

## Categorizing Salaries - Complete

This exercise involves categorizing salaries into four levels ('entry', 'mid', 'senior', 'exec') using defined salary ranges. The ranges are determined by the calculated 25th, 50th (median), and 75th percentiles, as well as the maximum salary value.

Learn / Courses / Exploratory Data Analysis In ...

Course Outline

Daily XP 1118

Exercise

### Categorizing salaries

Now it's time to make a new category! You'll use the variables `twenty_fifth`, `salaries_median`, and `seventy_fifth`, that you created in the previous exercise, to split salaries into different labels.

The result will be a new column called `"salary_level"`, which you'll incorporate into a visualization to analyze survey respondents' salary and at companies of different sizes.

`pandas` has been imported as `pd`, `matplotlib.pyplot` as `plt`, `seaborn` as `sns`, and the `salaries` dataset as a `pandas` `DataFrame` called `salaries`.

Instructions 3/4 25 XP 3 / 4

- Split `"Salary_USD"` based on the labels and ranges you've created.

Take Hint (-7 XP)

script.py

Light Mode

```
1 # Create salary labels
2 salary_labels = ["entry", "mid", "senior", "exec"]
3
4 # Create the salary ranges list
5 salary_ranges = [0, twenty_fifth, salaries_median,
6                  seventy_fifth, salaries["Salary_USD"].max()]
7
8 # Create salary_level
9 salaries["salary_level"] = pd.cut(salaries["Salary_USD"],
10                                  bins=salary_ranges,
11                                  labels=salary_labels)
```

Run Code Submit Answer

IPython Shell

Slides

<script.py> output:

	Salary_USD	salary_level
0	248256.840	exec
1	104099.820	senior
2	19096.680	entry
3	143225.100	senior
4	68748.048	mid
...	...	...
402	147044.436	exec
403	120309.084	senior
404	123173.592	senior
405	143225.100	senior
406	190966.800	exec

[407 rows x 2 columns]

In [1]:

## Answer

```
# Create salary labels
salary_labels = ['entry', 'mid', 'senior', 'exec']
```

```
# Create the salary ranges list
salary_ranges = [0, twenty_fifth, salaries_median, seventy_fifth,
salaries['Salary_USD'].max()]
```

```
# Categorize salaries into salary levels
salaries['salary_level'] = pd.cut(
    salaries['Salary_USD'],
    bins=salary_ranges,
    labels=salary_labels
)
```

```
# Print the resulting DataFrame to verify
print(salaries[['Salary_USD', 'salary_level']])
```

Explanation: The `salary\_labels` list defines four salary categories ('entry', 'mid', 'senior', 'exec'). The `salary\_ranges` list defines the bin boundaries using the calculated percentiles (25th, median, 75th) and the maximum salary. Using pandas `cut`, the 'Salary\_USD' values are assigned to these bins, and the new column 'salary\_level' is added to the DataFrame. The output shows the assigned salary levels for each salary value.