

Data overview and pre-processing

```
In [ ]: import pandas as pd
data=pd.read_excel("results-for-retirement-at-2023-08-23-1405.xlsx")
```

```
In [ ]: data.head()
```

```
Out[ ]:
```

	1. What is your age?	2. What is your gender? (assigned gender at birth)	3. Social status:	4. Number of children:	5. What is your employment status?	6. At what age do you plan to retire?	7. On a scale of 1 (extremely uncomfortable) to 5 (very comfortable), how comfortable are you with the idea of retirement?	7.1. Extremely uncomfortable vs Very comfortable	8. How are you planning to fund your retirement?	9. How would you rate your knowledge about retirement planning in general?	10. How would you rate your familiarity with your country's social security or pension system?	11. What is your primary source of information for retirement planning?
0	25 - 34	Female	Couple	0	Employed full-time	56 - 60	NaN	5	Personal savings	Poor	Not at all familiar	Family friend
1	35 - 44	Male	Single	1 - 2	Employed full-time	65 - 70	NaN	4	Pension	Fair	Somewhat familiar	Family friend
2	25 - 34	Male	Couple	3 - 4	Self-employed/Business owner	Not planning to retire at all	NaN	2	Personal savings	Poor	Somewhat familiar	Family friend
3	45 - 54	Female	Couple	0	Employed full-time	65 - 70	NaN	2	Retirement account (e.g. IRA, 401k)	Poor	Not very familiar	Family friend
4	25 - 34	Female	Single	0	Employed full-time	61 - 65	NaN	3	Pension	Poor	Not at all familiar	I am not sure, I have not retired yet

```
In [ ]: short_column_names = {
    '1. What is your age?': 'Age',
    '2. What is your gender? (assigned gender at birth)': 'Gender',
    '3. Social status:': 'Social Status',
    '4. Number of children:': 'Num Children',
    '5. What is your employment status?': 'Employment Status',
    '6. At what age do you plan to retire?': 'Retirement Age',
    '7. On a scale of 1 (extremely uncomfortable) to 5 (very comfortable), how comfortable are you with the idea of retirement?': 'Comfort Score',
    '7.1. Extremely uncomfortable vs Very comfortable': 'Comfort Score',
    '8. How are you planning to fund your retirement?': 'Funding Plan',
    '9. How would you rate your knowledge about retirement planning in general?': 'Knowledge Level',
    '10. How would you rate your familiarity with your country's social security or pension system?': 'Familiarity',
    '11. What is your primary source of information for retirement planning?': 'Information Source',
    '12. What was the impact of COVID-19 pandemic on your retirement plans?': 'COVID Impact',
    '13. How did COVID affect your income/employment status?': 'COVID Effect on Income',
    '14. Has your confidence in your retirement plan changed due to COVID-19 pandemic?': 'COVID Confidence Change',
    '15. Has the COVID-19 pandemic influenced how you save for retirement?': 'COVID Influence on Saving',
    '16. What do you expect to do during retirement?': 'Retirement Expectations',
    '16.a. If you plan to continue working (full/part-time) or start your own business. Why is that?': 'Reason for Continuing Work',
    '17. Once retired, do you expect to have same standards of living compared to pre-retirement?': 'Living Standards',
    '18. Which of the following are you concerned about after retirement (you may select more than one)?': 'Retirement Concerns'
}

# Rename columns using the dictionary
data.rename(columns=short_column_names, inplace=True)
```

```
In [ ]: #Data types
print(data.info())
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37 entries, 0 to 36
Data columns (total 20 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   Age                                   37 non-null     object
 1   Gender                               37 non-null     object
 2   Social Status                         37 non-null     object
 3   Num Children                         37 non-null     object
 4   Employment Status                   37 non-null     object
 5   Retirement Age                      37 non-null     object
 6   Comfort Level                       0 non-null      float64
 7   Comfort Score                       37 non-null     int64
 8   Funding Plan                        37 non-null     object
 9   Knowledge Level                     37 non-null     object
10   Familiarity Level                   37 non-null     object
11   Information Source                  37 non-null     object
12   COVID Impact                       37 non-null     object
13   COVID Effect on Income              37 non-null     object
14   COVID Confidence Change             37 non-null     object
15   COVID Influence on Saving           37 non-null     object
16   Retirement Expectations             37 non-null     object
17   Reason for Continued Work           28 non-null     object
18   Living Standards after Retirement   37 non-null     object
19   Retirement Concerns                37 non-null     object
dtypes: float64(1), int64(1), object(18)
memory usage: 5.9+ KB
None

```

```
In [ ]: data.describe()# Get summary statistics of the numeric columns
```

```
Out[ ]:
```

	Comfort Level	Comfort Score
count	0.0	37.000000
mean	NaN	3.486486
std	NaN	1.282827
min	NaN	1.000000
25%	NaN	3.000000
50%	NaN	4.000000
75%	NaN	4.000000
max	NaN	5.000000

– Brief description of variables

Age: The age of the participant.

Gender: The gender of the participant, assigned at birth.

Social Status: The social status of the participant.

Number of Children: The number of children the participant has.

Employment Status: The participant's current employment status.

Retirement Age: The age at which the participant plans to retire.

Comfort with Retirement: The participant's comfort level with the idea of retirement, rated on a scale of 1 (extremely uncomfortable) to 5 (very comfortable).

Comfort with Retirement (Categorized): A categorical representation of the participant's comfort level with retirement.

Funding for Retirement: The participant's planned funding source for retirement.

Knowledge about Retirement Planning: The participant's self-rated knowledge about retirement planning.

Familiarity with Social Security/Pension System: The participant's familiarity with the country's social security or pension system.

Primary Information Source: The participant's primary source of information for retirement planning.

Impact of COVID-19 on Retirement Plans: How the COVID-19 pandemic impacted the participant's retirement plans.

Impact of COVID-19 on Income/Employment: How the COVID-19 pandemic impacted the participant's income and employment status.

Change in Confidence due to COVID-19: Whether the participant's confidence in their retirement plan changed due to the COVID-19 pandemic.

Change in Retirement Saving due to COVID-19: Whether the participant changed their retirement saving habits due to the COVID-19 pandemic.

Expected Activities During Retirement: The participant's expectations for activities during retirement.

Plan to Work During Retirement: Whether the participant plans to work (full/part-time) or start a business during retirement and the reason for it.

Expected Standard of Living: The participant's expectation of their standard of living in retirement compared to pre-retirement.

Retirement Concerns: Specific concerns the participant has after retirement, selected from a list of options.

Age Group: A binned representation of the participant's age into groups.

In []:

- Describe and justify data cleaning and pre-processing

Data Cleaning: Data cleaning is like fixing mistakes and empty spots in your dataset. It makes sure your data is accurate for analysis.

Justification: Cleaning data is important because wrong or missing data can make your analysis wrong, too.

Data Pre-processing: Data pre-processing is getting your data ready for analysis by changing and organizing it.

Justification: Pre-processing helps make sure your data is set up right for finding useful things in it.

– Handling of missing or erroneous data

In []: `data.isnull().sum()`

```
Out[ ]: Age                0
Gender                0
Social Status        0
Num Children         0
Employment Status    0
Retirement Age      0
Comfort Level        37
Comfort Score        0
Funding Plan         0
Knowledge Level      0
Familiarity Level    0
Information Source   0
COVID Impact        0
COVID Effect on Income 0
COVID Confidence Change 0
COVID Influence on Saving 0
Retirement Expectations 0
Reason for Continued Work 9
Living Standards after Retirement 0
Retirement Concerns 0
dtype: int64
```

In []: `data=data.drop(columns={"Comfort Level"},axis=1)`

In []: `import numpy as np`

Question 16a has NO answer if the previous question (i.e., 16) was answered as “Enjoy my life with family and friends.”

```
In [ ]: # Identify rows where Question 16 is answered as "Enjoy my life with family and friends"
rows_to_clean = data[data['Retirement Expectations'] == 'Enjoy my life with family and friends']

# Set Question 16a to empty for the identified rows
data.loc[rows_to_clean.index, 'Reason for Continued Work'] = 'NA'
```

In []: `data`

Out[]:

	Age	Gender	Social Status	Num Children	Employment Status	Retirement Age	Comfort Score	Funding Plan	Knowledge Level	Familiarity Level	Information Source	COVID Impact
0	25 - 34	Female	Couple	0	Employed full-time	56 - 60	5	Personal savings	Poor	Not at all familiar	Family and friends	It did no change my plans

1	35 - 44	Male	Single	1 - 2	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did no change my plans
2	25 - 34	Male	Couple	3 - 4	employed/Business owner	Not planning to retire at all	2	Personal savings	Poor	Somewhat familiar	Family and friends	It did no change my plans
3	45 - 54	Female	Couple	0	Employed full-time	65 - 70	2	Retirement account (e.g. IRA, 401k)	Poor	Not very familiar	Family and friends	It did no change my plans
4	25 - 34	Female	Single	0	Employed full-time	61 - 65	3	Pension	Poor	Not at all familiar	I am not informed about retirement planning	It did no change my plans
5	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	1	Pension	Fair	Somewhat familiar	Family and friends	It did no change my plans
6	35 - 44	Male	Couple	1 - 2	employed/Business owner	Not planning to retire at all	1	Personal savings	Fair	Not very familiar	Other	It did no change my plans
7	25 - 34	Male	Single	0	Employed full-time	Not planning to retire at all	1	Inheritance	Excellent	Somewhat familiar	Family and friends	It delayed my retirement
8	25 - 34	Female	Single	0	Employed full-time	55 or less	4	Retirement account (e.g. IRA, 401k)	Fair	Somewhat familiar	Employer	It did no change my plans
9	45 - 54	Male	Couple	0	Employed full-time	55 or less	3	Personal savings	Excellent	Somewhat familiar	Internet	It did no change my plans
10	35 - 44	Male	Couple	0	employed/Business owner	55 or less	4	Other	Fair	Somewhat familiar	Other	It did no change my plans
11	55 - 64	Male	Couple	0	Employed full-time	61 - 65	4	Personal savings	Good	Somewhat familiar	I am not informed about retirement planning	It did no change my plans
12	35 - 44	Male	Couple	1 - 2	employed/Business owner	56 - 60	5	Pension	Fair	Very familiar	Internet	It did no change my plans
13	35 - 44	Male	Couple	0	Employed full-time	61 - 65	1	Pension	Fair	Somewhat familiar	Internet	It delayed my retirement
14	25 - 34	Female	Couple	0	employed/Business owner	56 - 60	5	Other	Good	Not at all familiar	Family and friends	It did no change my plans
15	45 - 54	Female	Couple	1 - 2	Employed full-time	61 - 65	4	Pension	Fair	Somewhat familiar	Financial advisor	It did no change my plans
16	25 - 34	Male	Single	0	Employed full-time	55 or less	4	Personal savings	Fair	Somewhat familiar	Internet	It did no change my plans
17	35 - 44	Male	Couple	1 - 2	Employed full-time	55 or less	5	Personal savings	Fair	Not very familiar	Family and friends	It delayed my retirement
18	45 - 54	Male	Couple	3 - 4	employed/Business owner	65 - 70	3	Retirement account (e.g. IRA, 401k)	Good	Somewhat familiar	Financial advisor	It did no change my plans

19	45 - 54	Male	Couple	3 - 4	Employed full-time	61 - 65	4	Pension	Poor	Somewhat familiar	Employer	It delayed my retirement
20	45 - 54	Male	Couple	1 - 2	Employed full-time	After 70	3	Not planning to retire	Good	Somewhat familiar	Internet	It did not change my plans
21	35 - 44	Male	Couple	1 - 2	Employed full-time	55 or less	4	Personal savings	Poor	Not very familiar	Family and friends	I decided to retire earlier
22	35 - 44	Female	Couple	3 - 4	Employed full-time	55 or less	4	Other	Good	Somewhat familiar	Workshops/Seminars	I decided to retire earlier
23	18 - 24	Male	Single	0	Employed full-time	Not planning to retire at all	1	Not planning to retire	Poor	Not at all familiar	I am not informed about retirement planning	It did not change my plans
24	35 - 44	Male	Couple	3 - 4	Employed full-time	55 or less	5	Pension	Good	Very familiar	Employer	It did not change my plans
25	45 - 54	Male	Couple	3 - 4	Employed full-time	61 - 65	3	Pension	Poor	Not very familiar	Family and friends	It did not change my plans
26	35 - 44	Female	Single	0	Employed full-time	After 70	3	Other	Not sure	Not very familiar	Family and friends	It did not change my plans
27	35 - 44	Male	Couple	5 - 6	Employed full-time	61 - 65	4	Social security	Good	Somewhat familiar	Family and friends	It did not change my plans
28	55 - 64	Male	Couple	5 - 6	Retired	I am already retired	4	Pension	Excellent	Very familiar	Employer	I decided to retire earlier
29	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	3	Personal savings	Good	Not very familiar	Family and friends	It did not change my plans
30	45 - 54	Female	Single	0	Employed full-time	56 - 60	5	Pension	Good	Very familiar	Employer	It did not change my plans
31	45 - 54	Female	Couple	3 - 4	Employed full-time	55 or less	4	Pension	Poor	Not very familiar	Employer	It did not change my plans
32	55 - 64	Male	Couple	> 6	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did not change my plans
33	45 - 54	Male	Couple	3 - 4	Employed full-time	56 - 60	5	Personal savings	Good	Somewhat familiar	Financial advisor	It did not change my plans
34	45 - 54	Male	Couple	5 - 6	Employed full-time	56 - 60	5	Pension	Excellent	Very familiar	Internet	It did not change my plans
35	35 - 44	Male	Couple	1 - 2	Employed full-time	After 70	4	Retirement account (e.g. IRA, 401k)	Fair	Somewhat familiar	Financial advisor	It did not change my plans
36	55 - 64	Male	Couple	3 - 4	Employed full-time	61 - 65	3	Social security	Not sure	Not very familiar	Other	It did not change my plans

```
In [ ]: data = data.replace('missing_value', np.nan)
```

Data summary statistics

```
In [ ]: data
```

```
Out[ ]:
```

	Age	Gender	Social Status	Num Children	Employment Status	Retirement Age	Comfort Score	Funding Plan	Knowledge Level	Familiarity Level	Information Source	COVID Impact
0	25 - 34	Female	Couple	0	Employed full-time	56 - 60	5	Personal savings	Poor	Not at all familiar	Family and friends	It did not change my plans
1	35 - 44	Male	Single	1 - 2	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did not change my plans
2	25 - 34	Male	Couple	3 - 4	Self-employed/Business owner	Not planning to retire at all	2	Personal savings	Poor	Somewhat familiar	Family and friends	It did not change my plans
3	45 - 54	Female	Couple	0	Employed full-time	65 - 70	2	Retirement account (e.g. IRA, 401k)	Poor	Not very familiar	Family and friends	It did not change my plans
4	25 - 34	Female	Single	0	Employed full-time	61 - 65	3	Pension	Poor	Not at all familiar	I am not informed about retirement planning	It did not change my plans
5	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	1	Pension	Fair	Somewhat familiar	Family and friends	It did not change my plans
6	35 - 44	Male	Couple	1 - 2	Self-employed/Business owner	Not planning to retire at all	1	Personal savings	Fair	Not very familiar	Other	It did not change my plans
7	25 - 34	Male	Single	0	Employed full-time	Not planning to retire at all	1	Inheritance	Excellent	Somewhat familiar	Family and friends	It delayed my retirement
8	25 - 34	Female	Single	0	Employed full-time	55 or less	4	Retirement account (e.g. IRA, 401k)	Fair	Somewhat familiar	Employer	It did not change my plans
9	45 - 54	Male	Couple	0	Employed full-time	55 or less	3	Personal savings	Excellent	Somewhat familiar	Internet	It did not change my plans
10	35 - 44	Male	Couple	0	Self-employed/Business owner	55 or less	4	Other	Fair	Somewhat familiar	Other	It did not change my plans
11	55 - 64	Male	Couple	0	Employed full-time	61 - 65	4	Personal savings	Good	Somewhat familiar	I am not informed about retirement planning	It did not change my plans
12	35 - 44	Male	Couple	1 - 2	Self-employed/Business owner	56 - 60	5	Pension	Fair	Very familiar	Internet	It did not change my plans
13	35 - 44	Male	Couple	0	Employed full-time	61 - 65	1	Pension	Fair	Somewhat familiar	Internet	It delayed my retirement

14	25 - 34	Female	Couple	0	Self-employed/Business owner	56 - 60	5	Other	Good	Not at all familiar	Family and friends	It did not change my plans
15	45 - 54	Female	Couple	1 - 2	Employed full-time	61 - 65	4	Pension	Fair	Somewhat familiar	Financial advisor	It did not change my plans
16	25 - 34	Male	Single	0	Employed full-time	55 or less	4	Personal savings	Fair	Somewhat familiar	Internet	It did not change my plans
17	35 - 44	Male	Couple	1 - 2	Employed full-time	55 or less	5	Personal savings	Fair	Not very familiar	Family and friends	It delayed my retirement
18	45 - 54	Male	Couple	3 - 4	Self-employed/Business owner	65 - 70	3	Retirement account (e.g. IRA, 401k)	Good	Somewhat familiar	Financial advisor	It did not change my plans
19	45 - 54	Male	Couple	3 - 4	Employed full-time	61 - 65	4	Pension	Poor	Somewhat familiar	Employer	It delayed my retirement
20	45 - 54	Male	Couple	1 - 2	Employed full-time	After 70	3	Not planning to retire	Good	Somewhat familiar	Internet	It did not change my plans
21	35 - 44	Male	Couple	1 - 2	Employed full-time	55 or less	4	Personal savings	Poor	Not very familiar	Family and friends	I decided to retire earlier
22	35 - 44	Female	Couple	3 - 4	Employed full-time	55 or less	4	Other	Good	Somewhat familiar	Workshops/Seminars	I decided to retire earlier
23	18 - 24	Male	Single	0	Employed full-time	Not planning to retire at all	1	Not planning to retire	Poor	Not at all familiar	I am not informed about retirement planning	It did not change my plans
24	35 - 44	Male	Couple	3 - 4	Employed full-time	55 or less	5	Pension	Good	Very familiar	Employer	It did not change my plans
25	45 - 54	Male	Couple	3 - 4	Employed full-time	61 - 65	3	Pension	Poor	Not very familiar	Family and friends	It did not change my plans
26	35 - 44	Female	Single	0	Employed full-time	After 70	3	Other	Not sure	Not very familiar	Family and friends	It did not change my plans
27	35 - 44	Male	Couple	5 - 6	Employed full-time	61 - 65	4	Social security	Good	Somewhat familiar	Family and friends	It did not change my plans
28	55 - 64	Male	Couple	5 - 6	Retired	I am already retired	4	Pension	Excellent	Very familiar	Employer	I decided to retire earlier
29	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	3	Personal savings	Good	Not very familiar	Family and friends	It did not change my plans
30	45 - 54	Female	Single	0	Employed full-time	56 - 60	5	Pension	Good	Very familiar	Employer	It did not change my plans
31	45 - 54	Female	Couple	3 - 4	Employed full-time	55 or less	4	Pension	Poor	Not very familiar	Employer	It did not change my plans
32	55 - 64	Male	Couple	> 6	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did not change

33	45 - 54	Male	Couple	3 - 4	Employed full-time	56 - 60	5	Personal savings	Good	Somewhat familiar	Financial advisor	It did not change my plans
34	45 - 54	Male	Couple	5 - 6	Employed full-time	56 - 60	5	Pension	Excellent	Very familiar	Internet	It did not change my plans
35	35 - 44	Male	Couple	1 - 2	Employed full-time	After 70	4	Retirement account (e.g. IRA, 401k)	Fair	Somewhat familiar	Financial advisor	It did not change my plans
36	55 - 64	Male	Couple	3 - 4	Employed full-time	61 - 65	3	Social security	Not sure	Not very familiar	Other	It did not change my plans

```
In [ ]: # Get summary statistics for numerical columns
summary_stats = data.describe()

print(summary_stats)
```

```

      Comfort Score
count      37.000000
mean       3.486486
std        1.282827
min        1.000000
25%        3.000000
50%        4.000000
75%        4.000000
max        5.000000
```

Get the number of survey responses

```
In [ ]: num_responses = data.shape[0]

print("Number of survey responses:", num_responses)

Number of survey responses: 37
```

```
In [ ]:
```

Number of observations in external data

```
In [ ]: num_obs = data.shape[0]

print("Number of observations:", num_obs)

Number of observations: 37
```

Summary of demographics and key variables

```
In [ ]:
```

```
In [ ]: data.columns
```

```
Out[ ]: Index(['Age', 'Gender', 'Social Status', 'Num Children', 'Employment Status',
              'Retirement Age', 'Comfort Score', 'Funding Plan', 'Knowledge Level',
              'Familiarity Level', 'Information Source', 'COVID Impact',
              'COVID Effect on Income', 'COVID Confidence Change',
              'COVID Influence on Saving', 'Retirement Expectations',
              'Reason for Continued Work', 'Living Standards after Retirement',
              'Retirement Concerns'],
              dtype='object')
```

```
In [ ]: # Select columns related to demographics and key variables
selected_columns = [
    'Age',
    'Gender',
    'Social Status',
    'Num Children',
    'Employment Status',
```



```

    'Comfort Score',
    'Funding Plan'
]

selected_data = data[selected_columns]
numerical_summary = selected_data.describe()
categorical_columns = [
    'Gender',
    'Social Status',
    'Employment Status',
    'Comfort Score',
    'Funding Plan'
]

for column in categorical_columns:
    value_counts = selected_data[column].value_counts()
    print(f"\nValue counts for {column}:")
    print(value_counts)

```

```

Value counts for Gender:
Male      27
Female    10
Name: Gender, dtype: int64

```

```

Value counts for Social Status:
Couple     29
Single      8
Name: Social Status, dtype: int64

```

```

Value counts for Employment Status:
Employed full-time      30
Self-employed/Business owner    6
Retired                  1
Name: Employment Status, dtype: int64

```

```

Value counts for Comfort Score:
4      14
5       8
3       8
1       5
2       2
Name: Comfort Score, dtype: int64

```

```

Value counts for Funding Plan:
Pension      14
Personal savings    10
Retirement account (e.g. IRA, 401k)    4
Other         4
Not planning to retire    2
Social security    2
Inheritance       1
Name: Funding Plan, dtype: int64

```

In []:

Univariate and multivariate analysis

```

In [ ]: # Univariate analysis of the 'Age' variable
age_data = data['Age']

# Preprocess the age column to calculate average age for each range
data['Average Age'] = data['Age'].apply(lambda x: sum(map(int, x.split(' - '))) / 2)

average_age_data = data['Average Age']

# Calculate basic statistics
mean_average_age = average_age_data.mean()
median_average_age = average_age_data.median()
std_deviation_average_age = average_age_data.std()

# Generate a histogram
average_age_data.plot(kind='hist', bins=10, title='Average Age Distribution')

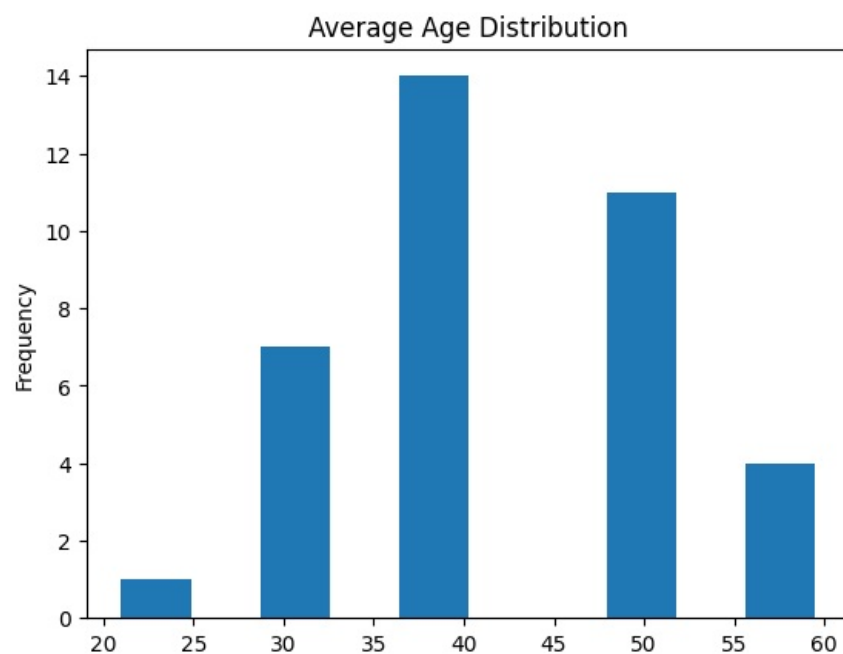
# Display basic statistics
print("Mean Average Age:", mean_average_age)
print("Median Average Age:", median_average_age)
print("Standard Deviation of Average Age:", std_deviation_average_age)

```

```

Mean Average Age: 42.24324324324324
Median Average Age: 39.5
Standard Deviation of Average Age: 9.809024771452409

```



Mean Average Age: 42.24

Median Average Age: 39.5

Standard Deviation of Average Age: 9.81

Mean Average Age represents the average age calculated from the midpoints of the age ranges. In this case, the average age of the respondents is approximately 42.24 years.

Median Average Age is the middle value when all the calculated average ages are arranged in ascending order. In this case, the median average age is 39.5 years.

Standard Deviation measures the dispersion or spread of the average ages around the mean. A higher standard deviation indicates greater variability in ages.

The histogram provides a visual representation of how the average ages are distributed across the dataset, showing how many respondents fall within each age range.

This analysis helps to understand the central tendency and variability of the average ages of the respondents in the dataset.

In []:

In []: `import matplotlib.pyplot as plt`

In []: `data.columns`

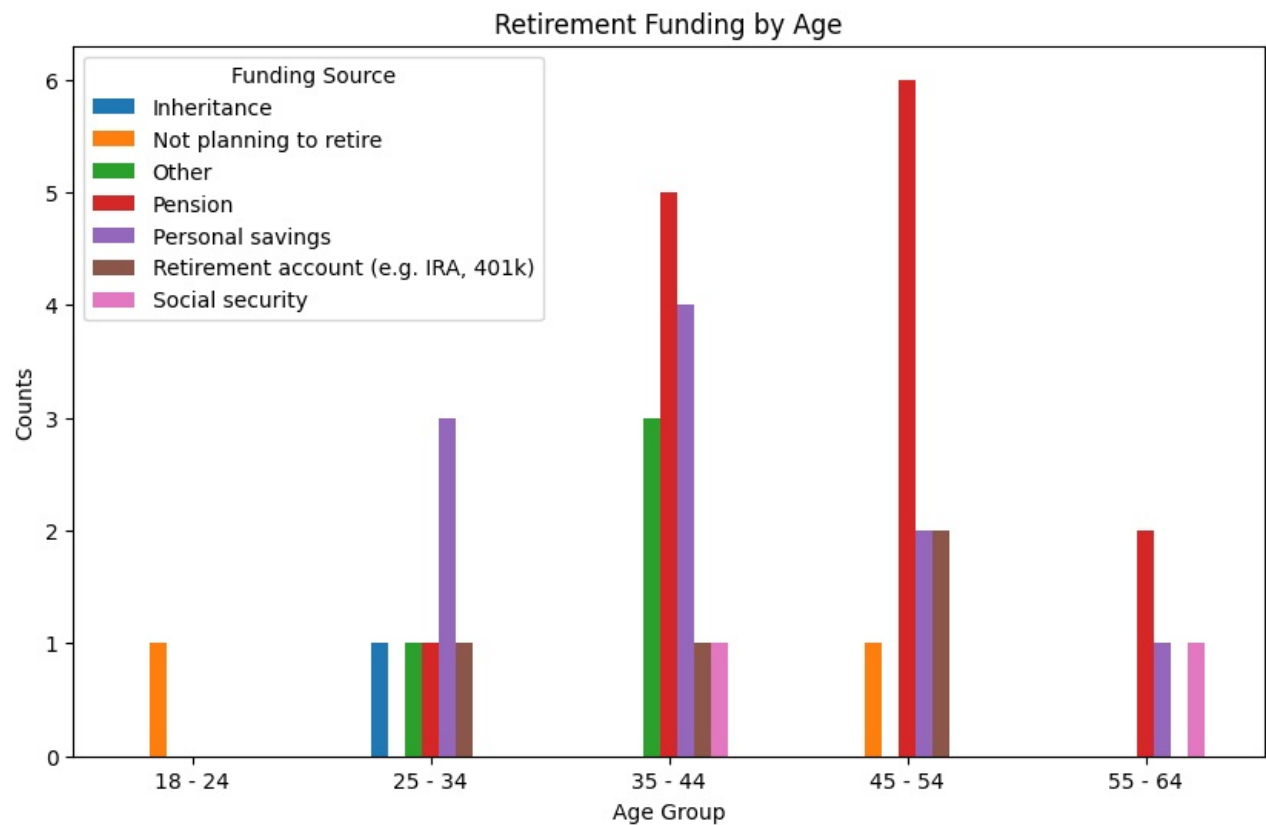
Out[]: Index(['Age', 'Gender', 'Social Status', 'Num Children', 'Employment Status', 'Retirement Age', 'Comfort Score', 'Funding Plan', 'Knowledge Level', 'Familiarity Level', 'Information Source', 'COVID Impact', 'COVID Effect on Income', 'COVID Confidence Change', 'COVID Influence on Saving', 'Retirement Expectations', 'Reason for Continued Work', 'Living Standards after Retirement', 'Retirement Concerns', 'Average Age'], dtype='object')

```
In [ ]: multivariate_data = data[['Age', 'Funding Plan']]

# Group the data by age and retirement funding and calculate counts
funding_by_age = multivariate_data.groupby(['Age', 'Funding Plan']).size().reset_index(name='Counts')

# Pivot the table for better visualization
pivoted_table = funding_by_age.pivot(index='Age', columns='Funding Plan', values='Counts')
pivoted_table.plot(kind='bar', figsize=(10, 6))
plt.title('Retirement Funding by Age')
plt.xlabel('Age Group')
plt.ylabel('Counts')
plt.xticks(rotation=0)
plt.legend(title='Funding Source')

plt.show()
```



- 1-The average amount of retirement funding increases with age.
- 2-People in their 50s and 60s have the most retirement funding.
- 3-Personal savings is the most common source of retirement funding.
- 4-Retirement accounts and pensions are also common sources of retirement funding.
- 5-It is important to start saving for retirement early in life.

In []:

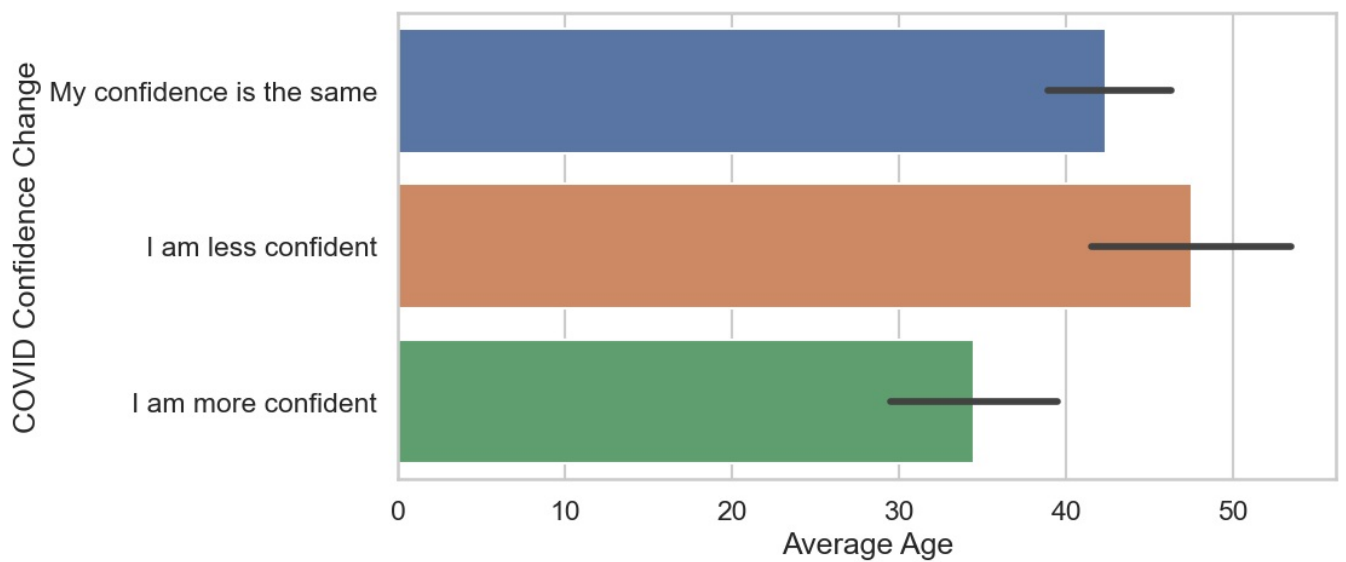
• Exploratory and explanatory visualisation

In []: `import seaborn as sns`

In []: `data.columns`

Out[]: Index(['Age', 'Gender', 'Social Status', 'Num Children', 'Employment Status', 'Retirement Age', 'Comfort Score', 'Funding Plan', 'Knowledge Level', 'Familiarity Level', 'Information Source', 'COVID Impact', 'COVID Effect on Income', 'COVID Confidence Change', 'COVID Influence on Saving', 'Retirement Expectations', 'Reason for Continued Work', 'Living Standards after Retirement', 'Retirement Concerns', 'Average Age'], dtype='object')

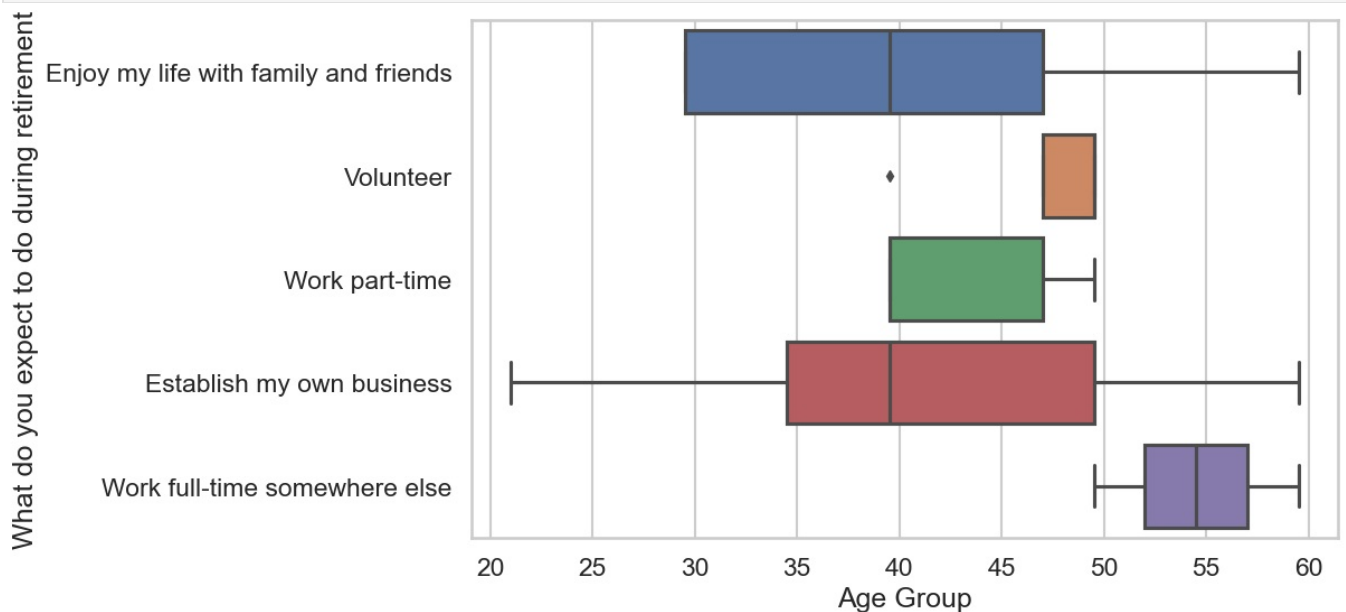
In []: `sns.set(style="whitegrid")`
`sns.set_context("talk")`
`plt.figure(figsize=(10, 5))`
`plot = sns.barplot(x='Average Age', y='COVID Confidence Change', data=data)`
`plt.show()`



The longest bar in the graph represents people who are “less confident”, followed by people whose “confidence is the same” and the shortest bar represents people who are “more confident”

```
In [ ]: plt.figure(figsize=(10, 6))
import seaborn as sns
sns.boxplot(x='Average Age', y='Retirement Expectations', data=data)

plt.xlabel('Age Group')
plt.ylabel('What do you expect to do during retirement')
plt.show()
```



The highest percentage of people expect to “Enjoy my life with family and friends” during retirement, followed by “Volunteer” and “Work part-time”. The lowest percentage of people expect to “Work full-time somewhere else” during retirement.

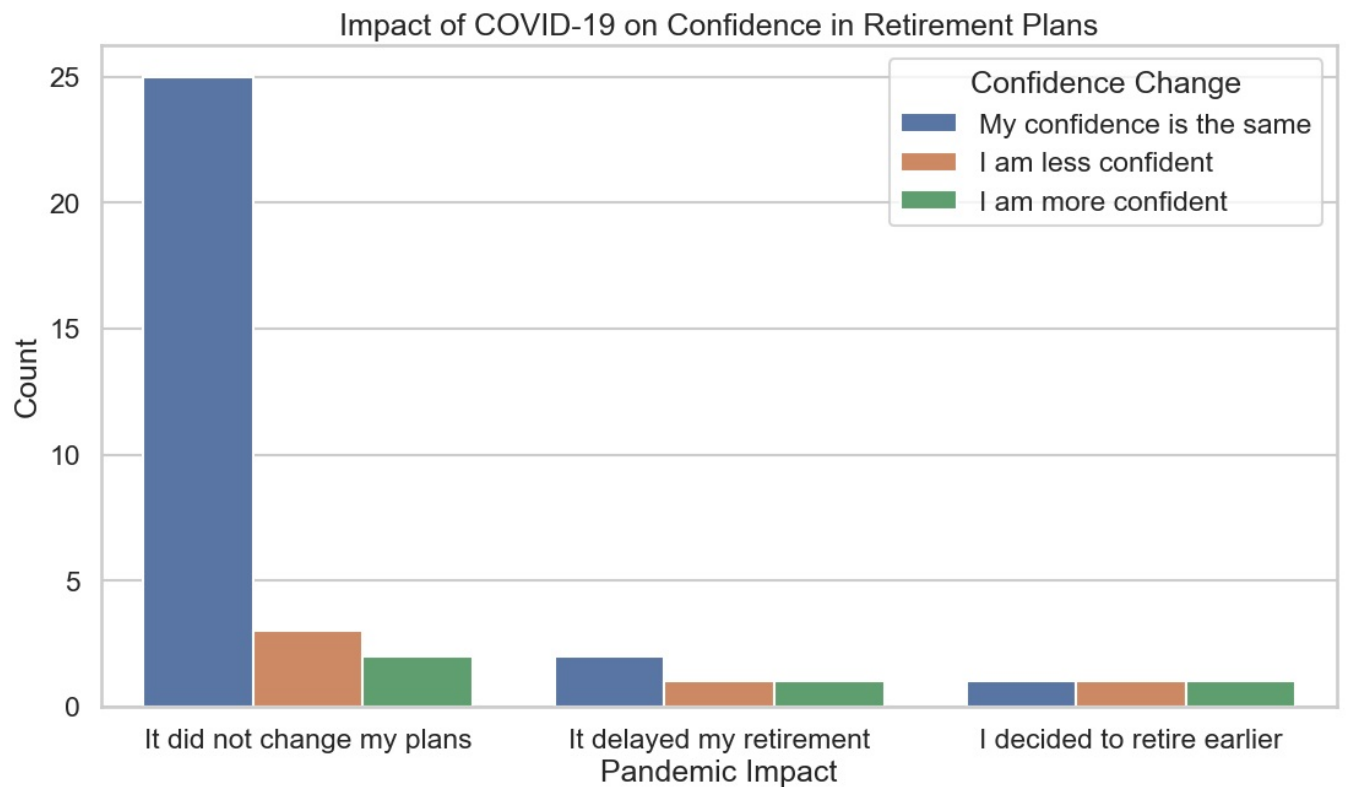
```
In [ ]:
```

Research Question: "How does the impact of the COVID-19 pandemic relate to changes in individuals' confidence in their retirement plans?"

```
In [ ]: sns.set(style="whitegrid")
sns.set_context("talk")
plt.figure(figsize=(13, 7))
sns.countplot(data=data, x='COVID Impact', hue='COVID Confidence Change')

plt.xlabel('Pandemic Impact')
plt.ylabel('Count')
```

```
plt.title('Impact of COVID-19 on Confidence in Retirement Plans')
plt.legend(title='Confidence Change')
plt.show()
```

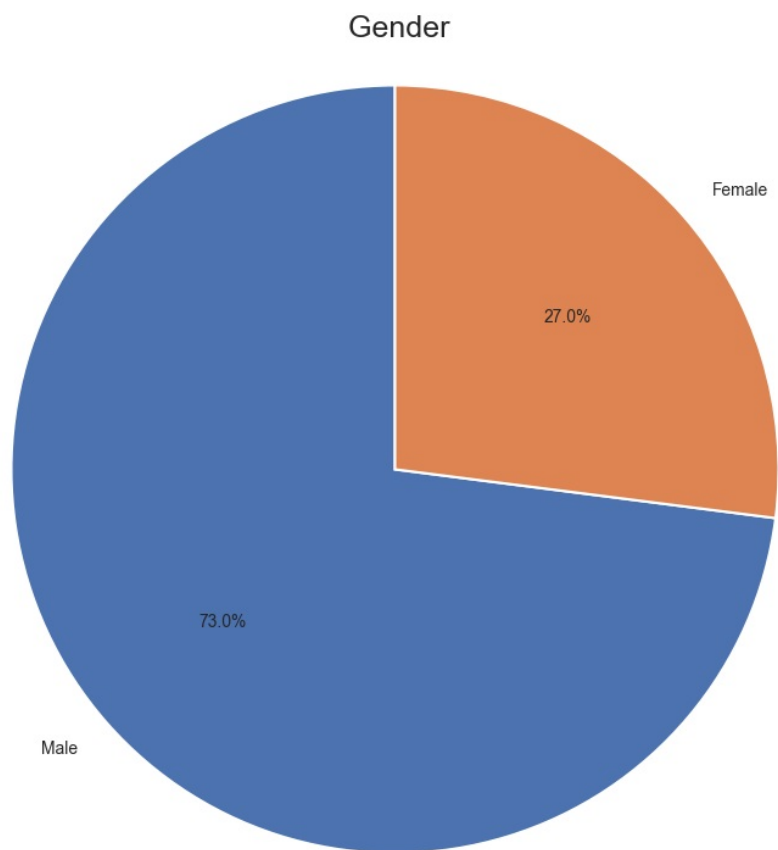
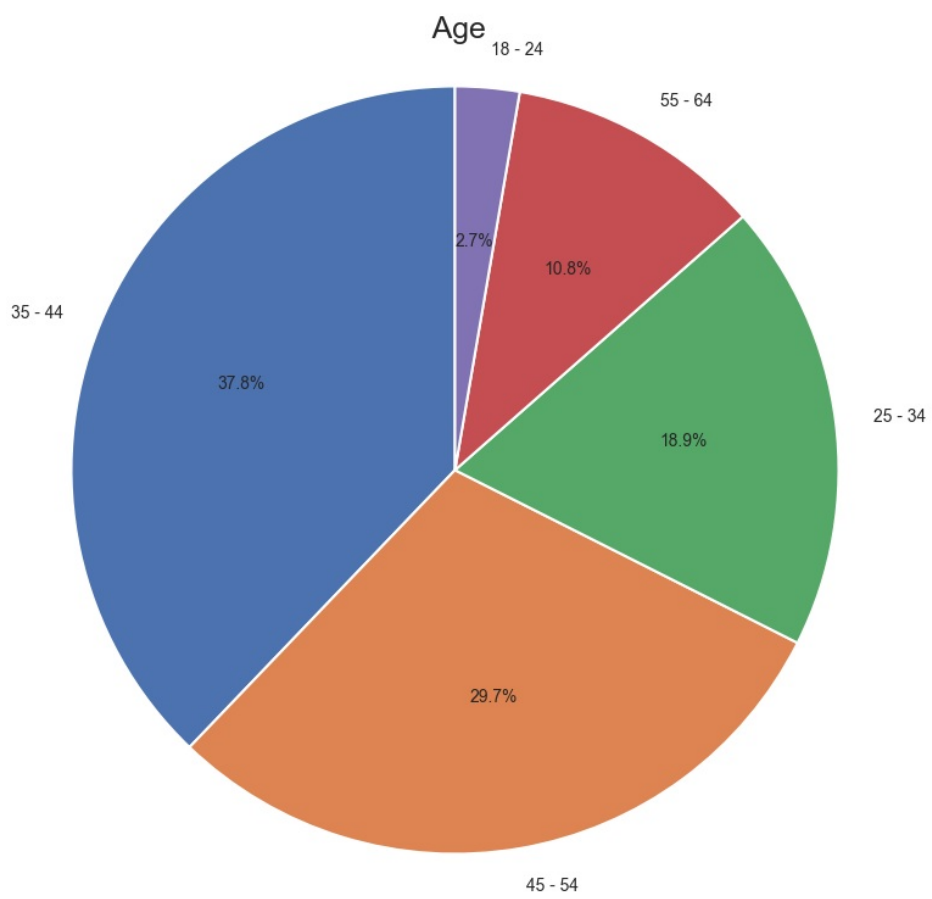


The bar graph titled “Impact of COVID-19 on Confidence in Retirement Plans” shows the results of a survey that asked people about the impact of the pandemic on their retirement plans and their confidence in those plans. The graph shows that for most people, the pandemic did not change their retirement plans, and their confidence in those plans remained the same. However, there were also people whose retirement plans were delayed by the pandemic, and they were less confident in their plans as a result. A small number of people decided to retire earlier due to the pandemic, and their confidence in their plans varied. Overall, the graph suggests that the pandemic had a mixed impact on people’s retirement plans and confidence in those plans

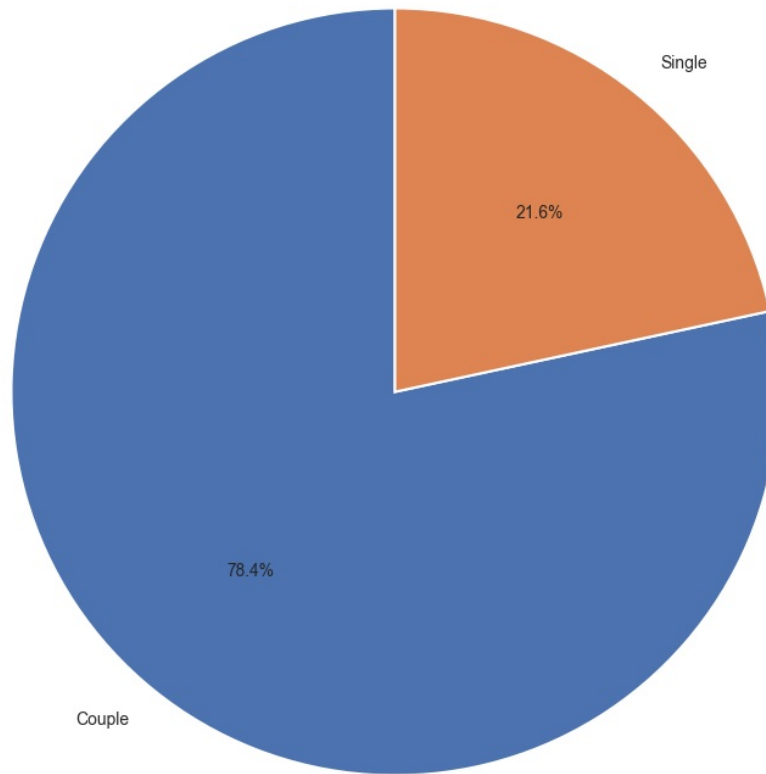
Visualise individual variables

```
In [ ]: for column in data.columns:
    try:
        plt.figure(figsize=(14, 9))

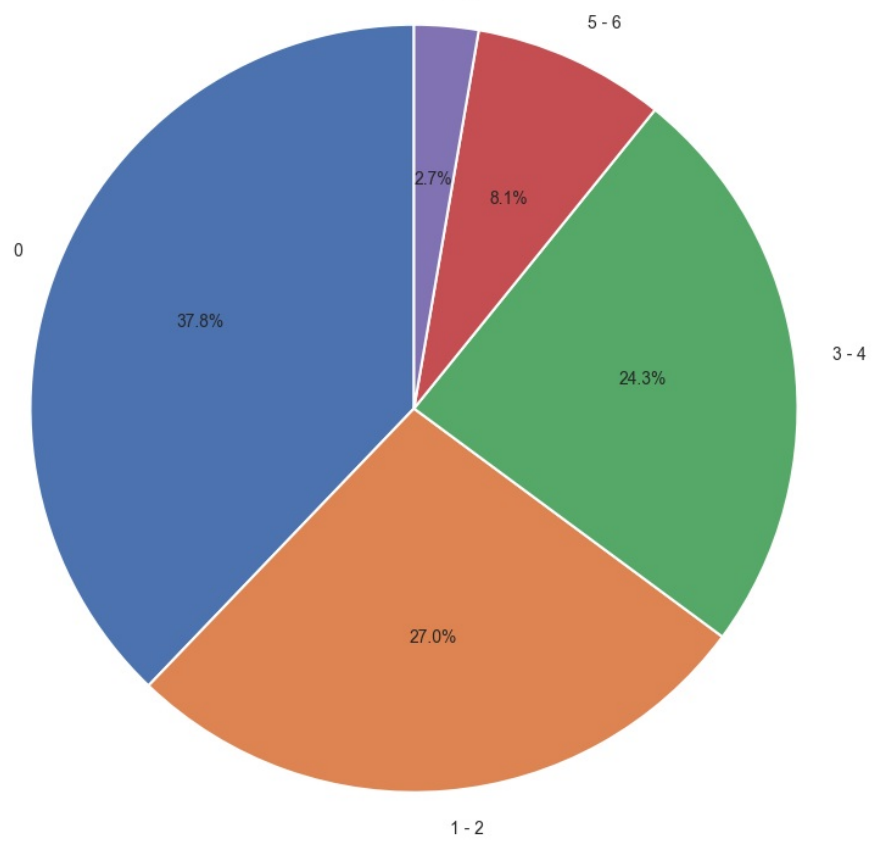
        # Create a new figure for each pie chart
        category_counts = data[column].value_counts()
        plt.pie(category_counts, labels=category_counts.index, autopct='%1.1f%%', startangle=90, textprops={'font':
        plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle
        plt.title(f' {column}')
        plt.show()
    except Exception as e:
        print(f"An error occurred: {str(e)}")
```



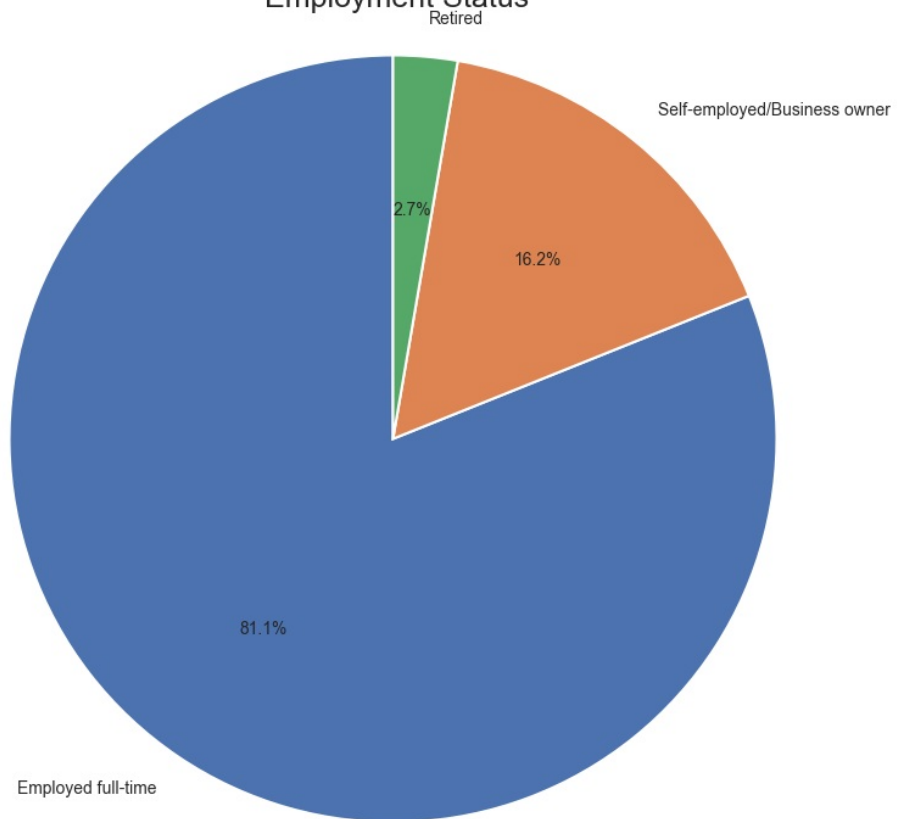
Social Status



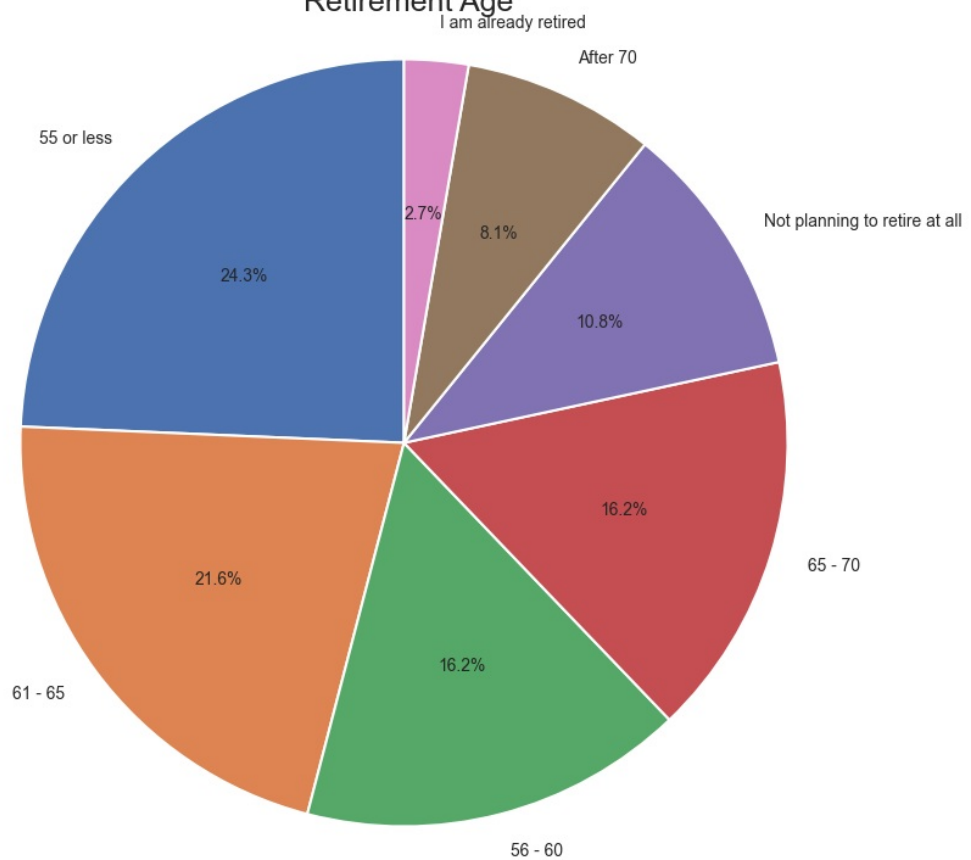
Num Children



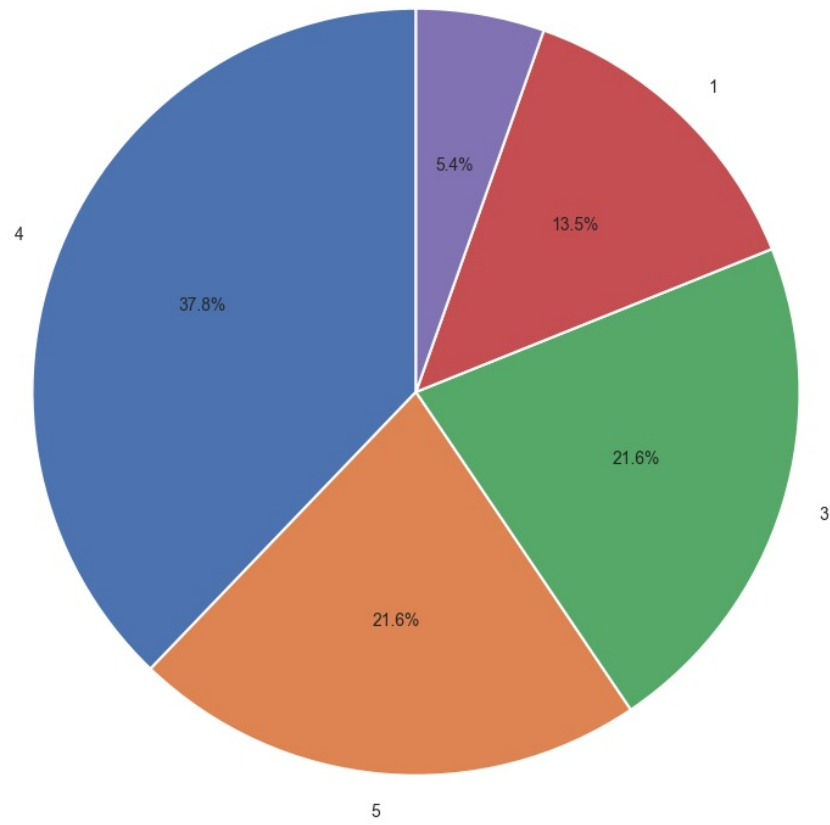
Employment Status



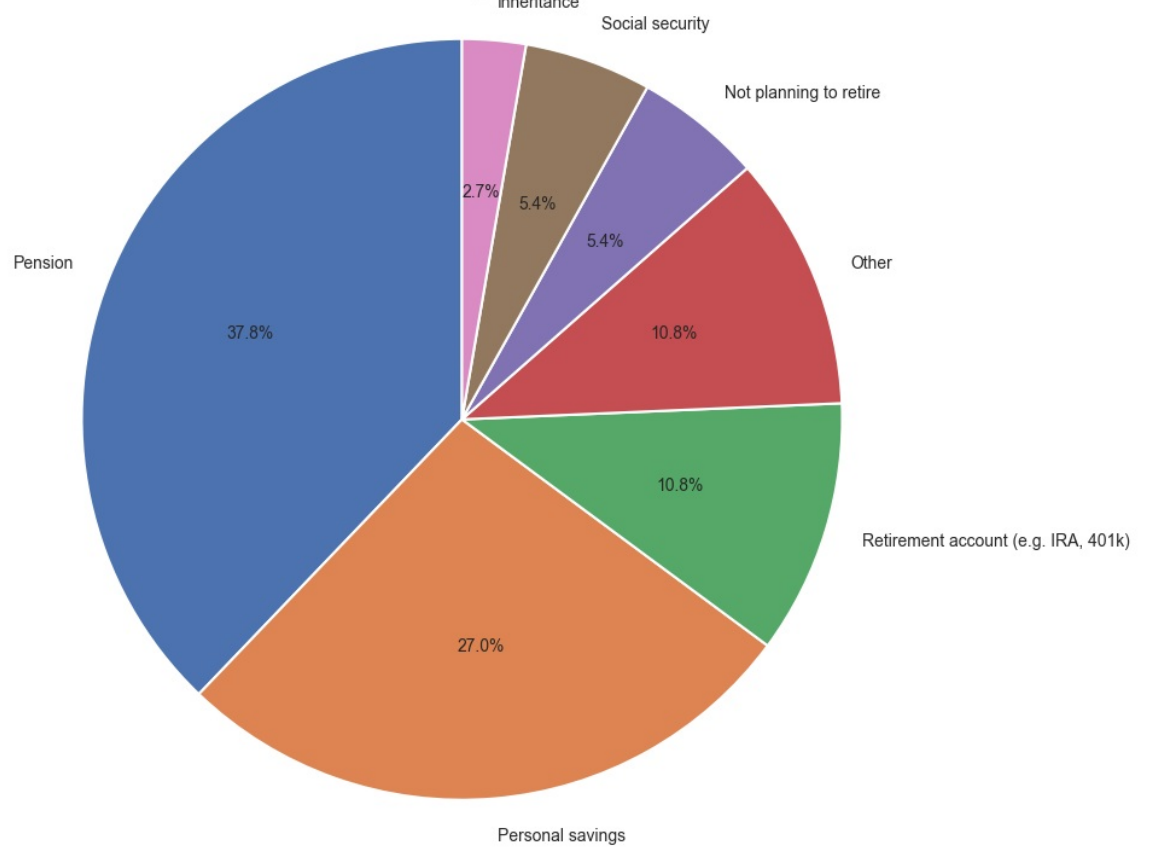
Retirement Age



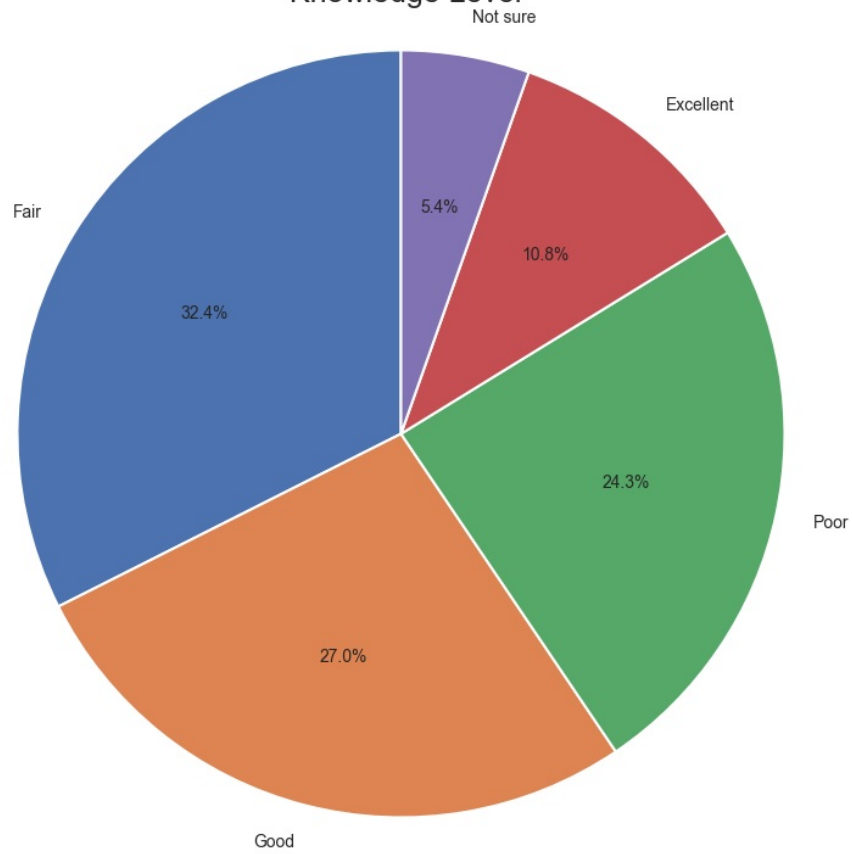
Comfort Score
2



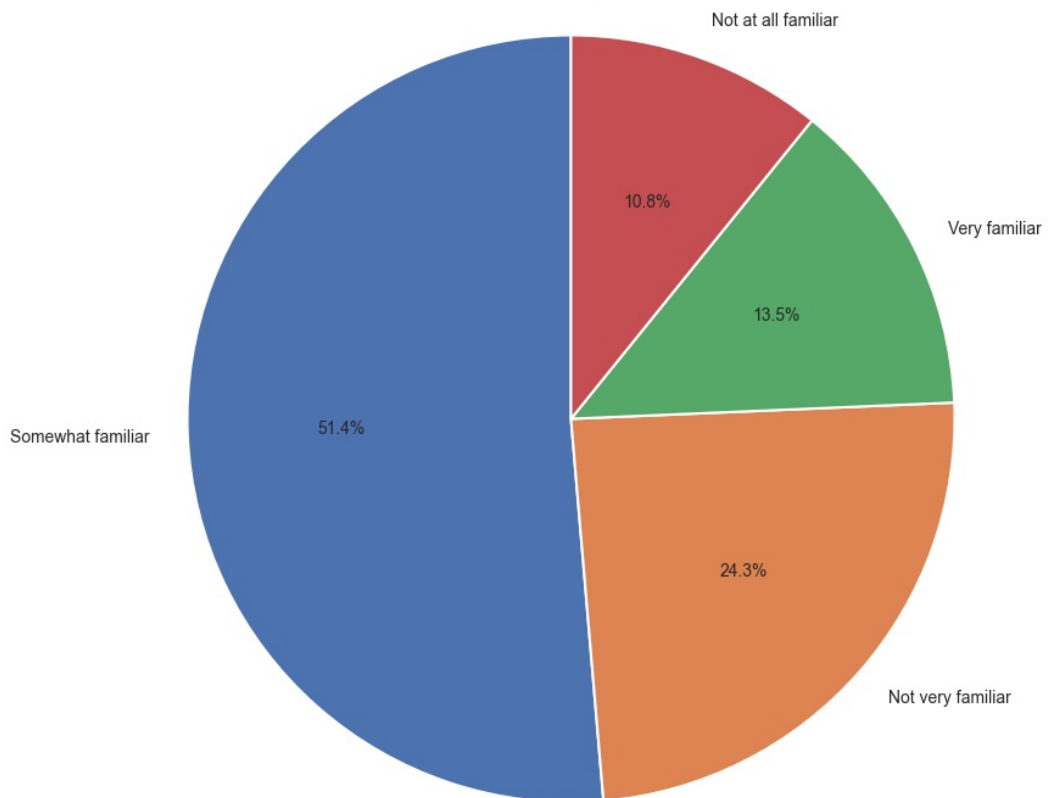
Funding Plan
Inheritance



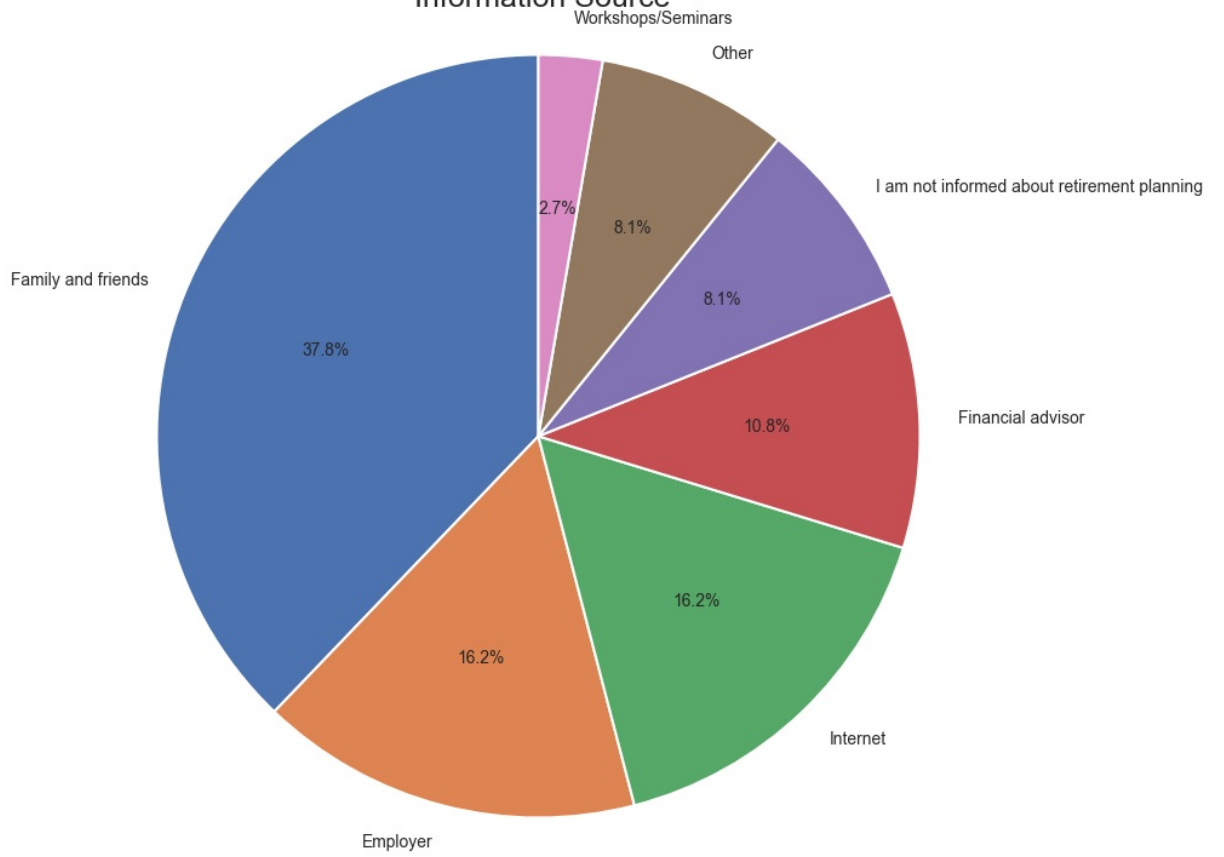
Knowledge Level



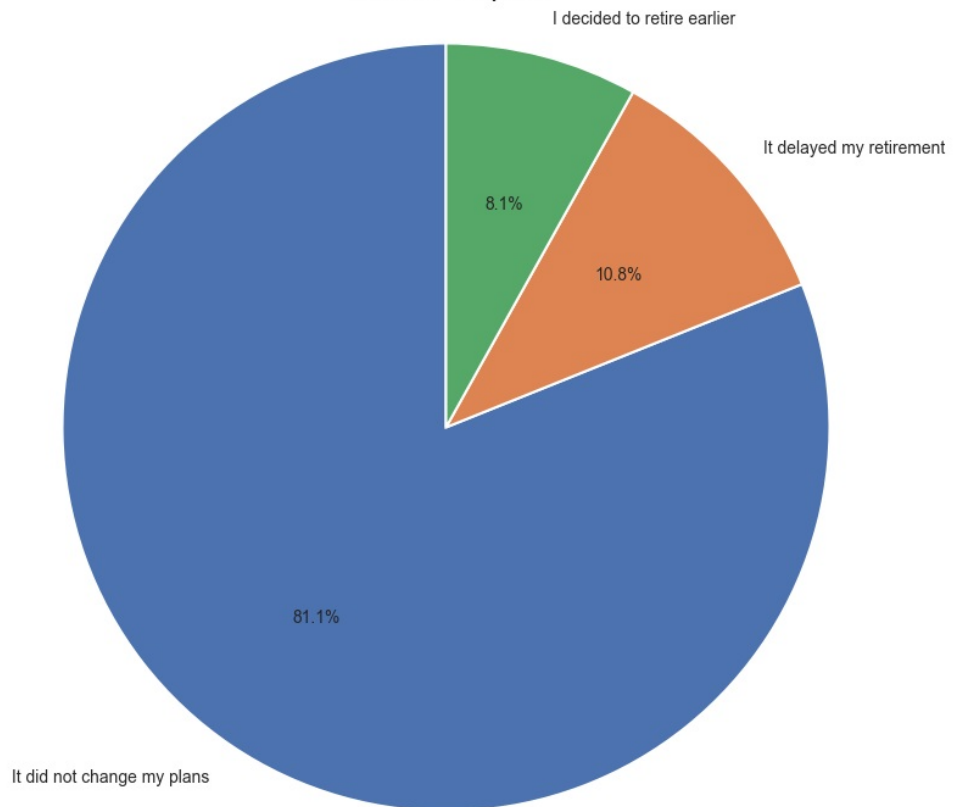
Familiarity Level



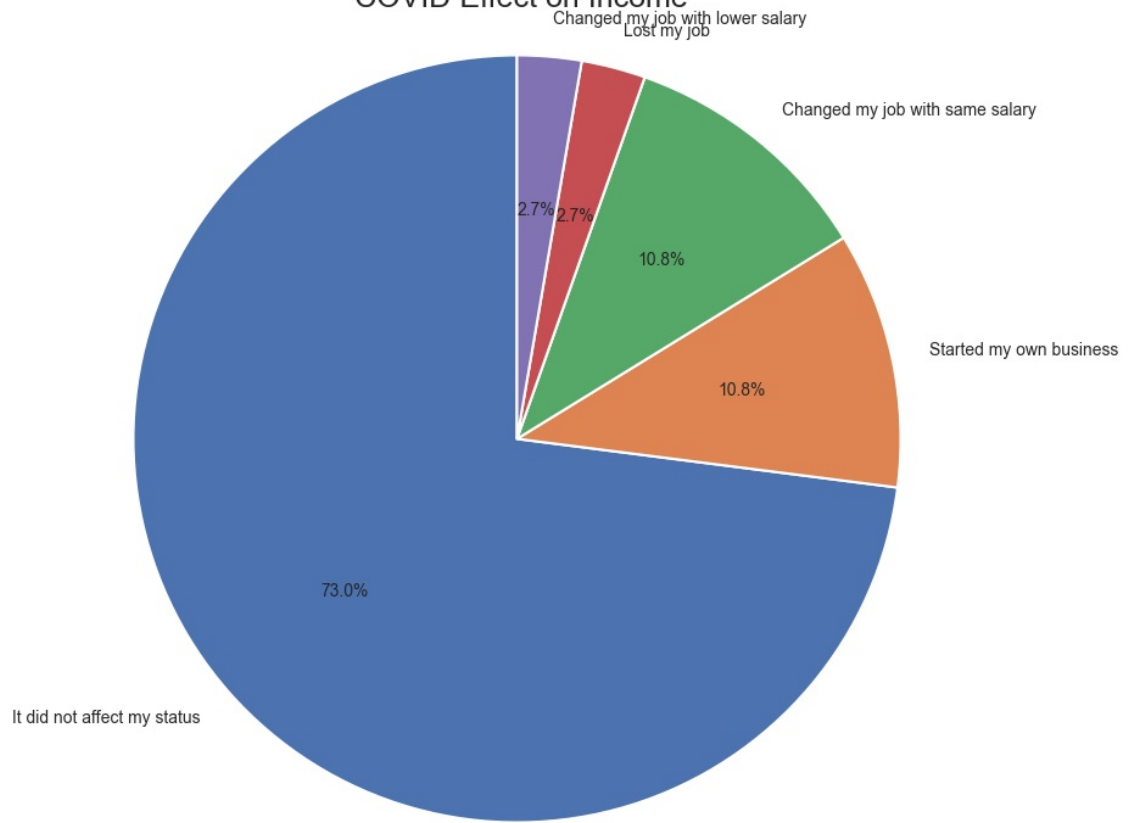
Information Source



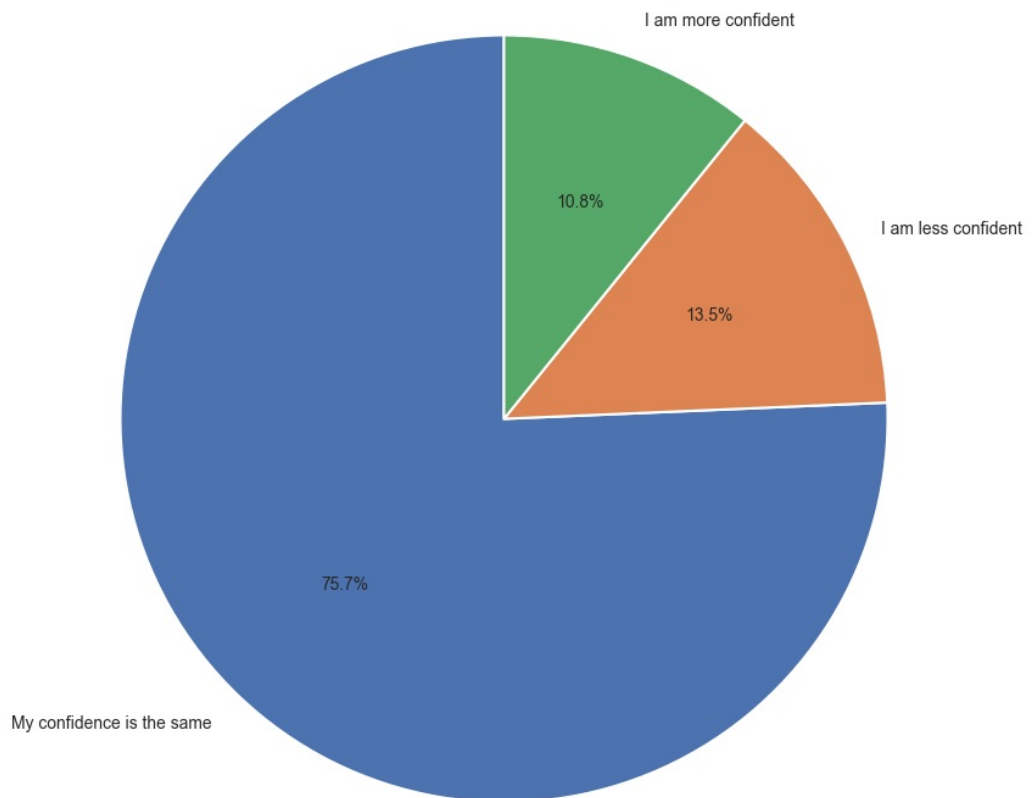
COVID Impact



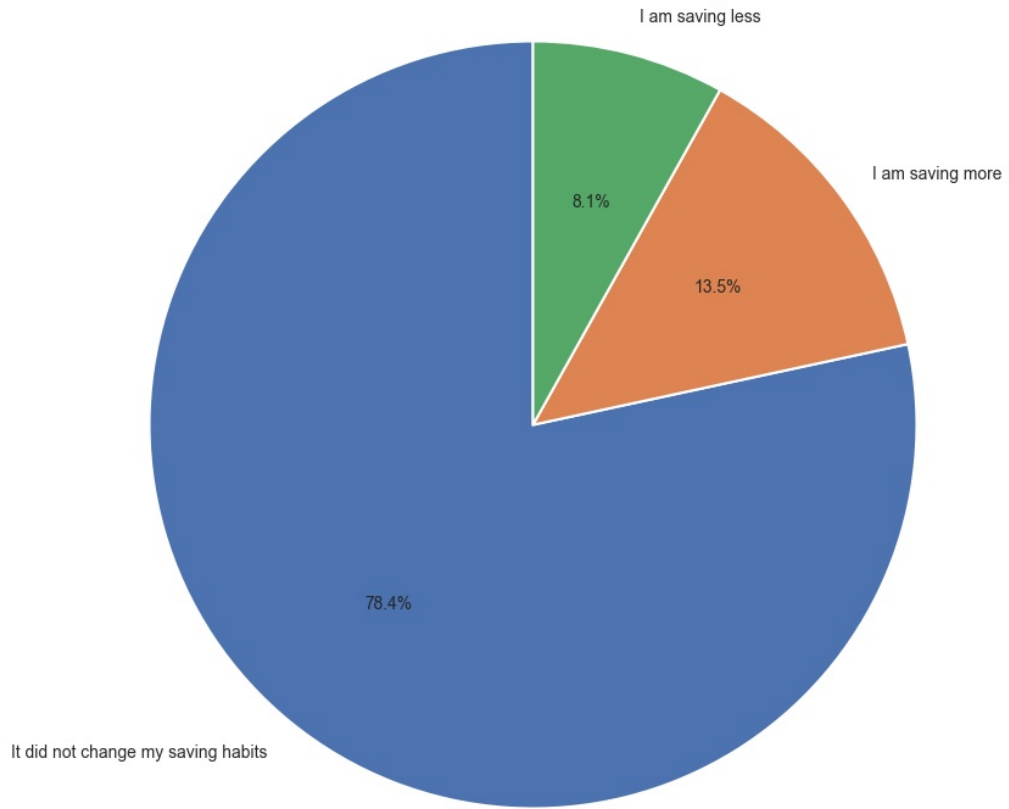
COVID Effect on Income



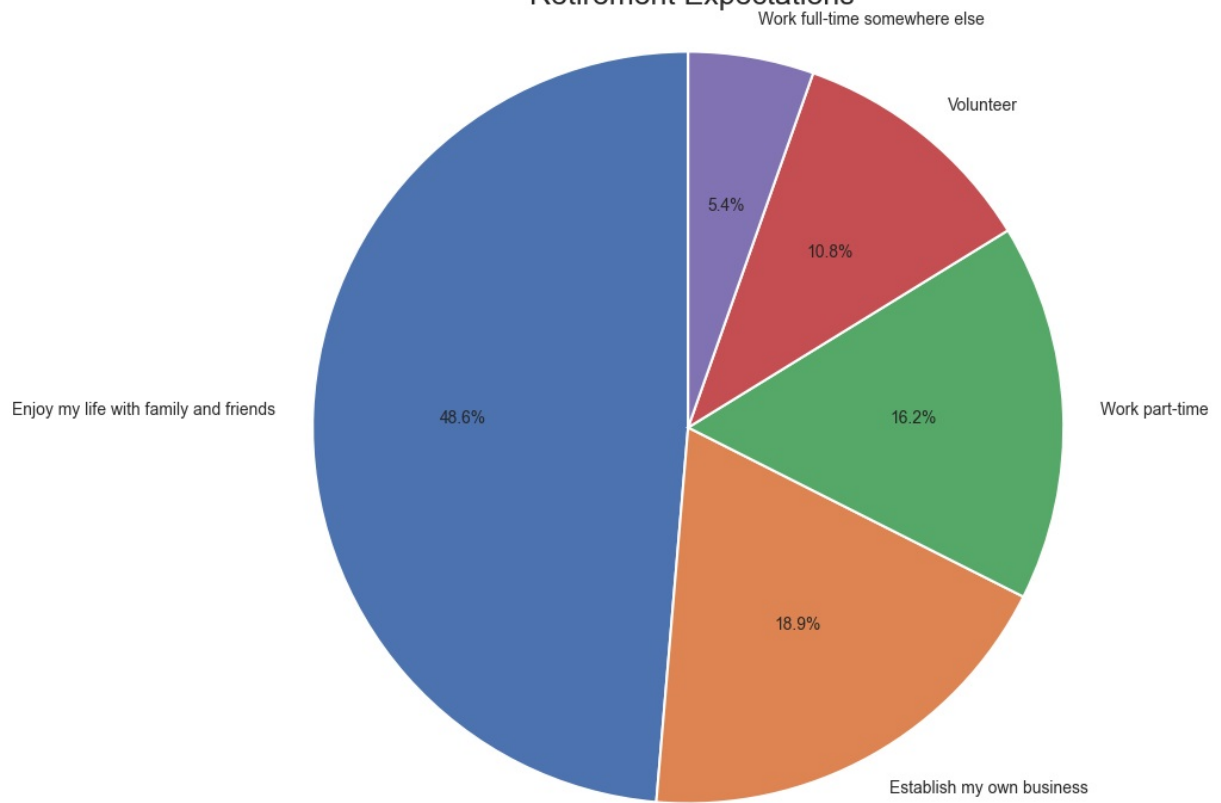
COVID Confidence Change



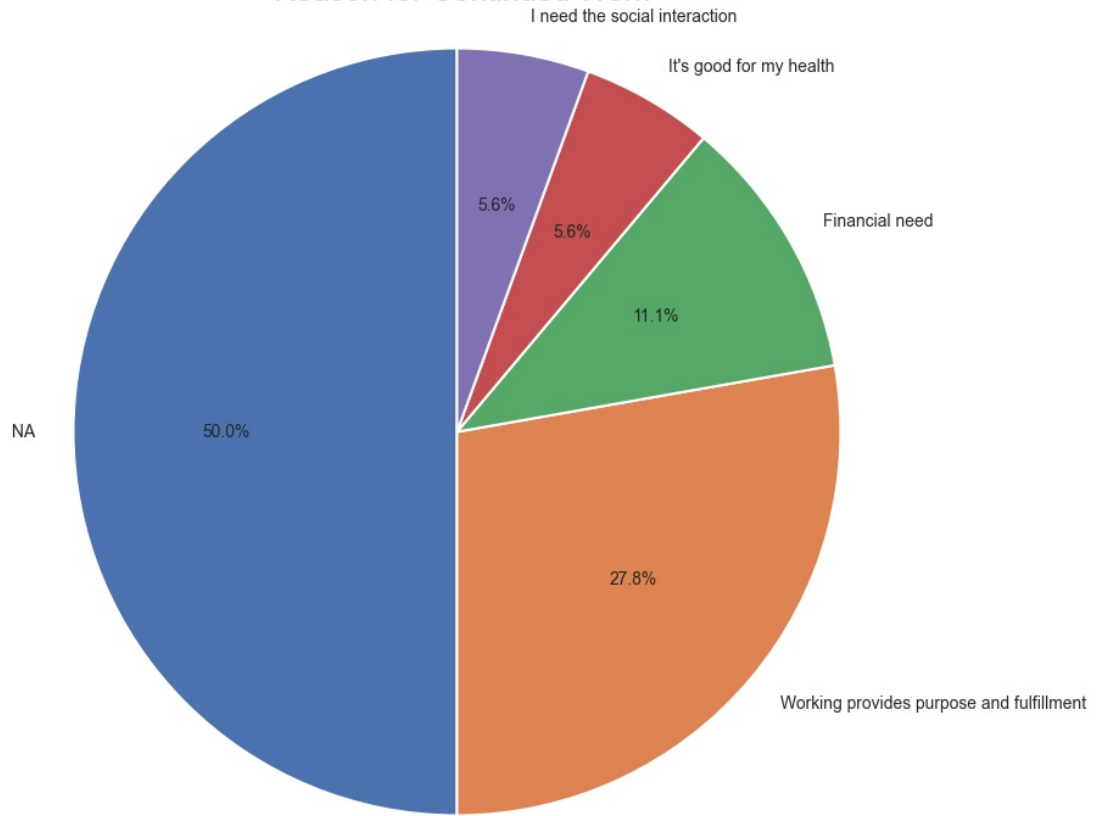
COVID Influence on Saving



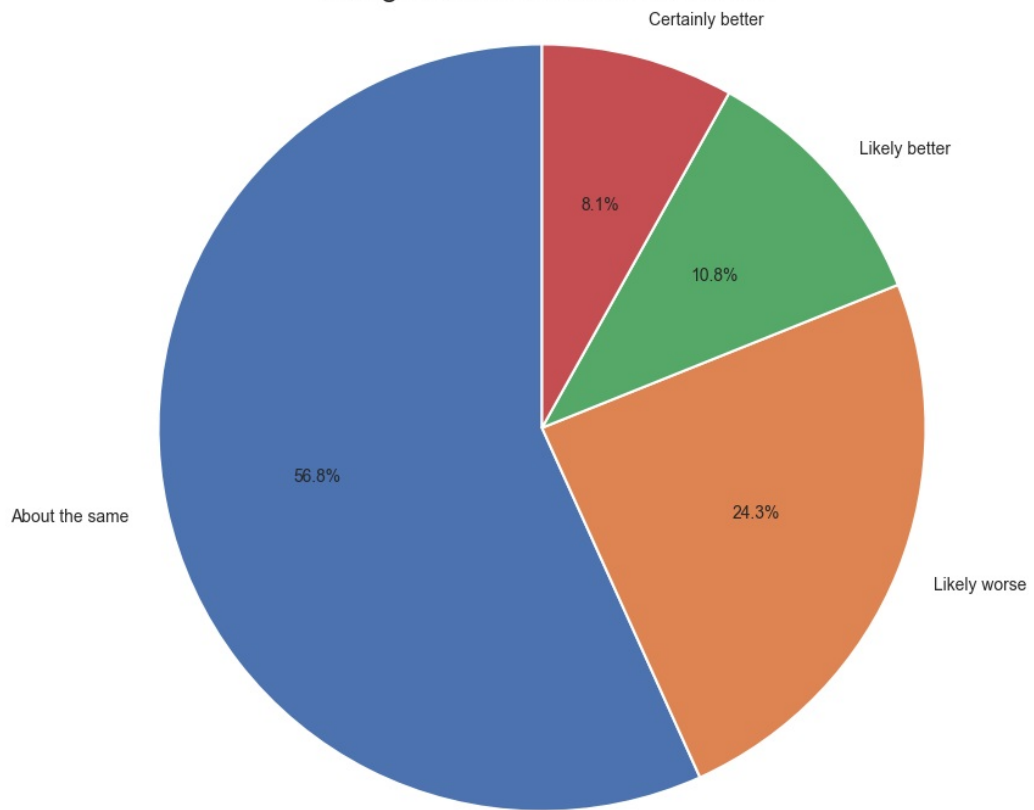
Retirement Expectations

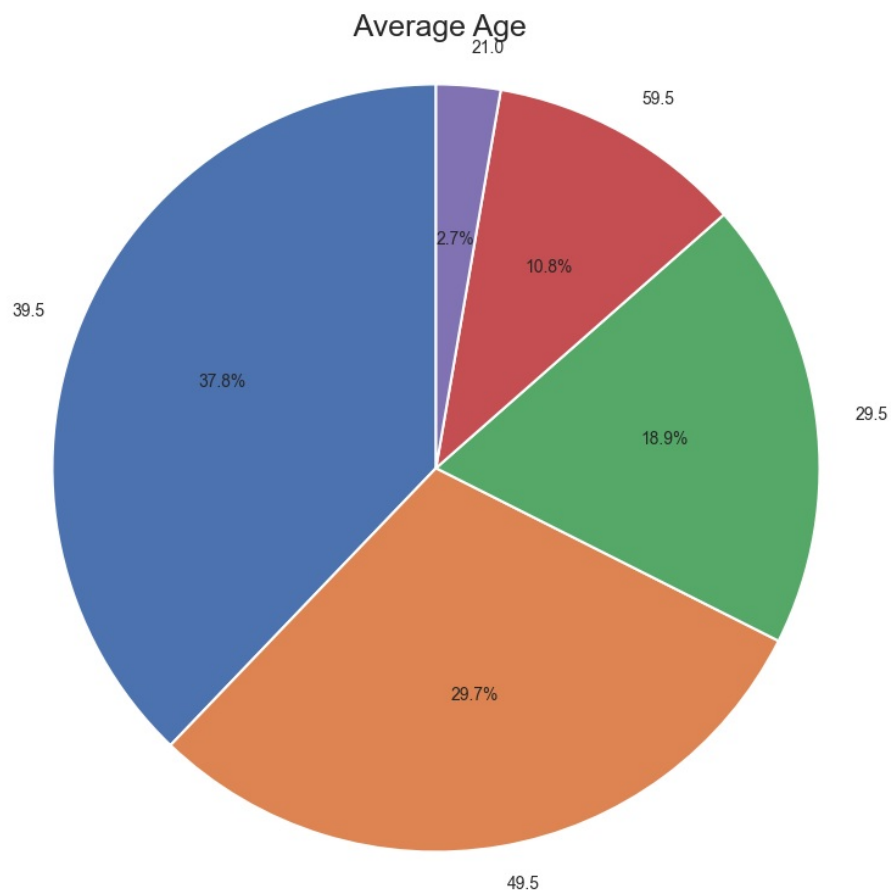
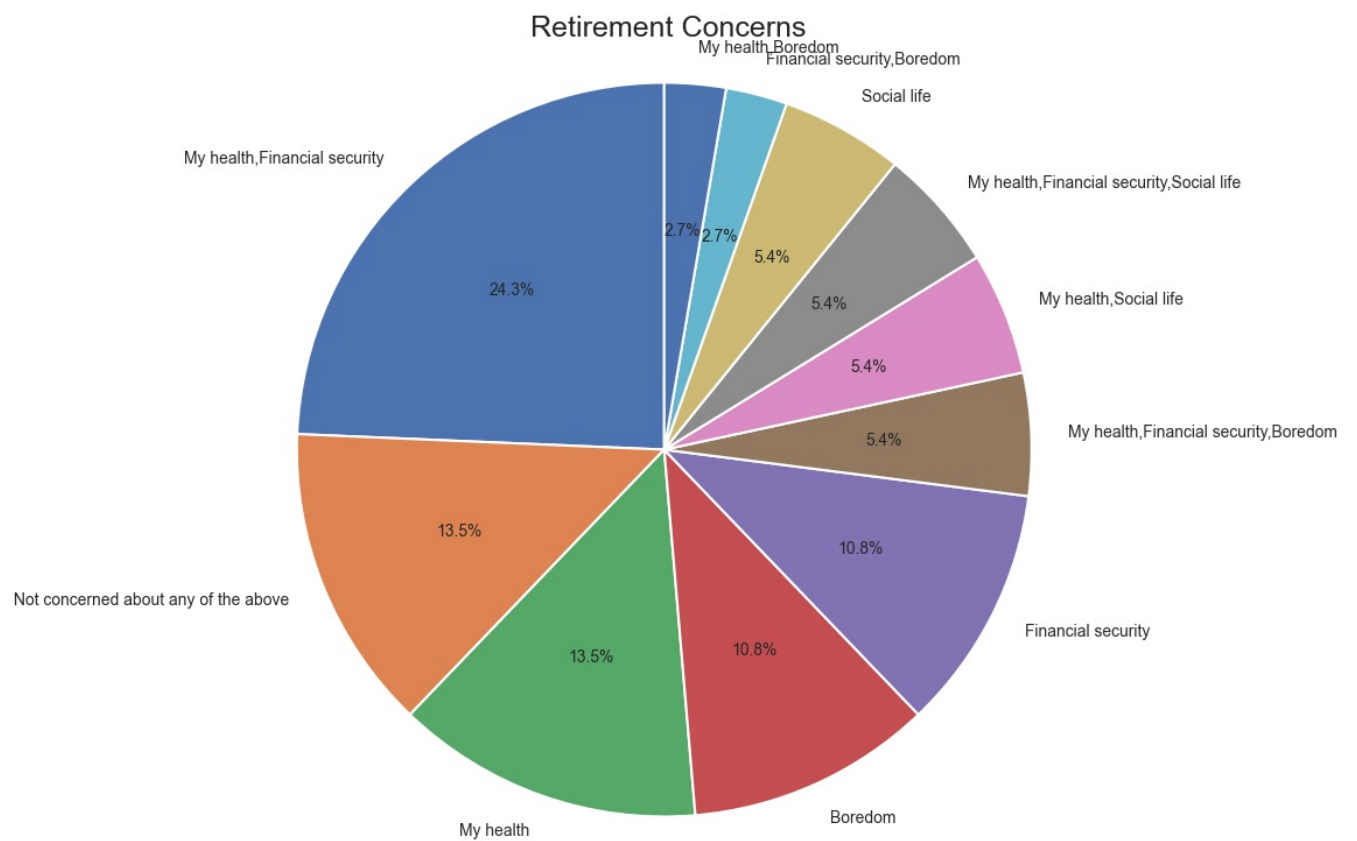


Reason for Continued Work



Living Standards after Retirement





Visualise relationships between variables

```
In [ ]: # Assuming you have a DataFrame df with a column 'column_name' containing null values
```

```
In [ ]: data
```

```
Out[ ]:
```

	Age	Gender	Social Status	Num Children	Employment Status	Retirement Age	Comfort Score	Funding Plan	Knowledge Level	Familiarity Level	Information Source	COVID Impac
0	25 -	Female	Couple	0	Employed full-time	56 - 60	5	Personal	Poor	Not at all	Family and friends	It did no change

34				savings				familiar				my plans
1	35 - 44	Male	Single	1 - 2	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did no change my plans
2	25 - 34	Male	Couple	3 - 4	employed/Business owner	Not planning to retire at all	2	Personal savings	Poor	Somewhat familiar	Family and friends	It did no change my plans
3	45 - 54	Female	Couple	0	Employed full-time	65 - 70	2	Retirement account (e.g. IRA, 401k)	Poor	Not very familiar	Family and friends	It did no change my plans
4	25 - 34	Female	Single	0	Employed full-time	61 - 65	3	Pension	Poor	Not at all familiar	I am not informed about retirement planning	It did no change my plans
5	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	1	Pension	Fair	Somewhat familiar	Family and friends	It did no change my plans
6	35 - 44	Male	Couple	1 - 2	employed/Business owner	Not planning to retire at all	1	Personal savings	Fair	Not very familiar	Other	It did no change my plans
7	25 - 34	Male	Single	0	Employed full-time	Not planning to retire at all	1	Inheritance	Excellent	Somewhat familiar	Family and friends	It delayec my retiremen
8	25 - 34	Female	Single	0	Employed full-time	55 or less	4	Retirement account (e.g. IRA, 401k)	Fair	Somewhat familiar	Employer	It did no change my plans
9	45 - 54	Male	Couple	0	Employed full-time	55 or less	3	Personal savings	Excellent	Somewhat familiar	Internet	It did no change my plans
10	35 - 44	Male	Couple	0	employed/Business owner	55 or less	4	Other	Fair	Somewhat familiar	Other	It did no change my plans
11	55 - 64	Male	Couple	0	Employed full-time	61 - 65	4	Personal savings	Good	Somewhat familiar	I am not informed about retirement planning	It did no change my plans
12	35 - 44	Male	Couple	1 - 2	employed/Business owner	56 - 60	5	Pension	Fair	Very familiar	Internet	It did no change my plans
13	35 - 44	Male	Couple	0	Employed full-time	61 - 65	1	Pension	Fair	Somewhat familiar	Internet	It delayec my retiremen
14	25 - 34	Female	Couple	0	employed/Business owner	56 - 60	5	Other	Good	Not at all familiar	Family and friends	It did no change my plans
15	45 - 54	Female	Couple	1 - 2	Employed full-time	61 - 65	4	Pension	Fair	Somewhat familiar	Financial advisor	It did no change my plans
16	25 - 34	Male	Single	0	Employed full-time	55 or less	4	Personal savings	Fair	Somewhat familiar	Internet	It did no change my plans
17	35 - 44	Male	Couple	1 - 2	Employed full-time	55 or less	5	Personal savings	Fair	Not very familiar	Family and friends	It delayec my retiremen
18	45 -	Male	Couple	3 - 4	Self-employed/Business	65 - 70	3	Retirement account	Good	Somewhat	Financial advisor	It did no change

	54	owner					(e.g. IRA, 401k)		familiar			my plans
19	45 - 54	Male	Couple	3 - 4	Employed full-time	61 - 65	4	Pension	Poor	Somewhat familiar	Employer	It delayed my retirement
20	45 - 54	Male	Couple	1 - 2	Employed full-time	After 70	3	Not planning to retire	Good	Somewhat familiar	Internet	It did not change my plans
21	35 - 44	Male	Couple	1 - 2	Employed full-time	55 or less	4	Personal savings	Poor	Not very familiar	Family and friends	I decided to retire earlier
22	35 - 44	Female	Couple	3 - 4	Employed full-time	55 or less	4	Other	Good	Somewhat familiar	Workshops/Seminars	I decided to retire earlier
23	18 - 24	Male	Single	0	Employed full-time	Not planning to retire at all	1	Not planning to retire	Poor	Not at all familiar	I am not informed about retirement planning	It did not change my plans
24	35 - 44	Male	Couple	3 - 4	Employed full-time	55 or less	5	Pension	Good	Very familiar	Employer	It did not change my plans
25	45 - 54	Male	Couple	3 - 4	Employed full-time	61 - 65	3	Pension	Poor	Not very familiar	Family and friends	It did not change my plans
26	35 - 44	Female	Single	0	Employed full-time	After 70	3	Other	Not sure	Not very familiar	Family and friends	It did not change my plans
27	35 - 44	Male	Couple	5 - 6	Employed full-time	61 - 65	4	Social security	Good	Somewhat familiar	Family and friends	It did not change my plans
28	55 - 64	Male	Couple	5 - 6	Retired	I am already retired	4	Pension	Excellent	Very familiar	Employer	I decided to retire earlier
29	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	3	Personal savings	Good	Not very familiar	Family and friends	It did not change my plans
30	45 - 54	Female	Single	0	Employed full-time	56 - 60	5	Pension	Good	Very familiar	Employer	It did not change my plans
31	45 - 54	Female	Couple	3 - 4	Employed full-time	55 or less	4	Pension	Poor	Not very familiar	Employer	It did not change my plans
32	55 - 64	Male	Couple	> 6	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did not change my plans
33	45 - 54	Male	Couple	3 - 4	Employed full-time	56 - 60	5	Personal savings	Good	Somewhat familiar	Financial advisor	It did not change my plans
34	45 - 54	Male	Couple	5 - 6	Employed full-time	56 - 60	5	Pension	Excellent	Very familiar	Internet	It did not change my plans
35	35 - 44	Male	Couple	1 - 2	Employed full-time	After 70	4	Retirement account (e.g. IRA, 401k)	Fair	Somewhat familiar	Financial advisor	It did not change my plans
	55 -							Social		Not very		It did not

In []: data.dtypes

Out[]:

Age	object
Gender	object
Social Status	object
Num Children	object
Employment Status	object
Retirement Age	object
Comfort Score	int64
Funding Plan	object
Knowledge Level	object
Familiarity Level	object
Information Source	object
COVID Impact	object
COVID Effect on Income	object
COVID Confidence Change	object
COVID Influence on Saving	object
Retirement Expectations	object
Reason for Continued Work	object
Living Standards after Retirement	object
Retirement Concerns	object
Average Age	float64
dtype:	object

In []: data.head()

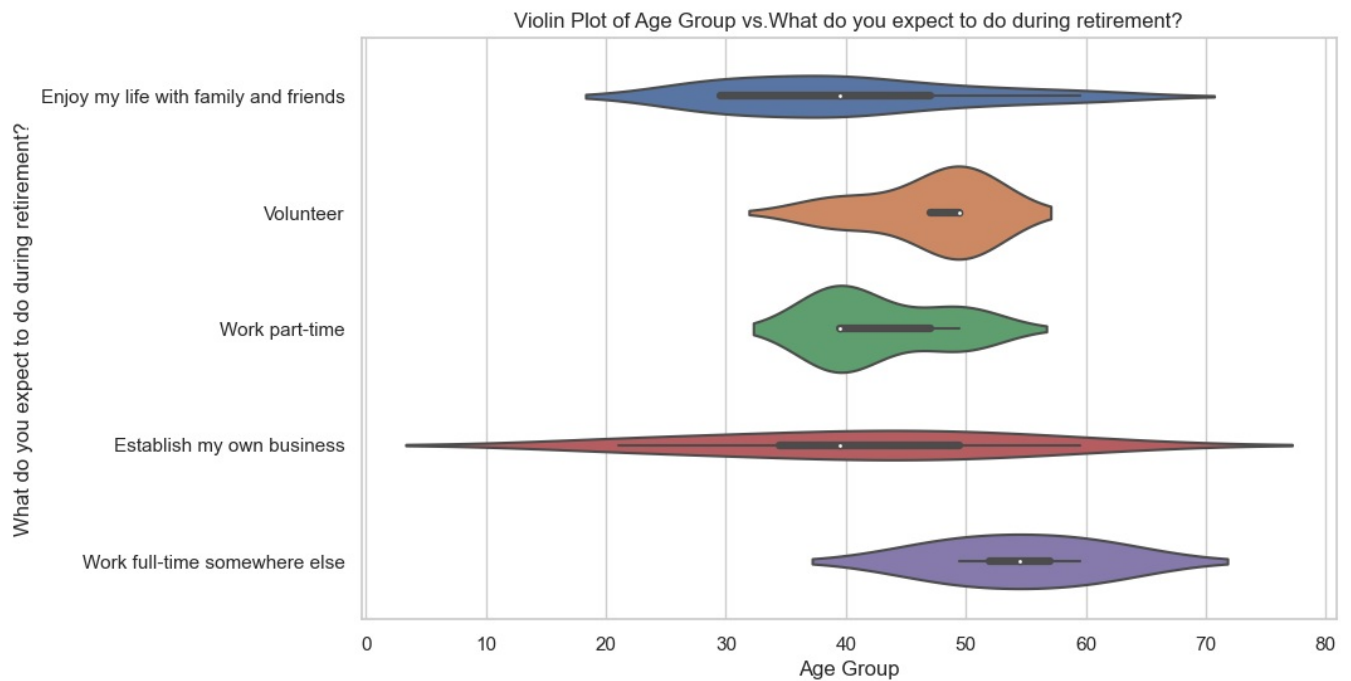
Out[]:

	Age	Gender	Social Status	Num Children	Employment Status	Retirement Age	Comfort Score	Funding Plan	Knowledge Level	Familiarity Level	Information Source	COVID Impact	COVID Effect on Income
0	25 - 34	Female	Couple	0	Employed full-time	56 - 60	5	Personal savings	Poor	Not at all familiar	Family and friends	It did not change my plans	It did not affect my status
1	35 - 44	Male	Single	1 - 2	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did not change my plans	It did not affect my status
2	25 - 34	Male	Couple	3 - 4	Self-employed/Business owner	Not planning to retire at all	2	Personal savings	Poor	Somewhat familiar	Family and friends	It did not change my plans	Started my own business
3	45 - 54	Female	Couple	0	Employed full-time	65 - 70	2	Retirement account (e.g. IRA, 401k)	Poor	Not very familiar	Family and friends	It did not change my plans	It did not affect my status
4	25 - 34	Female	Single	0	Employed full-time	61 - 65	3	Pension	Poor	Not at all familiar	I am not informed about retirement planning	It did not change my plans	It did not affect my status

In []:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Create a violin plot to visualize comfort level with retirement idea across age groups
plt.figure(figsize=(10, 6))
sns.set(style="whitegrid")
sns.violinplot(x='Average Age', y='Retirement Expectations', data=data)
plt.xlabel("Age Group")
plt.ylabel("What do you expect to do during retirement?")
plt.title("Violin Plot of Age Group vs.What do you expect to do during retirement?")
plt.show()
```



The most common expectation for retirement is to enjoy life with family and friends.

Volunteering is another popular expectation for retirement.

Working part-time and establishing a business are also common expectations.

Working full-time somewhere else is the least common expectation for retirement

```
In [ ]: reason_counts = data['Reason for Continued Work'].value_counts()
reason_counts
```

```
Out[ ]: NA                                18
Working provides purpose and fulfillment  10
Financial need                           4
It's good for my health                   2
I need the social interaction              2
Name: Reason for Continued Work, dtype: int64
```

```
In [ ]: data
```

	Age	Gender	Social Status	Num Children	Employment Status	Retirement Age	Comfort Score	Funding Plan	Knowledge Level	Familiarity Level	Information Source	COVID Impact
0	25 - 34	Female	Couple	0	Employed full-time	56 - 60	5	Personal savings	Poor	Not at all familiar	Family and friends	It did no change my plans
1	35 - 44	Male	Single	1 - 2	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did no change my plans
2	25 - 34	Male	Couple	3 - 4	Self-employed/Business owner	Not planning to retire at all	2	Personal savings	Poor	Somewhat familiar	Family and friends	It did no change my plans
3	45 - 54	Female	Couple	0	Employed full-time	65 - 70	2	Retirement account (e.g. IRA, 401k)	Poor	Not very familiar	Family and friends	It did no change my plans
4	25 - 34	Female	Single	0	Employed full-time	61 - 65	3	Pension	Poor	Not at all familiar	I am not informed about retirement planning	It did no change my plans
5	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	1	Pension	Fair	Somewhat familiar	Family and friends	It did no change my plans
6	35 -	Male	Couple	1 - 2	Self-employed/Business owner	Not planning to	1	Personal	Fair	Not very	Other	It did no change

25	45 - 54	Male	Couple	3 - 4	Employed full-time	61 - 65	3	Pension	Poor	Not very familiar	Family and friends	It did no change my plans
26	35 - 44	Female	Single	0	Employed full-time	After 70	3	Other	Not sure	Not very familiar	Family and friends	It did no change my plans
27	35 - 44	Male	Couple	5 - 6	Employed full-time	61 - 65	4	Social security	Good	Somewhat familiar	Family and friends	It did no change my plans
28	55 - 64	Male	Couple	5 - 6	Retired	I am already retired	4	Pension	Excellent	Very familiar	Employer	I decided to retire earlie
29	35 - 44	Male	Couple	1 - 2	Employed full-time	65 - 70	3	Personal savings	Good	Not very familiar	Family and friends	It did no change my plans
30	45 - 54	Female	Single	0	Employed full-time	56 - 60	5	Pension	Good	Very familiar	Employer	It did no change my plans
31	45 - 54	Female	Couple	3 - 4	Employed full-time	55 or less	4	Pension	Poor	Not very familiar	Employer	It did no change my plans
32	55 - 64	Male	Couple	> 6	Employed full-time	65 - 70	4	Pension	Fair	Somewhat familiar	Family and friends	It did no change my plans
33	45 - 54	Male	Couple	3 - 4	Employed full-time	56 - 60	5	Personal savings	Good	Somewhat familiar	Financial advisor	It did no change my plans
34	45 - 54	Male	Couple	5 - 6	Employed full-time	56 - 60	5	Pension	Excellent	Very familiar	Internet	It did no change my plans
35	35 - 44	Male	Couple	1 - 2	Employed full-time	After 70	4	Retirement account (e.g. IRA, 401k)	Fair	Somewhat familiar	Financial advisor	It did no change my plans
36	55 - 64	Male	Couple	3 - 4	Employed full-time	61 - 65	3	Social security	Not sure	Not very familiar	Other	It did no change my plans

In []: data.isnull().sum()

```
Out[ ]: Age                                0
Gender                                  0
Social Status                          0
Num Children                           0
Employment Status                      0
Retirement Age                        0
Comfort Score                          0
Funding Plan                           0
Knowledge Level                        0
Familiarity Level                      0
Information Source                     0
COVID Impact                           0
COVID Effect on Income                 0
COVID Confidence Change                0
COVID Influence on Saving              0
Retirement Expectations               0
Reason for Continued Work               1
Living Standards after Retirement       0
Retirement Concerns                   0
Average Age                            0
dtype: int64
```

```

In [ ]: unique_children = data['Gender'].unique()

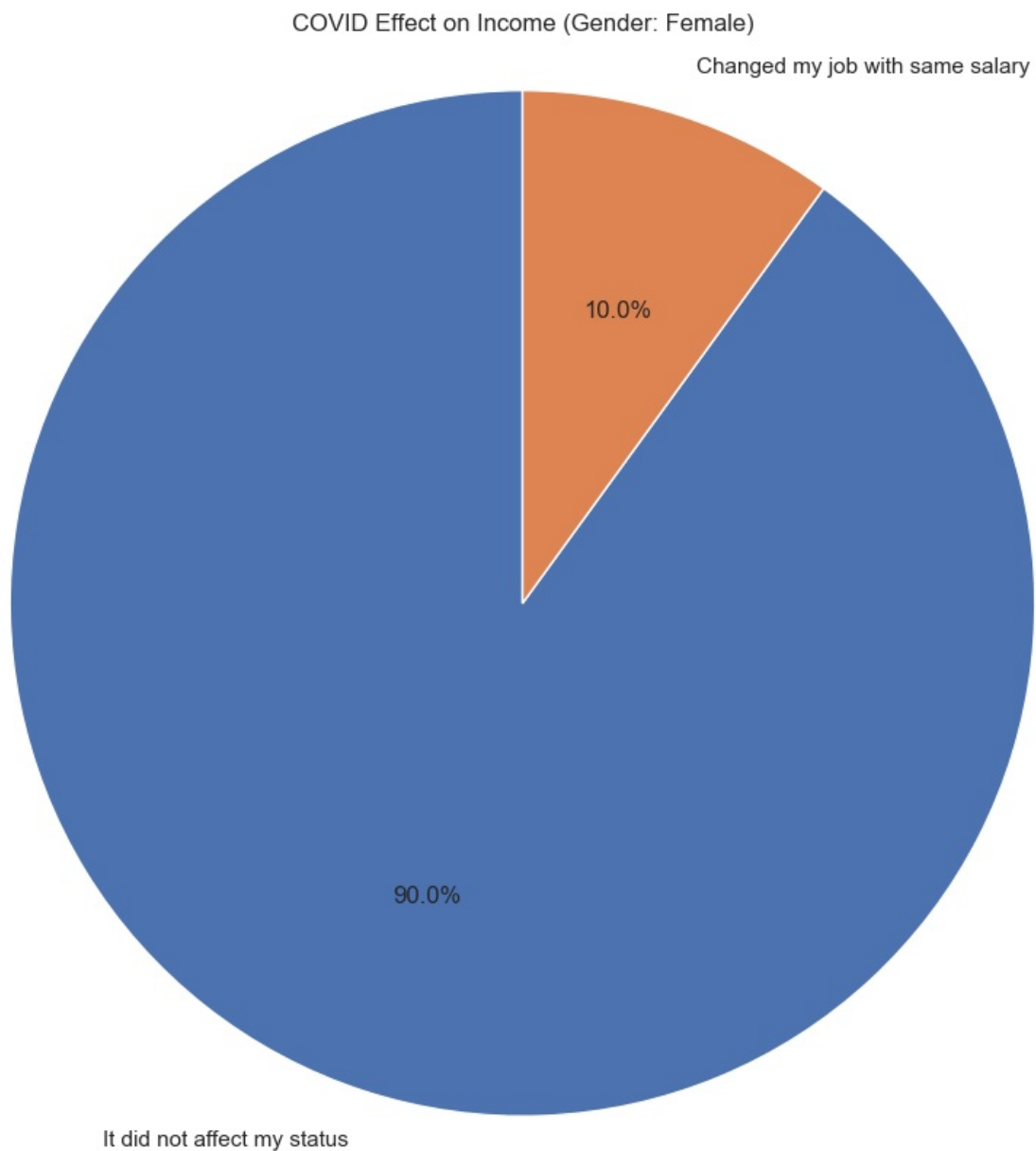
# Create a pie chart for each unique number of children
for num_children in unique_children:
    plt.figure(figsize=(10, 10))

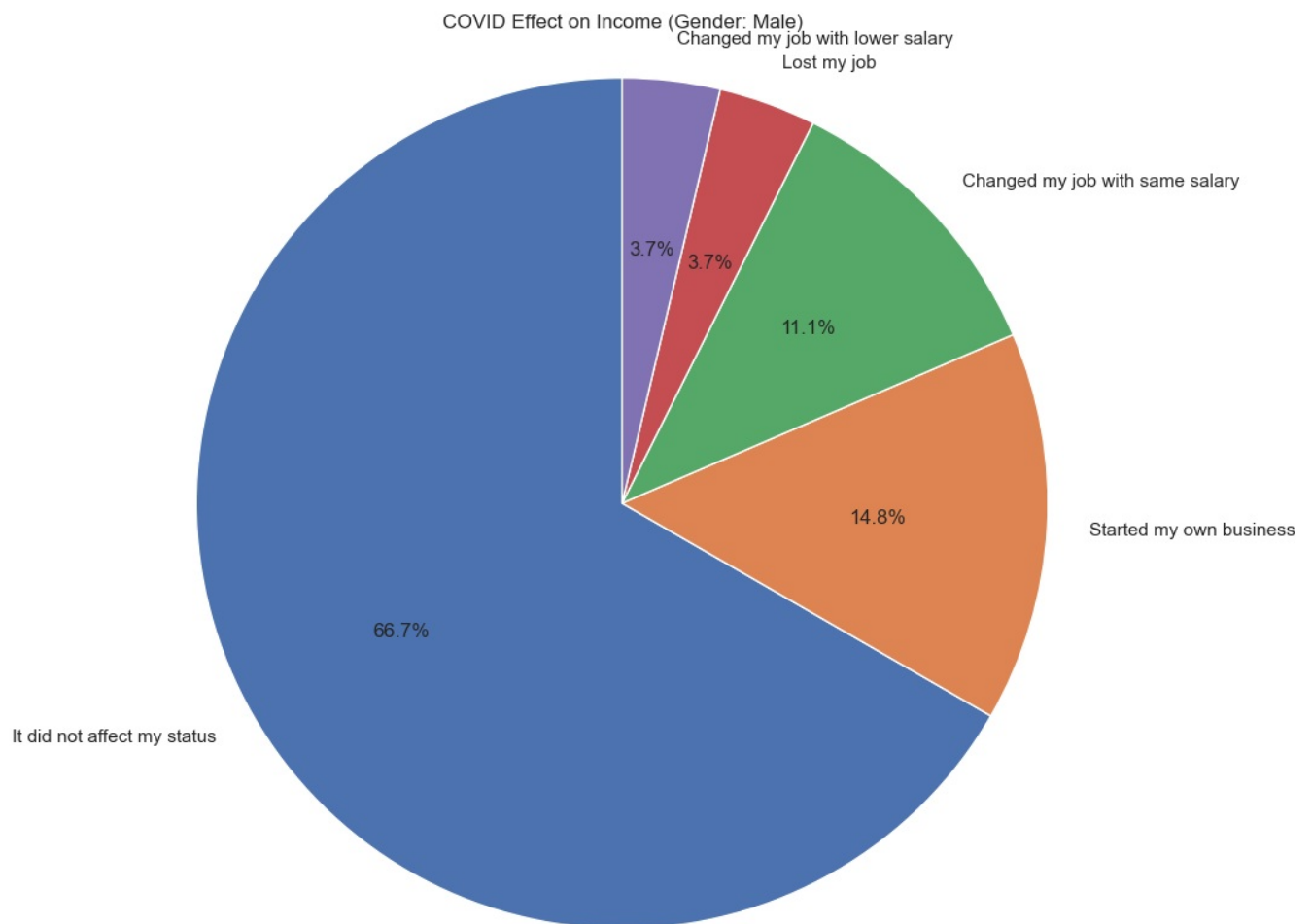
    # Filter the DataFrame for the current number of children
    filtered_data = data[data['Gender'] == num_children]

    # Count the occurrences of each reason
    reason_counts = filtered_data['COVID Effect on Income'].value_counts()

    # Create a pie chart
    plt.pie(reason_counts, labels=reason_counts.index, autopct='%1.1f%%', startangle=90)
    plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle
    plt.title(f"COVID Effect on Income (Gender: {num_children})")
    plt.show()

```





In a survey, when asked about the impact of a certain decision on their income:

Females:

90% of females reported that the decision didn't affect their income. 10% of females said they changed their job while maintaining the same income. Males:

66.7% of males stated that the decision had no impact on their income. 3.7% of males changed jobs with a lower salary. 3.7% of males lost their job. 11.1% of males changed jobs while maintaining the same salary. 14.8% of males decided to start their own business.

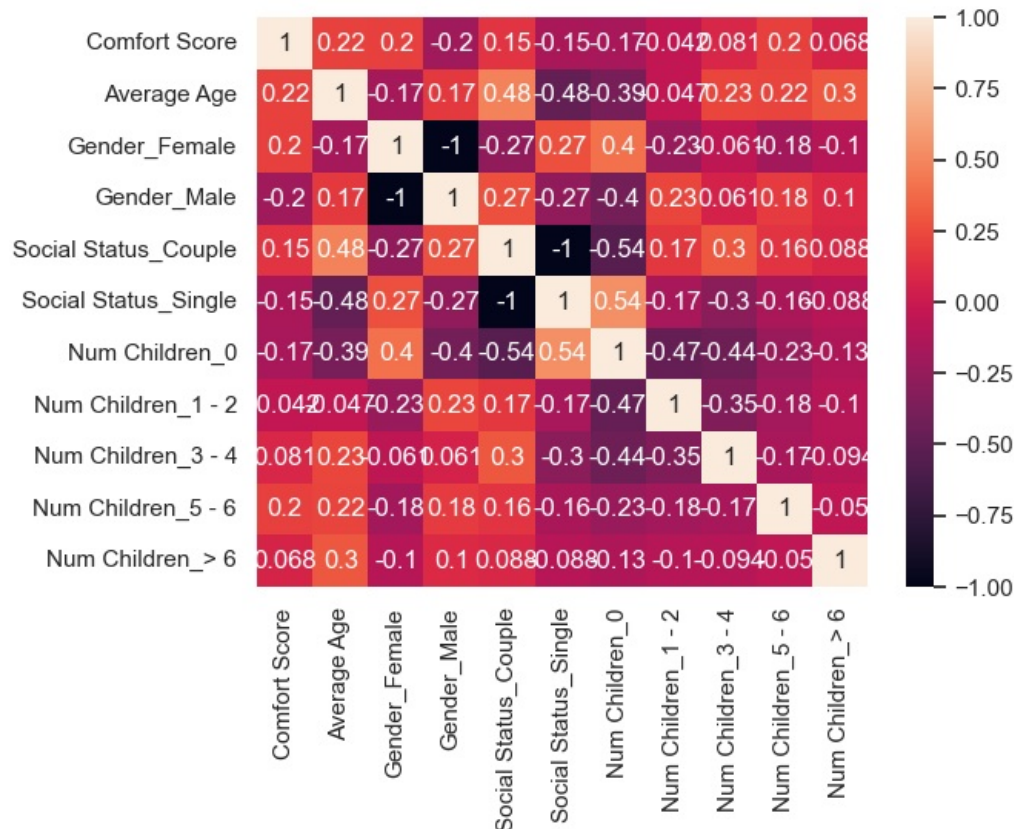
```
In [ ]: # Define the categorical columns to be encoded
categorical_columns = ['Gender',
                      'Social Status',
                      'Num Children']

# Perform one-hot encoding
encoded_data = pd.get_dummies(data, columns=categorical_columns)
```

```
In [ ]: corr=encoded_data.corr()
sns.heatmap(corr, annot=True)
```

C:\Users\Almadina computers\AppData\Local\Temp\ipykernel_10256\3538035621.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
corr=encoded_data.corr()

Out[]: <AxesSubplot: >



The correlation between "Extremely uncomfortable vs Very comfortable" and "Average age" is positive, but weak. This means that there is a slight tendency for older respondents to be more comfortable with their current living situation.

The correlation between "Extremely uncomfortable vs Very comfortable" and "Gender" is negative, but weak. This means that there is a slight tendency for female respondents to be more comfortable with their current living situation than male respondents.

The correlation between "Extremely uncomfortable vs Very comfortable" and "Social status" is positive, but weak. This means that there is a slight tendency for respondents who are in a couple to be more comfortable with their current living situation than respondents who are single.

In []: data.columns

Out[]: Index(['Age', 'Gender', 'Social Status', 'Num Children', 'Employment Status', 'Retirement Age', 'Comfort Score', 'Funding Plan', 'Knowledge Level', 'Familiarity Level', 'Information Source', 'COVID Impact', 'COVID Effect on Income', 'COVID Confidence Change', 'COVID Influence on Saving', 'Retirement Expectations', 'Reason for Continued Work', 'Living Standards after Retirement', 'Retirement Concerns', 'Average Age'], dtype='object')

```
In [ ]: from sklearn.preprocessing import LabelEncoder
# List of categorical columns to encode
categorical_columns = [
    'Age', 'Gender', 'Social Status', 'Num Children', 'Employment Status',
    'Funding Plan', 'Knowledge Level',
    'Familiarity Level', 'Information Source', 'COVID Impact',
    'COVID Effect on Income', 'COVID Confidence Change',
    'COVID Influence on Saving', 'Retirement Expectations',
    'Reason for Continued Work', 'Living Standards after Retirement',
    'Retirement Concerns', 'Average Age'
]

label_encoder = LabelEncoder()

# Encode categorical columns
for column in categorical_columns:
    data[column] = label_encoder.fit_transform(data[column])
```


Research Questions

1. Does one's knowledge about retirement planning and available schemes affect one's attitudes towards retirement? Any influence of age, sex, or social status/number of children?

```
In [ ]: import statsmodels.api as sm
knowledge = data['Knowledge Level']
attitudes = data['Comfort Score']
age = data['Average Age']
sex = data['Gender']
social_status = data['Social Status']
num_children = data['Num Children']

# Create a DataFrame with the independent variables
independent_vars = pd.DataFrame({
    'Age': age,
    'Sex': sex,
    'Social_Status': social_status,
    'Num_Children': num_children,
    'Knowledge': knowledge
})

independent_vars = sm.add_constant(independent_vars)

# Perform multiple linear regression
model = sm.OLS(attitudes, independent_vars).fit()

# Print the summary of the regression results
print(model.summary())
```

```
=====
                        OLS Regression Results
=====
Dep. Variable:          Comfort Score    R-squared:            0.187
Model:                  OLS              Adj. R-squared:       0.056
Method:                 Least Squares     F-statistic:          1.429
Date:                  Sat, 02 Sep 2023   Prob (F-statistic):    0.242
Time:                  09:35:45          Log-Likelihood:       -57.373
No. Observations:      37               AIC:                  126.7
Df Residuals:          31               BIC:                  136.4
Df Model:               5
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	4.0873	0.926	4.413	0.000	2.199	5.976
Age	0.1021	0.261	0.391	0.699	-0.431	0.635
Sex	-1.1010	0.527	-2.090	0.045	-2.176	-0.026
Social_Status	-0.2591	0.608	-0.426	0.673	-1.499	0.981
Num_Children	0.3254	0.239	1.363	0.183	-0.162	0.812
Knowledge	-0.1669	0.165	-1.010	0.321	-0.504	0.170

```
=====
Omnibus:                 2.099    Durbin-Watson:           1.702
Prob(Omnibus):            0.350    Jarque-Bera (JB):        1.695
Skew:                    -0.368    Prob(JB):                 0.429
Kurtosis:                 2.253    Cond. No.                 18.8
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Age (Age): The coefficient is positive (0.0096), but the p-value (0.719) is high, indicating that age is not statistically significant in predicting attitudes towards retirement.

Sex (Sex): The coefficient is negative (-1.1042), and the p-value (0.044) is below 0.05, suggesting that sex has a significant influence on attitudes towards retirement.

Social Status (Social_Status): The coefficient is negative (-0.2662), but the p-value (0.664) is high, indicating that social status is not statistically significant.

Number of Children (Num_Children): The coefficient is positive (0.3278), but the p-value (0.180) is relatively high, indicating that the number of children is not statistically significant.

Knowledge (Knowledge): The coefficient is negative (-0.1683), and the p-value (0.316) is relatively high, indicating that knowledge is not statistically significant.

2. Where do people place retirement planning on their list of important decisions? Is that also affected by their knowledge about retirement or by demographic factors?

```
In [ ]: data.columns
```

```
Out[ ]: Index(['Age', 'Gender', 'Social Status', 'Num Children', 'Employment Status',  
            'Retirement Age', 'Comfort Score', 'Funding Plan', 'Knowledge Level',  
            'Familiarity Level', 'Information Source', 'COVID Impact',  
            'COVID Effect on Income', 'COVID Confidence Change',  
            'COVID Influence on Saving', 'Retirement Expectations',  
            'Reason for Continued Work', 'Living Standards after Retirement',  
            'Retirement Concerns', 'Average Age'],  
           dtype='object')
```

```
In [ ]: retirement_priority = data['Comfort Score']  
knowledge = data['Knowledge Level']  
age = data['Average Age']  
gender = data['Gender']  
social_status = data['Social Status']  
num_children = data['Num Children']  
  
# Create a DataFrame with the independent variables  
independent_vars = pd.DataFrame({  
    'Knowledge': knowledge,  
    'Age': age,  
    'Gender': gender,  
    'Social_Status': social_status,  
    'Num_Children': num_children  
})  
  
# Add a constant column for the intercept  
independent_vars = sm.add_constant(independent_vars)  
  
# Perform multiple linear regression  
model = sm.OLS(retirement_priority, independent_vars).fit()  
  
# Print the summary of the regression results  
print(model.summary())
```

```

                    OLS Regression Results
=====
Dep. Variable:      Comfort Score    R-squared:                0.187
Model:              OLS             Adj. R-squared:           0.056
Method:             Least Squares   F-statistic:             1.429
Date:               Sat, 02 Sep 2023 Prob (F-statistic):       0.242
Time:               09:35:45        Log-Likelihood:          -57.373
No. Observations:   37              AIC:                       126.7
Df Residuals:       31              BIC:                       136.4
Df Model:           5
Covariance Type:    nonrobust
=====
                    coef    std err          t      P>|t|      [0.025    0.975]
-----
const                4.0873     0.926     4.413     0.000     2.199     5.976
Knowledge            -0.1669     0.165    -1.010     0.321    -0.504     0.170
Age                  0.1021     0.261     0.391     0.699    -0.431     0.635
Gender              -1.1010     0.527    -2.090     0.045    -2.176    -0.026
Social_Status       -0.2591     0.608    -0.426     0.673    -1.499     0.981
Num_Children         0.3254     0.239     1.363     0.183    -0.162     0.812
=====
Omnibus:                 2.099   Durbin-Watson:           1.702
Prob(Omnibus):            0.350   Jarque-Bera (JB):       1.695
Skew:                    -0.368   Prob(JB):               0.429
Kurtosis:                 2.253   Cond. No.               18.8
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

R-squared Value: The R-squared value of approximately 0.187 indicates that around 18.7% of the variability in the placement of retirement planning on the list of important decisions can be explained by the included independent variables.

Model Significance: The F-statistic's associated p-value is 0.243, suggesting that the overall model might not be statistically significant in predicting the placement of retirement planning as an important decision.

Coefficients: Among the included variables, the "Knowledge" variable has a coefficient of approximately -0.1683. This implies that there may be a potential negative relationship between individuals' knowledge about retirement and their placement of retirement planning on the list of important decisions. However, the p-value for this coefficient is 0.316, indicating that this relationship may not be statistically

significant.

Demographic Factors: The coefficients of demographic variables like "Gender," "Social_Status," and "Num_Children" indicate their potential influence on the placement of retirement planning on the list of important decisions. For instance, "Gender" has a coefficient of -1.1042, suggesting that gender may play a role in how individuals prioritize retirement planning.

Intercept: The intercept term of 3.9190 represents the estimated placement of retirement planning on the list of important decisions when all other independent variables are held constant.

Variable Significance: Some of the demographic variables, such as "Gender," appear to be statistically significant (p-value < 0.05), indicating that they might have an impact on individuals' prioritization of retirement planning.

3. What are the expectations after retirement in general and the main concerns after retirement, and how do these expectations and concerns affect retirement planning?

```
In [ ]: data=pd.read_excel("results-for-retirement-at-2023-08-23-1405.xlsx")

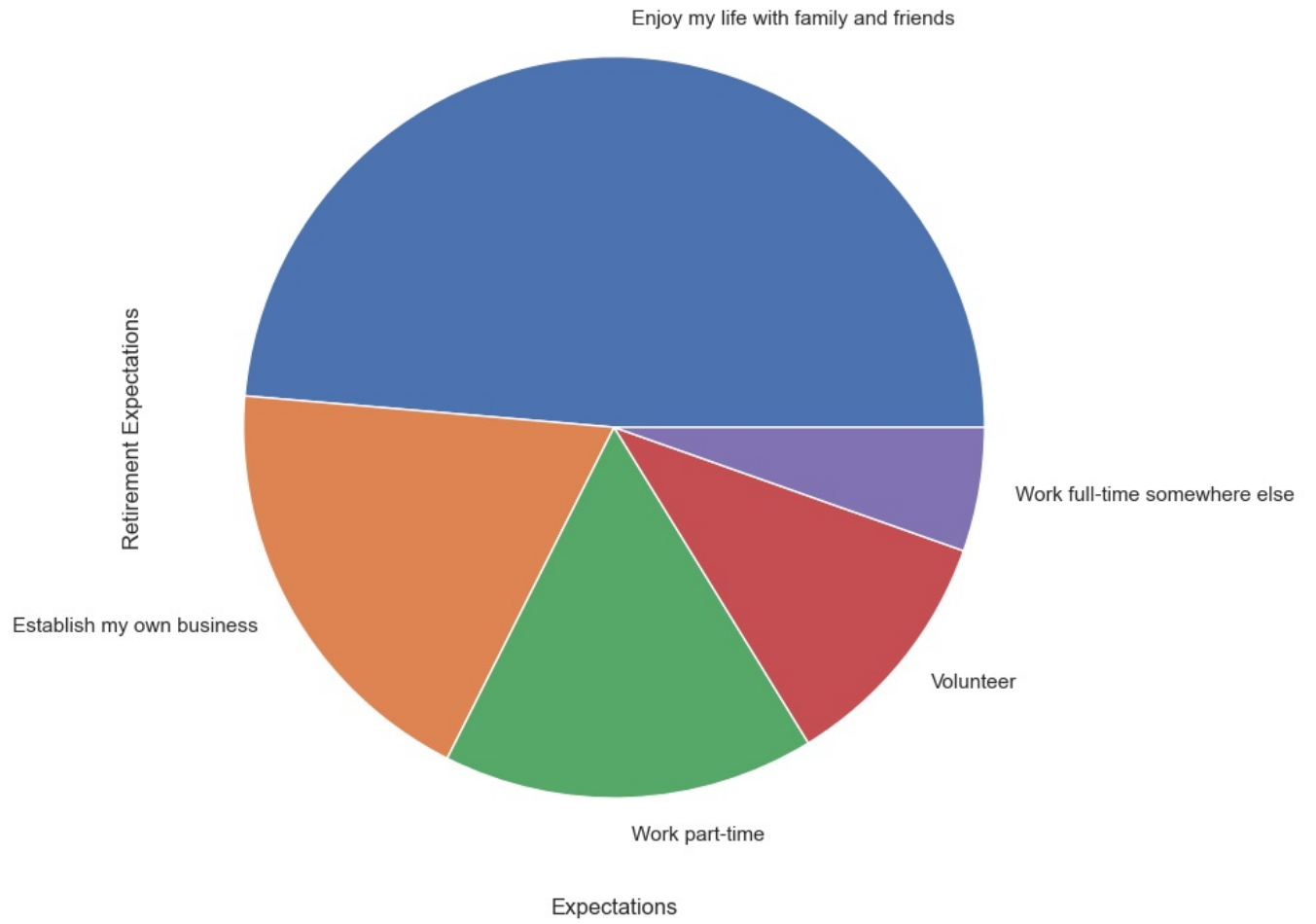
data.rename(columns=short_column_names, inplace=True)
expectations_decoded=data['Retirement Expectations']
concerns_decoded=data['Retirement Concerns']
expectations_counts = pd.Series(expectations_decoded).value_counts()
concerns_counts = pd.Series(concerns_decoded).value_counts()

plt.figure(figsize=(13, 9))
expectations_counts.plot(kind='pie')
plt.title('Distribution of Expectations After Retirement')
plt.xlabel('Expectations')

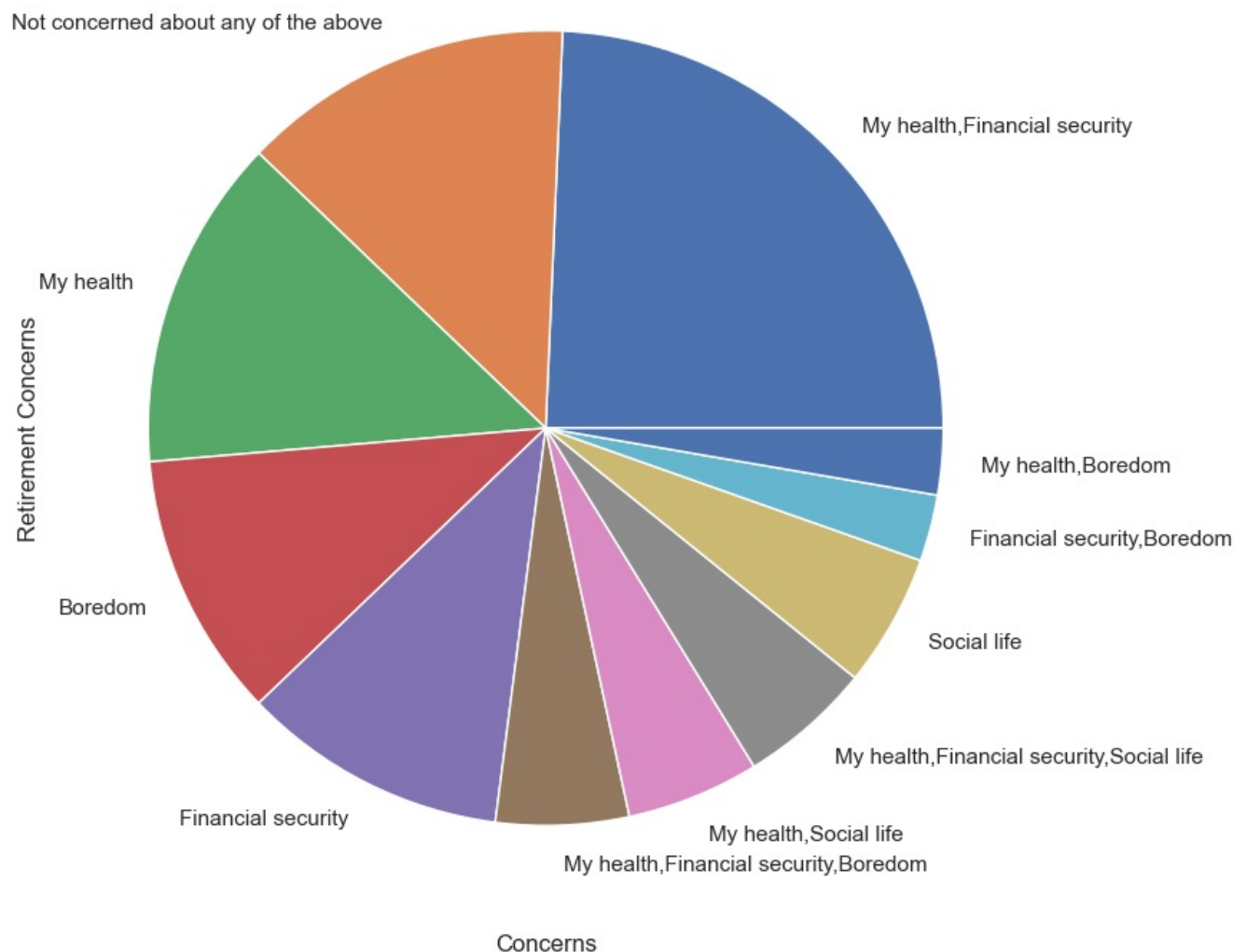
plt.xticks(rotation=0)
plt.show()

# Plot the distribution of concerns
plt.figure(figsize=(13, 9))
concerns_counts.plot(kind='pie')
plt.title('Distribution of Concerns After Retirement')
plt.xlabel('Concerns')
plt.show()
```

Distribution of Expectations After Retirement



Distribution of Concerns After Retirement



The largest expectation after retirement is to “Enjoy my life with family and friends”, which takes up more than half of the chart. The second largest expectation is to “Establish my own business”, which takes up about a quarter of the chart. The other expectations, in decreasing order of size, are “Work full-time somewhere else”, “Volunteer”, and “Work part-time”. The largest concern is “My health, Financial security”, followed by “My health, Financial security, Social life”. The other concerns, in decreasing order of size, are “Financial security”, “My health, Financial security, Boredom”, “My health”, “Boredom”, “My health, Boredom”, and “Not concerned about any of the above”.

```
In [ ]: data['Average Age'] = data['Age'].apply(lambda x: sum(map(int, x.split(' - '))) / 2)
```

```
In [ ]: for column in categorical_columns:
    data[column] = label_encoder.fit_transform(data[column])
# Extract relevant columns
expectations = data['Retirement Expectations']
concerns = data['Retirement Concerns']
retirement_planning = data['Funding Plan']
# Create a DataFrame with the encoded data
df = pd.DataFrame({
    'Expectations': expectations,
    'Concerns': concerns,
    'Retirement_Planning': retirement_planning
})

# Add a constant column for the intercept
df = sm.add_constant(df)

# Perform linear regression
model = sm.OLS(df[['Retirement_Planning']], df[['const', 'Expectations', 'Concerns']]).fit()

# Print the summary of the regression results
print(model.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:      Retirement_Planning    R-squared:                0.054
Model:              OLS                   Adj. R-squared:           -0.002
Method:             Least Squares         F-statistic:              0.9692
Date:               Sat, 02 Sep 2023      Prob (F-statistic):       0.390
Time:               09:35:45              Log-Likelihood:           -60.551
No. Observations:   37                   AIC:                     127.1
Df Residuals:       34                   BIC:                     131.9
Df Model:           2
Covariance Type:    nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	3.0487	0.409	7.459	0.000	2.218	3.879
Expectations	0.1905	0.146	1.304	0.201	-0.106	0.487
Concerns	0.0148	0.071	0.209	0.835	-0.129	0.159

```

=====
Omnibus:                0.497    Durbin-Watson:                1.966
Prob(Omnibus):           0.780    Jarque-Bera (JB):          0.043
Skew:                   -0.008    Prob(JB):                  0.979
Kurtosis:                3.166    Cond. No.:                 11.4
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Based on the analysis, there is insufficient evidence to conclude that Expectations and Concerns after retirement significantly affect individuals' retirement planning decisions. The regression model indicates that neither Expectations nor Concerns after retirement have a statistically significant impact on Retirement_Planning.

4. How did the COVID-19 pandemic affect retirement planning? Was that due to any change in job status

```

In [ ]: import pandas as pd
import statsmodels.api as sm

covid_impact_encoded = data['COVID Impact']

covid_effect_encoded = data['COVID Effect on Income']

# Create a DataFrame with the encoded columns and 'Retirement_Planning'
df = pd.DataFrame({
    'COVID_Impact': covid_impact_encoded,
    'COVID_Effect': covid_effect_encoded,
    'Retirement_Planning': data['Comfort Score']
})

# Add a constant column for the intercept
df = sm.add_constant(df)

# Perform multiple linear regression
model = sm.OLS(df[['Retirement_Planning']], df[['const', 'COVID_Impact', 'COVID_Effect']]).fit()

# Print the summary of the regression results
print(model.summary())

```

OLS Regression Results

Dep. Variable:	Retirement_Planning	R-squared:	0.063
Model:	OLS	Adj. R-squared:	0.008
Method:	Least Squares	F-statistic:	1.149
Date:	Sat, 02 Sep 2023	Prob (F-statistic):	0.329
Time:	09:35:45	Log-Likelihood:	-59.999
No. Observations:	37	AIC:	126.0
Df Residuals:	34	BIC:	130.8
Df Model:	2		
Covariance Type:	nonrobust		

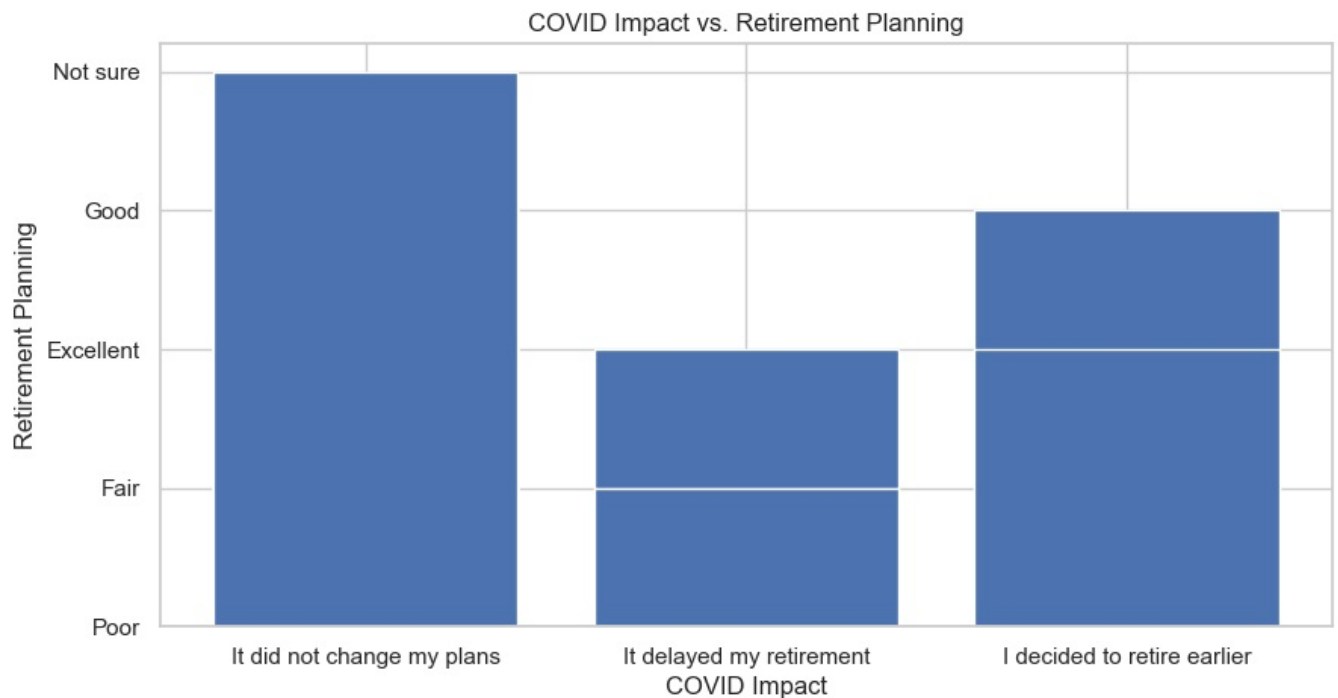
	coef	std err	t	P> t	[0.025	0.975]
const	4.1352	0.765	5.408	0.000	2.581	5.689
COVID_Impact	0.1039	0.358	0.290	0.774	-0.624	0.832
COVID_Effect	-0.3981	0.263	-1.516	0.139	-0.932	0.136

Omnibus:	2.825	Durbin-Watson:	1.782
Prob(Omnibus):	0.244	Jarque-Bera (JB):	2.529
Skew:	-0.558	Prob(JB):	0.282
Kurtosis:	2.373	Cond. No.	11.6

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```
In [ ]: df=pd.read_excel("results-for-retirement-at-2023-08-23-1405.xlsx")
plt.figure(figsize=(10, 5))
plt.bar(df['12. What was the impact of COVID-19 pandemic on your retirement plans?'], df['9. How would you rate
plt.title('COVID Impact vs. Retirement Planning')
plt.xlabel('COVID Impact')
plt.ylabel('Retirement Planning')
plt.show()
```



The linear regression model indicates that the encoded COVID Impact and COVID Effect variables are not statistically significant predictors of retirement planning attitudes. The analysis found that the encoded COVID Impact and COVID Effect variables were not statistically significant predictors of retirement planning attitudes. This means that the impact of COVID-19 on retirement planning attitudes was not clear-cut. Other factors, such as financial situation, health, and family circumstances, may have played a more important role in influencing retirement planning attitudes during the pandemic.

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js