**Idea #1 Kaggle completion “Yelp Restaurant Photo Classification”**

In this Kaggle competition, participants are asked to build a model that will automatically tags restaurants with multiple labels using photos of restaurants submitted by Yelp users. The competition provides an excellent opportunity to practise machine learning classification algorithms and image processing techniques.

**Idea #2 Bank Marketing Data set: will the client subscribe a term deposit?**

The data set is based on the marketing campaign conducted by a Portuguese banking institution (<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>). The campaign was based on phone calls. The task for the project is to build a classification model to determine whether a client subscribes a term deposit (binary: yes or no) based on client’s age, job, marital status, loan status etc. The data set contains 45211 instances and 17 attributes.

**Idea #3 London Underground data**

“Transport for London” (<https://tfl.gov.uk/info-for/open-data-users/our-feeds>) provides Oyster card journey information (5% sample of all Oyster card journeys performed in a week during November 2009 on bus, Tube, DLR and London Overground) and passenger counts data (passenger numbers entering and exiting London Underground stations). The potential project will aim to answer the following questions: “What are the busiest tube stations in London Underground?”, “During which times of day/weekday?” and help Londoners to plan their journeys around the city. The project will provide a good ground to practice data wrangling and data visualization skills. However the data is “patchy” in time. Therefore, the data set could be rather used for the “Data Story” assignment.