Desarrollo de un dashboard en Power BI que permita caracterizar los usuarios de una cadena de gimnasios.

Andrés Mauricio Guzmán Reina
Especialización en Big Data, Fundación Universitaria Compensar
Proyecto fin de grado

Jairo Cesar Gomez Acero

20 de Noviembre de 2022

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Desarrollo de un dashboard en Power BI que permita caracterizar los usuarios de una cadena de gimnasios.

Andrés Mauricio Guzmán Reina

Trabajo de grado presentado como requisito parcial para optar al título de:

Especialista en Big Data

Director (a):

M.Sc., M.Eng, Ingeniero Jairo Cesar Gomez Acero

Fundación Universitaria Compensar

Facultad de Ingeniería, Ingeniería de Especialización en Big Data

Bogotá, Colombia

2022

Resumen

En este trabajo se pretende, hacer un recorrido por el proceso para crear un dashboard que pueda generar un mejor conocimiento de los usuarios de una cadena de gimnasios, para de esta manera permitir a la compañía tomar mejores decisiones con respecto a sus compañas de mercadeo.
Se selecciona el informe del ERP de la compañía que se cree más información sobre el usuario arroja.
Se desarrolla paso por paso el proceso ETL de estos datos.
Para finalizar con la construcción del dashboard.

Palabras clave: ETL, EDA, Dashboard, Power Bi.

Abstract

In this work, it is intended to take a tour of the process to create a dashboard that can generate a better knowledge of the users of a chain of gyms, in order to allow the company to make better decisions regarding its marketing companies.

The company's ERP report that creates the most information about the user is selected.

The ETL process of these data is developed step by step.

To finish with the construction of the board.

Keywords: ETL, EDA, Dashboard, Power Bi.

Contenido

PÁG.

RESUMEN ..... **V**

INTRODUCCIÓN **1**

ANTECEDENTES Y JUSTIFICACIÓN 1
FORMULACIÓN DEL PROBLEMA 2
OBJETIVO GENERAL 2
Objetivos Específicos *2*
ALCANCES Y LIMITACIONES 3

1. MARCO TEÓRICO **5**

1.1 SUBTÍTULOS NIVEL 2 ..... **¡ERROR! MARCADOR NO DEFINIDO.**
1.1.1 Subtítulos nivel 3 ..... ***¡Error! Marcador no definido.***

2. DISEÑO METODOLÓGICO **7**

3. RESULTADOS **20**

4. DISCUSIÓN **25**

5. CONCLUSIONES Y RECOMENDACIONES **27**

5.1 CONCLUSIONES 27
5.2 RECOMENDACIONES 27

BIBLIOGRAFÍA **29**

Introducción

La cadena de gimnasios hace uso de un software para la gestión de sus establecimientos, que almacena toda la información de sus usuarios y desearía poder extraer información de valor al obtener al obtener la caracterización de sus usuarios.

La idea de la cadena de gimnasios es obtener un mayor conocimiento de sus usuarios y poder lograr un mayor tiempo promedio de la duración de los contratos mensuales. Ademas poder hacer una inversión mejor orientada y segmentada para sus posibles nuevos clientes. Al tener un mayor conocimiento de sus usuarios, el costo por adquisición de cada usuaria en términos de marketing se podría optimizar y disminuir su valor.

Antecedentes y Justificación

La cadena de gimnasios es una compañía de multinacional de procedencia escandinava, abrió operaciones en Colombia en el año 2017. En la actualidad cuenta con aproximadamente 257 gimnasios distribuidos en 5 países en el mundo.
En Colombia cuenta con 10 establecimientos operativos. La mayor promesa venta de la cadena de Gimnasios es que sus usuarios pueden hacer uso del servicio las 24 horas de los siete días de la semana. No se cierran en ningún momento del año.

En la actualidad para la gestión de sus gimnasios la compañía hace uso del software PerfectGym de procedencia polaco. El Software es muy completo y tiene variedad de módulos que soporta la gestión de todos los gimnasios. EL software provee módulos como el POS punto de venta, portal del cliente, la aplicación para el cliente y el back office.

Dentro del back office hay diversidad de módulos: gestión de los usuarios, gestión de los contratos, inventarios, configuración de planes y productos, un dashboard configurable para hacer BI. Ademas un módulo de reportes donde se dispone de 129 diferentes informes que se pueden descargar.

Uno de los reportes nos permite conocer algunos datos de los usuarios que tienen o tuvieron alguna vez una afiliación con la compañía.

Formulación del Problema

En la actualidad la compañía obtiene miles de registros de sus usuarios, pero estos simplemente quedan como registros en la base de datos del proveedor del software.
La compañía considera que en el estudio de estos registros podría encontrase información relevante que permita logra una mayor fidelización de sus usuarios, mejorar la propuesta de valor o servicio que da en la actualidad a sus usuarios.
Ademas conocer más a fondo sus usuarios y de esta manera poder optimizar la inversión en marketing logrando disminuir el costo por adquisición que la compañía tiene en la actualidad.

Objetivo General

Desarrollar de un dashboard en Power BI que permita caracterizar los usuarios de una cadena de gimnasios.

Objetivos Específicos

- Identificar el reporte necesario del software que utiliza en la actualidad la compañía, para poder desarrollar el dashboard.
- Determinar categorías o variables de interés dentro de afiliados.
- Determinar el proceso ETL que se va a realizar al reporte de usuarios del gimnasio.
- Diseñar el dashboard en Power BI.

Alcances y Limitaciones

El objetivo en este trabajo es desarrollar un primer bosquejo de un dashboard que la compañía pueda usar para caracterizar a sus usuarios y de esta manera poder tomar mejores decisiones con respecto a sus campañas de marketing.

La compañía, ha autorizado el uso de sus datos, pero no así el uso de su nombre, teniendo en cuenta que este documento se va a subir a una base de datos indexada en internet, no es del interés de la compañía que al buscar sus nombres se encuentren datos que podrían ser sensibles, si sus competidores directos los conocen.

Teniendo en cuenta, que la compañía ha dado vía libre para hacer uso de sus datos, que en el proceso llevado a cabo en la especialización se ha aprendido nuevas herramientas y técnicas. Que ademas del costo relacionado al tiempo dedicado a la realización del trabajo o de recursos como internet, el equipo de cómputo con el que ya se cuenta. No hay ningún otro impedimento que pueda limitar que no se logre el objetivo final de este trabajo.

1. Marco teórico

El ministerio de salud y protección social, define la caracterización como " La caracterización poblacional consiste en un análisis de índole esencialmente descriptivo, puede realizarse desde distintos enfoques metodológicos, haciendo uso de fuentes primarias y secundarias de información, con el propósito de aproximarse al conocimiento y comprensión del tamaño, estructura, características." (link)

Para poder caracterizar la población de usuarios del gimnasio primero debemos cono llevar a cabo un análisis exploratorio de los datos **EDA.**

1.1 EDA

El análisis exploratorio de datos es una aproximación haciendo uso de la estadística descriptiva y herramientas gráficas, para lograr un mejor entendimiento del conjunto de datos. Es mayormente utilizado para maximizar el entendimiento de un conjunto de datos, detectar datos fuera del rango, anomalías y probar los supuestos. Es un primer paso robusto antes de la aplicación de cualquier otro método estadístico. (Estelle Camizuli,Emmanuel John Carranza, 2018, The Encyclopedia of Archaeological Sciences)

1.2 ETL

Extracción, transformación y carga (ETL) es una canalización de datos que se usa para recopilar datos de varios orígenes. A continuación, transforma los datos según las reglas de negocio y los carga en un almacén de datos de destino. El trabajo de transformación en ETL tiene lugar en un motor especializado y, a menudo, implica el uso de tablas de almacenamiento provisional para conservar los datos temporalmente a medida que estos se transforman y, finalmente, se cargan en su destino.

La transformación de datos que tiene lugar a menudo conlleva varias operaciones como filtrado, ordenación, agregación, combinación de datos, limpieza de datos, desduplicación y validación de datos.

Frecuentemente, las tres fases del proceso ETL se ejecutan en paralelo para ahorrar tiempo. Por ejemplo, mientras se extraen datos, puede que esté funcionando un proceso de transformación sobre los datos ya recibidos y de preparación para la carga, y puede que empiece a funcionar un proceso de carga sobre los datos preparados, en lugar de tener que esperar a que termine todo el proceso de extracción. (link)

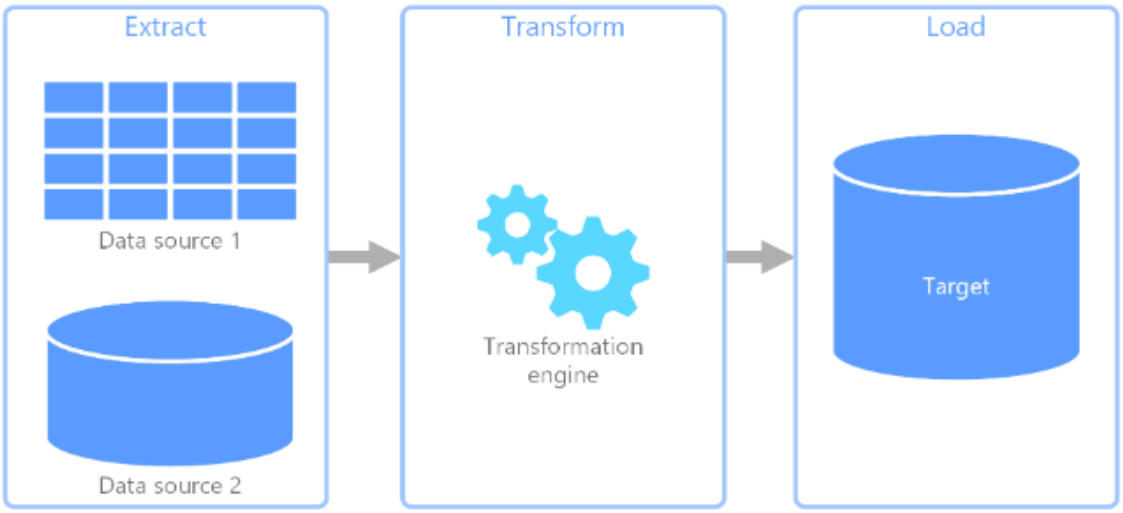


Figura 1. Proceso ETL, tomado de https://learn.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl

1.3 Power BI.

Power BI es un servicio de análisis de datos de Microsoft orientado a proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial (en inglés: business intelligence o BI) con una interfaz lo suficientemente simple como para que los usuarios finales puedan crear por sí mismos sus propios informes y paneles. (link)

2. Diseño Metodológico

2.1 Determinar el reporte a usar.

Al revisar el software para la gestión de usuarios del que hace uso la compañía. Se encuentran 129 diferentes reportes. Luego de examinar los reportes que tienen que ver con la información de usuarios. Se determina que el reporte que mayor cantidad de información relevante entrega sobre los usuarios es el reporte “All Members”.

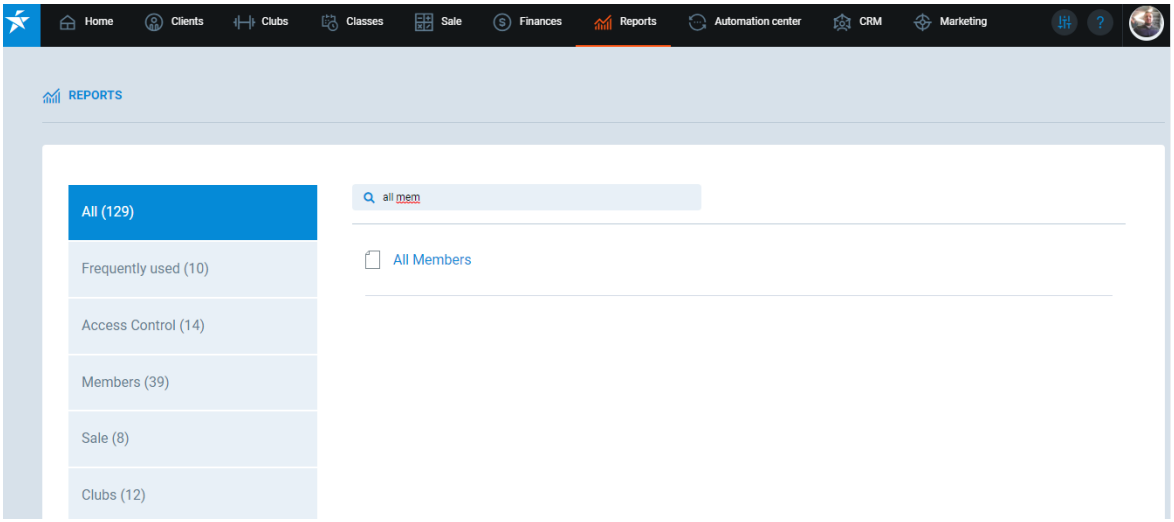


Figura 2. Reportes disponibles en PerfectGym. Construcción propia.

Se procede a generar el reporte sin aplicar ningún filtro de los disponibles.

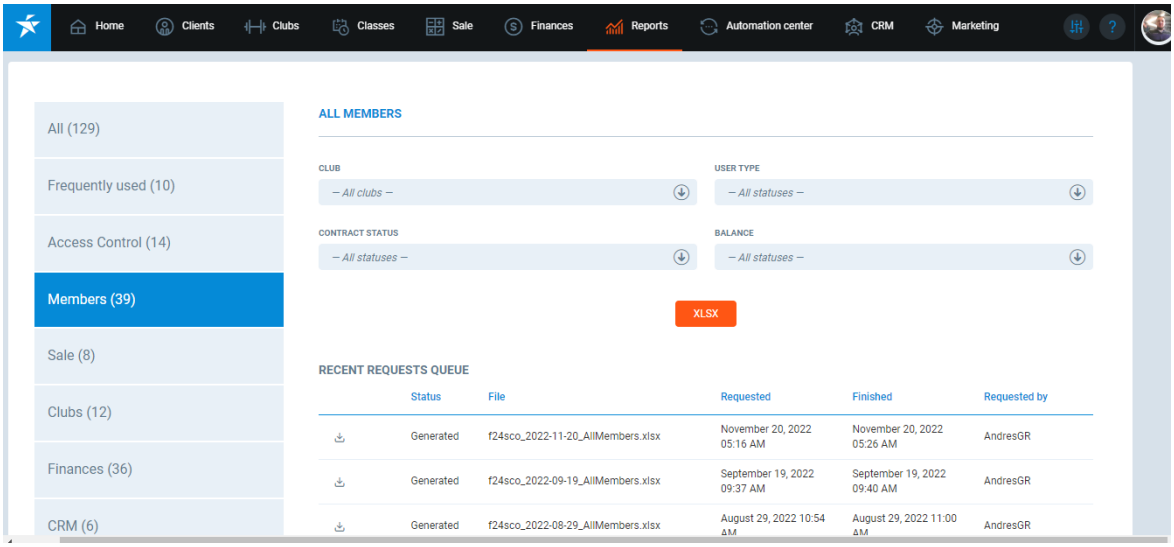


Figura 3. Generación del reporte de todos los usuarios. Construcción propia.

El reporte descargado en formato Excel tiene un tamaño de 43.199 KB y 104.554 registros

2.2 EDA de los datos en el reporte.

Para realizar una primera aproximación a los datos del reporte, convertimos el reporte de Excel a un archivo CSV, lo cargamos en Google colab y haciendo uso de Pandas. Obtenemos la siguiente información.

```
import pandas as pd
df=pd.read_csv("reporte.csv")
df.info()

RangeIndex: 104554 entries, 0 to 104553
Data columns (total 88 columns):
 #   Column                                  Non-Null Count   Dtype
---  ------                                  --------------   -----
 0   Club                                    104554 non-null  object
 1   Club Number                             104554 non-null  int64
 2   User number                             104554 non-null  int64
 3   UserId                                  104554 non-null  int64
 4   Full name                               104554 non-null  object
 5   First name                              104554 non-null  object
 6   SecondName                              226 non-null     object
 7   Last Name                               104554 non-null  object
 8   Birth date                              104506 non-null  object
 9   Age                                     104506 non-null
float64
 10  Gender                                  104480 non-null  object
 11  Email                                   104480 non-null  object
 12  PhoneMobile                             104502 non-null  object
13  Phone                                   12 non-null      float64
14  Street                                  103998 non-null  object
15  Additional address                      272 non-null     object
```

```
16  Post code                           32832 non-null   object
17  City                                104509 non-null  object
18  Country                             104517 non-null  object
19  Created date                        104554 non-null  object
20  Personal ID                         104298 non-null  object
21  Is active                           104554 non-null  object
22  User type                           104554 non-null  object
23  Is foreigner                        104554 non-null  object
24  Preferred language                  42281 non-null   object
25  Preferred contact                   900 non-null     object
26  Recommended by                      2755 non-null    object
27  Consultant employee                 103563 non-null  object
28  Designation                         0 non-null       float64
29  Status                              8 non-null       object
30  Last visit                          95766 non-null   object
31  Company name                        78 non-null      object
32  Tax Id                              7 non-null       object
33  Has photo                           104554 non-null  object
34  Corpo Portal Login                  1 non-null       object
35  GroupNames                          0 non-null       float64
36  Exclude From E-Payment              16 non-null      object
37  Exclude From E-Payment Until Date   7 non-null       object
38  Total days as a member              104554 non-null  int64
39  Perfect Member                      104554 non-null  int64
40  Source                              104554 non-null  object
41  Biometric agreement                 104554 non-null  object
42  Save in the club                    36701 non-null   object
43  Tags                                2 non-null       object
44  Nation                              0 non-null       float64
45  Number of related                   0 non-null       float64
46  Master member number                0 non-null       float64
47  Membership Status                   104162 non-null  object
48  Signup date                         104162 non-null  object
49  Start date                          104162 non-null  object
50  End date                            80739 non-null   object
51  Cancel date                         80738 non-null   object
52  Cancel reason                       78156 non-null   object
53  Payment interval                    104162 non-null  object
54  Min contract time                   104162 non-null  object
55  Min contract time date              104162 non-null  object
56  Membership fee                      104162 non-null  float64
57  Member. fee net                     104162 non-null  object
58  Access rules                        104162 non-null  object
59  Payment Plan Name                   104162 non-null  object
60  Payment plan type                   104162 non-null  object
61  Payment plan category name          0 non-null       float64
62  Contract discount                   43254 non-null   object
63  Cession from                        256 non-null     float64
64  Cession to                          269 non-null     float64
65  Upgrade                             0 non-null       float64
66  AddOns                              5 non-null       object
67  Payment method                      104554 non-null  object
68  Financial status                    104554 non-null  object
```

```
 69  Contract Balance                    104554 non-null  object
 70  PosBalance                          104554 non-null  int64
 71  POS Prepaid Amount                  104554 non-null  int64
 72  POS Debit Amount                    104554 non-null  int64
 73  DateOfDebt                          30151 non-null   object
 74  Intervals in debt                   104027 non-null  float64
 75  First debt                          30150 non-null   object
 76  Bank name                           44936 non-null   object
 77  Bank account                        1 non-null       object
 78  Credit card                         86611 non-null   object
 79  CreditCardPublicNumber              86611 non-null   float64
 80  CC Debit/credit                     104554 non-null  object
 81  Virtual account number              0 non-null       float64
 82  Last provider message               34334 non-null   object
 83  Is paid by different person         0 non-null       float64
 84  Substitute payer                    0 non-null       float64
 85  Debt collection stage               0 non-null       float64
 86  ChildrenCount                       104554 non-null  int64
 87  FamilyList                          0 non-null       float64
dtypes: float64(19), int64(9), object(60)
     memory usage: 70.2+ MB
```

Tenemos un total de 87 diferentes columnas con 104.554 registros.
Se decide hacer uso de las columnas, que en un primer análisis se cree podrían arrojar información valiosa para la caracterización de los usuarios.

- Club Number
- Age
- Gender
- City
- Total days as a member
- Access rules
- Financial status

```
df1 =df[["Club Number", "Age", "Gender", "City", "Total days as
 a member", "Access rules", "Financial status", ]]
df1.describe()
```

	Club Number	Age	Total days as a member
count	104554.000000	104506.000000	104554.000000
mean	57105.910888	31.476633	259.531706

std	3.460191	9.768980	243.098573
min	57100.000000	1.000000	0.000000
25%	57103.000000	24.000000	97.000000
50%	57106.000000	29.000000	181.000000
75%	57108.000000	36.000000	350.000000
max	57112.000000	122.000000	1827.000000

Tabla 1. Estadística descriptiva de las variables continuas del Dataframe Elegido.

Aunque la columna Club Number es de tipo numérico no es una variable continua, si no categórica, por lo tanto los resultados estadísticos obtenidos para esta variable no revelan nada.

- Para la edad de los 104506 usuarios tenemos que
    - El promedio de edad de los usuarios del Gym es de 31.4 años.
    - La edad mínima es de 1 año, lo cual debe ser un error teniendo en cuenta que el gimnasio no recibe menores de edad.
    - El 25 por ciento de los usuarios son menores de 24 años.
    - El 50 por ciento de los usuarios son menores de 29 años.
    - El 75 por ciento de los usuarios son menores de 36 años.

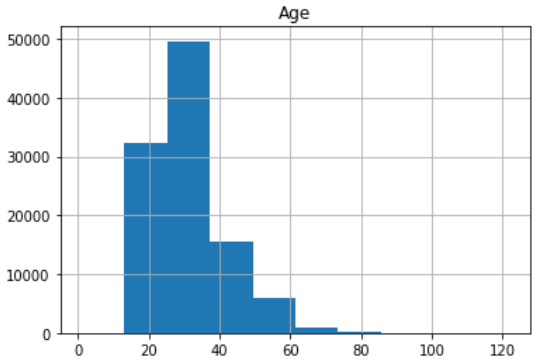


Figura 4, Histograma de la distribución de edades para los usuarios del gimnasio. Construcción propia.

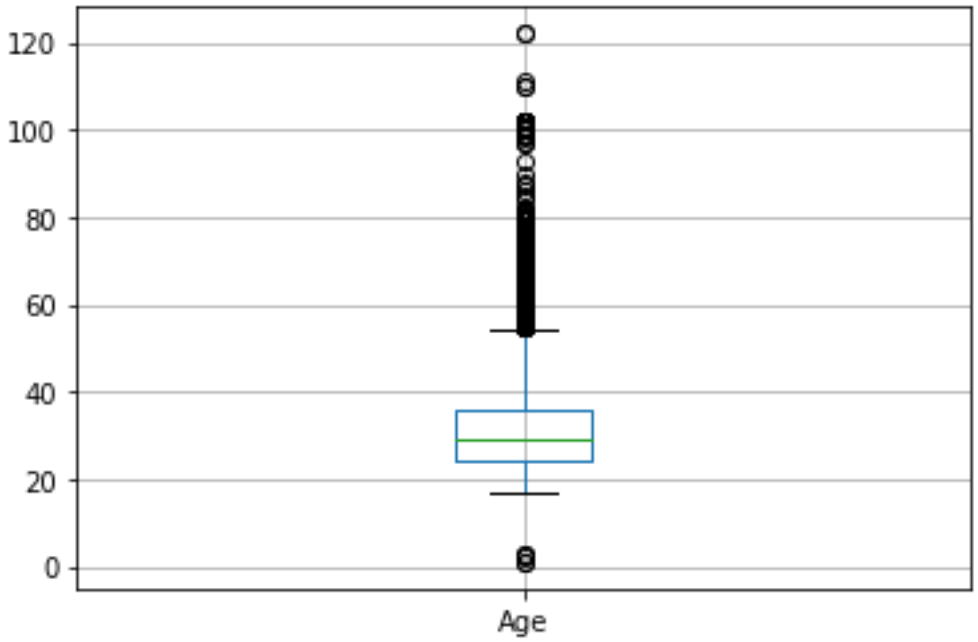


Figura 5. Diagrama de cajas y bigotes para la edad de los usuarios del gimnasio. Construcción propia.

- Para el tiempo que han sido miembros del gimnasio tenemos:
    - El tiempo promedio de los usuarios en el gimnasio es de 259 días.
    - El 25 por ciento de los usuarios ha durado en el gimnasio menos de 97 días.
    - El 50 por ciento de los usuarios ha durado en el gimnasio menos de 181 días.
    - El 75 por ciento de los usuarios ha durado en el gimnasio menos de 350 días.

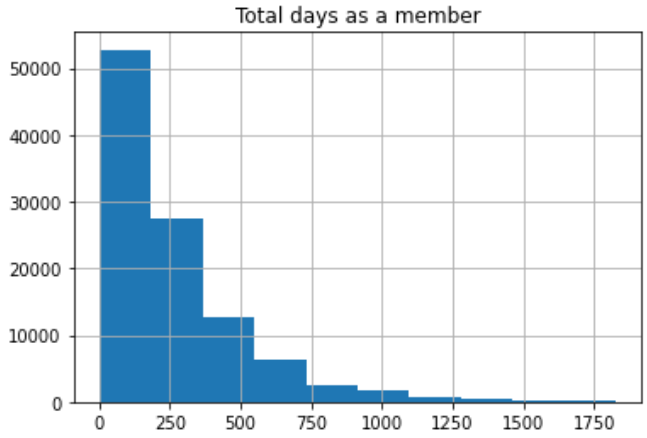


Figura 6. Histograma de la distribución de los días que los usuarios han durado con su contrato. Construcción propia.

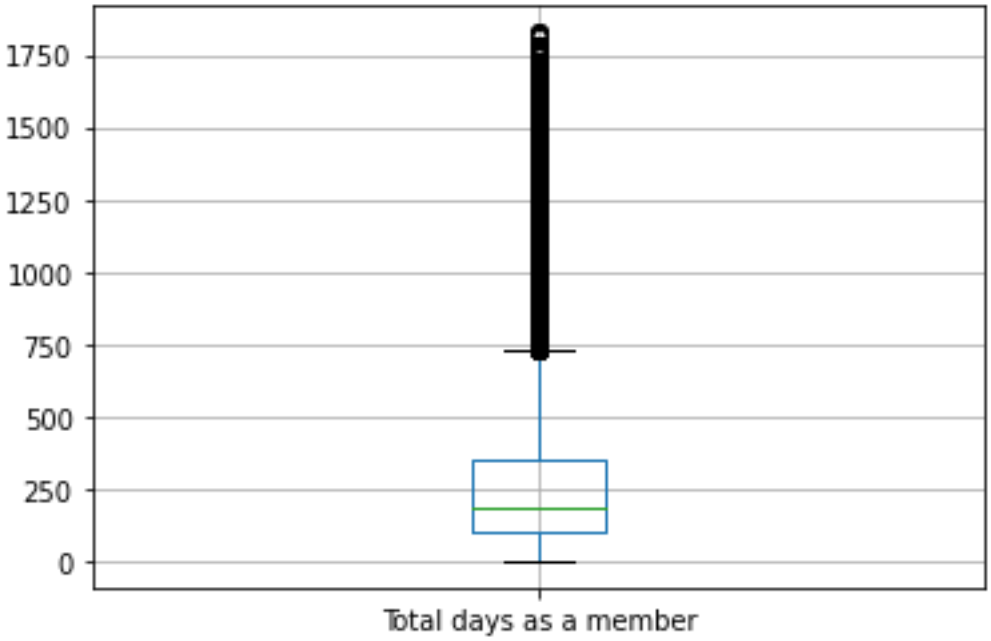


Figura 7. Diagrama de cajas y bigotes para los días que los usuarios han durado con su contrato. Construcción propia.

2.3 ETL

Teniendo en cuenta que la idea es hacer uso de Power Bi para generar el Dashboard se decide realizar el proceso de ETL en powerquery.

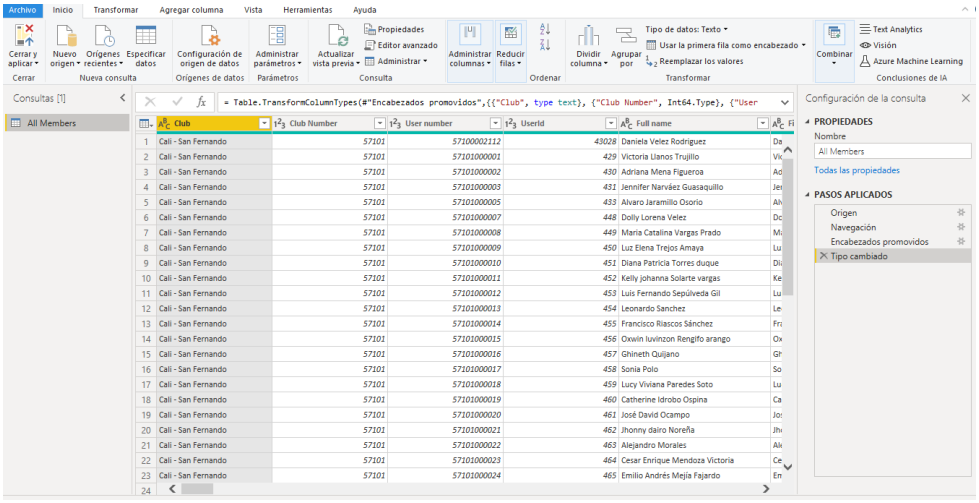


Figura 8. Primera carga de los datos en Power Query. construcción propia.

1. Ya que en el análisis exploratorio de los datos se había decidido cuales variables iban a ser usadas para la caracterización. Se eliminan el resto de columnas.

Quedándonos con las siguientes columnas.

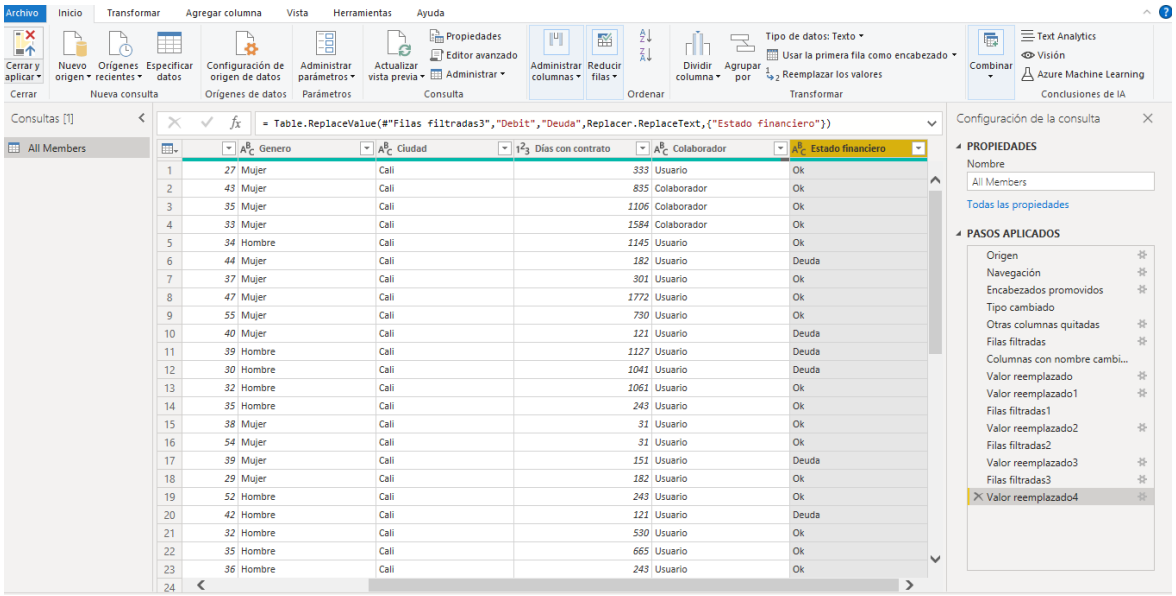
- Club Name. En esta columna está el nombre de la sede a la que asiste el usuario. Son 11 categorías.
- Age. En esta columna esta la edad del usuario.
- Gender. En esta columna esta el genero del usuario, Son 2 categorías Female y Male.
- City. En esta columna esta la ciudad del usuario.
- Total days as a member. En esta columna nos indica durante cuantos días ha tenido un contrato el usuario.
- Access rules. Esta columna de tipo categórica, nos permite determinar si el usuario es un colaborador de la compañía categoría “Staff Restricted” o cliente “24-7”.
- Financial status. Esta columna de tipo categórica nos permite determinar si el usuario tiene o no deuda con el gimnasio. Deuda “Debit” sin deuda “Ok”.

2. Como en el análisis exploratorio de datos se encontraron usuarios que supuestamente son menores de edad y el gimnasio por sus políticas no admite estos usuarios. Se eliminan estos usuarios de los datos.

3. Se cambia el nombre de las columnas a nombres que nos den mas información y en español.

Club Name - > Sede
Age -> Edad
Gender -> Genero
City - > Ciudad
Total days as a member -> Días con contrato.
Access Rules -> Colaborador.
Financial status-> Estado financiero.

4. Se cambian las categorías de genero de ingles a español. Male - > Hombre, Female -> Mujer.
5. Se cambian las categorías en la columna colaborador 24-7 -> Usuario , Staff restricted -> Colaborador .
6. Se transforman las categorías para el estado financiero de la siguiente manera. Debit -> Deuda.



	Genero	Ciudad	Días con contrato	Colaborador	Estado financiero
1	27 Mujer	Cali	333	Usuario	Ok
2	43 Mujer	Cali	835	Colaborador	Ok
3	35 Mujer	Cali	1106	Colaborador	Ok
4	33 Mujer	Cali	1584	Colaborador	Ok
5	34 Hombre	Cali	1145	Usuario	Ok
6	44 Mujer	Cali	182	Usuario	Deuda
7	37 Mujer	Cali	301	Usuario	Ok
8	47 Mujer	Cali	1772	Usuario	Ok
9	55 Mujer	Cali	730	Usuario	Ok
10	40 Mujer	Cali	121	Usuario	Deuda
11	39 Hombre	Cali	1127	Usuario	Deuda
12	30 Hombre	Cali	1041	Usuario	Deuda
13	32 Hombre	Cali	1061	Usuario	Ok
14	35 Hombre	Cali	243	Usuario	Ok
15	38 Mujer	Cali	31	Usuario	Ok
16	54 Mujer	Cali	31	Usuario	Ok
17	39 Mujer	Cali	151	Usuario	Deuda
18	29 Mujer	Cali	182	Usuario	Ok
19	52 Hombre	Cali	243	Usuario	Ok
20	42 Hombre	Cali	121	Usuario	Deuda
21	32 Hombre	Cali	530	Usuario	Ok
22	35 Hombre	Cali	665	Usuario	Ok
23	36 Hombre	Cali	243	Usuario	Ok

Figura 9. Captura de Power Query donde se evidencia que se realizaron 11 transformaciones al conjunto de datos.

2.4 Construcción del Dashboard.

- Se agrega un grafico de torta que representa el porcentaje de colaboradores o usuarios.

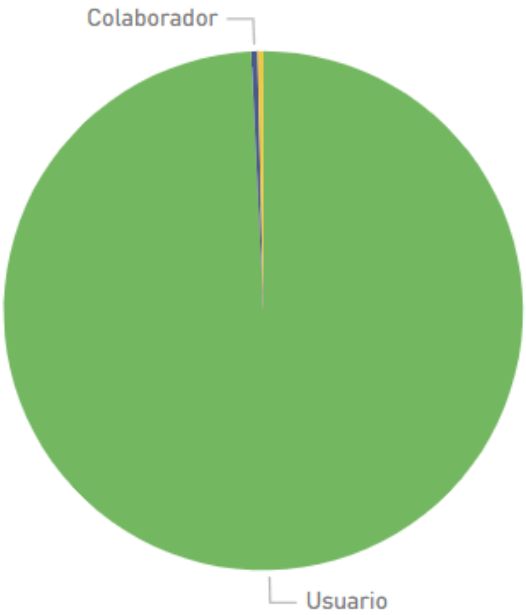


Figura 10. Visualización diagrama de torta para el tipo de usuario. Construcción propia.

- Se agrega un grafico de torta que representa el porcentaje de usuarios según se estado financiero.

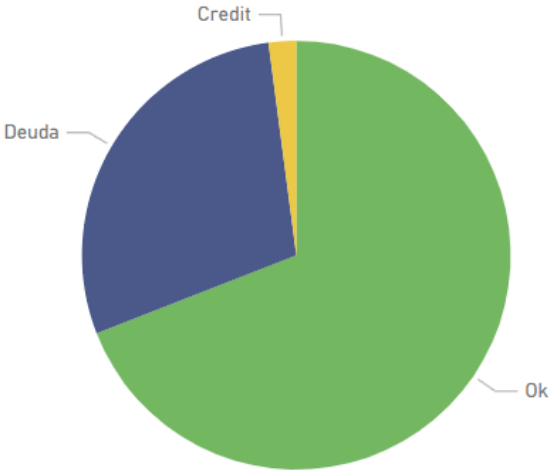


Figura 11. Visualización diagrama de torta para el estado financiero. Construcción propia.

- Se agrega un grafico de torta que representa el porcentaje de usuarios según el genero.

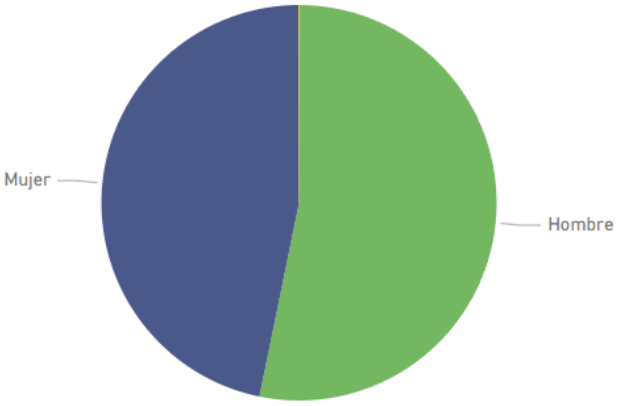


Figura 12. Visualización diagrama de torta para el genero. Construcción propia.

- Haciendo uso del objeto visual Script de R y previamente habiendo generado un índice, se genera un histograma para la distribución de la edad.

```
hist(dataset$Edad, main=" Histograma de la edad", xlab="Edad",
ylab="Frecuencia")
```

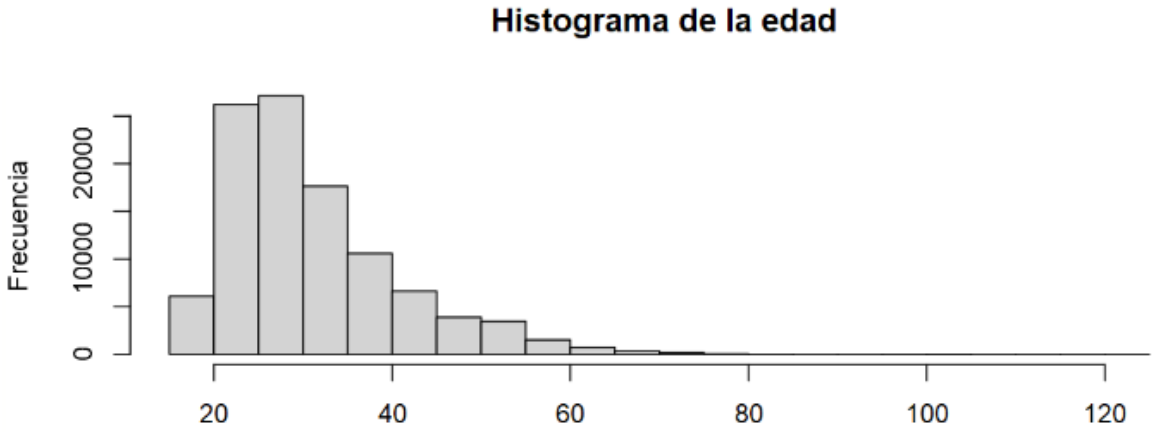


Figura 13 Visualización en Power Bi con script en R histograma para la edad de los usuarios. Construcción propia.

- Haciendo uso del objeto visual Script de R y previamente habiendo generado un índice, se genera un histograma para la distribución de los días con contrato.

```
hist(dataset$'Días con contrato', main=" Histograma Dias con
contrato", xlab="Días", ylab="Frecuencia")
```

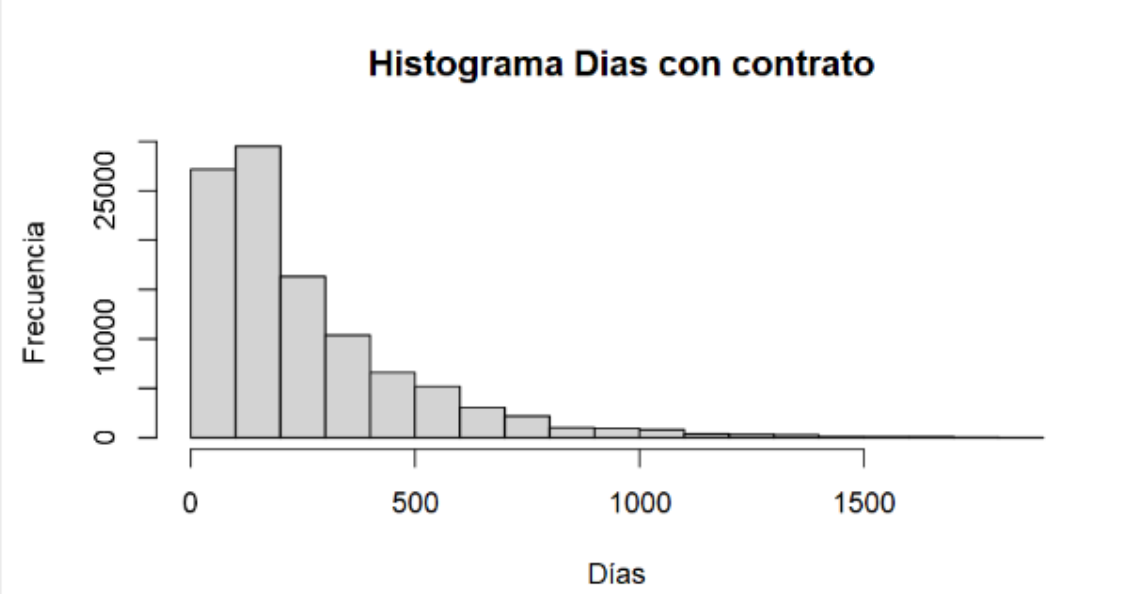


Figura 14. Visualización en Power Bi con script en R histograma para el tiempo con contrato. Construcción propia.

- Se agregaron segmentaciones de datos del tipo lista para las categorías.
  - Sede
  - Genero
  - Colaborador
  - Estado Financiero

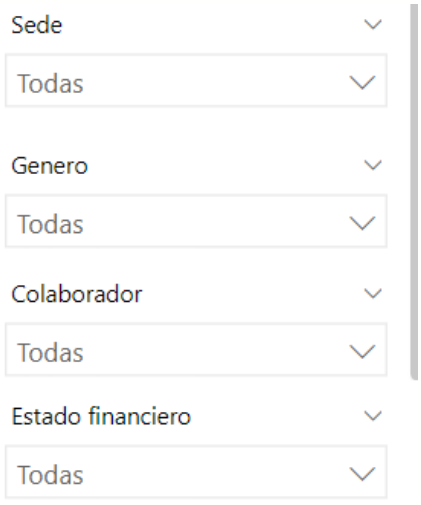


Figura 15. Segmentación de datos por categorías. Construcción propia.

- Haciendo uso de DAX se genero la medida para el promedio edad y este se agrego como un objeto de visualización tipo tarjeta al dasboard.

```
Promedio edad = AVERAGE('All Members'[Edad])
```



Figura 16. Visualización tipo tarjeta, con la medida creada. Construcción propia.

- Haciendo uso de DAX se genero la medida para el promedio de duración del contrato.

```
promedio días con contrato = AVERAGE('All Members'[Días con contrato])
```

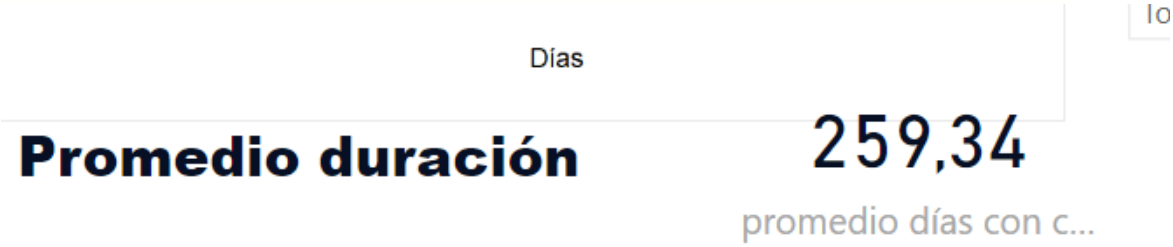


Figura 17. Visualización tipo tarjeta, con la medida creada. Construcción propia.

3. Resultados

Al hacer uso de la herramienta construida podemos rápidamente obtener los siguientes resultados.

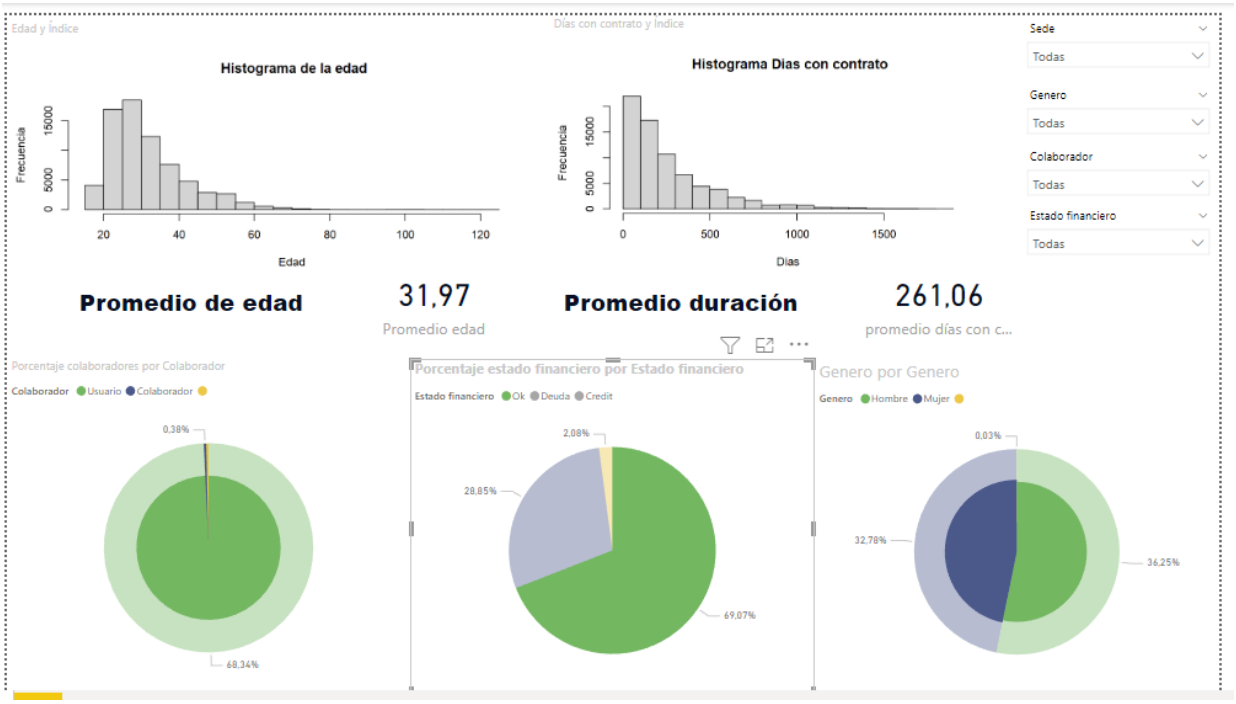


Figura 18. Dashboard. Construcción propia.

- La compañía de gimnasios:
    - Tiene un promedio de edad de sus usuarios de 31 años.
    - Promedio de duración de los usuarios 259 días.
    - 52.83 % de Hombres y 46.38 % de Mujeres.
    - 68.34% de los usuarios no tienen deuda y 28.85% tienen deuda.
    - 99.26 % de los usuarios son clientes y el 0.38% son colaboradores.

- La sede 20 de Julio
    - Tiene un promedio de edad de sus usuarios de 30 años.
    - Promedio de duración de los usuarios 237 días.
    - 53.34 % de Hombres y 46.1 % de Mujeres.

- 59.49% de los usuarios no tienen deuda y 36.29% tienen deuda.
- 99.52 % de los usuarios son clientes y el 0.31% son colaboradores.

- La sede Chapinero
  - Tiene un promedio de edad de sus usuarios de 32 años.
  - Promedio de duración de los usuarios 247 días.
  - 58.41 % de Hombres y 40.41 % de Mujeres.
  - 70.89% de los usuarios no tienen deuda y 26.67% tienen deuda.
  - 98.88 % de los usuarios son clientes y el 0.8% son colaboradores.

- La sede Metro Recreo
  - Tiene un promedio de edad de sus usuarios de 31 años.
  - Promedio de duración de los usuarios 28 días.
  - 52.27 % de Hombres y 47.54 % de Mujeres.
  - 95.27% de los usuarios no tienen deuda y 3.75% tienen deuda.
  - 98.86 % de los usuarios son clientes y el 0.14% son colaboradores.

- La sede Suba
  - Tiene un promedio de edad de sus usuarios de 30 años.
  - Promedio de duración de los usuarios 242 días.
  - 53.29 % de Hombres y 46.06 % de Mujeres.
  - 62.06% de los usuarios no tienen deuda y 36.49% tienen deuda.
  - 99.42 % de los usuarios son clientes y el 0.3% son colaboradores.

- La sede La 27
  - Tiene un promedio de edad de sus usuarios de 29 años.
  - Promedio de duración de los usuarios 268 días.
  - 52.46 % de Hombres y 46.56 % de Mujeres.
  - 69.16% de los usuarios no tienen deuda y 27.9% tienen deuda.
  - 99.06 % de los usuarios son clientes y el 0.67% son colaboradores.

- La sede La Florida
  - Tiene un promedio de edad de sus usuarios de 30 años.
  - Promedio de duración de los usuarios 273 días.
  - 50.8 % de Hombres y 48.75 % de Mujeres.
  - 71.88% de los usuarios no tienen deuda y 26.14% tienen deuda.
  - 99.56 % de los usuarios son clientes y el 0.24% son colaboradores.

- La sede Ciudad Jardín.
  - Tiene un promedio de edad de sus usuarios de 31 años.
  - Promedio de duración de los usuarios 287 días.
  - 52.91 % de Hombres y 47.04 % de Mujeres.
  - 74.12% de los usuarios no tienen deuda y 26.49% tienen deuda.

  - 99.17 % de los usuarios son clientes y el 0.44% son colaboradores.

- La sede Plaza Norte.
  - Tiene un promedio de edad de sus usuarios de 32 años.
  - Promedio de duración de los usuarios 283 días.
  - 54.58 % de Hombres y 45.38 % de Mujeres.
  - 68.98% de los usuarios no tienen deuda y 30% tienen deuda.
  - 99.3 % de los usuarios son clientes y el 0.5% son colaboradores.

- La sede San Fernando
  - Tiene un promedio de edad de sus usuarios de 33 años.
  - Promedio de duración de los usuarios 278 días.
  - 55.19 % de Hombres y 44.74 % de Mujeres.
  - 71.51% de los usuarios no tienen deuda y 27.24% tienen deuda.
  - 99.42 % de los usuarios son clientes y el 0.3% son colaboradores.

- La sede Santa Monica
  - Tiene un promedio de edad de sus usuarios de 33 años.
  - Promedio de duración de los usuarios 270 días.
  - 53.26 % de Hombres y 46.65 % de Mujeres.
  - 70.45% de los usuarios no tienen deuda y 28.23% tienen deuda.
  - 99.18 % de los usuarios son clientes y el 0.57% son colaboradores.

- La sede Chia.
  - Tiene un promedio de edad de sus usuarios de 30 años.
  - Promedio de duración de los usuarios 259 días.
  - 51.69 % de Hombres y 48.2 % de Mujeres.
  - 66.24% de los usuarios no tienen deuda y 31.97% tienen deuda.
  - 99.25 % de los usuarios son clientes y el 0.42% son colaboradores.

- La edad promedio de las mujeres es de 31 años.
- La edad promedio de los hombres es de 31 años.
- La edad promedio de los usuarios en deuda es de 30.07 años
- La edad promedio de los usuarios que están al día con la compañía 31.97 años.
- El porcentaje de mujeres en deuda es de 12.86 %
- El porcentaje de hombres en deuda es de 15.97%

4. Discusión

Al analizar los datos obtenidos haciendo uso de la herramienta diseñada para este ejercicio. Podemos plantear algunas hipótesis.

- Los deudores tienden a ser aquellas personas que tienen menor edad.
- La compañía tiene un nicho destendido y que podría ser un nicho objetivo en usuarios de edades mayores a los 36 años. El 75% de sus clientes se tiene 36 años o menos.
- En mayor porcentaje las mujeres cumplen con sus deudas.
- La compañía tiene áreas exclusivas para el uso de mujeres dentro de sus gimnasios, pero porcentualmente la cantidad de mujeres que se han afiliado es menor. La compañía esta desaprovechando la inversión grande que hace en generar gimnasios exclusivos de mujeres, si aun así es mas la cantidad de hombres que de mujeres en sus afiliaciones.
- Las sedes ubicadas en la región de Santander, tienen menor promedio de edad entre sus usuarios.
- Las sedes ubicadas en el Valle del Cauca tiene el mayor promedio de edad entre sus usuarios.
- La sede ciudad Jardín. en la cual sus usuarios han tenido sus contratos por mas tiempo 287 días en promedio, es también la sede que tiene el mayor porcentaje de contratos que no se encuentra en deuda.

5. Conclusiones y recomendaciones

5.1 Conclusiones

- Una vez realizado el proceso de verificar los distintos reportes disponibles en la compañía se define que el reporte, que permite de mejor manera caracterizar a los usuarios del gimnasio es el que en el software de gestión de los gimnasios tiene el nombre de “All members”.
- Luego de realizar el análisis exploratorio de datos se puede concluir que son de interés únicamente 7 variables de las 87 que contiene el informe.
- La elección de la o las herramientas de ETL está muy ligada al software que se va a usar para el análisis final de los datos. En este caso la mejor opción por compatibilidad con el software Power Bi, fue usar su modulo Power Query para hacer la Extracción, transformación, limpieza y carga de la información.
- El software Power Bi es muy versátil, al punto de permitirnos generar imágenes haciendo uso de Scripts de R.

5.2 Recomendaciones

Teniendo en cuenta el proceso realizado, que el software tiene otros reportes que podrían ser interesantes y que se podrían cruzar con el informe escogido para este trabajo, se podría profundizar y obtener nueva información, haciendo uso de informes como las asistencia al gimnasio, la asistencia a clases grupales.

Bibliografía

Murillo, Javier . *Recomendaciones para escribir un ensayo normas APA.*

MINTIC, (2018), *Herramientas de analítica para la explotación de datos*

Hoyos Jorge, Aponte Fredy (2019) *Caracterización de los estudiantes de una institución de educación superior mediante big data*

Histogramas y Boxplots con Power BI (https://www.youtube.com/watch?v=vRXR_gQr4gs)

How to stop Power BI from removing duplicates when creating a data.frame using R script visual? (https://stackoverflow.com/questions/43725941/how-to-stop-power-bi-from-removing-duplicates-when-creating-a-data-frame-using-r)

Zapata Jose. *PANDAS - Manipulacion de Datos con Python (link)*