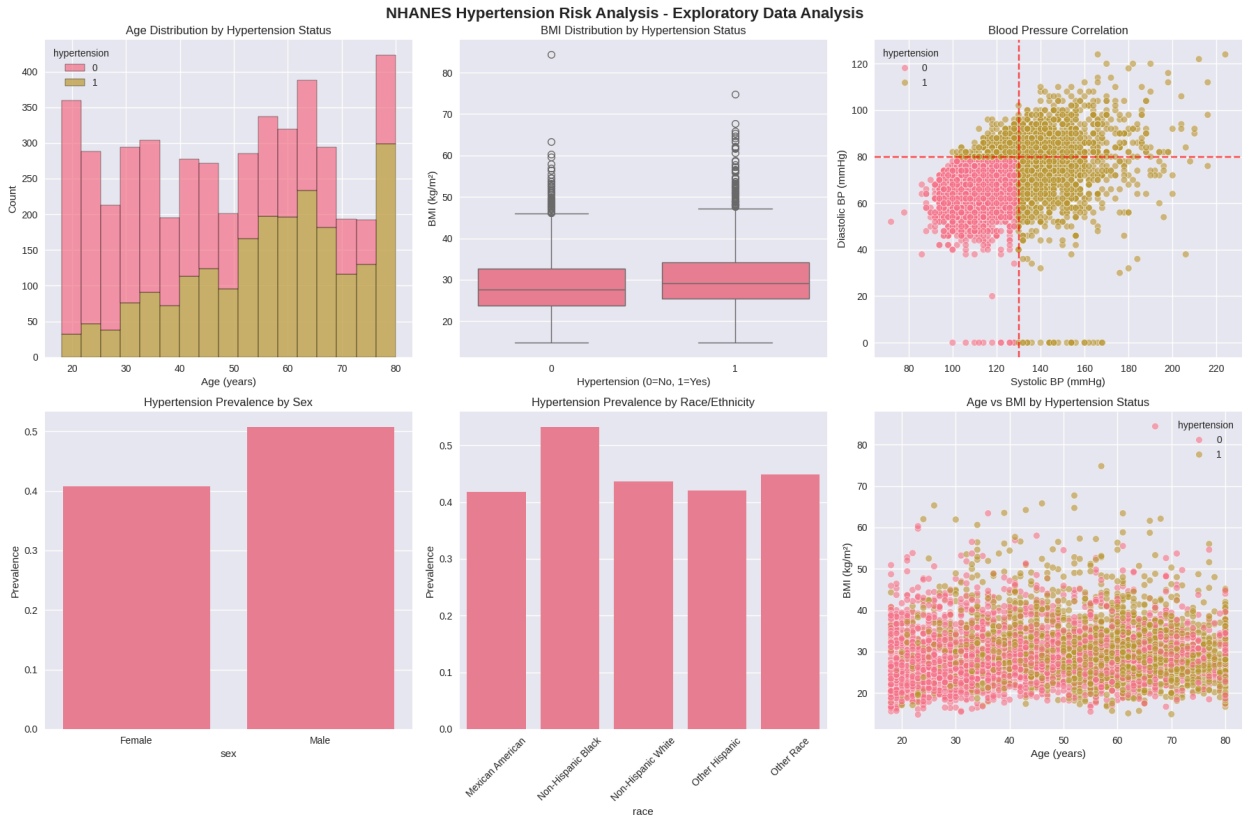
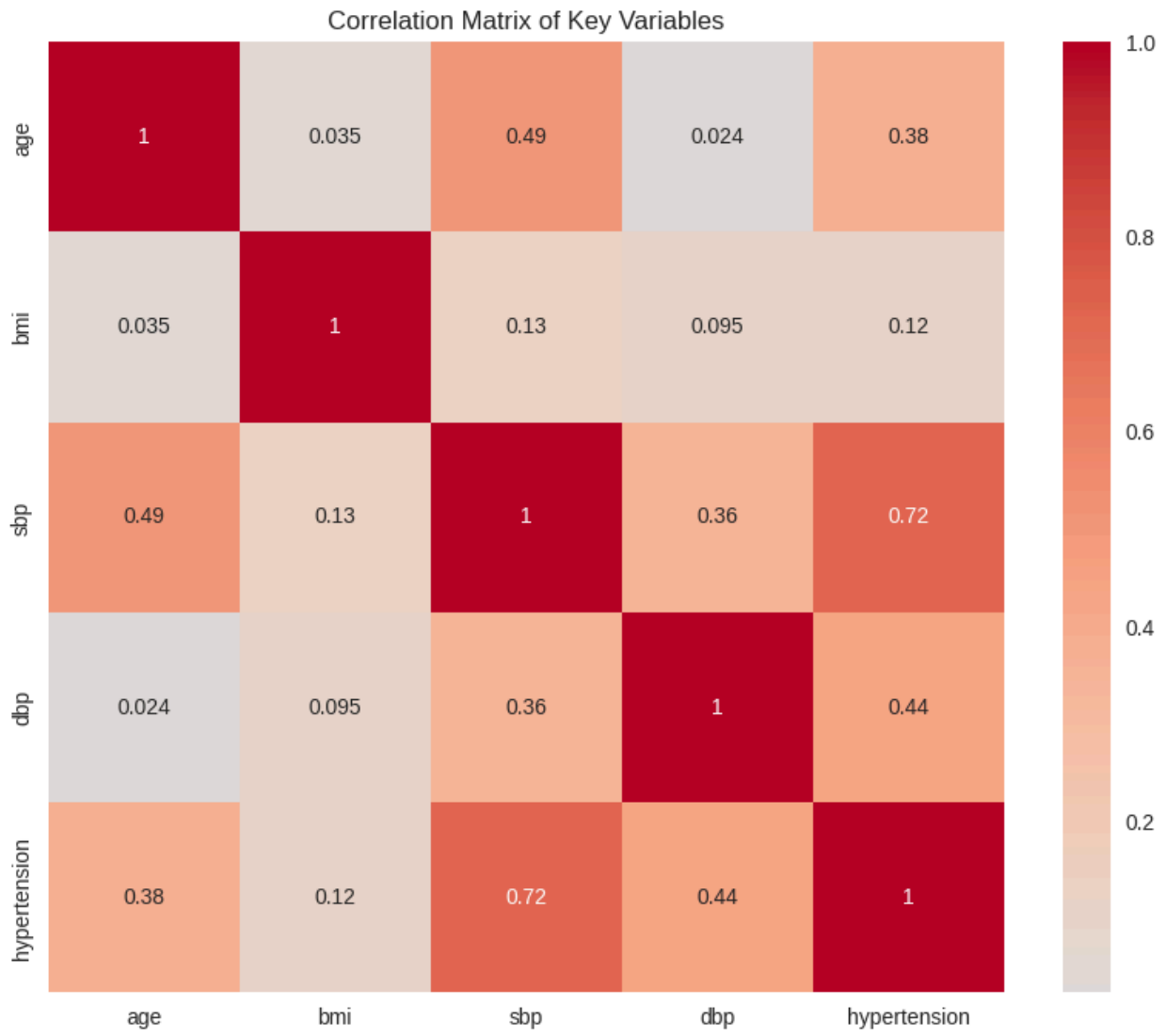
```
============================================================
NHANES HYPERTENSION RISK ANALYSIS
============================================================

==================== PHASE 1 ====================
Loading NHANES datasets...
Demographics: 9254 participants
Blood pressure: 8704 participants
Body measures: 8704 participants
Merged dataset: 8704 participants
Cleaning and preparing data...
Removed 700 participants with missing data
Final analysis sample: 4833 participants
Performing exploratory data analysis...

=== DESCRIPTIVE STATISTICS ===
              age          bmi          sbp          dbp
count  4833.000000  4833.000000  4833.000000  4.833000e+03
mean     49.457687    29.588537   125.793503  7.164287e+01
std      18.602931     7.296148    19.463786  1.348849e+01
min      18.000000    14.800000    72.000000  5.397605e-79
25%      33.000000    24.500000   112.000000  6.400000e+01
50%      51.000000    28.400000   124.000000  7.200000e+01
75%      64.000000    33.400000   136.000000  8.000000e+01
max      80.000000    84.400000   224.000000  1.240000e+02

=== HYPERTENSION PREVALENCE ===
Overall prevalence: 45.7%
By age group:
  18-39: 23.7%
  40-59: 51.8%
  60+: 64.6%
By sex:
  Female: 40.8%
  Male: 50.7%
By race/ethnicity:
  Mexican American: 41.8%
  Non-Hispanic Black: 53.2%
  Non-Hispanic White: 43.7%
  Other Hispanic: 42.0%
  Other Race: 44.9%
```


NHANES Hypertension Risk Analysis - Exploratory Data Analysis
Age Distribution by Hypertension Status
hypertension
0
1
Count
Age (years)
BMI Distribution by Hypertension Status
BMI (kg/m²)
Hypertension (0=No, 1=Yes)
Blood Pressure Correlation
hypertension
0
1
Diastolic BP (mmHg)
Systolic BP (mmHg)
Hypertension Prevalence by Sex
Prevalence
Female
Male
sex
Hypertension Prevalence by Race/Ethnicity
Prevalence
Mexican American
Non-Hispanic Black
Non-Hispanic White
Other Hispanic
Other Race
race
Age vs BMI by Hypertension Status
hypertension
0
1
BMI (kg/m²)
Age (years)



========== PHASE 2 ==========
Fitting logistic regression models...

Fitting MODEL_A...
Sample size: 4833
Features: ['age', 'bmi']
Hypertension prevalence: 0.457
AUC: 0.754
Accuracy: 0.696
CV AUC: 0.727 ± 0.021

Fitting MODEL_B...
Sample size: 4833
Features: ['age', 'bmi', 'sex_Male', 'race_Non-Hispanic Black', 'race_Non-Hispanic White', 'race_Other Hispanic', 'race_Other Race']

Hypertension prevalence: 0.457
AUC: 0.771
Accuracy: 0.694
CV AUC: 0.739 ± 0.016

Fitting MODEL_C...
Skipping MODEL_C - No data after filtering.

=== MODEL EVALUATION RESULTS ===

Model	AUC	Accuracy	CV AUC	Features
MODEL_A	0.754	0.696	0.727 ± 0.021	2
MODEL_B	0.771	0.694	0.739 ± 0.016	7

=== MODEL DIAGNOSTICS ===

MODEL_A Diagnostics:
Variance Inflation Factors:

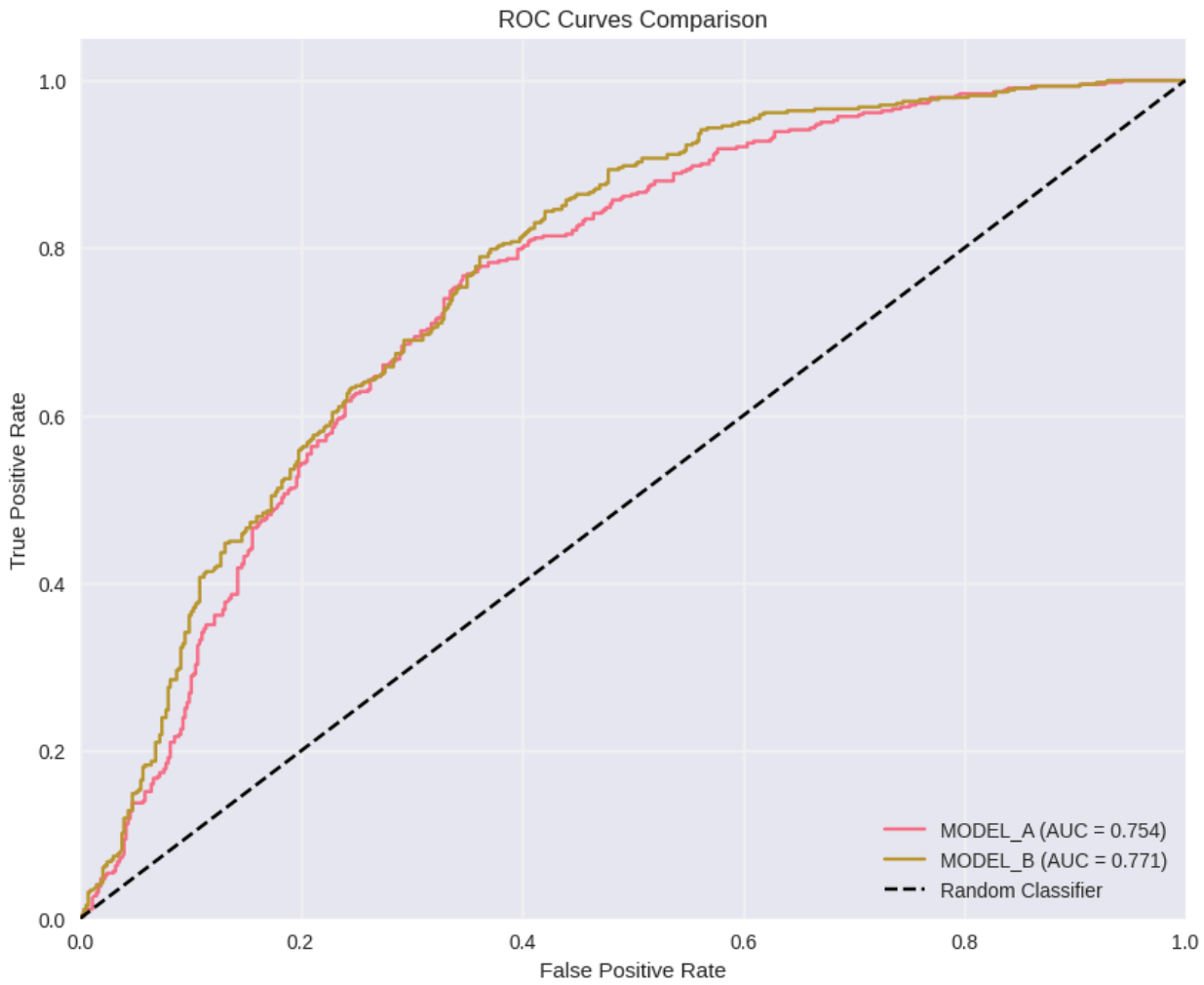
Feature	VIF
age	5.898102
bmi	5.898102

Model Summary:
Log-Likelihood: -2362.846
AIC: 4731.692
BIC: 4750.471

MODEL_B Diagnostics:
Variance Inflation Factors:

Feature	VIF
age	6.834831
bmi	8.018980
sex_Male	1.892862
race_Non-Hispanic Black	2.399830
race_Non-Hispanic White	3.092646
race_Other Hispanic	1.508864
race_Other Race	1.957358

Model Summary:
Log-Likelihood: -2322.746
AIC: 4661.493
BIC: 4711.573



=== STATISTICAL TESTS ===
Best performing model: MODEL_B
Best AUC: 0.771

=== MODEL_B INTERPRETATION ===
Coefficients and Odds Ratios:

Variable	Coefficient	OR	CI_Lower	CI_Upper	p_value
age	0.046	1.047	1.043	1.051	0.000
bmi	0.037	1.038	1.028	1.048	0.000
sex_Male	0.446	1.562	1.359	1.795	0.000
race_Non-Hispanic Black	0.325	1.384	1.092	1.753	0.007
race_Non-Hispanic White	-0.230	0.794	0.635	0.994	0.044
race_Other Hispanic	-0.089	0.914	0.680	1.230	0.554
race_Other Race	0.231	1.259	0.984	1.612	0.067

Clinical Interpretation:
- Age: Each additional year increases odds of hypertension by 4.7%

- BMI: Each unit increase in BMI increases odds by 3.8%
- Sex: Males have 56.2% higher odds compared to females

=== DECISION MATRIX ===

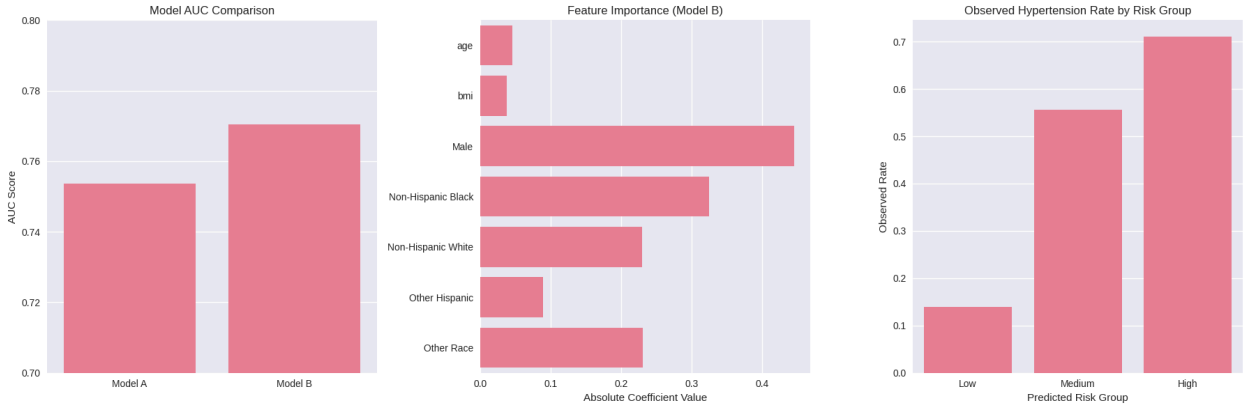
	AUC	Variable_Significance	Multicollinearity	Interpretability	\
model_a	3	5	5	5	
model_b	4	4	4	4	
model_c	4	3	3	3	

	Clinical_Use	Total_Score
model_a	4	22
model_b	5	21
model_c	3	16

Recommended model: MODEL_A
Total score: 22

========== PHASE 3 ==========

Example Prediction:
45-year-old Non-Hispanic White male with BMI 28:
Predicted hypertension probability: 0.373



Results exported to nhanes_hypertension_results.xlsx

==========
ANALYSIS COMPLETE
==========