**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

****

**PROJECT SUBMISSION 1**

Phạm Ngọc Thùy Trang 18127022

Ngô Thị Thanh Thảo 18127053

Hoàng Nguyên Trúc 18127055

**Môn học: Hệ thống thông tin phục vụ trí tuệ kinh doanh**

**Thành phố Hồ Chí Minh**

**Ộ MÔN HỆ**

# Mục lục

[I. Mục lục 1](#_Toc85261529)

[II. Bảng thông tin chi tiết nhóm 2](#_Toc85261530)

[III. Phân công công việc 3](#_Toc85261531)

[IV. Mô tả ý nghĩa các thuộc tính của nguồn dữ liệu 4](#_Toc85261532)

[1. Postcode 4](#_Toc85261533)

[2. LSOA 4](#_Toc85261534)

[3. Bảng Casualties: 10](#_Toc85261535)

[4. Bảng Accidents: 14](#_Toc85261536)

[5. Bảng Vehicles: 18](#_Toc85261537)

[V. References 21](#_Toc85261538)

[ http://doc.ukdataservice.ac.uk/doc/7961/mrdoc/pdf/7961\_nspl\_user\_guide\_2011.pdf 21](#_Toc85261539)

[ Lower Super Output Area - Area (findthatpostcode.uk) (lsoa) 21](#_Toc85261540)

[ Output area - Area (findthatpostcode.uk) (oa) 21](#_Toc85261541)

[ Middle Super Output Area - Area (findthatpostcode.uk) (msoa) 21](#_Toc85261542)

[ Local Authority - Area (findthatpostcode.uk) (ladcd) 21](#_Toc85261543)

[ Postcode to Output Area to Lower Layer Super Output Area to Middle Layer Super Output Area to Local Authority District (February 2018) Lookup in the UK - data.gov.uk 21](#_Toc85261544)

# Bảng thông tin chi tiết nhóm

|  |  |  |  |
| --- | --- | --- | --- |
| **Mã nhóm:** | Nhóm 2 | | |
| **Số lượng:** | 3 | | |
| **MSSV** | **Họ tên** | **Email** | **SĐT** |
| 18127022 | Phạm Ngọc Thùy Trang | 18127022@  student.hcmus.edu.vn |  |
| 18127053 | Ngô Thị Thanh Thảo | 18127053@  student.hcmus.edu.vn |  |
| 18127055 | Hoàng Nguyên Trúc | 18127055@  student.hcmus.edu.vn |  |

# Phân công công việc

|  |  |  |
| --- | --- | --- |
| **Họ tên thành viên** | **Nội dung công việc** | **Phần trăm hoàn thành (%)** |
| Phạm Ngọc Thùy Trang |  |  |
| Ngô Thị Thanh Thảo |  |  |
| Hoàng Nguyên Trúc |  |  |

# Mô tả ý nghĩa các thuộc tính của nguồn dữ liệu

## Postcode

**Các thuộc tính sử dụng:**

* **Postcode (Mã bưu chính):** là mã được quy định bởi liên hiệp bưu chính toàn cầu giúp định vị khi chuyển thư, bưu phẩm. Hiểu đơn giản, mã bưu chính là một “tọa độ” của địa chỉ trên “bản đồ bưu chính”. Mỗi quốc gia có một hệ thống ký hiệu mã bưu chính riêng. Mỗi mã bưu chính trong một quốc gia, ký hiệu cho một địa phận bưu điện duy nhất. Địa phận bưu điện có thể là một làng nhỏ, một thành phố nhỏ, một quận của một thành phố lớn.
* **Latitude:** là vĩ độ - một giá trị xác định vị trí của một điểm trên bề mặt Trái Đất ở phía bắc hay phía nam của xích đạo.
* **Longitude:** là kinh độ - một giá trị tọa độ địa lý theo hướng đông-tây, được sử dụng phổ biến nhất trong bản đồ học và hoa tiêu toàn cầu
* **City:** Thành phố
* **County (Hạt):** Là một khu vực nằm trong một bang và lớn hơn thành phố, có chính quyền riêng để xử lý các công việc địa phương. Ví dụ, Hạt El Paso nằm trong bang Colorado, có thành phố Colorado Springs và các cộng đồng dân cư khác.
* **Region\_code:** mã vùng (mã khu vực địa lý). Thông thường, một khu vực địa lý có thể bao gồm nhiều quốc gia khác nhau.
* **Region name:** Tên khu vực địa lý tương ứng với region\_code
* **Country\_code:** Mã quốc gia (một loại mã địa lý bằng chữ hoặc bằng số được dùng để đại diện cho một quốc gia hoặc một khu vực phụ thuộc)
* **Country\_name:** Tên quốc gia tương ứng với country\_code

## LSOA

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên thuộc tính** | **Ý nghĩa/Mô tả thuộc tính** | **Phạm vi giá trị của mã** | **Độ dài thuộc tính** |
| **pcd7** | Còn được gọi là unit postcode (mã bưu chính độc nhất) phiên bản dài 7 ký tự. Pdc7 chính là những mã bưu chính còn hoạt động cho đến thời điểm hiện tại nằm trong vương quốc Anh, Channel Islands và Isle of Man có nhận thư hàng tháng từ Royal Mail - công ty dịch vụ bưu chính và chuyển phát nhanh đa quốc gia của Anh, cũng như tất cả những mã bưu chính bị đóng và không được sử dụng lại bởi Royal Mail trong vương quốc Anh và bởi các cơ quan hành chính của Channel Island và Isle of Man | AB1Δ1AA-ZE999ZZ (trong đó Δ đại diện cho ký tự khoảng trắng) | 7 ký tự (bao gồm ký tự khoảng trắng) |
| **pcd8** | Như trên, chỉ khác ở điểm đây là mã bưu chính độc nhất phiên bản 8 ký tự | AB1ΔΔ1AAZE99Δ9ZZ (trong đó Δ là ký tự khoảng trắng) | 8 ký tự (bao gồm ký tự khoảng trắng) |
| **oac11cd** | Viết tắt của cụm từ Output Area và có tên gọi đầy đủ là 2011 Census Output Area (hoặc là SA - Small Area - những khu vực nhỏ trong vương quốc Anh). Trong đó, pseudo oac11cd code sẽ được dùng cho Channel Islands và Isle of Man. oac11cd code sẽ để trống đối với các mã bưu chính không nằm trong hệ thống lưới. Oac11d được dùng để định nghĩa một cấp độ địa lý có cấp cao hơn unit postcode. Một output area sẽ có thể có hoặc không có khu vực kế vị (cũng là một output area), nhưng sẽ phải có một khu vực cha - hay còn gọi là SOA (Lower Super Output Area) | E00XXXXXX, W00XXXXXX, S00XXXXXX, 95AA010001- 95ZZ519999, pseudo hoặc null (không có thông tin).  Cụ thể hơn:  E00000001 - E00165665 = England; W00000001 - W00009769 = Wales; S00000001 - S00042604 = Scotland; 95AA010001 - 95ZZ519999 = Northern Ireland; L99999999 (pseudo code) = Channel Islands; M99999999 (pseudo code) = Isle of Man | 9 - 10 ký tự (hoặc chỉ có 1 ký tự nào trong trường hợp không có thông tin) |
| **ladnmw** | Tên khu vực tương ứng với mã oac11cd, nếu không có thì để trống |  |  |
| **lsoa11cd** | Viết tắt của cụm từ Lower Super Output Area và có tên gọi đầy đủ là 2011 Census Lower Layer Super Output Area, hoặc là Data Zone (DZ). **lsoa11cd** đại diện cho mã 2011 LSOA của khu vực England và xứ Wales, đại diện cho mã SOA của khu vực Northern Ireland and DZ code cho Scotland. Trong đó, pseudo **lsoa11cd** code sẽ được dùng cho Channel Islands và Isle of Man. **lsoa11cd** code sẽ để trống đối với các mã bưu chính không có mã OA. Một LSOA có thể có nhiều hoặc không có các output area. Thông thường, một LSOA sẽ phải có một MSOA (middle super output area). **lsoa11cd** sẽ có tên gọi là DZONE1 dùng để chỉ những mã bưu chính nằm trong Scotland, và sẽ được gọi là SOA1 khi nó đại diện cho những mã bưu chính nằm trong England và Wales. | E01XXXXXX  S01XXXXXX  W01XXXXXX  Null (nếu không có thông tin hiển thị)  Cụ thể hơn:  E01000001- E01032482 = England;  W01000001- W01001896 = Wales;  S99999999 (pseudo) = Scotland;  N99999999 (pseudo) = Northern Ireland;  L99999999 (pseudo) = Channel Islands;  M99999999 (pseudo) = Isle of Man | 9 ký tự (hoặc chỉ 1 ký tự null nếu không có thông tin) |
| **lsoa11nm** | Tên khu vực tương ứng với mã **lsoa11cd**, nếu không có thì để trống |  |  |
| **msoa11cd** | Viết tắt của cụm từ Middle Super Output Area và có tên gọi đầy đủ là 2011 Middle Lower Layer Super Output Area đối với khu vực England và Wales hoặc có tên là Intermediate Zone(IZ) đối với khu vực Scotland. Trong đó, pseudo **msoa11cd** code sẽ được dùng cho Channel Islands và Isle of Man. **msoa11cd** code sẽ để trống đối với các mã bưu chính nằm trong hệ thống lưới. Một MSOA có thể có nhiều hoặc không có LSOA cũng như có thể có hoặc không có các workplace zone. Thông thường, một MSOA sẽ phải có một khu vực cha gọi là Local Authority - LA (Chính quyền địa phương). Tuy nhiên, nó có thể có hoặc không có LSOA cũng như workplace zone. | * **E02XXXXXX** * **S02XXXXXX** * **W02XXXXXX**   Cụ thể là:  E02000001- E02006781 = England; W02000001- W02000413 = Wales; S99999999 (pseudo) = Scotland; N99999999 (pseudo) = Northern Ireland; L99999999 (pseudo) = Channel Islands; M99999999 (pseudo) = Isle of Man | 9 ký tự (hoặc chỉ 1 ký tự null nếu không có thông tin) |
| **msoa11nm** | Tên khu vực tương ứng với mã **msoa11cd**, nếu không có thì để trống. |  |  |
| **ladcd** | Viết tắt của cụm từ Local Authority, **ladcd** chính là  Quận hiện tại mà mã bưu chính đã được chỉ định. LA có rất nhiều tên đầy đủ khác nhau, trong đó có Local Authority District (LAD), hoặc là unitary authority (UA), hoặc là metropolitan district (MD), hoặc London borough (LB), hoặc council area (CA) hoặc district council area (DCA). Trong đó, pseudo **ladcd** code sẽ được dùng cho Channel Islands và Isle of Man. **ladcd** code sẽ để trống đối với các mã bưu chính không nằm trong hệ thống lưới. Một LA có thể có nhiều hoặc không có các phường cũng như có thể có hoặc không có các khu vực cộng đồng. Tuy nhiên, một LA cần phải thuộc về một quốc gia nào đó. | E06XXXXXX  E07XXXXXX  E08XXXXXX  E09XXXXXX  N09XXXXXX  S12XXXXXX  W06XXXXXX  Cụ thể hơn:  E06000001 - E06000056 = England (UA); E07000001 - E07000239 = England (LAD); E08000001 - E08000036 = England (MD); E09000001 - E09000033 = England (LB); W06000001 - W06000024 = Wales (UA); S12000005 - S12000044 = Scotland (CA); 95A – 95Z = Northern Ireland (DCA); L99999999 (pseudo) = Channel Islands; M99999999 (pseudo) = Isle of Man |  |
| **ladnm** | Tên chính quyền địa phương tương ứng với mã ladcd, nếu không có thì để trống |  |  |

## Bảng Casualties:

**Các thuộc tính sử dụng:**

* **Accident\_Index:** mã số tai nạn
* **Sex\_of\_Casualty:** giới tính nạn nhân. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Nam |
| 2 | Nữ |
| 3 | Không rõ |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Age\_Band\_of\_Casualty:** nhóm độ tuổi thương vong (0-5, 6-10, 11-15,...hơn 75 tuổi). Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | 0-5 |
| 2 | 6-10 |
| 3 | 11-15 |
| ... | ... |

* **Casualty\_Severity:** mức độ nghiêm trọng gồm ba mức độ :gây tử vong,

nghiêm trọng, nhẹ. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Gây tử vong |
| 2 | Nghiêm trọng |
| 3 | Nhẹ |

* **Casualty\_Type:** loại nạn nhân. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 0 | Người dân |
| 1 | Người đi xe đạp |
| 2 | Hành khách hoặc những người đi xe moto có phân khối <= 50cc |
| ... | ... |

* **Age\_of\_Casualty:** độ tuổi của nạn nhân bị thương vong.
* **Age\_Band\_of\_Casualty**: các nhóm độ tuổi của nạn nhân bị thương vong. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | 0-5 |
| 2 | 6-10 |
| 3 | 11-15 |

* **Casualty\_Home\_Area\_Type**: loại khu vực nạn nhân sinh sống. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Khu vực thành thị |
| 2 | Thị trấn nhỏ |
| 3 | Nông thôn |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Casualty\_Class**: lớp nạn nhân (gồm 3 loại: người lái xe hoặc tài xế, cư dân, hành khách). Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Tài xế hoặc người lái xe |
| 2 | Hành khách |
| 3 | Cư dân |

* **Car\_passenger:** Vị trí của hành khách bị xảy ra tai nạn nếu có đi ô tô. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải hành khách đi ô tô |
| 1 | Hành khách ngồi ghế trước |
| 2 | Hành khách ngồi ghế sau |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Bus\_passenger:** Vị trí của hành khách bị xảy ra tai nạn nếu có đi xe buýt. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải là hành khách đi xe buýt hoặc xe khách |
| 1 | Lên xe |
| 2 | Xuống xe |
| 3 | Hành khách đứng trên xe |
| 4 | Hành khách ngồi trên xe |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* **Pedestrian\_Location**: Vị trí của người dân bị xảy ra tai nạn. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải là khu dân cư |
| 1 | Đi qua dải đi bộ |
| 2 | Đi qua đường zig zag |
| ... | ... |

* **Pedestrian\_Movement:** Chuyển động của người dân. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 0 | Không phải là người đi |
| 1 | Băng qua từ phía gần người lái |
| 2 | Băng qua đường gần - bị che bởi xe đang đỗ hoặc đang dừng |
| ... | ... |

* **Vehicle\_Reference:** mã phương tiện

## Bảng Accidents:

**Các thuộc tính sử dụng:**

* **Accident\_Index:** mã số của tai nạn
* **Latitude:** là vĩ độ - một giá trị xác định vị trí của một điểm trên bề mặt Trái Đất ở phía bắc hay phía nam của xích đạo.
* **Longitude:** là kinh độ - một giá trị tọa độ địa lý theo hướng đông-tây, được sử dụng phổ biến nhất trong bản đồ học và hoa tiêu toàn cầu
  + **Accident\_Severity:** mức độ nghiêm trọng. Các giá trị được quy định như sau

|  |  |
| --- | --- |
| Code | Label |
| 1 | Gây tử vong |
| 2 | Nghiêm trọng |
| 3 | Nhẹ |

* **Local\_Authority(District):** địa phương xảy ra tai nạn

|  |  |
| --- | --- |
| Code | Label |
| 1 | Westminster |
| 2 | Camden |
| 3 | Islington |
| ... | ... |

* **Date:** Ngày/tháng/năm xảy ra tai nạn
* **Number\_of\_Vehicles**: Số lượng phương tiện tham gia
* **Number\_of\_Casualties**: Số lượng thương vong
* **Speed\_limit:** tốc độ tối đa cho phép
* **Local\_Authority\_(District)** **(tính theo quận):** Các mã LA (chính quyền địa phương) tại các quận. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Westminster |
| 2 | Camden |
| 3 | Islington |
| …. | …. |

* **Local\_Authority\_(Highway):** Các mã LA (chính quyền địa phương) tại các đường cao tốc. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| S12000033 | Aberdeen City |
| S12000034 | Aberdeenshire |
| S12000041 | Angus |
| ... | ... |

* **Day\_of\_Week**: thứ (buổi) xảy ra tai nạn. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Chủ nhật |
| 2 | Thứ hai |
| 3 | Thứ bai |
| ... | ... |
| 7 | Thứ bảy |

* **Time:** Thời điểm xảy ra tai nạn (giờ:phút)
* **Weather\_Conditions**: Điều kiện thời tiết. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Thời tiết ổn và không có gió mạnh, gió giật |
| 2 | Trời mưa nhưng không có gió mạnh, gió giật |
| 3 | Tuyết rơi nhưng không có gió mạnh, gió giật |
| ... | …. |

* **Light\_Conditions:** Điều kiện ánh sáng. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Ánh sáng ban ngày |
| 2 | Trời tối, đèn sáng |
| 3 | Trời tối, đèn không sáng |
| ... | ... |

* + **Road\_Surface**: Bề mặt đường. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Khô ráo |
| 2 | Ẩm ướt |
| 3 | Tuyết phủ |
| ... | ... |

* + **Road\_Type:** Loại đường. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Vòng xoay |
| 2 | Đường một chiều |
| 3 | Xa lộ hai chiều |
| ... | ... |

* + **Urban\_or\_Rural\_Area:** Loại khu vực. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Đô thị |
| 2 | Nông thôn |
| 3 | Chưa phân bổ |

* **LSOA\_of\_Accident\_Location**: Mã vị trí nơi xảy ra tai nạn

## Bảng Vehicles:

**Các thuộc tính sử dụng:**

* + **Accident\_Index:** Mã số tai nạn
  + **Vehicle\_Reference:** Mã phương tiện
  + **Driver\_Home\_Area\_Type:** Loại khu vực nhà ở của tài xế. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Khu đô thị |
| 2 | Thị trấn nhỏ |
| 3 | Nông thôn |
| -1 | Không có dữ liệu |

* + **Sex\_of\_Driver:** Giới tính của tài xế. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Nam |
| 2 | Nữ |
| 3 | Không rõ |
| -1 | Dữ liệu bị thiếu hoặc nằm ngoài phạm vi dữ liệu có sẵn |

* + **Age\_Band\_of\_Driver:** Nhóm độ tuổi của tài xế (0-5, 6-10, 11-15,...hơn 75 tuổi). Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | 0-5 |
| 2 | 6-10 |
| 3 | 11-15 |
| ... | ... |

* **Journey\_Purpose\_of\_Driver:** Mục đích hành trình. Các giá trị được quy

định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Hành trình như một phần công việc |
| 2 | Đi làm/Tan làm |
| 3 | Đưa/đón học sinh đi học |
| ... | ... |

- **Was\_Vehicle\_Left\_Hand\_Drive:** Tay thuận của tài xế có là tay trái không

|  |  |
| --- | --- |
| Code | Label |
| 1 | Không |
| 2 | Có |
| -1 | Không có dữ liệu |

- **Vehicle\_Type:** Loại phương tiện. Các giá trị được quy định như sau:

|  |  |
| --- | --- |
| Code | Label |
| 1 | Xe đạp |
| 2 | Xe mô tô dưới 50 phân khối |
| 3 | Xe mô tô dưới 125 phân khối |
| ... | ... |

# References

## <http://doc.ukdataservice.ac.uk/doc/7961/mrdoc/pdf/7961_nspl_user_guide_2011.pdf>

## [Lower Super Output Area - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/lsoa11.html) (lsoa)

## [Output area - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/oa11.html) (oa)

## [Middle Super Output Area - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/msoa11.html) (msoa)

## [Local Authority - Area (findthatpostcode.uk)](https://findthatpostcode.uk/areatypes/laua.html?p=45) (ladcd)

## [Postcode to Output Area to Lower Layer Super Output Area to Middle Layer Super Output Area to Local Authority District (February 2018) Lookup in the UK - data.gov.uk](https://data.gov.uk/dataset/9b090605-9861-4bb4-9fa4-6845daa2de9b/postcode-to-output-area-to-lower-layer-super-output-area-to-middle-layer-super-output-area-to-local-authority-district-february-2018-lookup-in-the-uk)