Homework

This document outlines all homework problems for the Software Packages for Data Analysis course.

The document is split into three separate problem sets. Each problem set is work 3.33%.

You need to submit the solutions for ALL problem sets (one file per question) by the **end of the semester** using the link below.

<https://forms.office.com/e/W9f3ZsAN4m>

To successfully submit your homework, you need to send the **link to a GitHub public repository** containing all the solutions.

# Problem set 1

1. Analyse retail sales data to identify trends, seasonal patterns, and product performance.

Dataset: <https://archive.ics.uci.edu/dataset/352/online+retail>

1. Explore patient data to identify trends in diseases, treatment effectiveness, or healthcare resource utilization.

Dataset: <https://www.hcup-us.ahrq.gov/>

1. Analyze historical stock prices to identify market trends, volatilities, and potential investment opportunities.

Dataset: Yahoo Finance or Quandl

# Problem set 2

1. Use sentiment analysis on social media posts to gauge public opinion on various topics or brands.

Dataset: <https://www.kaggle.com/crowdflower/twitter-airline-sentiment>

1. Analyse crime data to identify patterns, hot spots, and effective prevention strategies.

Dataset: <https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/home>

# Problem set 3

1. Analyse customer behaviour to segment customers for targeted marketing.

Dataset: <https://archive.ics.uci.edu/ml/datasets/Online+Retail+II>

1. Build a machine learning model to predict the likelihood of a patient having heart disease based on various medical attributes.

Dataset: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

1. [EXTRA POINTS] Build a recommendation system that suggests movies to users based on their viewing history and preferences.

Dataset: <https://www.kaggle.com/datasets/dev0914sharma/dataset>