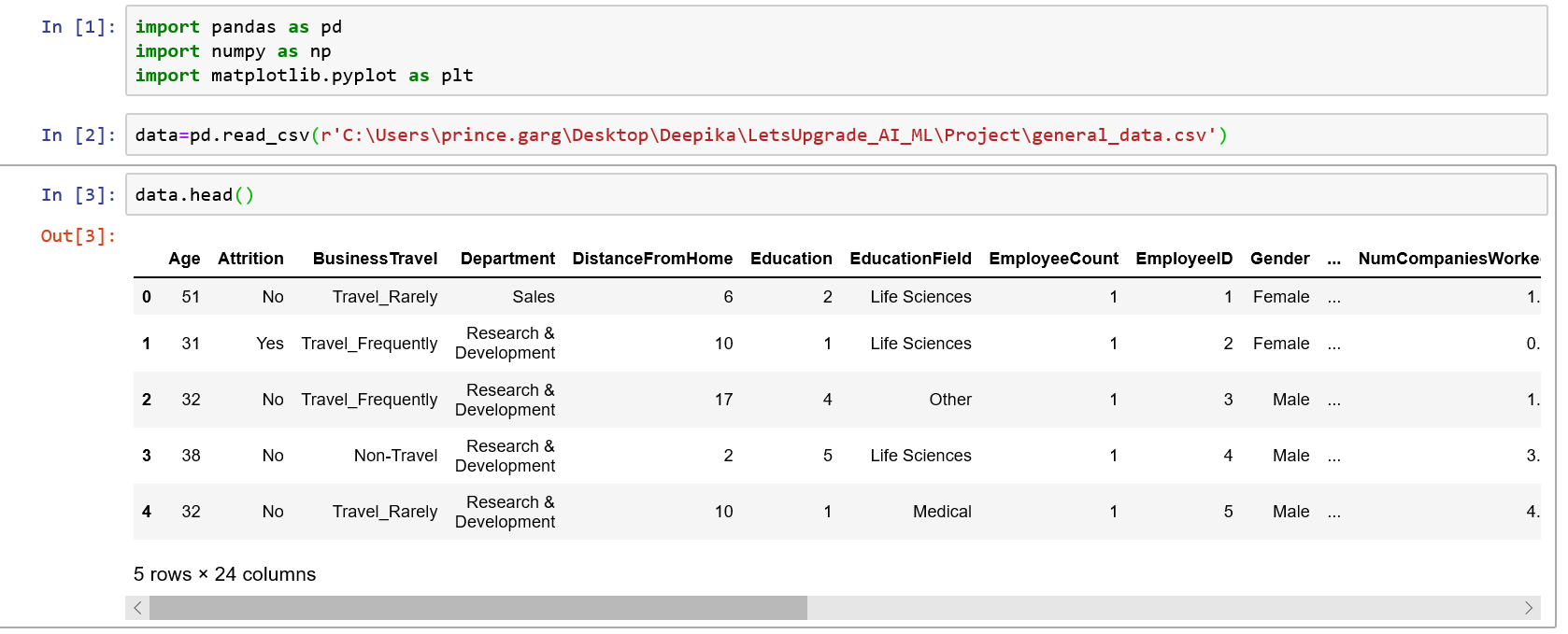
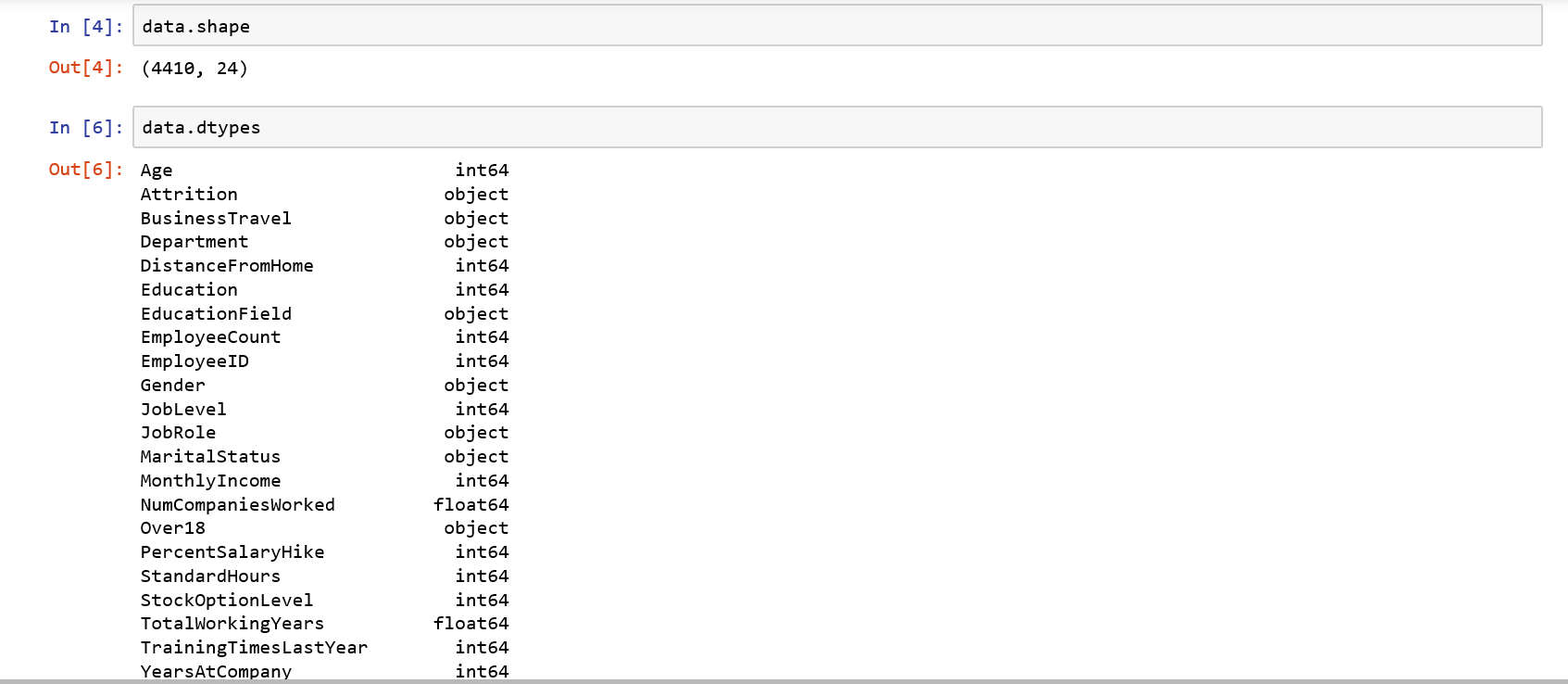
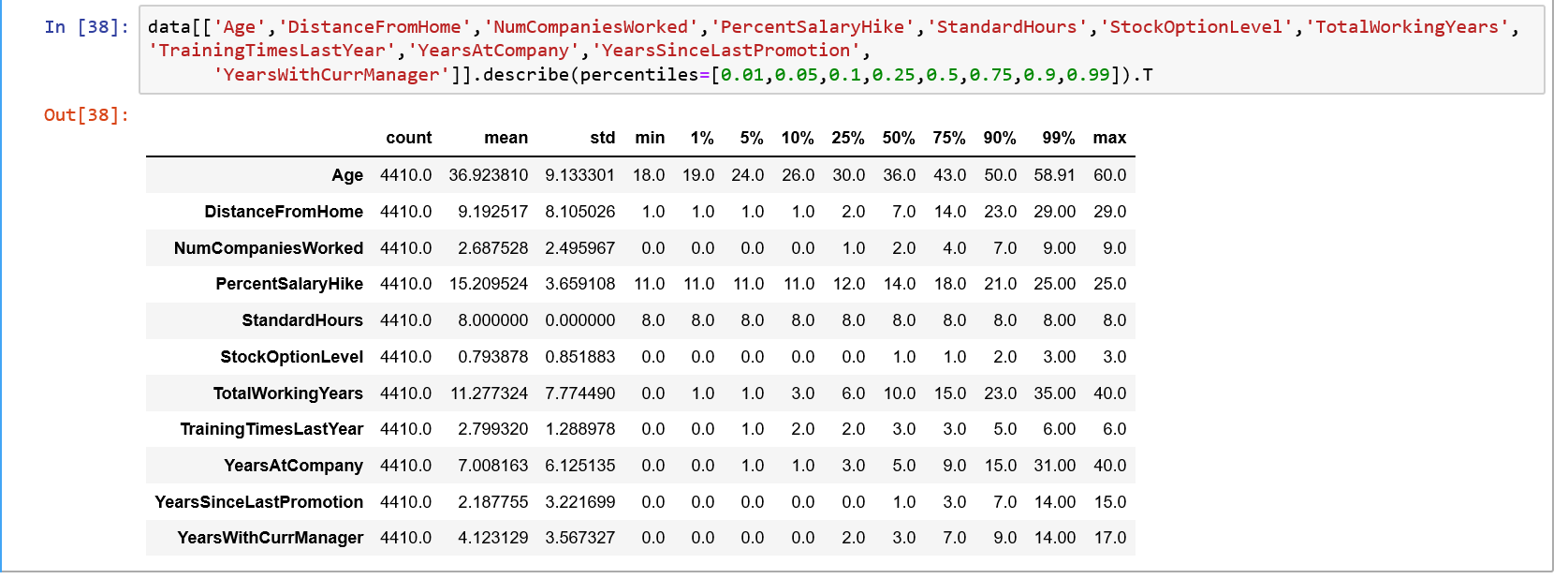
**Attrition Project**



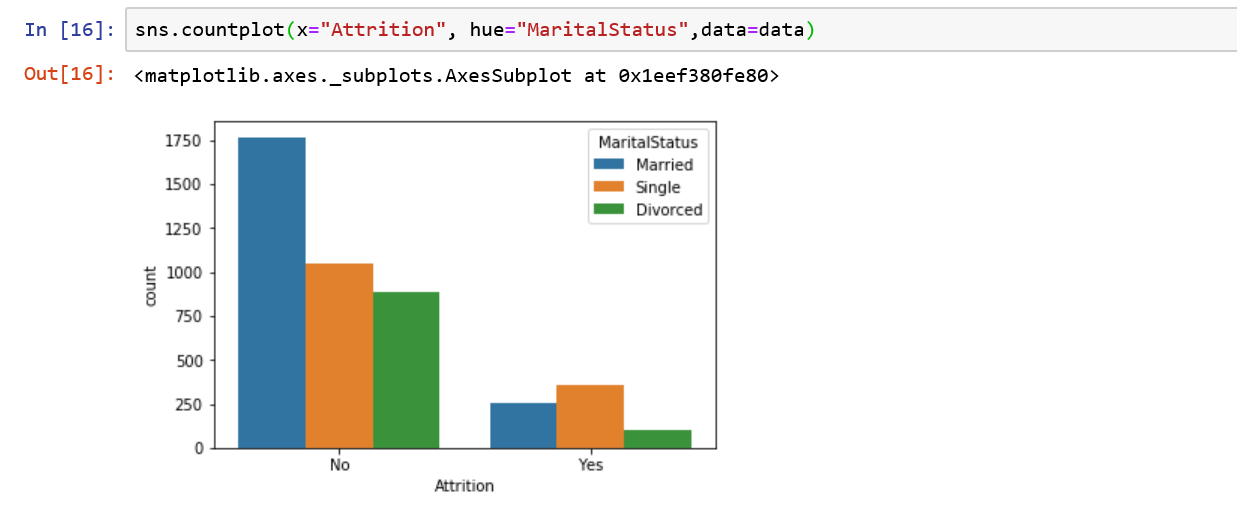
 No. of rows = 4416

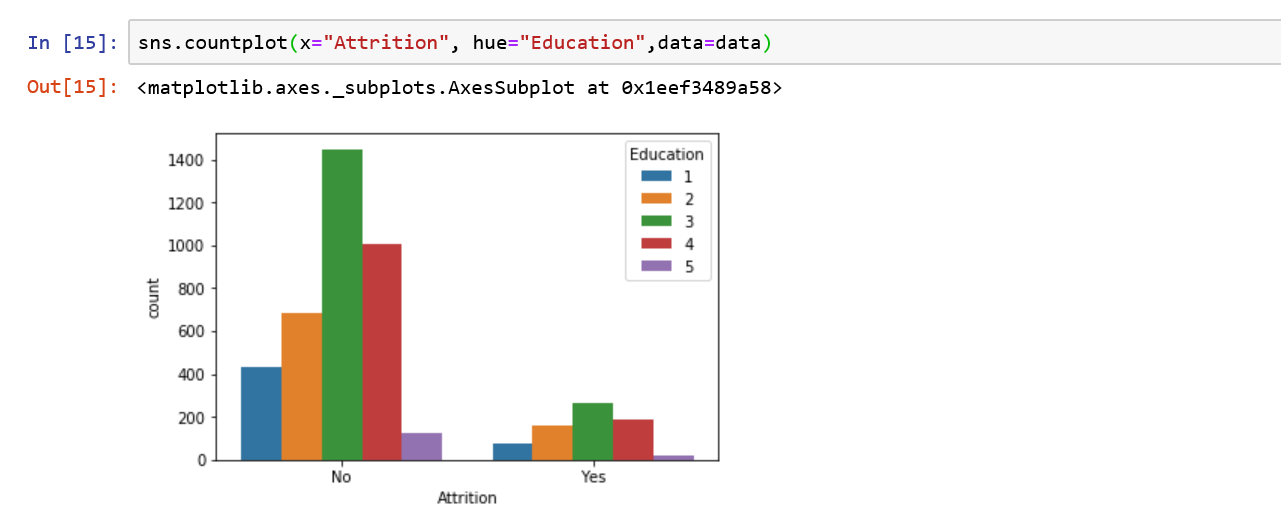
No. of columns = 24

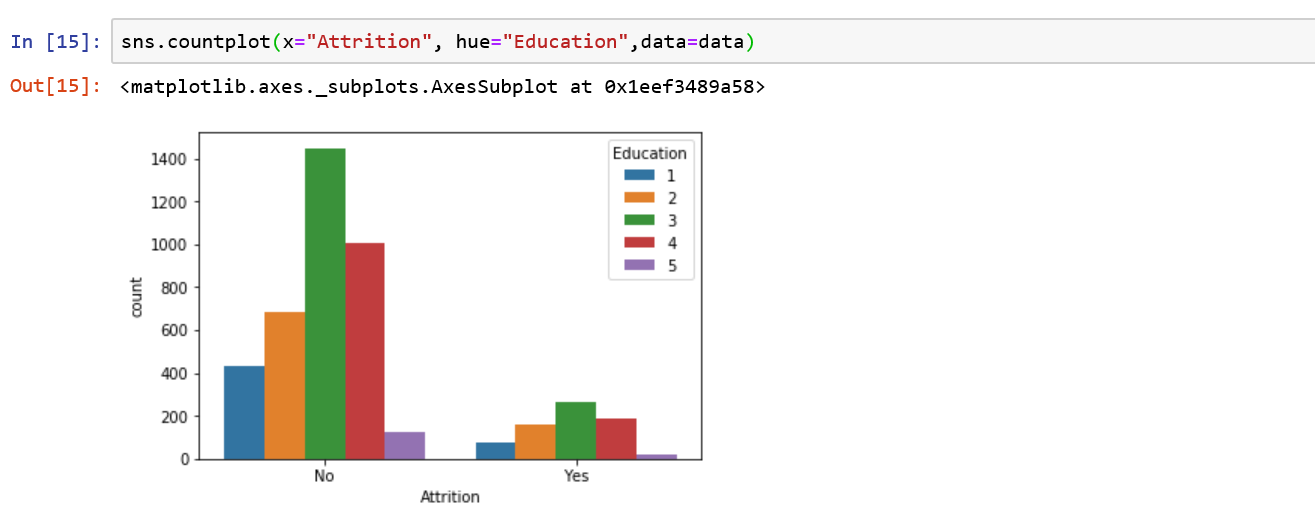
***Univariate Analysis – Numerical Features:***

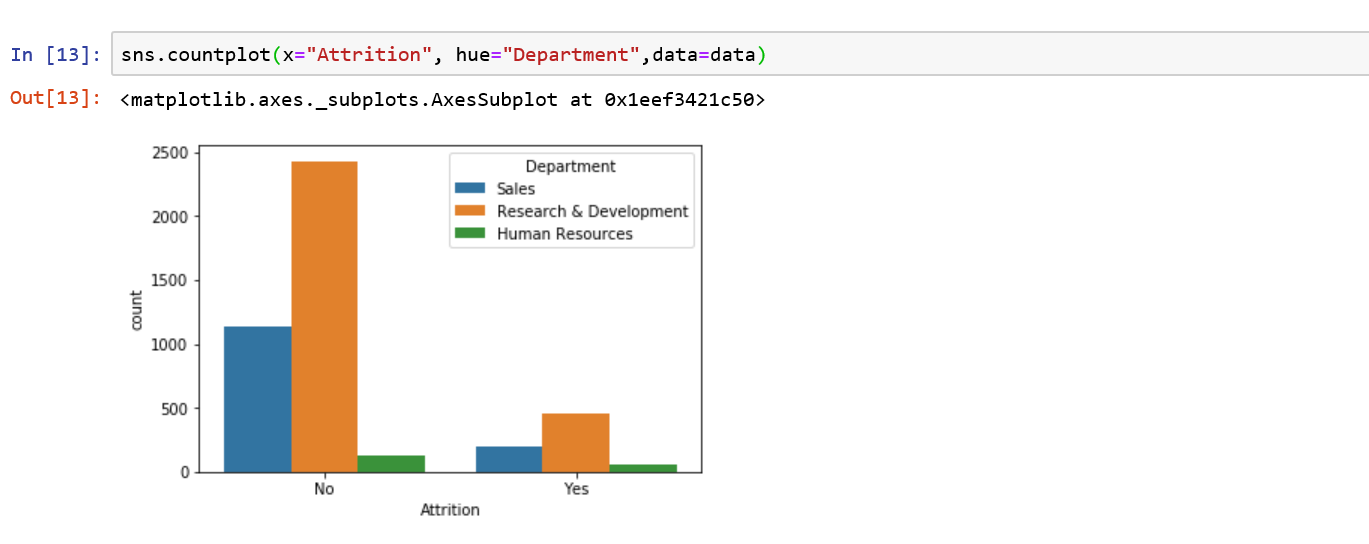
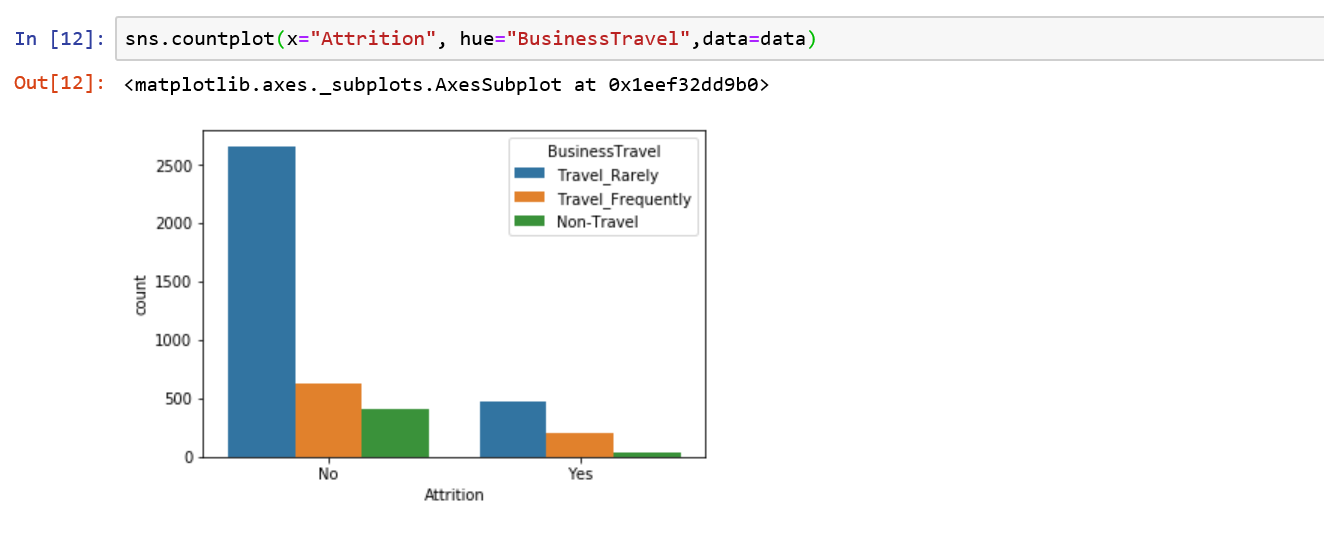


***Univariate Analysis – Categorical Features:***

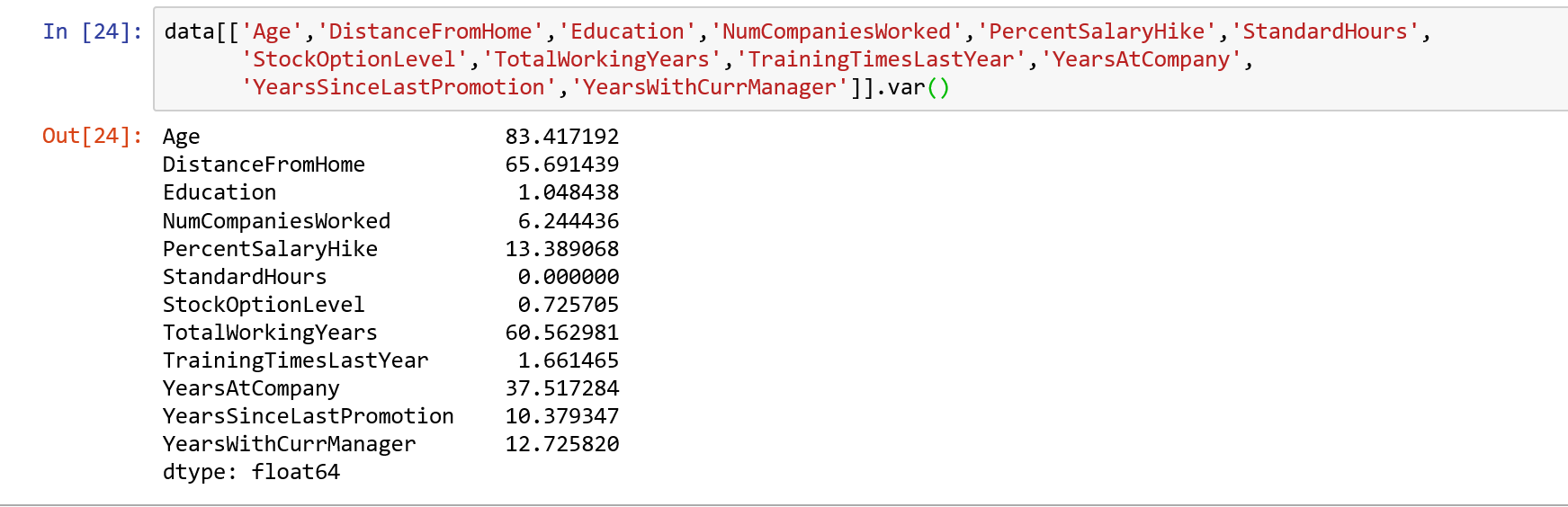


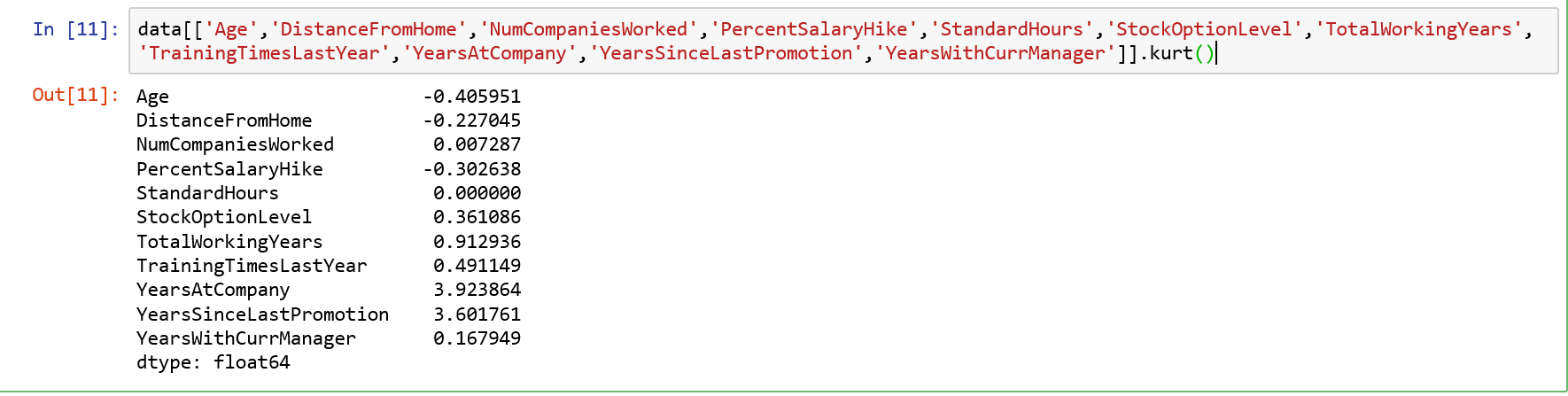
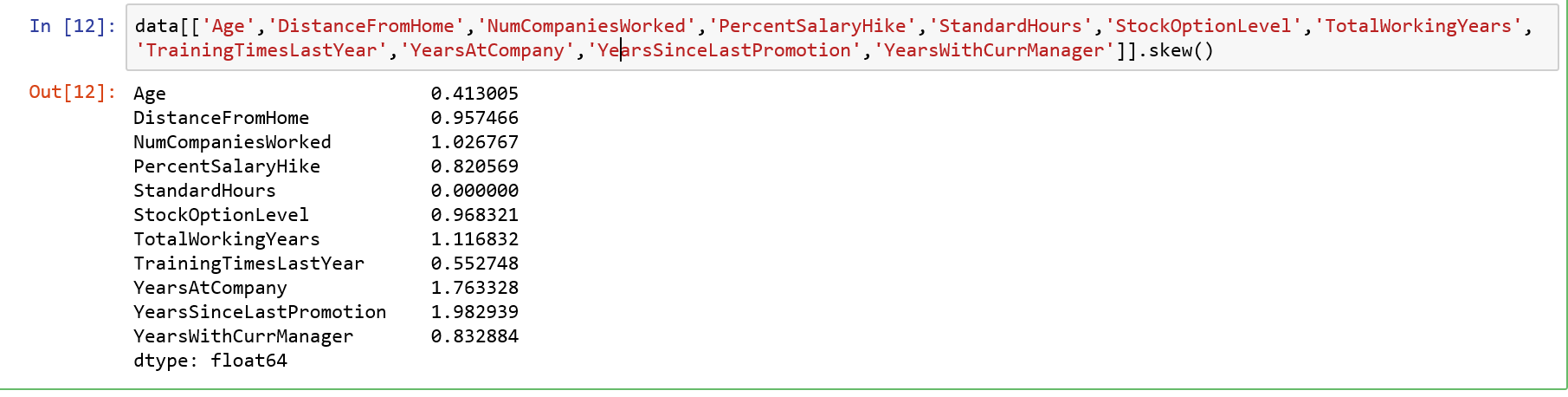








* *All above variables have positive values.*
* *This means all are positively skewed.*
* *mean > median*

***Checking Missing Values:***

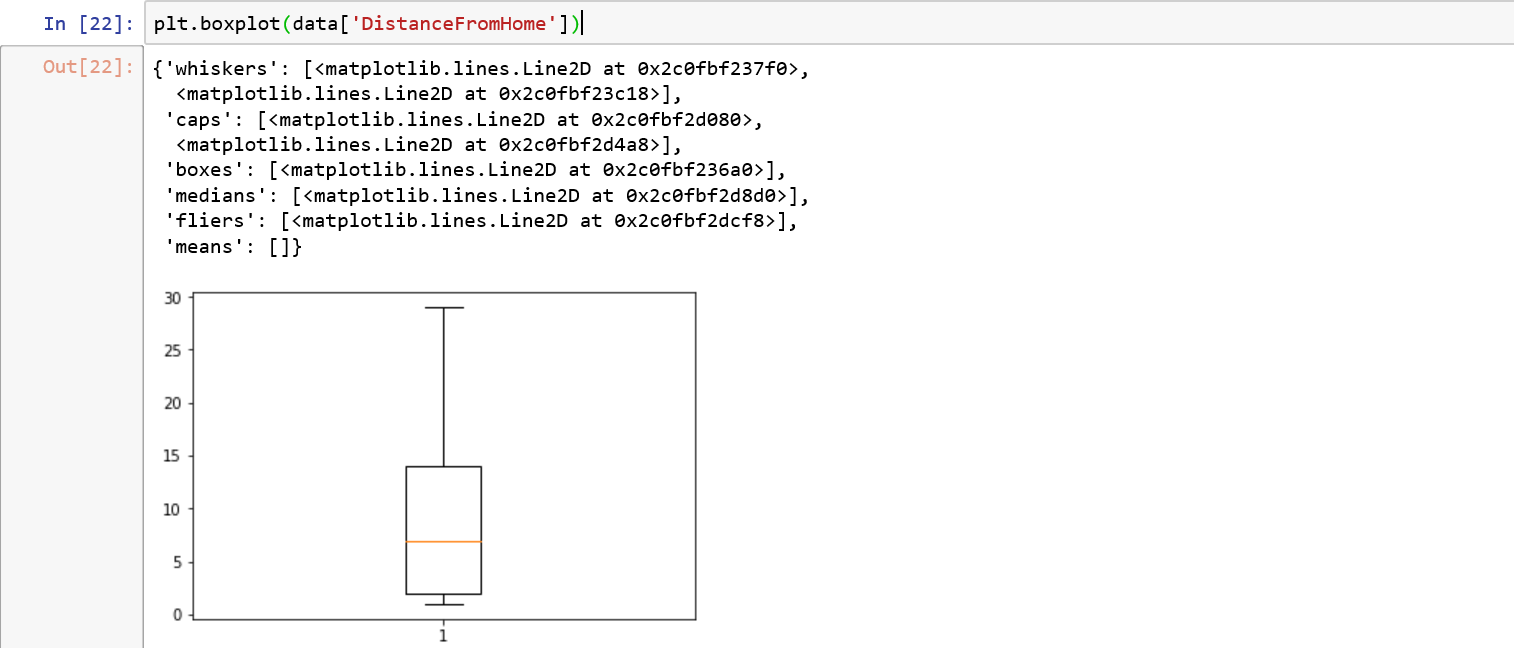


*As we can see only 2 variables has missing values.*

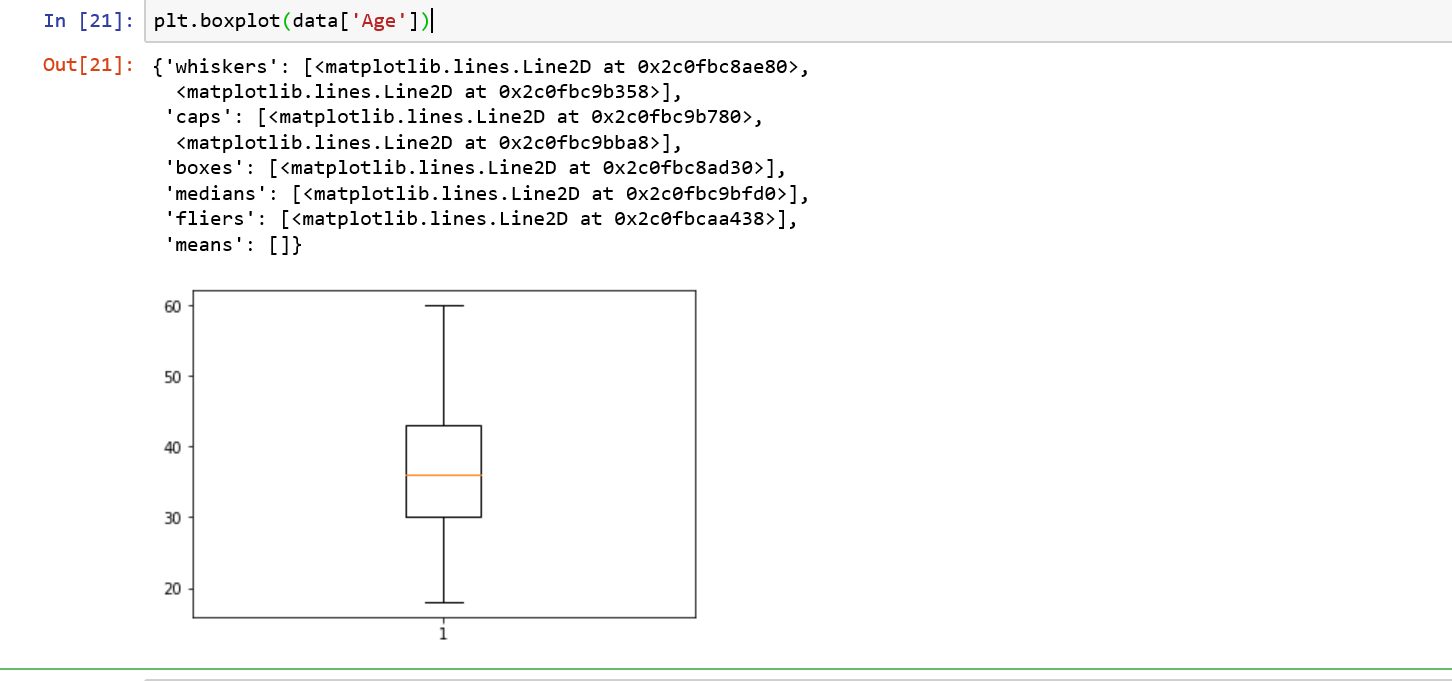
***Missing Values Imputation:***



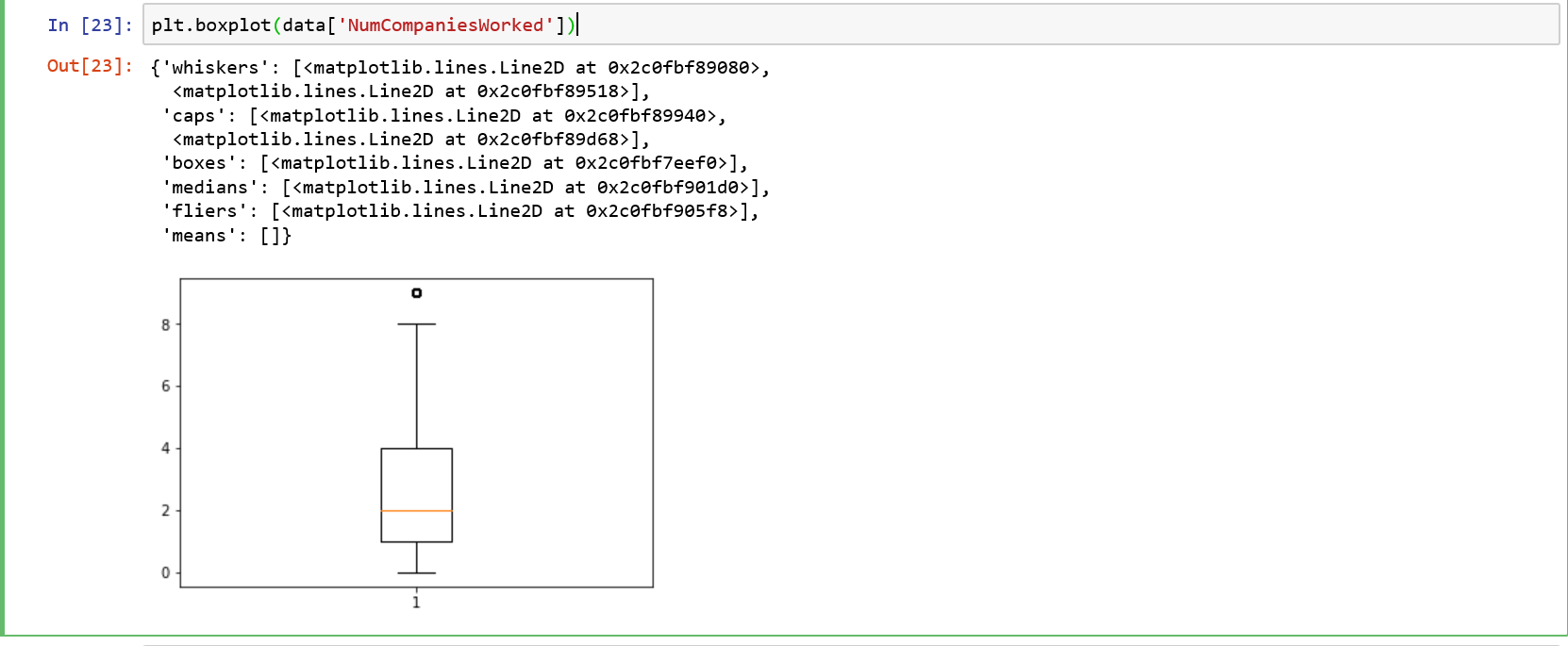
***Outlier Detection:***

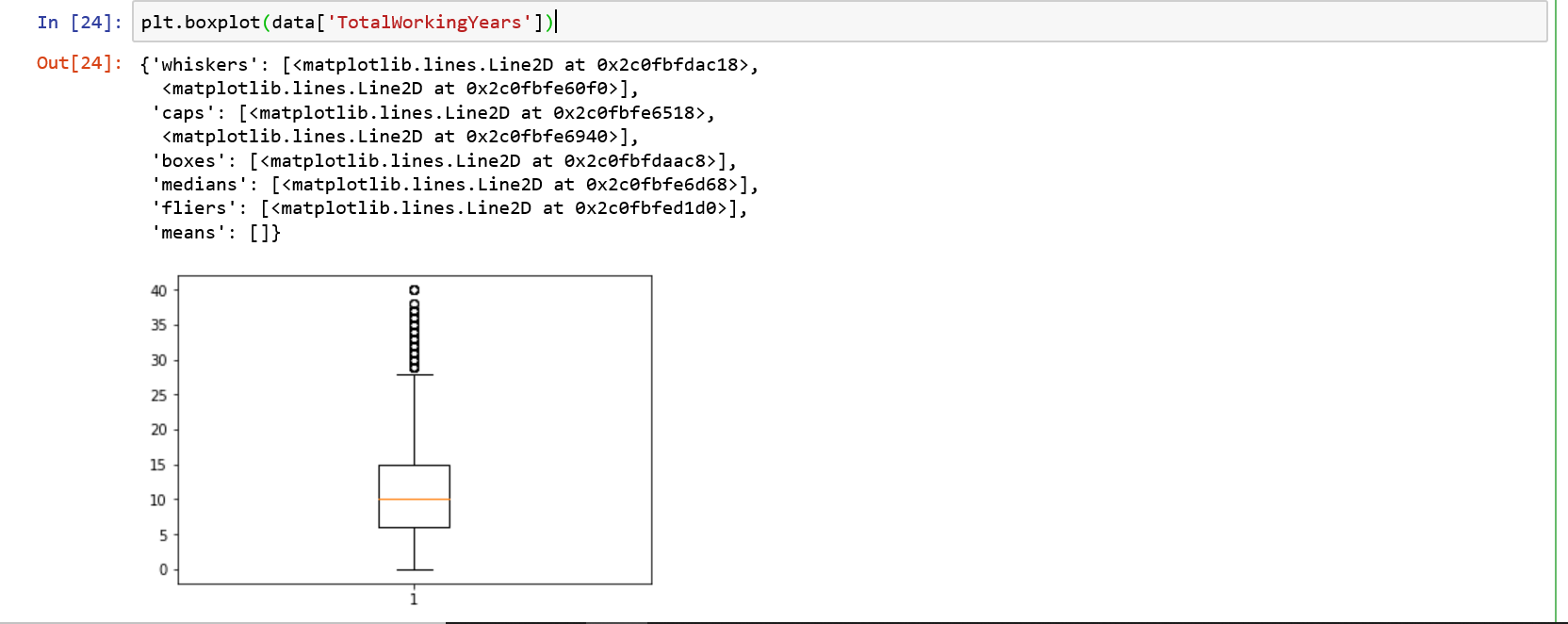


We can observe that for variable ‘DistanceFromHome’ there are no bubbles which means no outliers.

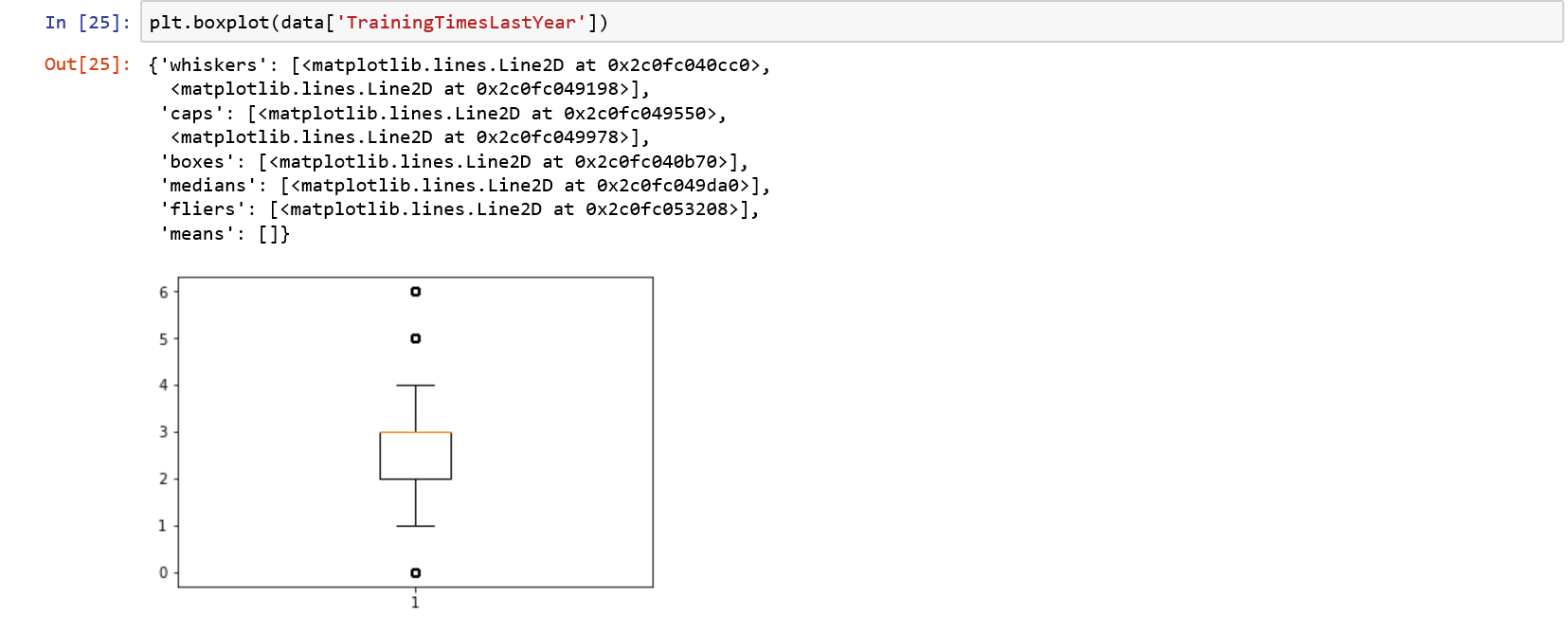


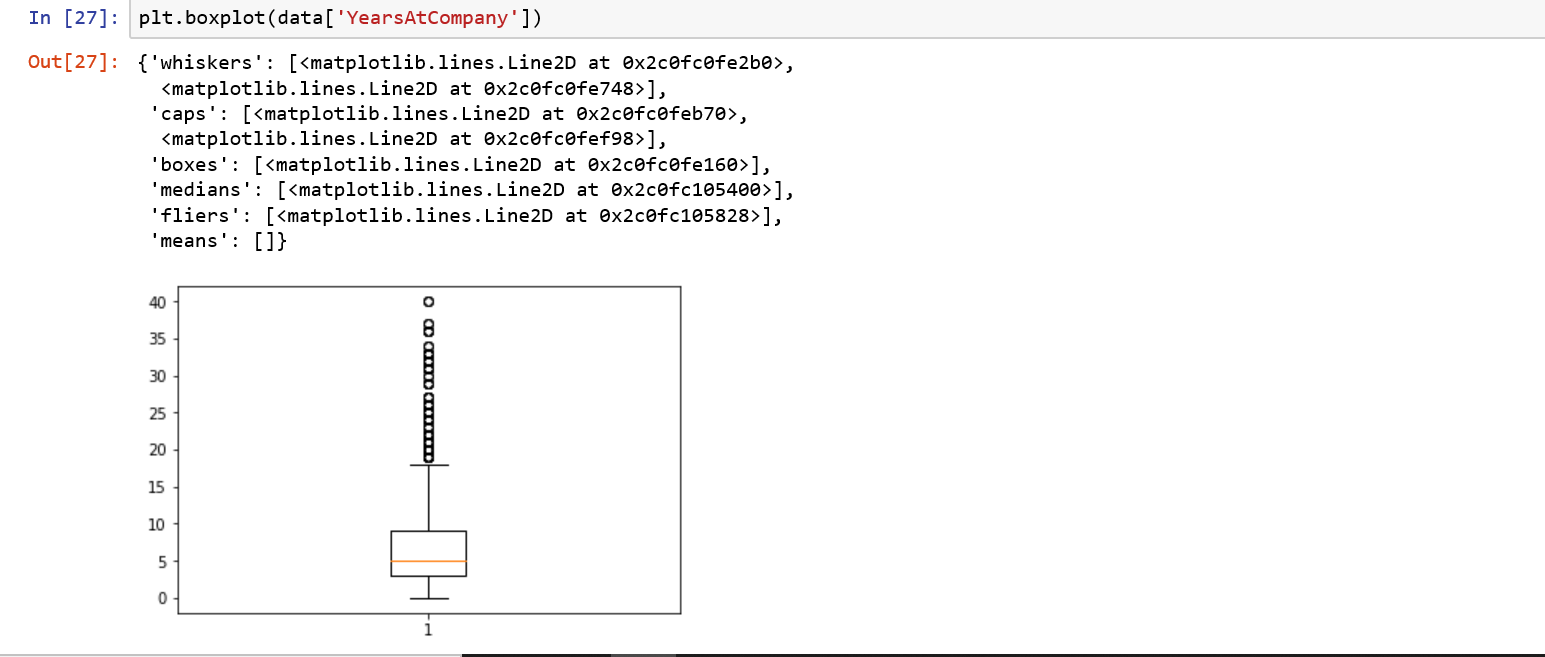
No outliers



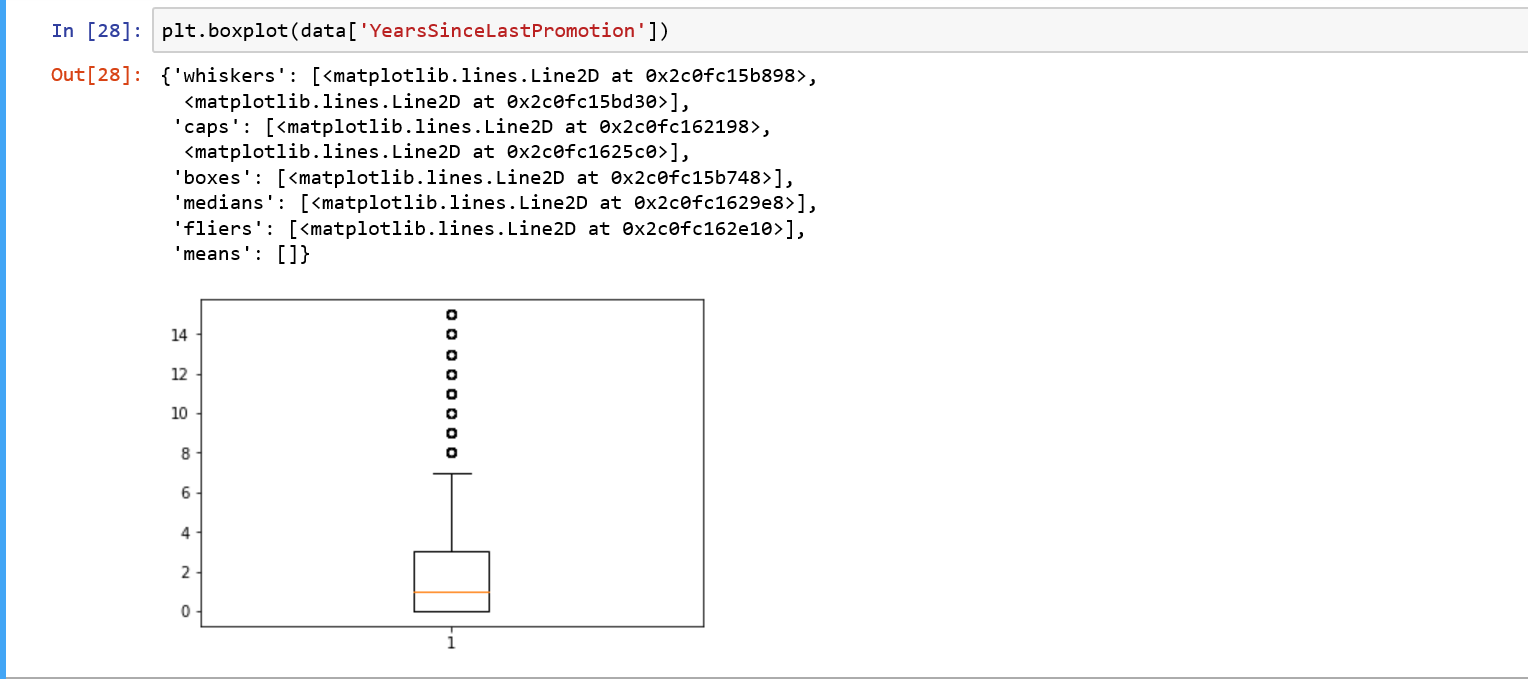


From above plot, we can see there are lot of bubbles which means the variable “Total Working Years” has outliers.

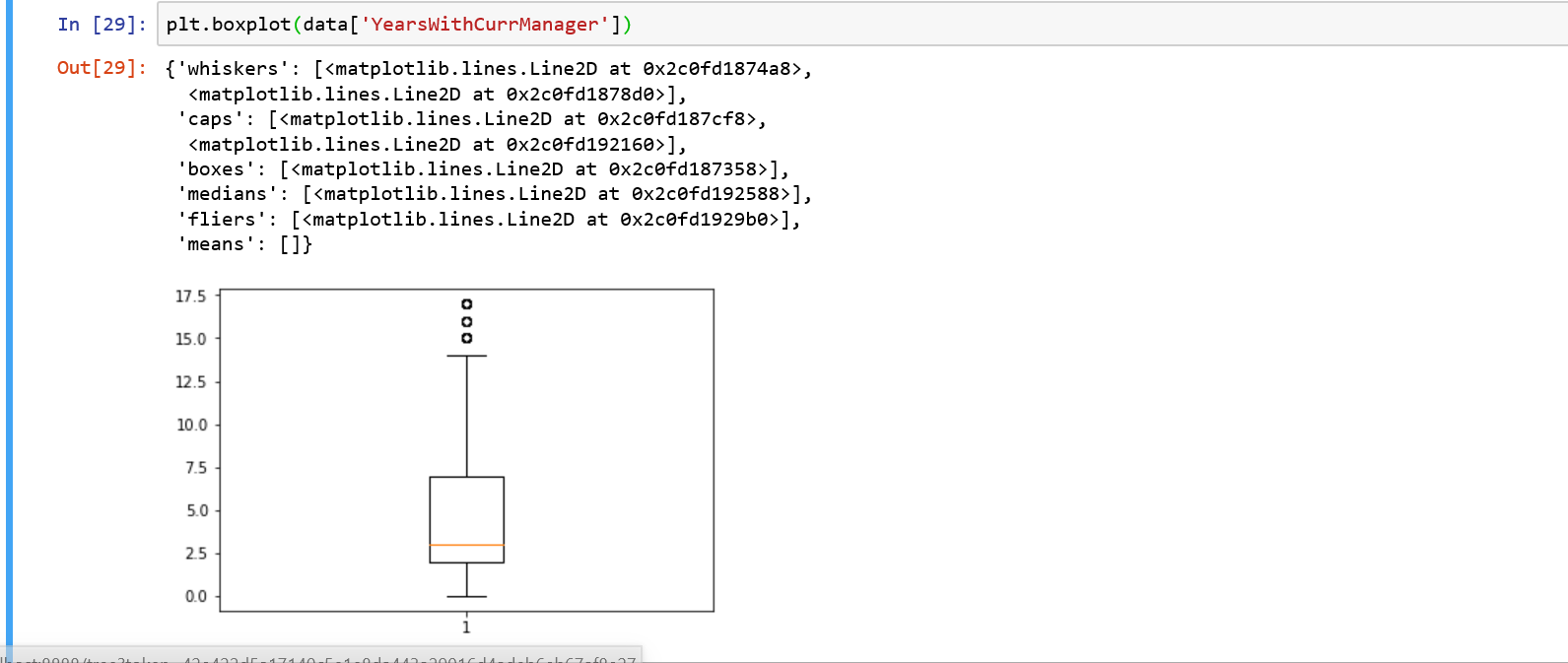




From above plot, we can see there are lot of bubbles which means the variable “Years at company” has outliers.

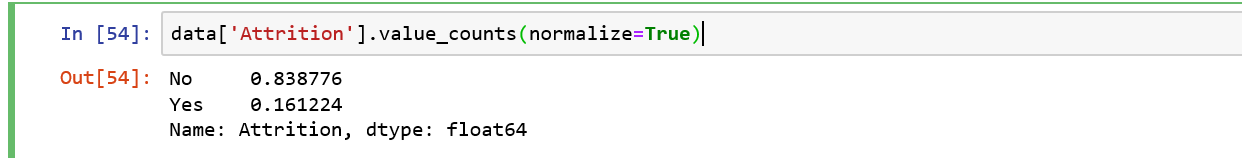


From above plot, we can see there are lot of bubbles which means the variable “Years Since last promotion” has outliers.



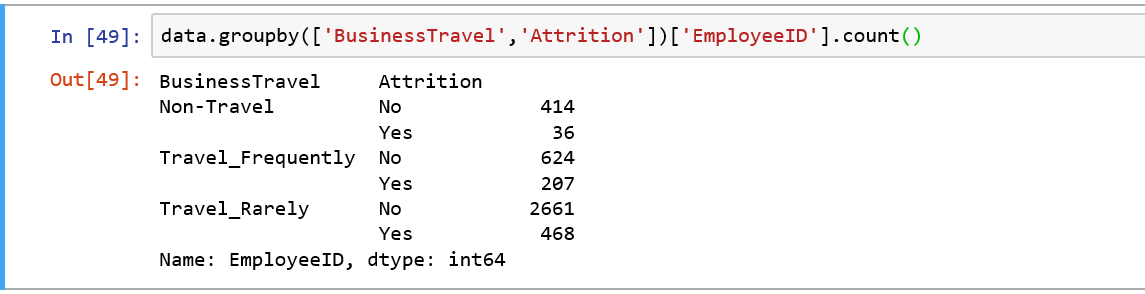
**Potential Factors for Attrition**

**Overall Attrition rate for the given population is 16.1%**



**Hypothesis 1:** Employee who **travels frequently** have high chances of attrition.

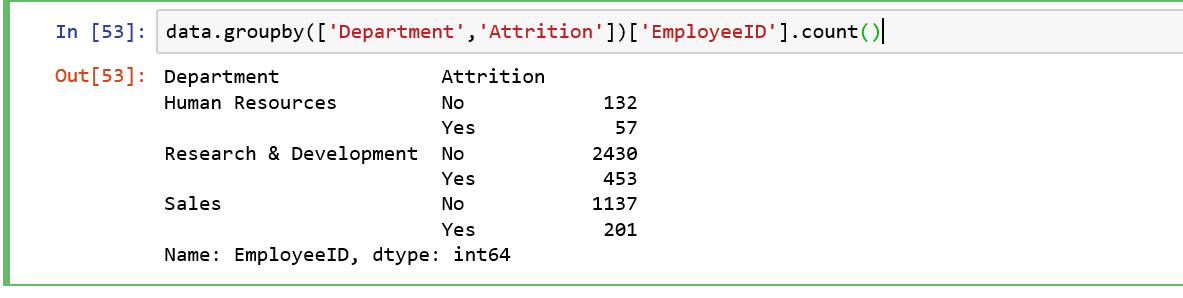
Attrition rate when travel frequently=207 / (207+624) = **24.9%**



**Hypothesis 2:** Employee from **Human Resources** department have double the chances of Attrition as compared to other departments, however the volume of Attrition is low.

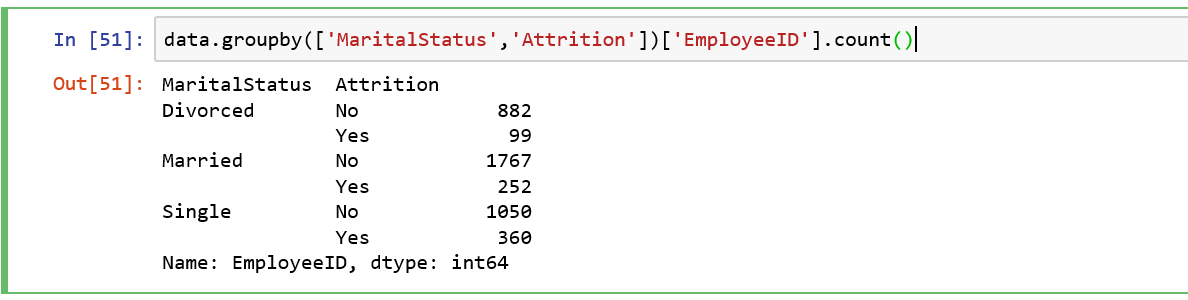
Attrition rate for HR department=57 / (57+132) = **30.1%**

Volume of Attrition for HR=57 / (57+453+201) = 8%



**Hypothesis 3**: **Marital Status** could be another factor for Attrition, **Single** Employees have higher attrition rate than others.

Attrition rate for Single Employees=360 / (360+1050)= **25.5%**



**Hypothesis 4**: Employees with **Educational Field** as **Human Resources** have higher chances of attrition, Technical degree and Other educational field have lower chances of Attrition.

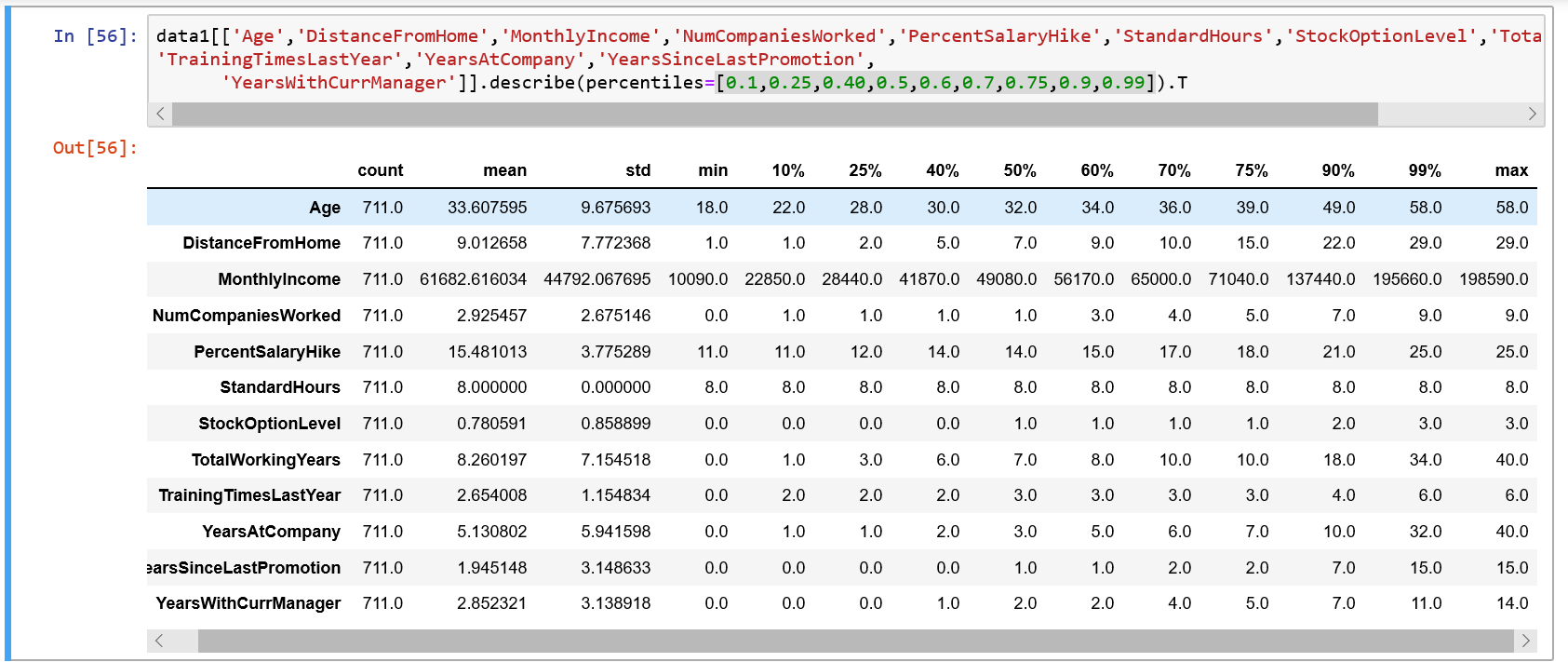
Attrition rate for HR=33 / (33+48) = **40.7%**

Attrition rate for Technical Degree=45 / (45+351) = **11.3%**

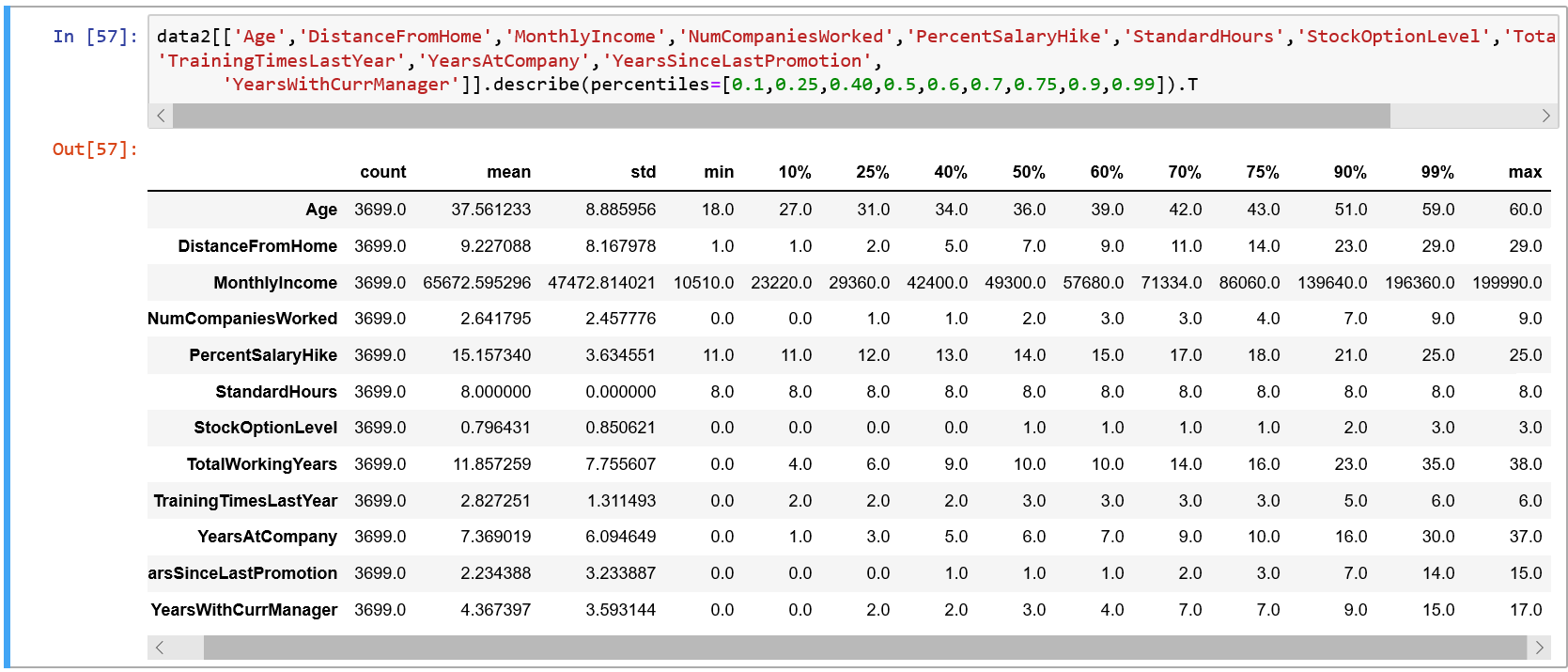
Attrition rate for other= 30 / (30+216) = **12.19%**



Univariate Analysis of all numerical variables where **Attrition = Yes**



Univariate Analysis for **Attrition = NO**



**Hypothesis 5: Age** could be one of the factor for Attrition.

Average age for Attrite Employees= **33.6**, Average age for non-Attrite Employees=**37.5**

**Hypothesis 6: Total working years** could justify reason for Attrition

Average of Total working years for Attrite Employees= **8.26**, Average of Total working years for non-Attrite Employees= **11.85**

**Hypothesis 7: Years at Company** could be another reason for Attrition

Average of Years at Company for Attrite Employees= **5.13**, Average of Years at Company for non-Attrite Employees= **7.36**

**Hypothesis 8: Years with current Manager** could show reason for Attrition

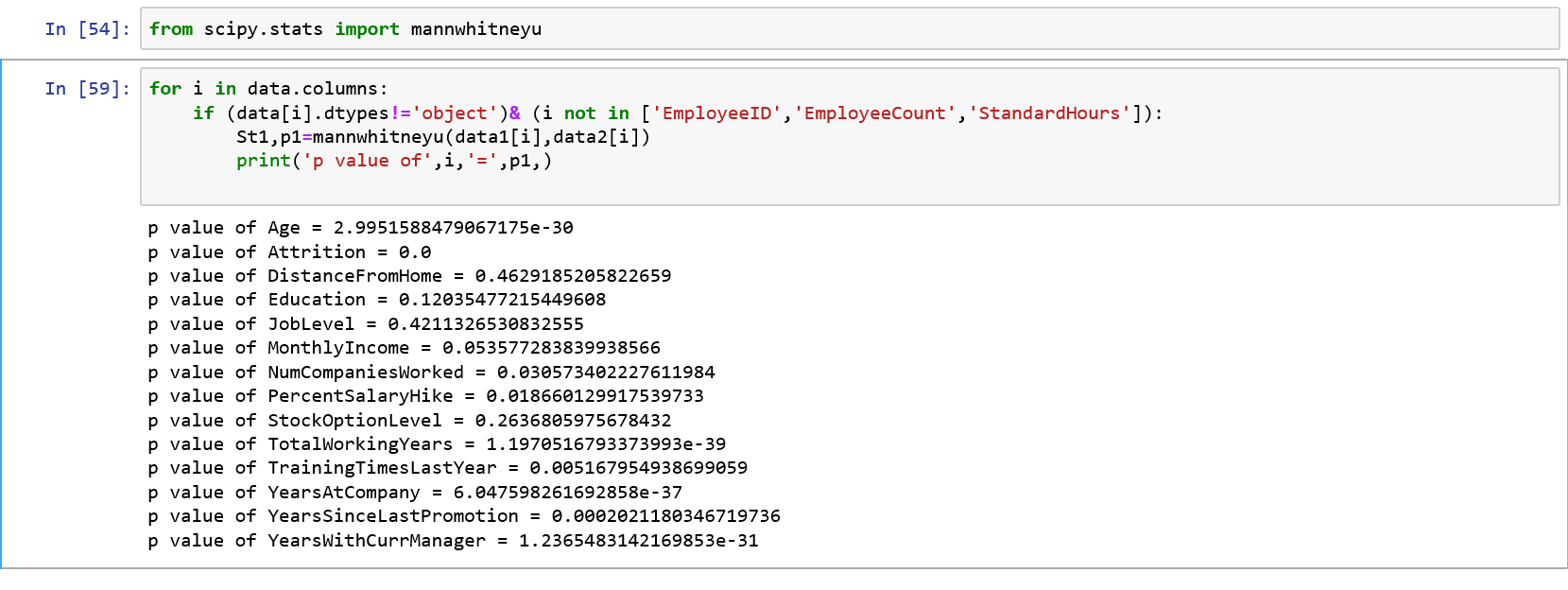
Average of Years with current Manager for Attrite Employees= 2.85, Average of Years with current Manager for non-Attrite Employees= 4.36

**Hypothesis 9:** **Gender** is not a significant factor for Attrition.

**Hypothesis 10:**

Variables like Distance from home, Monthly Income, Percent salary hike, Stock option level does not seem to have impact on Attrition because average for Attrite and non-attrite Employees are almost same.

**STASTICAL TESTS – Numerical Features**



**Group-1: Attrition=Yes Group-2: Attrition=No**

**Hypothesis 1: Age**

Null Hypothesis, Ho = There is no significant difference in Age for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Age for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 2: Distance from home**

Null Hypothesis, Ho = There is no significant difference in Distance from home for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Distance from home for Group-1 and Group-2

As p-value > 0.05, so we **accept** null hypothesis and **reject** Alternate Hypothesis.

**Hypothesis 3: Education**

Null Hypothesis, Ho = There is no significant difference in Education for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Education for Group-1 and Group-2

As p-value > 0.05, so we **accept** null hypothesis and **reject** Alternate Hypothesis.

**Hypothesis 4: Job level**

Null Hypothesis, Ho = There is no significant difference in Job level for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Job level for Group-1 and Group-2

As p-value > 0.05, so we **accept** null hypothesis and **reject** Alternate Hypothesis.

**Hypothesis 5: Monthly Income**

Null Hypothesis, Ho = There is no significant difference in Monthly Income for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Monthly Income for Group-1 and Group-2

As p-value >0.05, so we **accept** null hypothesis and **reject** Alternate Hypothesis.

**Hypothesis 6: Number of companies worked**

Null Hypothesis, Ho = There is no significant difference in Number of Companies worked for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Number of Companies worked for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 7: Percent Salary hike**

Null Hypothesis, Ho = There is no significant difference in Percent Salary hike for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Percent Salary hike for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 8: Stock Option Level**

Null Hypothesis, Ho = There is no significant difference in Stock Option level for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Stock Option level for Group-1 and Group-2

As p-value >0.05, so we **accept** null hypothesis and **reject** Alternate Hypothesis.

**Hypothesis 9: Total Working Years**

Null Hypothesis, Ho = There is no significant difference in Total Working Years for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Total Working Years for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 10: Training time last year**

Null Hypothesis, Ho = There is no significant difference in Training time last year for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Training time last year for Group-1 and Group-2

As p-value >0.05, so we **accept** null hypothesis and **reject** Alternate Hypothesis.

**Hypothesis 11: Years at Company**

Null Hypothesis, Ho = There is no significant difference in Years at Company for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Years at Company for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 12: Years since last promotion**

Null Hypothesis, Ho = There is no significant difference in Years since last promotion for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Years since last promotion for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 13: Years with current manager**

Null Hypothesis, Ho = There is no significant difference in Years with current manager for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Years with current manager for Group-1 and Group-2

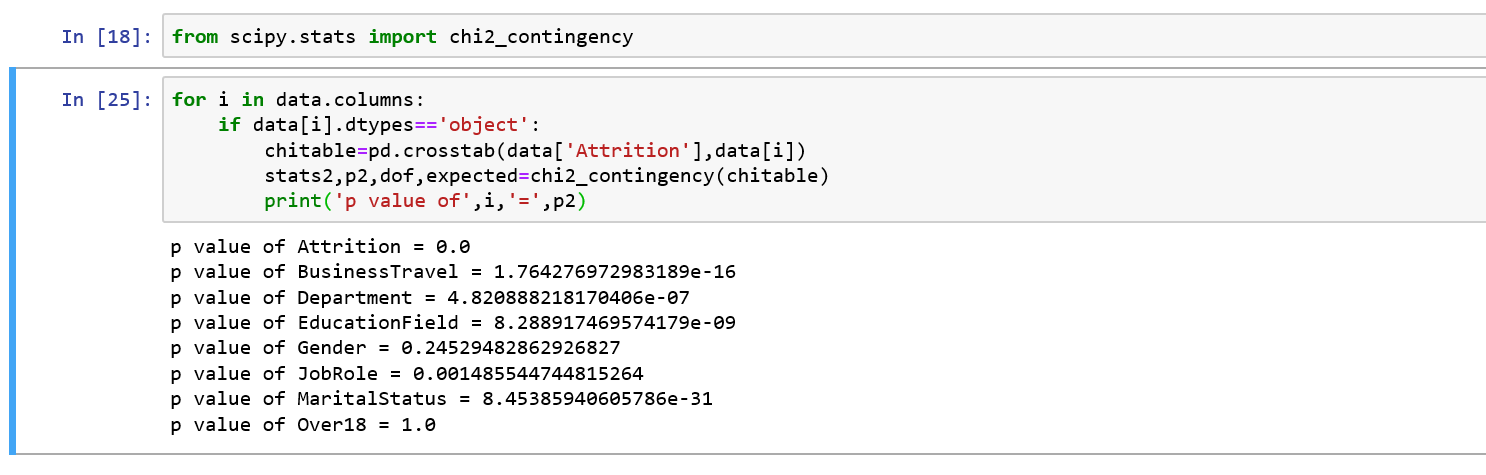
As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

The following features are statistically different for Attrition and non-Attrition group.

**Age, Number of Companies Worked, Percent Salary hike, Total Working Years, Years at Company, Years since last promotion, Years with current Manager.**

These features helps in finding **high Attrition** group.

**STASTICAL TESTS – Categorical Features**



**Hypothesis 1: Business Travel**

Null Hypothesis, Ho = There is no significant difference in Business Travel for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in Business Travel for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 2: Department**

Null Hypothesis, Ho = There is no significant difference in **Department** for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in **Department** for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 3: Education Field**

Null Hypothesis, Ho = There is no significant difference in **Education Field** for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in **Education Field** for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 4: Gender**

Null Hypothesis, Ho = There is no significant difference in **Gender** for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in **Gender** for Group-1 and Group-2

As p-value > 0.05, so we **Accept** null hypothesis and **Reject** Alternate Hypothesis.

**Hypothesis 5: Job Role**

Null Hypothesis, Ho = There is no significant difference in **Job Role** for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in **Job Role** for Group-1 and Group-2

As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

**Hypothesis 5: Marital Status**

Null Hypothesis, Ho = There is no significant difference in **Marital Status** for Group-1 and Group-2

Alternate Hypothesis, H1 = There is significant difference in **Marital Status** for Group-1 and Group-2

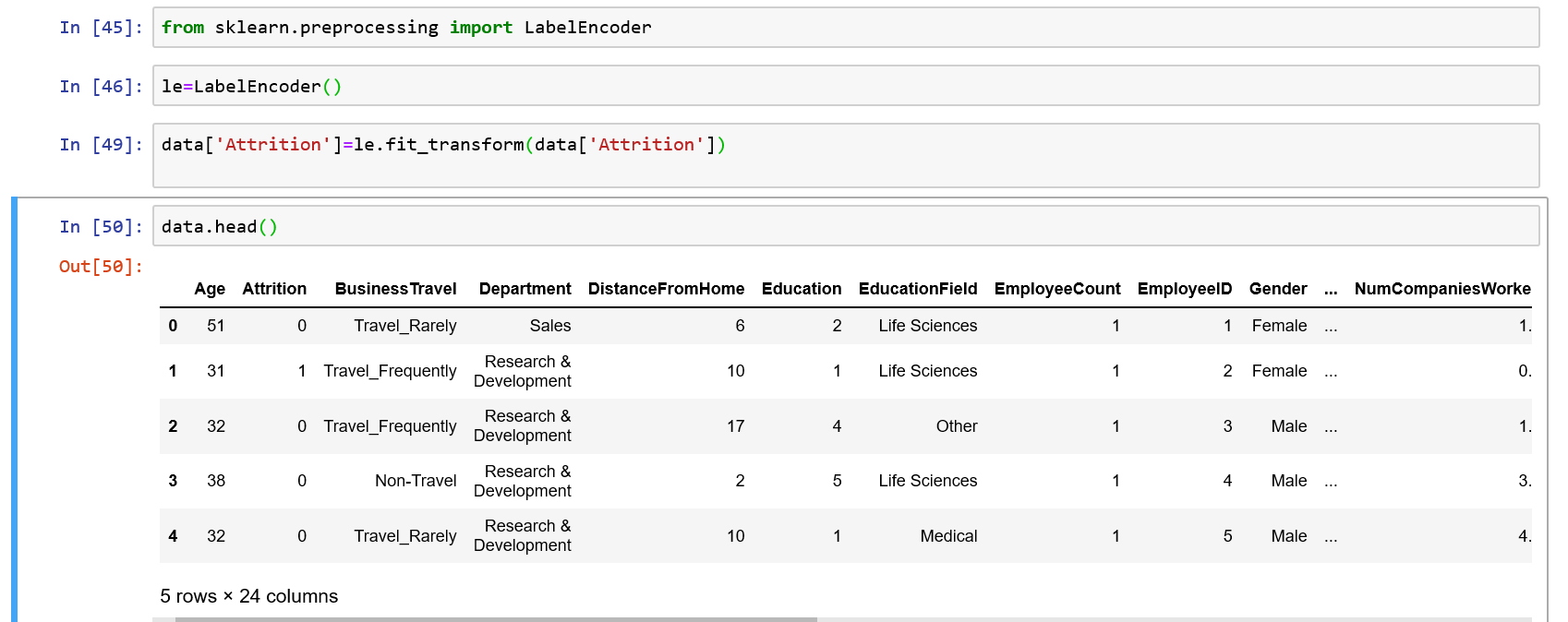
As p-value < 0.05, so we **reject** null hypothesis and **accept** Alternate Hypothesis.

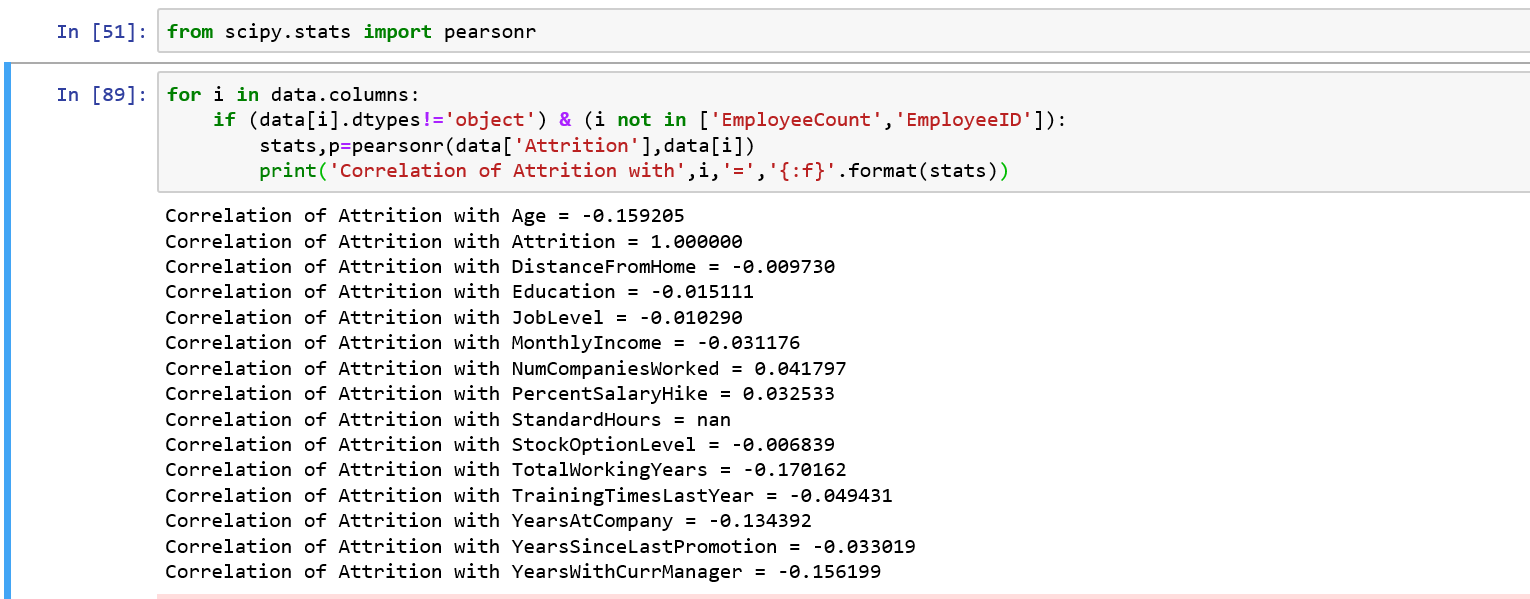
The following features are statistically different for Attrition and non-Attrition group.

**Marital Status, Job Role, Education Field, Business Travel, Department.**

These features helps in finding **high Attrition** group.

**Correlation**





All numerical variables are very less correlated with Attrition.