



**PHÂN TÍCH VÀ TRỰC QUAN HÓA DỮ LIỆU
TRONG NGÀNH **BẢO HIỂM Y TẾ** TẠI HOA KỲ**

TỔNG QUAN VỀ BỘ DỮ LIỆU

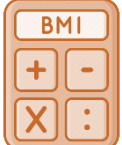
THUỘC TÍNH




AGE
Độ tuổi




SEX
Giới tính




BMI
Chỉ số BMI




SMOKER
Thói quen
Hút thuốc



CHILDREN
Số lượng con cái
Trong gia đình

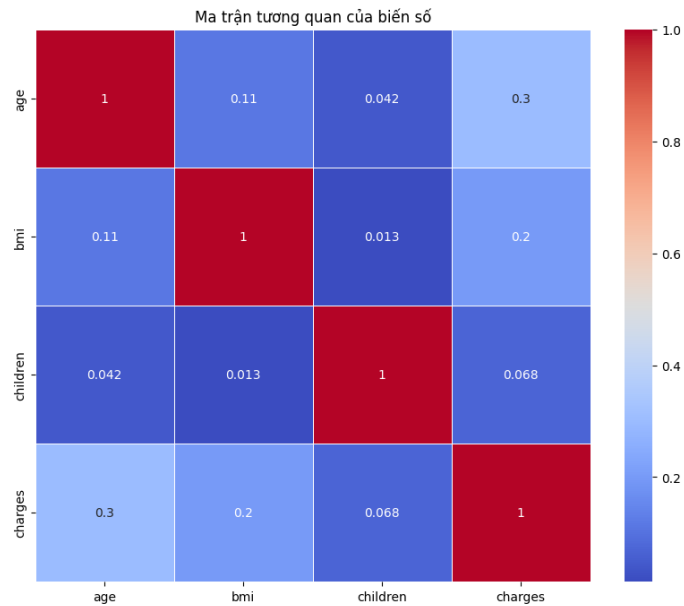


REGION
Khu vực



CHARGES TARGET
Tổng chi phí mà BHYT
Chi trả cho Y tế cá nhân

HẠN CHẾ CỦA BỘ DỮ LIỆU

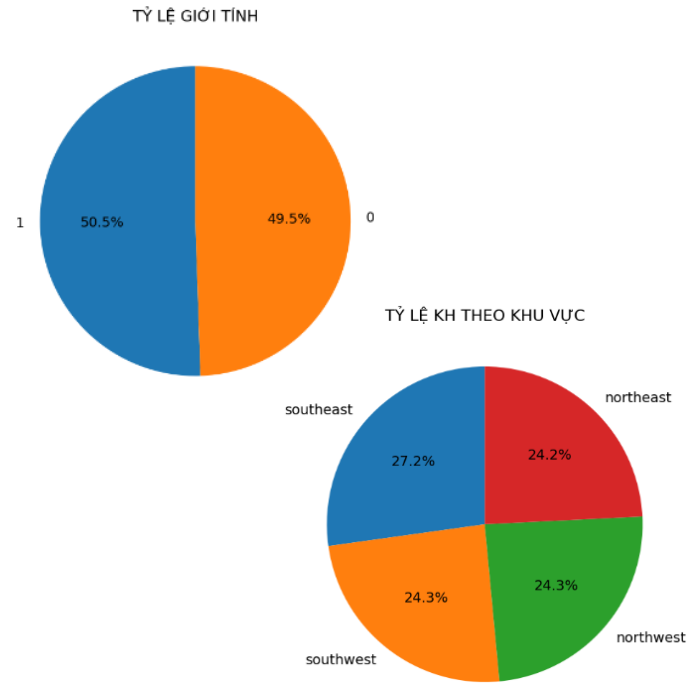


Các biến số có trong bộ dữ liệu ban đầu có độ tương quan cực kỳ thấp với biến mục tiêu, hay thậm chí với các biến số khác

Các biến phân loại đều có số lượng biến gần như là đồng đều nhau, nghĩa là không có một lớp nào nổi trội hơn các lớp khác, điều này có thể gây khó khăn trong việc đánh giá và nắm bắt thông tin.

**KHÓ NHẬN BIẾT SỰ KHÁC BIỆT
ẢNH HƯỞNG ĐẾN ĐỘ TIN CẬY CỦA BIỂU ĐỒ
KHÓ SO SÁNH GIỮA CÁC LỚP**

→ **KHÔNG ĐỦ THÔNG TIN ĐỂ ĐƯA RA NHẬN ĐỊNH
HAY KẾT LUẬN ĐỐI VỚI BIẾN MỤC TIÊU**



THAY ĐỔI Ý NGHĨA CỦA BỘ DỮ LIỆU

Ý NGHĨA BAN ĐẦU



CHARGES

TARGET

Tổng chi phí mà BHYT
Chi trả cho Y tế cá nhân

Ý NGHĨA SAU THAY ĐỔI



CHARGES

TARGET

Tổng chi phí mà Người dân Hoa Kỳ
Chi trả cho BHYT

BỔ SUNG THUỘC TÍNH CHO BỘ DỮ LIỆU

YẾU TỐ XÃ HỘI



STATE

Tiểu bang



MARRIED

Tình trạng
Hôn nhân



EMPLOY

Tình trạng
Việc làm

YẾU TỐ SỨC KHỎE



MOST_RECENT_INSUR_PUR

Thời điểm
Mua BHYT gần nhất

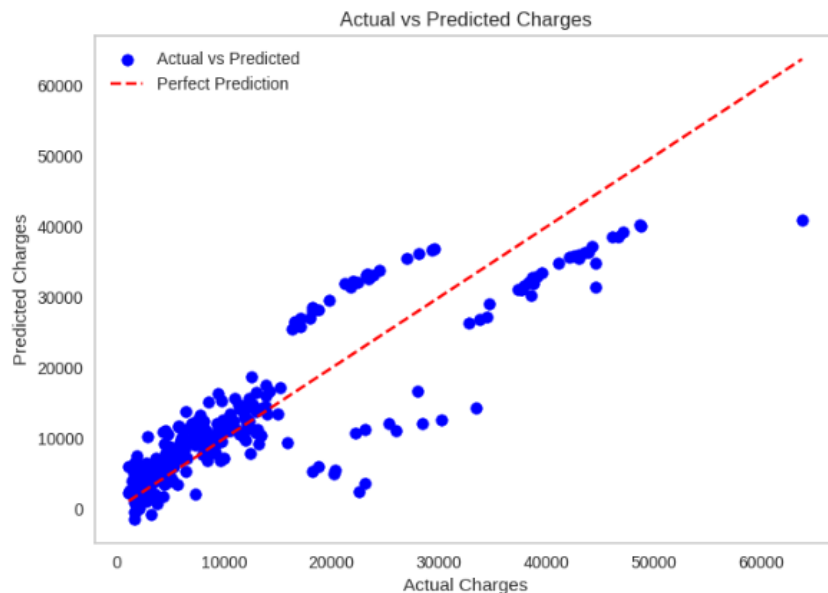


INSUR_RENEWALS

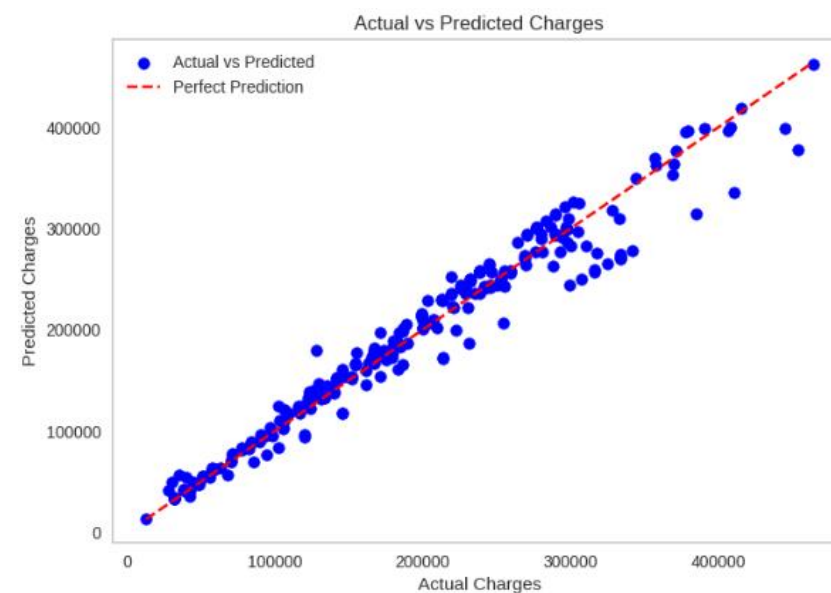
Số lần
Làm mới BHYT

KIỂM TRA HIỆU QUẢ THAY ĐỔI BIẾN

MÔ HÌNH HỒI QUY



BỘ DỮ LIỆU BAN ĐẦU



BỘ DỮ LIỆU SAU KHI THAY ĐỔI

Nhìn từ biểu đồ, bộ dữ liệu mới có sự gia tăng về khả năng giải thích sự biến thiên của biến mục tiêu (*charges*), ngoài ra xuất hiện ít điểm nằm xa đường hoàn hảo hơn với bộ dữ liệu ban đầu, tức sai số nhỏ hơn.

$$R^2 = 0,7811$$



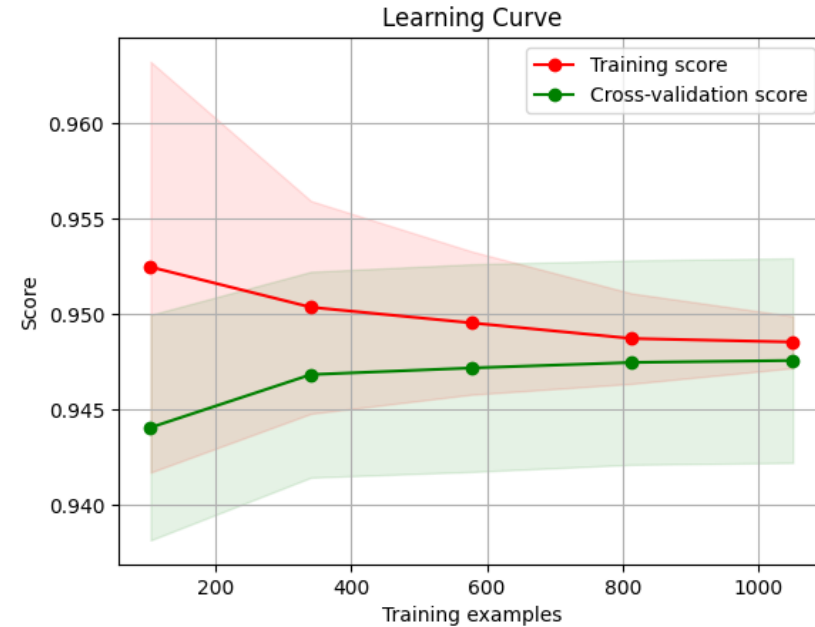
$$R^2 = 0,9599$$

XÉT Ở NHỮNG ĐIỂM GIÁ TRỊ CAO, CÓ SỰ KHÁC BIỆT LỚN GIỮA GIÁ TRỊ THỰC TẾ VÀ GIÁ TRỊ DỰ ĐOÁN. CHO THẤY, MÔ HÌNH HỒI QUY TUYẾN TÍNH KHÔNG THỂ DỰ ĐOÁN CHÍNH XÁC CHI PHÍ BẢO HIỂM CỦA KHÁCH HÀNG TRÊN DỮ LIỆU KIỂM TRA, DO MÔ HÌNH QUÁ PHỨC TẠP HOẶC KHÔNG PHÙ HỢP VỚI BỘ DỮ LIỆU. CHÍNH VÌ VẬY MÀ **BỘ DỮ LIỆU CÓ NGUY CƠ BỊ OVERFITTING.**

KIỂM TRA OVERFITTING

```
Cross-Validation Scores: [0.94708351 0.95151018 0.94562583 0.95274824 0.94818664 0.94675286
0.94420617 0.95280185 0.95007694 0.95001853 0.94866484 0.95365474
0.94402004 0.94866978 0.94817385 0.93872587 0.93522801 0.95125769
0.93867251 0.95035332 0.94961234 0.94551069 0.94517381 0.94003591
0.95628093 0.9487065 0.94941416 0.94650316 0.94429782 0.94820786
0.9503659 0.94168808 0.94164512 0.9470207 0.94459942 0.95154021
0.95354946 0.92777502 0.94250204 0.95644465 0.9532381 0.94348415
0.96087161 0.95142028 0.94336403 0.95216908 0.94540139 0.94379318
0.93785151 0.94791389 0.94501929 0.94692609 0.95634989 0.95250628
0.93840181 0.94585708 0.9493627 0.94863648 0.94769881 0.94988647
0.9567543 0.95188871 0.93828639 0.9454567 0.94638807 0.94942658
0.9452843 0.95361611 0.94567791 0.94268082 0.94993607 0.95348727
0.94205289 0.95542098 0.93774553 0.94254082 0.94212201 0.94524216
0.95273369 0.95128881 0.95402895 0.94629457 0.93810778 0.94448737
0.9416901 0.95133527 0.95012085 0.94981831 0.94849255 0.94990934
0.95152484 0.95019989 0.94791856 0.95647936 0.94718768 0.94525658
0.94315528 0.95314558 0.95285212 0.94727164]
Mean Cross-Validation Score: 0.9475476405319277
```

BẢNG ĐIỂM SỐ CROSS-VALIDATION



BIỂU ĐỒ LEARNING CURVE

Đường biểu diễn sai số huấn luyện (Training Error) và sai số kiểm tra (Validation Error) có xu hướng hội tụ khi kích thước tập huấn luyện (Training Size) tăng lên.

→ **MÔ HÌNH CÓ KHẢ NĂNG KHÁI QUÁT HÓA TỐT TRÊN CẢ TẬP HUẤN LUYỆN VÀ TẬP KIỂM TRA**

Sai số huấn luyện và sai số kiểm tra đều ở mức thấp, khoảng 0,02.

→ **MÔ HÌNH CÓ ĐỘ CHÍNH XÁC CAO, KHÔNG BỊ THIẾU DỮ LIỆU (UNDERFITTING) HAY QUÁ KHỚP DỮ LIỆU (OVERFITTING).**

→ **CỐ THỂ NỔI, MÔ HÌNH CÓ ĐỘ NHẤT QUẢN CAO VÀ KHÔNG BỊ ẢNH HƯỞNG NHIỀU BỞI CÁC YẾU TỐ NHIỀU HAY BIẾN ĐỘNG CỦA DỮ LIỆU.**

TIỀN XỬ LÝ DỮ LIỆU

XÓA GIÁ TRỊ RỖNG

```
Các giá trị bị thiếu:
CusID      0
age        0
is_male    0
married    0
num_of_children  0
bmi        0
smoker     0
state      0
region     0
employ     0
most_recent_insur_pur  0
insur_renewals  18
charges    20
dtype: int64
```

Giá trị rỗng xuất hiện của yếu ở hai cột **insur_renewals** và **charges**, để giải thích cho hiện tượng này chính do nhiều bang ở Hoa Kỳ không đưa ra hạn mức BHYT trung bình mà người dân phải đóng mỗi năm. Chính vì sự thiếu hụt thông tin đó mà nhóm không thể điền vào một số ô ở hai cột trên.

XÓA GIÁ TRỊ BẤT THƯỜNG

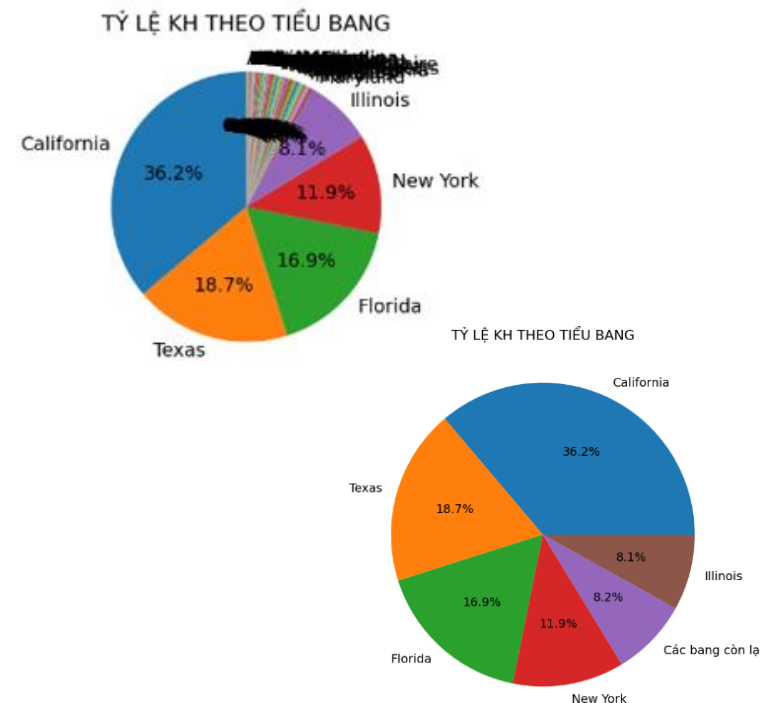
```
CusID most_recent_insur_pur
65    CS476      9/31/2021
97    CS861      6/31/2009
121   CS888      2/31/2017
136   CS669      2/31/2018
157   CS460      2/30/2007
259   CS289      2/31/2020
314   CS844      9/31/2021
353   CS801      2/31/2018
393   CS728      2/30/2017
453   CS260      4/31/2018
462   CS746      2/31/2014
464   CS994      2/31/2008
477   CS401      4/31/2008
621   CS865      2/30/2018

680   CS729      2/30/2020
936   CS858      4/31/2018
971   CS927      11/31/2022
997   CS995      2/29/2021
1011  CS315      2/29/2022
1016  CS445      6/31/2022
1059  CS136      9/31/2018
1070  CS282      2/31/2022
1101  CS430      2/31/2021
1105  CS949      2/31/2021
1169  CS386      11/31/2021
1203  CS456      2/31/2022

Số dòng bị xóa: 26
Số dòng còn lại: 1313
```

Trong quá trình khởi tạo giá trị cho cột này, nhóm đã sử dụng hàm **RANDBETWEEN()** của Excel, chính vì thế mà không thể tránh khỏi việc xuất hiện của những biến thời gian bất thường như **Tháng 2 có ngày 30** hay **Tháng 4 có ngày 31**. Chính vì vậy, nhóm cần thực hiện xóa những dòng dữ liệu không phù hợp như vậy ra khỏi bộ dữ liệu.

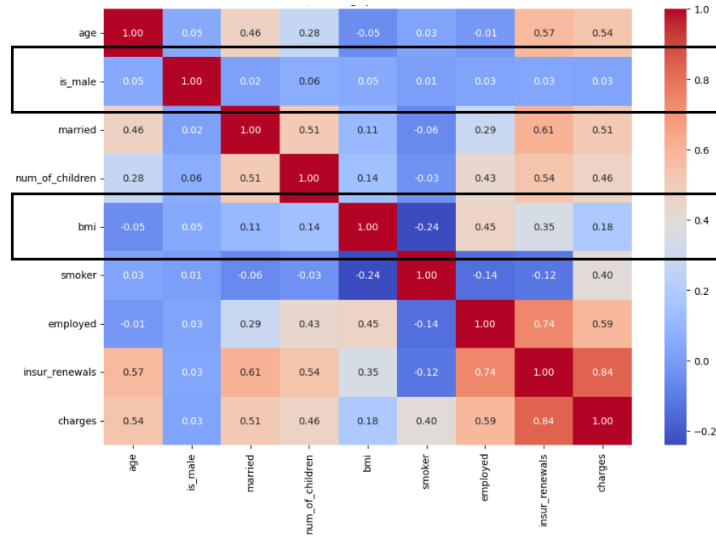
GOM NHÓM DỮ LIỆU



Số liệu chỉ tập trung phần đa vào 5 bang trung tâm của Hoa Kỳ bao gồm: Florida, California, Texas, New York và Illinois, các bang còn lại vì số liệu quá ít khiến cho biểu đồ bị **"đen đặc"** lại.

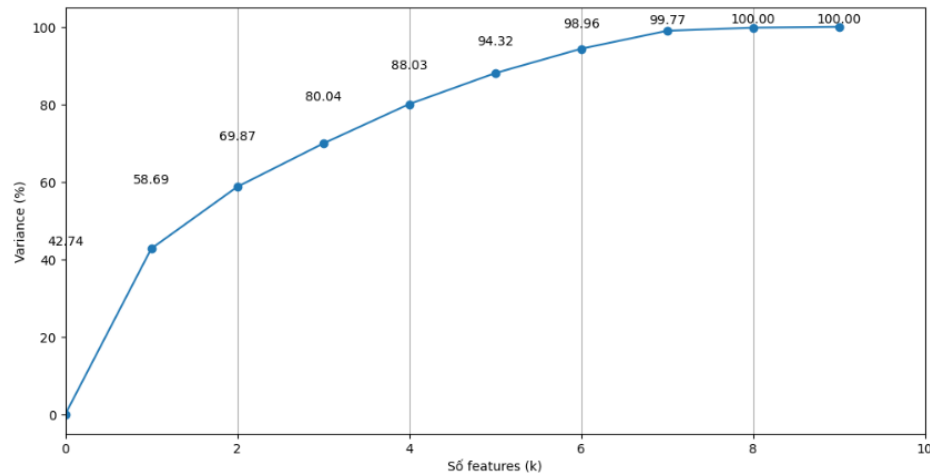
Để xử lý, nhóm thực hiện gom nhóm các tiểu bang khác ngoài các bang trung tâm thành một cụm và đặt tên cho giá trị là **"Các bang còn lại"**

GIẢM CHIỀU DỮ LIỆU



Nhìn từ ma trận tương quan, nhóm nhận thấy có một số biến thuộc tính có độ tương quan thấp với biến mục tiêu mà nhóm nghiên cứu (Ví dụ: **is_male**, **bmi** hay thậm chí là biến **smoker** chỉ có độ tương quan cao duy nhất đối với biến mục tiêu là **charges**, còn ở các biến khác thì rất thấp hay có thể nói là độc lập với nhau).

SỬ DỤNG PHƯƠNG PHÁP PCA



LỰA CHỌN K = 7 → LOẠI BỎ 2 BIẾN THUỘC TÍNH

CÂN NHẮC LOẠI BỎ BIẾN **IS MALE** VÀ **BMI**

XÂY DỰNG 2 KIỂM ĐỊNH:

KIỂM ĐỊNH 1: NAM VÀ NỮ SỬ DỤNG BHYT NHƯ NHAU, VỚI ĐỘ TIN CẬY 95%

$$\begin{cases} H_0: \mu_{\text{Male}}(\text{charges}) = \mu_{\text{Female}}(\text{charges}) \\ H_1: \mu_{\text{Male}}(\text{charges}) \neq \mu_{\text{Female}}(\text{charges}) \end{cases}$$

KIỂM ĐỊNH 2: CHỈ SỐ SỨC KHỎE (BMI) KHÔNG ẢNH HƯỞNG ĐẾN VIỆC LÀM MỚI BHYT, VỚI ĐỘ TIN CẬY 95%

$$\begin{cases} H_0: \mu_{\text{bmi_stand}}(\text{insur_renewals}) = \mu_{\text{bmi_stand}}(\text{insur_renewals}) \\ H_1: \mu_{\text{bmi_stand}}(\text{insur_renewals}) \neq \mu_{\text{bmi_stand}}(\text{insur_renewals}) \end{cases}$$

GIẢM CHIỀU DỮ LIỆU

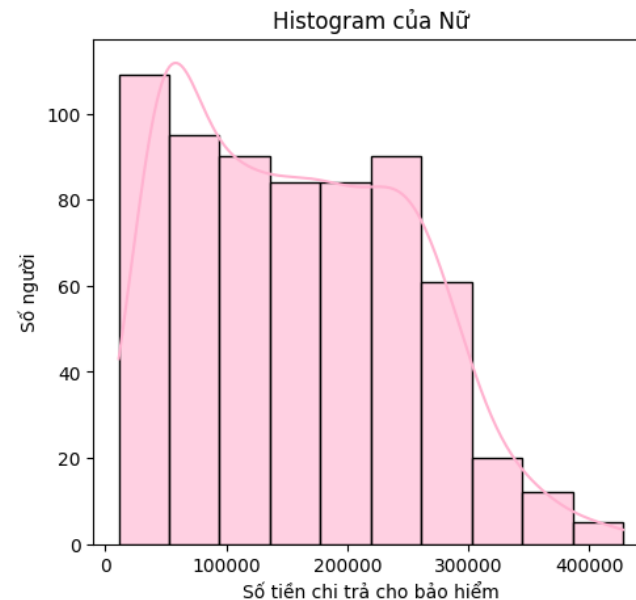
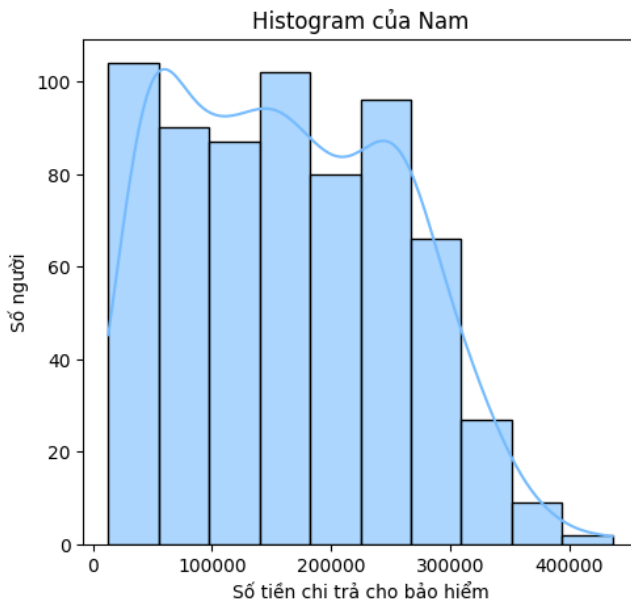
KIỂM ĐỊNH 1: NAM VÀ NỮ SỬ DỤNG BHYT NHƯ NHAU, VỚI ĐỘ TIN CẬY 95%

$$\begin{cases} H_0: \mu_{\text{Male}}(\text{charges}) = \mu_{\text{Female}}(\text{charges}) \\ H_1: \mu_{\text{Male}}(\text{charges}) \neq \mu_{\text{Female}}(\text{charges}) \end{cases}$$

```
1    663
0    650
Name: is_male, dtype: int64
```

```
Phương sai của giới tính nam: 8525114999.80
Phương sai của giới tính nữ: 8571391607.22
```

Trị số p = 0.2859 >= 0.05 nên chấp nhận H_0 .
=> $\mu[\text{Male}] == \mu[\text{Female}]$



Đối với biểu đồ Histogram, nhóm thấy được với số lượng mẫu nam nhiều hơn mẫu là nữ, tổng khoảng tiền nam giới chi trả cho bảo hiểm cũng chênh lệch hơn nữ giới khá nhỏ.

→ **TỪ KIỂM ĐỊNH VÀ BIỂU ĐỒ TRỰC QUAN, NHÓM ĐÃ CÓ CƠ SỞ ĐỂ LOẠI BỎ BIẾN IS_MALE RA BỘ DỮ LIỆU ĐỂ TIẾT KIỆM THỜI GIAN NGHIÊN CỨU.**

GIẢM CHIỀU DỮ LIỆU

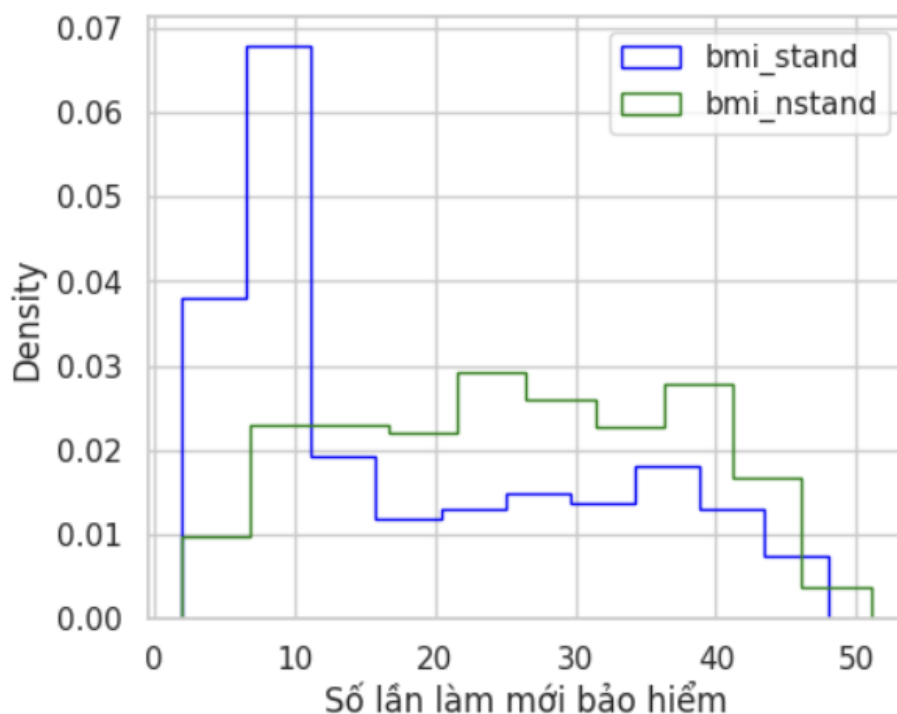
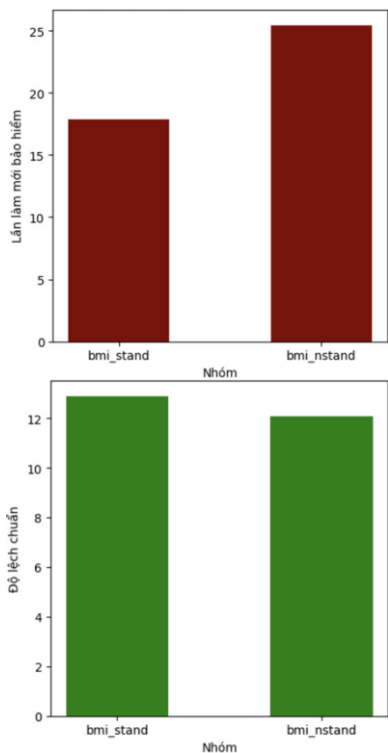
KIỂM ĐỊNH 2: CHỈ SỐ SỨC KHỎE (BMI) KHÔNG ẢNH HƯỞNG ĐẾN VIỆC LÀM MỚI BHYT, VỚI ĐỘ TIN CẬY 95%

$$\begin{cases} H_0: \mu_{\text{bmi_stand}} \text{ insur_renewals} = \mu_{\text{bmi_stand}} \text{ insur_renewals} \\ H_1: \mu_{\text{bmi_stand}} \text{ insur_renewals} \neq \mu_{\text{bmi_stand}} \text{ insur_renewals} \end{cases}$$

Số quan sát của mẫu bmi chuẩn: 349// Giá trị trung bình: 17.9083
Số quan sát của mẫu bmi không chuẩn: 964// Giá trị trung bình: 25.4481
Độ lệch chuẩn của tập chỉ số sức khỏe bình thường: 12.8949
Độ lệch chuẩn của tập chỉ số sức khỏe dưới và quá tiêu chuẩn: 12.0846
Độ chênh lệch của std: 0.810308

Trị số p = 8.915994860989547e-28 < 0.0500 bác bỏ H_0

=> Chỉ số bmi có ảnh hưởng đến quyết định làm mới bảo hiểm



Thông qua biểu đồ phân phối của hai biến, nhóm thấy được rằng nhóm người có chỉ số BMI không chuẩn có xu hướng làm mới BHYT khá đều ở từng mức độ, trong khi đó nhóm người có chỉ số BMI chuẩn lại hầu hết chỉ tập trung làm mới bảo hiểm từ 1 đến 10 lần, sau đó họ không còn có xu hướng làm mới bảo hiểm thêm nữa.

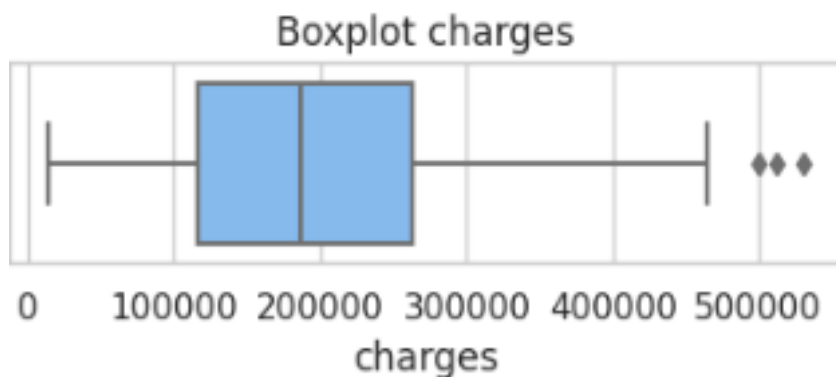
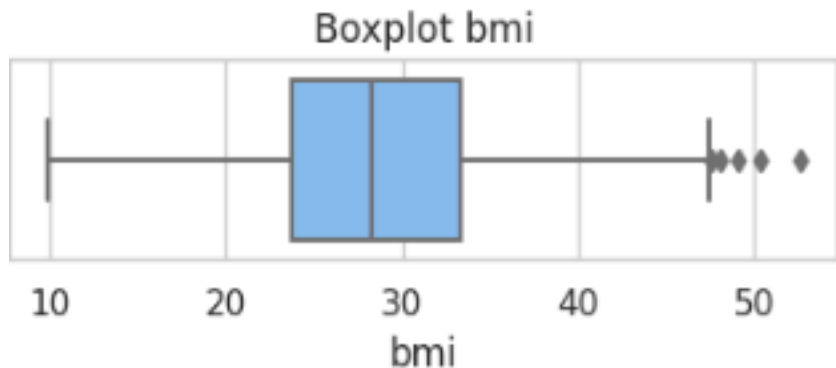
→ **TỪ KIỂM ĐỊNH VÀ BIỂU ĐỒ TRỰC QUAN, NHÓM VẪN CHƯA CÓ CƠ SỞ ĐỂ LOẠI BỎ BIẾN BMI RA BỘ DỮ LIỆU.**

XỬ LÝ GIÁ TRỊ OUTLIERS

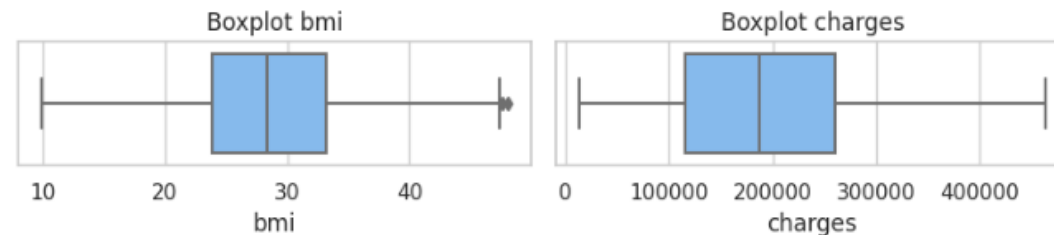
QUY TẮC 3-SIGMA

XÁC ĐỊNH & LOẠI BỎ OUTLIERS

```
Outliers cột age :[]  
Outliers cột bmi :[49.06, 52.58, 50.38]  
Outliers cột insur_renewals :[]  
Outliers cột charges :[530823.0, 512614.0, 499200.0]
```



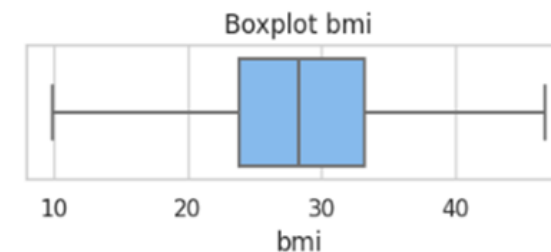
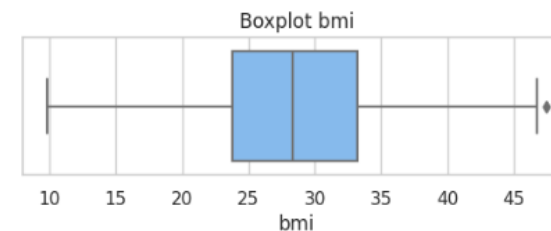
SAU KHI LOẠI BỎ OUTLIERS



KHOẢNG TỬ PHÂN VI (IQR)

XÁC ĐỊNH & LOẠI BỎ OUTLIERS CÒN LẠI

```
outliers cột bmi :[47.6, 48.07]
```

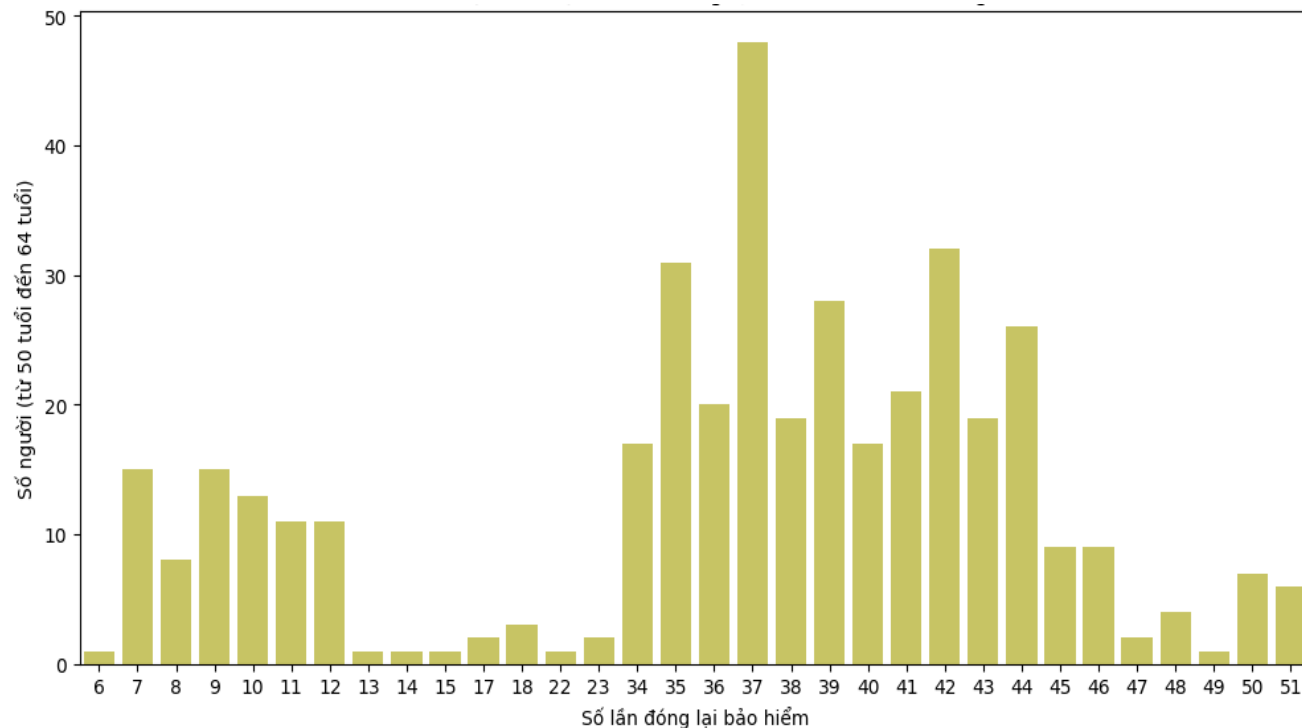


NHÂN KHẨU HỌC (DEMOGRAPHICS)

KIỂM ĐỊNH 3: SỐ LẦN TRUNG BÌNH ĐÓNG LẠI BHYT CỦA NHÓM KH CAO TUỔI (TỪ 50 TUỔI TRỞ LÊN) KHÔNG ÍT HƠN 39 LẦN. ĐỘ TIN CẬY 95%.

$$\begin{cases} H_0: 39 \leq \mu_{\text{Elder}} (\text{renewals}) \\ H_1: 39 > \mu_{\text{Elder}} (\text{renewals}) \end{cases}$$

Trị số $p = 3.9497637822051615e-17 < 0.0500$ nên bác bỏ H_0 .
 $\Rightarrow 39 > \mu$



Số lượng KH có số lần đóng lại bảo hiểm hơn 37 lần chiếm đa phần trong tổng số. Trong khi, lượng KH đóng bảo hiểm chỉ khi cần thiết tương đối thấp khi so với xu hướng này (khoảng 84 người).

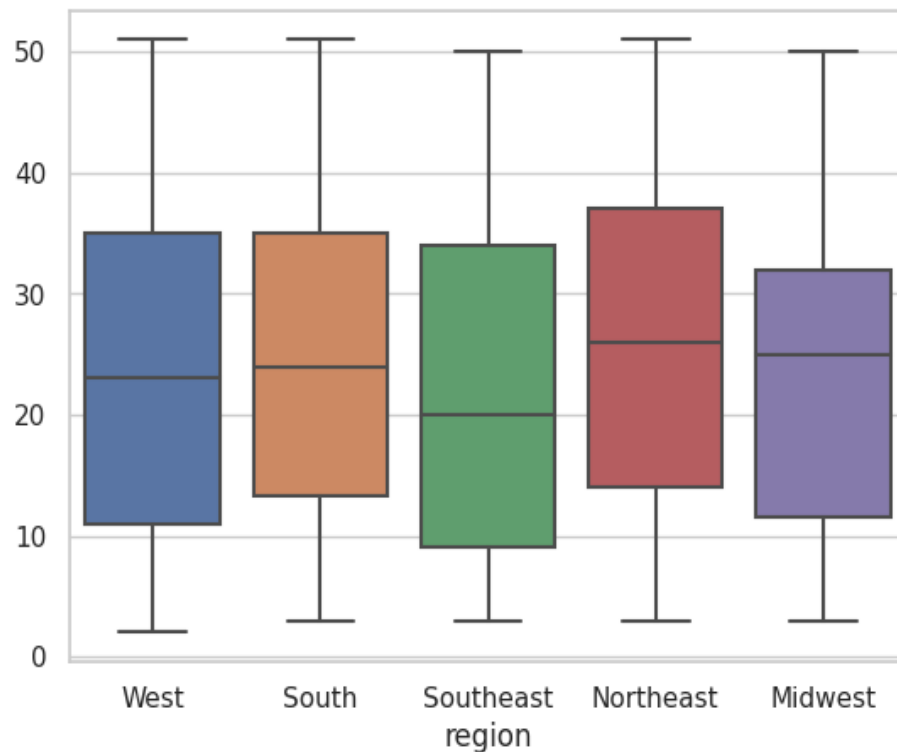
- Quy luật: chọn BHYT không quan trọng về giới tính, việc chọn mua loại BHYT nào được quyết định ở độ tuổi, tuổi càng cao thì mua bảo hiểm lại càng nhiều.
- Đề xuất: doanh nghiệp BHYT có thể tập trung đến khía cạnh đóng bảo hiểm thường niên của nhóm người cao tuổi để quảng bá các gói bảo hiểm BHYT theo kỳ.

ĐỊA LÝ (GEOGRAPHY)

KIỂM ĐỊNH 4: CÁC BANG Ở CÁC VÙNG KHÁC NHAU CÓ XU HƯỚNG CHI TIỀN CHO BHYT KHÁC NHAU. ĐỘ TIN CẬY 95%

$$\begin{cases} H_0: \mu_{\text{West}} = \mu_{\text{Midwest}} = \mu_{\text{Northeast}} = \mu_{\text{Southeast}} = \mu_{\text{South}} \\ H_1: \text{Có ít nhất một giá trị } \mu \text{ khác biệt} \end{cases}$$

* Trị số $p = 0.0097 < 0.0500$ cho nên bác bỏ H_0
==> có sự khác biệt giữa các khu vực



Sự phân bố của các vùng có sự khác biệt rất rõ ràng ở các vùng như Midwest và Southeast so với các khu vực còn lại khi các giá trị phân bố lệch hẳn sang hai bên so với giá trị trung bình trong khi 3 vùng còn lại thì không.

→ Công ty bảo hiểm nên có những chính sách marketing phù hợp theo từng vùng địa lý khác nhau.

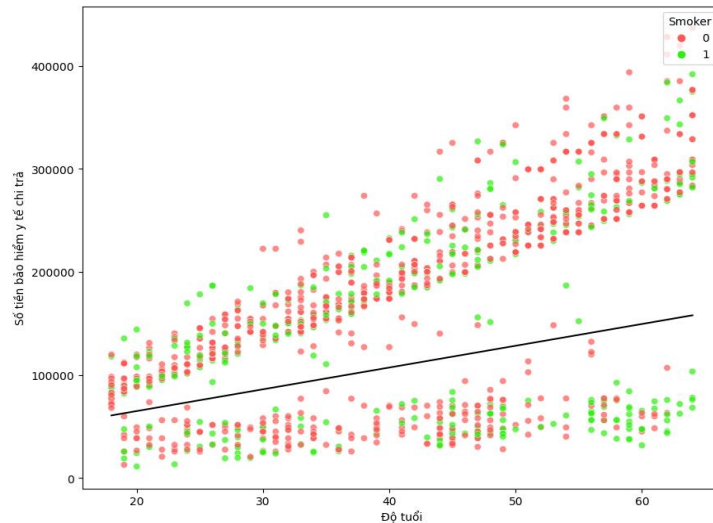
THÓI QUEN VÀ LỐI SỐNG

KIỂM ĐỊNH 5: SỐ TIỀN MÀ NGƯỜI SỬ DỤNG THUỐC LÁ CHỈ TRẢ CHO BHYT SO VỚI NGƯỜI BÌNH THƯỜNG CÓ GIỐNG NHAU KHÔNG? ĐỒNG THỜI, KIỂM ĐỊNH XEM THÓI QUEN HÚT THUỐC CÓ LIÊN QUAN ĐẾN GIỚI TÍNH CỦA KH HAY KHÔNG? ĐỘ TIN CẬY 95%

$$(1): \begin{cases} H_0: \mu_{\text{Smoker}}(\text{charges}) = \mu_{\text{non-Smoker}}(\text{charges}) \\ H_1: \mu_{\text{Smoker}}(\text{charges}) \neq \mu_{\text{non-Smoker}}(\text{charges}) \end{cases}$$

Phương sai của người nghiện: 7952246651.29
Phương sai của người bình thường: 8582136205.57

Trị số $p = 1.232674037705928e-06 < 0.05$ nên bác bỏ H_0 .
 $\Rightarrow \mu[\text{Smoker}] \neq \mu[\text{Nonsmoker}]$



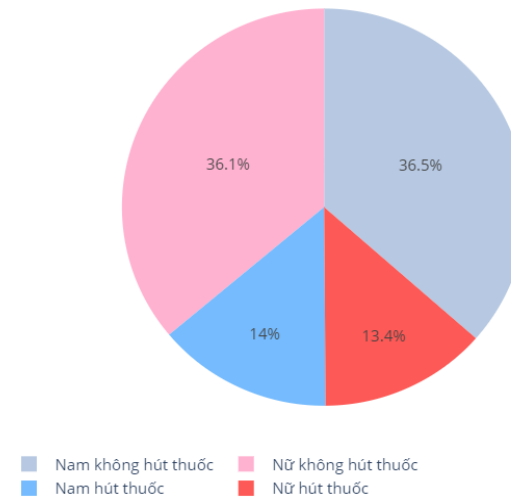
Khi nhìn theo khuynh hướng độ tuổi trên, nhóm xác định được rằng nhóm KH càng lớn tuổi thì có xu hướng trả tiền bảo hiểm càng nhiều hơn, bất chấp cho việc người đó có hút thuốc hay không.

$$(2): \begin{cases} H_0: \text{"smoker" và "is_male" là hai biến độc lập} \\ H_1: \text{"smoker" và "is_male" là hai biến phụ thuộc} \end{cases}$$

smoker	0	1	Tổng cộng
is_male			
0	474	176	650
1	479	184	663
Tổng cộng	953	360	1313

probability=0.9500
critical=9.4877
stat=0.075

Trị số $p = 0.9993 \geq 0.0500$ nên chấp nhận H_0 .
 \Rightarrow Giới tính và lối sống hút thuốc không ảnh hưởng đến nhau.



Số lượng nữ giới không có thói quen hút thuốc là 474, có tính tương đồng với 3 nhóm còn lại. Hay có thể nói, không có sự khác biệt về giới tính trong thói quen hút thuốc của KH.

BIẾN is_MALE KHÔNG MANG LẠI BẤT KỲ GIÁ TRỊ THÔNG TIN NÀO, HAY VIỆC LOẠI BỎ BIẾN KHÔNG LÀM ẢNH HƯỞNG ĐẾN BẤT CỨ BIẾN NÀO TRONG BỘ DỮ LIỆU.

KIỂM ĐỊNH THỐNG KÊ

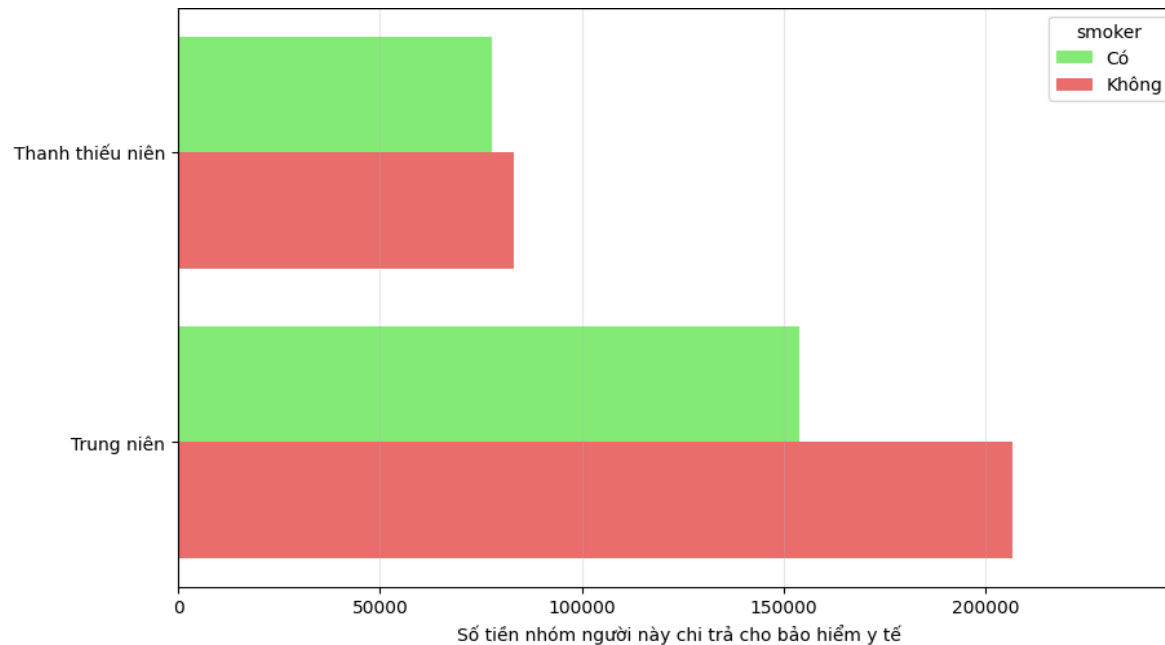
THỜI QUEN VÀ LỐI SỐNG

KIỂM ĐỊNH 6: SỐ LƯỢNG KH CỐ THỜI QUEN HÚT THUỐC Ở ĐỘ TUỔI VỊ THÀNH NIÊN (TỪ 18 TUỔI ĐẾN 29 TUỔI) SỬ DỤNG BHYT BẰNG NHÓM KH Ở ĐỘ TUỔI TRUNG NIÊN (TỪ 30 TUỔI ĐẾN 49 TUỔI) , ĐỘ TIN CẬY 95%

- H_0 : Sự khác biệt về chi phí đóng bảo hiểm giữa nhóm thanh thiếu niên sử dụng thuốc lá và nhóm trung niên sử dụng thuốc lá là không đáng kể.
- H_1 : Có sự khác biệt đáng kể về chi phí đóng bảo hiểm giữa nhóm thanh thiếu niên sử dụng thuốc lá và nhóm trung niên sử dụng thuốc lá.

Trị số p = 6.47966302207274e-11 < 0.05 nên bác bỏ H_0 .

=> Có sự khác biệt đáng kể về chi phí y tế giữa nhóm vị thành niên hút thuốc và nhóm trung niên hút thuốc.

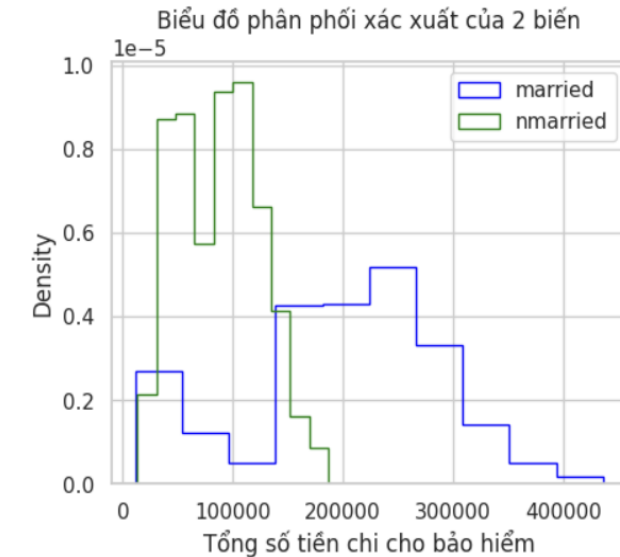


KIỂM ĐỊNH 7: YẾU TỐ HÔN NHÂN GIA ĐÌNH CÓ LÀM ẢNH HƯỞNG ĐẾN VIỆC CHI TRẢ CHO BHYT HAY KHÔNG?, ĐỘ TIN CẬY 95%

$$\begin{cases} H_0: \mu_{\text{married}} \text{ charges} \leq \mu_{\text{nmarrried}} \text{ charges} \\ H_1: \mu_{\text{married}} \text{ charges} > \mu_{\text{nmarrried}} \text{ charges} \end{cases}$$

Trị số p = 6.239814849628644e-307 < 0.0500 bác bỏ H_0

=> Yếu tố hôn nhân có ảnh hưởng đến quyết định làm mới bảo hiểm

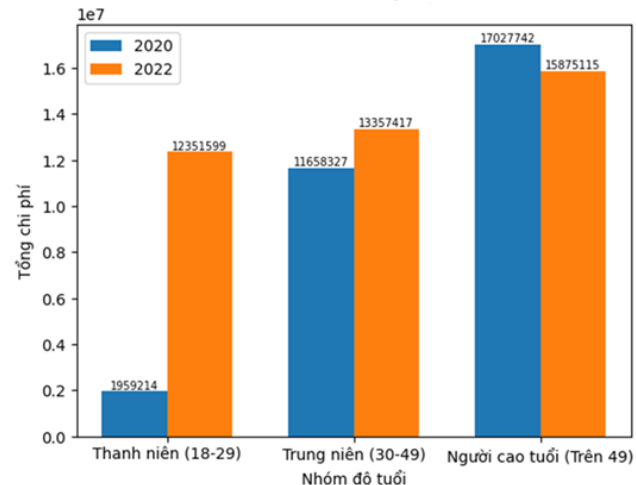
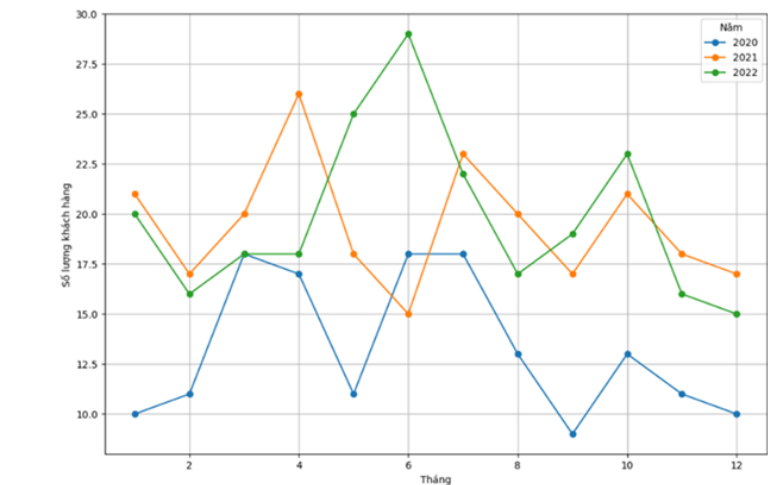
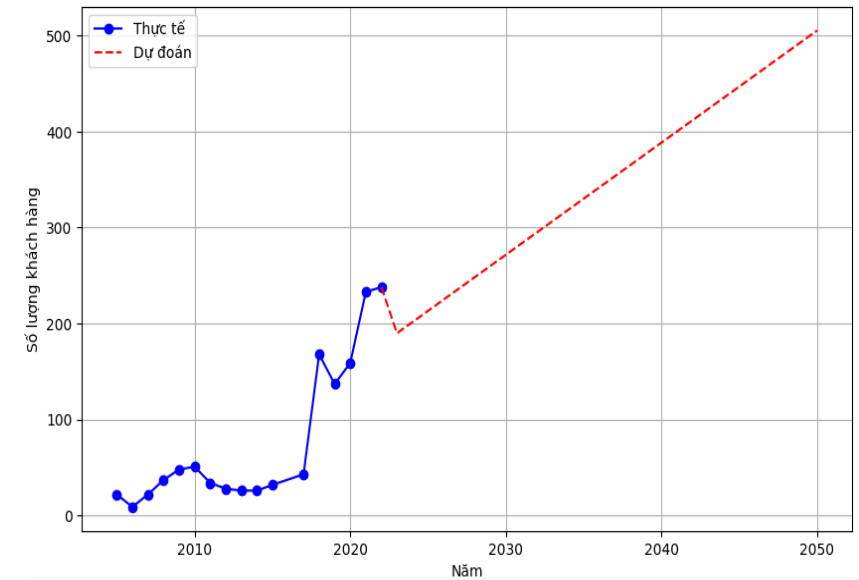
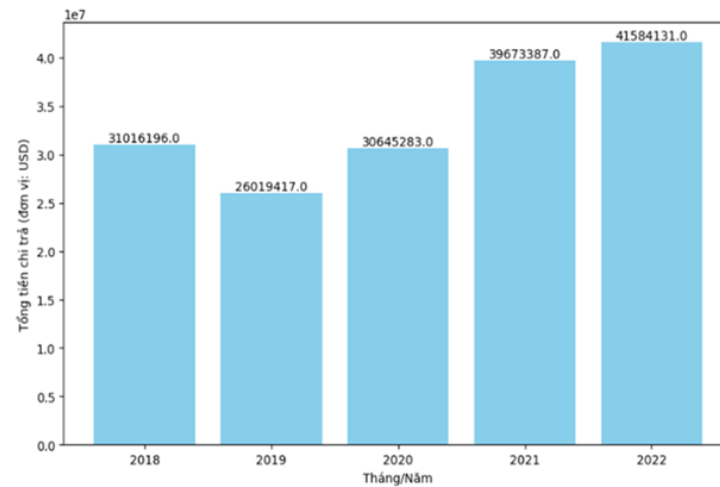
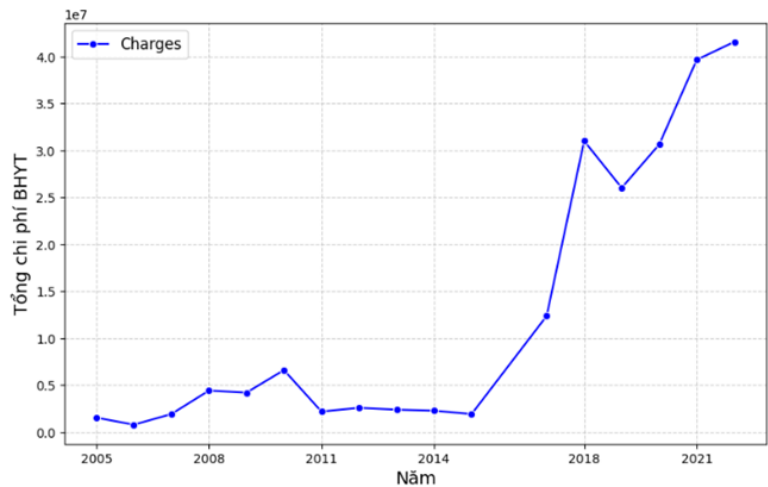


Quy luật: khi con người càng lớn tuổi dần và có nhiều mối quan tâm hơn thì họ có khả năng chi trả BHYT cao hơn.

Giải pháp cho các công ty bảo hiểm: tập trung vào nhóm người có xu hướng lớn tuổi hơn và đã có gia đình.

BIỂU DIỄN TRỰC QUAN

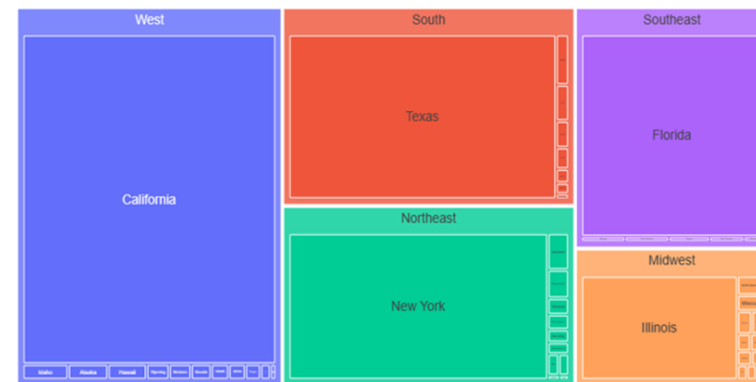
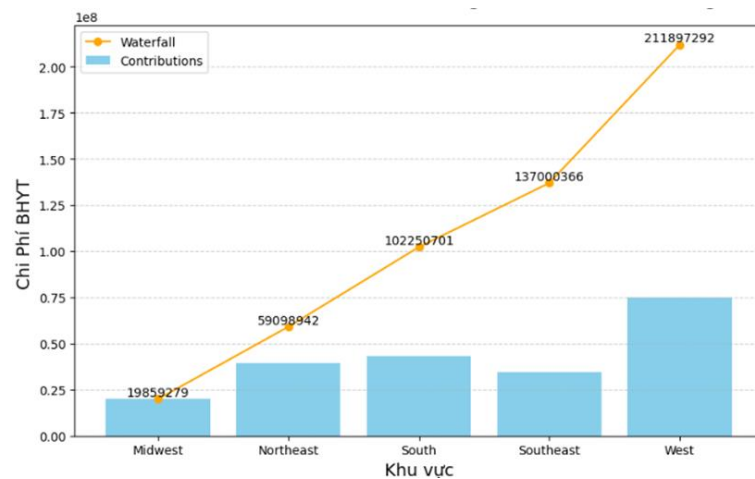
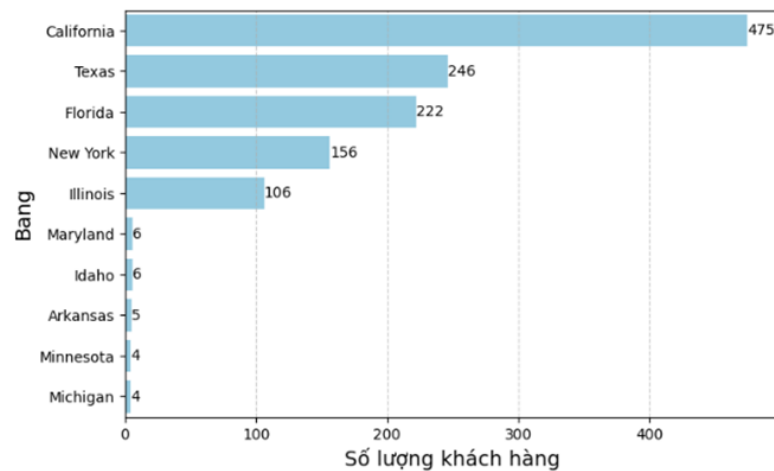
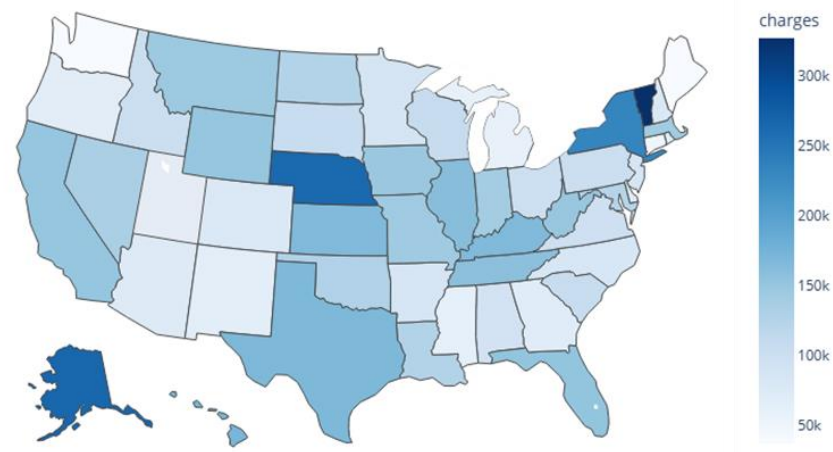
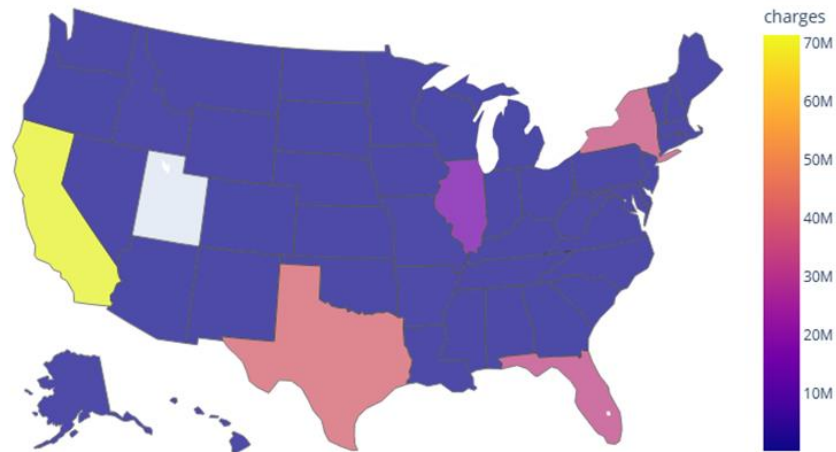
PHÂN TÍCH CHUỖI THỜI GIAN



Xu hướng tăng trưởng của số lượng KH tham gia BHYT đã và đang diễn ra một cách tích cực. Đặc biệt, với sự gia tăng mạnh mẽ trong nhóm Độ tuổi Thanh niên và Trung niên, cùng với sự điều chỉnh hợp lý của chính phủ đối với nhóm người Cao tuổi, nhóm có thể kỳ vọng sự gia tăng mạnh về số lượng KH tham gia BHYT trong tương lai.

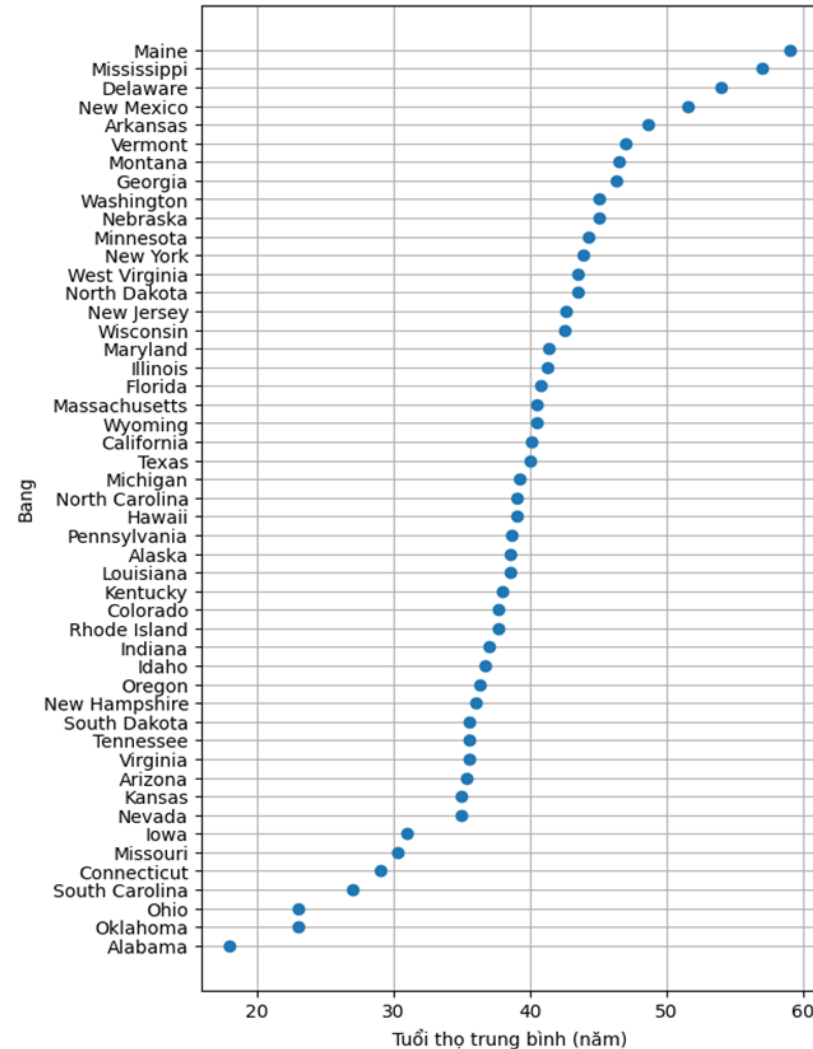
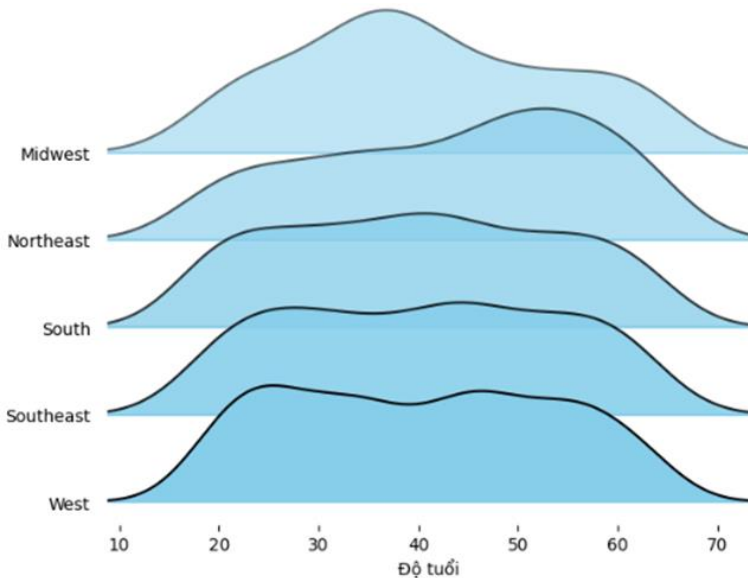
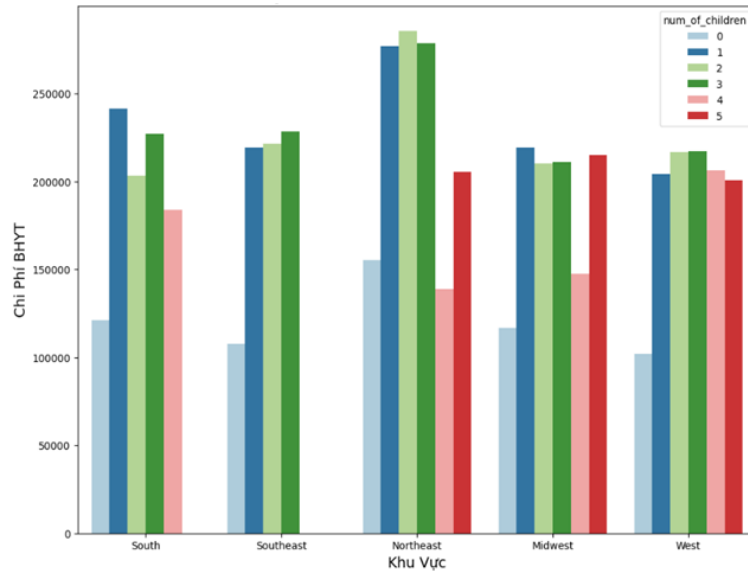
BIỂU DIỄN TRỰC QUAN

PHÂN TÍCH KHU VỰC/ MẬT ĐỘ



BIỂU DIỄN TRỰC QUAN

PHÂN TÍCH KHU VỰC/ MẬT ĐỘ



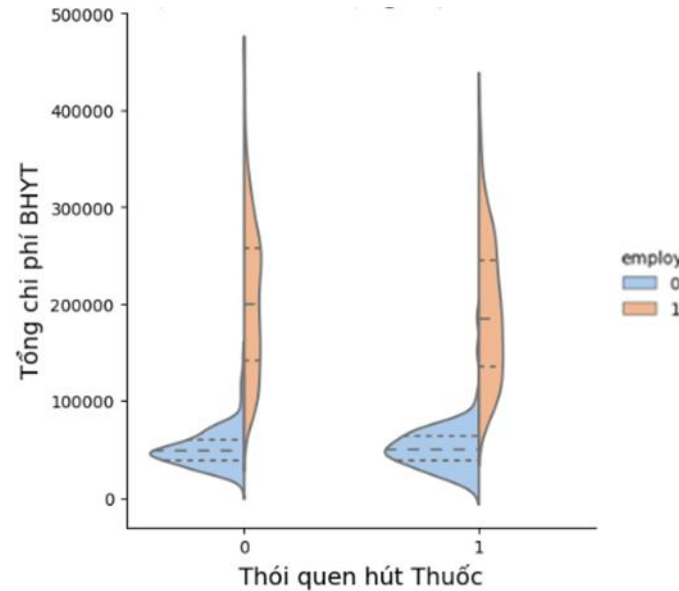
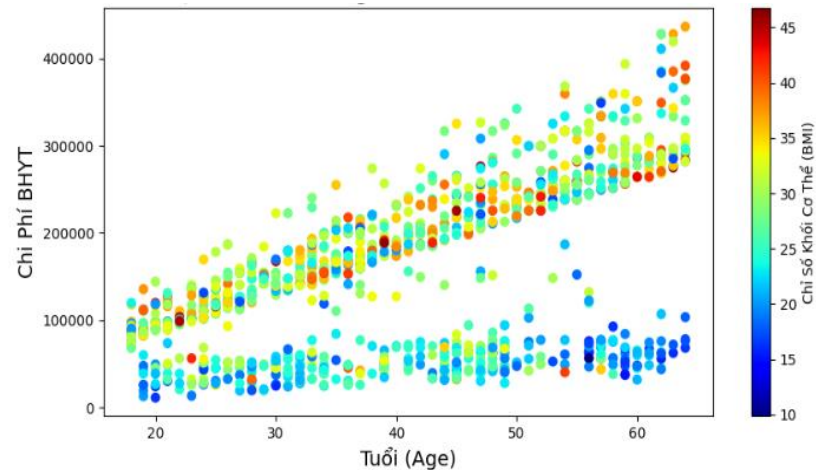
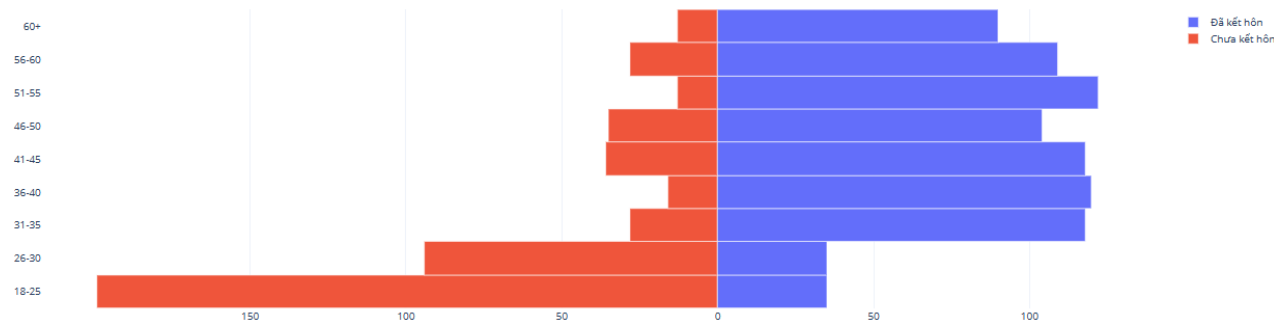
Mật độ dân số ảnh hưởng đến khả năng tiếp cận và chi phí bảo hiểm ở Hoa Kỳ. Các khu vực có mật độ dân số cao thường có nhiều lựa chọn bảo hiểm hơn và chi phí thấp hơn so với các khu vực có mật độ dân số thấp.

Các khu vực có mật độ dân số cao thường có nhiều lựa chọn bảo hiểm hơn. Điều này là do các khu vực này có nhiều người mua bảo hiểm tiềm năng, điều này khiến các công ty bảo hiểm có động lực cung cấp nhiều lựa chọn hơn.

Ngoài ra, các khu vực có mật độ dân số cao thường có nhiều nguồn lực hơn để hỗ trợ những người không có bảo hiểm. Điều này có thể bao gồm các chương trình Medicaid và Medicare, cũng như các chương trình trợ cấp của tiểu bang và địa phương.

BIỂU DIỄN TRỰC QUAN

PHÂN TÍCH SỨC KHỎE & GIA ĐÌNH



→ **QUẢN LÝ CHI PHÍ VÀ CHĂM SÓC Y TẾ ĐẶC BIỆT CHO NHÓM TUỔI 25-45 VÀ BMI 25-45.**

→ **TĂNG CƯỜNG CHIẾN LƯỢC THÚC ĐẨY BẢO HIỂM VÀ QUẢN LÝ CHI PHÍ CHO NHÓM KH CÓ THÓI QUEN HÚT THUỐC.**

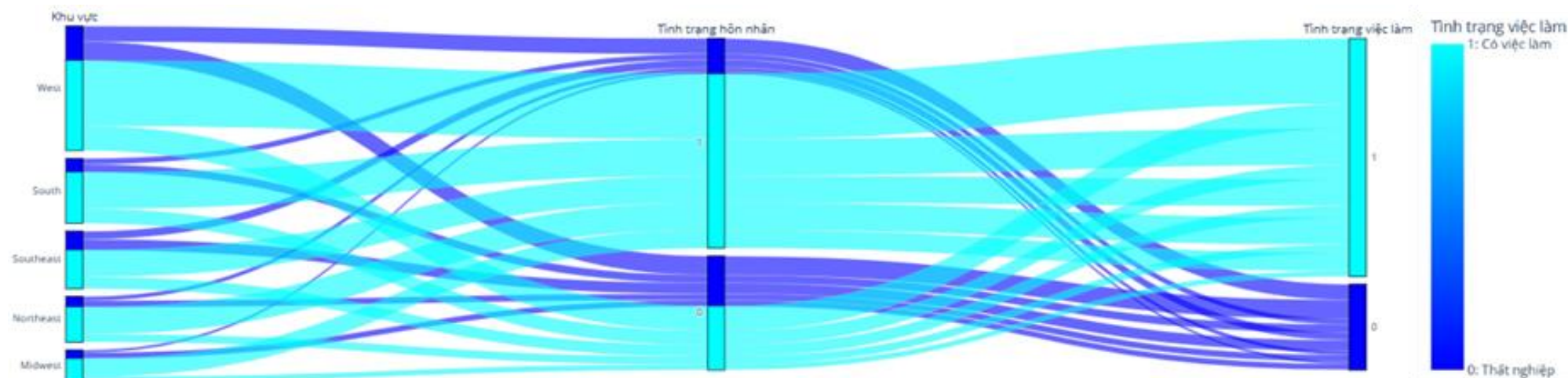
→ **TẬP TRUNG NHÓM ĐỐI TƯỢNG:**

Tăng cường dịch vụ tư vấn và hỗ trợ cho KH, giúp họ hiểu rõ hơn về lợi ích của việc sử dụng BHYT. Thực hiện ở nhóm KH có tuổi từ 26-35 và 36 trở đi, đặc biệt là những người đã kết hôn, vì họ có nhu cầu về BHYT cao hơn.

Đối với chính phủ từng bang, cần đưa ra các chương trình nhân đạo: ưu đãi hoặc tạo cơ hội dành riêng cho đối tượng KH kết hôn sớm ở độ tuổi dưới 26 trở đi để thu hút họ sử dụng BHYT.

BIỂU DIỄN TRỰC QUAN

PHÂN TÍCH TỔNG HỢP



Dựa trên xu hướng hiện tại, nhóm có thể dự đoán rằng tỷ lệ KH vẫn đang đi làm khi tham gia BHYT sẽ tiếp tục tăng trong tương lai, đặc biệt là trong nhóm KH đã kết hôn và KH ở khu vực phía Tây. Tuy nhiên, cần có thêm nghiên cứu để xác nhận và hiểu rõ hơn về những xu hướng này.

KIỂM ĐỊNH: CÓ SỰ KHÁC BIỆT VỀ CHI PHÍ BHYT GIỮA NHÓM CÓ VIỆC LÀM VÀ KHÔNG CÓ VIỆC LÀM, ĐỘ TIN CẬY 95%

μ_1 : Trung bình chi phí BHYT của nhóm có việc làm.

μ_2 : Trung bình chi phí BHYT của nhóm không có việc làm.

$H_0: \mu_1 = \mu_2$ (Không có sự khác biệt về chi phí BHYT giữa hai nhóm)

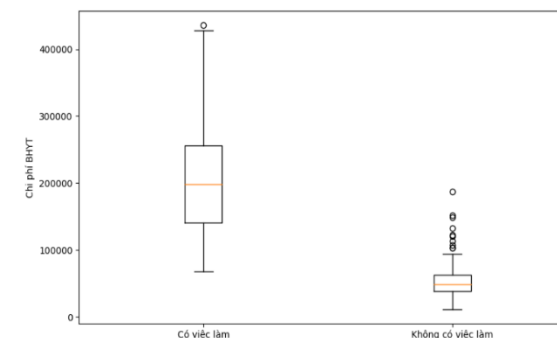
$H_1: \mu_1 \neq \mu_2$ (Có sự khác biệt về chi phí BHYT giữa hai nhóm)

Z-statistic: 36.55338036574715

P-value: 1.5757431907735459e-292

Vì $p_value < \alpha \Rightarrow$ Có cơ sở để bác bỏ giả thuyết H_0

\Rightarrow Có sự khác biệt về chi phí BHYT giữa nhóm có việc làm và không có việc làm.



XIN CHÂN THÀNH CẢM ƠN!