Module 5 Assignment — Sponsor Project

Pei-Yu Jheng, Shreeya Ambre, Sean Mclean, Theodore Smith

MPS Analytics, Northeastern University

ALY6980: Capstone

Professor Ken Parker

Feb 9th, 2025

Overview/Introduction

Big Sky Franchise Team manages 9,000 customer accounts with a small team of 15 employees, making it challenging to track account status, engage potential clients, and drive franchise growth. To address this, we will use R and Python to merge, clean, and analyze datasets, ensuring data integrity and usability. Our proposal includes building a management system using Streamlit, providing a user-friendly dashboard for tracking customer interactions and optimizing sales conversions. Through data-driven insights, we aim to streamline operations, improve client engagement, and help Big Sky focus on high-potential opportunities for growth.

Business question/Our proposal

After looking at the data and background research, we found that managing customer accounts is an opportunity for us to improve Big Sky's internal process. They have more than 9,000 accounts but have only 15 employees. It is difficult for them to keep track of all the accounts' status whether contacting potential customers, scheduling a call, or pushing franchise plan further for their customers. To avoid delay or losing the opportunities to grow their business, we want to help the Big Sky Franchise Team to manage their customer accounts and build a timeline for each account. Therefore, we planned to build an accounts management system by using Streamlit.

Big Sky needs to be forward-focused, but that does not mean they can forget about the data they previously collected. If the leads are dead, annotate them as such. Just because you have 5,000 email addresses on record does not mean that you need to spend time focusing on contacts that have no potential as a customer.

What tools and data we will use

- For dataset cleaning and research, R and Python will be utilized between the team members.
- For visual presentation, an interactive dashboard will be presented that is user friendly and effective for assessing and editing.
  Datasets provided from Big Sky and outside datasets for comparing industry standards.

Data wrangling (what we have done so far)

Cleaning processes:

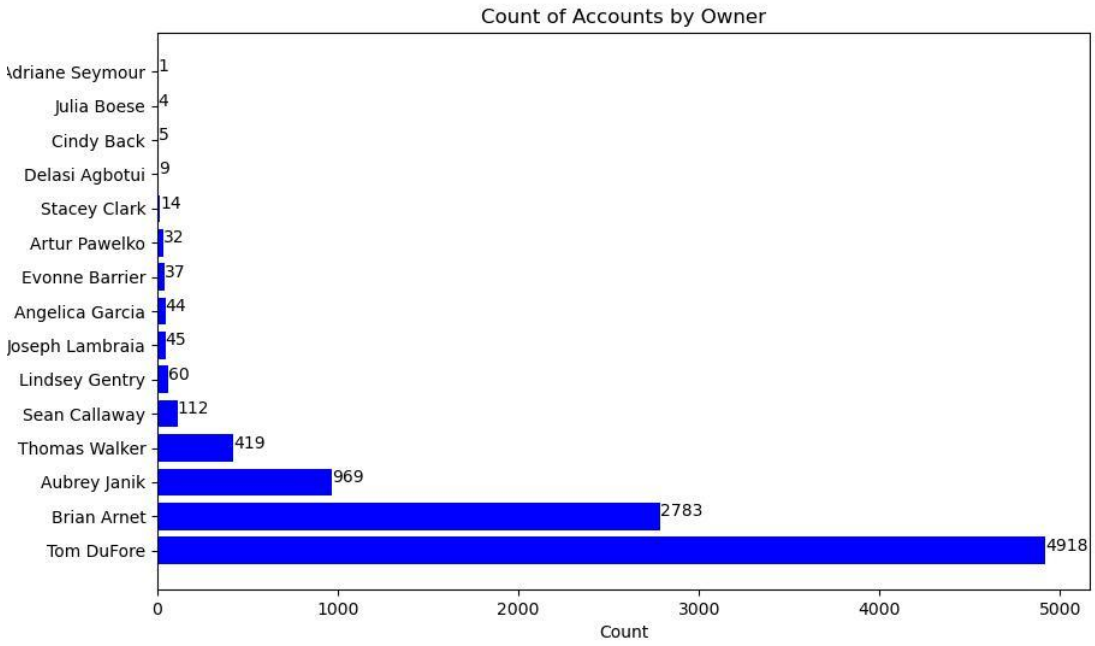
1. Print out table's stats and all missing value

```
Missing Value Percentage per Column:
record_id                  0.0 %
account_owner_id           0.0 %
account_owner              0.0 %
account_type              0.92 %
modified_by_id             0.0 %
modified_by                0.0 %
created_time               0.0 %
modified_time              0.0 %
last_activity_time        0.08 %
lead_received             9.35 %
lead_source               5.51 %
status                    2.02 %
follow-up_date           61.26 %
client_inquiry           31.89 %
mailing_zip              95.84 %
mailing_country          72.16 %
mailing_state            33.59 %
attended_presentation      0.0 %
partner_rep_id           92.66 %
partner_rep              92.66 %
change_log_time          97.29 %
locked                     0.0 %
last_enriched_time       100.0 %
enrich_status            98.87 %
new_customer_type         0.15 %
lead_sub_source_id        0.43 %
lead_sub_source           0.43 %
dtype: object
```

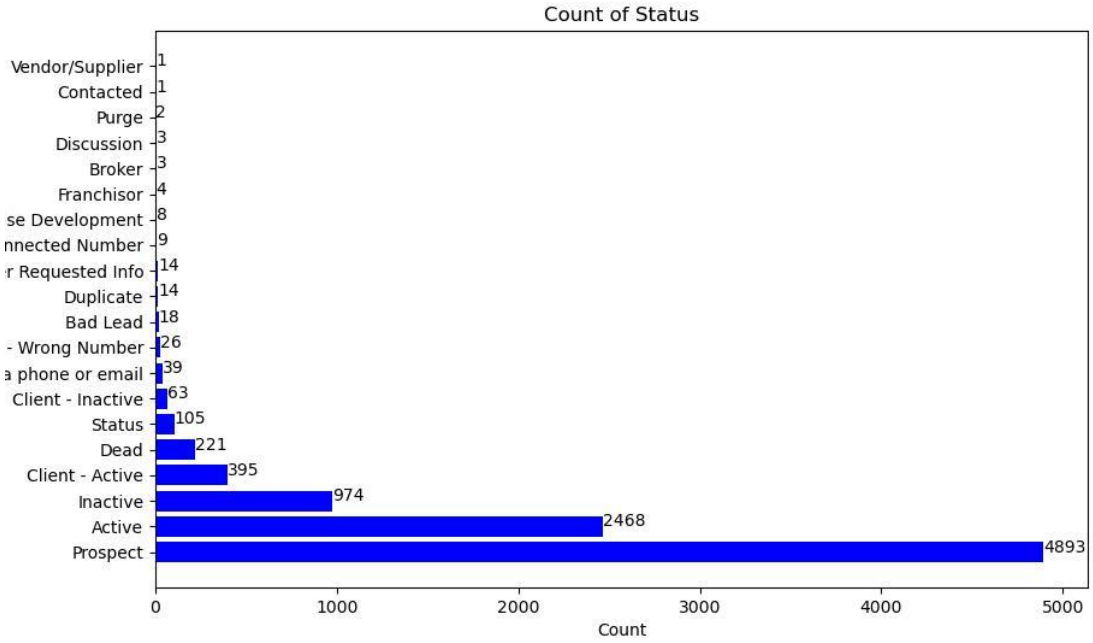
2. Merge datasets together and make them flow better.
3. Clean column names: convert to lowercase, replace space & dots with underscores
4. Distinguish duplicate column names
5. Drop unnecessary columns (eg. id columns or "last enriched time")
6. Fill blank cells with unknown (eg. "new customer type" or "lead sub source")
7. For column "last activity time," file NA with 'follow_up date' or 'modified time'
8. For "mailing country" column file NA based on mailing state or replaced with unknown
9. For the "mailing state" column file NA is based on the mailing zip or replaced with unknown. Edit cell types so it provides an accurate number for each state and province.
10. File NA with "unavailable" (eg. "Enrich status")
11. Change "probability" to an ordinal value.

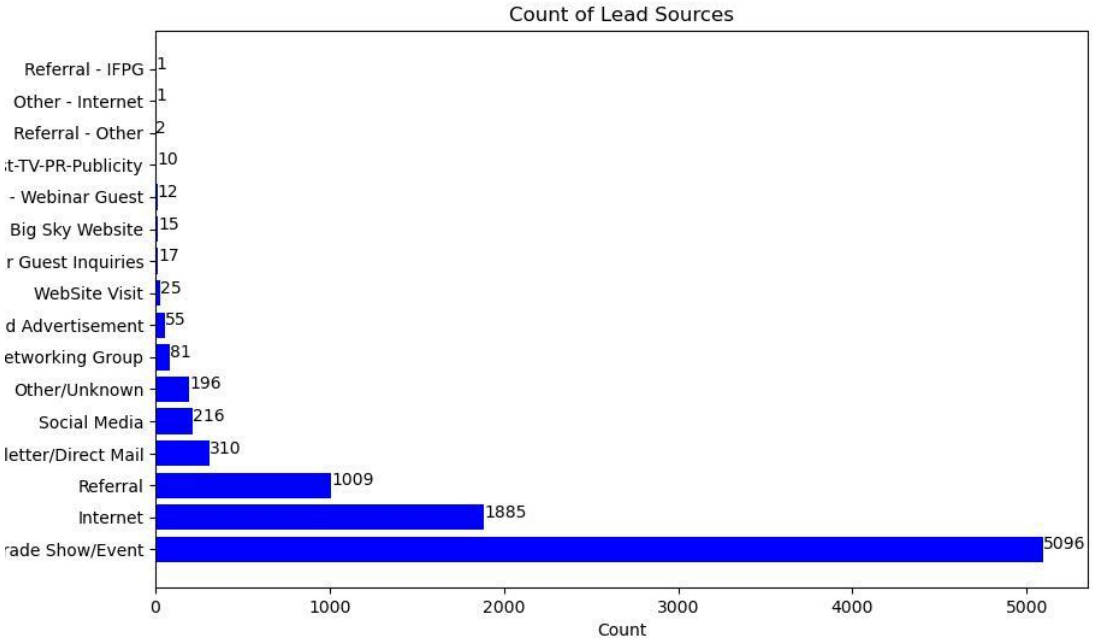
EDA and domain research (that is related to our proposal)

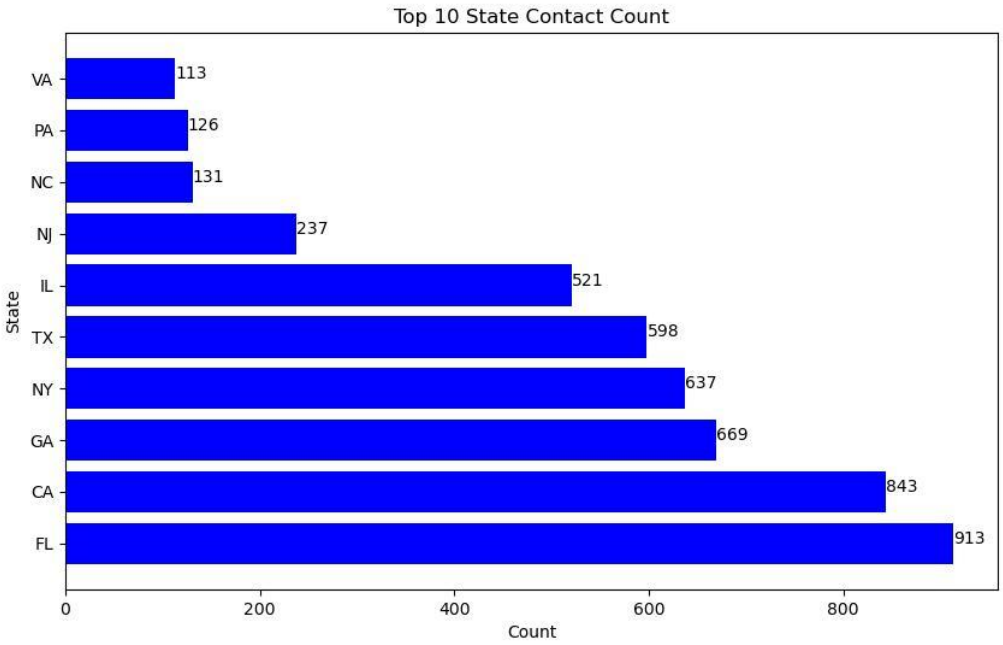
- Identify trends and patterns between attributes that will allow the company to land more potential clients.
- Evaluate and compare similar datasets from external sources that could benefit the analysis of Big Sky datasets.
- Examine relationships between variables in the datasets for potential correlations that are beneficial to answering business questions and address past concerns.

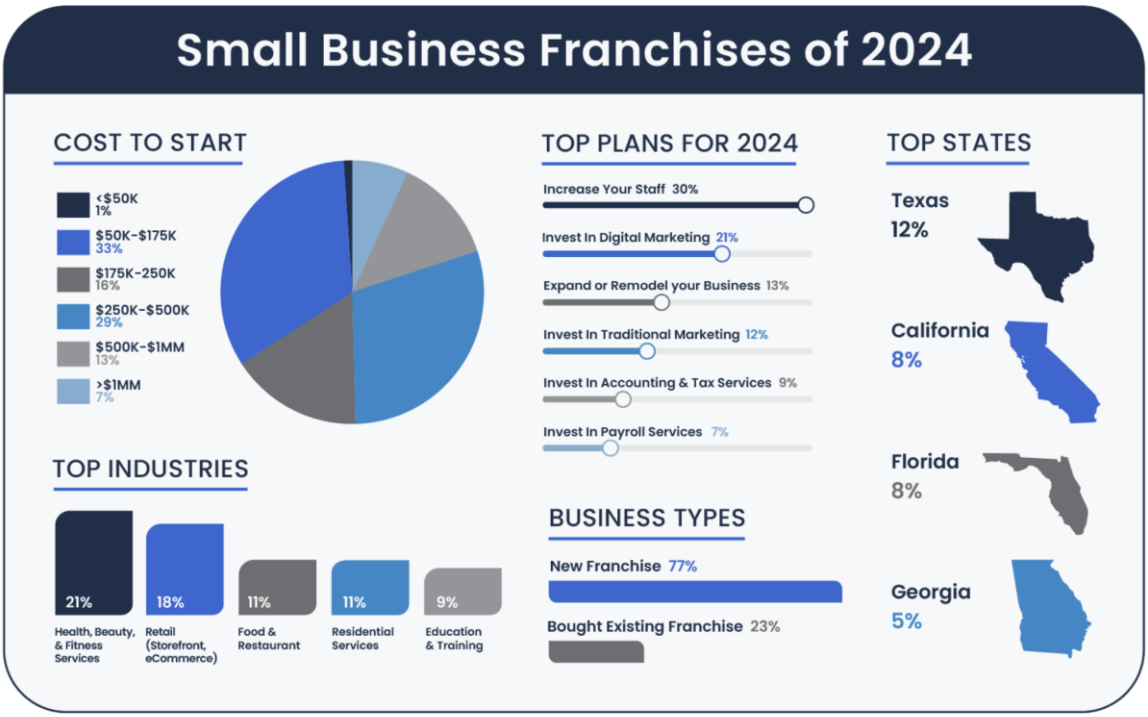


This chart shows the number of customers each employee has from 2020 to 2025. We can see that the top three have more than 5,000 customers combined, showing the difficulty of managing customer accounts.


Count of Status
Vendor/Supplier
Contacted
Purge
Discussion
Broker
Franchisor
se Development
nnected Number
r Requested Info
Duplicate
Bad Lead
- Wrong Number
a phone or email
Client - Inactive
Status
Dead
Client - Active
Inactive
Active
Prospect
1
1
2
3
3
4
8
9
14
14
18
26
39
63
105
221
395
974
2468
4893
0
1000
2000
3000
4000
5000
Count


Count of Lead Sources
Referral - IFPG
Other - Internet
Referral - Other
t-TV-PR-Publicity
- Webinar Guest
Big Sky Website
r Guest Inquiries
WebSite Visit
d Advertisement
etworking Group
Other/Unknown
Social Media
letter/Direct Mail
Referral
Internet
rade Show/Event
1
1
2
10
12
15
17
25
55
81
196
216
310
1009
1885
5096
0
1000
2000
3000
4000
5000
Count





This infographic provides a comprehensive overview of small business franchises in 2024.

Cost to Start (shown in pie chart):

- Most franchises (33%) require $50K-$175K to start
- 29% need $250K-$500K

- Only 1% can be started with less than $50K
- 7% require over $1 million

Top Industries:

1. Health, Beauty, & Fitness Services leads at 21%
2. Retail (including storefront and eCommerce) at 18%
3. Food & Restaurant and Residential Services tie at 11% each
4. Education & Training at 9%

Top Plans for 2024:

- Increasing staff is the highest priority (30%)
- Digital marketing investment follows at 21%
- Business expansion/remodeling at 13%
- Traditional marketing at 12%
- Lower priorities include accounting/tax services (9%) and payroll services (7%)

Business Types:

- New franchises dominate at 77%
- Only 23% are bought existing franchises

Top States for Franchises:

1. Texas leads with 12%
2. California and Florida tie at 8% each
3. Georgia at 5%
4. This result is the same as our analysis of the top five locations with the most customers.

This data suggests a robust franchise market with a focus on service-based businesses, particularly in the health and retail sectors, with significant initial investment requirements and a strong emphasis on growth through staffing and digital presence.

More questions and issues to address/answer

- Our analysis of your data will show you what to focus on in order to land more clients
- What is the average, min/max number of days from inquiry to calling back
- Add pictures of the franchises in the slides

- Goal is to triple the amount of conversions
- Count number of “probabilities” in dataset. I think these are really just stages in business relationship development.
- Add a year column (fiscal year?)
- Does certain states or regions of the country impact inquiries and higher probabilities from the dataset?

Final words/What can be expected

- A demo of managing the customer accounts system will be built.
    - Map indicates customers distribution
    - Timeline of customer life cycle
- Data collection methods to gain important metrics will be suggested.
- An Interactive dashboard on data insights will be provided.
- Explain why the analysis will be effective in long-term growth for the company.
- A big highlight at the beginning of the presentation will be bringing data governance to the forefront of their data collection efforts.
    - Why data governance is important
    - The current situation they are in
    - What they need to focus on
    - Why clean data makes for more efficient analysis

Reference

2024 Small Business franchise Trends - Guidant. (n.d.). Guidant. https://www.guidantfinancial.com/small-business-trends/franchise-business-trends/