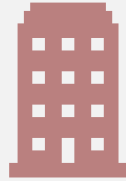# ANZ Assignment

Akshay Sharma

1. List of Customers

- Bank XYZ needs to choose the suppliers which are having contracts of high values.

- Data needs to be combined from the previous years to understand the consistent contractors/suppliers.

- Decision can be influenced by the duration of the contract.

Suppliers with respect to the contract worth in 2020

FMS ACCOUNT RESERVE BANK OF AUSTRALIA
2020 Q2

FMS ACCOUNT RESERVE BANK OF AUSTRALIA
2019 Q3

BROADSPECTRUM (AUSTRALIA) PTY LTD - GEMS PROJECT
2019 Q3

UNIVERSITY OF NSW ADFA UNSW@ADF
2019 Q4

JPO PROJECT PAMMANDI CBASS EMBASSAY
2019 Q4

RAYTHEON AUSTRALIA PTY LTD
2019 Q4

LENDLEASE BUILDING PTY LIMITED
2020 Q4

PMA-262 TRITON DPS CO-OP PROGRAM
2020 Q2

FMS ACCOUNT RESERVE BANK OF AUSTRALIA
2019 Q4

FMS ACCOUNT RESERVE BANK OF

SPOTLESS FACILITY SERVICES PTY LTD GEMS PROJECT
2019 Q3

BOEING DEFENCE AUSTRALIA LTD
2019 Q4

BOEING DEFENCE AUSTRALIA LTD
2020 Q2

DEPARTMENT OF DEFENSE OF THE UNITED STATES OF AMERICA
2019 Q3

LOCKHEED MARTIN AUSTRALIA PTY LTD
2019 Q1

CPB

CANSTRUCT

Clark

C S ENERGY LIMITED
2020 Q2

NAVAL SEA SYSTEMS COMMAND
2019 Q4

AUSTAL SHIPS PTY LTD
2020 Q2

Accenture Australia Pty Ltd
2019 Q3

MEGT

Pfizer

NAVAL GROUP 2020 Q1

RAYTHEON AUST PTY LTD
2019 Q3

QANTAS AIRWAYS LIMITED
2020 Q2

JONES LANG LASALLE (ACT) PTY LIMITED
2019 Q3

LEIDOS AUSTRALIA PTY

Deka

SERCO AUSTRALIA PTY LIMITED
2019 Q3

COBS TELSTRA
2019 Q4

THALES AUSTRALIA LIMITED
2020 Q1

PALADIN HOLDINGS PTE

Telstra Australia

LAING O'ROURKE AUSTRALIA CONSTRUCTION PTY LTD
2019 Q3

DMOJSF OFFICIAL AUSTRALIAN ACCOUNT PSFD
2019 Q4

CSL Behring (Australia) Pty Ltd
2020 Q1

NGJ COOPERATIVE

THALES AUSTRALIA

BROOKFIELD GLOBAL

2.a Optimum Facility Location

We can notice from data that most of the suppliers offices are located in Sydney, Melbourne & Canberra. So eventually office needs to be open on these particular locations.
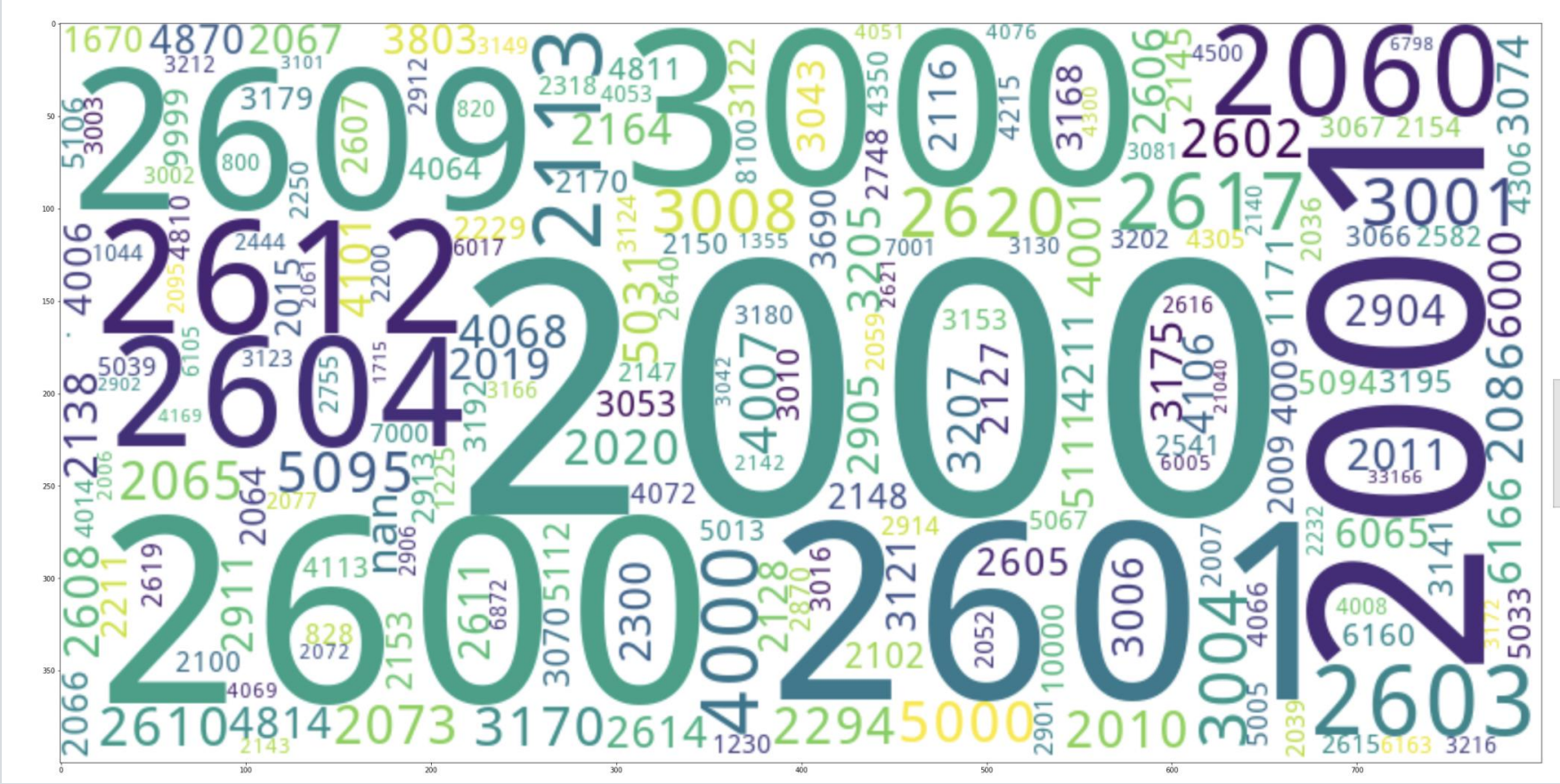
Secondly, Bank might already have major branches operating in these cities so we can collect the zips of the office location & zips of suppliers to find the optimum office location for day-to-day operations.
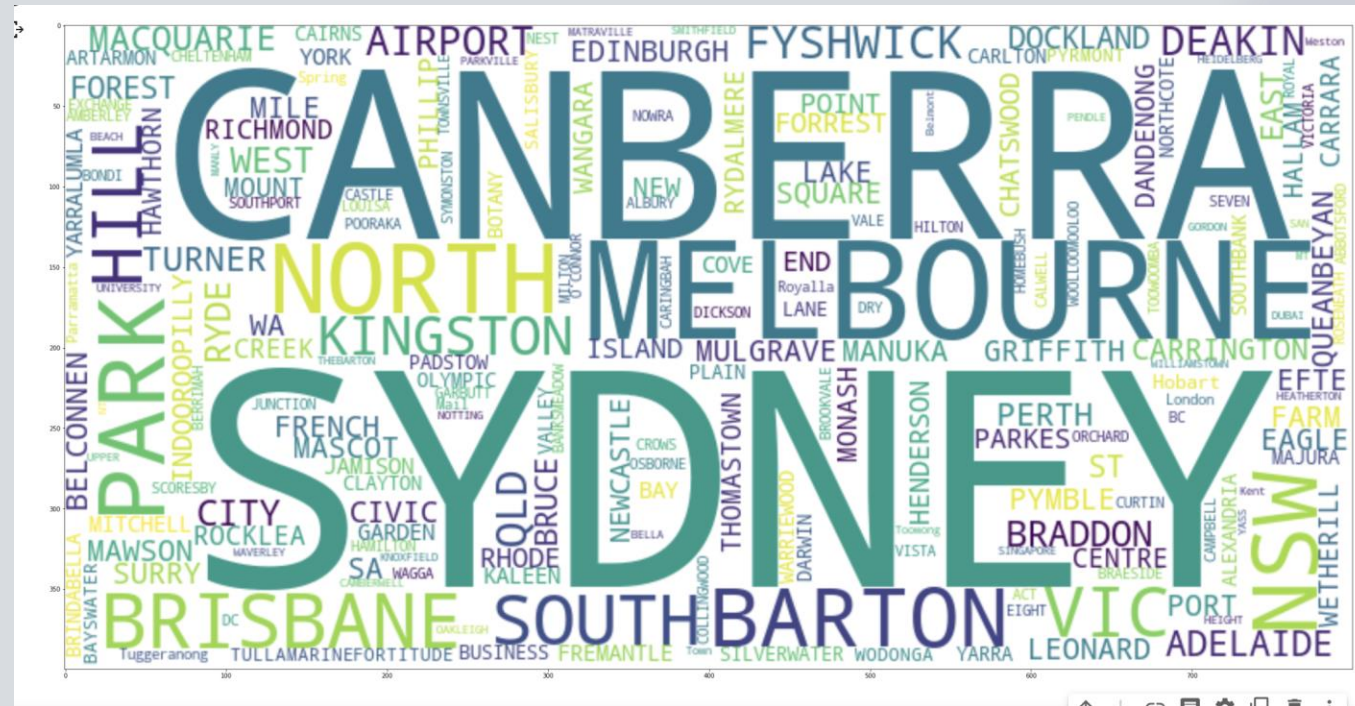
World Map for the optimum locations

Zip Codes containing maximum contracts

2b. Office Rent vs Revenue

- We need to create a metric which calculates the overall revenue generated by opening the office in city vs suburbs, if revenue from city is really high & can offset the rent cost then, we should open the offices in cities.

- If Rent is high & revenue generated is quite low then we can go for less popular location.

- Future Work: It can be solved using Facility Optimization Problem

3. Problem Statement

Categorizing clusters in order to focus on the clusters with high overall profits, so bank can target the right customers.

Proposed solution

- Based on data, clustering algorithm can create meaningful clusters & as we have location in the dataset we can create choropleth maps with the clusters predicted by our model & use them to visualizing the clusters of contracts & get the better understanding about opening the office at targetted cluster.

Machine learning algorithm

- KMEANS
- DBSCAN

Pros

- Easier to Implement.

- Highly scalable.

- Convergence guaranty .

- Centroid's warm-start.

- Adaptability to new instances.

- Generalizes to clusters of different shapes and sizes.

Cons

- **Choosing** k **manually.**

- **depends on initial values.**

- **Dragged by outliers which can be offset by using HDBSCAN.**

- **Curse of Dimensionality**

Sample Implementation

- As of now, we can see the clusters added in the DataFrame for further analysis & ready to be combined with Choropleth.

- After getting the clusters we can plot **geo sptial clusters** on the australia map using geopandas & folio libraries, it would result into the cluster colors with the map of australia but as of now geo JSON file is paid & I am trying to find a workaround it.

- Below is one of the sample examples, which was my inspiration for the geo spatial clustering. This crime pridiction using geo spatial clusters, we are doing the same as we have postcode & clusters defined(table on next page).

- Once we have the postcode & geo json we can join them & create a similar map for all the clusters to be visualized according to the post code.

DataFrame consisting the cluster information from KMeans & DBSCAN

- This data can be combined with the shape file to be fed for geo spatial analysis.

plicable nd Date	Applicable Value	Duration Years	Procurement Method	Panel Arrangement	Confidentiality Contract Flag	Confidentiality Outputs Flag	Consultancy Flag	Supplier Suburb	Supplier Postcode	Supplier State	Supplier Country	Supplier ABN Exempt	Office Postcode	predicted clusters	predicted clusters by Kmeans
2019-05-31	16163.40	0.222222	Limited tender	No	No	No	No	BEENLEIGH	4207	QLD	AUSTRALIA	No	2021	-1	4
2023-06-30	648000.00	3.097222	Open tender	Yes	No	No	No	Pennant Hills	1715	NSW	AUSTRALIA	No	2617	-1	2
2020-03-27	155997.60	0.027778	Open tender	Yes	No	No	No	QVB Post Office Sydney	1230	NSW	AUSTRALIA	No	2900	-1	4
2020-06-30	169542.12	0.530556	Open tender	Yes	No	No	No	NORTH SYDNEY	2060	NSW	AUSTRALIA	No	2606	-1	4
2019-07-31	13263.55	0.016667	Limited tender	No	No	No	No	SYDNEY	2000	NSW	AUSTRALIA	No	4000	-1	5

Future Work

Recognize Optimum Location by treating them as Facility Optimization Problem.

Geo Spatial clustering completion.

References

- https://towardsdatascience.com/visualizing-data-at-the-zip-code-level-with-folium-d07ac983db20

- https://www.datacamp.com/community/tutorials/wordcloud-python