

.....

CAN YOU
HAVE
HEART
DISEASE?



BEFORE

โรคหัวใจเป็นสาเหตุการตายอันดับต้น ๆ ของคนส่วนใหญ่ในสหรัฐอเมริกา (ชาวแอฟริกัน อเมริกัน ชาวอเมริกันอินเดียน ชาวอะแลสกา และคนผิวขาว) ประมาณครึ่งหนึ่งของชาว อเมริกันทั้งหมด (47%) มีปัจจัยเสี่ยงอย่างน้อย 1 ใน 3 ประการของโรคหัวใจ ได้แก่ ความดัน โลหิตสูง คอเลสเตอรอลสูง และการสูบบุหรี่ ตัวบ่งชี้ที่สำคัญอื่น ๆ ได้แก่ สถานะโรคเบาหวาน โรคอ้วน (ค่าดัชนีมวลกายสูง) ออกกำลังกายไม่เพียงพอ หรือดื่มแอลกอฮอล์มากเกินไป ถ้า ไม่มีการตรวจสอบปัจจัยความเสี่ยง เราอาจไม่รู้ได้เลยว่าตนเองมีโอกาสในการเป็นโรคหัวใจ

AFTER

- บุคคลที่ไม่สามารถประเมินความเสี่ยงในโรคหัวใจได้เองโดยไม่ต้องเสียเงินตรวจสุขภาพร่างกาย
- หมอยาบาลสามารถรู้ความเสี่ยงของคนไข้ได้และสามารถรักษาได้ทันท่วงทีและ อาจจะเพิ่มโอกาสในการรักษาโรคหัวใจให้หายได้ถ้าได้รับการรักษาตั้งแต่เนิ่นๆ
- อัตราการเกิดโรคหัวใจลดลง 10%



BRIDGE

ชุดข้อมูลที่ใช้

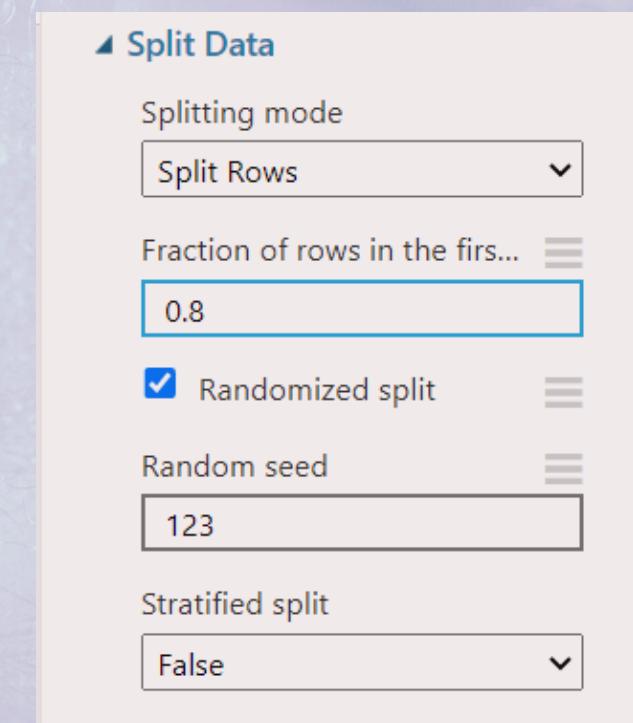
[HTTPS://WWW.KAGGLE.COM/DATASETS/KAMILPYTLAK/PERSONAL-KEY-INDICATORS-OF-HEART-DISEASE](https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease)

rows	columns											
319795	18											
HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWalking	Sex	AgeCategory	Race	Diabetic	PhysicalActivi
No	16.6	Yes	No	No	3	30	No	Female	55-59	White	Yes	Yes
No	20.34	No	No	Yes	0	0	No	Female	80 or older	White	No	Yes
No	26.58	Yes	No	No	20	30	No	Male	65-69	White	Yes	Yes
No	24.21	No	No	No	0	0	No	Female	75-79	White	No	No
No	23.71	No	No	No	28	0	Yes	Female	40-44	White	No	Yes
Yes	28.87	Yes	No	No	6	0	Yes	Female	75-79	Black	No	No
No	21.63	No	No	No	15	0	No	Female	70-74	White	No	Yes
No	31.64	Yes	No	No	5	0	Yes	Female	80 or older	White	Yes	No

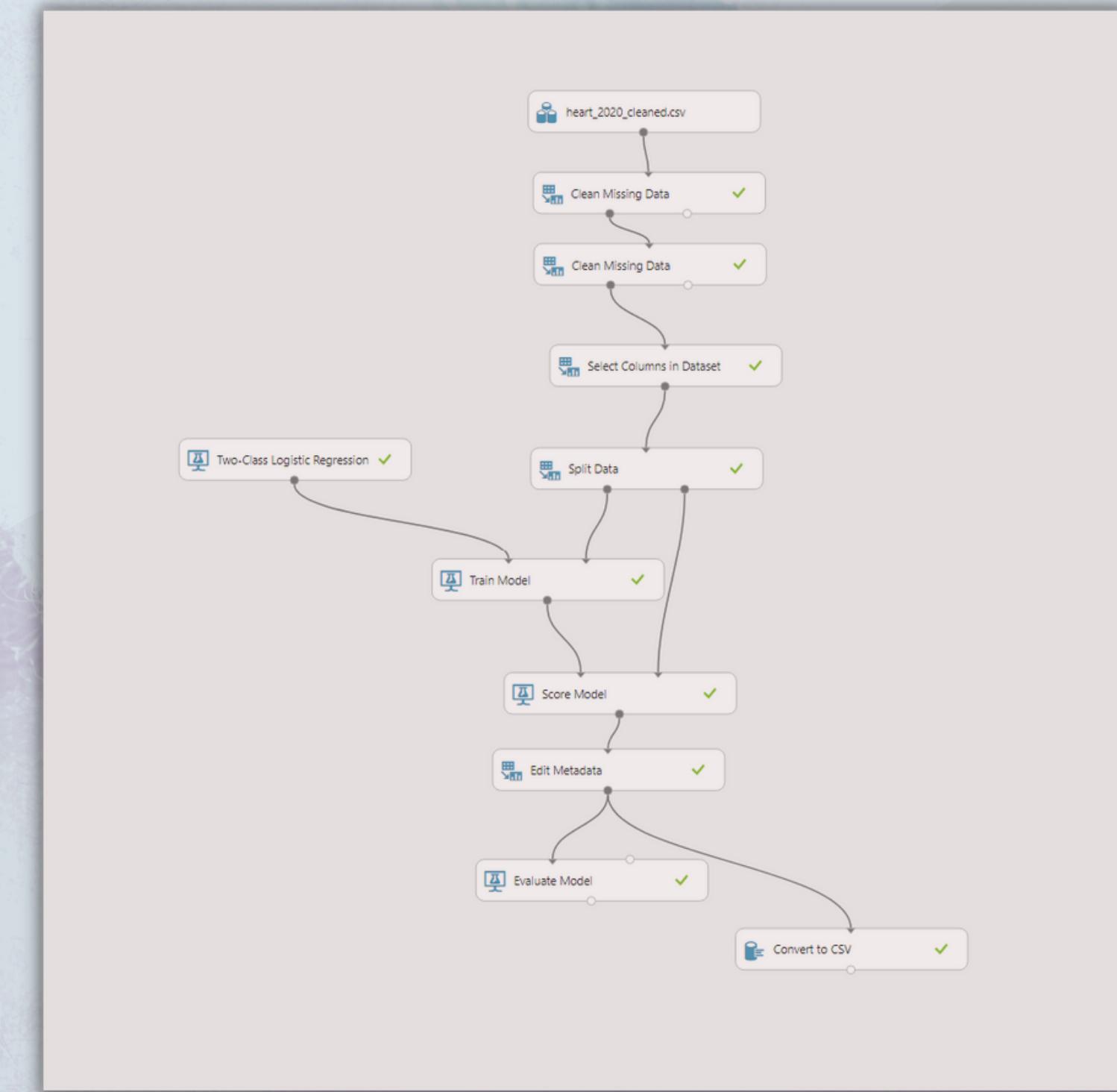
ข้อมูลที่ได้ทำการวิเคราะห์มีดังนี้

- HeartDisease: โรคหัวใจ
- BMI: ค่าดัชนีมวลกาย
- Smoking: สูบบุหรี่
- AlcoholDrinking: การดื่มเครื่องดื่มแอลกอฮอล์
- Stroke: โรคหลอดเลือดสมอง
- PhysicalHealth: สุขภาพกาย
- MentalHealth: สุขภาพจิต
- DiffWalking: เดินแบบ diff
- Sex: เพศ
- AgeCategory: หมวดหมู่อายุ
- Race: สีผิว
- Diabetic: เบาหวาน
- PhysicalActivity: การออกกำลังกาย

ใช้โมเดล Two-Class Logistic Regression เป็นการสร้างแบบจำลองแบบทํานายอนาคต ที่ได้จากข้อมูลเก่า



เลือกข้อมูลที่จะเป็น และแบ่งมา train 80% และ test 20%



เลือกที่จะทำนาย การเกิดโรคหัวใจ

ทำการ score-model และ
Visualization ด้วย Evaluate

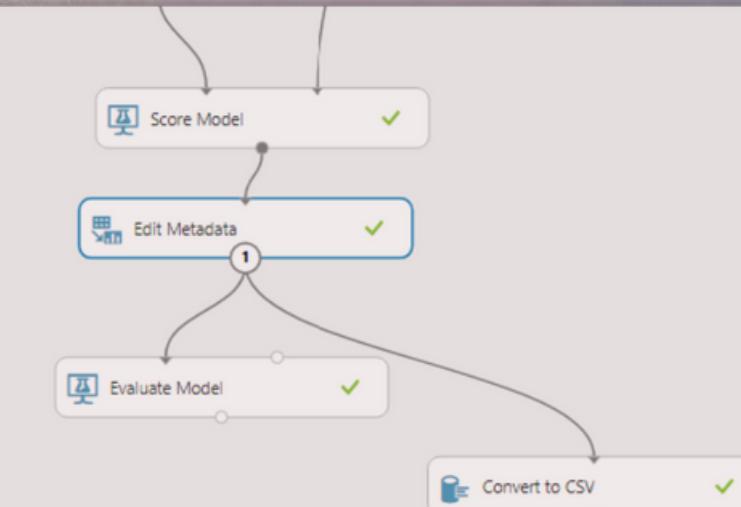
Label column

Selected columns:

Column names:

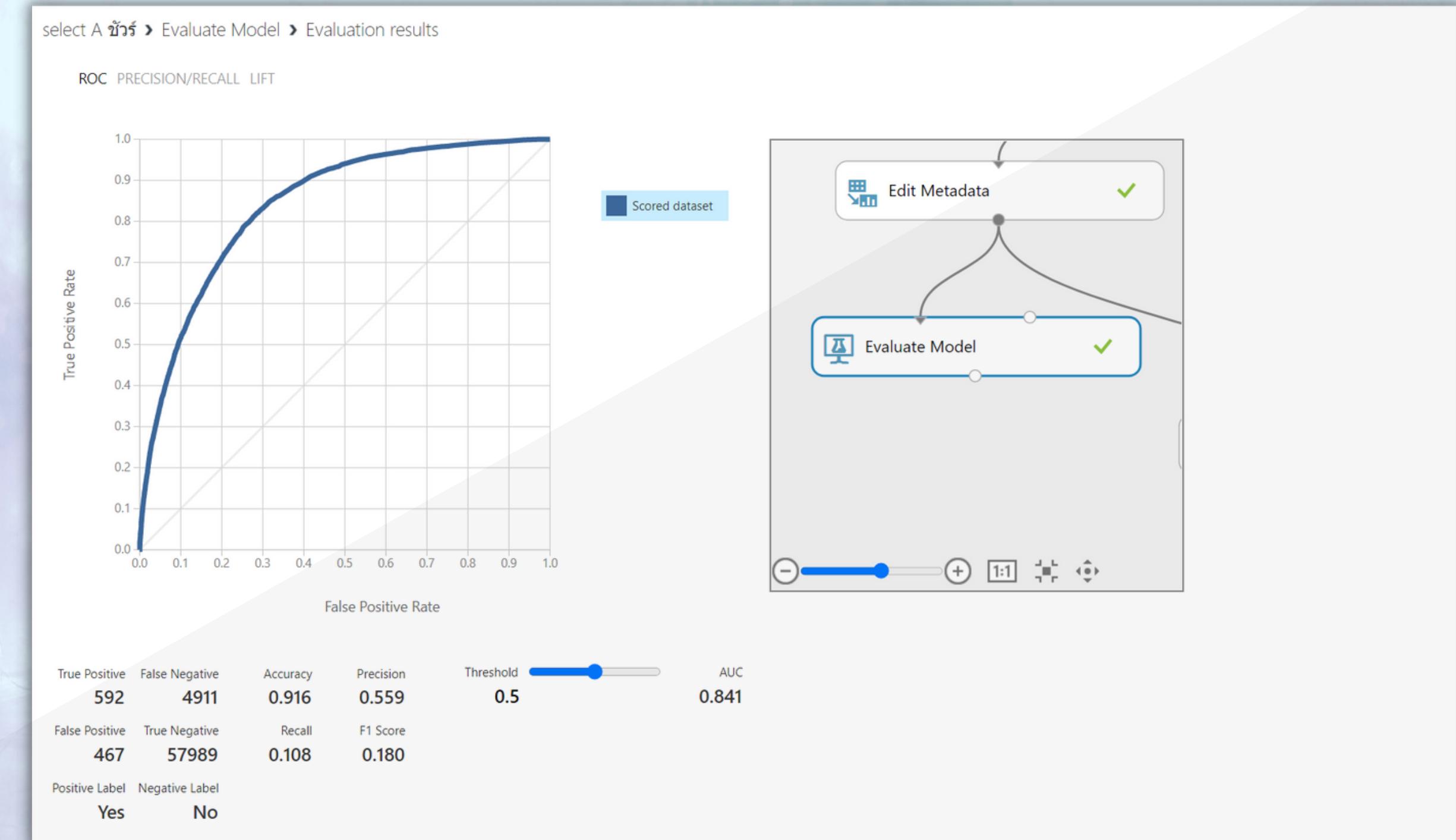
HeartDisease

[Launch column selector](#)



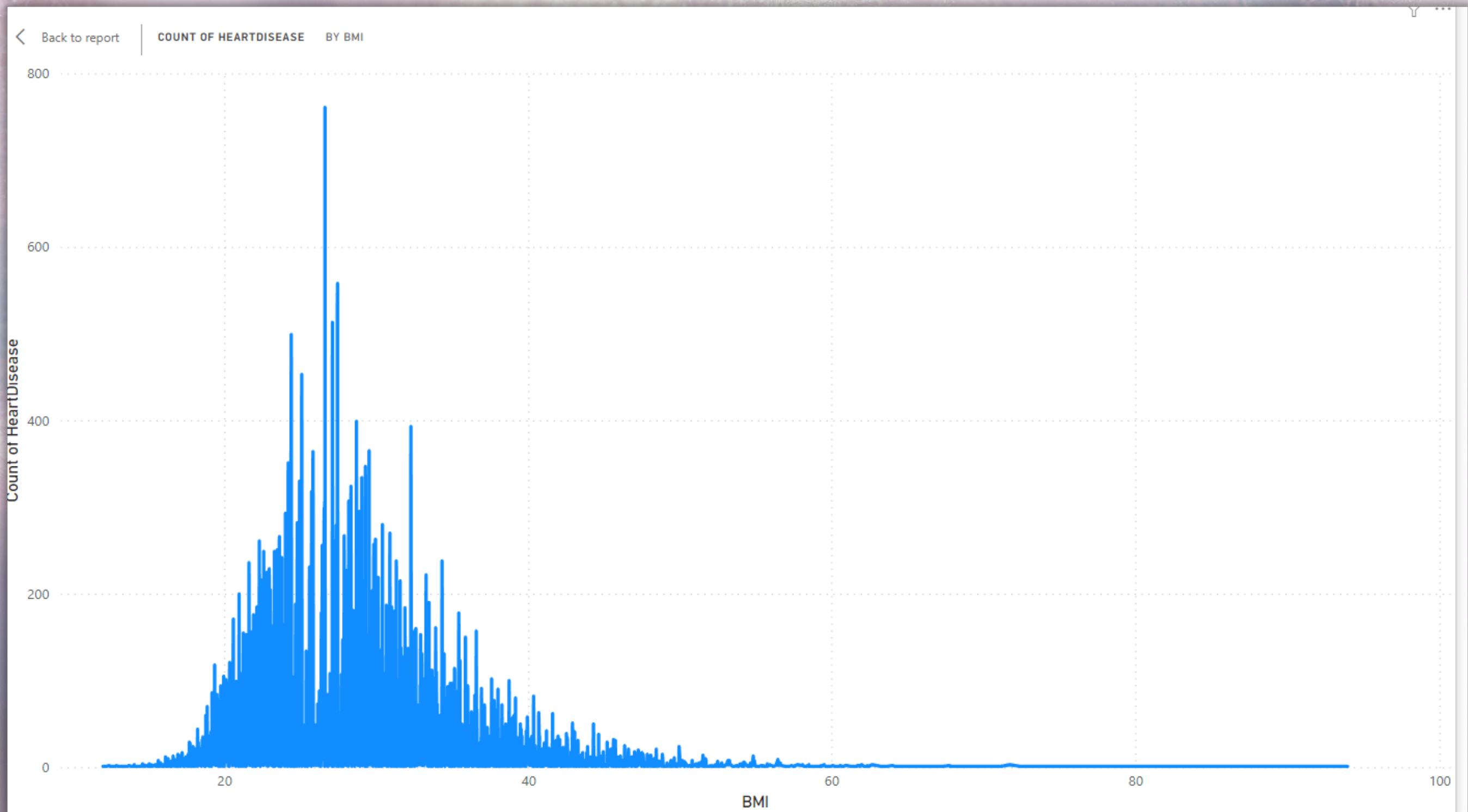
กราฟ roc ค่อนข้างดี

จากการทำนาย
เราได้ค่า Accuracy ถึง
0.916(91%)
มีค่า AUC = 0.841
ยิ่งเข้าใกล้ 1 ยิ่งดี

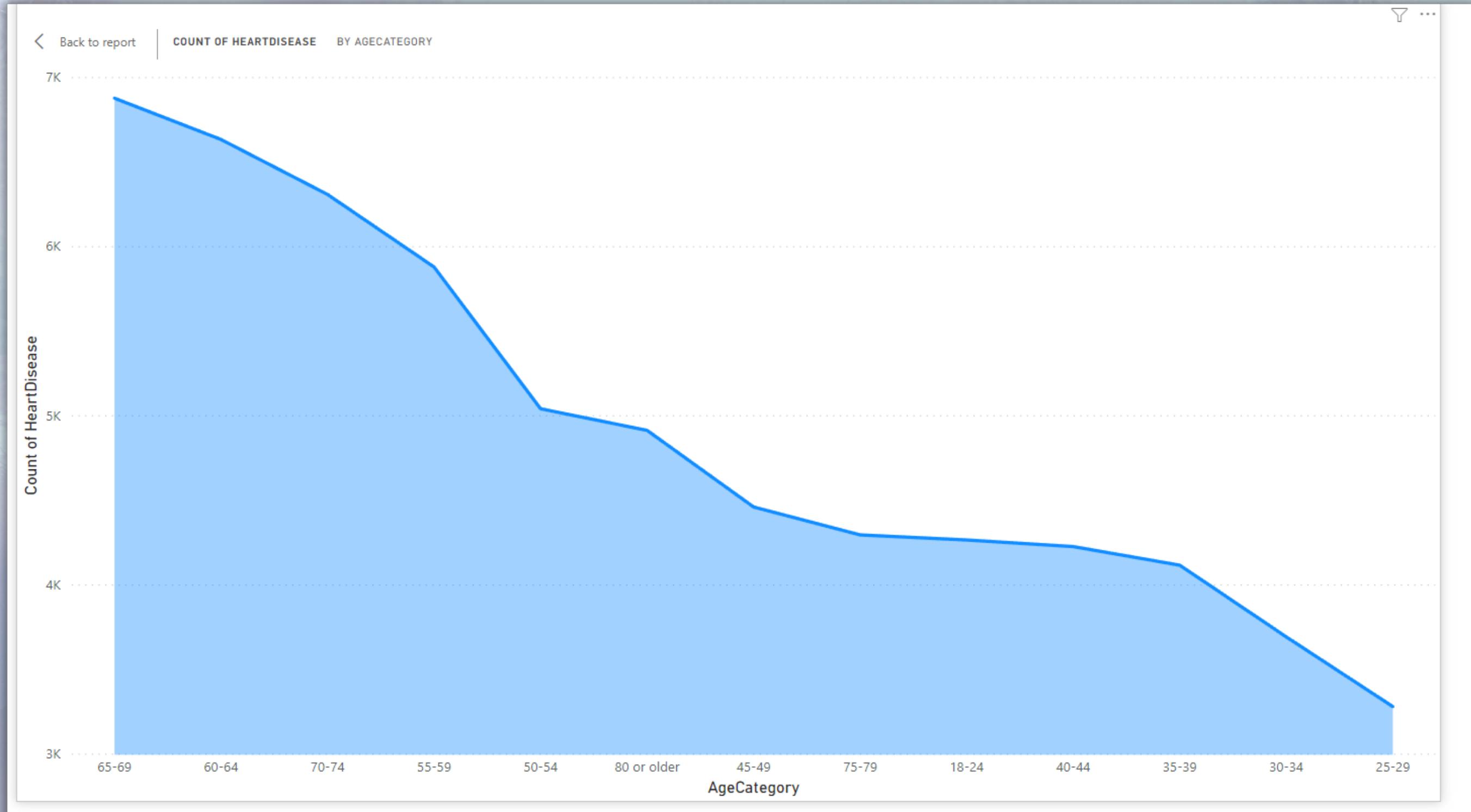


ค่า true positive และ true negative มีค่ามากกว่า ค่า false positive และ false negative

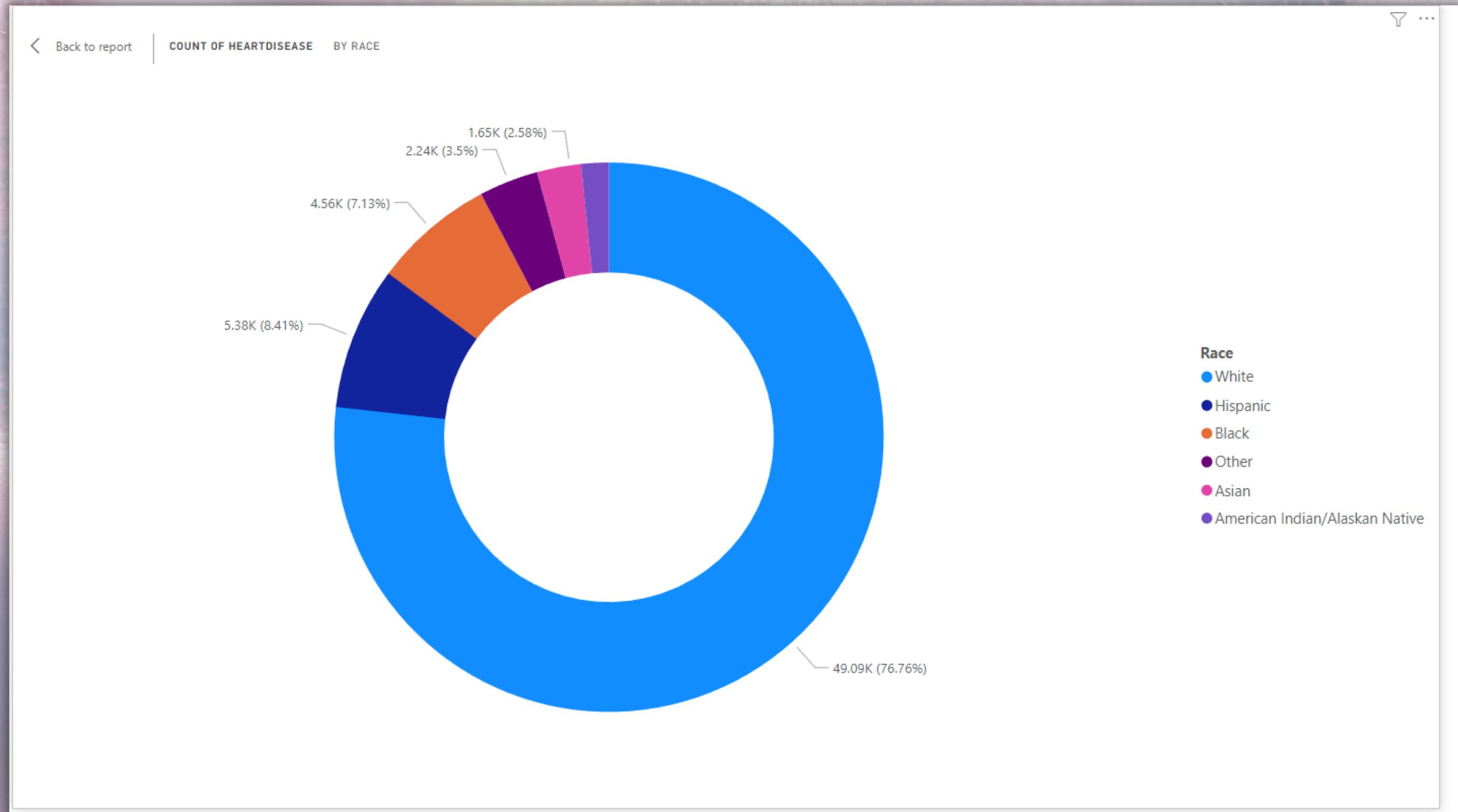
BMI



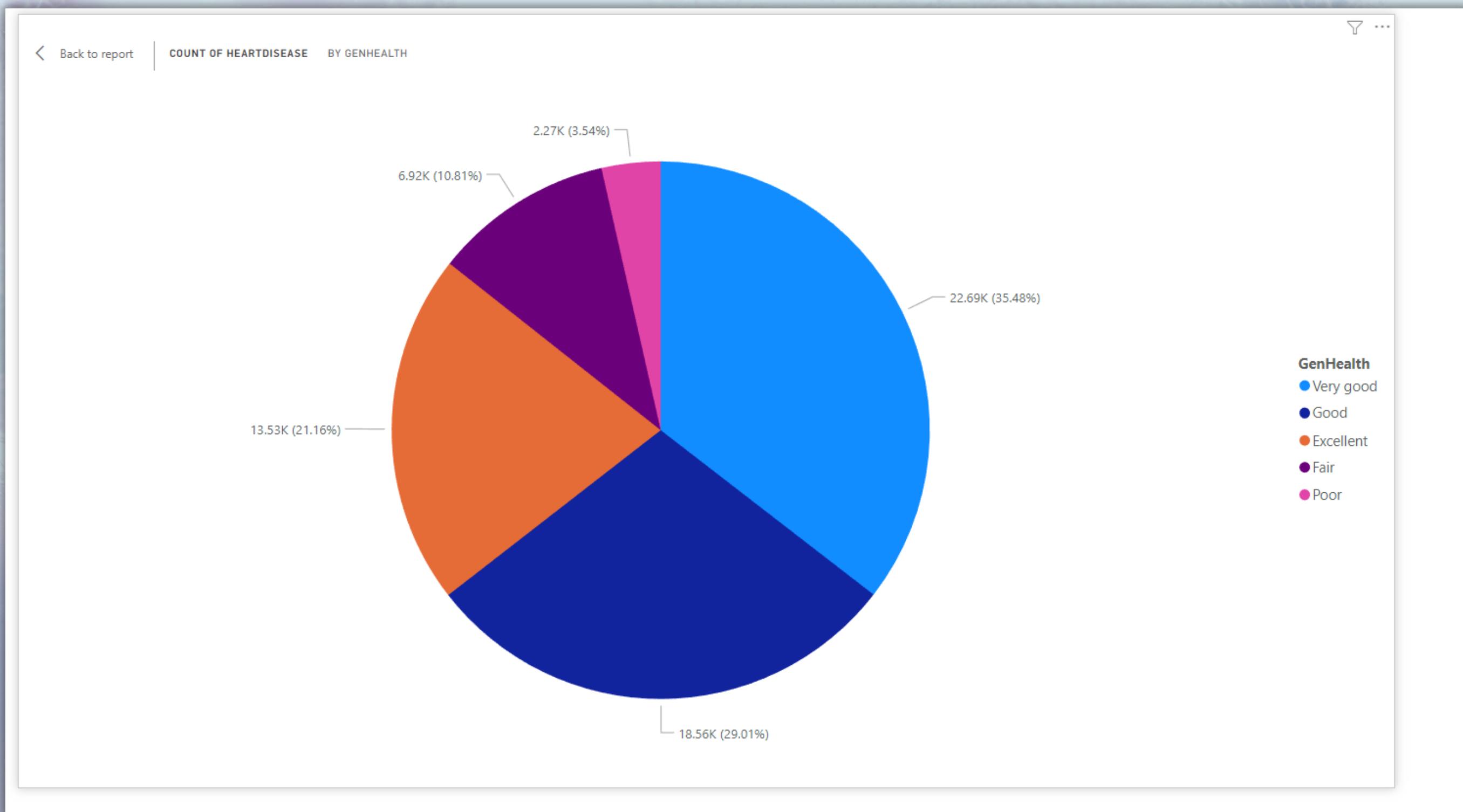
แสดงอายุที่มีการเกิดโรคหัวใจ



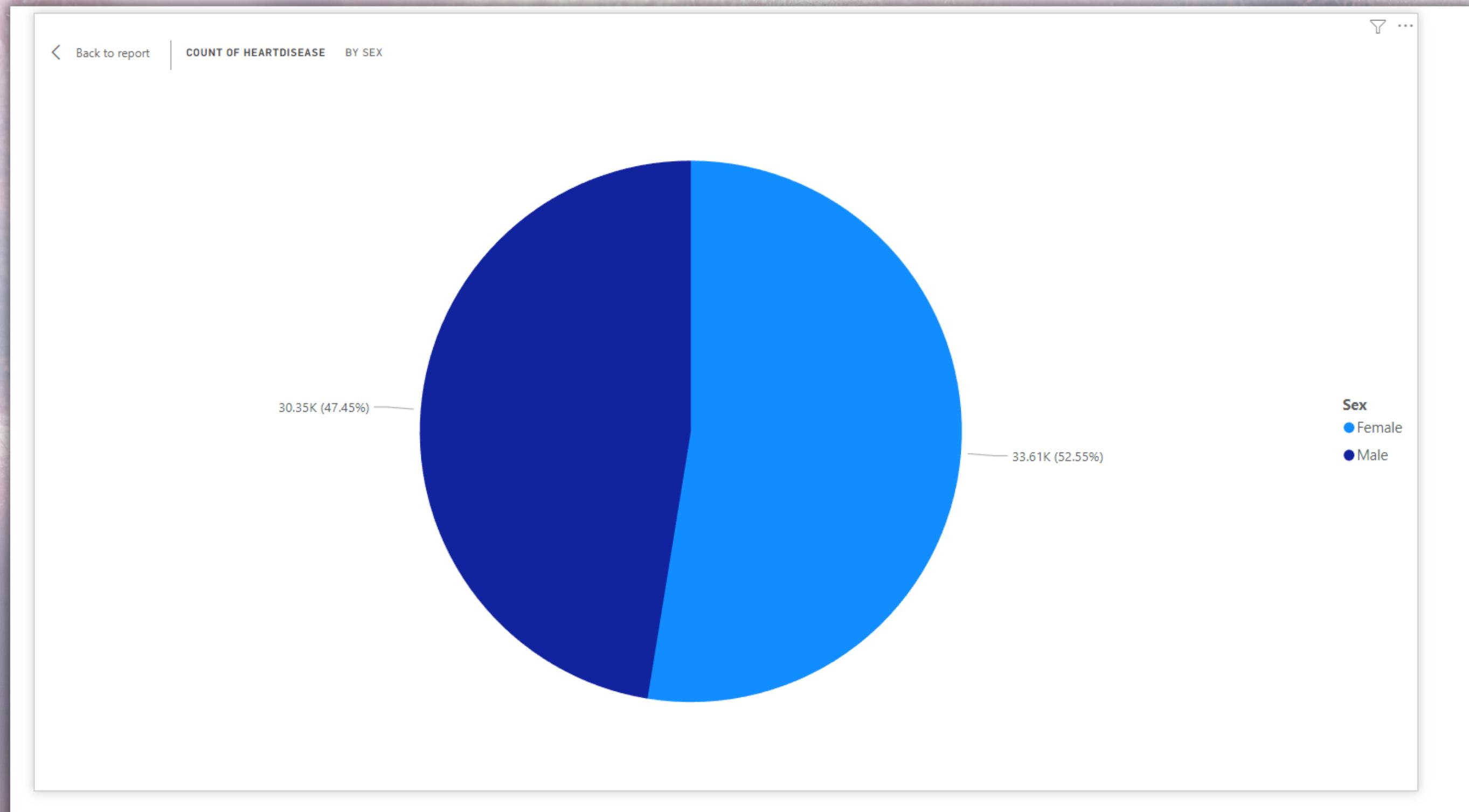
แสดงจำนวนชาติพันธุ์ที่เกิดโรคหัวใจ



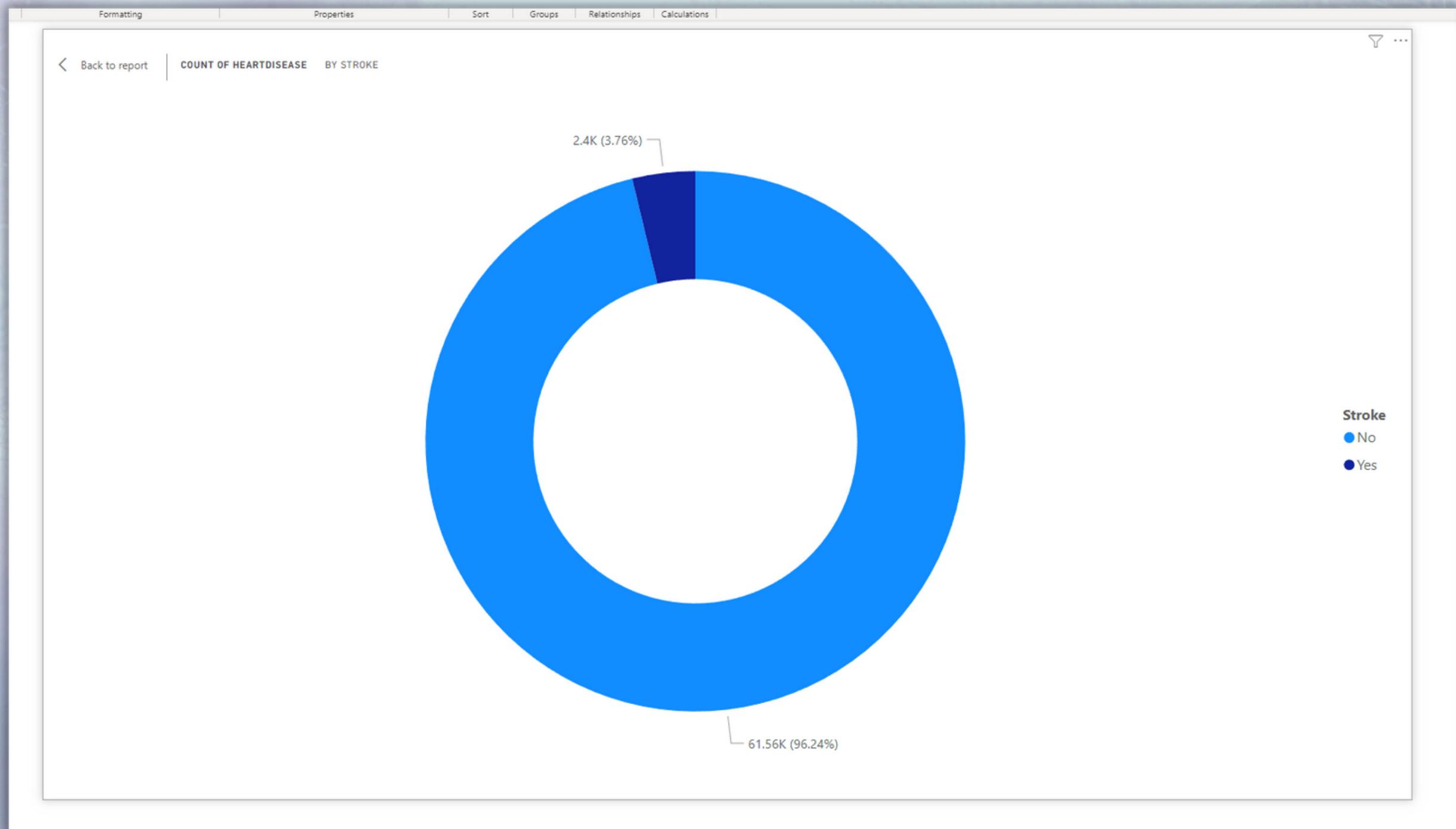
สุขภาพโดยทั่วไปของผู้ที่เป็นโรคหัวใจ



แสดงจำนวนเพศที่มีการเกิดโรคหัวใจ



แสดงจำนวนคนที่เป็นโรคหัวใจจากการมีโรคหลอดเลือดสมอง



สรุปผลการทดลอง

จากข้อมูลการ Train 80%(255836 ข้อมูล)

และ Test 20%(63959 ข้อมูล)

มีความแม่นยำ Accuracy(91%) และ

ค่า true positive และ true negative มีค่ามากกว่า ค่า false positive และ false negative มีค่า AUC ถึง 0.841 และ มีกราฟ roc ห่างจากเส้นตรงกลางค่อนข้างมาก

จากการทดลองโน้ตเดลของเรานั้นสามารถแก้ปัญหาผู้ป่วยที่มีความเสี่ยงของโรคภาวะหัวใจล้มเหลวได้ เพียงแค่กรอกข้อมูลเพียงเล็กน้อย ก็สามารถประเมินความเสี่ยงของโรคได้โดยไม่ต้องไปเสียเงินตรวจสุขภาพราคาแพง และค่าแม่นยำกว่า 90% ซึ่งทำให้ Model ที่ได้นั้น มีประโยชน์ต่อ ผู้ป่วยทำให้ผู้ป่วยสามารถรู้ตัวว่าเสี่ยงต่อโรคและรักษาได้กันท่วงที