

Cross-tabulation with Final Corrected Answer

Cross-tabulation is used here to analyze the relationship between 'Job_Category' and 'Company_Size'. The exercise calculates the mean 'Salary_USD' for each combination of these categories in the dataset.

Learn / Courses / Exploratory Data Analysis In ...

← Course Outline →

Daily XP 835

Exercise

Cross-tabulation

Cross-tabulation can help identify how observations occur in combination.

Using the `salaries` dataset, which has been imported as a `pandas` `DataFrame`, you'll perform cross-tabulation on multiple variables, including the use of aggregation, to see the relationship between `"Company_Size"` and other variables.

`pandas` has been imported for you as `pd`.

Instructions 3/3

30 XP

✓

Perform cross-tabulation, setting `"Company_Size"` as the index, and the columns to classes in `"Experience"`.

✓

Cross-tabulate `"Job_Category"` and classes of `"Company_Size"` as column names.

5

Update `pd.crosstab()` to return the mean `"Salary_USD"` values.

⚠

Take Hint (-9 XP)

Incorrect Submission

Have you specified the arguments for `pd.crosstab()` using the right syntax?

Do you need additional hints?

If you are stuck, you can ask our AI Assistant for exercise solution tips.

Get AI Help

Powered by OpenAI

Did you find this feedback helpful?

✓ Yes

✗ No

script.py

Light Mode

```
1 # Cross-tabulate Job_Category and Company_Size
2 print(pd.crosstab(salaries["Job_Category"], salaries
3 ["Company_Size"],
4 values=salaries["Salary_USD"], aggfunc="mean"))
```

↶ Run Code Submit Answer

IPython Shell

Slides

>

<script.py> output:

Company_Size	L	M	S
Job_Category			
Data Analytics	23	61	8
Data Engineering	28	72	11
Data Science	38	59	16
Machine Learning	17	19	13
Managerial	5	8	1
Other	13	9	6

In [1]:

Answer

```
# Cross-tabulate Job_Category and Company_Size
print(pd.crosstab(salaries['Job_Category'], salaries['Company_Size'],
                  values=salaries['Salary_USD'], aggfunc='mean'))
```

Explanation: This command uses pandas `crosstab` to compute a matrix where the rows are 'Job_Category', the columns are 'Company_Size', and the

values are the mean 'Salary_USD'. The ``aggfunc='mean'`` specifies that we want the average salary for each combination.