COMP3590
Programming for Artificial Intelligence

Assessment 4

Due: 23:55 12/12/2023

Please direct all queries to d.santry@kent.ac.uk

Please do your work in a single Jupyter notebook. Each question should be in an individual cell. Once you have finished the assessment please upload it to the Moodle submission area, "Assessment 4". If the format is anything other than a Jupyter Notebook then your work will not be marked. It is up to you to ensure that your work runs. Verify this on your Kent Jupyter account BEFORE you submit it. Clearly label the exercises.

Brevity and conciseness will be rewarded. Most questions can be answered 1-3 lines of code.

Please use the following dataset:

http://raptor.kent.ac.uk/~ds756/Data/kc_house_data_small.csv.

This is a dataset containing a sample of real-estate sales for King County, WA (it includes Seattle). Each row is a sale in the county.

Note: in the US, a postcode is called a zip code.

Task 1 (10 marks)

Create a feature in your dataframe, called "unit_price", that contains price per square foot for each sale. Use the "sqft_living" feature.

Task 2 (20 marks)

Create a dataframe that consists of two columns. The columns are "zipcode" and the average price per "sqft_living" in the corresponding zip code. Each zip

code will have its own row, and the second column will be the mean price of all the sales in said zip code.

Task 3 (20 marks)

Create a dataframe with two columns: bedrooms and mean price. Each unique count of bedrooms will have its own row, and the second column will be the mean price for properties with that many bedrooms. Use a bar plot to display your dataframe. Give each bar its own colour.

Task 4 (20 marks)

Graph “long” versus “lat”. Colour the waterfront properties red, and all others as blue. Title the plot and label the axis appropriately.

Task 5 (30 marks)

Create a time-series from the dataframe (set the index to the date). Plot the daily mean price. On the same plot, draw (overlay) a line from the smallest daily average to the highest daily average. The line should appear on top of the time-series.

Submission

The tasks should be completed in a single Jupyter Notebook. Each exercise should be clearly labelled. The notebook should be submitted on Moodle by