Project Report -1
Adult Census Data

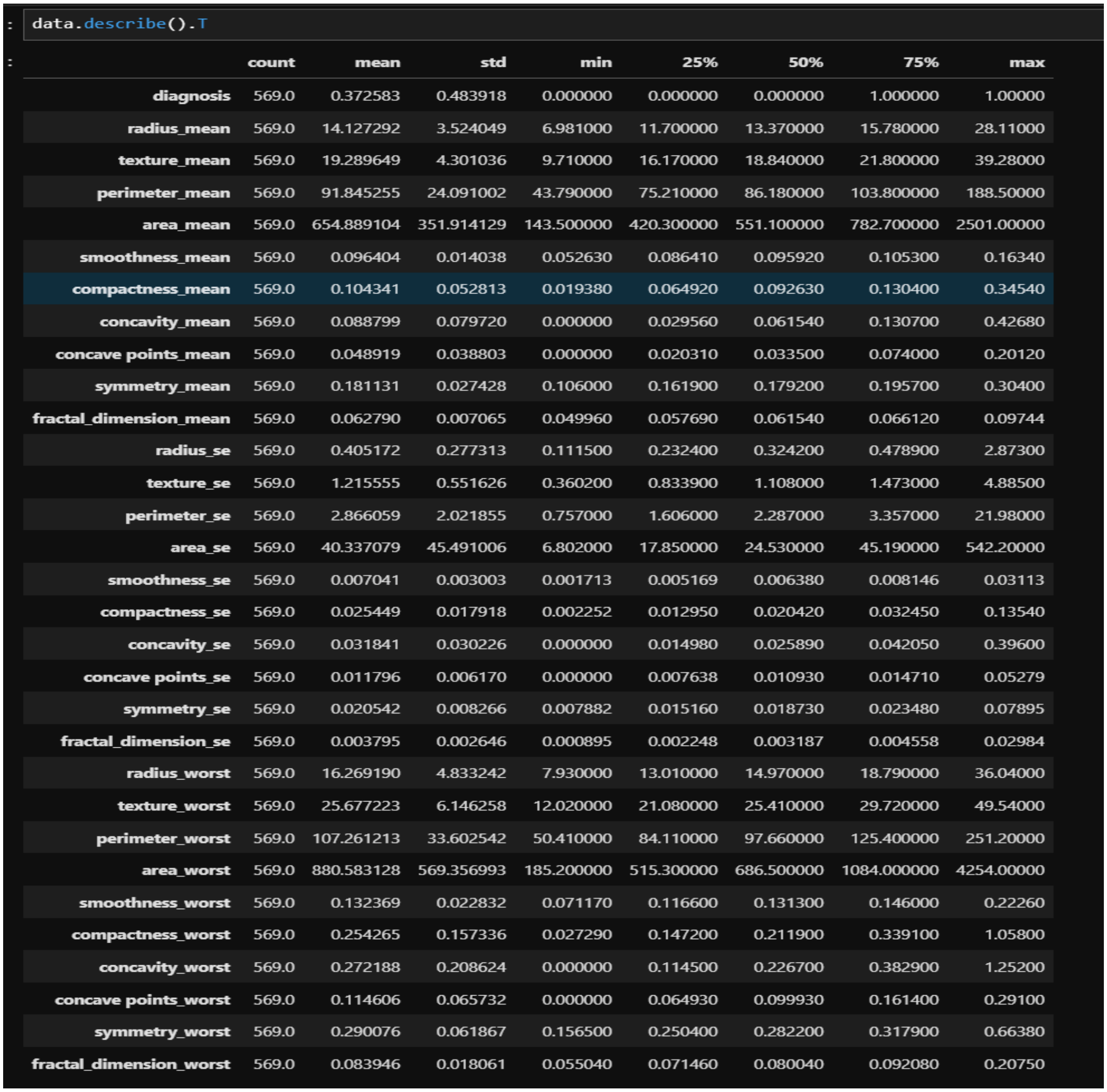
1. Provide brief details about the nature of your dataset. What is it about? What type of data are we encountering? How many entries and variables does the dataset comprise?

1. The WDBC dataset represent the characteristics of the cell nuclei present in the image of brest tissue . The end goal is to find whether it is benign or malignant

2. We are encountering with 31 numerical features like size , shape and texture of the cell , smoothness and 1 categorical features with B or M .

3. The dataset contains 569 entries and 32 columns .

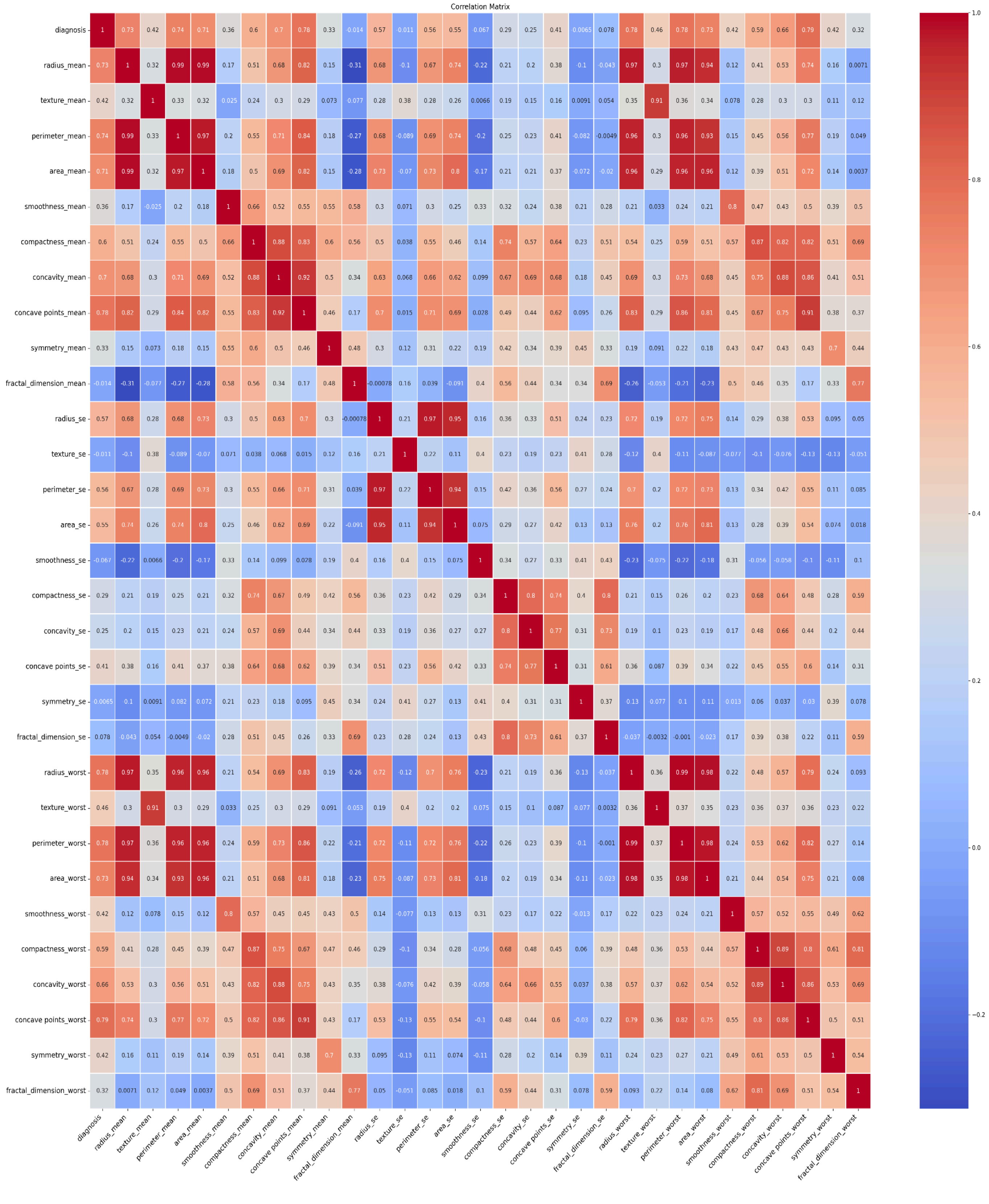
2. Provide the main statistics about the dataset entries (mean, std, number of missing values, etc.)



```
data.describe().T
```

	count	mean	std	min	25%	50%	75%	max
diagnosis	569.0	0.372583	0.483918	0.000000	0.000000	0.000000	1.000000	1.00000
radius_mean	569.0	14.127292	3.524049	6.981000	11.700000	13.370000	15.780000	28.11000
texture_mean	569.0	19.289649	4.301036	9.710000	16.170000	18.840000	21.800000	39.28000
perimeter_mean	569.0	91.845255	24.091002	43.790000	75.210000	86.180000	103.800000	188.50000
area_mean	569.0	654.889104	351.914129	143.500000	420.300000	551.100000	782.700000	2501.00000
smoothness_mean	569.0	0.096404	0.014038	0.052630	0.086410	0.095920	0.105300	0.16340
compactness_mean	569.0	0.104341	0.052813	0.019380	0.064920	0.092630	0.130400	0.34540
concavity_mean	569.0	0.088799	0.079720	0.000000	0.029560	0.061540	0.130700	0.42680
concave points_mean	569.0	0.048919	0.038803	0.000000	0.020310	0.033500	0.074000	0.20120
symmetry_mean	569.0	0.181131	0.027428	0.106000	0.161900	0.179200	0.195700	0.30400
fractal_dimension_mean	569.0	0.062790	0.007065	0.049960	0.057690	0.061540	0.066120	0.09744
radius_se	569.0	0.405172	0.277313	0.111500	0.232400	0.324200	0.478900	2.87300
texture_se	569.0	1.215555	0.551626	0.360200	0.833900	1.108000	1.473000	4.88500
perimeter_se	569.0	2.866059	2.021855	0.757000	1.606000	2.287000	3.357000	21.98000
area_se	569.0	40.337079	45.491006	6.802000	17.850000	24.530000	45.190000	542.20000
smoothness_se	569.0	0.007041	0.003003	0.001713	0.005169	0.006380	0.008146	0.03113
compactness_se	569.0	0.025449	0.017918	0.002252	0.012950	0.020420	0.032450	0.13540
concavity_se	569.0	0.031841	0.030226	0.000000	0.014980	0.025890	0.042050	0.39600
concave points_se	569.0	0.011796	0.006170	0.000000	0.007638	0.010930	0.014710	0.05279
symmetry_se	569.0	0.020542	0.008266	0.007882	0.015160	0.018730	0.023480	0.07895
fractal_dimension_se	569.0	0.003795	0.002646	0.000895	0.002248	0.003187	0.004558	0.02984
radius_worst	569.0	16.269190	4.833242	7.930000	13.010000	14.970000	18.790000	36.04000
texture_worst	569.0	25.677223	6.146258	12.020000	21.080000	25.410000	29.720000	49.54000
perimeter_worst	569.0	107.261213	33.602542	50.410000	84.110000	97.660000	125.400000	251.20000
area_worst	569.0	880.583128	569.356993	185.200000	515.300000	686.500000	1084.000000	4254.00000
smoothness_worst	569.0	0.132369	0.022832	0.071170	0.116600	0.131300	0.146000	0.22260
compactness_worst	569.0	0.254265	0.157336	0.027290	0.147200	0.211900	0.339100	1.05800
concavity_worst	569.0	0.272188	0.208624	0.000000	0.114500	0.226700	0.382900	1.25200
concave points_worst	569.0	0.114606	0.065732	0.000000	0.064930	0.099930	0.161400	0.29100
symmetry_worst	569.0	0.290076	0.061867	0.156500	0.250400	0.282200	0.317900	0.66380
fractal_dimension_worst	569.0	0.083946	0.018061	0.055040	0.071460	0.080040	0.092080	0.20750

3. Provide the correlation matrix and write your inference about it.



	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	fractal_dimension_mean	radius_se	texture_se	perimeter_se	area_se	smoothness_se	compactness_se	concavity_se	concave points_se	symmetry_se	fractal_dimension_se	radius_worst	texture_worst	perimeter_worst	area_worst	smoothness_worst	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	fractal_dimension_worst
diagnosis	1	0.73	0.42	0.74	0.71	0.36	0.6	0.7	0.78	0.33	-0.014	0.57	-0.011	0.56	0.55	-0.067	0.29	0.25	0.41	-0.0065	0.078	0.78	0.46	0.78	0.73	0.42	0.59	0.66	0.79	0.42	0.32
radius_mean	0.73	1	0.32	0.99	0.99	0.17	0.51	0.68	0.82	0.15	-0.31	0.68	-0.1	0.67	0.74	-0.22	0.21	0.2	0.38	-0.1	-0.043	0.97	0.3	0.97	0.94	0.12	0.41	0.53	0.74	0.16	0.0071
texture_mean	0.42	0.32	1	0.33	0.32	-0.025	0.24	0.3	0.29	0.073	-0.077	0.28	0.38	0.28	0.26	0.0066	0.19	0.15	0.16	0.0091	0.054	0.35	0.91	0.36	0.34	0.078	0.28	0.3	0.3	0.11	0.12
perimeter_mean	0.74	0.99	0.33	1	0.97	0.2	0.55	0.71	0.84	0.18	-0.27	0.68	-0.089	0.69	0.74	-0.2	0.25	0.23	0.41	-0.082	-0.0049	0.96	0.3	0.96	0.93	0.15	0.45	0.56	0.77	0.19	0.049
area_mean	0.71	0.99	0.32	0.97	1	0.18	0.5	0.69	0.82	0.15	-0.28	0.73	-0.07	0.73	0.8	-0.17	0.21	0.21	0.37	-0.072	-0.02	0.96	0.29	0.96	0.96	0.12	0.39	0.51	0.72	0.14	0.0037
smoothness_mean	0.36	0.17	-0.025	0.2	0.18	1	0.66	0.52	0.55	0.55	0.58	0.3	0.071	0.3	0.25	0.33	0.32	0.24	0.38	0.21	0.28	0.21	0.033	0.24	0.21	0.8	0.47	0.43	0.5	0.39	0.5
compactness_mean	0.6	0.51	0.24	0.55	0.5	0.66	1	0.88	0.83	0.6	0.56	0.5	0.038	0.55	0.46	0.14	0.74	0.57	0.64	0.23	0.51	0.54	0.25	0.59	0.51	0.57	0.87	0.82	0.82	0.51	0.69
concavity_mean	0.7	0.68	0.3	0.71	0.69	0.52	0.88	1	0.92	0.5	0.34	0.63	0.068	0.66	0.62	0.099	0.67	0.69	0.68	0.18	0.45	0.69	0.3	0.73	0.68	0.45	0.75	0.88	0.86	0.41	0.51
concave points_mean	0.78	0.82	0.29	0.84	0.82	0.55	0.83	0.92	1	0.46	0.17	0.7	0.015	0.71	0.69	0.028	0.49	0.44	0.62	0.095	0.26	0.83	0.29	0.86	0.81	0.45	0.67	0.75	0.91	0.38	0.37
symmetry_mean	0.33	0.15	0.073	0.18	0.15	0.55	0.6	0.5	0.46	1	0.48	0.3	0.12	0.31	0.22	0.19	0.42	0.34	0.39	0.45	0.33	0.19	0.091	0.22	0.18	0.43	0.47	0.43	0.43	0.7	0.44
fractal_dimension_mean	-0.014	-0.31	-0.077	-0.27	-0.28	0.58	0.56	0.34	0.17	0.48	1	-0.00078	0.16	0.039	-0.091	0.4	0.56	0.44	0.34	0.34	0.69	-0.26	-0.053	-0.21	-0.23	0.5	0.46	0.35	0.17	0.33	0.77
radius_se	0.57	0.68	0.28	0.68	0.73	0.3	0.5	0.63	0.7	0.3	-0.00078	1	0.21	0.97	0.95	0.16	0.36	0.33	0.51	0.24	0.23	0.72	0.19	0.72	0.75	0.14	0.29	0.38	0.53	0.095	0.05
texture_se	-0.011	-0.1	0.38	-0.089	-0.07	0.071	0.038	0.068	0.015	0.12	0.16	0.21	1	0.22	0.11	0.4	0.23	0.19	0.23	0.41	0.28	-0.12	0.4	-0.11	-0.087	-0.077	-0.1	-0.076	-0.13	-0.13	-0.051
perimeter_se	0.56	0.67	0.28	0.69	0.73	0.3	0.55	0.66	0.71	0.31	0.039	0.97	0.22	1	0.94	0.15	0.42	0.36	0.56	0.27	0.24	0.7	0.2	0.72	0.73	0.13	0.34	0.42	0.55	0.11	0.085
area_se	0.55	0.74	0.26	0.74	0.8	0.25	0.46	0.62	0.69	0.22	-0.091	0.95	0.11	0.94	1	0.075	0.29	0.27	0.42	0.13	0.13	0.76	0.2	0.76	0.81	0.13	0.28	0.39	0.54	0.074	0.018
smoothness_se	-0.067	-0.22	0.0066	-0.2	-0.17	0.33	0.14	0.099	0.028	0.19	0.4	0.16	0.4	0.15	0.075	1	0.34	0.27	0.33	0.41	0.43	-0.23	-0.075	-0.22	-0.18	0.31	-0.056	-0.058	-0.1	-0.11	0.1
compactness_se	0.29	0.21	0.19	0.25	0.21	0.32	0.74	0.67	0.49	0.42	0.56	0.36	0.23	0.42	0.29	0.34	1	0.8	0.74	0.4	0.8	0.21	0.15	0.26	0.2	0.23	0.68	0.64	0.48	0.28	0.59
concavity_se	0.25	0.2	0.15	0.23	0.21	0.24	0.57	0.69	0.44	0.34	0.44	0.33	0.19	0.36	0.27	0.27	0.8	1	0.77	0.31	0.73	0.19	0.1	0.23	0.19	0.17	0.48	0.66	0.44	0.2	0.44
concave points_se	0.41	0.38	0.16	0.41	0.37	0.38	0.64	0.68	0.62	0.39	0.34	0.51	0.23	0.56	0.42	0.33	0.74	0.77	1	0.31	0.61	0.36	0.087	0.39	0.34	0.22	0.45	0.55	0.6	0.14	0.31
symmetry_se	-0.0065	-0.1	0.0091	-0.082	-0.072	0.21	0.23	0.18	0.095	0.45	0.34	0.24	0.41	0.27	0.13	0.41	0.4	0.31	0.31	1	0.37	-0.13	-0.077	-0.1	-0.11	-0.013	0.06	0.037	-0.03	0.39	0.078
fractal_dimension_se	0.078	-0.043	0.054	-0.0049	-0.02	0.28	0.51	0.45	0.26	0.33	0.69	0.23	0.28	0.24	0.13	0.43	0.8	0.73	0.61	0.37	1	-0.037	-0.0032	-0.001	-0.023	0.17	0.39	0.38	0.22	0.11	0.59
radius_worst	0.78	0.97	0.35	0.96	0.96	0.21	0.54	0.69	0.83	0.19	-0.26	0.72	-0.12	0.7	0.76	-0.23	0.21	0.19	0.36	-0.13	-0.037	1	0.36	0.99	0.98	0.22	0.48	0.57	0.79	0.24	0.093
texture_worst	0.46	0.3	0.91	0.3	0.29	0.033	0.25	0.3	0.29	0.091	-0.053	0.19	0.4	0.2	0.2	-0.075	0.15	0.1	0.087	-0.077	-0.0032	0.36	1	0.37	0.35	0.23	0.36	0.37	0.36	0.23	0.22
perimeter_worst	0.78	0.97	0.36	0.96	0.96	0.24	0.59	0.73	0.86	0.22	-0.21	0.72	-0.11	0.72	0.76	-0.22	0.26	0.23	0.39	-0.1	-0.001	0.99	0.37	1	0.98	0.24	0.53	0.62	0.82	0.27	0.14
area_worst	0.73	0.94	0.34	0.93	0.96	0.21	0.51	0.68	0.81	0.18	-0.23	0.75	-0.087	0.73	0.81	-0.18	0.2	0.19	0.34	-0.11	-0.023	0.98	0.35	0.98	1	0.21	0.44	0.54	0.75	0.21	0.08
smoothness_worst	0.42	0.12	0.078	0.15	0.12	0.8	0.57	0.45	0.45	0.43	0.5	0.14	-0.077	0.13	0.13	0.31	0.23	0.17	0.22	-0.013	0.17	0.22	0.23	0.24	0.21	1	0.57	0.52	0.55	0.49	0.62
compactness_worst	0.59	0.41	0.28	0.45	0.39	0.47	0.87	0.75	0.67	0.47	0.46	0.29	-0.1	0.34	0.28	-0.056	0.68	0.48	0.45	0.06	0.39	0.48	0.36	0.53	0.44	0.57	1	0.89	0.8	0.61	0.81
concavity_worst	0.66	0.53	0.3	0.56	0.51	0.43	0.82	0.88	0.75	0.43	0.35	0.38	-0.076	0.42	0.39	-0.058	0.64	0.66	0.55	0.037	0.38	0.57	0.37	0.62	0.54	0.52	0.89	1	0.86	0.53	0.69
concave points_worst	0.79	0.74	0.3	0.77	0.72	0.5	0.82	0.86	0.91	0.43	0.17	0.53	-0.13	0.55	0.54	-0.1	0.48	0.44	0.6	-0.03	0.22	0.79	0.36	0.82	0.75	0.55	0.8	0.86	1	0.5	0.51
symmetry_worst	0.42	0.16	0.11	0.19	0.14	0.39	0.51	0.41	0.38	0.7	0.33	0.095	-0.13	0.11	0.074	-0.11	0.28	0.2	0.14	0.39	0.11	0.24	0.23	0.27	0.21	0.49	0.61	0.53	0.5	1	0.54
fractal_dimension_worst	0.32	0.0071	0.12	0.049	0.0037	0.5	0.69	0.51	0.37	0.44	0.77	0.05	-0.051	0.085	0.018	0.1	0.59	0.44	0.31	0.078	0.59	0.093	0.22	0.14	0.08	0.62	0.81	0.69	0.51	0.54	1

1.Features like radius_mean, texture_mean, perimeter_mean, and area_mean show higher mean values compared to other features.

2.In the correlation matrix concave_point_worst , concavity_point_mean are highly correlated with diagnosis . That is it strongly predict the malignancy .

3.Other than concave_point_worst, perimeter_mean , radius_mean, area_mean ,area_worst, perimeter_worst,radius worst are strongly correlated with malignancy .

4. From the row radius_mean the perimeter_mean , area_mean are highly correlated to each other . This indicates that it contains the redundant values .

To reduce the redundancy we need to take PCA to combine into one feature .

5. These redundant data contains in perimeter_mean , area_mean,radius_worst,perimeter_worst columns respectively .

6.concavity_se , compactness_se show weak correlations with diagnosis.

7. fractional_dimension_mean is likely associated with benign tumors .