Analysis Report

This report is structured as follows.

Contents

Sample Characteristics

The analysis started by examining the characteristics of the sample.

Firstly, examining the gender-based statistics reveals that the mean age for females is 40.4 years with a standard deviation of 9.4, while for males, the mean age is 40.0 years, also with a standard deviation of 9.4. This indicates a relatively similar age distribution between genders.

The overall average contribution period is approximately 3.98 years. When breaking down the descriptive statistics further, the mean age of the entire population is 40.5 years, the mean salary is 455,456.3 units, and the mean contribution rate is 14.8%. Gender-specific analysis shows that females have a mean age of 40.6 years, a mean salary of 398,051.1 units, and a mean contribution rate of 14.6%, whereas males have a mean age of 40.4 years, a mean salary of 568,419.5 units, and a mean contribution rate of 15.1%.

Longitudinal data from January 2018 to April 2024 highlight several trends (table below). The number of active customers consistently increased from 994 in January 2018 to 2400 in April 2024. The average age of customers remained stable around 40 years, with minor fluctuations. The proportion of female customers also remained fairly constant at around 66%, with occasional minor variations.

Year	Month	Active Customers	Avg Age	Proportion of Females	Avg Salary	Avg Contribution Rate
2018	01	994	39.8	66%	377743.2	14.3
2018	02	1007	39.7	66%	373624.4	14.3
2018	03	1012	39.7	66%	374941.6	14.3
2018	04	1017	39.8	66%	378251.7	14.4
2018	05	1031	39.8	67%	378863.1	14.4
2018	06	1037	39.9	67%	378362.4	14.3
2018	07	1043	39.9	67%	391375.4	14.3
2018	08	1052	39.9	67%	392918.3	14.4
2018	09	1055	40.0	67%	392862.4	14.4
2018	10	1068	40.0	67%	394427.2	14.4
2018	11	1079	40.0	66%	395987.1	14.5
2018	12	1096	40.0	66%	396213.5	14.4
2019	01	1108	40.0	66%	397140.5	14.4
2019	02	1130	39.9	66%	395812.0	14.4
2019	03	1158	39.9	67%	393975.6	14.4
2019	04	1186	39.8	66%	396367.5	14.5
2019	05	1205	39.8	66%	399282.4	14.5
2019	06	1221	39.7	66%	397147.0	14.5
2019	07	1234	39.8	66%	413268.9	14.5
2019	08	1246	39.8	66%	415950.0	14.6
2019	09	1262	39.7	66%	415238.9	14.6
2019	10	1275	39.7	66%	416697.7	14.6
2019	11	1293	39.8	66%	417685.3	14.6

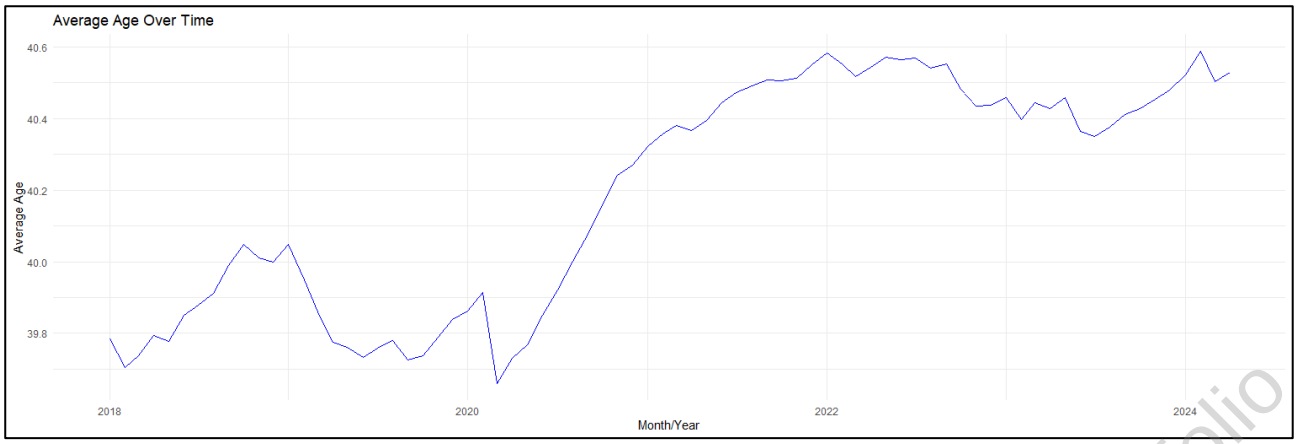
Year	Month	Active Customers	Avg Age	Proportion of Females	Avg Salary	Avg Contribution Rate
2019	12	1305	39.8	66%	417684.6	14.6
2020	01	1321	39.9	66%	419312.7	14.6
2020	02	1329	39.9	66%	419877.1	14.6
2020	03	1368	39.7	66%	413895.4	14.6
2020	04	1373	39.7	66%	414346.7	14.5
2020	05	1380	39.8	66%	415119.8	14.5
2020	06	1380	39.9	66%	414871.1	14.5
2020	07	1382	39.9	66%	420741.1	13.1
2020	08	1383	40.0	66%	422086.9	13.1
2020	09	1385	40.1	66%	422276.3	13.1
2020	10	1387	40.2	66%	424412.9	13.1
2020	11	1393	40.2	66%	424493.3	13.1
2020	12	1401	40.3	66%	424020.6	13.1
2021	01	1405	40.3	66%	424163.3	13.1
2021	02	1412	40.4	66%	424141.4	13.1
2021	03	1424	40.4	66%	422998.1	13.1
2021	04	1435	40.4	66%	422853.1	14.6
2021	05	1443	40.4	66%	424837.3	14.6
2021	06	1454	40.4	66%	426041.7	14.6
2021	07	1462	40.5	66%	425463.0	14.6
2021	08	1475	40.5	66%	425799.1	14.7
2021	09	1491	40.5	66%	426737.5	14.7
2021	10	1507	40.5	66%	426277.1	14.7
2021	11	1519	40.5	66%	427001.3	14.7
2021	12	1533	40.6	66%	426957.9	14.6
2022	01	1544	40.6	66%	426732.5	14.6
2022	02	1562	40.6	66%	431140.9	14.6
2022	03	1591	40.5	66%	426743.5	14.6
2022	04	1614	40.5	66%	443302.8	14.6
2022	05	1628	40.6	66%	442527.7	14.7
2022	06	1650	40.6	66%	441467.8	14.7
2022	07	1669	40.6	66%	444303.3	14.7
2022	08	1703	40.5	66%	442485.2	14.7
2022	09	1729	40.6	67%	441526.1	14.7
2022	10	1757	40.5	66%	441212.4	14.7
2022	11	1786	40.4	66%	440092.5	14.7
2022	12	1834	40.4	66%	439653.4	14.7
2023	01	1852	40.5	66%	440716.9	14.7
2023	02	1884	40.4	66%	439090.2	14.7
2023	03	1973	40.4	65%	437916.3	14.7
2023	04	2006	40.4	66%	440687.6	14.7
2023	05	2066	40.5	66%	438224.4	14.7
2023	06	2124	40.4	66%	436075.4	14.7
2023	07	2157	40.4	66%	458033.3	14.7
2023	08	2184	40.4	66%	456629.2	14.8
2023	09	2198	40.4	66%	456057.1	14.8

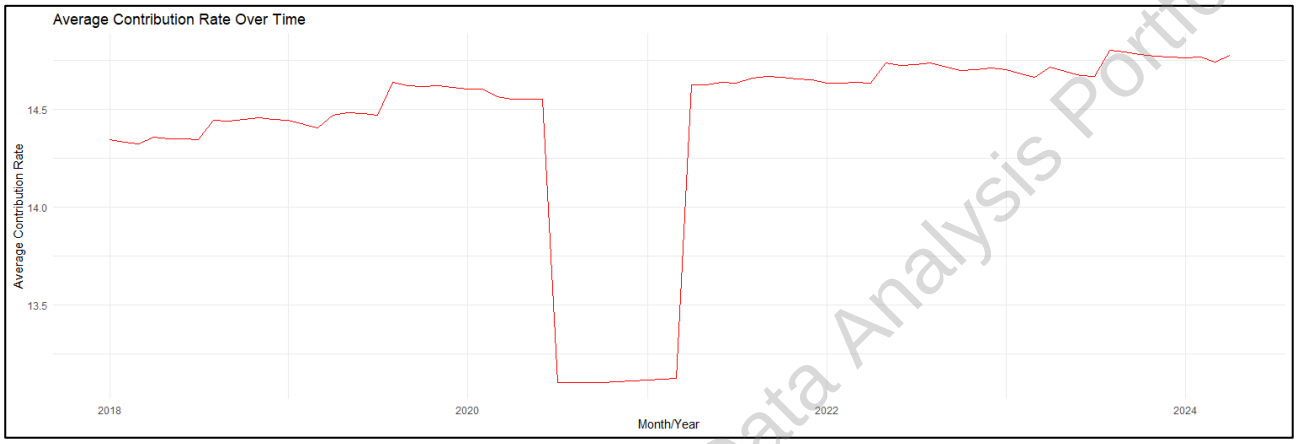
Year	Month	Active Customers	Avg Age	Proportion of Females	Avg Salary	Avg Contribution Rate
2023	10	2226	40.4	66%	456703.6	14.8
2023	11	2252	40.5	66%	456086.0	14.8
2023	12	2281	40.5	66%	456823.3	14.8
2024	01	2303	40.5	66%	456300.5	14.8
2024	02	2322	40.6	66%	458626.4	14.8
2024	03	2373	40.5	66%	456287.9	14.7
2024	04	2400	40.5	66%	455456.3	14.8

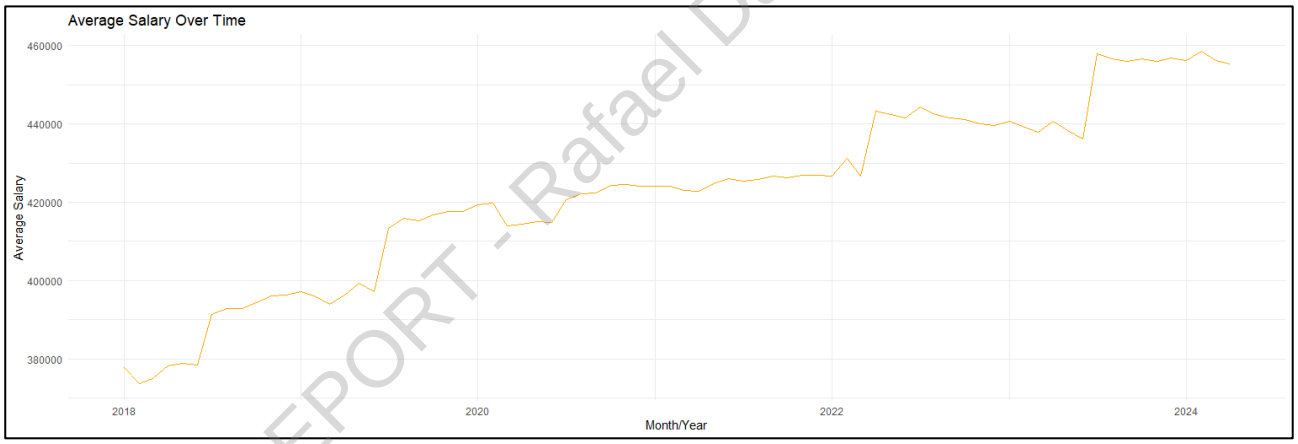
The average salary showed an increasing trend over the years. Starting at 377,743.2 units in January 2018, it gradually increased, reaching 458,626.4 units by February 2024. This increase in salary is indicative of a positive growth trend in the population's earnings over the observed period.

The average contribution rate exhibited some fluctuations, initially around 14.3% in early 2018. It remained relatively stable with minor variations until mid-2020, when it dipped to around 13.1% for a few months before returning to the previous levels. From 2021 onwards, the contribution rate stabilized around 14.7% to 14.8%, suggesting that the earlier dip was an anomaly rather than a trend.

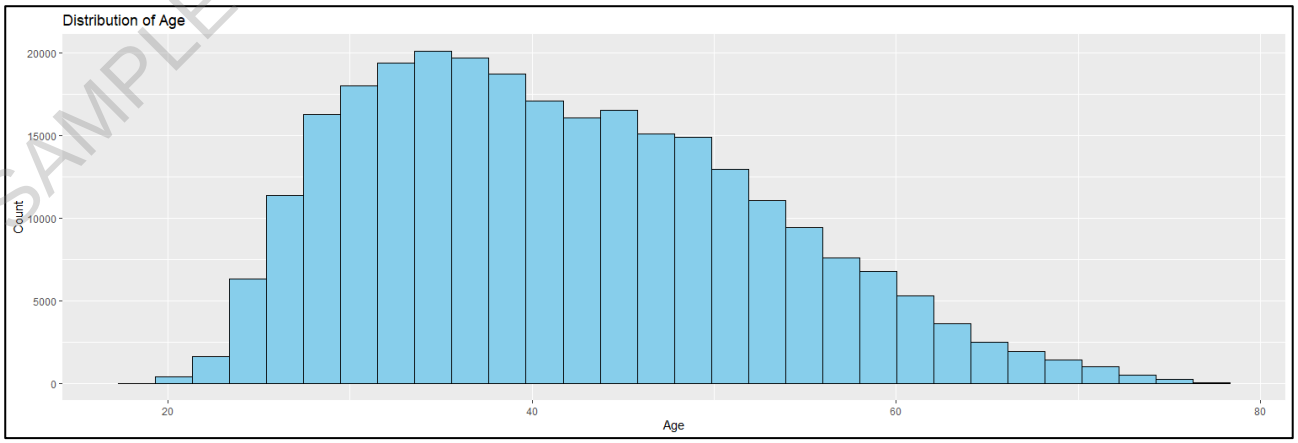
The following plots provide additional visualizations of the trends over time.







The histogram below shows the age distribution of the sample.



Linear Mixed Effects Models

Before proceeding to the statistical models to test if contribution rates have significantly change as a result of communication changes happening in August of every year, data was analyzed for outliers and normality.

Outliers were examined using standardized Z-scores for the variables 'Contribution Rate', 'Age' and 'Salary'. Z-scores were calculated for each customer on each month/year data point. Several cases showed Z-scores outside of the +-3 range, which can be considered outliers and were therefore deleted from the analysis. Not all the data from the subject were deleted, only those corresponding to the problematic variable. 24 cases were removed from 'member_age', 402 for 'annual_pensionable_salary' and 2241 for 'retirement_contribution_rate.

The examination of boxplots indicated that contribution rate and salary were highly skewed variable indicating lack of normality. These variables were log-transformed to attenuate the impacts in the subsequent models.

Dummy variables were created to determine the time periods when new communication took place (from August to July of each year). This allows us to evaluate the effect of each change in communication, whilst neutralizing the effect of changes from other years.

We employed Linear Mixed Models (LMM) due to its ability to handle nested data structures and account for both fixed and random effects. Our dataset included repeated measures for individual participants over several years, allowing us to capture within-subject correlations and between-subject variability.

The dataset comprised longitudinal data from January 2018 to April 2024 of 2,400 unique participants, including variables such as Change_2018, Change_2019, Change_2020, Change_2021, Change_2022, Change_2023, member_age, log_annual_pensionable_salary, and log_retirement_contribution_rate.

The evaluation was performed across three models: one considering annual changes alone, the second controlling for age, and the third controlling for both age and salary. The dependent variable in these models is the log-transformed contribution rate, and the exponentiated estimates (exp. Estimate) provide an interpretation of the effects in terms of percentage changes in the original scale of the contribution rate.

Model 1

The first model assessed the impact of annual changes from 2018 to 2023 on the log-transformed contribution rate and the results are below.

term	estimate	std.error	statistic	df	p.value	exp. Estimate	N
(Intercept)	2.717	0.003	834.906	2489.046	0.000	15.141	113583
Change_2018	-0.017	0.001	-22.742	111359.526	0.000	0.983	113583
Change_2019	-0.005	0.001	-6.428	111377.839	0.000	0.995	113583
Change_2020	-0.003	0.001	-3.555	111378.271	0.000	0.997	113583
Change_2021	0.003	0.001	4.777	111397.923	0.000	1.003	113583
Change_2022	0.011	0.001	17.024	111394.597	0.000	1.011	113583
Change_2023	0.020	0.001	25.656	111321.505	0.000	1.020	113583

The results indicate that the introduction of Change_2023 had a significant positive effect on the log-transformed contribution rate (estimate = 0.020, $p < 0.0001$). The exponentiated estimate suggests that Change_2023 increased the contribution rate by approximately 2.0%. Notably, other years also showed significant effects, with Change_2018, Change_2019, and Change_2020 having negative impacts on the contribution rate, while Change_2021 and Change_2022 showed positive impacts, although the magnitudes of these effects varied.

Model 2 – Controlling for Age

The second model included the same annual changes while controlling for member age.

term	estimate	std.error	statistic	df	p.value	exp. Estimate	N
(Intercept)	2.424	0.007	365.976	15217.180	0.000	11.293	113559
Change_2018	0.000	0.001	0.114	111785.469	0.910	1.000	113559
Change_2019	0.005	0.001	7.159	113455.043	0.000	1.005	113559
Change_2020	0.000	0.001	-0.493	111557.890	0.622	1.000	113559
Change_2021	-0.002	0.001	-2.266	112042.077	0.023	0.998	113559
Change_2022	0.000	0.001	-0.411	113523.957	0.681	1.000	113559
Change_2023	0.005	0.001	5.513	113159.524	0.000	1.005	113559
member_age	0.008	0.000	50.451	42237.565	0.000	1.008	113559

Controlling for age, Change_2023 continued to have a significant positive effect on the log-transformed contribution rate (estimate = 0.005, $p < 0.0001$). The exponentiated estimate suggests a 0.5% increase in the contribution rate due to Change_2023. Additionally, member age was positively associated with the contribution rate (exp. estimate = 1.008, $p < 0.0001$), indicating that older members tend to have higher contribution rates. Interestingly, the impact of other annual changes was less pronounced when controlling for age, with only Change_2019 remaining significantly positive.

Model 3 – Controlling for Age and Salary

The third model included the annual changes, controlling for both member age and log-transformed annual pensionable salary.

term	estimate	std.error	statistic	df	p.value	exp. Estimate	N
(Intercept)	2.368	0.024	99.698	62419.248	0.000	10.676	113111
Change_2018	0.000	0.001	0.043	111355.116	0.966	1.000	113111
Change_2019	0.005	0.001	7.010	113078.702	0.000	1.005	113111
Change_2020	0.000	0.001	-0.272	111048.615	0.786	1.000	113111
Change_2021	-0.001	0.001	-1.982	111837.315	0.047	0.999	113111
Change_2022	0.000	0.001	-0.407	113060.548	0.684	1.000	113111
Change_2023	0.004	0.001	5.350	112907.319	0.000	1.004	113111
member_age	0.007	0.000	39.045	30284.646	0.000	1.007	113111
Log annual pensionable salary	0.005	0.002	2.420	56665.880	0.016	1.005	113111

When controlling for both age and salary, Change_2023 remained a significant predictor of the log-transformed contribution rate (estimate = 0.004, $p < 0.0001$), with an exponentiated estimate indicating a 0.4% increase in the contribution rate. Both member age and log annual pensionable salary were positively associated with contribution rates, with older members and those with higher salaries contributing more. Notably, Change_2019 consistently showed a positive effect across all models, while the effects of Change_2018, Change_2020, Change_2021, and Change_2022 varied depending on the model specification.

Conclusion

The three models collectively demonstrate the robust effect of Change_2023 on increasing retirement contribution rates. The positive impact of Change_2023 remained significant across all models, albeit with varying magnitudes. This suggests that the introduction of the new communication strategy in 2023 effectively enhanced the contribution rates. The inclusion of age and salary as control variables helped to isolate the specific effect of Change_2023, confirming its positive influence on contribution rates.

The exponentiated estimates provide a clear interpretation in the context of the original percentage contribution rates. For instance, an exp. estimate of 1.020 for Change_2023 in the first model indicates a 2.0% increase in the contribution rate due to the new communication strategy. Similarly, the exp. estimate for member age in the controlled models suggests a slight increase in contribution rates with age.

The findings also highlight the varying impacts of annual changes from other years. Change_2019 consistently showed a positive effect on contribution rates, while other years exhibited mixed effects depending on the model specification. This indicates that the specific context and external factors associated with each year's changes play a crucial role in influencing contribution behaviors.