

**ĐẠI HỌC QUỐC GIA TP HCM**  
**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**



**MÔN: TRỰC QUAN HÓA DỮ LIỆU**  
**BÁO CÁO ĐỒ ÁN CUỐI KÌ**  
**LỚP HỌC PHẦN: 21KHDL – NHÓM 3**

**Thành viên nhóm:** 21127012 – Trần Huy Bân  
21127191 – Nguyễn Nhật Truyền  
21127430 – Nguyễn Huy Thành  
21127507 – Đinh Công Huy Hoàng  
21127592 – Nguyễn Minh Đạt

**Giảng viên hướng dẫn:** Thầy Bùi Tiên Lên  
Thầy Lê Ngọc Thành  
Thầy Lê Nguyễn Nhựt Trường

*Tp.HCM, 1 tháng 5 năm 2024*

## TABLE OF CONTENT

|  |    |
|--|----|
| <b>I. THÀNH VIÊN</b>                             | 3  |
| <b>II. DỮ LIỆU</b>                               | 3  |
| 1. Giới thiệu về bộ dữ liệu:                     | 3  |
| 2. Nguồn dữ liệu:                                | 3  |
| 3. Cách lấy dữ liệu từ Tổng Cục Thống Kê:        | 4  |
| 4. Nội dung dữ liệu thô:                         | 4  |
| 5. Xử lý dữ liệu:                                | 6  |
| 6. Phân tích dữ liệu:                            | 9  |
| <b>III. INSIGHTS:</b>                            | 18 |
| 1. Xét theo tiêu chí: Nguồn dữ liệu đáng tin cậy | 18 |
| 2. Xét theo tiêu chí: Phù hợp với mục đích       | 18 |
| 3. Xét theo tiêu chí: Rõ ràng và dễ hiểu         | 19 |
| 4. Xét theo tiêu chí: Sự tích hợp và liên kết    | 20 |
| 5. Xét theo tiêu chí: Phân tích dữ liệu          | 20 |
| 6. Xét theo tiêu chí: Tương tác và điều hướng    | 25 |
| 7. Thiết kế hấp dẫn                              | 26 |

## I. THÀNH VIÊN

| STT | HỌ VÀ TÊN           | MSSV     | CÔNG VIỆC  |
|-----|---------------------|----------|--|
| 1   | Trần Huy Bân        | 21127012 | - Tổng hợp dữ liệu<br>- Xử lý dữ liệu                                    |
| 2   | Nguyễn Nhật Truyền  | 21127192 | - Xây dựng dashboard   |
| 3   | Nguyễn Huy Thành    | 21127430 | - Viết report<br>- Phân tích dữ liệu                                     |
| 4   | Đinh Công Huy Hoàng | 21127507 | - Tổng hợp report<br>- Xây dựng bố cục của dashboard<br>- Chuẩn bị slide |
| 5   | Nguyễn Minh Đạt     | 21127592 | - Viết report<br>- Phân tích dữ liệu                                     |

## II. DỮ LIỆU

### 1. Giới thiệu về bộ dữ liệu:

Dữ liệu liên quan đến dân số và thu nhập của 63 tỉnh thành và các khu vực lớn cũng như cả nước Việt Nam từ năm 1995 đến năm 2022.

Trong đó dữ liệu thô ban đầu gồm có 3 tập dữ liệu như sau:

- [Diện tích, dân số và mật độ dân số phân theo địa phương](#)
- [Dân số trung bình phân theo địa phương, giới tính và thành thị nông thôn](#)
- [Thu nhập bình quân đầu người một tháng theo giá hiện hành phân theo nguồn thu và phân theo địa phương](#)

### 2. Nguồn dữ liệu:

Tất cả dữ liệu được lấy từ trang [Tổng Cục Thống Kê](#).

→ Xét về độ tin cậy thì dữ liệu là hoàn toàn đáng tin cậy do Tổng Cục Thống Kê là cơ quan trực thuộc Bộ Kế hoạch và Đầu tư, thực hiện chức năng tham mưu, giúp Bộ trưởng Bộ Kế hoạch và Đầu tư quản lý nhà nước về thống kê, điều phối hoạt động thống kê, tổ chức các hoạt động thống kê và cung cấp thông tin thống kê kinh tế - xã hội cho các cơ quan, tổ chức và cá nhân theo quy định của pháp luật.

### 3. Cách lấy dữ liệu từ Tổng Cục Thống Kê:

Vì vấn đề liên quan đến định dạng và format của các tập dữ liệu khi lấy từ Tổng Cục Thống Kê, nhóm quyết định thực hiện lấy dạng file excel (.xlsx)

Tuy nhiên, ở đây, dữ liệu vẫn chưa thể dùng được vì không thể mở được mà phải thông qua các công cụ trung gian để mở và lấy về các tập dữ liệu mới.

Ví dụ về cách lấy dữ liệu từ Tổng Cục Thống Kê được minh họa ở hình dưới như sau:

| Địa phương *   | Năm *   | Chỉ tiêu *   |
|--|---|--|
| <div><input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></div> <div>Tổng số 70 Đã chọn 70</div> <div>Kiên Giang<br/>Cần Thơ<br/>Hậu Giang<br/>Sóc Trăng<br/>Bạc Liêu<br/>Cà Mau</div> <div>Tìm kiếm <input type="text"/> <input type="button" value="➤"/></div> <div><input type="checkbox"/> Từ đầu của hàng</div> | <div><input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></div> <div>Tổng số 12 Đã chọn 12</div> <div>2017<br/>2018<br/>2019 (*)<br/>2020 (*)<br/>2021 (*)<br/>2022 (*)</div> <div>Tìm kiếm <input type="text"/> <input type="button" value="➤"/></div> <div><input type="checkbox"/> Từ đầu của hàng</div> | <div><input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/></div> <div>Tổng số 3 Đã chọn 3</div> <div>Diện tích(Km2)<br/>Dân số trung bình (Nghìn người)<br/>Mật độ dân số (Người/km2)</div> <div>Tìm kiếm <input type="text"/> <input type="button" value="➤"/></div> <div><input type="checkbox"/> Từ đầu của hàng</div> |

Số ô dữ liệu đã chọn **2.520** (maximum number allowed is 100.000)  
Hiển thị trên màn hình bị giới hạn bởi 1.000 dòng và 30 cột

Excel (xlsx)

Figure 1: Cách thức lấy dữ liệu về

→ Sau khi đã chọn tất cả xong thì người dùng chỉ cần chọn tiếp tục để lấy được tập dữ liệu. Vì ở đây có đến 3 tập dữ liệu nên chúng ta cần làm 3 lần.

### 4. Nội dung dữ liệu thô:

#### a. Đối với dữ liệu diện tích và dân số:

Dữ liệu gồm có **tiêu đề** ở dòng đầu tiên và các năm từ năm 2011 đến 2022.

Dữ liệu ứng với mỗi năm gồm các cột thuộc tính (3 cột):

- Diện tích (km<sup>2</sup>)
- Dân số trung bình (nghìn người)
- Mật độ dân số (người/km<sup>2</sup>)

Minh họa dữ liệu ban đầu như sau:

| Diện tích, dân số và mật độ dân số phân theo địa phương chia theo Địa phương, Năm và Chỉ tiêu |                |                                 |                           |                |                                 |                           |
|---|----------------|---------------------------------|---------------------------|----------------|---------------------------------|---------------------------|
|   | 2011           |                                 |                           | 2012           |                                 |                           |
|   | Diện tích(Km2) | Dân số trung bình (Nghìn người) | Mật độ dân số (Người/km2) | Diện tích(Km2) | Dân số trung bình (Nghìn người) | Mật độ dân số (Người/km2) |
| CẢ NƯỚC   | 330957.6       | 88145.83                        | 266.34                    | 330951.1       | 89202.89                        | 269.53                    |
| Đồng bằng sông Hồng   | 21068.1        | 20216.81                        | 959.59                    | 21050.9        | 20525.36                        | 975.03                    |
| Hà Nội  | 3328.9         | 6825.82                         | 2050.47                   | 3323.6         | 6991.43                         | 2103.57                   |
| Vĩnh Phúc   | 1236.5         | 1029.5                          | 832.59                    | 1236.5         | 1047.39                         | 847.06                    |
| Bắc Ninh  | 822.7          | 1072.38                         | 1303.49                   | 822.7          | 1104.69                         | 1342.76                   |
| Quảng Ninh  | 6102.4         | 1181.53                         | 193.62                    | 6102.3         | 1198.39                         | 196.38                    |
| Hải Dương   | 1656           | 1747.83                         | 1055.45                   | 1656           | 1767.83                         | 1067.53                   |
| Hải Phòng   | 1523.4         | 1886.2                          | 1238.15                   | 1523.9         | 1912.94                         | 1255.29                   |
| Hưng Yên  | 926            | 1150.74                         | 1242.7                    | 926            | 1162.91                         | 1255.84                   |
| Thái Bình   | 1570           | 1798.66                         | 1145.65                   | 1570           | 1806.48                         | 1150.63                   |
| Hà Nam  | 860.5          | 795.65                          | 924.63                    | 860.5          | 800.81                          | 930.63                    |
| Nam Định  | 1651.4         | 1816.81                         | 1100.16                   | 1652.6         | 1811.5                          | 1096.15                   |
| Ninh Bình   | 1390.3         | 911.69                          | 655.75                    | 1376.7         | 920.99                          | 668.98                    |
| Trung du và miền núi phía Bắc   | 95264.4        | 11311.41                        | 118.74                    | 95272.3        | 11441.84                        | 120.1                     |
| Hà Giang  | 7914.9         | 749.54                          | 94.7                      | 7914.9         | 763.5                           | 96.46                     |
| Cao Bằng  | 6707.9         | 514.02                          | 76.63                     | 6707.9         | 515.88                          | 76.91                     |
| Bắc Kạn   | 4859.4         | 298.69                          | 61.47                     | 4859.4         | 300.8                           | 61.9                      |

Figure 2: Dữ liệu minh họa ở tập đầu tiên

## b. Đối với dữ liệu dân số theo giới tính, thành thị và nông thôn:

Dữ liệu gồm có **tiêu đề** ở dòng đầu tiên và 5 cột thuộc tính ở dòng kế tiếp lần lượt là:

- Tổng số (tổng dân số)
- Nam (dân số là Nam)
- Nữ (dân số là Nữ)
- Thành thị (dân số sống ở thành thị)
- Nông thôn (dân số sống ở nông thôn)

Khác với dữ liệu về dân số và diện tích thì các cột thuộc tính sẽ ở phía trên và tiếp sau đó là các năm theo thứ tự từ năm 1995 đến 2022.

Minh họa dữ liệu ban đầu như sau:

| Dân số trung bình phân theo địa phương, giới tính và thành thị nông thôn chia theo Tỉnh, thành phố, Dân số trung bình và Năm |         |         |         |         |         |         |         |         |         |         |         |         |
|--|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|  | Tổng số |         |         |         |         |         |         |         |         |         |         |         |
|  | 1995.00 | 1996    | 1997    | 1998    | 1999    | 2000    | 2001    | 2002    | 2003    | 2004    | 2005    | 2006    |
| CẢ NƯỚC  | 71995.5 | 73156.7 | 74306.9 | 75456.3 | 76596.7 | 77630.9 | 78620.5 | 79537.7 | 80467.4 | 81436.4 | 82392.1 | 83311.2 |
| Đồng bằng sông Hồng  | 17078.4 | 17289.8 | 17494.8 | 17692.9 | 17877.8 | 18060.7 | 18249.9 | 18432.9 | 18617.5 | 18807.9 | 18976.7 | 19108.9 |
| Hà Nội   | 2431    | 2492.9  | 2556    | 2621.5  | 2685    | 2767.7  | 2852.9  | 2928.3  | 3000.3  | 3071.4  | 3133.4  | 3184.8  |
| Hà Tây   | 2299    | 2328    | 2353    | 2373.6  | 2391.7  | 2430.2  | 2475.5  | 2527    | 2587.3  | 2649.1  | 2702.9  | 2742.4  |
| Vĩnh Phúc  | 1048.2  | 1061.9  | 1075.8  | 1085.7  | 1094.3  | 1104.1  | 1114.1  | 1124.3  | 1134.1  | 1145.2  | 1157    | 1166    |
| Bắc Ninh   | 916     | 925.3   | 931.7   | 937.6   | 943     | 950.6   | 958.9   | 967.6   | 975.3   | 983.2   | 991.1   | 999.8   |
| Quảng Ninh   | 941.7   | 958     | 974.4   | 991.4   | 1007.2  | 1024.2  | 1039.3  | 1054.4  | 1068.4  | 1081.8  | 1096.1  | 1109.3  |
| Hải Dương  | 1609.1  | 1619.5  | 1630.5  | 1641.3  | 1651    | 1656.9  | 1662.7  | 1668.1  | 1673.7  | 1680.4  | 1685.5  | 1689.7  |
| Hải Phòng  | 1608.2  | 1625    | 1641.9  | 1659.5  | 1676.7  | 1691.5  | 1708.6  | 1725.3  | 1741    | 1757.8  | 1773.4  | 1789.1  |
| Hưng Yên   | 1033.2  | 1042.6  | 1052.2  | 1061.7  | 1071.4  | 1079    | 1086    | 1091.5  | 1097    | 1104.3  | 1111    | 1116.4  |
| Thái Bình  | 1752.3  | 1761    | 1769.5  | 1778.5  | 1788.1  | 1789.4  | 1790.3  | 1790.9  | 1790.9  | 1791.2  | 1790.5  | 1787.4  |
| Hà Nam   | 763.7   | 771.6   | 779.4   | 786.6   | 792.5   | 793.8   | 793.2   | 792.8   | 792.3   | 791.9   | 791.3   | 790     |
| Nam Định   | 1820.5  | 1838    | 1856.2  | 1874.4  | 1891.9  | 1886.5  | 1879.7  | 1872.8  | 1866.3  | 1859.4  | 1851    | 1839.4  |
| Ninh Bình  | 855.5   | 866     | 874.2   | 881.1   | 885     | 886.8   | 888.7   | 889.9   | 890.9   | 892.2   | 893.5   | 894.6   |
| Trung du và miền núi phía Bắc  | 9522.9  | 9679.7  | 9820.8  | 9951.2  | 10085.3 | 10204.4 | 10321.8 | 10440   | 10559.5 | 10681   | 10798.7 | 10904.3 |
| Hà Giang   | 550.3   | 564.2   | 577.7   | 590.4   | 605.9   | 619.2   | 631.6   | 643.9   | 656.2   | 669     | 681.7   | 693.4   |

Figure 3: Dữ liệu minh họa ở tập thứ hai

### c. Đối với dữ liệu dân số theo giới tính, thành thị và nông thôn:

Dữ liệu gồm có tiêu đề ở dòng đầu tiên và các năm từ năm 2010 đến 2022.

Dữ liệu ứng với mỗi năm gồm các cột thuộc tính (5 cột):

- Tổng số (thu nhập bình quân đầu người 1 tháng tính theo nghìn VNĐ)
- Thu từ tiền lương, tiền công
- Thu từ nông, lâm, thủy sản
- Thu từ phi nông, lâm, thủy sản
- Thu từ các khoản thu khác

Minh họa dữ liệu ban đầu như sau:

| Thu nhập bình quân đầu người một tháng theo giá hiện hành phân theo nguồn thu và phân theo địa phương chia theo Địa phương, Năm và Nguồn thu |         |                              |                                   |                                    |                    |         |
|--|---------|------------------------------|-----------------------------------|------------------------------------|--------------------|---------|
|  | 2010    |                              |                                   |                                    |                    | 2012    |
|  | Tổng số | Thu từ tiền lương, tiền công | Thu từ nông, lâm nghiệp, thủy sản | Thu phi nông, lâm nghiệp, thủy sản | Các khoản thu khác | Tổng số |
| TỔNG SỐ  | 1387    | 622                          | 279                               | 328                                | 158                | 923     |
| Đồng bằng sông Hồng  | 1580    | 798                          | 189                               | 392                                | 201                | 1216    |
| Hà Nội   | 2013    | 1165                         | 109                               | 467                                | 273                | 1741    |
| Vĩnh Phúc  | 1232    | 519                          | 256                               | 317                                | 140                | 784     |
| Bắc Ninh   | 1646    | 560                          | 190                               | 707                                | 189                | 875     |
| Quảng Ninh   | 1787    | 895                          | 176                               | 533                                | 183                | 1367    |
| Hải Dương  | 1306    | 600                          | 258                               | 286                                | 163                | 996     |
| Hải Phòng  | 1694    | 958                          | 115                               | 418                                | 204                | 1278    |
| Hưng Yên   | 1199    | 522                          | 261                               | 275                                | 141                | 785     |
| Thái Bình  | 1129    | 518                          | 251                               | 216                                | 143                | 878     |
| Hà Nam   | 1150    | 470                          | 264                               | 248                                | 168                | 710     |
| Nam Định   | 1237    | 477                          | 286                               | 322                                | 153                | 765     |
| Ninh Bình  | 1202    | 498                          | 264                               | 280                                | 160                | 762     |
| Trung du và miền núi phía Bắc  | 905     | 359                          | 314                               | 146                                | 86                 | 536     |
| Hà Giang   | 610     | 193                          | 317                               | 61                                 | 38                 | 274     |

Figure 4: Dữ liệu minh họa ở tập thứ ba

## 5. Xử lý dữ liệu:

Trước khi tiến hành vào các bước đọc dữ liệu và tiền xử lý dữ liệu, chúng ta sẽ điểm sơ qua về dữ liệu có trong từng tập dữ liệu. Trong đó, nhóm nhận thấy rằng:

- Đối với dữ liệu về dân số, mặc dù ở tập dữ liệu về dân số và diện tích chỉ có từ năm 2011 đến năm 2022 nhưng ở dữ liệu riêng về dân số thì lại có cả từ năm 1995 đến 2022.
- Yếu tố thứ hai mà nhóm nhận thấy là sự thay đổi về diện tích và dân số của thủ đô Hà Nội có vẻ khác thường từ năm 2008. Sau khi tìm hiểu thì nhóm thấy rằng là do vào tháng 1 năm 2008 thì tỉnh Hà Tây đã được sáp nhập vào thủ đô Hà Nội.

→ Dựa trên các yếu tố trên, nhóm quyết định tiến hành xử lý **2 tập dữ liệu** liên quan đến dân số trước rồi mới xử lý dữ liệu về thu nhập của từng tỉnh. Vì vậy, ở việc xử lý dữ liệu, nhóm sẽ thực hiện song song cho việc đọc dữ liệu và xử lý dữ liệu để có thể kết nối được các tập dữ liệu với nhau chứ không thể xử lý riêng lẻ được.

## a. Đọc dữ liệu

Đầu tiên, nhóm sẽ tiến hành đọc dữ liệu của tập dữ liệu về dân số và diện tích trước. Như đã trình bày ở phần trên, dữ liệu hiện tại đang nằm ở dạng không đúng ý muốn. Chúng ta sẽ chuyển nó về dạng dữ liệu gồm **5 cột chính** là:

- Tỉnh
- Diện tích
- Mật độ dân số
- Dân số
- Năm (từ 2011 đến 2022)

Bằng cách đọc bằng *pandas* và *skiprows = 2*, sau đó loại bỏ dòng đầu tiên là các tiêu đề. Dữ liệu bây giờ ta nhận được:

| Unnamed: 0           | 2011                | Unnamed: 2     | Unnamed: 3                      | 2012                      | Unnamed: 5     | Unnamed: 6                      | 2013                      | Unnamed: 8     | Unnamed: 9                      | ...                       | Unnamed: 27               | 2020           | Unnamed: 29                     | Unnamed: 30               | 2021           | Unnamed: 32                     |          |
|----------------------|---------------------|----------------|---------------------------------|---------------------------|----------------|---------------------------------|---------------------------|----------------|---------------------------------|---------------------------|---------------------------|----------------|---------------------------------|---------------------------|----------------|---------------------------------|----------|
| 0                    | NaN                 | Diện tích(Km2) | Dân số trung bình (Nghìn người) | Mật độ dân số (Người/km2) | Diện tích(Km2) | Dân số trung bình (Nghìn người) | Mật độ dân số (Người/km2) | Diện tích(Km2) | Dân số trung bình (Nghìn người) | Mật độ dân số (Người/km2) | Mật độ dân số (Người/km2) | Diện tích(Km2) | Dân số trung bình (Nghìn người) | Mật độ dân số (Người/km2) | Diện tích(Km2) | Dân số trung bình (Nghìn người) |          |
| 1                    | CÁ NƯỚC             | 330957.6       | 88145.83                        | 266.34                    | 330951.1       | 89202.89                        | 269.53                    | 330972.4       | 90191.42                        | 272.5                     | ...                       | 291            | 331317.1                        | 97582.69                  | 295            | 331344.27                       | 98506.19 |
| 2                    | Đồng bằng sông Hồng | 21068.1        | 20216.81                        | 959.59                    | 21050.9        | 20525.36                        | 975.03                    | 21059.3        | 20785.63                        | 987                       | ...                       | 1064           | 21254.6                         | 22920.18                  | 1078           | 21278.46                        | 23224.84 |
| 3                    | Hà Nội              | 3328.9         | 6825.82                         | 2050.47                   | 3323.6         | 6991.43                         | 2103.57                   | 3324.3         | 7128.39                         | 2144.33                   | ...                       | 2410           | 3359.84                         | 8246.54                   | 2454           | 3359.82                         | 8330.83  |
| 4                    | Vĩnh Phúc           | 1236.5         | 1029.5                          | 832.59                    | 1236.5         | 1047.39                         | 847.06                    | 1238.6         | 1061.5                          | 857.02                    | ...                       | 934            | 1236                            | 1171.2                    | 948            | 1236                            | 1191.78  |
| ...                  | ...                 | ...            | ...                             | ...                       | ...            | ...                             | ...                       | ...            | ...                             | ...                       | ...                       | ...            | ...                             | ...                       | ...            | ...                             | ...      |
| 66                   | Cần Thơ             | 1409           | 1195.93                         | 848.78                    | 1409           | 1200.93                         | 852.33                    | 1409           | 1203.23                         | 853.96                    | ...                       | 859            | 1440.4                          | 1240.73                   | 861            | 1440.4                          | 1246.99  |
| 67                   | Hậu Giang           | 1602.5         | 752.73                          | 469.72                    | 1602.5         | 750.73                          | 468.47                    | 1602.4         | 747.88                          | 466.73                    | ...                       | 451            | 1622.2                          | 729.78                    | 450            | 1622.23                         | 729.89   |
| 68                   | Sóc Trăng           | 3311.6         | 1273.93                         | 384.69                    | 3311.6         | 1266.42                         | 382.42                    | 3311.6         | 1254.98                         | 378.96                    | ...                       | 362            | 3298.2                          | 1195.74                   | 363            | 3298.2                          | 1206.82  |
| 69                   | Bạc Liêu            | 2468.7         | 866.87                          | 351.15                    | 2468.7         | 872.14                          | 353.28                    | 2468.7         | 877.04                          | 355.26                    | ...                       | 340            | 2667.9                          | 913.48                    | 342            | 2667.88                         | 918.51   |
| 70                   | Cà Mau              | 5294.9         | 1204.5                          | 227.48                    | 5294.9         | 1203.46                         | 227.29                    | 5294.9         | 1202.07                         | 227.02                    | ...                       | 229            | 5274.5                          | 1193.89                   | 226            | 5274.51                         | 1208.75  |
| 71 rows x 37 columns |                     |                |                                 |                           |                |                                 |                           |                |                                 |                           |                           |                |                                 |                           |                |                                 |          |

Figure 5: Dữ liệu sau khi thực hiện chuyển lại thành các tỉnh

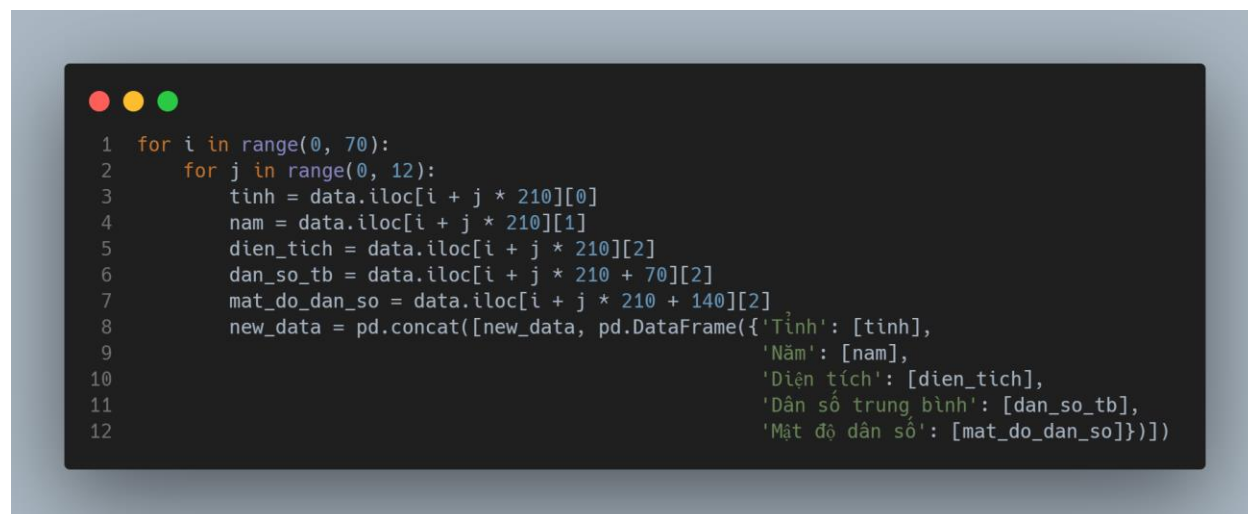
Sau đó, dùng phương thức *melt* dữ liệu thành 3 cột, ta được:

| Unnamed: 0 | Năm                 | Giá trị       |
|------------|---------------------|---------------|
| 0          | CÁ NƯỚC             | 2011 330957.6 |
| 1          | Đồng bằng sông Hồng | 2011 21068.1  |
| 2          | Hà Nội              | 2011 3328.9   |
| 3          | Vĩnh Phúc           | 2011 1236.5   |
| 4          | Bắc Ninh            | 2011 822.7    |
| ...        | ...                 | ...           |

Figure 6: Gom các dữ liệu lại

Nhận thấy, dữ liệu cột tỉnh đang nằm ở **cột đầu tiên**, các tên cột gồm các giá trị năm và các cột *unnamed 2* đến *unnamed 36* nằm ở **cột thứ hai**, các giá trị năm ở **cột thứ ba** với thứ tự là diện tích, dân số và mật độ dân số.

Bây giờ, chúng ta tạo ra một **DataFrame** mới gồm **5 cột** như đã nêu và tiến hành lặp qua 2 vòng lặp. Ý tưởng ở đây đó là 1 vòng lặp lặp qua **70** tỉnh và khu vực, một vòng lặp con bên trong lặp qua **12** năm từ 2011 đến 2002 với công thức như đoạn code sau:



```
1 for i in range(0, 70):
2     for j in range(0, 12):
3         tinh = data.iloc[i + j * 210][0]
4         nam = data.iloc[i + j * 210][1]
5         dien_tich = data.iloc[i + j * 210][2]
6         dan_so_tb = data.iloc[i + j * 210 + 70][2]
7         mat_do_dan_so = data.iloc[i + j * 210 + 140][2]
8         new_data = pd.concat([new_data, pd.DataFrame({'Tỉnh': [tinh],
9                                                         'Năm': [nam],
10                                                        'Diện tích': [dien_tich],
11                                                        'Dân số trung bình': [dan_so_tb],
12                                                        'Mật độ dân số': [mat_do_dan_so]})])
```

Figure 7: Gom các dữ liệu theo từng tỉnh và khu vực

Sau khi thực hiện vòng lặp trên, ta đã hoàn thành việc xử lý dữ liệu và đây cũng là cơ chế để xử lý cho **2 tập dữ liệu** còn lại nhưng ta cần chỉnh sửa tùy vào số lượng cột và số lượng năm.

#### b. Tiền xử lý với hai tập dữ liệu đầu

Với cách đọc dữ liệu đã nêu ở trên, sau mỗi lần đọc dữ liệu, chúng ta sẽ **merge** dữ liệu lại theo thứ tự như sau:

Trước hết, đọc dữ liệu dân số trước và đọc tiếp dữ liệu về dân số phân theo giới tính, thành thị/nông thôn. Sau đó tiến hành nối hai dữ liệu này lại.

Tuy nhiên, như đã trình bày ở trên, khi merge lại, chúng ta sẽ merge theo tỉnh, năm và dân số thì ở đây do dữ liệu thứ hai có nhiều năm hơn (từ 1995 đến 2022), vì vậy nó sẽ có những dữ liệu trống về diện tích và mật độ dân số.

Phương thức xử lý ở đây đó là với cột diện tích, sử dụng **backward fill** để fill các diện tích và sau đó dùng cột dân số chia cho các diện tích này để ra được mật độ dân số.



### c. Tiền xử lý với tập dữ liệu còn lại

Đối với dữ liệu liên quan đến thu nhập, ta sẽ thực hiện merge theo tỉnh và năm nhưng do dữ liệu thiếu các năm cũ nên ta cũng dùng *backward fill*.

Tuy nhiên, điểm mấu chốt gây khó khăn trong phần xử lý đó là có nhưng ô dữ liệu trống đến từ cả 3 tập dữ liệu thì dữ liệu được ghi bởi chuỗi “.” và ta cần xử lý nó trước trước khi thực hiện tiền xử lý.

Ngoài ra, đối với dữ liệu, có tỉnh đó là tỉnh **Hà Tây** và ta cần loại bỏ tỉnh này.

Cuối cùng, một điểm cần lưu ý đó là do format về kiểu chữ, một số tỉnh cần thay đổi lại tên trước khi merge các tập dữ liệu lại để có thể thực hiện được việc merge dữ liệu.

### 6. Phân tích dữ liệu:

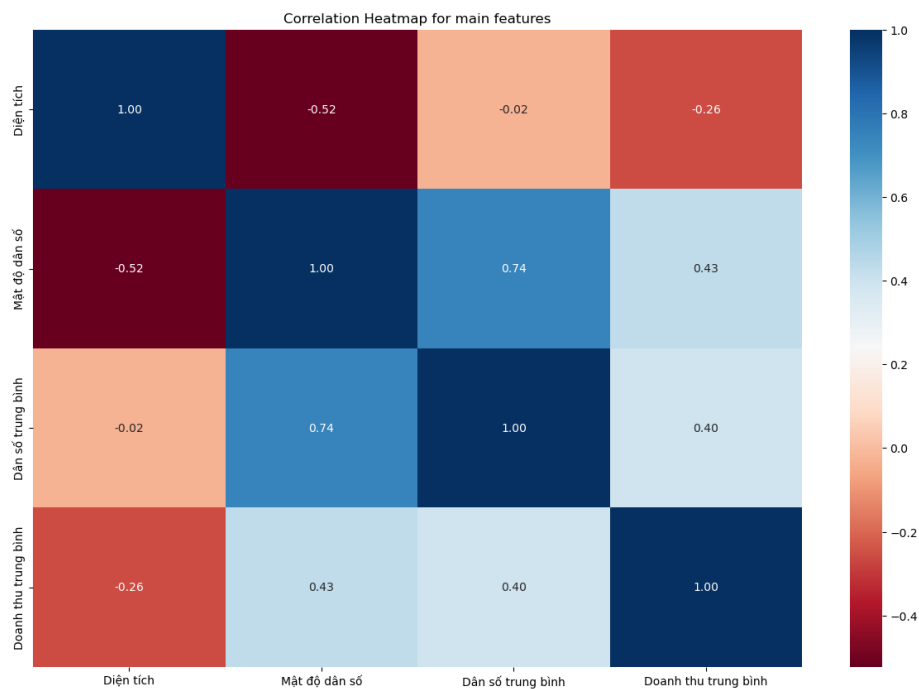
Dữ liệu liên quan về:

| TÊN CỘT           | LOẠI DỮ LIỆU | ĐƠN VỊ                | NHẬN XÉT  |
|-------------------|--------------|-----------------------|---|
| Tỉnh              | Object       |                       | 63 tỉnh thành<br>7 giá trị tổng hợp khác (1 FULL và 6 vùng kinh tế): <ul style="list-style-type: none"><li>- <u>CẢ NƯỚC</u></li><li>- Đồng bằng sông Hồng</li><li>- Đồng bằng sông Cửu Long</li><li>- Trung du và miền núi phía Bắc</li><li>- Bắc Trung Bộ và Duyên hải miền Trung</li><li>- Đông Nam Bộ</li><li>- Tây Nguyên</li></ul> |
| Diện tích         | Float64      | km <sup>2</sup>       |   |
| Mật độ dân số     | Float64      | người/km <sup>2</sup> |   |
| Năm               | Int64        |                       | 28 năm (1995 - 2022)  |
| Dân số trung bình | Float64      | nghìn người           | Tạo từ dân số <b>Nam + Nữ</b><br>Tạo từ dân số <b>Thành thị + Nông thôn</b>   |
| Nam               | Float64      | nghìn người           |   |
| Nữ                | Float64      | nghìn người           |   |
| Thành thị         | Float64      | nghìn người           |   |
| Nông thôn         | Float64      | nghìn người           |   |

|                                       |       |                                 |                          |
|---------------------------------------|-------|---------------------------------|--------------------------|
| Doanh thu trung bình                  | Int64 | nghìn VND/người tính theo tháng | Tạo từ 4 dữ liệu còn lại |
| Thu từ tiền lương, tiền công          | Int64 | nghìn VND/người tính theo tháng |                          |
| Thu từ nông, lâm nghiệp, thủy sản     | Int64 | nghìn VND/người tính theo tháng |                          |
| Thu từ phi nông, lâm nghiệp, thủy sản | Int64 | nghìn VND/người tính theo tháng |                          |
| Các khoản thu khác                    | Int64 | nghìn VND/người tính theo tháng |                          |

Phân tích sơ bộ về dữ liệu:

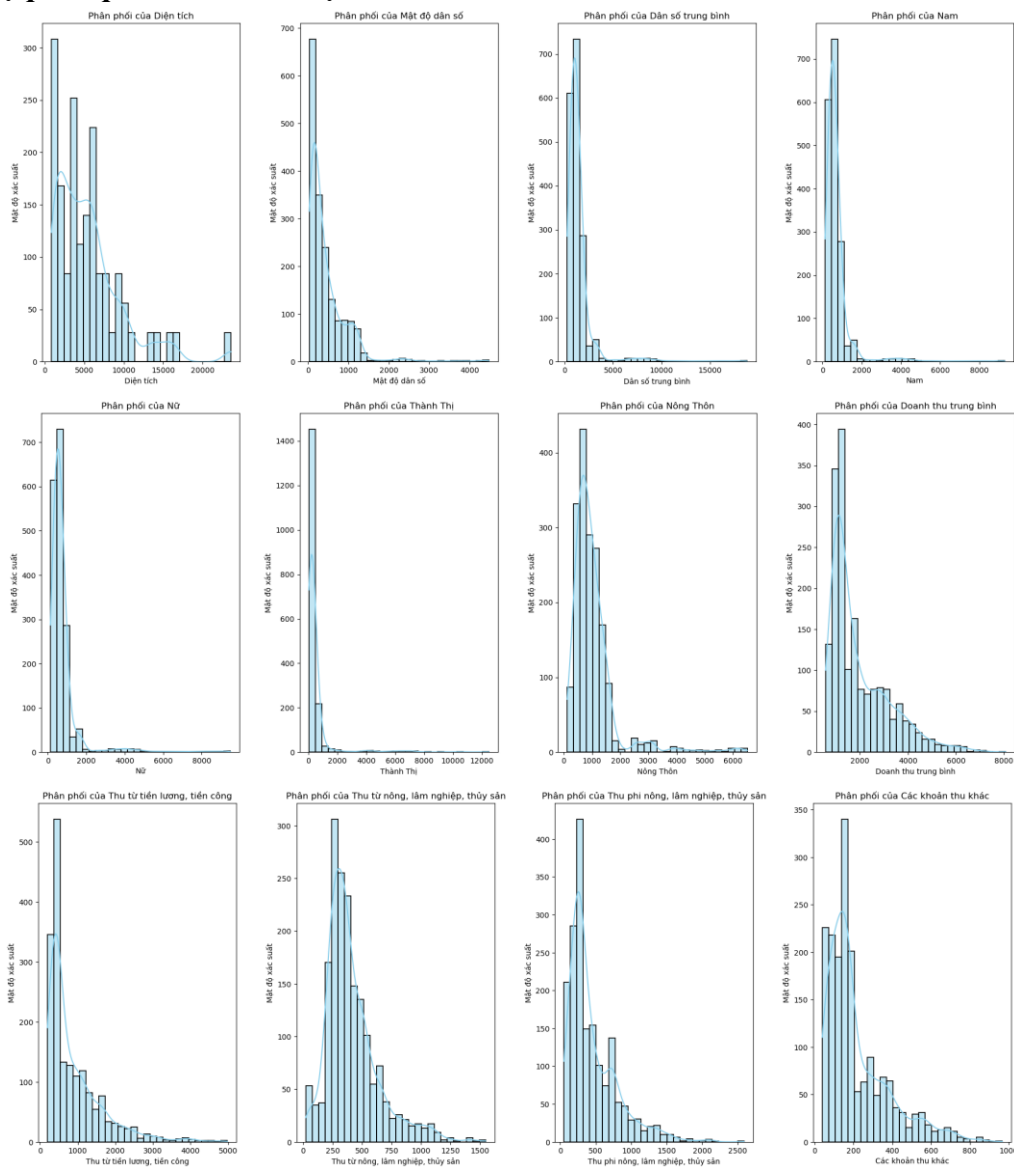
- **Mức độ tương quan của các thuộc tính chính:**



Các biến được phân tích bao gồm dân số trung bình, mật độ dân số, diện tích và doanh thu trung bình. Kết quả cho thấy mối tương quan dương mạnh giữa dân số trung bình và mật độ dân số, cũng như giữa dân số trung bình và doanh thu trung bình. Trong khi đó, diện tích có mối tương quan âm yếu với mật độ dân số và doanh thu trung bình.

Mật độ dân số cũng có tương quan âm yếu với doanh thu trung bình. Kết quả này cho thấy mối quan hệ phức tạp giữa các yếu tố dân số và kinh tế, với sự ảnh hưởng của mật độ dân số lên các chỉ số khác nhau.

- ***Sự phân phối của các thuộc tính:***



**Biểu đồ Phân phối Diện tích:**

- Biểu đồ có hình dạng lệch phải.
- Phần lớn tỉnh/thành có diện tích dưới 10,000 km<sup>2</sup>.
- Một số ít tỉnh/thành có các giá trị ngoại lệ bao gồm Nghệ An (gần 16,500 km<sup>2</sup>), Gia Lai (xấp xỉ 15,500 km<sup>2</sup>) và Sơn La (hơn 14,000 km<sup>2</sup>).

**Biểu đồ Phân phối Mật độ dân số:**

- Biểu đồ có hình dạng lệch phải.
- Phần lớn tỉnh/thành có mật độ dân số không quá 1,000 người/km<sup>2</sup>.
- Cụ thể, biểu đồ có 1 đỉnh chính ở phía ngoài cùng bên trái. Điều này cho thấy mật độ dân số phổ biến nhất là dưới 200 người/km<sup>2</sup>.
- Một số ít tỉnh/thành có các giá trị ngoại lệ bao gồm Tp.HCM (gần 4,500 người/km<sup>2</sup>) và Hà Nội (trên 2,500 người/km<sup>2</sup>). Điều này phản ánh hiện tượng “đất chật người đông” tại một số ít các thành phố lớn trên cả nước.

**Biểu đồ Phân phối Dân số trung bình:**

- Biểu đồ có hình dạng lệch phải, gần như tương đồng với Biểu đồ Phân phối Giới tính Nam và Biểu đồ Phân phối Giới tính Nữ. Điều này thể hiện sự cân bằng về giới tính trên phạm vi cả nước.
- Phần lớn tỉnh/thành có dân số trung bình từ 800,000 đến dưới 2,000,000 người.
- Một số ít tỉnh/thành có các giá trị ngoại lệ bao gồm Tp.HCM (xấp xỉ 9,000,000 người), Hà Nội (xấp xỉ 8,000,000 người) và Thanh Hoá (xấp xỉ 3,600,000 người).

**Biểu đồ Phân phối Giới tính Nam:**

- Biểu đồ có hình dạng lệch phải, gần như tương đồng với Biểu đồ Phân phối Dân số trung bình và Biểu đồ Phân phối Giới tính Nữ, từ đó cho thấy sự cân bằng về giới tính trên phạm vi cả nước.
- Phần lớn tỉnh/thành có từ 400,000 đến dưới 900,000 nam giới, ngoại trừ một số tỉnh/thành đông dân như Tp.HCM, Hà Nội, Thanh Hoá, ...

**Biểu đồ Phân phối Giới tính Nữ:**

- Biểu đồ có hình dạng lệch phải, gần như tương đồng với Biểu đồ Phân phối Dân số trung bình và Biểu đồ Phân phối Giới tính Nam, từ đó cho thấy sự cân bằng về giới tính trên phạm vi cả nước.

- Phần lớn tỉnh/thành có từ 400,000 đến dưới 900,000 nữ giới, ngoại trừ một số tỉnh/thành đông dân như Tp.HCM, Hà Nội, Thanh Hoá, ...

#### **Biểu đồ Phân phối của Thành thị:**

- Biểu đồ có hình dạng lệch phải.
- Số dân thành thị nằm phần lớn phía bên trái biểu đồ, rơi vào khoảng dưới 1,000,000 người.
- Cụ thể, biểu đồ có 1 đỉnh chính nằm phía ngoài cùng bên trái. Điều này cho thấy số dân thành thị phổ biến nhất là dưới 250,000 người.
- Tuy nhiên, một vài trường hợp ngoại lệ là các đô thị, thành phố lớn có từ 3,000,000 đến 7,000,000 dân số thành thị.
- Điều này phản ánh thực tế rằng tỉ lệ nông thôn vẫn áp đảo so với tỉ lệ thành thị, ngoại trừ một số đô thị lớn như Tp.HCM, Đà Nẵng, ...

#### **Biểu đồ Phân phối của Nông thôn:**

- Biểu đồ có hình dạng lệch phải.
- Số dân nông thôn chủ yếu rơi vào khoảng từ 600,000 đến 1,400,000 người.
- Điều này một lần nữa khẳng định tỉ lệ nông thôn vẫn lớn hơn nhiều so với tỉ lệ thành thị trên phạm vi cả nước.

#### **Biểu đồ Phân phối Doanh thu trung bình:**

- Biểu đồ có hình dạng lệch phải.
- Doanh thu trung bình nằm phần lớn ở phía bên trái biểu đồ, chủ yếu rơi vào khoảng dưới 5,000,000 VND/tháng.
- Cụ thể, biểu đồ có 1 đỉnh chính, cho biết doanh thu trung bình phổ biến nhất là xung quanh mức 1,000,000 VND/tháng.
- Một số trường hợp ngoại lệ có doanh thu trung bình trong khoảng từ 6,000,000 đến 8,000,000 VND/tháng. Tuy nhiên, tỉ lệ này chiếm rất ít trên tổng dân số. Nhìn chung, biểu đồ thể hiện xu hướng doanh thu trung bình càng cao thì chiếm tỉ lệ càng thấp.

#### **Biểu đồ Phân phối Doanh thu từ Tiền lương, Tiền công:**

- Biểu đồ có hình dạng lệch phải, gần như tương đồng với Biểu đồ Phân phối Doanh thu trung bình.
- Doanh thu từ tiền lương, tiền công nằm phần lớn ở phía bên trái biểu đồ, chủ yếu rơi vào khoảng dưới 3,000,000 VND/tháng.

- Một số trường hợp ngoại lệ có doanh thu từ tiền lương, tiền công trong khoảng từ 3,000,000 đến 5,000,000 VND/tháng. Tuy nhiên, tỉ lệ này chiếm rất ít và có xu hướng tiền lương, tiền công càng cao thì chiếm tỉ lệ càng thấp.

**Biểu đồ Phân phối Doanh thu từ Nông, Lâm, Thủy sản:**

- Biểu đồ có hình dạng lệch phải.
- Doanh thu từ Nông, Lâm, Thủy sản chủ yếu rơi vào khoảng từ 200,000 đến 600,000 VND/tháng.
- Một số trường hợp ngoại lệ có doanh thu từ Nông, Lâm, Thủy sản lên tới 1,000,000 đến 1,500,000/tháng. Tuy nhiên tỉ lệ này chiếm không nhiều.
- Nhìn chung, doanh thu từ Nông, Lâm, Thủy sản là thấp hơn so với doanh thu từ tiền công, tiền lương.

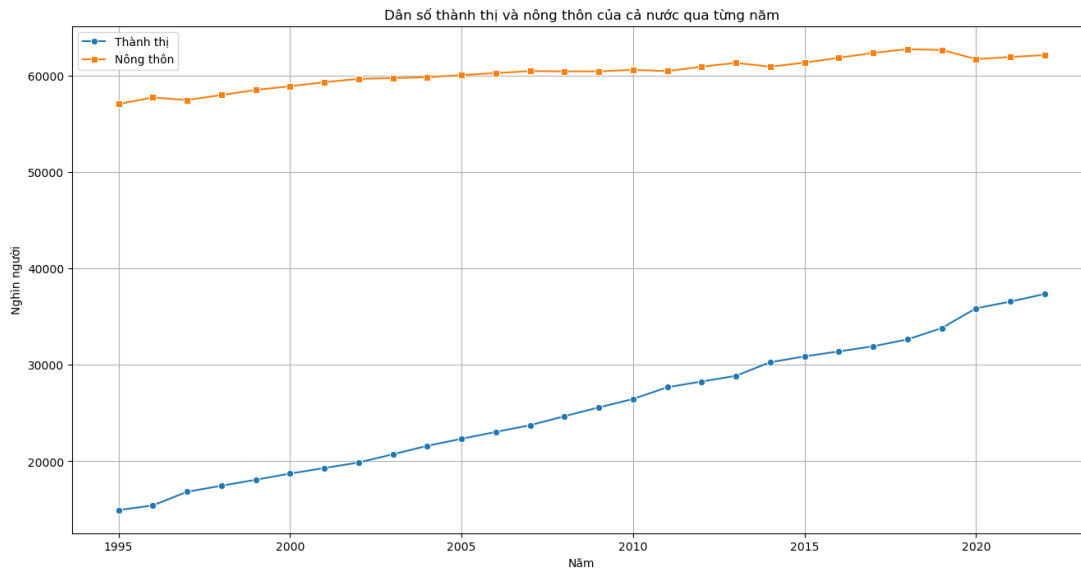
**Biểu đồ Phân phối Doanh thu từ Phi Nông, Lâm, Thủy sản:**

- Biểu đồ có hình dạng lệch phải.
- Doanh thu từ các hoạt động Phi Nông, Lâm, Thủy sản chủ yếu rơi vào khoảng từ 200,000 đến dưới 1,000,000 VND/tháng.
- Tuy nhiên, một vài trường hợp ngoại lệ có doanh thu Phi Nông, Lâm, Thủy sản lên tới từ 1,500,000 đến 2,500,000 VND/tháng.
- Như vậy, doanh thu từ các hoạt động Phi Nông, Lâm, Thủy sản vẫn thấp hơn so với doanh thu từ tiền công và tiền lương.
- Tuy nhiên nếu xét các trường hợp ngoại lệ thì các hoạt động Phi Nông, Lâm, Thủy sản lại có doanh thu cao hơn so với doanh thu từ Nông, Lâm, Thủy sản.

**Biểu đồ Phân phối của các khoản thu khác:**

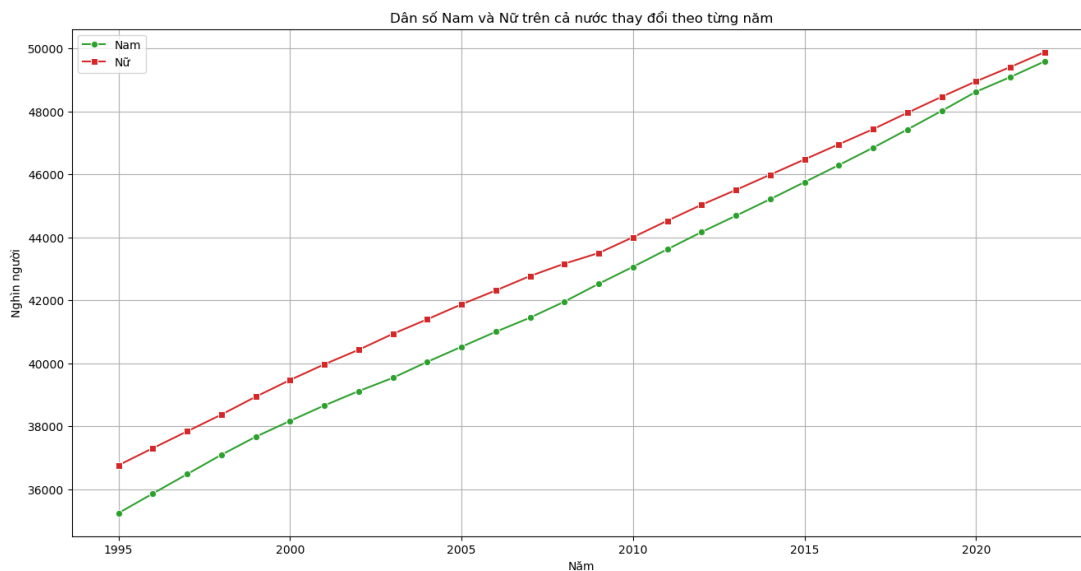
- Biểu đồ có hình dạng lệch phải.
- Doanh thu từ các khoản thu khác nằm phần lớn bên trái biểu đồ, chủ yếu rơi vào khoảng dưới 500,000 VND/tháng.
- Một số trường hợp ngoại lệ có thể kiếm từ 600,000 đến 1,000,000 VND/tháng từ các khoản thu khác, tuy nhiên, tỉ lệ này không nhiều.
- Nhìn chung, doanh thu từ các khoản thu khác là thấp nhất trong tổng doanh thu.

- ***Sự biến động của các giá trị thuộc tính qua từng năm:***



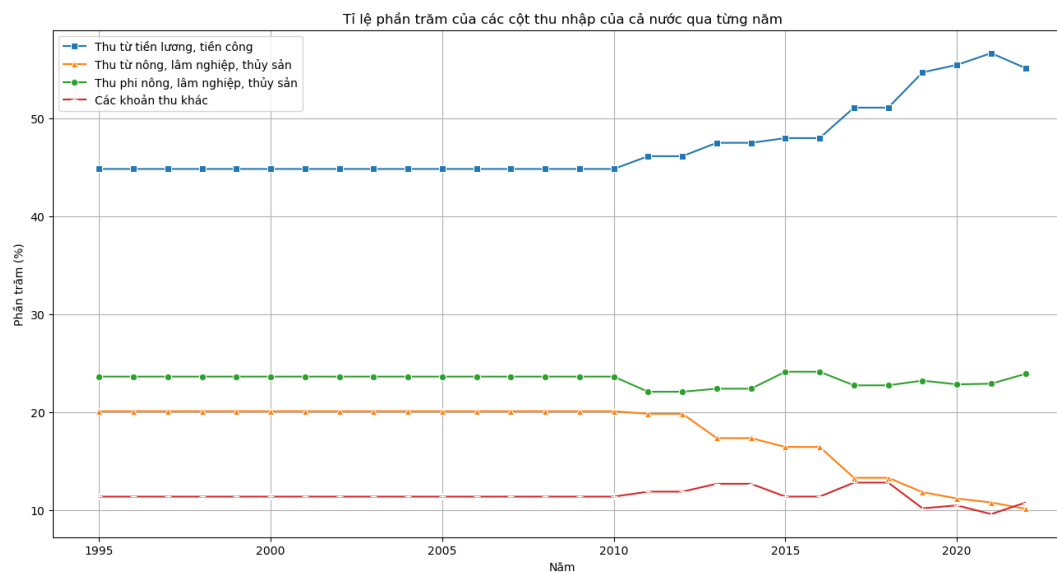
Chúng ta có thể dễ dàng nhận thấy rằng dân số Việt Nam đa phần là ở nông thôn, nhưng trong những năm gần đây dân số Việt Nam có xu hướng chuyển đến các vùng thành thị và đặc biệt tăng mạnh qua các năm từ 2000 **khoảng 19 nghìn người** đến năm 2022 thì con số này dường như đạt gấp đôi khoảng **38 nghìn người**.

Trong khi đó dân cư nông thôn, cũng thể hiện xu hướng phát triển mình như không thể bằng được so với dân cư ở thành thị, vào các năm 2003 đến năm 2011 dân cư nông thôn hầu như không có sự thay đổi đáng kể cho đến năm 2019 thì đạt đỉnh với hơn **63 nghìn người**. Sau đó giảm dân qua các năm còn lại.



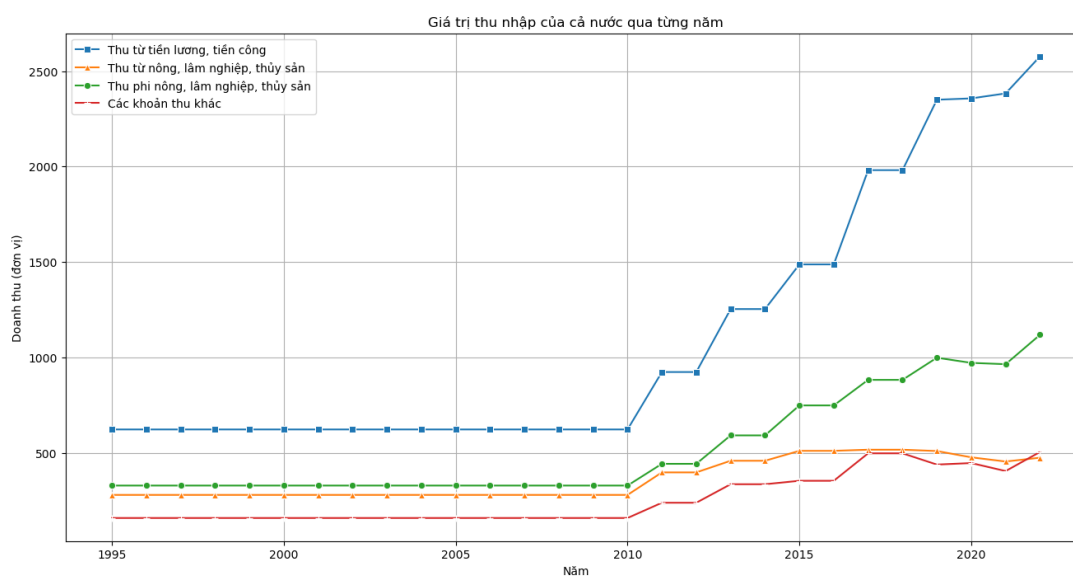
Tỷ lệ nam và nữ ở Việt Nam qua các năm từ 1995 đến 2022 đều tăng theo mỗi quan hệ tuyến tính và đạt giá trị lớn nhất khoảng hơn **49 nghìn người** đối với nam và **50 nghìn người** đối với nữ.

Bên cạnh đó, số lượng nam và nữ đang có xu hướng cân bằng khi trong các năm từ 1995 đến 2010 thì số lượng nữ luôn hơn nam 1 khoảng gần **10 nghìn người**.



Các nguồn thu nhập khác đang dần hạ tỉ trọng của mình xuống và có xu hướng chuyển thành nguồn thu nhập tiền công, tiền lương qua các năm.

Lí giải cho việc thu nhập từ tiền lương, tiền công luôn tăng qua các năm và các nguồn thu nhập còn lại đang dần chuyển hướng thành thu nhập tiền công, tiền lương.





Có thể thấy doanh thu đến từ tiền lương và tiền công luôn đứng đầu so với các nguồn thu nhập khác qua các năm và tăng mạnh từ năm 2010 đến năm 2022 từ **khoảng 600** đến **hơn 2500**. Cũng giống như thu nhập từ lương và tiền công, thu nhập từ thu phí nông, lâm nghiệp, thủy sản tăng mạnh từ năm 2010 **khoảng 480 đến khoảng 1300** năm 2022.

Trong khi đó ta có thể thấy rằng, các nguồn thu nhập từ nông, lâm nghiệp, thủy sản và các khoản thu khác đều có sự dao động mặc dù trước đó đã có một khoảng thời gian 2 nguồn thu nhập này rất ổn định.

## 7. Xây dựng:

Dashboard của nhóm được xây dựng dựa trên cấu trúc sau:

**Dropbox chính:** điều khiển tổng thể (**áp dụng cho tất cả**):

- **Dropbox cho mùa màu:**
  - + Bình thường,
  - + Mùa màu đỏ - xanh lá,
  - + Mùa màu vàng - xanh dương

- **Dropbox theo năm:**

**1 Map và 3 Biểu đồ tròn:**

- **Điều khiển:**
  - + **Dropbox 1:** Diện tích, Dân số, Doanh thu.
  - + **Dropbox 2:** Chọn giữa CẢ NƯỚC, Đồng bằng sông Hồng, Đồng bằng sông Cửu Long, Trung du và miền núi phía Bắc. Bắc Trung Bộ và Duyên hải miền Trung, Đông Nam Bộ, Tây Nguyên.
- **Biểu đồ tròn:**
  - + **Biểu đồ tròn 1:** Cho Nam và Nữ.
  - + **Biểu đồ tròn 2:** Cho Nông thôn và Thành thị.
  - + **Biểu đồ tròn 3:** Cho các khoản thu.

**Biểu đồ so sánh** (dùng biểu đồ cột chồng, hoặc biểu đồ cột liền cho các thông số phụ):

- **Checkbox so sánh:** Chọn các tỉnh để lấy ra so sánh.
- **Dropbox so sánh:** Chọn các mảng (Diện tích, Nam - nữ, Thành thị - Nông thôn, Doanh thu) để so sánh.

**Trang lấy màu:** [ImageColorPicker](#)

### III. INSIGHTS:

#### 1. Xét theo tiêu chí: Nguồn dữ liệu đáng tin cậy

**Nguồn dữ liệu:** Được lấy từ trang [Tổng Cục Thống Kê](#).

Tổng Cục Thống Kê là cơ quan trực thuộc **Bộ Kế hoạch và Đầu tư**, thực hiện chức năng tham mưu, giúp **Bộ trưởng Bộ Kế hoạch và Đầu tư** quản lý nhà nước về **thống kê**, điều phối, tổ chức các hoạt động thống kê và cung cấp **thông tin** thống kê **kinh tế - xã hội** cho các cơ quan, tổ chức và cá nhân theo quy định của pháp luật.

Như vậy, chúng ta có thể kết luận rằng dữ liệu là hoàn toàn đáng tin cậy.

#### 2. Xét theo tiêu chí: Phù hợp với mục đích

Ta xem xét đối tượng cần được trực quan và đánh giá sự phù hợp của biểu đồ được lựa chọn.

**Đối tượng:** Phân bố dân cư

**Loại biểu đồ:** Bản đồ

**Đánh giá:**

- Biểu đồ bản đồ biểu diễn chính xác các thông tin về địa lý, cho phép người dùng **định vị** và **tìm kiếm** tỉnh/thành một cách dễ dàng, nhanh chóng.
- Độ đậm nhạt của màu sắc ứng với mật độ dân số của từng khu vực, cho phép người dùng dễ dàng **so sánh**, đối chiếu các thông tin về dân số của từng tỉnh/thành với nhau.

**Đối tượng:** Tỷ lệ giới tính

**Loại biểu đồ:** Biểu đồ tròn

**Đánh giá:**

- Dữ liệu tỷ lệ giới tính là rời rạc và chỉ có **2 giá trị** (nam/nữ).
- Biểu đồ tròn được lựa chọn vì cho phép **so sánh** tỷ lệ giữa nam và nữ một cách dễ dàng và trực quan.

**Đối tượng:** Tỷ lệ nông thôn và thành thị

**Loại biểu đồ:** Biểu đồ tròn

**Đánh giá:**

- Tương tự như tỷ lệ giới tính, dữ liệu về tỷ lệ nông thôn và thành thị là rời rạc và chỉ có **2** giá trị (nông thôn/thành thị).
- Do đó, biểu đồ tròn được lựa chọn vì cho phép **so sánh** tỷ lệ giữa nông thôn và thành thị một cách dễ dàng và trực quan.

**Đối tượng:** Tỷ lệ các khoản thu nhập

**Loại biểu đồ:** Biểu đồ tròn

**Đánh giá:**

- Dữ liệu về tỷ lệ các khoản thu nhập là rời rạc và có **4** giá trị:
  - + Thu từ tiền lương, tiền công
  - + Thu từ nông, lâm nghiệp, thủy sản
  - + Thu phi nông, lâm nghiệp, thủy sản
  - + Các khoản thu khác
- Biểu đồ tròn được lựa chọn vì cho phép **so sánh** tỷ lệ giữa các khoản thu nhập một cách dễ dàng và trực quan.
- Tuy nhiên, biểu đồ cột cũng là một lựa chọn tốt để trực quan đối tượng này.

**Đối tượng:** Doanh thu các khu vực qua các năm

**Loại biểu đồ:** Biểu đồ đường

**Đánh giá:**

- Dữ liệu về doanh thu có sự thay đổi theo thời gian và được chia theo **6** khu vực:
  - + Bắc Trung Bộ và Duyên hải miền Trung
  - + Đồng bằng sông Cửu Long
  - + Đồng bằng sông Hồng
  - + Đông Nam Bộ
  - + Tây Nguyên
  - + Trung du và miền núi phía Bắc
- Biểu đồ đường cho phép quan sát xu hướng tăng/giảm về doanh thu của một khu vực bất kỳ theo thời gian, đồng thời so sánh doanh thu của các khu vực với nhau.

### **3. Xét theo tiêu chí: Rõ ràng và dễ hiểu**

Chúng ta xem xét một số khía cạnh liên quan đến thiết kế.

**Bố cục của bảng điều khiển:** Bao gồm **5** biểu đồ, được phân chia rõ ràng bằng các đường viền xám và khoảng trắng. Trong đó, cả 5 biểu đồ đều được lựa chọn phù hợp với dữ liệu cần trực quan, có tiêu đề và chú thích rõ ràng.

**Màu sắc:**

- Biểu đồ bản đồ sử dụng chung 1 gam màu, với độ đậm nhạt khác nhau thể hiện mật độ dân số dày đặc/thưa thớt của từng tỉnh/thành.
- Biểu đồ tròn và biểu đồ đường sử dụng các gam màu tương phản để làm nổi bật sự khác nhau giữa các thành phần đối tượng, từ đó tối ưu hoá hiệu quả so sánh.

**Kích thước:** Kích thước của mỗi biểu đồ được căn chỉnh phù hợp với hàm lượng thông tin mà biểu đồ cung cấp.

- Biểu đồ bản đồ chứa thông tin về dân số của **63** tỉnh/thành trên cả nước, do đó có kích thước lớn nhất.
- Biểu đồ đường cung cấp thông tin về doanh thu của **6** khu vực lớn trên cả nước qua các năm, do đó có kích thước lớn nhì.
- Cuối cùng là 3 biểu đồ tròn cung cấp các thông tin về tỉ lệ giới tính, tỉ lệ nông thôn/thành thị và tỉ lệ các khoản thu nhập. Cả 3 biểu đồ này có kích thước bằng nhau.
- Như vậy, các chi tiết liên quan tới **bố cục, màu sắc, kích thước** được lựa chọn và kết hợp với nhau nhằm tối ưu hóa công năng của bảng điều khiển, làm nổi bật thông điệp và cho phép người dùng có thể sử dụng một cách nhanh chóng, dễ dàng.

**4. Xét theo tiêu chí: Sự tích hợp và liên kết**

Bảng điều khiển cung cấp một lượng lớn thông tin về dân cư, giới tính và thu nhập trên phạm vi cả nước, cũng như trên từng khu vực và tỉnh/thành cụ thể.

Tuy nhiên, bảng điều khiển vẫn duy trì được thiết kế rõ ràng, dễ hiểu, không bị quá tải về số lượng biểu đồ hay màu sắc. Điều này có được là do tính liên kết cao giữa các biểu đồ với nhau. Cụ thể:

- Bản đồ vừa thể hiện phân bố dân cư vừa cho phép lựa chọn tỉnh/thành cụ thể để trực quan các thông tin về tỉ lệ giới tính, tỉ lệ nông thôn/thành thị và tỉ lệ các khoản thu nhập.
- Như vậy, cả 3 biểu đồ tròn (biểu đồ tỉ lệ giới tính, biểu đồ tỉ lệ nông thôn/thành thị, biểu đồ tỉ lệ các khoản thu nhập) có mối quan hệ liên kết, phụ thuộc chặt chẽ vào bản đồ.

**5. Xét theo tiêu chí: Phân tích dữ liệu**

**Câu 1: Tương quan giữa doanh thu của các khu vực và mật độ dân số qua các năm?**

**Trả lời:** quan sát biểu đồ đường “Doanh thu các khu vực qua các năm” kết hợp biểu đồ “Phân bố dân cư” và “Tỉ lệ các khoản thu nhập”

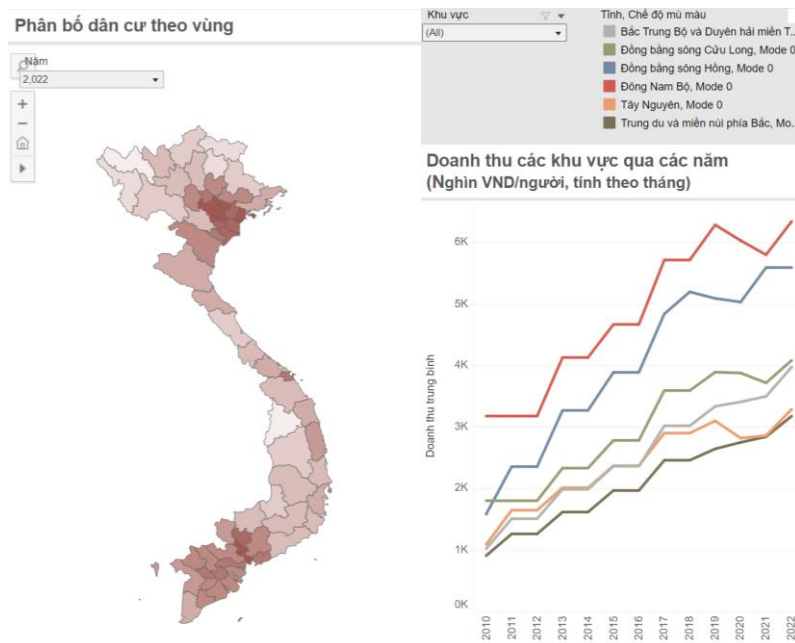


Figure 8: Dữ liệu thể hiện phân bố dân cư và tỉ lệ các khoản thu nhập

### Nhận xét:

Dựa vào biểu đồ ta có thể rút ra được mật độ dân số cả nước có xu hướng tập trung vào các khu vực có doanh thu cao trên cả nước.

Có thể thấy hầu hết dân số tập trung chủ yếu ở 2 khu vực có nền kinh tế trọng điểm của cả nước là ĐBSCL và ĐBSH điển hình là Tp Hồ Chí Minh và thủ đô Hà Nội nơi có mật độ dân số trung bình lần lượt là 4481 và 2511 (nghìn người), doanh thu của 2 tỉnh lần lượt là 6392 và 6423 (nghìn đồng/người theo tháng).

Đa phần các khu vực có doanh thu cao có tỉ lệ tiền lương công cao nhất trong các thành phần còn lại.

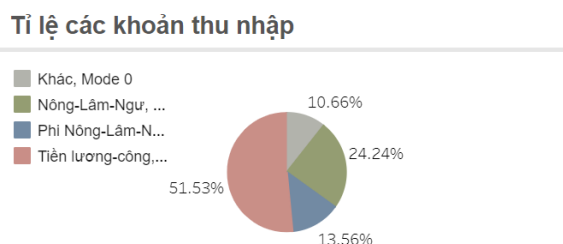


Figure 9: Tỉ lệ các khoản thu nhập

**Mục tiêu của câu hỏi:** Mục tiêu chính của câu hỏi này là hiểu rõ mối liên hệ giữa doanh thu và mật độ dân số, và cung cấp thông tin hữu ích cho việc đưa ra quyết định chiến lược, chính sách

phát triển kinh tế và xã hội cho các khu vực. Một trong những mục tiêu cụ thể có thể là đề xuất các biện pháp phát triển kinh tế và xã hội phù hợp với điều kiện cụ thể của mỗi khu vực dựa trên nhận định về mối tương quan giữa doanh thu và mật độ dân số.

**Câu 2: Doanh thu của các khu vực qua các năm thay đổi như thế nào?**

**Trả lời:** Quan sát biểu đồ “**Doanh thu các khu vực qua các năm**”

**Nhận xét:**

Giai đoạn từ năm 2010-2012: nền kinh tế ĐNB và ĐBSCL hầu như không có sự phát triển và giữ nguyên ở mức 3100 và 1800 (nghìn đồng/người theo tháng).

Tuy nhiên trong giai đoạn này các khu vực còn lại có sự phát triển mạnh về doanh thu từ năm 2010-2011 và giữ nguyên đến năm 2012. Trong giai đoạn 2012-2018 nền kinh tế cả nước có xu hướng phát triển mạnh bởi vì trong giai đoạn này chính phủ đề ra chiến lược phát triển kinh tế - xã hội 10 năm.

Giai đoạn sau đó đến hiện tại doanh thu cả nước có khuynh hướng suy giảm do đại dịch thế kỉ COVID 19 đến năm 2021 mới bắt đầu phục hồi và phát triển trở lại.

**Mục tiêu của câu hỏi:** Xác định xu hướng tăng trưởng hoặc suy giảm của doanh thu của các khu vực qua các năm. Điều này có thể giúp ta nhận ra các ngành công nghiệp, lĩnh vực hoặc khu vực đặc biệt nổi bật trong việc tạo ra doanh thu.

**Đánh giá yếu tố ảnh hưởng:** Xác định các yếu tố ảnh hưởng đến sự biến động trong doanh thu của các khu vực. Các yếu tố này có thể bao gồm sự phát triển kinh tế, chính sách công, biến động thị trường, và các yếu tố về môi trường kinh doanh.

**Dự báo và kế hoạch chiến lược:** Dựa trên phân tích của thay đổi trong doanh thu, ta có thể dự báo xu hướng tương lai và phát triển các kế hoạch chiến lược phù hợp. Điều này giúp các nhà quản lý và chính phủ có thông tin cần thiết để đưa ra các quyết định chiến lược và chính sách phát triển.

**Tối ưu hóa tài nguyên:** Hiểu rõ sự biến động trong doanh thu giúp các doanh nghiệp và tổ chức tối ưu hóa việc sử dụng tài nguyên, đầu tư vào các lĩnh vực có tiềm năng tăng trưởng và thích ứng với những thay đổi trong môi trường kinh doanh.



Figure 10: Biểu đồ biến thiên doanh thu theo các khu vực riêng

Phân tích sự thay đổi và xu hướng theo thời gian:

Quan sát biểu đồ “**Doanh thu các khu vực qua các năm**”, chúng ta nhận thấy:

- Doanh thu của tất cả các khu vực đều gia tăng theo thời gian.
- Cụ thể, sau 12 năm (từ năm 2010 đến năm 2022), doanh thu của tất cả các khu vực gia tăng từ 2 đến 4 lần.
- Trong đó, khu vực Đông Nam Bộ có doanh thu cao nhất.
- Ngược lại, khu vực Trung du và miền núi phía Bắc có doanh thu thấp nhất.

Kết hợp quan sát các biểu đồ “**Phân bố dân cư**” và “**Tỉ lệ các khoản thu nhập**”, ta nhận thấy:

- Tại các thành phố lớn như Hà Nội, Đà Nẵng, TP.HCM, thu nhập đến từ **tiền lương và tiền công** là phổ biến nhất, chiếm từ 55% đến hơn 60% tổng tỉ lệ các khoản thu nhập.



Figure 11: Tỉ lệ các khoản thu nhập tại Hà Nội và TPHCM năm 1995

- Trái lại, các tỉnh/thành như Bắc Ninh, Bà Rịa-Vũng Tàu lại có thu nhập đến từ **nông – lâm – ngư nghiệp** chiếm tỉ lệ cao nhất, cụ thể từ 40% đến 45% trên tổng tỉ lệ thu nhập.



Figure 12: Tỉ lệ các khoản thu nhập tại Bắc Ninh và Bà Rịa – Vũng Tàu năm 1995

- Cuối cùng, các tỉnh/thành như Sơn La, Đắk Nông có thu nhập chủ yếu đến từ các hoạt động **phi nông – lâm – ngư nghiệp**, chiếm từ 50% đến 60% trên tổng tỉ lệ các khoản thu nhập.



Figure 13: Tỉ lệ các khoản thu nhập tại Sơn La và Đắk Nông năm 1995

## Kết luận:

Bảng điều khiển cho phép quan sát, phân tích xu hướng về cả mặt **địa lý** và **thời gian**. Trong đó:

- Các thay đổi về mặt thời gian có thể quan sát thông qua biểu đồ đường.
- Các thay đổi về mặt địa lý có thể quan sát thông qua biểu đồ bản đồ.

Hơn nữa, bảng điều khiển cung cấp các thông tin về dân số, giới tính, thu nhập, v.v. **từ tổng quát đến chi tiết**, thông qua việc tích hợp, liên kết các biểu đồ lại với nhau. Cụ thể:

- Ở chế độ mặc định, bảng điều khiển thể hiện thông tin về tỉ lệ giới tính, tỉ lệ nông thôn/thành thị, và tỉ lệ các khoản thu nhập trên phạm vi cả nước.
- Khi người dùng nhấn chọn 1 tỉnh/thành bất kì trên bản đồ, bảng điều khiển lập tức thể hiện thông tin về tỉ lệ giới tính, tỉ lệ nông thôn/thành thị, và tỉ lệ các khoản thu nhập của riêng tỉnh/thành đó.



## 6. Xét theo tiêu chí: Tương tác và điều hướng

Bảng điều khiển cho phép tương tác thông qua các cử chỉ:

### Tùy chỉnh khu vực địa lý trên bản đồ:

- Người dùng có thể phân tích các thông tin liên quan đến mật độ dân số, giới tính, thu nhập của riêng tỉnh/thành mình đang sinh sống.
- Điều này tạo ra tính cá nhân hóa, giúp người dùng gắn kết với thiết kế và nâng cao trải nghiệm người dùng.

### Tùy chỉnh khu vực để quan sát doanh thu:

- Người dùng có thể lựa chọn khu vực bất kỳ từ checkbox để quan sát doanh thu của khu vực đó theo thời gian.
- Điều này giúp người dùng tập trung quan sát doanh thu của một hay một nhóm các khu vực nhất định, từ đó dễ dàng so sánh, đối chiếu, và phân tích các xu hướng quan trọng.

### Di chuyển và click chuột để highlight thông tin:

- Thao tác này giúp người dùng tập trung vào dữ liệu cụ thể một cách dễ dàng, nhằm lôi cuốn và tăng tính gắn kết với người dùng, khiến trải nghiệm của họ trở nên thú vị hơn.

**Mục tiêu:** Điều hướng sự chú ý của người dùng đi từ những thông tin quan trọng đến những thông tin chi tiết. Cụ thể:

- Trước hết, hướng sự chú ý của người dùng vào biểu đồ “Phân bố dân cư theo vùng”.
- Kế tiếp, hướng chú ý của người dùng vào biểu đồ “Doanh thu các khu vực qua các năm”.
- Cuối cùng, hướng sự chú ý của người dùng vào các biểu đồ tròn **“Tỉ lệ giới tính”, “Tỉ lệ Nông thôn & Thành thị”** và **“Tỉ lệ các khoản thu nhập”**.

### Phương pháp:

- Thể hiện thông qua bố cục của bảng điều khiển:
  - + Phân chia các biểu đồ một cách rõ ràng, tránh gây rối loạn thông tin.
  - + Hai biểu đồ quan trọng chiếm diện tích lớn: **“Phân bố dân cư theo vùng”** và **“Doanh thu các khu vực qua các năm”**.
- Thể hiện thông qua màu sắc và kích thước:
  - + Các biểu đồ chính sử dụng màu sắc nổi bật.
  - + Không sử dụng thêm màu cho văn bản.
  - + Các khoảng trắng giúp cân bằng và làm nổi bật thông tin trong các biểu đồ.
- Thể hiện thông qua kiểu chữ:

- + Tiêu đề dashboard có kích thước lớn nhất.
- + Tiếp theo là tên biểu đồ.
- + Cuối cùng là thông tin chi tiết trong mỗi biểu đồ và kí hiệu.

## 7. Thiết kế hấp dẫn

### Bố cục:

- Bảng điều khiển bao gồm **5** biểu đồ, được phân chia rõ ràng bằng các đường viền và khoảng trắng. Trong đó, cả 5 biểu đồ đều có tiêu đề và chú thích rõ ràng.
- Việc sử dụng khoảng trắng giữa các yếu tố giúp tạo ra một tổng thể sạch sẽ và tôn vinh thông tin trong mỗi biểu đồ.

### Văn bản:

- Sử dụng chung 1 font chữ để đảm bảo tính nhất quán và dễ đọc.
- Màu chữ được lựa chọn để tạo ra sự tương phản với nền và dễ nhìn.
- Kích thước văn bản được điều chỉnh khác nhau giữa tiêu đề, tên biểu đồ, chú thích, v.v. nhằm đảm bảo sự nhất quán và dễ đọc.

### Màu sắc:

- Biểu đồ bản đồ sử dụng chung 1 gam màu, với độ đậm nhạt khác nhau thể hiện mật độ dân số dày đặc/thưa thớt của từng tỉnh/thành.
- Biểu đồ tròn và biểu đồ đường sử dụng các gam màu tương phản để làm nổi bật sự khác nhau giữa các thành phần đối tượng, từ đó tối ưu hoá hiệu quả so sánh.

### Cụ thể, các màu sử dụng:

#### - Bình thường:

- + **Dải màu:** #912c22
- + **4 màu cho thu nhập:** #708aa5 #cc8d86 #939d6f #b2b3ac
- + **2 màu đối nghịch:** #708aa5 #cc8d86
- + **6 màu so sánh khu vực:** #b2b3ac #939d6f #708aa5 #da564d #ee9c6c #787154

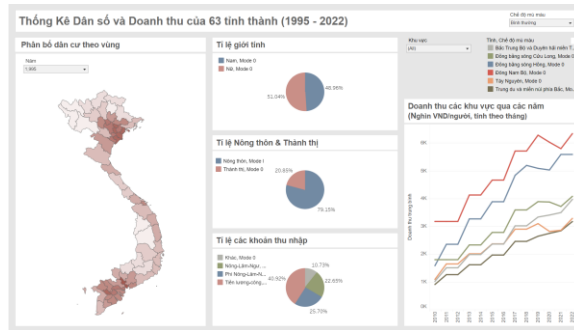


Figure 14: Dashboard ở người bình thường

- **Màu màu đỏ - xanh lá:** khó nhìn màu đỏ và xanh lá, có xu hướng chuyển sang màu vàng.

- + **Dải màu:** #f5251c
- + **4 màu cho thu nhập:** #3cd8fc #f6251c #c4fd40 #3b1e3c
- + **2 màu đối nghịch:** #3cd8fc #f6251c
- + **6 màu so sánh khu vực:** #3b1e3c #4e60fd #3cd8fc #f6251c #f9a655 #004da9

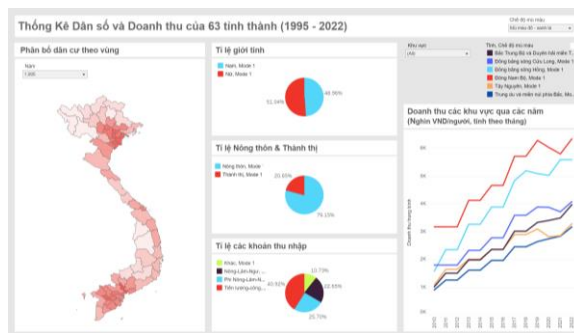


Figure 15: Dashboard màu đỏ - xanh lá khi người bình thường nhìn vào

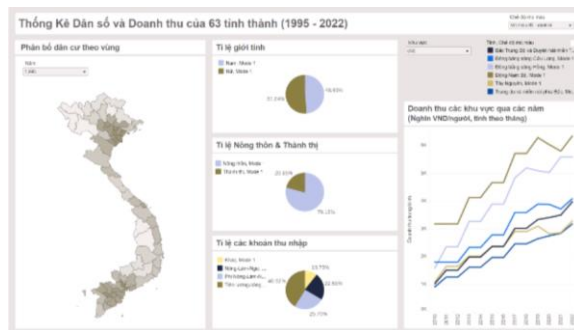


Figure 16: Dashboard màu đỏ - xanh lá khi người mù màu đỏ nhìn vào

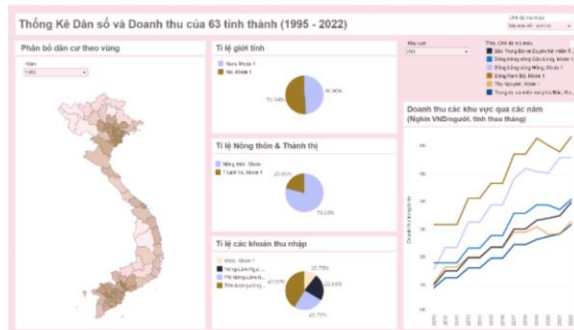


Figure 17: Dashboard mùa màu đỏ - xanh lá khi người mù màu xanh lá nhìn vào

**Màu màu vàng - xanh dương:** nhìn thấy thế giới qua một tông màu hồng hồng xanh xanh.

- + **Dải màu:** #bc1b14
- + **4 màu cho thu nhập:** #c11b17 #0036f8 #660d90 #b5fd3f
- + **2 màu đối nghịch:** #c11b17 #0036f8
- + **6 màu so sánh khu vực:** #c11b17 #0036f8 #660d90 #b5fd3f #febe00 #00ffc3

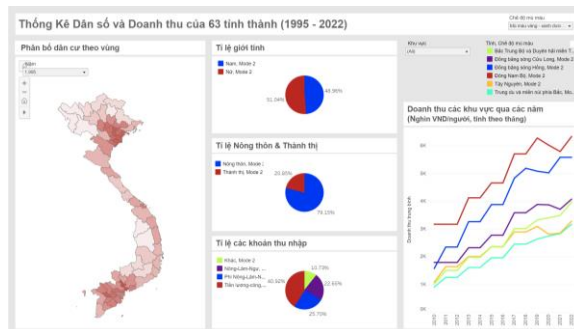


Figure 18: Dashboard mùa màu vàng - xanh dương khi người bình thường nhìn vào

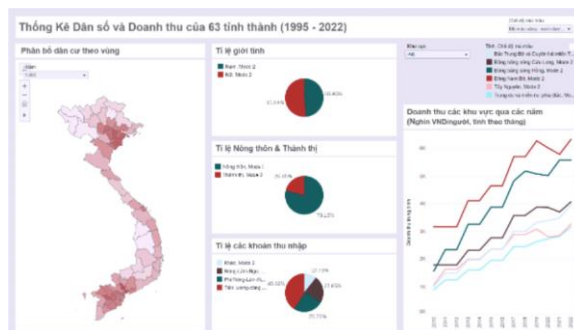


Figure 19: Dashboard mùa màu vàng - xanh dương khi người mù màu xanh dương nhìn vào

Như vậy, các chi tiết thiết kế được lựa chọn phù hợp với mục đích trực quan, giúp bảng điều khiển trở nên vừa sinh động, thu hút, vừa rõ ràng, dễ sử dụng.