East-West airlines problem

Business Objective:- Business objective of this project would be to cluster customers of similar type together, so that better business decisions can be taken.

Constraints:-

Dataset Description:-

|  |  |  |  |
| --- | --- | --- | --- |
| Name of feature | Description | Type | Relevance |
| ID# | index | Quantitative,Nominal | index |
| Balance | Balance miles | Quantitative,Nominal | Number of miles eligible for award travel |
| Qual\_Miles | Qualifing miles | Quantitative,Nominal | Number of miles counted as qualifying for Topflight status |
| cc1\_miles | Miles earned | Categorical | Number of miles earned with freq. flyer credit card in the past 12 months |
| cc2\_miles | Miles earned | Categorical | Number of miles earned with Rewards credit card in the past 12 months: |
| cc3\_miles | Miles earned | Categorical | Number of miles earned with Small Business credit card in the past 12 months: |
| Bonus\_miles | Miles earned | Quantitative | Number of miles earned from non-flight bonus transactions in the past 12 months |
| Bonus\_trans | Bonus transactions | Quantitative | Number of non-flight bonus transactions in the past 12 months |
| Flight\_miles\_12mo | Flight miles | Quantitative | Number of flight miles in the past 12 months |
| Flight\_trans\_12 | Flight transactions | Quantitative | Number of flight transactions in the past 12 months |
| Days\_since\_enroll | Number of days | Quantitative | Number of days since Enroll\_date |
| Award? | Dummy variable | Quantitative | Dummy variable for Last\_award (1=not null, 0=null) |

Data Pre-processing:-

(1) Shape and Description of data

(3999, 12)

ID# Balance Qual\_miles cc1\_miles cc2\_miles \

count 3999.000000 3.999000e+03 3999.000000 3999.000000 3999.000000

mean 2014.819455 7.360133e+04 144.114529 2.059515 1.014504

std 1160.764358 1.007757e+05 773.663804 1.376919 0.147650

min 1.000000 0.000000e+00 0.000000 1.000000 1.000000

25% 1010.500000 1.852750e+04 0.000000 1.000000 1.000000

50% 2016.000000 4.309700e+04 0.000000 1.000000 1.000000

75% 3020.500000 9.240400e+04 0.000000 3.000000 1.000000

max 4021.000000 1.704838e+06 11148.000000 5.000000 3.000000

cc3\_miles Bonus\_miles Bonus\_trans Flight\_miles\_12mo \

count 3999.000000 3999.000000 3999.00000 3999.000000

mean 1.012253 17144.846212 11.60190 460.055764

std 0.195241 24150.967826 9.60381 1400.209171

min 1.000000 0.000000 0.00000 0.000000

25% 1.000000 1250.000000 3.00000 0.000000

50% 1.000000 7171.000000 12.00000 0.000000

75% 1.000000 23800.500000 17.00000 311.000000

max 5.000000 263685.000000 86.00000 30817.000000

Flight\_trans\_12 Days\_since\_enroll Award?

count 3999.000000 3999.00000 3999.000000

mean 1.373593 4118.55939 0.370343

std 3.793172 2065.13454 0.482957

min 0.000000 2.00000 0.000000

25% 0.000000 2330.00000 0.000000

50% 0.000000 4096.00000 0.000000

75% 1.000000 5790.50000 1.000000

max 53.000000 8296.00000 1.000000

(2) Deleted duplicates, there were none as observed

(3) variance

ID# 1.347374e+06

Balance 1.015573e+10

Qual\_miles 5.985557e+05

cc1\_miles 1.895907e+00

cc2\_miles 2.180060e-02

cc3\_miles 3.811896e-02

Bonus\_miles 5.832692e+08

Bonus\_trans 9.223317e+01

Flight\_miles\_12mo 1.960586e+06

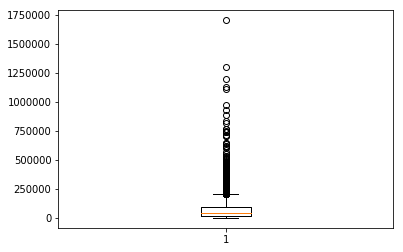
Flight\_trans\_12 1.438816e+01

Days\_since\_enroll 4.264781e+06

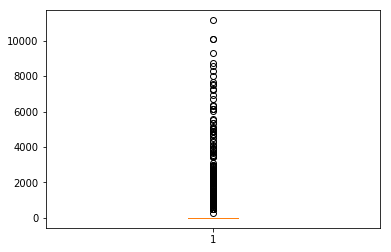
Award? 2.332473e-01

(4) Outliers

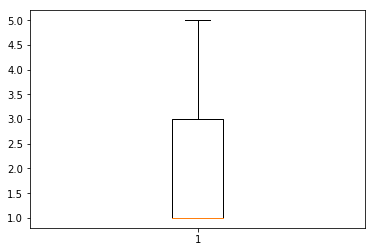
Balance column



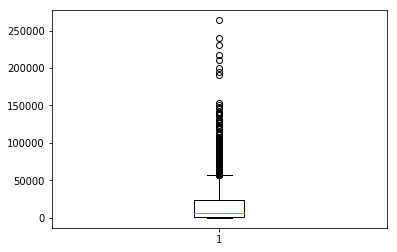
Qual\_miles



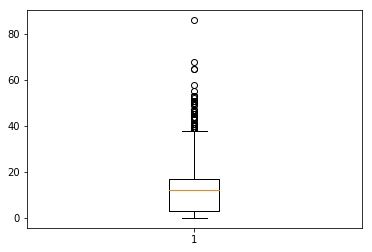
Cc1\_miles



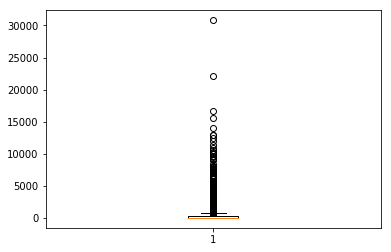
Bonus\_miles



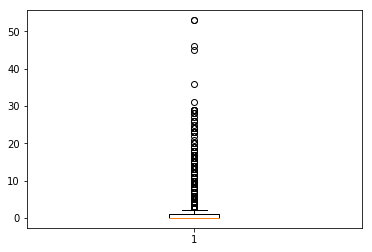
Bonus\_trans



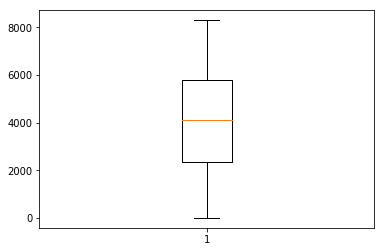
Flight\_miles\_12mo



Flight\_trans\_12



Days\_since\_enroll



All outliers were removed and normalization was performed, which resulted in

Days\_since\_enroll Balance ... cc2\_miles cc3\_miles

count 3999.000000 3999.000000 ... 3999.000000 3999.000000

mean 0.496330 0.313793 ... 0.007252 0.003063

std 0.248991 0.288722 ... 0.073825 0.048810

min 0.000000 0.000000 ... 0.000000 0.000000

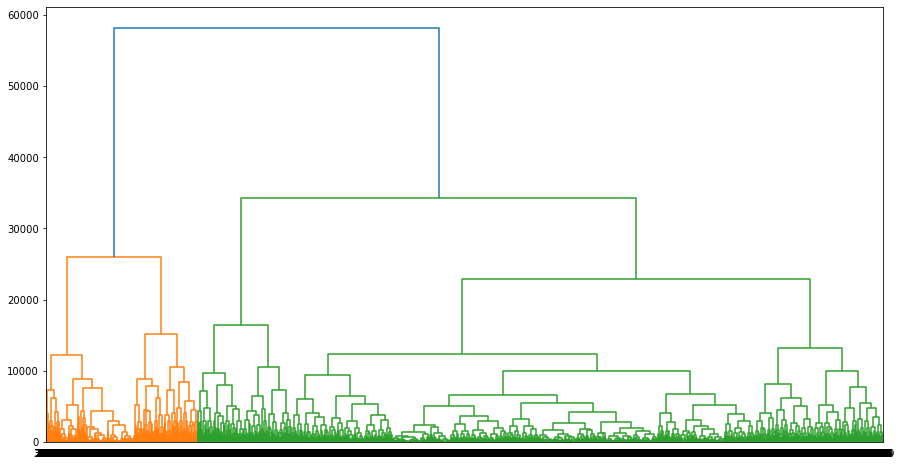
25% 0.280685 0.091170 ... 0.000000 0.000000

50% 0.493610 0.212072 ... 0.000000 0.000000

75% 0.697914 0.454702 ... 0.000000 0.000000

max 1.000000 1.000000 ... 1.000000 1.000000

(6)Dendrogram was plotted to get understanding of clustering



(7) CLustering was performed and clusters column was added to original dataframe

Days\_since\_enroll clust Qual\_miles ... Bonus\_miles cc2\_miles cc3\_miles

0 7000.0 1 0.0 ... 174.0 1.0 1.0

1 6968.0 1 0.0 ... 215.0 1.0 1.0

2 7034.0 1 0.0 ... 4123.0 1.0 1.0

3 6952.0 1 0.0 ... 500.0 1.0 1.0

4 6935.0 0 0.0 ... 43300.0 1.0 1.0

(8) New dataframe was stored as an excel file