

# Project Autumn 2023

COMP1013 Analytics Programming

Due Friday 2 June 2023 (Week 13)

## 1 Project Description

In this assignment there are 4 parts. For each part you should:

- Explain your rationales behind choices made answering the tasks.
- Write the appropriate R code.
- Include comments within the code to explain the algorithm.
- Test the code to ensure its correctness.
- Format and structure the code to maximise its readability.

A report must be submitted containing a cover page, the solutions to each of the four parts, and your code, as a **PDF**, to the vUWS submission site. The cover page must contain your name, student number, subject code and name, and the declaration below.

Submissions in other formats or without cover pages will have marks deducted.

Submission is due by Friday of week 13. Late submissions will receive a 10% reduction in marks for each day late.

## 2 Marking Criteria

This assignment is worth 40% of the subject assessment tasks. There four problems to investigate and 10 marks available for each of the four problems. The marking criteria for each question is given in Table 1.

Criteria	Q1	Q2	Q3	Q4
Rationales of algorithm choices (1 mark)				
Code Correctness (4 marks)				
Comments explaining code (2 marks)				
Code Testing (1 mark)				
Code Style and Readability (2 marks)				
Total (10 marks)				

Table 1: Marking criteria for each part of this project.

When writing the solutions to each of the four parts, make sure to consult the marking criteria and check that you have covered them. The project will be marked using this criteria.

For each task, there are a maximum of 2 bonus marks available for answers above and beyond the subject content. For example, extra analysis.

## 3 Declaration

Before submitting the assignment, include the following declaration in a clearly visible and readable place on the cover page of your project report.

\*\*\*

By including this statement, we the authors of this work, verify that:

---

© Copyright: Western Sydney University, 2023. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without the prior written permission from the Dean, School of Computer, Data and Mathematical Sciences. Copyright for acknowledged materials reproduced herein is retained by the copyright holder. All readings in this publication are copied under licence in accordance with Part VB of the Copyright Act 1968.

- We hold a copy of this assignment that we can produce if the original is lost or damaged.
- We hereby certify that no part of this assignment/product has been copied from any other student's work or from any other source except where due acknowledgement is made in the assignment.
- No part of this assignment/product has been written/produced for us by another person except where such collaboration has been authorised by the subject lecturer/tutor concerned.
- We are aware that this work may be reproduced and submitted to plagiarism detection software programs for the purpose of detecting possible plagiarism (**which may retain a copy on its database for future plagiarism checking**).
- We hereby certify that we have read and understand what the School of Computing, Engineering and Mathematics defines as minor and substantial breaches of misconduct as outlined in the learning guide for this unit.

\*\*\*

Note: An examiner or lecturer/tutor has the right not to mark this project report if the above declaration has not been added to the cover of the report.

## 4 Project Tasks

Your company, Dawson Steward Analytics, a consultancy firm specialised in AI analytics is tasked by a retail chain in the northern hemisphere to analyse their sales data. The data are contained in three different sets:

**sales:** daily sales data over a seven day period, containing the following columns:

- store\_id – the unique identifier of a store
- product\_id – the unique identifier of a product
- date – sales date (YYYY-MM-DD)
- sales – sales quantity
- revenue – daily total sales revenue
- price – product sales price
- promo\_type\_1 – type of promotion applied on product
- promo\_bin\_1 – binned promotion rate for promotion type
- promo\_discount\_2 – discount rate for applied promotion
- promo\_discount\_type\_2 – type of discount applied

**product\_hierarchy:** data containing the hierarchy and sizes of products:

- product\_id – the unique identifier of a product
- product\_length – length of product
- product\_depth – depth of product
- product\_width – width of product
- hierarchy1\_id – barcode hierarchy of product. The most general hierarchy. For example: fords and beverages.
- hierarchy2\_id – This is the second level hierarchy. For example: beverages
- hierarchy3\_id – Third level hierarchy, e.g. Cola
- hierarchy4\_id – Forth level hierarchy, e.g. Pepsi cola without sugar
- hierarchy5\_id – Fifth level hierarchy, e.g. pepsi cola without sugar 300ML

**store\_cities:** data containing the city, type and size information of the stores

- store\_id – the unique identifier of a store
- storetype\_id – type of store
- store\_size
- city\_id

Your tasks are:

1. Write the code to compute the total revenue of each store at the end of each day. Is there a noted difference between the days? Write also the code to calculate the total revenue over the seven day period. Plot the latter on a graph.
2. What's the most popular product type (hierarchy 1) sold in all stores over a week? How much revenue did the stores receive for that product during the week? How does that compare with the second most popular product? Provide a table that shows the product type ranked from most to least popular. For each product type provide: how many subtypes (hierarchy 2) are there, how many products are in this product type, what's the sales quantity, and the revenue generated.
3. Compare the sales volumes between the two most common store types in the data set. How do they compare in terms of total revenue? Is there a relationship between a store's size and its revenue?
4. Several different types of promotions were applied to the products during the period with various level of promotion rates. For each promotion type, display the different levels of promotion used during the period. Analyse the effectiveness of the promotion on the sales of the products.

Write a PDF report containing your code and all required analysis and results. The report is being marked using the marking criteria, so make sure that each piece of analysis covers all of the criteria.