



Exercise 3- Report

GROUP 2:

Amber Sethi (101328584)
Joshua Dmello (101346654)
Mohamed Bennis (101276333)
Muhammad Attique (101292217)
Tanvi Sharma (101265693)

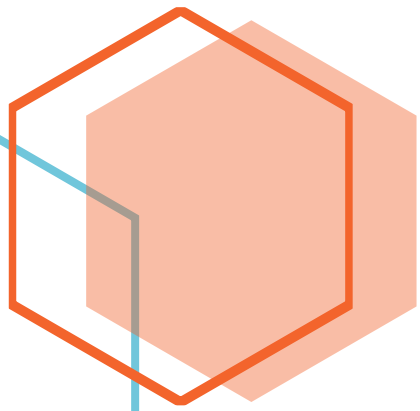




Table of Contents

DATA LOADING	2
THE DATA DIAGNOSTIC REPORT.....	2
FREQUENCY DISTRIBUTION REPORT	4
KEY FINDINGS & INSIGHTS	7
APPENDIX	8
REFERENCES	10

DATA LOADING

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

- First 5 records of the dataset

THE DATA DIAGNOSTIC REPORT

VARIABLE	# OF RECORDS	DATA FIELD FORMAT	# OF UNIQUE VALUES	# OF MISSING VALUES
Pregnancies	768	Character	0	0
Glucose	768	Character	0	0
Blood Pressure	768	Character	0	0
Skin Thickness	768	Character	0	0
Insulin	768	Character	0	0
BMI	768	Character	0	0
Diabetes Pedigree Fn	768	Character	0	0
Age	768	Character	0	0
Outcome	768	Character	0	0

DATA DIAGNOSTIC- OTHER STATS

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

- The above-mentioned data can further be interpreted as follows:
- There are 768 people as the population sample.
- The population sample consists of females (Pregnancy is a variable) with an average of 3 pregnancies (The lowest absolute number).
- The minimum age of the population sample is 21 and the maximum age is 81 with the average age of the population at 33.
- BMI is on average very close to 32 which corresponds to the category Obese (BMI 30 and over)

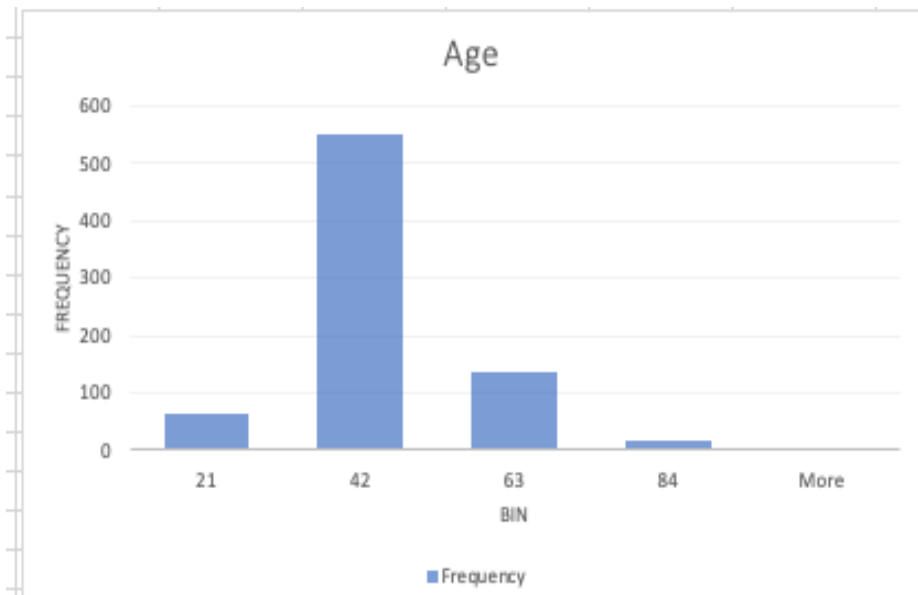
DATA USEFULNESS

- This data exploration gives us quantifiable dimensions to our prediction model.

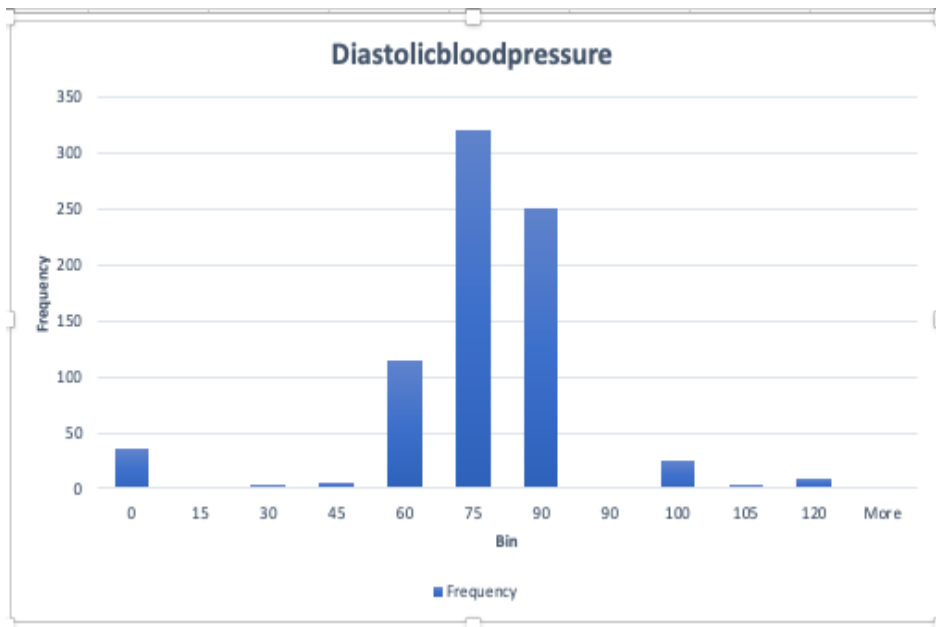
DATA FALLACIES

- We do not have any non-numeric data.

FREQUENCY DISTRIBUTION REPORT



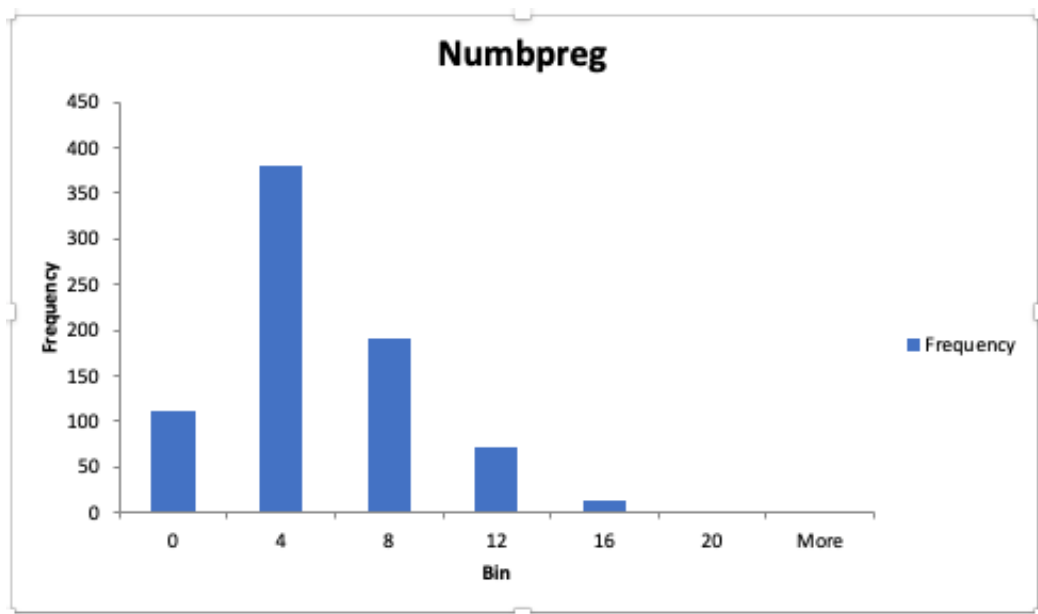
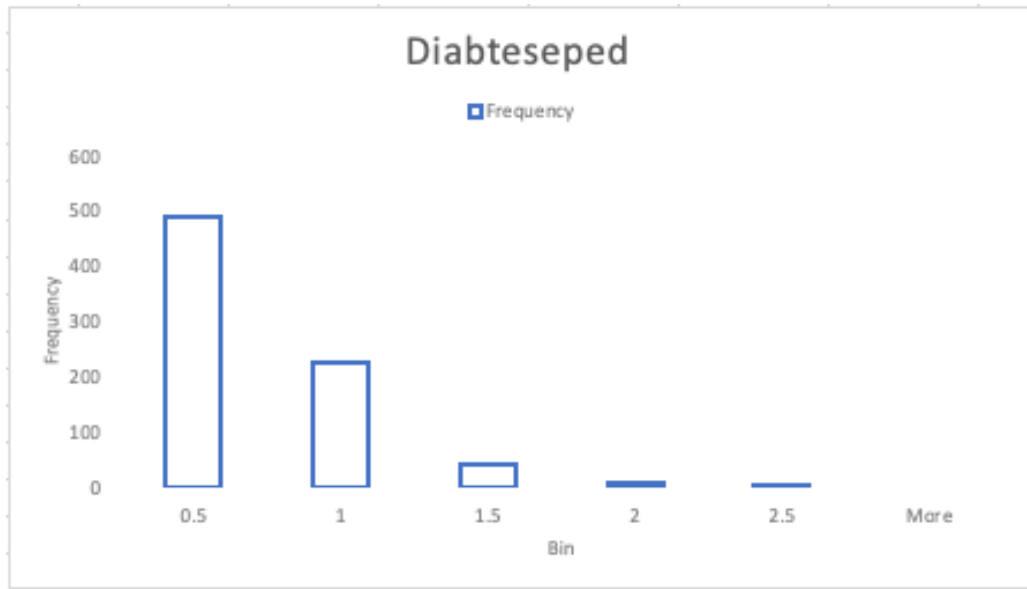
- If we create age ranges and distribute frequency counts amongst them, we can start analyzing demographic-based indicators.
- More than 500 people diagnosed were of age 42. The report further says that almost only 50 were of age 21 and around 120 were of age 63 and the least were of age 84.



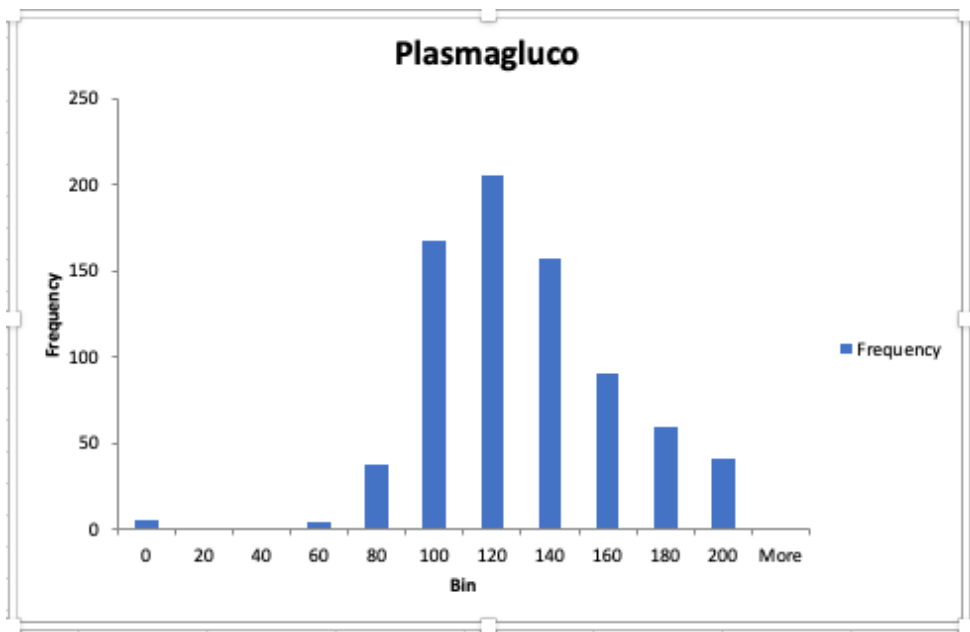
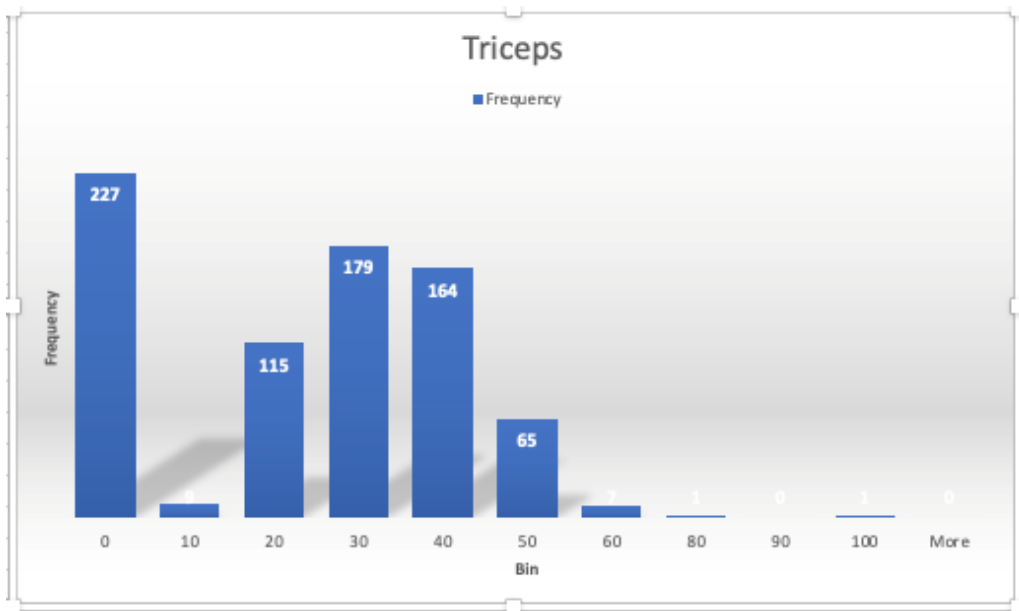
Exercise 3- Report



- The majority of Diastolic blood pressure patients lie at 75 which is almost close to the normal range that should be 80. The patients with a high diastolic blood pressure range are less than 50.

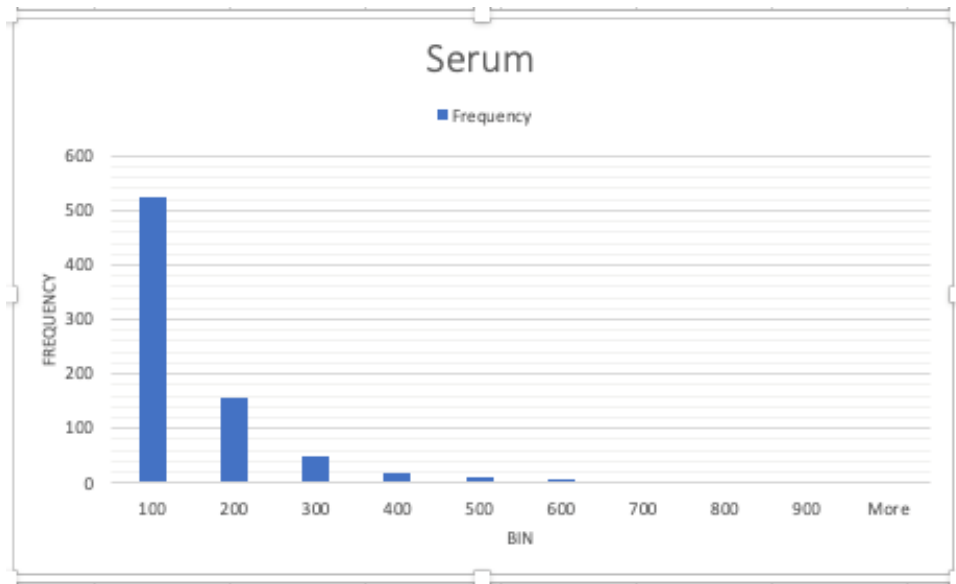


Exercise 3- Report



- The plasma glucose level reports being 120 for most of the patients which is on the risk side as the normal plasma glucose level is 70.

Exercise 3- Report



More than 500 patients had an almost normal serum level which was calculated around 100. Around 50 patients had a high range of serum levels.

KEY FINDINGS & INSIGHTS

- Most of the diabetic patients recorded were in their 40's. The patients recorded in their old age were very few.
- The diastolic blood pressure range was recorded to be normal for most diabetic patients. However, only a few of them were recorded to be at a high diastolic blood pressure range.
- The research shows that majority of the patients had glucose levels ranging between 100 -140, which is slightly concerning as the normal range of glucose levels is 70. The patients on a slightly riskier side with high glucose levels were about 200 which is a bit alarming.
- As for the serum levels, more than 500 patients were recorded to have a normal range of serum levels and only 50 of them were recorded to have high serum levels.

APPENDIX

Appendix A

	Numbpreg	Plasmagluucose	Diastolicbloodpressure	Triceps	seruminsulin	bodymass	Diabetespedigree	Age	Classvariable
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1

Appendix B

	mean	sd	min	max	median
Numbpreg	3.845052	3.369578	0.000	17.00	3.0000
Plasmagluucose	120.894531	31.972618	0.000	199.00	117.0000
Diastolicbloodpressure	69.105469	19.355807	0.000	122.00	72.0000
Triceps	20.536458	15.952218	0.000	99.00	23.0000
seruminsulin	79.799479	115.244002	0.000	846.00	30.5000
bodymass	31.992578	7.884160	0.000	67.10	32.0000
Diabetespedigree	0.471876	0.331329	0.078	2.42	0.3725
Age	33.240885	11.760232	21.000	81.00	29.0000
Classvariable	0.348958	0.476951	0.000	1.00	0.0000

Appendix C

	mean	sd	min	max	median	miss.val	uniq.val	length
Numbpreg	3.845052	3.369578	0.000	17.00	3.0000	0	0	768
Plasmagluucose	120.894531	31.972618	0.000	199.00	117.0000	0	0	768
Diastolicbloodpressure	69.105469	19.355807	0.000	122.00	72.0000	0	0	768
Triceps	20.536458	15.952218	0.000	99.00	23.0000	0	0	768
seruminsulin	79.799479	115.244002	0.000	846.00	30.5000	0	0	768
bodymass	31.992578	7.884160	0.000	67.10	32.0000	0	0	768
Diabetespedigree	0.471876	0.331329	0.078	2.42	0.3725	0	0	768
Age	33.240885	11.760232	21.000	81.00	29.0000	0	0	768
Classvariable	0.348958	0.476951	0.000	1.00	0.0000	0	0	768

REFERENCES

- Mayo Foundation for Medical Education and Research. (2020, October 30). *Diabetes*. Mayo Clinic. <https://www.mayoclinic.org/diseases-conditions/diabetes/diagnosis-treatment/drc-20371451>.
- 15th January 2019 By Editor, & Editor. (2020, April 30). *Normal blood sugar ranges and blood sugar ranges for adults and children with type 1 diabetes, type 2 diabetes, and blood sugar ranges to determine people with diabetes*. Diabetes. https://www.diabetes.co.uk/diabetes_care/blood-sugar-level-ranges.html.