

Analysis of Young Professional Magazine Survey Data

Zachary Picard

Purpose: Using the Young Professional Magazine survey data, develop a profile of current subscribers. By doing this, Young Professional Magazine can expand its current audience by targeting and marketing to the same demographics of people who are currently subscribed. Additionally, aim to use the demographics of the magazine's subscribers to comment on topics that could be of interest for current/future readers.

Background: Young professional magazine was developed for a target audience of recent college graduates in their first 10 years in a business/professional career. While the first two years of publication have been fairly successful, it is time to expand the magazine's advertising base. This is why the survey results are so important. The survey results give us the power to find key demographics of the subscribers. By doing this, we can target marketing and advertising efforts to those demographics, more so than other demographics because they would be considered the "target audience."

Scope: The study will begin showing measures of central tendency and variability within certain variables. It will then show measures of variability for other variables. Furthermore, the investigation will dive into the distribution shapes of the continuous Random Variables, along with visual and verbal explorations. Then, the study goes into a visual exploration of the discrete random variables through the use of histograms and comments. The study goes on to explore various confidence intervals, and try to find inferred answers to the questions: "Would Young Professional be a good advertising outlet for online brokers?" and "Would this magazine be a good place to advertise for companies selling educational software and computer games for young children?" After these two investigations the analysis dives into other possible article topics based on the established statistical demographic and includes statistical evidence as to why those article topics are a good choice. Then, the article dives into a conclusion offering a summary of findings, the implications of those findings and additional steps.

Methodology: This analysis was created with the use of Python in collaboration with the software package Anaconda Navigator, specifically using Jupyter Notebook or Pandas to explore and model data.

Descriptive Statistics:

Central Tendency:

Mean Age of Subscribers: 30.11 years old

Median Age of Subscribers: 30.0

Mean Household Income: \$74,450

Median Household Income: \$66,050

Variability:

Range of Age: 23

Variance of Age: 16.19

SD of Age: 4.02

Range of HouseholdIncome: 306300

Variance of HouseholdIncome: 1212307794.382491

SD of HouseholdIncome: 34818.2106717518

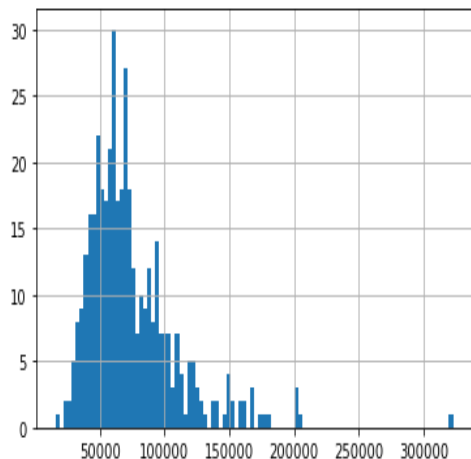
Range of ValueInvestments: 133400

Variance of ValueInvestments: 249982368.71608347

SD of ValueInvestments: 15810.830740858732

Distribution Shape:

HouseHold Income(Continuous RV)



The Distribution has a right skew, which means that many more of the subscribers to Young Professional Magazine have a lower income as compared to a higher income.

This makes sense because the magazine's target audience is recent college graduates, so you would expect the majority to be making a lower income.

A right skew also makes sense because the mean is greater than the median, so when analyzing the central tendency, you could have made an informed prediction of the histogram for Household income having a right skew.

In terms of developing a target demographic for new subscribers, this histogram shows that current subscribers are more likely to have a household income below \$74,450

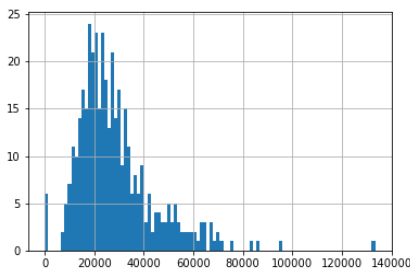
Comments: While this histogram does give us a grasp on the target household income, it is not necessarily

Kurtosis Value: 7.389

This tells us that the data has heavy tails and a very peaked center, it means that the data may have more outliers than a normal distribution

accurate as "Household income" refers to one's income as well as their spouse (if they have one) This implies that some of these subscribers income could be coming from themselves alone (the subscriber) OR could be coming from them and their spouse. Making it more difficult to get a grasp on just the subscribers income

Value of Investments(Continuous RV)



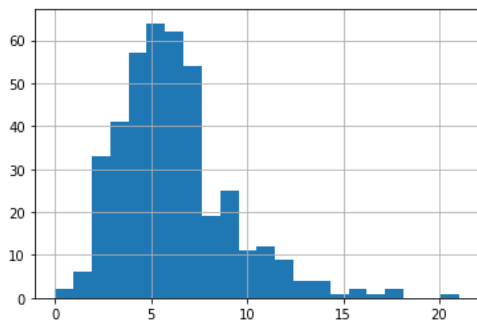
Kurtosis Value: 5.562

This tells us that the data also has fairly heavy tails and a very peaked center, it means that the data may have more outliers than a normal distribution

The distribution once again has a right skew, which means that many more of the subscribers to Young professional magazine have a smaller value of investments as compared to a larger value of investments.

Prematurely, there could be a correlation between both the household income and value of investments, and this would make sense logically as well. To test this I calculated the correlation by calculating the r value between the two columns, to see if there was any linear relationship between the two variables. This came to an R value of 0.003, indicating that the premature prediction was incorrect and there was no correlation between household income as well as the value of investments.

Number of Transactions(Continuous RV)



	Age	Value of Investments (\$)	Number of Transactions	Household Income (\$)
Age	1.000000	-0.029727	0.021208	0.010395
Value of Investments (\$)	-0.029727	1.000000	0.085912	0.003026
Number of Transactions	0.021208	0.085912	1.000000	0.070633
Household Income (\$)	0.010395	0.003026	0.070633	1.000000

Kurtosis Value: 2.415

This tells us that the data has tails that are slightly lighter than those of a Normal distribution

The distribution here is also skewed to the right, however this time it was a little more difficult to visualize, because at a premature look, the data looks somewhat uniform. To confirm, the skew value which comes from the Python Scipy library in my analysis was 1.214, indicating a skew to the right, and taking it even a step further.

This was a more difficult measure of a distribution shape because looking at mean and median, in this case the median was actually larger than the mean and not the other way around, in fact the two were very close to each other. This usually indicates a somewhat uniform distribution, but once again the skew value confirms that the histogram is skewed to the right.

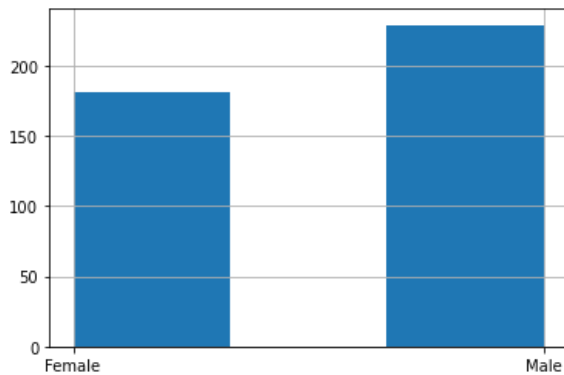
Additionally, furthering the investigation of the current demographic of subscribers, there needs to be analysis done on possible correlation between the number of transactions with the value of investments and the household income, since all of these relate to the subscribers financial status, and could give further insight into the target audience.

This continued analysis led to the use of the built in corr() function in python which allowed me to see the correlation between all of the continuous variables (including age which is technically discrete when rounded to a whole number so in this case let it be

ignored) When looking at this you can see that there is a small but weak correlation between both the number of transactions and value of investments, as well as a correlation between the number of transactions and household income. While these correlation r values are still not considered "a strong correlation" since they are both below 0.1, this could still be useful information for developing a target demographic

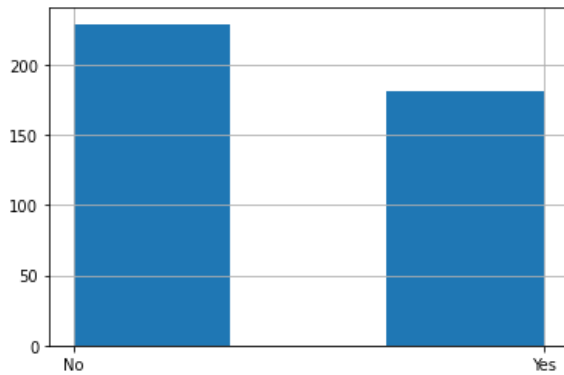
Frequency Distributions

Gender Frequency Distribution



The frequency distribution shows us that there are significantly more males subscribed to the magazine than females, and diving deeper approximately 56% of the subscribers from the survey are males while only 44% of the subscribers are female.

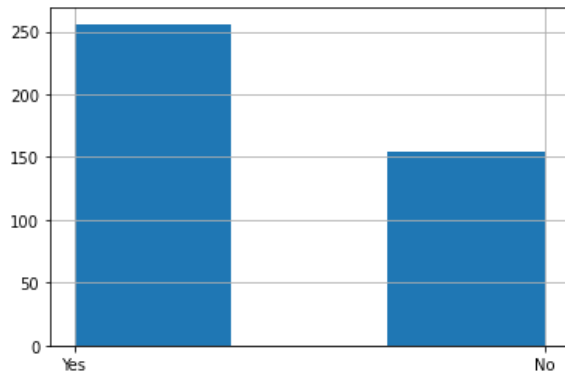
Real Estate Purchases Distribution



The frequency distribution shows us that there are significantly more subscribers who don't plan on making real estate purchases, than those who do. This once again fits into the target demographic of the magazine in the first place, as you wouldn't expect recent college graduates to have plans on real estate purchases.

Diving into percentages, about 56% of the subscribers from the survey don't have any plans for real estate purchases while about 44% of subscribers do have real estate purchase plans.

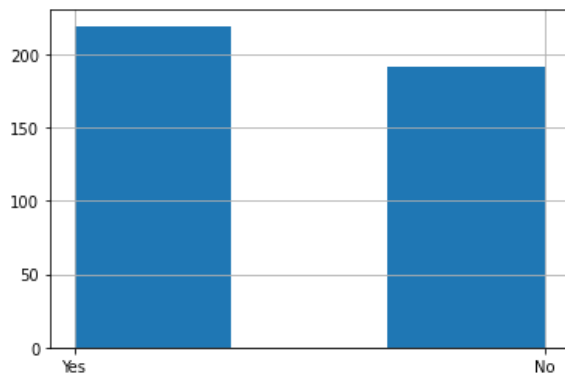
BroadBand Access Distribution



The frequency distribution shows us that more subscribers have broadband access to the internet at home than those who don't.

Specifically, about 62.4% of subscribers from the survey had access to broadband access while about 37.6% of subscribers did not.

Have Children Distribution



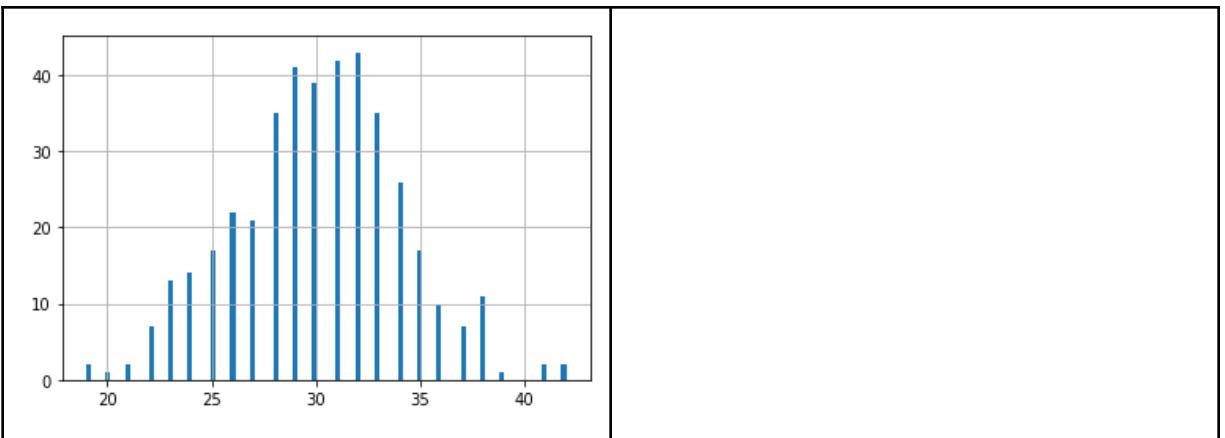
You can see here that it is approximately an even split between subscribers having children vs. subscribers not having children, but slightly more subscribers do have children. This even split makes sense because if you reference the histogram for age below, you can see that the distribution for age looks relatively uniform, and you would expect this to result in a relatively even frequency distribution between subscribers having children and not having children because a subscriber is equally likely to be older as younger, so that most likely correlates with having children or not.

Specifically, 53.4% of subscribers have children and 46.6% do not have children.

Age Distribution

Age technically classifies as a discrete variable in this case and not a continuous one because there is a countable number of values over a specific range. If age was looked at as a decimal in this case it would be continuous, however in the survey age is listed as an integer.

As you can see in the distribution, the age of subscribers looks relatively uniform.



95% Confidence Interval of mean Age

If we were to repeat this process, 95% of the time, the interval 29.72-30.50 years would contain the true population mean. This measure is relatively narrow because the sample size is relatively large. To calculate this confidence interval we used a normal distribution tool within python because the sample size is clearly over 30.

95% Confidence Interval of Household Income

If we were to repeat this process, 95% of the time, the interval \$71,089.25 - \$77,829.76 would contain the true population mean.

95% Confidence Interval Broadband Access?

Yes

If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.577 - 0.671 would contain the population proportion of saying yes to having broadband access as compared to saying no

No

If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.328 - 0.422 would contain the population proportion of saying no to having broadband access as compared to saying yes

95% Confidence Interval Have Children?

Yes

If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.485 - 0.582 would contain the population proportion of saying yes to having children as compared to saying no

No

If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.417 - 0.514 would contain the population proportion of saying no to having children as compared to saying yes

Would Young Professional be a good advertising outlet for online brokers?

While there isn't a direct answer to this question, there are multiple areas of this survey data that we can provide a closer analysis on, to help make inferences as to the answer to this.

The variables in this survey that could help create a hypothetical answer:

Plans to purchase real estate?	Generalizing real estate as an investment, we can infer that people without plans to purchase real estate should have more financial room in their budget to invest in an online brokerage system.
How many stock/bond/mutual fund transactions have you made in the past year?	By getting a grasp of this variable, we can get a gauge on people who would be interested in online brokers. Hypothetically, someone with more stock market interactions should logically be more inclined as to invest in online brokers, as they have done it so many times before.

Using these two variables to conduct our investigation: first let the data set be divided into two subsets of data points. These two datasets are subjects with plans to purchase real estate and subjects without plans to purchase real estate.

Subset without Plans to purchase real estate(Dataset 2)					Subset with plans to purchase real estate (Dataset 3)				
	Age	Value of Investments (\$)	Number of Transactions	Household Income (\$)		Age	Value of Investments (\$)	Number of Transactions	Household Income (\$)
count	229.000000	229.000000	229.000000	229.000000	count	181.000000	181.000000	181.000000	181.000000
mean	30.030568	28140.174672	5.825328	75695.196507	mean	30.215470	29041.988950	6.160221	72896.132597
std	4.078606	16182.467011	3.129523	36417.196675	std	3.962741	15357.535408	3.062636	32716.411724
min	19.000000	0.000000	0.000000	16200.000000	min	19.000000	0.000000	0.000000	24300.000000
25%	28.000000	17700.000000	4.000000	52200.000000	25%	28.000000	18700.000000	4.000000	49600.000000
50%	30.000000	24800.000000	5.000000	66200.000000	50%	31.000000	24700.000000	6.000000	65700.000000
75%	33.000000	33100.000000	7.000000	88400.000000	75%	33.000000	36000.000000	8.000000	89500.000000
max	42.000000	133400.000000	18.000000	322500.000000	max	41.000000	95200.000000	21.000000	201700.000000

This separation is done to get a better look at the descriptive statistics of the two datasets and compare. Specifically, this investigation aimed to provide reason as to why the subjects in dataset 2 (contains the majority of the subjects in the dataset) should be more interested in online brokers than dataset 3, provided the inference that the subjects in dataset 2 have more financial room to invest.

First we conclude that the subjects without plans to purchase real estate on average make more money than the subjects who do plan on purchasing real estate. Thus, logically having even more buying power/financial room to invest in online brokers. This could very well mean that the subjects in dataset 2 would be a great fit for online broker advertising, and given that the subjects in dataset 2 are the proportional majority in the dataset, this supports the claim that Young Professional Magazine would be a good place for online broker advertisements.

The next variable of interest in the dataset was clearly how many stock/bond/mutual fund transactions the subjects had made in the past year. If the investigation could show that subjects with a larger number of recent transactions over the last year, had more of a reason to invest additional money, then that would also support the claim that Young Professional

Magazine would be a good place to advertise online brokers.

To begin, we pick the under/over separation mark at 6 recent transactions. So we split the original dataset into two subsets, 1 subset containing the subjects with 6 or more recent transactions (subset 4), and 1 subset containing the subjects with less than 6 recent transactions (subset 5).

Subset containing subjects with 6 more recent transactions (subset 4)						Subset containing subjects with less than 6 recent transactions (subset 5)					
Correlations between variables:						Correlations between variables:					
	Age	Value of Investments (\$)	Number of Transactions	Household Income (\$)	U		Age	Value of Investments (\$)	Number of Transactions	Household Income (\$)	U
Age	1.000000	0.034765	0.10781	0.048185		Age	1.000000	-0.094214	-0.060912	-0.035542	
Value of Investments (\$)	0.034765	1.000000	0.17363	0.051299		Value of Investments (\$)	-0.094214	1.000000	0.049151	-0.053889	
Number of Transactions	0.107810	0.173630	1.00000	0.087310		Number of Transactions	-0.060912	0.049151	1.000000	0.058933	
Household Income (\$)	0.048185	0.051299	0.08731	1.000000		Household Income (\$)	-0.035542	-0.053889	0.058933	1.000000	

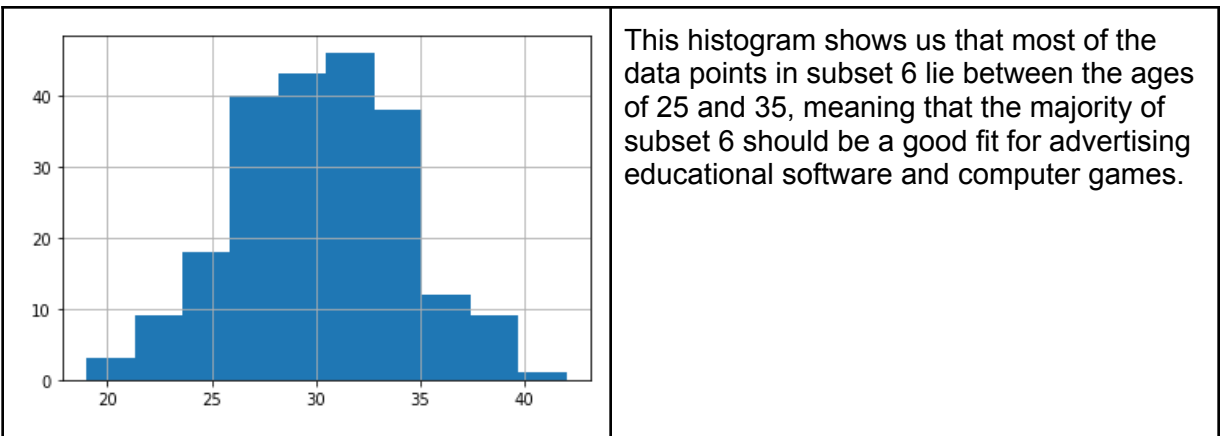
To help support this claim, we can look at the correlation between variables in the two different subsets and compare. This comparison shows that subjects in subset 4 have a much higher correlation between the number of transactions and the corresponding value of investments than the subjects in subset 3. This means that the people who make more transactions in the stock market, earn more money on those transactions on average, and at the same time we can infer that those people are more likely to invest in additional opportunities such as online brokers, because they on average have seen a larger ROI then those who have invested less.

Furthermore, both of these subsets of interest in this particular investigation, that we believe would be more likely to invest in online brokers make up 330/410 entries in the dataset (about 80%). These findings lead me to conclude that yes, Young Professional Magazine would be a good place to advertise online brokers.

Would this magazine be a good place to advertise for companies selling educational software and computer games for young children?

To begin this investigation, we should obviously first filter the data by subjects that answered yes to having children since they would be the main demographic of interest. Let's call this data subset, subset 6. Subset 6 contains a little more than half of the whole dataset, so this is a preliminary good sign.

Logically, we would want the parents to be older than 25, and younger than 35, as these are possible cutoff ages as to when the kids could be too young or too old to be using educational software and computer games. To visualize the ages of the parents of children, using subset 6, we create a histogram.



In addition, these parents would need broadband access to the internet at home in order to use the software and video games, so we filter the data to see how many of the subjects from subset 6, fall between the ages of 25 and 35, and also have broadband access to the internet at home. This takes subset 6 which has 219 subjects and brings it to only 114 subjects. This means that only 114 of the 410 subjects in the dataset would be a good fit for this type of advertising. This is only 27% of the dataset, so this concludes that Young People Magazine would not be a good place to advertise for companies selling educational software and computer games for young children, and that another advertisement could do much better on this demographic of people.

Types of articles of interest to readers

-"How To Make Sure You Are Paid Fairly"
 An article that talked about making sure you are compensated fairly would be a potentially very capturing article for the audience. As you get older and correspondingly either hold more and more positions at organizations, or work at one organization for a longer period of time; you should be making more and more money, as your worth/experience grows in the workfield. Most organizations/businesses give raises or other monetary benefits, the longer an employee works for them, and when moving to a new position at a new company, most individuals will be doing this for an increased salary.

Our dataset however shows quite the opposite. The mean household income for the subjects of the dataset over the age of 30 is about \$73,500, while the mean household income for the subjects of the dataset under the age of 30 is significantly higher: \$75,500. This leads us to believe that there is significant evidence that the older subjects in this dataset need to make sure they are doing a proper job ensuring a fair compensation for their work, given their value and experience.

-"Gender Equality in The Workforce"
 An article that talks about making sure that women are paid equally to men in the workforce, could also be of interest. Within this survey, it can be shown that women make a mean household income of \$72,846 while men make a mean household income of \$75,734. Provided that in this audience of subscribers to the magazine, men make on average 4% more money than women.

-"The Truth About Womens Investments"

Despite women on average making 4% less than men in this study, if we dive deeper into the monetary disparities of the filtered subsets (men and women). We find that the correlation between the number of transactions in stocks/bonds/stock market and the corresponding value of those investments is 30% higher in women than in men's investing habits. This type of article could be empowering for women and could also make men think more about the impulsiveness and type of investments that they are making. This is almost astounding considering the pay gap between men and women in the study as well.

Conclusion

Findings:

The key findings of the investigation reveal that on average the subscribers have a lower income based on the given range, it also reveals that many more of the subscribers have less value of investments in the given range. We also find that there is a correlation between the value of investments and the number of transactions in stocks/bonds, etc. We find basic frequency distributions, revealing that 56% of the subscribers are males, while 44% are females. We also find that 56% of subscribers do not have plans to purchase real estate in the next two years, while 44% do.

Along with other measures of descriptive statistics, we find that out of the 56% of subscribers that don't plan on making real estate purchases on average make 3000 more than the subscribers that do plan on making real estate purchases. We also find that people who make more transactions in the stock market, earn more money on those transactions on average.

Further findings reveal that only 27% of the subjects of the dataset are between the ages of 25-35 and also have access to broadband internet at home, the mean household income of subscribers over the age of 30 is significantly lower as compared to the subscribers under the age of 30, and that women are on average paid less than men for their household income.

Lastly, we find that the correlation between the number of transactions in stocks/bonds/stock market and the corresponding value of those investments is 30% higher in women than in men's investing habits. In fact, men actually have a negative correlation between the number of transactions they make and the corresponding value of those investments/transactions.

Strategic Implications:

Based on the statistical findings of the survey, we conclude that Young Professional Magazine could very well be a great spot for online brokers to advertise online brokers. We can also conclude that the magazine would not be a great advertising spot for companies selling educational software and computer games.

Advertising Recommendations:

Based on the studied demographic, the survey helps us pick some articles that could be of interest to future/current readers. Articles that revolve around gender equality in the workforce would definitely be fitting for the demographic based on gender discrepancies.

Additionally, articles helping readers make sure they are getting paid fairly based on their worth/experience could be of interest given the discrepancies on household income based on age.

Finally, articles uncovering the truth behind investing as a woman and how women make smarter investments with greater ROI than men.

Further Steps:

Further steps could involve articles really showing the discrepancies of financial success based on age/gender/having children, etc. There seems to be a lot of different keys to the financial success of the subjects of the study, so articles emphasizing the discrepancies would help current/future readers learn from these discrepancies and optimize their financial success based on their investments, and plans for the future.