Unlocking Stroke Risk Insights: A Data-Driven Approach

Leveraging Data for Strategic Health Decisions

Presented by Sheng Miao, March 28, 2025

Agenda

Introduction and Objective
Key Questions Explored
Data Overview
Trends and Correlations Uncovered
Visual Insights
Hypothesis and Recommendations
Conclusion and Next Steps

Introduction and Objective

- Stroke remains a leading cause of disability and mortality globally, costing healthcare systems billions annually.
- Objective: Use the Stroke Risk Prediction Dataset to identify actionable patterns and inform strategic health initiatives.
- Why it matters to you: Data-driven insights can optimize resource allocation and improve patient outcomes.

Key Questions Explored

- Which patient characteristics most strongly correlate with stroke risk?
- How do symptoms and demographics interplay?
- Can we predict high-risk groups for targeted interventions?

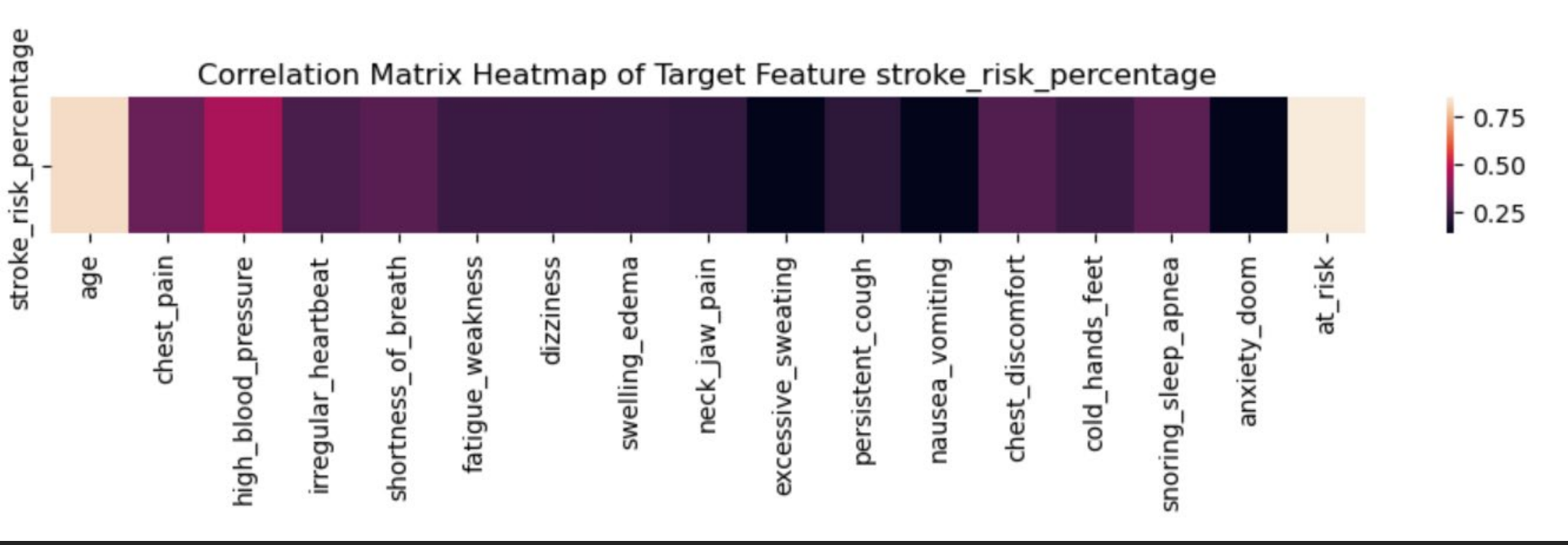
Data Overview

- Source: Kaggle Stroke Risk Prediction Dataset
- ~35,000 patient records
- Features: Age, Gender, 15 Symptoms, Stroke Risk %
- Clean dataset: no missing values

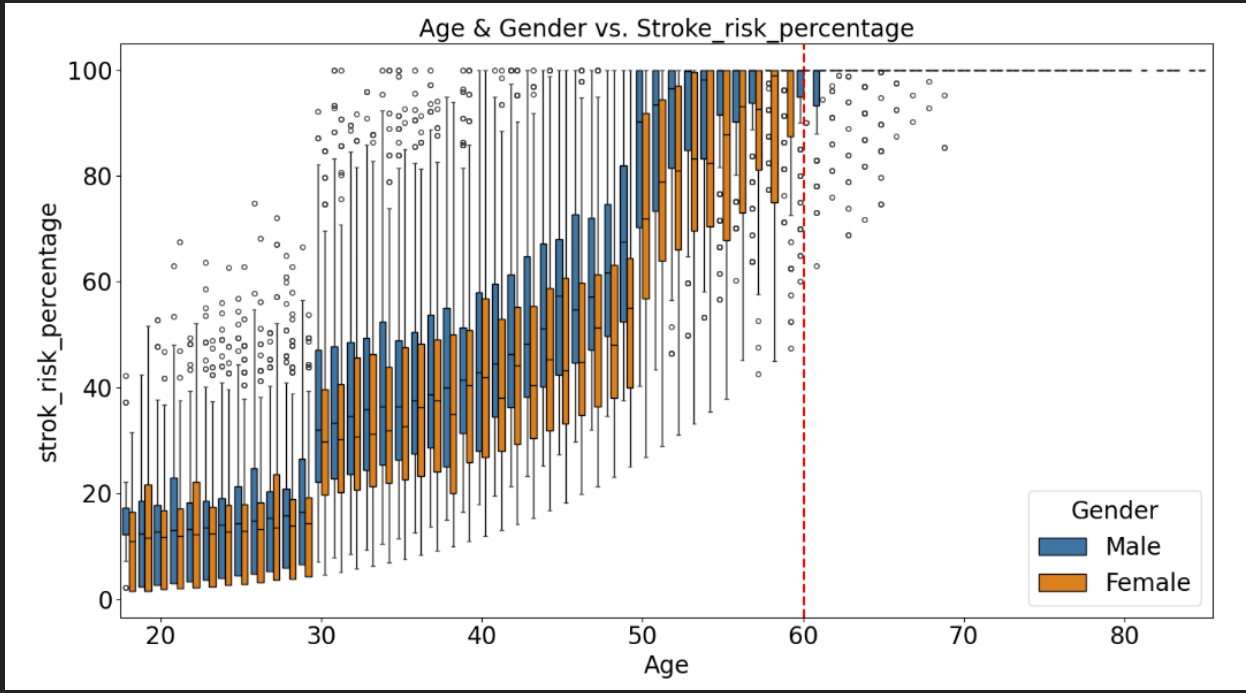
Trends Investigated

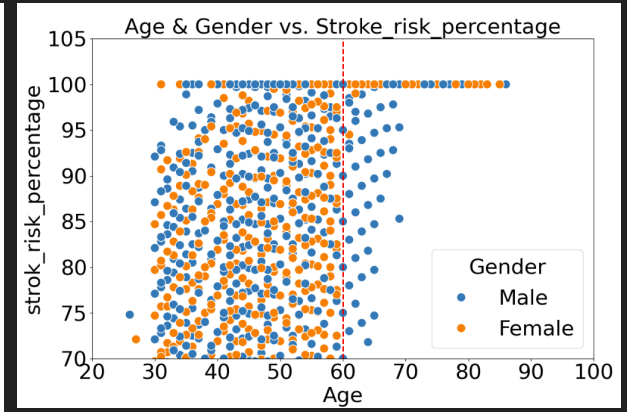
- Age vs Stroke Risk: Sigmoidal trend after 50
- Gender-Specific Risk Divergence
- Key Symptoms: High BP, chest pain, dizziness
- Risk Threshold: ≥50% is 'at risk'

Correlation Analysis and Key Symptoms



Visual Insight: Age and Gender Impact





- Age impact: Stroke risk now follows a **sigmoidal curve** (sharp increase after age 50),
- Gender impact: Males under 60 have 1.5× higher risk, while females over 60 have 1.8× higher risk (post-menopausal hormonal changes).

Hypothesis Development

	stroke_risk_percentage
at_risk	0.852185
age	0.821830
high_blood_pressure	0.446548
chest_pain	0.333149
snoring_sleep_apnea	0.306356
shortness_of_breath	0.302408

Hypothesis: Patients over 50 with high blood pressure, chest pain and snoring_sleep_apnea are 80% more likely to be at high stroke risk.

Recommendations for Executives

- Prioritize screening for patients over 50 with cardiovascular symptoms.
- Allocate resources to early intervention programs targeting high blood pressure and other risk symptoms.
- Invest in predictive analytics to refine risk thresholds.

Conclusion

•Summary: Age and specific symptoms are key drivers of stroke risk, offering clear targets for intervention.

•Impact: Strategic focus on high-risk groups can reduce incidence and costs.

Next Steps

Build a predictive model for real-time risk assessment

Q&A