Q1.

(a) A healthy individual, enrolled in the study at age 30, did not develop breast cancer during the study.

Right censoring. Because the individual was not diagnosed by the time the study ended. And we do not know if she had breast cancer afterwards, so it's right censored.

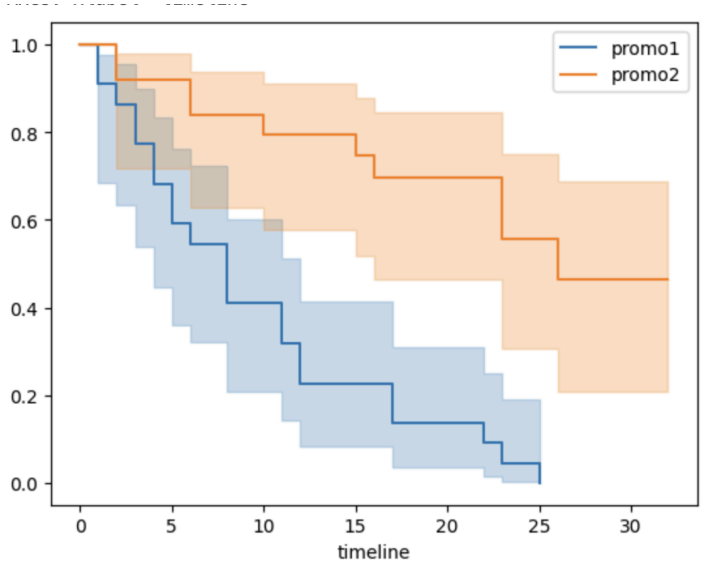
(b) A healthy individual, enrolled in the study at age 40, was diagnosed with breast cancer at the fifth exam after enrollment (i.e., the disease started sometime between 12 and 15 years after enrollment).

Interval Censoring. Because the event happens during a time interval(12-15 years) and not a exact specific time.

(c) An individual, enrolled in the study at age 42, moved away from the community at age 55 and was never diagnosed with breast cancer during the period of observation.

Right censoring. Because the individual was not diagnosed by the time he or she left. And the aftertime is unknown.

Q2.



Survival Function

`kmf1.survival_function_`

timeline	KM_estimate
0.0	1.000000
1.0	0.909091
2.0	0.863636
3.0	0.772727
4.0	0.681818
5.0	0.590909
6.0	0.545455
8.0	0.409091
11.0	0.318182
12.0	0.227273
17.0	0.136364
22.0	0.090909
23.0	0.045455
25.0	0.000000

`kmf2.survival_function_`

timeline	KM_estimate
0.0	1.000000
2.0	0.920000
6.0	0.840000
9.0	0.840000
10.0	0.795789
14.0	0.795789
15.0	0.746053
16.0	0.696316
18.0	0.696316
19.0	0.696316
20.0	0.696316
22.0	0.696316
23.0	0.557053
25.0	0.557053
26.0	0.464211
28.0	0.464211
30.0	0.464211
31.0	0.464211
32.0	0.464211

Confidence Interval

`kmf1.confidence_interval_`

	KM_estimate_lower_0.95	KM_estimate_upper_0.95
0.0	1.000000	1.000000
1.0	0.682979	0.976457
2.0	0.634438	0.953863
3.0	0.537434	0.898476
4.0	0.446186	0.833802
5.0	0.361036	0.762104
6.0	0.320677	0.723945
8.0	0.208521	0.600735
11.0	0.141759	0.511081
12.0	0.082726	0.414451
17.0	0.034130	0.308715
22.0	0.015598	0.251079
23.0	0.003205	0.189448
25.0	0.000000	0.000000

`kmf2.confidence_interval_`

	KM_estimate_lower_0.95	KM_estimate_upper_0.95
0.0	1.000000	1.000000
2.0	0.716390	0.979371
6.0	0.628056	0.936733
9.0	0.628056	0.936733
10.0	0.576216	0.909694
14.0	0.576216	0.909694
15.0	0.517122	0.877972
16.0	0.462752	0.843650
18.0	0.462752	0.843650
19.0	0.462752	0.843650
20.0	0.462752	0.843650
22.0	0.462752	0.843650
23.0	0.305494	0.749244
25.0	0.305494	0.749244
26.0	0.209159	0.686332
28.0	0.209159	0.686332
30.0	0.209159	0.686332
31.0	0.209159	0.686332
32.0	0.209159	0.686332

The confidence interval of promotion 2 is generally wider than CI for promotion1. And the overlapping part is small, saying the test of difference is significant.

Log Rank Test.

	time	risk1	risk2	case1	case2	totalrisk	totalcase	exp1	exp2
0	1	22	25	2	0	47	2	0.936170	1.063830
1	2	20	25	1	2	45	3	1.333333	1.666667
2	3	19	23	2	0	42	2	0.904762	1.095238
3	4	17	23	2	0	40	2	0.850000	1.150000
4	5	15	23	2	0	38	2	0.789474	1.210526
5	6	13	23	1	2	36	3	1.083333	1.916667
6	8	12	20	3	0	32	3	1.125000	1.875000
7	9	9	20	0	0	29	0	0.000000	0.000000
8	10	9	19	0	1	28	1	0.321429	0.678571
9	11	9	17	2	0	26	2	0.692308	1.307692
10	12	7	17	2	0	24	2	0.583333	1.416667
11	14	5	17	0	0	22	0	0.000000	0.000000
12	15	5	16	0	1	21	1	0.238095	0.761905
13	16	5	15	0	1	20	1	0.250000	0.750000
14	17	5	14	2	0	19	2	0.526316	1.473684
15	18	3	14	0	0	17	0	0.000000	0.000000
16	19	3	13	0	0	16	0	0.000000	0.000000
17	20	3	12	0	0	15	0	0.000000	0.000000
18	22	3	11	1	0	14	1	0.214286	0.785714
19	23	2	10	1	2	12	3	0.500000	2.500000
20	25	1	8	1	0	9	1	0.111111	0.888889
21	26	0	6	0	1	6	1	0.000000	1.000000
22	28	0	5	0	0	5	0	0.000000	0.000000
23	30	0	4	0	0	4	0	0.000000	0.000000
24	31	0	3	0	0	3	0	0.000000	0.000000
25	32	0	1	0	0	1	0	0.000000	0.000000

$exp_1 = risk1 * total\,case / total\,risk$

$exp_2 = risk2 * total\,case / total\,risk$

test statistics

$$= \frac{\left(\sum^{t} case1 - \sum^{t} exp_1\right)^2}{\sum^{t} exp_1} + \frac{\left(\sum^{t} case2 - \sum^{t} exp_2\right)^2}{\sum^{t} exp_2}$$

$$= \frac{(22 - 10.4589)^2}{10.4589} + \frac{(10 - 21.54)^2}{21.54}$$

$$= 18.918 > 3.84$$

The log rank test is significant shows there's a difference in effects of promotions 1 and 2. Promo 1 makes customers more likely to revisit.