

Project: Analyze Employee Exit Surveys

Introduction

In this project, I'll be working with exit surveys from employees of the Department of Education, Training and Employment (DETE) and the Technical and Further Education (TAFE) institute in Queensland, Australia.

You can find the [DETE exit survey data here](#)

The TAFE survey is no longer available for download.

In this project, we'll play the role of data analyst and pretend our stakeholders want to know the following:

- Are employees who only worked for the institutes for a short period of time resigning due to some kind of dissatisfaction?
- What about employees who have been there longer?
- Are younger employees resigning due to some kind of dissatisfaction? What about older employees?

They want us to combine the results for both surveys to answer these questions.

Lets start with importing and exploring the datasets:

```
In [243... import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use('fivethirtyeight')

dete_survey = pd.read_csv('dete_survey.csv')
tafe_survey = pd.read_csv('tafe_survey.csv')

print (dete_survey.info())
print (dete_survey.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 822 entries, 0 to 821
Data columns (total 56 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   ID                                    822 non-null    int64
1   SeparationType                       822 non-null    object
2   Cease Date                           822 non-null    object
3   DETE Start Date                      822 non-null    object
4   Role Start Date                      822 non-null    object
5   Position                             817 non-null    object
6   Classification                       455 non-null    object
7   Region                              822 non-null    object
8   Business Unit                       126 non-null    object
9   Employment Status                   817 non-null    object
10  Career move to public sector         822 non-null    bool
11  Career move to private sector        822 non-null    bool
12  Interpersonal conflicts               822 non-null    bool
13  Job dissatisfaction                  822 non-null    bool
14  Dissatisfaction with the department  822 non-null    bool
15  Physical work environment            822 non-null    bool
16  Lack of recognition                  822 non-null    bool
17  Lack of job security                 822 non-null    bool
18  Work location                        822 non-null    bool
```

19	Employment conditions	822	non-null	bool
20	Maternity/family	822	non-null	bool
21	Relocation	822	non-null	bool
22	Study/Travel	822	non-null	bool
23	Ill Health	822	non-null	bool
24	Traumatic incident	822	non-null	bool
25	Work life balance	822	non-null	bool
26	Workload	822	non-null	bool
27	None of the above	822	non-null	bool
28	Professional Development	808	non-null	object
29	Opportunities for promotion	735	non-null	object
30	Staff morale	816	non-null	object
31	Workplace issue	788	non-null	object
32	Physical environment	817	non-null	object
33	Worklife balance	815	non-null	object
34	Stress and pressure support	810	non-null	object
35	Performance of supervisor	813	non-null	object
36	Peer support	812	non-null	object
37	Initiative	813	non-null	object
38	Skills	811	non-null	object
39	Coach	767	non-null	object
40	Career Aspirations	746	non-null	object
41	Feedback	792	non-null	object
42	Further PD	768	non-null	object
43	Communication	814	non-null	object
44	My say	812	non-null	object
45	Information	816	non-null	object
46	Kept informed	813	non-null	object
47	Wellness programs	766	non-null	object
48	Health & Safety	793	non-null	object
49	Gender	798	non-null	object
50	Age	811	non-null	object
51	Aboriginal	16	non-null	object
52	Torres Strait	3	non-null	object
53	South Sea	7	non-null	object
54	Disability	23	non-null	object
55	NESB	32	non-null	object

dtypes: bool(18), int64(1), object(37)

memory usage: 258.6+ KB

None

	ID	SeparationType	Cease Date	DETE	Start Date	\
0	1	Ill Health Retirement	08/2012		1984	
1	2	Voluntary Early Retirement (VER)	08/2012		Not Stated	
2	3	Voluntary Early Retirement (VER)	05/2012		2011	
3	4	Resignation-Other reasons	05/2012		2005	
4	5	Age Retirement	05/2012		1970	

	Role	Start Date	Position	\
0		2004	Public Servant	
1	Not Stated		Public Servant	
2		2011	Schools Officer	
3		2006	Teacher	
4		1989	Head of Curriculum/Head of Special Education	

	Classification	Region	Business Unit	\
0	A01-A04	Central Office	Corporate Strategy and Performance	
1	A05-A07	Central Office	Corporate Strategy and Performance	
2	NaN	Central Office	Education Queensland	
3	Primary	Central Queensland	NaN	
4	NaN	South East	NaN	

	Employment Status	Career move to public sector	\
0	Permanent Full-time	True	
1	Permanent Full-time	False	
2	Permanent Full-time	False	
3	Permanent Full-time	False	

4	Permanent Full-time	False				
	Career move to private sector	Interpersonal conflicts			\	
0		False		False		
1		False		False		
2		False		False		
3		True		False		
4		False		False		
	Job dissatisfaction	Dissatisfaction with the department			\	
0		True		False		
1		False		False		
2		False		False		
3		False		False		
4		False		False		
	Physical work environment	Lack of recognition		Lack of job security		\
0		False		True		False
1		False		False		False
2		False		False		False
3		False		False		False
4		False		False		False
	Work location	Employment conditions		Maternity/family		Relocation \
0		False		False		False
1		False		False		False
2		False		False		False
3		False		False		False
4		False		False		False
	Study/Travel	Ill Health	Traumatic incident		Work life balance	Workload \
0	False	False		False	False	False
1	False	False		False	False	False
2	False	False		False	False	False
3	False	False		False	False	False
4	False	False		False	True	False
	None of the above			Professional Development Opportunities for promotion		\
0		True		A		A
1		False		A		A
2		True		N		N
3		False		A		N
4		False		A		A
	Staff morale	Workplace issue		Physical environment		Worklife balance \
0		N		N		A
1		N		N		N
2		N		N		N
3		N		N	A	A
4		N		N	D	D
	Stress and pressure support		Performance of supervisor		Peer support \	
0			A		A	A
1			A		A	A
2			N		N	N
3			N		N	A
4			N		A	A
	Initiative	Skills	Coach	Career Aspirations		Feedback Further PD \
0	N	N	N		A	A
1	N	N	N		A	A
2	N	N	N		N	N
3	A	A	A		A	A
4	A	A	A		A	SA
	Communication My say			Information Kept		Wellness programs \

0	N	A	A	N	N
1	N	A	A	N	N
2	A	A	N	N	N
3	A	A	A	A	N
4	D	D	A	N	A

	Health & Safety	Gender	Age	Aboriginal	Torres Strait	South Sea	\
0	N	Male	56-60	NaN	NaN	NaN	
1	N	Male	56-60	NaN	NaN	NaN	
2	N	Male	61 or older	NaN	NaN	NaN	
3	A	Female	36-40	NaN	NaN	NaN	
4	M	Female	61 or older	NaN	NaN	NaN	

	Disability	NESB
0	NaN	Yes
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

```
In [244.. print(dete_survey.isnull().sum())
```

ID	0
SeparationType	0
Cease Date	0
DETE Start Date	0
Role Start Date	0
Position	5
Classification	367
Region	0
Business Unit	696
Employment Status	5
Career move to public sector	0
Career move to private sector	0
Interpersonal conflicts	0
Job dissatisfaction	0
Dissatisfaction with the department	0
Physical work environment	0
Lack of recognition	0
Lack of job security	0
Work location	0
Employment conditions	0
Maternity/family	0
Relocation	0
Study/Travel	0
Ill Health	0
Traumatic incident	0
Work life balance	0
Workload	0
None of the above	0
Professional Development	14
Opportunities for promotion	87
Staff morale	6
Workplace issue	34
Physical environment	5
Worklife balance	7
Stress and pressure support	12
Performance of supervisor	9
Peer support	10
Initiative	9
Skills	11
Coach	55
Career Aspirations	76
Feedback	30
Further PD	54
Communication	8

My say	10
Information	6
Kept informed	9
Wellness programs	56
Health & Safety	29
Gender	24
Age	11
Aboriginal	806
Torres Strait	819
South Sea	815
Disability	799
NESB	790

dtype: int64

```
In [245... print (dete_survey['Aboriginal'].value_counts())
print ()
print (dete_survey['Business Unit'].value_counts())
```

Yes 16
Name: Aboriginal, dtype: int64

Education Queensland	54
Information and Technologies	26
Training and Tertiary Education Queensland	12
Other	11
Human Resources	6
Corporate Strategy and Performance	5
Early Childhood Education and Care	3
Policy, Research, Legislation	2
Infrastructure	2
Corporate Procurement	1
Finance	1
Pacific Pines SHS	1
Calliope State School	1
Indigenous Education and Training Futures	1

Name: Business Unit, dtype: int64

```
In [246... print (dete_survey['Classification'].value_counts())
```

Primary	161
Secondary	124
A01-A04	66
A05-A07	46
Special Education	33
A08 and Above	14
P01-P04	8
Middle	3

Name: Classification, dtype: int64

Notes:

As we can see the Dete Survey data is broken up into 53 columns and we have 822 entries. Looking at the aboriginal column I can see not many in the survey identify as an aboriginal Australian or just a lot of this data is just missing.

There is also a lot of missing information in the classification type which is a system categorizing where the employee worked. If in primary school, special education, administration, etc.

Let's take a look at the tafe survey.

Exploring Tafe Survey:

```
In [247... tafe_survey.info()
```

```
tafe_survey.head()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 702 entries, 0 to 701  
Data columns (total 72 columns):  
#    Column
```

		Non-Null Cou
nt	Dtype	
---	-----	-----
0	Record ID	702 non-null
	float64	
1	Institute	702 non-null
	object	
2	WorkArea	702 non-null
	object	
3	CESSATION YEAR	695 non-null
	float64	
4	Reason for ceasing employment	701 non-null
	object	
5	Contributing Factors. Career Move - Public Sector	437 non-null
	object	
6	Contributing Factors. Career Move - Private Sector	437 non-null
	object	
7	Contributing Factors. Career Move - Self-employment	437 non-null
	object	
8	Contributing Factors. Ill Health	437 non-null
	object	
9	Contributing Factors. Maternity/Family	437 non-null
	object	
10	Contributing Factors. Dissatisfaction	437 non-null
	object	
11	Contributing Factors. Job Dissatisfaction	437 non-null
	object	
12	Contributing Factors. Interpersonal Conflict	437 non-null
	object	
13	Contributing Factors. Study	437 non-null
	object	
14	Contributing Factors. Travel	437 non-null
	object	
15	Contributing Factors. Other	437 non-null
	object	
16	Contributing Factors. NONE	437 non-null
	object	
17	Main Factor. Which of these was the main factor for leaving?	113 non-null
	object	
18	InstituteViews. Topic:1. I feel the senior leadership had a clear vision and direct	

ion 608 non-null
 object
 19 InstituteViews. Topic:2. I was given access to skills training to help me do my job
 better 613 non-null
 object
 20 InstituteViews. Topic:3. I was given adequate opportunities for personal developmen
 t 610 non-null
 object
 21 InstituteViews. Topic:4. I was given adequate opportunities for promotion within %I
 nstitute]Q25LBL% 608 non-null
 object
 22 InstituteViews. Topic:5. I felt the salary for the job was right for the responsibi
 lities I had 615 non-null
 object
 23 InstituteViews. Topic:6. The organisation recognised when staff did good work
 607 non-null
 object
 24 InstituteViews. Topic:7. Management was generally supportive of me 614 non-null
 object
 25 InstituteViews. Topic:8. Management was generally supportive of my team 608 non-null
 object
 26 InstituteViews. Topic:9. I was kept informed of the changes in the organisation whi
 ch would affect me 610 non-null
 object
 27 InstituteViews. Topic:10. Staff morale was positive within the Institute 602 non-null
 object
 28 InstituteViews. Topic:11. If I had a workplace issue it was dealt with quickly
 601 non-null
 object
 29 InstituteViews. Topic:12. If I had a workplace issue it was dealt with efficiently
 597 non-null
 object
 30 InstituteViews. Topic:13. If I had a workplace issue it was dealt with discreetly
 601 non-null
 object
 31 WorkUnitViews. Topic:14. I was satisfied with the quality of the management and sup
 ervision within my work unit 609 non-null
 object
 32 WorkUnitViews. Topic:15. I worked well with my colleagues 605 non-null
 object
 33 WorkUnitViews. Topic:16. My job was challenging and interesting 607 non-null
 object
 34 WorkUnitViews. Topic:17. I was encouraged to use my initiative in the course of my
 work 610 non-null
 object
 35 WorkUnitViews. Topic:18. I had sufficient contact with other people in my job
 613 non-null
 object
 36 WorkUnitViews. Topic:19. I was given adequate support and co-operation by my peers
 to enable me to do my job 609 non-null
 object
 37 WorkUnitViews. Topic:20. I was able to use the full range of my skills in my job
 609 non-null
 object
 38 WorkUnitViews. Topic:21. I was able to use the full range of my abilities in my jo
 b. ; Category:Level of Agreement; Question:YOUR VIEWS ABOUT YOUR WORK UNIT] 608 non-nul
 l object
 39 WorkUnitViews. Topic:22. I was able to use the full range of my knowledge in my job
 608 non-null
 object
 40 WorkUnitViews. Topic:23. My job provided sufficient variety

	object	611 non-null
41	WorkUnitViews. Topic:24. I was able to cope with the level of stress and pressure in my job	610 non-null
	object	
42	WorkUnitViews. Topic:25. My job allowed me to balance the demands of work and family to my satisfaction	611 non-null
	object	
43	WorkUnitViews. Topic:26. My supervisor gave me adequate personal recognition and feedback on my performance	606 non-null
	object	
44	WorkUnitViews. Topic:27. My working environment was satisfactory e.g. sufficient space, good lighting, suitable seating and working area	610 non-null
	object	
45	WorkUnitViews. Topic:28. I was given the opportunity to mentor and coach others in order for me to pass on my skills and knowledge prior to my cessation date	609 non-null
	object	
46	WorkUnitViews. Topic:29. There was adequate communication between staff in my unit	603 non-null
	object	
47	WorkUnitViews. Topic:30. Staff morale was positive within my work unit	606 non-null
	object	
48	Induction. Did you undertake Workplace Induction?	619 non-null
	object	
49	InductionInfo. Topic:Did you undertake a Corporate Induction?	432 non-null
	object	
50	InductionInfo. Topic:Did you undertake a Institute Induction?	483 non-null
	object	
51	InductionInfo. Topic: Did you undertake Team Induction?	440 non-null
	object	
52	InductionInfo. Face to Face Topic:Did you undertake a Corporate Induction; Category:How it was conducted?	555 non-null
1	object	
53	InductionInfo. On-line Topic:Did you undertake a Corporate Induction; Category:How it was conducted?	555 non-null
	object	
54	InductionInfo. Induction Manual Topic:Did you undertake a Corporate Induction?	555 non-null
	object	
55	InductionInfo. Face to Face Topic:Did you undertake a Institute Induction?	530 non-null
	object	
56	InductionInfo. On-line Topic:Did you undertake a Institute Induction?	555 non-null
	object	
57	InductionInfo. Induction Manual Topic:Did you undertake a Institute Induction?	553 non-null
	object	
58	InductionInfo. Face to Face Topic: Did you undertake Team Induction; Category?	555 non-null
	object	
59	InductionInfo. On-line Topic: Did you undertake Team Induction?process you undertook and how it was conducted.]	555 non-null
	object	
60	InductionInfo. Induction Manual Topic: Did you undertake Team Induction?	555 non-null
	object	
61	Workplace. Topic:Did you and your Manager develop a Performance and Professional Development Plan (PPDP)?	608 non-null
	object	
62	Workplace. Topic:Does your workplace promote a work culture free from all forms of	


```
unlawful discrimination?
    object
63 Workplace. Topic:Does your workplace promote and practice the principles of employm
ent equity?
    object
64 Workplace. Topic:Does your workplace value the diversity of its employees?
    object
65 Workplace. Topic:Would you recommend the Institute as an employer to others?
    object
66 Gender. What is your Gender?
    object
67 CurrentAge. Current Age
    object
68 Employment Type. Employment Type
    object
69 Classification. Classification
    object
70 LengthofServiceOverall. Overall Length of Service at Institute (in years)
    object
71 LengthofServiceCurrent. Length of Service at current workplace (in years)
    object
dtypes: float64(2), object(70)
memory usage: 395.0+ KB
```

Out[247]:

	Record ID	Institute	WorkArea	CESSATION YEAR	Reason for ceasing employment	Contributing Factors. Career Move - Public Sector	Contributing Factors. Career Move - Private Sector	Contributing Factors. Career Move - Self-employment	Co
0	6.341330e+17	Southern Queensland Institute of TAFE	Non-Delivery (corporate)	2010.0	Contract Expired	NaN	NaN	NaN	
1	6.341337e+17	Mount Isa Institute of TAFE	Non-Delivery (corporate)	2010.0	Retirement	-	-	-	
2	6.341388e+17	Mount Isa Institute of TAFE	Delivery (teaching)	2010.0	Retirement	-	-	-	
3	6.341399e+17	Mount Isa Institute of TAFE	Non-Delivery (corporate)	2010.0	Resignation	-	-	-	
4	6.341466e+17	Southern Queensland Institute of TAFE	Delivery (teaching)	2010.0	Resignation	-	Career Move - Private Sector	-	

```
In [248... pd.set_option('display.max_rows', None)
print (tafe_survey.shape)
print()
print(tafe_survey.isnull().sum())
```

```
(702, 72)
```

```
Record ID
Institute
WorkArea
CESSATION YEAR
Reason for ceasing employment
Contributing Factors. Career Move - Public Sector
Contributing Factors. Career Move - Private Sector
Contributing Factors. Career Move - Self-employment
Contributing Factors. Ill Health
Contributing Factors. Maternity/Family
Contributing Factors. Dissatisfaction
Contributing Factors. Job Dissatisfaction
Contributing Factors. Interpersonal Conflict
Contributing Factors. Study
Contributing Factors. Travel
Contributing Factors. Other
Contributing Factors. NONE
Main Factor. Which of these was the main factor for leaving?
InstituteViews. Topic:1. I feel the senior leadership had a clear vision and direction
InstituteViews. Topic:2. I was given access to skills training to help me do my job better
InstituteViews. Topic:3. I was given adequate opportunities for personal development
InstituteViews. Topic:4. I was given adequate opportunities for promotion within %Institute]Q25LBL%
InstituteViews. Topic:5. I felt the salary for the job was right for the responsibilities I had
InstituteViews. Topic:6. The organisation recognised when staff did good work
InstituteViews. Topic:7. Management was generally supportive of me
InstituteViews. Topic:8. Management was generally supportive of my team
InstituteViews. Topic:9. I was kept informed of the changes in the organisation which would affect me
InstituteViews. Topic:10. Staff morale was positive within the Institute
InstituteViews. Topic:11. If I had a workplace issue it was dealt with quickly
```

InstituteViews. Topic:12. If I had a workplace issue it was dealt with efficiently	105
InstituteViews. Topic:13. If I had a workplace issue it was dealt with discreetly	101
WorkUnitViews. Topic:14. I was satisfied with the quality of the management and supervision within my work unit	93
WorkUnitViews. Topic:15. I worked well with my colleagues	97
WorkUnitViews. Topic:16. My job was challenging and interesting	95
WorkUnitViews. Topic:17. I was encouraged to use my initiative in the course of my work	92
WorkUnitViews. Topic:18. I had sufficient contact with other people in my job	89
WorkUnitViews. Topic:19. I was given adequate support and co-operation by my peers to enable me to do my job	93
WorkUnitViews. Topic:20. I was able to use the full range of my skills in my job	93
WorkUnitViews. Topic:21. I was able to use the full range of my abilities in my job. ; Category:Level of Agreement; Question:YOUR VIEWS ABOUT YOUR WORK UNIT]	94
WorkUnitViews. Topic:22. I was able to use the full range of my knowledge in my job	94
WorkUnitViews. Topic:23. My job provided sufficient variety	91
WorkUnitViews. Topic:24. I was able to cope with the level of stress and pressure in my job	92
WorkUnitViews. Topic:25. My job allowed me to balance the demands of work and family to my satisfaction	91
WorkUnitViews. Topic:26. My supervisor gave me adequate personal recognition and feedback on my performance	96
WorkUnitViews. Topic:27. My working environment was satisfactory e.g. sufficient space, good lighting, suitable seating and working area	92
WorkUnitViews. Topic:28. I was given the opportunity to mentor and coach others in order for me to pass on my skills and knowledge prior to my cessation date	93
WorkUnitViews. Topic:29. There was adequate communication between staff in my unit	99
WorkUnitViews. Topic:30. Staff morale was positive within my work unit	96
Induction. Did you undertake Workplace Induction?	83
InductionInfo. Topic:Did you undertake a Corporate Induction?	270
InductionInfo. Topic:Did you undertake a Institute Induction?	219
InductionInfo. Topic: Did you undertake Team Induction?	262
InductionInfo. Face to Face Topic:Did you undertake a Corporate Induction; Category:How it was conducted?	147
InductionInfo. On-line Topic:Did you undertake a Corporate Induction; Category:How it was conducted?	147
InductionInfo. Induction Manual Topic:Did you undertake a Corporate Induction?	147
InductionInfo. Face to Face Topic:Did you undertake a Institute Induction?	172
InductionInfo. On-line Topic:Did you undertake a Institute Induction?	147
InductionInfo. Induction Manual Topic:Did you undertake a Institute Induction?	149
InductionInfo. Face to Face Topic: Did you undertake Team Induction; Category?	147
InductionInfo. On-line Topic: Did you undertake Team Induction?process you undertook and how it was conducted.]	147
InductionInfo. Induction Manual Topic: Did you undertake Team Induction?	147
Workplace. Topic:Did you and your Manager develop a Performance and Professional Development Plan (PPDP)?	94

```

Workplace. Topic:Does your workplace promote a work culture free from all forms of unlawful discrimination? 108
Workplace. Topic:Does your workplace promote and practice the principles of employment equity? 115
Workplace. Topic:Does your workplace value the diversity of its employees? 116
Workplace. Topic:Would you recommend the Institute as an employer to others? 121
Gender. What is your Gender? 106
CurrentAge. Current Age 106
Employment Type. Employment Type 106
Classification. Classification 106
LengthofServiceOverall. Overall Length of Service at Institute (in years) 106
LengthofServiceCurrent. Length of Service at current workplace (in years) 106
dtype: int64

```

```
In [249... print (tafe_survey.columns)
```

```

Index(['Record ID', 'Institute', 'WorkArea', 'CESSATION YEAR',
      'Reason for ceasing employment',
      'Contributing Factors. Career Move - Public Sector ',
      'Contributing Factors. Career Move - Private Sector ',
      'Contributing Factors. Career Move - Self-employment',
      'Contributing Factors. Ill Health',
      'Contributing Factors. Maternity/Family',
      'Contributing Factors. Dissatisfaction',
      'Contributing Factors. Job Dissatisfaction',
      'Contributing Factors. Interpersonal Conflict',
      'Contributing Factors. Study', 'Contributing Factors. Travel',
      'Contributing Factors. Other', 'Contributing Factors. NONE',
      'Main Factor. Which of these was the main factor for leaving?',
      'InstituteViews. Topic:1. I feel the senior leadership had a clear vision and direction',
      'InstituteViews. Topic:2. I was given access to skills training to help me do my job better',
      'InstituteViews. Topic:3. I was given adequate opportunities for personal development',
      'InstituteViews. Topic:4. I was given adequate opportunities for promotion within %Institute]Q25LBL%',
      'InstituteViews. Topic:5. I felt the salary for the job was right for the responsibilities I had',
      'InstituteViews. Topic:6. The organisation recognised when staff did good work',
      'InstituteViews. Topic:7. Management was generally supportive of me',
      'InstituteViews. Topic:8. Management was generally supportive of my team',
      'InstituteViews. Topic:9. I was kept informed of the changes in the organisation which would affect me',
      'InstituteViews. Topic:10. Staff morale was positive within the Institute',
      'InstituteViews. Topic:11. If I had a workplace issue it was dealt with quickly',
      'InstituteViews. Topic:12. If I had a workplace issue it was dealt with efficiently',
      'InstituteViews. Topic:13. If I had a workplace issue it was dealt with discreetly',
      'WorkUnitViews. Topic:14. I was satisfied with the quality of the management and supervision within my work unit',
      'WorkUnitViews. Topic:15. I worked well with my colleagues',
      'WorkUnitViews. Topic:16. My job was challenging and interesting',
      'WorkUnitViews. Topic:17. I was encouraged to use my initiative in the course of my work',
      'WorkUnitViews. Topic:18. I had sufficient contact with other people in my job',
      'WorkUnitViews. Topic:19. I was given adequate support and co-operation by my peers']
)

```

```

rs to enable me to do my job',
    'WorkUnitViews. Topic:20. I was able to use the full range of my skills in my job',
    'WorkUnitViews. Topic:21. I was able to use the full range of my abilities in my job. ; Category:Level of Agreement; Question:YOUR VIEWS ABOUT YOUR WORK UNIT]',
    'WorkUnitViews. Topic:22. I was able to use the full range of my knowledge in my job',
    'WorkUnitViews. Topic:23. My job provided sufficient variety',
    'WorkUnitViews. Topic:24. I was able to cope with the level of stress and pressure in my job',
    'WorkUnitViews. Topic:25. My job allowed me to balance the demands of work and family to my satisfaction',
    'WorkUnitViews. Topic:26. My supervisor gave me adequate personal recognition and feedback on my performance',
    'WorkUnitViews. Topic:27. My working environment was satisfactory e.g. sufficient space, good lighting, suitable seating and working area',
    'WorkUnitViews. Topic:28. I was given the opportunity to mentor and coach others in order for me to pass on my skills and knowledge prior to my cessation date',
    'WorkUnitViews. Topic:29. There was adequate communication between staff in my unit',
    'WorkUnitViews. Topic:30. Staff morale was positive within my work unit',
    'Induction. Did you undertake Workplace Induction?',
    'InductionInfo. Topic:Did you undertake a Corporate Induction?',
    'InductionInfo. Topic:Did you undertake a Institute Induction?',
    'InductionInfo. Topic: Did you undertake Team Induction?',
    'InductionInfo. Face to Face Topic:Did you undertake a Corporate Induction; Category:How it was conducted?',
    'InductionInfo. On-line Topic:Did you undertake a Corporate Induction; Category:How it was conducted?',
    'InductionInfo. Induction Manual Topic:Did you undertake a Corporate Induction?',
    'InductionInfo. Face to Face Topic:Did you undertake a Institute Induction?',
    'InductionInfo. On-line Topic:Did you undertake a Institute Induction?',
    'InductionInfo. Induction Manual Topic:Did you undertake a Institute Induction?',
    'InductionInfo. Face to Face Topic: Did you undertake Team Induction; Category?',
    'InductionInfo. On-line Topic: Did you undertake Team Induction?process you undertook and how it was conducted.]',
    'InductionInfo. Induction Manual Topic: Did you undertake Team Induction?',
    'Workplace. Topic:Did you and your Manager develop a Performance and Professional Development Plan (PPDP)?',
    'Workplace. Topic:Does your workplace promote a work culture free from all forms of unlawful discrimination?',
    'Workplace. Topic:Does your workplace promote and practice the principles of employment equity?',
    'Workplace. Topic:Does your workplace value the diversity of its employees?',
    'Workplace. Topic:Would you recommend the Institute as an employer to others?',
    'Gender. What is your Gender?', 'CurrentAge. Current Age',
    'Employment Type. Employment Type', 'Classification. Classification',
    'LengthofServiceOverall. Overall Length of Service at Institute (in years)',
    'LengthofServiceCurrent. Length of Service at current workplace (in years)'],
dtype='object')

```

Notes:

The Tafe survey has 72 columns and 702 entries. The Tafe survey is similar to the Dete Survey with similar columns such as employee classification, employment length, age, gender, etc . Also each dataset is very detailed on finding out why employees are leaving their employer.

Identify Missing Values and Drop Unnecessary Columns:

Now that I explored both data sets and familiarized myself with the data. I will start to do some housekeeping and clean the data sets a bit. I will drop the unnecessary columns that I wont be using. I will also convert the non stated values in as NaN.

```
In [250...] dete_survey= pd.read_csv('dete_survey.csv', na_values='Not Stated')
```

```
In [251...] pd.reset_option("display.max_rows")
dete_survey_updated = dete_survey.drop(dete_survey.columns[28:49], axis=1)
print(dete_survey_updated)
```

	ID	SeparationType	Cease Date	DETE Start Date	\
0	1	Ill Health Retirement	08/2012	1984.0	
1	2	Voluntary Early Retirement (VER)	08/2012	NaN	
2	3	Voluntary Early Retirement (VER)	05/2012	2011.0	
3	4	Resignation-Other reasons	05/2012	2005.0	
4	5	Age Retirement	05/2012	1970.0	
..	
817	819	Age Retirement	02/2014	1977.0	
818	820	Age Retirement	01/2014	1980.0	
819	821	Resignation-Move overseas/interstate	01/2014	2009.0	
820	822	Ill Health Retirement	12/2013	2001.0	
821	823	Resignation-Move overseas/interstate	12/2013	NaN	

	Role Start Date	Position	\
0	2004.0	Public Servant	
1	NaN	Public Servant	
2	2011.0	Schools Officer	
3	2006.0	Teacher	
4	1989.0	Head of Curriculum/Head of Special Education	
..	
817	1999.0	Teacher	
818	1980.0	Teacher	
819	2009.0	Public Servant	
820	2009.0	Teacher	
821	NaN	Teacher Aide	

	Classification	Region	\
0	A01-A04	Central Office	
1	AO5-A07	Central Office	
2	NaN	Central Office	
3	Primary	Central Queensland	
4	NaN	South East	
..	
817	Primary	Central Queensland	
818	Secondary	North Coast	
819	A01-A04	Central Office	
820	Secondary	Darling Downs South West	
821	NaN	Metropolitan	

	Business Unit	Employment Status	\
0	Corporate Strategy and Performance	Permanent Full-time	
1	Corporate Strategy and Performance	Permanent Full-time	
2	Education Queensland	Permanent Full-time	
3	NaN	Permanent Full-time	
4	NaN	Permanent Full-time	
..	
817	NaN	Permanent Part-time	
818	NaN	Permanent Full-time	
819	Education Queensland	Permanent Full-time	
820	NaN	Permanent Full-time	
821	NaN	NaN	

	Career move to public sector	Career move to private sector	\
0	True	False	
1	False	False	
2	False	False	
3	False	True	
4	False	False	
..	

817	False	False
818	False	False
819	False	False
820	False	False
821	False	False

	Interpersonal conflicts	Job dissatisfaction \
0	False	True
1	False	False
2	False	False
3	False	False
4	False	False
..
817	False	False
818	False	False
819	False	False
820	False	False
821	False	False

	Dissatisfaction with the department	Physical work environment \
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
..
817	False	False
818	False	False
819	False	False
820	False	False
821	False	False

	Lack of recognition	Lack of job security	Work location \
0	True	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
..
817	False	False	False
818	False	False	False
819	False	False	False
820	False	False	False
821	False	False	False

	Employment conditions	Maternity/family	Relocation	Study/Travel \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	False	False
..
817	False	True	False	False
818	False	False	False	False
819	False	True	True	False
820	False	False	False	False
821	False	False	False	False

	Ill Health	Traumatic incident	Work life balance	Workload \
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	False	False
4	False	False	True	False
..
817	True	False	False	True

818	False	False	False	False
819	False	False	True	False
820	True	False	False	False
821	False	False	False	False

	None of the above	Gender	Age	Aboriginal	Torres Strait	\
0	True	Male	56-60	NaN	NaN	
1	False	Male	56-60	NaN	NaN	
2	True	Male	61 or older	NaN	NaN	
3	False	Female	36-40	NaN	NaN	
4	False	Female	61 or older	NaN	NaN	
..	
817	False	Female	56-60	NaN	NaN	
818	True	Male	51-55	NaN	NaN	
819	False	Female	31-35	NaN	NaN	
820	False	Female	41-45	NaN	NaN	
821	False	NaN	NaN	NaN	NaN	

	South Sea	Disability	NESB
0	NaN	NaN	Yes
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	NaN
4	NaN	NaN	NaN
..
817	NaN	NaN	NaN
818	NaN	NaN	NaN
819	NaN	NaN	NaN
820	NaN	NaN	NaN
821	NaN	NaN	NaN

[822 rows x 35 columns]

```
In [252... tafe_survey_updated = tafe_survey.drop(tafe_survey.columns[17:66], axis=1)
print(tafe_survey_updated)
```

	Record ID	Institute	\
0	6.341330e+17	Southern Queensland Institute of TAFE	
1	6.341337e+17	Mount Isa Institute of TAFE	
2	6.341388e+17	Mount Isa Institute of TAFE	
3	6.341399e+17	Mount Isa Institute of TAFE	
4	6.341466e+17	Southern Queensland Institute of TAFE	
..	
697	6.350668e+17	Barrier Reef Institute of TAFE	
698	6.350677e+17	Southern Queensland Institute of TAFE	
699	6.350704e+17	Tropical North Institute of TAFE	
700	6.350712e+17	Southbank Institute of Technology	
701	6.350730e+17	Tropical North Institute of TAFE	

	WorkArea	CESSATION YEAR	Reason for ceasing employment	\
0	Non-Delivery (corporate)	2010.0	Contract Expired	
1	Non-Delivery (corporate)	2010.0	Retirement	
2	Delivery (teaching)	2010.0	Retirement	
3	Non-Delivery (corporate)	2010.0	Resignation	
4	Delivery (teaching)	2010.0	Resignation	
..	
697	Delivery (teaching)	2013.0	Resignation	
698	Non-Delivery (corporate)	2013.0	Resignation	
699	Delivery (teaching)	2013.0	Resignation	
700	Non-Delivery (corporate)	2013.0	Contract Expired	
701	Non-Delivery (corporate)	2013.0	Resignation	

	Contributing Factors. Career Move - Public Sector	\
0	NaN	
1	-	
2	-	

3		-
4		-
..		...
697	Career Move - Public Sector	
698	Career Move - Public Sector	
699		-
700		NaN
701		-

	Contributing Factors. Career Move - Private Sector	\
0		NaN
1		-
2		-
3		-
4	Career Move - Private Sector	
..		...
697		-
698		-
699		-
700		NaN
701		-

	Contributing Factors. Career Move - Self-employment	\
0		NaN
1		-
2		-
3		-
4		-
..		...
697		-
698		-
699		-
700		NaN
701	Career Move - Self-employment	

	Contributing Factors. Ill Health	Contributing Factors. Maternity/Family	\
0	NaN		NaN
1	-		-
2	-		-
3	-		-
4	-		-
..
697	-		-
698	-		-
699	-		-
700	NaN		NaN
701	-		-

	Contributing Factors. Dissatisfaction	\
0	NaN	
1	-	
2	-	
3	-	
4	-	
..	...	
697	-	
698	-	
699	-	
700	NaN	
701	-	

	Contributing Factors. Job Dissatisfaction	\
0	NaN	
1	-	
2	-	
3	-	

4	-
..	...
697	-
698	-
699	-
700	NaN
701	-

	Contributing Factors. Interpersonal Conflict	Contributing Factors. Study \
0	NaN	NaN
1	-	-
2	-	-
3	-	-
4	-	-
..
697	-	-
698	-	-
699	-	-
700	NaN	NaN
701	-	-

	Contributing Factors. Travel	Contributing Factors. Other \
0	NaN	NaN
1	Travel	-
2	-	-
3	Travel	-
4	-	-
..
697	-	-
698	-	-
699	-	Other
700	NaN	NaN
701	Travel	-

	Contributing Factors. NONE	Gender. What is your Gender? \
0	NaN	Female
1	-	NaN
2	NONE	NaN
3	-	NaN
4	-	Male
..
697	-	Male
698	-	NaN
699	-	Female
700	NaN	Female
701	-	Female

	CurrentAge. Current Age	Employment Type. Employment Type \
0	26 30	Temporary Full-time
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	41 45	Permanent Full-time
..
697	51-55	Temporary Full-time
698	NaN	NaN
699	51-55	Permanent Full-time
700	41 45	Temporary Full-time
701	26 30	Contract/casual

	Classification. Classification \
0	Administration (AO)
1	NaN
2	NaN
3	NaN
4	Teacher (including LVT)

```

..      ...
697      Teacher (including LVT)
698      NaN
699      Teacher (including LVT)
700      Professional Officer (PO)
701      Administration (AO)

LengthofServiceOverall. Overall Length of Service at Institute (in years) \
0      1-2
1      NaN
2      NaN
3      NaN
4      3-4
..      ...
697      1-2
698      NaN
699      5-6
700      1-2
701      3-4

LengthofServiceCurrent. Length of Service at current workplace (in years)
0      1-2
1      NaN
2      NaN
3      NaN
4      3-4
..      ...
697      1-2
698      NaN
699      1-2
700      1-2
701      1-2

[702 rows x 23 columns]

```

```
In [253...] dete_survey_updated.columns
```

```
Out[253]: Index(['ID', 'SeparationType', 'Cease Date', 'DETE Start Date',
      'Role Start Date', 'Position', 'Classification', 'Region',
      'Business Unit', 'Employment Status', 'Career move to public sector',
      'Career move to private sector', 'Interpersonal conflicts',
      'Job dissatisfaction', 'Dissatisfaction with the department',
      'Physical work environment', 'Lack of recognition',
      'Lack of job security', 'Work location', 'Employment conditions',
      'Maternity/family', 'Relocation', 'Study/Travel', 'Ill Health',
      'Traumatic incident', 'Work life balance', 'Workload',
      'None of the above', 'Gender', 'Age', 'Aboriginal', 'Torres Strait',
      'South Sea', 'Disability', 'NESB'],
      dtype='object')
```

```
In [254...] tafe_survey_updated.columns
```

```
Out[254]: Index(['Record ID', 'Institute', 'WorkArea', 'CESSATION YEAR',
      'Reason for ceasing employment',
      'Contributing Factors. Career Move - Public Sector ',
      'Contributing Factors. Career Move - Private Sector ',
      'Contributing Factors. Career Move - Self-employment',
      'Contributing Factors. Ill Health',
      'Contributing Factors. Maternity/Family',
      'Contributing Factors. Dissatisfaction',
      'Contributing Factors. Job Dissatisfaction',
      'Contributing Factors. Interpersonal Conflict',
      'Contributing Factors. Study', 'Contributing Factors. Travel',
      'Contributing Factors. Other', 'Contributing Factors. NONE',
      'Gender. What is your Gender?', 'CurrentAge. Current Age',
      'Employment Type. Employment Type', 'Classification. Classification',
```

```
'LengthofServiceOverall. Overall Length of Service at Institute (in years)',  
'LengthofServiceCurrent. Length of Service at current workplace (in years)'],  
dtype='object')
```

Notes:

Now that I have removed the columns I dont need we can clean up the column names. The reasoning to clean the column names is that I will eventually combine the 2 datasets. They share alot of the same information but just with different column names. So it will look more uniform and easier to work with when combined.

Clean Column Names:

```
In [255...] dete_survey_updated.columns
```

```
Out[255]: Index(['ID', 'SeparationType', 'Cease Date', 'DETE Start Date',  
      'Role Start Date', 'Position', 'Classification', 'Region',  
      'Business Unit', 'Employment Status', 'Career move to public sector',  
      'Career move to private sector', 'Interpersonal conflicts',  
      'Job dissatisfaction', 'Dissatisfaction with the department',  
      'Physical work environment', 'Lack of recognition',  
      'Lack of job security', 'Work location', 'Employment conditions',  
      'Maternity/family', 'Relocation', 'Study/Travel', 'Ill Health',  
      'Traumatic incident', 'Work life balance', 'Workload',  
      'None of the above', 'Gender', 'Age', 'Aboriginal', 'Torres Strait',  
      'South Sea', 'Disability', 'NESB'],  
      dtype='object')
```

```
In [256...] dete_survey_updated.columns = dete_survey_updated.columns.str.strip().str.replace(' ','_')
```

```
In [257...] dete_survey_updated.columns
```

```
Out[257]: Index(['id', 'separationtype', 'cease_date', 'dete_start_date',  
      'role_start_date', 'position', 'classification', 'region',  
      'business_unit', 'employment_status', 'career_move_to_public_sector',  
      'career_move_to_private_sector', 'interpersonal_conflicts',  
      'job_dissatisfaction', 'dissatisfaction_with_the_department',  
      'physical_work_environment', 'lack_of_recognition',  
      'lack_of_job_security', 'work_location', 'employment_conditions',  
      'maternity/family', 'relocation', 'study/travel', 'ill_health',  
      'traumatic_incident', 'work_life_balance', 'workload',  
      'none_of_the_above', 'gender', 'age', 'aboriginal', 'torres_strait',  
      'south_sea', 'disability', 'nesb'],  
      dtype='object')
```

```
In [258...] tafe_survey_updated.columns
```

```
Out[258]: Index(['Record ID', 'Institute', 'WorkArea', 'CESSATION YEAR',  
      'Reason for ceasing employment',  
      'Contributing Factors. Career Move - Public Sector ',  
      'Contributing Factors. Career Move - Private Sector ',  
      'Contributing Factors. Career Move - Self-employment',  
      'Contributing Factors. Ill Health',  
      'Contributing Factors. Maternity/Family',  
      'Contributing Factors. Dissatisfaction',  
      'Contributing Factors. Job Dissatisfaction',  
      'Contributing Factors. Interpersonal Conflict',  
      'Contributing Factors. Study', 'Contributing Factors. Travel',  
      'Contributing Factors. Other', 'Contributing Factors. NONE',  
      'Gender. What is your Gender?', 'CurrentAge. Current Age',  
      'Employment Type. Employment Type', 'Classification. Classification',  
      'LengthofServiceOverall. Overall Length of Service at Institute (in years)',
```

```
'LengthofServiceCurrent. Length of Service at current workplace (in years)'],
dtype='object')
```

```
In [259... new_columns = {'Record ID': 'id',
                'CESSATION YEAR': 'cease_date',
                'Reason for ceasing employment': 'separationtype',
                'Gender. What is your Gender?': 'gender',
                'CurrentAge. Current Age': 'age',
                'Employment Type. Employment Type': 'employment_status',
                'Classification. Classification': 'position',
                'LengthofServiceOverall. Overall Length of Service at Institute (in years)',
                'LengthofServiceCurrent. Length of Service at current workplace (in years)'}

tafe_survey_updated = tafe_survey_updated.rename(new_columns, axis=1)

print(tafe_survey_updated.columns)
print(tafe_survey_updated.shape)

Index(['id', 'Institute', 'WorkArea', 'cease_date', 'separationtype',
       'Contributing Factors. Career Move - Public Sector ',
       'Contributing Factors. Career Move - Private Sector ',
       'Contributing Factors. Career Move - Self-employment',
       'Contributing Factors. Ill Health',
       'Contributing Factors. Maternity/Family',
       'Contributing Factors. Dissatisfaction',
       'Contributing Factors. Job Dissatisfaction',
       'Contributing Factors. Interpersonal Conflict',
       'Contributing Factors. Study', 'Contributing Factors. Travel',
       'Contributing Factors. Other', 'Contributing Factors. NONE', 'gender',
       'age', 'employment_status', 'position', 'institute_service',
       'role_service'],
      dtype='object')
(702, 23)
```

Filter the Data:

Now that we cleaned the column names. We will filter the data on separation type for each dataset. One of our main questions is to see why people are resigning? We are not really interested in retirements.

```
In [260... dete_survey_updated['separationtype'].value_counts()
```

```
Out[260]: Age Retirement                285
Resignation-Other reasons             150
Resignation-Other employer             91
Resignation-Move overseas/interstate   70
Voluntary Early Retirement (VER)       67
Ill Health Retirement                 61
Other                                 49
Contract Expired                      34
Termination                           15
Name: separationtype, dtype: int64
```

```
In [261... tafe_survey_updated['separationtype'].value_counts()
```

```
Out[261]: Resignation                340
Contract Expired                    127
Retrenchment/ Redundancy            104
Retirement                         82
Transfer                            25
Termination                         23
Name: separationtype, dtype: int64
```

```
In [262... dete_index = dete_survey_updated['separationtype'].str.contains('Resignation')
dete_resignations= dete_survey_updated[dete_index].copy()
```

```
print(dete_resignations['separationtype'].value_counts())
print(dete_resignations['separationtype'].value_counts().sum())
```

```
Resignation-Other reasons          150
Resignation-Other employer         91
Resignation-Move overseas/interstate 70
Name: separationtype, dtype: int64
311
```

```
In [263... tafe_index = tafe_survey_updated['separationtype']=='Resignation'
tafe_resignations= tafe_survey_updated[tafe_index].copy()

print(tafe_resignations['separationtype'].value_counts())
```

```
Resignation      340
Name: separationtype, dtype: int64
```

```
In [264... dete_other = dete_survey_updated['separationtype']=='Other'
dete_other_resignations= dete_survey_updated[dete_other].copy()

pd.set_option('display.max_columns', None)
dete_other_resignations
```

Out[264]:									
	id	separationtype	cease_date	dete_start_date	role_start_date	position	classification	region	
38	39	Other	2012	1998.0	1998.0	Teacher Aide	NaN	Metropolitan	
52	53	Other	2012	1982.0	1983.0	Teacher	Secondary	South East	
66	67	Other	2012	1969.0	2007.0	Teacher	Secondary	NaN	
67	68	Other	2012	NaN	NaN	Teacher Aide	NaN	Central Office	
149	150	Other	2012	2008.0	2008.0	Public Servant	A01-A04	Central Office	
152	153	Other	2012	1993.0	2012.0	Public Servant	AO5-A07	Central Office	
156	157	Other	2012	2007.0	2010.0	Technical Officer	NaN	Central Office	
157	158	Other	2012	2009.0	2011.0	Public Servant	AO5-A07	Central Office	
162	163	Other	2012	2007.0	2008.0	Public Servant	A01-A04	Metropolitan	
166	167	Other	2012	1997.0	2005.0	Public Servant	AO5-A07	Central Office	
167	168	Other	2012	2004.0	2010.0	Public Servant	AO5-A07	Central Office	
168	169	Other	2012	2008.0	2008.0	Public Servant	AO5-A07	Central	

								Office
170	171	Other	2012	2009.0	2009.0	Public Servant	AO8 and Above	Central Office
172	173	Other	2012	1975.0	2008.0	Public Servant	A01-A04	Central Office
173	174	Other	2012	2002.0	2007.0	Public Servant	AO8 and Above	Central Office
182	183	Other	2012	2010.0	2010.0	Professional Officer	PO1-PO4	Metropolitan
184	185	Other	2012	2006.0	2006.0	Public Servant	A01-A04	Central Office
186	187	Other	2012	1992.0	2007.0	Public Servant	AO5-A07	Central Office
188	189	Other	2012	2006.0	2009.0	Public Servant	AO5-A07	Central Office
191	192	Other	2012	1995.0	2009.0	Public Servant	A01-A04	Central Office
203	204	Other	2012	1974.0	1995.0	Other	NaN	South East
213	214	Other	2012	NaN	NaN	Cleaner	NaN	Central Queensland
232	233	Other	2012	1979.0	2005.0	Public Servant	A01-A04	Central Office
275	276	Other	2012	2001.0	2005.0	Public Servant	AO5-A07	Central Office
298	299	Other	2013	2012.0	2012.0	Teacher	Primary	Central Queensland
315	316	Other	2012	NaN	NaN	Cleaner	NaN	South East
336	337	Other	2012	NaN	NaN	School Administrative Staff	NaN	NaN
353	354	Other	2012	2004.0	2011.0	Teacher Aide	NaN	NaN
364	365	Other	2012	NaN	NaN	Teacher	Primary	NaN
393	394	Other	2012	2008.0	2008.0	Teacher	Secondary	North Queensland
424	425	Other	2013	2000.0	2006.0	Teacher Aide	NaN	NaN
425	426	Other	2013	2009.0	2009.0	Teacher Aide	NaN	NaN
426	427	Other	2013	2010.0	2012.0	Teacher Aide	NaN	Darling Downs South West
430	431	Other	2012	2003.0	2004.0	Teacher Aide	NaN	Metropolitan

455	456	Other	2013	2009.0	2009.0	Teacher	Secondary	Central Queensland
469	470	Other	2013	2010.0	2013.0	Teacher	Primary	Darling Downs South West
493	494	Other	2013	1996.0	1996.0	Teacher Aide	NaN	NaN
510	511	Other	2013	2013.0	2013.0	Other	NaN	South East
546	547	Other	2013	2009.0	2009.0	Cleaner	NaN	Darling Downs South West
570	571	Other	06/2013	2005.0	2005.0	Teacher Aide	NaN	Central Queensland
640	641	Other	08/2013	2007.0	2007.0	Teacher Aide	NaN	North Coast
648	649	Other	10/2013	2010.0	2010.0	Cleaner	NaN	NaN
656	657	Other	10/2013	2011.0	2011.0	Teacher Aide	NaN	NaN
702	704	Other	09/2013	1998.0	2006.0	Cleaner	NaN	Far North Queensland
710	712	Other	12/2013	1990.0	1990.0	Teacher Aide	NaN	South East
761	763	Other	12/2013	NaN	NaN	Teacher	Secondary	Darling Downs South West
792	794	Other	NaN	1996.0	2006.0	Teacher Aide	NaN	Metropolitan
799	801	Other	NaN	1981.0	2003.0	Teacher	Special Education	North Coast
800	802	Other	01/2014	2011.0	2011.0	Teacher	Primary	NaN

```
In [265... pd.reset_option("display.max_rows")
```

Verify the Data:

I will look at few columns to make sure the data looks good and that there are no major issues. I specifically look at 'cease_date', 'dete_start_date', role_start_date, 'age'.

```
In [266... dete_resignations.columns
```

```
Out[266]: Index(['id', 'separationtype', 'cease_date', 'dete_start_date',
      'role_start_date', 'position', 'classification', 'region',
      'business_unit', 'employment_status', 'career_move_to_public_sector',
      'career_move_to_private_sector', 'interpersonal_conflicts',
      'job_dissatisfaction', 'dissatisfaction_with_the_department',
      'physical_work_environment', 'lack_of_recognition',
      'lack_of_job_security', 'work_location', 'employment_conditions',
```



```
'maternity/family', 'relocation', 'study/travel', 'ill_health',  
'traumatic_incident', 'work_life_balance', 'workload',  
'none_of_the_above', 'gender', 'age', 'aboriginal', 'torres_strait',  
'south_sea', 'disability', 'nesb'],  
dtype='object')
```

```
In [267]: dete_resignations[['dete_start_date', 'role_start_date', 'cease_date', 'age']].head(20)
```

Out[267]:	dete_start_date	role_start_date	cease_date	age
3	2005.0	2006.0	05/2012	36-40
5	1994.0	1997.0	05/2012	41-45
8	2009.0	2009.0	07/2012	31-35
9	1997.0	2008.0	2012	46-50
11	2009.0	2009.0	2012	31-35
12	1998.0	1998.0	2012	36-40
14	2007.0	2010.0	2012	31-35
16	NaN	NaN	2012	61 or older
20	1982.0	1982.0	2012	56-60
21	1980.0	2009.0	2012	51-55
22	1997.0	1998.0	2012	46-50
23	1973.0	2012.0	2012	61 or older
25	1995.0	2002.0	2012	41-45
27	2005.0	2011.0	2012	21-25
33	2003.0	2003.0	2012	36-40
34	2006.0	2009.0	2012	61 or older
37	2011.0	2011.0	2012	21-25
39	NaN	NaN	2012	21-25
40	1977.0	1980.0	2012	56-60
41	1974.0	1994.0	2012	51-55

```
In [268]: dete_resignations[['dete_start_date', 'role_start_date', 'cease_date', 'age']].tail(20)
```

Out[268]:	dete_start_date	role_start_date	cease_date	age
774	2005.0	2005.0	NaN	41-45
784	2013.0	2013.0	10/2013	21-25
786	1994.0	2008.0	01/2014	41-45
788	1990.0	2010.0	NaN	41-45
789	1983.0	1999.0	01/2014	51-55
790	2008.0	2008.0	01/2014	36-40
791	2007.0	2007.0	NaN	46-50
794	NaN	NaN	11/2013	36-40
797	2000.0	2013.0	NaN	36-40

798	1995.0	NaN	NaN	36-40
802	NaN	NaN	12/2013	NaN
803	2004.0	2007.0	01/2014	46-50
804	2008.0	2013.0	01/2014	26-30
806	2005.0	2005.0	12/2013	41-45
807	2004.0	2004.0	12/2013	36-40
808	2010.0	2010.0	12/2013	26-30
815	2012.0	2012.0	01/2014	21-25
816	2012.0	2012.0	01/2014	21-25
819	2009.0	2009.0	01/2014	31-35
821	NaN	NaN	12/2013	NaN

Notes:

1) Preliminary findings so far show that the dete start date doesnt always align with the role start date. This could be the fact employees transfer to another position during their tenure.

2) The cease date has 2 different date formats. Month and year and just year.

3) The age column has a 4 year range but once it gets to the age of 61 it is entered at 61 and older

```
In [269]: dete_resignations['dete_start_date'].value_counts().sort_index()
```

```
Out[269]: 1963.0    1
1971.0    1
1972.0    1
1973.0    1
1974.0    2
1975.0    1
1976.0    2
1977.0    1
1980.0    5
1982.0    1
1983.0    2
1984.0    1
1985.0    3
1986.0    3
1987.0    1
1988.0    4
1989.0    4
1990.0    5
1991.0    4
1992.0    6
1993.0    5
1994.0    6
1995.0    4
1996.0    6
1997.0    5
1998.0    6
1999.0    8
2000.0    9
2001.0    3
2002.0    6
2003.0    6
2004.0   14
```

```

2005.0    15
2006.0    13
2007.0    21
2008.0    22
2009.0    13
2010.0    17
2011.0    24
2012.0    21
2013.0    10
Name: dete_start_date, dtype: int64

```

```
In [270...] dete_resignations['cease_date'].value_counts().sort_index()
```

```

Out[270]: 01/2014    22
          05/2012     2
          05/2013     2
          06/2013    14
          07/2006     1
          07/2012     1
          07/2013     9
          08/2013     4
          09/2010     1
          09/2013    11
          10/2013     6
          11/2013     9
          12/2013    17
          2010        1
          2012     126
          2013        74
Name: cease_date, dtype: int64

```

```
In [271...] dete_resignations['age'].value_counts().sort_index()
```

```

Out[271]: 20 or younger    1
          21-25           29
          26-30           35
          31-35           29
          36-40           41
          41-45           48
          46-50           42
          51-55           32
          56-60           26
          61 or older     23
Name: age, dtype: int64

```

Notes:

As I can see from further review the start date column is just in year. The cease date is both year and months and year. I have confirmed age has 2 data inputs outside of a range. Before I make any changes lets explore the tafe dataframe.

```
In [272...] tafe_resignations.columns
```

```

Out[272]: Index(['id', 'Institute', 'WorkArea', 'cease_date', 'separationtype',
      'Contributing Factors. Career Move - Public Sector ',
      'Contributing Factors. Career Move - Private Sector ',
      'Contributing Factors. Career Move - Self-employment',
      'Contributing Factors. Ill Health',
      'Contributing Factors. Maternity/Family',
      'Contributing Factors. Dissatisfaction',
      'Contributing Factors. Job Dissatisfaction',
      'Contributing Factors. Interpersonal Conflict',
      'Contributing Factors. Study', 'Contributing Factors. Travel',
      'Contributing Factors. Other', 'Contributing Factors. NONE', 'gender',
      'age', 'employment_status', 'position', 'institute_service',

```

```
'role_service'],  
dtype='object')
```

lets explore the 'cease_date', 'age', columns in tafe. Tafe doesnt have a start date to work with.

```
In [273]: tafe_resignations['cease_date'].astype(float)  
tafe_resignations['cease_date'].value_counts().sort_index()
```

```
Out[273]: 2009.0      2  
2010.0     68  
2011.0    116  
2012.0     94  
2013.0     55  
Name: cease_date, dtype: int64
```

```
In [274]: tafe_resignations['age'].value_counts().sort_index()
```

```
Out[274]: 20 or younger      9  
21  25      33  
26  30      32  
31  35      32  
36  40      32  
41  45      45  
46  50      39  
51-55      39  
56 or older      29  
Name: age, dtype: int64
```

Notes:

As we can see the tafe cease date follows just the year format. While the age range is similar to dete it is missing the dash. What I will do now is reduce the dete cease date range to just the year so it matches the format for start dates and the tafe cease date. There is no start date month to work with so it wouldnt make sense to add an arbitrary month.

The age range is not required but for my own curiosity I will bring this into a matching range so it can be combined. I would like to see if I can identify any patterns on the age range.

Cleaning Cease Date and Age Columns:

```
In [275]: dete_resignations['cease_date'] = dete_resignations['cease_date'].str.split('/').str.get  
det_resignations['cease_date'] = dete_resignations['cease_date'].astype('float')  
det_resignations['cease_date'].value_counts()
```

```
Out[275]: 2013.0     146  
2012.0     129  
2014.0      22  
2010.0       2  
2006.0       1  
Name: cease_date, dtype: int64
```

```
In [276]: dete_resignations['age'] = dete_resignations['age'].str.replace('20 or younger','16-20')  
det_resignations['age'].value_counts()
```

```
Out[276]: 56-75      49  
41-45      48  
46-50      42  
36-40      41  
26-30      35  
51-55      32  
31-35      29  
21-25      29
```

```
16-20      1
Name: age, dtype: int64
```

```
In [277]: tafe_resignations['age'] = tafe_resignations['age'].str.replace('20 or younger', '16-20')
tafe_resignations['age'].value_counts()
```

```
Out[277]: 41  45      45
46  50      39
51-55      39
21  25      33
36  40      32
31  35      32
26  30      32
56-75      29
16-20       9
Name: age, dtype: int64
```

```
In [278]: tafe_resignations['age'] = tafe_resignations['age'].str.replace('41  45', '41-45').str.re
tafe_resignations['age'].value_counts()
```

```
Out[278]: 41-45      45
46-50      39
51-55      39
21-25      33
36-40      32
31-35      32
26-30      32
56-75      29
16-20       9
Name: age, dtype: int64
```

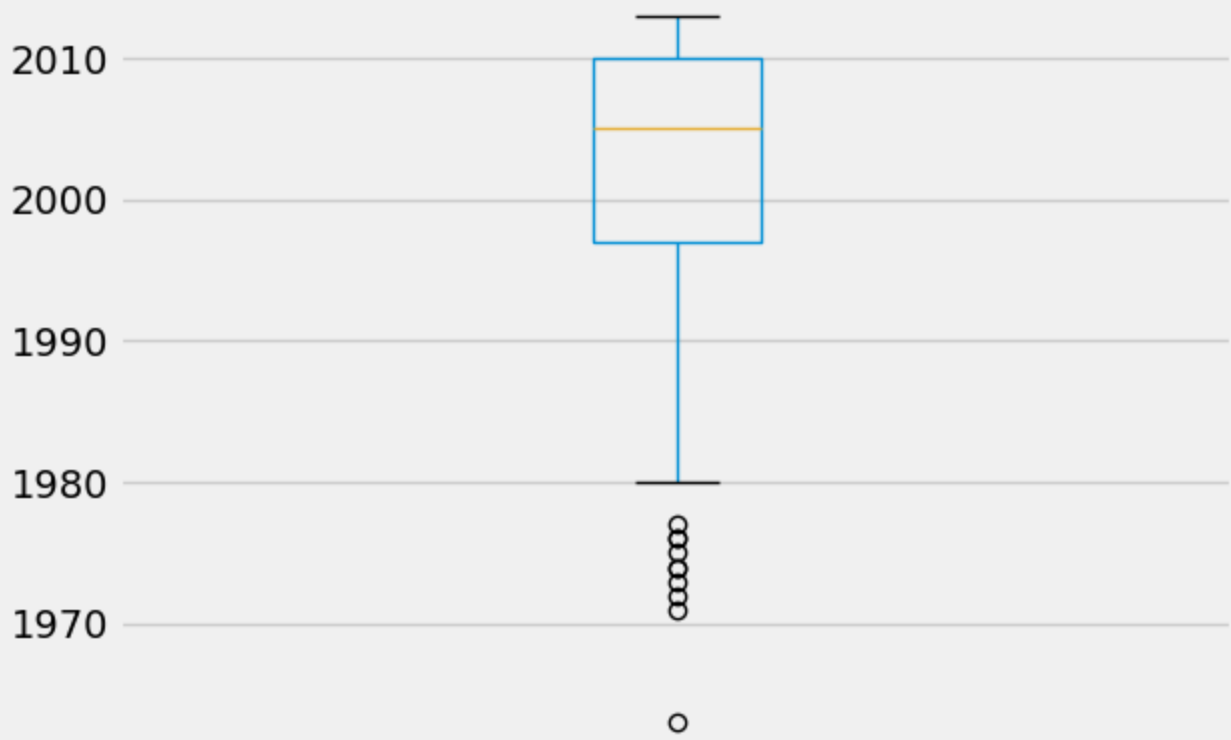
Notes:

Now I have the cease_date's and the start dates matching. I also have the age ranges matching. I changed the 20 or younger to 16-20. I did this to stay in the 5 year range as the other columns. Even though you can work in Australia at the age of 15 I dont foresee this impacting any results since it only accounts for 10 total. Also i made the last range larger in dete to match tafe. Tafe only had a 56 or older range. While dete had 56-60 and 61 and older. I combined the 2 dete ranges and made them both 56-75.

Lets now view the cease date in a box plot.

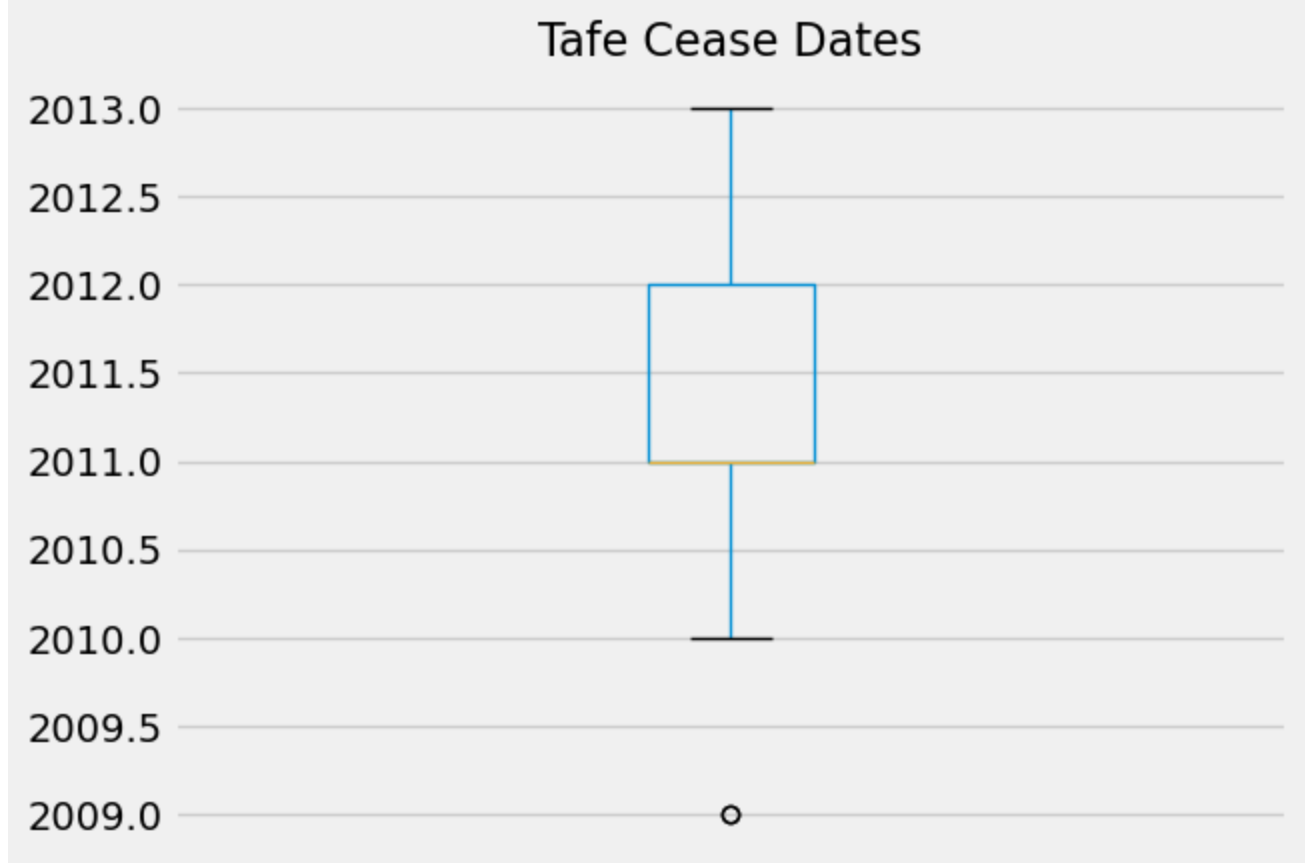
```
In [279]: %matplotlib inline
dete_resignations.boxplot(column='dete_start_date')
plt.title("Dete Start Dates", fontsize=16)
plt.xticks([])
plt.show()
dete_resignations.boxplot(column='cease_date')
plt.title("Dete Cease Dates", fontsize=16)
plt.xticks([])
plt.show()
tafe_resignations.boxplot(column='cease_date')
plt.title("Tafe Cease Dates", fontsize=16)
plt.xticks([])
plt.show()
```

Dete Start Dates



Dete Cease Dates





It seems a lot of the start dates according to the Dete dataset are after 1980 with a cease date mostly between 2012 and 2014. The Tafe dataset shows a lot of cease dates between 2010 and 2013.

Creating a New Column (institute_service):

In the Human Resources field, the length of time an employee spent in a workplace is referred to as their years of service. Tafe_resignations dataframe already contains a "service" column, which we renamed to institute_service. In order to analyze both surveys together, we'll have to create a corresponding institute_service column in dete_resignations. As dete dataframe does not have a length of time spent at the employer.

```
In [280]: dete_resignations['institute_service'] = dete_resignations['cease_date'] - dete_resignations['start_date']
det_resignations['institute_service'].value_counts()
```

```
Out[280]:
```

5.0	23
1.0	22
3.0	20
0.0	20
6.0	17
4.0	16
2.0	14
9.0	14
7.0	13
13.0	8
8.0	8
20.0	7
15.0	7
17.0	6
10.0	6
22.0	6
12.0	6
14.0	6
18.0	5
16.0	5

```

24.0    4
23.0    4
11.0    4
32.0    3
39.0    3
21.0    3
19.0    3
25.0    2
28.0    2
26.0    2
30.0    2
36.0    2
33.0    1
35.0    1
49.0    1
38.0    1
41.0    1
27.0    1
42.0    1
29.0    1
34.0    1
31.0    1
Name: institute_service, dtype: int64

```

Identifying Dissatisfied Employees:

One of the main questions of the project is to identify if employees are leaving due to some dissatisfaction with their employer. Each dataset has columns that correspond to reasons for resigning.

In dete_resignations I will use the following reasons for dissatisfied employees:

'interpersonal_conflicts','job_dissatisfaction', 'dissatisfaction_with_the_department',
 'physical_work_environment', 'lack_of_recognition','lack_of_job_security', 'work_location',
 'employment_conditions','work_life_balance', 'workload',

In tafe_resignations I will use the following reasons for dissatisfied employees:

'Contributing Factors. Dissatisfaction',
 'Contributing Factors. Job Dissatisfaction',
 'Contributing Factors. Interpersonal Conflict',

I will use these columns to create a new column called **dissatisfied** which will be a boolean of True, False, Nan.

```
In [281]: dete_resignations.columns
```

```

Out[281]: Index(['id', 'separationtype', 'cease_date', 'dete_start_date',
      'role_start_date', 'position', 'classification', 'region',
      'business_unit', 'employment_status', 'career_move_to_public_sector',
      'career_move_to_private_sector', 'interpersonal_conflicts',
      'job_dissatisfaction', 'dissatisfaction_with_the_department',
      'physical_work_environment', 'lack_of_recognition',
      'lack_of_job_security', 'work_location', 'employment_conditions',
      'maternity/family', 'relocation', 'study/travel', 'ill_health',
      'traumatic_incident', 'work_life_balance', 'workload',
      'none_of_the_above', 'gender', 'age', 'aboriginal', 'torres_strait',

```



```
        'south_sea', 'disability', 'nesb', 'institute_service'],
        dtype='object')
```

```
In [282... tafe_resignations.columns
```

```
Out[282]: Index(['id', 'Institute', 'WorkArea', 'cease_date', 'separationtype',
        'Contributing Factors. Career Move - Public Sector ',
        'Contributing Factors. Career Move - Private Sector ',
        'Contributing Factors. Career Move - Self-employment',
        'Contributing Factors. Ill Health',
        'Contributing Factors. Maternity/Family',
        'Contributing Factors. Dissatisfaction',
        'Contributing Factors. Job Dissatisfaction',
        'Contributing Factors. Interpersonal Conflict',
        'Contributing Factors. Study', 'Contributing Factors. Travel',
        'Contributing Factors. Other', 'Contributing Factors. NONE', 'gender',
        'age', 'employment_status', 'position', 'institute_service',
        'role_service'],
        dtype='object')
```

```
In [283... tafe_resignations['Contributing Factors. Dissatisfaction'].unique()
```

```
Out[283]: array(['-', 'Contributing Factors. Dissatisfaction ', nan], dtype=object)
```

```
In [284... tafe_resignations['Contributing Factors. Job Dissatisfaction'].unique()
```

```
Out[284]: array(['-', 'Job Dissatisfaction', nan], dtype=object)
```

```
In [285... tafe_resignations['Contributing Factors. Interpersonal Conflict'].unique()
```

```
Out[285]: array(['-', nan, 'Interpersonal Conflict'], dtype=object)
```

```
In [286... print (dete_resignations['interpersonal_conflicts'].unique())
print ()
print(dete_resignations['job_dissatisfaction'].unique())
print ()
print(dete_resignations['workload'].unique())
```

```
[False True]
```

```
[False True]
```

```
[False True]
```

```
In [287... def update_vals(value):
    if pd.isnull(value)==True:
        return np.nan
    elif value == '-':
        return False
    else:
        return True

tafe_resignations['dissatisfied'] = tafe_resignations[['Contributing Factors. Dissatisfac
print(tafe_resignations['dissatisfied'].value_counts())
print(tafe_resignations['dissatisfied'].unique())

dete_resignations['dissatisfied'] = dete_resignations[['interpersonal_conflicts','job_di
'physical_work_environment', 'lack_of_recognition','lack_of_job_security', 'work_locatio
'employment_conditions','work_life_balance', 'workload',]].any(axis=1, skipna=False)

print ()
```

```
print (dete_resignations['dissatisfied'].value_counts())
print (dete_resignations['dissatisfied'].unique())
```

```
False    235
True      105
Name: dissatisfied, dtype: int64
[False  True]

False    157
True     154
Name: dissatisfied, dtype: int64
[False  True]
```

```
In [288... dete_resignations_up = dete_resignations.copy()
tafe_resignations_up = tafe_resignations.copy()
```

Combining the Datasets:

I'll create a column identifying each row from its source dataset with a column called **institute**. Then I will combine the datasets and do some exploration to see how it came out.

```
In [289... dete_resignations_up['institute'] = 'DETE'
tafe_resignations_up['institute'] = 'TAFE'
```

```
In [290... combined = pd.concat([dete_resignations_up, tafe_resignations_up],axis=0, ignore_index=True)

print (combined.shape)
print ()
print (combined.columns)
print ()
print (combined.head())
```

```
(651, 53)
```

```
Index(['id', 'separationtype', 'cease_date', 'dete_start_date',
      'role_start_date', 'position', 'classification', 'region',
      'business_unit', 'employment_status', 'career_move_to_public_sector',
      'career_move_to_private_sector', 'interpersonal_conflicts',
      'job_dissatisfaction', 'dissatisfaction_with_the_department',
      'physical_work_environment', 'lack_of_recognition',
      'lack_of_job_security', 'work_location', 'employment_conditions',
      'maternity/family', 'relocation', 'study/travel', 'ill_health',
      'traumatic_incident', 'work_life_balance', 'workload',
      'none_of_the_above', 'gender', 'age', 'aboriginal', 'torres_strait',
      'south_sea', 'disability', 'nesb', 'institute_service', 'dissatisfied',
      'institute', 'Institute', 'WorkArea',
      'Contributing Factors. Career Move - Public Sector ',
      'Contributing Factors. Career Move - Private Sector ',
      'Contributing Factors. Career Move - Self-employment',
      'Contributing Factors. Ill Health',
      'Contributing Factors. Maternity/Family',
      'Contributing Factors. Dissatisfaction',
      'Contributing Factors. Job Dissatisfaction',
      'Contributing Factors. Interpersonal Conflict',
      'Contributing Factors. Study', 'Contributing Factors. Travel',
      'Contributing Factors. Other', 'Contributing Factors. NONE',
      'role_service'],
      dtype='object')
```

	id	separationtype	cease_date	dete_start_date	\
0	4.0	Resignation-Other reasons	2012.0	2005.0	
1	6.0	Resignation-Other reasons	2012.0	1994.0	
2	9.0	Resignation-Other reasons	2012.0	2009.0	

3	10.0	Resignation-Other employer	2012.0	1997.0
4	12.0	Resignation-Move overseas/interstate	2012.0	2009.0

	role_start_date	position	classification	region	\
0	2006.0	Teacher	Primary	Central Queensland	
1	1997.0	Guidance Officer	NaN	Central Office	
2	2009.0	Teacher	Secondary	North Queensland	
3	2008.0	Teacher Aide	NaN	NaN	
4	2009.0	Teacher	Secondary	Far North Queensland	

	business_unit	employment_status	career_move_to_public_sector	\
0	NaN	Permanent Full-time	False	
1	Education Queensland	Permanent Full-time	False	
2	NaN	Permanent Full-time	False	
3	NaN	Permanent Part-time	False	
4	NaN	Permanent Full-time	False	

	career_move_to_private_sector	interpersonal_conflicts	job_dissatisfaction	\
0	True	False	False	
1	True	False	False	
2	True	False	False	
3	False	True	True	
4	False	False	False	

	dissatisfaction_with_the_department	physical_work_environment	\
0	False	False	
1	False	False	
2	False	False	
3	True	False	
4	False	False	

	lack_of_recognition	lack_of_job_security	work_location	\
0	False	False	False	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	

	employment_conditions	maternity/family	relocation	study/travel	ill_health	\
0	False	False	False	False	False	
1	True	True	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	True	True	False	False	

	traumatic_incident	work_life_balance	workload	none_of_the_above	gender	\
0	False	False	False	False	Female	
1	False	False	False	False	Female	
2	False	False	False	False	Female	
3	False	False	False	False	Female	
4	False	False	False	False	Male	

	age	aboriginal	torres_strait	south_sea	disability	nesb	institute_service	\
0	36-40	NaN	NaN	NaN	NaN	NaN	7.0	
1	41-45	NaN	NaN	NaN	NaN	NaN	18.0	
2	31-35	NaN	NaN	NaN	NaN	NaN	3.0	
3	46-50	NaN	NaN	NaN	NaN	NaN	15.0	
4	31-35	NaN	NaN	NaN	NaN	NaN	3.0	

	dissatisfied	institute	Institute	WorkArea	\
0	False	DETE	NaN	NaN	
1	True	DETE	NaN	NaN	
2	False	DETE	NaN	NaN	
3	True	DETE	NaN	NaN	
4	False	DETE	NaN	NaN	

Contributing Factors. Career Move - Public Sector \		
0		NaN
1		NaN
2		NaN
3		NaN
4		NaN

Contributing Factors. Career Move - Private Sector \		
0		NaN
1		NaN
2		NaN
3		NaN
4		NaN

Contributing Factors. Career Move - Self-employment \		
0		NaN
1		NaN
2		NaN
3		NaN
4		NaN

Contributing Factors. Ill Health Contributing Factors. Maternity/Family \		
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

Contributing Factors. Dissatisfaction \		
0		NaN
1		NaN
2		NaN
3		NaN
4		NaN

Contributing Factors. Job Dissatisfaction \		
0		NaN
1		NaN
2		NaN
3		NaN
4		NaN

Contributing Factors. Interpersonal Conflict Contributing Factors. Study \		
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

Contributing Factors. Travel Contributing Factors. Other \		
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

Contributing Factors. NONE role_service		
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

```
In [291...] print (combined['age'].value_counts())
print (combined['cease_date'].value_counts())
```

```

41-45      93
46-50      81
56-75      78
36-40      73
51-55      71
26-30      67
21-25      62
31-35      61
16-20      10
Name: age, dtype: int64
2012.0      223
2013.0      201
2011.0      116
2010.0       70
2014.0       22
2009.0        2
2006.0        1
Name: cease_date, dtype: int64

```

```
In [292...] combined_updated= combined.dropna(thresh=500, axis=1)
combined_updated.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 651 entries, 0 to 650
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                     651 non-null   float64
1   separationtype         651 non-null   object
2   cease_date            635 non-null   float64
3   position              598 non-null   object
4   employment_status     597 non-null   object
5   gender                592 non-null   object
6   age                   596 non-null   object
7   institute_service     563 non-null   object
8   dissatisfied          651 non-null   bool
9   institute             651 non-null   object
dtypes: bool(1), float64(2), object(7)
memory usage: 46.5+ KB

```

I had a bunch of null data from my original exploration so I decided to clean the dataset to keep the non null that are over 500.

Cleaning the Service Column:

Now I will clean the service column and categorize the institute_service column into these categories:

New: Less than 3 years at a company

Experienced: 3-6 years at a company

Established: 7-10 years at a company

Veteran: 11 or more years at a company

The reasoning behind it is the service column has exact years and ranges. I want to make it more uniform for analysis. Since we have ranges and exact years the best method would be to use ranges and label them by years spent at the employer.

```
In [293...] print(combined_updated['institute_service'].value_counts())
print()
print(combined_updated['institute_service'].shape)
```

Less than 1 year	73
1-2	64
3-4	63
5-6	33
11-20	26
5.0	23
1.0	22
7-10	21
3.0	20
0.0	20
6.0	17
4.0	16
9.0	14
2.0	14
7.0	13
More than 20 years	10
13.0	8
8.0	8
15.0	7
20.0	7
10.0	6
14.0	6
12.0	6
17.0	6
22.0	6
18.0	5
16.0	5
11.0	4
23.0	4
24.0	4
32.0	3
39.0	3
19.0	3
21.0	3
36.0	2
25.0	2
30.0	2
26.0	2
28.0	2
49.0	1
41.0	1
27.0	1
42.0	1
29.0	1
34.0	1
31.0	1
33.0	1
35.0	1
38.0	1

Name: institute_service, dtype: int64

(651,)

```
In [294.. combined_updated['institute_service_up'] = (combined_updated['institute_service'].astype
print(combined_updated['institute_service_up'].value_counts())
print()
print(combined_updated['institute_service_up'].shape)
```

1.0	159
3.0	83
5.0	56
7.0	34
11.0	30
0.0	20
20.0	17

```

6.0      17
4.0      16
9.0      14
2.0      14
13.0     8
8.0      8
15.0     7
17.0     6
10.0     6
22.0     6
14.0     6
12.0     6
18.0     5
16.0     5
24.0     4
23.0     4
21.0     3
19.0     3
39.0     3
32.0     3
25.0     2
28.0     2
26.0     2
36.0     2
30.0     2
34.0     1
27.0     1
29.0     1
42.0     1
33.0     1
41.0     1
35.0     1
49.0     1
38.0     1
31.0     1
Name: institute_service_up, dtype: int64

```

```
(651,)
```

```

C:\Users\marko\AppData\Local\Temp\ipykernel_78696\2285758770.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
combined_updated['institute_service_up'] = (combined_updated['institute_service'].astype('str').str.extract(r"(\d+)").astype('float'))

```

```

In [295... def update_years(value):
    if pd.isnull(value)==True:
        return np.nan
    elif (value < 3):
        return 'New'
    elif (value >= 3 and value < 7):
        return 'Experienced'
    elif (value >=7 and value < 11):
        return 'Established'
    elif (value >= 11):
        return 'Veteran'

combined_updated['service_cat'] = combined_updated['institute_service_up'].apply(update_years)
print (combined_updated['service_cat'].value_counts())

New      193
Experienced  172

```

```

Veteran      136
Established   62
Name: service_cat, dtype: int64
C:\Users\marko\AppData\Local\Temp\ipykernel_78696\152098440.py:13: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
combined_updated['service_cat'] = combined_updated['institute_service_up'].apply(update_years)

```

Performing Analysis:

Now that I finally have the dataset in good shape some analysis can be performed on the data.

```

In [296...] print (combined_updated.columns)

Index(['id', 'separationtype', 'cease_date', 'position', 'employment_status',
      'gender', 'age', 'institute_service', 'dissatisfied', 'institute',
      'institute_service_up', 'service_cat'],
      dtype='object')

In [297...] print (combined_updated['dissatisfied'].value_counts(dropna=False))

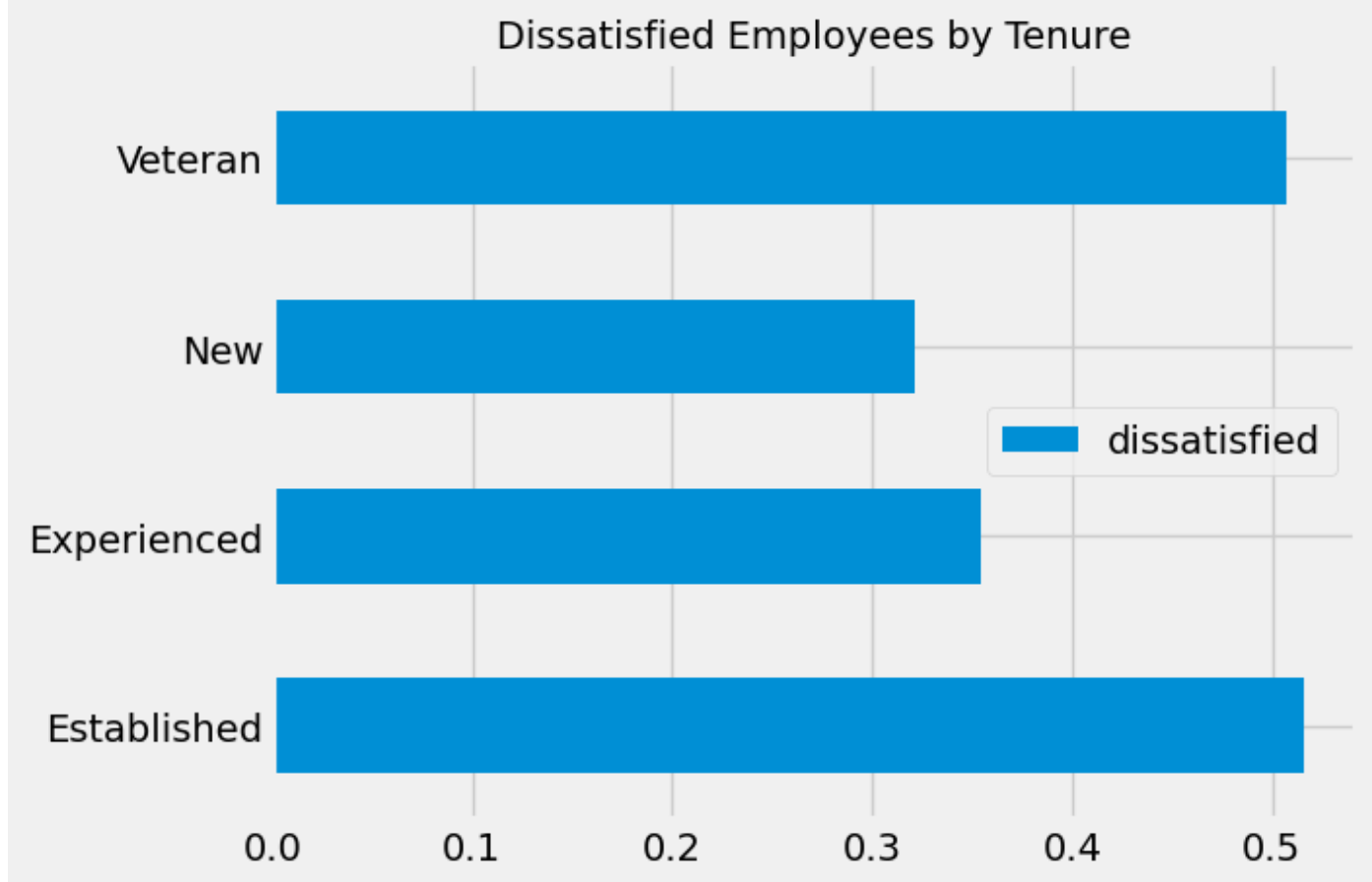
False      392
True       259
Name: dissatisfied, dtype: int64

In [298...] combined_pivot= pd.pivot_table(combined_updated, values= 'dissatisfied', index='service_cat')
print (combined_pivot)

              dissatisfied
service_cat
Established      0.516129
Experienced      0.354651
New              0.321244
Veteran          0.507353

In [300...] combined_pivot.plot(kind='barh')
plt.ylabel('')
plt.title("Dissatisfied Employees by Tenure", fontsize=14)
plt.show()

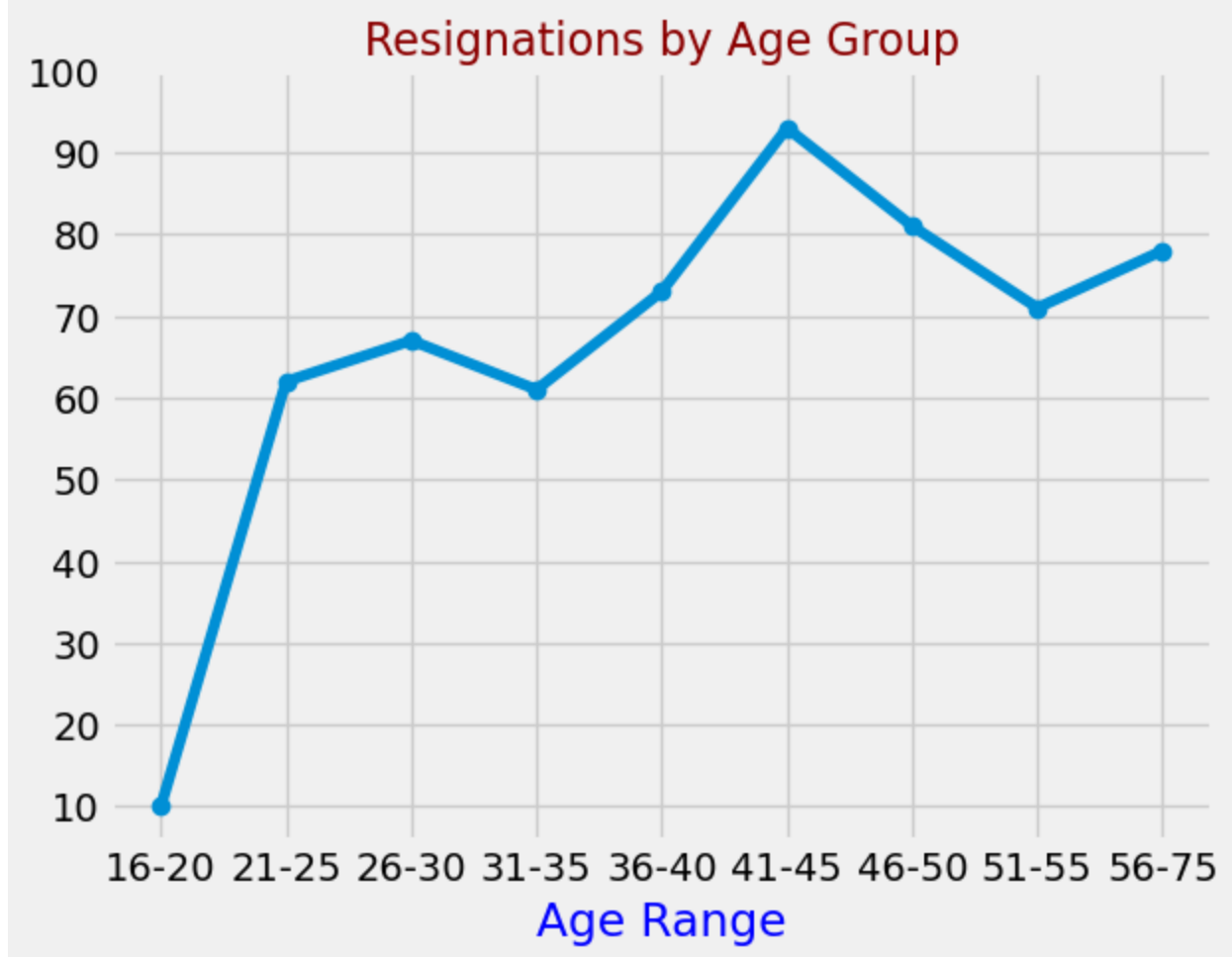
```

If we look at it as percentages Veterans and Established employees have a 50% chance of being dissatisfied with their employer. According to the graph it seems to increase over the years.

Resignations by age:

```
In [336... plt.plot(combined_updated['age'].value_counts().sort_index(), marker='o')
plt.yticks(np.arange(10, 110, 10))
plt.xlabel('Age Range', color='blue')
plt.title('Resignations by Age Group', fontsize=16, color='darkred')
plt.show()
```



```
In [337... print(combined_updated['age'].value_counts(dropna=False))
```

```
41-45    93
46-50    81
56-75    78
36-40    73
51-55    71
26-30    67
21-25    62
31-35    61
NaN       55
16-20     10
Name: age, dtype: int64
```

Notes:

The Age range with the most resignations is 41-45. The age range of 40-50 seems to be the largest age range of resignations. Lets see how the Age ranges compare to dissatisfaction.

- How many people in each age group resigned due to some kind of dissatisfaction?
- How many people in each career stage resigned due to some kind of dissatisfaction?
- Instead of analyzing the survey results together, analyze each survey separately. Did more employees in the DETE survey or TAFE survey end their employment because they were dissatisfied in some way?

```
In [304... truea_dissatisfied= combined_updated['age'] == '16-20'
resultsa=combined_updated.loc[truea_dissatisfied, 'dissatisfied']
print(resultsa.value_counts())
```

```
trueb_dissatisfied= combined_updated['age'] == '21-25'
```

```

resultsb=combined_updated.loc[trueb_dissatisfied, 'dissatisfied']
print(resultsb.value_counts())

truec_dissatisfied= combined_updated['age'] == '26-30'
resultsc=combined_updated.loc[truec_dissatisfied, 'dissatisfied']
print(resultsc.value_counts())

trued_dissatisfied= combined_updated['age'] == '31-35'
resultsd=combined_updated.loc[trued_dissatisfied, 'dissatisfied']
print(resultsd.value_counts())

truee_dissatisfied= combined_updated['age'] == '36-40'
resultse=combined_updated.loc[truee_dissatisfied, 'dissatisfied']
print(resultse.value_counts())

truef_dissatisfied= combined_updated['age'] == '41-45'
resultsf=combined_updated.loc[truef_dissatisfied, 'dissatisfied']
print(resultsf.value_counts())

trueg_dissatisfied= combined_updated['age'] == '46-50'
resultsg=combined_updated.loc[trueg_dissatisfied, 'dissatisfied']
print(resultsg.value_counts())

trueh_dissatisfied= combined_updated['age'] == '51-55'
resultsh=combined_updated.loc[trueh_dissatisfied, 'dissatisfied']
print(resultsh.value_counts())

truei_dissatisfied= combined_updated['age'] == '56-75'
resultsi=combined_updated.loc[truei_dissatisfied, 'dissatisfied']
print(resultsi.value_counts())

```

```

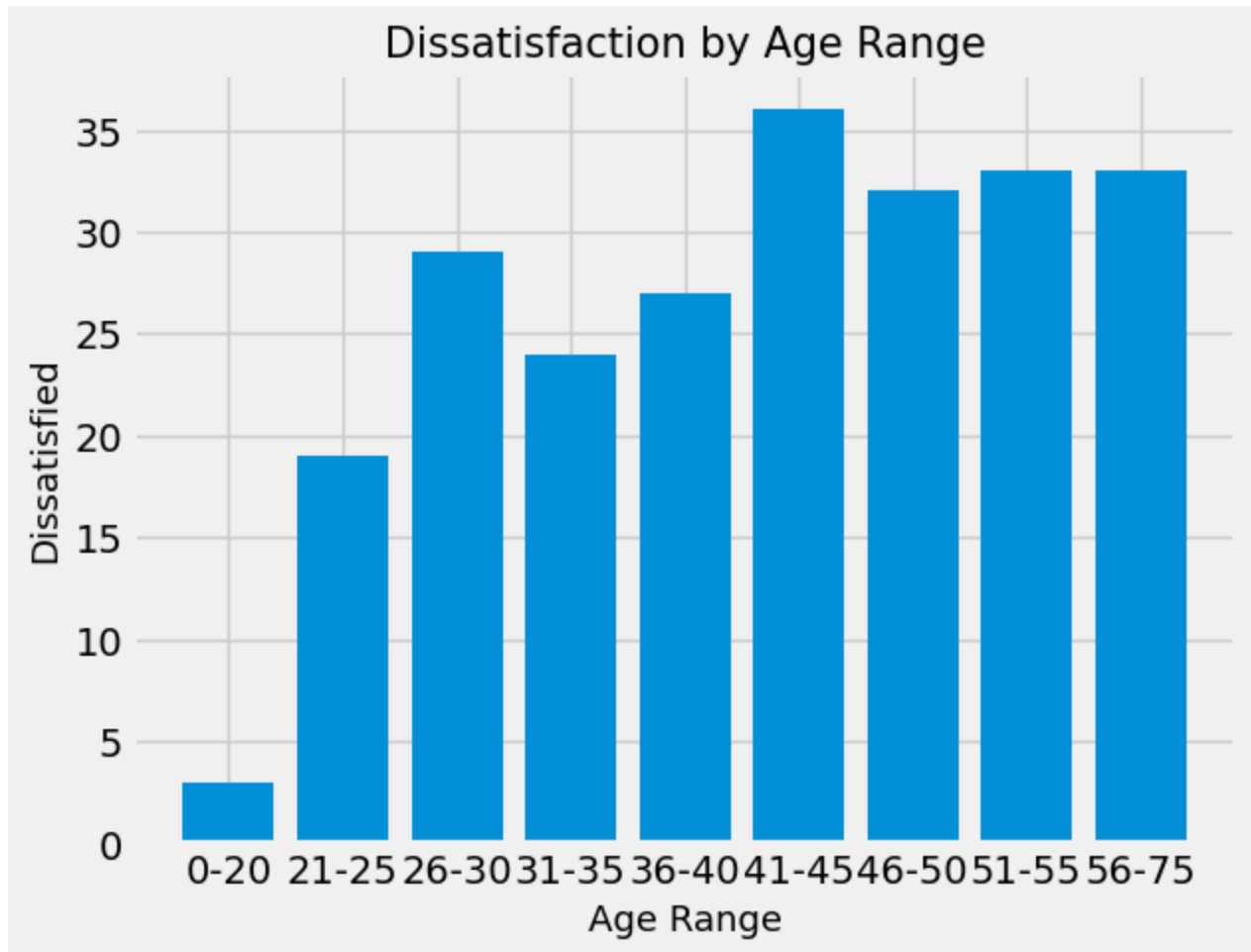
False      7
True       3
Name: dissatisfied, dtype: int64
False     43
True      19
Name: dissatisfied, dtype: int64
False     38
True      29
Name: dissatisfied, dtype: int64
False     37
True      24
Name: dissatisfied, dtype: int64
False     46
True      27
Name: dissatisfied, dtype: int64
False     57
True      36
Name: dissatisfied, dtype: int64
False     49
True      32
Name: dissatisfied, dtype: int64
False     38
True      33
Name: dissatisfied, dtype: int64
False     45
True      33
Name: dissatisfied, dtype: int64

```

```

In [348... x = ['0-20', '21-25', '26-30', '31-35', '36-40', '41-45', '46-50', '51-55', '56-75']
y = [3, 19, 29, 24, 27, 36, 32, 33, 33]
plt.bar(x,y)
plt.xlabel('Age Range', fontsize=13)
plt.ylabel('Dissatisfied', fontsize=13)
plt.title('Dissatisfaction by Age Range', fontsize=15)
plt.show()

```



Notes:

The most dissatisfied seem to be the age range of 41-45. The age range of 41-75 are more dissatisfied than people under the age of 40.

Lets see how dissatisfied vs not dissatisfied stack up against each other.

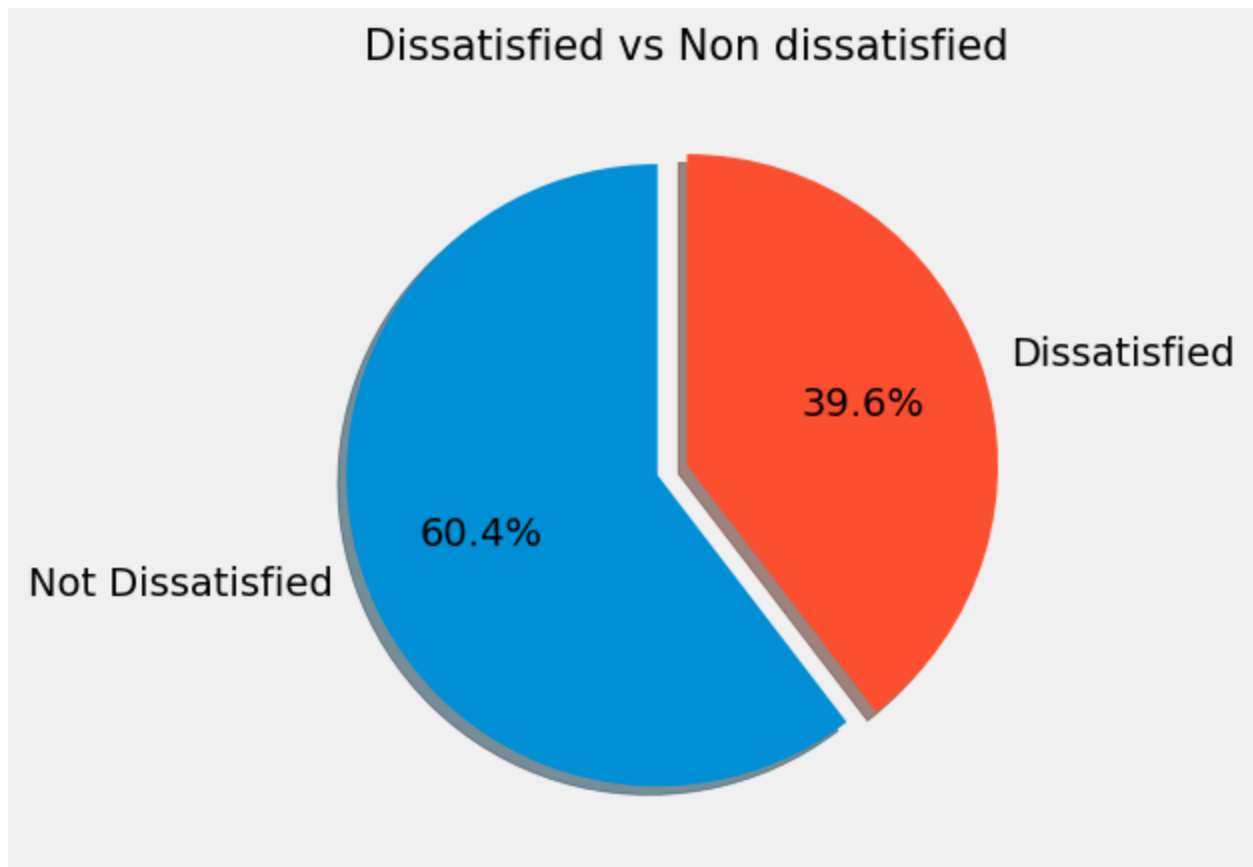
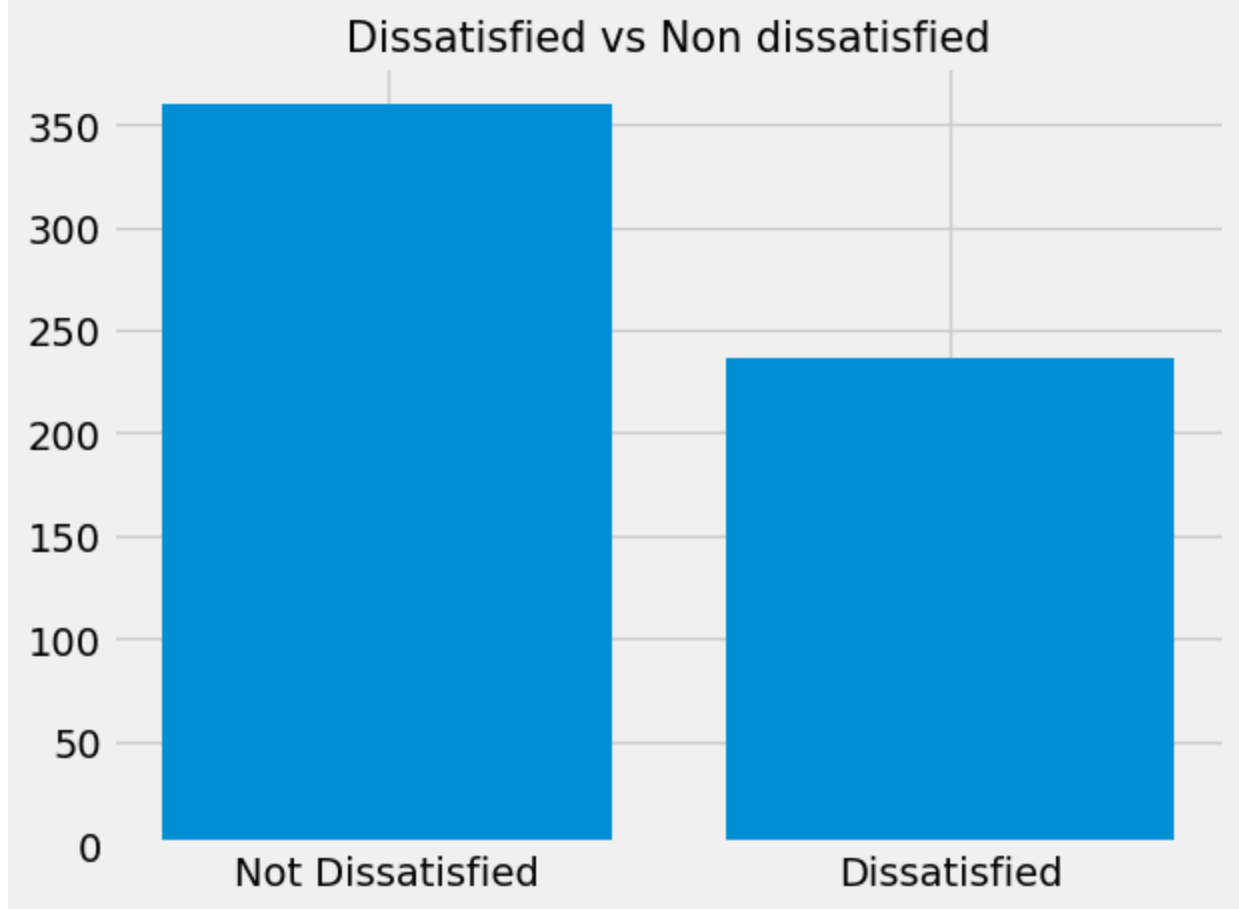
```
In [393... not_dissatisfied = 7+43+38+37+46+57+49+38+45
dissatisfied = 3+19+29+24+27+36+32+33+33

the_percentage_a = ( dissatisfied/(not_dissatisfied + dissatisfied)*100)
the_percentage_b = ( not_dissatisfied/(not_dissatisfied + dissatisfied )*100)

dis_per = pd.DataFrame({'percentage': [the_percentage_b,the_percentage_a]},index=['Not D

x = ['Not Dissatisfied', 'Dissatisfied']
y= [not_dissatisfied, dissatisfied]
plt.bar(x,y)
plt.title('Dissatisfied vs Non dissatisfied', fontsize=15)
plt.show()

dis_per.plot.pie(legend= None, subplots=True, explode=(0.1,0), shadow=True, startangle=
plt.xticks(rotation=0)
plt.title('Dissatisfied vs Non dissatisfied', fontsize=15)
plt.ylabel('')
plt.show()
```



Notes:

As we can see about 40% of people that resign from their job are dissatisfied. This is expected that a large portion of people that leave an employer would be dissatisfied. The other 60% are leaving for other reasons and possibly other opportunities.

```
In [307... print(combined_updated.columns)
print(combined_updated['service_cat'].value_counts())
print(combined_updated['dissatisfied'].value_counts())

Index(['id', 'separationtype', 'cease_date', 'position', 'employment_status',
      'gender', 'age', 'institute_service', 'dissatisfied', 'institute',
      'institute_service_up', 'service_cat'],
      dtype='object')
New          193
Experienced  172
Veteran      136
Established   62
Name: service_cat, dtype: int64
False        392
True         259
Name: dissatisfied, dtype: int64
```

```
In [396... newa_dissatisfied= combined_updated['service_cat'] == 'New'
resultsna=combined_updated.loc[newa_dissatisfied, 'dissatisfied']
print(resultsna.value_counts())

newb_dissatisfied= combined_updated['service_cat'] == 'Experienced'
resultsnb=combined_updated.loc[newb_dissatisfied, 'dissatisfied']
print(resultsnb.value_counts())

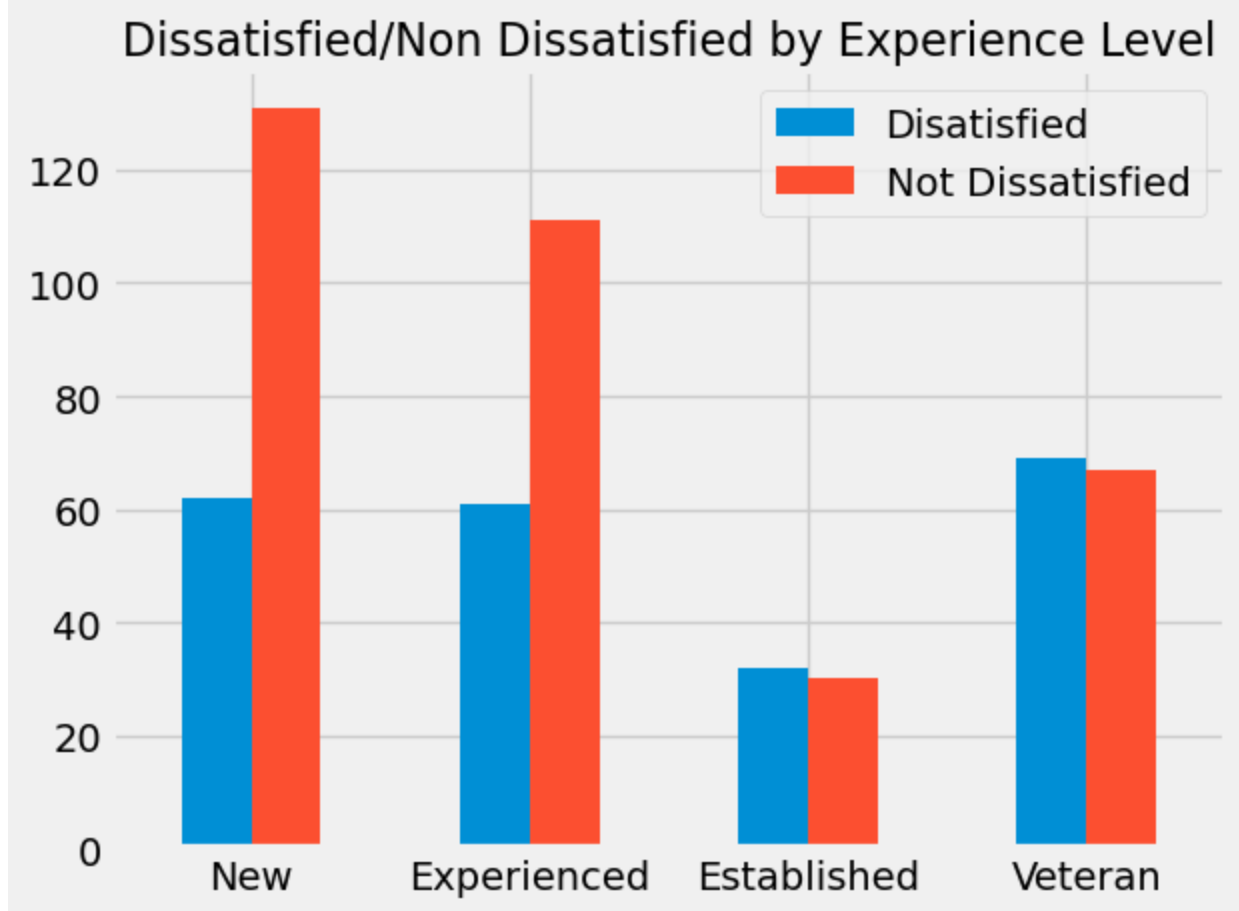
newc_dissatisfied= combined_updated['service_cat'] == 'Established'
resultsnc=combined_updated.loc[newc_dissatisfied, 'dissatisfied']
print(resultsnc.value_counts())

newd_dissatisfied= combined_updated['service_cat'] == 'Veteran'
resultsnd=combined_updated.loc[newd_dissatisfied, 'dissatisfied']
print(resultsnd.value_counts())

False        131
True          62
Name: dissatisfied, dtype: int64
False        111
True          61
Name: dissatisfied, dtype: int64
True          32
False         30
Name: dissatisfied, dtype: int64
True          69
False         67
Name: dissatisfied, dtype: int64
```

```
In [414... diss_t = pd.DataFrame({'Disatisfied':[62,61,32,69], 'Not Dissatisfied':[131,111,30,67]},
                        index=['New', 'Experienced', 'Established', 'Veteran'] )

diss_t.plot.bar()
plt.xticks(rotation=0)
plt.title('Dissatisfied/Non Dissatisfied by Experience Level', fontsize=16)
plt.show()
```



New: Less than 3 years at a company

Experienced: 3-6 years at a company

Established: 7-10 years at a company

Veteran: 11 or more years at a company

As we can see above dissatisfaction tends to increase over time.

Conclusion:

What we have gathered in this project:

- The age range of 40-50 seem to be most dissatisfied with their employer
- Dissatisfaction increases with tenure on the job.
- Dissatisfaction increases based on experience.
- Overall 60% of people that resign are not dissatisfied with their employer