

DATA ANALYTICS APPLICATIONS

ASSIGNMENT SEMESTER 2 2022



Preamble

The main purpose of this assignment is to help you to:

- consider the business environment in which a problem is to be solved;
- apply data analytics techniques to solve a business problem; and
- communicate the outcomes of your analysis to business stakeholders.

The specific skills that are being developed and assessed in the assignment are the ability to:

- assess the influence of domain knowledge on all stages of a data analytics project;
- evaluate how well data describes business activity;
- distinguish between different data sources and formats;
- apply each step in the natural language processing pipeline to solve a variety of business problems;
- develop and evaluate solutions to a range of classification problems using GLMs, tree-based models, ensembling and neural networks;
- evaluate the key drivers of successful model deployment; and
- communicate relevant points in language appropriate to the audience, in a logical and coherent manner.

You will be required to apply knowledge to specific situations in the time-constrained end of semester examination. This assignment provides an opportunity for you to think more deeply and spend more time preparing a detailed answer. This assignment will also help you self-reflect on your writing and presentation skills. Whilst there is ample time to write your answers for the assignment, you should ask yourself if you need to spend more time improving your writing skills to help you pass the examination.

The assignment requires you to build models and select appropriate parameters for those models. Consequently, there is no single right answer meaning you will be assessed on your reasoning and process. You therefore need to demonstrate *how* you chose parameters for your models and derived your answers. It is important that you describe what you did as the marker will want to understand if you are able to apply knowledge to the specific situation described in this assignment. We are also looking for you to demonstrate that you can deal with uncertainty in a reasonable way.



A key actuarial skill is to obtain a grasp of the qualitative nature of outputs from models and describe them in a non-technical manner. This assignment is designed to test how well you can explain your model(s) and outputs in a straight-forward way to a non-technical audience.

Marking Guide

This assignment represents 50% of the available marks for the Data Analytics Applications subject. Your assignment mark will be combined with your exam mark to determine your overall result for the subject.

It is anticipated that you will spend around 40 to 50 hours to complete the assignment. This is a guide as some students will spend more than this.

A detailed rubric is provided with the assignment question and will be used by the markers to assess your performance. The rubric has been posted on the Assignments page of Canvas to guide you as to what is required to achieve full marks for each part of the assignment. You should check that each of your answers covers the items specified in the rubric.

Submission

Deadline

The deadline for submission is **12:00 pm (AEST) on Friday, 23rd September 2022.**

Submit your assignment via the Assignments page in Canvas. If you experience technological issues when submitting your assignment, please send a copy of your assignment by email to education@actuaries.asn.au.

Penalties apply for late submissions (see below). You should anticipate potential delays by preparing and submitting your work in advance of the deadline.

Should circumstances arise that mean you cannot submit your assignment on time, you should contact the subject's Chief Examiner and education@actuaries.asn.au in advance of the deadline and apply for special consideration.



File format

The submitted documents must consist of one pdf file and one Jupyter notebook. Files in other formats will not be marked. The naming convention for both files is:

DAA 2022 S2 Assignment *candidate number*.¹

Your candidate number will be sent to you during the semester.

Coversheet

A coversheet for the assignment is provided on the Assignments page in Canvas. Complete and attach this coversheet as the front page of your pdf file.

Video summary

As part of this assignment, you are required to record a 5-minute video summary of your analysis findings. Advice about how to record an effective video summary is provided in Appendix 1. You should submit your video by following these steps:

- create a video recording using the naming convention 'DAA 2022 S2 Assignment *candidate number*';
- use your video recording to create an 'unlisted' YouTube video (see instructions in Appendix 2)²; and
- insert your YouTube video URL as a hyperlink in your assignment pdf file.

Jupyter notebook

The Jupyter notebook should use the assignment notebook template provided. The notebook must be capable of running successfully in Google Colab as markers will use this platform to view and access the notebooks. Within the notebook you should:

- explain each step taken in your analysis in a text cell above your code; and
- evaluate and comment on the output from each step in a text cell below the output.

¹ Please note that if you resubmit an assessment, Canvas automatically adds a suffix to the file name (such as '-1' for the first resubmission). You do not have to make any adjustment for this.

² Appendix 2 provides advice for students who do not have access to YouTube due to their location.



Please note that while there is no word limit for the comments that you include in your notebook, markers will look more favourably on students who provide **clear** and **succinct** commentary, compared to those who provide no commentary or those who provide too much commentary, including those who repeat large sections of the subject materials in their comments. This latter approach makes it very difficult for a marker to assess your understanding of the step being taken or the output being produced.

Word or time limit

Some questions in the assignment have a specific word or time limit. Markers will not read or watch any part of your answer that exceeds this limit. Keep your word count within any word limits that are specified. The word count includes any text within tables, text boxes or images consisting primarily of text. The word count does not include:

- contents table or index; or
- references to sources used.

Keep in mind one of the key principles taught in the Communication, Modelling and Professionalism subject: always write as clearly and succinctly as possible, while still including enough information that will be useful for your audience. With that in mind, consider whether each word, sentence, or paragraph you include in your assignment adds to or detracts from the message you are trying to convey. Importantly, know that 'more' is usually not 'best'.

Plagiarism

By submitting your assignment, you are implicitly stating that the work is your own.

Remember that an important aspect of being a professional actuary is to always act with integrity. Committing plagiarism by copying another person's work or not properly referencing other sources used in your assignment is a breach of the Integrity principle under the Actuaries Institute's Code of Conduct.

Any suspected plagiarism will be referred to the Institute's Executive General Manager, Education for review. Depending on findings, a complaint regarding the member may be made to the Institute's Conduct Committee. Subject marks may not be released until the matter is resolved.



Penalties

Deadline

Penalties will be applied to late submissions without prior approval.

If you submit an assessment after the due date (whether that is the original due date or any extended due date you have been granted), the following penalties apply:

- within 24 hours of due date and time: 10% x maximum mark available;
- 1 to 2 days late: 20% x maximum mark available;
- 2 to 3 days late: 30% x maximum mark available;
- 3 to 4 days late: 40% x maximum mark available;
- 4 to 5 days late: 50% x maximum mark available;
- >5 days late: 100% x maximum mark available (i.e. assessment score = 0).

Please note that 'days' above refers to calendar days, not working days.

Incorrectly formatted submissions

There is no direct penalty if an assessment is submitted in a format with an incorrect file name or an incorrect format (e.g. submitted as a word document when a pdf document was required). However, if a submission does not include a relevant identifier, then it may take time to identify the student and you might be asked to resubmit your work with an appropriate identifier. This may cause you to submit late and hence incur the late submission penalties outlined above. You should therefore follow all assessment instructions provided.



Assignment Context

You are an actuary who has a friend in the real estate industry. During a dinner last week, your friend joked about setting up a business together, combining your data analytics skills and their expertise in property sales. Since then, you have been starting to take this idea more seriously and have come up with the following business proposal to share with your friend.

Your friend is always telling you stories about properties that sell for prices well above their advertised price range. Your friend has explained that this often leaves potential buyers frustrated because they arrive at an auction expecting to be able to make a serious bid for the property, only to find that the final sale price is well above their budget.

Your proposal is to build a web app that potential property buyers can access, which lets them enter their:

- budget in fixed-range bands (e.g. \$0.5m-\$0.6m, \$0.6m-\$0.7m etc); and
- preferred postcode(s).

The app would then bring up a list of properties predicted to sell within their budget and postcode preferences. The key sales point for your app will be that it can give buyers a significantly more accurate indication of properties within their budget than they might get by searching based on properties' advertised prices.

You have sourced a dataset 'DAA 2022 S2 Assignment data.xlsx' (the property dataset) to help you build a proof-of-concept model to power your proposed app. This dataset contains details of properties that sold in New South Wales and Victoria in the three months to 30 June 2021. The dataset contains the features set out in Table 1.



Table 1 - Data dictionary for property dataset

Column name	Data type	Values	Description
Address	free text	various	The full address of the property
Suburb	category	various	The suburb of the property
Postcode	category	2000 to 3996	The postcode of the property
Beds	integer	≥ 0	The number of bedrooms in the property
Bathrooms	integer	≥ 0	The number of bathrooms in the property
Carspaces	integer	≥ 0	The number of car parking spaces on the property.
PropertyType	category	25 different categories	The type of property being sold.
SoldDate	date	1/4/21 to 30/6/21	The date that the property was sold.
SoldPrice	float or range	≥ 0	The price that the property was sold for. Note that for some properties, this is shown as a range that the sales price fell within, rather than the exact amount the property sold for.
FirstAdDate	date	3/4/20 onwards	The date that the property was first advertised.
AdPrice	float	≥ 0	The advertised price of the property.
AdHeading	free text	various	The heading of the advertisement for the property.
AdBody	free text	various	The body of the advertisement for the property.
AgencyName	category	various	The real estate agent who managed the sale of the property.



Assignment Questions (Total 100 marks)

Questions 1, 2, 6 and 7 must be answered in your pdf file. These do not need to be provided in a report format but should be written or presented using language suitable for communication with your friend.

Questions 3 to 5 must be answered in your Jupyter notebook using the assignment template provided.

Different questions may be reviewed by different markers, so your answer to each question should be self-contained. No marks will be awarded for answers to a question that are contained in your answers to other questions.

1. Explain to your friend, in 1,500 words or less, how well your property dataset captures the key drivers of property sales prices. To help you answer this question, you should apply at least two of the strategies outlined in Module 2 for gaining domain knowledge about the real estate industry. Your answer should include the source(s) of your information. Answer this question in your assignment pdf file. **(15 marks)**
2. Discuss, in 1,000 words or less and in language suitable for your friend, the deployment decisions that you will need to make when planning the roll-out of the app. Your discussion should be specific to this problem context. Answer this question in your assignment pdf file. **(10 marks)**



Answer Questions 3 to 5 in your assignment Jupyter notebook, using the notebook template provided.

3.

a. Examine the property dataset to give you a better understanding of the data that you have to work with. Your examination should include a summary of your key observations, written in language suitable for sharing with your friend. **(5 marks)**

b. Apply data cleaning and pre-processing techniques to the property dataset, to prepare it for use in predicting property sales prices. You should exclude the 'AdHeading' and 'AdBody' features in this step as these will be cleaned and pre-processed in Question 4. **(5 marks)**

4. Calculate vectorised features that represent the AdHeading and AdBody features in the property dataset. These vectorised features can be used in building your sales price prediction model. **(10 marks)**

5.

a. Calculate a response variable to indicate the sales price band that each property falls within. **(5 marks)**

b. Construct one additional feature that is not already available in the property dataset, which you believe will help predict the sales price band that a property will fall within. This additional feature should be sourced from an external dataset, not the dataset provided for the assignment. **(10 marks)**

c. Construct a classification model to predict the sales price band that a property will fall within. Your classifier should use any features in the property dataset, the additional constructed feature, or otherwise, that you believe will be beneficial in building a model that meets its purpose. **(15 marks)**

d. Evaluate how good your selected model's predictions are in meeting its purpose. **(10 marks)**



6. Prepare a five-minute video to send to your friend. Answer this question in your assignment pdf file as a YouTube hyperlink to your video recording. This video should use effective communication skills **(5 marks)** and outline the following **(5 marks)**:
- a. your business idea to build an app;
 - b. your success (or otherwise) to date in building a model to power the app;
 - c. any limitations of the analysis you have completed, including steps you could take to overcome these limitations; and
 - d. a recommendation on whether to proceed further with the business idea.

After watching your video, your friend has suggested that you could refine the app by allowing users to enter any range for their budget. For example, rather than being restricted to choosing a budget within a \$100,000 band (e.g. \$0.5m-\$0.6m, \$0.6m-\$0.7m, \$0.7m-\$0.8m etc), they might, for example, enter their budget as \$1.0m to \$1.35m.

7. Explain to your friend, in 500 words or less, the advantages and disadvantages of their proposed refinement to the app. Answer this question in your assignment pdf file.
- (5 marks)**

END OF ASSIGNMENT



APPENDIX 1 – VIDEO ADVICE

The following advice is provided to help make your video effective and easy for markers to find and understand your key points.

Your video should:

- feature a full or upper body shot of you to help you engage with your audience;
- have an appropriate volume and be free of background noise such that the marker can clearly hear what you are saying;
- not exceed the time limit; and
- not be sped up to fit within the required time limit - if your video is too long then you should consider removing some content.

To create an effective video, you should also remember to:

- plan the video to suit its intended audience and aim;
- apply structure to your presentation, with a clear start, middle and end;
- use transition statements to indicate movement between each of your key topics;
- make speaking notes to remind you of what to say on each key point;
- use visual aids to support your key messages;
- practise;
- engage your audience with your body language and voice; and
- be confident when delivering your message.

Please note that your video does not have to be 'perfect' to score full marks for it. The rubric provides more information about the exact criteria on which your video will be marked.



APPENDIX 2 – YOUTUBE INSTRUCTIONS

Students who do not have access to YouTube

Some students may not have access to YouTube due to their location. FOR THESE STUDENTS ONLY, please upload your video files directly to Canvas (preferably in an mp4 or mov format). We will then create an unlisted YouTube video for you.

In this case, your submitted video file should use the same naming convention as outlined in the submission section of this document.

Creating an unlisted YouTube video

An unlisted YouTube video is one that will not show up in YouTube search results and can only be seen by people you give the link to.³

To create an unlisted YouTube video, you need a Google account. If you don't already have a Google account, the following link provides instructions for setting one up:

<https://support.google.com/youtube/answer/161805?co=GENIE.Platform%3DDesktop&hl=en>

Once you have access to YouTube via a Google account, you are ready to create an unlisted YouTube video. The following YouTube video upload guide provides information about the basic steps required to upload a video to YouTube from either your computer or mobile device:

<https://support.google.com/youtube/answer/57407?co=GENIE.Platform%3DDesktop&hl=en>


When uploading your video, please choose the settings shown in the screen shots below.

³ Information about YouTube's privacy settings can be found at:
<https://support.google.com/youtube/answer/157177?co=GENIE.Platform%3DDesktop&hl=en>.



Data Analytics Applications

Assignment Semester 2 2022

EXAMPLESaved as draft×

1 Details

2 Video elements

3 Visibility


Details

Title (required)
EXAMPLE

Description ⓘ
Tell viewers about your video

Thumbnail

Select or upload a picture that shows what's in your video. A good thumbnail stands out and draws viewers' attention. [Learn more](#)

 ⓘ
Upload thumbnail

Playlists

Add your video to one or more playlists. Playlists can help viewers discover your content faster. [Learn more](#)

Playlists

Select

Audience

Is this video made for kids? (required)

6% uploaded 14 minutes left

Uploading video...

Video link
<https://youtu.be/smRMxAPgVEU>

Filename
EXAMPLE.mov


1. TITLE HERE

2. SCROLL DOWN FOR MORE OPTIONS

VIDEO URL

NEXT

© 2022 The Institute of Actuaries of Australia
Page 14 of 19

 Education
that Counts

AA



EXAMPLESaved as draft

1 Details2 Video elements3 Visibility

Audience

Is this video made for kids? (required)

Regardless of your location, you're legally required to comply with the Children's Online Privacy Protection Act (COPPA) and/or other laws. You're required to tell us whether your videos are made for kids. [What's content made for kids?](#)

☐ Yes, it's made for kids

☒ No, it's not made for kids

3. SELECT NOT MADE FOR KIDS

☒ Age restriction (advanced)

Uploading video...

Video link
<https://youtu.be/smRMxAPgVEU>

Filename
EXAMPLE.mov

Paid promotion

If another party paid to show a product or service in your video, let us know. Paid promotions need to follow our ad policies and any applicable laws. [Learn more](#)

☐ My video contains paid promotion like a product placement or endorsement

☐ Add a message to my video to inform viewers of paid promotion ?

4. SCROLL DOWN FOR MORE OPTIONS

Tags

Tags can be useful if content in your video is commonly misspelled. Otherwise, tags play a minimal role in helping viewers find your video. [Learn more](#)

Add tag

Enter a comma after each tag 0/500

Language, subtitles, and closed captions (CC)

Select your video's language and, if needed, a caption certification

Video languageCaption certification ?

10% uploaded 12 minutes left

NEXT



EXAMPLESaved as draft✕

1 Details

2 Video elements

3 Visibility

Select

This content has never aired...

UPLOAD SUBTITLES/CC ?

Recording date and location
Add when and where your video was recorded. Viewers can search for videos by location.

Recording date
None

Video location
None

License and distribution
Learn about [license types](#) and [distribution](#).

License
Standard YouTube License

5.UNCHECK "ALLOW EMBEDDING"

☒ Allow embedding ?

☒ Publish to subscriptions feed and notify subscribers **6.UNCHECK**

Category **"PUBLISH TO SUBSCRIPTION FEED AND NOTIFY SUBSCRIBERS"**
Add your video to a category so viewers can find it more easily

Education

**7.SELECT
DISABLE
COMMENTS**

Comments and ratings
Choose if and how you want to show comments

Comment visibility
Disable comments

Sort by
Top



8. SELECT NEXT

☒ Show how many viewers like and dislike this video

18% uploaded 10 minutes left

NEXT



EXAMPLESaved as draft


✓ Details

2 Video elements


3 Visibility

Video elements

Use cards and an end screen to show viewers related videos, websites, and calls to action. [Learn more](#)




You can complete this step after the standard definition (SD) version of your video has been processed. While you wait, you can close this screen or go to the next step.



Add an end screen
Promote related content at the end of your video


ADD



Add cards
Promote related content during your video

ADD

9. SELECT NEXT



21% uploaded 10 minutes left

BACK **NEXT**



EXAMPLESaved as draft

✓ Details

2 Video elements

3 **Visibility**

Visibility

Choose when to publish and who can see your video

☒ **Save or publish**
Make your video **public**, **unlisted**, or **private**

☐ **Public**
Everyone can see your video
☐ Set as instant Premiere

☒ **Unlisted**
Anyone with the video link can see your video

☐ **Private**
Only you and people you choose can see your video

☐ **Schedule**
Select a date to make your video **public**

Before you publish, check the following:

Do kids appear in this video?
Make sure you follow our policies to protect minors from harm, exploitation, bullying, and violations of labor law. [Learn more](#)

Looking for overall content guidance?
Our Community Guidelines can help you avoid trouble and ensure that YouTube remains a safe and vibrant community. [Learn more](#)

Uploading video...

EXAMPLE
Video link
<https://youtu.be/smRMxAPgVEU>

10. SELECT UNLISTED

11. TAKE NOTE OF VIDEO URL

12. SELECT SAVE

24% uploaded 10 minutes left

BACK **SAVE**



Video uploading

Your video is still uploading. Keep this browser tab open until uploading completes. Your video will be **unlisted** once uploading and processing finishes.

EXAMPLE

**13. KEEP BROWSER WINDOW OPEN
UNTIL VIDEO UPLOAD COMPLETE**

 29% uploaded 9 minutes left

CLOSE

Once your video has finished uploading, you should copy the video URL (see step 11 in the diagrams above) and paste this into your assignment file as instructed.

Optional step: using a 'brand channel' to hide your name

You will not be anonymous in your video as your face will be visible.⁴ However, it is preferable that your name does not appear in your video or in the YouTube channel that you upload your video to. The following link provides information about how to create a new channel in YouTube using a brand name rather than your personal name:

<https://support.google.com/youtube/answer/1646861?hl=en>.

Please use these instructions to create a new channel that does not include your personal name. The actual name you choose does not matter.⁵

⁴ There is a process in place to ensure that markers do not mark videos for students that they know.

⁵ You will not be penalised if you do not follow this optional step when uploading your video to YouTube.