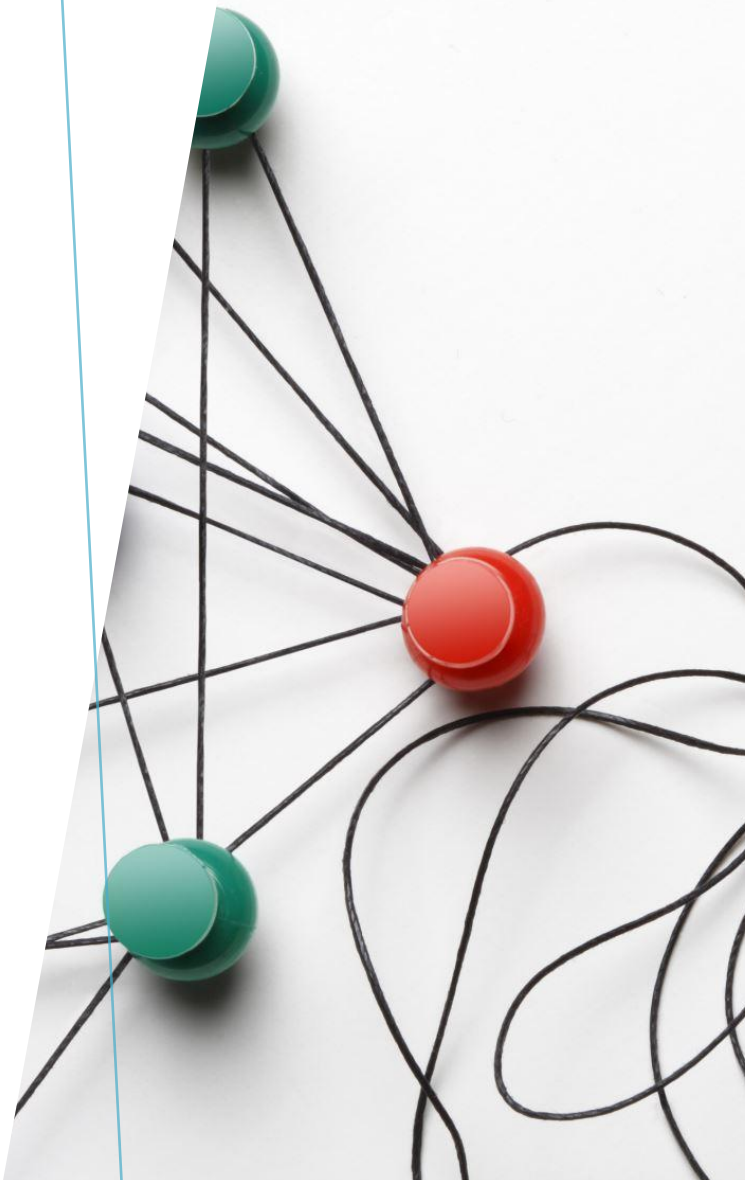
BY, ABDALRHMAN SAED, ALEXIS, JEROME, MARDHIAT, KYLE.

UNDERSTANDING HOW DEMOGRAPHIC FEATURES AFFECT HOMEOWNERSHIP.



OVERVIEW OF THE WORKFLOW

- 1. Hypothesize
- 2. Acquire
- 3. Explore
- 4. Deep Dive
- 5. Communicate

Introduction

The 2020 Census File was collected by the Federal Housing Agency as they documented information on mortgages of single-family properties in 2020. Documentation of the housing market is extremely important in understanding the construction of local environments and whether there is any trend or pattern in the industry. Year after year, federal housing institutions must release information on the market to the public to maintain clarity on the state of housing.

Research Questions

- Can we discover a relationship between demographic columns and homeownership?

- Does representation in the data correlate with whether or not the homeowner is a first time homebuyer or a non first time homebuyer?

DATA SCIENCE WORKFLOW: HYPOTHESIS

- Null Hypothesis: First-time homebuyers between the ages of 35-44 that are white, are overrepresented by 40% in owning property.

- Alternate Hypothesis: Non First-time homebuyers between the ages of 35-44 that are white, are overrepresented by 40% in owning property.

DATA SCIENCE WORKFLOW: ACQUIRE & EXPLORE

- The file that we acquired our data from is the file 2020_Single_Family_Census_Tract_File, which is data gathered from the government via surveys on single family homes.
- We explored the file using exploratory python functions such as .describe, .info, and .shape In order to gain enough general information about the file before delving further into it.
- At first glance our group was able to see that the rows in our data set did not have any missing values or any missing non null counts

DATA SCIENCE WORKFLOW: ACQUIRE & EXPLORE CONT…

	ENTFLAG	REC_NUM	UPSSTCODE	MSA_CODE	FIPS_CNTY_CODE	CENSUS_TRACT_CODE	PERCENT_MINORITY_CENTRACT	MEDIAN_INCOME_C
count	3779973.0	3779973.0	3779973.0	3779973.0	3779973.0	3779973.0	3779973.00	
unique	3.0	2001.0	107.0	746.0	502.0	28530.0	14005.00	
top	2.0	787.0	6.0	99999.0	13.0	200.0	13.57	
freq	3771781.0	3772.0	587276.0	307109.0	164948.0	11430.0	3074.00	
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

```
Mf.shape
```

```
(3779973, 70)
```

DATA SCIENCE WORKFLOW: ACQUIRE & EXPLORE CONT…

```
RangeIndex: 3779973 entries, 0 to 3779972
Data columns (total 70 columns):
 #   Column                                Dtype
---  ------                                -----
 0   ENTFLAG                               object
 1   REC_NUM                               object
 2   UPSSTCODE                             object
 3   MSA_CODE                              object
 4   FIPS_CNTY_CODE                        object
 5   CENSUS_TRACT_CODE                     object
 6   PERCENT_MINORITY_CENTRACT             object
 7   MEDIAN_INCOME_CENTRACT                object
 8   LOCAL_AREA_MEDIA_INCOME               object
 9   TRACT_LOCAL_MEDIAN_INC_RATIO          object
 10  BORROWER_ANNUAL_INCOME                object
 11  MSA_MEDIAN_INCOME                     object
 12  BORROWER_AREA_MEDIAN_FAM_INCOME       object
 13  ACQ_UNPAID_PRINCIPAL_BAL              object
 14  LOAN_PURPOSE                          object
 15  FED_GUARANTEE_TYPE                    object
 16  NUMBER_OF_BORROWERS                   object
 17  FIRST_TIME_HOME_BUYERS                object
 18  BORROWER_RACE_NATIONALORIG19          object
 19  BORROWER_RACE_NATIONALORIG20          object
 20  BORROWER_RACE_NATIONALORIG21          object
 21  BORROWER_RACE_NATIONALORIG22          object
 22  BORROWER_RACE_NATIONALORIG23          object
 23  BORROWER_ETHNICITY                    object
 24  COBORROWER_RACE_NATIONALORIG25        object
 25  COBORROWER_RACE_NATIONALORIG26        object
 26  COBORROWER_RACE_NATIONALORIG27        object
 27  COBORROWER_RACE_NATIONALORIG28        object
 28  COBORROWER_RACE_NATIONALORIG29        object
 29  COBORROWER_ETHNICITY                  object
 30  BORROWER_GENDER                       object
 31  COBORROWER_GENDER                     object
 32  BORROWER_AGE                          object
 33  COBORROWER_AGE                        object
 34  OCCUPANCY_CDE                         object
 35  RATE_SPREAD                           object
 36  HOPEA_STATUS                          object
 37  PROPERTY_TYPE                         object
 38  LIEN_STATUS                           object
 39  BORROWER_AGE_62OVER                   object
 40  COBORROWER_AGE_62OVER                 object
 41  ORIGINATION_LTV                       object
 42  MORTGAGE_NOTE_DATE                    object
 43                                        float64
 44  MORTGAGE_TERM_ORIGINATION             object
 45  INTEREST_RATE_ORIGINATION             object
 46  NOTE_AMT                              object
 47  PREAPPROVAL_CDE                       object
 48  APPLICATION_CHANNEL                   object
 49  AUTOMATED_UNDERWRITING_SYS_CDE        object
 50  BORROWER_CREDIT_SCORE_MODEL           object
 51  COBORROWER_CREDIT_SCORE_MODEL         object
 52  DEBT_TO_INCOME_RATIO                  object
 53  DISCOUNT_POINTS                       object
 54  INTRO_RATE_PERIOD                     object
 55  MANUFACTURED_HOME_LAND_PROP_INTREST   object
 56  PROPERTY_VALUE_AMT                    object
 57  RURAL_CENSUS_TRACT                    object
 58  LOWER_MISSISSIPPI_DELTA_CNTY          object
 59  MIDDLE_APPALACHIA_CNTY                object
 60  PERSISTENT_POVERTY_CNTY               object
 61  AREA_OF_CONCENTRATED_POVERTY          object
 62  HIGH_OPPORTUNITY_AREA                 object
 63  QOZ_CENSUS_TRACT                      object
 64  MSANAME                               object
 65  MSApos                                object
 66  MSaname1                              object
 67  MSA                                   object
 68  STATE                                 object
 69  FIPS_CNTY_NAME                        object
dtypes: float64(1), object(69)
```

DATA SCIENCE WORKFLOW: ACQUIRE & EXPLORE CONT…

- We summarized that there could be 2 potential reasons for this. The first being, that when we merged our 3 txt files with the Census file, the three txt files filled in any potential missing values. The 2nd reason and the most likely to be true, is the fact that this data is acquired by the government, so It wouldn't make any sense for the government to have any missing (possibly vital) information.
- In addition, We identified demographic columns that we believed tied directly in with First-Time home ownership as well as the values of the properties that were bought. This is also how we came to our Null Hypothesis.

DATA SCIENCE WORKFLOW: DEEP DIVING/MODELING, Null Pt1

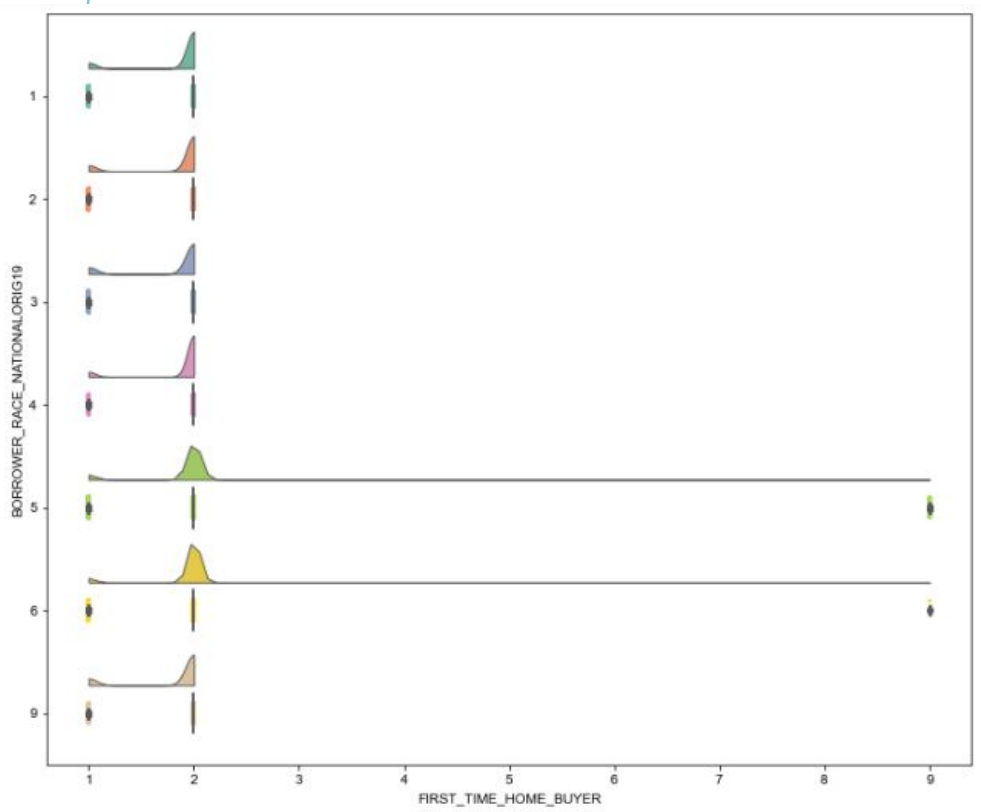
The first thing we decided to do was to check if First-Time Homebuyers alone are overrepresented in this dataset. After writing some code to do an overview of the counts between First-Time Home buyers and Non First-Time Homebuyers, we were able to see that First time Homebuyers are in the significant minority in comparison to Non first time home buyers.

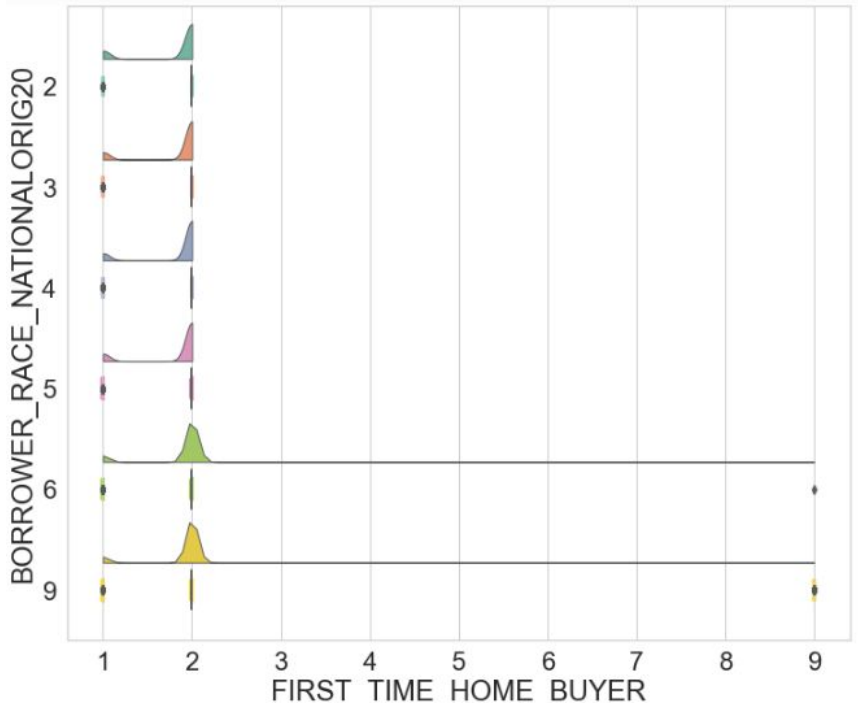
```
In [10]: Mf.groupby('FIRST_TIME_HOME_BUYERS').FIRST_TIME_HOME_BUYERS.describe(include = 'all').transpose()
```

Out[10]:

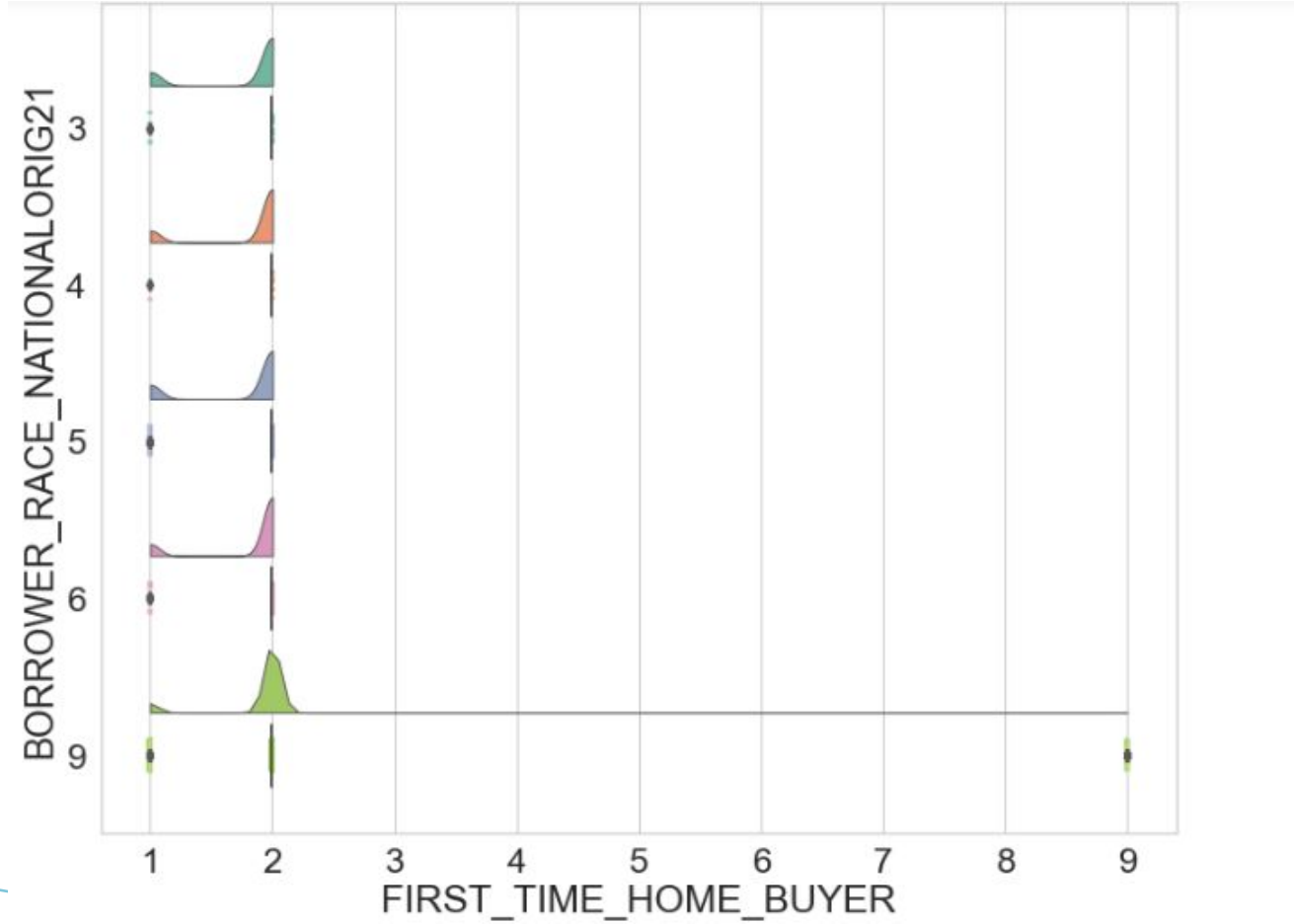
FIRST_TIME_HOME_BUYERS	1	2	9
count	453751.00	3325723.00	498.00
mean	1.00	2.00	9.00
std	0.00	0.00	0.00
min	1.00	2.00	9.00
25%	1.00	2.00	9.00
50%	1.00	2.00	9.00
75%	1.00	2.00	9.00
max	1.00	2.00	9.00

Graphs that Support Findings





Graphs that Support Finding Cont.



DATA SCIENCE WORKFLOW: DEEP DIVING/ MODELING, Null Pt2

What does this mean when comparing to our Null Hypothesis? Essentially the numbers shown in the table, already signify that the First-Time Homebuyers are the significant minority when compared to Non First-Time Homebuyers. This means that we can already reject our null hypothesis, because there is no way that First Time Homebuyers that are white and are aged between the ages of 35-44 can be overrepresented since they are the minority. On the other hand however, this signaled to us that our alternative hypothesis has passed the first step of being proven, because Non First-Time Home Buyers in our alternate hypothesis are the majority.

```
multilevel3 =pd.crosstab(index=[Mf.BORROWER_RACE_NATIONALORIG19,Mf.BORROWER_AGE], columns=Mf.FIRST_TIME_HOME_BUYERS)
multilevel3
```

DATA SCIENCE WORKFLOW: DEEP DIVING/ MODELING Alternate Pt1

BORROWER_RACE_NATIONALORIG19

5	1	34437	14444	5
	2	165208	406110	115
	3	65121	685837	123
	4	30069	571153	82
	5	19440	428314	71
	6	8293	225264	41

BORROWER_RACE_NATIONALORIG20

5	1	921	319	0
	2	6945	12526	4
	3	3028	22847	1
	4	1044	16427	0
	5	472	9751	2
	6	176	4477	1
	7	34	956	0
	9	0	2	0

BORROWER_RACE_NATIONALORIG21

5	1	28	10	0
	2	244	397	0
	3	98	642	0
	4	36	429	0
	5	8	267	0
	6	3	119	0
	7	2	25	0

BORROWER_RACE_NATIONALORIG22

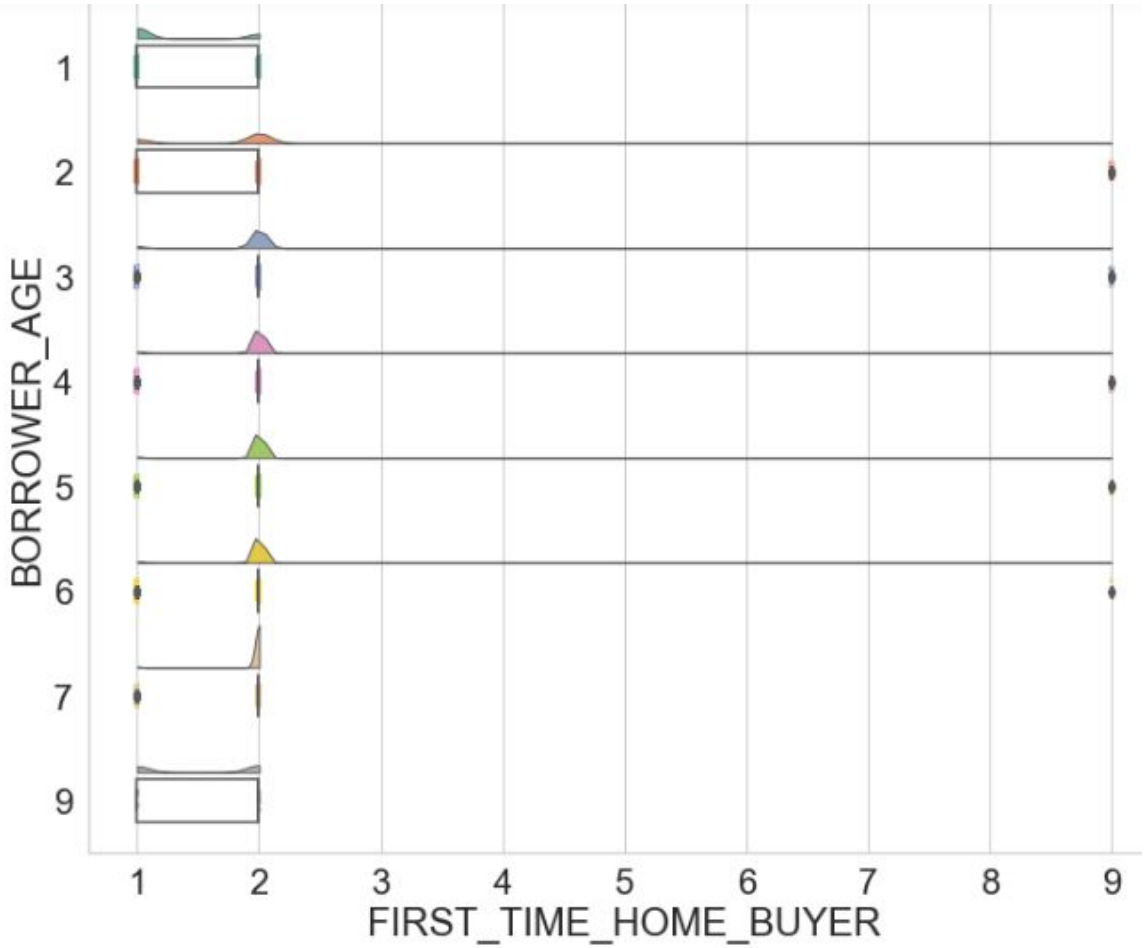
5	1	3	0	0
	2	13	29	0
	3	6	31	0
	4	5	22	0
	5	1	16	0
	6	2	6	0
	7	0	1	0

RACE_NATIONALORIG23

5	1	1	0	0
	2	1	8	0
	3	2	15	0
	4	1	9	0
	5	1	8	0
	6	0	3	0
	7	0	2	0

709,372 Non First-Time homebuyers that are white between ages 35-44

Graphs that Support Finding.



DATA SCIENCE WORKFLOW: DEEP DIVING/ MODELING Alternate Pt2

There are 709,372 Non First-TIme homebuyers that are white between ages 35-44 out of all 3,779,474 homebuyers. Using division we can see that they are only representative of about 18%. Meaning that we also reject our alternate hypothesis, due to the fact that although they represent a significant portion of the data they do not represent the 40% that we hypothesized.

Therefore, the conclusion to our project is that both our Null and our Alternative hypothesis were both rejected due to them both failing to reach the 40% benchmark.