To Approve or Not to Approve?

Desiree Garcia

http://github.com/dezgarci

Wallpaperaccess.com

Background

- Aspiring Data Scientist, Entity Academy
- Born in New Mexico, grew up in rural southern Colorado
- 2010 Colorado State University graduate with a BS in Biology
- Taught Upward Bound high school students about credit
- Employed with Halliburton Energy Services since 2011
  - Work in the laboratory performing testing and data analysis
  - Work closely with engineering and other stakeholders to bring solutions and maximize value for the customer

The History and Use of Credit

An administrative clay tablet from the Mesopotamian/Sumerian era. (Photo by Photo12/Universal Images Group Getty Images)

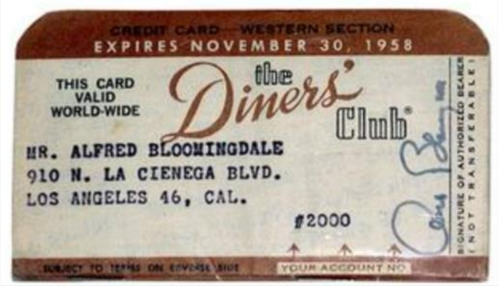
3000 B.C. — As early as 3,000 B.C., Mesopotamian and Harappan civilizations utilized clay tablets to track their trade and transactions

1850 — The earliest form of the modern credit card emerged in the 19th century as credit coins and charge plates to extend credit to local farmers until the harvest came

1950 — The credit card as we know it today can be traced back to the 1950s McNamara Diners Club card, where an increasing amount of loyal patrons could charge their meal at a 7% transaction fee with a $3 annual fee

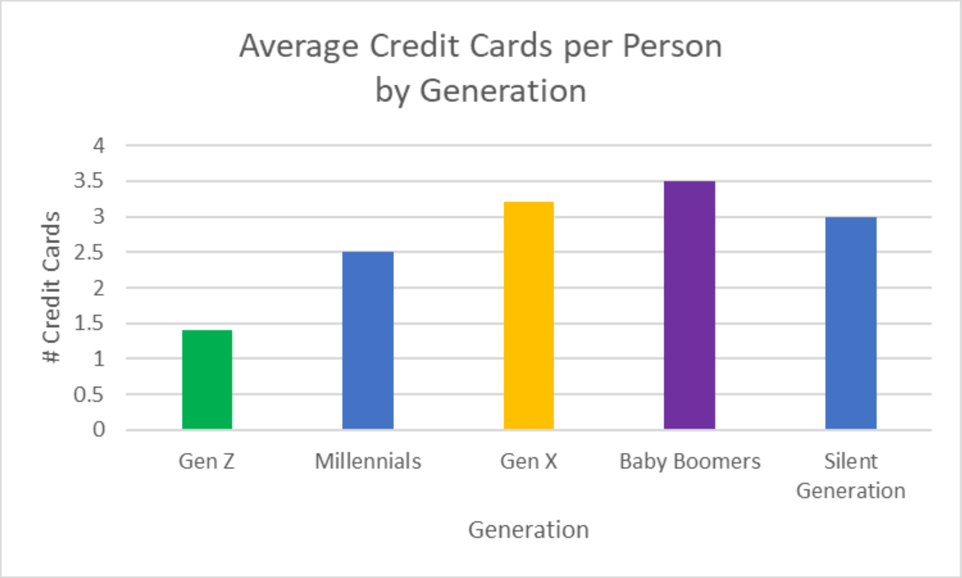
1960s — Since the 1960s, measures have been put into place to ensure the security of the now-familiar plastic cards, from magnetic stripes for over 40 years, to RFID chips in the early 2000s

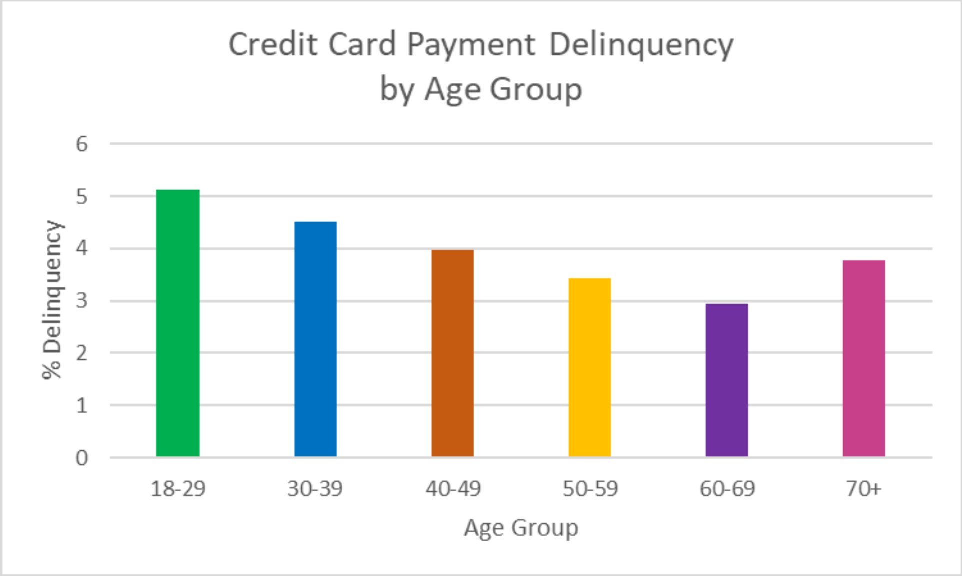
2020 — By 2020, new features like biometric identification and mobile technology allow users to access credit quickly and conveniently



Credit Card Statistics

+ The average American has 3 credit cards
+ Average credit score in the US is 710
+ US consumers utilize 25.3% of their credit card limits on average.
+ The most important factors for credit card approval include:
    + Credit Score
    + Number of Delinquencies
    + Hard Inquiries
    + Credit Utilization Rate
    + Income
    + Credit History





Why It's Important

Determining a borrower's ability to meet their debt obligations is central to estimating the probability of default and subsequent risk of financial loss to the lender.

Methods

1. Select the Dataset
2. Explore and Process the Data
3. Define Approval
4. Evaluate Each Applicant Based on Their History
5. Build a Model To Predict an Applicant's Approval
6. Determine Factors of Most Importance

Dissecting The Dataset

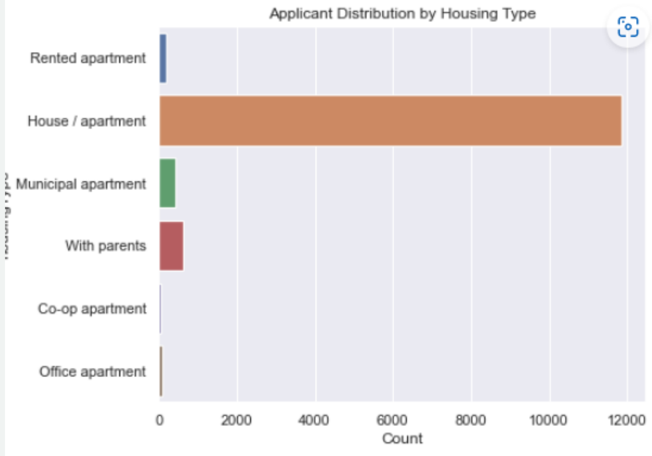
+ Kaggle website browsing led to a dataset with 13,311 unique customer (applicant) information from an undisclosed financial institution containing up to 5 years of monthly data for each applicant

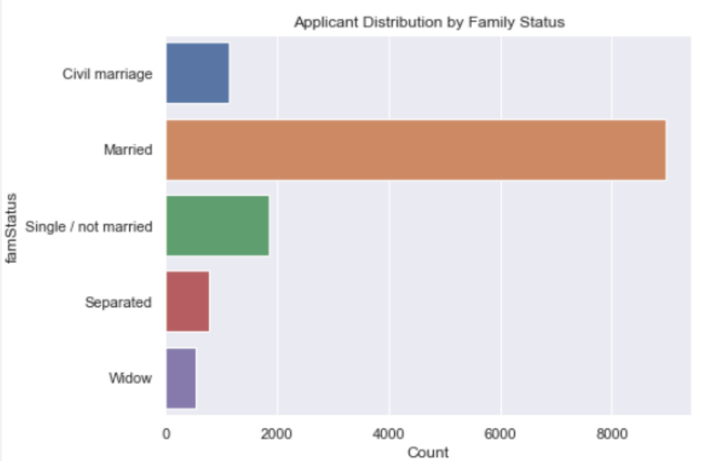
Information for each customer includes:

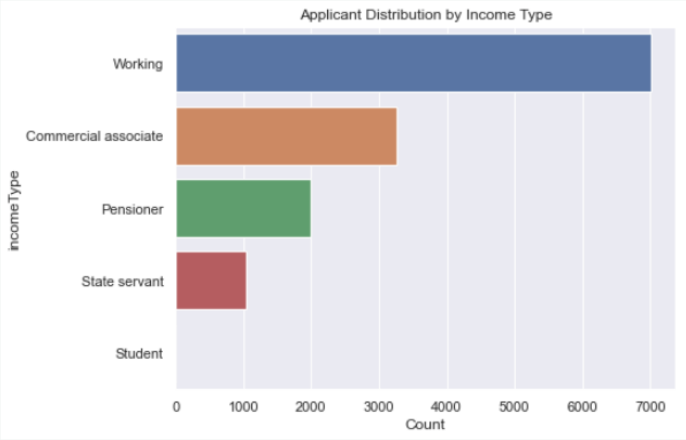
ID	Income Type
Gender	Education Level
Car Ownership	Family Status
Realty Ownership	Housing Type
Number of Children in Household	Occupation
Total Annual Income	Family Size
Age	Monthly Account Status

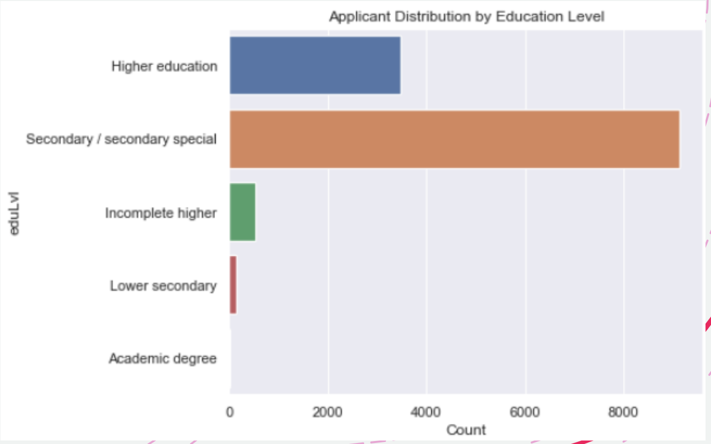
Dissecting The Dataset

- Stats summary
  - Females represent 64% of the applicants, while males represent 36%
  - A high percentage don’t own a car, but most of them own realty.
  - Annual income range: $27,000 to $1,575,000 (Average = $190,000)
  - Family Size: 1 to 20 members
  - Age: 21 to 69 years
  - Years Employed: 0 to 43
  - Account Status History per Applicant: 1 to 60 months







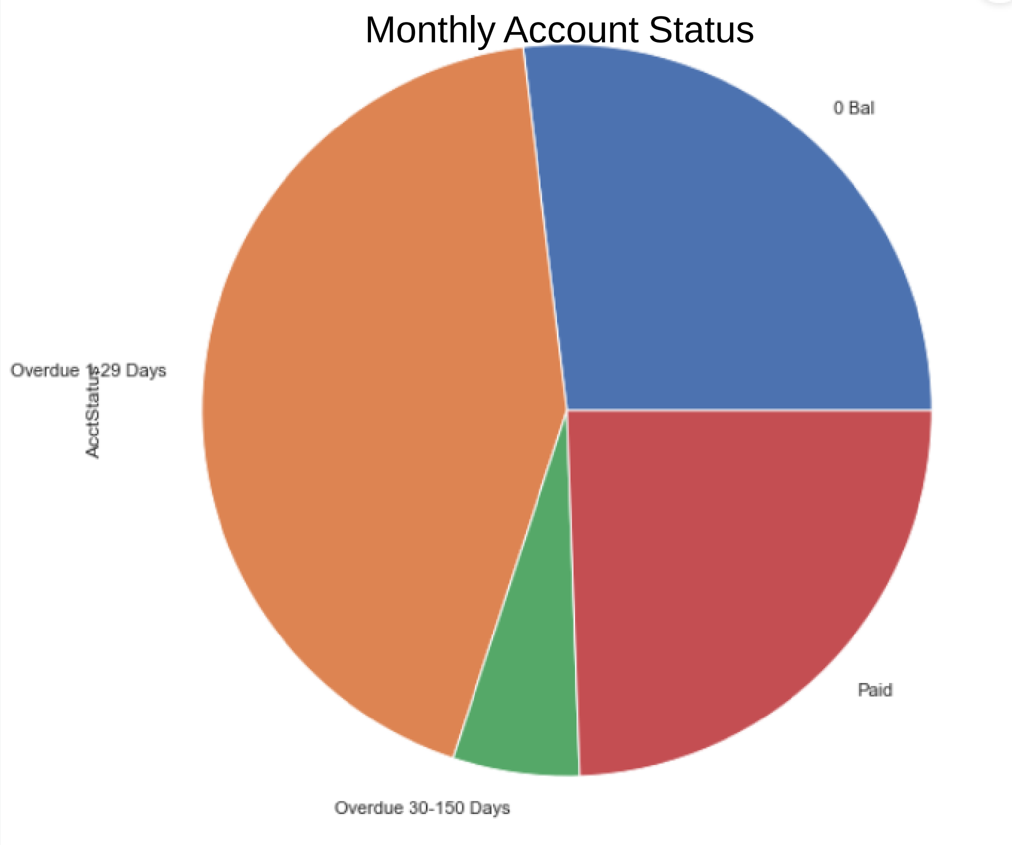


Defining Approval Status

- Approval/Rejection was not given as part of the dataset.
- A method for dividing the applicants into “good” and “bad” was developed by determining the number of months an applicant had a zero balance and/or paid in full and compared it to the number of months an applicant was late on payment:

$$\frac{Good}{Bad} \geq 1 \longrightarrow \text{Approve}$$

$$\frac{Good}{Bad} < 1 \longrightarrow \text{Reject}$$

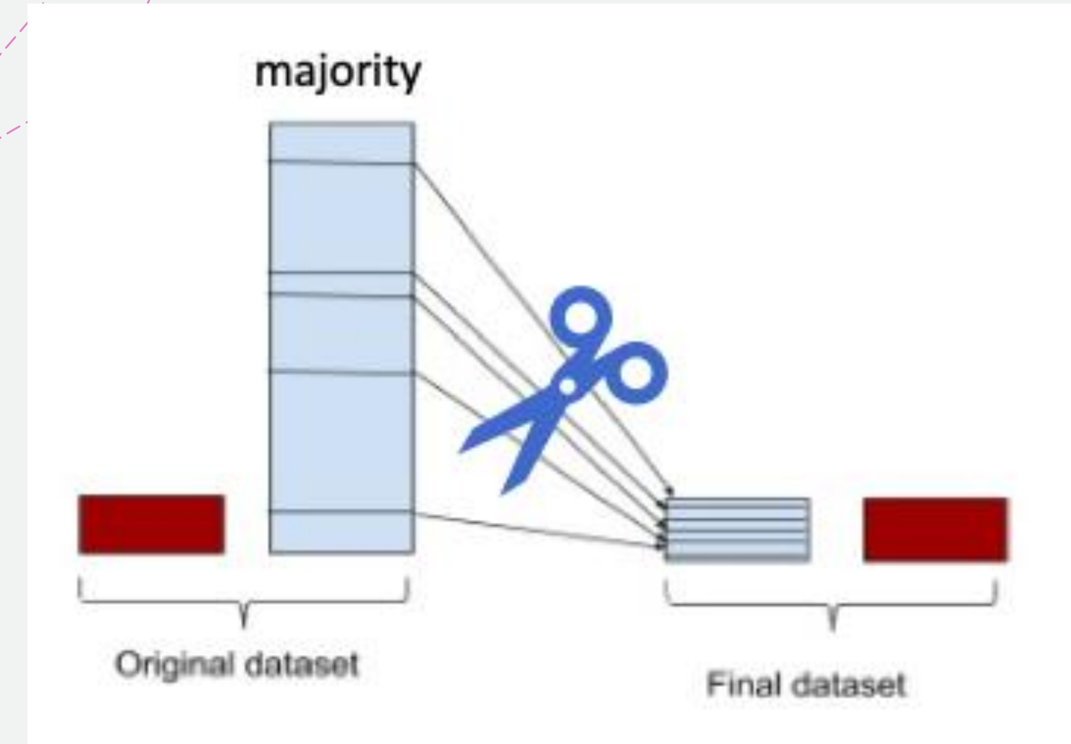


Very close to 50% of these applicants would be approved for a credit card under the proposed Policy.

New Categories

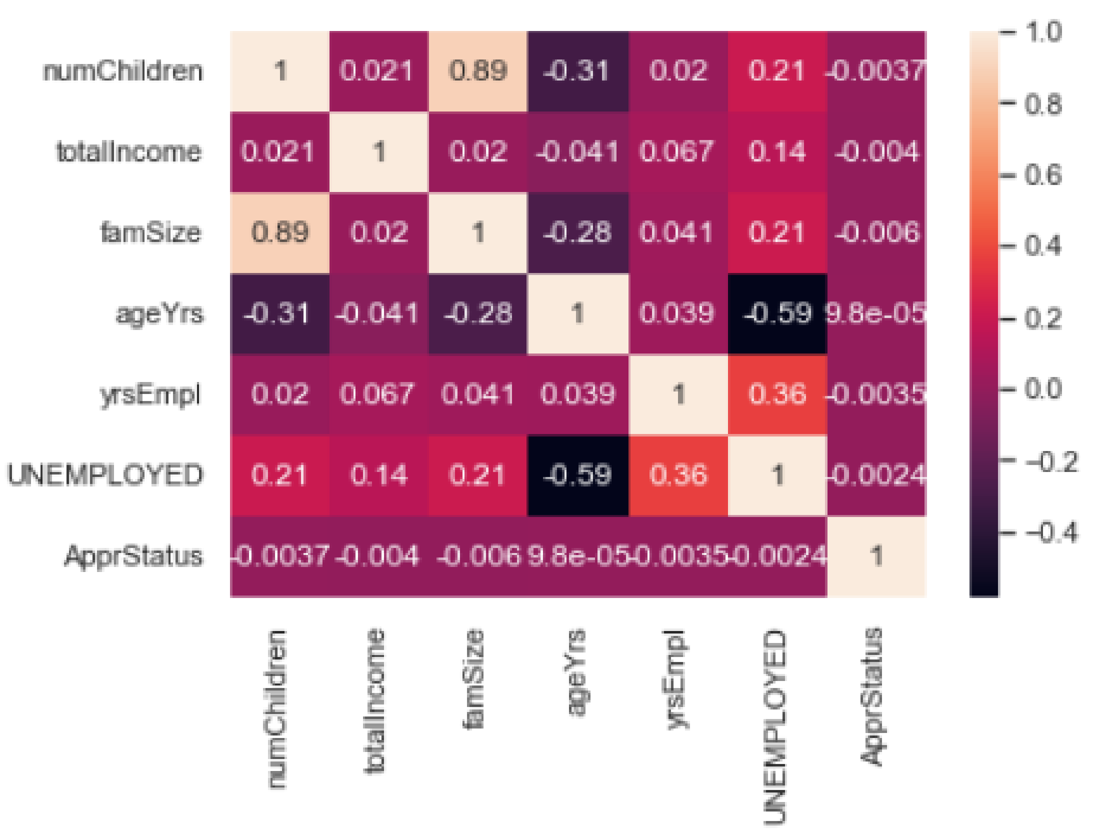
+ Three categories were created to capture additional information gleaned from the original dataset:
  - Approval Status (Response Variable)
  - Unemployment (Predictor Variable)
  - Rate (measures each applicant’s tendency to keep their account in ‘good’ standing; used to compare groups to each other during Exploratory Analysis)

Tackling Bias



- Females are represented nearly 2:1 compared to their male counterparts
- In order to make more accurate comparisons and limit bias in favor of the majority class, a method was used to “balance” the dataset by gender.
- Random resampling (via undersampling the majority class while keeping all of the data in the minority class) allowed the creation of a balanced dataset without compromising the overall data shape.
  - Con: Could lead to loss of information which may be important
- The resulting data frame with which subsequent analyses were performed yields 9500 records (4750 male, 4750 female)

Machine Learning

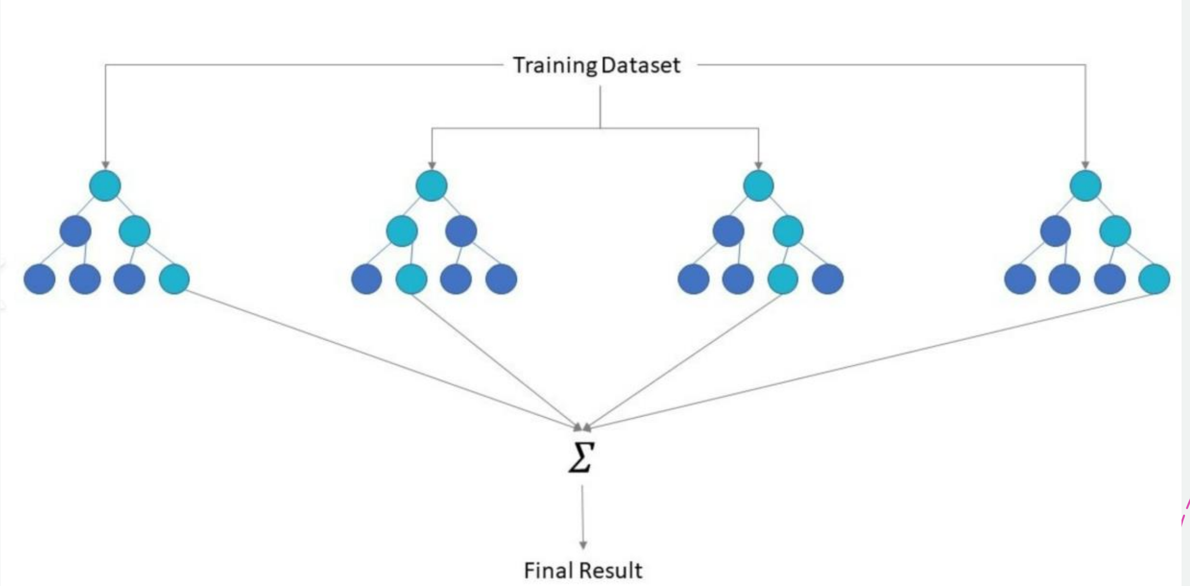


Goal: Create a model that accurately predicts approval status outcomes

- Step 1: Determine highly correlated factors
  - Unemployment and age of the applicant were negatively correlated
  - Family size and number of children highly positively correlated
    - Removed number of children factor to avoid redundancy in the model

Machine Learning

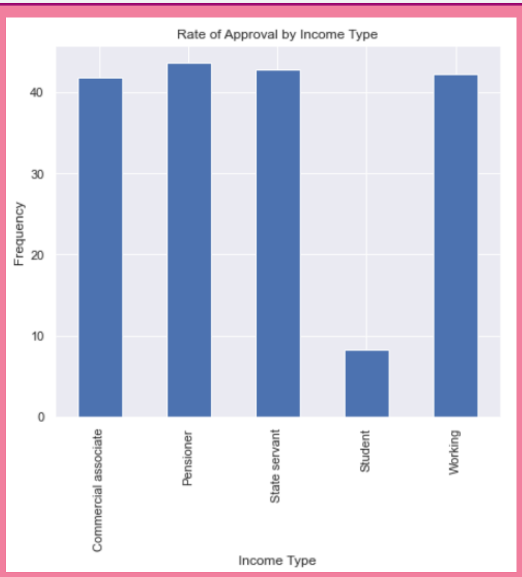
- Step 2: Define the target variable (Approval Status) and predictors (factors in the dataset)
- Step 3: Split the data into training and testing sets and then run the linear regression model on the training data
- Step 4: Determine the accuracy of the model
- Step 5: Features engineering: determine which factors are most relevant in the prediction of credit card approval status
- Step 6: Train a model using the features of most importance and evaluate its accuracy

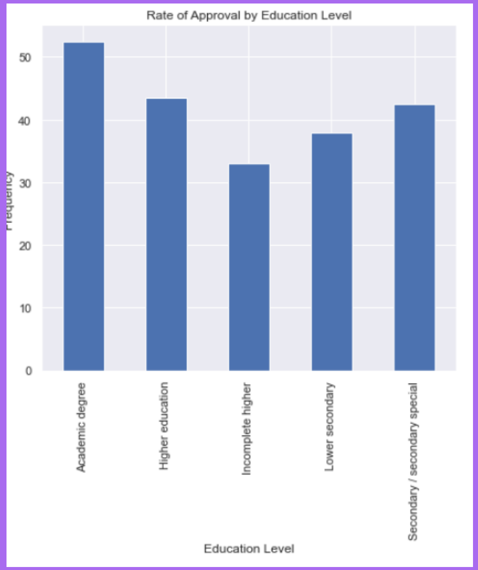


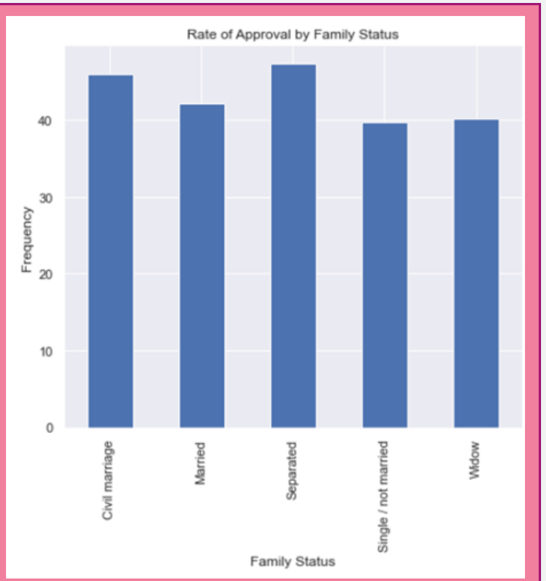
Results

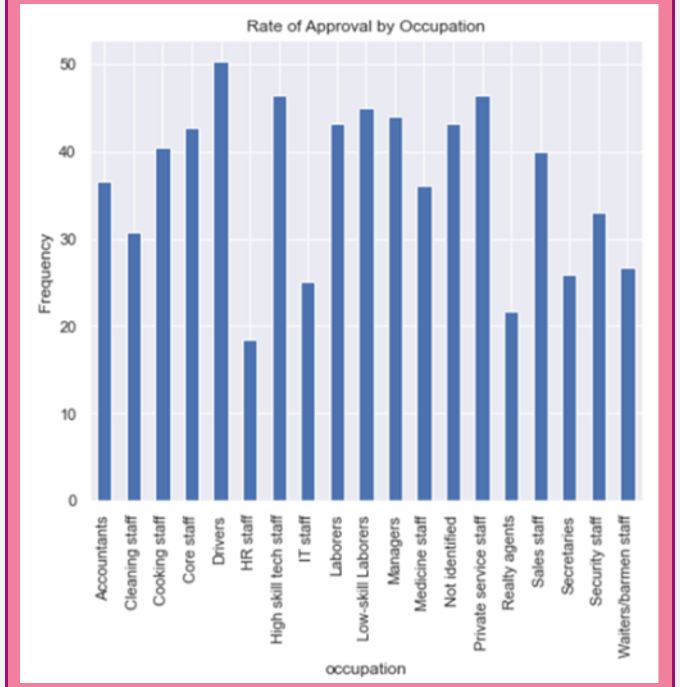
Exploratory Analysis

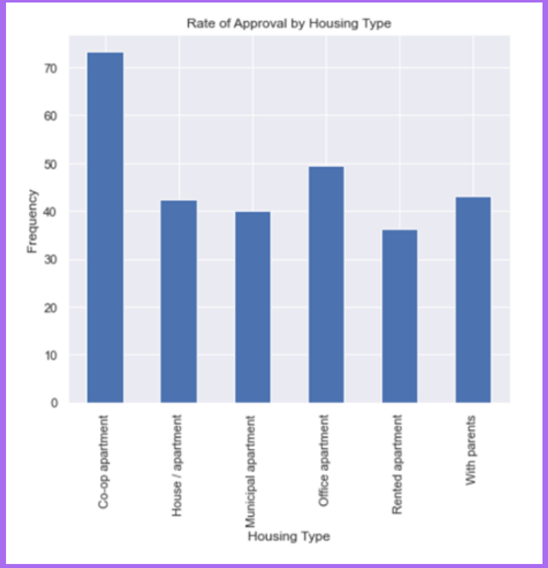
- Did certain demographics of applicants have better or worse rates of approval?
  - Students had the lowest approval rate
    - Once they complete college, they tend to keep up with their credit accounts and can boost their approval chances.
  - HR, real estate, and IT staff came up short compared to drivers, high skill techs, and private service staff.
  - Married individuals had slightly higher approval rates than single individuals
  - Applicants living in apartments have a higher approval rate than their peers who own homes or live with their parents.
  - Owning a car and owning realty has a negligible effect

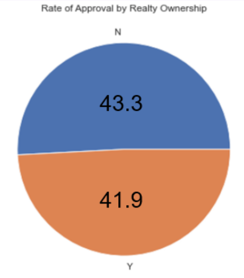


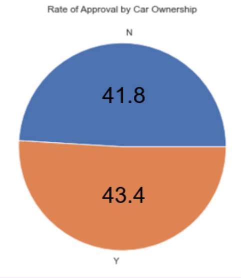








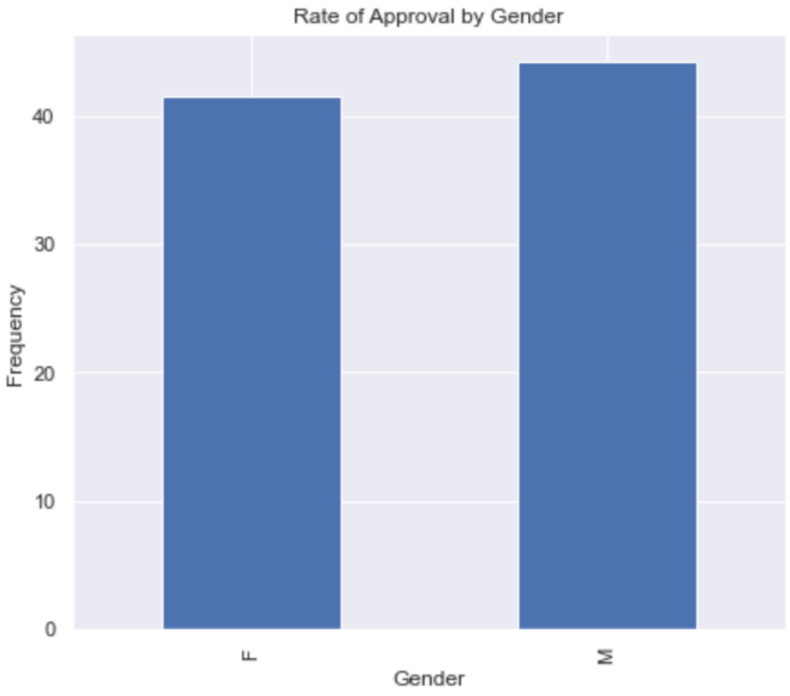




Battle of the Sexes!

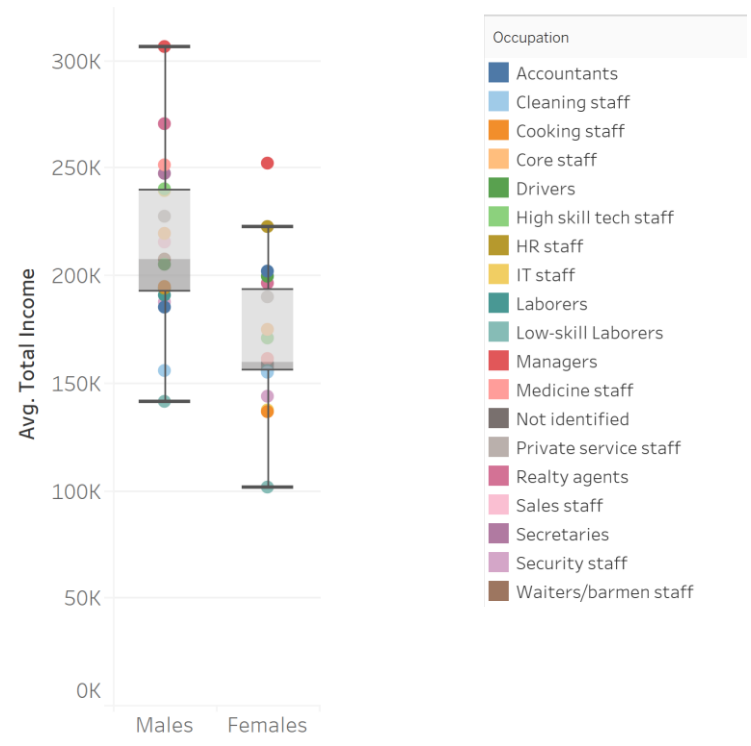
Is gender a good predictor of whether an applicant will be approved?

- An Independent Chi Squares analysis determined that there was no significant difference in approval rates between males and females.



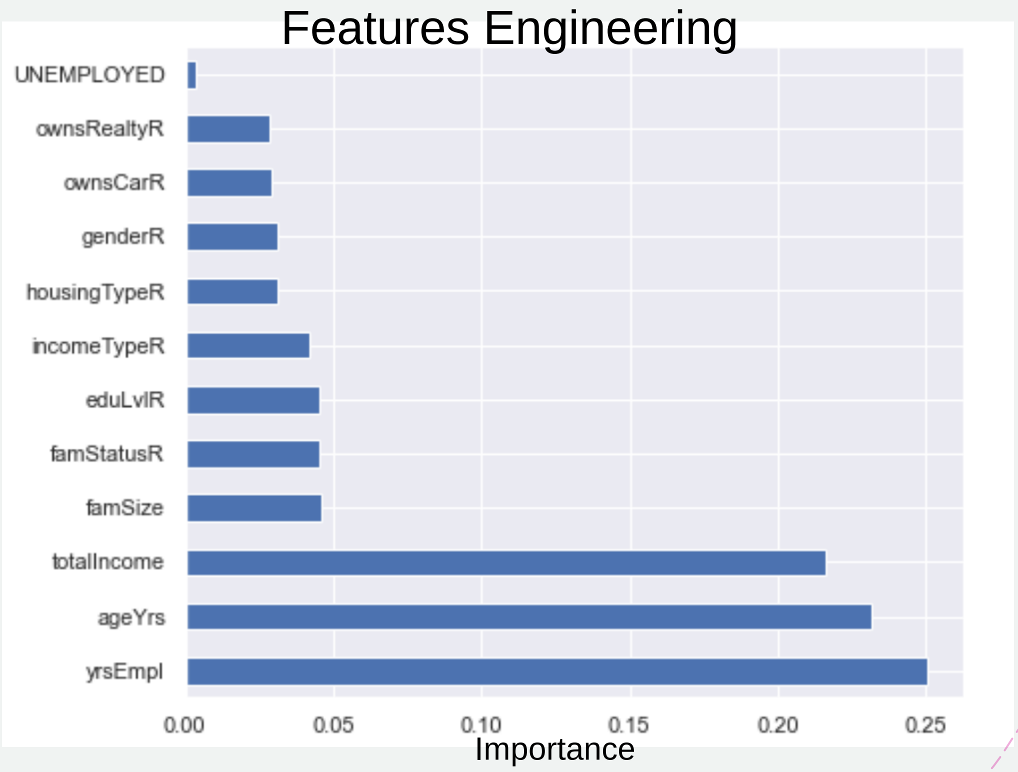
What about income disparity between males and females? Is there a difference?

- An independent t-Test revealed that there was a significant difference in income between males and females in this dataset.
- Test the theory of males dominating in “high paying” careers:



Results –
Modeling

+ Supervised Machine Learning Modeling
    + The initial model predicted credit card approval with 32% accuracy
    + By hypertuning the model, the overall accuracy increased to 39%
    + The features of most importance were determined to be: **age, income**, and **years employed**
    + Upon training a model with the factors of age, income, and years employed, the accuracy of approval prediction did not improve.



Summary

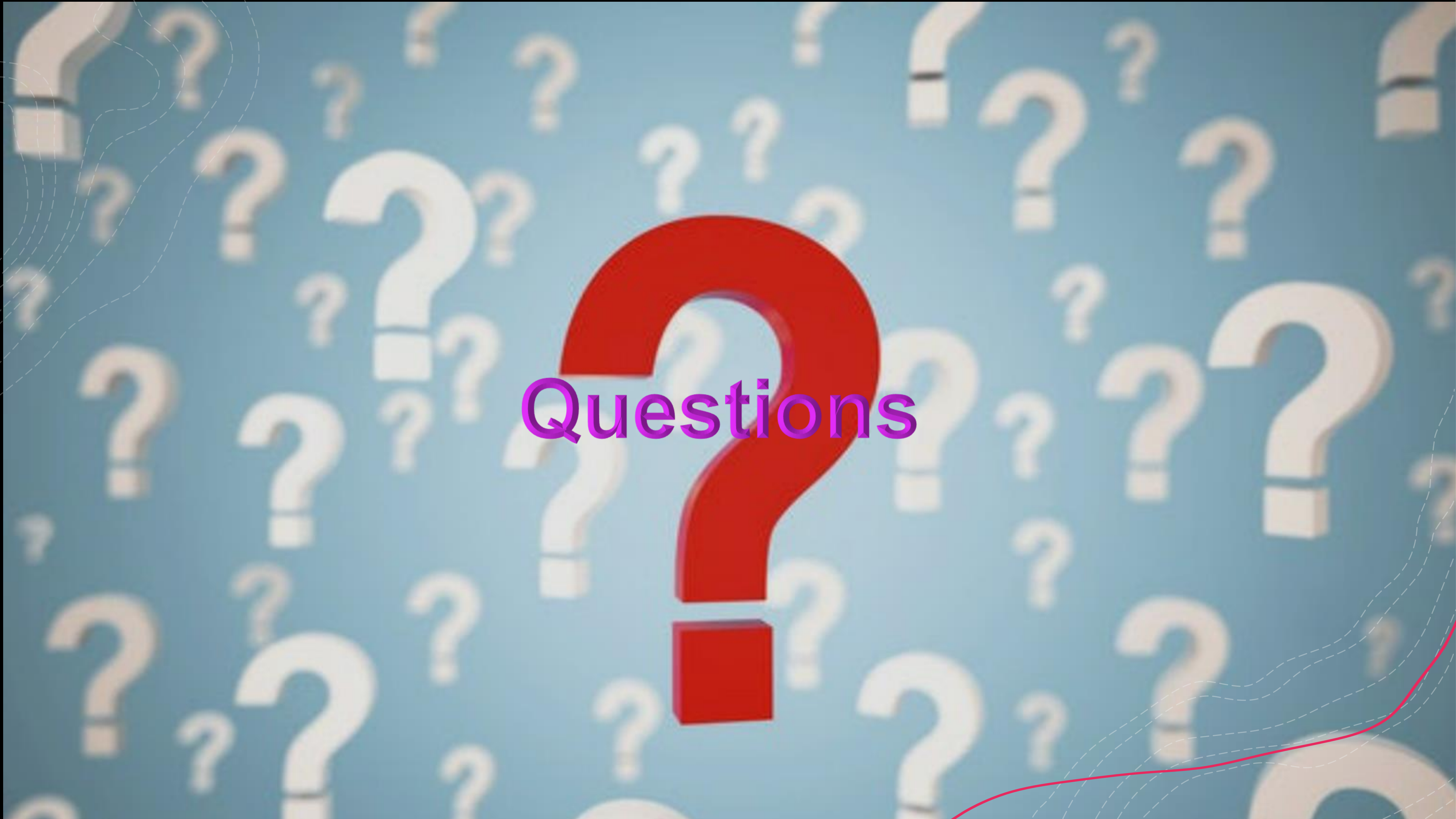
+ Failed to find a model that predicted approval status with an accuracy higher than 39%
+ Independent Chi-squares aided in finding out that gender did not affect approval rate
+ An Independent t-Test revealed that there was a significant “pay gap” between Males and Females - even within the same job categories

Conclusions

- The challenge for credit card issuers is finding the best method to assess an applicant in a way that decreases their financial risk.
- Significant losses in the financial industry can lead to catastrophic economic consequences as seen during the 2008 financial crisis.
- Age, income, and years employment were the factors of most importance for this dataset, which coincides with research of credit card use mentioned at the start of this presentation.

- Limitations Encountered
  - Dataset did not include important factors that card issuers usually look for, such as credit score or debt utilization rate.

- Surprising Findings
  - The data used in this project came from a snapshot of over 13,000 actual client accounts and it was found that not only do men make more money on average than women, but they consistently make more money even within the same job titles and industries.


Questions

Citations

- Kaggle.com
- Wallpaperaccess.com
- Thestreet.com
- Creditkarma.com
- Creditcards.com
- Ibm.com
- Cio.com
- Medium.com
- Fool.com
- Forbes.com