 'i' is in Australia and 0 otherwise; $NZ_i = 1$ if company 'i' is in NZ, and $UK_i = 1$ if company 'i' is in the UK. **Caution!** Beware of the dummy variable trap! In the regression, you must include only two out of the three dummies.

Choose one option to capture the effect of the country of jurisdiction.

Estimate the FOUR regression models in (1) and (2) above. You can use a Regression Table to present the results.

Table 2: Between-Country Regressions

Operating Revenues Growth Rate (Y)

	Model (AU)	Model (NZ)	Model (UK)	Pooled Model (1)	Pooled Model (2)
X1	b_1^{***} (std)	b_1 (std)	b_1 (std)	b_1 (std)	b_1 (std)
X2	b_2 (std)	b_2 (std)	b_2 (std)	b_2 (std)	b_2 (std)
...
Country	-	-	-	$b_{country}$ (std)	-
NZ	-	-	-	-	b_{NZ}
UK	-	-	-	-	b_{UK}
Constant	b_0 (std)	b_0 (std)	b_0 (std)	b_0 (std)	b_0 (std)
Observations	n_{AU}	n_{NZ}	n_{UK}	$n_{AU} + n_{NZ} + n_{UK}$	$n_{AU} + n_{NZ} + n_{UK}$
SE of regression
R-squared

Note: Your Table will have either pooled model (1) or pooled Model (2)

(3) **Analysis:** Using the estimated regressions in (1) and (2) (presented in Table 2):

- (i) Discuss differences in the significance of the determinant factors you have identified in Model (1) (Task (A)) with their effects on profits in the other countries.
- (ii) Interpret the effect of the dummy in your chosen pooled model. What does this effect mean when comparing the effect of the regressors between the countries?
- (iii) What insights do you get from the pooled regression model you selected compared to the option you have not selected? Here, no estimation of the other option is required. But your insights of the difference in interpreting the slope coefficients on the country indicator dummy(s).

Team PwC: For PwC three countries are still enough for the analysis. However, they seem to want to see what happens over the year. You need to consider a few years instead of just one year. Looking at all 5 or 6 years, you can plot bar charts of the country effect coefficients and see whether it changes over the years. Please also continue the analysis using the classification of companies you have chosen. Some insights may be found by looking at how industries are affected by different factors depending on the country and over the years.

Ethics Considerations

Considering what you have learnt in Week 8, are there any ethical considerations around data collection, data analysis and use/implementation of the report's recommendations? Refer to the PwC code of Ethics:

<https://www.pwc.com/gx/en/ethics-business-conduct/pdf/pwc-code-of-conduct-april-2021-v2.pdf>

Can you recommend to PwC how to make ethically-informed decisions related to your report's statistical analysis and results?

Executive Report

The final report unifies the insights you have gathered in M1 and M2 and this final analysis. It showcases all the statistical techniques you have learnt in COMM5000. This report is a final take on this problem and should include the key results you wish the PwC team to get from the data. Your report will consist of the following core components. **You can restructure the section as you see fit for your report.**

- **Executive summary** This is the punch line of this case study. This should tell PwC what you have found in clear and precise language. This shouldn't be technical. It should refer to PwC description of the problem and their expectations.
- **Introduction:** The introduction should present the problem analysed in this case study. The background and the expected outcomes or target questions of the analysis. Your introduction should include some key conclusions from M2 (and M1) and how these conclusions provide some basis for developing the model for wholesale company profit characteristics. This is not a copy/paste of all your reports for the first two milestones. It should report the take-home message(s) from the detailed analysis you have completed—no need to re-report the tables and graphs from those milestones.
- **Section 1:** Data considerations, including what countries you are analysing, any major limitations in the data like a significant % of missing points for some variables/countries. This will inform the reader that you will restrict the analysis according to data availability.
- **Section 2:** Analysis of the determinants of profitability (growth rate of operating revenues) in Australia, including your analysis of the covid19 effect. Is there a Covid19 premium? Was the move from pre-Covid19 to Post Covid19 costly for the company/or industry in your analysis? Can you express it in the per cent change of growth in the profit variable?
- **Section 3:** Analysis of the countries' differences in profitability determinants. How important is the country's effect? Is it the same size across the three countries you have analysed? Is there a premium or penalty to a company located in a specific country compared to Australia?
In this section, you can put your news commentator hat on and add a comment on how you believe the country effect may have be driven by the

- **Section 4:** Robustness analysis and model limitations. Any statistical analysis is based on assumptions that ensure that the inference you perform and the estimation of the model parameters is also correct. Some points to discuss in your analysis are the key assumptions:
 - **Zero conditional means:** does your model satisfies the exogeneity condition? This will ensure that no confounding factors will bias the estimates of your key target parameters in the model. A brief explanation of what this assumption implies for the context of the model(s) you are estimating. Whether

you believe it holds and why? What can be done to ensure it is satisfied if you had more data and more time to develop the model? If it isn't satisfied, what are the implications on the model estimation and inference?

- **Multicollinearity:** With the selected set of regressors in your models, check whether the regressors satisfy the assumption of no perfect collinearity. Explain what it means in your context and the implications of the regressors failing the condition on the validity of the inference.
- Concluding remarks. In this section, you should summarise the data's key take-home messages. Your view on ethics considerations and how this impacts the collection of the data and the use of your results can be discussed in conclusion.
(Team PwC: recommendations to PwC on what they should look at when evaluating wholesale company profits reports. You should articulate clearly what the data has to say about the project aims as described in the original project description. Ensure you review the Q&A recording and that your analysis covers PwC main concerns. If there are aspects you are not investigating, explain why. It could be a lack of data, time, etc)

For each section, you should summarise the key points and state briefly the statistical evidence/argument that support those key points.

Your report should tell a story that uses all the key conclusions you have come to from the first two milestones and which will also inform your analysis for the last milestone.

Please defer all the Excel regression outputs and any other data analysis output to the appendix. Keep the body of the report following with only some storytelling charts that must be included in the body. If you choose the regression tables to summarise the results of the regressions, include these in the analysis of the results in the report.