ISOM3530 HW2
Due: 5pm 28 Mar

- You can collaborate with your classmates for the assignments. If you work in groups, please list the names of your group members in the report. Submit by one member only.
- You need to submit both the report and the source code.

There will be new faces on the city's roads later this month. Hong Kong will roll out five new taxi fleets, a regulatory new regime designed to enhance taxi services. Analyzing taxi ridership could present valuable insights to city planners and taxi dispatchers, such as how to position cabs where they are most needed, how many taxis to dispatch, and how ridership varies over time. A dataset (taxi.csv) contains:

TimeMin	The time of a day (in minutes)
DayOfWeek	Day of the week; 1 = Monday…7 = Sunday
PickupCount	The pickup count

Preliminary study
Draw a scatter plot for 'PickupCount' (y-axis) and 'TimeMin' (x-axis). Is it a good choice to use linear regression? Explain.

Feature engineering
1. Divide the 'TimeMin' into 4 uniform time intervals. Name it as 'TimeQ'.
2. Classify the 'WeekOfDay' in to an indicator of weekday. Name it as 'isWeekDay'.

Modeling
1. According to the scatter plot in preliminary study, is it reasonable to model PickupCount and TimeMin by a polynomial regression? If yes, what is a reasonable choice of the order? Then, fit such model (model.1) and report the Adjusted Rsquare
2. Fit a cubic spline (model.2) for 'PickupCount' by 'TimeMin' with 3 evenly spaced knots (report the location of each knot). Report the Adjusted Rsquare.
3. Fit a multiple linear regression model (model.3) for 'PickupCount' by 'TimeQ' and 'isWeekDay'. Report the Adjusted Rsquare.
4. Fit a multiple linear regression model (model.4) with 'TimeQ', 'isWeekDay' and their interaction term. Report the Adjusted Rsquare.
5. Fit a GAM (model.5) with the following features:
   a. Cubic spline by 'TimeMin' (with the same knots of model.2)
   b. 'TimeQ' , 'isWeekDay' and their interaction terms.
   Report the Adjusted Rsquare.

Business Insights
1. Pick the model with the highest Adjusted Rsquare as the final model. Give a short business insight from the regression model (~ 50 words).