**ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**Khoa Công nghệ Thông tin**

**Chuyên ngành Hệ thống Thông tin**



**ĐỒ ÁN THỰC HÀNH**

**MÔN**

**HTTT PHỤC VỤ**

**TRÍ TUỆ KINH DOANH**

***Thành viên nhóm TTKD-19:***

1. *Nguyễn Lê Ngọc Tần MSSV: 18120553*
2. *Lê Hữu Thanh MSSV: 18120560*
3. *Nguyễn Thanh Thi MSSV: 18120570*
4. *Lê Việt Tiến MSSV: 18120590*

*Tháng 9 năm 2021*

**Thông tin thành viên nhóm**

|  |  |
| --- | --- |
| **Mã nhóm** | TTKD\_19 |
| **Tên nhóm** | Trí tuệ kinh doanh 19 |
| **Số lượng** | 4 thành viên |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Họ tên | MSSV | Email cá nhân | SDT |
| 1 | Nguyễn Lê Ngọc Tần | 18120553 | nguyenlengoctan.ltp@gmail.com | 0968473091 |
| 2 | Lê Hữu Thanh | 18120560 | thanhle.16012000@gmail.com | 0367012721 |
| 3 | Nguyễn Thanh Thi | 18120570 | thinguyen146@gmail.com | 0932953252 |
| 4 | Lê Việt Tiến | 18120590 | l.v.tien112@gmail.com | 0936616348 |

**Phân công công việc**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên công việc | Phụ trách | Mức độ hoàn thành |
| 1 | Mô tả dữ liệu UK Car Accidents 2011 - 2014  Bảng Accident (8 thuộc tính đầu)  Bảng Casualties (7 thuộc tính đầu) | Nguyễn Lê Ngọc Tần | 100% |
| 2 | Mô tả dữ liệu UK Car Accidents 2011 - 2014  Bảng Accident (thuộc tính 9 -16)  Bảng Casualties (8 thuộc tính sau) | Nguyễn Thanh Thi | 100% |
| 3 | Mô tả dữ liệu UK Car Accidents 2011 - 2014  Bảng Accident (thuộc tính 17 -24)  Bảng Vehicles (11 thuộc tính đầu) | Lê Hữu Thanh | 100% |
| 4 | Mô tả dữ liệu UK Car Accidents 2011 - 2014 (8 thuộc tính cuối)  Bảng Accident (8 thuộc tính cuối)  Bảng Vehicles (11 thuộc tính cuối) | Lê Việt Tiến | 100% |
| 5 | Mô tả dữ liệu LSOA-Postcode mapping | Làm nhóm chung qua Zoom | 100% |
| 6 | Mô tả dữ liệu UK-Postcodes | Làm nhóm chung qua Zoom | 100% |

**MỤC LỤC**

[I. MÔ TẢ DỮ LIỆU 5](#_Toc7613)

[1. Dữ liệu UK Car Accidents 2011 - 2014 5](#_Toc9073)

[2. Dữ liệu LSOA-Postcode mapping 11](#_Toc16823)

[3. Dữ liệu UK-Postcodes 12](#_Toc14064)

[II. THIẾT KẾ KDL - TỔNG HỢP, NẠP DỮ LIỆU CÁC NGUỒN VÀO KDL - THIẾT KẾ, XÂY DỰNG CUBE 13](#_Toc3222)

[III. OLAP VÀ REPORT 13](#_Toc15204)

[IV. DATA MINING 13](#_Toc28842)

[V. KẾT LUẬN CHUNG 14](#_Toc22773)

[LINK BÁO CÁO 15](#_Toc24624)

[DANH MỤC TÀI LIỆU THAM KHẢO 15](#_Toc16598)

**NỘI DUNG**

# MÔ TẢ DỮ LIỆU

## Dữ liệu UK Car Accidents 2011 - 2014

**Bảng dữ liệu Accidents0515.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu dữ liệu có khả năng trong SQL SERVER** | **Ý nghĩa** | **Dữ liệu minh họa** |
| 1 | Accident\_Index | VARCHAR(15) | Mã vụ định danh một vụ tai nạn giao thông | 200501BS00001;  200501BS70019 |
| 2 | Location\_Easting\_OSGR | INT | Tọa độ sử dụng Grid reference system để tham chiếu tìm vị trí tai nạn | 525890;  524220 |
| 3 | Location\_Northing\_OSGR | INT | 179710;  180830 |
| 4 | Longitude | FLOAT | Kinh độ xảy ra tai nạn | -0.206458  -0.17157 |
| 5 | Latitude | FLOAT | Vĩ độ xảy ra tai nạn | 51.516353  51.493429 |
| 6 | Police\_Force | TINYINT | Lực lượng cảnh sát chịu trách nhiệm tại hiện trường tai nạn hoặc lực lượng cảnh sát được giao xử lý vụ tai nạn đó. Có 98 lực lượng cảnh sát, đánh số từ 1..98 | 1  98 |
| 7 | Accident\_Severity | TINYINT | Mức độ nghiêm trọng của tai nạn:   1. ít nghiêm trọng   2 - nghiệm trọng  3 - rất nghiêm trọng | 1  2 |
| 8 | Number\_Of\_Vehicles | TINYINT | Số lượng phương tiện trong vụ tai nạn | 1  2 |
| 9 | Number\_Of\_Casualties | TINYINT | Số người thương vong (bị thương hoặc tử vong) | 1  2 |
| 10 | Date | DATE | Ngày tháng năm xảy ra tai nạn (1,2,…7) | 04/01/2005  03/03/2005 |
| 11 | Date of Week | TINYINT | Thứ trong tuần xảy ra tai nạn | 1  5 |
| 12 | Time | TIME | Thời gian xảy ra tai nạn | 11:20  20:48 |
| 13 | Local\_Authority\_(District) | INT | Quận/thị trấn xảy ra tai nạn. Có 941 quận/thị đánh mã từ 1..941 | 1 - Westminster  91 - Liverpool |
| 14 | Local\_Authority\_(Highway) | VARCHAR(12) | Cơ quan chịu trách nhiệm vận hành, quản lý, bảo trì tuyến đường giao thông. Có 207 cơ quan có tên khác nhau có mã để phân biệt | S12000033 - Aberdeen City  E09000003 - Barnet |
| 15 | 1st\_Road\_Class | TINYINT | Phân loại đường (đường nhà nước, đường do địa phương quản lý,...) [giá trị là 1, 2, 3, 4, 5, 6] | 1  3 |
| 16 | 1st\_Road\_Number | FLOAT | Tên đường xảy ra tai nạn (dạng mã số) | 1  3220 |
| 17 | Road\_Type | TINYINT | Loại đường (1 chiều, nhánh,...). Đánh các mã loại đường là 1, 2, 3, 6, 7, 9, 12, -1 | 1 - Bùng binh  12 - Đường 1 chiều |
| 18 | Speed\_Limit | TINYINT | Tốc độ tối đa cho phép | 30  70 |
| 19 | Junction\_Detial | TINYINT | Chi tiết nút giao thông (bùng binh, ngã tư,...) đánh mã [-1, 0, 1, 2, ...9] | 2 - Bùng binh nhỏ  6 - Ngã tư |
| 20 | Junction\_Control | SMALLINT | Kiểm soát nút giao thông (đèn tín hiệu, tự do di chuyển,...) gồm các mã [-1, 0, 1, 2, 3, 4] | 3 - Biển báo Stop  1 - Người được ủy quyền điều khiển |
| 21 | 2nd\_Road\_Class | TINYINT | Phân loại đường như 1st\_Road\_Class nhưng nếu tai nạn xảy ra nơi giao 2 con đường | 1  3 |
| 22 | 2nd\_Road\_Number | FLOAT | Tương tự, tên 2 đường xảy ra tai nạn | 23  99 |
| 23 | Pedestrian\_Crossing-Human\_Control | SMALLINT | Điều khiển vượt làn người đi bộ quá 50m hay không? Có các giá trị [0 - Không có trong vòng 50km, 1 - kiểm soát bằng tuần tra qua đường, 2 - kiểm soát bởi người được ủy quyền, -1 - Thiếu dữ liệu hoặc ngoài phạm vi] | 2  1 |
| 24 | Pedestrian\_Crossing-Physical\_Facilities | SMALLINT | Cơ sở vật chất nơi xảy ra tai nạn (có vạch kẻ đường, cầu vượt,...) Các giá trị [-1 - không có hoặc thiếu, 0 - Không có, 1 - gờ giảm tốc, 4 - vật chất cho người đi bộ, 5 - Pha dành cho người đi bộ tại ngã ba đèn tín hiệu giao thông, 7 - Cầu đi bộ hoặc tàu điện ngầm, 8 - Nơi ẩn náu trung tâm] | 1  2 |
| 25 | Light\_Conditions | SMALLINT | Điều kiện ánh sáng nơi xảy ra tai nạn. Mức độ [1, 4, 5, 6, 7, -1] | 1 - Ánh sáng ban ngày  6 - Chỗ tối - không chiếu sáng |
| 26 | Weather\_Conditions | SMALLINT | Điều kiện thời tiết nơi xảy ra tai nạn [1, 2, 3, .. 9] | 1  2 |
| 27 | Road\_Surface\_Conditions | SMALLINT | Điều kiện mặt đường nơi xảy ra tai nạn. [-1, 0, ... 7] | 0 |
| 28 | Special\_Condition\_at\_Site | SMALLINT | Điều kiện đặc biệt khác nơi xảy ra tai nạn [-1, 0, ... 7] | 0 |
| 29 | Carriageway\_Hazards | SMALLINT | Vật cản tại nơi xảy ra tai nạn [-1, 0, ... 7] | 0 |
| 30 | Urban\_or\_Rual\_Area | TINYINT | Nội ô hay ngoại ô [1 - ngoại ô, 2 - nội ô, 3 - chưa phân bổ] | 1 |
| 31 | Did\_Police\_Officer\_Attend\_Scene\_of\_Accident | SMALLINT | Cảnh sát có ở hiện trường tai nạn [-1, 1, 2, 3] | 1  2 |
| 32 | LOSA\_of\_Accident\_Location | VARCHAR(12) | Vị trí cho vùng nhỏ trên bản đồ của Anh | E01002835 |

**Bảng dữ liệu Casualties0515.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu dữ liệu có khả năng trong SQL SERVER** | **Ý nghĩa** | **Dữ liệu minh họa** |
| 1 | Accident\_Index | VARCHAR(20) | Mã định danh một vụ tai nạn | 201101BS70001 |
| 2 | Vehicle\_Reference | SMALLINT | Phân loại phương tiện | 1 |
| 3 | Casualty\_Reference | SMALLINT | Phân loại thương vong | 1 |
| 4 | Casualty\_Class | TINYINT | Mức độ thương vong, có 3 mức:   1. Người điều khiển xe đạp, xe hơi 2. Hành khách 3. Người đi bộ | 3 |
| 5 | Sex\_of\_Casualty | TINYINT | Giới tính của người gặp tai nạn. (1 - male, 2 - Female, 3 - Không xác định, 4 - Thiếu dữ liệu hoặc khác) | 2 |
| 6 | Age\_of\_Casualty | SMALLINT | Tuổi của người gặp tai nạn. | 25 |
| 7 | Age\_Band\_of\_Casualty | SMALLINT | Nhóm tuổi của người gặp tai nạn. Có 11 nhóm tuổi cách nhau 5 năm.Ví dụ: nhóm 1 (0 - 5), nhóm 2 (6 - 10), nhóm 11 (75 - trở lên) | 5 |
| 8 | Casualty\_Severity | TINYINT | Mức thương vong nghiêm trọng. (1 - tử vong, 2 - Bị thương nghiêm trọng, 3 - Bị thương nhẹ) | 3 |
| 9 | Pedestrian\_Location | SMALLINT | Vị trí của người đi bộ: Băng qua đường dành cho người đi bộ sang đường, trên đường đi bộ hoặc vỉa hè,… (đánh mã từ -1 đến 10) | 5 - Trên đường xe ngựa băng qua nơi khác |
| 10 | Pedestrian\_Movement | SMALLINT | Hành động của người đi bộ (băng qua đường từ phía gần tài xế; Trên đường xe chạy, đứng yêu không băng qua. (đánh số từ -1 đến 9) | 1 - băng qua đường từ phía gần tài xế |
| 11 | Car\_Passenger | SMALLINT | Số lượng xe khách trong vụ tai nạn. (-1, 0, 1, 2) | 0 |
| 12 | Bus\_or\_Coach\_Passenger | SMALLINT | Loại hành khách xe Buýt hoặc xe khách trong vụ tai nạn | 0 |
| 13 | Pedestrian\_Road\_Maintenance\_Worker | SMALLINT | Thuộc vào loại công nhân bảo trì đường hay người đi bộ không? | 2 |
| 14 | Casualty\_Type | TINYINT | Loại người tai nạn (người đi bộ, người đi xe đạp, người lái hoặc chở người trên moto từ 50cc trở xuống,…) được đánh số [0, 1, 2, 3, 4, 5, 8, 9, 10, 11, 16, 17, 18, 19, 20, 21, 22, 23, 90, 97, 98] | 0 - người đi bộ |
| 15 | Casualty\_Home\_Area\_Type |  | Loại khu vực nhà có nạn nhân | 1 |

**Bảng dữ liệu Vehicles0515.csv**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu dữ liệu có khả năng trong SQL SERVER** | **Ý nghĩa** | **Dữ liệu minh họa** |
| 1 | Accident\_Index | VARCHAR(20) | Mã số định danh vụ tai nạn | 201101BS70001 |
| 2 | Vehicle\_Reference | SMALLINT | Phân loại phương tiện [-1, 0,...] | 1 |
| 3 | Vehicle\_Type | SMALLINT | Loại phương tiện: xe moto 50cc trở xuống, taxi/oto thuê riêng,…đánh số gồm các giá trị [-1, 1, 2, 3, 4, 5, 8, 9, 10, 11, 16, 17, 18, 19, 20, 21, 22, 23, 90, 97, 98] | 9 - xe ô tô |
| 4 | Towing\_and\_Articulation | SMALLINT | Loại xe kéo, có khớp nối [-1,0,1,...5] | 0 |
| 5 | Vehicle\_Manoeuvre | TINYINT | Hành động của phương tiện: rẽ trái, dừng lại, đậu, chuyển sang làn phải,…Được đánh số là [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] | 18 |
| 6 | Vehicle\_Location | SMALLINT | Vị trí phương tiện: trên đường chính không ở làn đường cấm, trên làn xe bus,... được đánh số gồm [-1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] | 0 - trên đường chính không ở làn đường cấm |
| 7 | Junction\_Location | SMALLINT | Vị trí giao lộ: không ở hoặc trong vòng 20m của đường giao nhau; đến gần đường giao nhau hoặc chờ/đỗ ở gần đường giao nhau,... gồm các giá trị [-1, 0, 1, 2, 3, 4, 5, 6, 7, 8] | 8 - Giao lộ giữa trên bùng binh hoặc trục đường chính |
| 8 | Skidding\_and\_Overturning | SMALLINT | Phương tiện bị trượt hoặc lật [-1, 0, 1, 2, 5] | 0 |
| 9 | Hit\_Object\_in\_Carriageway | SMALLINT | Đâm trúng vật khác trên đường | 0 |
| 10 | Vehicle\_Leaving\_Carriageway | SMALLINT | Phương tiện rời khỏi hiện trường [-1, 0, 1, 7] | 0 |
| 11 | Hit\_Object\_off\_Carriageway | SMALLINT | Đâm trúng vật khác bên ngoài đường [-1,0,1,2,...12] | 0 |
| 12 | 1st\_Point\_of\_Impact | SMALLINT | Điểm tác động đầu tiên. Đánh mã gồm [-1, 0, 1, 2, 3, 4] | 1 |
| 13 | Was\_Vehicle\_Left\_Hand\_Drive | SMALLINT | Tay lái bên trái [-1,1,2] | 1 |
| 14 | Journey\_Purpose\_of\_Driver | SMALLINT | Mục đích lưu thông trên đường của người lái: đi làm việc, chở học sinh tới trường,... Được đánh số [-1, 1, 2, 3, 4, 5, 6, 15] | 6 - không rõ |
| 15 | Sex\_of\_Driver | TINYINT | Giới tính người lái. (1 - male, 2 - Female, 3 - Không xác định, 4 - Thiếu dữ liệu hoặc khác) | 1 - Nam |
| 16 | Age\_of\_Driver | SMALLINT | Tuổi của người lái | -1 - Không xác định |
| 17 | Age\_Band\_of\_Driver | TINYINT | Nhóm tuổi của người lái. Có 11 nhóm tuổi cách nhau 5 năm.Ví dụ: nhóm 1 (0 - 5), nhóm 2 (6 - 10), nhóm 11 (75 - trở lên) | -1 - Không xác định |
| 18 | Engine\_Capacity\_ | INT | Công suất động cơ [-1,2,4,8,...91000] | -1 - không xác định |
| 19 | Propulsion\_Code | SMALLINT | Mã lực đẩy. Gồm các giá trị [-1, 0, 1, .. 12] | -1 - không xác định |
| 20 | Age\_of\_Vehicle | SMALLINT | Năm sử dụng của phương tiện | -1 - không xác định |
| 21 | Driver\_IMD\_Decile | SMALLINT | Mức độ thiếu hụt tương đối của khu vực mà người lái xe liên quan đến vụ va chạm sinh sống | -1 |
| 22 | Driver\_Home\_Area\_Type | SMALLINT | Loại khu vực nhà xe của tài xế | -1 |

## Dữ liệu LSOA-Postcode mapping

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu dữ liệu có khả năng trong SQL SERVER** | **Ý nghĩa** | **Dữ liệu minh họa** |
| 1 | pcd7 | VARCHAR(8) | (PostCode 7) Phiên bản 7 ký tự của mã bưu điện | AB1 0AB |
| 2 | pcd8 | VARCHAR(8) | (PostCode 8) Phiên bản 8 ký tự của mã bưu điện | AB1 0AB |
| 3 | pcds | VARCHAR(8) | Khoảng cách giữa một quận/thị trấn với một đơn vị phân chia theo mã bưu điện | AB1 0AB |
| 4 | dointr | VARCHAR(6) | (Date of introduce: The date when the postcode was introduced). Ngày mà mã bưu điện được sử dụng | 199103 -> tháng 03 năm 1991 |
| 5 | doterm | VARCHAR(6) | (The date when the postcode was terminated). Ngày mà mã bưu điện chấm dứt sử dụng | 199207 -> tháng 07 năm 1992 |
| 6 | usertype | BIT | Loại người dùng của mã bưu điện (0: người dùng nhỏ, 1: người dùng lớn). Mã bưu chính của người dùng lớn được phân bổ cho các địa chỉ nhận 25 thư mỗi ngày. | 0; 1 |
| 7 | oa11cd | VARCHAR(20) | (Output Area Code). Mã khu vực đầu ra  dựa trên các cụm liền kề của mã bưu chính và những đo lường tiêu chuẩn của điều tra dân số. Vào link [https://www.doogal.co.uk/](https://www.doogal.co.uk/COA.php?code=S00090303)  Để xem chi tiết kinh độ, vĩ độ,… của khu vực | S00090303 |
| 8 | lsoa11cd | VARCHAR(10) | (Lower Layer Super Output Area). Mã vùng địa lý (quy mô nhỏ) | S01006514 |
| 9 | msoa11cd | VARCHAR(10) | (Middle Layer Super Output Area). Mã vùng địa lý (quy mô vừa) | S02001237 |
| 10 | ladcd | VARCHAR(10) | (Large Layer Super Output Area). Mã vùng địa lý (quy mô lớn) như khu vực/thành phố | S12000033 |
| 11 | lsoa11nm | VARCHAR(50) | Tên vùng địa lý (quy mô nhỏ) | Cults, Bieldside and Milltimber West - 02 |
| 12 | msoa11nm | VARCHAR(50) | Tên vùng địa lý (quy mô vừa) | Cults, Bieldside and Milltimber Wes |
| 13 | ladnm | VARCHAR(20) | Tên vùng địa lý (quy mô lớn) như khu vực/thành phố | Cardiff |
| 14 | ladnmw | VARCHAR(50) | Tên nhưng ghi theo thống kê bằng tiếng Welsh | Caerdydd |

## Dữ liệu UK-Postcodes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Kiểu dữ liệu có khả năng trong SQL SERVER** | **Ý nghĩa** | **Dữ liệu minh họa** |
| 1 | postcode | VARCHAR(8) | Mã bưu điện | S36 - [Sheffield](https://en.wikipedia.org/wiki/Sheffield)  BD10 - Bradford |
| 2 | easting | INT | Cách về phía Đông bao nhiêu mét | 440200  405600 |
| 3 | northing | INT | Cách về phía Bắc bao nhiêu mét | 284800  288200 |
| 4 | latitude | FLOAT | Vĩ độ | 53.50053  52.22985 |
| 5 | longitude | FLOAT | Kinh độ | 51.50316  -1.96275 |
| 6 | city | NVARCHAR(50) | Tên thành phố | Brierley  Nempnett Thrubwell |
| 7 | county | NVARCHAR(50) | Tên quận/huyện/thị trấn | Bath and North East Somerset  Birmingham |
| 8 | country\_code | VARCHAR(5) | Mã quốc gia | ENG  NIR |
| 9 | country\_name | NVARCHAR(50) | Tên quốc gia | England  Northern Ireland |
| 10 | iso3166-2 | VARCHAR(10) | Phần thứ 2 của tiêu chuẩn ISO3166, cho biết mã địa lý xác phân khu đơn vị hành chính quốc gia | GB-ENG  GB-SCT |
| 11 | region\_code | VARCHAR(10) | Mã vùng/khu vực | NORTHWEST  WESTMID |
| 12 | region\_name | NVARCHAR(50) | Tên vùng/khu vực | North West  Yorkshire and the Humber |

# THIẾT KẾ KDL - TỔNG HỢP, NẠP DỮ LIỆU CÁC NGUỒN VÀO KDL - THIẾT KẾ, XÂY DỰNG CUBE

# OLAP VÀ REPORT

# DATA MINING

# KẾT LUẬN CHUNG

# LINK BÁO CÁO

Link môn học của nhóm:

*<https://drive.google.com/drive/folders/1YcCjtJHqpSfbz3RAWGZD_p1V6InZCYMc?usp=sharing>*

Đường dẫn đến thư mục chứa video báo cáo:

*<https://drive.google.com/drive/folders/1-HO15z3e55suICYzTO4P1D0V3xtoW2cg?usp=sharing>*

# DANH MỤC TÀI LIỆU THAM KHẢO

*<https://ocsi.uk/2019/03/18/lsoas-leps-and-lookups-a-beginners-guide-to-statistical-geographies/?fbclid=IwAR1fTtyJdKUZjW0Cm6FlgQUONuRJidNLbW7WgToB6NPeuZZyViawqyst55Q>*

*<https://datadictionary.nhs.uk/nhs_business_definitions/middle_layer_super_output_area.html?fbclid=IwAR3y6wRVHUKq_D4xRmyI7xySxqgoaNMTeyZJRIpLuZekeYiPhOl6nPEHucw>*

*<https://datadictionary.nhs.uk/nhs_business_definitions/lower_layer_super_output_area.html?hl=lower%2Clayer%2Csuper%2Coutput%2Carea&fbclid=IwAR1UMDg64f-SmGCQRfPlywzKGR03cRt3AElUdrZ-1FjCi0v4pCORdt2_kjQ>*

*<https://www.datadaptive.com/docs/UK%20Political%20Maps%20(February%202017)%20-%20User%20Guide.pdf?fbclid=IwAR3LKJDNr5m1v09wSu27_UIzcJuIch37NOwf3RTBiKk4hDS7r8RvkROxx-Y>*