

Introduction

Participant Strategy Experience (PSX) strives to create a world-class investment experience for our retirement plan participants. One area of focus is to provide industry leading advice access to millions of defined contribution (DC) participants.

Scenario

You are a newly hired analyst in Client Experience Analytics (CxA), a department in PSX, aligned to our Institutional Advice offering. Your job is to explore the demographic data that was used to run a marketing campaign before you were hired. The campaign was completed via email with messages sent between 7/01/2020 through 1/31/2021.

Campaign Details: One of the methods used to encourage participants to enroll in Vanguard's Digital Advisor (DA) is to send them an email 30 days after their company begins to offer the DA service. Emails contained an invitation to use the site and a link directing them to the enrollment process. Participants received the emails through an automated process.

There are two data sets for this case study (metadata is provided on page 2):

1. demos.csv contains the demographic data for participants
2. email.csv contains open, and click through information for the emails sent

Objectives

As an Advice analyst, your team has tasked you with analyzing the demographic data used to create the campaign population. Please address the following as you review the data:

1. Describe the steps you took to analyze & clean the demographic data.
2. Provide an overview of the current participant user base.
3. Based on the email data:
 - a. What characteristics correlate with a participant being more likely to interact with our email?
 - b. Are there other trends in the demographics which could indicate a more effective communication strategy with other participants?

You should plan to prepare your insights as you would for a 20-minute presentation for an audience consisting of technical analysts and product managers (non-technical).

Please include in your final submission:

1. A power point presentation that articulates the objectives above regarding the campaign
2. All of the code used in your analysis. You must include your work using the advanced analysis tool/language of your choice submitted as a separate file. Our team predominantly uses Python and SQL today.

Metadata

Note: the data provided for this case study is fictitious and is only meant for the purposes of evaluating a candidate's skill set.

demos

Data Field	Description
part_id	Unique identifier for the participant
fund_ct	The number of funds in the participant's account
logon_ct	The number of times the participant logged on the website within the last 12 months
call_ct	The number of times the participant called within the last 12 months
balance	The total balance of all accounts for the participant
tenure	The tenure of the participant expressed as a decimal
age	The age of the participant expressed as a decimal
gender	The gender of the participant 0 = male 1 = female 2 = unknown
status	The employment status of the participant relative to the employer sponsoring their retirement plan. 0 = Active 1 = Terminated
campaign	0 = Email A 1 = Email B 2 = Did not receive an email

email

Data Field	Description
part_id	Unique identifier for the participant
click_dt	The first date the emailed link was clicked
bounce_dt	The date the email bounced (was not received)
open_dt	The first date the email itself was opened