Capstone Project Ideas

Idea 1: London Bike Sharing

Potential client(s):
Transport for London
Londoners looking to save time on their commute

Issue:
Santander cycles (formerly Barclays Cycle Hire and popularly known as Boris Bikes) started operating in London in July 2010 with 5,000 bikes and 315 docking stations, this has grown to 11,500 bikes and over 750 docking stations.
There's a £2 access fee, then journeys under 30 minutes are free, or £90 for annual access.
Previous user complaints have included a lack of bikes and/or docking stations near major railway stations, and poor coverage in relatively central areas (e.g. SE London) and the suburbs.

Aim:
Exploratory analysis and visualisation of the available dataset. Given daily bike rental and weather records predicting future daily bike rental demand (or possibly hourly). Identifying potential areas for new (or replacement) docking stations.

Obtaining data:
The monthly and annual data is provided by Transport for London in .csv format beginning from 4th January 2012. Fields included are: Rental Id, Duration, Bike Id, Start Date, Start Station Name, Start Station Id, End Date, End Station Name & End Station Id.
Link to main dataset

This provides a live feed of bike availability including the number of docking points at each station. Web scraping may need to be utilised to obtain the dock locations.

Link to docking stations page

There are a significant number of possible complimentary datasets which could be used to investigate such as Met Office weather data & annual public and school holiday calendars.

Historical rainfall data

Potential issues:

Utilisation of time series data and associated analysis. Dataset size is within the big data realm - approx. 9 million hires per year.

Idea 2: The Lidl Effect?

Potential client(s):

Home Owners

Home Buyers

Estate Agents

Issue:

Early in 2017 the national newspapers reported on the 'Waitrose efffect', the idea that a property being located close to a Waitrose supermarket (a high end British supermarket chain) can add over £36,000 to a property price See.

Aim:

Build a testable hypothesis and explore whether it can be shown that being close to a Lidl (a popular discount supermarket chain) has a statistically significant affect on house prices.

Obtaining data:

The Land Registry provides access to price paid data, published on a monthly basis, going back to 1995. Fields included are: unique Id, price, date of transfer, postcode, property type, old/new, duration (freehold/leasehold), address.

Link to main dataset

A comprehensive list of Lidl store locations is not readily available and would need to be scraped from their website.

Potential issues:

There are many other factors affecting the price of a house, on top of proximity to supermarkets. In London particularly, supermarkets are often located in close proximity to tube/railway stations for convenience, therefore it may be difficult to identify the supermarket affect in isolation.

Idea 3: The Three P's - Putney's Pollution Problem

Potential client(s):

Wandsworth Council

Local Residents

Issue:

At 7am on Friday 8th January 2016, Putney High Street in West London breached ANNUAL limits for nitrogen dioxide (NO_2), a toxic gas produced by diesel vehicles that has been linked to respiratory and heart problems. Under EU rules, sites are only allowed to breach hourly limits of 200 micrograms of NO_2 per cubic metre of air 18 times in a year, but Putney broke that limit for the 19th time. A report commissioned by mayor Boris Johnson in 2015 found that nearly

9,500 people die prematurely each year because of the capital's dirty air.See

Wandsworth council have persuaded TFL to replace or up-grade polluting buses in Putney High Street. Daytime delivery vehicles have been banned from stopping and measures introduced to ease traffic flow. Putney High Street also became the first Low Emission Bus Zone on the 9th March. The clean bus zone, which runs a total of 145 buses on seven scheduled routes, is now serviced by cleaner buses in a bold move to cut harmful nitrogen oxide (NOx) emissions. See

Aim:

Predict the air quality on a given day based on previous years' pollution.

Produce a dashboard or app that provides access to prediction and a warning if levels are high to take alternative routes for pedestrians/cyclists.

Obtaining data:

This would require accessing the London Air API in association with King's College. Information is requested using web page (http) protocol, and the exact request allows the definition of variables such as dates, that specify exactly what is wanted

Link to API

Potential issues:

Unknown how easy it will be to access the API. Since measures are actively being taken to reduce the air pollution, it may be difficult to come up with accurate predictions.