**Analyzing Factors Influencing Cryptocurrency Investment:**

**A Statistical Approach Using Predictive Models**

**Dataset**: American Trends Panel Wave 111 from Pew Research Centre

**Group Name: VARIABLES**

**Group Members**

Shiv Karthee Janardhanan Sarvenaz Dirakvand Eshwar Vennam Danish Basha Shaikh

Sarah Sarika Gangavarapu Goutam Kalyan Kumar

Fall - December 09, 2024

California State University, Long Beach

**TABLE OF CONTENTS**

| **SN. NO** | **CONTENT** |
| --- | --- |
| 1 | INTRODUCTION |
| 2 | OVERVIEW |
| 3 | PROBLEM OF STATEMENT |
| 4 | DESCRIPTION OF DATASET |
| 5 | TARGET VARIABLE |
| 6 | FEATURE VARIABLES |
| 7 | DESCRIPTIVE STATISTICS |
| 8 | HYPOTHESES |
| 9 | MODEL TRAINING AND VALIDATION |
| 10 | SIGNIFICANT VARIABLES |
| 11 | HYPOTHESIS RESULTS |
| 12 | USE CASES |
| 13 | CHALLENGES |
| 14 | LIMITATIONS |
| 15 | CONCLUSION AND FUTURE STUDY |
| 16 | REFERENCES |

**INTRODUCTION**

Probably among the most transformative phenomena of the modern-day financial world has been that of cryptocurrencies. Cryptocurrencies, such as Bitcoin and Ethereum, among others, are the first kind of decentralized currency and investments in history and thus attracted the interest of millions of investors all around the world.A Pew Research Center study estimates that 16% of Americans had used or invested in cryptocurrency as of 2021, and the percentage is still rising.. On the other hand, most of them have investments in cryptocurrency without proper awareness of its risks, the underlying technology, or how markets work.

Aggravating this is usually a path of uninformed decisions, financial losses, and increased vulnerability to market volatility. Addressing these gaps is critical for improving financial literacy and enabling better decision-making among potential investors.

This study seeks to predict the likelihood of a person investing in cryptocurrency based on a few key factors. The goal is to create a predictive model from survey data that identifies those most likely to invest in cryptocurrencies. By knowing these factors, this study will try to provide insight into who is likely to invest, why they are investing, and how to empower them with better knowledge and tools for decision-making.

The problem we intend to solve: "Why will anyone invest in cryptocurrency when they aren't fully familiar with it, and which factors trigger investment in crypto?" As we try to answer this, we will examine features against the age, income level, education attainment, application of technology tools, social network effect, and past investment experience against chances of one investing in Cryptocurrency.

**OVERVIEW**

**Motivation and Importance**

Crypto users fall unevenly across demographic groups, significantly influenced by age, gender, and financial literacy. According to a report provided by Statista (2020), more than 40% of users of cryptocurrency are 18-34 years, and the trend is highly predisposed towards tech-savviness and individuals with higher incomes. Lack of education and awareness of risks of cryptocurrency is a very important barrier to wider adoption. These statistics show that targeted marketing and education campaigns are necessary to bridge this knowledge gap for responsible investment to be understood.

Furthermore, the volatility in cryptocurrency, along with a lack of clarity from regulators, further contributes to the risks emanating from uninformed investors. For example, the 2018 crash in the price of Bitcoin saw many investors scramble to get out of the market predicament that could have been avoided if those investors had better understood the associated risks and motivations. This work is supposed to help further educational campaigns, financial literacy programs, and targeted marketing in an effort to build a better-educated and more confident investor.

**Objective of the Study**

The major focus of this research work is to analyze survey data through regression analysis and other predictive modeling techniques to identify factors that influence the adoption of cryptocurrencies. By developing a predictive model, we are going to establish the quantitative relationship between demographic variables, behavioral patterns, and investment decisions.

The study shall fill the gap in understanding the psychological, social, and economic drivers of investment behavior in cryptocurrencies.

The ultimate goal is to develop a practical tool for predicting the likelihood of investment in cryptocurrency using verified, trustworthy sources of data and robust statistical methods. Its findings will help financial institutions, marketers, and policymakers create strategies to educate potential investors in designing better financial products and establishing effective regulations.

**PROBLEM STATEMENT**

**Problem**

As much as cryptocurrency has become a popular investment alternative, people invest in them without knowing the risks involved or the underlying technology and how the markets work. This dearth of knowledge culminates into uninformed decisions leading to losses and heightened exposure to volatility.

**Scope**

This research tries to fill the gap in understanding through the analysis of survey data, focusing on the main factors influencing one's decision to invest in cryptocurrency. The factors are going to give insights into the behavioral patterns and demographic trends driving cryptocurrency adoption.

**Impact**

The conclusions derived from this project are assured to give insight into meaningful action that could improve the state of financial literacy, further better decision-making, and thus give people a chance to face investments in cryptocurrency with more confidence and knowledge. This will reduce risks involving uninformed investments and build more knowledgeable investors.

**DESCRIPTION OF DATASET**

The **American Trends Panel Wave 111 Survey** is part of the Pew Research Center's ongoing effort to understand social, political, and technological trends in the United States. This wave covers a range of topics, including purchasing behaviors, social media usage, cryptocurrency awareness, online shopping, and gambling.

Cleaning raw survey data consists of several steps. The structured approach to cleaning this American Trends Panel Wave 111 dataset for the task of predicting cryptocurrency adoption should go as follows:

**Understanding the Raw Data**

First of all, the dataset preparation involves familiarization with raw data, and studying the questionnaire PDF to identify what variables and response values are used and which may be considered invalid or system-missing. In other words, proper interpretation of variable definitions should be carried out and the ways of representation of missing or invalid responses in the dataset determined. Common examples of invalid values are placeholder values, like -9, -8, or system-missing values such as SYSMIS in SPSS. By looking at the questionnaire and comparing it with the dataset, we may establish a correct understanding of data that will be cleaned or transformed in later steps.

**Handling Missing Values**

* **Subset A** involves removing all rows with missing or invalid values. In this approach, we first identify missing data, particularly looking for values such as -9, -8, or other placeholders that signify missing responses. We also account for system-missing values (SYSMIS) in SPSS. Instead of replacing these missing values with mean or mode, we remove the entire row containing any missing data to ensure that only complete cases remain for analysis.
* **Subset B** follows a more flexible approach, where we remove rows if they have more than 30% of missing values. For the remaining missing data, we apply imputation methods. Continuous variables are imputed with the mean of the column, while categorical variables are imputed with the mode. This ensures that missing data does not reduce the sample size drastically and helps maintain statistical power while minimizing potential bias introduced by imputation.

**Recoding Categorical Variables**

After addressing missing data, we recode categorical variables in the dataset to prepare them for analysis. This step involves transforming the values in the original dataset into new variables that are more suitable for modeling. For example, if a variable contains multiple categories, we might recode it into binary or dummy variables to make it more accessible for algorithms such as logistic regression or KNN.

**Combining and Splitting Variables**

* Combining Variables involves aggregating multiple related variables into a single composite variable to simplify the analysis. For instance, the variables ONLSHOP1\_a\_W111, ONLSHOP1\_b\_W111, and ONLSHOP1\_c\_W111 can be combined into a single variable called Online\_Shopping. Similarly, variables like ONLSHOP2\_a\_W111, ONLSHOP2\_b\_W111, and ONLSHOP2\_c\_W111 can be combined into Online\_Shopping\_Freq, making the dataset more manageable.
* Splitting Variables involves creating new binary or dummy variables to capture specific categories. For example, the variable F\_AGECAT can be split into Age\_18\_29 and Age\_30\_49, allowing us to analyze age groups separately. Similarly, F\_USR\_SELFID can be split into Rural and Urban to categorize respondents based on their location.

**Checking for Outliers**

Outlier detection is another important step in data cleaning. We examine numeric variables like age and income to identify extreme values that may distort the results of statistical models. Visualization tools such as box plots or histograms are useful for detecting outliers and understanding the distribution of data. In this case, it was found that there are no extreme values in the dataset, meaning the data is likely well-behaved for analysis.

**Validating the Data**

Once the data has been cleaned, we perform validation steps to ensure its integrity. We check for duplicate entries in the dataset, ensuring that multiple entries for the same respondent are justified. Additionally, we verify that the variable types are correct and align with the intended purpose. For example, age should be numeric, while gender should be categorical. Ensuring that the data is free from errors and inconsistencies is crucial for performing reliable analysis.

Example Cleaning Steps for Cryptocurrency Adoption

For the cryptocurrency adoption analysis, the raw data presents several cleaning challenges. The dependent variable, Crypto\_Invest, is the target of our analysis and needs to be processed accordingly. The independent variables include age, gender, income, and social media use.

* **Age**: Age ranges are converted into midpoints or dummy variables for more precise modeling. This conversion helps to capture age-related trends more effectively.
* **Gender**: Gender values are re-coded, with 1 representing male, 2 representing female, and -9 representing missing data. These missing values will be handled either by removal or imputation.
* **Income**: Income values of -9, which represent refusal to disclose income, are replaced with NaN and imputed using appropriate methods.
* **Social Media Use**: Variables like SNSUSE1, which indicates the use of social media for news, are combined into a single feature to simplify the analysis. This consolidation enables a more straightforward assessment of how social media use impacts cryptocurrency adoption.

Through these cleaning and transformation steps, the dataset is made suitable for use in models like logistic regression and KNN, ensuring that the analysis is both accurate and meaningful.

**TARGET VARIABLE**

When taking Crypto\_Invest as the target variable in SPSS, the objective is to identify factors that influence whether respondents choose to invest in cryptocurrency. This variable is likely binary, allowing for a logistic regression analysis or other classification techniques to explore relationships with predictors.

Description of Analysis

By setting Crypto\_Invest as the target variable, we can investigate which demographic, behavioral, or psychological factors significantly predict cryptocurrency investment. Predictors might include variables such as age, marital status, education, risk tolerance, and income levels, as well as behavioral metrics like social media usage, online shopping preferences, or betting experience. These variables can be tested for their contribution to the likelihood of cryptocurrency investment.

Importance of the Target Variable

Using Crypto\_Invest as the target variable provides actionable insights into cryptocurrency adoption patterns. This analysis could help organizations, policymakers, or marketers understand the barriers and motivations for different demographic groups, enabling tailored strategies to promote or regulate cryptocurrency usage effectively.

**FEATURE VARIABLES**

When analyzing Crypto\_Invest as the target variable, several feature variables can be used to understand its predictors. These variables can be grouped into demographic, behavioral, psychological, and economic categories.

1. **QKEY**
   1. Definition: Unique identifier for each respondent
   2. Scale: Nominal
   3. Values: Numbers
2. **Me\_too\_Support**
   1. Definition: Respondent's attitude towards the #MeToo movement
   2. Scale: Ordinal
   3. Values: 1 (Have not heard) to 6 (Strongly support)
3. **Online\_Shopping\_PC**
   1. Definition: Use of desktop/laptop for online shopping
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)
4. **Online\_Shopping\_Phone**
   1. Definition: Use of phone for online shopping
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)
5. **Prefer\_Shopping\_Online**
   1. Definition: Preference for online shopping
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)
6. **Use\_Social\_Media**
   1. Definition: Social media usage
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)
7. **Following\_Influencers**
   1. Definition: Following influencers on social media
   2. Scale: Nominal (Binary)
   3. Values: 0 (No/Not sure), 1 (Yes)
8. **Influenced\_Purchase**
   1. Definition: Made purchase based on an influencer post
   2. Scale: Nominal (Binary)
   3. Values: 0 (No/Not sure), 1 (Yes)
9. **Influenced\_by\_Influencers**
   1. Definition: Impact of influencers on purchase decisions
   2. Scale: Ordinal
   3. Values: 0 (Not at All), 1 (A Little), 2 (A lot)
10. **Married**
    1. Definition: Marital status
    2. Scale: Nominal (Binary)
    3. Values: 0 (Not married), 1 (Married)
11. **Used\_Dating\_Site**
    1. Definition: Ever used a dating site
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
12. **Currently\_Using\_Dating\_Site**
    1. Definition: Currently using a dating site
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)

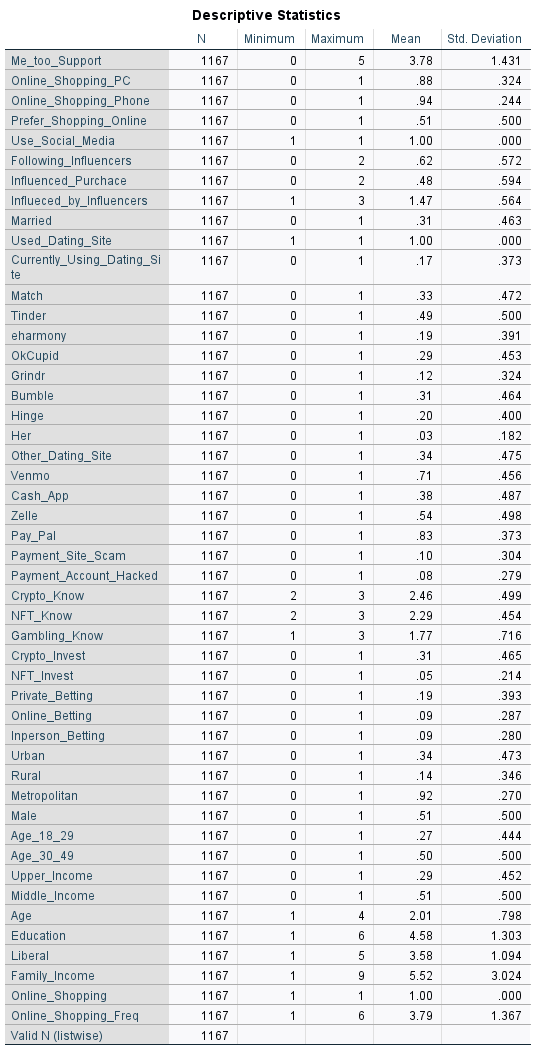
**13-20. Match, Tinder, eharmony, OkCupid, Grindr, Bumble, Hinge, Her**

1. Definition: Ever used a specific dating site
2. Scale: Nominal (Binary)
3. Values: 0 (No), 1 (Yes)
4. **Other\_Dating\_Site**
   1. Definition: Used other dating site
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)

**22-25. Venmo, Cash\_App, Zelle, Pay\_Pal**

1. Definition: Ever used a specific payment app
2. Scale: Nominal (Binary)
3. Values: 0 (No), 1 (Yes)
4. **Payment\_Site\_Scam**
   1. Definition: Victim of payment site scam
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)
5. **Payment\_Account\_Hacked**
   1. Definition: Payment account hacked
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)
6. **Crypto\_Know**
   1. Definition: Knowledge about cryptocurrency
   2. Scale: Ordinal
   3. Values: 1 (Nothing at all) to 3 (A lot)
7. **NFT\_Know**
   1. Definition: Knowledge about NFTs
   2. Scale: Ordinal
   3. Values: 1 (Nothing at all) to 3 (A lot)
8. **Gambling\_Know**
   1. Definition: Knowledge about gambling
   2. Scale: Ordinal
   3. Values: 1 (Nothing at all) to 3 (A lot)
9. **NFT\_Invest**
   1. Definition: Invested in NFTs
   2. Scale: Nominal (Binary)
   3. Values: 0 (No), 1 (Yes)
10. **Private\_Betting**
    1. Definition: Bet money on sports with friends/family
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
11. **Online\_Betting**
    1. Definition: Bet money on sports online
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
12. **Inperson\_Betting**
    1. Definition: Bet money on sports in person
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
13. **Urban**
    1. Definition: Lives in an urban area
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
14. **Rural**
    1. Definition: Lives in a rural area
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
15. **Metropolitan**
    1. Definition: Lives in a metropolitan area
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
16. **Male**
    1. Definition: Gender is male
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
17. **Age\_18\_29**
    1. Definition: Age between 18-29
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
18. **Age\_30\_49**
    1. Definition: Age between 30-49
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
19. **Upper\_Income**
    1. Definition: Upper-income tier
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 3 (Yes)
20. **Middle\_Income**
    1. Definition: Middle-income tier
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 2 (Yes)
21. **Age**
    1. Definition: Age category
    2. Scale: Ordinal
    3. Values: 1 (18-29) to 4 (65+)
22. **Education**
    1. Definition: Education level
    2. Scale: Ordinal
    3. Values: 1 (Less than high school) to 6 (Postgraduate)
23. **Liberal**
    1. Definition: Political ideology
    2. Scale: Ordinal
    3. Values: 1 (Very conservative) to 5 (Very liberal)
24. **Family\_Income**
    1. Definition: Family income range
    2. Scale: Ordinal
    3. Values: 1 (Less than $30,000) to 9 ($100,000 or more)
25. **Online\_Shopping**
    1. Definition: Uses any device for online shopping
    2. Scale: Nominal (Binary)
    3. Values: 0 (No), 1 (Yes)
26. **Online\_Shopping\_Freq**
    1. Definition: Frequency of online shopping
    2. Scale: Ordinal
    3. Values: 1 (Less often than monthly) to 6 (Every day)

**DESCRIPTIVE STATISTICS**



This table presents descriptive statistics for various variables in the dataset. Here's an explanation of the key elements:

1. N: This represents the number of valid responses for each variable. All variables have 1167 responses, in this case, indicating no missing data.
2. Minimum and Maximum: These show the lowest and highest values for each variable. For binary variables (0/1), the minimum is usually 0 and the maximum is 1.
3. Mean: This is the average value for each variable. For binary variables, it represents the proportion of "Yes" responses. For example, the mean of 0.88 for Online\_Shopping\_PC indicates that 88% of respondents use a PC for online shopping.
4. Standard Deviation (Std. Deviation): This measures the data spread around the mean. A higher value indicates more variability in responses.

**Key observations:**

1. Some variables (Use\_Social\_Media, Used\_Dating\_Site, Online\_Shopping) have a mean of 1.00 and a standard deviation of 0.000, indicating that all respondents answered "Yes" to these questions.
2. Me\_too\_Support has a mean of 3.78 on a scale of 0-5, suggesting generally positive support for the movement.
3. Online shopping is very common, with high means for PC (0.88) and Phone (0.94) usage.
4. About 51% of respondents prefer shopping online (mean of 0.51 for Prefer\_Shopping\_Online).
5. 31% of respondents are married (mean of 0.31 for Married).
6. Dating app usage varies, with Tinder being the most popular (49% usage) among those listed.
7. PayPal is the most commonly used payment app (83% usage), followed by Venmo (71%).
8. Cryptocurrency knowledge (Crypto\_Know) is relatively high (mean 2.46 out of 3), but only 31% have invested in crypto.
9. Gambling activities are less common, with means around 0.09-0.19 for various betting types.
10. The sample is fairly evenly split between males and females (51% male).
11. Age distribution shows 27% in the 18-29 range and 50% in the 30-49 range.
12. The average education level is high (4.58 out of 6), indicating many respondents have college degrees or higher.
13. The sample leans slightly liberal (mean of 3.58 on a 1-5 scale).

**HYPOTHESES**

H1: Married people are less likely to invest in cryptocurrency.

Married individuals are less likely to invest in cryptocurrency due to financial priorities, risk aversion, and social influences. Shared responsibilities like household expenses, childcare, and retirement planning lead to a preference for stable, low-risk investments. Psychologically, married individuals often prioritize financial security, as their decisions impact their families. Demographic factors, such as being in older age groups or higher income brackets, further reinforce financial conservatism, favoring traditional investments over speculative ventures like cryptocurrency.

H2: Respondents who had their accounts hacked are less likely to invest in cryptocurrency.

Individuals who have experienced account hacks are less likely to invest in cryptocurrency due to security concerns and reduced trust in digital platforms. A hacking incident erodes confidence, making cryptocurrency exchanges appear vulnerable. This amplifies risk aversion, as cryptocurrencies lack regulatory protections and are highly volatile. The psychological impact of such incidents, including feelings of vulnerability and distrust, further discourages speculative investments, leading individuals to prefer secure financial options over risky digital assets.

H3: Men are more likely to invest in crypto.

Men are more likely to invest in cryptocurrency due to differences in risk tolerance, financial behavior, and social influences. Men typically exhibit higher risk appetite, aligning with cryptocurrency’s volatility and potential for significant returns. They often have greater exposure to networks and influencers promoting crypto investments, emphasizing innovation and financial independence. Psychological factors, such as higher confidence in navigating complex financial systems, further drive male participation in speculative markets. Together, these factors contribute to the gender disparity in cryptocurrency investment.

H4: Age 18 - 29 will have more significance than Age 30 - 49.

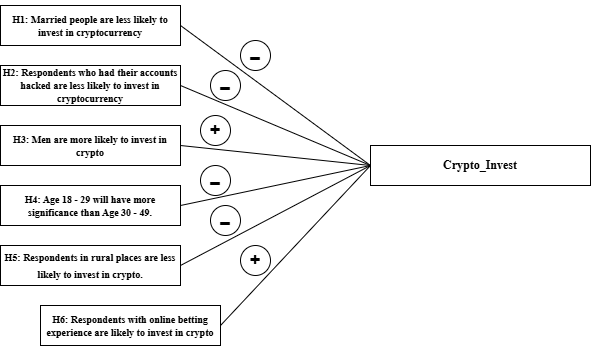
Younger individuals aged 18–29 are more significant in cryptocurrency adoption compared to the 30–49 age group due to their greater tech-savviness, risk tolerance, and social media engagement. This demographic is more comfortable with digital platforms and frequently exposed to cryptocurrency promotions through influencers and online communities. With fewer financial responsibilities and a higher willingness to take risks, they are attracted to the potential for high returns. In contrast, individuals aged 30–49 often prioritize financial stability and long-term investments, making them less likely to engage in speculative ventures like cryptocurrency. These generational differences drive the disparity in cryptocurrency adoption.

H5: Respondents in rural places are less likely to invest in crypto.

Respondents in rural areas are less likely to invest in cryptocurrency due to limited access to technology, financial infrastructure, and crypto-related information. Poor internet access and fewer digital resources hinder participation in cryptocurrency markets, while traditional investments like farming or real estate are preferred for their familiarity and stability. Lower income levels in rural areas also limit discretionary funds for speculative investments, and reduced exposure to online communities and crypto discussions further decreases adoption. Together, these factors make cryptocurrency investment less prevalent in rural regions compared to urban counterparts.

H6: Respondents with online betting experience are likely to invest in crypto.

Respondents with betting experience are more likely to invest in cryptocurrency due to their comfort with risk and speculative behavior. Betting involves risk-taking for potential rewards, aligning with the volatile and high-reward nature of cryptocurrency markets. Familiarity with financial loss and the thrill of uncertainty further motivates them to engage in speculative ventures like crypto trading. Additionally, betting communities often share insights about high-risk opportunities, including cryptocurrencies, encouraging individuals to explore digital assets as an extension of their risk-taking tendencies.



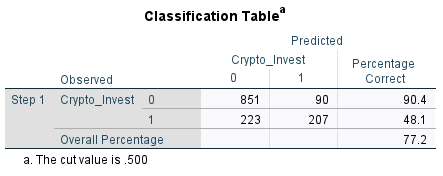
**MODEL TRAINING AND VALIDATION**

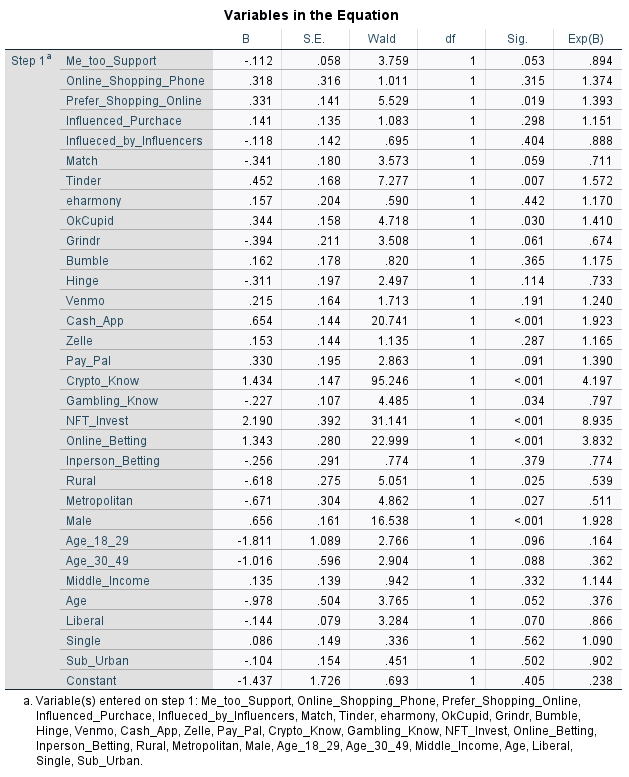
The comparison between Subset A and Subset B illustrates how different approaches to handling missing data can influence the performance of predictive models for cryptocurrency investment. Both models used Logistic Regression and K-Nearest Neighbors (KNN), but their handling of missing values had a significant impact on the results.

Subset A

**Removed all rows with missing values**

* Logistic Regression - 77.2%
* KNN - 75.2%



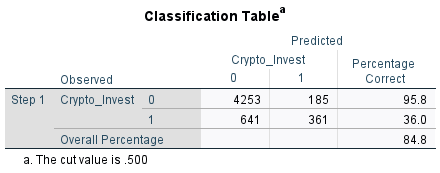


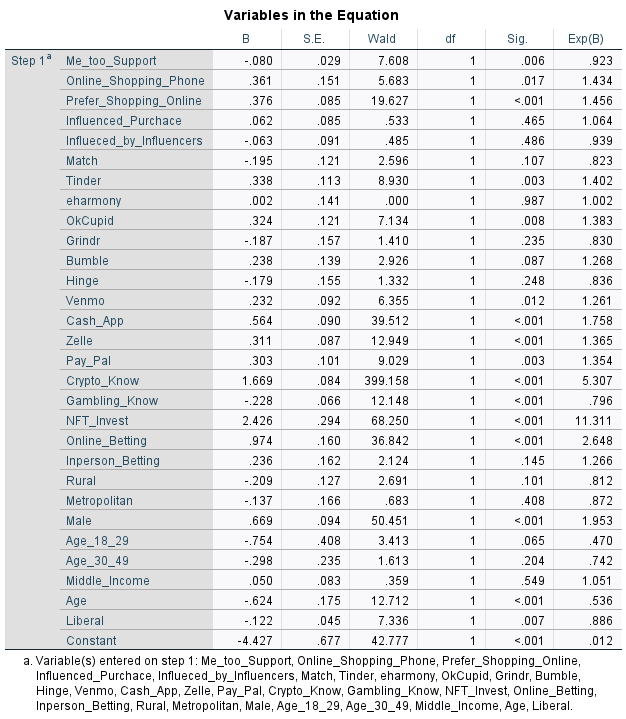
In **Subset A**, all rows with any missing values were removed, resulting in a smaller and cleaner dataset. While this ensured data integrity, it also reduced the sample size, potentially excluding valuable information that could improve subset performance. The logistic regression model achieved an accuracy of **77.2%**, and KNN achieved **75.2%**. Although these results were reasonable, the limited dataset likely constrained the models' ability to generalize effectively.

Model B

**Removed all rows which have more than 30% of missing values**

* Logistic Regression - 84.8%
* KNN - 83.6%





In contrast, **Subset B** adopted a more lenient approach by removing only rows with more than 30% missing values, preserving more of the dataset. This approach maintained a balance between data quality and quantity, allowing the subsets to leverage a broader range of information. Consequently, logistic regression achieved a significantly higher accuracy of **84.8%**, and KNN achieved **83.6%**. The improved performance demonstrates that retaining more data, even with some missing values, can provide models with a more comprehensive understanding of underlying patterns. The number of significant variables has also increased.

These results highlight the importance of thoughtful data preprocessing in machine learning. By striking a balance between cleaning data and preserving valuable information, Subset B achieved superior predictive accuracy, showcasing the benefits of a less restrictive approach to handling missing data.

**SIGNIFICANT VARIABLES**

(Assuming Alpha level = 0.05)

| Variable | B | S.E. | Wald | df | Sig. | Exp(B) |
| --- | --- | --- | --- | --- | --- | --- |
| Cash\_App | 0.654 | 0.144 | 20.741 | 1 | <0.001 | 1.923 |
| Crypto\_Know | 1.434 | 0.147 | 95.246 | 1 | <0.001 | 4.197 |
| NFT\_Invest | 2.19 | 0.392 | 31.141 | 1 | <0.001 | 8.935 |
| Online\_Betting | 1.343 | 0.28 | 22.999 | 1 | <0.001 | 3.832 |
| Male | 0.656 | 0.161 | 16.538 | 1 | <0.001 | 1.928 |
| Tinder | 0.452 | 0.168 | 7.277 | 1 | 0.007 | 1.572 |
| Prefer\_Shopping\_Online | 0.331 | 0.141 | 5.529 | 1 | 0.019 | 1.393 |
| Rural | -0.618 | 0.275 | 5.051 | 1 | 0.025 | 0.539 |
| Metropolitan | -0.671 | 0.304 | 4.862 | 1 | 0.027 | 0.511 |
| OkCupid | -0.344 | 0.158 | 4.718 | 1 | 0.03 | 1.41 |
| Gambling\_Know | -0.227 | 0.107 | 4.485 | 1 | 0.034 | 0.797 |

The analysis reveals several significant predictors of cryptocurrency investment at an alpha level of 0.05. These variables, along with their corresponding odds ratios, highlight the factors that increase or decrease the likelihood of investing in cryptocurrency. Here is a summary of the key findings:

**Positive Predictors**

1. **Cash\_App (Exp(B) = 1.923, p < 0.001):** Respondents using Cash App are nearly twice as likely to invest in cryptocurrency, likely due to the app's integration with crypto trading features, making it more accessible.
2. **Crypto\_Know (Exp(B) = 4.197, p < 0.001):** Knowledge about cryptocurrency strongly increases the likelihood of investment, with a 4.2 times higher chance for those with greater awareness, underscoring the importance of familiarity with digital assets.
3. **NFT\_Invest (Exp(B) = 8.935, p < 0.001):** Respondents who invest in NFTs are significantly more likely to invest in cryptocurrency, as both involve similar digital finance ecosystems and speculative behaviors.
4. **Online\_Betting (Exp(B) = 3.832, p < 0.001):** Those with online betting experience are over 3.8 times more likely to invest, reflecting a higher risk tolerance and familiarity with speculative financial activities.
5. **Male (Exp(B) = 1.928, p < 0.001):** Male respondents are nearly twice as likely to invest in cryptocurrency, which aligns with existing research indicating higher risk tolerance and interest in speculative investments among men.
6. **Tinder (Exp(B) = 1.572, p = 0.007):** Use of Tinder is positively associated with cryptocurrency investment, potentially reflecting a younger, tech-savvy demographic that is open to new digital platforms.
7. **Prefer\_Shopping\_Online (Exp(B) = 1.393, p = 0.019):** Respondents who prefer online shopping are 1.39 times more likely to invest, likely due to greater comfort with digital transactions and technology.

**Negative Predictors**

1. **Rural (Exp(B) = 0.539, p = 0.025):** Living in rural areas decreases the likelihood of investing in cryptocurrency by nearly half, likely due to limited access to digital resources and financial infrastructure.
2. **Metropolitan (Exp(B) = 0.511, p = 0.027):** Surprisingly, living in metropolitan areas also slightly reduces the likelihood of investment. This could indicate that residents of metropolitan areas may focus more on traditional financial systems or have competing investment priorities.
3. **OkCupid (Exp(B) = 1.41, p = 0.03):** The odds of investing decrease slightly for OkCupid users, which may reflect demographic or behavioral differences compared to users of other platforms.
4. **Gambling\_Know (Exp(B) = 0.797, p = 0.034):** Greater knowledge of gambling is associated with a slight decrease in the likelihood of cryptocurrency investment, which could indicate caution from individuals who are aware of speculative risks.

These significant variables provide valuable insights into the demographic, behavioral, and contextual factors influencing cryptocurrency investment. Positive predictors emphasize the role of digital familiarity, risk-taking behaviors, and gender, while negative predictors highlight the impact of geographical and platform-specific factors. This analysis underscores the multifaceted nature of cryptocurrency adoption.

**HYPOTHESIS RESULTS**

H1: Married people are less likely to invest in cryptocurrency.



The coefficient for "Married" is 0.240 with a p-value of 0.270, which is greater than the standard significance threshold (α = 0.05). This result indicates that marital status is not a statistically significant predictor of cryptocurrency investment. Therefore, the hypothesis that married people are less likely to invest in cryptocurrency is not supported by the data.

H2: Respondents who had their accounts hacked are less likely to invest in cryptocurrency.



The coefficient for "Payment\_Account\_Hacked" is 0.177, with a p-value of 0.504, which is also above the significance threshold. This suggests that having experienced an account hack does not significantly affect the likelihood of investing in cryptocurrency. Thus, the hypothesis that respondents who had their accounts hacked are less likely to invest is not supported.

H3: Men are more likely to invest in crypto.



The coefficient for "Male" is 0.675, with a p-value of <0.001, which is highly significant. The odds ratio (Exp(B)) is 1.963, indicating that men are nearly twice as likely as women to invest in cryptocurrency. This supports the hypothesis that men are more likely to invest in cryptocurrency.

H4: Age 18 - 29 will have more significance than Age 30 - 49.



The coefficient for "Age\_18–29" is -1.851, and its p-value is 0.091, while the coefficient for "Age\_30–49" is -1.064, with a p-value of 0.077. Both p-values are greater than 0.05, indicating that neither age group is a statistically significant predictor of cryptocurrency investment. However, the coefficient for "Age\_18–29" is larger in magnitude, suggesting a stronger, albeit non-significant, effect compared to "Age\_30–49."

H5: Respondents in rural places are less likely to invest in crypto.



The coefficient for "Rural" is -0.528, with a p-value of 0.043, which is statistically significant. The odds ratio (Exp(B)) is 0.590, indicating that respondents in rural areas are about 41% less likely to invest in cryptocurrency compared to their counterparts. This supports the hypothesis that respondents in rural places are less likely to invest.

H6: Respondents with online betting experience are likely to invest in crypto.



Engaging in online betting is a strong, statistically significant predictor of the outcome. The coefficient (B = 1.343) indicates that individuals who bet online are more likely to experience the outcome, and the p-value (< .001) confirms the result is unlikely due to chance. The odds ratio (Exp(B) = 3.832) means that participating in online betting raises the odds of the event by nearly fourfold compared to those who do not bet online, after controlling for other factors in the model.

**USE CASES**

1. **Policy Development and Consumer Protection**

Governments and regulatory bodies can use the insights from this study to develop policies aimed at protecting consumers from the risks associated with cryptocurrency investments. Understanding which groups are more likely to invest in cryptocurrency can help in designing educational campaigns or implementing protective measures for at-risk demographics.

1. **Risk Assessment for Financial Institutions**

Banks or financial services providers looking to assess the risk of customers investing in highly volatile assets like cryptocurrency could use this model to predict customer behavior. Financial institutions can better understand the likelihood of their clients investing in crypto, which would help them tailor their risk assessment and mitigation strategies.

**CHALLENGES**

1. **Data Quality and Completeness**

Survey datasets often contain missing, incomplete, or inconsistent responses. Missing data can reduce the model's accuracy and reliability if not handled properly.

1. **Bias in Data Collection**

The survey sample may not represent the entire population, leading to demographic or geographic biases. Insights may not generalize to all potential cryptocurrency investors.

1. **Feature Selection**

Identifying the most relevant variables for predicting cryptocurrency investment can be challenging. Including irrelevant or redundant features may lead to overfitting and reduced model performance.

1. **Interpreting Behavioral Variables**

Variables like "trust in technology" or "gambling behavior" are subjective and may not always be reliable. Results may vary based on how these variables are measured and interpreted.

**LIMITATIONS**

1. **Limited Scope of Variables**

The survey may not include all relevant factors influencing cryptocurrency investment, such as psychological traits or external economic conditions. The model's predictive power may be limited by the variables available in the dataset.

1. **Binary Outcome**

Focusing on a binary target variable (e.g., invest or not invest) may oversimplify the complexity of investment behavior.:The model may not capture nuances like investment amount or frequency.

1. **Dependence on Assumptions**

Regression and predictive models rely on assumptions such as linear relationships or independent features.Violations of these assumptions could reduce the model’s effectiveness.

1. **Overfitting Risk**

Using a complex model with limited data can lead to overfitting, where the model performs well on the training data but poorly on unseen data. Reduced generalizability of the model.

**CONCLUSION AND FUTURE STUDY**

This project successfully identified the key demographic, behavioral, and socio-economic factors influencing an individual's likelihood to invest in cryptocurrency. Among the significant predictors were knowledge of cryptocurrency, participation in online betting, and investment in NFTs, which greatly increased the probability of investment. Demographic factors such as gender and living in rural or metropolitan areas showed significant effects as well. Behavioral patterns, such as the use of financial apps like Cash App, also strongly correlated with cryptocurrency investment.

1. **Knowledge and Awareness Drive Investments**:

Knowledge about cryptocurrency and related digital assets (e.g., NFTs) had the highest positive impact on the likelihood of investing.

1. **Behavioral Indicators Matter**:

Behavioral factors such as online betting and usage of certain financial tools were significant predictors.

1. **Demographics Influence Adoption**:

Men and individuals in suburban areas were more likely to invest compared to women and those in metropolitan/rural areas.

1. **Mixed Influence of Social Factors**:

Factors like influence by social media or influencers had weaker associations, indicating that direct knowledge and behavior outweigh social persuasion.

**Ideas for Improvement**

1. **Enhance Data Quality**: Gather a larger dataset with balanced representation across demographics to improve generalizability.
2. **Address Model Bias**: Use techniques like oversampling or superior machine learning models to improve prediction accuracy for minority classes (e.g., folks who do not invest).
3. **Incorporate Real-Time Data**: Combine real-time financial or market trends to enhance the model's relevance in predicting future behaviors.
4. **Refine Predictive Variables**: Consists of additional features such as risk tolerance, investment history, or exposure to financial education for deeper insights.
5. **Expand Study Scope**: Extend the study to include other regions or cultures to observe variations in cryptocurrency adoption patterns globally.

**REFERENCES**

[1] Faverio, Michelle. Dawson, Wyatt. and Sidoti, Olivia. (October 24, 2024) The majority of Americans aren’t confident in the safety and reliability of cryptocurrency. <https://www.pewresearch.org/short-reads/2024/10/24/majority-of-americans-arent-confident-in-the-safety-and-reliability-of-cryptocurrency/#:~:text=By%20age%20and%20gender&text=42%25%20of%20men%20ages%2018,women%20in%20this%20age%20group.>

[2] Bryman, A., & Cramer, D. (2011). *Quantitative data analysis with SPSS for Windows: A guide for social scientists*. Routledge.

[3] Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

[4] Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

[5] Pew Research Center. (2021). *16% of Americans say they have ever invested in, traded or used cryptocurrency*. Retrieved from<https://www.pewresearch.org/>

[6] Statista. (2020). *Share of cryptocurrency users worldwide by age group*. Retrieved from <https://www.statista.com/statistics/1201463/cryptocurrency-users-by-age-worldwide/>

[7] Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

[8] Van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press.

[9] Zhang, Z., & Wang, L. (2019). Handling missing data in social science research: A methodological review. *Social Science Research*, 78, 48–61. <https://doi.org/10.1016/j.ssresearch.2018.12.002>

[10] Zhang, X., & Zhao, R. (2020). *Understanding cryptocurrency adoption: A behavioral perspective*. Journal of Financial Technology, 7(2), 121–134. <https://doi.org/10.1016/j.jfintech.2020.10.003>

[11] Zou, K., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>