COMP41680/COMP47670 Class Test - Spring 2024

Guidelines*:*

- Complete all three tasks. All tasks carry equal marks.
- All tasks should be completed in a single Jupyter Notebook (IPYNB file). A template Jupyter notebook file has been provided on Brightspace.
- Use Markdown cells to briefly explain your work.
- Limit of 90 minutes allowed for the test. Once time has elapsed, you must submit your Jupyter notebook. Please monitor the clock to ensure you are not late.
- No collaboration is allowed. All submissions will be subject to plagiarism checking. Any evidence of plagiarism can result in a 0 grade.
- This is an open book test (i.e., you can use module lecture/lab material and internet resources).
- ChatGPT or similar generative AI / code assistance / chatbot tools cannot be used during the test. Development environments such as VSCode or Colab also cannot be used.

Summary:

This test involves analysing store product sales data from different retail stores. The data is stored in CSV format, with the following fields:

- *sale_date:* date on which the product was sold
- *prod_cat:* high-level category type of the product
- *store_loc:* store location where the product was sold
- *amount:* the sale value in euros of the product

Task 1: Data Loading and Initial Characterisation

a) Use Python to download the retail sales dataset from the link below and load it into a Pandas DataFrame:

http://mlg.ucd.ie/modules/python/test/retail.csv

b) Create a plot to show the distribution of all product sale *amount* values.

c) What is the number of product sales for

1. Each product category?
2. Each retail store location?

Task 2: Analysis of Feature Associations

a) Create a plot showing how the *number of product sales* in the full dataset relates to the product *category*.

b) Create a plot showing how the total product sales *amount* in the full dataset relates to the store *location* in which the sale took place.

c) What is the relationship between store *location* and product *category*, across all product sales in the dataset?

Task 3: Time Series Analysis

a) For each product *category*, create a monthly time series showing the total sale amount per month for products in that category. Plot these series together on a single figure.

b) From the series created in Task 3(a), calculate the 3 month rolling average for each category and plot these on a single figure.