**BỘ MÔN HỆ THỐNG THÔNG TIN – KHOA CÔNG NGHỆ THÔNG TIN**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN THÀNH PHỐ HỒ CHÍ MINH, ĐẠI HỌC QUỐC GIA TP HCM**

TRỰC QUAN HÓA DỮ LIỆU



GV phụ trách: Tiết Gia Hồng

SV thực hiện: Nhóm 9

TRỰC QUAN HÓA DỮ LIỆU

HỌC KỲ II – NĂM HỌC 2021-2022

**MỤC LỤC**

[1. Tập dữ liệu 2](#_Toc99139199)

[2. Xác định 3](#_Toc99139200)

[3. Phát biểu vấn đề cần biết trên tập dữ liệu 5](#_Toc99139201)

**Thông Tin Thành Viên Nhóm**

|  |  |  |
| --- | --- | --- |
| **Mã nhóm**: 9 | | |
| Thông tin các thành viên | | |
| MSSV | Họ và tên | Mail |
| 18120158 | Lý Ngọc Bình | 18120158@student.hcmus.edu.vn |
| 18120510 | Cao Xuân Hồng Phúc | cxhphuc@gmail.com |
| 18120518 | Phạm Thị Bích Phượng | ptbphuong2k@gmail.com |
| 18120588 | Nguyễn Nhị Cẩm Tiên | 18120588@student.hcmus.edu.vn |

# 

# Tập dữ liệu

Nhóm chúng em quyết định chọn tập dữ liệu [Atlanta City Employee Salaries - dataset by brentbrewington | data.world](https://data.world/brentbrewington/atlanta-city-employee-salaries) từ trang data.world

Mô tả chung về tập dữ liệu: Tập dữ liệu thể hiện thông tin lương của các nhân viên trong 1 công ty theo hàng năm.

Bảng dưới đây thể hiện một phần dữ liệu được trích từ tập dữ liệu:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| name | age | sex | ethnic.origin | job.title | organization | annual.salary |
| Aaron, Dayan M | 38 | Male | Black or African American (Not Hispanic or Latino) | ATL311 Team Lead | EXE Executive Offices | 45999.99 |
| Aaron, Tyrone M | 44 | Male | Black or African American (Not Hispanic or Latino) | Watershed Crew Supervisor (D) | DWM Department of Watershed Management | 46575.01 |
| Abazenab, Kokeb | 42 | Female | Black or African American (Not Hispanic or Latino) | Benefits Representative, Sr | DHR Department of Human Reources | 42866.89 |

# Xác định

Loại tập dữ liệu(dataset type) là Table, bao gồm các item là nhân viên và attribute là thông tin của nhân viên tạo thành, được minh họa trong hình 1.

Table

Description automatically generated

Hình 1: *Item và attribute của tập dữ liệu*

* Attribute: name, age, sex, ethnic.origin, job title, organization, annual salary.
* Item: Employee salary.

Bảng 1 thể hiện ý nghĩa cũng như loại của từng thuộc tính trong tập dữ liệu.

|  |  |  |  |
| --- | --- | --- | --- |
| Attribute | Data type | Data semantic | Attribute type |
| Name | Attribute | Họ và tên của nhân viên | Categorical |
| Age | Attribute | Tuổi của nhân viên | Quantitative |
| Sex | Attribute | Giới tính của nhân viên | Ordinal |
| Ethnic origin | Attribute | Nguồn gốc chủng tộc của nhân viên | Categorical |
| Job title | Attribute | Vị trí làm việc của nhân viên trong công ty | Categorical |
| Organization | Attribute | Các phòng ban làm việc | Categorical |
| Annual salary | Attribute | Lương hàng năm | Quantitative |

Bảng 1: *Ý nghĩa và loại của từng thuộc tính*

# Phát biểu vấn đề cần biết trên tập dữ liệu

|  |  |  |
| --- | --- | --- |
| Vấn đề | Đặt ra câu hỏi liệu rằng ethnic.origin có quyết định mức lương (annual salary) của người lao động hay không? | Tìm kiếm các nhân viên thuộc phòng nhân sự có lương hàng năm trên 20000 |
| Abstract task | Discover, Present | Query |
| Trả lời cho câu hỏi What | Dữ liệu được trình bày dưới dạng bảng chứa các bản ghi dữ liệu | Dữ liệu được trình bày dưới dạng bảng chứa các bản ghi dữ liệu |
| Xác định target | All data(features) | Attribute(Many-> dependency) |
| Xác định task | Discover( tạo ra giả thuyết -> kiểm chứng giả thuyết đưa ra là đúng hay sai?)  Present ( trình bày thông qua biểu đồ thể hiện mối tương quan) | Lookup (Tìm kiếm các nhân viên thuộc phòng ban nhân sự) → Compare (So sánh lương xem ai có lương > 5000) |

## 

## 