

# Introduction to Data Science

Bộ môn Khoa học dữ liệu  
Khoa Công nghệ thông tin  
Trường Đại học Công nghiệp thành phố Hồ Chí Minh-IUH

## Practice:

### 1- US Baby Name

US\_BaBy\_Names.csv data contains information of babies born from 2004 - 2014 in the US. Based on the above data set, do the exercises as follows:

## LEARNING DATA

- 1,1. Read data
- 1,2. Display the first 10 lines of data
- 1,3. Show the data size
- 1,4. Delete 2 columns 'Unnamed: 0' and 'Id'
- 1,5. Tell me the number of boys and girls?
- 1,6. How many different names are there in the dataset?
- 1,7. Which name appears the least?
- 1,8. Top 10 most used names?

## DATA VISUALIZATION

- 1.9. Visualize the top 10 popular names in 2014
- 1.10. Visualize the ratio of men and women over the years 2004-2014

## 2- WHAT AFFECTS TO TIPS?

Read tips.csv data, perform descriptive statistics (learn quantities and visualizations) to learn the factors that affect tips

### DATA

Includes columns

- total\_bill
- tip (in USD)
- sex
- smoker who does/does not smoke
- day of the week
- time moment
- size number of people

Please perform descriptive statistics to learn about the tip behavior of customers at the restaurant

## **UNDERSTANDING THE DATA**

- 2.1. Read data
- 2.2. Delete the column `Unnamed: 0`
- 2.3. How many rows and attributes the data has.
- 2.4. Among the attributes, which attribute is the categorical attribute, which is the quantitative attribute?
- 2.5. How many men and women. What rate?
- 2.6. How many people smoke. What rate?
- 2.7. What day of the week do customers come to the restaurant most? What rate?
- 2.8. Which restaurant serves the most meals (Breakfast, Lunch, Dinner)?
- 2.9. What is the largest group of customers at the restaurant?
- 2.10. On average, how many guests does a restaurant serve in a meal?

## **FACTORS AFFECTING TO TIP**

In this section, we try to learn about tipping at restaurants and the factors that can affect it

- 2.11. What is the restaurant's tip rate?
- 2.12. What is the tip to total bill?
- 2.13. Between men and women, who tips more money, who tips more often?
- 2.14. Between smokers and non-smokers, who tips more, who tips more often?
- 2.15. Statistics of the total amount of tips received by day.

## **DATA VISUALIZATION**

Let's visualize with the questions as follows:

- 2.16. Visualize total tips by day
- 2.17. Visualize total tip by meal (time)
- 2.18. Plot histogram of the total invoice
- 2.19. Plot the boxplot of the total invoice
- 2.20. Visualize the relationship between total bill and tip.
- 2.21. Compare how much customers have spent between days of the week.