Final Practice I

DSA2101

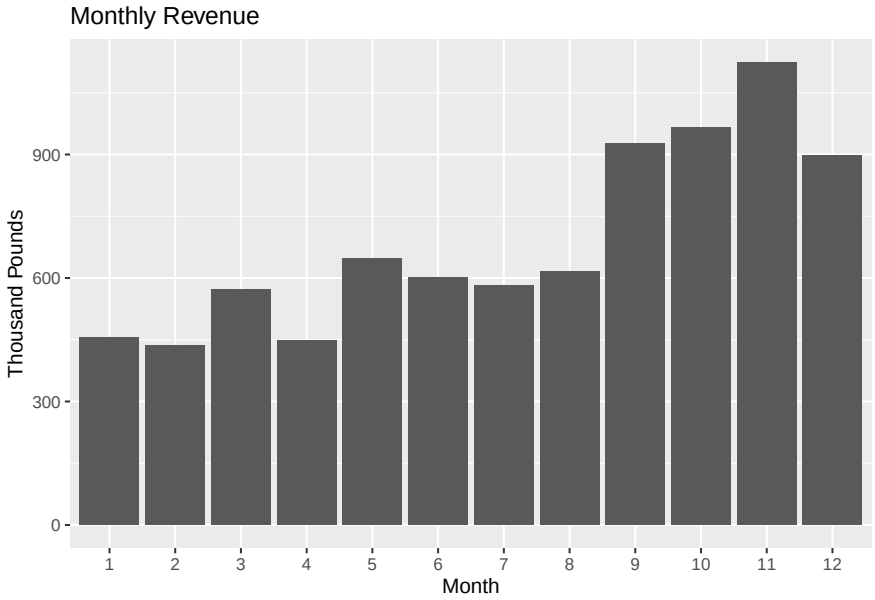
All questions in this exam relate to purchases made from an online retailer, situated in the United Kingdom, over a period of approximately one year. Please ensure that you have downloaded the following files:

- retail_clean.rds
- cty_coords.rds
- cust_features.rds
- column_description.txt

Question 1: Monthly Revenue (10 Marks)

The monthly revenue is computed as `sum(UnitPrice * Quantity)`. Comment on the quality of the following plot, which is meant to depict the monthly revenue that the company receives. Is it accurate? *Hint: Check the range of invoice dates.*

If you feel that there are mistakes in this plot, produce a corrected one. On the other hand, if you feel this plot cannot be improved, create a variation of it using at least one different geom or aesthetic.



Question 2: Christmas Shopping (15 marks)

The company would like to figure out when people begin Christmas shopping, so that they can start preparing for it. In the box below, write the R code that will perform the following tasks:

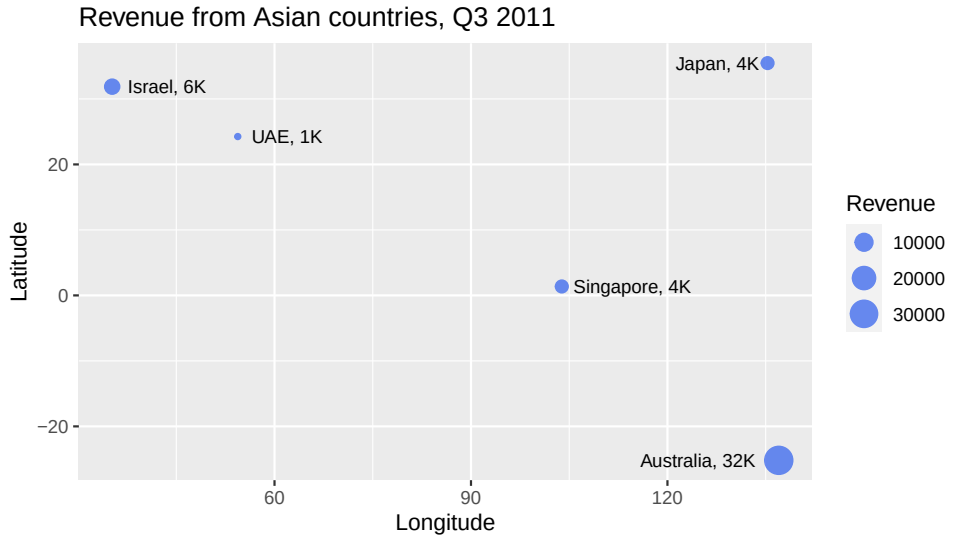
In this question, a line item refers to a record in the dataset. For instance, the first line item refers to StockCode 85123A bought in Invoice 536365.

1. Create a new column from `InvoiceDate`, that is of class `Date`.
2. Use a regular expression to count the number of *line items* each day, that are related to Christmas. Look for the following keywords in the Description of each item: CHRISTMAS, XMAS, SANTA or REINDEER.
3. Create a visualisation of this daily count, and use it to conclude when Christmas shopping begins.

Growth outside United Kingdom (10 marks)

The company in question is located within the United Kingdom (UK). Naturally, most of the revenue comes from within the UK. In this question, we shall investigate the change in revenue to countries in Asia.

Use the country table `cty_coords.rds` to create the following chart, where point size indicates the revenue from Asian countries in the *third* quarter of 2011.



Customer features (15 marks)

The table `cust_features.rds` contains the following features for each customer id in the retail table:

1. Country: The country in which the customer is located.
2. CustomerID: The customer id.
3. m_amt: The average revenue per invoice to this customer.
4. m_item: The average number of items ordered by this customer, per invoice.
5. m_uq: The average number of unique items ordered by this customer, per invoice.
6. n_inv: The number of invoices issued to this customer.
7. group: A grouping of countries into continents, except for United Kingdom.
8. grp2: Same as `group`, except for an abbreviation of United Kingdom to UK.

Explore the data, and create a visualisation that uses at least 3 of the variables in this dataset. You can of course join it with the other two tables if you need to.