

**Coursera Capstone**  
**IBM Applied Data Science Capstone**

***Open a new Coffeehouse in Sydney, Australia***

By: Sapna Yadav

July, 2020



# Introduction

---

Sydney (Australia) is the state capital of New South Wales and the most populous city in Australia.

In this modern world where most of our communications are online, there is a growing desire amongst people to create more personal moments in their lives. Coffee outlets offer the perfect meeting place

The coffee shop offers a uniquely calm atmosphere where people can gather with friends to catch up over coffee or focus on work in a relaxed environment. Coffee outlets also make the perfect place for informal meetings and discussions, thanks to the relaxed ambience. Coffee shops have become so popular over the last decade due to their relaxed and open feel. This is especially prevalent when you compare this establishment to other competing locations, such as restaurants, pubs and fast-food chains.

Coffee is an intrinsic part of Sydneysider's life where many of us turn to for both working and social occasions.

## Business Problem

---

One of the popular retail coffeehouse chain company is planning to open a new coffee shop in Sydney. It is trying to decide which suburb of Sydney would be the best one to locate the Coffee shop.

The Company is much concerned about the location. It wants to prefer those location only where the coffee shops are the most visited venue so that will be good opportunity for them to open a coffee shop in that area.

The objective of this capstone project is to analyse and select the best suburb location in Sydney to open a coffee house or café. Using Data Science methodologies like data collection, data cleaning, data analysis, this project aims to provide solutions to answer the business problem:

***In the city of Sydney, Australia, if a retail coffeehouse company is looking to open a new café or coffee-shop, where would you recommend that they open it?***

# Target Audience

---

This project is particularly useful to investors, business owners looking to expand their business in one of the most liveable city of Australia.

## Data

---

The following two datasets are used to solve the above problem-

- (a) **List of suburbs in Sydney:** This defines the scope of this project. The first dataset is used to retrieve the name with postcodes of all suburbs in Sydney. This data is further helpful to get the coordinates of each location in order to plot the map and to get the venue data. This is done by using the following link.

<https://www.intosydneydirectory.com.au/sydney-postcodes.php>

In this link all suburbs of Sydney with their post-codes are given.

- (b) **List of venues suburb wise:**  
The second dataset is used to get all information about venues like category, venue location etc. located in each suburb. And after that Coffee shop venue is filtered out. This dataset is retrieved with Foursquare Location API