**Springboard Capstone Project Ideas**

Based on the Springboard Project guidelines and some real-world problems that are of interest to me, I have attempted to identify three Capstone project ideas that are real world, medium complex, not too fancy and may have potential to be developed and expanded into a more elaborate, stand-alone data analytics program that could service multiple stakeholders. The datasets were from Kaggle and official Melbourne government website with detailed information such as data characteristics, data sources and their context. Datasets selection are listed below in order of preference based on my personal interests and the industries where I plan to search for new job opportunities in the near future.

**Springboard Capstone Project Idea #1**

**Industry:** Real Estate, Mortgage Lending

**Potential Stakeholders:** Bankers/Mortgage Lenders, Buyers, Investors, Property Owners

**Dataset name:** Melbourne Housing Market – housing clearance data from Jan 2016

**Link:** <https://www.kaggle.com/anthonypino/melbourne-housing-market>

**Data Type:** CSV

**Data Size** (FULL dataset): 34858 rows, 21 columns; (LESS dataset): 63024 rows, 13 columns

This dataset is my preferred choice for the Capstone project as I live in Melbourne and always keenly interested in the real estate market. This dataset is from Kaggle and has been extracted from a well-known retail estate website (domain.com.au). Also, as I work in a top national bank, where data analytics in lending business has immense opportunity and this potential project would give me a good exposure to the kind of data analytics projects undertaken by banks and lending institutions. This dataset has a huge potential for various types of stakeholders, as well as analysis from varied perspectives. Some of the questions that could potentially be addressed are: comparison of real estate selling clearance rates over the years, is there an increase/decrease of real estate sales over the years, is there a trend, which suburbs sell better than the others, which is the best time to buy or sell etc.

**Springboard Capstone Project Idea #2**

**Industry:** Business/Retail

## **Dataset name:** Business establishments, with trading name and industry classification, Melbourne (2002-2017)

# Link: <https://data.melbourne.vic.gov.au/Economy/Business-establishments-with-trading-name-and-indu/bs7n-5veh>

**Data Type:** CSV

**Data Size:** 266797 rows, 11 columns

This dataset is from the official government website of the City of Melbourne’s open data platform and presents the census data of various business establishments in a particular locality in Melbourne over the years. The dataset seems interesting with potential answers to Retail industries and businesses to determine the demographics and presence of various industries in a given suburb/area. The stakeholder can determine the potential of setting up a business/retail in a particular location and analyze the demand and competition in the area. I am interested in this data set as I am looking for opportunities in the Retail IT sector and this project would be an interesting case study to present during my interviews. The dataset does not look very complex though and the nature of questions that can be answered is not very obvious at this stage.

**Springboard Capstone Project Idea #3**

**Industry:** Telecom, Cyber security, e-business

## **Dataset name:** Household use of Information Technology, Australia, 2016-17

# Link: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/8146.0Main+Features12016-17?OpenDocument>

**Data Type:** CSV/Excel

This seems to be an elaborate dataset spread over multiple spreadsheets. Data clean-up and data wrangling will be a major portion of the work in order to obtain a unified data set for analysis. This dataset can address various questions such as the trend in the use of internet, which age group uses internet and for what purposes, how would this data benefit e-businesses and e-marketing.