

## Module 5: Data Manipulation

---

### Case Study – 2

edureka!

**edureka!**

© Brain4ce Education Solutions Pvt. Ltd.

## Case Study – 2

### Domain – HR

focus – Insights from data

#### Business challenge/requirement

SFO Public Department - referred to as SFO has captured all the salary data of its employees from the year 2011-2014. Now we are in the year 2015 and the organization is facing a financial crisis. As the first step HR wants to rationalize employee costs to save the payroll budget. You have to do data manipulation and analysis on the salary data to answer specific questions for cost savings.

#### Key issues

Cost can be saved by figuring out the key pockets of high salaries

#### Considerations

NONE

#### Data volume

- Approx 150K records across files

#### additional information

- NA

#### Business benefits

Save at least 10% of employee costs by identifying and letting them go

#### Approach to Solve

You have to use the fundamentals of Pandas covered in module 6 and answer the following 5 Questions

1. Compute how much total salary cost has increased from the year 2011 to 2014
2. Which Job Title in the Year 2014 has the highest mean salary?

3. How much money could have been saved in the Year 2014 by stopping OverTimePay?
4. Which are the top 5 common jobs in the Year 2014 and how much do they cost SFO?
5. Who was the top earning employee across all the years?

### Enhancements for code

You can try these enhancements in code

1. Which are the last 5 common jobs in the Year 2014 and how much do they cost SFO?
2. In year 201 OverTimePay was what percentage of TotalPayBenefits
3. Which Job Title in the Year 2014 has the lowest mean salary?

edureka!