DIABETIC PATIENTS

Exploring the intricacies of diabetic patient data to identify trends and patterns that can aid in improving healthcare outcomes.

OUTLINE

- ➔ Introduction
- ➔ Data sources
- ➔ Documentation of cleaning and manipulations
- ➔ Summary of data analysis
- ➔ Key visualizations and findings
- ➔ Recommendations

Introduction

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether a patient has diabetes based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

Data Source

Dataset was gotten from meriskill Internship program.

Data contained 769 rows and 9 columns of sales data from an organization pertsining human resource department for a period of 40 years.

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases.

This dataset is deemed credible.

Documentation of cleaning and manipulations

Data was downloaded and stored using the appropriate file naming convention.

Tools employed for analysis were Excel and PowerBI.

I opened the dataset and attempted to understand the type of analysis expected. Upon examination, I discovered that the dataset contains data suitable for predictive analysis.

Excel was used to check for consistency and clean the empty columns.

I uploaded the dataset on PowerBI using the 'Get Data' option, and then proceeded to transform the data.

Predictive analysis is a type of analysis where this means understanding the probable future trends and behavior

To derive predictions from the given dataset, I partitioned the data into independent and dependent variables. Upon analysis, I observed that all seven variables i.e. Pregnancies, Glucose, Blood pressure, skin thickness, Insulin, BMI , Diabetes, Age are independent, while 'Outcome' serves as the dependent variable.

From the given dataset, 0 represents that the person is not diabetic and 1 shows that the person is diabetic.

Then moved on to creating visualizations while uncovering trends and gathering insights

Summary of data analysis

Analyzed data from 768 patients with the following averages; age 33.24, BMI 31.99, BP 69.11, Skin thickness 20.54, Pregnancies 3.85, Glucose 120.89, and Insulin 79.80.

I subdivided some variables

BMI

- 0 < 18.5 - Underweight
- 18.5 < 25 - Healthy weight
- 25 < 30 - Overweight
- 30 and above – Obese
- 30-34.9 - Obese class I
- 35-39.9 - Obese class II
- ≥ 40 Obese class III

Summary of data analysis

Glucose

- <100 - Normal fasting glucose
- 100-125 - Impaired fasting glucose
- ≥126 - Provisional diagnosis of Diabetes

Age

- 17-30 - Young Adults
- 31-45 - Middle Aged Adults
- Above 45 - Old Adults

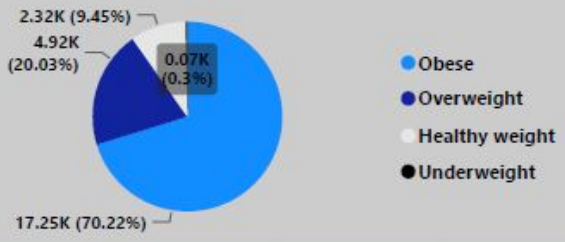
Key visualizations and findings

DIABETES DASHBOARD

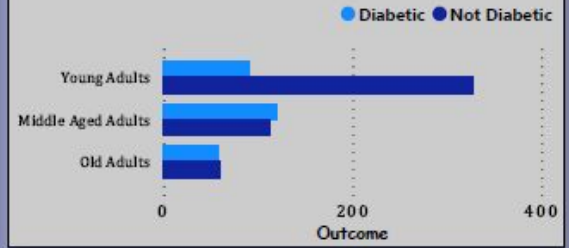
Average Age	Total Patients
33.24	768

Average Blood Pressure	Average SkinThickness	Average BMI	Average Pregnancies	Average Glucose	Average Insulin
69.11	20.54	31.99	3.85	120.89	79.80

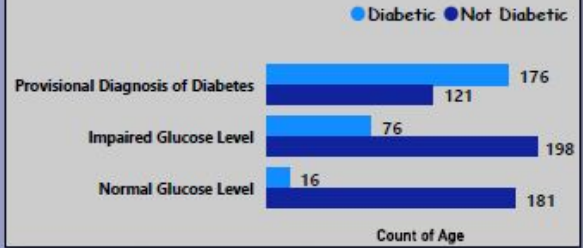
BMI Classification



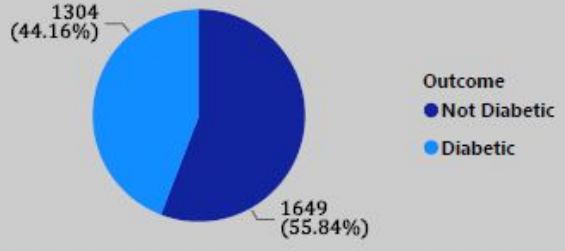
Outcome by Age



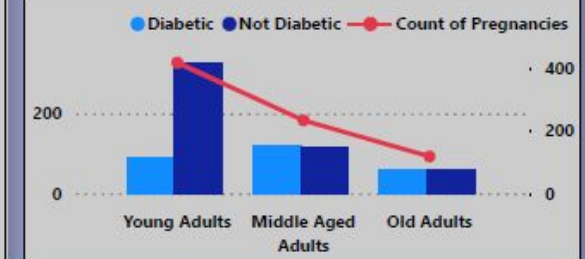
Count of Age by Glucose and Outcome



Pregnancy Distribution by Outcome



DiabetesPedigreeFunction and Pregnancies by Age and Outcome



Average of Insulin by BMI and Age

Recommendations

By harnessing data insights in healthcare, disease patterns can be predicted, adequate medical resources allocated and public health campaigns designed and implemented which in turn transforms the entirety of healthcare operations.