<https://www.kaggle.com/uciml/student-alcohol-consumption>

The data were obtained in a survey of students math and Portuguese language courses in secondary school. It contains a lot of interesting social, gender and study information about students. You can use it for some EDA or try to predict students final grade.

I like this because it has many interesting variables and it would be interesting to see which of them are key influencers on grades and also investigate other correlations. This info could be useful to student entry officers in evaluating potential strong students and those who may require additional support based on their demographics.

<https://www.kaggle.com/chasebank/bank-deposits>

This shows deposits for Chase Bank branches over a number of years. I would go about looking for trends in deposits and location , establishment dates and predict deposits for the next year ( though historic data here is small). Questions I could ask would potentially be ‘What would potentially be the total deposits In the next period for a particular branch?’ or ‘ Do the newly formed branches generate more deposits?’

<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>

This shows data from 4 locations of patients with potential heart disease symptoms and their underlying conditions.

I can use this dataset to ask questions like ‘which factors highly influence the presence of heart disease?’, ‘which factors least influence heart disease?’, ‘which conditions show strong correlations?’ .I can split this data and predict those the which ones have heart disease and also ask questions of the data (EDA)

Road Collisions In Camden

<https://data.gov.uk/dataset/road-collisions-in-camden>

This dataset contains road collisions in Camden dating back to 2005. Data is provided by TfL (Transport for London) on an annual basis, in three parts: attendant, casualty and vehicle – these are joined by the Accident Reference before being uploaded to the open data platform; therefore where multiple attendants/casualties/vehicles exists for one Accident Reference there will be multiple records for the same accident. To avoid anomalies it was suggested any data analysis undertaken should use three years of data.

Data could be used to determine the when collisions mostly occured and also what some of the potential key influencers were.