

STAT 656

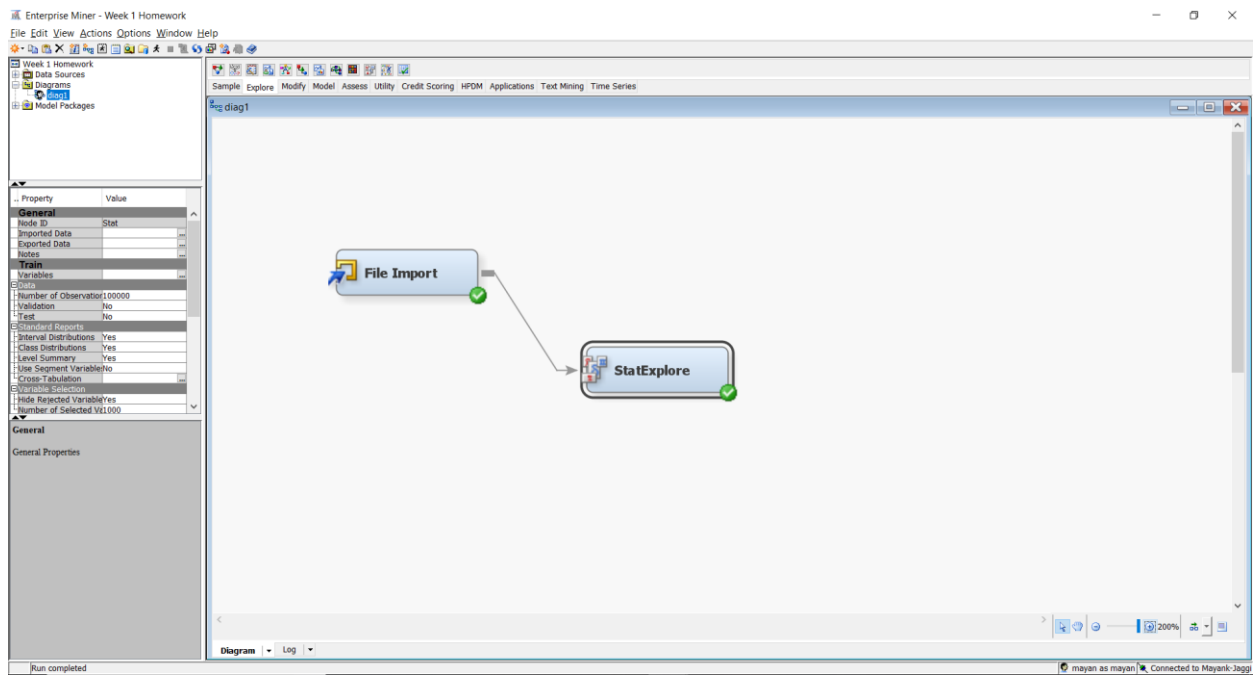
Homework 1

Name-Mayank Jaggi

UIN-526005299

PART 1: SAS ENTERPRISE MINER

Screenshot of Project Window



Output Table

Note: The outliers and attributes are encircled

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
R1	INPUT	0.029226	0.023029	207	1	0.0015	0.0228	0.1371	2.123538	5.551496
R10	INPUT	0.208356	0.134734	207	1	0.0113	0.181	0.7106	1.276265	1.794883
R11	INPUT	0.236013	0.132705	208	0	0.0289	0.2245	0.7342	0.987813	1.190935
R12	INPUT	0.250928	0.14012	206	2	0.0236	0.2484	0.706	0.590043	0.034427
R13	INPUT	0.273305	0.140962	208	0	0.0184	0.2624	0.7131	0.736584	0.400126
R14	INPUT	0.296568	0.164474	208	0	0.0273	0.2803	0.997	1.022369	1.5768
R15	INPUT	0.321163	0.20906	208	0	0.0031	0.2794	1.2	0.906731	0.803433
R16	INPUT	0.378487	0.23265	208	0	0.0162	0.3019	0.9988	0.694671	-0.51315
R17	INPUT	0.408291	0.283104	208	0	-1	0.3052	1.2	0.079517	2.058875
R18	INPUT	0.445986	0.282101	207	1	-1	0.3657	1.2	-0.09959	2.213002
R19	INPUT	0.505773	0.260206	208	0	0.0494	0.4348	1.2	0.324212	-1.04624
R2	INPUT	0.03855	0.032999	207	1	0.0006	0.0308	0.2339	2.149882	7.062024
R20	INPUT	0.564008	0.264619	208	0	0.0656	0.542	1.2	-0.04838	-1.1334
R21	INPUT	0.612906	0.265047	208	0	0.0512	0.6079	1.2	-0.12294	-0.8346
R22	INPUT	0.625237	0.257671	208	0	0.0219	0.6628	1.2	-0.3562	-0.89965
R23	INPUT	0.639283	0.277094	208	0	-1	0.6967	1.2	-1.33026	4.740771
R24	INPUT	0.654385	0.289246	208	0	-1	0.6933	1.2	-2.07752	8.856107
R25	INPUT	0.668693	0.274693	208	0	-1	0.7159	1.2	-1.52203	5.636179
R26	INPUT	0.661405	0.32998	208	0	-1	0.7474	1	-2.57683	10.30843
R27	INPUT	0.615616	0.418254	208	0	-1	0.6997	1	-2.46634	7.173519
R28	INPUT	0.655563	0.329267	208	0	-1	0.708	1	-2.54732	10.28732
R29	INPUT	0.635343	0.269697	208	0	-1	0.679	1.2	-1.25387	5.346339
R3	INPUT	0.067776	0.345778	208	0	0.0015	0.0344	5	14.1574	202.8304
R30	INPUT	0.508028	0.220749	208	0	0.0613	0.6069	1	-0.13239	-0.68977
R31	INPUT	0.504475	0.213992	208	0	0.0402	0.4901	0.9657	0.259797	-0.64397
R32	INPUT	0.43904	0.213237	208	0	0.0404	0.4289	0.9306	0.31342	-0.66427
R33	INPUT	0.41722	0.206513	208	0	0.0477	0.3903	1	0.423864	-0.51471
R34	INPUT	0.403233	0.231242	208	0	0.0212	0.3497	0.9647	0.563296	-0.55821
R35	INPUT	0.383917	0.27573	208	0	-1	0.31	1.2	-0.02399	2.133442
R36	INPUT	0.385809	0.266723	208	0	0.008	0.3195	1.2	0.670385	-0.46741
R37	INPUT	0.363807	0.239912	208	0	0.0351	0.3039	0.9497	0.676994	-0.55324
R38	INPUT	0.339657	0.212973	208	0	0.0383	0.3104	1	1.033366	0.623908
R39	INPUT	0.3258	0.199075	208	0	0.0371	0.2829	0.9857	0.908835	0.457845
R4	INPUT	0.048842	0.08663	208	0	-1	0.0432	0.4264	-8.07841	105.9171
R40	INPUT	0.311207	0.178662	208	0	0.0117	0.2771	0.9297	0.86128	0.761169
R41	INPUT	0.305282	0.171411	171	37	0.0488	0.2707	0.8995	0.858536	0.42353
R42	INPUT	0.278929	0.168887	207	1	0.0056	0.2458	0.8246	0.852803	0.521339
R43	INPUT	0.247733	0.138262	207	1	0.0159	0.224	0.7733	0.930042	1.099881
R44	INPUT	0.215109	0.132775	207	1	0.0255	0.1778	0.7762	1.252617	1.434399
R45	INPUT	0.19873	0.151935	205	3	0.0046	0.1473	0.7034	1.361257	1.040761
R46	INPUT	0.157999	0.138392	173	35	0.0025	0.1124	0.7292	1.733716	2.9245
R47	INPUT	0.123044	0.086743	207	1	0.0073	0.1018	0.5522	1.805143	4.253426
R48	INPUT	0.091866	0.062241	207	1	0.0041	0.0785	0.3339	1.288905	1.867497
R49	INPUT	0.05218	0.035858	207	1	0.0021	0.0449	0.1981	1.283602	1.799271
R5	INPUT	0.075202	0.055552	208	0	0.0067	0.062	0.401	2.018141	7.044579
R50	INPUT	0.020523	0.013624	207	1	0.0006	0.0179	0.0825	1.781092	4.452958
R51	INPUT	0.01616	0.012013	206	2	0.0009	0.014	0.1004	2.728749	14.36888
R52	INPUT	0.01342	0.009634	208	0	0.0008	0.0113	0.0709	2.093976	7.006616
R53	INPUT	0.010709	0.00706	208	0	0.0005	0.0095	0.039	1.060572	1.248923
R54	INPUT	0.010941	0.007301	208	0	0.001	0.0093	0.0352	1.09309	0.9359
R55	INPUT	0.009321	0.007091	207	1	0.0006	0.0075	0.0447	1.786432	4.518894
R56	INPUT	0.008222	0.005736	208	0	0.0004	0.0068	0.0394	1.780805	5.367747
R57	INPUT	0.00782	0.005785	208	0	0.0003	0.0059	0.0355	1.65309	3.611717
R58	INPUT	0.007949	0.00647	208	0	0.0003	0.0058	0.044	2.09833	6.762729
R59	INPUT	0.007941	0.006181	208	0	0.0001	0.0063	0.0364	1.737506	3.740889
R6	INPUT	0.10457	0.059105	208	0	0.0102	0.0921	0.3823	1.248166	2.572759
R60	INPUT	0.006507	0.005031	208	0	0.0006	0.0053	0.0439	2.775754	14.74067
R7	INPUT	0.121685	0.06197	206	2	0.0033	0.1056	0.3729	0.973608	1.455237
R8	INPUT	0.134799	0.085152	208	0	0.0055	0.1119	0.459	1.481107	2.802407
R9	INPUT	0.178003	0.118387	208	0	0.0075	0.1522	0.6828	1.63387	3.592943
object	TARGET	0.533654	0.50007	208	0	0	1	1	-0.1359	-2.00086

PART 2: PYTHON

PYTHON PROGRAM

```
import pandas as pd
path="sonar_hw1.csv"    #data file name
df=pd.read_csv(path)
df1=df.drop(columns='object') # removing the column object as its the target

missing=df1.shape[0]-pd.DataFrame(df1.count(),columns=['Missing'])
# missing values is difference between the number of rows and the count of non
missing observations

median=pd.DataFrame(df1.median(),columns=['Median'])
maximum=pd.DataFrame(df1.max(),columns=['Maximum'])
minimum=pd.DataFrame(df1.min(),columns=['Minimum'])

outlier_high=df1[df1>1].count()    #count the number of observations above 1
outlier_low=df1[df1<0].count()    #count the number of observations below 0
outlier=pd.DataFrame(outlier_high+outlier_low,columns=['Outlier']) #combine the high
and low outliers

pd.concat([missing,outlier,minimum,median,maximum], axis=1) # combining the
dataframes by column
```

OUTPUT

Out[91]:

	Missing	Outlier	Minimum	Median	Maximum
R1	1	0	0.0015	0.02280	0.1371
R2	1	0	0.0006	0.03080	0.2339
R3	0	1	0.0015	0.03450	5.0000
R4	0	1	-1.0000	0.04360	0.4264
R5	0	0	0.0067	0.06250	0.4010
R6	0	0	0.0102	0.09215	0.3823
R7	2	0	0.0033	0.10695	0.3729
R8	0	0	0.0055	0.11210	0.4590
R9	0	0	0.0075	0.15225	0.6828
R10	1	0	0.0113	0.18100	0.7106
R11	0	0	0.0289	0.22480	0.7342
R12	2	0	0.0236	0.24905	0.7060
R13	0	0	0.0184	0.26395	0.7131
R14	0	0	0.0273	0.28110	0.9970
R15	0	1	0.0031	0.28170	1.2000
R16	0	0	0.0162	0.30470	0.9988
R17	0	3	-1.0000	0.30600	1.2000
R18	1	3	-1.0000	0.36570	1.2000

R19	0	1	0.0494	0.43495	1.2000
R20	0	1	0.0656	0.54250	1.2000
R21	0	4	0.0512	0.61770	1.2000
R22	0	1	0.0219	0.66490	1.2000
R23	0	3	-1.0000	0.69770	1.2000
R24	0	3	-1.0000	0.69435	1.2000
R25	0	4	-1.0000	0.71800	1.2000
R26	0	4	-1.0000	0.74750	1.0000
R27	0	9	-1.0000	0.70030	1.0000
R28	0	4	-1.0000	0.70920	1.0000
R29	0	4	-1.0000	0.67905	1.2000
R30	0	0	0.0613	0.60715	1.0000
R31	0	0	0.0482	0.49035	0.9657
R32	0	0	0.0404	0.42960	0.9306
R33	0	0	0.0477	0.39120	1.0000
R34	0	0	0.0212	0.35105	0.9647
R35	0	2	-1.0000	0.31040	1.2000
R36	0	1	0.0080	0.32115	1.2000
R37	0	0	0.0351	0.30630	0.9497
R38	0	0	0.0383	0.31270	1.0000
R39	0	0	0.0371	0.28350	0.9857
R40	0	0	0.0117	0.27805	0.9297
R41	37	0	0.0488	0.27070	0.8995
R42	1	0	0.0056	0.24580	0.8246
R43	1	0	0.0159	0.22400	0.7733
R44	1	0	0.0255	0.17780	0.7762
R45	3	0	0.0046	0.14870	0.7034
R46	35	0	0.0025	0.11360	0.7292
R47	1	0	0.0073	0.10180	0.5522
R48	1	0	0.0041	0.07850	0.3339
R49	1	0	0.0021	0.04490	0.1981
R50	1	0	0.0006	0.01790	0.0825
R51	2	0	0.0009	0.01400	0.1004
R52	0	0	0.0008	0.01140	0.0709
R53	0	0	0.0005	0.00955	0.0390
R54	0	0	0.0010	0.00930	0.0352
R55	1	0	0.0006	0.00750	0.0447
R56	0	0	0.0004	0.00685	0.0394
R57	0	0	0.0003	0.00595	0.0355
R58	0	0	0.0003	0.00580	0.0440
R59	0	0	0.0001	0.00640	0.0364
R60	0	0	0.0006	0.00530	0.0439