EDA ON HR ANAYLTICS

Identifing the best source of recruitment for a tech startup, based on previous data of candidate sources and recruitment strategies

Steps for Analysing the HR Analytics Data

- Problem statement understanding
- Data understanding
- Data preparation

Problem statement understanding:

- Identify the best source of recruitment for a tech startup, based on previous data of candidate sources and recruitment strategies.

Data Understanding:

1. The data is very small
2. The data is less imbalanced
3. The HR Analytics data is having only 4 columns and 446 rows
4. The attrition column is a continuous column and having only 0 and 1 values i.e '0' is about employees who are surviving in their company and '1' is who resigned or left their job in their company.

5.The performance rating column is a continuous column and having rating like 1,2,3,4,5.The highest no.of employees who got '3'rating in their performance in job and lowest no.of employees who got '5' in their performance in job.

6.The sales column is a float data type column and Sales quota attainment is how much a salesperson sold last year relative to their quota. An employee whose sales_quota_pct equals 75 sold 75% of their quota.

Data preparation:

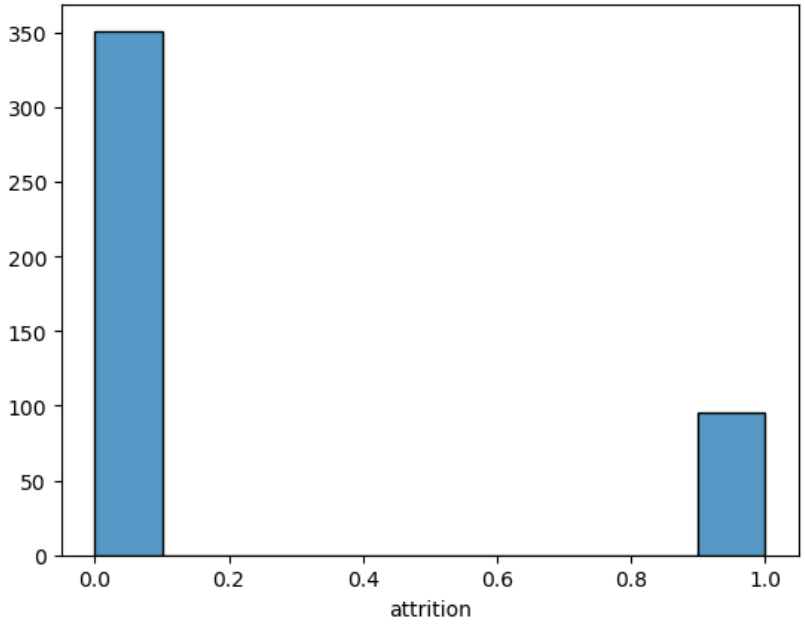
- --- Sub-contents:
- Data cleaning and Missing value Treatment
- Performed Analysis:
- ---Univariate
- ---Bivariate
- ---Multivariate
- Identified groups in the dataset
- Checking Average Sales Numbers and Attrition Numbers

- **Missing value Treatment:**
- The recruiting_source column contains more missing values, the missing values is filled with

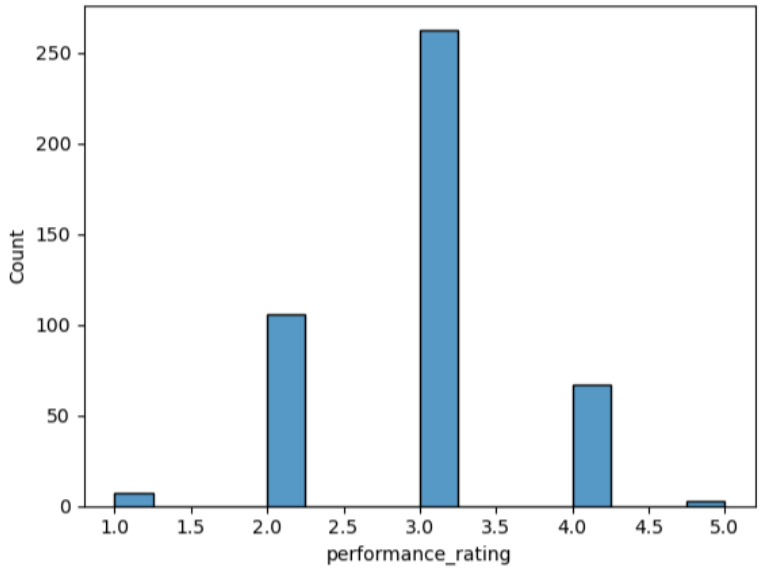
  "other_source" in the HR Analytics Data frame.
- **ANALYSIS:**
- Univariate Analysis:
- Bivariate Analysis:
- Multivariate Analysis:

Univariate Analysis-Frequency distribution

‘attrition’

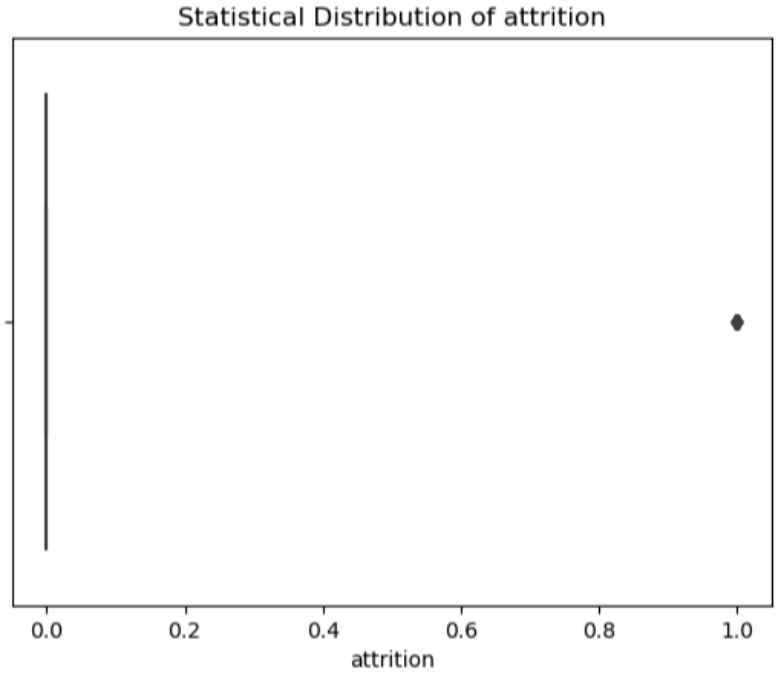


‘performance_rating’

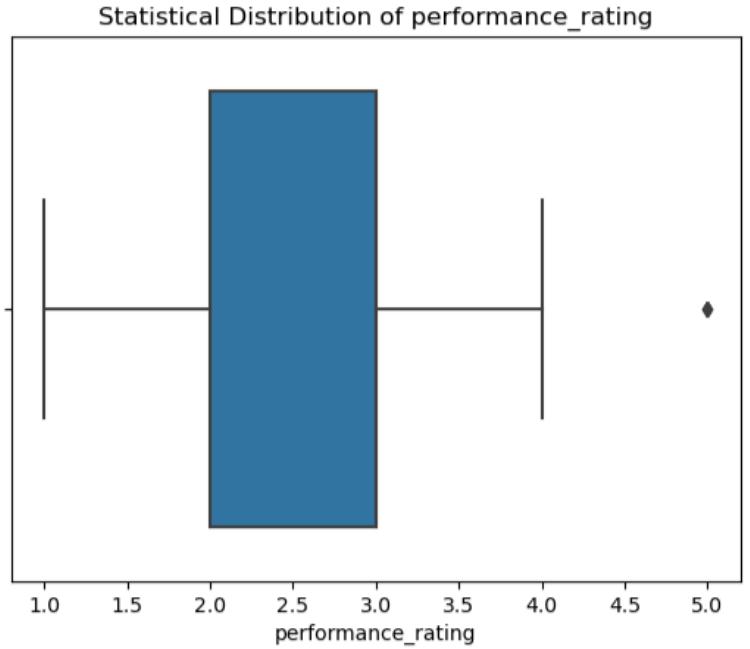


Statistical distribution

‘attrition'



‘performance _rating’

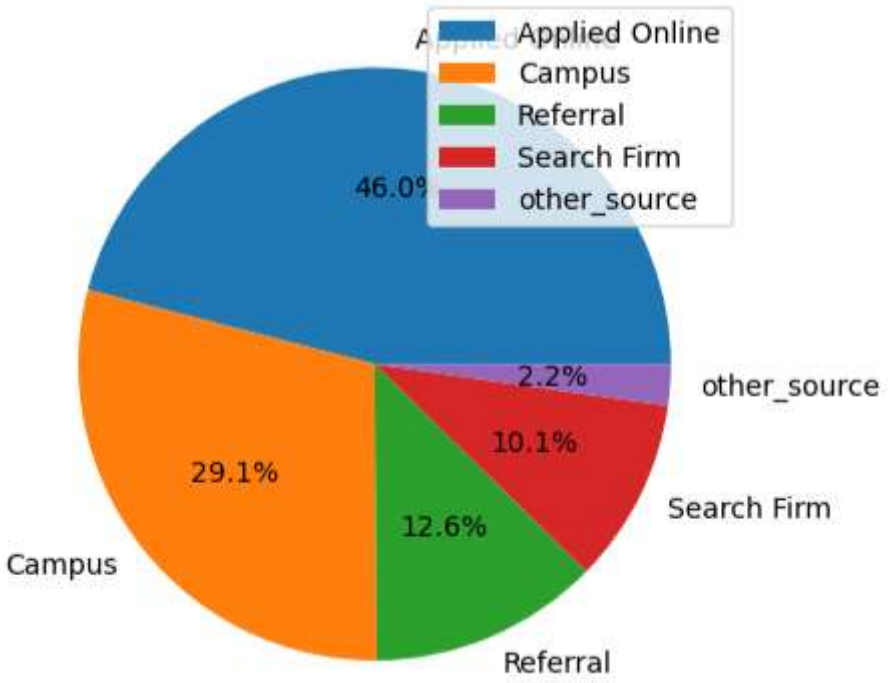


- **Insights:**
- In both frequency distributon and statistical distribution:
- The “attrition” contains high 'o' and low '1' values in HR Analytics dataframe.
- By this we can get that there are highest number of employees who are surviving in their company i.e '0' and lowest number of employees who are resigned their job in their company.
- In "performance_rating" column the more no.of employees having performance rating '3' and the less no.of employees having performance rating '5' in HR Analytics data.
- The outliers for 'attrition'column is resigned employees in their company.
- The outliers for 'performance_rating' column is having less employees with rating '5‘.

Pie -chart for categorical column.

Insights:

By this pie-chart we get to know that there are highest no.of employees who are “Applied online” and lowest number of employees who are taken” Search Firm” source.

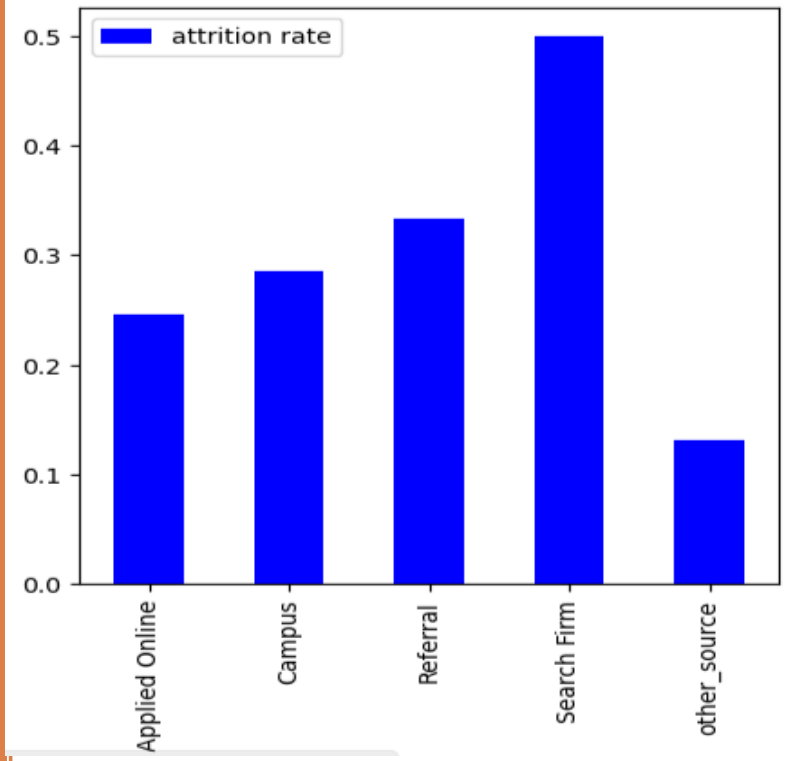


Bivariate Analysis
The Average comparision between 'recruiting_source' column and 'attrition'column

Another quality of hire metric you can consider is the attrition rate, or how often hires leave the company.

Insights:

The only “search Firm” source having highest 'attrition' rate when compared to other sources.while lowest for hires coming from "Applied Online".

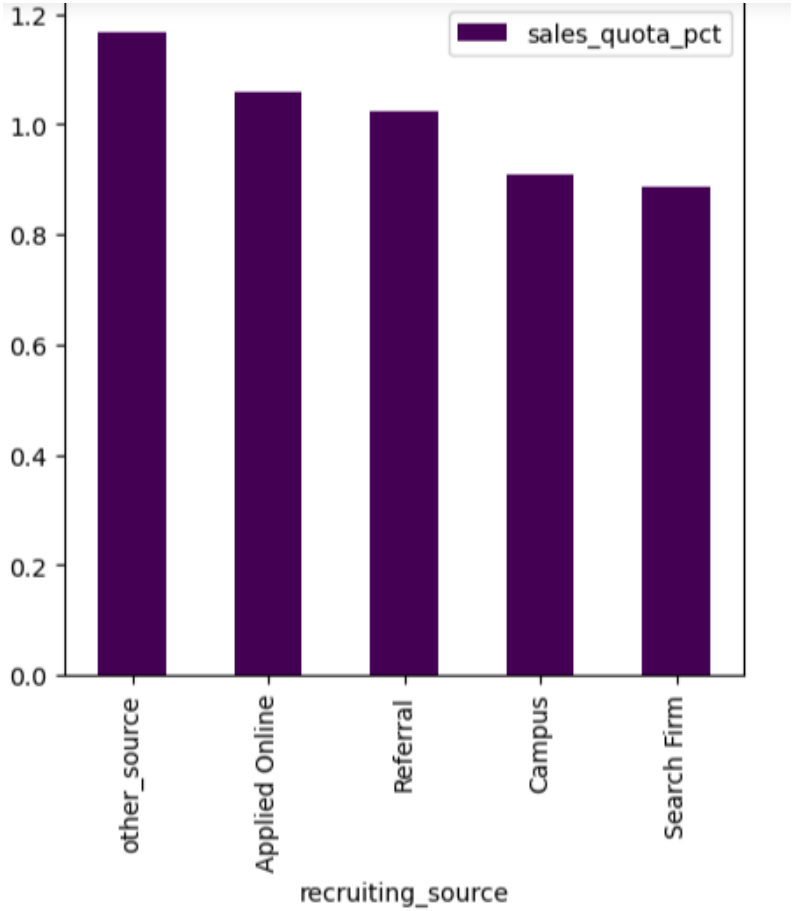


The Average comparision between 'recruiting_source' and 'sales_quota_pct' column

Which recruiting source resulted highest sales quota retainment Sales quota attainment is how much a salesperson sold last year relative to their quota. An employee whose sales_quota_pct equals 75 sold 75% of their quota.

Insights:

The 'Applied online' having highest rate in sales_quota_pct when compared to others sources in recruiting_source.



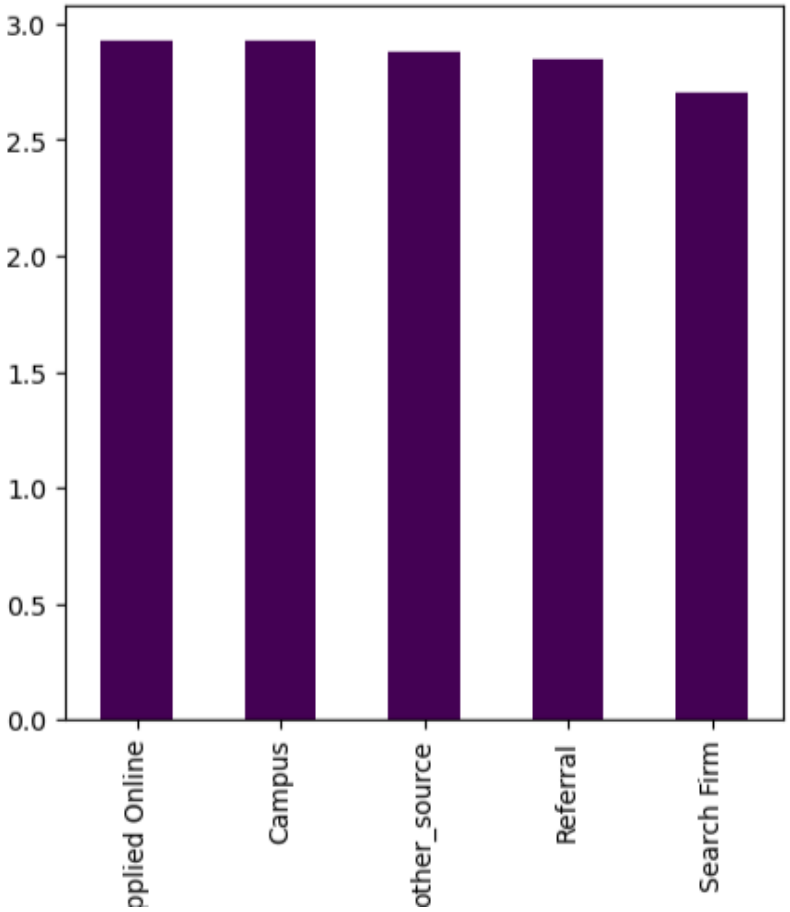
The Average comparision between 'recruiting_source' column and 'performance_rating' column

Insights:

This shows that the performance rating of the employees who were hired via "Applied Online" channel is higher than the rest. However the differences are small and

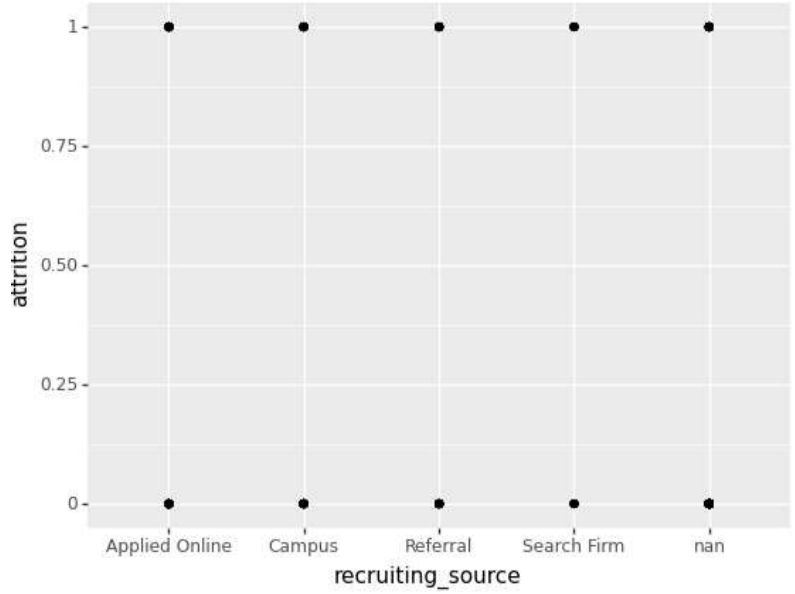
a statistical test is required to study the significance of the differences.

The Employees who are taken these sources having performance rating nearly equal to '3'.

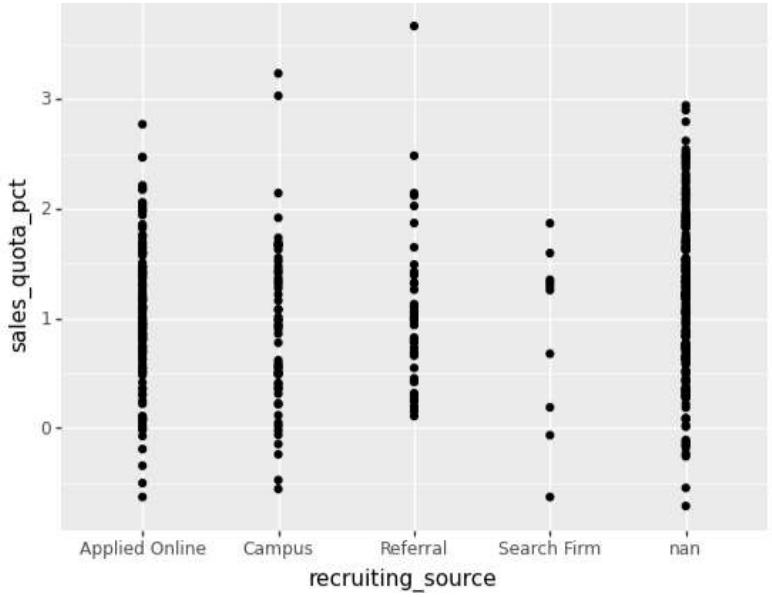


Visualization with plotnine

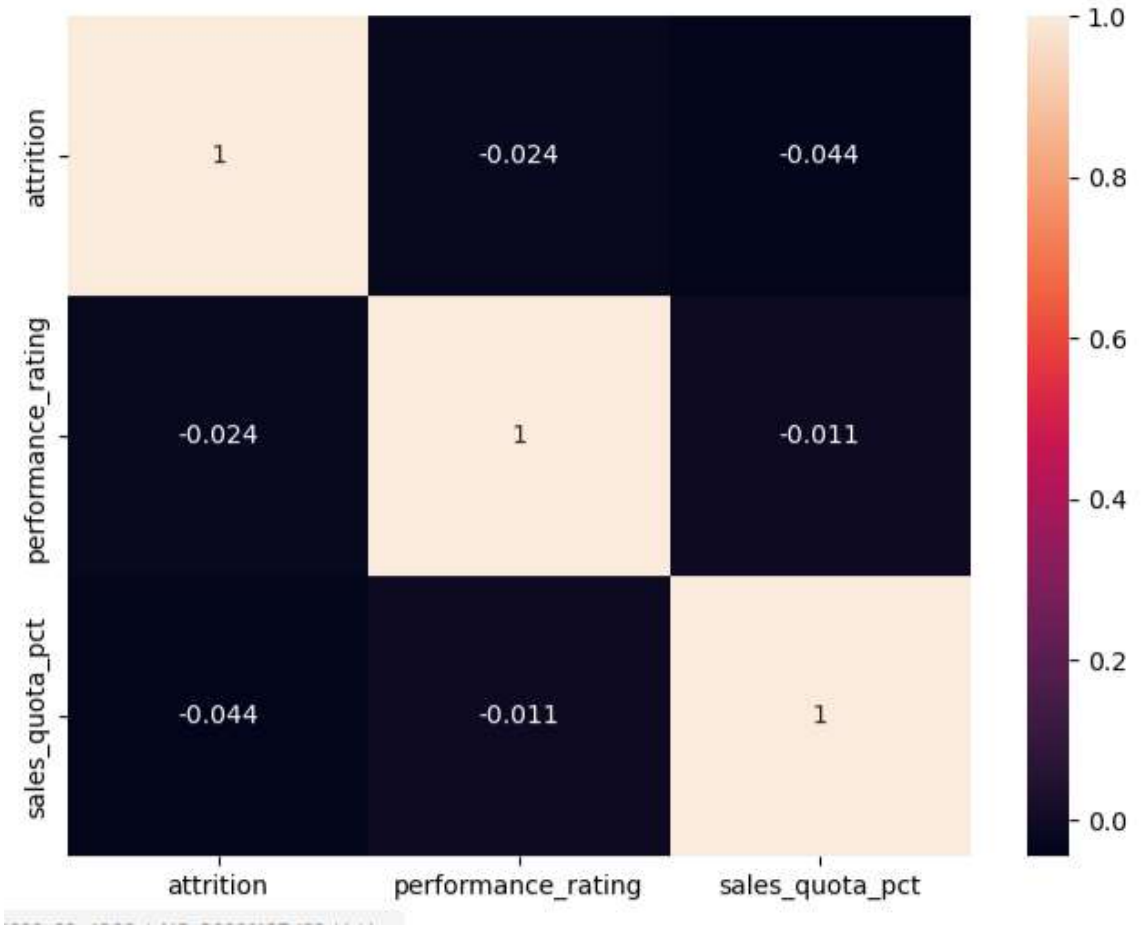
Visualizing Attrition differences by Recruiting Source



Visualizing Sales differences by Recruiting Source



Multivariate Analysis



Checking Average Sales Numbers and Attrition Numbers

Printing out the average Sales Number grouped by Recruiting Source

```
recruiting_source
Applied Online    1.058590
Campus            0.908035
Referral          1.023198
Search Firm       0.886960
other_source      1.168109
Name: sales_quota_pct, dtype: float64
```

Printing out the average Attrition Number grouped by Recruiting Source

```
recruiting_source
Applied Online    0.246154
Campus            0.285714
Referral          0.333333
Search Firm       0.500000
other_source      0.131707
Name: attrition, dtype: float64
```

Identified groups in the dataset

recruiting_source	attrition	performance_rating	sales_quota_pct
Applied Online	0.246154	2.930769	1.058590
Campus	0.285714	2.928571	0.908035
Referral	0.333333	2.844444	1.023198
Search Firm	0.500000	2.700000	0.886960
other_source	0.131707	2.882927	1.168109

Conclusion:

The sources that have high Sales numbers and low Attrition numbers.

Sales Numbers indicate the performance of employees recruited from a source.

Which of the recruiting sources in this dataset produced the best hires, measured by attrition and sales? Which source produced the worst hires?

Best source in HR Analytics: Applied Online,

Worst source in HR Analytics: Search Firm.