Business Analysis

Data Science Communications Intern
Tung-Ting(Nelson) Wang
tungtingwang15@gmail.com

OVERVIEW

- **Introduction**

- **Data Wrangling**

- **Data Analysis pt.1 : Device Usage**

- **Data Analysis pt.2 : Potential Valuable Customers**

- **Data Analysis pt.2 (contd.): Supervised Learning**

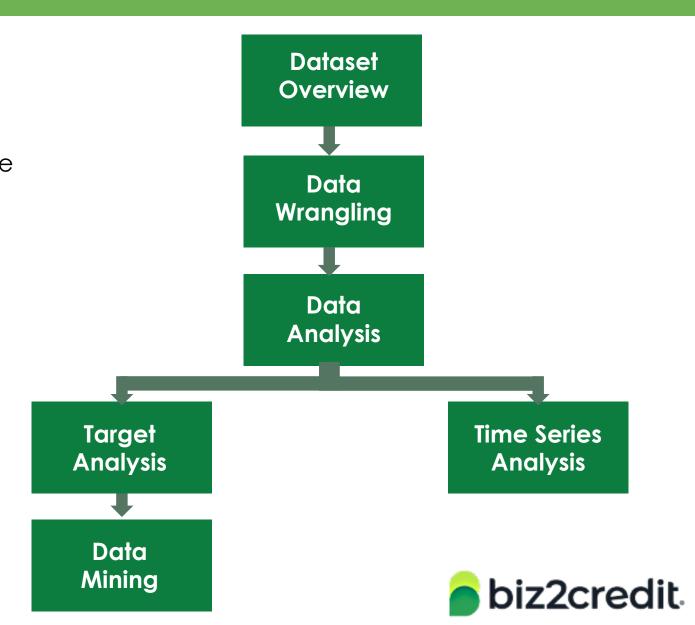- **Data Analysis pt.2 (contd.): Prediction**

biz2credit

Introduction

To find out how devices usage interacts with the market and what kind of companies might be the potential valuable customers for Biz2Credit, I utilized data analytic tools to achieve the goal. The analysis is separated into two parts. (Completed code has been uploaded on Github or GoogleDrive.)

Dataset Overview

Data Wrangling
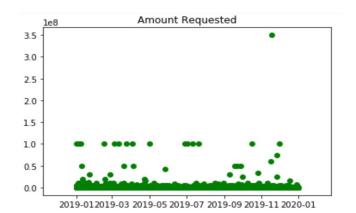
Data Analysis

Target Analysis

Data Mining

Time Series Analysis

Data Wrangling

Null Value / Outlier

Null Value / Outlier
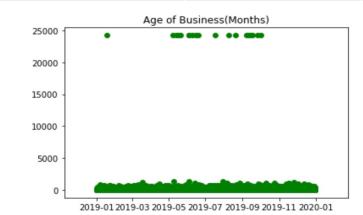
📎 Null Value

Attributes	Quantity	Percentage	Treatment
Industry	21	<<1%	Assign a new category, "Unknown"
Credit Score	5170	12%	Replaced with mean
Annual Revenue	5732	13%	Replaced with mean

```
Industry              41551 non-null object
Personal CreditScore  41567 non-null float64
```
＋
Credit Score of the application. "0" means data is missing

Annual Revenue . "0" means data is missing

📎 Outlier

Attributes	Quantity	Percentage	Treatment
Amount Requested	1	<<1%	Drop the row
Age of Business (month)	19	<<1%	Drop the rows

biz2credit

Data Analysis pt.1

Device Usage

- **Which device contributed to the max number of applications in 12 months period?**

- **How did the application distribution change over time?**

- **Did "application channel" have any impact on the trend of device usage?**

- **Which Industry contributed the most in 2019?**

biz2credit

- **Which device contributed to the max number of applications in 12 months period?**

Device

After grouping up the device and counting the applications, we can easily find out that "Phone" has the maximum number among all.

Device	Total number of application
Phone	19,235
Computer	16,591
Unspecified	4,850
Tablet	727
Mobile App	149

biz2credit

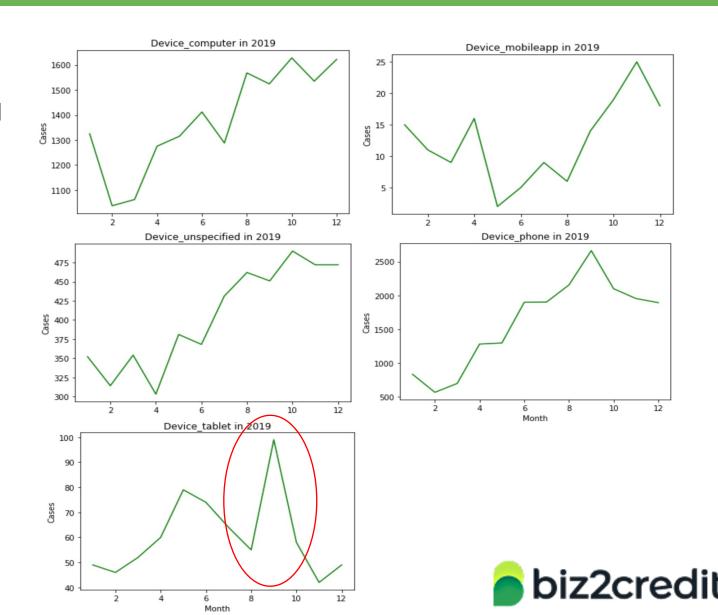- **How did the application distribution change over time?**

Trend

Only "tablet" has a dramatic increase and decrease in applications between August and October.
 In my point of view, it is probably caused by the launch of the iPad 7 on September 25 in 2019, which suddenly reminded the customers to try their iPad to see if it functions as usual.
 After that, they switched back to the device they are more comfortable with, making the usage drop dramatically.
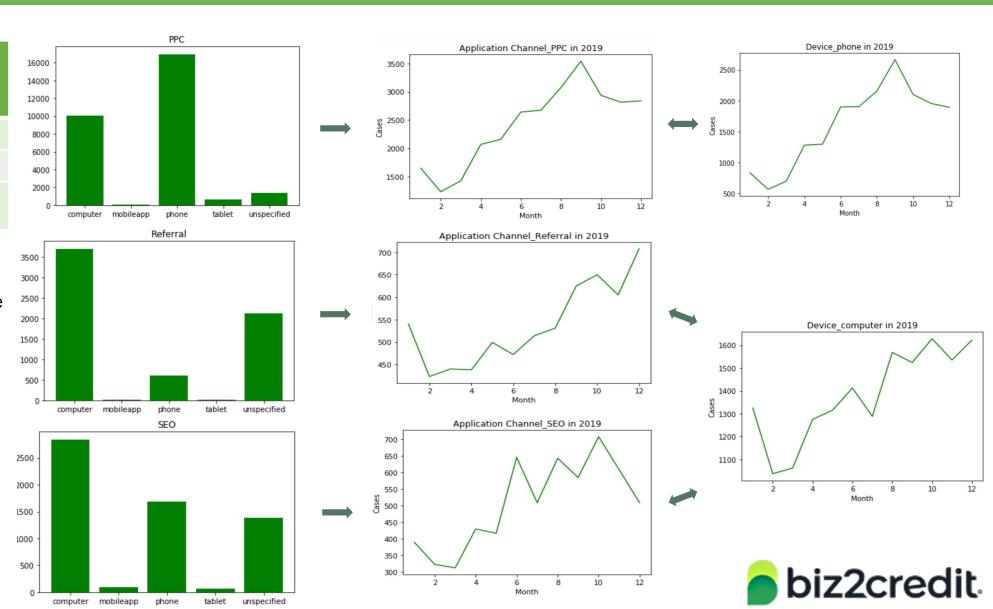
- **Did "application channel" have any impact on the trend of device usage?**

Channel	Most used device (channel)
PPC	Phone
Referral	Computer
SEO	Computer

Two ==similar track== of trend line implies that "application channel" certainly has impacts on device usage.

Which Industry contributed the most in 2019?

Industry

The industry that has the maximum number of applications contributed the most in 2019. Although "Other industry" has the maximum number of applications and "Retail" is the second-highest, "Other Industry" is consisted of many other industries that cannot be categorized. Since "Other Industries" might contain a hundred industries, it is not fair to compare it with a single industry. Thus, the "Retail" industry contributed the most in 2019.

Industry (top 10)	The number of applications
Other Services (except Public Administration)	7,823
Retail Trade	6,640
Accommodation and Food Services'	5,175
Construction	4,505
Transportation and Warehousing	3,248
Arts, Entertainment, and Recreation	2,232
Health Care and Social Assistance	2,064
Professional, Scientific, and Technical Services	1,825
Real Estate and Rental and Leasing	1,527
Wholesale Trade	1,256

biz2credit.

Data Analysis pt.2

- **Potential Valuable Customers**

  - What's the relationship between attributes?

  - How much amount of annual revenue can make the company a valuable customer?

  - What kind of customers can be potential valuable customers to Biz2Credit?

- **Supervised Learning**
  (Decision Tree)

- **Prediction**

biz2credit

- **What's the relationship between attributes?**

Correlation

In this diagram, we can easily find out that "Amount Requested" has the weakest relationship with any other attributes. Moreover, "Self-Directed" and "Annual Revenue" have the strongest relationship with others, which is represented in dark green or dark red.

Weakest

	Amount Requested	Age of Business(Months)	Personal CreditScore	Customer Type_b	Application Channel_b	Industry_b	Census Region_b	Business Legal Structure_b	Device_b	Self Directed_b	Self Submit_b	Annual Revenue
Amount Requested	1											
Age of Business(Months)	0.028583526	1										
Personal CreditScore	0.025107064	0.208531453	1									
Customer Type_b	-0.006612222	0.185976447	0.180434868	1								
Application Channel_b	0.008453873	0.235407322	0.262127942	0.401354721	1							
Industry_b	0.010952414	0.020171914	-0.014215115	-0.138683927	-0.084509787	1						
Census Region_b	0.001997256	0.021111242	0.009303319	0.053672263	0.064105333	-0.030201835	1					
Business Legal Structure_b	0.007035584	0.13728605	0.292347716	0.246656108	0.25602935	0.011491738	0.02657836	1				
Device_b	0.017283957	0.019467168	0.105684034	-0.493184898	-0.048992612	0.080841111	-0.035162101	0.040356128	1			
Self Directed_b	-0.001993327	-0.260836071	-0.310038565	-0.689752771	-0.584398044	0.1150753	-0.055854	-0.332830221	0.188614413	1		
Self Submit_b	-0.007166536	0.023485271	0.103096631	-0.120871813	-0.038198551	0.0563634	-0.029044479	0.074221538	0.153896619	0.190381328	1	
Annual Revenue	0.0769529 13	0.232033462	0.3821663	0.297973358	0.39415659	-0.037868899	0.023870367	0.292089862	0.005908969	-0.459697514	-0.019804208	1

: Positively strong
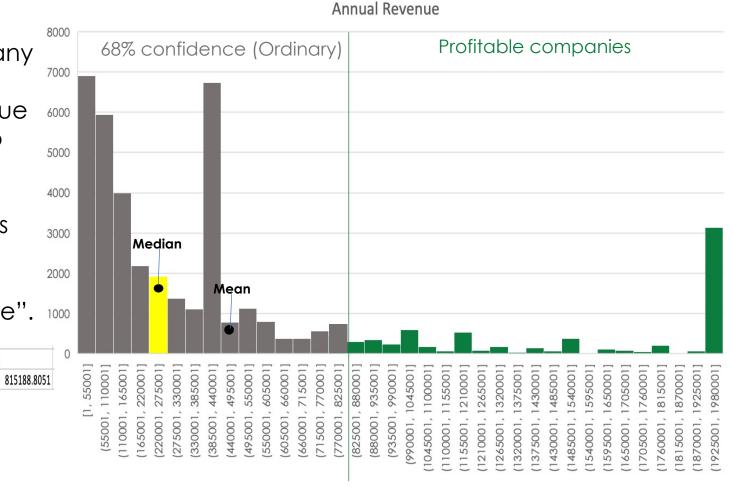
: Negatively strong

biz2credit

How much amount of annual revenue can make the company a valuable customer?

Target

In my point of view, the bigger the company is the larger amount of loan it needs. Thus, the customer that has more annual revenue than most of the companies is valuable to Biz2Credit. Since the distribution is skewed severely, I used "Median" to calculate, instead of "Mean". Any company that has annual revenue beyond 68% confidence, which means the annual revenue is more than $815,188, will be defined as "Valuable".

Annual Revenue

68% confidence (Ordinary) | Profitable companies

Annual Revenue	
Mean	459279.7672
Standard Error	2690.085498
Median	266833.5
Mode	403569.1412
Standard Deviation	548355.3051
Sample Variance	3.00694E+11
Kurtosis	1.94960058
Skewness	1.727486475
Range	1949999
Minimum	1
Maximum	1950000
Sum	19083992888
Count	41552

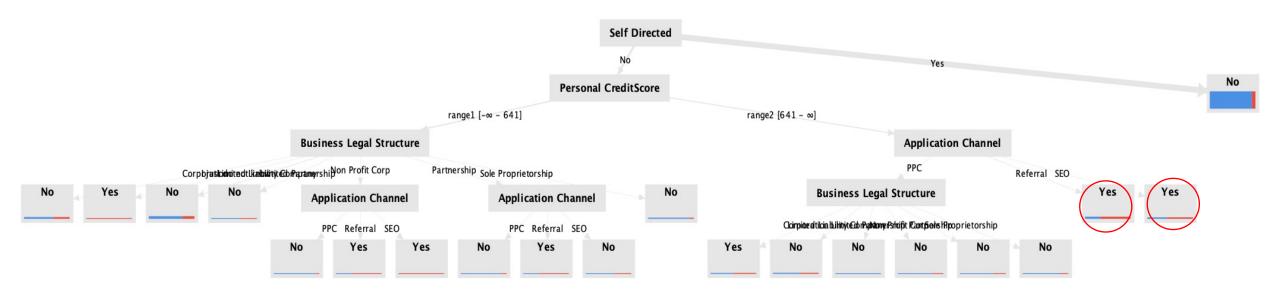|	From:	To:
68% 1 standard deviation:	0	815188.8051

biz2credit

- **What kind of customers can be potential valuable customers to Biz2Credit?**

Decision Tree

In many routes, only seven routes have "Yes", which means they have earned more annual revenue than $815,188. However, there are only two routes that have a significant number of "Yes", other "Yes" only contain less than 10 items. The only difference between the two routes is "Application Channel". One of our target audiences is through " Referral", which has 1186 out of 1798 "valuable" companies. The other one is through "SEO", which has 516 out of 894 "valuable" companies.

Conclusion

In my point of view, those applicants that have the potential to become a big company are too busy with the bigger deal. They are less likely to engage in the earlier tier of the application process. Plus, financially stable companies that deal with banks frequently will make the applicants' credit scores higher. At last, I believe those outstanding companies would not select their partner only through advertisement. Surely, they tend to have more connections to ensure their partner is trustworthy and profitable. Therefore, applying for the loan through "referral" is the only best way. As to "SEO", companies might know Biz2Credit from other business places or from their employees. With the lack of connection to Biz2Credit, the only way in three of the "application channel" is through "SEO", which is searching Biz2credit online. The company that has the criteria, which are "not self-directed", "higher credit score", "through referral" or "SEO", can be a potential valuable customer for Biz2Credit.

Independent Variable(Hierarchy)	Criteria
Self-Directed	No
Personal Credit Score	> 641
Application Channel	Referral, SEO

Self Directed

No — Yes

No

Personal CreditScore

range2 [641 – ∞]

Application Channel

PPC

Referral — SEO

Yes

Yes

Yes
Distribution: 612*No*, 1186*Yes*
Number of items: 1798
Ratio of total: 6.18%

Yes
Distribution: 378*No*, 516*Yes*
Number of items: 894
Ratio of total: 3.07%

biz2credit.