

# Exit Analysis

## Table of contents

1. [Introduction](#)
2. [Data Collection](#)
3. [Data Wrangling](#)
4. [Explanatory Data Analysis](#)
5. [Questions & Answers](#)
6. [Exploring relations between variables](#)
7. [Conclusion](#)
8. [Final Notes](#)

---

## **\*\*Introduction\*\***

**Note: After the first analysis, the company was interested in in-depth-analysis, so they provided more accurate comprehensive dataset**

In XYZ company, many employees leave each month. The stackholders want to know facts about employees who leave, so the recruitment will be ready and will keep enough supply for hot jobs that employees leave, this is on one hand. On the other hand, they want to know what rules have less lifespan more than others as well as other questions to answer.

Imported Necessary libs

## **DATA Wrangling**

### **DATA Collection process**

**The Data was provided by the company from their system.**

Out[3]:

	code	nat	dept	job	join	leave	marital_status	birth_date	qualifi
0	21	Pakistani	Consumer Products	Sales Supervisor	01/07/1989	31/03/2014	Married	09/06/1954	High
1	26	Indian	General Products	Strategic New Business Dev. Manager	01/07/1998	31/12/2017	Married	16/04/1949	
2	27	Indian	Office Furniture & Furnishing	Assistant Manager - Office Furniture	01/07/1990	28/02/2015	Married	28/10/1947	Bac
3	29	Indian	Home Appliances Service Center	Team Leader	01/10/1998	29/11/2012	Married	05/02/1949	
4	31	Indian	HR & Administration	Supervisor - Payroll	01/07/1988	28/02/2015	Married	25/03/1946	Bac

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 887 entries, 0 to 886
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   code                   887 non-null    int64
1   nat                    887 non-null    object
2   dept                   887 non-null    object
3   job                    887 non-null    object
4   join                   887 non-null    object
5   leave                  887 non-null    object
6   marital_status         887 non-null    object
7   birth_date             887 non-null    object
8   qualification          751 non-null    object
dtypes: int64(1), object(8)
memory usage: 69.3+ KB
```

As we notice we have some missing values in qualificaiton column, however we will check each column for any wrong data as we have a few number of columns

```
Out[5]: code          0
nat            0
dept           0
job            0
join           0
leave          0
marital_status 0
birth_date     0
qualification   0
dtype: int64
```

No duplicates in the dataframe

```
Out[7]: Nepali      366
        Indian      164
        Sri Lankan  161
        Filipino    113
        Egyptian    22
        Kenyan      18
        Pakistan    9
        Bangladeshi 6
        Tunisian    6
        Jordanian   5
        Dutch       2
        Moroccan    2
        Lebanese    2
        Netherland   2
        Ghana       2
        Italian     1
        South African 1
        Syrian      1
        Pakistani   1
        American    1
        Algerian    1
        Nepalese    1
        Name: nat, dtype: int64
```

```
Out[9]: Nepalese    367
        Indian      164
        Sri Lankan  161
        Filipino    113
        Egyptian    22
        Kenyan      18
        Pakistani   10
        Bangladeshi 6
        Tunisian    6
        Jordanian   5
        Name: nat, dtype: int64
```

Now the values look ok, we move to the next column

```
Out[10]: Consumer Products      190
         Logistics               190
         Consumer Electronics & Home Appliances 109
         General Products        90
         Home Appliances Service Center      76
         Office Equipment Service Center     68
         Office Furniture & Furnishing       65
         Office Equipment & Business Solutions 51
         Finance                   23
         HR & Administration           16
         Directors Office            5
         IT                          4
         Name: dept, dtype: int64
```

Departments column is ok

```
Out[11]: Merchandiser      197
         Helper           61
         Sales Executive   50
         Driver           43
         Technician       33
         Asst Technician   24
         Senior Sales Executive 23
         Van Salesman     21
         Warehouse Asst   20
         Technician - Air Cons 18
         Name: job, dtype: int64

Out[12]: array(['Sales Supervisor', 'Strategic New Business Dev. Manager',
               'Assistant Manager - Office Furniture', 'Team Leader',
               'Supervisor - Payroll', 'Storekeeper',
               'Supervisor - Small Appl & Electronics', 'Foreman',
               'Senior Technician', 'Supervisor', 'Senior Sales Executive',
               'Asst Service Manager', 'Driver', 'Upholster', 'Asst Storekeeper',
               'Asst Manager Projects', 'Helper', 'Furnishing Sales Specialist',
               'Senior Receivable Accountant', 'Asst Technician',
               'Supervisor - Appliances', 'Fitter', 'Tailor',
               'Supervisor Accounts Payable', 'Store Supervisor',
               'Executive Secretary', 'Skilled Helper', 'Security Guard',
               'Sales Executive', 'Upholster - Team Leader',
               'Supervisor - Projects', 'Skilled Driver - Fitter',
               'Carpet Fitter', 'Manager - OFD', 'Driver Cum Asst',
               'Sr. Interior Designer', 'Senior Electrician', 'Technician',
               'Forklift Operator', 'Office Boy', 'Manager - OABS',
               'Warehouse Manager', 'Merchandiser',
               'Supervisor Banking Relations', 'Secretary',
               'Field Support Supervisor', 'Supervisor - MT',
               'Driver Cum Asst Technician', 'Technician - Air Cons',
               'Administrative Asst', 'Assistant Manager Sales', 'Accountant',
               'Sr. Consultant / Proj. Manager', 'Assistant Sales Executive',
               'Senior Cashier', 'Carpenter',
               'Senior Sales Executive (Projects)', 'Senior Support Engineer',
               'Carpet Fitter -Driver', 'Senior Merchandiser',
               'Field Sales Manager - MT', 'Assistant Manager - Office Equipment',
               'Skilled Driver', 'Documentation Asst.', 'Cashier',
               'Product Specialist', 'Sales Executive - Showroom',
               'Junior Technician', 'Merchandising Team Captain', 'Site Engineer',
               'Heavy Duty Driver', 'Technician - Projects', 'Welder',
               'Senior Functional Consultant', 'Warehouse Asst', 'Van Salesman',
               'Personnel - Asst', 'IT Support Executive', 'Office Assistant',
               'Debt Collector', 'Telephone / Fax Operator',
               'Technician - Electronics', 'Functional Consultant',
               'Software Developer', 'CSR', 'Asst Merchandiser', 'Manager OESC',
               'AutoCAD Operator', 'Senior Technician - Air Cons',
               'Senior Sales Supervisor', 'C&F Supervisor', 'Manager Logistics',
               'Sales Coordinator', 'Supervisor - QF', 'Sr. Admin Assistant',
               'Clerk', 'Cleaner', 'Field Sales Manager - GT', 'Asst Manager',
               'Merchandiser / Promoter', 'Sales & Inventory Accountant',
               'C&F Coordinator', 'Logistic Clerk', 'Route Planner',
```

```
'C&F Asst Supervisor', 'Manager Finance',
'Sales Executive - Perfect Store', 'Head of Support Services',
'Contract Administrator', 'Sr. Accountant Sales & Inventory',
'Receivable Supervisor', 'Fitter - Furniture', 'Invoice Clerk',
'Business Development Manager', 'Senior Technician - Appliances',
'Showroom Incharge', 'Collections Officer',
'Sales Executive - IT Products', 'Administration Executive',
'Recruitment & Training Officer', 'Assistant IT Support',
'Credit Controller', 'Electrician', 'Inventory Coordinator',
'OPS Consultant', 'Operations Manager', 'Senior Accountant',
'Manager - CPD', 'Junior Accountant', 'Administration Supervisor',
'MIS Assistant', 'Accounts Assistant',
'Head of Tenders & Technical', 'Manager Special Projects',
'Recruitment Assistant', 'Sr. Supervisor Personnel Affairs',
'Warehouse Supervisor', 'Application & Network Support Executive',
'Sales Support Executive', 'Assistant Manager HASC',
'Chief Executive Officer', 'Asst Manager - Key Account',
'Inventory Analyst', 'Accounts Receivable Accountant',
'Sales Support Supervisor', 'Head of Digital & Retail',
'Business Unit Manager - FMCG', 'Sales Promotor',
'Logistics Officer', 'Interior Designer',
'Business Development Executive', 'Support Services Officer',
'Tele Sales Representative', 'Senior Designer',
'Merchandising Supervisor', 'Product Executive',
'Trade Marketing Manager', 'Trade Marketing Officer'], dtype=object
)
```

I think we are ok with jobs column, at least for the top 10

we have joining date, birth date and leave date as objects, we should convert them to the right format in order to be beneficial for us

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 887 entries, 0 to 886
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   nat                    887 non-null   object
1   dept                   887 non-null   object
2   job                    887 non-null   object
3   join                   887 non-null   datetime64[ns]
4   leave                  887 non-null   datetime64[ns]
5   martial_status         887 non-null   object
6   birth_date             887 non-null   datetime64[ns]
7   qualification          751 non-null   object
dtypes: datetime64[ns](3), object(5)
memory usage: 62.4+ KB
```

- now our columns were converted to the valid format of date time, however we can use this in analysis, but also we need to make some numbers from them, so we can improve our dataset
- we will first make a service columns with all service years by usbtracting the leave date from

joining date

Out[16]:

	nat	dept	job	join	leave	martial_status	birth_date	qualification	servi
431	Pakistani	Finance	Receivable Supervisor	2013-09-01	2013-03-28	Married	1977-11-05	Bachelors	-0.4301
434	Lebanese	Consumer Products	Senior Sales Executive	2013-01-26	2013-01-04	Married	1984-08-22	Masters	-0.0602
438	Indian	Consumer Products	Sales Coordinator	2013-02-03	2013-01-09	Single	1984-10-04	Bachelors	-0.0684
529	Indian	Consumer Electronics & Home Appliances	Sales Executive	2014-08-03	2014-08-03	Single	1989-01-12	NaN	0.0000
551	Nepalese	Consumer Electronics & Home Appliances	Merchandiser	2014-04-28	2014-01-06	Single	1987-10-23	NaN	-0.3068

- we can see that there are some wrong dates, because we cannot consider that the employee left the work before joining, so we can fix the dates, simply by removing them. So in order to avoid any problems we will drop them.

now we can create a column using the birth date and joining date to know the age of employee when joined the company

Out[19]:

nat	dept	job	join	leave	martial_status	birth_date	qualification	service	join_age
-----	------	-----	------	-------	----------------	------------	---------------	---------	----------

- for joining age we dont have wrong values

Out[20]: Married 496  
Single 361  
Divorced 5  
Widow 1  
Name: martial\_status, dtype: int64

- martial status columns looks fine

Out[21]: High School 264  
Secondary School 183  
Bachelors 155

```
Diploma 70
Primary School 24
Masters 22
Information Not Available 18
Chartered Accountant 1
Name: qualification, dtype: int64
```

We have some employees that their qualification information is not available, so we will replace them with null values in order to decide how to deal with this later

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 863 entries, 0 to 885
Data columns (total 1 columns):
#   Column          Non-Null Count  Dtype
---  -
0   qualification    719 non-null   object
dtypes: object(1)
memory usage: 13.5+ KB
```

- We have high school and secodnary school, which is the same, so we are going to change secondary school to high school as well.
- One final step in wrangling is to set the index as the leave date in order to help us answer the below questions.

Out[27]:

	nat	dept	job	join	marital_status	birth_date	qualification	service	joi
leave									
2012-03-10	Nepalese	Consumer Products	Merchandiser	2008-05-03	Single	1986-01-29	NaN	3.854795	22
2012-05-09	Filipino	Consumer Products	Assistant Sales Executive	2002-07-20	Married	1978-08-20	NaN	9.810959	23
2012-05-11	Nepalese	Consumer Products	Merchandiser	2010-05-08	Single	1985-10-13	NaN	2.010959	24
2012-06-12	Nepalese	Home Appliances Service Center	Helper	2007-03-12	Single	1987-12-01	NaN	5.257534	19
2012-07-10	Nepalese	Consumer Electronics & Home Appliances	Merchandiser	2008-01-04	Single	1982-06-08	NaN	4.517808	25

# Explanatory Data Analysis

## Questions & Answers

**What are the minimum, maximum and average service period for employee and their ages at joining time?**

Out [ 28 ] :

	service	join_age
count	863.000000	863.000000
mean	6.005333	29.639734
std	6.325528	7.122875
min	0.005479	17.980822
25%	1.893151	24.190411
50%	3.556164	28.709589
75%	7.646575	33.745205
max	31.298630	58.460274

### **\*\*Explaining the service summary\*\***

- The number of records is **863** employees.
- We can see that the average service period for employee is **6** years.
- We can see that the standard deviation is **6.3** years > which is considered high and means that data includes wide range of values.
- The minimum service period is **0.00547** year which means 5 month.
- The first quartile is **1.89** years.
- The Second quartile - Median is **3.5** years.
- The third quartile is **7.65** years.
- The interquartile range (IQR) is  $7.65 - 1.89 = 5.76$  years.
- The maximum service period is **31** years..

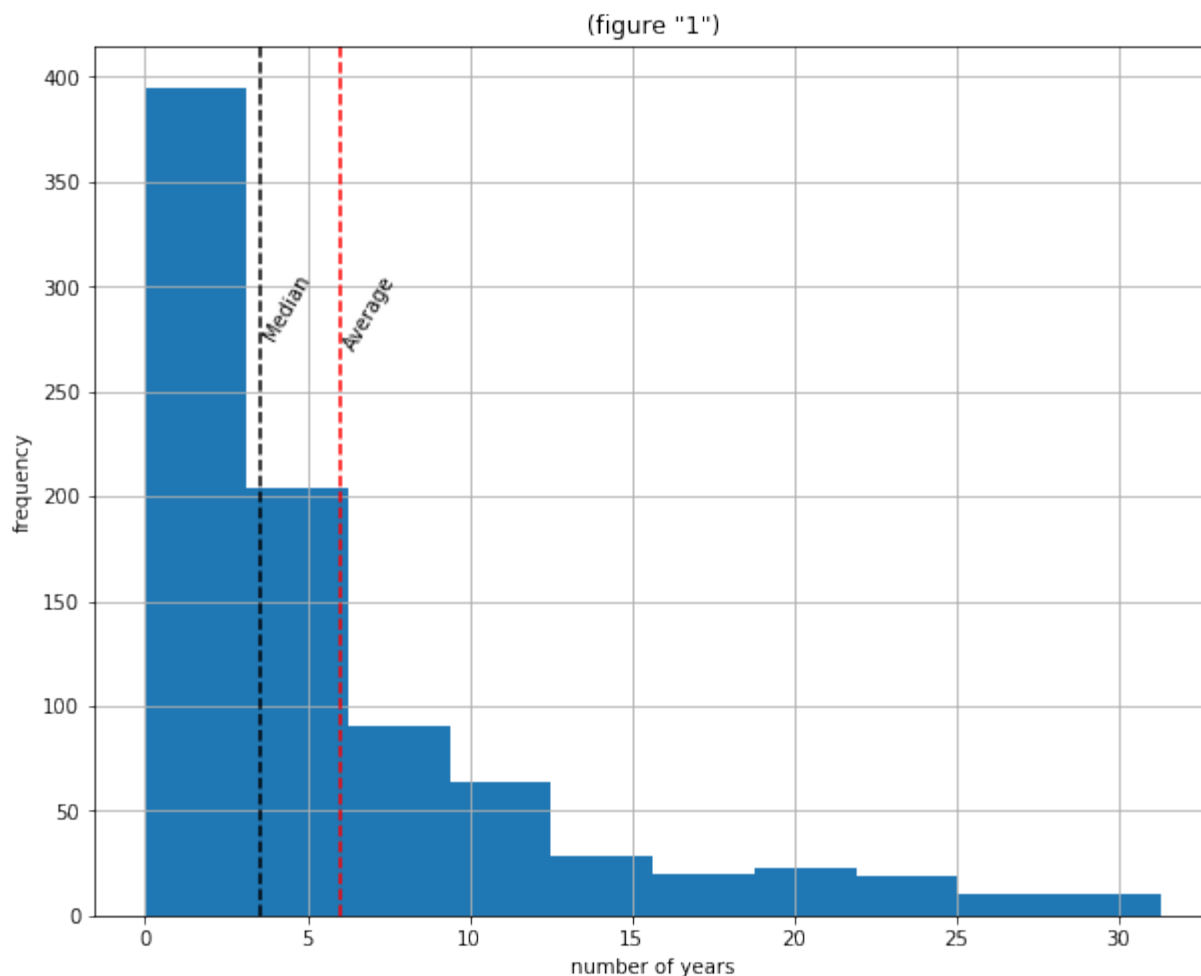
### **\*\*Explaining the joining age summary\*\***

- The number of records is **863** employees.
- We can see that the average joining age for employee is **29.6** years.
- We can see that the standard deviation is **7** years > which is considered high and means that data includes wide range of values.



- The minimum age is **17.9** year which means 5 month.
- The first quartile is **24.19** years.
- The Second quartile - Median is **28.7** years.
- The third quartile is **33.75** years.
- The interquartile range (IQR) is  $33.75 - 24.19 = 9.56$  years.
- The maximum age of joining is **58** years.

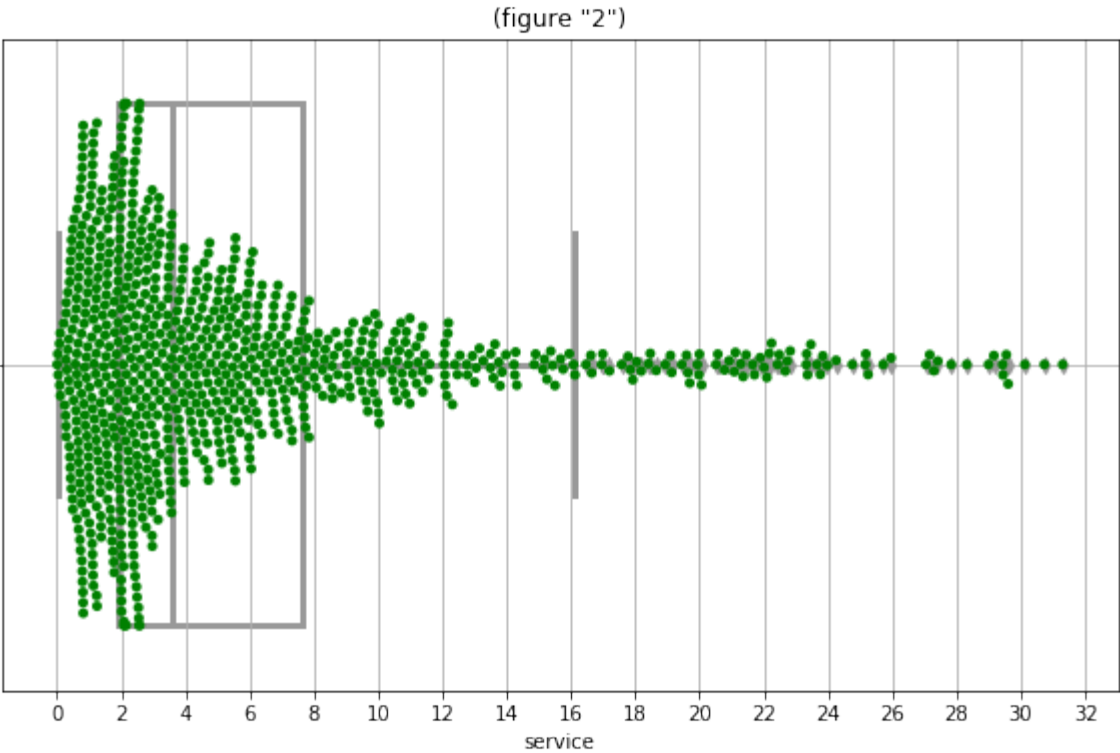
Out[29]: Text(6.005333417990762, 300, 'Average')



### explaining histogram (figure 1)

- We notice that the sample has positive skewness.
- More than 350 employees did not complete 5 years at least in the company
- we can see on the right side, there is a few number of employees they worked for more than 30 years, they are outliers. ##### As we have outliers, using The Average (mean) ~6.6~ will be misleading, it is better to use median **4.4** as a less misleading value for the average of service period for the employee

Out[30]: Text(0.5, 1.0, '(figure "2")')

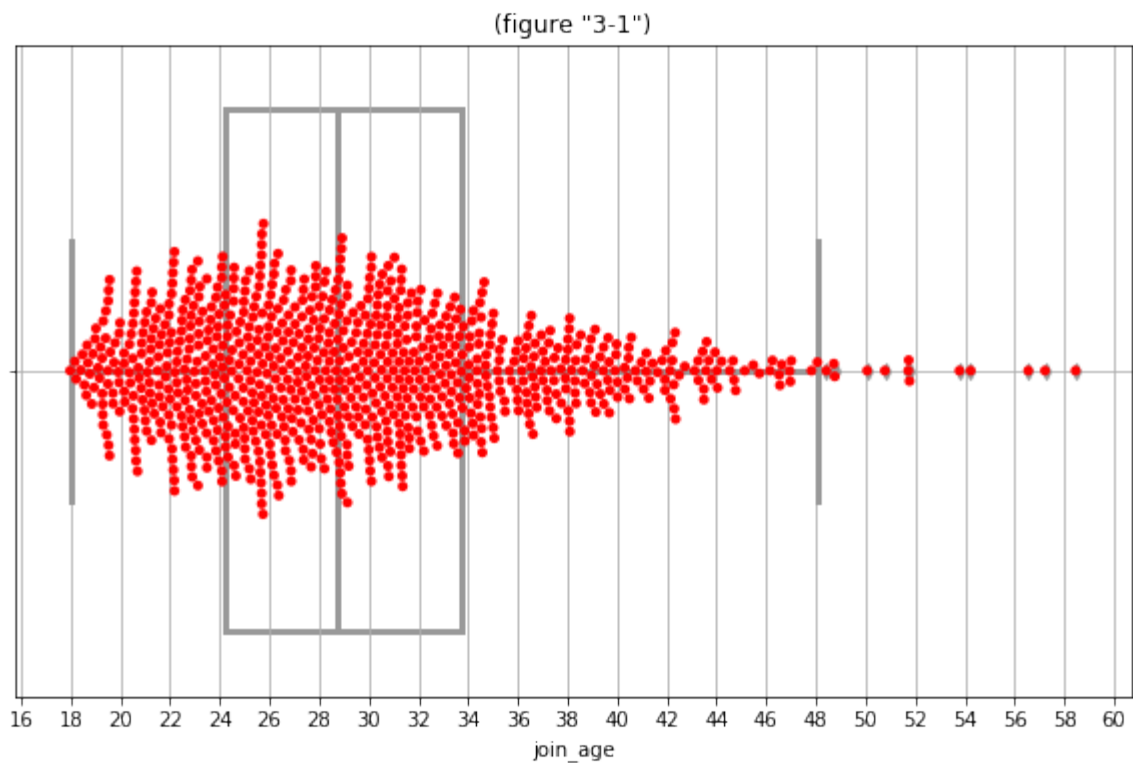


**Explaining Service Box plot (figure 2)**

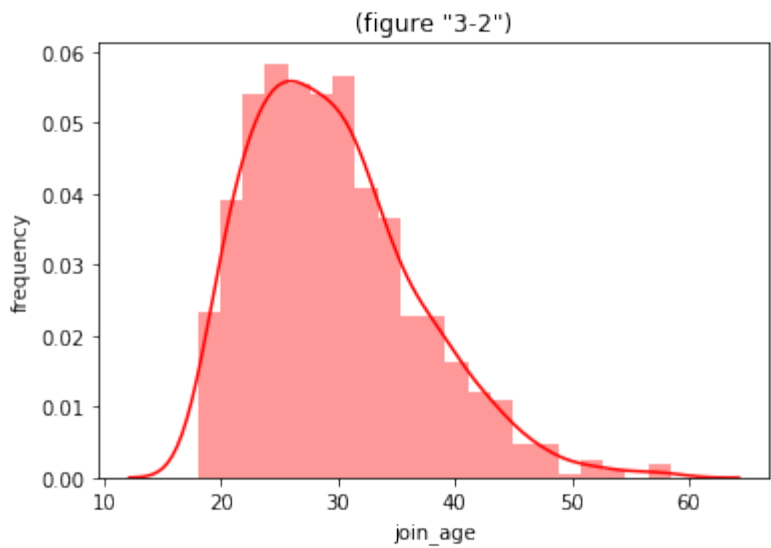
**We notice that most of population is between 0 and 8 years.**

**Employees who spent more than 16 years are considered outliers (too far from the average employees).**

```
Out[31]: Text(0.5, 1.0, '(figure "3-1")')
```



```
Out[32]: Text(0, 0.5, 'frequency')
```



**Explaining Join age Box plot (figure 4)**

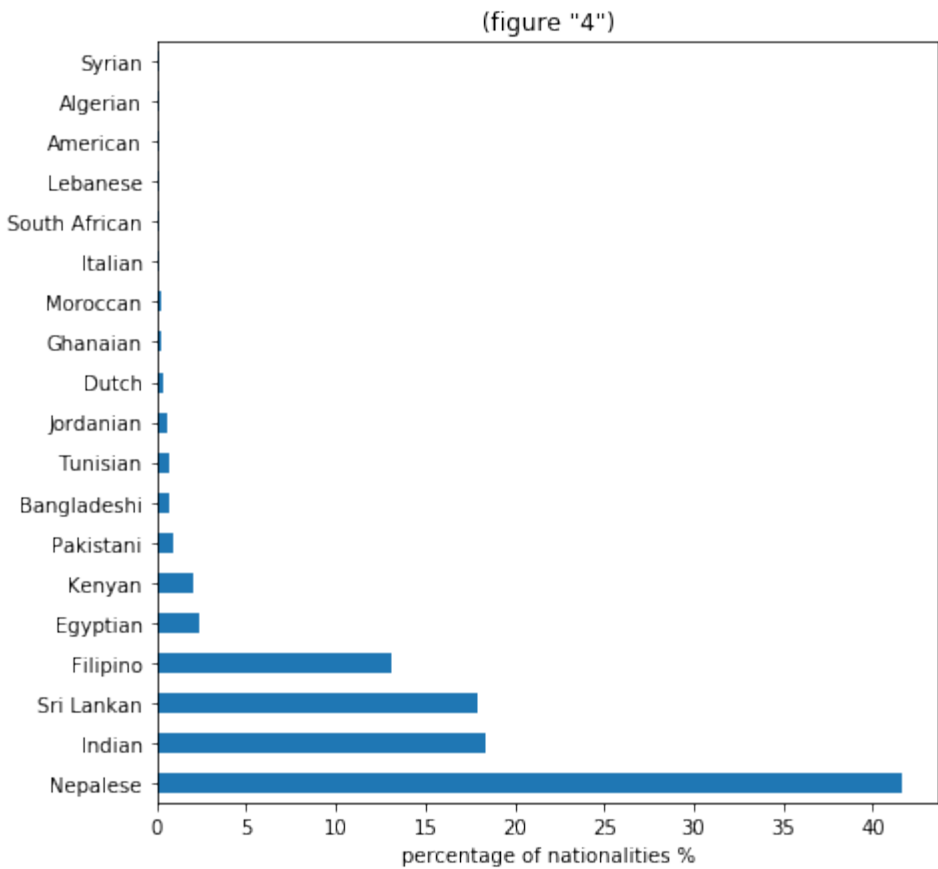
We notice that most of population age is between 18 and 48 years.

Employees who are more than 48 years are considered outliers (too far from the average employees).

**What are the nationalities that tend to leave the company more than the others?**

Out[33]:

	percentages	counts
Nepalese	41.599073	359
Indian	18.424102	159
Sri Lankan	17.960603	155
Filipino	13.093859	113
Egyptian	2.433372	21
Kenyan	2.085747	18
Pakistani	0.926999	8
Bangladeshi	0.695249	6
Tunisian	0.695249	6
Jordanian	0.579374	5
Dutch	0.347625	3
Ghanaian	0.231750	2
Moroccan	0.231750	2
Italian	0.115875	1
South African	0.115875	1
Lebanese	0.115875	1
American	0.115875	1
Algerian	0.115875	1
Syrian	0.115875	1



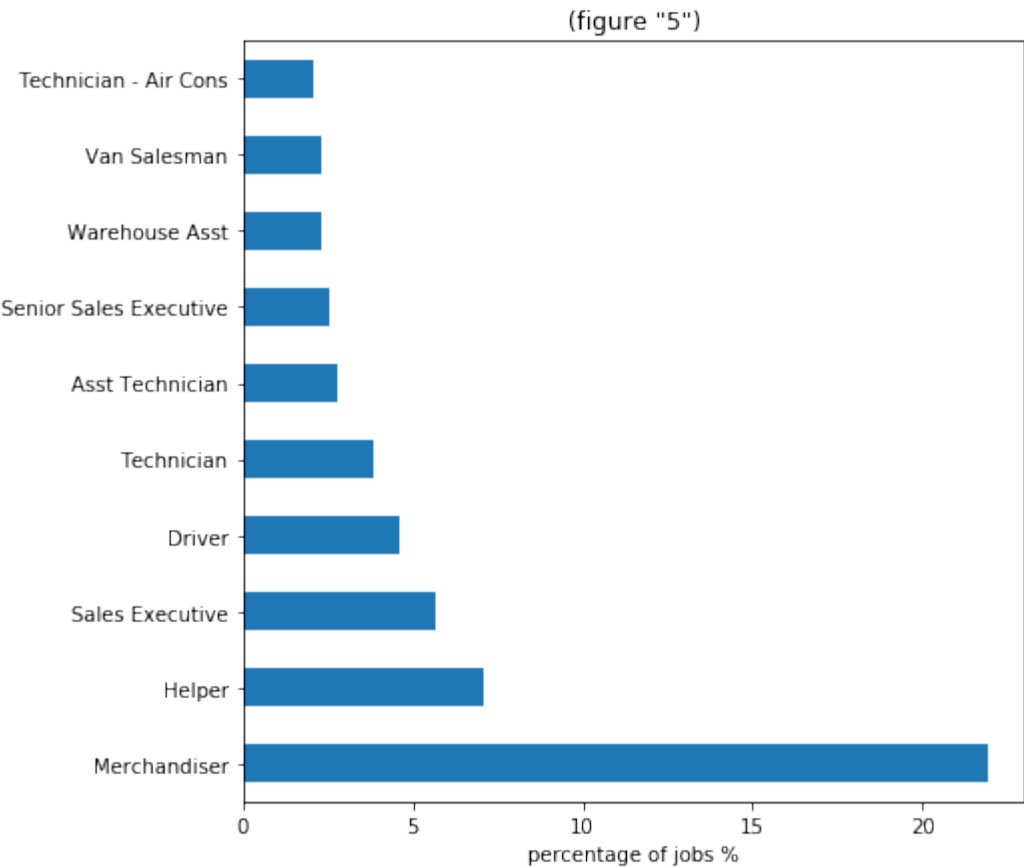
**Nationalities column statistics (figure 4)**

We notice that the dominant nationalities of the sample are Nepalese, Indians, Sri-Lankans and Filipinos (As they represent the most of the company's population)

**What are the top 10 positions that they tend to leave the company more than others?**

Out[34]:

	percentages	counts
Merchandiser	21.900348	189
Helper	7.068366	61
Sales Executive	5.677868	49
Driver	4.634994	40
Technician	3.823870	33
Asst Technician	2.780997	24
Senior Sales Executive	2.549247	22
Warehouse Asst	2.317497	20
Van Salesman	2.317497	20



**Positions/jobs column statistics (figure 5)**

There are too many jobs, however we choose to show only the top 10 jobs in order to display them properly

We notice that most of percentage of the sample is Merchanidisers.

Knowing the hot jobs that they require recruting on fast pace will make the company always ready to face these changes without interrupting the workflow.

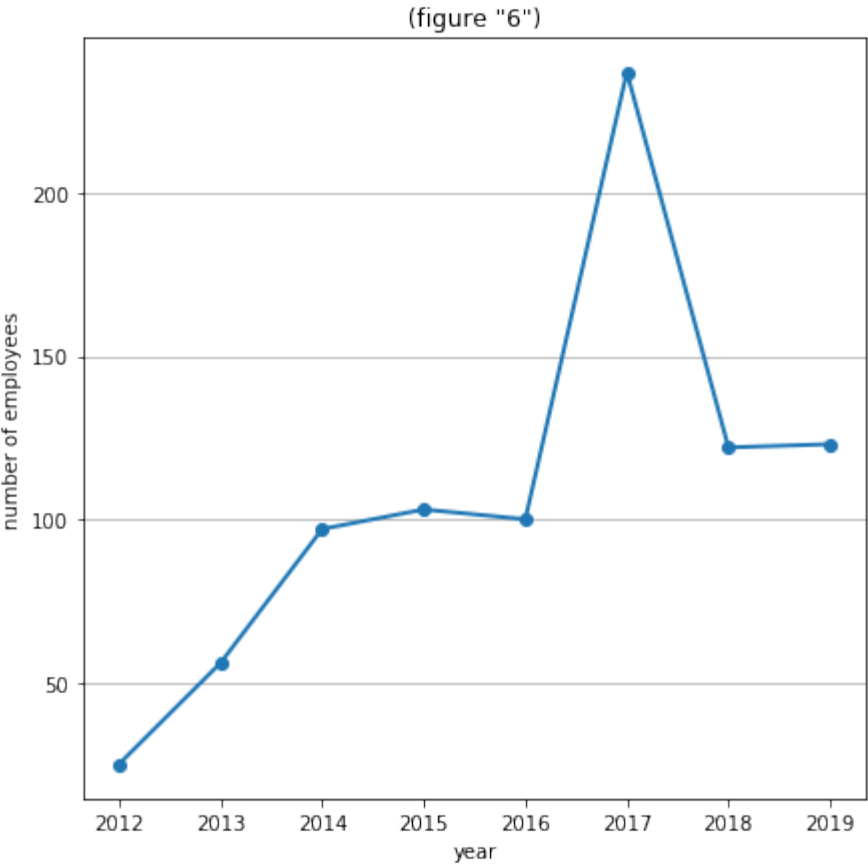
**What is the average number of people they are leaving each month/quarter/year?**

The expected amount of employees to leave per month are: 10.0  
per Quarter: 30.0  
per half year: 56.0  
per year 120.0

**General year-month Statistics**

	nat												Total_per_year
leave	1	2	3	4	5	6	7	8	9	10	11	12	
leave													
2012	0.0	0.0	1.0	0.0	2.0	1.0	2.0	2.0	1.0	2.0	6.0	8.0	25
2013	4.0	4.0	5.0	3.0	3.0	5.0	3.0	4.0	8.0	6.0	4.0	7.0	56
2014	16.0	14.0	6.0	10.0	2.0	7.0	11.0	6.0	5.0	11.0	4.0	5.0	97
2015	18.0	8.0	8.0	8.0	7.0	12.0	3.0	8.0	13.0	8.0	7.0	3.0	103
2016	7.0	8.0	9.0	6.0	14.0	4.0	5.0	5.0	14.0	12.0	3.0	13.0	100
2017	32.0	14.0	16.0	22.0	32.0	6.0	13.0	29.0	22.0	17.0	16.0	18.0	237
2018	19.0	13.0	16.0	12.0	7.0	12.0	16.0	4.0	4.0	9.0	8.0	2.0	122
2019	25.0	20.0	12.0	11.0	11.0	6.0	9.0	6.0	11.0	6.0	4.0	2.0	123

```
Out[36]: Text(0.5, 1.0, '(figure "6")')
```



**Year statistics (figure 6)**

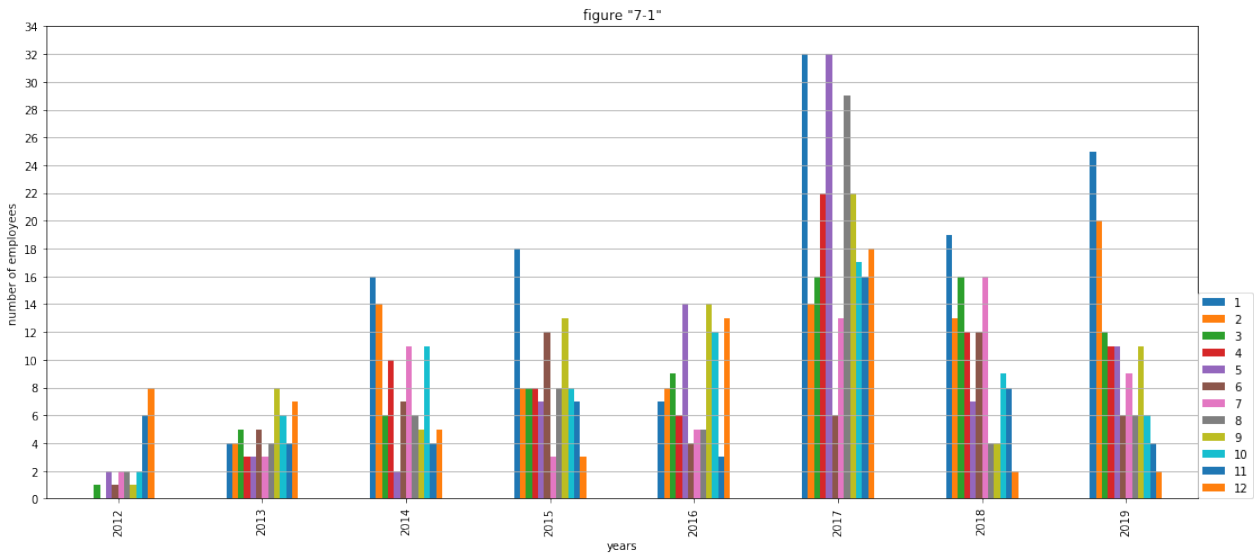
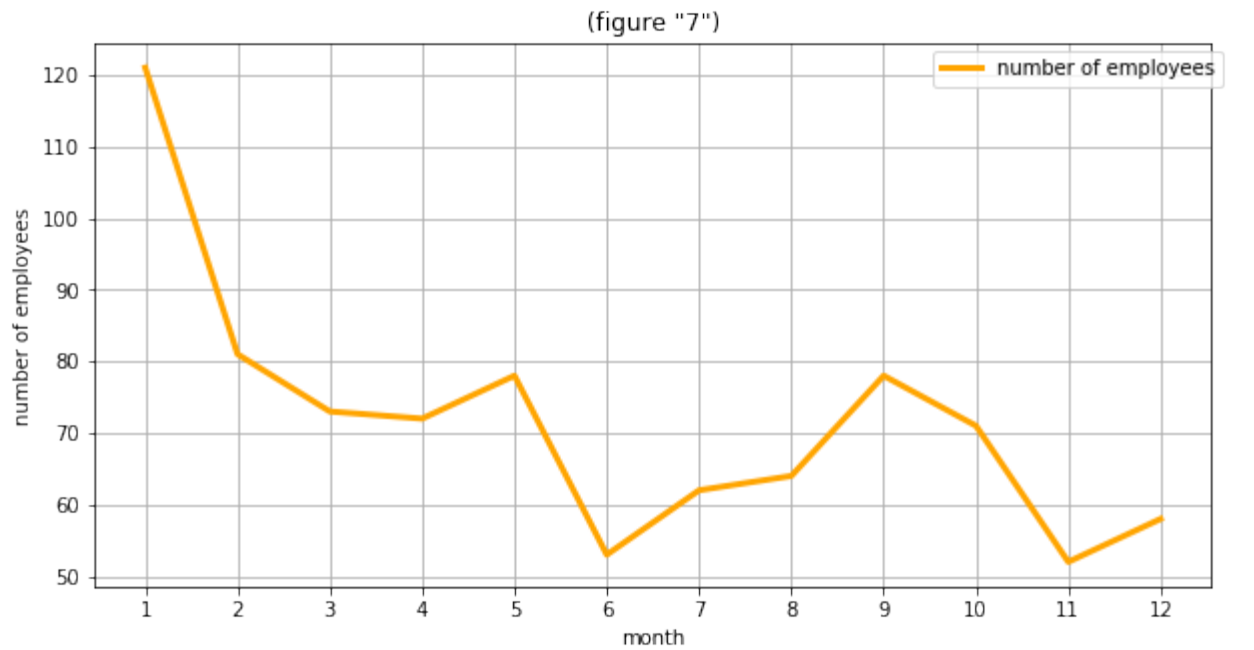
**We notice from the table that 2012 the data is not complete.**

**The years that have the full data of all months are 2013:2019.**

As the system was implemented in 2012 and this data was given for analysis by the end of Nov. 2019, December is not complete.

2017 has the peak if we compared it to the other full years in which may abnormal trend in exit rate of employees, they it continued normally in 2018 & 2019.

Month over years	1	2	3	4	5	6	7	8	9	10	11	12
number of employees	121	81	73	72	78	53	62	64	78	71	52	58



Out [38]:

leave	1	2	3	4	5	6	7	8	9	10	11	12
2012	0.0	0.0	1.0	0.0	2.0	1.0	2.0	2.0	1.0	2.0	6.0	8.0
2013	4.0	4.0	5.0	3.0	3.0	5.0	3.0	4.0	8.0	6.0	4.0	7.0



2014	16.0	14.0	6.0	10.0	2.0	7.0	11.0	6.0	5.0	11.0	4.0	5.0
2015	18.0	8.0	8.0	8.0	7.0	12.0	3.0	8.0	13.0	8.0	7.0	3.0
2016	7.0	8.0	9.0	6.0	14.0	4.0	5.0	5.0	14.0	12.0	3.0	13.0
2017	32.0	14.0	16.0	22.0	32.0	6.0	13.0	29.0	22.0	17.0	16.0	18.0
2018	19.0	13.0	16.0	12.0	7.0	12.0	16.0	4.0	4.0	9.0	8.0	2.0
2019	25.0	20.0	12.0	11.0	11.0	6.0	9.0	6.0	11.0	6.0	4.0	2.0

Month statistics (figure 7)

We notice from the table that january is the highest month at all, we notice also that january always higher in average than other months. However, if we considered that employees resigns before the leave date with one month at least as a notice, then we can see that january is high because many people provided their resignation on December for example which changes our prospective for high and low months.

Month statistics (figure 6-1)

providing a chart of all months in each year in order to compare them to each other

```
Out[39]: '(df_main["job"] == "Merchandiser") | (df_main["job"] == "Helper") | (df_m
ain["job"] == "Sales Executive") | (df_main["job"] == "Driver") | (df_main
["job"] == "Technician") | (df_main["job"] == "Asst Technician") | (df_mai
n["job"] == "Senior Sales Executive") | (df_main["job"] == "Warehouse Asst
") | (df_main["job"] == "Van Salesman") | (df_main["job"] == "Technician -
Air Cons") '
```

Out[41]:

job	Assistant Sales Executive	Driver	Helper	Merchandiser	Sales Executive	Senior Sales Executive	Senior Technician	Technician
1	2.480000	4.1300	8.260000	17.360000	5.790000	1.650000	0.830000	6.610000
2	1.230000	4.9400	4.940000	14.810000	9.880000	0.000000	2.470000	2.470000
3	2.740000	6.8500	8.220000	28.770000	5.480000	4.110000	2.740000	1.370000
4	2.780000	1.3900	5.560000	15.280000	5.560000	1.390000	1.390000	2.780000
5	2.560000	1.2800	5.130000	26.920000	8.970000	1.280000	1.280000	3.850000
6	0.000000	7.5500	9.430000	18.870000	1.890000	9.430000	1.890000	1.890000
7	1.610000	6.4500	11.290000	22.580000	8.060000	3.230000	1.610000	6.450000
8	0.000000	3.1200	6.250000	31.250000	3.120000	1.560000	0.000000	4.690000
9	0.000000	2.5600	11.540000	23.080000	3.850000	1.280000	3.850000	3.850000
10	4.230000	7.0400	7.040000	26.760000	4.230000	2.820000	4.230000	2.820000
11	0.000000	7.6900	1.920000	25.000000	1.920000	5.770000	1.920000	3.850000

12	0.000000	5.1700	3.450000	15.520000	6.900000	1.720000	0.000000	3.450000
Minimum	0.000000	1.2800	1.920000	14.810000	1.890000	0.000000	0.000000	1.370000
Average	1.469167	4.8475	6.919167	22.183333	5.470833	2.853333	1.850833	3.673333
Maximum	4.230000	7.6900	11.540000	31.250000	9.880000	9.430000	4.230000	6.610000

Explaining table of top 10 jobs percentages

- The table shows the percentage of each job that left in each month over all the years.
  - The minimum and maximum and average are the expected rate for each of these jobs to leave each month
- ##### for example if there is 100 employees leave the company per month, it is expected that the average around 22% of them will be Merchandisers and up to 34% or more of total number of employees will be Merchandisers.
- ### As we saw above the average employees they leave the company every month around 10 employees, we expect 2:3 to be Merchandisers.

What are the highest and lowest months?

'The lowest month'

nat												
leave	1	2	3	4	5	6	7	8	9	10	11	12
leave												
2012	NaN	NaN	1.0	NaN	NaN	1.0	NaN	NaN	1.0	NaN	NaN	NaN
2013	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2014	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2015	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2016	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2017	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2018	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2019	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

'The highest month'

nat												
leave	1	2	3	4	5	6	7	8	9	10	11	12
leave												

2012	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2013	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2014	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2015	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2016	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2017	32.0	NaN	NaN	NaN	32.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2018	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2019	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

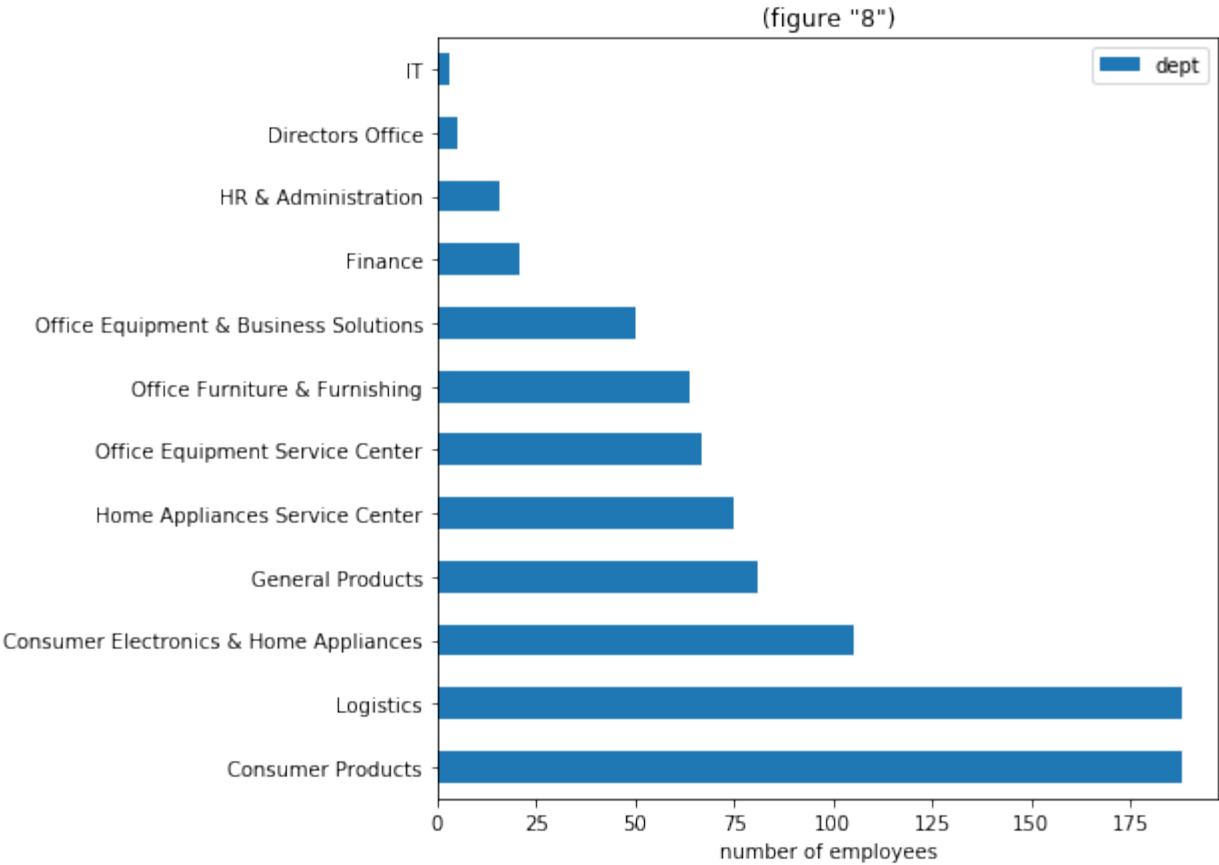
We see that the lowest month is March 2012.

The maximum number was in January & May 2017.

What is the highest department in employees left and what are the departments that they have less service periods?

	dept
Consumer Products	188
Logistics	188
Consumer Electronics & Home Appliances	105
General Products	81
Home Appliances Service Center	75
Office Equipment Service Center	67
Office Furniture & Furnishing	64
Office Equipment & Business Solutions	50
Finance	21
HR & Administration	16
Directors Office	5
IT	3

```
Out[43]: Text(0.5, 1.0, '(figure "8")')
```



We notice (figure 8) that Consumer Products & Logistics departments have the highest exit rate.

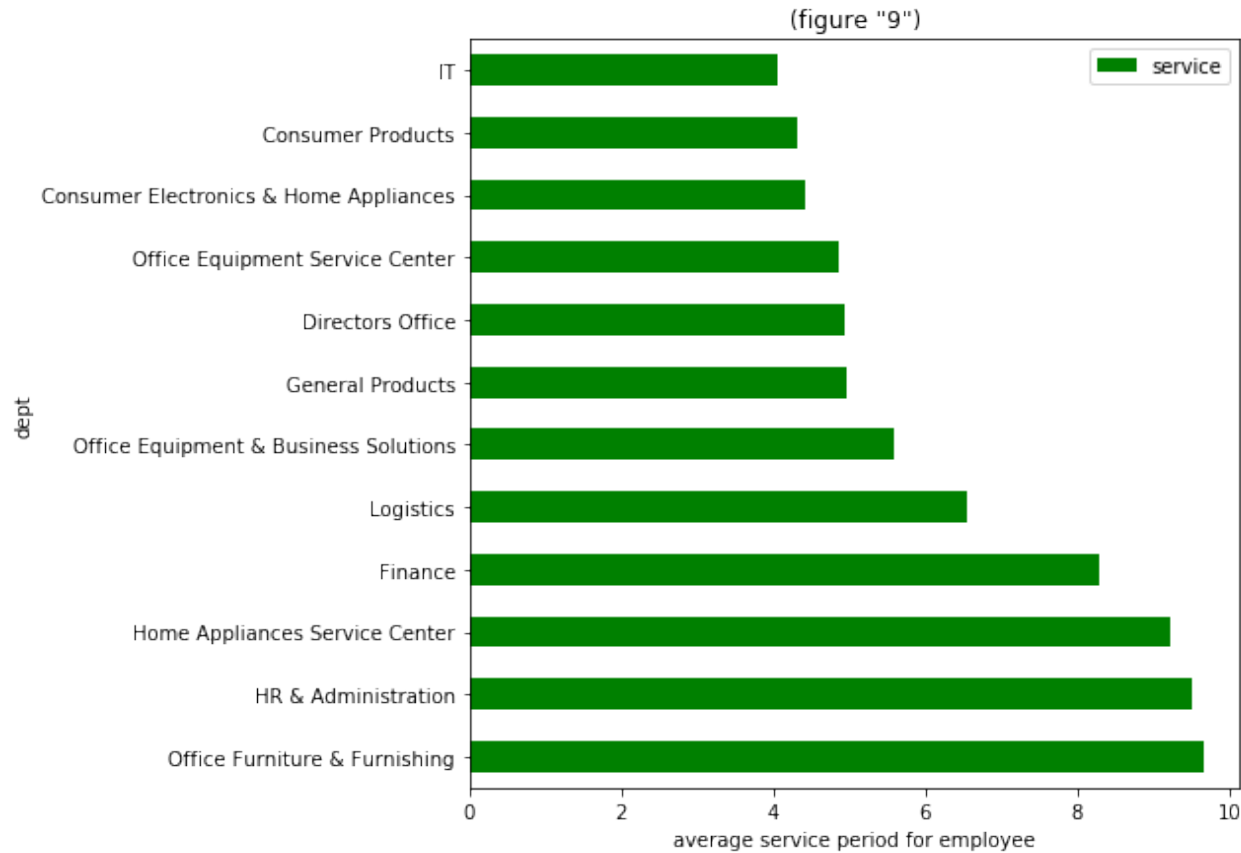
The lowest rate in Directors office and IT as the total number of employees in these departments are extremely low.

We notice also that HR and Finance Departments are stable and have less number of employees.

dept	service
Office Furniture & Furnishing	9.658048
HR & Administration	9.503082
Home Appliances Service Center	9.215379
Finance	8.293020
Logistics	6.539172
Office Equipment & Business Solutions	5.573753
General Products	4.966447
Directors Office	4.930411
Office Equipment Service Center	4.853854

<b>Consumer Electronics &amp; Home Appliances</b>	4.420587
<b>Consumer Products</b>	4.311615
<b>IT</b>	4.063927

```
Out[44]: Text(0.5, 1.0, '(figure "9")')
```

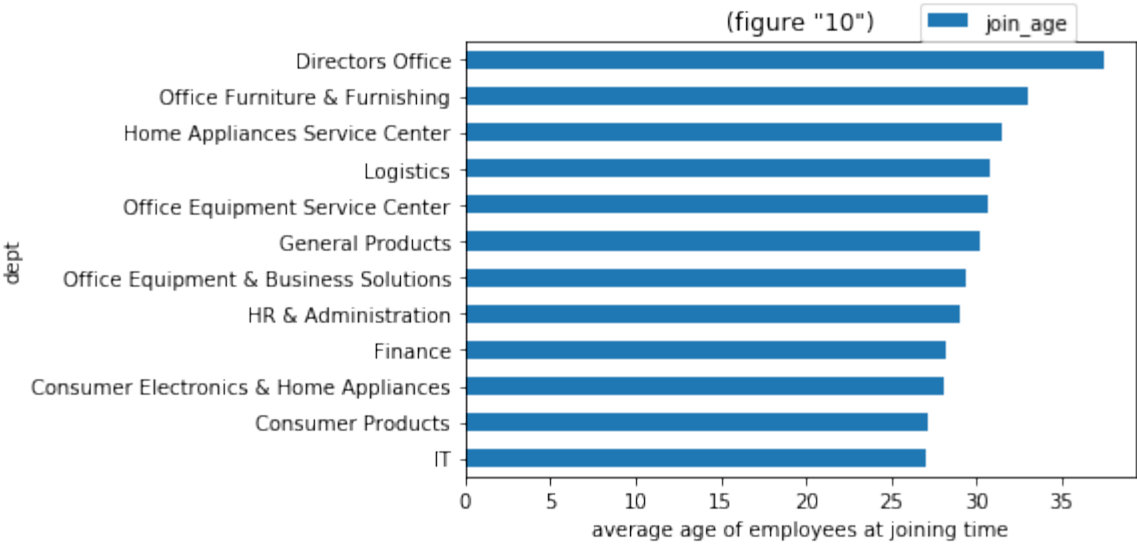


In last table and plot (figure 9), if we excluded IT and Directors for their low numbers, we will find that Finance department has higher rates of employee service and stability and on the other hand Consumer Products Department is not stable and it has one of hot positions that requires continuous hiring.

	join_age
<b>dept</b>	
<b>IT</b>	27.062100
<b>Consumer Products</b>	27.082177
<b>Consumer Electronics &amp; Home Appliances</b>	28.052368
<b>Finance</b>	28.132942
<b>HR &amp; Administration</b>	28.966096
<b>Office Equipment &amp; Business Solutions</b>	29.345699
<b>General Products</b>	30.220193
<b>Office Equipment Service Center</b>	30.629605

<b>Logistics</b>	30.737846
<b>Home Appliances Service Center</b>	31.514265
<b>Office Furniture &amp; Furnishing</b>	32.968921
<b>Directors Office</b>	37.421370

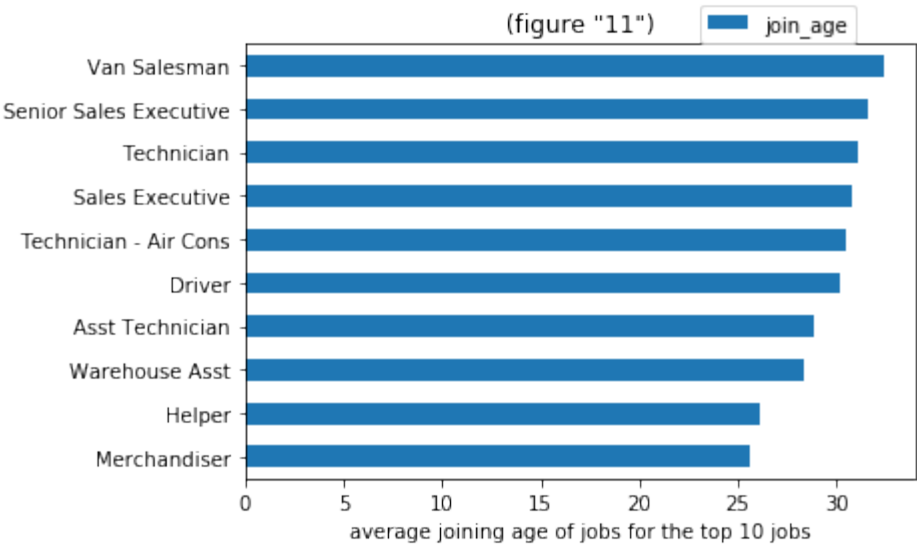
```
Out[45]: Text(0.5, 1.0, '(figure "10")')
```



From the above plot (Figure 10): The youngest people joined where in IT and Consumer Products 27 years and the oldest people joined were in directors office at the age of 37 years.

	join_age
<b>Merchandiser</b>	25.620686
<b>Helper</b>	26.094049
<b>Warehouse Asst</b>	28.316301
<b>Asst Technician</b>	28.815068
<b>Driver</b>	30.195959
<b>Technician - Air Cons</b>	30.438356
<b>Sales Executive</b>	30.775231
<b>Technician</b>	31.044832
<b>Senior Sales Executive</b>	31.585803
<b>Van Salesman</b>	32.359863

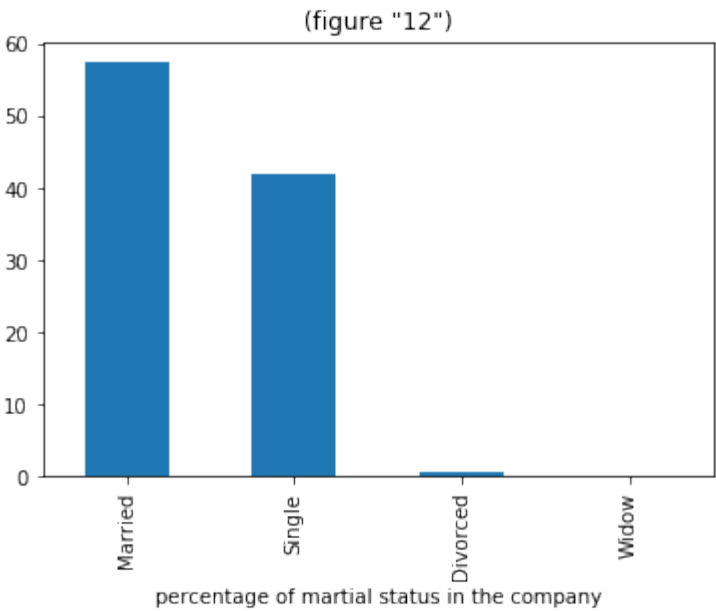
```
Out[46]: Text(0.5, 1.0, '(figure "11")')
```



From the previous plot figure 11: we can observe that Merchandiser job has the youngest ages with 25 years and van salesman has the highest 32 years

marital_status	
Married	57.473928
Single	41.830823
Divorced	0.579374
Widow	0.115875

```
Out[47]: Text(0.5, 1.0, '(figure "12")')
```

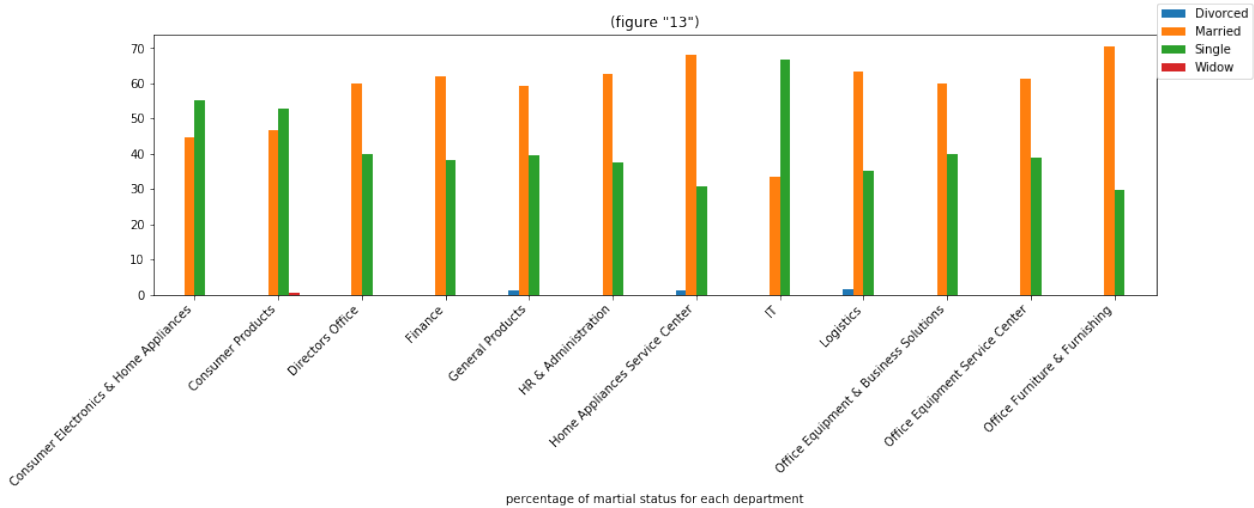


We can see in figure 12 that most of employees are married **57.5%** and the remaining are singles

41.8% with very low percentage for Divorced and widow.

	Divorced	Married	Single	Widow
dept				
Consumer Electronics & Home Appliances	0.000000	44.761905	55.238095	0.000000
Consumer Products	0.000000	46.808511	52.659574	0.531915
Directors Office	0.000000	60.000000	40.000000	0.000000
Finance	0.000000	61.904762	38.095238	0.000000
General Products	1.234568	59.259259	39.506173	0.000000
HR & Administration	0.000000	62.500000	37.500000	0.000000
Home Appliances Service Center	1.333333	68.000000	30.666667	0.000000
IT	0.000000	33.333333	66.666667	0.000000
Logistics	1.595745	63.297872	35.106383	0.000000
Office Equipment & Business Solutions	0.000000	60.000000	40.000000	0.000000
Office Equipment Service Center	0.000000	61.194030	38.805970	0.000000
Office Furniture & Furnishing	0.000000	70.312500	29.687500	0.000000

```
Out[48]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11]),
<a list of 12 Text xticklabel objects>)
```



In figure 13: we see that most of departments dominated by married employees except Consumer Electronics, Consumer products and IT

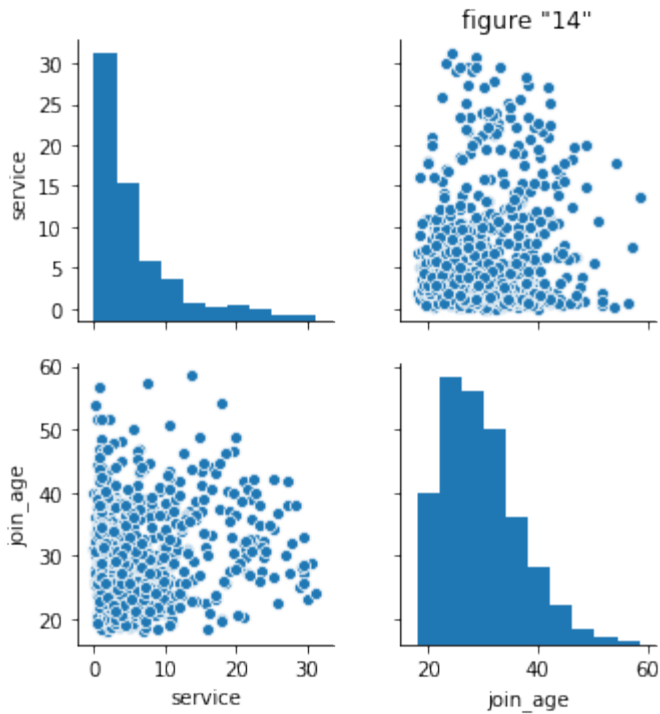
## Exploring relationships between different



**variables:**

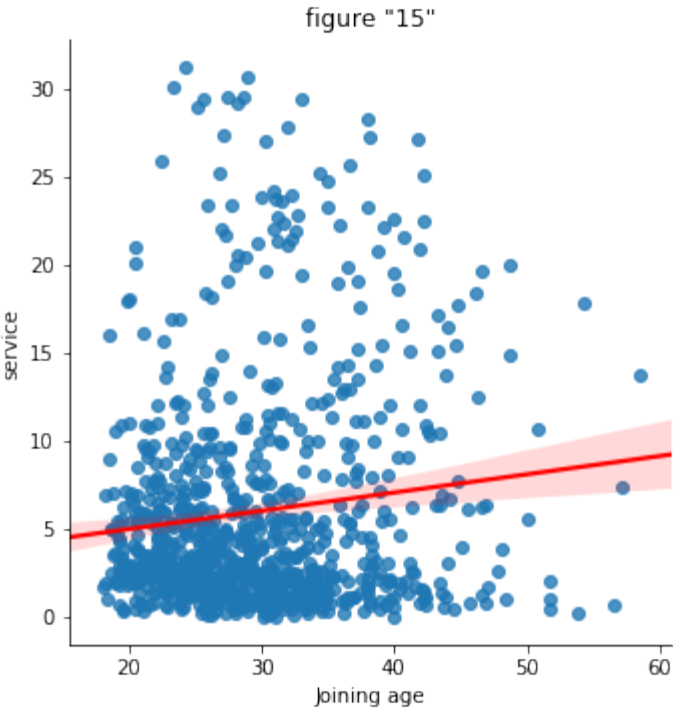
- The relation between joining age and service

```
Out[49]: Text(0.5, 1.0, 'figure "14"')
```



```
Out[50]:
```

	service	join_age
service	1.000000	0.116607
join_age	0.116607	1.000000



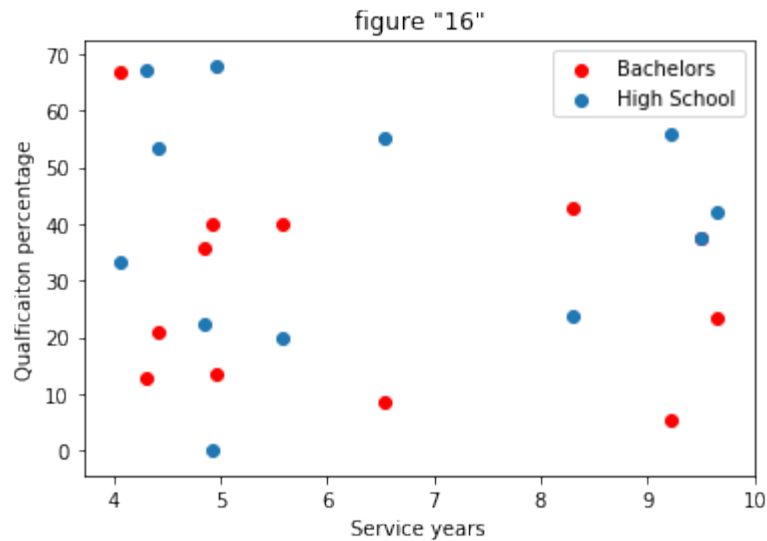
From the corrleation coefficient matrix and plot figure 14 & 15 , we notice that there is a weak relationship between joining age and service years

Out[51]:

	Bachelors	Diploma	High School	Masters	Primary School	service
dept						
Consumer Electronics & Home Appliances	20.952381	7.619048	53.333333	0.952381	0.0	4.420587
Consumer Products	12.765957	2.127660	67.021277	3.191489	0.0	4.311615
Directors Office	40.000000	0.000000	0.000000	20.000000	0.0	4.930411
Finance	42.857143	4.761905	23.809524	14.285714	0.0	8.293020
General Products	13.580247	4.938272	67.901235	3.703704	0.0	4.966447
HR & Administration	37.500000	6.250000	37.500000	6.250000	0.0	9.503082
Home Appliances Service Center	5.333333	18.666667	56.000000	0.000000	4.0	9.215379
IT	66.666667	0.000000	33.333333	0.000000	0.0	4.063927
Logistics	8.510638	2.127660	55.319149	1.063830	11.0	6.539172
Office Equipment & Business Solutions	40.000000	18.000000	20.000000	4.000000	0.0	5.573753
Office Equipment Service Center	35.820896	31.343284	22.388060	0.000000	1.0	4.853854
Office Furniture & Furnishing	23.437500	6.250000	42.187500	4.687500	8.0	9.658048

```
The correlation between service and bachelor degree:
[[ 1.          -0.20074178]
 [-0.20074178  1.          ]]
The correlation between service and high school degree:
[[1.          0.02212787]
 [0.02212787 1.          ]]
```

```
Out[53]: Text(0.5, 1.0, 'figure "16"')
```



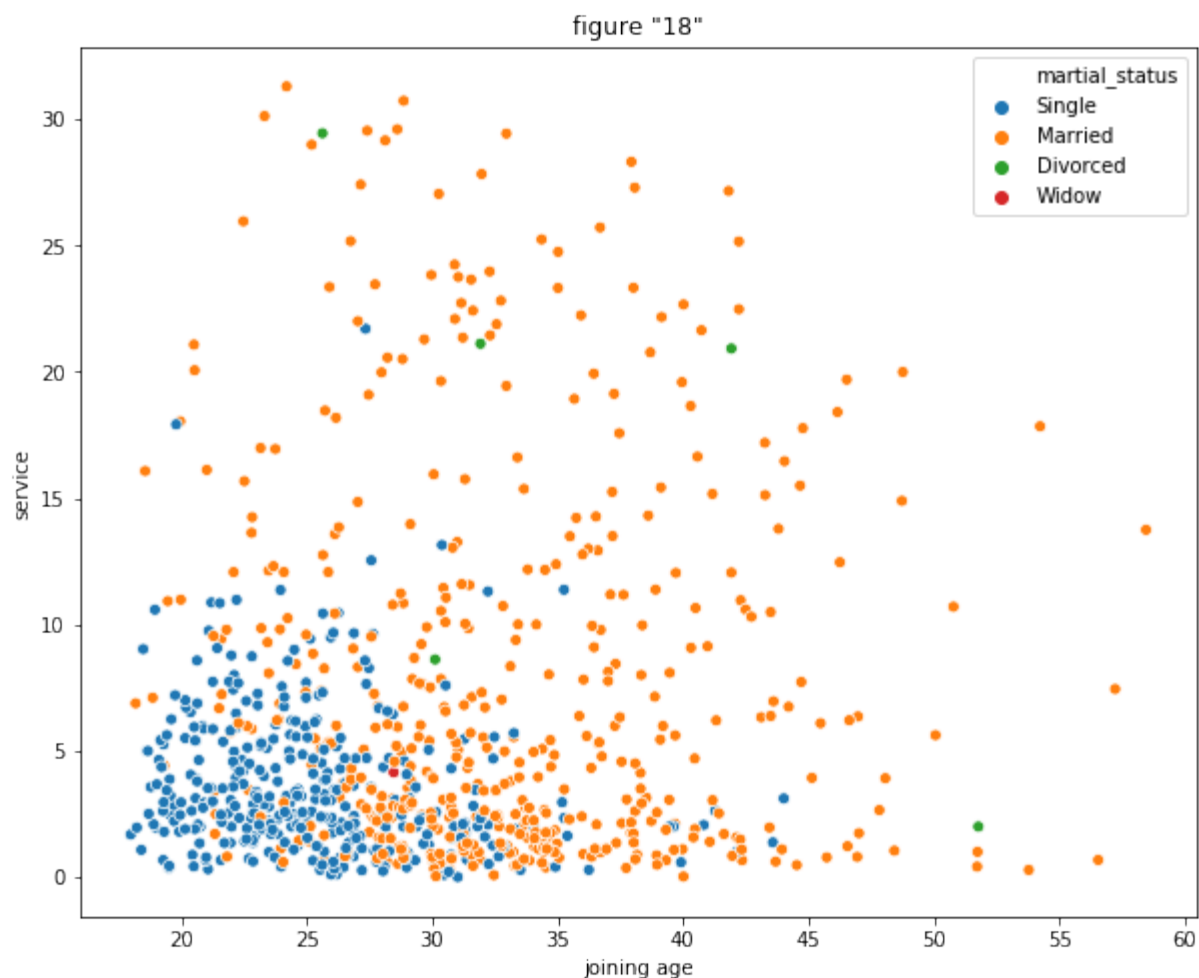
**From the corrlaetion coefficient matrix and plot figure 16 , we notice that there is a weak relationship between qualification and service years**

```
Out[54]: Text(0.5, 1.0, 'figure "17"')
```



**We observe in figure 17: the people who join at young age mostly they have high schiik and diploma, specially at the age of 22 or less, then after the age of 25 we see significant increase in the number of employees that have Bachelor degree and Masters**

```
Out[55]: Text(0.5, 1.0, 'figure "18"')
```

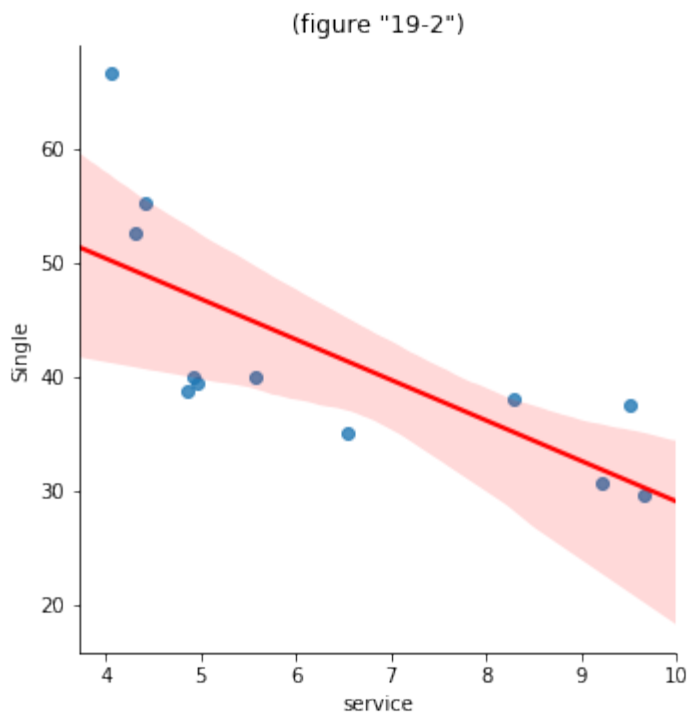
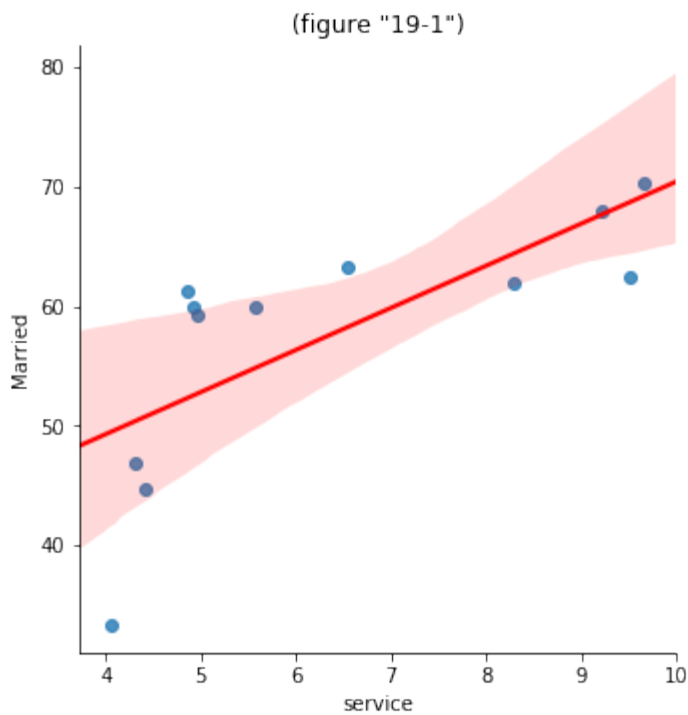


From the previous figure 18, it is obvious that people who join early they are single and their service period looks lower than those who are married. The thing that may indicate a relation between service period and marital status. We should dig deeper for more about this relation.

we will join the average service period for department and the department marital status percentages

we will see if there is a relation between higher avg service for each department associated with marital status percentage and we will do it with married and single columns

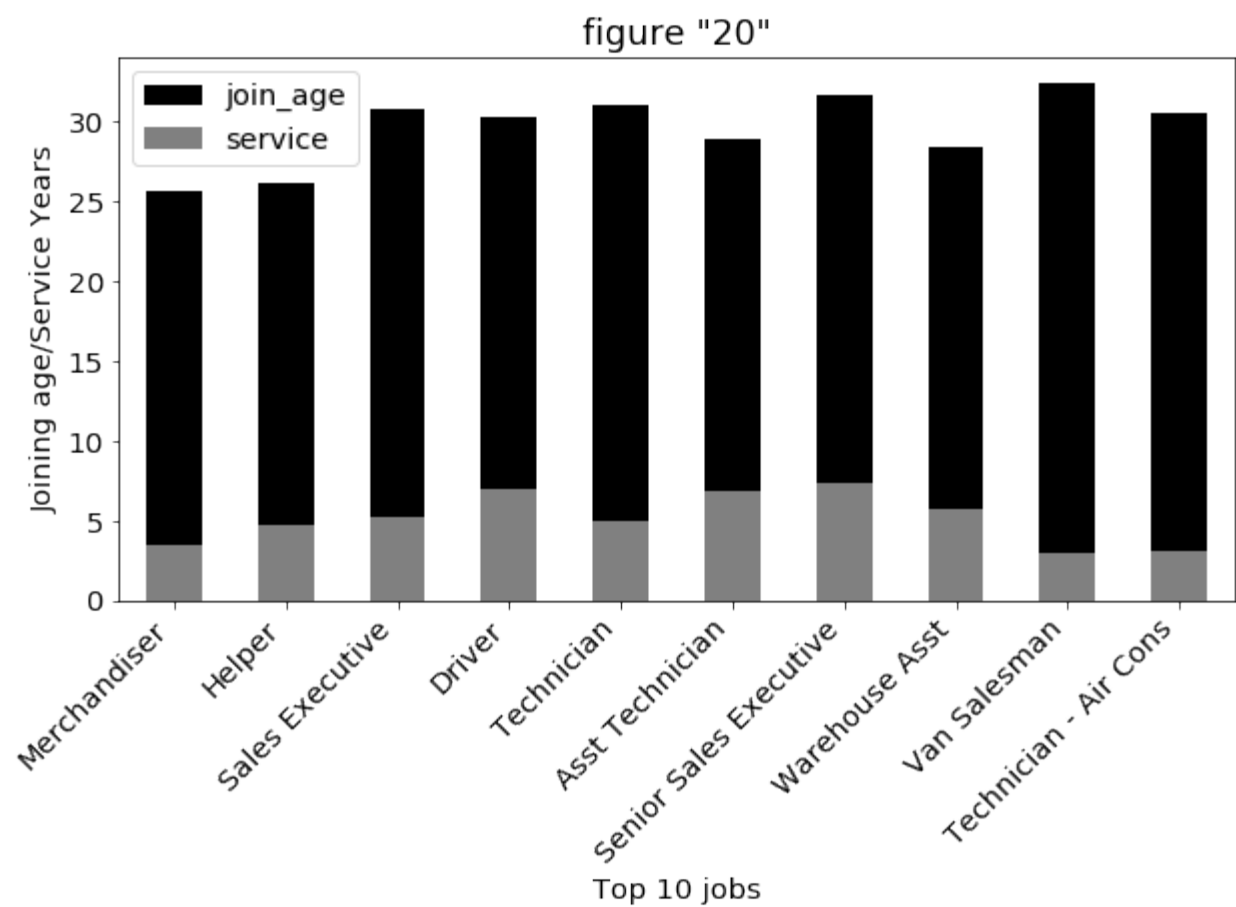
```
Marriage correlation [[ 1.          -0.36802241]
 [-0.36802241  1.          ]]
Single correlation [[1.          0.37096525]
 [0.37096525  1.          ]]
```



**From the previous figures 19-1 & 19-2, there is a moderate relationship between service period and martial status.**

**Department with higher average service period tend to have higher married rate of employees and the opposite is true, that departments with high rate single employees tend to have less average service period.**

```
Out[62]: Text(0.5, 1.0, 'figure "20"')
```



## **Conclusion:**

- There are some jobs which require quick actions and to be ready for them with immediate candidates and there are some jobs that employees spend longer time in them.
- We can notice that there are some departments that have active jobs and people move in and out quickly and there are some departments that they stand still.

## **Final Notes:**

- We did not analyse joining years as the analysis is about exit and also we do not have all needed data for employees who joined because not all employees joined left, however, this may give us a better idea about the joining age for each job and the employees who leave.
- This analysis can be improved by collecting more features like proposed salary, benefits and more features that may give better understanding for the people who leave because of low salary for example.