Técnicas de Programación

Actividad 2: Análisis de datos de navegación y conversión usando R

Máster en Gestión y Análisis de Grandes Volúmenes de Datos:

Big Data

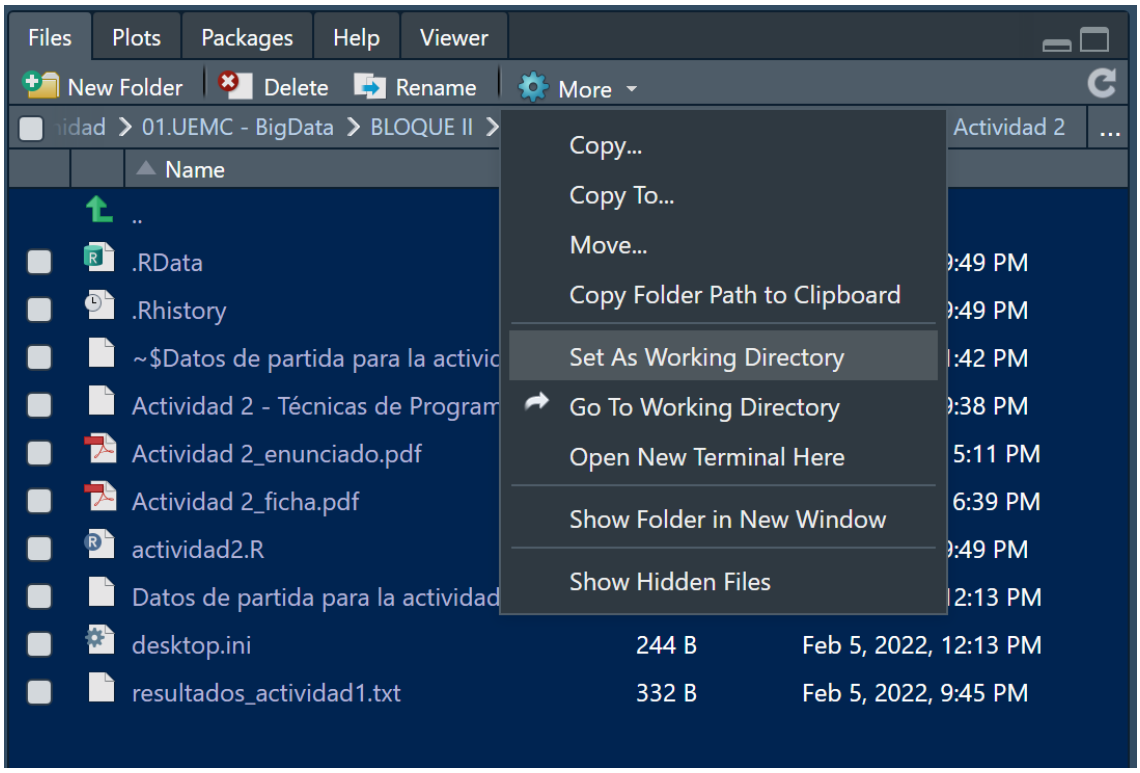
07/03/2022

Cristina Varas Menadas

Índice

1. Configuración del entorno e instalación de librerías

Para trabajar cómodamente en RStudio, se ha establecido la carpeta del proyecto donde se alojarán los scripts de R y los datasets de entrada como directorio de trabajo para trabajar cómodamente y poder utilizar rutas relativas.



Además, se han instalado e importado las librerías necesarias:

```
# Librerias
install.packages("lubridate")
install.packages("tidyverse")
install.packages("rriskDistributions")
install.packages("readxl")

library(lubridate)
library(tidyverse)
library(rriskDistributions)
library(readxl)
```

2. Lectura de los datos de los dataset de navegación y conversión

El dataset utilizado ha sido el dataset preparado en la actividad 1, ya que la limpieza y los datos obtenidos fueron los correctos como se indican en las notas de la actividad. Este dataset contiene toda la información de aquellos clientes que SÍ han convertido.

```
# 1. Lectura de los datos de los dataset de navegación y conversión
df <- read_excel("Datos de partida para la actividad 2.xlsx")
```

	day	month	year	hour	uuid	id_user	gclid	user_re...	url_land...	URL_ba...	glid	iduser	uuid.1	camp	adg	device	sl	adv	rec	conv...	id_lead	lead_type	result	hour_conversion	day_conversion	month_conversion	year_conversion
1	6	9	2021	13:44:21.471	54bfdfe1...	59257b75-5...	Cj0KCQ...	True	https://...	https://w...	Cj0K...	idUser...	uuid=54...	camp=16...	adg=625...	device=m	sl=	adv=494...	rec=true	1	7959	CALL	No le int...	13:45	9	9	2021
2	6	9	2021	0:16:49.530	a2bb63f...	215134e8-3...	EAIaIQo...	False	https://...	https://w...	Cj0K...	idUser...	uuid=37...	camp=10...	adg=639...	device=m	sl=	adv=528...	rec=false	1	7917	FORM	No le int...	10:54	8	9	2021
3	6	9	2021	11:56:55.657	e70ae10...	8fb314e8-b...	Cj0KCQ...	False	https://...	https://w...	Cj0K...	idUser...	uuid=d5...	camp=13...	adg=118...	device=m	sl=	adv=525...	rec=true	1	7914	CALL	llocalizable	11:56	8	9	2021
4	6	9	2021	4:5:37.360	aff513a7...	cefd8f05-8d...	Cj0KCQ...	False	https://...	https://w...	Cj0K...	idUser...	uuid=aff...	camp=73...	adg=467...	device=m	sl=	adv=533...	rec=true	1	7900	FORM	llocalizable	9:06	6	9	2021
5	6	9	2021	7:0:16.755	8c89b4b...	1d792022-7...	Cj0KCQ...	True	https://...	https://w...	Cj0K...	idUser...	uuid=90...	camp=73...	adg=467...	device=c	sl=	adv=383...	rec=true	1	7851	CALL	llocalizable	11:37	7	9	2021
6	6	9	2021	17:35:39.374	39e46fb...	ff88164-ce...	EAIaIQo...	False	https://...	https://w...	EAIaI...	idUser...	uuid=b0...	camp=73...	adg=467...	device=c	sl=	adv=533...	rec=false	1	7989	CALL	No le int...	17:36	10	9	2021
7	6	9	2021	11:56:25.928	c4748bc...	2f616f73-6d...	Cj0KCQ...	False	https://...	https://w...	EAIaI...	idUser...	uuid=0a...	camp=73...	adg=467...	device=c	sl=	adv=533...	rec=false	1	7843	CALL	llocalizable	17:58	7	9	2021
8	6	9	2021	13:17:22.388	38773db...	e942cf56-2...	Cj0KCQ...	False	https://...	https://w...	EAIaI...	idUser...	uuid=f9...	camp=73...	adg=467...	device=c	sl=	adv=533...	rec=true	1	7889	FORM	Positivo	14:10	6	9	2021
9	6	9	2021	14:29:27.778	ceb8bd7...	ebde412f-7...	Cj0KCQ...	True	https://...	https://w...	EAIaI...	idUser...	uuid=23...	camp=73...	adg=467...	device=t	sl=	adv=477...	rec=true	1	7922	FORM	Positivo	18:35	8	9	2021
10	6	9	2021	16:20:0.787	89780e6...	765f65aa-d...	Cj0KCQ...	True	https://...	https://w...	EAIaI...	idUser...	uuid=7b...	camp=16...	adg=585...	device=m	sl=	adv=497...	rec=false	1	7886	FORM	Positivo	18:20	6	9	2021
11	6	9	2021	17:7:36.217	e195aa6...	193f16ad-6...	Cj0KCQ...	False	https://...	https://w...	Cj0K...	idUser...	uuid=e5...	camp=73...	adg=467...	device=m	sl=	adv=499...	rec=false	1	7884	FORM	llocalizable	19:11	6	9	2021
12	6	9	2021	17:42:51.638	3b9610c...	2f59a840-2...	EAIaIQo...	False	https://...	https://w...	Cj0K...	idUser...	uuid=ed...	camp=73...	adg=467...	device=m	sl...	adv=477...	rec=false	1	7881	FORM	llocalizable	20:02	6	9	2021
13	6	9	2021	20:01:18.455	427545b...	fbee18ca-6...	Cj0KCQ...	True	https://...	https://w...	EAIaI...	idUser...	uuid=12...	camp=73...	adg=467...	device=m	sl=	adv=477...	rec=false	1	7836	CALL	No le int...	20:02	6	9	2021
14	6	9	2021	20:20:58.821	6ce9652...	3f336d07-2...	Cj0KCQ...	False	https://...	https://w...	Cj0K...	idUser...	uuid=dd...	camp=16...	adg=625...	device=m	sl=	adv=481...	rec=false	1	7879	FORM	Positivo	20:40	6	9	2021
15	6	9	2021	20:21:0.489	6e37691...	7de715da-3...	Cj0KCQ...	False	https://...	https://w...	Cj0K...	idUser...	uuid=7e...	camp=13...	adg=126...	device=m	sl=	adv=525...	rec=false	1	8003	FORM	Positivo	12:27	10	9	2021

3. Importar los datos generados en la actividad 1

```
# 2. Importar los datos generados en la actividad 1
metricas <- read.table("resultados_actividad1.txt", header = TRUE, sep = ',', dec = '.')
```

Dado que en la actividad 1 solo se pedía la media que tarda el call center en contestar, independientemente del tipo de "result" (Positivo, No le interesa, Ilocalizable), se hicieron estos cálculos previos de estas medias sobre el código Python de la actividad 1, obteniéndose los siguientes resultados:

```
1721.0
LA MEDIA DE TIEMPO QUE EL CALL CENTER TARDA EN CONTESTAR PARA RESULT = Positivo ES DE: 28.683333333333334 HORAS Y 41.0 MINUTOS
3518.0
LA MEDIA DE TIEMPO QUE EL CALL CENTER TARDA EN CONTESTAR PARA RESULT = No le interesa ES DE: 58.63333333333333 HORAS Y 38.0 MINUTOS
188.33333333333334
LA MEDIA DE TIEMPO QUE EL CALL CENTER TARDA EN CONTESTAR PARA RESULT = Ilocalizable ES DE: 3.138888888888889 HORAS Y 8.333333333333343 MINUTOS
1409.7777777777778
LA MEDIA DE TIEMPO QUE EL CALL CENTER TARDA EN CONTESTAR ES DE: 23.496296296296297 HORAS Y 29.77777777777783 MINUTOS
```

También se han calculado los ratios por producto

Estos resultados junto a otros datos de la actividad 1 se han añadido a un csv para trabajar en R con mayor comodidad. Contiene los siguientes datos:

metrica	valor1	valor2
media_call_center	23:29:00	1409.0
media_call_center_positivo	28:41:00	1721.0
media_call_center_noleinteresa	58:38:00	3518.0
media_call_center_ilocalizable	03:08:00	188.33
media_ratio_campana	0.3816888179265104	-
camp=1648174978	0.07692307692307693	-
camp=1042446156	0.008403361344537815	-
camp=13352768134	0.05	-
camp=732187328	0.001885014137606032	-
camp=732401031	0.2	-
camp=732401028	0.012	-
camp=1646744098	0.016666666666666666	-
camp=732187355	0.004016064257028112	-
camp=1648648995	0.0009250693802035153	-
camp=13352855428	0.010869565217391304	-
media_ratio_adgroup	0.47719938514674987	-
adg=62589482065	0.1111111111111111	-
adg=63912889335	0.008403361344537815	-
adg=118216881250	0.1111111111111111	-
adg=46724581628	0.0020964360587002098	-
adg=46724585508	0.2	-
adg=46724585188	0.012	-
adg=58527617970	0.016666666666666666	-
adg=46724587148	0.004016064257028112	-
adg=62589383945	0.0009250693802035153	-
adg=126733863807	0.010869565217391304	-
media_ratio_sitelink	0.04463465479471415	-
sl=*vacío*	0.002967988128047488	-
sl=43115966789	0.041666666666666664	-
media_ratio_adv	0.8711688578755147	-
adv=494939238432	0.1111111111111111	-

4. Calcular las desviaciones típicas

1. De la media que tarda el call center en ponerse en contacto con el usuario

Los datos se han obtenido de los resultados de la actividad 1.

- La media que tarda el call center en ponerse en contacto con el usuario sin tener en cuenta el resultado es 23 horas y 29 minutos (23:29:00), equivalente a 1409 minutos.

Si se quiere tener en cuenta la media que tarda en ponerse en contacto con el usuario por cada "result":

- Media que tarda el call center en ponerse en contacto con el usuario para ***result = "Positivo"***: 28 horas y 41 minutos (28:41:00), equivalente a 1721 minutos.
- Media que tarda el call center en ponerse en contacto con el usuario para ***result = "No le interesa"***: 58 horas y 38 minutos (58:38:00), equivalente a 3518 minutos.
- Media que tarda el call center en ponerse en contacto con el usuario para ***result = "Ilocalizable"***: 3 horas y 8 minutos (03:08:00), equivalente a 188.33 minutos.

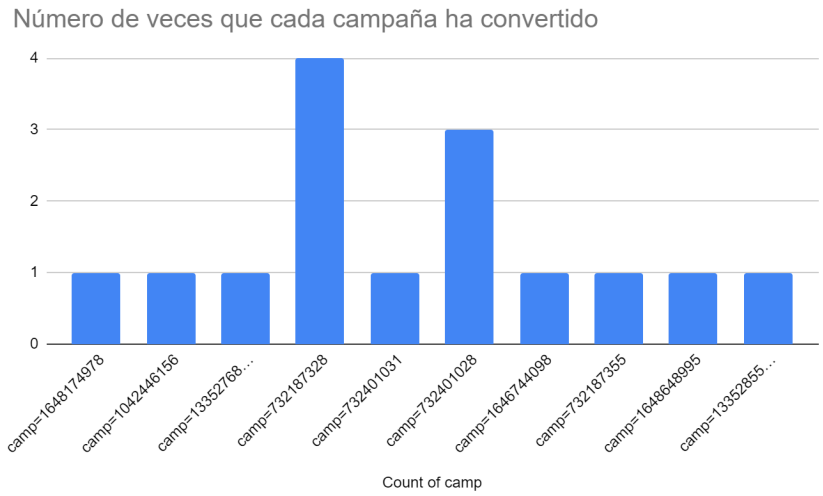
```
18  # 3. Desviacion tipica de la media de tiempo que tarda el call center en ponerse en contacto
19  #    con el usuario en caso de que el tipo de lead sea FORM por tipo de respuesta
20  #    ('Positivo', 'No le interesa', 'Ilocalizable')
21  medias_call_center <- c (1721.0, 3518.0, 188.33)
22  desviacion_tipica_medias <- sd(medias_call_center)
```

```
> desviacion_tipica_medias
[1] 1666.583
```

2. De la media de ratio de conversión por campaña, adgroup, sitelink y anuncio

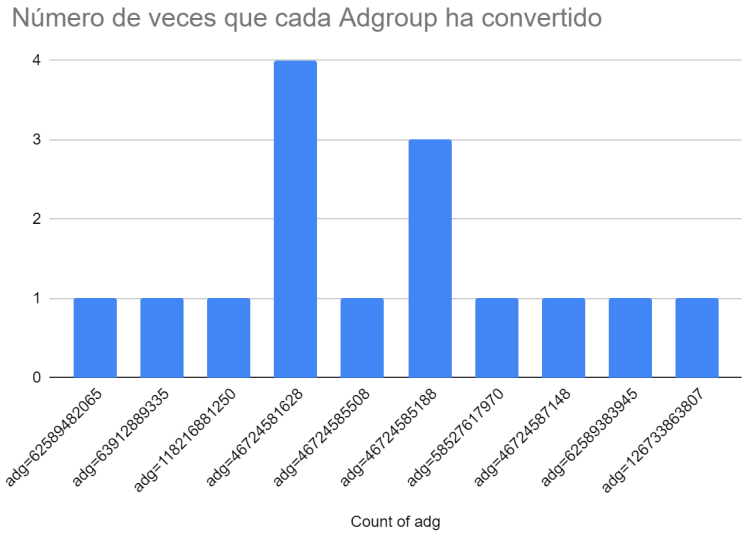
Para el cálculo del ratio de conversión se dividió el número total de clientes totales registrados entre el número de clientes que convirtieron. El resultado fue 0,0025, es decir un 0,25% de ratio de conversión.

Para hacer este cálculo por campaña, adgroup, sitelink y anuncio, se recogieron los siguientes gráficos:



Se obtuvieron los siguientes ratios por campaña:

- camp=1648174978: 0.07692307692307693
- camp=1042446156: 0.008403361344537815
- camp=13352768134: 0.05
- camp=732187328: 0.001885014137606032
- camp=732401031: 0.2
- camp=732401028: 0.012
- camp=1646744098: 0.016666666666666666
- camp=732187355: 0.004016064257028112
- camp=1648648995: 0.0009250693802035153
- camp=13352855428: 0.010869565217391304



Se obtuvieron los siguientes ratios por adgroup:

- adg=62589482065: 0.1111111111111111
- adg=63912889335 : 0.008403361344537815
- adg=118216881250 : 0.1111111111111111
- adg=46724581628 : 0.0020964360587002098
- adg=46724585508 : 0.2
- adg=46724585188 : 0.012
- adg=58527617970 : 0.016666666666666666
- adg=46724587148 : 0.004016064257028112
- adg=62589383945 : 0.0009250693802035153
- adg=126733863807: 0.010869565217391304

Se obtuvieron los siguientes ratios por sitelink:

- sl=*vacío* : 0.002967988128047488
- sl=43115966789 : 0.041666666666666664

- adv=494939238432 : 0.1111111111111111
- adv=528158594826 : 0.009009009009009009
- adv=525403074131: 0.1111111111111111
- adv=533655604703: 0.002233389168062535
- adv=383441982861 : 0.25
- adv=477211734920: 0.024
- adv=497204330125 : 0.3333333333333333
- adv=499223878287: 0.004132231404958678
- adv=481025151787: 0.0018484288354898336
- adv=525407334167: 0.024390243902439025

Por último se han calculado los ratios por marca de coche, lo cual no se pedía en la actividad 1:

- cea: 0.0037593984962406013
- dep30: 0.006493506493506494
- clin200: 0.005208333333333333
- clin400: 0.014035087719298246
- cea-electrico: 0.00819672131147541
- tria: 0.0008764241893076249

Calculando la desviación típica de estos ratios de conversión:

```
#    Desviacion tipica de la media de ratio de conversion por campaña, adgroup, sitelink y anuncio
medias_ratio_conversion <- c ( 0.3816888179265104, 0.47719938514674987,  0.04463465479471415, 0.8711688578755147)
desviacion_tipica_medias_ratio_conversion <- sd(medias_ratio_conversion)
```

Obtenemos:

```
> desviacion_tipica_medias_ratio_conversion
[1] 0.3400737
```

5. Calcular el intervalo de confianza al 95% de las desviaciones típicas del punto 1

Para el cálculo del intervalo de confianza de las desviaciones típicas del punto anterior se ha utilizado el paquete "BSA".

Para la desviación típica de las medias de call center se ha obtenido lo siguiente:

```
# 4. Intervalo de confianza al 95% de las desviaciones tipicas del punto anterior
nivel_confianza <- 0.95
media_medias_call_center <- mean(medias_call_center)
zsum.test(mean.x=media_medias_call_center,sigma.x=desviacion_tipica_medias, n.x=length(df),conf.level=nivel_confianza)
```

```
        One-sample z-Test

data:  Summarized x
z = 4.4757, p-value = 7.616e-06
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1016.881 2601.339
sample estimates:
mean of x
  1809.11
```

Esta función nos indica directamente el intervalo de confianza para el tamaño medio poblacional. Vemos que es I = [1016.881, 2601.339] y por lo tanto sabemos que $P(1016.881 < \mu < 2601.33) = 0.95$, o lo que es lo mismo, el intervalo I contendrá el verdadero valor de la media poblacional, μ, con una probabilidad del 95%.

Para la desviación típica de las medias de ratio de conversión:

```
media_medias_ratio_conversion <- mean(medias_ratio_conversion)
zsum.test(mean.x=media_medias_ratio_conversion,sigma.x=desviacion_tipica_medias_ratio_conversion, n.x=length(df),conf.level=nivel_confianza)
```

```
        One-sample z-Test

data:  Summarized x
z = 5.3792, p-value = 7.483e-08
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.2820151 0.6053307
sample estimates:
mean of x
0.4436729
```

Vemos que el intervalo de confianza es I = [0.2820151, 0,6053307] y por lo tanto sabemos que $P(0.2820151 < \mu < 0,6053307) = 0.95$, o lo que es lo mismo, el intervalo I contendrá el verdadero valor de la media poblacional, μ, con una probabilidad del 95%.

6. Calcular la previsión de número de llamadas que debe soportar en los próximos días el call center

ARIMA es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

Para llevar a cabo este apartado, lo primero que se ha hecho es tomar los datos de las conversiones que han tenido lugar por medio de CALL. Son las siguientes:

	Unna...	day	month	year	hour	uuid	id_user	gclid	user_r...	url_landing	URL_b...	glid	iduser	uuid.1	camp	adg	device	sl	adv	rec	conversiones	id_l...	lead_type	result	hour_conversion	day_conversion	month_conversion	year_conversion	tiempo_en_contestar
1	5	6	9	2021	13:44:21.471	54bf...	59257...	Cj0...	True	https://ww...	https://...	Cj0K...	idUs...	uuid=...	camp=...	adg=...	devic...	sl=	adv=494...	rec=true	1	7959	CALL	No le interesa	13:45	9	9	2021	4321
2	71	6	9	2021	11:56:55.657	e70...	8fb31...	Cj0...	False	https://ww...	https://...	Cj0K...	idUs...	uuid=...	camp=...	adg=...	devic...	sl=	adv=525...	rec=true	1	7914	CALL	llocalizable	11:56	8	9	2021	2880
3	371	6	9	2021	7:0:16.755	8c8...	1d792...	Cj0...	True	https://ww...	https://...	Cj0K...	idUs...	uuid=...	camp=...	adg=...	devic...	sl=	adv=383...	rec=true	1	7851	CALL	llocalizable	11:37	7	9	2021	1717
4	854	6	9	2021	17:35:39.374	39e...	fff881...	EA...	False	https://ww...	https://...	EAIa...	idUs...	uuid=...	camp=...	adg=...	devic...	sl=	adv=533...	rec=fal...	1	7989	CALL	No le interesa	17:36	10	9	2021	5761
5	1794	6	9	2021	11:56:25.928	c47...	2f616f...	Cj0...	False	https://ww...	https://...	EAIa...	idUs...	uuid=...	camp=...	adg=...	devic...	sl=	adv=533...	rec=fal...	1	7843	CALL	llocalizable	17:58	7	9	2021	1802
6	4246	6	9	2021	20:01:18.455	427...	fbee1...	Cj0...	True	https://ww...	https://...	EAIa...	idUs...	uuid=...	camp=...	adg=...	devic...	sl=	adv=477...	rec=fal...	1	7836	CALL	No le interesa	20:02	6	9	2021	1

Lo que buscamos predecir es cuánto tiempo tardarán en dar respuesta cuando los clientes realicen una consulta de tipo CALL. Para ello, restando la diferencia de tiempo entre la hora de la consulta y la respuesta, se han calculado los minutos que tardan en dar respuesta.

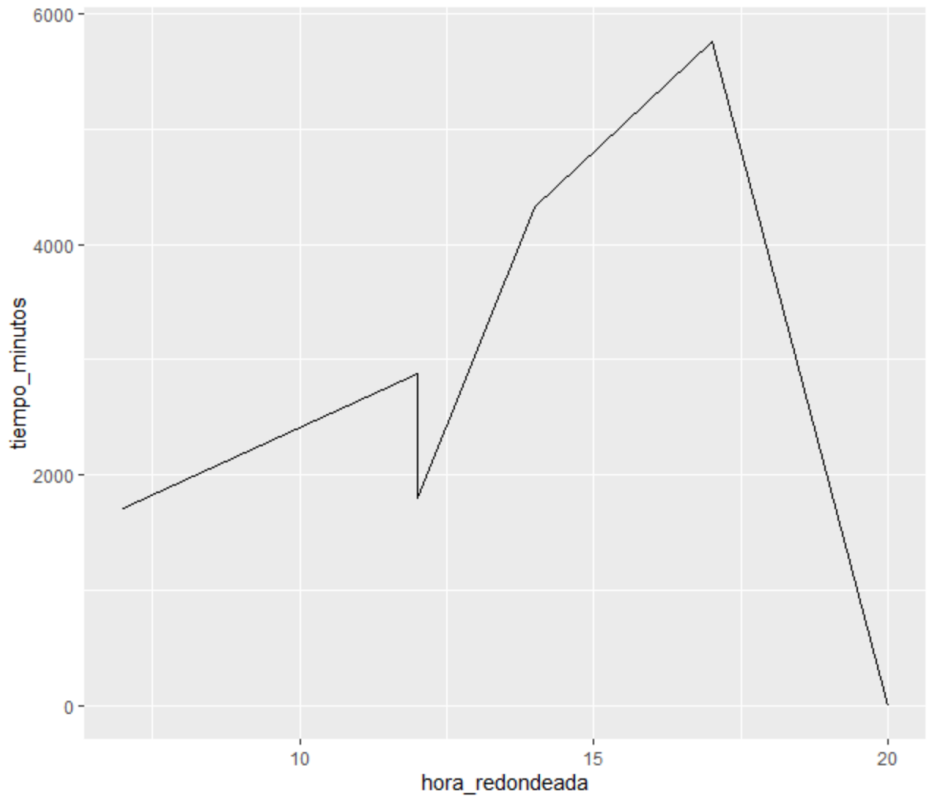
Dado que todas estas conversiones realizadas por medio de CALL son en el **mismo día**, se ha utilizado el campo de la hora para hacer predicciones sobre ella. Con el dato del día, mes y año no obtendríamos ninguna predicción porque el campo temporal no varía.

Por lo tanto, el data frame construido para ejecutar ARIMA es el siguiente:

	id_conversion	hora	hora_redondeada	tiempo_minutos
1	5	13:44:21	14	4321
2	71	11:56:56	12	2880
3	371	7:00:17	7	1717
4	854	17:35:39	17	5761
5	1794	11:56:26	12	1802
6	4246	20:01:18	20	1

Aclarar que, dado que las conversiones son tan solo 15, específicamente de tipo CALL son menos aún, 6, y son muy pocos datos para poder hacer una correcta predicción con un modelo ARIMA.

A continuación vemos el gráfico de líneas de representación de los datos (tiempo que tarda en contestar el CALL center en minutos en función de la hora del día.



Ejecutando el siguiente código:

```
CALL_conversiones_con_tiempo_en_contestar <- read.table("CALL_contact_con_tiempo_en_contestar.csv", header = TRUE, sep = ',', dec = '.')
df_arima <- data.frame("id_conversion" = CALL_conversiones_con_tiempo_en_contestar$Column1,
                       "hora" = CALL_conversiones_con_tiempo_en_contestar$hour,
                       "tiempo_minutos" = CALL_conversiones_con_tiempo_en_contestar$tiempo_en_contestar)
#Gráfica de lineas inicial
ggplot(data = df_arima, aes(x = hora, y = tiempo_minutos)) +
  geom_line()

#Creando objeto ts para podelo
tiempo_minutos_ts <- ts(df_arima$tiempo_minutos,
                        start = 1,
                        frequency = 24)
# Ajuste del modelo
ajuste <- auto.arima(y = tiempo_minutos_ts)
summary(ajuste)
```

Obtenemos el siguiente resumen de nuestro modelo:

```
> summary(ajuste)
Series: tiempo_minutos_ts
ARIMA(0,0,0) with non-zero mean

Coefficients:
           mean
      2747.0000
s.e.   765.1112

sigma^2 = 4214760:  log likelihood = -53.73
AIC=111.46   AICc=115.46   BIC=111.04

Training set error measures:
                        ME     RMSE      MAE       MPE     MAPE MASE       ACF1
Training set -4.547474e-13 1874.113 1573.667 -45769.84 45800.97  NaN -0.1558963
```

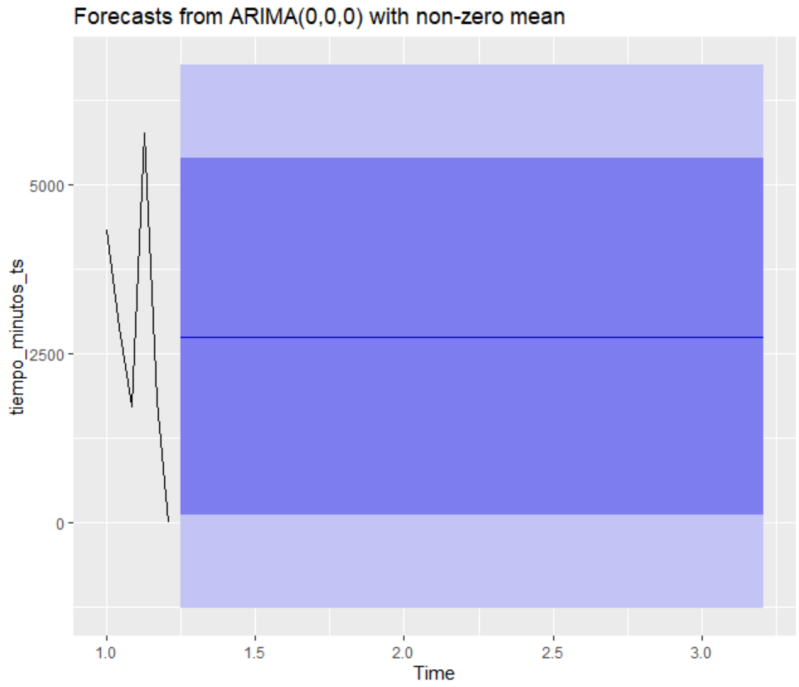
Obtenemos ARIMA (0,0,0), por lo que parece que el modelo no está correctamente ajustado. Estos datos no son los más apropiados para predecir con ARIMA.

Realizando las predicciones obtenemos lo siguiente:

```
# Predicciones
predicciones <- forecast(ajuste)
min(predicciones[['lower']])
min(predicciones[['upper']])

p_predict <- autoplot(predicciones)
p_predict
```

```
> predicciones <- forecast(ajuste)
> min(predicciones[['lower']])
[1] -1276.783
> min(predicciones[['upper']])
[1] 5378.01
```



En la gráfica vemos en negro la línea de tiempo de los datos introducidos durante un día (extremadamente poco) y en morado la predicción. Como vemos, el modelo no se ajusta ni predice nada más allá de la media de minutos, por lo que con este modelo y con tan pocos datos es muy difícil predecir cuánto tardará el CALL center en responder dependiendo de la hora a la que se haga la consulta.

7. Analizar el tipo de distribución que siguen los ratios de conversión por producto

Teniendo en cuenta las medias de los ratios de conversión por marca de coche:

- cea: 0.0037593984962406013
- dep30: 0.006493506493506494
- clin200: 0.005208333333333333
- clin400: 0.014035087719298246
- cea-electrico: 0.00819672131147541
- tria: 0.0008764241893076249

Y utilizando el paquete *rriskDistributions*, el cual permite realizar el diagnóstico de las diferentes distribuciones sobre un rango de datos:

```
tipo_distribucion <- fit.cont(medias_ratio_conversion_marcas$media_ratio_conversion)
```

Se han obtenido los siguientes resultados:

Distribution

- (•) Normal
- ( ) Cauchy
- ( ) Logistic
- ( ) Beta
- ( ) Exponential
- ( ) Chi-square
- ( ) Uniform
- ( ) Gamma
- ( ) Lognormal
- ( ) Weibull

Family	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)	H(AD)	KS(value)	H(KS)
Normal	24.48	-44.96	-45.38	0.62	NULL	0.24	{not rejected}	0.17	NULL
Cauchy	23.79	-43.57	-43.99	0.14	NULL	0.15	{not rejected}	0.13	NULL
Logistic	24.43	-44.86	-45.28	0.36	NULL	0.19	{not rejected}	0.14	NULL
Beta	24.9	-45.81	-46.22	0.16	NULL	0.22	NULL	0.17	NULL
Exponential	24.28	-46.56	-46.77	0.79	0.37	0.40	{not rejected}	0.28	NULL
Chi-square	15.65	-29.3	-29.51	15.44	0	2.01	NULL	0.55	NULL
Uniform	NULL	NULL	NULL	0.94	NULL	Inf	NULL	0.21	NULL
Gamma	24.9	-45.8	-46.21	0.16	NULL	0.22	{not rejected}	0.17	NULL
Lognormal	24.35	-44.69	-45.11	0.24	NULL	0.36	{not rejected}	0.22	NULL
Weibull	25.02	-46.05	-46.46	0.21	NULL	0.20	{not rejected}	0.14	NULL
F	14.56	-25.12	-25.54	19.37	NULL	2.55	NULL	0.62	NULL
Gompertz	25.07	-46.13	-46.55	0.47	NULL	0.19	NULL	0.16	NULL

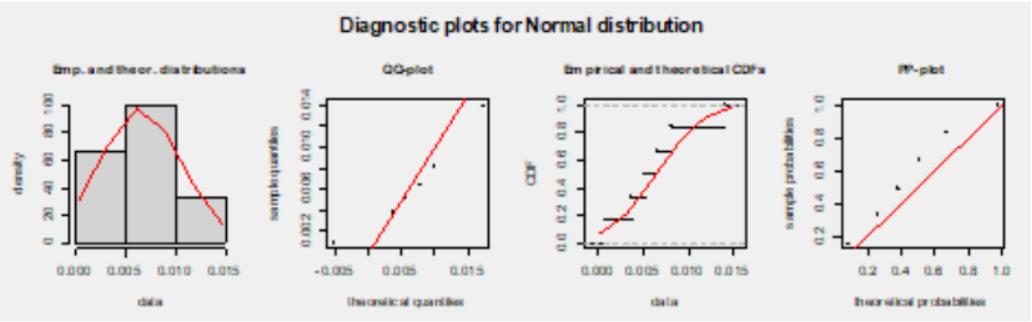
Siguiendo esta tabla, sabemos que el modelo que tenga menor valor de AIC es el mejor modelo. No solo para modelos probabilísticos, sino para series de tiempo y de regresión. Vemos que todas las distribuciones tienen un AIC muy bajo.

La prueba que más fuerza tiene para ver si hay un ajuste de curva en los datos es H(AD). Podemos ver que no rechaza las siguientes distribuciones:

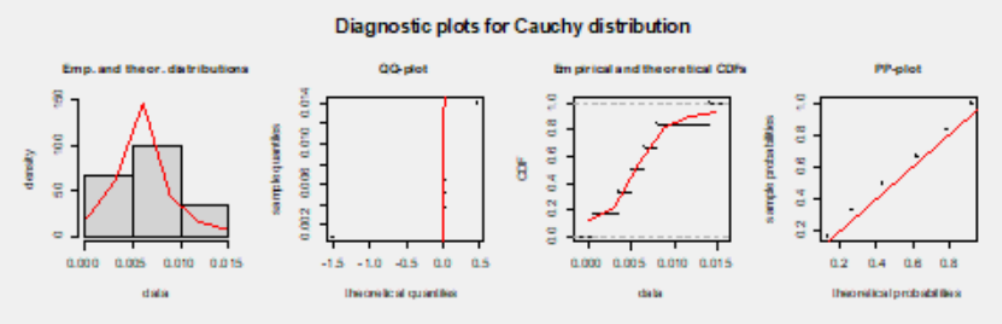
- Normal
- Cauchy
- Logistica
- Exponential
- Gamma
- Lognormal
- Weibull

A continuación de muestran los gráficos que mejor se adaptan a la distribución de los datos:

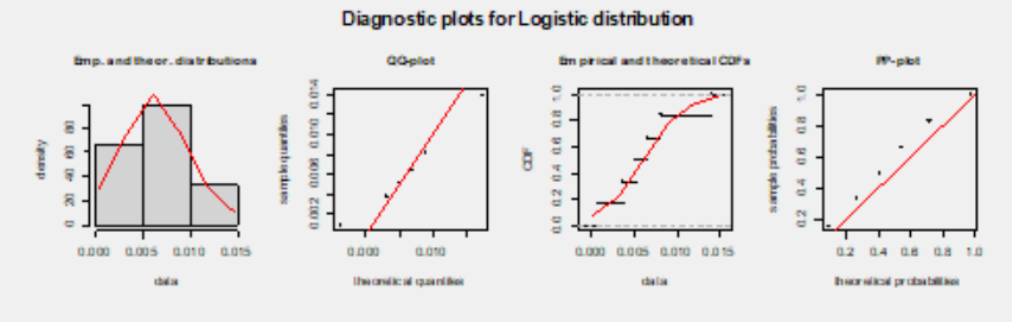
- Normal



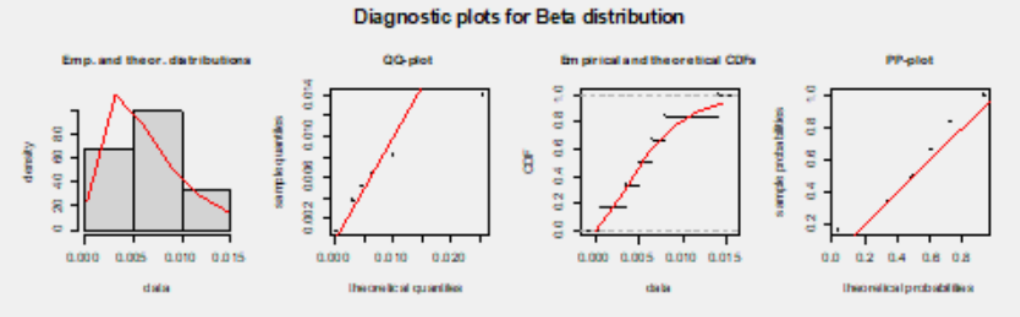
- Cauchy



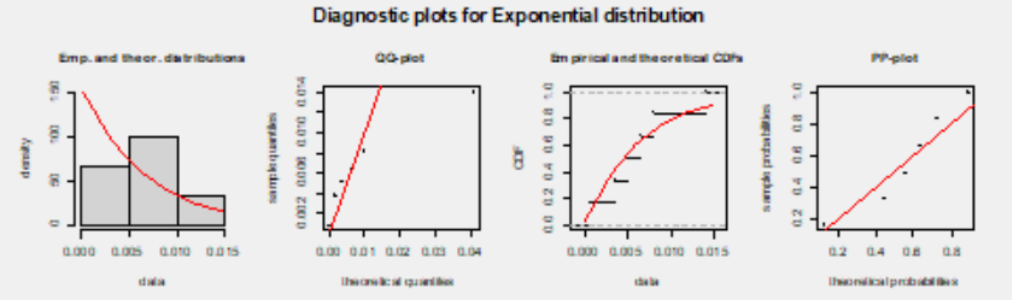
- Logistica



- Beta



- Exponencial



Vemos que la distribución de los datos se ajusta bastante bien a las rectas en diferentes distribuciones.