# COMP646 Special Topic in Data Science with R
## *Mock Exam*

David Gabauer

May 6, 2025

You are applying for a Data Analyst position at a renowned fashion organization operating in over 20 countries. As part of the application process, you are required to demonstrate your programming and analytical skills by analyzing one of their datasets and answering the following questions. Each question is worth a maximum of 5 marks. You are required to submit the R project folder as a **ZIP file**, along with a **5- to 10-minute presentation** explaining your code and results.

1. Create an R project. Save the dataset and your R code within this project. Load the dataset: *comp*646_*sales.xlsx*.
2. Compute the revenue stream for each item (row) using $unit\_price * quantity$. Then aggregate the revenues by country and sort them in descending order by gross revenue.
3. Create a bar chart using `ggplot2` that displays sales by country in descending order. Which country generates the highest revenue for the company?
4. What does the following function do, and what is its expected output?

```
function(data, selected_country='New Zealand') {
    data$days = as.character(data$days)
    data |>
        filter(country == selected_country) |>
        select(country, year, month, days, price) |>
        group_by(days) |>
        summarize(daily_revenue = sum(price)) |>
        mutate(days = parse_number(days)) |>
        arrange(days)
}
```

5. Use the function and create a `ggplot` chart for output with respect to the United Kingdom.
6. Calculate the total revenue generated by each customer. Who is our top customer - the one generating the highest revenue?
7. Calculate the revenue for each invoice. Which invoice generated the highest revenue?
8. It seems unusual that the highest invoice revenue exceeds the revenue from our top customer. Why does this seem odd? Filter the data for:
   - the invoices of the customer who generated the highest overall revenue, and
   - the invoice with the highest individual revenue.

   Then explain why this outcome actually makes sense - that an individual invoice can have higher revenue than the total revenue from a single customer.