The data-cleaning methods applied on this dataset are limited to the personal project, in real-time I would communicate with the business owners for understanding the data and then decide the required changes. Finally, before changing the dataset, I would ask my manager to proceed further.

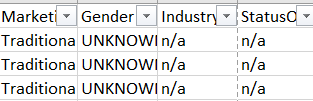
CC = Current rows after executing the cleaning step.

Initially total rows in the dataset are 114k

CC-114326

**Step:1**

Removed entire columns of industry, Gender and Status of employment.



CC-114326

**Step:2**

As I’m analyzing the data with approved loans only, I’ve excluded the **30k** records.

CC-83506

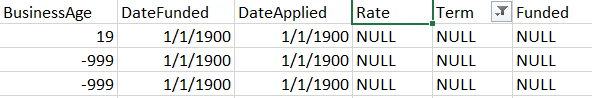
**Step:3**

As I’m trying to analyze data to make the conclusions related to the Revenue tied with the marketing campaign (Digital/Traditional) and Risk rate. My necessity lies in these three variables (**Rate**, **Funded amount** & **Term**).

There were the rows, where I’ve found the NULL values for the necessary attributes **Rate**| **Term**| **Funded Amount**, One more reason to remove this data is that the year this loan was funded in was **1900**, which is completely unhelpful for analyzing this dataset.

Excluded the rows which are not having the NULL values in Funded, Term and Rate. Removed **34.5k** rows.

CC-49015



**Blank Rows Clean-Up**

* Removed the blank rows generated because of applying **STEPs [2 & 3].**

It took more than 20 minutes to remove the blank lines.

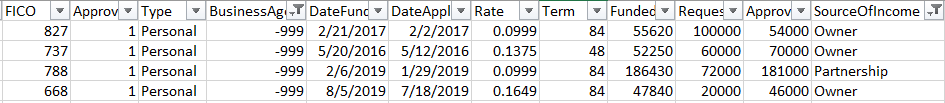
Optimal Solution: I would make an ETL pipeline using Python, so that data just gets in with all applied filters.

**Step: 4**

Basically, **8k** records are having business age registered as ‘-999’. As, the number is so high and all other attributes are non-negligible.

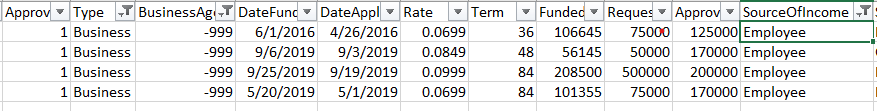
Instead if removing the fields, I’d considered the business age as ‘0’ changed it from ‘-999’. Below are various reasons for the changes made related to the Business Age ‘-999’. Total changes made to, 8671 rows.

By observing source of Income and Type of loan approved. I’ve **assumed** that, despite having employment Owner and Partner, specifically these clients are requesting for the personal loan and they don’t want to disclose their business Age or any other factor. So, we should consider their business age as 0 or may be in real-time we could keep it as ‘Do not want to disclose’[DNWD].

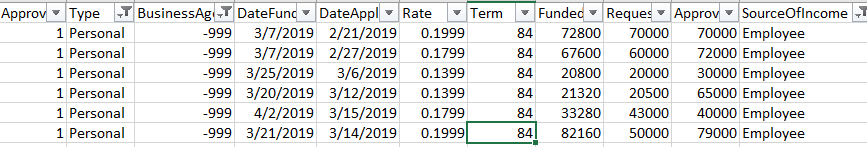


Other 4 which has requested the loan for the Business type. They’re employees and they are new in the business. May be just doing a startup. So, currently they’re employees and they have requested the business loan. Simply we can consider their business age ‘0’.

[**NOTE:** In real-time, for more accurate data modification I would refer to the other data. These are just the assumptions.]



Rest of the entries with business age ‘-999’, are employees and they have requested the personal loan. So we can assume that they’re not related with any kind of business. And changing all the entries for Business Age ‘-999’ with ‘0’.



CC-49015

**Step: 5**

Added the revenue column, using the interest rate percentage.

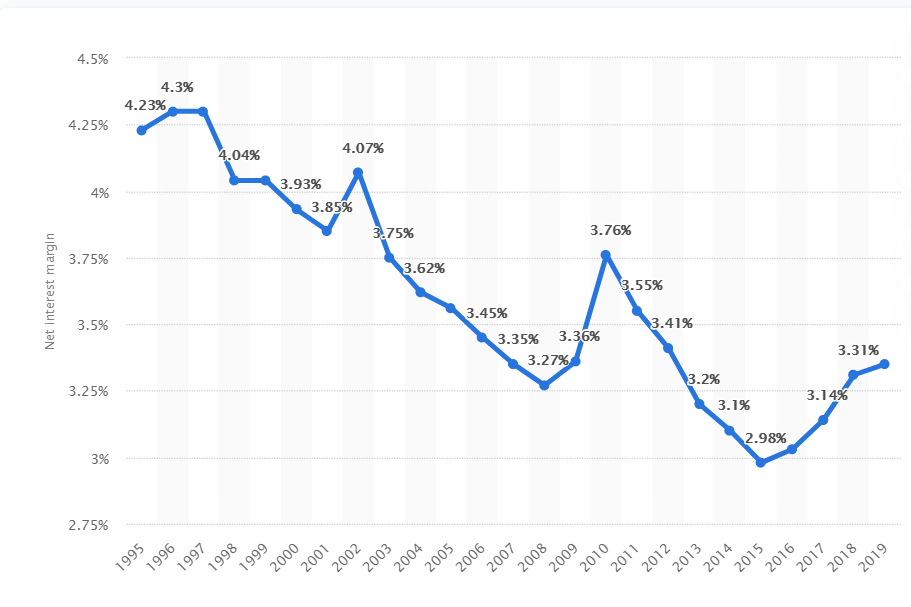
Formula: Monthly Revenue =( Funded amount/Term)\*Rate

**Step: 6**

**NIM: Net Interest Margin.**

Average-NIM-Rate varies with the year. As the Dataset is anonymous, I didn’t have the exact NIM-rate. So, I’ve calculated the margin with average NIM for each of the years.

**NIM** Formula: Monthly NIM = Funded Amount\*(Average-NIM-Rate)



<https://www.investopedia.com/ask/answers/061715/what-net-interest-margin-typical-bank.asp>

2014: 3.1% 0.031

2015: 2.98 % 0.0298

2016: 3.03% 0.0303

2017: 3.14% 0.0314

2018: 3.31% 0.0331

2019: 3.35% 0.0335

2020: 2.72% 0.0272

**Step: 7**

**NIM: Percentage.**

Added up the gross margin percentage, using the generated NIM column by [Step: 6].

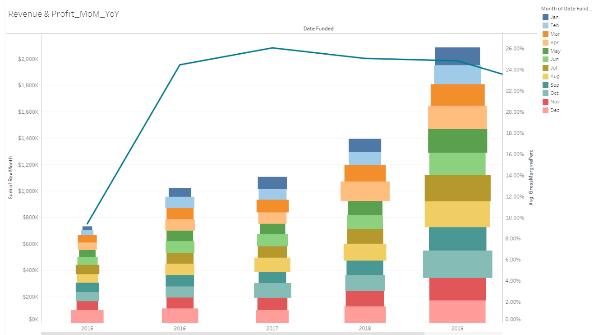
Formula: Percentage of Monthly NIM = (NMI\*100/Revenue)

Analysis: **Marketing** **by Profit** **&** **Revenue**

For further analysis, I’ve used Tableau. To follow along the with summary. Please use the **link attached with each worksheet,** it will redirect you to the corresponding graph**. Please use the Full Screen mode for thorough visuals**.

**Please Use the Legend on the side bar.**

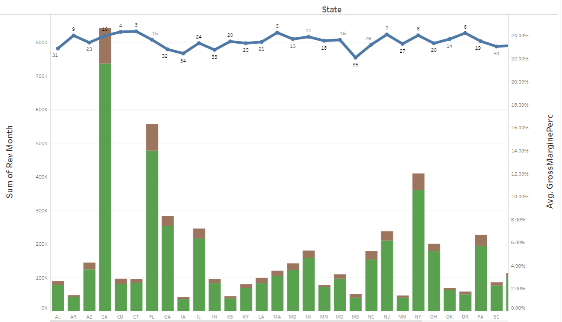
1. **Revenue & Profit\_MoM\_YoY** - <https://tabsoft.co/3jgbOvl>

* This graph is showing the total revenue and Profit in each year.
* Size of each pile is the month showing the total revenue.
* Size of the pile is been determined by the sum of revenue.
* Color of the pile represents the Months.
* The Line represents the Profit percentage based on NIM for each year.
* Height of the Bar represents the total revenue in each year.

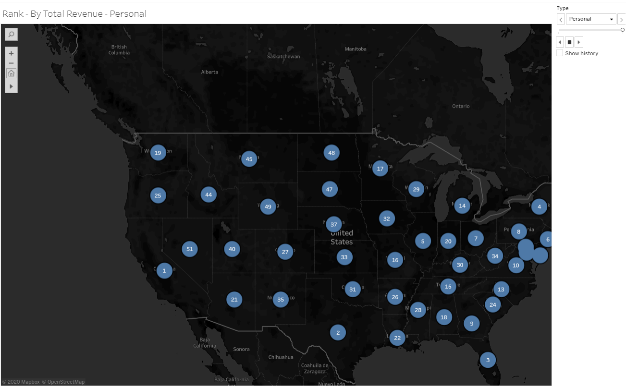
Observation: While studying the revenue and profit for year 2018 and 2019, we have results.

Total Revenue for year 2019 was way higher than the year 2018, on the other side profit made in Year 2019 was 24.82% while in 2018 it was 25.05%. Despite having high higher revenue in 2019, company had made less profit (%) than year 2018.

1. **Revenue & Profit(%)\_Marketing Type & State** - <https://tabsoft.co/3cJlJqN>

* Similar to 1st graph, this graph also represents the Total revenue and profit (%) for different states.
* The color of the Bars represent the revenue generated from the type of marketing, [Green: Traditional | Brown: Digital]
* Line represents the overall Profit (%) made in each state.
* Numbers on the Profit line represents the Rank of each state based on the Profit made.

Observation: If we look at the Bars of **California, Florida, Texas & New York**, these are the states having total revenue more than 450K $. But still their Profit-Ranks are **10, 15, 8 & 17**. And top 5 states based on Profit-rankings are **Washington, Utah, Connecticut, Colorado & Massachusetts** [1 to 5]. Despite having revenue, even less than 150k $ these states make highest profit. We can find the reason behind the success and, make according changes in the states which are not able to make good profits.

1. **Rank - By Total Revenue – Personal -** <https://tabsoft.co/3cF4C9R> 

* Here is the Geographical graph which shows the Total Revenue generated from the Type of Loan [**Business** and **Personal**].
* To observe both the type of rankings, please use the side legend as mentioned.

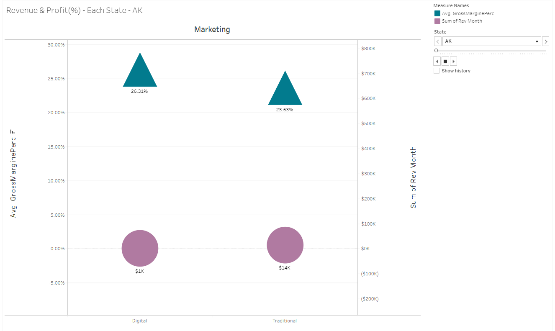
Observation: By observing the revenue in each state, we can identify that, in which state or states we should improve in our marketing campaigns.

**Please Use the Legend.**

**Loan Type & State**

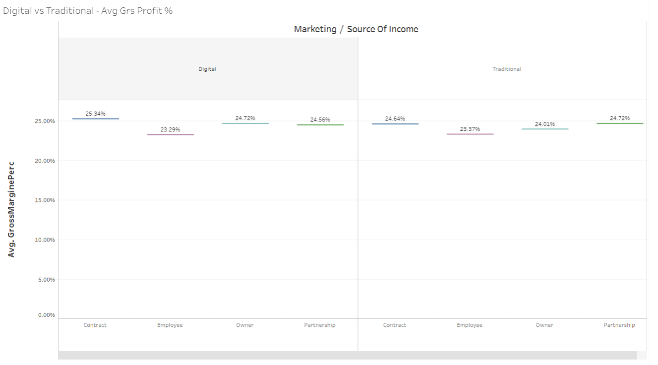
More specifically, by observing the revenue generated by the type of loan, we can decide that what type of loan should be marketed in what state, for filling the deficiency in particular type of market.

1. **Revenue & Profit(%) - Each State – AK** - <https://tabsoft.co/3jdsSlW>

This graph is been made to individually analyze the Revenue and Profit (%) for each state.

Observation: After looking at bigger picture from the previous Geographical graph, now our interest would be in analyzing the situation thoroughly. And here we can examine the Total Revenue and Profit generated by type of marketing. Which gives us the strong foundation to make the required changes in our marketing campaigns.

1. **Digital vs Traditional - Avg Grs Profit %** : <https://tabsoft.co/349lQYT>

This graph is to inspect the difference between the Traditional and Digital marketing.

* We will consider the Average Gross Margin and Designation.
* Gantt bars are representing the average profit we have gained from the people from diverse designations.

Observation: If we compare the Profit-gained by Designations, in Digital [Left] & Traditional [Right]. We can observe that they are pretty close to each other.

So, type of marketing does not matter.

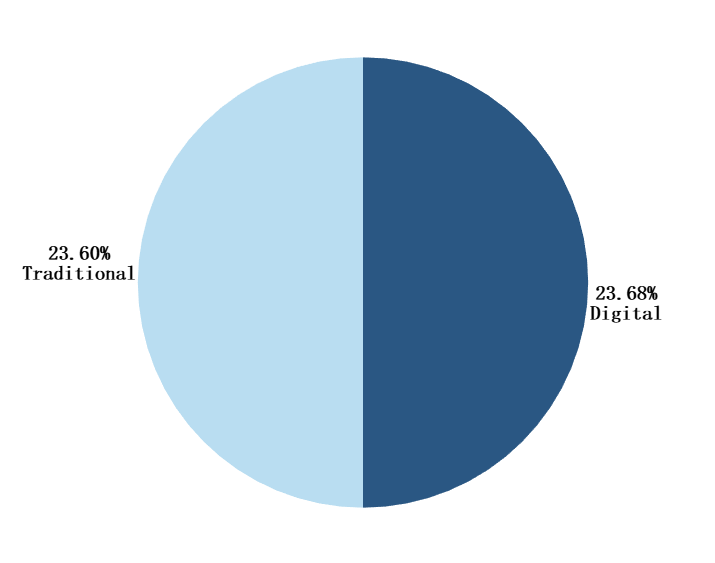
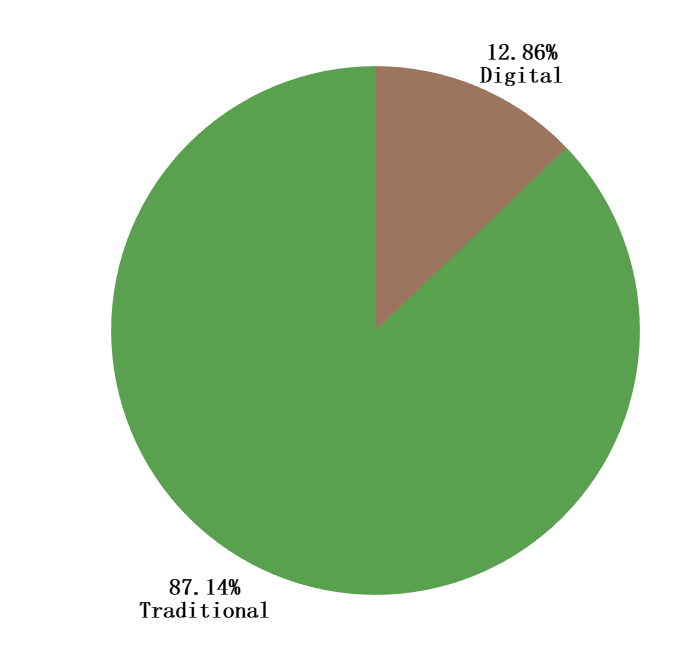
**Conclusion**

If we look at the pie chart, the average profits we have made from Digital marketing and avg-profit made from Traditionally marketing are similar, in fact average profit (%) by Digital marketing is slightly higher than Traditional.

But Conversion rate of Digital Marketing is very low 12.86%. Which indicates that we should improve the digital process for reaching out the clients. Through out the Analysis, we can consider our KPI to be:

**KPI**: 1. Quality of Digital marketing. [Should Increase]

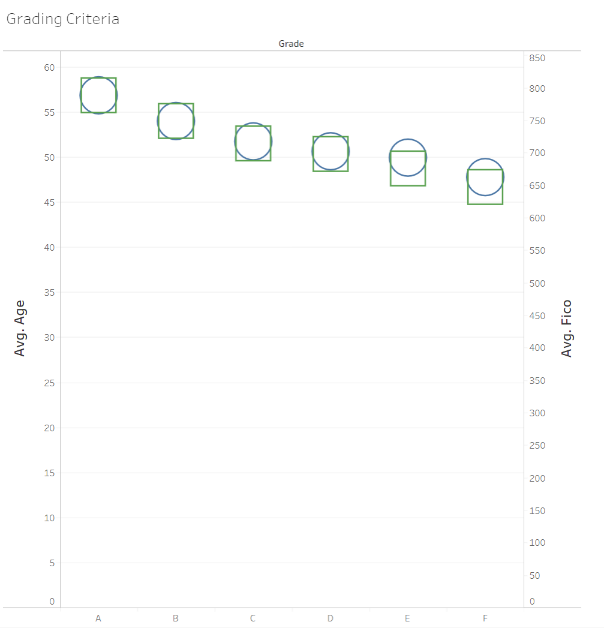
2. Required efforts in the Approval Processes. [Should Decrease]



Profit by Marketing

Conversion Rate

Analysis: **Risk Factor**



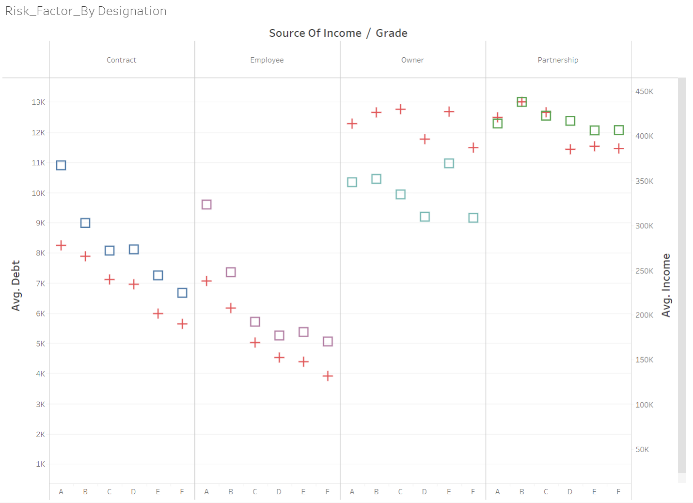
We’re just analyzing that how were the internal grades [A-F] were assigned.

1. **Grading Criteria**: <https://tabsoft.co/33ecjkf>

* This graph clearly explains that how the grade had been decided for each client.

Observation: The Age (states maturity) and the FICO (credit score), both together determines the Grade from **A** to **F**.

1. **Risk\_Factor\_By Designation**: <https://tabsoft.co/3kZiyyh>



* Here the Red plus (+) sign stands for the Average Debt.
* Square boxes stand for Average Income.
* We have also included the third factor, Source of Income [Designation]
* These three factors significantly contribute into deciding the Risk level of the client.

Observation: By observing the graph, we can conclude that

Client who works as an Employee or a Contract, is having very low debt. It displays the low risk.

On the other end, if client is an Owner or a Partner, then s/he is having more Debt. Despite having higher income, these type of clients have higher Debt amount.

**Conclusion**

Our consideration for evaluating the risk of our client should not be dependent only on the Grade, Fico and Age. We should also consider client’s Income, Designation, and the Debt. All the factors are contributing significant role.

Yes, Partners and Owners are having higher Debts but they also have higher income, which can increase the revenue exponentially and so the Profit as well.

Evaluating client’s risk may require various other factors such as person’s Tax-returns, Expenses, etc. As I had limited attributes to work with, I have tried my best to identify the facts.