- Hello everyone, thank you for attending our presentation today. This presentation is for anyone who is considering buying or renting a home in San Francisco.
- My name is Isabelle, and I am here with my teammates, Mwanga and Ivy.
- We are excited to talk to you about our research proposal analyzing the financial outcomes of renting vs buying a home in San Francisco over a four year period from 2018 to 2021. We will go through the study design details, including data, samples, variables & statistical methods. We will also review the potential risks and deliverables.
- We hope this study will provide clarity in the decision to rent or buy.

Determining whether to buy or rent a home involves a complex decision-making process that can be further convoluted by an array of conflicting research. While companies like NerdWallet and Realtor.com have created user-friendly calculators to compare the cost of renting or buying over time, these sites have been shown to miss key financial factors involved in each process.

Our research aims to build on these existing tools by creating and analyzing a more complete dataset on the total costs of renting or owning a home using survey data. For the purpose of this study, we focus solely on the financial costs of renting vs. buying a home in order to limit the dependent variables and produce a single, tangible outcome.

To conduct this analysis, we will compare the total costs accumulated after a period of four years for renters and homeowners in order to align with a single phase of the real estate life cycle. Additionally, we will limit our study to San Francisco to allow us to perform an in-depth analysis within our budget and time constraints. San Francisco is an ideal location for this analysis as it is a population-dense, diverse, metropolitan location with a wide variety of housing options.

This research will act as an essential resource for individuals who are in the process of deciding to buy or rent a home in San Francisco, while also providing a framework

for those looking to analyze the rent vs. buy decision in other locations.

To that end, we hope to answer the question: does renting a home cost more than buying a home in San Francisco over a period of four years? Throughout this analysis, we hope to discover what factors play the largest role in cost differences and how these trends have changed over time.
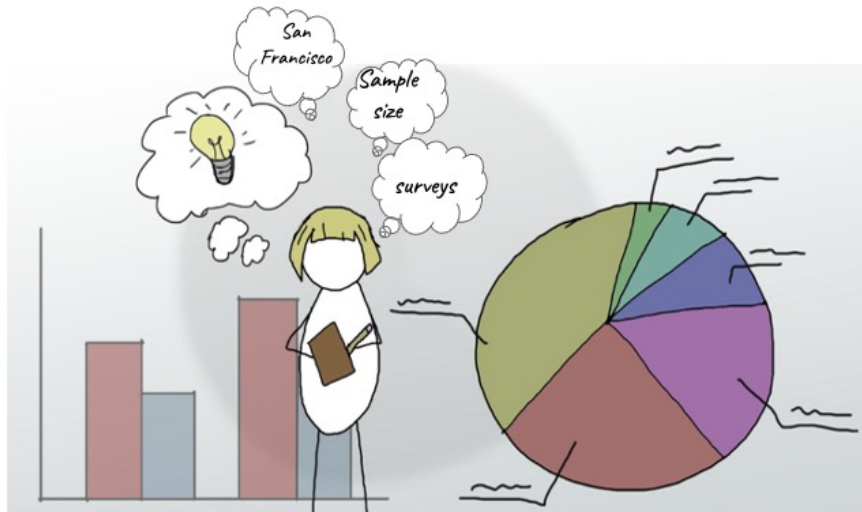
- In the study we will partner with zillow.com and apartments.com.
- Our study is limited to residents living in San Francisco county.
- Analyse the data sources we get from them that covers our timeframe of interest of 4 years.
- Sample size will be from with in this data and selection will be randomized.
- For data, we cannot get from our partners, we will send out surveys, by way of mail.
    - Explain what survey is about
    - The type of data to be collected and time frame the encrypted data will be stored
    - Compensation for participation(gift card)
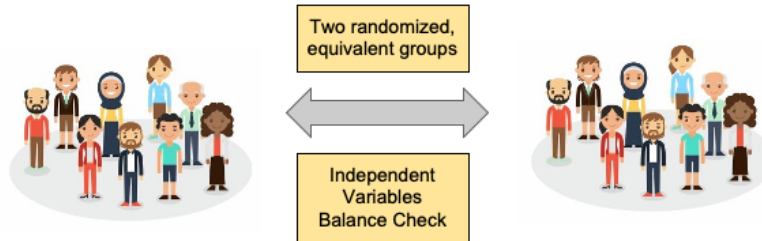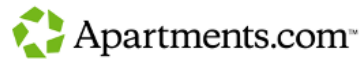   - Data from survey will be converted from analog to digital to allow for automated data analysis

- For homebuyer samples, we will partner with zillow.com to obtain randomized homebuyers samples in San Francisco.
- For renters, we will partner with apartment.com to obtain randomized renter samples in San Francisco.
- By using an online sample size calculator, we estimate:
  - the sample size of 384 based on the population of San Francisco - 874,000 people,
  - the standard margin of error 5%, and
  - the confidence level 95%.
  - (Please see report for details: 62% of households rent homes in San Francisco - 238 renters and 146 homeowners).
- To establish causality between rent/buy and the outcome variable (cost of housing), we will conduct an independent variable balance check to confirm these two groups are equivalent in various characteristics.
- Next slide, please.
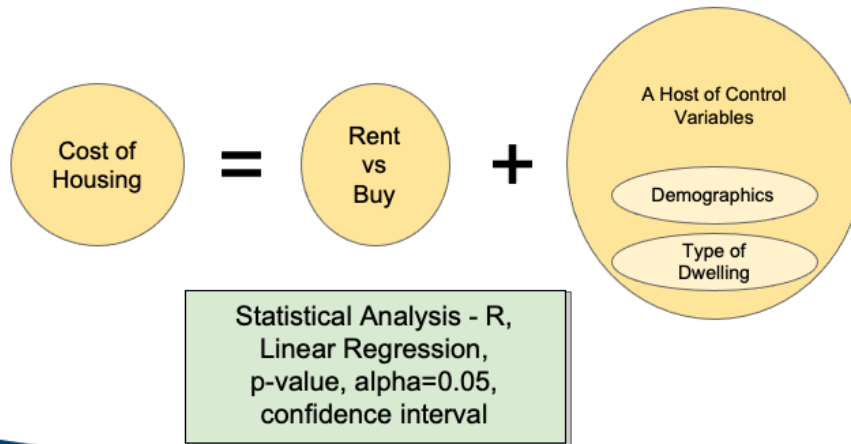
- We will use standard linear regression to model the result of the experiment.
- On the left hand side of the equation. We have the observed outcome of the study, the cost of housing.
- For renters, the cost of housing is calculated by monthly rent, renter insurance, utilities and additional rental fees.
- For homebuyers, it is calculated by mortgage payment, mortgage insurance, utilities, HOA, property tax, repairs, and other cost related to owning a home.
- On the right hand side of the linear regression equation, there is a variable in research (Rent vs Buy) and a host of control variables.
- These control variables improve the precision of the study and helps us identify conditional effects for future experiments.
- They also help us ensure that we have two equivalent samples for comparison.
- These control variables include demographics variable such as age, marital status, size of household, household income, etc.
- It also contains dwelling related variables such as type of housing (e.g. single family home, condo, apartment, townhouse), number of bedrooms, number of bathrooms, square footage, etc.
- Finally, we will calculate the p-value of each variable and determine if they are statistically significant or p-value is less than 0.05 (5%).
- Next slide, please
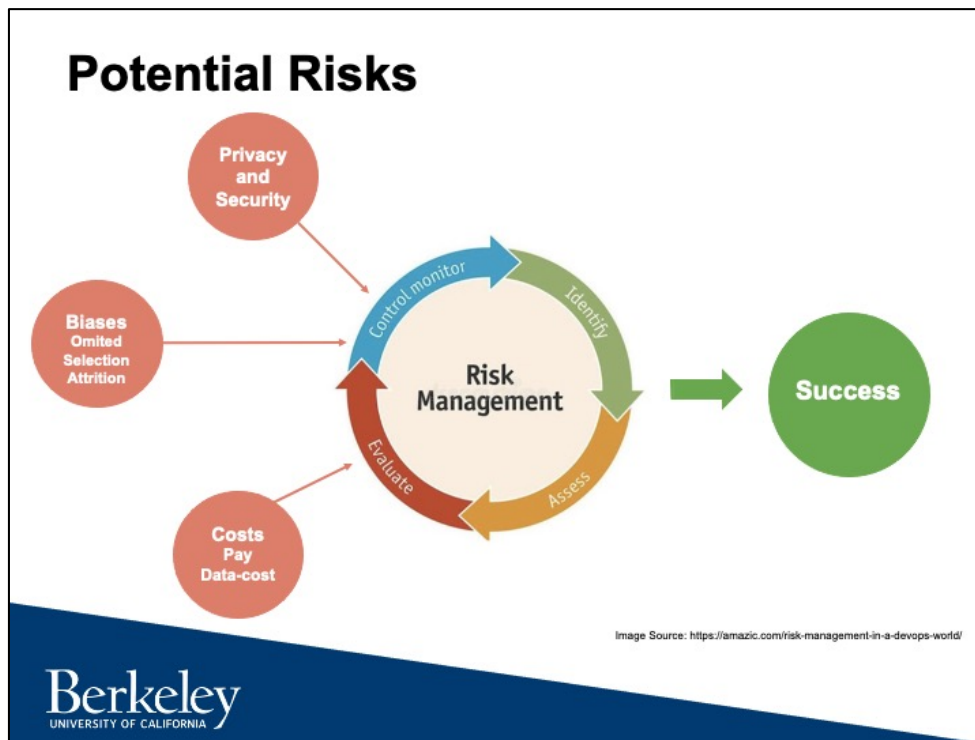
- Privacy and security
    - we are tasked to keep data safe.
    - Because we collect data from different sources and link it together there is a risk of de-identification.
    - Some ways of controlling this is through data stratification and for hacks encryption.
    - To conform to the CPRA that will be effective Jan 1 2023 we advise participants early in the survey process about the intended time of storage.
- Biases
    - Omitted variable bias(due to an observational study design we aim to compare two groups on similar variables but there are some we will miss that some may see as important) Mitigate by looking at several papers and articles
    - Selection bias we can not possibly compare apples for apples for all rentals and owned homes so we assume a general measure of square footage.
    - Attrition bias, due to privacy concerns some survey participants may still opt out of the experiment. These can be countered through reduction of participation effort by going online, more compensation , reminders
- Cost of surveys
    - Fair compensation

- Cost of the data collection, analog to digital conversion and analysis
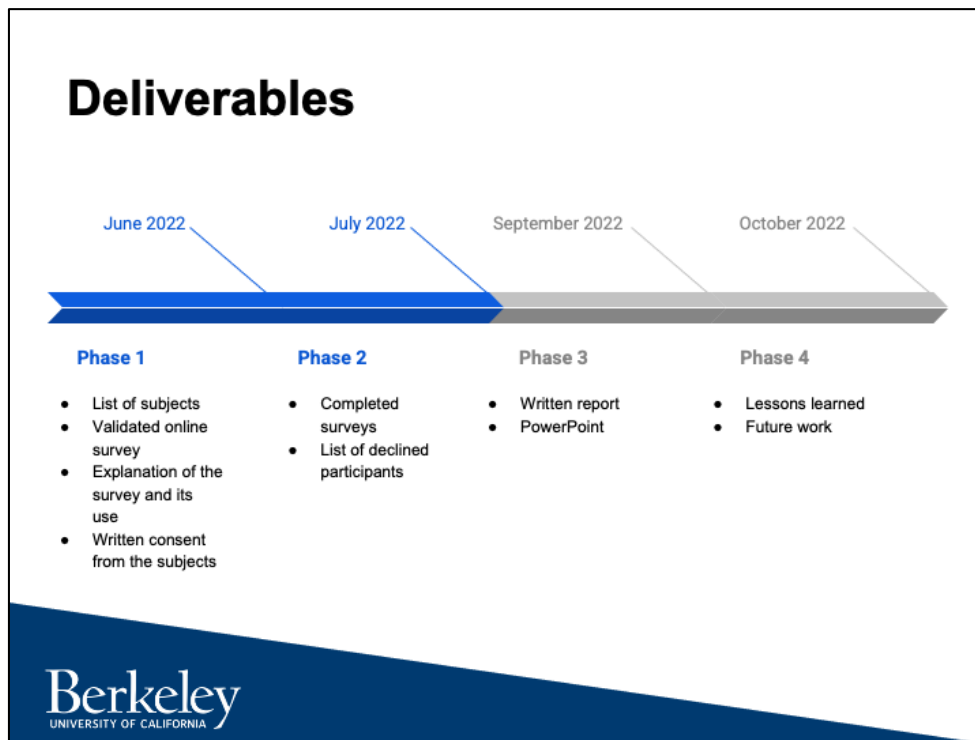
Phase 1 - Participant Recruitment and Survey Design (2 months)

In Phase 1, we will complete the participant recruitment for our survey while finalizing the survey design. Deliverables for this phase include:
- An up-to-date list of 238 renters' and 146 homeowners' addresses and emails in the San Francisco area.
- A validated survey.
- A written explanation of the survey and its use.
- Written consent from the subjects.

Phase 2 - Data Collection (3 weeks)

In Phase 2, we will complete the data collection by way of an online survey. Deliverables for this phase include:
- Completed housing surveys.
- A comprehensive list of participants who declined to participate.

Phase 3 - Data Analysis and Report (2 months)

In Phase 3, we will complete the data analysis and compile a report of our findings. Deliverables for this phase include:

- A written report with the sections:
  - Introduction
  - Method
  - Results
  - Discussion
  - References
- A PowerPoint presentation summarizing the report.

Phase 4 - Post-Analysis (2 weeks)

Following Phase 3, we will complete a post-mortem analysis of our study.
Deliverables for this phase include:
- A summary of lessons learned.
- Opportunities for future work from this study.

We hope this study can act as:

- A resource for you when considering buying or renting a home in San Francisco, and
- we also look forward to work from our colleagues who will build off of this study to analyze the rent vs. buy decision in other locations.

Thank you all for your time.