

1 Repositorio
2 Presentación y problemática (desafío)
3 Descripción del Dataset
4 Limpieza y preprocesamiento de datos
5 Modelo de Predicción de costos
6 Conclusiones

Modelo Predictivo para entidades que contratan servicios de Call Center con terceros basado en datos de Normas Estandar COPC

Code ▾

Adrian Felipe Pinzon Hurtado (apinzonh)



1 Repositorio

https://github.com/adrite20/PEC3_TFM_adpinzon

2 Presentación y problemática (desafío)

La gran mayoría de las empresas contratan los servicios de call center con un externo (proveedor). Actualmente, con el auge de los bots, muchos de estos proveedores ofrecen un paquete que incluye atención de agentes humanos y bots. Estos nuevos paquetes de facturación se ofrecen en rangos de contactos mensuales llamados líneas bases, los cuales tienen un precio dependiendo del rango que estés dispuesto a pagar y también los promedio de contactos que se recibe mensualmente.

La mayoría de las empresas necesitan un modelo estadístico que les pueda dar ese precio ideal teniendo en cuenta los datos con los que cuentan, que es el costo actual de su call center, los indicadores COPC y lo que está atendiendo por bots. En la mayoría de estos contratos, las empresas contratan el servicio sin tener en cuenta número de asesores humanos, como sí lo hace el proveedor, quien tiene para eso modelos muy maduros en el mercado como Erlang o ARIMA. Esto debido que la administración de ese personal no es core de las empresas, más sí de los call center.

El dataset cuenta con un histórico de estas facturaciones mensuales por call center, junto con todos los indicadores COPC, eventos masivos, número de contactos por bots, casos por masivos y disponibilidad de los canales. El proveedor contractualmente jamás revela su número de asesores, porque es una estrategia de ahorro de costos que utilizan. Por ejemplo, pueden poner un asesor para que conteste hasta tres chats simultáneos o una llamada y un chat simultáneo, ya que este último canal tiene un mayor ASA, es decir, el usuario puede esperar un poco más a que le contesten (en esta data, está convenido hasta 60 segundos para contestar por chat).

El reto es poder predecir una nueva tabla de precios ideales para negociar un RFP (Request for Proposals) o solicitud de propuestas.

3 Descripción del Dataset

Se relacionan a continuación la descripción de cada variable:

TABLA RESUMEN VARIABLES DATA 'TEST'

Variable	Descripción
Mesa	(tipo factor) Mesa (tecnología u operaciones). Variable categórica discreta
Fecha	(tipo fecha) Fecha del registro de la información. Ej.: 08/2016. Variable continua
Contactos_Atendidos_telefono (*)	(tipo entero) Contactos atendidos en el canal teléfono. Ej.: 45739. Variable continua
Contactos_Atendidos_Chat (*)	(tipo entero) Contactos atendidos en el canal Chat. Ej.: 15013. Variable continua
Contactos_Atendidos_Bot (*)	(tipo entero) Contactos atendidos por el bot. Ej.: 2618. Variable continua
Contactos_Atendidos_autoreporte_gestion_indirecta (*)	(tipo entero) Contactos atendidos por autoreporte (web). Ej.: 4298. Variable continua
Contactos_ofrecidos (*)	(tipo entero) Contactos que entraron (contestados o no). Ej.: 79066. Variable continua
Contestados_Chat < Umbral (*)	(tipo entero) Contactos contestados en Chat < Umbral, es decir, contestados antes de los 60 seg desde que entró al chat. Ej.: 9667. Variable continua
Contestado_Tel < Umbral (*)	(tipo entero) Contactos contestados en teléfono < Umbral, es decir, contestados antes de los 25 seg desde que entró al chat. Ej.: 34158 Variable continua
AHT_Tel_min (*)	(tipo date) AHT en teléfono (tiempo que duró una llamada contestada en MM:SS). Ej.: 12:15 . Variable continua
AHT_Chat_min (*)	(tipo date) AHT en Chat (tiempo que duró una llamada contestada en MM:SS). Ej.: 16:15. Variable continua
SD06_ASA_Correo (*)	(tipo entero) Tiempo de respuesta (Average Speed of Answer). Indica cuánto tiempo esperó un cliente para ser atendido. Ej: 11.79 segundos. Variable continua
gestion_autoreporte (*)	(tipo doble) Tiempo en segundos para Gestión de autoreporte. Ej: 13. Variable continua
SD05_Disponibilidad_de_la_mesa_de_servicio (**)	(tipo doble) Indicador que mide cuánto estuvo un componente tecnológico indisponible. Ejemplo: canal de teléfono, canal de chat, bot, web autoreporte. La meta debe ser 100%. Fórmula: [1-sumatoria(Tiempo componente indisponible*peso componente)/total tiempo]. Ej: 100% o 1. Variable continua
Numero de masivos	(tipo entero) Número de eventos masivos que afectaron el servicio expresado en contactos. Ej: 1740 contactos por masivos
Línea Base	(tipo entero) Número de contactos base para ese periodo (lo que se esperaba atender contractualmente). Ej: 64308 contactos. Variable continua
AHT Tel en segundos (*)	(tipo entero) (Average Handle Time o Average Handling Time) Teléfono en segundos, es un KPI que nos ayuda a entender el tiempo medio de las interacciones con clientes en un contact center. Ej: 725 segundos. Variable continua
AHT Chat en segundos (*)	(tipo entero) (Average Handle Time o Average Handling Time) Chat en segundos, es un KPI que nos ayuda a entender el tiempo medio de las interacciones con clientes en un contact center. Ej: 975 segundos. Variable continua
Costo Mesa	(tipo doble) Costo del servicio. Ej: COP 660.029.876.86. Variable continua

() Corresponden a norma COPC. (**) Corresponde a metodología ITIL.

Variable	Descripción
Costo adicionales en mesa	(tipo doble) Costo por otros conceptos (Ej: mejora bot, nueva infra). Ej: COP 717.560.654,99. Variable continua
AHT Autoreporte(web) (*)	(tipo entero) (Average Handle Time o Average Handling Time) de web en segundos. Ej: 722 segundos. Variable continua

() Corresponden a norma COPC. (**) Corresponde a metodología ITIL.

4 Limpieza y preprocesamiento de datos

ID	Mesa	Fecha	Contactos_Atendidos_telefono	Contactos_Atendidos_Chat	Contactos_Atendidos_Chatbot	Contactos_autoreporte_gestion_indirecta	Contactos_ofrecidos	Contestados
1	Tecnología	2016-08	45739	15013	2618	4295	79066	
2	Tecnología	2016-09	49778	13613	2429	3066	77648	
3	Tecnología	2016-10	45791	12776	1760	2624	66346	
4	Tecnología	2016-11	48796	15540	2208	2366	86326	
5	Tecnología	2016-12	44182	12541	1874	2394	63829	
6	Tecnología	2017-01	50021	15315	1873	2618	77217	
7	Tecnología	2017-02	47534	13635	1295	2946	72188	
8	Tecnología	2017-03	50579	15403	1751	3325	75397	
9	Tecnología	2017-04	46743	13670	1304	2835	74187	
10	Tecnología	2017-05	56268	17046	1608	3570	91969	

Muestra de variables tomadas en cuenta para el modelo

4.1 Análisis Inicial

```

## ID Mesa Fecha
## Min. : 1.00 Length:100 Length:100
## 1st Qu.: 25.75 Class :character Class :character
## Median : 50.50 Mode :character Mode :character
## Mean : 50.50
## 3rd Qu.: 75.25
## Max. :100.00
##
## Contactos_Atendidos_telefono Contactos_Atendidos_Chat
## Min. :2182 Min. : 90.0
## 1st Qu.: 3030 1st Qu.: 444.8
## Median :16430 Median : 6610.5
## Mean :23521 Mean : 8069.6
## 3rd Qu.:44345 3rd Qu.:15544.0
## Max. :56268 Max. :19775.0
##
## Contactos_Atendidos_Chatbot Contactos_autoreporte_gestion_indirecta
## Min. : 93.0 Min. : 0.00
## 1st Qu.: 371.5 1st Qu.: 34.75
## Median : 510.5 Median : 1319.50
## Mean : 739.9 Mean : 3119.73
## 3rd Qu.: 870.8 3rd Qu.: 6239.50
## Max. :2618.0 Max. :10807.00
##
## Contactos_ofrecidos Contestado_Chat_men_Umbral Contestado_Tel_men_Umbral
## Min. :3055 Min. : 87.0 Min. :1046
## 1st Qu.: 4302 1st Qu.: 428.8 1st Qu.:2582
## Median :32341 Median :1221.0 Median :4228
## Mean :40304 Mean :4327.4 Mean :14098
## 3rd Qu.:75995 3rd Qu.:8486.2 3rd Qu.:27942
## Max. :99648 Max. :14244.0 Max. :43036
##
## SD06_ASA_Correo gestion_autoreporte
## Min. : 0.100 Min. :1.630
## 1st Qu.: 8.582 1st Qu.: 7.872
## Median :10.420 Median : 9.845
## Mean :10.657 Mean :11.838
## 3rd Qu.:12.852 3rd Qu.:13.120
## Max. :27.130 Max. :69.890
## NA's :2 NA's :10
## SD05_Disponibilidad_de_la_mesa_de_servicio No_masivos Linea_Base
## Min. :0.9892 Min. : 0.00 Min. :2671
## 1st Qu.:0.9986 1st Qu.: 37.75 1st Qu.:3520
## Median :0.9999 Median : 300.00 Median :24140
## Mean :0.9985 Mean :1117.83 Mean :30960
## 3rd Qu.:1.0000 3rd Qu.:2058.00 3rd Qu.:59084
## Max. :1.0000 Max. :5353.00 Max. :71498
##
## AHT_Tel_seg AHT_chat_seg Costo Costo_adicion_bot
## Min. : 378.0 Min. : 739.0 Min. : 37351316 Min. :700775152
## 1st Qu.: 445.0 1st Qu.: 880.0 1st Qu.: 61670405 1st Qu.:700775152
## Median : 609.5 Median : 928.0 Median : 64401586 Median :717560655
## Mean : 580.1 Mean : 955.5 Mean :188437832 Mean :726293125
## 3rd Qu.: 654.5 3rd Qu.: 985.8 3rd Qu.: 72285934 3rd Qu.:751131662
## Max. :1029.0 Max. :1393.0 Max. :713780346 Max. :765928956
## NA's :37 NA's :87
## AHT_Web_min
## Min. : 0.0
## 1st Qu.: 0.0
## Median :570.0
## Mean :465.3
## 3rd Qu.:722.0
## Max. :900.0
##

```

Se observan 37 valores nulos en costos, puesto que era otro RFP con otras condiciones. En ASA Correo y gestión de autoreporte también existen datos nulos.

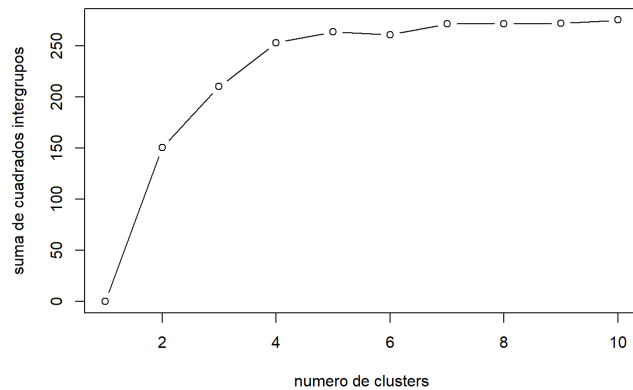
Se observa que la mesa de operaciones sólo se ha movido entre una sola línea base, mientras que la mesa de tecnología ha tenido varias líneas bases.

4.2 Agrupaciones

Realizamos el estudio, sin tener en cuenta la variable costo:

[Code](#)

De acuerdo a este resultado, un k=5, sería el ideal.

[Code](#)

Aunque se vieron variables altamente correlacionadas, como Disponibilidad mesa de servicio, Contactos atendidos Bot, ASA Correo, Gestion Autoreporte, AHT Tel, AHT chat y Costo, se decide eliminar todas excepto mas tres últimas, ya que el modelo de cobro depende de estas para su diseño.

4.3 Valores perdidos

Se identifican variables con valores perdidos y se realiza con la media.

Se toma el argumento de KNN con la habilitación del parámetro numFun = median. Se realiza con k=5, de acuerdo al gráfico de agrupamiento k-means anterior.

Variables con datos nulos:

SD06_ASA_Correo gestion_autoreporte

Para estos valores perdidos (NA) se recorre en el dataframe con las nuevas variables imputadas por el algoritmo de Gower.

Se contaban con un contrato desde 2016 hasta agosto de 2019, donde inicia un nuevo RFP. Por lo tanto, no se deben tomar precios de antes de esta fecha, eran incoherentes porque el contrato anterior tenía otras condiciones, y no habían variables en el precio como IPC u otras indicativas de aumento gradual.

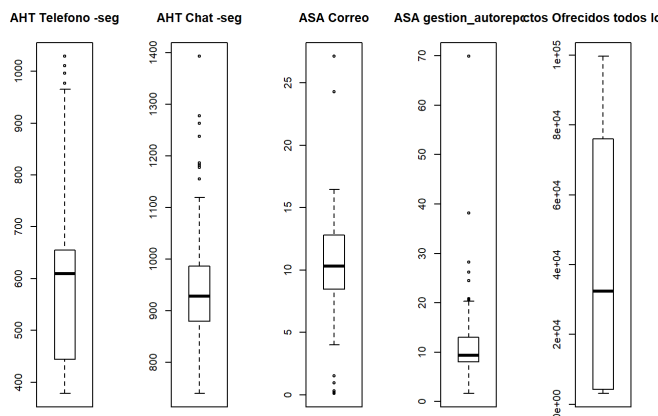
No se identifican por los índices cuáles eran los valores perdidos (NA), sólo se harán para los precios antes de agosto de 2019, se recorren en la base con las nuevas variables Costo_Imputada, las cuales se muestran como TRUE o FALSE, que significa que fueron o no imputadas por el algoritmo de Gower.

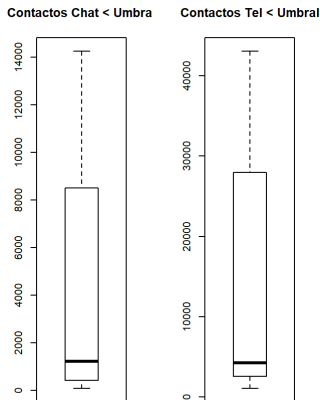
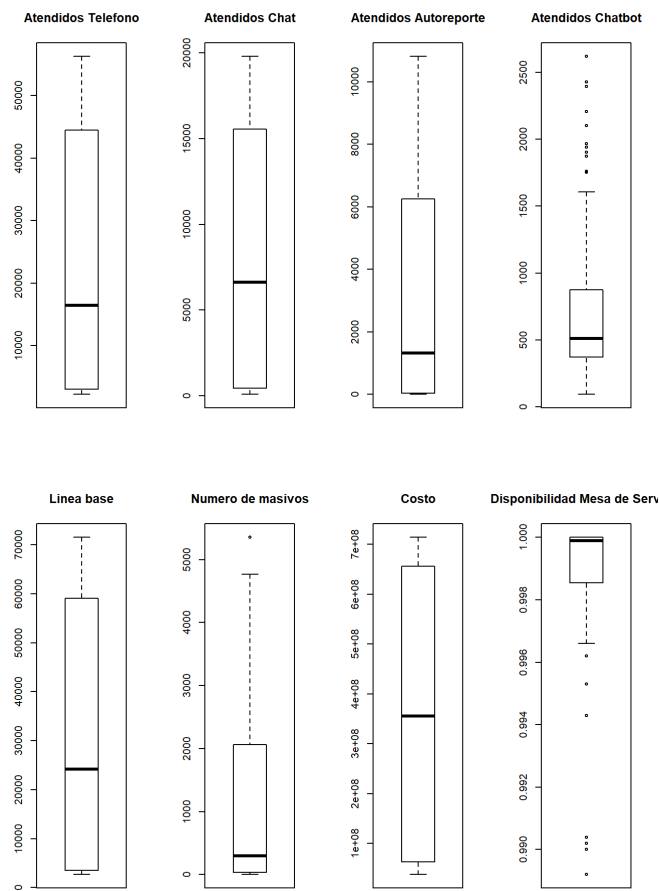
4.4 Valores extremos

Se analiza la presencia de posibles valores extremos (outliers) en las distintas variables:

- Contactos_Atendidos_telefono,
- Contactos_Atendidos_Chat,
- Contactos_Atendidos_Chatbot,
- Contactos_autoreporte_gestion_indirecta,
- Contactos_ofrecidos, Contestados_Chat_men_Umbral, Contestado_Tel_men_Umbral,
- SD06_ASA_Correo,
- SD05_Disponibilidad_de_la_mesa_de_servicio,
- ASA_gestion_autoreporte,
- No_masivos,
- Línea_Base,
- AHT_Tel_seg,
- AHT_chat_seg
- Costo

Se tienen valores nulos para los registros antes de agosto de 2019, que era otro RFP



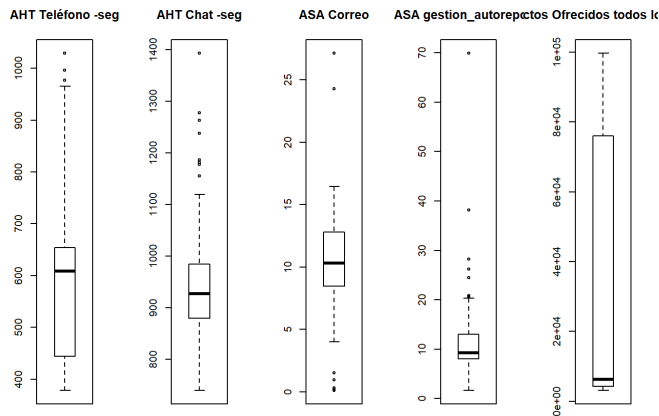


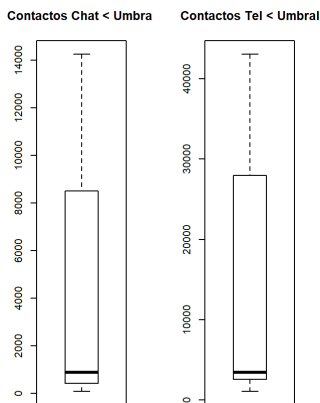
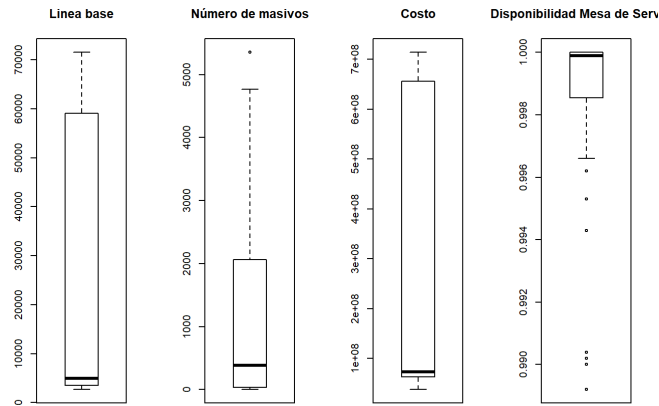
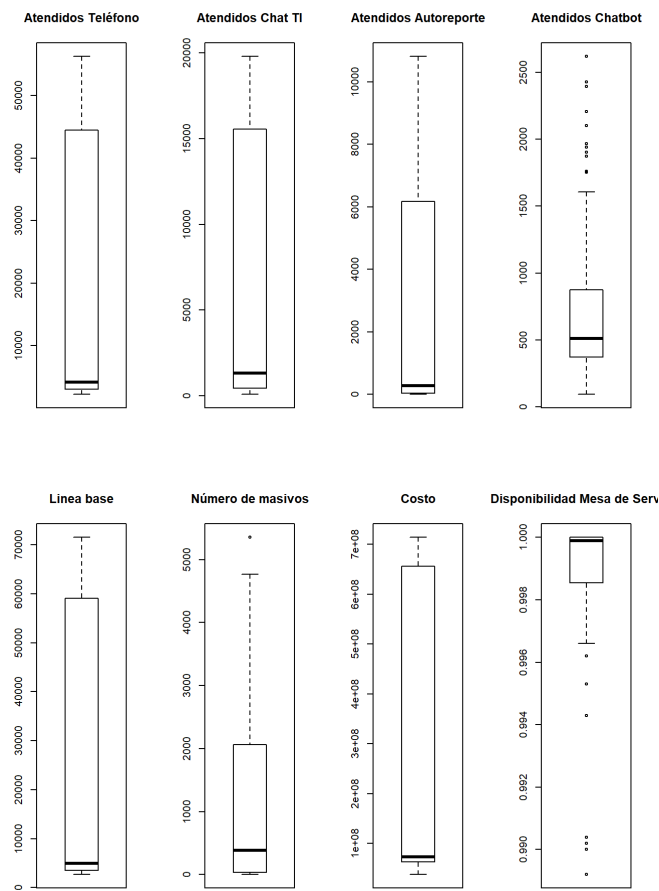
Como se puede observar en la gráfica, los puntos son valores extremos. Estos están en AHT teléfono, AHT Chat, ASA Correo, ASA Gestión autoreporte, número de masivos, Disponibilidad mesa de servicios. Vemos valores extremos en AHT Chat, Se muestran los valores en la siguiente salida:

Se observa que para el AHT Chat, todos los outliers están en Operaciones y en los últimos meses, por lo que se borran estos datos. El único AHT Chat fuera de lo común en Tecnología, fue del mes de abril-2020, con 1182 seg (19,7 min).

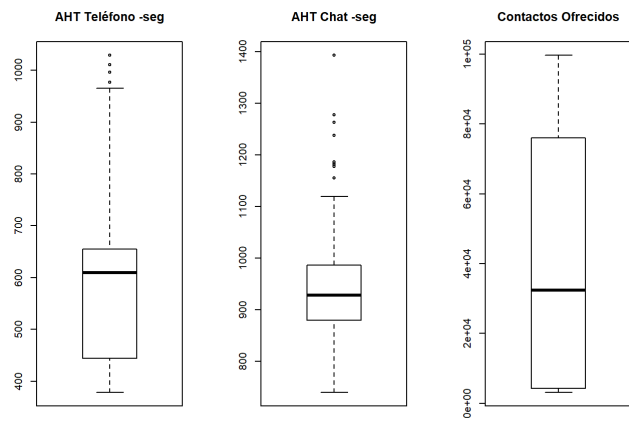
Vemos valores extremos en AHT Teléfono, Se muestran los valores en la siguiente salida:

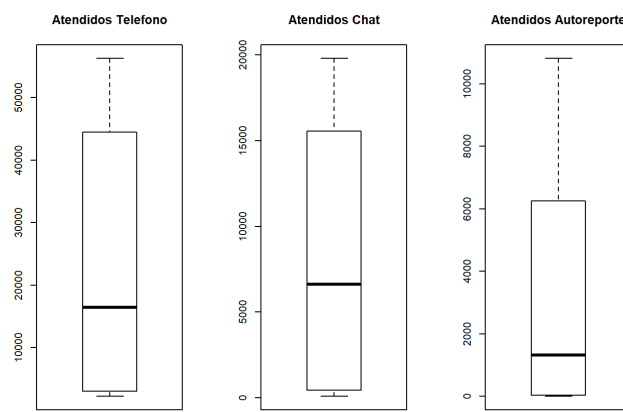
Vemos valores extremos en AHT Teléfono para los meses de 03,05,06,07,08,09 de 2020, sin tener otra variable que los correlacione, por lo que se eliminan, importante para el modelo.



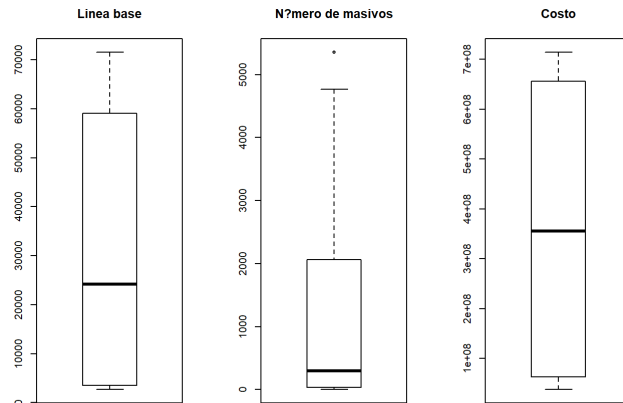


Outliers de masivos ya borrados. Ver nuevos graficos con outliers eliminados:

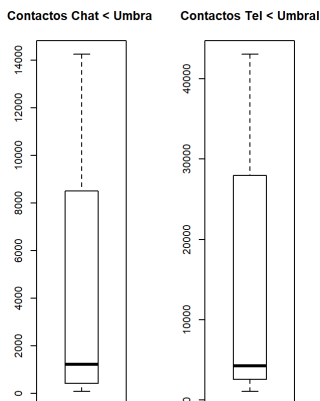




Code



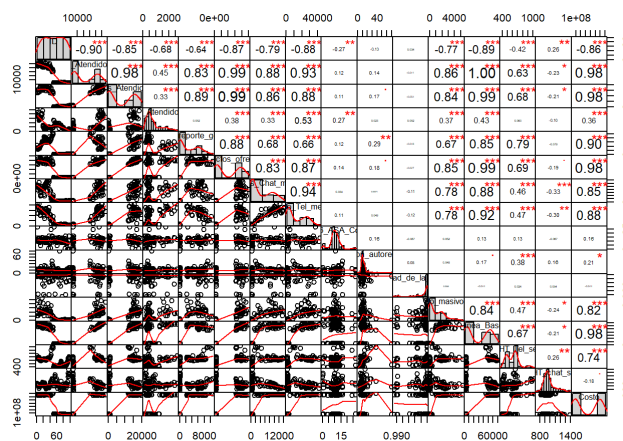
Code



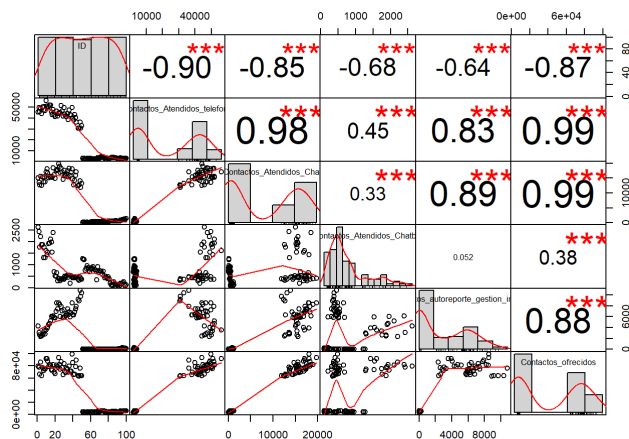
4.5 Correlaciones

Nos proponemos analizar las relaciones entre las diferentes variables del conjunto de datos para ver cómo es esta correlación con los históricos que se tienen.

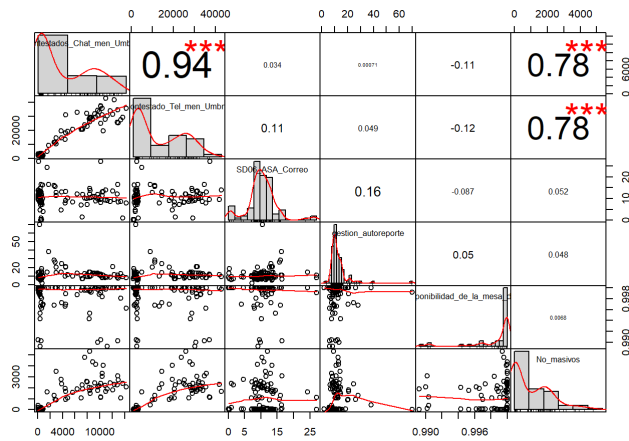
Code



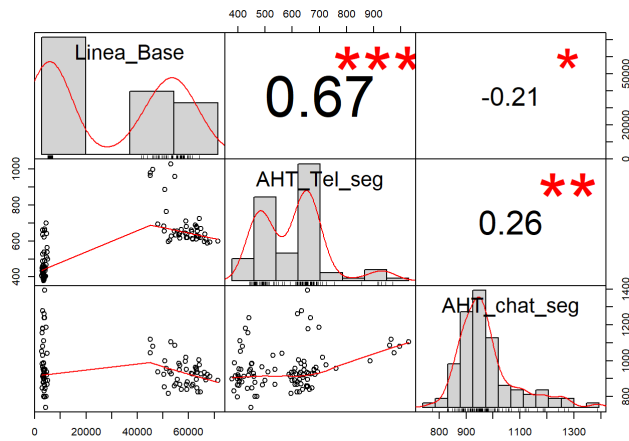
Code



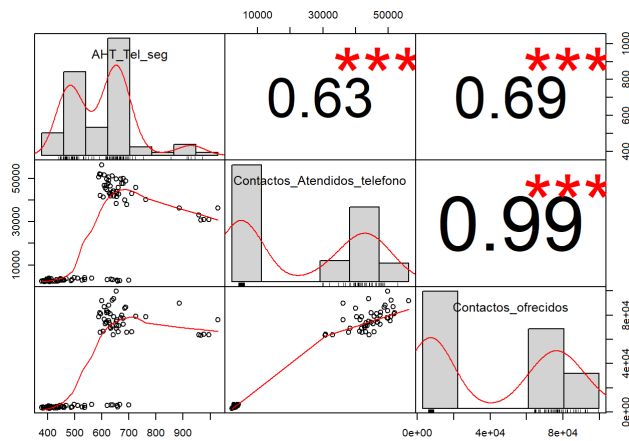
Code



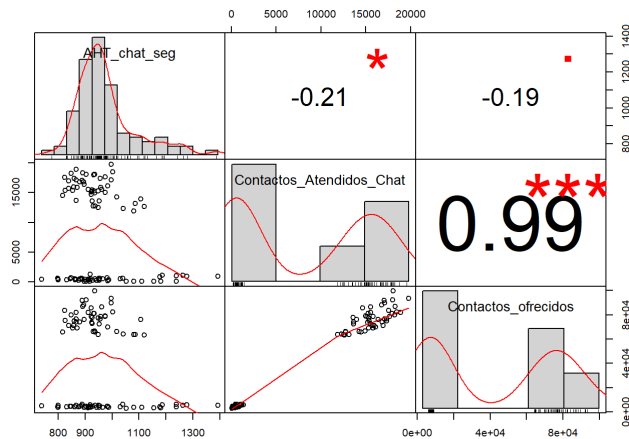
Code



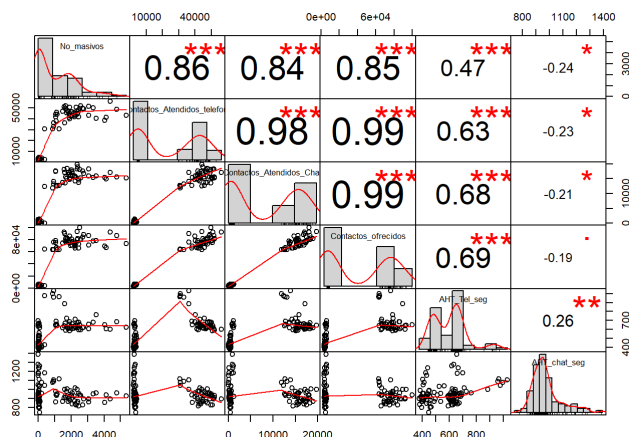
Code



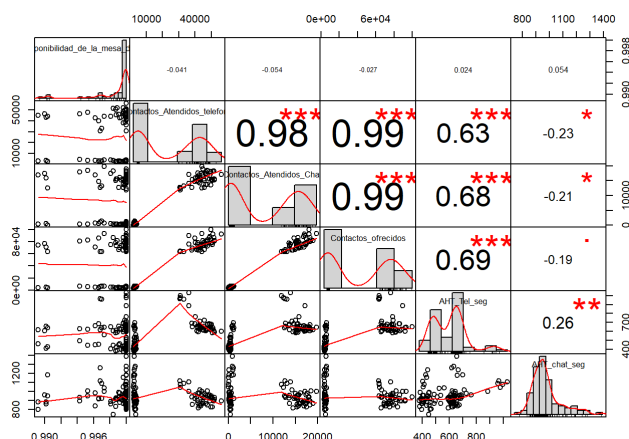
Code



Code



Code



4.6 Variables Correlacionadas para eliminar

```
## 'data.frame': 99 obs. of 4 variables:
## $ SD06_ASA_Correo : num 11.79 14.27 13.10 32.95 ...
## $ gestion_autoreporte : num 13.94 11.09 12.58 12.21 ...
## $ SD05_Disponibilidad_de_la_mesa_de_servicio: num 1 1 0.999 1 0.99 ...
## $ AHT_chat_seg : num 975 926 930 961 949 948 952 826 867 890 ...
```

La distribución de cada variable está en los cuadros diagonales. Por ejemplo, tienen buena distribución los contactos telefónicos antes del umbral, las variables AHT y los No_masivos.

En la parte de abajo de la diagonal, está el gráfico de dispersión. Se observa que Línea base tiene una alta correlación con los contactos telefónicos < umbral.

En la parte de arriba de la diagonal está el valor de correlación y el nivel de significación con número de estrellas. Cada nivel de significación está asociado al símbolo p-values(0, 0.001, 0.01, 0.05, 0.1, 1) <=> símbolos(" ", " ", " ", " ", " ", " "). Por ejemplo, un dato interesante es que el AHT de Teléfono tiene una baja correlación positiva con el AHT de chat (p-value de 0,26), x aumenta ligeramente cuando y lo hace.

También, los contactos telefónicos antes del umbral y línea base están altamente correlacionados. Igualmente, los contactos atendidos de teléfono están altamente correlacionados con los de chat (p-value 0.98). Por último, los contactos ofrecidos y atendidos están altamente relacionados (p-value 0-99).

Disponibilidad del servicio NO tiene correlación con los contactos ofrecidos, atendidos y AHTs.

Masivos Los masivos definitivamente afectan los contactos ofrecidos y atendidos.

Bots Es interesante encontrar que la atención del chatbot "tiene correlación de -0.73 con la gestión indirecta, es decir, que muchos de los casos que no resuelve el bot, tienen que ser gestionados indirectamente por los asesores humanos.

4.7 Valores Nulos

Se encontraron que las variables altamente correlacionadas eran las que tenían también valores nulos, por lo que se decide eliminar (gestión autoreporte y asa correo).

La variable costo por mejora bot o infraestructura, no es relevante para el modelo, por esto también se elimina.

Code

5 Modelo de Predicción de costos

5.1 Modelo de Regresión Lineal Múltiple

- a. Se estima el modelo de predicción de costo, tomando las medias bases de AHT de teléfono (580 segundos= 9,9 minutos) y chat (955 segundos= 15,91 min)
- a. Se toma la línea base.

5.1.1 Modelo 1: Regresión Lineal Múltiple con AHT de Referencia de teléfono y chat

Contractualmente el AHT referencia o de base está definido en 580 segundos para teléfono y 955 segundos para chat. Se toma este para un nuevo modelo:

[Code](#)

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + AHT_chat_seg + AHT_Tel_seg +
##   No_masivos, data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -99200163 -31297564  1276228 22942852 95400390
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.468e+07  3.650e+07  -0.676  0.50066
## Linea_Base   8.709e+03  3.532e+02  24.660 < 2e-16 ***
## AHT_chat_seg -1.329e+05  4.394e+04  -3.024  0.00322 **
## AHT_Tel_seg  4.031e+05  4.636e+04  8.696 1.08e-13 ***
## No_masivos   7.628e+03  5.940e+03   1.284  0.20225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41020000 on 94 degrees of freedom
## Multiple R-squared:  0.9821, Adjusted R-squared:  0.9813
## F-statistic: 1286 on 4 and 94 DF, p-value: < 2.2e-16
```

5.1.2 Modelo 2: Regresión Lineal Múltiple con AHT de Referencia de teléfono y chat + Contactos atendidos

[Code](#)

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + AHT_Chat_Ref + AHT_Tel_seg +
##   No_masivos + Contactos_Atendidos_telefono + Contactos_Atendidos_Chat,
##   data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -88196627 -25231369  1052419 20354208 97621506
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -118384894  27390323  -4.322 3.91e-05 ***
## Linea_Base         9564     2940   3.253  0.00160 **
## AHT_Chat_Ref1      17232265  10522846  1.638  0.10492
## AHT_Tel_seg       312657    49623   6.301 1.01e-08 ***
## No_masivos        6744     6959   0.969  0.33503
## Contactos_Atendidos_telefono -4020     3673  -1.094  0.27660
## Contactos_Atendidos_Chat     9547     3374   2.830  0.00572 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.1e+07 on 92 degrees of freedom
## Multiple R-squared:  0.9825, Adjusted R-squared:  0.9813
## F-statistic: 858.4 on 6 and 92 DF, p-value: < 2.2e-16
```

5.1.3 Modelo 3: Regresión Lineal Múltiple con AHT de Referencia de teléfono y chat + Contactos atendidos + masivos

[Code](#)

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + AHT_Chat_Ref + AHT_Tel_Ref +
##   No_masivos + Contactos_Atendidos_telefono + Contactos_Atendidos_Chat,
##   data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -88879633 -23932864  -3893616  6294150 159951698
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    45904305  9980001   4.600 1.33e-05 ***
## Linea_Base      17189     3188   5.391 5.28e-07 ***
## AHT_Chat_Ref1  -7031711  11653557  -0.603  0.54771
## AHT_Tel_Ref1      NA         NA      NA      NA
## No_masivos       7786     8279   0.940  0.34943
## Contactos_Atendidos_telefono -13318     4002  -3.328 0.00125 **
## Contactos_Atendidos_Chat     11275     4002  2.817  0.00591 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48790000 on 93 degrees of freedom
## Multiple R-squared:  0.9749, Adjusted R-squared:  0.9735
## F-statistic: 721.8 on 5 and 93 DF, p-value: < 2.2e-16
```

5.1.4 Modelo 4: Regresión Lineal Múltiple Costo: con AHTs, Disponibilidad, contactos, masivos y Línea base

[Code](#)

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + AHT_chat_seg + AHT_Tel_seg +
## SD05_Disponibilidad_de_la_mesa_de_servicio + Contactos_Atendidos_telefono +
## Contactos_Atendidos_Chat + Contactos_Atendidos_Chatbot +
## No_masivos + Contactos_autoreporte_gestion_indirecta + Contactos_ofrecidos +
## Contestados_Chat_men_Umbral + SD06_ASA_Correo, data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -86978534 -17059777  2844569 19406640 72247482
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)      360124525 1286960469  0.280
## Linea_Base          -1047       3146  -0.333
## AHT_chat_seg       -86204      38594  -2.234
## AHT_Tel_seg        287595      53632  5.362
## SD05_Disponibilidad_de_la_mesa_de_servicio -401134994 1287774596 -0.311
## Contactos_Atendidos_telefono      12609      4272  2.951
## Contactos_Atendidos_Chat        -1827       5799  -0.315
## Contactos_Atendidos_Chatbot      18357     11321  1.621
## No_masivos           4197       5813  0.722
## Contactos_autoreporte_gestion_indirecta  26311      4386  5.999
## Contactos_ofrecidos        -1369       1402  -0.976
## Contestados_Chat_men_Umbral        1206       2418  0.499
## SD06_ASA_Correo      1602283      820055  1.954
##
##              Pr(>|t|)
## (Intercept)      0.78028
## Linea_Base        0.74003
## AHT_chat_seg      0.02810 *
## AHT_Tel_seg       6.81e-07 ***
## SD05_Disponibilidad_de_la_mesa_de_servicio 0.75618
## Contactos_Atendidos_telefono      0.00408 **
## Contactos_Atendidos_Chat        0.75350
## Contactos_Atendidos_Chatbot      0.10857
## No_masivos           0.47228
## Contactos_autoreporte_gestion_indirecta 4.55e-08 ***
## Contactos_ofrecidos        0.33160
## Contestados_Chat_men_Umbral        0.61926
## SD06_ASA_Correo      0.05397 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33570000 on 86 degrees of freedom
## Multiple R-squared:  0.989, Adjusted R-squared:  0.9875
## F-statistic: 644.5 on 12 and 86 DF,  p-value: < 2.2e-16
```

5.1.5 Modelo 5: Regresión Lineal Múltiple para predecir Costo y Línea base

Code

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + AHT_chat_seg + No_masivos +
## Contactos_Atendidos_telefono + Contactos_Atendidos_Chat,
## data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -88971319 -22908341 -3591938  7083811 161165152
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      16545498  45401559  0.364 0.716367
## Linea_Base        17368       3120  5.567 2.49e-07 ***
## AHT_chat_seg       25767      45389  0.568 0.571608
## No_masivos         8018       8258  0.971 0.334061
## Contactos_Atendidos_telefono -13673      3828 -3.572 0.000563 ***
## Contactos_Atendidos_Chat      11607      3971  2.923 0.004357 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48800000 on 93 degrees of freedom
## Multiple R-squared:  0.9749, Adjusted R-squared:  0.9735
## F-statistic: 721.4 on 5 and 93 DF,  p-value: < 2.2e-16
```

5.1.6 Modelo 6: Regresión Lineal Múltiple para predecir Costo y Línea base con AHT Chat de Referencia

Este presenta una muy buena predicción.

Code

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + AHT_chat_seg, data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -128428863 -22241179 -2982110  6193061 196009358
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.403e+07  4.823e+07  -0.913  0.364
## Linea_Base   1.061e+04  2.016e+02  52.626 <2e-16 ***
## AHT_chat_seg  7.857e+04  4.847e+04  1.621  0.108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54530000 on 96 degrees of freedom
## Multiple R-squared:  0.9676, Adjusted R-squared:  0.9669
## F-statistic: 1433 on 2 and 96 DF,  p-value: < 2.2e-16
```

5.1.7 Modelo 7: Regresión Lineal Múltiple para predecir Costo y Línea base con AHTS Referencia y contactos

Este tambien presenta una buena predicción:

Code

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + Contactos_Atendidos_telefono +
##   Contactos_Atendidos_Chat + AHT_chat_seg + AHT_Tel_seg + No_masivos,
##   data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -93332076 -26477609 -1186385  22098685  93173228
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -9555338   37282334  -0.256  0.79829
## Linea_Base         8710      2842   3.065  0.00286 **
## Contactos_Atendidos_telefono  -2711    3509  -0.773  0.44176
## Contactos_Atendidos_Chat    8273    3280   2.522  0.01339 *
## AHT_chat_seg    -122353    42873  -2.854  0.00534 **
## AHT_Tel_seg      357304    51920   6.882 7.06e-10 ***
## No_masivos       6265     6751   0.928  0.35583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39860000 on 92 degrees of freedom
## Multiple R-squared:  0.9834, Adjusted R-squared:  0.9823
## F-statistic: 908.8 on 6 and 92 DF,  p-value: < 2.2e-16
```

Code

5.1.8 Modelo 8: Regresión Lineal Múltiple para predecir con un costo referencia

El costo promedio de la factura es de 660029877, para tomar de referencia en nuevo modelo:

Code

```
##
## Call:
## lm(formula = Costo_Ref ~ Linea_Base + Contactos_Atendidos_telefono +
##   Contactos_Atendidos_Chat + AHT_chat_seg + AHT_Tel_seg + No_masivos,
##   data = df6.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70623 -0.05926 -0.01063  0.06191  0.40922
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.316e+00  1.329e-01  9.900 3.71e-16 ***
## Linea_Base     -5.928e-05  1.013e-05  -5.850 7.46e-08 ***
## Contactos_Atendidos_telefono  6.954e-05  1.251e-05  5.557 2.65e-07 ***
## Contactos_Atendidos_Chat    4.013e-05  1.170e-05  3.431 0.000903 ***
## AHT_chat_seg     2.363e-04  1.529e-04  1.546 0.125557
## AHT_Tel_seg     -1.205e-03  1.851e-04  -6.509 3.92e-09 ***
## No_masivos     -3.051e-05  2.407e-05  -1.268 0.208161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1422 on 92 degrees of freedom
## Multiple R-squared:  0.6702, Adjusted R-squared:  0.6487
## F-statistic: 31.16 on 6 and 92 DF,  p-value: < 2.2e-16
```

5.1.9 Modelo 9: Regresión Lineal Múltiple con costo referencia sin masivos ni AHT chat

Este nuevo modelo se hace sin variables altamente relacionadas:

Code

```
##
## Call:
## lm(formula = Costo_Ref ~ Linea_Base + Contactos_Atendidos_telefono +
##   Contactos_Atendidos_Chat + AHT_Tel_seg, data = df6.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.70950 -0.05666 -0.00660  0.06604  0.40619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.488e+00  8.082e-02  18.406 < 2e-16 ***
## Linea_Base     -5.443e-05  9.219e-06  -5.904 5.61e-08 ***
## Contactos_Atendidos_telefono  6.258e-05  1.114e-05  5.616 1.98e-07 ***
## Contactos_Atendidos_Chat    3.471e-05  1.134e-05  3.061 0.00288 **
## AHT_Tel_seg     -1.072e-03  1.615e-04  -6.637 2.04e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1436 on 94 degrees of freedom
## Multiple R-squared:  0.6561, Adjusted R-squared:  0.6414
## F-statistic: 44.83 on 4 and 94 DF,  p-value: < 2.2e-16
```

5.1.10 Modelo 10: Regresión Lineal Múltiple para calcular Costo y Línea base sólo con AHT

Este nuevo modelo se hace sin variables altamente relacionadas:

Code

```
##
## Call:
## lm(formula = Costo ~ Linea_Base + AHT_chat_seg + AHT_Tel_seg,
##   data = df5.1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -93936854 -31088959  1065598  24434876  95892434
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.079e+07  3.650e+07  -0.569  0.57038
## Linea_Base   9.046e+03  2.372e+02  38.135 < 2e-16 ***
## AHT_chat_seg -1.324e+05  4.408e+04  -3.003  0.00341 **
## AHT_Tel_seg  3.925e+05  4.577e+04  8.576 1.81e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41160000 on 95 degrees of freedom
## Multiple R-squared:  0.9817, Adjusted R-squared:  0.9812
## F-statistic: 1702 on 3 and 95 DF,  p-value: < 2.2e-16
```

Con regresión logística, el mejor modelo es el número 7.

5.2 Modelo 11: Random Forest

Este nuevo modelo se hace sin variables altamente relacionadas:

```
##      Length Class Mode
## call      5 -none- call
## type      1 -none- character
## predicted  71 -none- numeric
## mse       1000 -none- numeric
## rsq        1000 -none- numeric
## oob.times  71 -none- numeric
## importance 6 -none- numeric
## importanceSD 0 -none- NULL
## localImportance 0 -none- NULL
## proximity  0 -none- NULL
## ntree      1 -none- numeric
## mtry       1 -none- numeric
## forest     11 -none- list
## coefs      0 -none- NULL
## y          71 -none- numeric
## test       0 -none- NULL
## inbag      0 -none- NULL
```

Comparación de actual y predicho en este modelo:

Code

```
##      Predicho
## Actual 54300570.8077684 56783185.6320269 58962014.2506336
## 53819531.6716      0      0      1
## 58024182.5856      0      0      0
## 58452678.5600419      0      1      0
## 59986353.0112      1      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      0      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      0
## 70596310.68      0      0      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      0
## 655850438.67      0      0      0
## 660029876.86      0      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual 63213458.6444973 63674061.4232186 63987646.4319076
## 53819531.6716      0      0      0
## 58024182.5856      1      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      1      0
## 63069763.68      0      0      1
## 65000834.04      0      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      0
## 70596310.68      0      0      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      0
## 655850438.67      0      0      0
## 660029876.86      0      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual 64115262.8447835 64312687.4873473 64490651.0556336
## 53819531.6716      0      0      0
## 58024182.5856      0      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      1      0      0
## 62941379.46      0      1      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      0      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      1
## 70596310.68      0      0      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      0
## 655850438.67      0      0      0
## 660029876.86      0      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual 64617627.8533125 64735813.470967 64800843.4655685
## 53819531.6716      0      0      0
## 58024182.5856      0      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      1      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      0
## 70596310.68      0      0      0
## 71469039.88      0      0      1
## 72881118.5      0      1      0
## 642750969.4144      0      0      0
## 655850438.67      0      0      0
## 660029876.86      0      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual 65654452.9012951 65910943.1129547 534258437.604351
## 53819531.6716      0      0      0
## 58024182.5856      0      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      0      0      0
## 67419011.24      1      0      0
## 68525825.04      0      0      0
## 70596310.68      0      1      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      0
## 655850438.67      0      0      0
## 660029876.86      0      0      0
## 712799261.1856      0      0      1
##      Predicho
```

```

## Actual      651228089.044309 651910528.912875 652287958.111282
## 53819531.6716      0      0      0
## 58024182.5856      0      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      0      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      0
## 70596310.68      0      0      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      0
## 655850438.67      1      1      1
## 660029876.86      0      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual      652644689.254412 652776686.489717 652935374.226758
## 53819531.6716      0      0      0
## 58024182.5856      0      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      0      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      0
## 70596310.68      0      0      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      0
## 655850438.67      1      1      1
## 660029876.86      0      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual      654602630.559655 654855067.376786 654984206.383917
## 53819531.6716      0      0      0
## 58024182.5856      0      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      0      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      0
## 70596310.68      0      0      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      0
## 655850438.67      1      1      1
## 660029876.86      0      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual      655206243.8154 655464763.828192 657647060.986131
## 53819531.6716      0      0      0
## 58024182.5856      0      0      0
## 58452678.5600419      0      0      0
## 59986353.0112      0      0      0
## 62434537.28      0      0      0
## 62941379.46      0      0      0
## 62963928.57      0      0      0
## 63069763.68      0      0      0
## 65000834.04      0      0      0
## 67419011.24      0      0      0
## 68525825.04      0      0      0
## 70596310.68      0      0      0
## 71469039.88      0      0      0
## 72881118.5      0      0      0
## 642750969.4144      0      0      1
## 655850438.67      0      1      0
## 660029876.86      1      0      0
## 712799261.1856      0      0      0
##      Predicho
## Actual      661685086.916481
## 53819531.6716      0
## 58024182.5856      0
## 58452678.5600419      0
## 59986353.0112      0
## 62434537.28      0
## 62941379.46      0
## 62963928.57      0
## 63069763.68      0
## 65000834.04      0
## 67419011.24      0
## 68525825.04      0
## 70596310.68      0
## 71469039.88      0
## 72881118.5      0
## 642750969.4144      0
## 655850438.67      1
## 660029876.86      0
## 712799261.1856      0

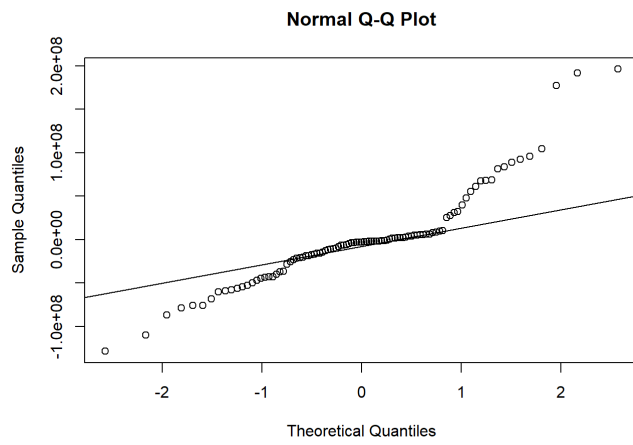
```

5.2.1 Normalidad, Shapiro y Residuos

Se toman los siguientes modelos para realizar este proceso:

5.2.1.1 Modelo 6

Code

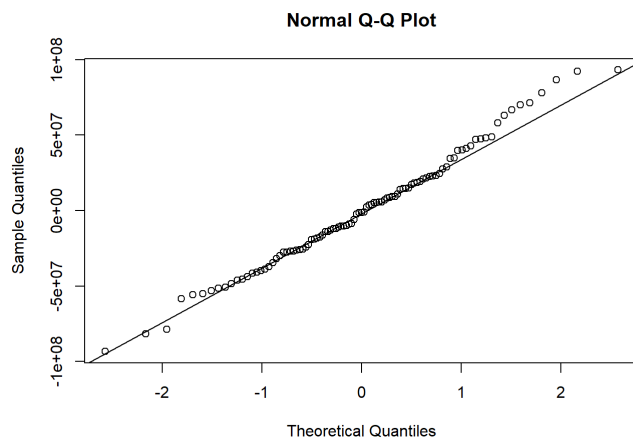


Code

```
##
## Shapiro-Wilk normality test
##
## data: m6_Costo$residuals
## W = 0.88856, p-value = 4.746e-07
```

5.2.1.2 Modelo 7

Code

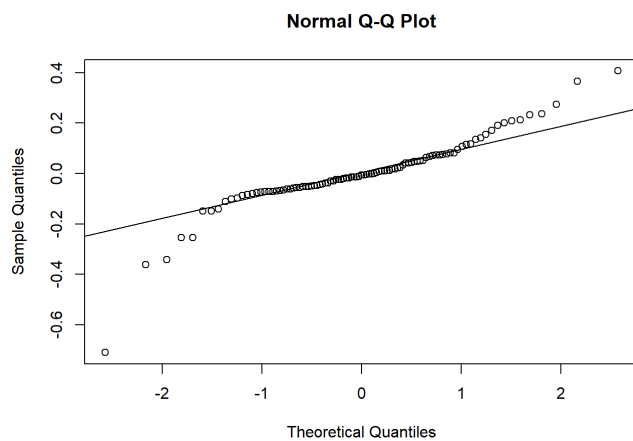


Code

```
##
## Shapiro-Wilk normality test
##
## data: m7_Costo$residuals
## W = 0.98928, p-value = 0.6135
```

5.2.1.3 Modelo 8

Code

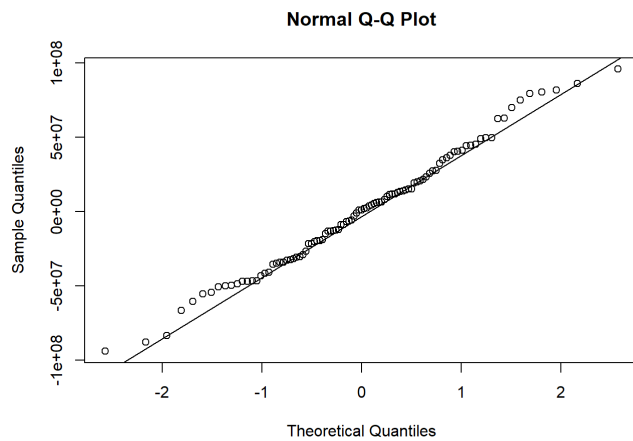


Code

```
##
## Shapiro-Wilk normality test
##
## data: m8_Costo$residuals
## W = 0.88, p-value = 2.031e-07
```

5.2.1.4 Modelo 10

Code



```
##
## Shapiro-Wilk normality test
##
## data: m10_Costo$residuals
## W = 0.98958, p-value = 0.6375
```

5.3 Predicción de precio para Modelos con mejor comportamiento

Se toman los modelos 7 y modelos 10 para las predicciones.

5.3.1 Predicción modelo 7, superior rango 61020, AHT ponderado 968

```
##      fit      lwr      upr
## 1 776640198 676991764 876288631
```

5.3.2 Predicción modelo 7, superior rango 68979, AHT ponderado 1148

```
##      fit      lwr      upr
## 1 880810288 833210149 928410428
```

5.3.3 NUEVA TABLA DE PRECIOS : Predicción completa modelo 7

PREDICCIÓN MODELO 7

Línea base	AHT Teléfono	AHT Chat	Masivos	Contactos atendidos Tel	Contactos Atendidos Chat	LB	fit	lwr	upr	Segmento_Línea_Base
3153.0	404.0	843.0	13.8	2589.6	282.6	3153.0	54516770	39350332	69683208	LB 0-3153
3438.6	431.2	867.0	31.8	2879.4	405.2	3438.6	64127754	50333421	77922087	LB 3439-3667
3667.8	453.0	890.4	53.2	3124.8	512.2	3667.8	71404102	58576778	84231426	LB 3668-4052
4052.4	529.8	912.2	78.2	3421.6	643.8	4052.4	99968113	85487619	114448606	LB 4053-4976
4976.0	609.0	927.0	387.0	4167.0	1333.0	4976.0	140115219	121010529	159219909	LB 4053-4976
52550.2	625.6	947.2	1261.2	40235.0	13663.0	52550.2	567626817	552735920	582517715	LB 52551-56545
56545.0	642.2	964.0	1903.2	43163.2	15293.4	56545.0	615867113	603144991	628589235	LB 52551-56545
61376.8	659.4	1008.6	2256.0	46096.2	16009.4	61376.8	658821125	643813784	673828467	LB 61377-63727
63728.4	694.2	1111.8	2725.0	49784.8	17393.0	63728.4	683494180	661148003	705840357	NA

Las clasificaciones fit, lwr y upr indican el ajuste, el mínimo y máximo estimado respectivamente.

5.3.4 Resumen Predicción completa modelo 7

```
##      fit      lwr      upr
## Min. : 54516770 Min. : 39350332 Min. : 69683208
## 1st Qu.: 71404102 1st Qu.: 58576778 1st Qu.: 84231426
## Median :140115219 Median :121010529 Median :159219909
## Mean :328437910 Mean :312844597 Mean :344031224
## 3rd Qu.:615867113 3rd Qu.:603144991 3rd Qu.:628589235
## Max. :683494180 Max. :661148003 Max. :705840357
```

6 Conclusiones

Las variables que se han tomado para el modelo de precios, son las ampliamente utilizadas para la mayoría de modelos de cobro: línea base, AHT de teléfono, AHT de chat, número de masivos (o servicios nuevos), contactos atendidos y contado atendidos en chat.

Los modelos indican que la variable de # de masivos afecta el resultado final de la línea base predicha.

El AHT también afecta considerablemente la línea base y por lo tanto el costo.

Se toma el modelo 7 para la nueva lista de precios con un p-value: 3.774e-12, mostró una distribución normal y su precisión es aceptable.

Aunque la data muestra algunas variables altamente correlacionadas que se podrían eliminar, como el AHT Chat, este es necesario para el modelo de predicción. La data histórica no mostró esto, pero en la práctica todo call center se afecta por su tiempo de conversación. Por ejemplo, con un tiempo de conversación de 10 minutos se atiende una llamada, pero si el asesor demora 20 minutos con una llamada cuando debería atender dos en este tiempo, pues ahí se confirma que es bastante importante, por lo que se deben tomar en cuenta otras variables.

Una de las variables que más afecta el AHT es un evento masivo (no entregado), que no se puede predecir, pero que ocasiona que el asesor se demore más por un tema que no tiene entregado o no sabe qué hacer. La data recogida mostró la relevancia de los eventos masivos.

Los contactos por bots estaban relacionados negativamente (-0.73) con la gestión indirecta. Esto se interpreta como que entre más contactos atiende el bot, menos contactos van a llegar a una asesor para atender indirectamente.

En la tabla de precios predichos no se tuvo en cuenta estos contactos de bots, por tratarse de un medio bastante barato que no afecta en gran medida el costo. Sin embargo,

entre más contactos haya en este canal, los costos se disminuyen, pero esto aplica para el proveedor de servicios, mas no para el cliente contratante si no lo tiene implementado como core.

Es tambien bastante importante resaltar que aunque se podría aumentar la afectividad y precisión del modelo, se habría tenido que sacrificar importantes variables como el AHT o los contactos atendidos.

Fuentes:

<http://umh3067.edu.umh.es/wp-content/uploads/sites/240/2013/02/Modelos-Lineales-Aplicados-en-R.pdf>

https://rpubs.com/Joaquin_AR/226291

https://rpubs.com/Cristina_Gil/Regresion_Lineal_Simple

https://rpubs.com/Joaquin_AR/229736

<https://thstatsgeek.com/2014/02/16/the-hosmer-lemeshow-goodness-of-fit-test-for-logistic-regression/>

<https://jllopisperez.com/2014/04/02/aplicacion-del-test-de-hosmer-lemeshow-en-medicina/>

<https://www.youtube.com/watch?v=MYW8gA1EQCQ>

https://www.youtube.com/watch?edufilter=NULL&v=HJB6XFkmezM&ab_channel=JuanGabrielGomilaSalas