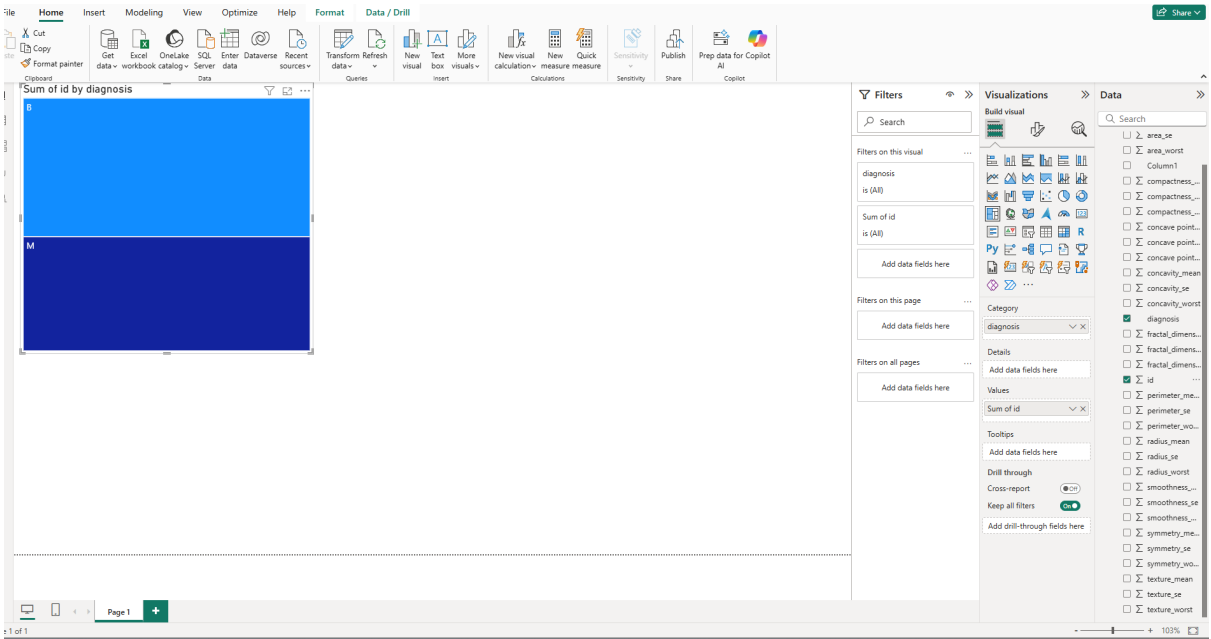
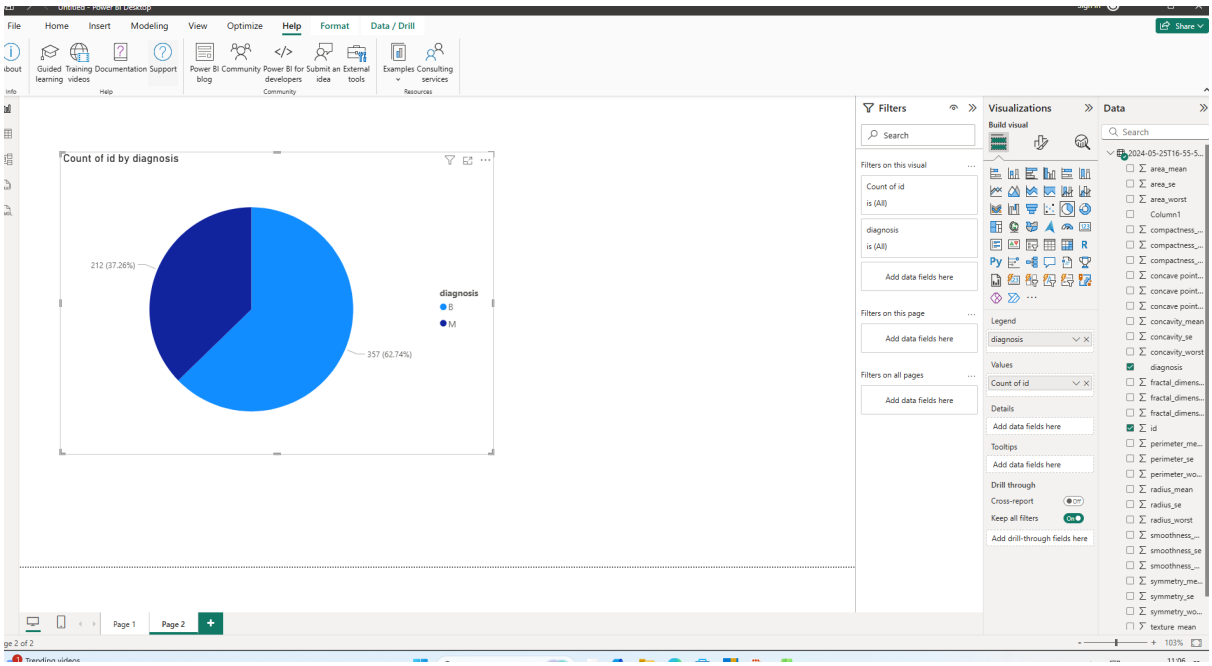
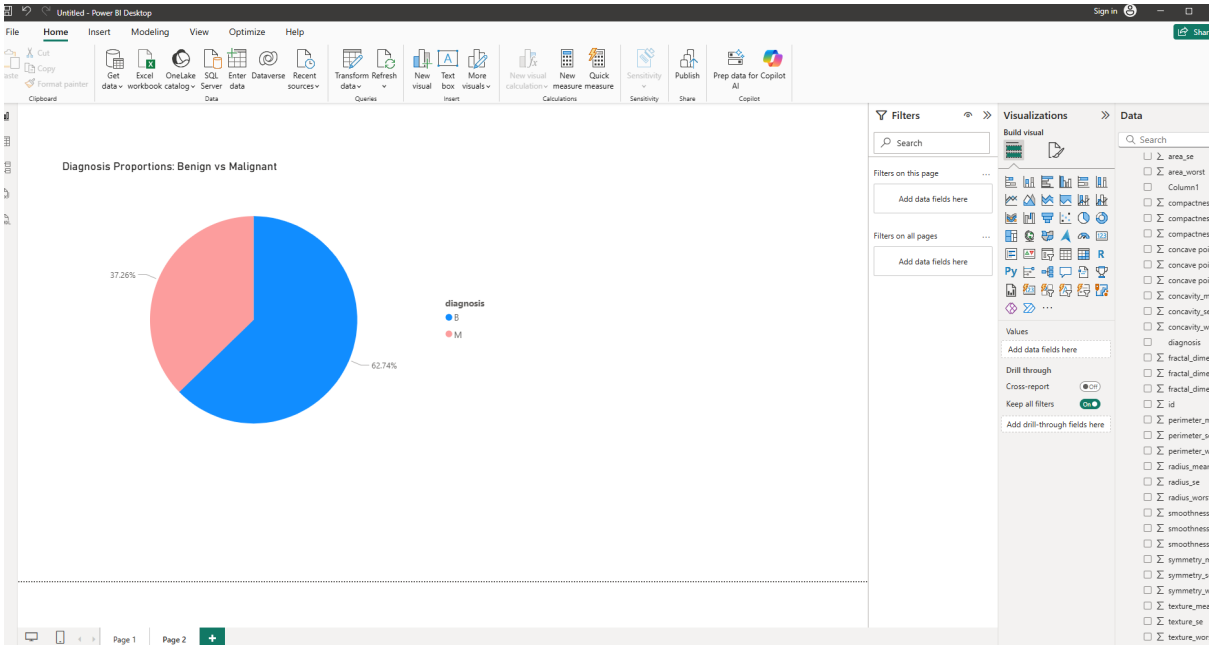
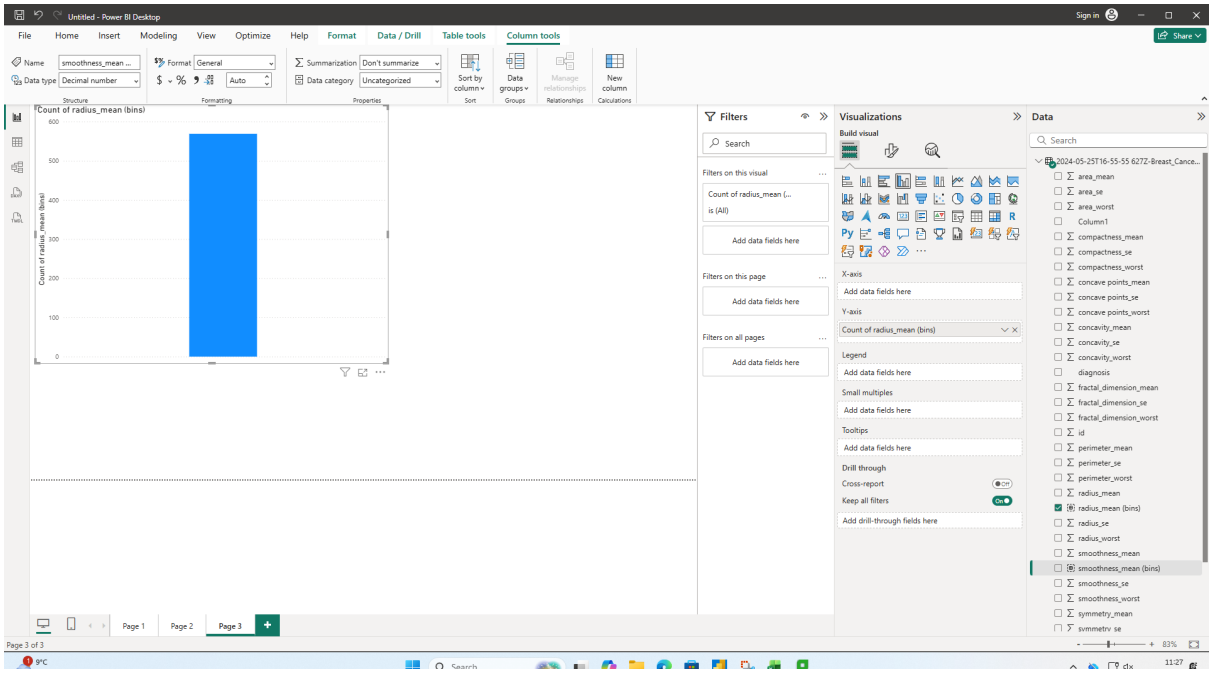
Supporting Info Upload – Group Project

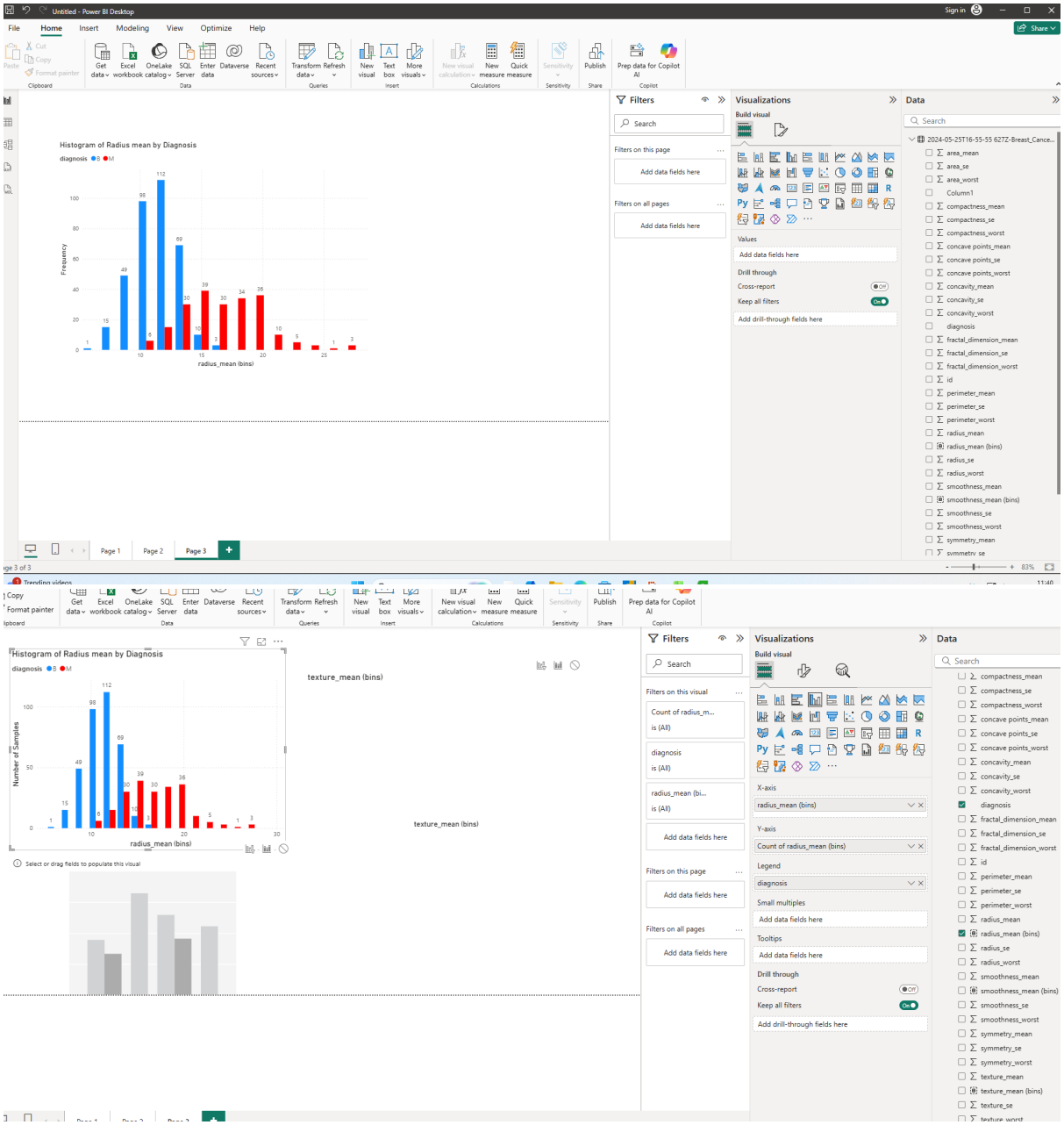
These are the images of the progress of few Visualizations.

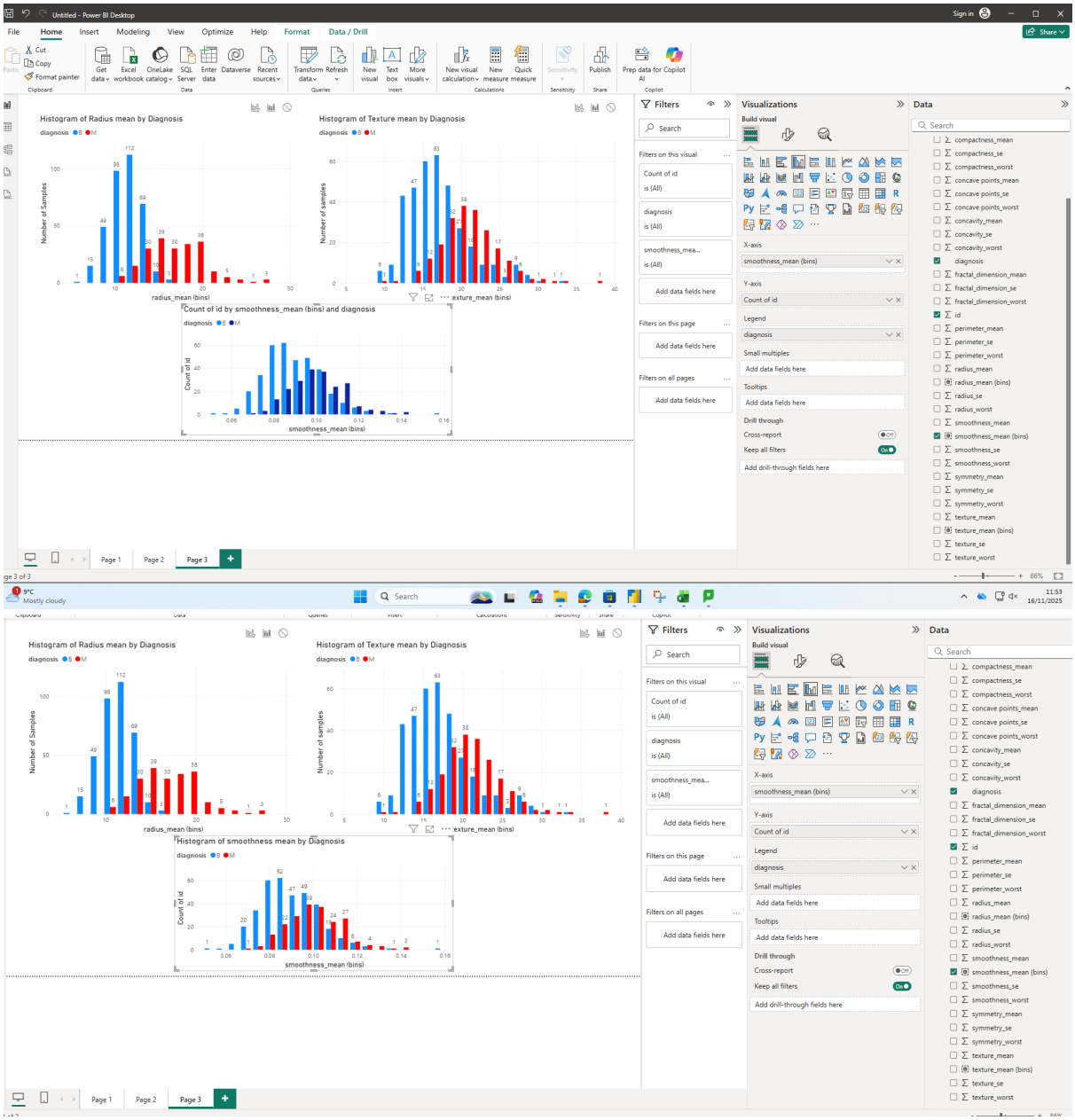


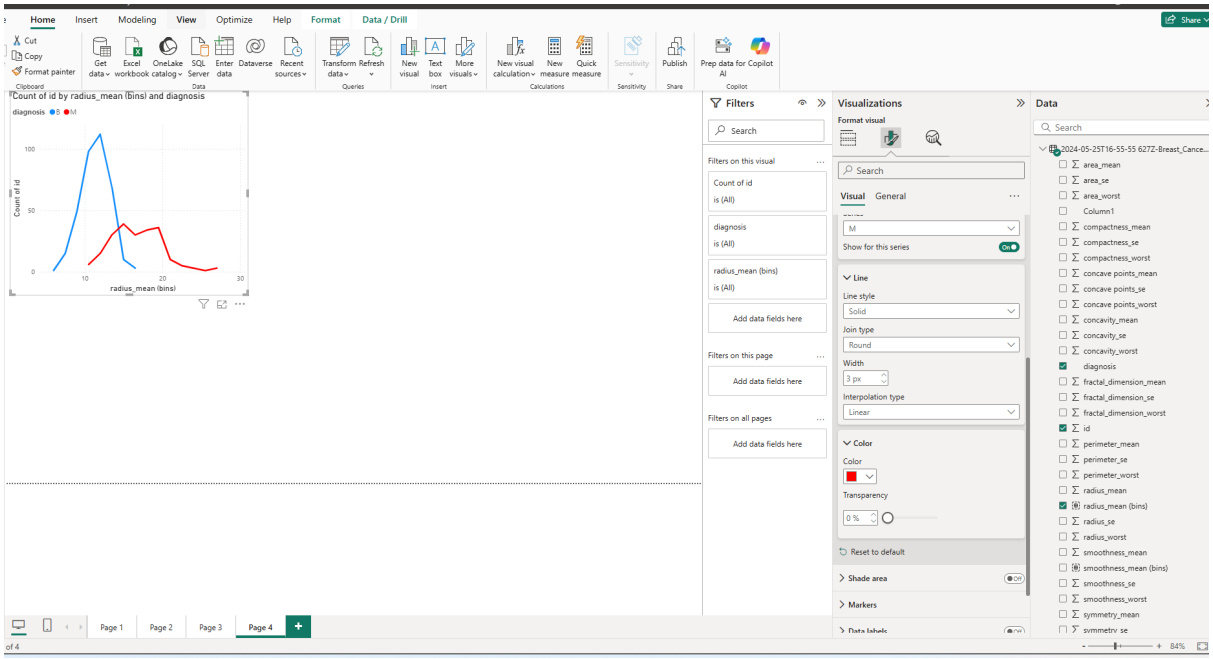


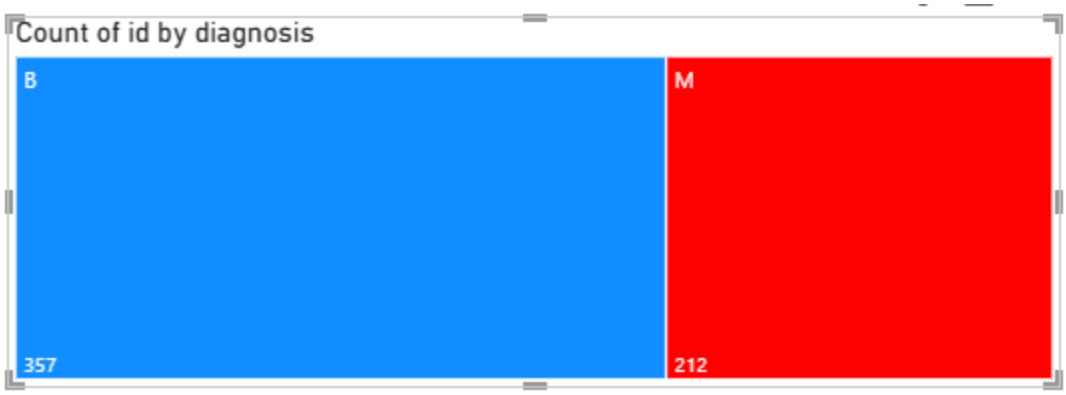

Untitled - Power BI Desktop
Diagnosis Proportions: Benign vs Malignant
37.26%
62.74%
diagnosis
B
M
Filters
Visualizations
Data
Page 1
Page 2


Untitled - Power BI Desktop
Count of radius_mean (bins)
Count of radius_mean (bins)
Filters
Visualizations
Data
Page 1
Page 2
Page 3
Page 3 of 3

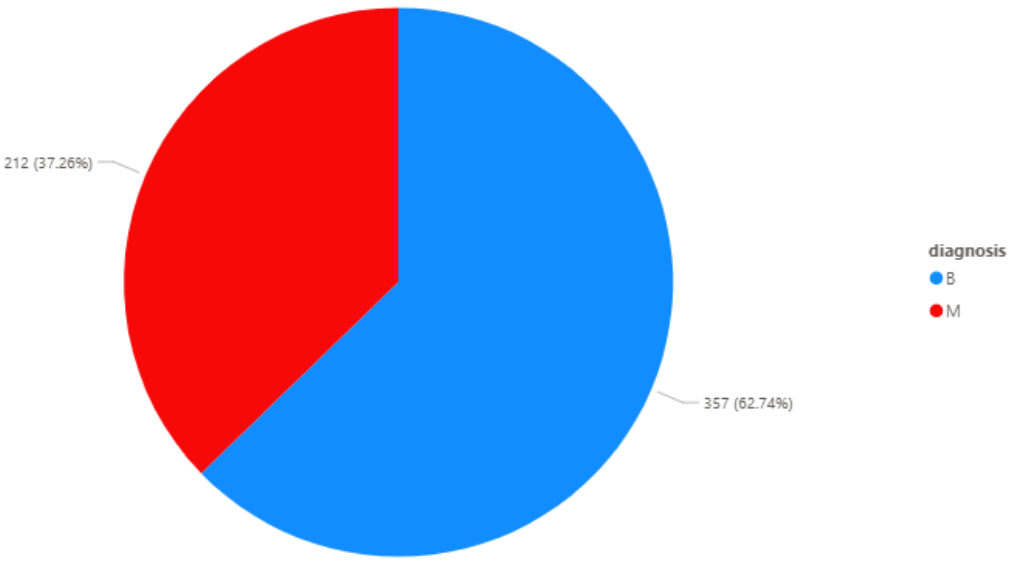

Untitled - Power BI Desktop
Histogram of Radius mean by Diagnosis
diagnosis B M
Frequency
radius_mean (bins)
Filters
Visualizations
Data
Page 1
Page 2
Page 3
texture_mean (bins)
Number of Samples
Select or drag fields to populate this visual

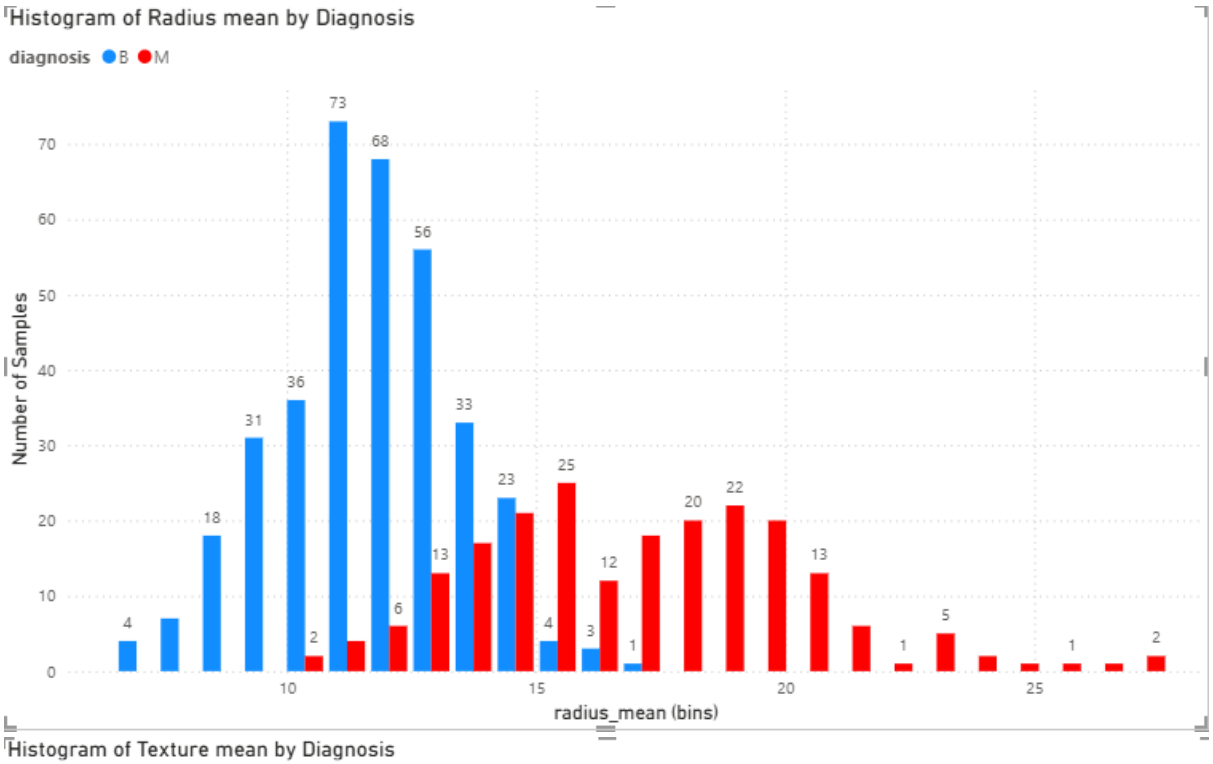

Untitled - Power BI Desktop
Histogram of Radius mean by Diagnosis
Histogram of Texture mean by Diagnosis
Count of id by smoothness_mean (bins) and diagnosis
Filters
Visualizations
Data
Page 1
Page 2
Page 3

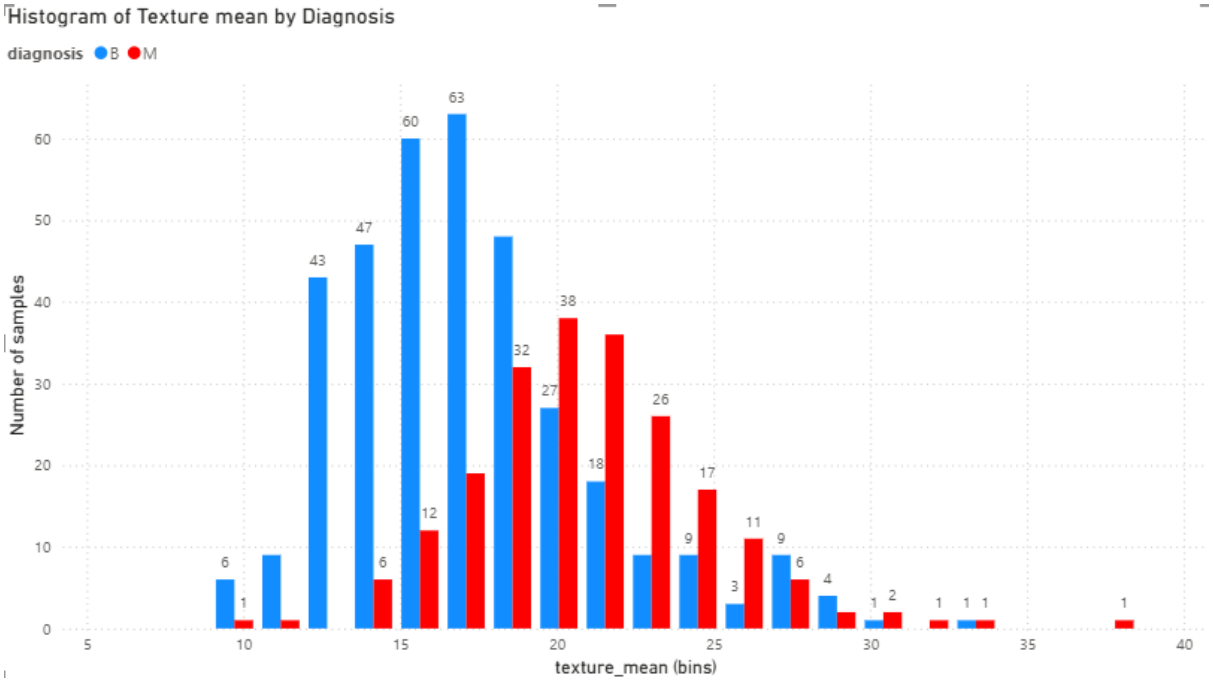


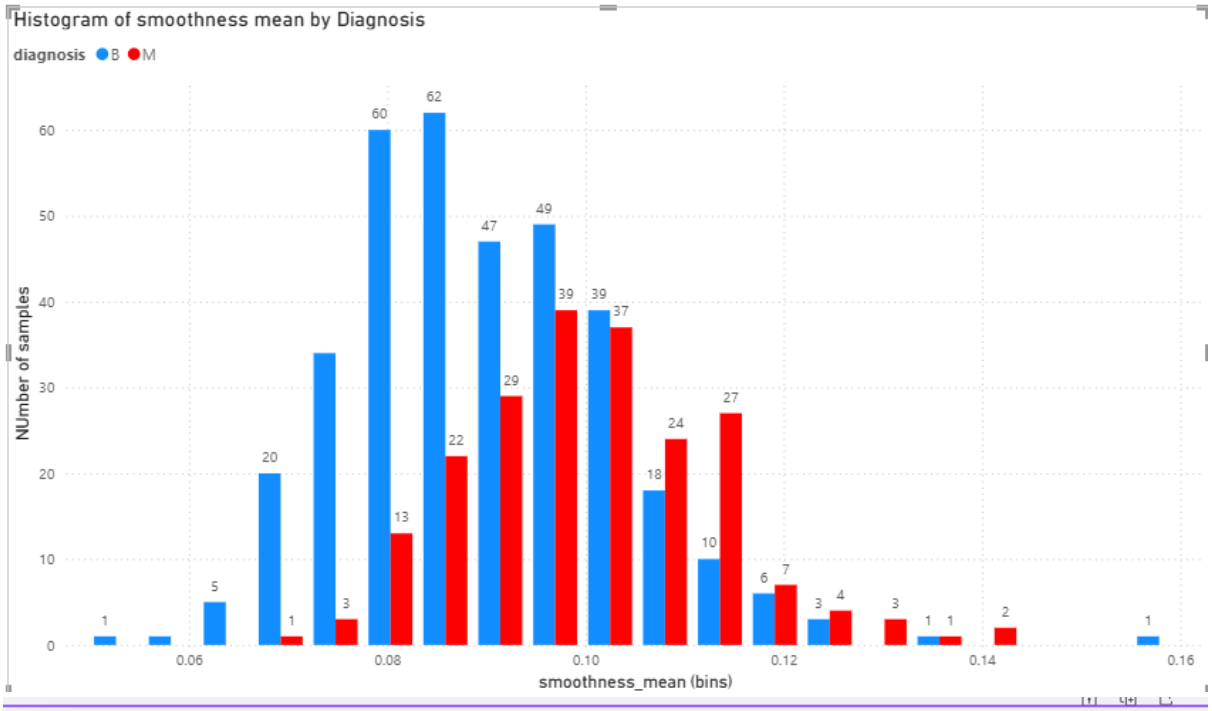


Diagnosis Proportions: Benign vs Malignant




Histogram of Radius mean by Diagnosis
diagnosis B M
Number of Samples
radius_mean (bins)
0
10
20
30
40
50
60
70
10
15
20
25
4
18
31
36
73
68
56
33
23
4
3
1
2
6
13
25
12
20
22
13
1
5
1
2


Histogram of Texture mean by Diagnosis
diagnosis B M
Number of samples
texture_mean (bins)
0
10
20
30
40
50
60
5
10
15
20
25
30
35
40
6
1
43
47
6
60
12
63
32
27
38
18
26
9
17
3
11
9
6
4
1
2
1
1
1
1



University of South Wales Prifysgol De Cymru

Comparative Analysis of Cell Radius, Texture, and Smoothness in Malignant vs Benign Breast Tumors

Introduction

Breast cancer remains one of the leading causes of mortality among women worldwide. Early diagnosis is critical for improving outcomes, and data-driven approaches can support clinical decision-making. We analyzed over a dataset named "Breast Cancer Wisconsin Dataset" which is created by William H. Wolberg. We researched over this dataset whether a tumor is benign or malignant (Cancer) using total 569 records of patients. Specifically, from cell nucleus features extracted from digitized images. Our goal is to

Dataset

The dataset contains 569 patient records, each identified by a unique ID. It includes 30 numerical features extracted from digitized images of cell nuclei, with 357 benign and 212 malignant cases.

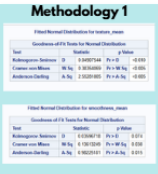
diagnosis	N Obs	Variable	Mean	Std Dev	Minimum	Maximum	Median	N	N Miss
B	357	radius_mean	12.1465238	1.7805115	6.9810000	17.8500000	12.2000000	357	0
		texture_mean	17.9147619	3.9891245	9.7100000	33.8100000	17.3900000	357	0
		smoothness_mean	0.0924778	0.0134481	0.0628300	0.1634000	0.0907600	357	0
M	212	radius_mean	17.4628302	3.2099711	10.9500000	28.1100000	17.3250000	212	0
		texture_mean	21.6046207	3.7794486	10.3800000	39.2800000	21.4650000	212	0
		smoothness_mean	0.1028865	0.0120082	0.0737100	0.1447000	0.1023000	212	0

Hypothesis

H_0: There is no significant difference in cell radius, texture, or smoothness between malignant and benign tumors.
H_1: At least one of these features differs significantly between the two groups.

Methodology 1

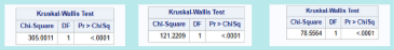
Normality Test



All three variables (radius_mean, texture_mean, smoothness_mean) failed normality tests ($p < 0.05$), indicating non-normal distributions. Therefore, non-parametric methods were selected for further analysis.

Methodology 2

Non-Parametric One-Way ANOVA



Non-parametric analysis reveals a statistically significant difference in radius_mean between malignant and benign cases (Wilcoxon Z = 17.464, $p < 0.0001$; Kruskal-Wallis $\chi2$ = 305.0011, $p < 0.0001$). This suggests that radius_mean is a very strong discriminative feature in breast cancer diagnosis.

texture_mean differs significantly between malignant and benign diagnoses (Wilcoxon Z = 11.0098, $p < 0.0001$; Kruskal-Wallis χ^2 = 121.2209, $p < 0.0001$). This suggests that texture_mean is a strong discriminative feature in breast cancer diagnosis.

smoothness_mean shows a significant difference between malignant and benign diagnoses (Wilcoxon Z = 8.8629, $p < 0.0001$; Kruskal-Wallis χ^2 = 78.5564, $p < 0.0001$). This suggests that smoothness_mean is a moderate-strong discriminative feature in breast cancer diagnosis

Methodology 3

T-Test

An independent samples T-test revealed a significant difference in radius_mean between malignant and benign cases (t = -22.21, $p < 0.0001$). The mean radius was substantially higher in malignant tumors (17.46) compared to benign (12.15), confirming its strong diagnostic value.

A Two samples T-test revealed a significant difference in texture_mean between malignant and benign cases (t = -11.02, $p < 0.0001$). Malignant tumors had a higher mean texture (21.60) compared to benign (17.91), supporting its role as a discriminative feature in breast cancer diagnosis.

An independent samples T-test revealed a significant difference in smoothness_mean between malignant and benign cases (t = -9.30, $p < 0.0001$). Although the mean difference is smaller than other features, it remains statistically meaningful, supporting its diagnostic relevance

Interpretation & Conclusion

All three features—radius, texture, and smoothness—show statistically significant differences between malignant and benign tumors. Among them, radius_mean is the strongest discriminator, followed by texture_mean, with smoothness_mean showing moderate but meaningful separation. These findings support the diagnostic relevance of these features and their potential use in breast cancer classification.