# Major assignment 1: Your submission

This is your assignment template for [AnalyticsX Major assignment 1](https://courses.edx.org/courses/course-v1:AdelaideX+AnalyticsX+1T2021/courseware/e0b375053b6441a08d658133752b3531/b06e32d08342432b8ccd0f2c738f5bb8/1?activate_block_id=block-v1%3AAdelaideX%2BAnalyticsX%2B1T2021%2Btype%40vertical%2Bblock%40457a6f9c5f7444c286c6f3a3e2851bc9). Save this document on our local machine and include all of your work within the relevant sections. Once you’ve completed all five parts of the assignment, upload the document via the submission area on the “[Submit your assignment](https://courses.edx.org/courses/course-v1:AdelaideX+AnalyticsX+1T2021/courseware/e0b375053b6441a08d658133752b3531/b06e32d08342432b8ccd0f2c738f5bb8/7?activate_block_id=block-v1%3AAdelaideX%2BAnalyticsX%2B1T2021%2Btype%40vertical%2Bblock%404de5ad6a293e4d6092d1d5756facfc9d)” page at the end of Major assignment 1.

# Checklist

* Have you shown all of your working?
* Have you given all numbers to the correct number of decimal places?
* Have you included all R output and plots to support your answers where necessary?
* Have you included all of your R code – input and output?
* Have you made sure that all plots and tables each have a meaningful caption?

**Quick links:**

[Major assignment 1: Part 1](#_Major_assignment_1:)

[Major assignment 1: Part 2](#_Major_assignment_1:_1)

[Major assignment 1: Part 3](#_Major_assignment_1:_2)

[Major assignment 1: Part 4](#_Major_assignment_1:_3)

[Major assignment 1: Part 5](#_Major_assignment_1:_4)

# Major assignment 1: Part 1

1. Load in the flights dataset from the nycflights13 package [1 point]

Your input code and output code for the flights dataset from the nycflights13 package go here:

1. Produce a scatterplot of dep\_delay against arr\_delay [3 points]

Your plot, code and caption go here:

1. Fit a linear regression model with arr\_delay as the response variable and dep\_delay as the predictor, and obtain the intercept and slope of the fitted model. [3 points]

Your code goes here:

1. Produce a residual versus fitted plot for the fitted model [1 point]

Your code and plot go here:

1. Produce a normal QQ\_plot for the fitted model [1 point]

Your code and plot go here:

6. Predict the expected arrival delay for flights with a departure delay of 100 minutes [1 point]

Your code (as evidence of your answer to Part 1, Question 6) goes here:

# Major assignment 1: Part 2

1. Load in the flights dataset from the nycflights13 package [1 point]

Your input code and output code for the flights dataset from the nycflights13 package go here:

1. Produce a scatterplot of arr\_delay against dep\_delay with colours for the different origins. Add a linear regression line to the plot for each origin. [2 points]

Your plot and code go here:

1. Fit a linear model with arr\_delay as the response variable; dep\_delay as a predictor; origin as another predictor; and an interaction term between dep\_delay and origin. [2 points]

Your input and output code goes here:

1. Produce a table of the coefficients for each term in the model. [1 point]

Your input and output code go here:

# Major assignment 1: Part 3

1. Load in the flights dataset from the nycflights13 package [1 point]

Your input code and output code for the flights dataset from the nycflights13 package go here:

1. Produce a nested dataframe with a row for each carrier [2 points]

Your input and output code goes here:

1. Fit a linear model, for each carrier, of arrival delay regressed on departure delay. [5 points]

Your input and output code goes here:

1. Produce a table of the coefficients for each carrier model, and display the portion of the table for carrier YV. [3 points]

Your input and output code goes here:

# Major assignment 1: Part 4

1. Load in the flights dataset from the nycflights13 package [1 point]

Your input code and output code for the flights dataset from the nycflights13 package go here:

1. Create a new variable called delayed that has the value 1 if the arrival delay is greater than zero, and value 0 if the arrival delay is less than or equal to zero. Count the number of delayed flights. [4 points]

Your input and output code goes here:

1. Fit a logistic regression model with delayed as the response variable and origin as the predictor. [3 points]

Your input and output code goes here:

1. Produce a table of the coefficients for the logistic regression. [5 points]

Your input and output code goes here:

# Major assignment 1: Part 5

1. Set the random seed to be 19, then load the flights dataset from the nycflights13 package. Count the number of flights in the dataset that arrived at their destination early. [3 points]

Your input and output code goes here:

1. Build a model to predict arrival delay using origin, departure delay, flight time, carrier, distance travelled, year, month, day and hour. This is a "full" model compared to those in Parts 1-4 of this assignment. [3 points]

Your input and output code goes here:

1. Use backward elimination starting from this model to create a better linear regression model to predict arrival delay. [3 points]

Your input and output code goes here:

1. Apply 10-fold cross-validation, to the model. [2 points]

Your input and output code goes here:

1. Compute the average mean-square prediction error for arrival delay using this model. [3 points]

Your input and output code goes here: