

---

**Assessment 1 – Data Analytics Notebook**

---

<b>Name</b>	Assessment 1 – Data Analytics Notebook
<b>Due</b>	Sun 19 Apr 11:59pm
<b>Weight</b>	40% (indicative weighting)
<b>Submit</b>	Jupyter Notebook via Blackboard

---

## Rationale and Description

Foundational to asking good questions, and answering them with data analytics, is an understanding of potential data sources, the kinds of techniques that may be used to process and analyse those data, and an ability to present the final analytics in a way that is meaningful for the stakeholders.

This assessment will involve the creation of two Jupyter notebooks, demonstrating your understanding of the technical process required to ask good questions and obtain meaningful answers using data analytics.

You will use your knowledge from the lectures together with the techniques practiced in the tutorial sessions, and apply both to completing a basic skills notebook, and creating another notebook that answers 2 questions. You will not only perform the necessary analysis steps, but also provide an explanation of your decision process.

## Learning Outcomes

Successful completion of this task will demonstrate:

1. An understanding of how a variety of analysis techniques can be used to take data from different sources and analyse it in a way that is meaningful to a specific question.
2. How the question shapes the decision-making process in data analytics.
3. An ability to select, prepare, and use appropriate data, analysis techniques, and visualisations.
4. An understanding of a variety of data sources and the way that the data is structured.

## Essential Elements

You must submit 2 Jupyter notebooks which together will:

1. Demonstrate an understanding of:
  - a. Selecting and processing data appropriate for required analysis
  - b. Selecting and performing analysis techniques appropriate to a specific question
  - c. Addressing a specific question through visualisation of analysis
2. Document your decision making with explanations of your choices

You will use the code cells of the notebook to demonstrate your grasp of analysis techniques, and you will use the markdown cells to (a) craft a narrative linking the analysis to the question, and (b) document your decision making.

Further detail on the steps required to produce the notebooks is outlined in the 'detailed instructions' section below.

## Two Parts: Basic Techniques (Part A) and Addressing Questions (Part B)

The assessment will be completed in 2 parts:

- **PART A** – Complete the provided Jupyter notebook to demonstrate your basic skills. This part will be computer marked, and you will be able to submit multiple times for marking until you are successful with all tasks. Marking will occur every Friday from the 13<sup>th</sup> March through to 17<sup>th</sup> April, so you will need to submit your attempt to Blackboard on the Thursday night in order to have it marked the following morning. Feedback from your attempt will be provided to you via email. Success on all tasks will guarantee you a minimum of a 4 for this assessment. It is advisable that you aim for success early on Part A, so that you can focus most of your time on Part B of the assignment.
- **PART B** – In a single Jupyter notebook, address 2 of the 3 questions provided (below). The link provides data that you can start with, but depending on how you address the question, you may need to augment this data or use a different data set. You should clearly document your analysis and demonstrate the full data analytics cycle for each question.

## Questions

- **Question:** How does the frequency of mental health illness and attitudes towards mental health vary by geographic location, and what are the strongest predictors of mental health illness and specific attitudes towards mental health in the workplace?  
**Starting data:** <https://www.kaggle.com/osmi/mental-health-in-tech-survey>  
<https://www.kaggle.com/osmi/mental-health-in-tech-2016>
- **Question:** What do housing-market indicators say about the socio-economic conditions in different geographical locations, and how can they be used to provide forecasts of future economic conditions?  
**Starting data:** <https://fred.stlouisfed.org/>
- **Question:** What were the top Australian news topics over the last decade, and what can these say about the national conversation?  
**Starting Data:** <https://www.kaggle.com/therohk/million-headlines>

## Detailed Instructions – Part A

1. Sync GitHub to your cloud based Jupyter environment using the link in the welcome notebook.
2. Open the notebook: assignment1-partA

3. Address all of the tasks in the notebook, ensuring that you write code in the appropriate cell and run each cell after doing so. Do not change the order of the cells, and do not add or delete cells.
4. When you have a notebook that runs without errors, download it from your Jupyter environment, and upload it to the Blackboard submission link for Assignment1 – Part A. Ensure that you submit prior to 11:59 Thursday (starting the 12<sup>th</sup> March) for it to be marked the following Friday morning.
5. Feedback on your attempt will be returned to you via email. If your notebook has errors in it, you need to fix those tasks and resubmit for remarking. You need to ensure that you have a fully correct notebook by the 17<sup>th</sup> April (final marking). Once you have a fully correct notebook, you don't need to do anything further, you will receive at least a grade of 4 for this assignment.

## Detailed Instructions – Part B

From the 3 questions listed above, choose 2 and address the question by analysing one or more open data. You should ensure that your choice of questions enables you to best demonstrate your grasp of the required data processing, analysis and visualisation techniques.

In your analysis and visualisation for each question, you must at least draw on one or more techniques that have been covered in the lectures. You may, if you choose, supplement these techniques with other techniques that you have investigated yourself.

**It is critical that you document your thinking behind the selection of techniques as you will be assessed on your critical thinking and your ability to execute that thinking within the Jupyter notebook environment. You need to demonstrate that you understand the techniques that you are using, therefore it is better to use simpler techniques and show understanding than it is to employ complex techniques and show limited or no understanding.**

You will NOT be assessed on the quality of your code apart from its utility to perform the necessary tasks.

The notebook should tell a story (narrative) based on each selected question, that starts with the data selection, moves through the analysis, visualisation of the analytics, and concludes with connecting insights to the question under consideration. The story should make sense to potential stakeholders of the question/data.

For each step, you must document your decision making and explain why you did what you did. This description of thinking should align with the overall narrative.

1. **Question:** State the question, describe its significance, and identify the key stakeholders who have an interest in the question. A description of how you interpret your question should be provided with the question itself.
2. **Data:** Select data source/s appropriate to your selected questions, and write the necessary code to obtain the data and make it available for analysis in your notebook. We have provided starting data for each question, but you may supplement this with additional data

if you wish. The data may need to be processed prior to analysis, including cleaning in certain cases depending on the analysis techniques being employed.

3. **Analysis:** You will need to select analysis techniques that are appropriate to each question. You should include techniques learnt in this unit, but you may choose to supplement them with additional analysis techniques that you have learnt yourself.
4. **Visualisation:** You will need to create a visualisation that is appropriate to both the results of your analysis, and answering the question. You should have at least one visualisation for each question, but you may choose to include more if you decide that it is important to provide the necessary insight.
5. **Insight:** You need to answer the question in a way that is meaningful to the potential stakeholders. This may involve providing additional descriptive text that explains how the analytics and accompanying visualisation/s address the question.

You must submit a completed Jupyter notebook that has been run completely (all cells run in order), and which has no errors. You need to include data used by your notebook, so that the marker can re-run the notebook and verify the cell results. Both the notebook and the data should be zipped together and submitted via blackboard. Ensure that you name all files with your student number and name.

## Marking Criteria

This assessment is criteria referenced, meaning that your grade for the assessment will be given based on your ability to satisfy key criteria. Refer to the attached Criteria Sheet and ensure that you understand the detailed criteria.

It is important to realise that the assessment does not only require that you know or understand, but also that you demonstrate or provide evidence of your understanding. This means that you are making your knowledge and understanding clear to the person marking your assignment.

You will *not* receive marks or percentages for this assessment. You will receive an overall grade (e.g. pass - 4, high distinction - 7) based on the extent to which you meet the criteria. In general, the most important criteria (criteria 1-5) will be essential to the grade, and the least important (criteria 6-7) will affect the grade when important criteria results conflict or are ambiguous.

## Resources

The following resources may assist with the completion of this task:

- Refer to the workshop and lab notebooks for techniques
- Use Slack to exchange code and discuss detail of the task

## Questions

Questions related to the assessment should be directed initially to your tutor during the lab session or on the appropriate slack channel. Your tutor may address these for the benefit of the whole class.

The teaching team will not be available to answer questions outside business hours, nor immediately before the assessment is due.

## Criteria Sheet – Assessment 1 Data Analytics Notebook – IFN619 Data Analytics for Information Professionals

Criteria	7	6	5	4	3	2
[1] Evidence of meaningful connections between data analytics and significant questions.	Consistently makes meaningful connections between data analytics and significant questions with clear narratives that are interesting and engaging.	Mostly establishes meaningful connections between data analytics and significant questions but lacks some consistency in the clarity of the narratives.		Sufficiently connects the data analytics to questions to establish a meaningful relationship through the use of adequate narratives.	Some elements of the narratives make it difficult to see a meaningful connection between the data analytics and questions.	
[2] Demonstration of appropriate techniques for addressing significant questions with analytics.	All techniques are clearly appropriate and are consistently implemented in an exemplary way.		All techniques are appropriate but some implementations could be improved.		Techniques are either inappropriate and/or are used incorrectly.	
[3] Evidence of understanding analytics visualisation and its significance answering questions.	Provides exemplary evidence of a deep understanding of analytics visualisation and its significance.	Mostly provides evidence of an understanding of analytics visualisation and its significance.		Provides evidence of a basic understanding of analytics visualisation and its significance.	There is a lack of evidence of understanding analytics visualisation and/or its significance.	
[4] Evidence of an understanding of data selection and analysis techniques and their importance to the data analytics.	Provides evidence of a deep understanding of data selection and analysis techniques and their importance.	Mostly provides evidence of an understanding of data selection and analysis techniques and their significance.		Provides evidence of a basic understanding of data selection and analysis techniques and their significance.	There is a lack of evidence of understanding of data selection and/or analysis techniques and/or their significance.	There is little or no evidence of understanding of data selection and analysis techniques.
[5] Demonstration of appropriate data selection, processing and analysis techniques in order to yield a desired result.	Data selection is well suited to the task and all techniques are appropriate and implemented well.		Data selection, processing and analysis is mostly appropriate and suitable to the task. Most are implemented well.		Some processes or techniques are missing, incomplete and/or are insufficient to achieve a required result.	
[6] Demonstration of effective English expression and use of markdown.	Very good English expression and use of markdown.		Generally good English expression and use of markdown.		English expression and/or use of markdown is insufficient for the tasks.	
[7] <b>PART A only</b> Demonstrated ability to use foundational data analytics techniques with Python in a Jupyter notebook environment.	Successful completion of all tasks in Jupyter notebook for Part A – minimum grade of 4.				Jupyter notebook for Part A not completed.	