Indicators of unemployment

36103 - STATISITCAL THINKING FOR DATA SCIENCE

Assessment Task 2 – Part A, Project Proposal

28 April 2019

Group Name – The Magnificent 7

Htet Naing Aung 13385531, Susannah Gynther 95059489, Robert Kell , William Kent 13285337, James Tesoriero , Reasmey Tith 10845345, Xiaojun Zeng 13331145.

Project Proposal

Indicators of unemployment

Rationale

The Australian Bureau of Statistics (ABS) uses the internationally agreed standards in defining unemployment. To be classified as unemployed a person needs to meet the following three criteria (Abs.gov.au, 2019):

- Not working more than one hour in the reference week;
- Actively looking for work in previous four weeks; and
- Be available to start work in the reference week.

Gleeson (2019) states that unemployment impacts on the economic, social and mental health of not only the person who is unemployed but their family and community in the short-term and can have an impact for decades to come. Additionally, the longer a person remains unemployed the more difficult it can become to find employment, as skills and abilities deteriorate over time. Hudson (2019) found that unemployment can cause a ripple effect across the economy. As the proportion of unemployed persons increases, less tax is collected, and government spending will rise accordingly to pay more in unemployment benefits, potentially affecting the ongoing financial stability of the economy.

Research has revealed several misnomers about unemployment and resulting factors such as crime and domestic violence. Janko and Popli (2015) in their analysis of Canadian data showed that there was no relationship between unemployment and crime. Another study showed that gender-based unemployment played a part in the increase of domestic violence although did not increase domestic violence over all (Anderberg et al, 2013).

Our research aims to broaden the scope for factors that could affect unemployment in NSW. The unemployment rate in Australia for March 2019 is 5.0% (Tradingeconomics.com, 2019), while in NSW it was 4.3% for the same period (Taffa, 2019). To maintain NSW's low unemployment and resulting high prosperity, detailed and region specific research can potentially reveal important characteristics regarding each area and how they impact unemployment rates. The outcomes of our research are targeted towards governmental policy makers and social welfare groups in NSW. Any new information can be used to assess existing services and their effectiveness, as well as highlight new areas where more services are needed.

Our questions

What factors predict unemployment rates in New South Wales? Of these factors, are there any that are unique or unexpected? What social demographics are related to unemployment? Does location affect unemployment and if so what is it about that location that contributes to unemployment?

Data Sources

This analysis will bring together a range of data sources and information covering geographical, educational and biographical data. Data will predominantly be obtained from Australian State and Federal agencies and departments.

Unemployment figures will be obtained from the ABS Census of Population and Housing conducted in 2016 (Census 2016). This research will further explore data from the Census 2016 and a range of datasets from other areas of the ABS and non-ABS sources summarised in Table 1.

Table 1 – Sources of Data

Indicator/variable	Description of dataset	Source	Geographic Level
Unemployment	Census of Population and Housing	ABS	SA2
Socio economic status	Socio- economic indexes for areas (SEIFA) 2016	ABS	SA2
Crime Statistics	Annual incident counts, rates per 100,000 population and ranks for selected offences (2011-2018)	Bureau of Crime and Statistics research (BOSCAR), NSW Department of Justice	LGA
Education Level	Census of Population and Housing: Reflecting Australia - Stories from the Census, 2016	ABS	SA2
Age and Gender	Estimated Resident Population (ERP) by SA2 (ASGS 2016) Age and Sex, 2001 Onwards	ABS	SA2
Dwelling type	Census 2016, T24 Dwelling Structure by Dwelling Type	ABS	SA2
Household composition	Census 2016, T23 Household Composition by Number of Persons Usually Resident	ABS	SA2
Air quality readings	Site Air Quality Index	NSW Office of Environment and Heritage	Reading station
Family Status	Marriages and divorces, Australia, 2016	ABS	SA2
Commute to work	Census of Population and Housing: Commuting to Work - More Stories from the Census, 2016	ABS	SA2
Access to Green Space	NSW Mesh blocks ASGS Edition 2016	ABS	Mesh Block

This project will collate data at the Australian Statistical Geographic Standard (ASGS) Statistical Area 2 (SA2) level. An SA2 has an average population of 10,000 people and can include one or more related suburbs that interact socially and economically (Abs.gov.au, 2018).

Using data at the SA2 level will allow analysis of over 570 distinct geographical areas in NSW. This will allow interpretation of any trends found to answer the research question. Data from the year 2016 was chosen as it was the year with the most data available. Coding examples of how we have acquired and merged our data are included in the appendices.

Modelling

An individual that wants to work is either employed or unemployed; this is a binary outcome. For an SA2 the unemployment rate is a proportion of the population that is unemployed and seeking employment compared with the total number of people employed or seeking employment. This unemployment rate is considered grouped data as individual's data is aggregated to form one observation per SA2. As part of this project, a multivariate logistic regression on grouped data will be performed in order to help answer the research question.

Issues

Issues currently experienced and expected include:

- **Diverse datasets** – Datasets have been gathered in a variety of formats making merging more difficult.
- **Level of Granularity** – Data is not always held at the SA2 level. It may be captured at a more granular level and will need to be aggregated, or data may be captured at larger geographical areas and need to be broken down to the SA2 level.
- **Data inconsistencies** – Data from government agencies is often input by humans with different levels of understanding and standards with can lead to data inconsistencies.
- **Data Reliability** – For some predictors the data will rely heavily on the ABS Census 2016. The census was completed by individuals who may provide false or non-sensical answers due to not understanding the question, systems and staff failed to interpret an answer, or an individual did not want to provide an answer due to data privacy concerns.

Summary

This research aims to provide new information about the factors affecting unemployment in NSW. It is hoped that this additional information will aid those making decisions that impact the provision of support and services for the unemployed, as well as better help those involved in improving employment prospects.

References

Abs.gov.au. (2019). *6105.0 - Australian Labour Market Statistics, July 2014*. [online] Available at: http://www.abs.gov.au/ausstats/abs@.nsf/products/FBE517ECA9B07F63CA257D0E001AC7D4 [Accessed 26 Apr. 2019].

Crowe, L. and Butterworth, P. (2016). The role of financial hardship, mastery and social support in the association between employment status and depression: results from an Australian longitudinal cohort study. *BMJ Open*, [online] 6(5), p.e009834. Available at: https://bmjopen.bmj.com/content/6/5/e009834?utm_source=trendmd&utm_medium=cpc&utm_campaign=jnis&trendmd-shared=1&utm_term=TrendMDPhase4&utm_content=Journalcontent.

Anderberg, D., Rainer, H., Wadsworth, J. and Wilson, T. (2013). *Unemployment and Domestic Violence: Theory and Evidence (Discussion Paper 7515)*. [online] Ftp.iza.org. Available at: http://ftp.iza.org/dp7515.pdf [Accessed 26 Apr. 2019].

Gleeson PhD, P. (2019). *The Overall Effects of Unemployment*. [online] Smallbusiness.chron.com. Available at: https://smallbusiness.chron.com/overall-effects-unemployment-37104.html [Accessed 26 Apr. 2019].

Hudson, P. (2019). *How Unemployment Rates Affect The Economy*. [online] Elite Daily. Available at: https://www.elitedaily.com/news/business/how-unemployment-rates-affect-the-economy [Accessed 26 Apr. 2019].

Taffa, V. (2019). *New South Wales Unemployment Rate At 4.3 % March 2019 | The Southern Thunderer*. [online] Southernthunderer.com.au. Available at: https://www.southernthunderer.com.au/new-south-wales-unemployment-rate-at-4-3-march-2019/ [Accessed 26 Apr. 2019].

Tradingeconomics.com. (2019). *Australia Unemployment Rate | 2019 | Data | Chart | Calendar | Forecast*. [online] Available at: https://tradingeconomics.com/australia/unemployment-rate [Accessed 26 Apr. 2019].

Abs.gov.au (2018). *Australian Statistical Geography Standard (ASGS),* July 2018.[online] Available at: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS) [Accessed 24 Apr. 2019]

Hanrahan, C. (2017), *Census results are out, but can we trust the data?,* [online] ABC News, Available at: https://www.abc.net.au/news/2017-06-26/census-results-are-coming-out-can-we-trust-them/8594132 [Accessed 27 April 2019]

Appendices

Appendix 1 - Example of Australian Bureau of Statistics Census 2016 data extract R code

```
library(rsdmx)
library(tidyverse)

# getwd()
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

# Check to make sure the ABS folder is available
# and, if not, create it. Saving file to right
# location will fail without the required folder
if (!dir.exists("../Data Files/ABS")) {
create.dir("../Data Files/ABS")
}
# Get the ABS Census 2016 Data on Dwelling Type
dwelling_data <- as.data.frame(readSDMX(providerId = "ABS",
resource = "data", flowRef = "ABS_C16_T24_SA",
key =
"TOT.TOT+11+21+22+31+32+33+34+91+92+93+94+Z+NA.0+1+2+3+4+5+6+7+8+9.SA2",
key.mode = "SDMX", start = 2016, end = 2016))
summary(dwelling_data)
head(dwelling_data)
str(dwelling_data)

# MISSING 9 SA2 Codes
dwelling_data %>% distinct(ASGS_2016)

# Distinct dimension values
dwelling_data %>% distinct(DWTD_2016)
## Retrieve Metadata to help with decoding values.
ds_url =
"http://stat.data.abs.gov.au/restsdmx/sdmx.ashx/GetDataStructure/ABS_C16_T
24_SA"
dataStructure <- readSDMX(ds_url)
codeList <- slot(dataStructure, "codelists")

# Dwelling Type
dwelling_type <- as.data.frame(codeList, codelistId =
"CL_ABS_C16_T24_SA_STRD_2016")

# Get Required Data and put in meaningful
# descriptions
dwelling_data_final <- dwelling_data %>% inner_join(dwelling_type,
by = c(STRD_2016 = "id")) %>% select(SA2_CODE = ASGS_2016,
DWELLING_TYPE = label.en, obsValue)

# getwd()
write_csv(dwelling_data_final, "../Data
Files/ABS/Dwelling_Type_SA2_2016.csv")
```

Appendix 2 - Example of NSW Government Air Quality data download using R

```
# getwd()
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
if (!dir.exists("../Data Files/NSWGovt/")) {
dir.create("../Data Files/NSWGovt/")
}

## Download NSW Air Quality File if it doesn't
## already exist
if (!file.exists("../Data Files/NSWGovt/AirQuality_Data.xls")) {
aq =
"https://airquality.environment.nsw.gov.au/aquisnetnswphp/tmp/tmp_table_21
553_1555911469.xls"
download.file(aq, destfile = "../Data Files/NSWGovt/AirQuality_Data.xls",
mode = "wb")
}

## Download NSW Air Quality Stations if it doesn't
## already exist
if (!file.exists("../Data Files/NSWGovt/AirQuality_Station_Data.xlsx")) {
stations = paste0("https://datasets.seed.nsw.gov.au/dataset/",
"ee5fd225-ab54-49c4-8c91-930219018cd0/resource/",
"e09a1918-af2b-4375-ad04-00fabce72a10/download/",
"air-quality-monitoring-sites-summary.xlsx")
download.file(stations, destfile = "../Data
Files/NSWGovt/AirQuality_Stations_Data.xlsx",
mode = "wb")
}
```

Appendix 3 - Example of Australian Bureau of Statistics Socio-Economic Indexes for Areas data extract using R

```
library(rsdmx)

# getwd()
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
if (!dir.exists("../Data Files/ABS/")) {
dir.create("../Data Files/ABS/")
}

data <- as.data.frame(readSDMX(providerId = "ABS",
resource = "data", flowRef = "ABS_SEIFA2016_SA2",
key.mode = "SDMX", start = 2016, end = 2016))
write.csv(data, "../Data Files/ABS/SEIFA_2016_Data.csv")
```

Appendix 4 - Example of Australian Bureau of Statistics Census 2016 LGA data cleaning using R

```
library(readxl)
library(tidyverse)

# getwd()
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

raw_data <- read_excel("../../Raw Data/Data
Files/NSWGovt/LgaRankings_27_offences.xlsx",
    sheet = "Assault - domestic violence",
    range = "A6:P137")

# Get Local Government Area to Mesh Block
# Data
lga_data_raw <- read_csv("../../Raw Data/Data
Files/ABS/Mesh_Blocks/LGA_2016_NSW.csv")

# Data Cleansing required - remove (A), (C), (NSW) etc and rename
# some merged/changed LGA's
lga_data <- lga_data_raw %>% filter(LGA_NAME_2016 !=
    "No usual address (NSW)") %>% mutate(LGA = gsub("*\\([A-Z]+\\)",
    "", LGA_NAME_2016)) %>% select(LGA, MB_CODE = MB_CODE_2016) %>%
    mutate(LGA = replace(LGA, LGA == "Botany Bay"|
        LGA == "Rockdale", "Bayside")) %>%
    mutate(LGA = replace(LGA, LGA == "Western Plains Regional",
        "Dubbo Regional")) %>% mutate(LGA = replace(LGA,
    LGA == "Unincorporated NSW", "Unincorporated Far West")) %>%
    mutate(LGA = replace(LGA, LGA == "Gundagai",
        "Cootamundra-Gundagai"))

# Get Mesh Block to SA2 Data
mb_data <- read_csv("../../Raw Data/Data
Files/ABS/Mesh_Blocks/MB_2016_NSW.csv")

clean_data <- raw_data %>% select(LGA = `Local Government Area`,
    Total_2016 = Total__2, Rate_Per_100k_2016 = `Rate per 100,000
population__2`,
    Rank_2016 = Rank__2) %>% filter(LGA !=
    "Total NSW*") %>% transform(Rate_Per_100k_2016 =
as.numeric(Rate_Per_100k_2016),
    Rank_2016 = as.integer(Rank_2016)) %>%
    replace_na(list(Total_2016 = 0, Rate_Per_100k_2016 = 0,
        Rank_2016 = 0)) %>% left_join(lga_data,
    by = c("LGA")) %>% left_join(mb_data,
    by = c("MB_CODE" = "MB_CODE_2016")) %>%
    distinct(LGA, SA2_CODE = SA2_MAINCODE_2016,
        Total_2016, Rate_Per_100k_2016, Rank_2016)

# Data Checks
clean_data %>% group_by(SA2_CODE) %>% summarise(cnt = n())

clean_data %>% filter(SA2_CODE == 101021011)

#Create output folder if it doesn't exist
if (!dir.exists("../Data Files/NSWGovt/")) {
    dir.create("../Data Files/NSWGovt/")
}

write.csv(clean_data, "../Data
Files/NSWGovt/Domestic_Violence_Offenders_LGA.csv",
    row.names = FALSE)
```

Appendix 5 - Example of Australian Bureau of Statistics Census 2016 data cleaning using R

```
library(tidyverse)
library(janitor)

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
erp <- read_csv("../../Raw Data/Data
Files/ABS/ERP/ABS_ERP_ASGS2016_25042019132433480.csv")
names(erp)

erp <- erp %>% select(-c("MEASURE", "Measure",
"SEX_ABS", "AGE", "FREQUENCY", "Frequency",
"TIME", "Flag Codes", "Flags", "REGIONTYPE",
"Geography Level"))

erp <- erp %>% rename(sa2_code = ASGS_2016) %>%
clean_names()

erp_by_sex <- erp %>% group_by(sa2_code,
sex) %>% summarise(total_value = sum(value)) %>%
spread(sex, total_value) %>% clean_names()

erp_by_age <- erp %>% group_by(sa2_code,
age) %>% summarise(total_value = sum(value)) %>%
spread(age, total_value) %>% clean_names()

erp <- erp_by_sex %>% left_join(erp_by_age)

write_csv(erp, "../Data Files/ABS/ERP_SA2_2016.csv")
```

Appendix 6 - Example of finding percent of SA2 area covered by parkland using R

```
library(tidyverse)

# getwd()
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

# Create destination folder if it doesn't
# already exist
if (!dir.exists("../Data Files/ABS/")) {
dir.create("../Data Files/ABS/")
}
# Read the raw data csv
mesh_blocks <- read.csv("../../Raw Data/Data
Files/ABS/Mesh_Blocks/MB_2016_NSW.csv")
str(mesh_blocks)
mesh_blocks %>% distinct(MB_CATEGORY_NAME_2016)

# Find % of space allocated to Parkland
# for each mesh block
open_space <- mesh_blocks %>% filter(STATE_NAME_2016 ==
"New South Wales") %>% select(MB_CODE_2016,
MB_CATEGORY_NAME_2016, SA2_CODE = SA2_MAINCODE_2016,
AREA_SQKM = AREA_ALBERS_SQKM) %>% group_by(SA2_CODE,
MB_CATEGORY_NAME_2016) %>% summarise(SUM_AREA_SQKM = sum(AREA_SQKM)) %>%
spread(MB_CATEGORY_NAME_2016, SUM_AREA_SQKM,
fill = 0) %>% mutate(PERC_OPEN_SPACE = Parkland/(Commercial +
Education + `Hospital/Medical` + Industrial +
MIGRATORY + NOUSUALRESIDENCE + OFFSHORE +
Other + Parkland + `Primary Production` +
Residential + SHIPPING + Transport +
Water))

# Write data to csv
write_csv(open_space, "../Data Files/ABS/Open_Space_SA2.csv")
```

Appendix 7 - Example of Australian Bureau of Statistics Socio-Economic Indexes for Areas data cleaning using R

```
library(tidyverse)

# Set directory to my the location where
# this file is
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

# getwd()
seifa_data <- read_csv("../../Raw Data/Data
Files/ABS/SEIFA_2016_Data.csv")

# Review data - AsGS_2016 field is an INT
head(seifa_data)

# Summarise the data - obsValue has 48
# NA's but non of these have a
# SEIFA_MEASURE == SCORE
summary(seifa_data)
seifa_nas <- seifa_data %>% filter(is.na(obsValue) == TRUE)

# Create a clean data set for use - only
# want high level scores and remove
# records with an obsValue of NA
clean_seifa_data <- seifa_data %>% filter(SEIFA_MEASURE ==
"SCORE") %>% select(SA2_CODE = ASGS_2016,
SEIFAINDEXTYPE, obsValue) %>% spread(SEIFAINDEXTYPE,
obsValue) %>% select(SA2_CODE, SEIFA_Edu_Occ_Index = IEO,
SEIFA_Economic_Res_Index = IER, SEIFA_Rel_SocioEco_Adv_Disadv_Index =
IRSAD,
SEIFA_Rel_SocioEco_Disadv_Index = IRSD)

# Write Clean Data to disk
write_csv(clean_seifa_data, "../Data Files/ABS/SEIFA_2016_Data.csv")

# Check for duplicates - nope none, only
# 1 record per SA2
clean_seifa_data %>% group_by(SA2_CODE) %>%
mutate(total = n()) %>% filter(total >
1)
```

Appendix 8 - Example of Exploratory Data Analysis of Australian Bureau of Statistics Census 2016 data

```
library(tidyverse)
library(Hmisc)
library(corrplot)

# getwd()
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

# read cleaned data set
dwelling_type <- read_csv("../Data Files/ABS/Dwelling_Type_SA2.csv")
mesh_blocks <- read_csv("../../Raw Data/Data
Files/ABS/Mesh_Blocks/MB_2016_NSW.csv")
str(mesh_blocks)

# Get mesh block data at SA2 level
sa2_data <- mesh_blocks %>% distinct(SA2_MAINCODE_2016,
SA2_NAME_2016, STATE_CODE_2016, STATE_NAME_2016)

# No duplicate SA2 Codes
sa2_data %>% group_by(SA2_MAINCODE_2016) %>% summarise(cnt = n()) %>%
filter(cnt > 1)

# Some SA2's don't have any dwellings - positive
# skew
dwelling_type %>% mutate(TOTAL = DWELLING_HOUSE + DWELLING_FLAT +
DWELLING_SEMI + DWELLING_OTHER) %>% ggplot() +
geom_histogram(aes(x = TOTAL), bins = 50)

# 62 SA2's have no dwellings - 7 in NSW, a military
# base, centennial park, a NP, a cemetry, and
# Industrial area, Banksmeadow is whaves and
# industry
dwelling_type %>% mutate(TOTAL = DWELLING_HOUSE + DWELLING_FLAT +
DWELLING_SEMI + DWELLING_OTHER) %>% filter(TOTAL ==
0) %>% left_join(sa2_data, by = c(SA2_CODE = "SA2_MAINCODE_2016")) %>%
select(SA2_CODE, TOTAL, SA2_NAME_2016, STATE_NAME_2016) %>%
filter(between(SA2_CODE, 1e+08, 2e+08))

# There are a couple of areas with high numbers of
# dwellings - Waterloo/Beaconsfield in NSW is high
# density
dwelling_type %>% mutate(TOTAL = DWELLING_HOUSE + DWELLING_FLAT +
DWELLING_SEMI + DWELLING_OTHER) %>% filter(TOTAL >
15000) %>% left_join(sa2_data, by = c(SA2_CODE = "SA2_MAINCODE_2016")) %>%
select(SA2_CODE, TOTAL, SA2_NAME_2016, STATE_NAME_2016) %>%
filter(between(SA2_CODE, 1e+08, 2e+08))

# Remove SA2's with no dwellings and only show NSW
# SA2's
dwelling_type_filtered <- dwelling_type %>% mutate(TOTAL = DWELLING_HOUSE
+
DWELLING_FLAT + DWELLING_SEMI + DWELLING_OTHER) %>%
filter(TOTAL != 0) %>% inner_join(sa2_data, by = c(SA2_CODE =
"SA2_MAINCODE_2016"))

## DWELLING HOUSE Some areas in NSW have no houses -
## data may be slightly skewed
dwelling_type_filtered %>% ggplot() + geom_histogram(aes(x =
DWELLING_HOUSE),
bins = 50)

# Standardise data and confirm data has a long tail
house_std <- scale(dwelling_type_filtered$DWELLING_HOUSE)
qqnorm(house_std)
abline(a = 0, b = 1, col = "grey")

## DWELLING FLAT There are 21 no flat SA2's in NSW -
## industrial areas, offshore shipping, Rural areas
```

```
dwelling_type_filtered %>% filter(DWELLING_FLAT ==
0) %>% select(SA2_CODE, SA2_NAME_2016, PERC_DWELLING_HOUSE,
PERC_DWELLING_FLAT, PERC_DWELLING_SEMI, PERC_DWELLING_OTHER)

# Some areas in NSW have no flats - data skewed
dwelling_type_filtered %>% ggplot() + geom_histogram(aes(x =
DWELLING_FLAT),
bins = 100)

# Standardise data and confirm data is not normally
# distributed
flat_std <- scale(dwelling_type_filtered$DWELLING_FLAT)
qqnorm(flat_std)
abline(a = 0, b = 1, col = "grey")

## DWELLING SEMI There are 19 no semi SA2's in NSW -
## industrial areas, airport, offshore shipping,
## rural areas
dwelling_type_filtered %>% filter(DWELLING_SEMI ==
0) %>% select(SA2_CODE, SA2_NAME_2016, PERC_DWELLING_HOUSE,
PERC_DWELLING_FLAT, PERC_DWELLING_SEMI, PERC_DWELLING_OTHER)

# Some areas in NSW have no semis - data skewed
dwelling_type_filtered %>% ggplot() + geom_histogram(aes(x =
DWELLING_SEMI),
bins = 100)

# Standardise data and confirm data is not normally
# distributed
semi_std <- scale(dwelling_type_filtered$DWELLING_SEMI)
qqnorm(semi_std)
abline(a = 0, b = 1, col = "grey")

# Check correlation between variables
dwelling_matrix <- dwelling_type_filtered %>% select(DWELLING_FLAT,
DWELLING_HOUSE, DWELLING_OTHER, DWELLING_SEMI) %>%
as.matrix()

# Show values - nothing really high
rcorr(dwelling_matrix, type = "pearson")

# And a plot for good measure
corrplot(cor(dwelling_matrix), method = "ellipse")
```

Appendix 9 - Example of merging cleaned datasets for modelling using R

```
library(tidyverse)

# getwd()
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

# Read csv's
dwelling_type <- read_csv("../../Clean Data/Data
Files/ABS/Dwelling_Type_SA2.csv")
hh_composition <- read_csv("../../Clean Data/Data
Files/ABS/HouseHold_Composition_SA2.csv")
place_of_birth <- read_csv("../../Clean Data/Data
Files/ABS/Place_Of_Birth_SA2.csv")
seifa <- read_csv("../../Clean Data/Data Files/ABS/SEIFA_2016_Data.csv")
mesh_blocks <- read_csv("../../Raw Data/Data
Files/ABS/Mesh_Blocks/MB_2016_NSW.csv")

# Get mesh block data at SA2 level
sa2_data <- mesh_blocks %>% distinct(SA2_MAINCODE_2016,
SA2_NAME_2016, STATE_CODE_2016, STATE_NAME_2016)

# Join Datasets together
model_data <- dwelling_type %>% inner_join(hh_composition,
by = c("SA2_CODE")) %>% inner_join(place_of_birth,
by = c("SA2_CODE")) %>% inner_join(seifa, by = c("SA2_CODE")) %>%
semi_join(sa2_data, by = c(SA2_CODE = "SA2_MAINCODE_2016")) %>%
select(-starts_with("PERC_"))
```