INTRODUCTION
There are three major concerns in the retirement industry:
• Very few people are focused on saving for retirement. People may need to take money out early for emergencies which reduces retirement corpus.
• More money is being spent on services and products impacted by inflation.

One-third of Americans feel unprepared or unsure of whether they are on track for retirement, one of the main reasons being the lack of engagement between employers/plan providers and employees. The purpose of this study is to examine these challenges and provide insights into the ways that employers and plan providers could help improve the contribution rate.

PROBLEM STATEMENT
A significant percentage of employees do not contribute regularly, fail to take advantage of employer match, or are not aware of various incentives that come with regular contributions to a retirement plan. Many individuals are putting savings towards short-term expenditures rather than focusing on significant long-term investments like retirement.

PROJECT GOALS
Develop new methods to:
• Increase participant engagement with retirement plans
• Optimize participant contributions and/or withdrawals to promote long-term financial wellness

Focus is on two types of plan participants:
• Active Contributors
• Retirees

MATERIALS & METHODS
We were provided with a small dataset (100 rows) containing qualitative and quantitative attributes like age, gender, marital status, occupation, remitting percentage, lifestyle score, and fitness score. We performed exploratory data analysis (EDA) using:
• K-Means Clustering: Learn more about the persona characteristics.
• Pairwise Correlation: Understand relationships between attributes.
• Decision Tree (DT): Predict a person’s remitting amount and learn more about important variables.

DATA ANALYSIS RESULTS
The pairwise correlation matrix revealed several pairs of factors that are highly correlated. Some interesting examples are:
• Positive: Salary and Remitting Amount
• Negative: Remitting Percentage and Digital Adoption Score

K-means clustering revealed three groups within our data:
Cluster 1: Retired (70+, long-time contributor, low income)
Cluster 2: Early Career (21+, short contribution window, good digital adoption score)
Cluster 3: Mid-Career (40+, average contribution window, high salary, high remitting amount)

Our decision tree provided insights on which variables may have a larger impact on how much a person can contribute to their retirement. For example: men with < 3 dependents making $20,000+ may contribute approx. 13%. However, those with 3+ dependents are predicted to contribute approx. 7%.

CONCLUSIONS
Despite various challenges that impacted the construction of a robust machine learning model, the Synechron team is very happy with the final product.

Key takeaways about the retirement industry:
• Shifting focus from financial wellness score to personalized actions and recommendations
• Inflation and salary growth affect retirement savings ideas on next steps:
  • Collect/extrapolate additional data to build a more robust model
  • Enhance the UI to include peer comparison and more personalized recommendations derived using NLP

This dive deep into retirement data gave us an increased awareness about the options and incentives available for our own personal retirement savings. We are grateful for the opportunity to have worked with Synechron!