03. EXPLORATORY DATA ANALYSIS

A. QUANTITATIVE DATA ANALYSIS

<mark>A1. QUANTITATIVE – UNIVARIATE ANALYSIS</mark>

Q1. What is the distribution of passenger age?

[57]:		Age_Range_Start	Age_Range_End	Frequency
	0	15.5	20.7	122
	1	20.7	25.9	629
	2	25.9	31.1	7063
	3	31.1	36.3	11030
	4	36.3	41.5	13004
	5	41.5	46.7	11232
	6	46.7	51.9	8526
	7	51.9	57.1	6552
	8	57.1	62.3	2644
	9	62.3	67.5	2186

Q2. What is the distribution of flight count?

[59]:		FLIGHT_COUNT Range Start	FLIGHT_COUNT Range End	Frequency
	0	2.000000	3.033333	16871
	1	3.033333	4.066667	5494
	2	4.066667	5.100000	4321
	3	5.100000	6.133333	3676
	4	6.133333	7.166667	3072
	5	7.166667	8.200000	2800
	6	8.200000	9.233333	2298
	7	9.233333	10.266667	2145
	8	10.266667	11.300000	1854
	9	11.300000	12.333333	1703
	10	12.333333	13.366667	1522
	11	13.366667	14.400000	1389
	12	14.400000	15.433333	1180
	13	15.433333	16.466667	1109
	14	16.466667	17.500000	1008
	15	17.500000	18.533333	888

	FLIGHT_COUNT Range Start	FLIGHT_COUNT Range End	Frequency
16	18.533333	19.566667	817
17	19.566667	20.600000	744
18	20.600000	21.633333	682
19	21.633333	22.666667	638
20	22.666667	23.700000	596
21	23.700000	24.733333	578
22	24.733333	25.766667	489
23	25.766667	26.800000	446
24	26.800000	27.833333	404
25	27.833333	28.866667	412
26	28.866667	29.900000	364
27	29.900000	30.933333	327
28	30.933333	31.966667	313
29	31.966667	33.000000	4848

Q3. What is the distribution of total flight distance?

	SEG_KM_SUM Range Start	SEG_KM_SUM Range End	Frequency
0	368.0000	1890.9875	2629
1	1890.9875	3413.9750	6930
2	3413.9750	4936.9625	6875
3	4936.9625	6459.9500	5678
4	6459.9500	7982.9375	4485
5	7982.9375	9505.9250	3771
6	9505.9250	11028.9125	3361
7	11028.9125	12551.9000	2716
8	12551.9000	14074.8875	2435
9	14074.8875	15597.8750	2205
10	15597.8750	17120.8625	1927
11	17120.8625	18643.8500	1741
12	18643.8500	20166.8375	1461
13	20166.8375	21689.8250	1399
14	21689.8250	23212.8125	1219
15	23212.8125	24735.8000	1179
16	24735.8000	26258.7875	1022
17	26258.7875	27781.7750	895
18	27781.7750	29304.7625	834
19	29304.7625	30827.7500	739
20	30827.7500	32350.7375	686
21	32350.7375	33873.7250	574
22	33873.7250	35396.7125	533
23	35396.7125	36919.7000	501
24	36919.7000	38442.6875	482
25	38442.6875	39965.6750	486
26	39965.6750	41488.6625	380
27	41488.6625	43011.6500	363
28	43011.6500	44534.6375	369
29	44534.6375	46057.6250	5113

Q4. What is the distribution of average discount?

	avg_discount Range Start	avg_discount Range End	Frequency
0	0.315779	0.342109	752
1	0.342109	0.368440	341
2	0.368440	0.394770	506
3	0.394770	0.421101	941
4	0.421101	0.447431	741
5	0.447431	0.473762	1044
6	0.473762	0.500092	1489
7	0.500092	0.526423	1614
8	0.526423	0.552753	2153
9	0.552753	0.579084	2405
10	0.579084	0.605415	3007
11	0.605415	0.631745	3438
12	0.631745	0.658076	3877
13	0.658076	0.684406	4297
14	0.684406	0.710737	4686
15	0.710737	0.737067	4621
16	0.737067	0.763398	4499
17	0.763398	0.789728	3981
18	0.789728	0.816059	3596
19	0.816059	0.842389	2862
20	0.842389	0.868720	2414
21	0.868720	0.895050	1849
22	0.895050	0.921381	1727
23	0.921381	0.947711	1095
24	0.947711	0.974042	931
25	0.974042	1.000372	1179
26	1.000372	1.026703	219
27	1.026703	1.053033	265
28	1.053033	1.079364	175
29	1.079364	1.105694	2284

Q5. What is the distribution of loyalty points?

	Points_Sum Range Start	Points_Sum Range End	Frequency
0	0.00	6318.75	31471
1	6318.75	12637.50	13586
2	12637.50	18956.25	6587
3	18956.25	25275.00	3688
4	25275.00	31593.75	7656

Q6. What is the distribution of exchange count?

	EXCHANGE_COUNT Range Start	EXCHANGE_COUNT Range End	Frequency
0	-0.5	-0.3	0
1	-0.3	-0.1	0
2	-0.1	0.1	62988
3	0.1	0.3	0
4	0.3	0.5	0
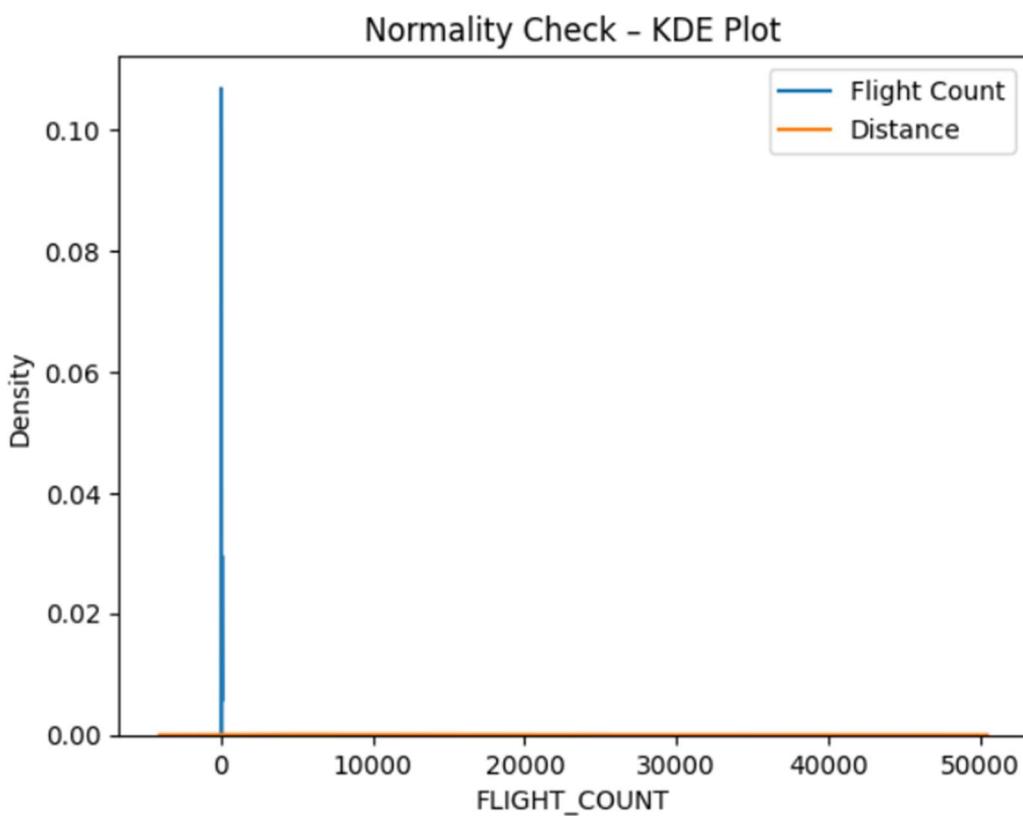
Q7. What is the distribution of membership duration?

	Range	Frequency
0	0 – 687 days	20012
1	687 – 1374 days	16290
2	1374 – 2061 days	11252
3	2061 – 2748 days	9964
4	2748 – 3436 days	5049

Q9. Are there outliers in numerical variables?
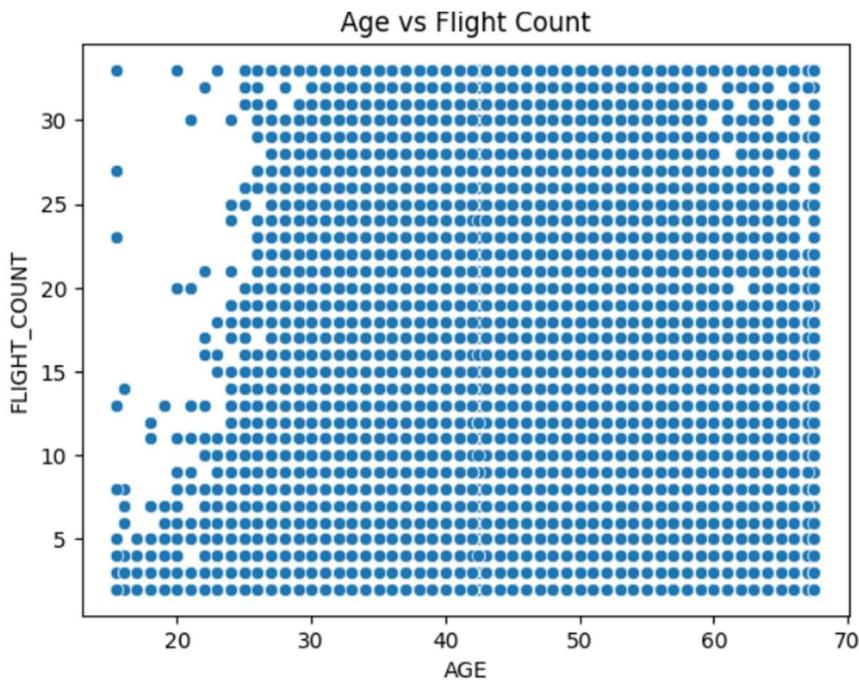
|	Feature	Q1	Q3	IQR	Lower Bound	Upper Bound	Outlier Count
0	AGE	35.0	48.0	13.0	15.5	67.5	0
1	FLIGHT_COUNT	3.0	15.0	12.0	-15.0	33.0	0
2	SEG_KM_SUM	4732.0	21206.0	16474.0	-19979.0	45917.0	4780
3	Points_Sum	2764.0	14249.0	11485.0	-14463.5	31476.5	5494

Q10. Are numerical variables normally distributed?
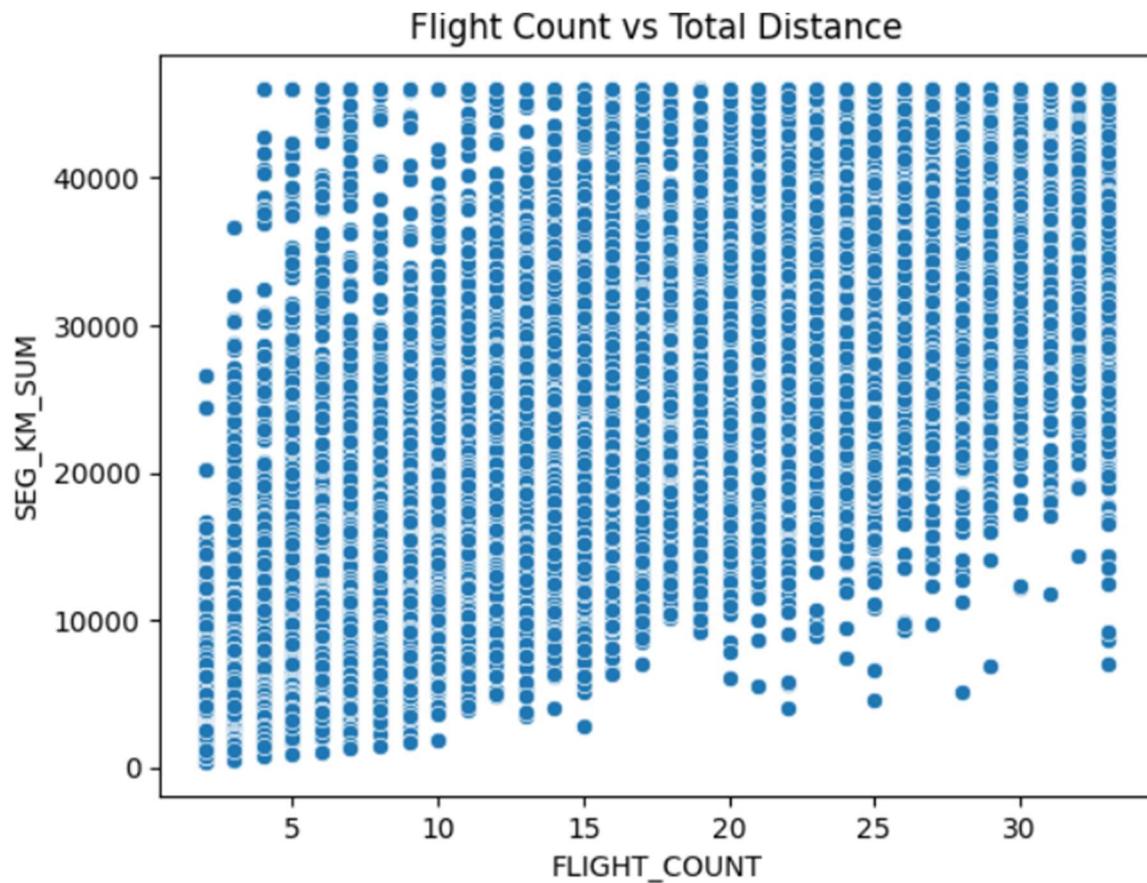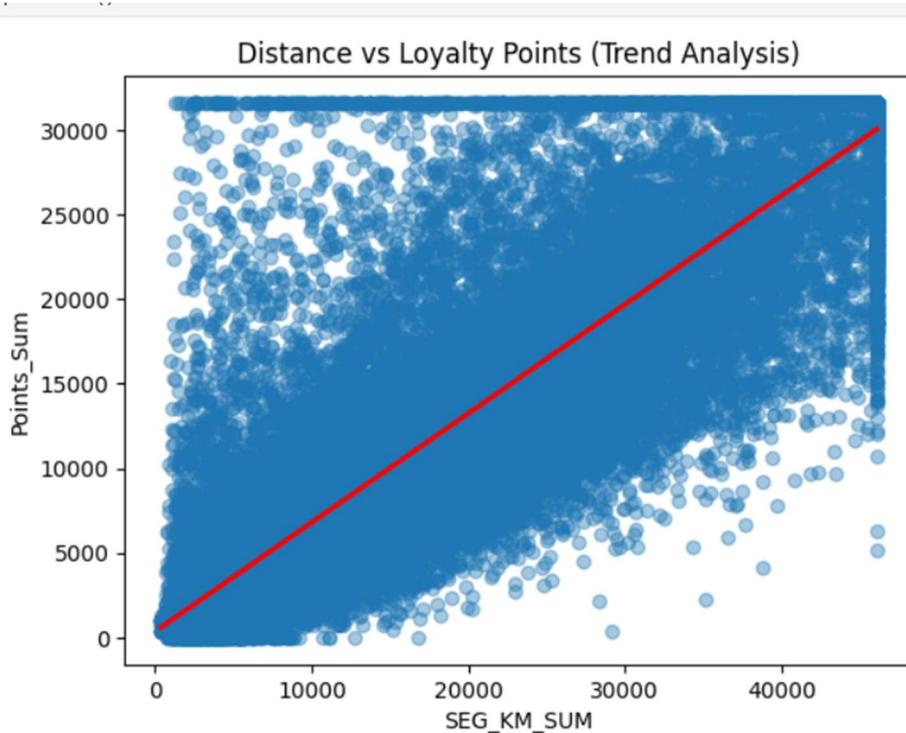
Q11. What is the relationship between age and flight count?
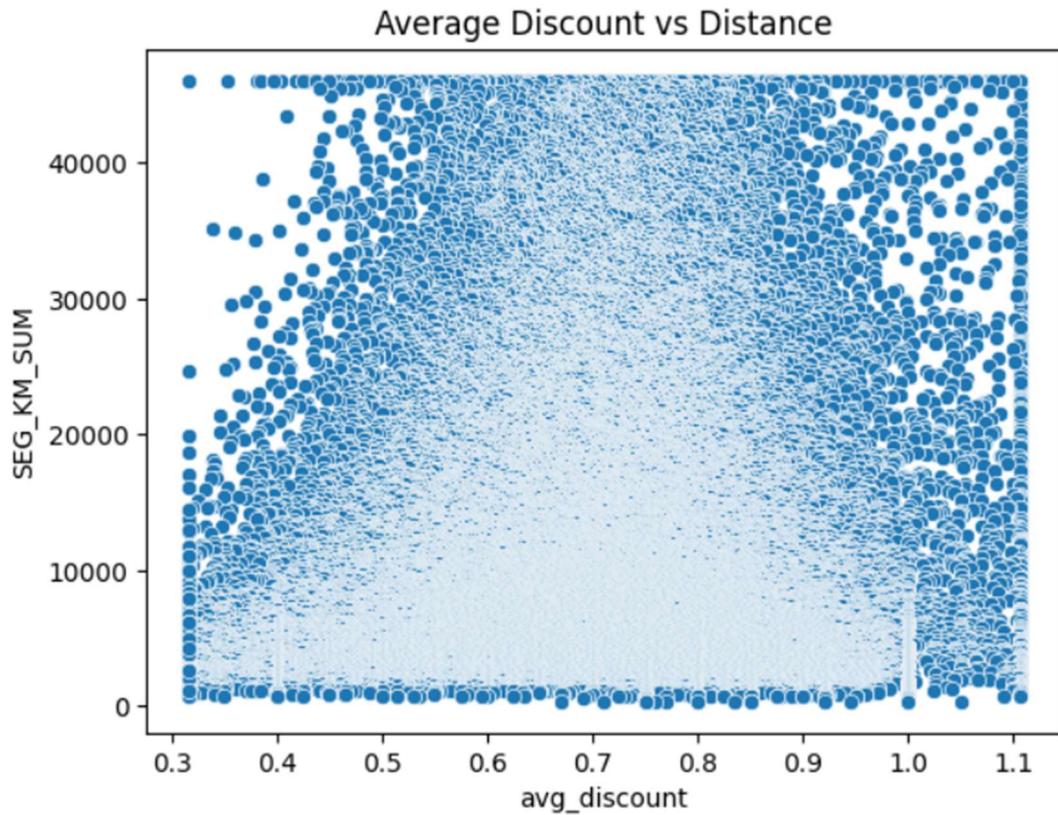
Q12. What is the relationship between flight count and distance?

Flight Count vs Total Distance
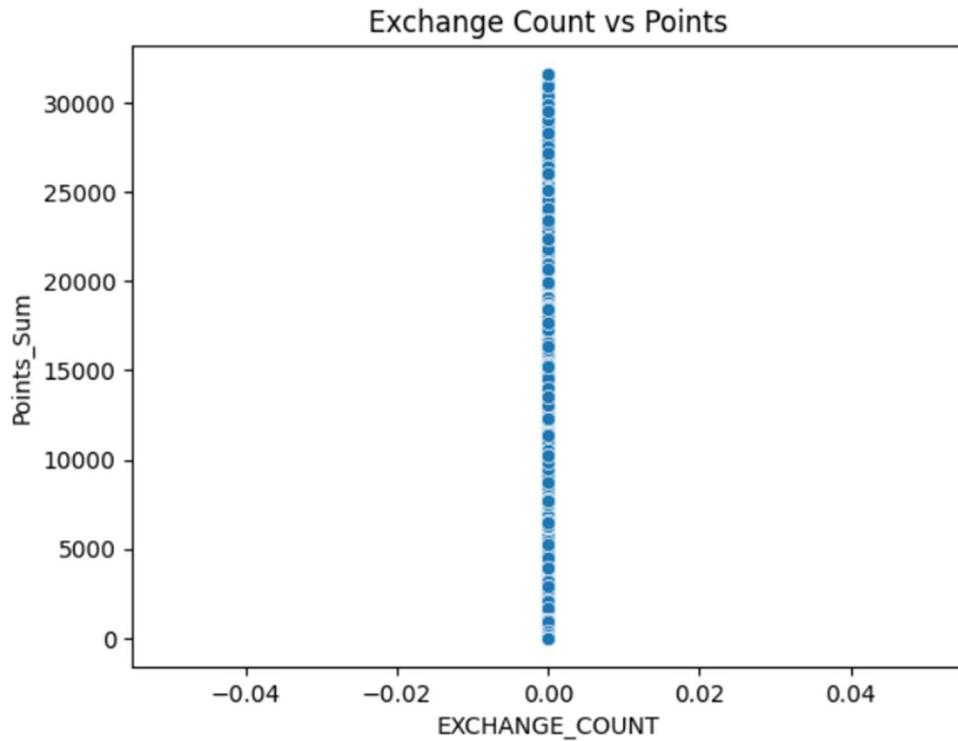
Q13. What is the relationship between points and distance?

Distance vs Loyalty Points (Trend Analysis)
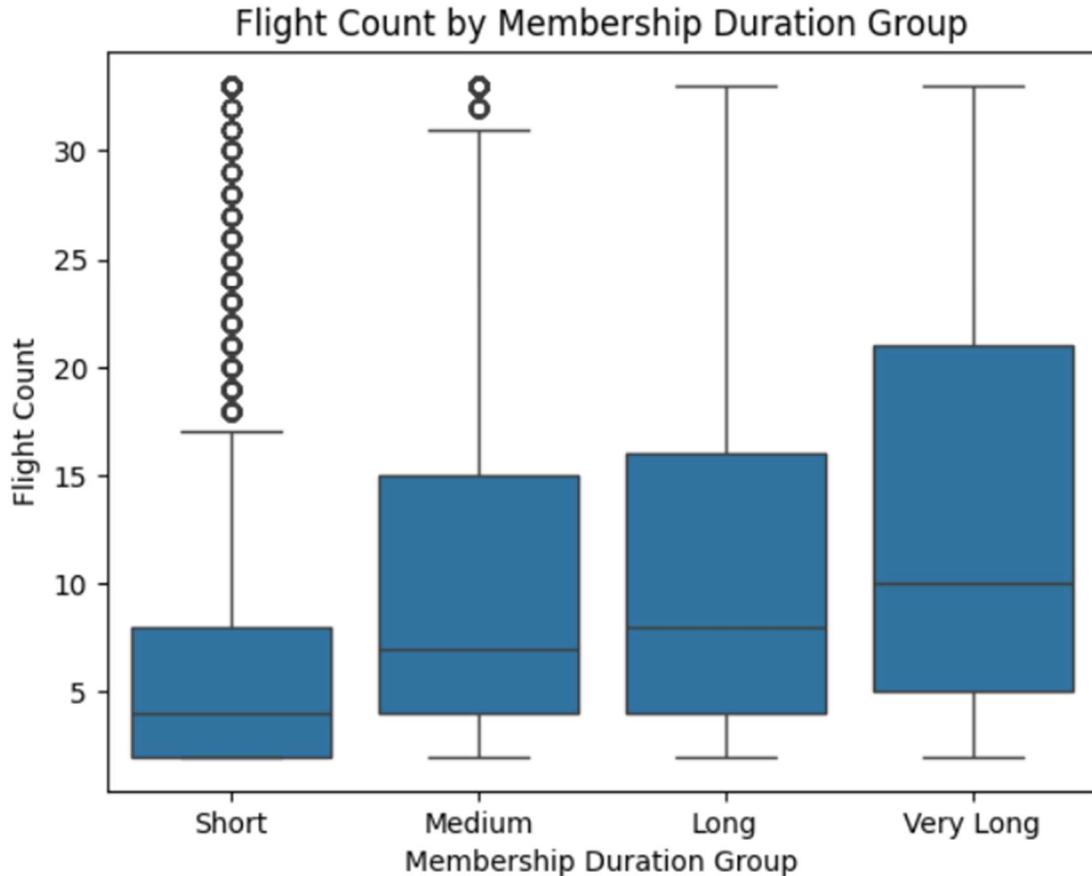
Q14. How does discount vary with distance?

Q15. What is the relationship between exchange count and points?

Q16. Does membership duration affect flight count?

Flight Count by Membership Duration Group

Q17. What are the correlations among numerical features?

[92]: corr_matrix

[92]:

	MEMBER_NO	FFP_TIER	AGE	FLIGHT_COUNT	BP_SUM	SUM_YR_1	SUM_YR_2	SEG_KM_SUM	LAST_TO_END	AVG_INTERVAL	MAX_INTERVA
MEMBER_NO	1.000000	NaN	0.001183	0.003769	0.000154	-0.001511	0.001545	0.001143	-0.002729	-0.000404	0.00415
FFP_TIER	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
AGE	0.001183	NaN	1.000000	0.089889	0.117615	0.127564	0.089455	0.101277	-0.030373	-0.027486	0.01134
FLIGHT_COUNT	0.003769	NaN	0.089889	1.000000	0.875178	0.762203	0.814429	0.894220	-0.488771	-0.442824	-0.16226
BP_SUM	0.000154	NaN	0.117615	0.875178	1.000000	0.804745	0.850352	0.942029	-0.438436	-0.387965	-0.13274
SUM_YR_1	-0.001511	NaN	0.127564	0.762203	0.804745	1.000000	0.536663	0.783154	-0.194355	-0.357502	-0.12801
SUM_YR_2	0.001545	NaN	0.089455	0.814429	0.850352	0.536663	1.000000	0.838308	-0.579832	-0.354433	-0.13068
SEG_KM_SUM	0.001143	NaN	0.101277	0.894220	0.942029	0.783154	0.838308	1.000000	-0.457385	-0.396191	-0.13220
LAST_TO_END	-0.002729	NaN	-0.030373	-0.488771	-0.438436	-0.194355	-0.579832	-0.457385	1.000000	-0.126451	-0.35037
AVG_INTERVAL	-0.000404	NaN	-0.027486	-0.442824	-0.387965	-0.357502	-0.354433	-0.396191	-0.126451	1.000000	0.77359
MAX_INTERVAL	0.004151	NaN	0.011342	-0.162265	-0.132744	-0.128014	-0.130688	-0.132202	-0.350375	0.773593	1.00000
EXCHANGE_COUNT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Na
avg_discount	0.002708	NaN	0.117892	0.144293	0.325106	0.261388	0.207277	0.099582	-0.017897	-0.056564	-0.01666
Points_Sum	-0.007575	NaN	0.093338	0.844974	0.955670	0.770011	0.820031	0.904916	-0.429277	-0.371137	-0.12530
Point_NotFlight	-0.031918	NaN	-0.065193	0.408043	0.391698	0.338882	0.359359	0.393531	-0.220125	-0.172274	-0.06782
MEMBERSHIP_DAYS	-0.045863	NaN	0.273400	0.284712	0.269542	0.221061	0.265812	0.267223	-0.322706	-0.006507	0.10373

Q18. Is there a relationship between gender and tier?