IMRAD METHODOLOGY & DATA PREPROCESSING

Pima Diabetes Data Optimization

ROLE
Lead Methodologist & Data Processor

TOOLS
Microsoft Excel & JASP

01 The Challenge

THE CONSTRAINT

To satisfy the rigorous standards of our IMRAD research design, the data needed to meet strict assumptions for **Normality and Homogeneity.**

THE PROBLEM

The dataset contained massive gaps—specifically **48.7% missingness in Insulin**—and significant outliers that threatened the validity of our group's findings.

TECHNICAL STACK UTILIZED

Statistical Methods
Median Imputation, IQR Outlier Detection, Z-Score Standardization

Software
Microsoft Excel (Analysis Toolpak), JASP (Statistical Computing)

02 Methodology: Audit & Repair

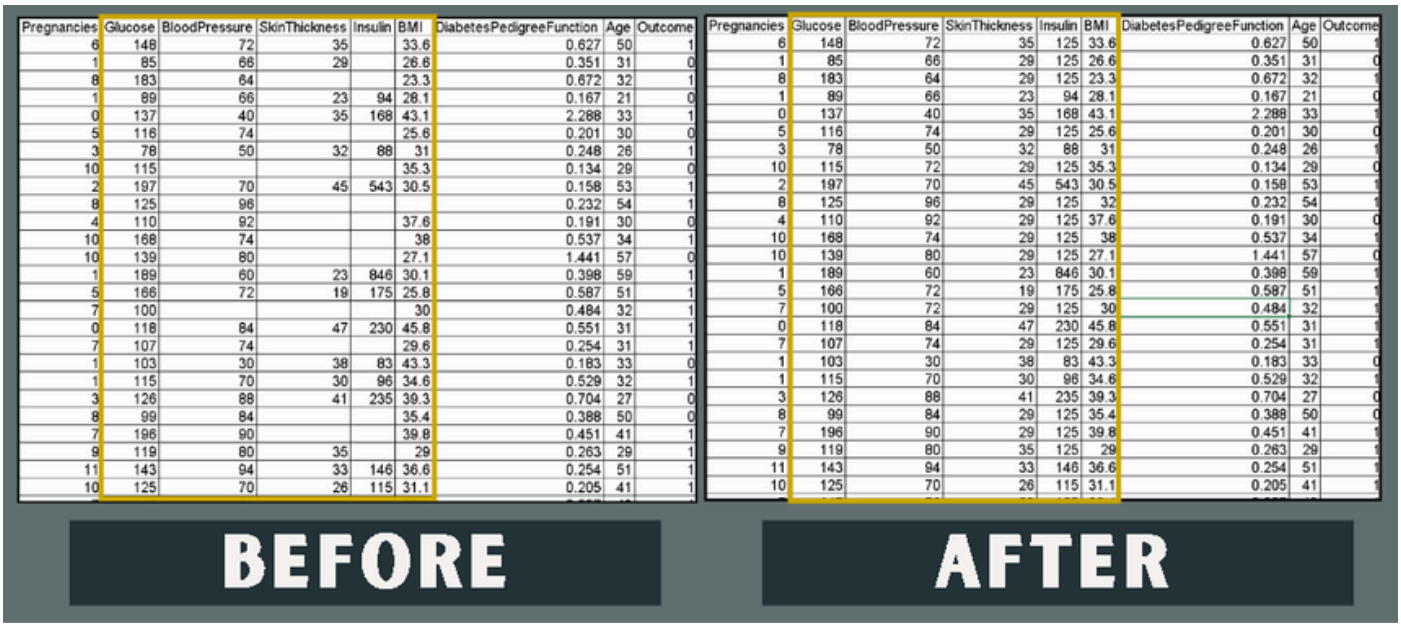
A THE AUDIT CONTEXT

Initial audit identified severe gaps in Insulin and Skin Thickness. This step was crucial to calculate the risk before applying imputation strategies.

Variable	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Pregnancies	Diabetes Pedigree Function	Age
Missing Values (%)	0.65	4.56	29.56	48.70	1.43	0	0	0
Imputed Method	Median	Median	Median	Median	Median	N/A	N/A	N/A

B DATA INTEGRITY (IMPUTATION)

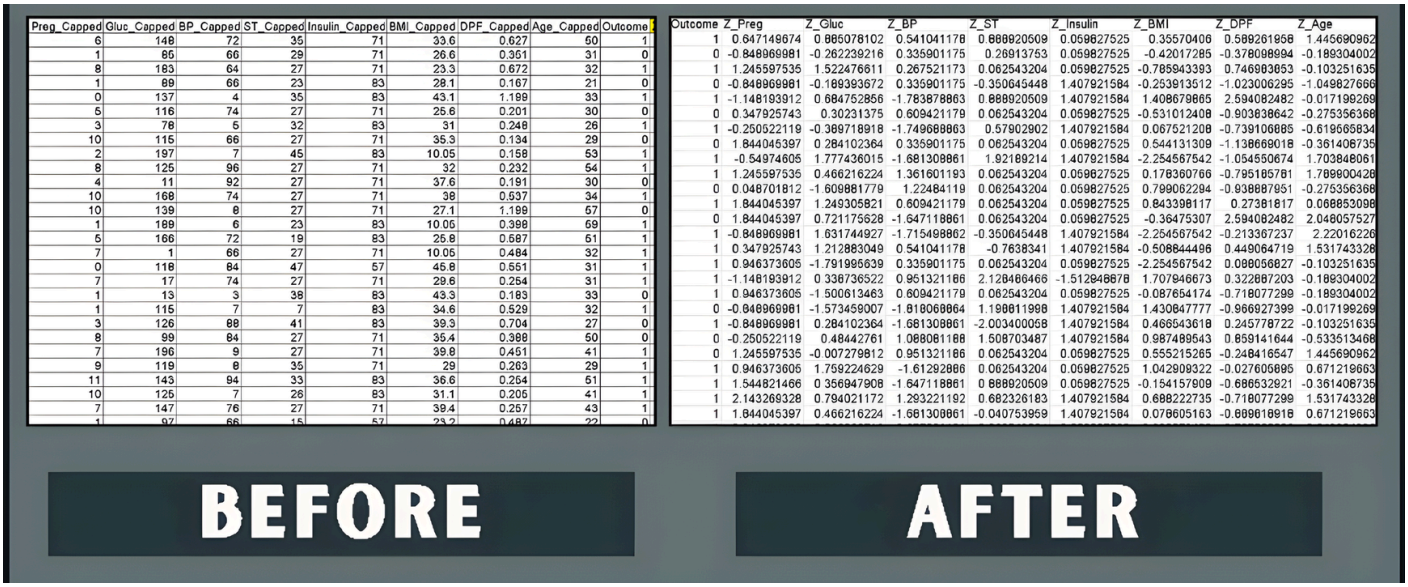
"Successfully salvaged 48.7% of the dataset (Insulin Records) using Median Imputation, preventing massive data loss."



... Methodology (Continued)

C STATISTICAL READINESS (Z-SCORES)

Standardized variables into Z-Scores (Mean=0, SD = 1). This transformation was mathematically necessary to satisfy assumptions for the group's Multivariate Analysis.



03 Research Impact

Data Integrity Preserved

Prevented the deletion of nearly half the sample size by successfully imputing 100% of missing **Insulin** records.

Statistical Robustness

The cleaned data set satisfied assumptions for the team's JASP analysis, leading to higher confidence intervals in our findings.