# Orbital Insight Data Set Challenge

*Please do not distribute this document or the accompanying data.*

**Context**

      A client has asked us to provide a daily data feed that forecasts car traffic to a particular parking lot. We need to provide a time series of historical data, as well as updates to this time series in the form of one new data point each day. This new data point is to be a forecast of the number of cars that we expect to show up tomorrow. We have historical imagery of this parking lot going back ~5 years, from which we have counted the daily number of cars in each image using a computer vision algorithm. Additionally, each day we receive a new image of this parking lot from which we will count the number of cars that are actually in the parking lot (this information will be used to update our time series). The included CSV file contains the relevant historical data we have extracted from the output of this algorithm. Please answer the questions below and include your code and supporting figures along with your answers. **We are looking to evaluate your answers, the evidence you use to justify them, the quality of your code (particularly for part 2), and how you structure your code. Please use python for this exercise; while you are free to use whatever Python packages you like, we recommend that you only leverage numpy, pandas, and matplotlib (more advanced things are not necessary to solve these problems).** We want to see how you solve the problem and how well you understand the techniques you leverage.

*Recommended time: 2-3 hours*

**Questions**
1. Data exploration
   a. What are some characteristics of the data you notice?
   b. How are the car counts related to the cloud indicator variable? What do you think explains this relationship? Is this data reliable, and why or why not?
   c. Is there a significant dependence on the number of cars with time? If so, on what timescales? Are there trends with month of year or day of week? Why or why not?
   d. How are car counts related to weather? How significant do you think the weather will be in our final model?
   e. How does what you found from this data exploration influence your path ahead in working with this data?
2. Forecasting

a. Construct a simple model to fit the car counts over time and predict car traffic one day in advance. Discuss the parameters of the model, and justify their inclusion. How do you guard against over-fitting? How did you treat the cloudy data?
b. Are the residuals from your best fit model stationary? How can we test this, and what does this tell us?
c. How can you validate/test how well your model will perform out of sample?
d. How could we generate uncertainties on our forecasted values?