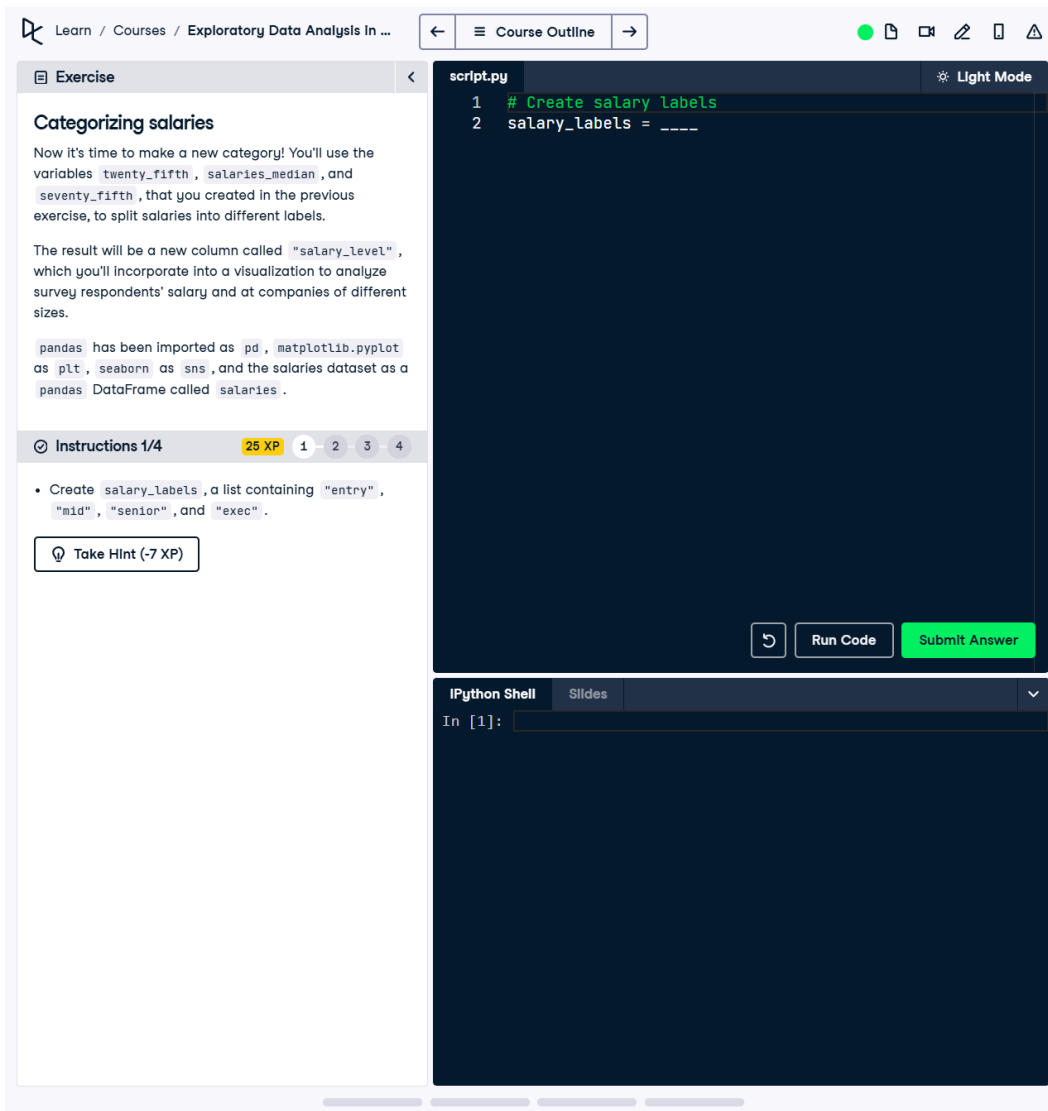


## Categorizing Salaries - Final

This task involves defining salary labels and creating salary ranges to categorize salaries into four levels: 'entry', 'mid', 'senior', and 'exec'. The categorization is based on the calculated percentiles and the maximum salary value.



The screenshot shows a JupyterLab interface. On the left, there's a sidebar with a breadcrumb trail: 'Learn / Courses / Exploratory Data Analysis In ...'. Below it, a tab labeled 'Exercise' contains the title 'Categorizing salaries'. The text in the exercise tab explains the task: '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`.

Below the text, there's a section 'Instructions 1/4' with a progress bar showing 25 XP and four steps. The first step is selected: 'Create `salary_labels`, a list containing "entry", "mid", "senior", and "exec".' A 'Take Hint (-7 XP)' button is visible.

On the right, a code editor window titled 'script.py' shows a Python script in 'Light Mode'. The script contains two lines of code:

```
1 # Create salary labels
2 salary_labels = ----
```

At the bottom of the code editor, there are buttons for 'Run Code' and 'Submit Answer'. Below the code editor is an 'IPython Shell' window with a prompt 'In [1]:' and a cursor.

## 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()]
```

```
# Use pandas cut to categorize salaries
salaries['salary_level'] = pd.cut(
    salaries['Salary_USD'],
    bins=salary_ranges,
    labels=salary_labels
)
```

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

Explanation: The `salary\_labels` list defines the four categories for salary levels. The `salary\_ranges` list uses the calculated percentiles (25th, median, 75th) and the maximum salary to define bin boundaries. Using the `cut` function from pandas, 'Salary\_USD' values are segmented into these bins and assigned corresponding labels ('entry', 'mid', 'senior', 'exec'). The new 'salary\_level' column maps each salary to its respective category.