Assignment 3: For 20% module credit

Assignments must be completed individually (you may not work together) and must be submitted on time (late assignments are not accepted).

Submission deadline: 1.00pm 3rd May 2022 uploaded to the Brightspace module as a single pdf document only. [NB: multiple documents/other document formats are not accepted.]

Customer churn is the loss of customers to a company. Data on the length of time customers remained with a telecommunication company will be considered. These data include details of the customer (e.g. gender, household type etc.) and details of the services consumed (e.g. monthly charges, internet service, payment method etc.). The purpose of this analysis will be to consider predictors (risk factors) of customer churn.

The dataset includes the following variables:

Variable name	Details
customerid	Unique customer ID.
gender	Male or Female.
partner	Live with a partner (yes or no).
monthlycharges	average charge on monthly bill.
paymentmethod	Method of paying their bill. One of: Bank transfer; Credit card; Electronic check; Mailed check
internetservice	One of: DSL; Fibre Optic; No
churn (binary response variable)	Whether the customer was lost (i.e. churned) within two years. 1=churned, 0=did not churn.

1. Write an R programme to read in the telecoms_churn.csv dataset and analysis these data. The response is churn and you should fit a model with all predictors. Use R to calculate an estimate of the probability of churning for: a male, living with his partner, monthly charges=70, internet service=fibre optic, payment method= credit card. Use R to get a 95% confidence interval for this estimate.

2. Write a report (4 pages maximum, excluding the R-code) on your conclusions from this analysis. You might include discussions of the following: Which predictors are significantly related to the response? How is each predictor related to the response (odds ratios etc). How do different levels of categorical predictors compare to each other?

NB: Reports exceeding the 4 page maximum will be penalised.

Methods expected to be used:
Methods covered in the module up to and including video Regression_8A and lecture notes 'Lecture_8'.