THE TRACK FITNESS GYM CASE MEMO

Roll Number: 21f2000426

Name: Arnav Kumar

Executive Summary:

I had approached the fitness studio's responsible staff member and with the owner's permission collected the client invoices, client personal information and punch-in punch-out digital logs which are recorded from the start of this year, January 2022 to the end of august 2022.The collected data can help the business in identifying its shortcomings and finding an effective solution to deal with them.

To solve the problem of unhealthy retention rate, which can be looked into via the invoices generated and collected between the said period of time, the necessary data has been cleaned and transformed into a clean spreadsheet which will aid me to draw some insights and gain a basic understand of the problem at hand which is the primary reason for dynamic and unstable revenue

The business second concern, Overcrowding can be clearly seen at specific slots in the graphs I have attached in appropriate sections below. From the first look it is evident to me that most people in this fitness studio prefers working out in the morning and evening mostly between 7 am to 9 am and 6 pm to 8 pm respectively, and that's when the gym is congested and client might leave because of absence of Equipment and even less ground to step the foot on.

Proof of originality of the Data:

1. Primary Data

I have shared the raw / primary data which I had collected from the business with the exception of few features like client phone number, name and other sensitive records for privacy concerns. The Link to these data files is provided below and is already share with the IITM-Panel

Data Files Link:

https://drive.google.com/drive/folders/1ATU8yRkMy70g9QYP1xgWggp9ohzALWCX

2. Photographs

I have also shared some images of the Fitness Studio which spans across a huge area and consists of a ground floor mainly focused for weight lifting and the first floor which mostly has cardio focused equipment's. All the pictures that are included gives the viewer a crisp idea about the Fitness studio and the facilities it provides. All the pictures are shared with the IITM group and can be accessed from the below Link.

Image Link:

https://drive.google.com/drive/folders/1AVclbxlB9IkEVup4COjK54sMn0ClrD1X

3. Organization Letter:

As the part of the proof I had also asked the owner to notarize a letter for me which mentions all the necessary clauses which I thought will indeed be beneficial for my case. The letter can be accessed from the below link and is share with the IITM group.

organization Letter Link:

https://drive.google.com/drive/folders/1ATSBMzFCBBk0Km6it0CMiVS9ixtyPYgr

4. Interaction Video:

I was not able to capture a video as most of my talks with the business Owner happened over a phone call due to his unavailability most of the times as the owners handles multiple businesses spanned across different regions in the district. The staff also denied for a video proof because of personal privacy concerns.

Metadata and Descriptive statistics:

The client invoices, client personal data (Personally Identifiable Information, P.I.I) and the client logs or simply Punch in Punch Out (P.I.P.O) are the 3 files which are collected, with the invoice sheet the revenue generated for every month can be identified. After data cleaning transformation and elimination of unavailing data columns, I am left with some of the most important data features which I believe to be highly corelated with my analysis and have been compiled in one single file.

Basic operations like changing the data type of the 'supposedly numbers' from text or a string of text to numbers, dropping all the unusable columns which bears no consequences over the analysis, getting rid of columns that carry redundant information and adding a few meaningful columns to further simplify my own analysis has been done.

1. Invoice Sheet

In this sheet Featured columns like "CGST", "SGST"," tax"," Executive", "Mode of Payment"," Balance" were all dropped due to nature of the values they contained, for example the studio has tax included package prices, hence no dedicated tax is levied from the clients at the time of entry as a consequence all the tax columns had nothing but a zero as its every entry, other columns like 'payment mode' and balance didn't have much to contribute either way to the analysis and had to be dropped. I have also added a few columns of my own to the polished data which helped make my analysis easier. Below is the picture of the systematized data I came up with.

#Rno	PaymentDate	package	validity	subscription	Client joining_month	Discount	Paid Amount	Revenue
943	01 January 2022	3	01 April 2022	Expire	1	0.00	2700.00	2700.00
953	01 January 2022	1	31 January 2022	Expire	2021	0.00	1000.00	1000.00
1018	01 January 2022	1	31 January 2022	Expire	1	0.00	1000.00	1000.00
1019	01 January 2022	1	31 January 2022	Expire	1	0.00	1000.00	1000.00
1020	01 January 2022	1	31 January 2022	Expire	2021	0.00	1000.00	1000.00
1021	01 January 2022	1	31 January 2022	Expire	1	0.00	1000.00	1000.00
1024	01 January 2022	1	31 January 2022	Expire	2021	0.00	1000.00	1000.00
1025	01 January 2022	1	31 January 2022	Expire	2021	0.00	1000.00	1000.00
1026	01 January 2022	1	31 January 2022	Expire	1	0.00	1000.00	1000.00
1027	01 January 2022	3	01 April 2022	Expire	1	0.00	2700.00	2700.00
1032	01 January 2022	1	31 January 2022	Expire	2021	0.00	1000.00	1000.00
1036	01 January 2022	1	31 January 2022	Expire	1	0.00	1000.00	1000.00
1044	01 January 2022	3	01 April 2022	Expire	2021	0.00	2700.00	2700.00
1045	01 January 2022	3	01 April 2022	Expire	2021	0.00	2700.00	2700.00
1046	01 January 2022	3	01 April 2022	Expire	1	0.00	2700.00	2700.00
1047	01 January 2022	3	01 April 2022	Expire	1	0.00	2700.00	2700.00
1048	01 January 2022	6	30 June 2022	Expire	1	0.00	4500.00	4500.00
1049	01 January 2022	6	30 June 2022	Expire	2021	0.00	4500.00	4500.00
1050	01 January 2022	6	30 June 2022	Expire	1	0.00	4500.00	4500.00
1052	01 January 2022	6	30 June 2022	Expire	2021	0.00	4500.00	4500.00
1053	01 January 2022	3	01 April 2022	Expire	1	0.00	2700.00	2700.00
1054	01 January 2022	12	01 January 2023	Active	1	0.00	8500.00	8500.00
1055	01 January 2022	12	01 January 2023	Active	2021	0.00	8500.00	8500.00
1057	01 January 2022	6	30 June 2022	Expire	2021	0.00	4500.00	4500.00
1058	01 January 2022	6	30 June 2022	Expire	2021	0.00	4500.00	4500.00
904	01 January 2022	1	31 January 2022	Expire	2021	0.00	1000.00	1000.00
967	02 January 2022	3	02 April 2022	Expire	2021	0.00	2700.00	2700.00
1039	02 January 2022	1	01 February 2022	Expire	1	0.00	1000.00	1000.00
1060	02 January 2022	6	01 July 2022	Expire	2021	0.00	4500.00	4500.00
1063	02 January 2022	1	01 February 2022	Expire	1	0.00	1000.00	1000.00
1065	02 January 2022	1	01 February 2022	Expire	1	0.00	1000.00	1000.00
1066	02 January 2022	3	02 April 2022	Expire	1	0.00	2700.00	2700.00
968	03 January 2022	12	03 January 2023	Active	1	0.00	8500.00	8500.00
983	03 January 2022	1	02 February 2022	Expire	2021	0.00	1000.00	1000.00
1071	03 January 2022	1	02 February 2022	Expire	1	0.00	1000.00	1000.00
1073	03 January 2022	1	02 February 2022	Expire	1	0.00	1000.00	1000.00
1076	03 January 2022	1	02 February 2022	Expire	2021	0.00	1000.00	1000.00
1077	03 January 2022	3	03 April 2022	Expire	2021	0.00	2700.00	2700.00
1078	03 January 2022	3	03 April 2022	Expire	1	0.00	2700.00	2700.00

Metadata:

1.	#Rno	Upon registration, every client is given a unique identification number which is also called as Registration Number or #Rno
2.	Payment Date	This column consist of the dates on which invoices were generated
3.	Package	Package column consists of the duration (in month) for which the client has purchased the gym subscription for
4.	Validity	Indicates the date beyond which the client gym membership expires
5.	Subscription	Shows the status of the client membership, either Active or inactive, the status may change with every consecutive recharge after membership expiration
6.	Discount	This column shows the discount offered by the business
7.	Paid Amount	Shows the amount paid by the client to buy the membership for any specific duration
8.	Revenue	Shows the revenue, this column can be pivoted with other columns to find revenue across different criteria's, like revenue across every month and etc.
9.	Client joining month	Indicates the month in which a client joins the gym for the first time

Descriptive statistics:

- The maximum revenue of 277500 Rs was generated in the month of January, which is primarily because of the new year resolutions which people take. The revenue dropped in the month of February and march and then kept slowly rising, as of 31st August 2022 the gym is sitting at a revenue of approximately 2 lacs.

- The gym added 319 new clients in the span of 8 months, out of which 140 are still active hence retention rate among new clients is 43 %

- A total of 494 unique members were registered with the track fitness club across the period of 8 months from January to august, out of which the gym has managed to retain 188 of them, that is there are currently 188 Active subscriptions, thus the overall retention rate is 38 %

- Throughout the period of 8 months 803 one-month packages, 477, three months packages and 144, 1-year subscription packages were bough by the clients, which proves that an average client rather pay up 1000 Rs upfront every month but not recharge for 3 or 6 months at once for a lot less money.

- The business loses 50 percent of its new client every 3 months, which is evident from the month of January where 71 new clients joined out of which only 32 clients had an

active subscription at the start of April, February had seen a registration of 35 new members which dropped to 12 after 3 months which is about 34 %

- February and march were the least performing month for the business, upon further enquiry it was discovered that this opening of a new gym (SJ gym) in the neighborhood affected the revenue of track Fitness gym the new gym had lucrative sign up offers which took away some of the current and new members of the business

- A total of 15555 RS of Discount was given in the month of June as a part of the monsoon offers which shot off the new registrations by a huge margin of 40%, June had 2nd highest new client registrations after January

The above statistics is very helpful in understanding the business revenue fluctuations and current retention rates.

2. Client Information File (P.I.I):

This File contains the personal information of the clients except their phone number and names basic data cleaning involved changing data types, removing outliers which are mostly the typing errors made at the billing counter for.
Below is the image for the file

#Rno	Instructor	Age	Gender	member Goal	Occupation
943	NA	37	M	cardio and flexibility	working professional
953	NA	19	M	strength gain	student
1018	NA	25	M	Weight gain	working professional
1019	NA	37	M	weight loss	working professional
1020	NA	39	M	cardio and flexibility	self employed
1021	NA	18	F	Weight loss	student
1024	trainer nikhil	39	M	strength gain	job
1025	NA	19	M	Fat loss Muscle Gain	student
1026	NA	28	M	Weight loss	job
1027	trainer sachin	50	M	weight loss	working professional
1032	NA	33	M	weight loss	working professional
1036	trainer nikhil	28	M	strength gain	working professional
1044	trainer kiran	16	M	strength gain	student
1045	NA	17	M	Weight loss	student
1046	NA	28	F	Weight loss	job

Metadata:

The instructor column indicates whether or not a member opts for a personal trainer, remaining all the column labels are very self-explanatory.

Descriptive statistics:

Upon combining the personal information of 494 unique clients and the invoice data file which had over 1000 observations into a one single sheet, I came across few findings which are explained below

- Out of 173 active members (321-member subscription has expired as of 31st august) only 13 percent (23) are Females and the rest are males.
- 65% of the active Female clients wants to lose weight, followed by 21% females who mentioned cardio and flexibility as their goal followed by 9 percent females who want to lose fat and also gain some muscle.
- About 40% of active males mentioned losing fat and gaining muscle as their primary goal, which resonates with 38.65% males who shared the same goal among overall males (not just the active ones). The data indicates that the % of males who share this particular goal did not toppled over a course of time.
- The data also indicates that most of the female quitters were the ones who wanted to lose weight, the percentage drop between females who wanted to lose weight (61%) and females who wanted to lose weight and are active (48%) is 13% which is the highest among females.
- In the history of 8 months, about 30 % of the clients opted for personal training, the clients who showed a least drop in the % across 8 months were fat loss and muscle gain focused clients.

3. Punch-in punch-out file (PIPO):

This file contains every digital attempt to enter the workout premises including when a fingerprint is misread and labelled as illegal fingerprint due to the finger being sweaty or finger moved too soon from the scanner.

After making necessary changes like removing all the digital records where the member failed to pass the scanner also removing all the records of the staff who went in and out as they pleased and extracting necessary values from the existing columns, I cultivated a datasheet which is photographed below.

#Rno	Punch-In	Punch-Out	Day	Hour
943	2022-01-01 07.36.07	2022-01-01 09.25.18	Sat	7
1046	2022-01-01 12.15.48	2022-01-01 13.28.38	Sat	12
1024	2022-01-01 16.01.54	2022-01-01 17.33.04	Sat	16
1052	2022-01-03 06.20.05	2022-01-03 08.07.01	Mon	6
767	2022-01-03 06.38.20	2022-01-03 09.14.46	Mon	6
1049	2022-01-03 07.45.13	2022-01-03 10.03.05	Mon	7
1060	2022-01-03 08.07.10	2022-01-03 08.51.39	Mon	8
1019	2022-01-03 09.45.40	2022-01-03 10.23.31	Mon	9
1053	2022-01-03 10.03.31	2022-01-03 12.18.47	Mon	10
1066	2022-01-03 10.59.06	2022-01-03 11.04.00	Mon	10
1050	2022-01-03 11.26.01	2022-01-03 12.26.11	Mon	11
1027	2022-01-03 11.26.11	2022-01-03 12.18.40	Mon	11
1036	2022-01-03 11.38.03	2022-01-03 14.07.41	Mon	11
1065	2022-01-03 11.49.46	2022-01-03 11.58.19	Mon	11
1073	2022-01-03 14.58.56	2022-01-03 15.54.51	Mon	14
1079	2022-01-03 15.58.23	2022-01-03 18.26.45	Mon	15

Metadata:

1.	Punch-In	This column indicates the exact instance of time and date when a member enters the premises to start with their workouts
2.	Punch-out	Instances of time along with the date when a member leaves the premises after their workout
3.	Day	Represents the day, which is extracted from the date of the Punch-In column
4.	Hour	Represents the hour (in 24-hour format) at which the client logs in the premises

Descriptive Statistics:

- The frequency of active people in the gym gradually falls over a week, with Friday, Saturday and Sunday nights being the least
- Most workouts happened between 7 am – 8 am, followed by 6 pm – 7 pm slot, which are the prime time of congestion, also the workout frequencies are the least at 3pm. Most of the clients who are involved in the overcrowding period are males who are working professionals
- The client with the highest Adherence rate (workout frequency in a given period of time) had worked out for about 87% of their total active duration.

Detailed Explanation of Analysis Process / Method:

I have been plainly asked to work upon Two paramount problem the business suffers with, below is the detailed process I will use to approach these problems

1. Membership Retention:

After the obligatory data preprocessing, now I can proceed towards identification of all of the high AR clients via tracing each observation of the PIPO dataset via a for loop and keeping a count of every in-outs the respective client made through its subscription's validity. The clients who are working out since more than 4 months and worked out more than 75% of their total valid duration will be labelled as high AR clients.

I will import the dataset as a DataFrame via the Pandas module and keep tracks of the workout frequencies for all the candidates in a key-value pair, the unique registration number of the client will act as the key and the value to those keys will be the number indicating the count of their workouts (in days), thus making a dictionary.

The dictionary will be sorted in the descending order according to its values (workout frequencies) and their corresponding keys (Rno.). All the values will be then divided by the total duration of their corresponding key's subscription validity (in days) to get the percentage. In the end the all the clients who satisfy the established criteria will be labelled as high AR client, all the features of such clients will be minutely observed via MS excel and/or python so as to establish any common link or feature these high AR clients have in common. Upon the identification of the common link, the business will be approached with a suggestion

2. Overcrowding:

To approach this problem, I will have to again trace the PIPO dataset via converting it into a pandas DataFrame. The first step will be to create three slots namely morning, noon and evening slot, the morning slots will encapsulate all the clients who logged in the gym between 6 am to 11:59 am, The noon slot will have clients between 12 pm to 5:59 pm and the Evening slot will consist of all the members from 6 pm to 10 pm.

Clients which are found in every slot will be identified by tracing the list of unique clients containing the unique Registration number and maintaining and updating a dictionary only this time, the key to this dictionary will be the client Rno. and the value to this key will be the number amounting to the times a client switches / jumps from its current slot. A counter will be set to zero and will be incremented by 1 every time the client changes slot, this process will be carried out for each and every unique client

For example, if there exist a client whose corresponding key value in the dictionary is 270 (where 270 amounts to the number of times the client switched slots this means the client is probably flexible with the timing, higher the value more the possibility of the client being flexible so. Hence the flexible members can be identified who can be pushed into afternoon slots via some marketing techniques like personalized offers to prevent congestion

It goes without saying that first and foremost all the observations generated on Saturday and Sundays will be erased (since most people may become flexible on the weekend due to holiday at work, schools and colleges) so that it does not interfere with the analysis, the elimination of

such records will be done with the help of Microsoft excel, where a simple filter on the "Day" column in the PIPO data will do the job.

If the n number of slots are not able to catch much variation leading to minimal number of flexible clients then the number of slots will go from n to n + 1, to a maximum of 6 slots

Results And Findings:

a. Below is the pictorial Representation of the workout frequencies vs hour of the day, the spikes can be observed near the 7th to 8th hour mark and near 18th to 19th hour mark.

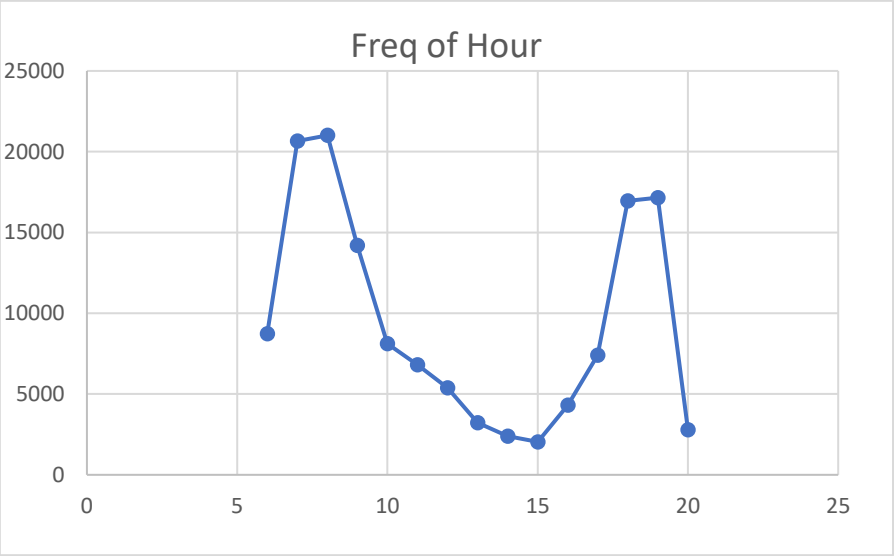

Fig (a)

b. The following is the column chart showing the month wise break of Revenue, which is highest in the month of January and lowest in the march, the new competitor to the business tanked its revenue in the February and which took a further dive in the month march, it was also noted that the revenue is significantly lower throughout the summer season

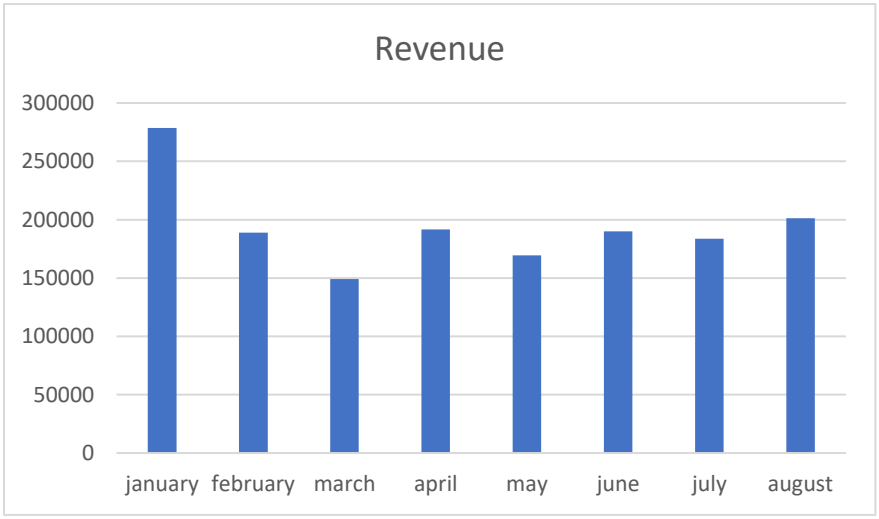

Fig (b)

c. The breakdown of number of new registerations every month from january to august can be observed below, the major spike is in the months of january and june, june was the time when monsoon discount were offered to the new members.The deepest descent is observed in the February which furthur plunged into summers.

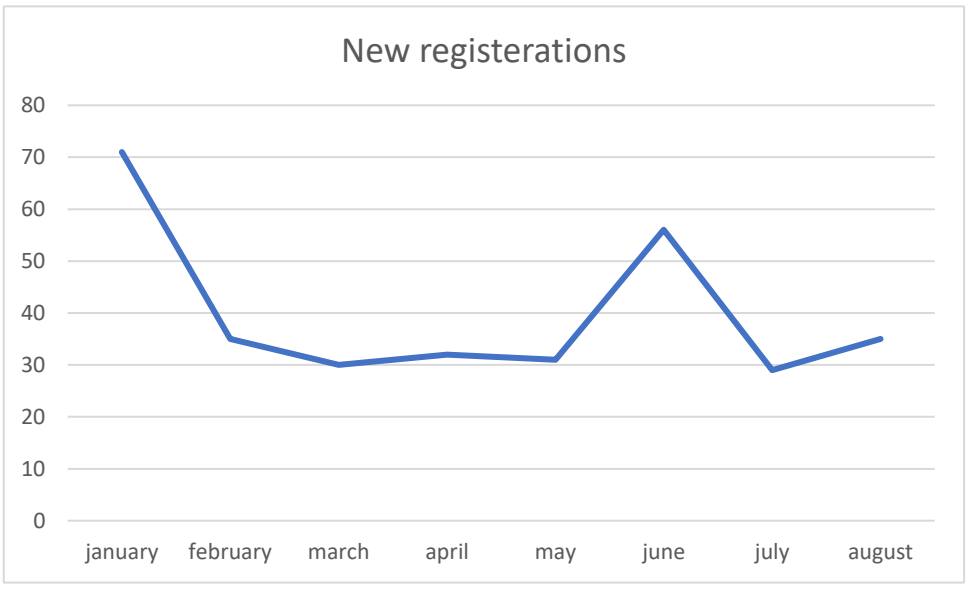

Fig (c)