ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

**KHOA CÔNG NGHỆ THÔNG TIN**

**\_\_\_\_\_\_\_\_\_\_**

****

**BUSINESS INTELIGENT**

**ĐỒ ÁN**

**XÂY DỰNG VÀ KHAI THÁC DỮ LIỆU**

* **Thông tin nhóm:**

Nguyễn Chí Lập – 18120435

Huỳnh Hoàng Long – 18120446

Lê Hoàng Long – 18120447

Nguyễn Đại Long – 18120448

* **Thông tin giảng viên:**

Cô Hồ Thị Hoàng Vy

Cô Nguyễn Thị Như Anh

Cô Tiết Gia Hồng

TP. HCM, Ngày 1 tháng 10 năm 2021

|  |  |  |
| --- | --- | --- |
| **BẢNG PHÂN CÔNG** | | |
| Mô tả dữ liệu | | |
| 18120435 | Nguyễn Chí Lập | Dữ liệu UK Car Accidents (Accident, Vehicles) |
| 18120446 | Huỳnh Hoàng Long | Dữ liệu UK-Postcodes |
| 18120447 | Lê Hoàng Long | Dữ liệu LSOA-Postcode mapping: |
| 18120448 | Nguyễn Đại Long | Dữ liệu UK Car Accidents (Casualties) |

# **Mô tả dữ liệu :**

## Dữ liệu UK Accidents (2011 – 2014):

|  |  |  |
| --- | --- | --- |
| Accidents | | |
| **Tên thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| Accident\_Index | Number | Mã xác định duy nhất một vụ tai nạn |
| Police\_Force | Number | Mã lực lượng cảnh sát phụ trách vụ tai nạn |
| Accident\_Severity | Number | Mức độ nghiêm trọng của vụ tai nạn  Có 3 mức độ : 1 – Fatal, 2 – Serious, 3 – Slight |
| Number\_of\_Vehicles | Number | Số lượng phương tiện gây ra tai nạn |
| Number\_of\_Casualties | Number | Số lượng thương vong tai nạn gây ra |
| Date | Date | Ngày xảy ra tai nạn |
| Date\_of\_Week | Number | Thứ ngày trong tuần (Giá trị từ 1 : CN -> 7 : Thứ bảy) |
| Local\_Authority\_(District) | Number | Mã quản lý khu vực |
| Local\_Authority\_(Highway) | String | Mã quản lý tuyến đường cao tốc |
| 1­­st Road Class | Number | Đường cấp 1, thuộc sở hữu của nhà nước |
| Road\_Type | Number | Loại tuyến đường |
| Junction Detail | Number | Chi tiết giao lộ ( 0 – không thuộc giao lộ hoặc cách giao lộ 20m, -1 data missing/out of range) |
| Junction Control | Number | Chủ thể điều khiển giao lộ ( 0 – không thuộc giao lộ hoặc cách giao lộ 20m, -1 data missing/out of range) |
| 2nd Road Class | Number | Đường cấp 2, thuộc sở hữu của khu vực |
| **Vehicles** | | |
| **Tên thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| Vehicle\_Reference | Number | Số phương tiện liên quan đến tai nạn |
| Vehicle\_Type | Number | Loại xe |
| Vehicle\_Manoeuve | Number | Cách di chuyển của phương tiện khi gây tai nạn |
| Junction\_Location | Number | Vị trí phương tiện trên giao lộ |
| Sex\_of\_Driver | Number | Giới tính tài xế |
| Age\_of\_Driver | Number | Tuổi tài xế |
| Age\_Band\_of\_Driver | Number | Độ tuổi tài xế |
| Age\_of\_Vehicle | Number | Tuổi của phương tiện |
| **Calsualties** | | |
| **Tên thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| Accident\_Index | string | Mã thứ tự các vụ tai nạn |
| Vehicle\_Reference | Int | Số phương tiện liên quan |
| Casualty\_Reference | Int | Số nạn nhân liên qua |
| Casualty\_Class | Int | Nạn nhân thuộc lớp người nào: 1 người lái xe, 2 hành khác, 3 đi bộ |
| Sex\_of\_Casualty | Int | Giới tính của nạn nhân |
| Age\_of\_Casualty | Int | Tuổi của nạn nhân |
| Age\_Band\_of\_Casualty | Int | Nhóm tuổi của nạn nhân( có giá trị từ -1 đến 11) |
| Casualty\_Severity | Int | Mức độ nghiêm trọng của vụ tai nạn |
| Pedestrian\_Location | Int | Vị trí đi bộ của nạn nhân  (Có 12 giá trị khác nhau) |
| Pedestrian\_Movement | Int | Nạn nhân đi bộ di chuyển như thế nào |
| Car\_Passenger | Int | Nạn nhân ngồi ở vị trí nào trên xe |
| Bus\_or\_Coach\_Passenger | Int | Nạn nhân đi xe bus hay xe khách |
| Pedestrian\_Road\_Maintenance\_Worker | Int | Nạn nhân có phải là công nhân bảo trì đường dành cho người đi bộ không |
| Casualty\_Type | Int | Loại tai nạn( tai nạn với phương tiện gì) |
| Casualty\_Home\_Area\_Type | Int | Nạn nhân sinh sống ở loại khu vực nào |

## Dữ liêu LSOA-Postcode mapping:

|  |  |  |
| --- | --- | --- |
| **Tên thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| pcd7 | string | Mã bưu điện gồm có 7 ký tự |
| pcd8 | string | Mã bưu điện gồm có 8 ký tự |
| pcds | string | Mã bưu điện có số ký tự tùy chỉnh |
| dointr | number | Ngày bắt đầu |
| Doterm | number | Ngày kết thúc |
| usertype | bool | Loại người dùng  0 = small user; 1 = large user |
| oa11cd | string | Khu vực đầu ra 2011 |
| lsoa11cd | string | Mã khu vực siêu đầu ra lớp dưới 2011 |
| msoa11cd | string | Mã khu vực siêu đầu ra lớp giữa 2011 |
| ladcd | string | Mã quận chính quyền đại phương |
| lsoa11nm | string | Địa chỉ khu vực siêu đầu ra lớp dưới 2011 |
| msoa11nm | string | Địa chỉ khu vực siêu đầu ra lớp giữa 2011 |
| ladnm | string | Địa chỉ quận chính quyền đại phương |
| ladnmw |  |  |

## Dữ liệu UK-Postcodes:

|  |  |  |
| --- | --- | --- |
| **Tên thuộc tính** | **Kiểu dữ liệu** | **Mô tả** |
| postcode | nvarchar | Mã bưu điện, là một chuỗi ký tự viết bằng chữ, hoặc bằng số hay tổ hợp của số và chữ, được viết bổ sung vào địa chỉ nhận thư với mục đích tự động xác định điểm đến cuối cùng của thư tín, bưu phẩm. |
| easting | int | khoảng cách đã đi hoặc đo về phía đông(meter) |
| Northing | int | khoảng cách đã đi hoặc đo về phía bắc (meter) |
| latitude | float | Vĩ độ, giá trị xác định vị trí của một điểm trên bề mặt Trái Đất ở phía bắc hay phía nam của xích đạo |
| longitude | float | Kinh độ, là giá trị tọa độ địa lý theo hướng đông-tây, với quy mô toàn cầu. |
| city | nvarchar | Tên của thành phố |
| county | nvarchar | Hạt, những khu vực sử dụng cho các mục đích phân định ranh giới hành chính, địa lý và chính trị. |
| country\_code | nvarchar | Mã quốc gia, những mã địa lý bằng chữ hoặc số được phát triển để đại diện cho quốc gia và khu vực phụ thuộc, dùng trong xử lý dữ liệu và thông tin liên lạc |
| country\_name | nvarchar | Tên quốc gia |
| iso3166-2 | nvarchar | ISO 3166-2 là phần thứ hai của tiêu chuẩn ISO 3166. Nó là hệ thống mã địa lý được tạo ra để mã hóa các tên gọi của các đơn vị hành chính thuộc các quốc gia và các vùng lãnh thổ độc lập |
| region\_code | nvarchar | Mã vùng, |
| region\_name | nvarchar | Tên vùng |
| postcode | nvarchar | Mã bưu điện, là một chuỗi ký tự viết bằng chữ, hoặc bằng số hay tổ hợp của số và chữ, được viết bổ sung vào địa chỉ nhận thư với mục đích tự động xác định điểm đến cuối cùng của thư tín, bưu phẩm. |
| easting | int | khoảng cách đã đi hoặc đo về phía đông(meter) |
| Northing | int | khoảng cách đã đi hoặc đo về phía bắc (meter) |