CAPSTONE PROJECT

Applied Data Science by IBM/Coursera

ABC Bookstore

1. Introduction

As Nelson Mandela says "Education is the most powerful weapon which you can use to change the world". Education not only makes a person knowledgeable but also helps them grow and grab better opportunities in their life to be successful. It also makes an individual more responsible, take the right decisions and develop them personally and socially. It also helps one to do their daily activities in the best possible ways. It helps in acquiring knowledge and new skills that will impact their personal development and for their country. To study, one needs to have access to good resources of books and a book store.

2. Business Problem

ABC company is located in New York, is a known and independent book store focused on staff curated book selections. Stores are generally located near a school mainly helping students with academic books and stationaries. The bookstore will offer a wide selection of books including children's literature, modern fiction, true crime, cookbooks, foreign language titles and art books. The cosy bookstore will also offer a book club, author events and children's story hour and gift-wrapping section. ABC company is looking to help students and community by opening a new branch in Sydney, Australia.

To open a bookstore, a few things that will be considered while finalizing the location. Below are the important things that need to be taken into consideration:

1. The bookstore should be located within a 2 km radius of a school.
2. The school rating should be 7 or above out of 10.
3. There should not be more than 1 bookstore within a 2 km radius of the school.
4. There should be a bank/atm in the vicinity.

3. Data Location Gathering:

In Sydney, there are a lot of schools but very limited bookstore available. In order to finalize an optimum location, data from Foursquare.com will be gathered. It will also help us grab the rating for different schools along with the number of bookstores or bank available in its approximation areas. Foursquare has rich information available and above all, all of this information can be extracted by using their API service. This API service will gather a list of all the schools available in Sydney along with their rating. Once all the schools' location (latitude and longitude) and filtered with rating 7 or above, we will look for other criteria such as the number of other bookstores and bank/atm availability.

In order to generate the best location, we would need to find

- schools in Sydney
- schools average rating
- number of existing books stores around the school
- other facilities such as bank/atm

We will look for schools within a radius of 30 Km from Sydney. For the existing bookstore or bank/atm we will search within 2 km of radius from the school.

4. Data Acquisition and cleaning

4.1 Data Sources

Retrieving the data is the most important aspect of any data analysis. For our analysis, we used Foursquare API to pull a list of all the schools and list of all book stores, banks and ATM's near each school that was required for the data analysis. There are various sources that provide the data but either they do not have any API service to use or have a very limited number of calls we can make through their website for free. Foursquare helped us by letting us scrap most of the data in the first couple of days.

4.2 Data Cleaning

Data scraped from various sources were combined, cleaned and stored in various dataframe for analysis. There were several problems with the data available.
Firstly, there were missing data that we couldn't use, hence needs to be dropped.
Secondly, the datatypes were converted to a proper format
Thirdly, some of them were having null values that were replaced will 0.

4.3 Data Selection

While considering the data, we need to make sure we use the data from the correct location. For that, we used a radius of 30 km of Sydney location. It retrieved 100 schools. Later, we also need to retrieve their average rating. And finally, we need to pull a list of all book stores, banks, atm within 2 km radius of each school.

5. Exploratory Data Analysis

Once we retrieved and data was pre-processed, we plotted all our schools, book stores, banks and atm's on the map to visualize it further (Figure 1).
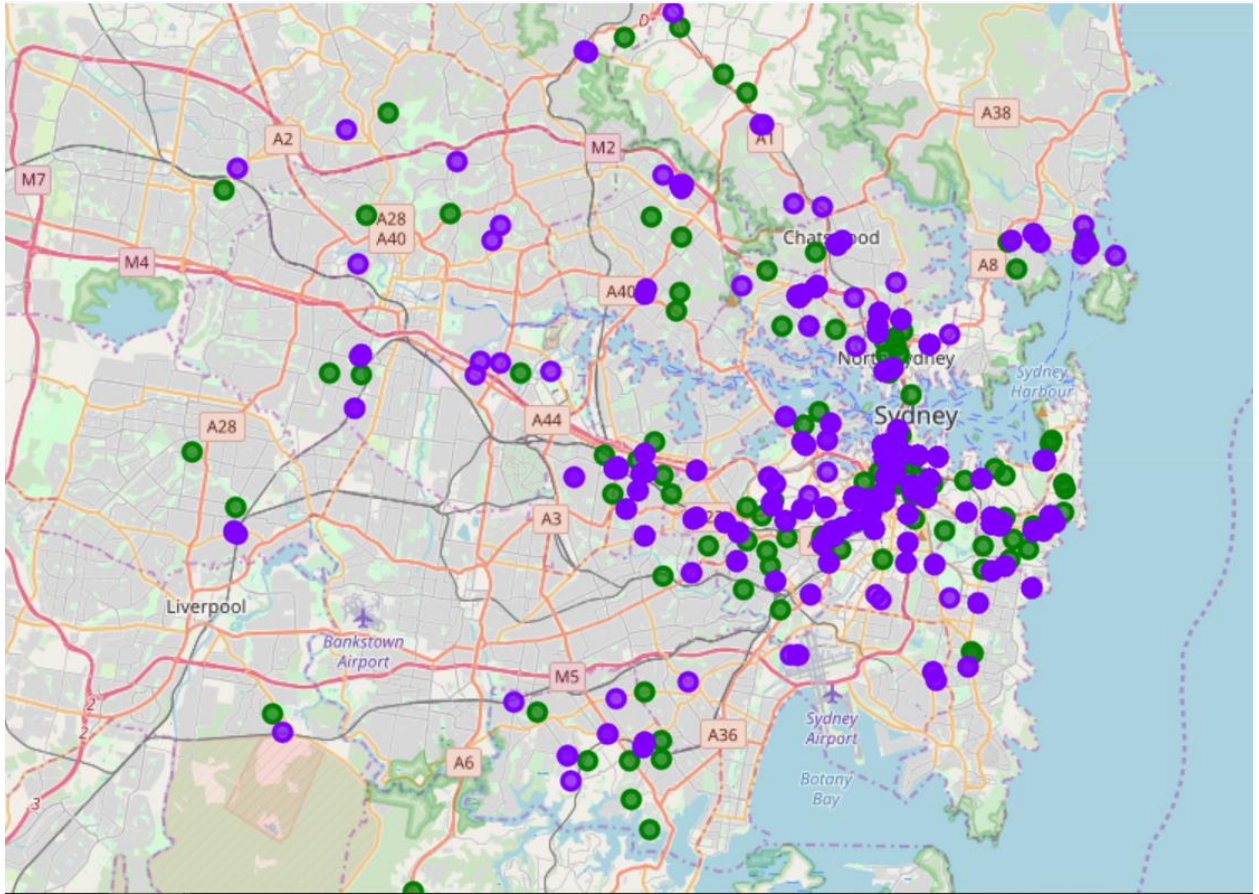
Fig 1: Location of all 100 schools, ~800 book stores and ~400 Banks/ATM's in Sydney

This image has around 800 book stores, 400 banks/atm and 100 schools.

There were few places which have a staggering number of book stores within 2 km from these schools. Some of the schools have 35 book stores around them. This was again plotted to see where these were located. There were 8 schools around which more than 30 book stores were present within its 2 km radius (Figure 2).
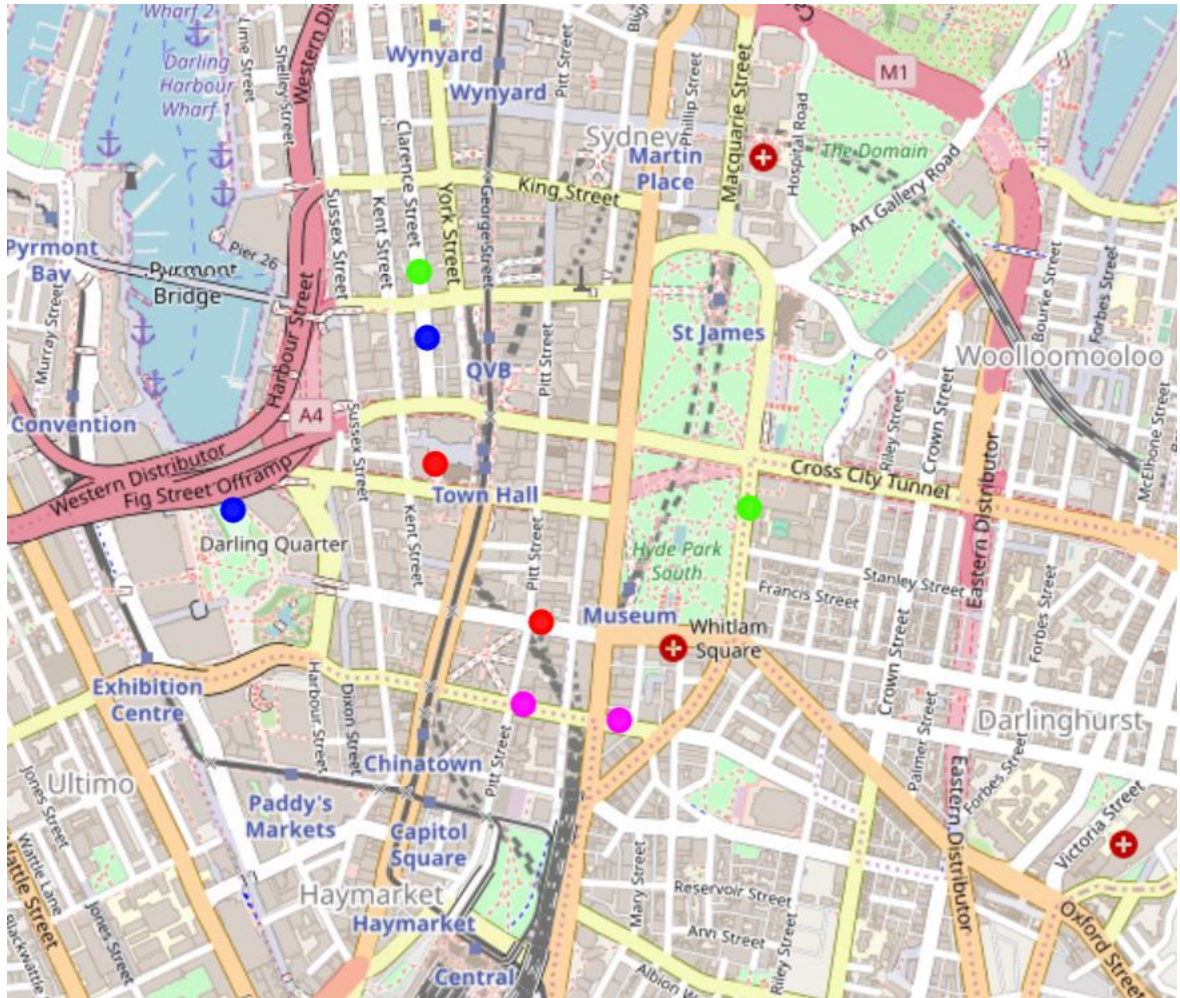
Fig 2: Location of 8 schools having 30+ book store existing around each school in Sydney

Finally, after sorting and filtering our data, we were able to gather our prime candidates with least number of book stores in its vicinity and have at least 1 bank/atm with its 2 km radius. Figure 3 shows the locations of these schools.
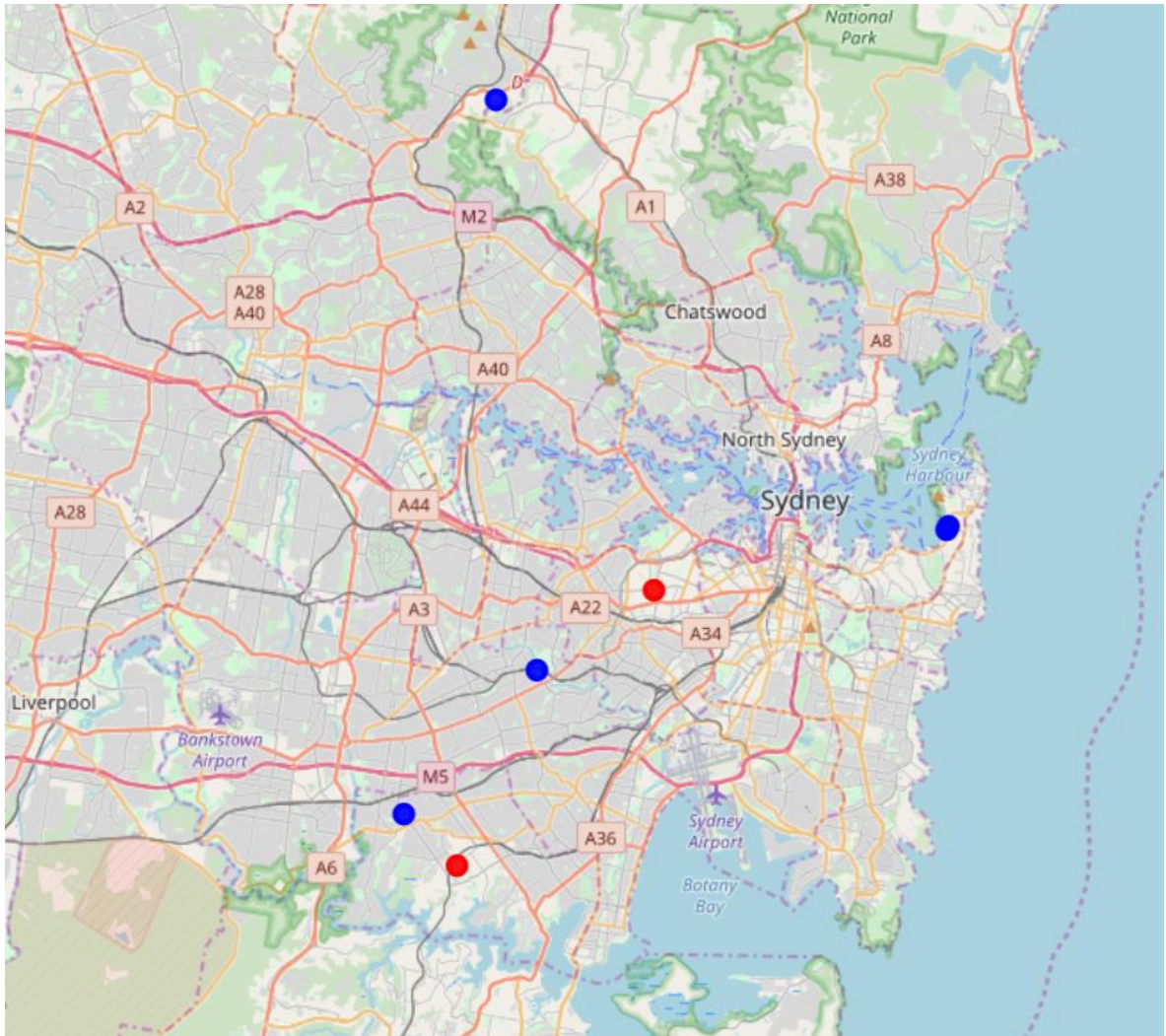
Fig 3: Location of 7 schools having 1 book store and at least 1 bank/atm around each school in Sydney

6. Results and Discussions

Our results show that although we had 100 schools with around 800 books store and approximately 400 atm's/banks within a radius of 2 km from the schools, there are few schools where a number of books store was minimal. There were few schools around which a staggering number of book stores existed (35 book stores). Most of these schools were located in the Central/Financial district of Sydney. Our potential location candidate for book stores will be in the suburbs of Sydney.

With narrowing down all the available schools and as per the requirement, there were 7 schools that have at least 1 atm/bank and at max 1 book store within its vicinity. These schools are located in:

1. one school in the upper north suburb of Sydney
2. two schools in the western suburb of Sydney
3. two schools in the south-west suburb of Sydney and
4. two schools in the eastern suburb of Sydney

7. Conclusions

Purpose of this project was to identify an optimal location for opening a book store with the least number of bookstores around a school and have at least one bank or atm near it. Identifying these schools, existing bank stores and banks or ATM's according to with the help for Foursquare API's has helped achieve our requirement by the exploration of these data points and can be shared with the stakeholders.

Based on the results, a stakeholder can take an appropriate decision based on other specific characteristics such as real estate price and availability, number of people living in that neighbourhood, accessibility to parking, public transportation in each recommended zone. They can also take into considerations like accessibility to highway or types of lease available for the shop.