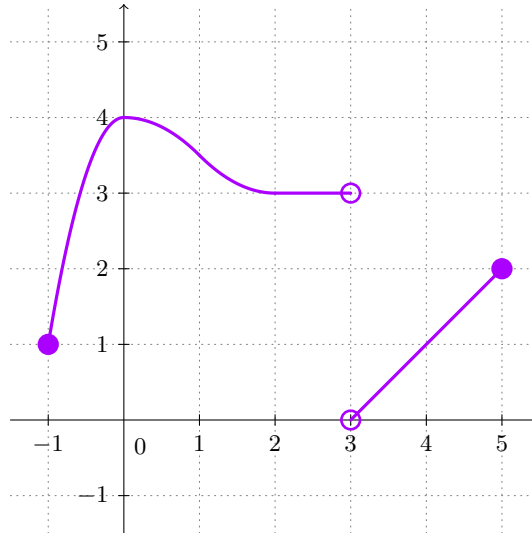


# Mathematical Foundations of Data Science

## Assignment 1

Trimester 1, 2023

1. Consider the function  $g$  whose graph is below



- (a) What is the domain of  $g$ ?
  - (b) What is the range of  $g$ ?
  - (c) Is  $g$  a one-to-one function? Explain your answer.
2. In total, how much money do the people in Adelaide spend buying cups of coffee every year?
- Note that you are not expected to get an exact answer! This is an exercise in *estimation* - there is no one correct number that will get you full marks. Rather, you must give a reasonable argument with plausible estimates following the principles presented in the course materials. You need to justify your assumptions and make it clear how you arrived at your answer.
3. Let  $J = \{x \in \mathbb{N} \mid -7 \leq 3x - 4 < 9\}$ ,  $K = (0, 3]$ , and  $L = \{0, 1/2, 3\}$ . Determine the following, giving reasoning:
- (a)  $J \cap K$
  - (b)  $K \setminus L$
4. You should complete this question using a Jupyter Notebook. All of the code you will need to complete this question can be taken directly or generalised from the week 1 computer exercise, or will be given to you in the question.

Download the file `movie_data.csv` from MyUni (this file comes from the [imdb-5000-movie-dataset](#) on Kaggle). Then use `Python` to do the following:

- (a) Using `pandas`, read the data into a dataframe and print out its `tail()`.
- (b) Create a histogram of `duration` for all movies.  
*Hint: Remember to add axis labels.*
- (c) Calculate the mean `duration` of all movies. Print out a statement showing the mean duration.  
*Hint: In the week 1 computer exercise, we saw how to print out a line containing text and numbers.*
- (d) Among movies longer than 180 minutes, find the movie with the highest `gross`.  
*Hint: Try creating a new dataframe that only contains movies with `duration` greater than 180.*
- (e) Find the Western movie with the highest `gross` among movies longer than 180 minutes.  
*Hint: Many movies have multiple genres. We want to include all movies with Western in the genre, not just movies where Western is the only genre. The string method `.str.contains()` might be useful here.*

Present your answers as a full Jupyter Notebook. Your notebook must include code to find the results, and text answering the questions based on the output of your code. Download this notebook and convert to a PDF and submit with your assignment.

**Please note: You must include the code you used to find results. Each answer submitted without code will receive a mark of 0.**

*Hint for submitting: You can "Download As PDF" in Jupyter, but that may not work on your computer. If it doesn't, you can download as HTML and convert that to a PDF. Make sure you join it to your assignment to make a single PDF when submitting! You might want to try googling things like "convert html to pdf" and "combine multiple pdfs". There is also a video in the Python Module on MyUni demonstrating how to save a Jupyter Notebook as a PDF.*