A. Problem Description

Overview

I have an amazing opportunity from my company that will set up its primary IT services operations in New Castle, New South Wales. I have been asked to lead the effort of setting up operations and ensure we identify the right location for the office space and also a great neighborhood for the pool of talent we will be hiring.

As a part of this major project I will be taking you through my journey of exploration and give you an amazing glimpse of Newcastle and its surroundings. Are you ready to come along and find a great place to work and live in Newcastle?

Our company also intends to use this study to share the details with our prospective employees who we are currently recruiting. It is very critical to choose the right location for the office as it will serve as a hub for our operations in the New South Wales region. It is important to be cost effective but also be closer to bigger cities like Sydney which is the capital of New South Wales to ensure we attract the right pool of talent. Our company has major offices in Adelaide, South Australia and Brisbane, Queensland and is now planning to expand in New Castle, New South Wales as part of a major planned growth and expansion opportunities due to the increased revenue projections of our service offerings.

Audience

Primary Audience will be our Corporate and Real Estate Strategy of our company who will be evaluating the final results and recommendations of

this study to determine the site and office space for our New South Wales office in New Castle.

B. Data Description

Data Sources

1. https://simplemaps.com/static/data/country-cities/au/au.json
2. https://github.com/srisatakopan/coursera_capstone/blob/master/newcastle_neighborhood.csv

JSON Data Structure (au.json)


```
▼ root: [] 219 items
  ▼ 0:
      city: "Sydney"
      admin: "New South Wales"
      country: "Australia"
      population_proper: "3641422"
      iso2: "AU"
      capital: "admin"
      lat: "-33.861481"
      lng: "151.205475"
      population: "4630000"
```

JSON Data Structure processed into a Pandas Dataframe

	City	Admin	Country	Population Proper	ISO2	Capital	Lat	Lng	Population
0	Sydney	New South Wales	Australia	3641422	AU	admin	-33.861481	151.205475	4630000
1	South Melbourne	Victoria	Australia	93625	AU		-37.833333	144.966667	4170000
2	Brisbane	Queensland	Australia	926353	AU	admin	-27.47101	153.024292	1860000
3	Perth	Western Australia	Australia	880216	AU	admin	-31.95224	115.861397	1532000
4	Adelaide	South Australia	Australia	836354	AU	admin	-34.928661	138.598633	1145000
5	Newcastle	New South Wales	Australia	497955	AU		-32.927792	151.784485	1134616
6	Gold Coast	Queensland	Australia	332249	AU		-28.00029	153.430878	527660
7	Cranbourne	Victoria	Australia	39419	AU		-38.1	145.283333	460491
8	Canberra	Australian Capital Territory	Australia	234032	AU	primary	-35.27603	149.13435	327700
9	Wollongong	New South Wales	Australia	141725	AU		-34.424	150.893448	260914
10	Geelong	Victoria	Australia	137681	AU		-38.153461	144.358093	160991
11	Cairns	Queensland	Australia	109989	AU		-16.92304	145.766251	154225
12	Townsville	Queensland	Australia	119470	AU		-19.26639	146.805695	138954
13	Albury	New South Wales	Australia	32810	AU		-36.074823	146.924006	104258

Newcastle Neighborhoods (Manually created from Wikipedia)

However I needed more granular information about Newcastle and its neighborhoods that I was not able to locate a CSV or JSON file. So I built my own dataset of Newcastle Neighborhoods by looking up Wikipedia and getting each of it neighborhoods coordinates-

https://en.wikipedia.org/wiki/List_of_suburbs_in_Greater_Newcastle,_New_South_Wales

City of Newcastle suburbs [edit]

- Adamstown
- Adamstown Heights[n 1]
- Bar Beach
- Beresfield
- Birmingham Gardens
- Black Hill[n 2]
- Broadmeadow
- Callaghan (University)
- Carrington
- Cooks Hill
- Elermore Vale
- Fletcher
- Georgetown
- Hamilton
- Hamilton East
- Hamilton North
- Hamilton South
- Hexham
- The Hill
- Islington
- Jesmond
- The Junction
- Kooragang
- Kotara
- Lambton
- Lenaghan
- Maryland
- Maryville
- Mayfield
- Mayfield East
- Mayfield North
- Mayfield West
- Merewether
- Merewether Heights
- Minmi
- New Lambton
- New Lambton Heights[n 1]
- Newcastle
- Newcastle East
- Newcastle West
- North Lambton
- Rankin Park[n 1]
- Sandgate
- Shortland
- Stockton
- Tarro
- Tighes Hill
- Wallsend
- Warabrook
- Waratah
- Waratah West
- Wickham

df_newcastle_neighborhood = pd.read_csv('newcastle_neighborhood.csv')

```
df_newcastle_neighborhood
```

	Neighborhood	City	Lat	Lng
0	Adamstown	Newcastle	-32.9344	151.7258
1	Adamstown Heights	Newcastle	-32.9477	151.7152
2	Bar Beach	Newcastle	-32.9400	151.7678
3	Beresfield	Newcastle	-32.7990	151.6350
4	Birmingham Gardens	Newcastle	-32.8937	32.8937
5	Black Hill	Newcastle	-32.8400	151.6120
6	Broadmeadow	Newcastle	-32.9252	151.7342
7	Callaghan	Newcastle	-32.8900	151.7010
8	Carrington	Newcastle	-32.9150	151.7680

Use of Foursquare data

Foursquare API for Newcastle, New South Wales and its neighborhoods will be used to analyze nearby venues, get the coordinates of the venues to get more information of the neighborhood areas as shown below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adamstown	-32.9344	151.7258	Adams Ribs & Pizza	-32.934490	151.729330	Pizza Place
1	Adamstown	-32.9344	151.7258	Cupcake Espresso Adamstown	-32.935927	151.725734	Cupcake Shop
2	Adamstown	-32.9344	151.7258	Domino's Pizza	-32.934716	151.725785	Pizza Place
3	Adamstown	-32.9344	151.7258	7 Eleven	-32.934659	151.729345	Convenience Store
4	Adamstown	-32.9344	151.7258	The Nags Head Hotel	-32.935290	151.725370	Pub
5	Adamstown	-32.9344	151.7258	The Gates Hotel	-32.933370	151.721400	Beer Garden
6	Adamstown Heights	-32.9477	151.7152	Doors 4U Garage Doors	-32.948633	151.710709	Home Service
7	Adamstown Heights	-32.9477	151.7152	Mimco Kotara	-32.948681	151.719802	Jewelry Store
8	Bar Beach	-32.9400	151.7678	Bar Beach	-32.941137	151.768059	Beach

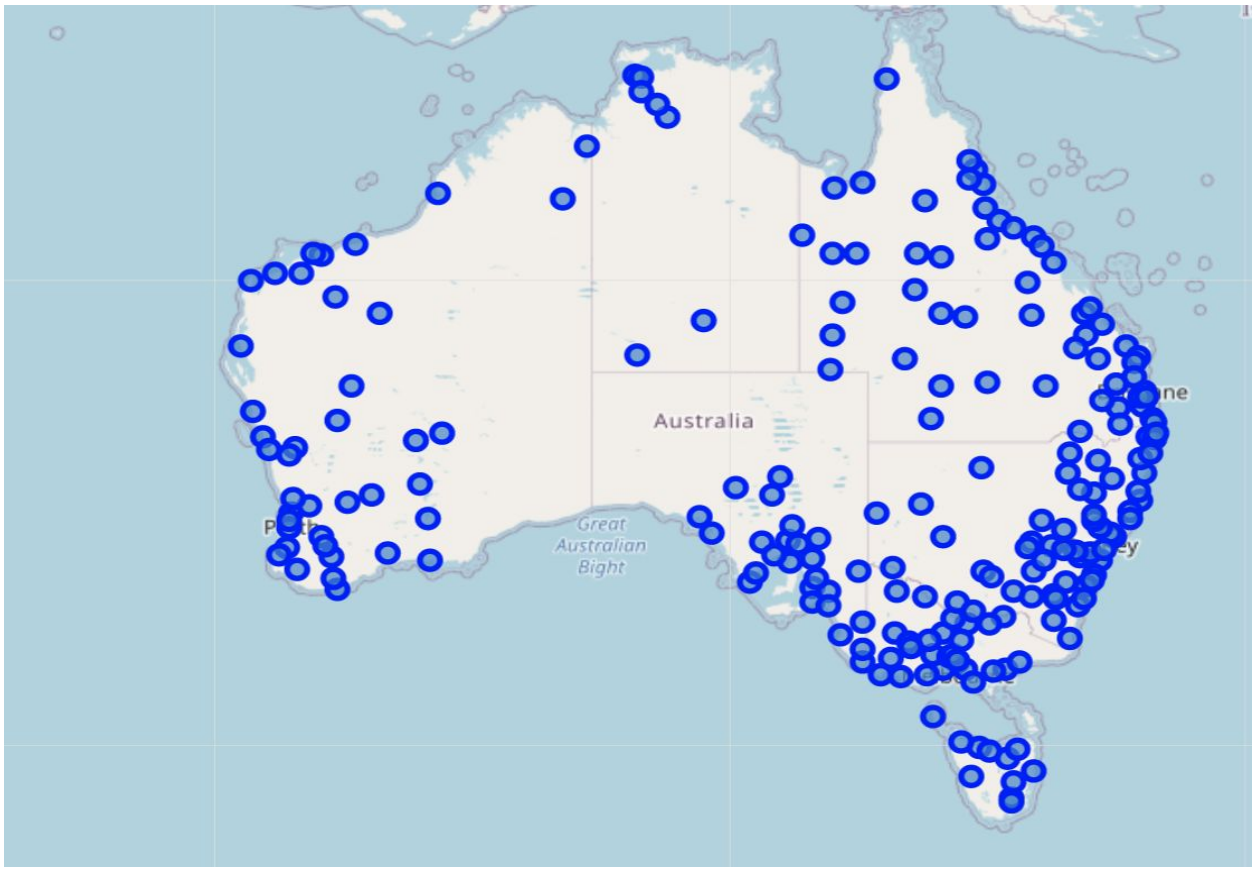
C. Methodology

Step 1 - Understanding Australian Territories and Cities

1. Unpacking the Australian Cities JSON File and extracting key information attributes
    a. 'City', 'Admin', 'Country', 'Population Proper','ISO2', 'Capital','Lat','Lng','Population'

2. Uploading the data from JSON to a dataframe

3. To understand the locations of the different territories and the geographical coordinates I used a folium geocoders python library to plot the coordinates of the Australian cities.



Exploring cities where Admin is New South Wales

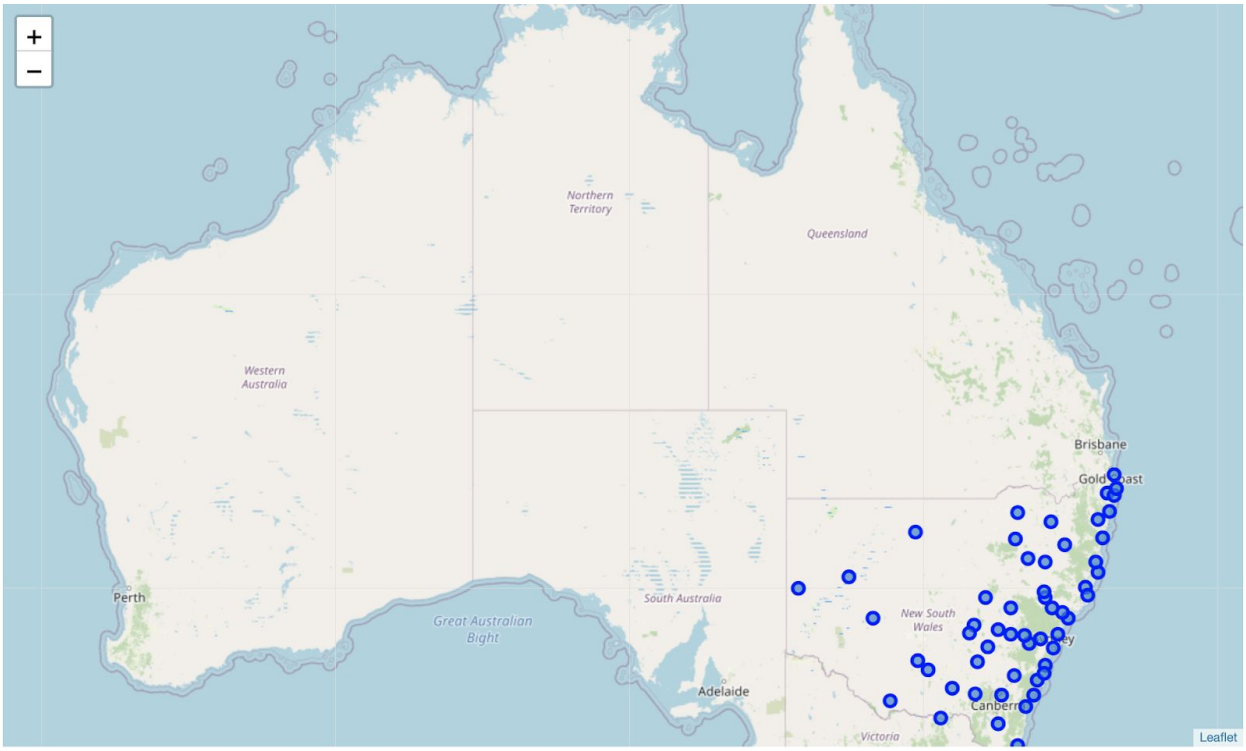
1. Used the Dataframe filtering method to filter where Admin = 'New South Wales' and started to explore deeper into this geographical area of Australia where our company wanted to set up their next major presence to expand the business operations.

[27]:

	City	Admin	Country	Population Proper	ISO2	Capital	Lat	Lng	Population
0	Sydney	New South Wales	Australia	3641422	AU	admin	-33.861481	151.205475	4630000
1	Newcastle	New South Wales	Australia	497955	AU		-32.927792	151.784485	1134616
2	Wollongong	New South Wales	Australia	141725	AU		-34.424	150.893448	260914
3	Albury	New South Wales	Australia	32810	AU		-36.074823	146.924006	104258
4	Nowra	New South Wales	Australia	27292	AU		-34.88422	150.600357	94781
5	Coffs Harbour	New South Wales	Australia	34944	AU		-30.296261	153.11351	62978
6	Wagga Wagga	New South Wales	Australia	35717	AU		-35.125771	147.353745	55381
7	Port Macquarie	New South Wales	Australia	35593	AU		-31.43084	152.908936	48547
8	Taree	New South Wales	Australia	16081	AU		-31.91099	152.453873	44182
9	Orange	New South Wales	Australia	34087	AU		-33.283089	149.100006	39329
10	West Tamworth	New South Wales	Australia	31609	AU		-31.10625	150.909356	38551
11	Tweed Heads	New South Wales	Australia	33065	AU		-28.17561	153.542007	33065
12	Queanbeyan	New South Wales	Australia	32214	AU		-35.354931	149.231995	32602
13	Dubbo	New South Wales	Australia	29361	AU		-32.24295	148.604843	31574

2. In the above data we can see that Sydney is the admin and the capital city of New South Wales.

3. Looking at a map of New South Wales would help in our exploration of this area so plotted this region on the map as shown below.



Exploring Neighborhoods of Newcastle

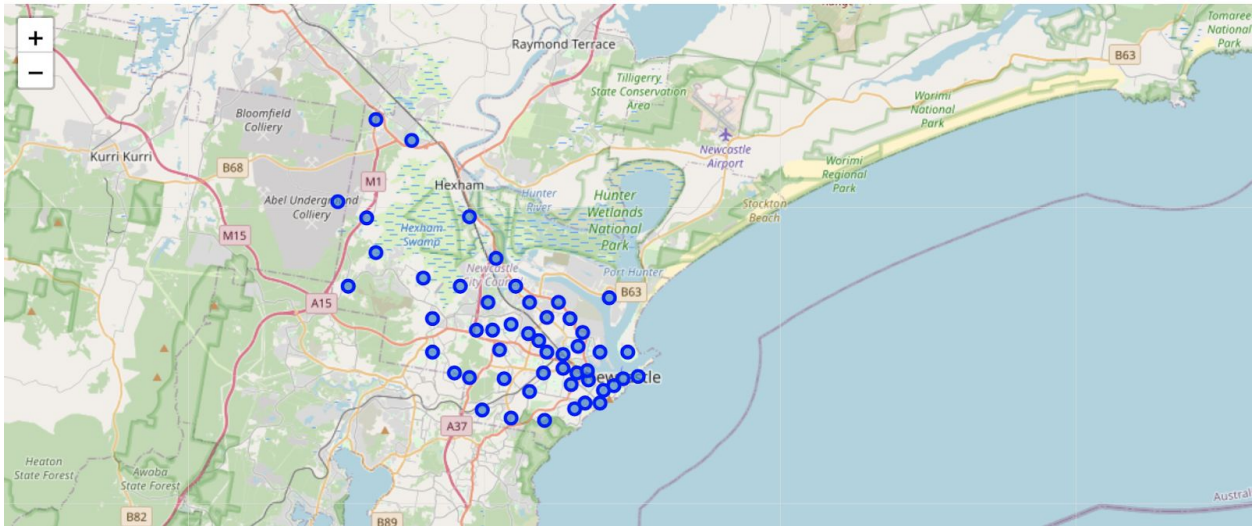
a. Using Wikipedia and searches I had to then create a CSV file for all the neighbourhoods of Newcastle and upload it into a pandas dataframe.

```
[18]: df_newcastle_neighborhood
```

[18]:

	Neighborhood	City	Lat	Lng
0	Adamstown	Newcastle	-32.9344	151.7258
1	Adamstown Heights	Newcastle	-32.9477	151.7152
2	Bar Beach	Newcastle	-32.9400	151.7678
3	Beresfield	Newcastle	-32.7990	151.6350
4	Birmingham Gardens	Newcastle	-32.8937	32.8937
5	Black Hill	Newcastle	-32.8400	151.6120
6	Broadmeadow	Newcastle	-32.9252	151.7342
7	Callaghan	Newcastle	-32.8900	151.7010

b. To get a better perspective of the neighborhood, I created a map view using the coordinate information.



Exploring Newcastle Neighborhood with Foursquare App

```
df_australia_nsw_reindexed.loc[1, 'City']
```

```
'Newcastle'
```

```
neighborhood_latitude = df_newcastle_neighborhood.loc[37, 'Lat'] # neighborhood latitude value
neighborhood_longitude = df_newcastle_neighborhood.loc[37, 'Lng'] # neighborhood longitude value
neighborhood_name = df_newcastle_neighborhood.loc[37, 'Neighborhood'] # neighborhood name
neighborhood_city = df_newcastle_neighborhood.loc[37, 'City'] # neighborhood name

print('Latitude and longitude values of {} in {} are {}, {}.'.format(neighborhood_name,
                                                               neighborhood_city,
                                                               neighborhood_latitude,
                                                               neighborhood_longitude))
```

```
Latitude and longitude values of Newcastle in  Newcastle are -32.9283, 151.7817.
```

```
# type your answer here

LIMIT = 100
radius = 1000
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
url
```

```
'https://api.foursquare.com/v2/venues/explore?&client_id=OXJOWOXEE34SFMUOHDR1MJUXUPZ3ZDHO542DK22REOXNXUAK&client_secret=IZ0UOVFNIL3PYSGNIGG
S4J0X3G4ODDREQO3AK1FJ2SXX2XJI&v=20180605&ll=-32.9283,151.7817&radius=1000&limit=100'
```

```
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '5eaab8a7b4b684001b4a2674'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Newcastle',
  'headerFullLocation': 'Newcastle',
  'headerLocationGranularity': 'city',
  'totalResults': 82,
  'suggestedBounds': {'ne': {'lat': -32.91929999099999,
    'lng': 151.79240255458265},
   'sw': {'lat': -32.93730000900001, 'lng': 151.77099744541735}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '51ddf6cb498e4db6b96b4862',
       'name': 'Xtraction',
       'location': {'address': 'Bolton Street',
        'lat': -32.928444644897596,
        'lng': 151.78319952369563,
        'labeledLatLngs': [{'label': 'display',
          'lat': -32.928444644897596,
          'lng': 151.78319952369563}],
        'distance': 141,
        'postalCode': '2300',
        'cc': 'AU',
        'city': 'Newcastle',
        'state': 'NSW',
        'country': 'Australia',
        'formattedAddress': ['Bolton Street',
         'Newcastle NSW 2300',
         'Australia']},
```

Getting the Venues in Newcastle from Foursquare Venue EndPoint

```
# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']
```

```
venues = results['response']['groups'][0]['items']

nearby_venues = pd.json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues
```

	name	categories	lat	lng
0	Xtraction	Café	-32.928445	151.783200
1	Reserve Wine Bar	Wine Bar	-32.927360	151.783070
2	Good Brother	Café	-32.928656	151.783220
3	One Penny Black	Breakfast Spot	-32.926633	151.779735
4	Estabar	Café	-32.929019	151.786629
5	Newcastle Beach	Beach	-32.929221	151.786746

Getting the coordinates of each of the venues in the Neighborhoods

[39]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adamstown	-32.9344	151.7258	Adams Ribs & Pizza	-32.934490	151.729330	Pizza Place
1	Adamstown	-32.9344	151.7258	Cupcake Espresso Adamstown	-32.935927	151.725734	Cupcake Shop
2	Adamstown	-32.9344	151.7258	Domino's Pizza	-32.934716	151.725785	Pizza Place
3	Adamstown	-32.9344	151.7258	7 Eleven	-32.934659	151.729345	Convenience Store
4	Adamstown	-32.9344	151.7258	The Nags Head Hotel	-32.935290	151.725370	Pub
5	Adamstown	-32.9344	151.7258	Ritz Cafe	-32.934840	151.726150	Fast Food Restaurant
6	Adamstown	-32.9344	151.7258	The Gates Hotel	-32.933370	151.721400	Beer Garden
7	Adamstown Heights	-32.9477	151.7152	Doors 4U Garage Doors	-32.948633	151.710709	Home Service
8	Adamstown Heights	-32.9477	151.7152	Mimco Kotara	-32.948681	151.719802	Jewelry Store
9	Bar Beach	-32.9400	151.7678	Bar Beach	-32.941137	151.768059	Beach
10	Bar Beach	-32.9400	151.7678	Bar Beach Kiosk	-32.941350	151.767556	Fast Food Restaurant
11	Bar Beach	-32.9400	151.7678	Howzat	-32.935996	151.768287	Gym / Fitness Center

Step 7 - Grouping and Sorting to understand Neighborhood Venues

[40]:
```
newcastle_venues_grouped = newcastle_venues.groupby('Neighborhood').count()
```

[41]:
```
newcastle_venues_grouped.sort_values(by='Venue',ascending=False)
```

[41]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Newcastle	40	40	40	40	40	40
Hamilton	38	38	38	38	38	38
Cooks Hill	28	28	28	28	28	28
The Hill	25	25	25	25	25	25
Mayfield	22	22	22	22	22	22
Wickham	21	21	21	21	21	21
Newcastle West	19	19	19	19	19	19
Jesmond	18	18	18	18	18	18
Newcastle East	17	17	17	17	17	17
The Junction	17	17	17	17	17	17
Waratah	14	14	14	14	14	14
Islington	10	10	10	10	10	10

One Hot Encoding

What is One Hot Encoding?
One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction.

Venue Category is used to perform One Hot Encoding

```
# one hot encoding
newcastle_onehot = pd.get_dummies(newcastle_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
newcastle_onehot['Neighborhood'] = newcastle_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [newcastle_onehot.columns[-1]] + list(newcastle_onehot.columns[:-1])
newcastle_onehot = newcastle_onehot[fixed_columns]

newcastle_venues_grouped = newcastle_onehot.groupby('Neighborhood').mean().reset_index()
newcastle_venues_grouped
```

[33]:

	Neighborhood	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	Auto Dealership	Automotive Shop	Bakery	Bar	Baseball Field	Basketball Court	Beach	Bed & Breakfast	B Gar
0	Adamstown	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.000	0.000	0.0000	0.142
1	Adamstown Heights	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.000	0.000	0.0000	0.000
2	Bar Beach	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.000	0.200	0.0000	0.000
3	Black Hill	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.000	0.000	0.0000	0.000
4	Broadmeadow	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.125	0.000	0.0000	0.000
5	Callaghan	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.111111	0.222222	0.000000	0.000	0.000	0.0000	0.000
6	Carrington	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.000	0.000	0.0000	0.000
7	Cooks Hill	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.068966	0.000000	0.000	0.000	0.0000	0.000
8	Elermore Vale	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.000	0.000	0.0000	0.000
9	Fletcher	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.000000	0.000000	0.000000	0.000	0.000	0.0000	0.000
10	Georgetown	0.000000	0.000000	0.000000	0.00	0.0000	0.000000	0.0	0.200000	0.000000	0.000000	0.000	0.000	0.0000	0.000
11	Hamilton	0.000000	0.000000	0.052632	0.00	0.0000	0.000000	0.0	0.026316	0.026316	0.000000	0.000	0.000	0.0000	0.000

Determining the top 10 Venues in each Neighborhood

[35]:
```
num_top_venues = 10

for hood in newcastle_venues_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = newcastle_venues_grouped[newcastle_venues_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 3})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Adamstown----
                  venue   freq
0           Pizza Place  0.286
1                   Pub  0.143
2           Beer Garden  0.143
3  Fast Food Restaurant  0.143
4     Convenience Store  0.143
5          Cupcake Shop  0.143
6   Japanese Restaurant  0.000
7                  Lake  0.000
8                  Pool  0.000
9                Plaza  0.000


----Adamstown Heights----
                         venue  freq
0                  Music Venue   1.0
1                  Art Gallery   0.0
2                         Pool   0.0
3                   Playground   0.0
4                  Pizza Place   0.0
5                     Pharmacy   0.0
6                    Pet Store   0.0
7                         Park   0.0
8  Paper / Office Supplies Store   0.0
9                  Noodle House   0.0
```

Analyzing the TOP 10 venues in each neighborhood

[37]:
```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

[38]:
```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = newcastle_venues_grouped['Neighborhood']

for ind in np.arange(newcastle_venues_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(newcastle_venues_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted
```

[38]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adamstown	Pizza Place	Fast Food Restaurant	Pub	Convenience Store	Cupcake Shop	Beer Garden	Gay Bar	Gas Station	College Cafeteria	Gym
1	Adamstown Heights	Music Venue	Wine Bar	College Bookstore	Convenience Store	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Drugstore	Falafel Restaurant
2	Bar Beach	Gym / Fitness Center	General Entertainment	Café	Beach	Fast Food Restaurant	Wine Bar	Flea Market	Cupcake Shop	Department Store	Dim Sum Restaurant
3	Black Hill	Furniture / Home Store	Wine Bar	Flea Market	Convenience Store	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Drugstore	Falafel Restaurant

D. OBSERVATIONS

Observing the Top 10 Venues in each of the neighborhoods, it has become very clear that Newcastle/ Newcastle East/Newcastle West is the right location for the New South Wales Office location. Neighboring areas have venues that will provide ample opportunities for a great residential and recreational, multi-cultural cuisine and outdoor living for the employees of the company.

1. Cooks Hill
2. Hamilton
3. Jesmond
4. Mayfield
5. The Hill

6. The Junction
7. Waratah
8. Wickham

Clustering for further Analysis

k-means Clustering (Number of clusters = 5)

[40]:
```
# set number of clusters
kclusters = 5

newcastle_grouped_clustering = newcastle_venues_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(newcastle_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

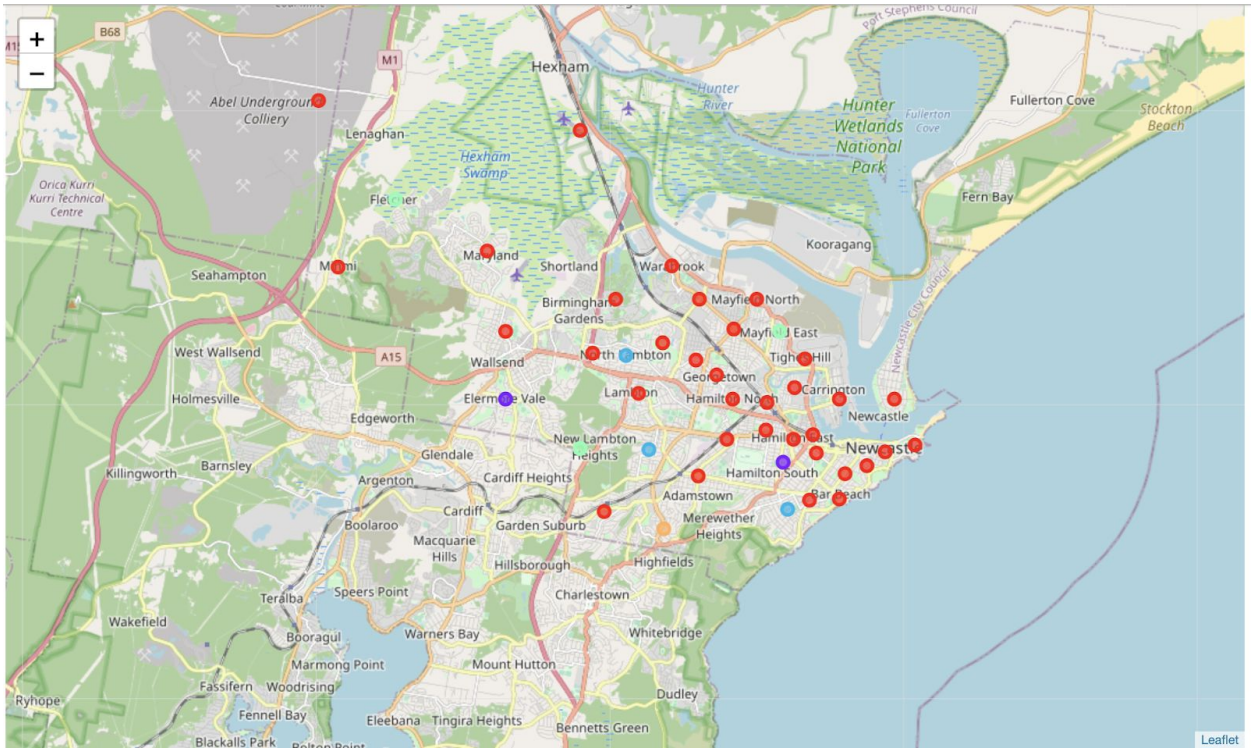
[40]:
```
array([0, 4, 0, 0, 0, 0, 0, 0, 1, 3], dtype=int32)
```

[41]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
```

[42]:
```
newcastle_merged = df_newcastle_neighborhood

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
newcastle_merged = newcastle_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

newcastle_merged # check the last columns!
newcastle_merged.dropna(axis=0,inplace=True)
```

Mapping the Clusters

```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=10)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(newcastle_merged['Lat'], newcastle_merged['Lng'], newcastle_merged['Neighborhood'], newcastle_merged['Clus
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [float(lat), float(lon)],
        radius=5,
        popup=label,
        color=rainbow[int(cluster)-1],
        fill=True,
        fill_color=rainbow[int(cluster)-1],
        fill_opacity=0.7).add_to(map_clusters)
map_clusters
```



E. Results

Analyzing each Cluster

Cluster 0

```
[51]: newcastle_merged.loc[newcastle_merged['Cluster Labels'] == 0, newcastle_merged.columns[[1] + list(range(5, newcastle_merged.shape[1]))]]
```

[51]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Newcastle	Jewelry Store	Trail	Café	Wine Bar	Fried Chicken Joint	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market
2	Newcastle	Park	Gym / Fitness Center	Café	Fast Food Restaurant	Beach	Wine Bar	Fried Chicken Joint	Dim Sum Restaurant	Donut Shop	Falafel Restaurant
5	Newcastle	Furniture / Home Store	Wine Bar	Fried Chicken Joint	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant
7	Newcastle	College Cafeteria	Coffee Shop	Bar	College Bookstore	Juice Bar	Bakery	Sandwich Place	Food & Drink Shop	Dim Sum Restaurant	Donut Shop
8	Newcastle	Café	Mexican Restaurant	Park	Thai Restaurant	Sports Bar	Wine Bar	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant
9	Newcastle	Café	Pub	Vietnamese Restaurant	Restaurant	Bar	Pizza Place	Fruit & Vegetable Store	Park	Coffee Shop	Falafel Restaurant
12	Newcastle	Ice Cream Shop	Park	Bakery	Sandwich Place	Fast Food Restaurant	Food & Drink Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant
13	Newcastle	Café	Pub	Mexican Restaurant	Falafel Restaurant	Italian Restaurant	Chinese Restaurant	Snack Place	Dim Sum Restaurant	Racetrack	Playground
15	Newcastle	Park	Pub	Grocery Store	Rental Car Location	Café	Sandwich Place	Fried Chicken Joint	Farmers Market	Wine Bar	Fish & Chips Shop
17	Newcastle	Trail	Salad Place	Fast Food Restaurant	Gas Station	Wine Bar	Food & Drink Shop	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop
18	Newcastle	Café	Bar	Vietnamese Restaurant	Restaurant	Art Gallery	Hotel Bar	Indian Restaurant	New American Restaurant	Park	Coffee Shop

Cluster 1

```
[52]: newcastle_merged.loc[newcastle_merged['Cluster Labels'] == 1, newcastle_merged.columns[[1] + list(range(5, newcastle_merged.shape[1]))]]
```

[52]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Newcastle	Park	Wine Bar	Fried Chicken Joint	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Fish & Chips Shop
50	Newcastle	Concert Hall	Park	Fried Chicken Joint	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Fish & Chips Shop

Cluster 2

```
]: newcastle_merged.loc[newcastle_merged['Cluster Labels'] == 2, newcastle_merged.columns[[1] + list(range(5, newcastle_merged.shape[1]))]]
```

]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	Newcastle	Sandwich Place	Wine Bar	Concert Hall	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant

Cluster 3

```
[56]: newcastle_merged.loc[newcastle_merged['Cluster Labels'] == 3, newcastle_merged.columns[[1] + list(range(5, newcastle_merged.shape[1]))]]
```

[56]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Newcastle	Pizza Place	Convenience Store	Cupcake Shop	Park	Pub	Fast Food Restaurant	Beer Garden	Flower Shop	Department Store	Dim Sum Restaurant
6	Newcastle	Food & Drink Shop	Train Station	Pub	Tennis Court	Bus Station	Fast Food Restaurant	Flower Shop	Wine Bar	Cupcake Shop	Department Store
14	Newcastle	Pub	Furniture / Home Store	Arts & Crafts Store	Bus Station	Train Station	Gas Station	Thrift / Vintage Store	Auto Dealership	Sporting Goods Shop	Fast Food Restaurant
24	Newcastle	Pub	Convenience Store	Park	Coffee Shop	Café	Music Venue	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop
26	Newcastle	Pub	Grocery Store	Shopping Mall	Pizza Place	Wine Bar	Flea Market	Convenience Store	Cupcake Shop	Department Store	Dim Sum Restaurant
27	Newcastle	Pub	Gym	Bakery	Liquor Store	Convenience Store	Wine Bar	Food & Drink Shop	Department Store	Dim Sum Restaurant	Donut Shop
31	Newcastle	Hotel	Pub	Burger Joint	Motel	Wine Bar	Flower Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant
32	Newcastle	Pub	Italian Restaurant	Café	Grocery Store	Sports Bar	Wine Bar	Food & Drink Shop	Department Store	Dim Sum Restaurant	Donut Shop
35	Newcastle	Pub	Park	Grocery Store	Wine Bar	Flower Shop	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant
40	Newcastle	Pub	Wine Bar	Concert Hall	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant
45	Newcastle	Pub	Fish & Chips Shop	Music Venue	Wine Bar	Concert Hall	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant

```
[58]: #### Cluster 4
```

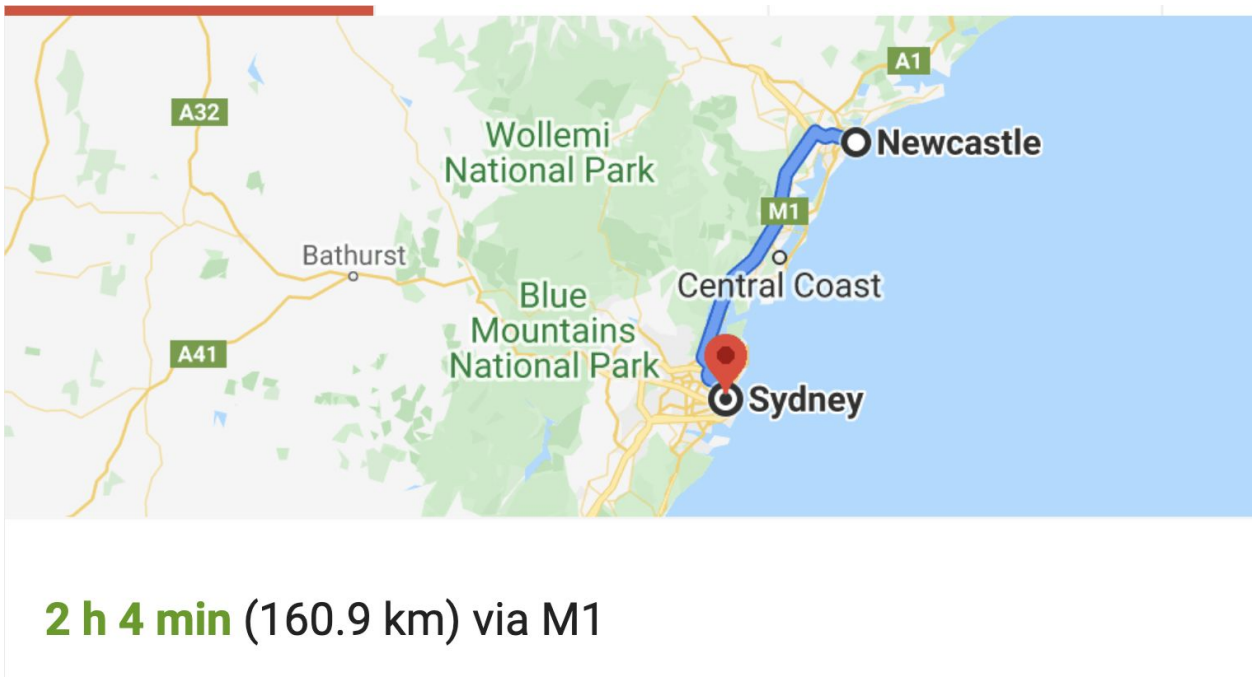
```
[59]: newcastle_merged.loc[newcastle_merged['Cluster Labels'] == 4, newcastle_merged.columns[[1] + list(range(5, newcastle_merged.shape[1]))]]
```

[59]:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
29	Newcastle	Home Service	Fast Food Restaurant	Wine Bar	Fried Chicken Joint	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market
36	Newcastle	Hostel	Home Service	Wine Bar	Food & Drink Shop	Cupcake Shop	Department Store	Dim Sum Restaurant	Donut Shop	Falafel Restaurant	Farmers Market

F. Conclusion

The study and analysis has provided significant details into Newcastle neighborhoods (Newcastle / Newcastle East / Newcastle West) and conclusively determined that Newcastle with its diverse and multicultural venues is undoubtedly the best location for the company operations and best positioned to attract and retain the best talent for the company and also establish its presence near the capital city of Sydney.



Sydney is 161 Km (100 miles) away from New Castle that will provide a significant boost to the revenue impact for the company. The unique and niche IT Service operations of the company will be able to provide a valued presence and a significant profitable revenue generator for the company.

There were other significant positive impacts of this study. Newcastle is surrounded by neighborhood areas like Cooks Hill, Hamiltion, Jesmond, Mayfield, The Hill, The Junction, Waratah and Wickham that provides great

locations for residential needs and recreational entertainments and outdoor experience that will delight the employees.

G. References

1. Australia JSON
2. Newcastle Neighborhood Wiki
3. Foursquare Developer
4. Google Maps