Data Science For All Presentation

**Introduction**

**Thank you all for joining us today[NS]**

Background – Teneika - 30 sec

Women-owned businesses represent **42% of all businesses, employ 9.4 million workers and generate 1.9 trillion in revenue**. Yet while growing 2X faster than all businesses nationwide, we only receive 2.8% of venture capital (VC) funding according to industry reports.

Our goal is to help women entrepreneurs understand the VC landscape by producing a model to predict success in reaching a VC, while also investigating gender bias, and providing recommendations based on the model.**[NS]**

Qualitative Surveys Intro – Tina - 60 sec

As we began this project, we realized there were a lot of factors that could not necessarily be quantified, so we decided to split our approach into two directions: one being qualitative surveys, and one being the quantitative analysis of the datasets we had. First, we will discuss our qualitative surveys.

**Qualitative Surveys:**

We interviewed 5 Investors and 3 founders of different genders to discuss their experience with funding and biases that exist. We found these surveys as a whole to be meaningful because, **during this effort, all participants were extremely engaged and intrigued by our questions and what the data showed for the industry and their firms as well.**

Qualitative Surveys – Shea

After conducting our surveys, we took time to reflect on a few key points:

How much do qualitative attributes matter?

Again, in our conversations, our interviewees described how women VCs are more aware when they are speaking to Women Founders, so they are more invested. If there are not people in positions that look like you or think like you, there is a lack of ability to connect to their background, starting point, etc…. It becomes subjective vs objective. VCs are also looking at your network as well as the success of your previous companies and startups, so if you never had the opportunity to succeed or form close relationships in the first place, it makes your future ventures also less likely to succeed

Another question we looked at was: Does having a woman on the VC board affect the company’s estimated revenue?

According to our conversations, we learned that there are not enough women venture partners. This is a problem because if a product caters towards women, perhaps in the beauty or lifestyle industries, such as Spanx, women on the VC board can understand its importance and market potential, while men may not. **[NS]**

Quantitative Analysis Intro – Tina - 60 sec

**Target Dataset**: Now for the quantitative analysis, we graciously received access to Crunchbase from a mentor. We extracted data into CSV files and hosted it on a Shared Google Drive to provide continuity for everyone. The three datasets that we used were:

1. Sample Crunchbase Companies (110K records)

2. Crunchbase Investors (57.3K records)

3. Crunchbase 2019 Funding Rounds (15.3K records)

Based on this data, we decided the scope of our project would be the United States in 2019.

Cleaning Effort: In order to begin our analysis, we did an initial cleaning and filtering of our data in Python and Tableau. We created Boolean fields for Women Led/Founded companies so that we could easily filter based on whether a company was women-led or not. We also changed the column “Estimated Revenue Range” to Estimated Revenue >= 1M so the column would give a binary result. Finally, we removed fields that were unrelated to our main focus, such as Company Description etc.

Data Modeling – Tina

Through surveying our data, we found that a Many to Many relationship was needed to successfully evaluate all factors, so we kept this in mind as we continued our analysis.**[NS]**

BIAS in VC FUNDING – Teneika

50.8% of the US population are women and 42% of business owners are women. Yet women only make up 11 percent of investment partners (VCs) in the US and are only included in 17% of VC deals.

In our sample we found that 7.7% of funding went to women owned or led businesses. This is higher than the industry average of 2.8% because we’ve included women-led businesses in our analysis. **[NS]**

What we also discovered was that there was a disproportionately low number of deals received by women in comparison to men in states like Minnesota, Washington, Texas, Virginia and Illinois. Even when women estimate their revenues higher, they continue to receive only a small part of the investment. Texas and Washington for example have a large disparity in the number of deals received by women vs men, at almost 1000% for women who’ve raised less than $500M. **[NS]**

Women Led companies also report 18-69% higher profits in the same industries as compared to men. **[NS]**

Because VCs overlook the opportunity to invest in women they are missing out on potential profits. Women generate more revenue while receiving less than half the funding of men. According to a report by Morgan Stanley, the U.S. Census Bureau’s Survey of Business Owners and the Bureau of Labor Statistics shows had the number of women and minority-owned businesses and the portion of revenues matched their percentage in the labor force–56%–gross receipts would have increased to $6.8 trillion, suggesting a missed opportunity of up to $3 trillion. **[NS]**

KNN IMPUTATION OF FEATURES AND SMOTE – Hima

To prepare our data for modeling, we filled the Null values by implementing KNN (K Nearest Neighbour) imputation.**[NS]**

Our classification model outputs 3 classes ‘low’, ‘moderate’ and ‘high’ each corresponding to the chance for a woman owned business to get through to the next round of funding.

We faced the imbalance Class Problem here: There were many 'low' and 'moderate' classes in the dataset. So we used Synthetic Minority Oversampling Technique, or SMOTE to upsample minority - 'high' class before we could go ahead with implementing the model.

CLASSIFICATION MODEL (ML) – Shermeen

We decided to use a decision tree classification model for better interpretability of features. The input features we chose for the model are:

* Estimated Revenue Range
* Number of Investments
* Number of Exits (IPO)

Our output classes are ‘low’, ‘moderate’ and ‘high’ chances of getting to the next round of funding, which we inferred from the funding column of our dataset. Random Forest classifier gave us our best accuracy. It constructs a multitude of decision trees at training time and outputs the mode of the classifications.

Based on our conversations with various VC (Venture Capital) investors and founders, these features seemed to be important to predict the chance of a business getting to proceed to the next round of funding. Our results from the model solidify and accept this hypothesis.

RECOMMENDATIONS - Lauren

After this analysis, we came up with recommendations for women-owned businesses to help them improve their chances of getting venture capital funding.

Low:

From our graphs, since we noticed that there were states where there was a disproportionately low number of deals received by women in comparison to men, we would suggest women-owned businesses to have their company headquarters in other states where the ratio is higher, such as California, New Jersey, and New York. Some states like New Jersey also have startup incubator or accelerator programs. There are also ecosystem building programs to build networks and traction (Female Founders Collective, DivInc, etc.). This would increase the probability that the startup receives a higher amount of funding.

Moderate:

According to the order of decision criteria produced by the model, it is vitally important for women entrepreneurs to confidently and positively estimate their revenues instead of underestimating the expected revenue, especially when they pitch their startups before VC investors.

High:

Analysis of the number of investments to women owned businesses showed that women-owned businesses with IPOs or Investments raise more funding and have more exits and have a high probability of receiving additional funding. We would recommend after passing the series A funding rounds to pitch in the fourth quarter to generate a higher amount of capital. If you have not generated enough investments or exit opportunities.

For venture capitalists, we recommend diversifying the venture partners by adding more women so that they can provide their unique perspectives in funding new startups.

CONCLUSION -

From our analysis, we were able to conclude that the gender disparity in funding does exist, but there are ways for women-owned businesses to increase their VC funding.

In order to improve our work, we would use more data and more qualitative data because there is a huge deficit in the availability of data. There was a significant amount of null values in numerous fields.

As for next steps, our model currently only categorizes the funding that a female founder would receive as low, moderate, or high. We plan to improve the model so that we can quantify the estimated funding amount that women-owned businesses will receive. That way, we can provide more targeted recommendations for the businesses.

Demo - Teneika

California is where I would want to incorporate but the tool would inform me on various decisions, like I should specifically headquarter in the Bay Area not Silicon Valley because I have higher opportunities to resolve the recommendations provided in the tool. **[NS]**

For example, Silicon Valley received almost 300% less deals than the Bay Area than Silicon Valley. Silicon Valley also has relatively low amounts of Firms and VC Partners investing in Women in comparison to those headquartered in the Bay Area. I also learned that Silicon Valley companies report higher revenue across the top industries for women but the highest investment deals occur in Q4 but in the Bay Area the deals are widely distributed.

Women owned businesses would leverage our tool to evaluate where they should incorporate their business, what VCs or accelerator programs they can join to build a network, raise capital and when the best times to approach VCs based on funding rounds.