## **Analysis Questions**

### Selected Cost of Living Index Dataset

1. Proportion of expenditure spent on healthcare from Jun 2007 to Sept 2023, compared to other goods and services.
2. Is there an emerging trend in healthcare index?
3. Determine yoy change from 2007 to 2023 in healthcare index? Is the percentage change increased or decreased?
4. What household groups have the highest expenditure in health and impacted the most?

Assumptions:

1. Australians are spending large proportion on healthcare in 2022-2023 than before.
2. Between 2021 and 2023, proportion of change in healthcare index yoy has increased significantly.

### Census Datasets

1. Average / median personal and household income by state and SAL3
2. Average / median personal and household income by gender by state and SAL3
3. Average / median personal and household income by age
4. Check any increase in trends of median income by state, gender, age
5. Calculate % cost of out of pocket. Has that increased (expected yes)

Assumptions:

1. There will be a percentage increase out of pocket cost between 2014 to 2022
2. Females have a lower median income than male, possibly are paying higher cost of pocket-> requiring more services and lower median incomes than males.

### MBS Datasets

## **Data Cleaning & Transformation**

### Selected Cost of Living Index Dataset

1. Convert into CSV for importing.
2. Calculate missing value percentage.
3. Calculate summary statistics of index
4. Remove unwanted rows 1-10 (*summary of column)*
5. Remove data for years 1988 to 1999. Data observations missing for individual goods and services across all household groups.

### Census Datasets

1. 2011 census:
   1. combine income, age sex datasets
   2. combine SAL3 2011 edition with combined census dataset using SAL3 code
   3. find codes that exists in census but don’t exist in SAL3
   4. find codes that exists in SAL3 but not in census (requires investigation)
2. 2016 census:
   1. combine income, age sex datasets
   2. combine SAL3 2016 edition with combined census dataset using SAL3 code
   3. find codes that exists in census but don’t exist in SAL3
   4. find codes that exists in SAL3 but not in census (requires investigation)
3. 2021 census:
   1. combine income, age sex datasets
   2. combine SAL3 edition 3 with combined census dataset using SAL3 code
   3. find codes that exists in census but don’t exist in SAL3
   4. find codes that exists in SAL3 but not in census (requires investigation)

### MBS Datasets

1. MBS 2013-19, 2019-2021, 2021-2022
   1. Remove columns Name, Data source and Code
   2. Remove rows where worksheet = Private Health Insurance as this is out of context
   3. Extract and save dataset where worksheet = PHN
      1. Convert the values corresponding to below data points to column:
         1. Percentage of people who had the service (%)
         2. Services per 100 people
         3. Medicare benefits per 100 people ($)
         4. No. of patients
         5. No. of services
         6. Total Medicare benefits paid ($)
         7. Total provider fees ($)
      2. Investigate State Territory blanks.
   4. Extract and save dataset where worksheet – SA3
      1. Convert the values corresponding to below data points to column:
         1. Percentage of people who had the service (%)
         2. Services per 100 people
         3. Medicare benefits per 100 people ($)
         4. No. of patients
         5. No. of services
         6. Total Medicare benefits paid ($)
         7. Total provider fees ($)
2. SAL3 Checks
   1. Compare SAL3, PHN list of 3 datasets to ensure they match correctly.
      1. Check for changed / missing SAL3
      2. Check for changed / missing PHN
3. Combine the 3 datasets for SAL3 and PHN into single datatset so it contains years 2013-2022.
4. Analysis of combined datasets
   1. Investigate Flag & Flag explanation #, n.p
   2. Calculate the summary statistics.

### Patient Experience

1. Extract “*Persons 15 years and over, Experience of GP services in the last 12 months by age and sex: Proportion of persons (%)*”and “*Persons 15 years and over, Experience of after hours GP care in the last 12 months by selected characteristics:: Proportion of persons (%)*” excel sheets for each year
   1. Segment data into age, remoteness and sex columns and extract corresponding data points
      1. Did not need to see a GP
      2. Needed to see a GP
      3. Needed to but did not see a GP at all
      4. Needed to and saw a GP
      5. At no time delayed seeing or did not see a GP when needed
      6. At least once delayed seeing or did not see a GP when needed
      7. At least once delayed seeing or did not see a GP when needed - cost a reason
      8. At least once delayed seeing or did not see a GP when needed - cost not a reason
2. Extract “*Persons 15 years and over, Experience of after hours GP care in the last 12 months by age and sex: Proportion of persons (%)”* and “*Persons 15 years and over, Experience of after hours GP care in the last 12 months by selected characteristics: Proportion of persons “*excel sheets for each year.
   1. Segment data into age, remoteness and sex columns and extract corresponding data points:
      1. Did not need to see an after hours GP
      2. Needed to see an after hours GP
      3. Needed to but did not see an after hours GP at all
      4. Needed to and saw an after hours GP
      5. Always saw an after hours GP when needed
      6. At least once did not see an after hours GP when needed
      7. At least once did not see an after hours GP when needed - cost main reason
      8. At least once did not see an after hours GP when needed - cost not main reason