IDS-ASSIGNMENT

Data Set Link : https://www.kaggle.com/mehdidag/pima-indians

Team Member 1:

Name: R SIVA GIRISH

SRN: PES1201700159

EMAIL ID : sivagirish81@gmail.com

CONTACT NO.:8884749969

Team Member 2:

Name: Anirudh Avadhani

SRN: PES1201701526

EMAIL ID : anirudh99a@gmail.com

CONTACT NO.:8217496489

ABSTRACT

-Pima, North American **Indians** who traditionally lived along the Gila and Salt rivers in Arizona, U.S .

-The Pima Indians of Arizona have the highest reported prevalence of Diabetes in the World.

-This dataset is taken from Kaggle. The objective of the dataset is to diagnostically determine the probable reasons behind women of a particular tribe getting diabetes and hypertension, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances. In particular, all patients here are females at least 21 years old of Pima Indian heritage. Therefore we performed a descriptive analysis on the given dataset to arrive at certain Conclusions regarding the health of the women.

Data Set

-The data set consists of several medical variables such as number of pregnancies the women have had, their BMI, insulin level, age, and so on. It is a small Data set Consisting of 392 rows and 9 columns. It has 9 columns which are:

- ❖Pregnant
  - → It represents the number of times the woman got pregnant during her life.
- ❖Glucose
  - → It represents the plasma glucose concentration at 2 hours in an oral glucose tolerance test.
- ❖Diastolic
  - → The blood pressure is a very well-known way to measure the health of the heart of a person, there are too measure in fact, the diastolic and the systolic. In this data set, we have the diastolic which is in the fact the pressure in (mm/Hg) when the heart relaxed after the contraction.
- ❖Triceps
  - → It is a value used to estimate body fat (mm) which is measured on the right arm halfway between the olecranon

process of the elbow and the acromial process of the scapula.

❖ Insulin

→ It represents the rate of insulin 2 hours serum insulin (mu U/ml).

❖ Body Mass Index

→ It represents the Body Mass Index (weight in kg / (height in meters squared), and is an indicator of the health of a person.

❖ Diabetes

→ It is an indicator of history of diabetes in the family.

❖ Age

→ It represents the age in years of the Pima's woman.

❖ Test

→ It can take only 2 values ('negatif' or 'positif') and represents if the woman shows signs of diabetes.

	pregnant	glucose	diastolic	triceps	insulin	bmi	diabetes	age
count	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000	392.000000
mean	3.301020	122.627551	70.663265	29.145408	156.056122	33.086224	0.523046	30.864796
std	3.211424	30.860781	12.496092	10.516424	118.841690	7.027659	0.345488	10.200777
min	0.000000	56.000000	24.000000	7.000000	14.000000	18.200000	0.085000	21.000000
25%	1.000000	99.000000	62.000000	21.000000	76.750000	28.400000	0.269750	23.000000
50%	2.000000	119.000000	70.000000	29.000000	125.500000	33.200000	0.449500	27.000000
75%	5.000000	143.000000	78.000000	37.000000	190.000000	37.100000	0.687000	36.000000
max	17.000000	198.000000	110.000000	63.000000	846.000000	67.100000	2.420000	81.000000

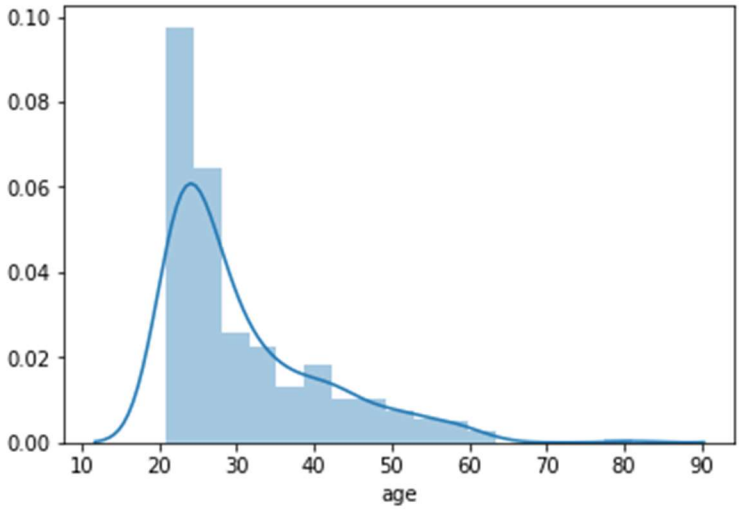
Introduction

-We chose this dataset inorder to study the health of pima Indian women and analyse the reason for the high rate of prevalence of Diabetes amongst them.

-Based on our analysis we intend to draw reasonable conclusions and give suitable suggestions as to how to minimize the number of women suffering from diabetes amongst them.

-This Data set facilitated us with many important medical parameters which suited our analysis and helped us to see a correlation between many of the medical parameters hence we chose this dataset.

Processing of Data



-The above graph is a distplot which shows the distribution of data in the dataset.

-Women above age 65 and below the age 21 have been removed from the dataset as women above the age 65 can have diabetes due to old age and women below the age 21 suffering diabetes is an extremely rare case.

-We cleaned the glucose column based on the fact that the glucose levels range from 35 to 165 in all cases including the cases of people having diabetes.

-So the values close to the upper limit and lower limit were rounded of to the closest value which could be possible.

-The values which deviated a lot from the range mentioned above were replaced with their respective medians.

-Similarly we cleared the triceps column as well as the diastolic column.

-Glucose:35-165milli-gram per deci litre(mg/dl)

-Triceps:10-40 milli metre(mm)

-Diastolic:40-100 milli metre per merqury(mm/hg)

*IT is not possible for a human being to have values for the above parameters which are not in the mentioned range.

Exploratory Analysis

Assumptions

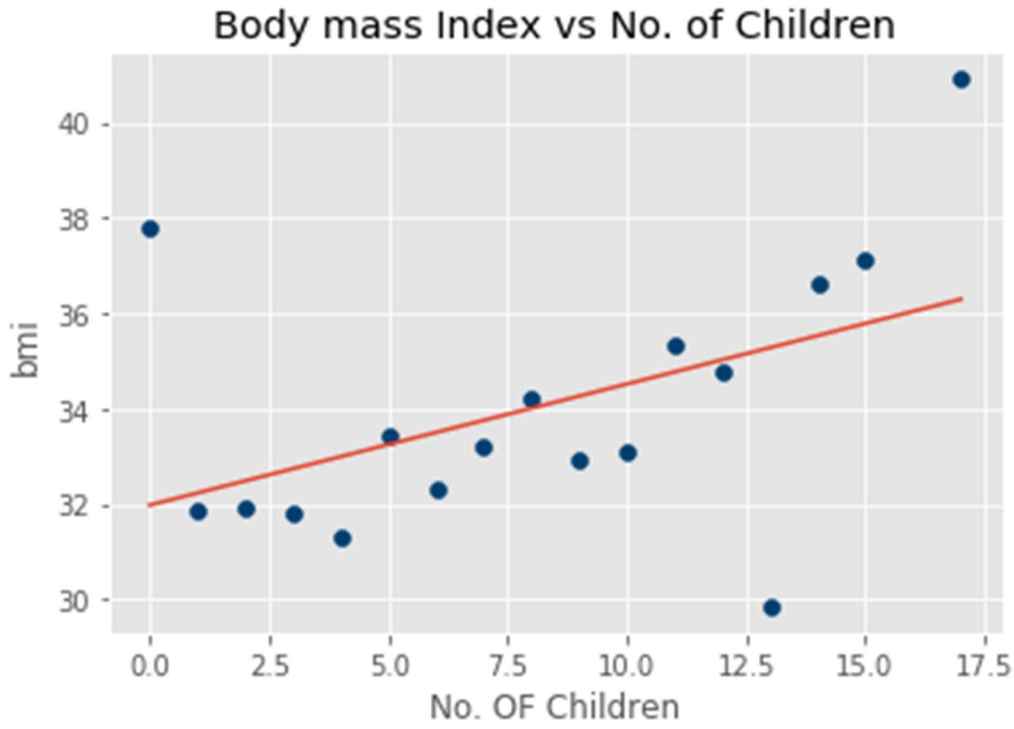
-All the women in the sample belong to same tribe .Thus genetically all the women are almost the same.If one of the women in the tribe has diabetes then all the Women in the tribe should have it and vice versa.Thus genetics cannot be taken into consideration.

-As all the women belong to the same tribe

We assume that all of them have the same food

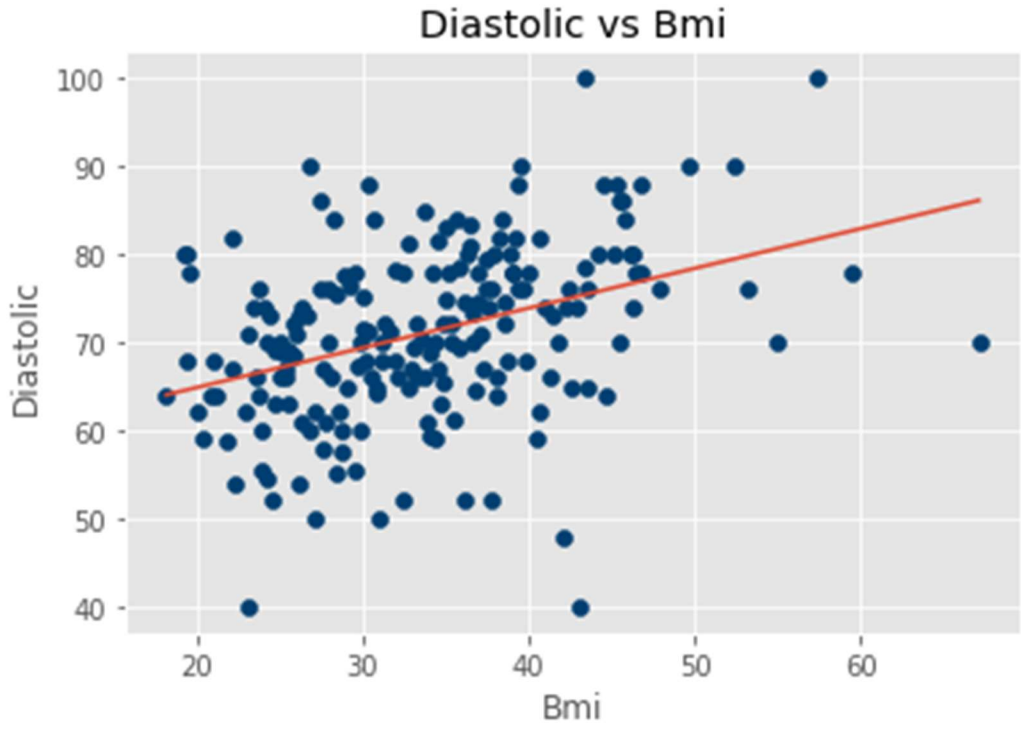
Habits.

BMI Vs Number of children



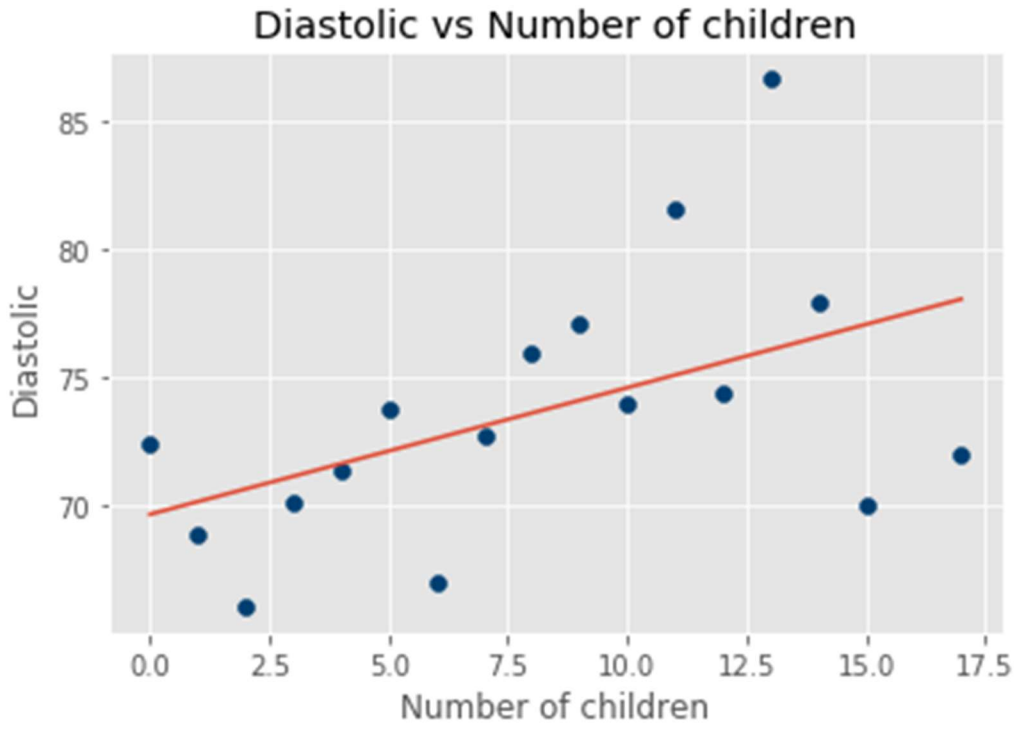
- After childbirth there are hormonal changes in the woman due to which there is increase in the weight of the woman.
- Stress: Taking care of a newborn baby is stressful. This stress results in release of Certain hormones like cortisol which increases. The appetite thus causing the body to store more Fat.

Diastolic Vs BMI



As the bmi increases,The fat deposits in the walls Of the arteries due to which the arteries narrow down,Thus increasing the diastolic pressure.

Diastolic Vs Number of Children



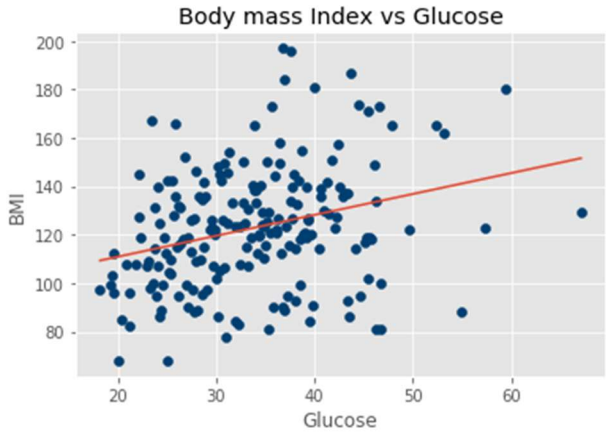
Based on the insights of the previous 2 graphs we can infer that as the

a)As the number of children a woman bears increases

The bmi increases.

b)As bmi increases the diastolic pressure increases.

Thus we can conclude that as the number of children a woman bears increases diastolic pressure increases.



As Bmi increases insulin resistance

Also increases which results in

increased blood glucose level in

body

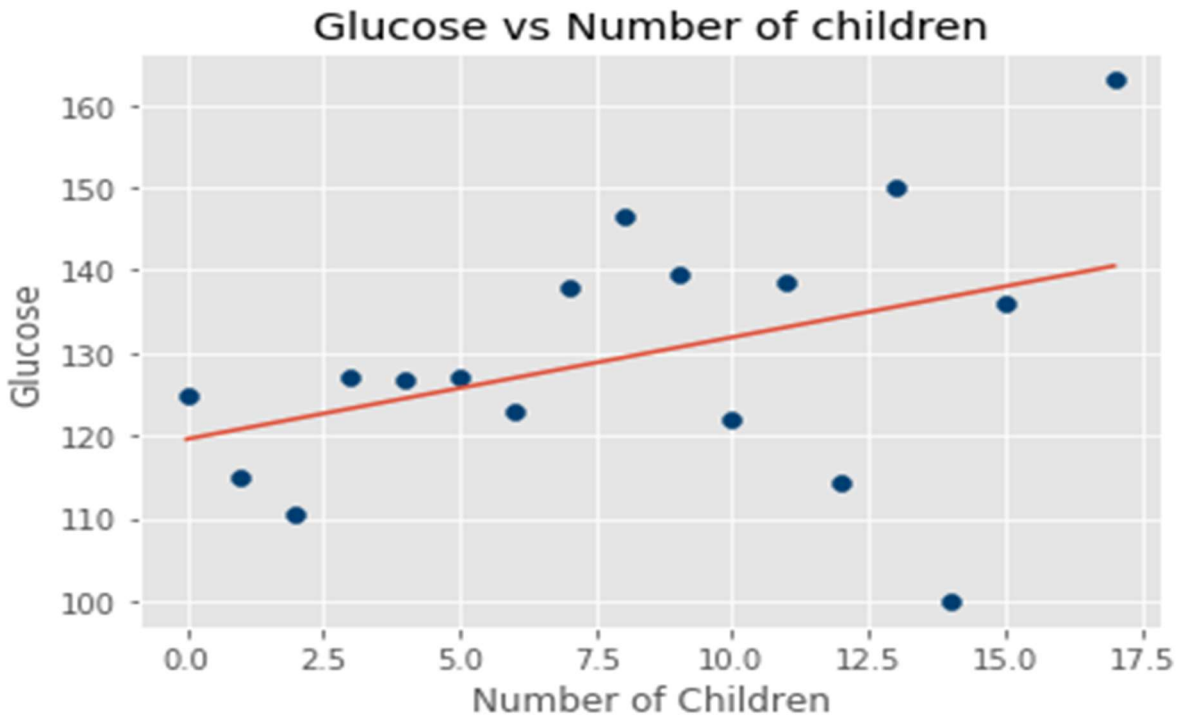
Since body weight is associated

with bmi

It may be expected that bmi should

Correlate with blood glucose levels.

Glucose Vs Number of Children



Based on the insights of the previous

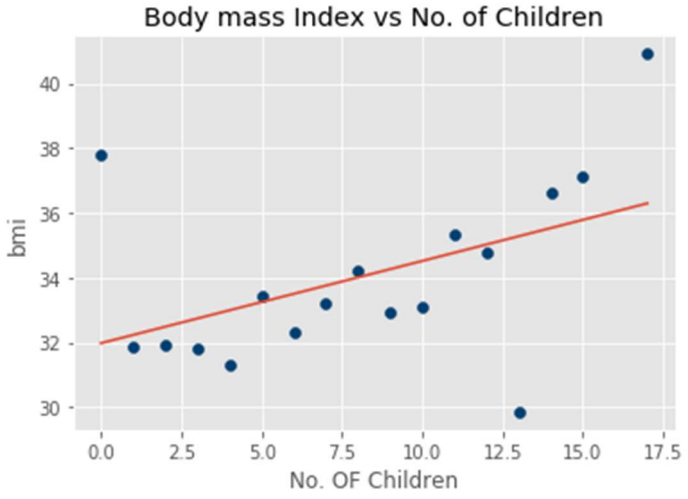
graphs we can infer that as the

a)As the number of children a woman bears increases

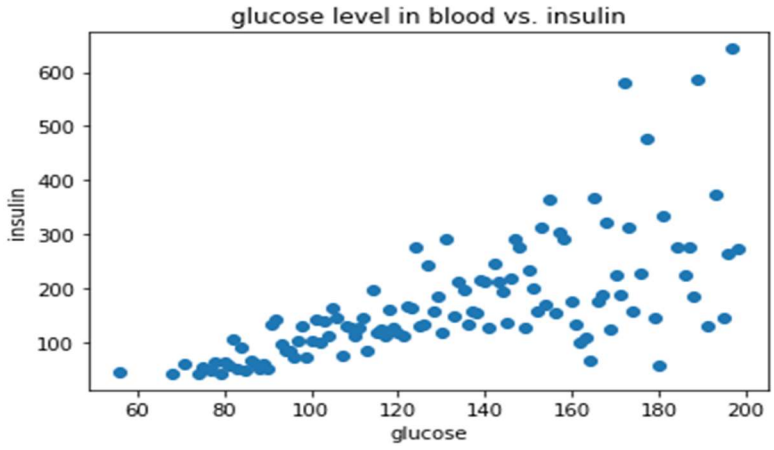
The bmi increases.

b)Bmi increases with increase in glucose levels in the blood.Thus we can conclude that as the number of children increases blood glucose level also increases.

1)As the number of children a woman bears increases her bmi Increases.



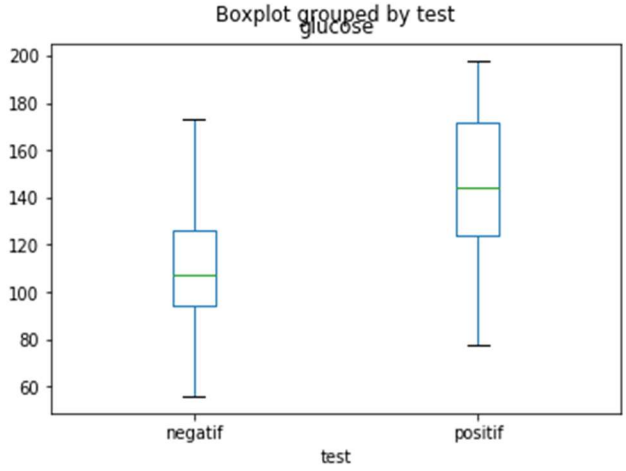
2) More the carbohydrate level in the blood more is the requirement for Pancreases to produce insulin to break it down into smaller Units called glucose.

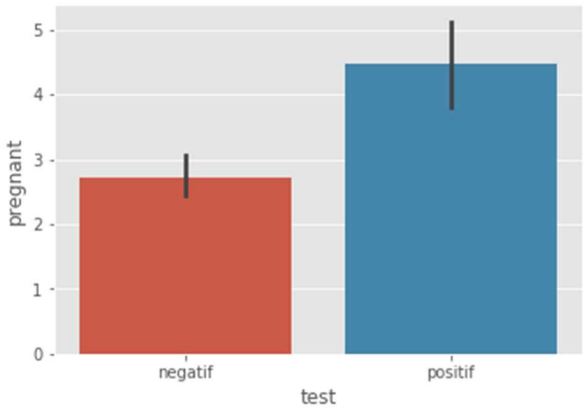


From the previous graph we know that as blood glucose level Increases bmi increases.

If the woman has high bmi (obese),more amount of insulin is Required by the body . Suppose there is slight increase in the sugar

Consumption then extra amount of insulin will not be available As most of the insulin produced will already be used.Thus sugar level in the blood increases and the person is tested positive.





Conclusion:

Thus based on the analysis done we can successfully conclude that the women having more number of children have risk of getting diabetes and high blood pressure.

Thus,

- They must be educated on Birth control.
- Family welfare programs must be conducted for them by ngo’ s or the government .
- More medical assistance must be provided to the women having more children. For example regular health checkups etc.
- Organizations must raise awareness amongst them as to the seriousness of diabetes and high blood pressure.