

# Determining an Association between Family History of Cardiometabolic Disease and the Incidence of Preeclampsia in Pregnant Women of Different Racial and Age Groups using Artificial Intelligence and Machine Learning

Yara Yaghi, Omar Barabandi,  
Zahra Izzi, Ayham Elayan

# The Team

Yara Yaghi  
Sophomore  
George Mason University  
Computational Data Science

Ayham Elayan  
Junior  
George Mason University  
Bioengineering

Omar Barabandi  
Senior  
George Washington University  
Computer Science

Zahra Izzi  
Sophomore  
Georgetown University  
Healthcare Management  
and Policy

# Background

- Maternal cardiometabolic health observes mothers' health throughout pregnancy and childbirth
- Correlation between family history of cardiometabolic diseases with development of preeclampsia
  - hypertension, cholesterol, diabetes, cardiovascular disease
- Cardiometabolic Correlates Synthetic Dataset
  - Family History
  - Visits
  - Diagnosis
  - Person



# Main Factors

## **Hypertension/Cholesterol:**

High blood pressure (R03.0)

Hypertension (I10)

High  
cholesterol/Hypercholesterol  
emia (E78.0)

Hyperlipidemia (E78.5)

## **Diabetes Conditions:**

Diabetes/Diabetes Mellitus  
(E8-E13)

Diabetes Mellitus Type 1  
(E10)

## **Cardiovascular Diseases:**

Heart Attack/Myocardial  
Infarction (I21.9)

Stroke/Cerebral Infarction  
(I63.9)

Heart Disease (I51.9)

# Hypothesis

Minorities with family history in cardiovascular disease and aged 35 and older are more likely to develop preeclampsia during pregnancy.

# Lit Review: Family History of CVD

## 1. The Association of Familial Hypertension and Risk of Gestational Hypertension and Preeclampsia

- 912 women; 12.4% with GH, 2.6% with PE
- Paternal hypertension -> GH: OR = 1.98 (1.2-3.28)
  - increased with pre-pregnancy smoking or overweightness
- Maternal hypertension -> PE: OR = 3.26 (1.3-8.16)
  - increased with obesity and with non-smokers
- No relation to GH/PE in women with normal BMI

## 2. Family History of Hypertension, Cardiovascular Disease, or Diabetes and Risk of Developing Preeclampsia: A Systematic Review

- Observational study of 4 databases (MEDLINE, EMBASE, Cochrane library, CINAHL)
- Family history of CVD linked to increased risk of preeclampsia
  - Women with familial CVD - OR 1.7 (1.0-3.0)
  - Women with familial CVD in first-degree relatives - OR 2.82 (1.22-6.51)



# Lit Review: Racial Disparities

## 3. “Racial Disparities in Comorbidities, Complications, and Maternal and Fetal Outcomes in Women with Preeclampsia/Eclampsia”

- 1,175,046 weighted patient discharges with PE/eclampsia
  - 6.04% African American, 3.75% white, 2.58% Hispanic
    - 16.69% of AA, 9.33% of white, 8.53% of Hispanic with HT
- Increased risk of adverse pregnancy outcomes (maternal and infant mortality - OR 2.85 and 2.45) for AA women

## 4. Hypertensive Disorders in Pregnancy and Mortality at Delivery Hospitalization — United States, 2017–2019

- According to the CDC, the prevalence for hypertensive disorders in pregnant women hospitalized for delivery is highest for Black women (20.9%) and American Indian/Alaskan Native women (16.4%).

# Lit Review: Age Disparities

## 5. “Pregnancy at Age 35 Years or Older”

- Risk of preeclampsia rises with maternal age.
  - Women aged 40-44 have a 30% higher risk ( $RR = 1.32$ , 95% CI: 1.25–1.4) compared to those aged 35-39
  - Women aged 45-59, the risk more than doubles ( $RR = 2.21$ , 95% CI: 1.89–2.58)



# Methodology

Logistic Regression Analysis:

Objective: Model the relationship between family history of cardiometabolic conditions and preeclampsia incidence, adjusting for confounders.

# Methodology Data

## PHASE 1

Data  
Exploration  
and  
Preparation

Import and examine datasets (`new_person`, `new_family_hx`, `new_diagnosis`, `new_visits`).

Create binary features for family history conditions (e.g., hypertension, diabetes).

Analyze distributions for `RACE`, `AGE_GROUP`, `FACILITY_TYPE`, etc.

## PHASE 2

Data Merging  
& Final  
Preparation

Combine datasets on `PERSON_ID` to form a complete feature set.

Apply `pd.get_dummies()` to categorical variables (e.g., `REASON_FOR_VISIT`, `FACILITY_TYPE`).

Define `Preeclampsia` based on relevant conditions from the diagnosis data.

## PHASE 3

Model  
Training &  
Evaluation

Split data into training and test sets using `train_test_split()`.

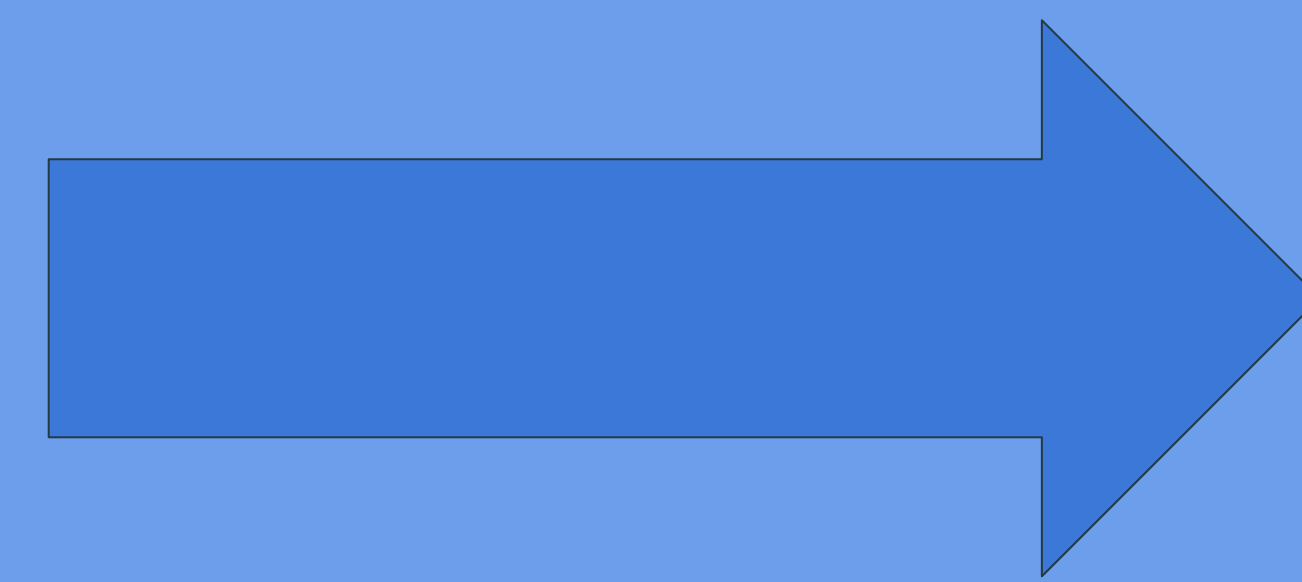
Train models (Logistic Regression, XGBoost, RandomForest, MLP) on the training set.

Assess model performance using accuracy, ROC AUC, and classification reports.



# Methodology

- Person Dataset: 32,193
- Visits Dataset: 205,742
- Diagnosis Dataset: 1,227,375
- Family History Dataset: 59,012



- Person Dataset: 15,229
- Visits Dataset: 15,229
- Diagnosis Dataset: 15,229
- Family History Dataset: 15,229

## Predictor Variables:

- Person ID
- Race
- Age
- Ethnicity
- Language
- Marital Status
- Zip Type
- Admit Type
- Facility Type
- Insurance Type
- Encounter Type
- Reason For Visit (Top 50)
- Family History
  - Hypertension/Cholesterol
  - Diabetes
  - Cardiovascular Disease

## Dependant Variables:

- Preeclampsia
- Hypertension
- High Blood Pressure

70%

Training data used to build a probability model between the predictor and dependent variables.

30%

Testing data to evaluate the performance of the model.

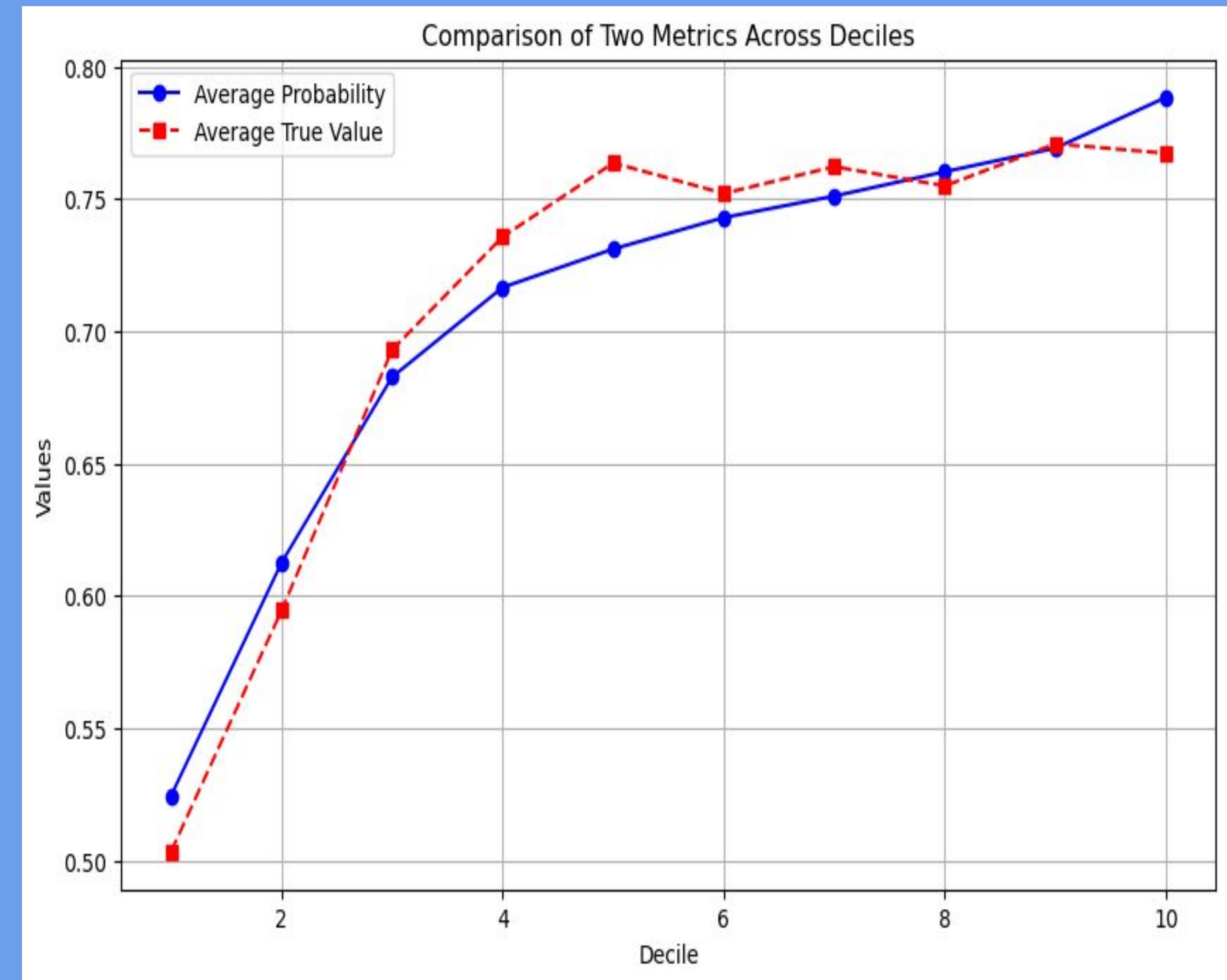
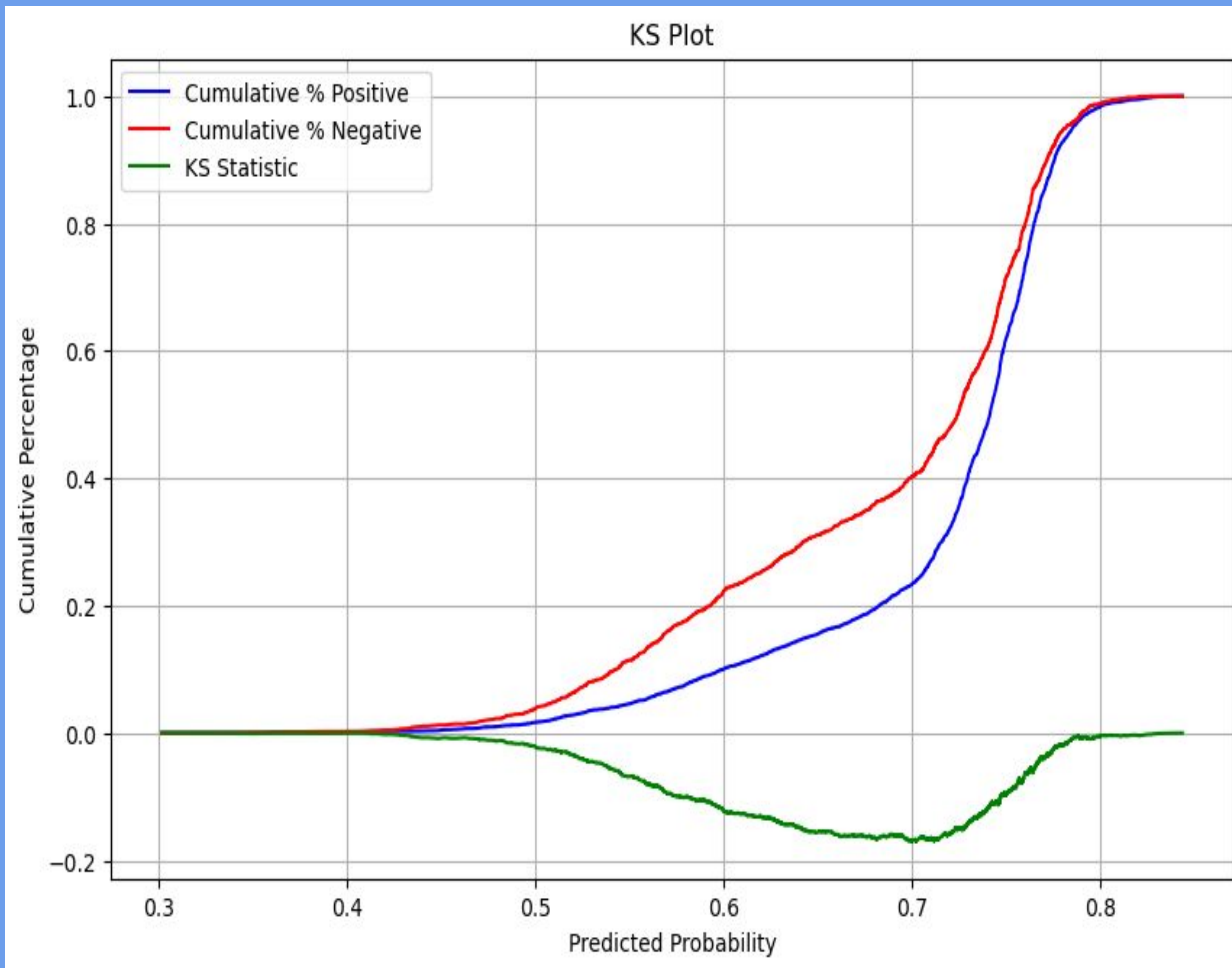


# Results From Data Model

	precision	recall	f1-score	support
0	0.47	0.07	0.12	1769
1	0.72	0.97	0.82	4323
accuracy			0.71	6092
macro avg	0.60	0.52	0.47	6092
weighted avg	0.65	0.71	0.62	6092



# Results From Data Model





# Results

Rates for Patients with Family History Cardiometabolic Risk and Incidence of Preeclampsia for Race

Race	Total Count	Preeclampsia Count	Preeclampsia Rate
White	4699	2597	55.267078
African American	2924	1733	59.268126
Other	1718	991	57.683353
Unknown	527	289	54.838710
Asian	512	306	59.765625

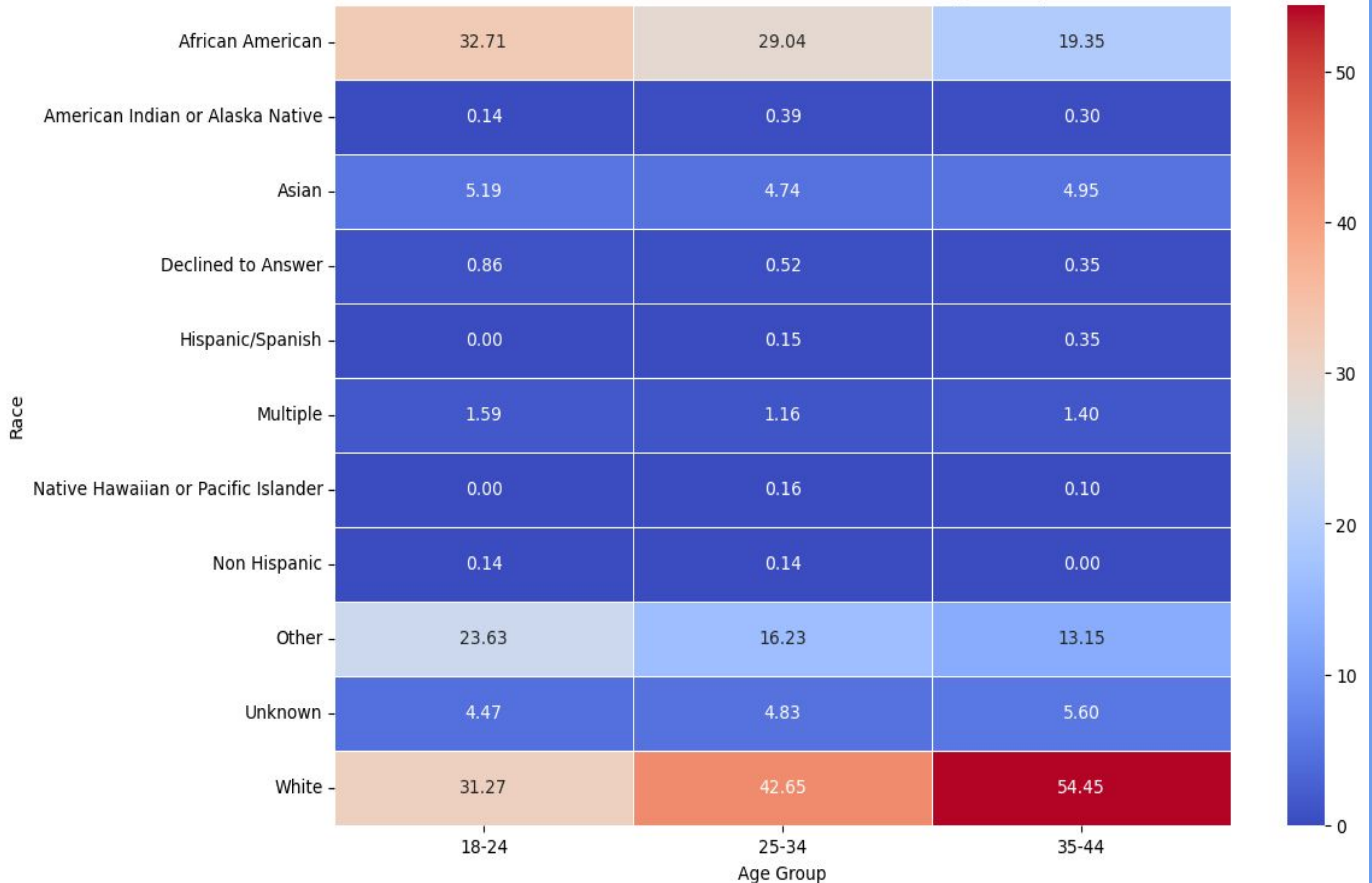
Rates for Patients with Family History Cardiometabolic Risk and Incidence of Preeclampsia for Age

Age Group	Total Count	Preeclampsia Count	Preeclampsia Rate
18-24	694	412	59.365994
25-34	7955	4555	57.259585
35-44	2000	1110	55.500000



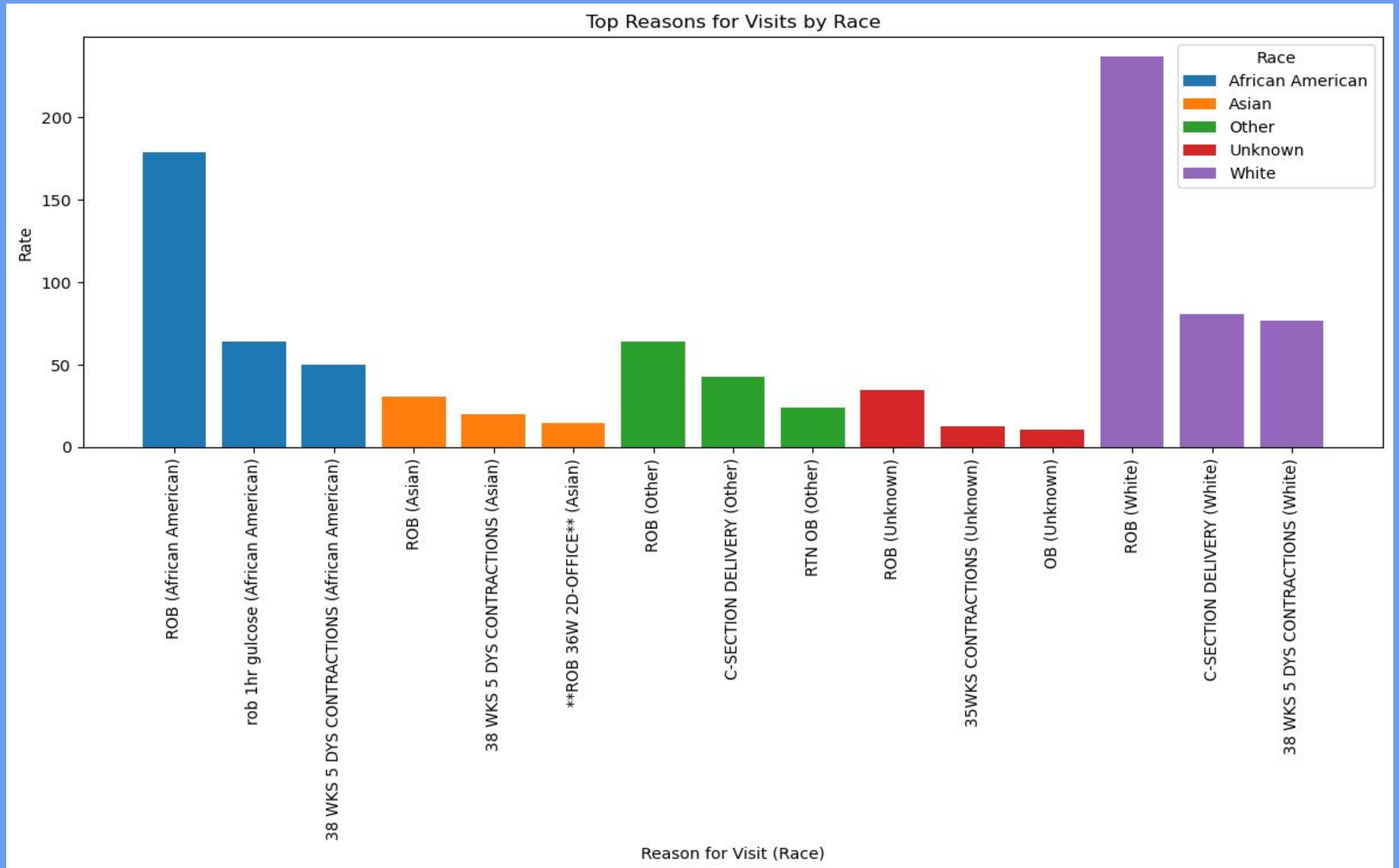
# Results

Cross-tabulation Rates between Race and Age Group





# Results





# Conclusion

Our analysis has highlighted significant variations in preeclampsia rates across different racial and age groups, particularly the decrease in rates for African American as they age, contrasted with an increase for White women. These findings emphasize the critical role that healthcare visit patterns, such as routine obstetric care and delivery-related visits, play in indicating health outcomes.

For future studies we suggest consider exploring the underlying causes of the observed differences in preeclampsia rates across racial and age groups. Additionally, expanding the study to include a larger and more diverse sample could help to validate these findings and identify other factors that may contribute to preeclampsia risk.



# Determining an Association between Family History of Cardiometabolic Disease and the Incidence of Preeclampsia in Pregnant Women of Different Racial and Age Groups using Artificial Intelligence and Machine Learning



<sup>1</sup> MedStar-Georgetown Collaborative Center for Artificial Intelligence in Healthcare Research and Education (AI CoLab)



## BACKGROUND AND OBJECTIVE:

Maternal cardiometabolic health encompasses the health of women pre, during, and post childbirth, and is important to monitor throughout the course of pregnancy. In order to pinpoint risk factors for various cardiometabolic diseases in mothers, we researched the correlation between mothers with a family history of cardiometabolic diseases, such as hypertension, cholesterol, diabetes, and heart failure, and looked for an association with mothers who develop preeclampsia. In order to conduct this, we used the Cardiometabolic Correlates Synthetic Dataset, specifically the Family History dataset and the New Event dataset.

## METHODS AND ANALYSIS:

We used four key datasets: the Person dataset, the Visits dataset, the Diagnosis dataset, and the Family History dataset. After cleaning the data by removing duplicates and ensuring only unique variables were included, we merged these datasets. This data cleaning and merging process left us with a final cohort of 15,229 unique patients. For our analysis, we identified several predictor variables, including demographic factors such as race, age, ethnicity, language, and marital status, as well as clinical variables including admit type, facility type, insurance type, and encounter type. We also included specific family history conditions such as hypertension, diabetes, and cardiovascular disease, and the top 50 reasons for visits. Our primary outcomes of interest were preeclampsia, hypertension, and high blood pressure. We then split our data, using 70% to train our logistic regression model and reserving 30% for testing the model's performance. This approach ensured that the model we developed can generalize well to new data, allowing us to predict health outcomes effectively in the population.

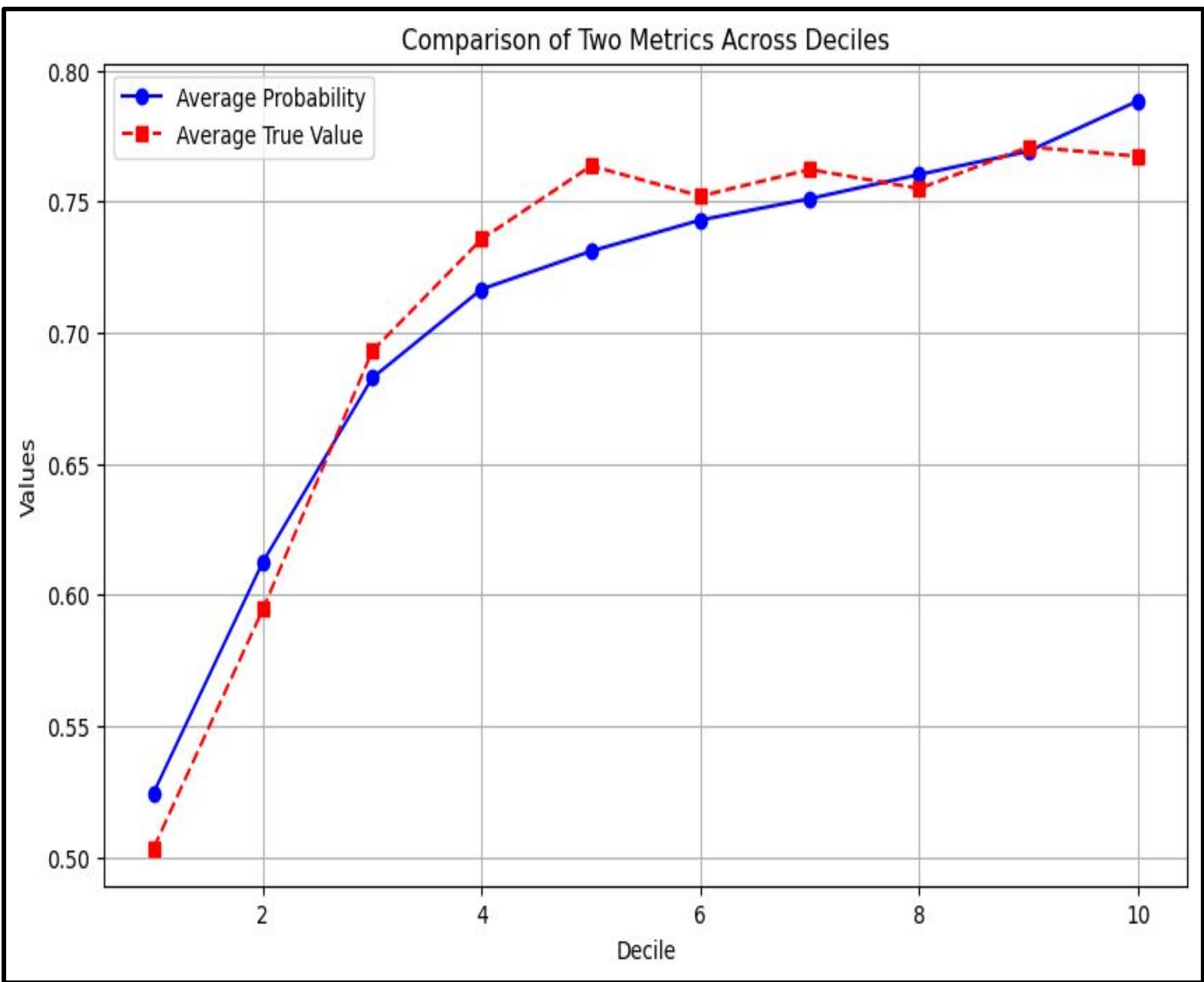


Figure 1. Comparison of two metrics across deciles

Race	Total Count	Preeclampsia Count	Preeclampsia Rate
White	4699	2597	55.267078
African American	2924	1733	59.268126
Other	1718	991	57.683353
Unknown	527	289	54.838710
Asian	512	306	59.765625

Figure 2. Rates for Patients with Family History Cardiometabolic Risk and Incidence of Preeclampsia for Race

Age Group	Total Count	Preeclampsia Count	Preeclampsia Rate
18-24	694	412	59.365994
25-34	7955	4555	57.259585
35-44	2000	1110	55.500000

Figure 3. Rates for Patients with Family History Cardiometabolic Risk and Incidence of Preeclampsia for Age

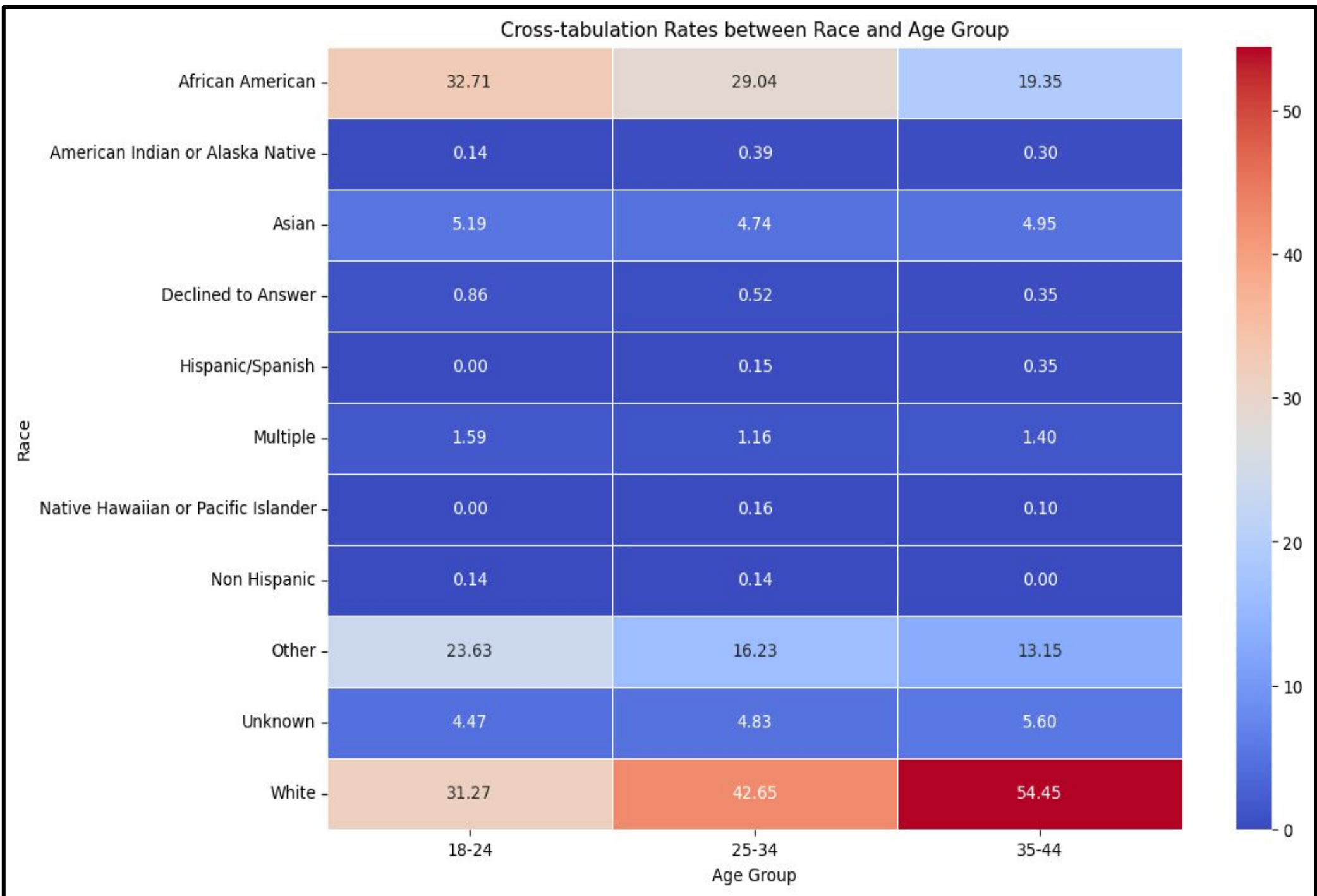


Figure 4. Cross-tabulation rates between race and age groups

## RESULTS

**Model Accuracy:** The logistic regression model achieved 71% accuracy, with good discrimination and calibration, especially at higher probabilities.

**Racial Impact:** African American, Asian, and 'Other' racial groups have a 4-5% higher preeclampsia risk than White patients.

**Age Trends:** Younger women (18-24) face the highest preeclampsia risk (~59%), which decreases with age, with the lowest rate (~55.5%) in the 35-44 age group.

**Visit Reasons:** Routine obstetric care dominates across races, but higher cesarean delivery rates in White women may explain their increased preeclampsia risk with age.

## CONCLUSION

Our analysis has highlighted significant variations in preeclampsia rates across different racial and age groups, particularly the decrease in rates for African American as they age, contrasted with an increase for White women. These findings emphasize the critical role that healthcare visit patterns, such as routine obstetric care and delivery-related visits, play in indicating health outcomes. For future studies we suggest consider exploring the underlying causes of the observed differences in preeclampsia rates across racial and age groups. Additionally, expanding the study to include a larger and more diverse sample could help to validate these findings and identify other factors that may contribute to preeclampsia risk.



# Works Cited

1. Lewandowska, Małgorzata. (2021). The Association of Familial Hypertension and Risk of Gestational Hypertension and Preeclampsia. *International Journal of Environmental Research and Public Health*. 18. 7045.  
10.3390/ijerph18137045.
2. Kay, V. R., Wedel, N., & Smith, G. N. (2021). Family History of Hypertension, Cardiovascular Disease, or Diabetes and Risk of Developing Preeclampsia: A Systematic Review. *Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC*, 43(2), 227–236.e19.  
<https://doi.org/10.1016/j.jogc.2020.08.010>
3. Shahul, S., Tung, A., Minhaj, M., Nizamuddin, J., Wenger, J., Mahmood, E., Mueller, A., Shaefi, S., Scavone, B., Kociol, R. D., Talmor, D., & Rana, S. (2015). Racial Disparities in Comorbidities, Complications, and Maternal and Fetal Outcomes in Women with Preeclampsia/Eclampsia. *Hypertens Pregnancy*, 34(4), 506–515.
4. Ford ND, Cox S, Ko JY, et al. Hypertensive Disorders in Pregnancy and Mortality at Delivery Hospitalization — United States, 2017–2019. *MMWR Morb Mortal Wkly Rep* 2022;71:585–591. DOI:  
<http://dx.doi.org/10.15585/mmwr.mm7117a1>
5. American College of Obstetricians and Gynecologists. (2022, August). Pregnancy at Age 35 Years or Older.