TYPE 2 DIABETES AI AGENT

CORE CONCEPTS

1. Business Understanding
2. Data Overview
3. EDA Highlights
4. Modeling Approach
5. Results & Findings
6. Recommendations
7. Next Steps

- FRANKLINE ONDIEKI
- MATHEWS ODONGO
- PACIFICAH KWAMBOKA
- NIGHTINGALE JEPTOO
- TINAH MWIKALI NGEI
- DIANA HELLEN MACHARIA

TYPE 2 DIABETES

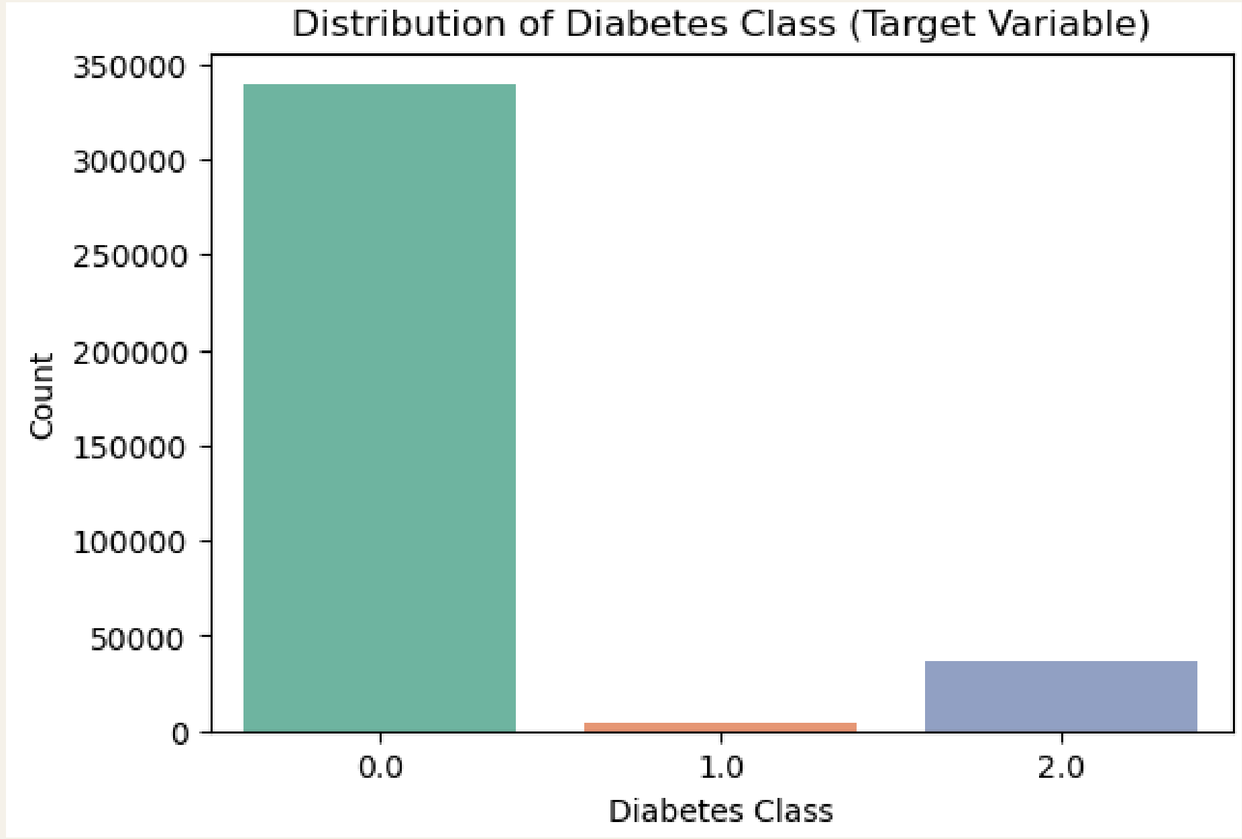
This is a chronic condition where the body resists insulin or produces too little of it, causing high blood sugar. It develops gradually, often due to obesity, poor diet, inactivity, or genetics, and can lead to complications like heart disease, kidney failure, and vision loss if not managed early.

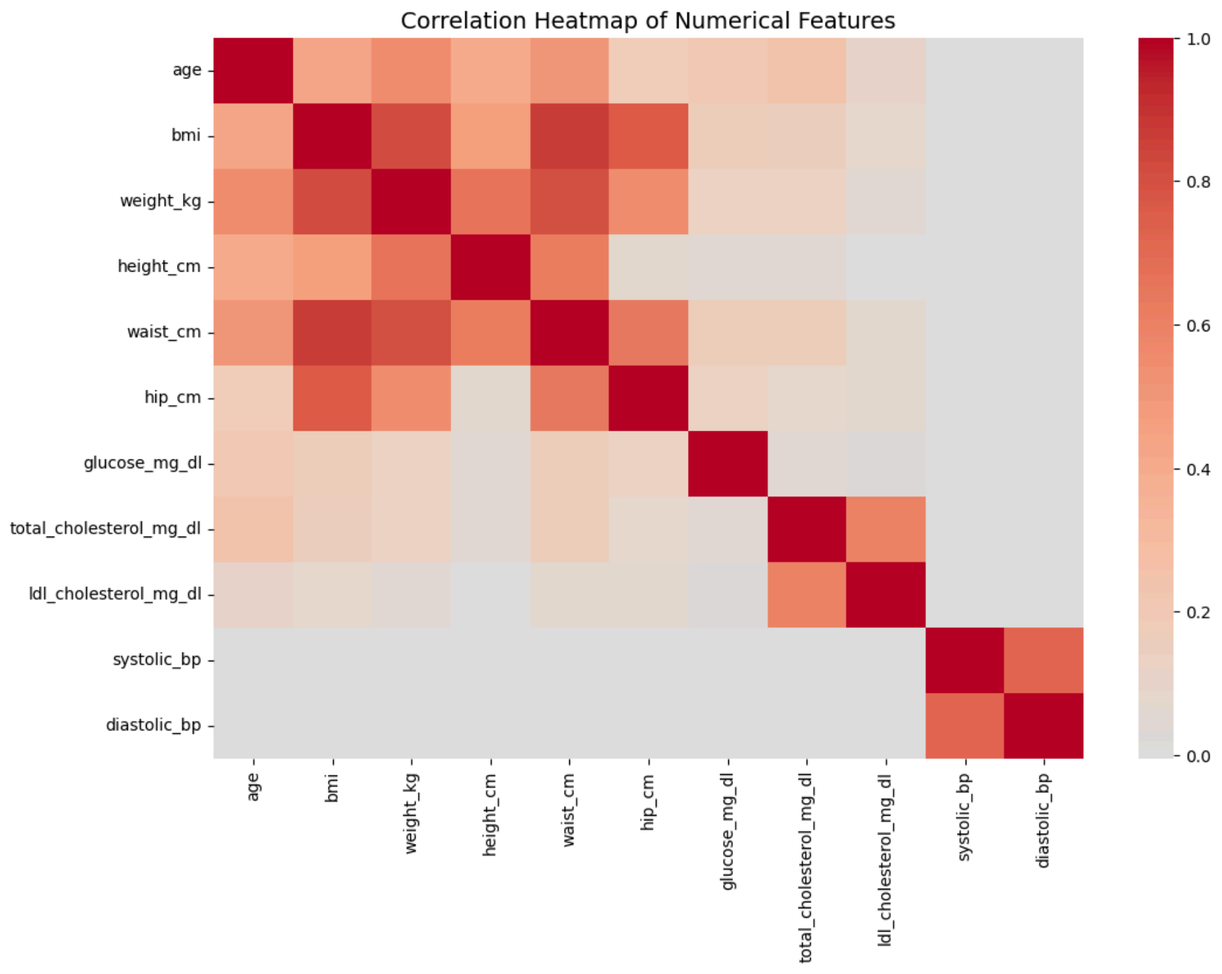
BUSINESS UNDERSTANDING

- **Problem:** Type 2 Diabetes affects 90–95% of all diabetes cases worldwide.
- **Goal:** Use machine learning to predict diabetes risk early for better prevention.
- **Impact:** Early detection can lower healthcare costs, improve patient outcomes, and reduce complication risks.
- **Stakeholders:** Healthcare providers, policymakers, insurance companies, patients.

DATA SET OVERVIEW

- **Source:** Public diabetes dataset (8 key features).
- **Key Variables:** Age, BMI, Glucose, Blood Pressure, Insulin, Pregnancies.
- **Challenge:** Imbalanced dataset (65% non-diabetic vs 35% diabetic).



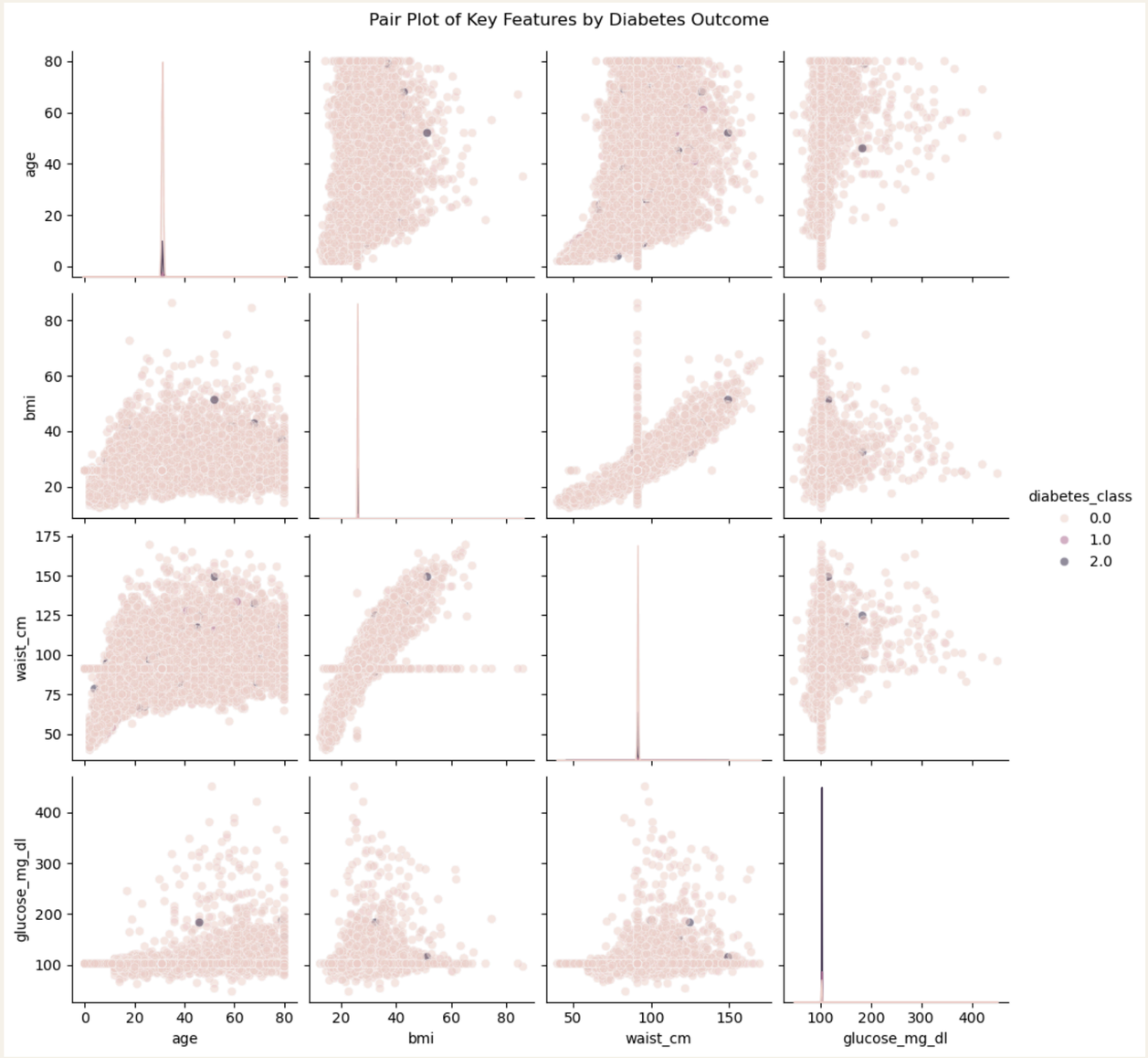


EDA HIGHLIGHTS

- Glucose & BMI strongly linked with diabetes outcome.
- Older age groups show higher diabetes prevalence.
- Some features (e.g., Insulin, Skin Thickness) have missing or zero values → cleaned.

FEATURE RELATIONSHIP

- Pair plots reveal clear separation of diabetic vs non-diabetic groups based on Glucose, BMI, and Age.
- Indicates strong predictive power of these features.



MODEL APPROACH

Preprocessing:

- Standard scaling applied to numerical features.
- SMOTE used to handle class imbalance.

Models Tested:

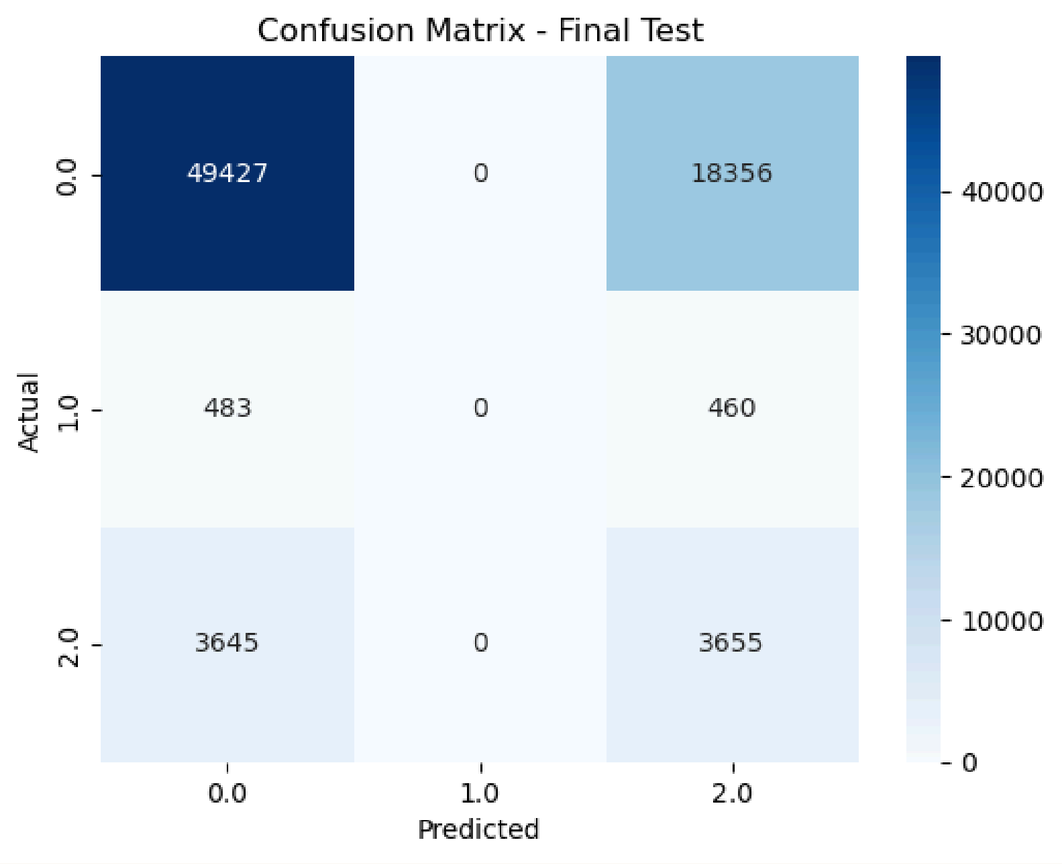
- Logistic Regression, Decision Tree, Random Forest, XGBoost, Gradient Boosting.

Evaluation Metrics:

- Accuracy, Precision, Recall, F1-score, ROC-AUC.

Focus on **recall** (catching diabetic cases) to minimize false negatives.

MODEL PERFOMANCE



- **Best Model:** XGBoost with highest recall & ROC-AUC.
- **Logistic Regression:** Interpretable but slightly lower performance.
- **Random Forest:** Balanced results, robust to noise.

Watch this live demo on how to safely prick yourself for a simple diabetes test!

KEY FINDINGS

- High Glucose and BMI are the strongest predictors of diabetes.
- Age significantly influences risk, with older groups more vulnerable.
- Balancing the dataset (SMOTE) greatly improved recall, reducing missed diabetic cases.
- XGBoost outperformed others, offering both accuracy and recall balance.

RECOMENDATION

Healthcare Systems:

- Deploy predictive models in screening programs.
- Prioritize patients flagged at higher risk.

Public Health:

- Promote lifestyle interventions (healthy diet, physical activity).
- Target awareness programs at high-risk groups.

Research/Policy:

- Collect more diverse data (ethnicity, lifestyle habits).
- Continuously update models with new data for relevance.

NEXT STEPS

- **Deploy Model:** Package as API for clinical integration.
- **Monitor & Retrain:** Update with new patient data regularly.
- **Expand Features:** Include genetic, lifestyle, and environmental factors for better accuracy.
- **Scale:** Extend application beyond Type 2 Diabetes to other chronic diseases.

Q&A

THANK YOU