CAPSTONE PRESENTATION TECHNICAL PRESENTATION

CONTENTS

BACKGROUND AND BUSINESS PROBLEM

BACKGROUND AND INTRODUCTION OF THE BUSINESS

- Started in 1999 as a small venture but has grown to more than 300 monthly active users and a total clientele of around 3.000 and yearly revenue of around CAD$100,000.
- Owner experience with yoga: 30+ years (Including fitness instructor and personal trainer for more than ten years in several Ontario colleges and schools)
- In 2019, they pushed the digital side of the business. They started to develop courses and content, which you can observe today in all their digital media presence. In mid-May 2022, they formally incorporated the company.

THE BUSINESS PROBLEM - **DIAGNOSIS**

- After several conversations with the client owner, we identified business problems not related to data but that should be solved to create a better data environment.

  - Target client definition.
    - Which information should I ask the new clients.
    - How to track the old ones
  - Value proposition
    - Using data as leverage.
  - Marketing strategy
    - Should create a marketing strategy with all its components to obtain better results.

THE BUSINESS PROBLEM – **KEY CHALLENGES**

- Software have been changing based on the business needs but it's highly fragmented and not "talking" to each other.
  - Square / Google Analytics / Newsletters / Excel / Trello / Accounting Software
  - Solutions: Mindbody (LINK), Wellness Living (LINK)
- Inexistence of a properly-structured DB or infrastructure to host it.
- Solutions: Use the proposed software or implement a complete tailor-made solution.
- The owner's idea is to go global, and we expect the amount of data will grow exponentially, but its hard to analyze all this data and choose the metrics to do so without implementing a complete solution.

ANALYTICAL APPROACH & METHODOLOGY

ANALYTICAL APPROACH

- Analyze customers using loyalty rank to understand who are most frequent customers and their preference for yoga type
- Segment the customers based on loyalty rank.
- Create RF Model to categorize customer into High Loyalty and Low Loyalty.
- Build logistic regression model to understand the loyalty level based on:

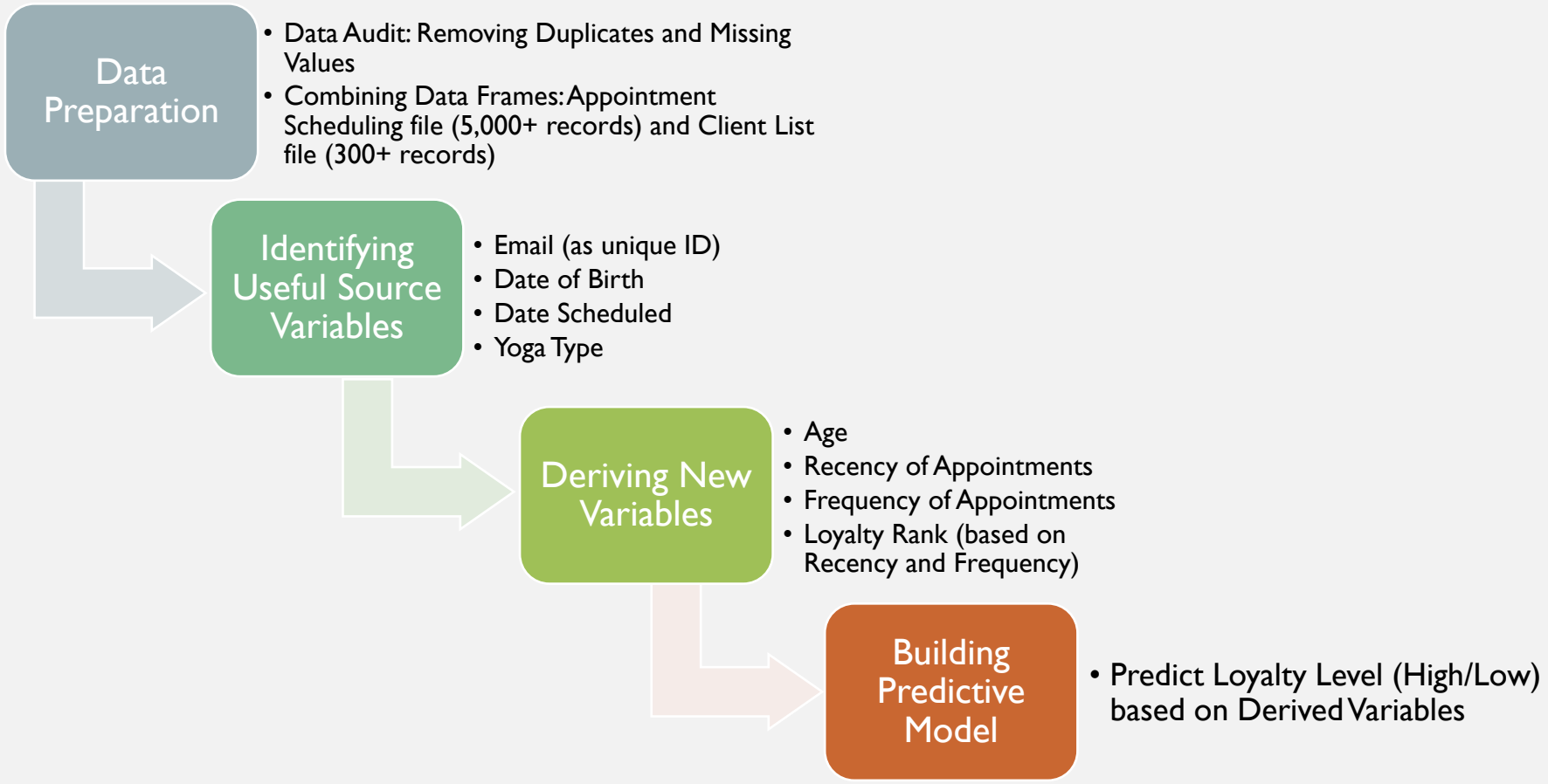
❑ Age

❑ Recency

❑ Frequency

❑ Loyalty Rank

ANALYTICAL APPROACH



METHODOLOGY

Excel: To perform data audit, Data cleaning, storing data and visualize data.

SQL: Joining two different dataset and creating a table.

Python: To analyze data, perform RF analysis, Logistic regression and visualizing data.

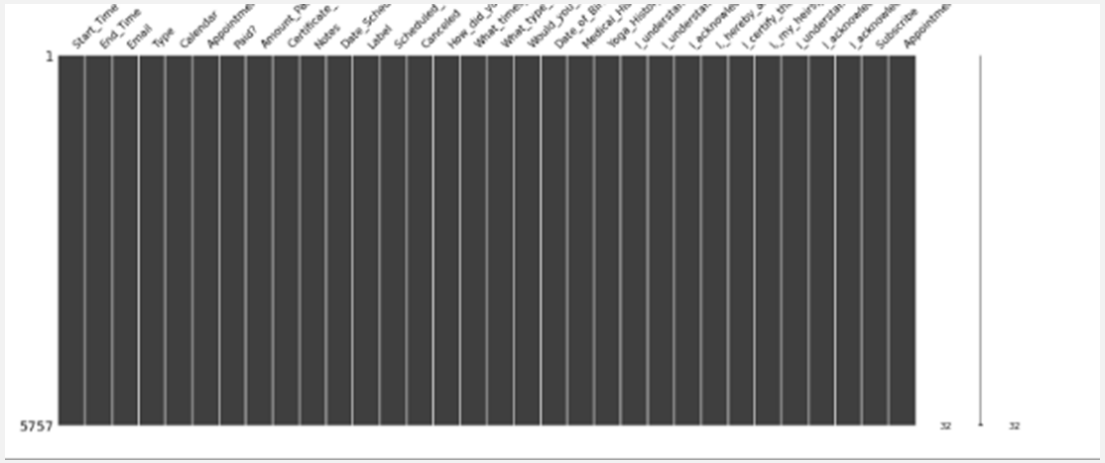
KEY FINDINGS OF DATA EXPLORATION

SUMMARY OF KEY FINDINGS OF DATA AUDIT

- From the working files and the process analysis, we can summarize the findings in:
  - The intake form is manually filled, so it creates duplicates, typos, blank fields.
  - We couldn't obtain raw data from Google Analytics. It will need separate analysis.
  - Active clients appeared as 341 but after removing duplicates, incompletes and not real, we were around 300 clients.
    - Going deeper we find there is actually 28 duplicates
    - Others might be missing data, no emails, or data entry errors

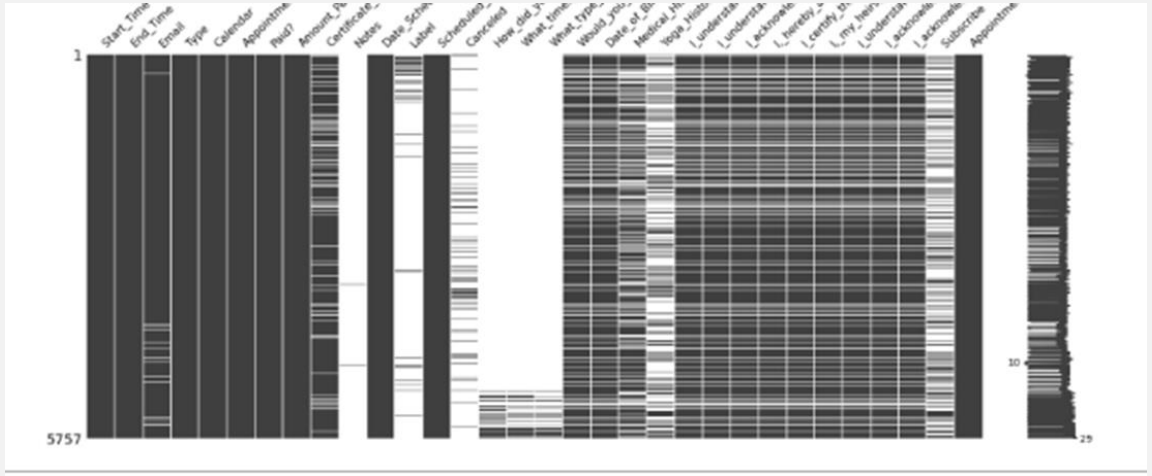
SUMMARY OF KEY FINDINGS OF DATA AUDIT

A complete dataset should look like this one...without any white spots that represent empty records



BUT

Our data set, has several records empty and that caused interference when we tried to perform the analysis



SUMMARY OF KEY FINDINGS OF DATA AUDIT

	variable	%_Missing	Unique_Values	Data_Type
0	Start_Time	0	1465	object
1	End_Time	0	1481	object
2	Email	1.56	214	object
3	Type	0	16	object
4	Calendar	0	2	object
5	Appointment_Price	0	2	float64
6	Paid?	0	2	object
7	Amount_Paid_Online	0	3	float64
8	Certificate_Code	13.34	191	object
9	Notes	99.58	21	object
10	Date_Scheduled	0	469	object
11	Label	92.37	5	object
12	Scheduled_By	0	65	object
13	Canceled	89.11	1	object
14	How_did_you_hear_about_us?	94.51	9	object
15	What_times_of_day_do_you_prefer_to	95.97	42	object
16	What_type_of_yoga_do_you_prefer?_(	96.63	7	object
17	Would_you_like_to_be_on_our_inform	19.21	2	object
18	Date_of_Birth_(YYYY/MM/DD):	19.25	222	object
19	Medical_History._Please_list_all_healt	33.77	365	object
20	Yoga_History__(If_new_to_yoga_what_	63.99	257	object
21	I_understand_there_is_an_inherent_ri	19.21	1	object
22	I_understand_and_am_aware_that__th	19.21	1	object
23	I_acknowledge_that_I_have_either_ha	19.21	1	object
24	I,_hereby_assume__all_responsibility_	19.21	1	object
25	I_certify_that_I_am_physically_well_ar	19.21	1	object
26	I,_my_heirs,_or_legal_representatives,	19.21	1	object
27	I_understand_that_Atlas_Yoga_studio_	19.21	1	object
28	I_acknowledge_that_I_have_read_this	19.21	1	object
29	I_acknowledge_that_I_am_signing_this	19.21	1	object
30	Subscribe	67.6	1	object
31	Appointment_ID	0	5757	int64

Important information to understand the clients is missing:

More than 90% :
How they heard about the studio, the type of yoga they prefer or what it's the best time to practice it.

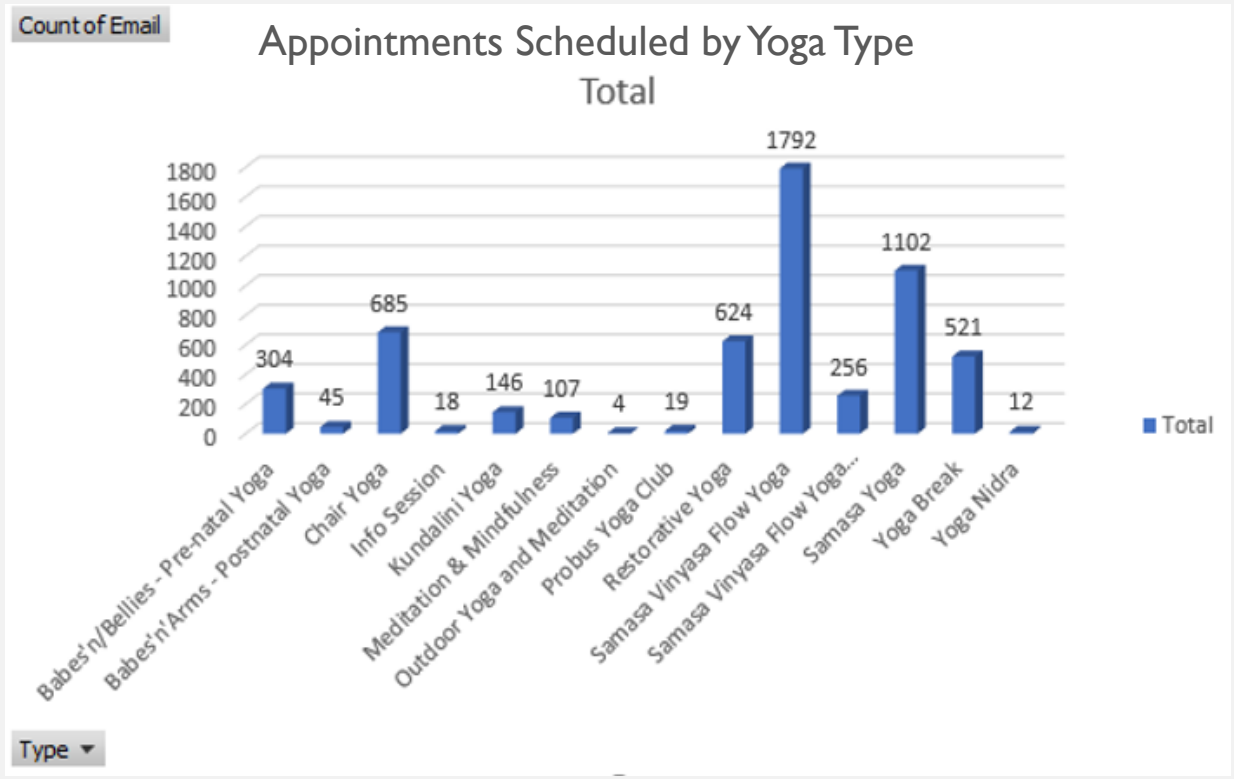
Between 30% and 60%: What is the client's experience or if they have any medical condition worth mentioning.

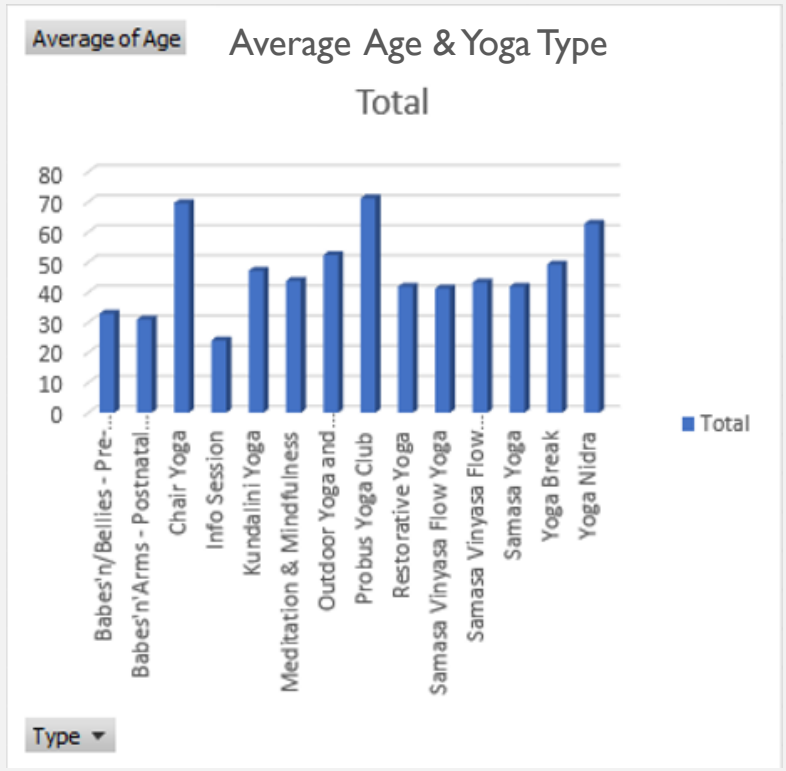
One in five records regarding health disclaimers and responsibilities are skipped.

Two thirds don't subscribe to the newsletters or any bulletins

KEY FINDINGS OF DATA EXPLORATION

- Samasa Vinyasa Flow Yoga class is the most frequently booked class and it's frequented by people with an average age of 40 years.
- Outdoor Yoga and Meditation class are the least booked classes with only 4 bookings in the last year and a half.





ANALYTICAL FILE

ANALYTICAL FILE VARIABLES

- RF Model for customer segmentation
- Logistic regression model to understand the loyalty level based on derived variable.
- The key source variable used to create a RF Model (Recency Frequency) are:
- Date of Birth
- Email Address
- Date Scheduled

SOURCE VARIABLES

Source Variable	Description
Date Scheduled	It will help us in calculating Recency.
Email	Email id of customer, will be used to calculate frequency of customer through scheduling file.
Date of Birth	Date of birth of customer. Use to calculate age of customer.
Medical History	It describe the different medical history/condition of customer. It will be use as a segmentation criteria
Yoga Type	Describe different yoga type practiced at AYS. Used for segmentation criteria.
Yoga History	It describe since when the client has been doing yoga.

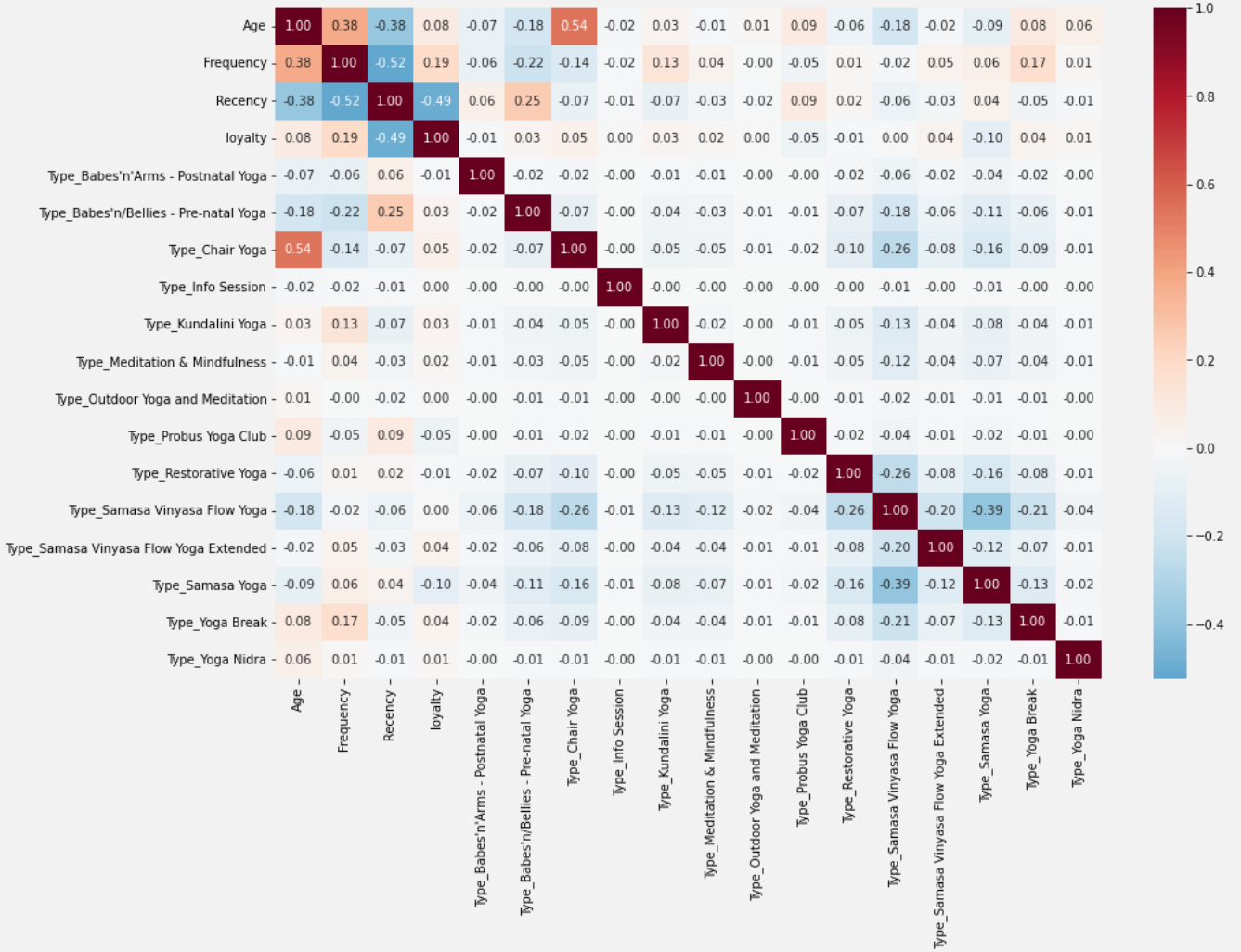
DERIVED VARIABLES

Derived Variable	Description
Age	Calculated from DOB. Useful for segmenting customers by demographics. It will help make a specific intense level plan for different age groups
Recency	It will help in understanding customer's last visit to the atlas yoga studio website..
Frequency	It will help in knowing their most regular client based on how often they take yoga class
Loyalty Rank	It classify customers into different segment Platinum, Gold, Silver, and Bronze based on RF model.

TARGET VARIABLES

Target Variable	Description
High Loyalty Rank	Platinum and Gold.
Low Loyalty Rank	Silver and Bronze.

DETERMINING KEY VARIABLES FOR THE MODEL



EXPLORING KEY VARIABLES WITHIN EACH SEGMENT (HIGH LOYALTY & LOW LOYALTY)

```
High
count    4180.000000
mean       45.096411
std        14.416531
min        19.000000
25%        35.000000
50%        43.000000
75%        57.000000
max        82.000000
Name: Age, dtype: float64
```

Symmetric

```
Low
count    125.000000
mean      37.048000
std        9.598266
min       21.000000
25%       29.000000
50%       36.000000
75%       45.000000
max       67.000000
Name: Age, dtype: float64
```

Symmetric

```
High
count    4180.000000
mean       62.435646
std       100.729958
min         0.000000
25%         0.000000
50%         3.000000
75%       107.000000
max       448.000000
Name: Recency, dtype: float64
```

Highly Skewed

```
Low
count    125.000000
mean      413.616000
std       105.770429
min       110.000000
25%       406.000000
50%       451.000000
75%       482.000000
max       536.000000
Name: Recency, dtype: float64
```

Roughly Symmetric

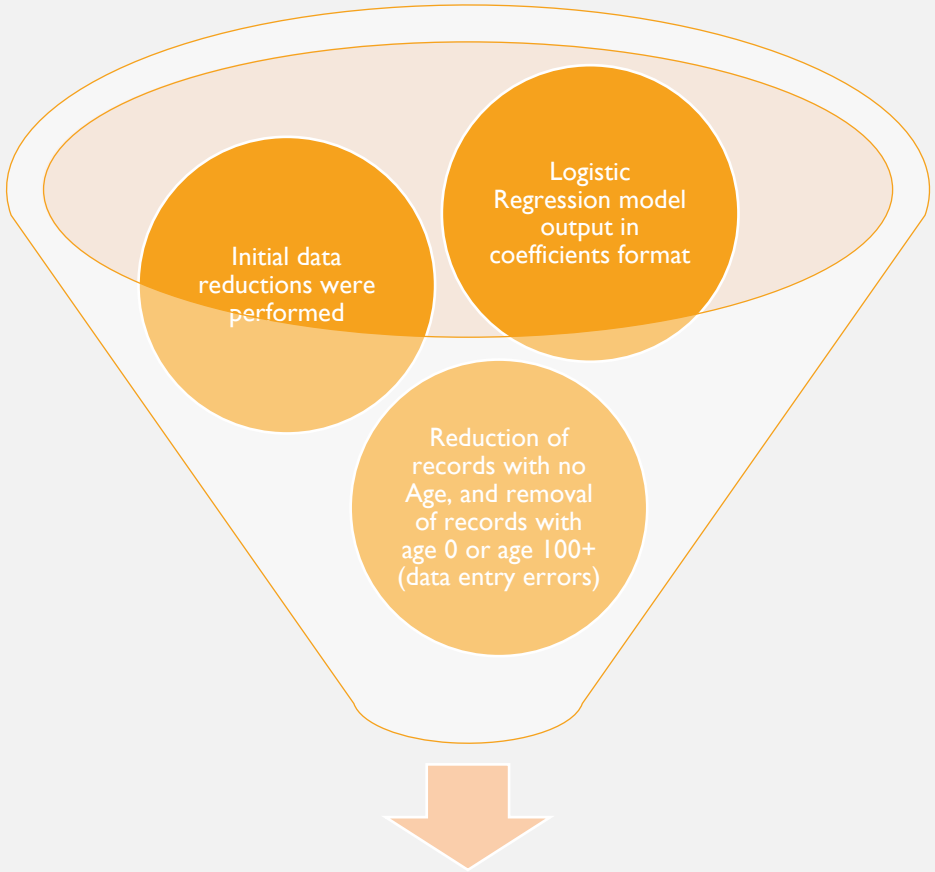
```
High
count    4180.000000
mean      210.519139
std       177.920805
min         1.000000
25%        71.000000
50%       126.000000
75%       333.000000
max       556.000000
Name: Frequency, dtype: float64
```

Highly Skewed

```
Low
count    125.000000
mean       4.080000
std        3.946415
min        1.000000
25%        1.000000
50%        2.000000
75%        7.000000
max       13.000000
Name: Frequency, dtype: float64
```

Highly Skewed

MODEL REPORT



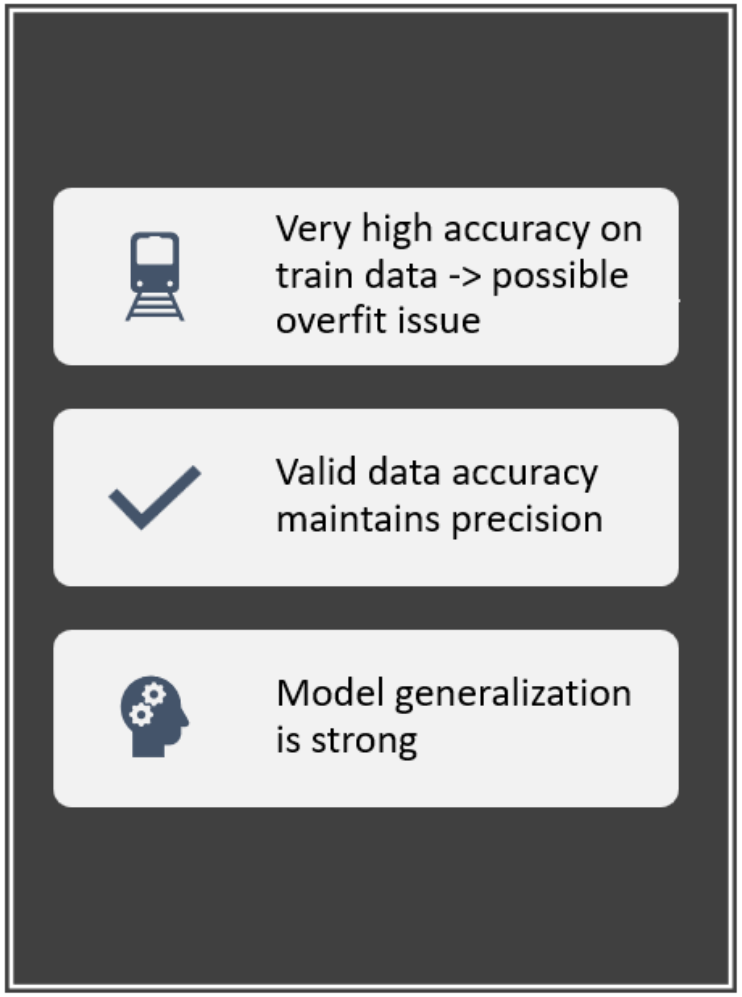
Num records reduced from 5000+ to approx. 4000.

intercept 4.2943473255291815

	coeff
Age	1.276418e-03
Frequency	5.458094e-01
Recency	-2.416836e-02
Type_Babes'n'Arms - Postnatal Yoga	1.045936e+00
Type_Babes'n/Bellies - Pre-natal Yoga	2.244114e+00
Type_Chair Yoga	5.814966e-01
Type_Info Session	0.000000e+00
Type_Kundalini Yoga	1.000384e+00
Type_Meditation & Mindfulness	-2.888379e+00
Type_Outdoor Yoga and Meditation	-2.009402e-07
Type_Probus Yoga Club	4.767678e+00
Type_Restorative Yoga	-1.793997e+00
Type_Samasa Vinyasa Flow Yoga	1.450857e-01
Type_Samasa Vinyasa Flow Yoga Extended	8.417978e-02
Type_Samasa Yoga	-6.259603e-03
Type_Yoga Break	-9.040906e-01
Type_Yoga Nidra	1.819973e-02

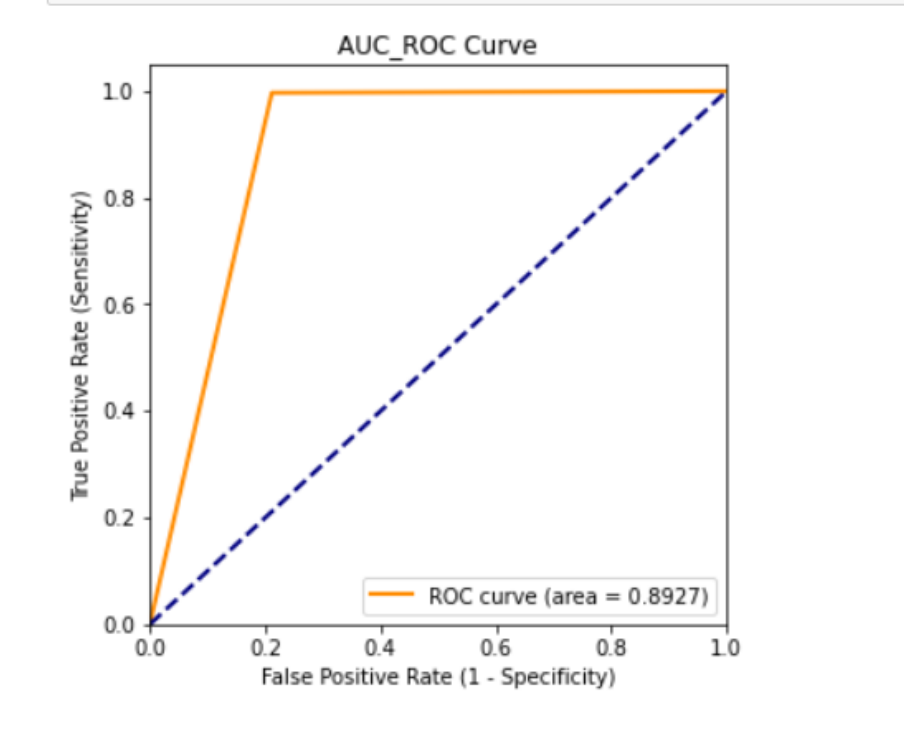
AIC -3131.816093494983

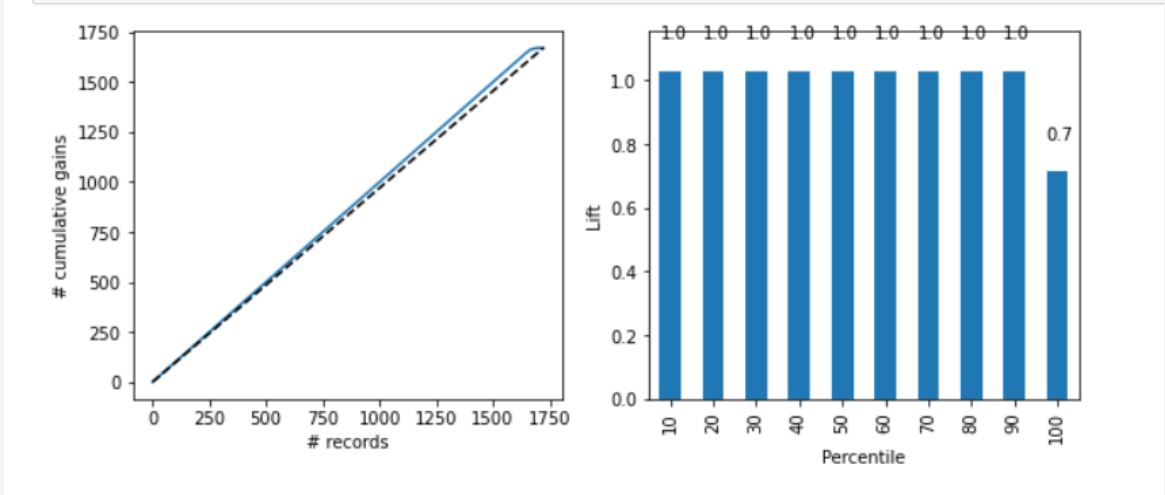
FINAL MODEL REPORT




```
In [20]: classificationSummary(train_y,logit_reg.predict(train_X))

         Confusion Matrix (Accuracy 0.9938)

                Prediction
         Actual    0    1
              0   62   11
              1    5 2505

In [21]: classificationSummary(valid_y,logit_reg.predict(valid_X))

         Confusion Matrix (Accuracy 0.9907)

                Prediction
         Actual    0    1
              0   41   11
              1    5 1665
```

DECILE GAINS & AUC CHART





- AUC_ROC curve indicating a 0.89 area
- Strong model performance indicator
- Gains are minimal, and lift doesn't follow a step like pattern.

KEY INSIGHTS: CLASSIFICATION MODEL

PREDICTORS

Strong predictors of high loyalty

- Samasa Vinyasa Flow Yoga Extended
- Chair Yoga
- Frequency of visits

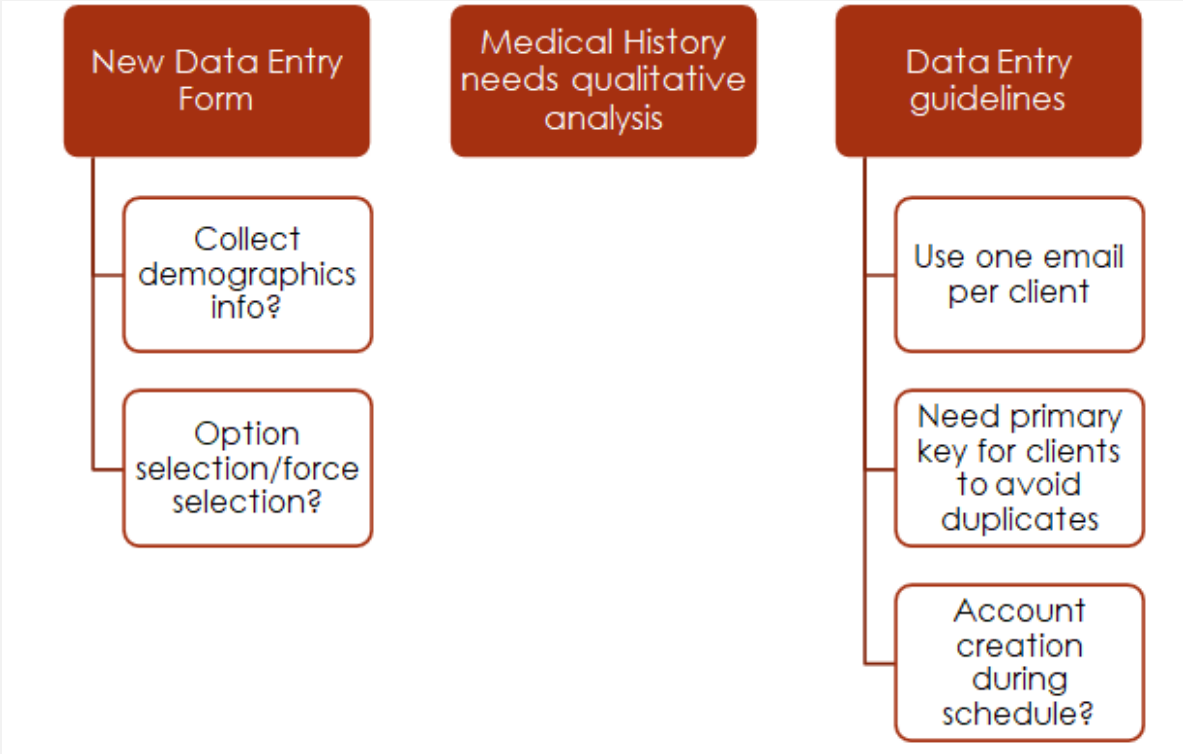
Strong predictors of low loyalty

- Yoga Break
- Samasa Yoga

RECOMMENDATION TO IMPROVE DATA QUALITY

1. Define what you can measure and how you can use it.

2.



3. Improve website to make the class selection easier (based on client's experience or preferences)