

jupyter exploratory\_analysis Last Checkpoint: Yesterday at 15:04 (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 Logout

## What make developers stay longer in our current jobs ?

```
In [1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from data_helpers import bucketize_last_hire_date
from data_helpers import transform_coding_experience
from data_helpers import filter_employed_developers

%matplotlib inline
%config InlineBackend.figure_format = 'retina'
sns.set(color_codes=True)
```

### Loading data

```
In [2]: df = pd.read_csv('data/stackoverflow-2019/survey_results_public.csv')
df.head()
```

Out[2]:

	Respondent	MainBranch	Hobbyist	OpenSourcer	OpenSource	Employment	Country	Student	EdLevel	UndergradMajor	...	WelcomeChange
0	1	I am a student who is learning to code	Yes	Never	The quality of OSS and closed source software ...	Not employed, but looking for work	United Kingdom	No	Primary/elementary school	NaN	...	Just as welcome now as I felt last year
1	2	I am a student who is learning to code	No	Less than once per year	The quality of OSS and closed source software ...	Not employed, but looking for work	Bosnia and Herzegovina	Yes, full-time	Secondary school (e.g. American high school, G...	NaN	...	Just as welcome now as I felt last year
2	3	I am not primarily a developer, but I write co...	Yes	Never	The quality of OSS and closed source software ...	Employed full-time	Thailand	No	Bachelor's degree (BA, BS, B.Eng., etc.)	Web development or web design	...	Just as welcome now as I felt last year
3	4	I am a developer by profession	No	Never	The quality of OSS and closed source software ...	Employed full-time	United States	No	Bachelor's degree (BA, BS, B.Eng., etc.)	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year
4	5	I am a developer by profession	Yes	Once a month or more often	OSS is, on average, of HIGHER quality than pro...	Employed full-time	Ukraine	No	Bachelor's degree (BA, BS, B.Eng., etc.)	Computer science, computer engineering, or sof...	...	Just as welcome now as I felt last year

5 rows × 85 columns

```
In [3]: df.shape
```

Out[3]: (88883, 85)

```
In [4]: df2 = pd.read_csv('data/stackoverflow-2019/survey_results_schema.csv')
```

Out[4]:

Column	QuestionText
0 Respondent	Randomized respondent ID number (not in order ...
1 MainBranch	Which of the following options best describes ...
2 Hobbyist	Do you code as a hobby?
3 OpenSourcer	How often do you contribute to open source?
4 OpenSource	How do you feel about the quality of open sour...

### Part I: Does the job role make professionals stay longer in a job ?

First, we need to find in the data how can we calculate how long each professional stays in a job.

```
In [5]: df2[df2.QuestionText == 'When was the last time that you took a job with a new employer?']['Column'].values[0]
```

Out[5]: 'LastHireDate'

```
In [6]: df.LastHireDate.value_counts()
```

```
Out[6]: Less than a year ago           25844
1-2 years ago                      21365
More than 4 years ago                14594
3-4 years ago                      11597
NA - I am an independent contractor or self employed    3426
I've never had a job                  3028
Name: LastHireDate, dtype: int64
```

```
In [7]: df2[df2.QuestionText == 'Which of the following best describes your current employment status?']['Column'].values[0]
```

Out[7]: 'Employment'

```
In [8]: df.Employment.value_counts()
```

```
Out[8]: Employed full-time            64440
Independent contractor, freelancer, or self-employed  8511
Not employed, but looking for work                   5600
Employed part-time                                4469
Not employed, and not looking for work               3803
Retired                                              358
Name: Employment, dtype: int64
```

As we are interested only in analyzing professional who current work for a employer, we are going to filter our data set.

```
In [9]: p1_df = df.copy()
p1_df = filter_employed_developers(p1_df)
p1_df.shape
```

Out[9]: (68063, 85)

Now, we also have to find how to extract different job titles from the interviewed professionals.

```
In [10]: df2[df2.QuestionText == 'Which of the following describe you? Please select all that apply.']['Column'].values[0]
```

Out[10]: 'DevType'

```
In [11]: p1_df.DevType.value_counts().head()
```

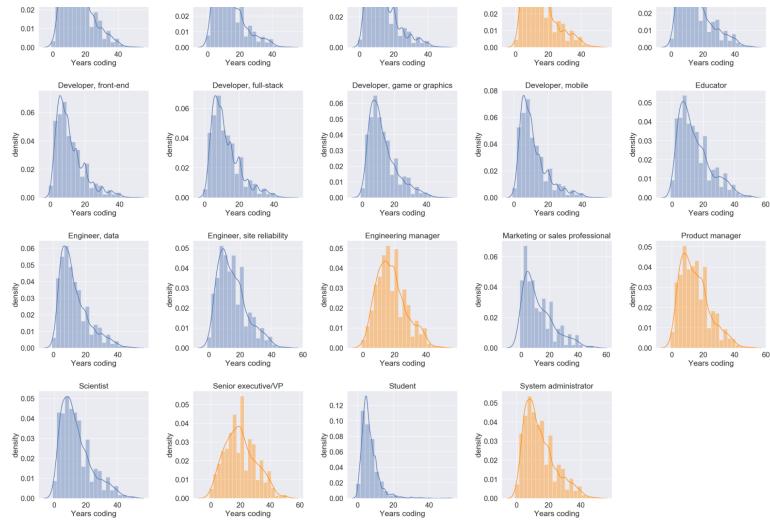
```
Out[11]: Developer, full-stack          7550
Developer, back-end                     4550
Developer, back-end;Developer, front-end;Developer, full-stack 2072
Developer, front-end                     1961
Developer, mobile                        1938
Name: DevType, dtype: int64
```

We definitely started to understand our data! We have information about how long professionals are in their current jobs and we can analyze it by job title (different professionals). However, we need to transform the data in order to use it properly.

### Data preparation

First, we need to convert `LastHireDate` to discrete values:





#### Conclusions

- Years coding distribution is very similar for job roles in which most of professionals are from **less than 1 year** in their jobs
- For professionals who are longer in their jobs (orange plots), we can observe for management roles, the interviewed professionals are more experienced.

## Part II: Wishing to become a manager makes professionals stay longer in a job ?

```
In [17]: df2[df2.QuestionText == 'Do you want to become a manager yourself in the future?']['Column'].values[0]
Out[17]: 'MgrWant'

In [18]: pl_df['MgrWant'].value_counts()
Out[18]: Not sure    60504
No        54326
Yes      44987
I am already a manager    9434
Name: MgrWant, dtype: int64

In [19]: manager_df = pl_df[pl_df['MgrWant'].isin(['Yes'])].groupby(['JobRole']).size() / pl_df.groupby(['JobRole']).size()
manager_df.sort_values(ascending=False).head()
Out[19]: JobRole
Developer, mobile           0.284702
Designer                     0.276505
Data or business analyst     0.269758
Academic researcher          0.261189
Marketing or sales professional 0.250000
dtype: float64
```

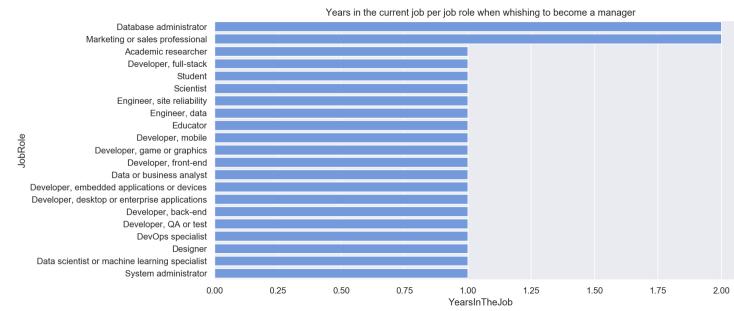
We can observe, professionals who are **Developer mobile**, **Designer**, **Data or business analyst**, **Academic researcher** or **Marketing or sales professional** are more willing to become a manager. Let's check then weather they are more willing to stay in their current jobs.

Let's get started by filtering our current result set by professionals who current are not a manager but want to become one.

```
In [20]: p2_df = pl_df[(pl_df.MgrWant == 'Yes') &
                   (~pl_df.JobRole.isin(['Senior executive/VP', 'Product manager', 'Engineering manager']))]
p2_df.shape
Out[20]: (44343, 87)

In [21]: agg_p2_df = p2_df.groupby(['JobRole']).agg({'YearsInTheJob': lambda x: x.value_counts().index[0]})\
          .sort_values(by='YearsInTheJob', ascending=False)\
          .reset_index()

plt.figure(figsize=(12,6))
sns.barplot(x='YearsInTheJob', y='JobRole', orient='h', data=agg_p2_df, color='cornflowerblue');
plt.title('Years in the current job per job role when whishing to become a manager');
```



#### Conclusions

It seems that wishing to become a manager in the future does not make professionals stay longer in their current jobs. The mode for most of professionals also concentrate in **1 (less than 1 year)**. Maybe this is different for experienced professionals. Let's analyze professionals with at least 10 years of experience

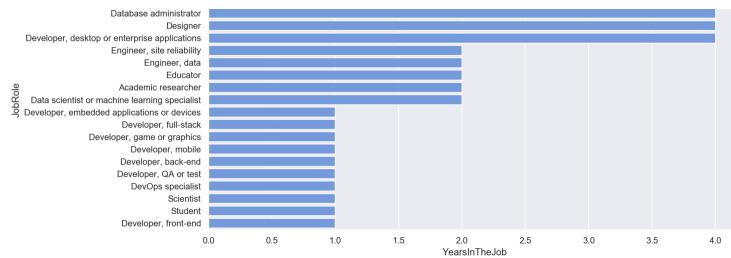
```
In [22]: df2[df2.QuestionText == 'Including any education, how many years have you been coding?']['Column'].values[0]
Out[22]: 'YearsCode'

In [23]: p2a_df = p2_df[p2_df.YearsCode >= 10]
p2a_df.shape
Out[23]: (17886, 87)

In [24]: agg_p2a_df = p2a_df.groupby(['JobRole']).agg({'YearsInTheJob': lambda x: x.value_counts().index[0]})\
          .sort_values(by='YearsInTheJob', ascending=False)\
          .reset_index()

plt.figure(figsize=(12,6))
sns.barplot(x='YearsInTheJob', y='JobRole', orient='h', data=agg_p2a_df, color='cornflowerblue');
plt.title('Years in the current job per job role when whishing to become a manager (> 10 years coding)');
```





#### Conclusions

It seems that more experienced professionals who wish to become a manager in the future stay longer in their jobs!