

# Memoria\_TFM

## 1. Introduccion

El Trabajo Fin de Máster consistirá en analizar el impacto de diversos factores en el retraso de los vuelos. Se añadirán datos climáticos al dataset de entrada y se estudiará el impacto del clima y del resto de variables sobre el retraso de los vuelos.

Para realizar el presente trabajo se han usado los lenguajes Python (con los notebooks de jupyter) y R.

## 2. Análisis del dataset de entrada

A continuación se muestra un breve análisis de las variables del dataset de entrada:

```
str(vuelos)
```

```
## 'data.frame':    222012 obs. of  65 variables:
## $ snapshot_date      : Factor w/ 821 levels "2011-01-10","2012-01-09",...: 515 522 529 536 ...
## $ snapshot_time      : Factor w/ 17543 levels "00:00:11","00:00:19",...: 11747 11378 11391 ...
## $ airline_code       : Factor w/ 3 levels "OR","QN","XX": 3 3 3 3 3 3 3 3 ...
## $ flight_number      : int  5354 5354 5354 5354 5354 5354 5354 5354 ...
## $ flight_date        : Factor w/ 835 levels "2011-01-08","2011-01-10",...: 526 533 540 547 ...
## $ board_point        : Factor w/ 159 levels "AAE","AAL","ADB",...: 73 73 73 73 73 73 73 73 ...
## $ board_lat          : num  42.3 42.3 42.3 42.3 42.3 ...
## $ board_lon          : num  1.41 1.41 1.41 1.41 1.41 ...
## $ board_country_code : Factor w/ 55 levels "AE","AL","AM",...: 21 21 21 21 21 21 21 21 ...
## $ departure_date     : Factor w/ 835 levels "2011-01-08","2011-01-10",...: 526 533 540 547 ...
## $ arrival_date       : logi  NA NA NA NA NA NA ...
## $ off_point          : Factor w/ 161 levels "AAE","AAL","ADB",...: 51 51 51 51 51 51 51 51 ...
## $ off_lat            : num  59.4 59.4 59.4 59.4 59.4 ...
## $ off_lon            : num  5.27 5.27 5.27 5.27 5.27 ...
## $ off_country_code   : Factor w/ 56 levels "AE","AL","AM",...: 39 39 39 39 39 39 39 39 ...
## $ distance           : int  1917 1917 1917 1917 1917 1917 1917 1917 ...
## $ scheduled_time_of_departure : Factor w/ 103286 levels "2011-01-08 10:20:00",...: 54769 55929 57099 ...
## $ estimated_time_of_departure : Factor w/ 26833 levels "", "2011-01-08 10:50:00",...: 12656 1 1 1 139 ...
## $ actual_time_of_departure : Factor w/ 184295 levels "", "2011-01-08 10:53:00",...: 88596 91000 93 ...
## $ scheduled_time_of_arrival : Factor w/ 107739 levels "2011-01-08 12:15:00",...: 57032 58231 59442 ...
## $ estimated_time_of_arrival : Factor w/ 26926 levels "", "2011-01-08 13:05:00",...: 12612 1 1 1 139 ...
## $ actual_time_of_arrival : Factor w/ 185622 levels "", "2011-01-08 13:58:00",...: 1 91317 93815 ...
## $ departure_delay     : int  33 9 10 15 23 5 4 17 53 2 ...
## $ arrival_delay       : int  45 16 13 27 18 19 14 26 42 -5 ...
## $ sched_blocktime     : int  225 225 225 225 225 225 225 225 ...
## $ est_blocktime       : int  240 NA NA NA 205 NA NA 225 235 NA ...
## $ act_blocktime       : int  NA 232 228 237 220 239 235 234 214 218 ...
## $ aircraft_type       : Factor w/ 19 levels "319","320","321",...: 11 11 11 11 11 11 11 11 ...
## $ aircraft_registration_number: Factor w/ 97 levels "", "EF1FQ", "EF2FQ",...: 74 75 75 74 75 74 73 75 ...
## $ general_status_code : Factor w/ 5 levels "Cancelled","Departed",...: 2 2 2 2 2 2 2 2 ...
## $ routing             : Factor w/ 357 levels "AAE-LEU","AAL-LEU",...: 143 143 143 143 143 143 ...
## $ cabin_1_code        : Factor w/ 1 level "Y": 1 1 1 1 1 1 1 1 ...
## $ cabin_1_fitted_configuration: int  227 227 227 227 227 227 227 227 ...
## $ cabin_1_saleable     : int  227 227 227 227 227 227 227 227 ...
## $ cabin_1_pax_boarded  : int  220 222 209 202 226 189 226 226 ...
## $ cabin_1_rpk          : int  421759 425593 400671 387252 433262 362329 433262 433262 ...
```

```
## $ cabin_1_ask          : int  435179 435179 435179 435179 435179 435179 435179 435179 435179
## $ cabin_2_code        : Factor w/ 2 levels "", "J": 1 1 1 1 1 1 1 1 1 1 ...
## $ cabin_2_fitted_configuration: int  NA NA NA NA NA NA NA NA NA NA ...
## $ cabin_2_saleable    : int  NA NA NA NA NA NA NA NA NA NA ...
## $ cabin_2_pax_boarded : int  NA NA NA NA NA NA NA NA NA NA ...
## $ cabin_2_rpk         : int  NA NA NA NA NA NA NA NA NA NA ...
## $ cabin_2_ask         : int  NA NA NA NA NA NA NA NA NA NA ...
## $ cabin_3_code        : logi  NA NA NA NA NA NA NA ...
## $ cabin_3_fitted_configuration: logi NA NA NA NA NA NA NA ...
## $ cabin_3_saleable    : logi  NA NA NA NA NA NA NA ...
## $ cabin_3_pax_boarded : logi  NA NA NA NA NA NA NA ...
## $ cabin_3_rpk         : logi  NA NA NA NA NA NA NA ...
## $ cabin_3_ask         : logi  NA NA NA NA NA NA NA ...
## $ cabin_4_code        : logi  NA NA NA NA NA NA NA ...
## $ cabin_4_fitted_configuration: logi NA NA NA NA NA NA NA ...
## $ cabin_4_saleable    : logi  NA NA NA NA NA NA NA ...
## $ cabin_4_pax_boarded : logi  NA NA NA NA NA NA NA ...
## $ cabin_4_rpk         : logi  NA NA NA NA NA NA NA ...
## $ cabin_4_ask         : logi  NA NA NA NA NA NA NA ...
## $ total_rpk           : int  421759 425593 400671 387252 433262 362329 433262 433262 429428
## $ total_ask           : int  435179 435179 435179 435179 435179 435179 435179 435179 435179
## $ load_factor         : num  0.969 0.978 0.921 0.89 0.996 ...
## $ total_pax           : int  220 222 209 202 226 189 226 226 224 214 ...
## $ total_no_shows      : int  0 2 0 6 10 0 4 0 0 0 ...
## $ total_cabin_crew     : int  5 5 5 5 5 5 5 5 5 ...
## $ total_technical_crew : int  2 2 2 2 2 2 2 2 2 ...
## $ total_baggage_weight : int  2737 2615 2613 2302 2680 2282 2527 2598 2593 2367 ...
## $ number_of_baggage_pieces : int  202 192 182 168 195 166 194 201 191 179 ...
## $ file_sequence_number : Factor w/ 4 levels "1","2","3","P": 4 4 4 4 4 4 4 4 4 ...
```

Como se puede apreciar, el dataset de entrada esta compuesto por 222012 objetos y 65 variables. Los tipos de las variables son variados, existiendo los siguientes: logi, int, num y factor.

`summary(vuelos)`

```
##      snapshot_date      snapshot_time      airline_code      flight_number
## 2016-09-07: 508 13:23:08: 88 OR: 4 Min. : 0
## 2016-09-06: 506 17:16:07: 88 QN: 32278 1st Qu.: 242
## 2016-09-21: 503 17:07:07: 85 XX:189730 Median : 880
## 2016-06-14: 501 06:28:06: 84 Mean : 2594
## 2016-08-31: 499 05:02:06: 81 3rd Qu.: 5465
## 2016-09-13: 499 05:03:06: 81 Max. : 9999
## (Other) :218996 (Other):221505
##      flight_date      board_point      board_lat      board_lon
## 2016-09-01: 508 LEU :110727 Min. : -22.56 Min. : -165.256
## 2016-09-08: 507 MCM : 5612 1st Qu.: 42.34 1st Qu.: 1.409
## 2016-06-09: 504 EOI : 5399 Median : 42.34 Median : 1.409
## 2016-09-02: 500 SYU : 4766 Mean : 42.94 Mean : 1.554
## 2016-09-29: 498 TLN : 4295 3rd Qu.: 43.91 3rd Qu.: 5.070
## 2016-09-15: 497 GNB : 3274 Max. : 70.98 Max. : 91.100
## (Other) :218998 (Other): 87939
##      board_country_code      departure_date      arrival_date      off_point
## ES :120026 2016-09-01: 508 Mode:logical LEU :111285
## FR : 30382 2016-09-08: 507 NA's:222012 MCM : 5566
## GB : 17299 2016-06-09: 504 EOI : 5389
```

```

## DZ      : 6595      2016-09-02: 500      SYX      : 4737
## MC      : 5612      2016-09-29: 498      TLN      : 4306
## (Other): 41269      2016-09-15: 497      GNB      : 3272
## NA's    : 829      (Other) :218998      (Other): 87457
## off_lat      off_lon      off_country_code      distance
## Min.      :-22.56      Min.      :-165.256      ES      :121512      Min.      : 123
## 1st Qu.: 42.34      1st Qu.: 1.409      FR      : 29552      1st Qu.: 389
## Median : 42.34      Median : 1.409      GB      : 17200      Median : 925
## Mean      : 42.93      Mean      : 1.567      DZ      : 6231      Mean      :1742
## 3rd Qu.: 43.73      3rd Qu.: 5.070      MC      : 5566      3rd Qu.:1897
## Max.      : 70.98      Max.      : 91.100      (Other): 41136      Max.      :9275
## NA's      : 815
## scheduled_time_of_departure      estimated_time_of_departure
## 2016-04-17 19:20:00: 14      :190593
## 2016-04-10 19:20:00: 13      2016-05-27 16:45:00: 7
## 2016-04-24 19:20:00: 13      2016-08-19 18:15:00: 6
## 2016-09-25 19:20:00: 13      2016-11-27 19:25:00: 6
## 2016-04-03 19:20:00: 12      2016-01-26 16:40:00: 5
## 2016-04-28 19:20:00: 12      2016-02-05 15:45:00: 5
## (Other)      :221935      (Other)      : 31390
## actual_time_of_departure      scheduled_time_of_arrival
##      : 1765      2016-01-20 21:05:00: 14
## 2016-10-17 16:35:00: 6      2016-02-17 21:05:00: 12
## 2016-11-30 11:35:00: 6      2016-02-24 21:05:00: 12
## 2015-12-31 06:36:00: 5      2016-01-21 21:05:00: 11
## 2016-01-26 19:01:00: 5      2016-01-29 21:05:00: 11
## 2016-01-26 19:12:00: 5      2016-02-04 21:05:00: 11
## (Other)      :220220      (Other)      :221941
## estimated_time_of_arrival      actual_time_of_arrival
##      :191102      : 5535
## 2016-12-06 23:10:00: 6      2016-09-28 18:32:00: 7
## 2016-01-28 20:10:00: 5      2016-03-08 18:28:00: 6
## 2016-02-02 21:55:00: 5      2016-03-23 17:54:00: 6
## 2016-06-10 20:30:00: 5      2016-10-20 14:12:00: 6
## 2016-08-12 21:40:00: 5      2015-12-30 18:24:00: 5
## (Other)      : 30884      (Other)      :216447
## departure_delay      arrival_delay      sched_blocktime      est_blocktime
## Min.      :-59.000      Min.      :-59.000      Min.      :-171.0      Min.      :-1245.0
## 1st Qu.: -4.000      1st Qu.: -9.000      1st Qu.: 115.0      1st Qu.: 120.0
## Median : 0.000      Median : -2.000      Median : 145.0      Median : 160.0
## Mean      : 2.536      Mean      : -0.171      Mean      : 233.8      Mean      : 236.5
## 3rd Qu.: 6.000      3rd Qu.: 7.000      3rd Qu.: 235.0      3rd Qu.: 235.0
## Max.      : 59.000      Max.      : 59.000      Max.      :2385.0      Max.      : 2075.0
## NA's      :1647      NA's      :5078      NA's      :191102
## act_blocktime      aircraft_type      aircraft_registration_number
## Min.      :-1220.0      320      :41043      MFQAT      : 5067
## 1st Qu.: 116.0      319      :38539      MFXAT      : 5038
## Median : 147.0      AT7      :34269      MFTAT      : 5000
## Mean      : 233.2      E90      :27172      MFVAT      : 4977
## 3rd Qu.: 229.0      ATR      :23544      MFAAT      : 4911
## Max.      : 2875.0      321      :18568      MFBAT      : 4826
## NA's      :5535      (Other):38877      (Other):192193
## general_status_code      routing      cabin_1_code
## Cancelled: 1601      MCM-LEU: 5612      Y:222012

```

```

## Departed :220247      LEU-MCM: 5566
## Locked   :      3      EOI-LEU: 5399
## Open     :    154      LEU-EOI: 5389
## Suspended:      7      SYU-LEU: 4766
##                               LEU-SYU: 4737
##                               (Other):190543
## cabin_1_fitted_configuration cabin_1_saleable cabin_1_pax_boarded
## Min.      : 48.0              Min.      : 48.0      Min.      : 0.00
## 1st Qu.: 72.0              1st Qu.: 72.0      1st Qu.: 45.00
## Median :137.0              Median :117.0      Median : 82.00
## Mean      :134.4              Mean      :126.1      Mean      : 92.69
## 3rd Qu.:165.0              3rd Qu.:165.0      3rd Qu.:129.00
## Max.      :297.0              Max.      :297.0      Max.      :273.00
##                               NA's      :4484
## cabin_1_rpk      cabin_1_ask      cabin_2_code
## Min.      :      0      Min.      : 5914      : 97468
## 1st Qu.: 19053      1st Qu.: 33279      J:124544
## Median : 84542      Median : 120857
## Mean      : 253912      Mean      : 312432
## 3rd Qu.: 208701      3rd Qu.: 272924
## Max.      :2467231      Max.      :2457955
## NA's      :4484
## cabin_2_fitted_configuration cabin_2_saleable cabin_2_pax_boarded
## Min.      : 1.00              Min.      : 0.00      Min.      : 0.00
## 1st Qu.: 1.00              1st Qu.:12.00      1st Qu.: 3.00
## Median : 1.00              Median :14.00      Median : 6.00
## Mean      : 7.69              Mean      :18.07      Mean      : 9.22
## 3rd Qu.: 2.00              3rd Qu.:22.00      3rd Qu.:12.00
## Max.      :45.00              Max.      :62.00      Max.      :49.00
## NA's      :97468              NA's      :97514      NA's      :97957
## cabin_2_rpk      cabin_2_ask      cabin_3_code
## Min.      :      0      Min.      :      0      Mode:logical
## 1st Qu.: 3099      1st Qu.: 11755      NA's:222012
## Median : 8868      Median : 22764
## Mean      : 41795      Mean      : 68232
## 3rd Qu.: 22161      3rd Qu.: 40628
## Max.      :417389      Max.      :417389
## NA's      :97957      NA's      :97514
## cabin_3_fitted_configuration cabin_3_saleable cabin_3_pax_boarded
## Mode:logical      Mode:logical      Mode:logical
## NA's:222012      NA's:222012      NA's:222012
##
##
##
##
## cabin_3_rpk      cabin_3_ask      cabin_4_code      cabin_4_fitted_configuration
## Mode:logical      Mode:logical      Mode:logical      Mode:logical
## NA's:222012      NA's:222012      NA's:222012      NA's:222012
##
##
##
##

```

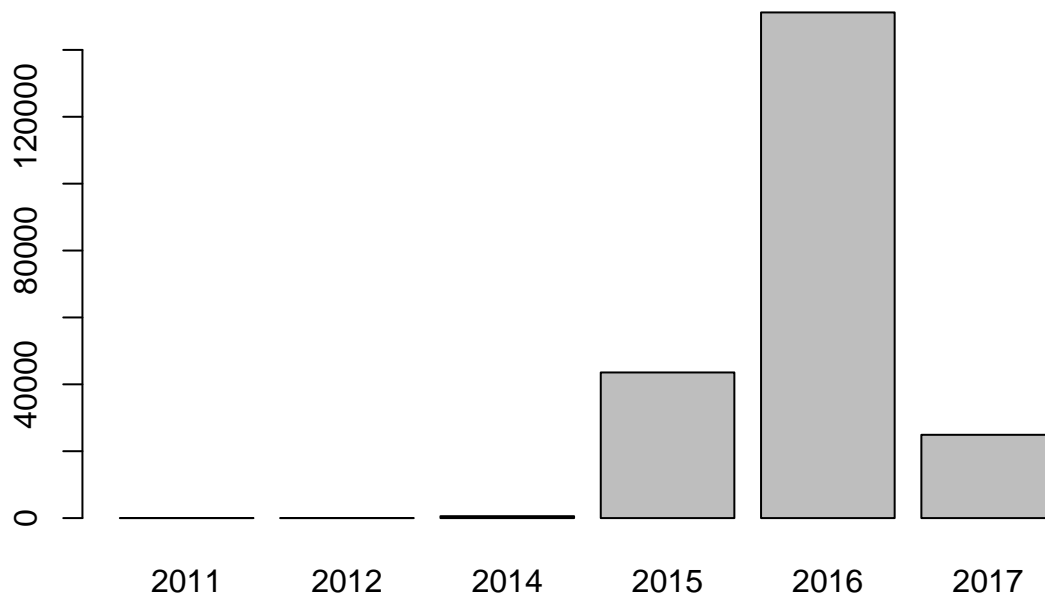
```
## cabin_4_saleable cabin_4_pax_boarded cabin_4_rpk cabin_4_ask
## Mode:logical      Mode:logical      Mode:logical      Mode:logical
## NA's:222012      NA's:222012      NA's:222012      NA's:222012
##
##
##
##
##
## total_rpk      total_ask      load_factor      total_pax
## Min. : 0      Min. : 5914      Min. :0.000      Min. : 0.00
## 1st Qu.: 19108      1st Qu.: 33279      1st Qu.:0.500      1st Qu.: 46.00
## Median : 88179      Median : 134193      Median :0.725      Median : 86.00
## Mean : 277748      Mean : 350694      Mean :0.680      Mean : 97.95
## 3rd Qu.: 221951      3rd Qu.: 299729      3rd Qu.:0.908      3rd Qu.:135.00
## Max. :2764040      Max. :2754765      Max. :1.027      Max. :305.00
## NA's :4484      NA's :4484      NA's :4484
## total_no_shows      total_cabin_crew      total_technical_crew
## Min. : 0.000      Min. : 0.000      Min. :1.000
## 1st Qu.: 0.000      1st Qu.: 2.000      1st Qu.:2.000
## Median : 0.000      Median : 3.000      Median :2.000
## Mean : 1.114      Mean : 3.488      Mean :2.083
## 3rd Qu.: 2.000      3rd Qu.: 4.000      3rd Qu.:2.000
## Max. :171.000      Max. :23.000      Max. :5.000
## NA's :1096      NA's :1096
## total_baggage_weight      number_of_baggage_pieces      file_sequence_number
## Min. : 0      Min. : 0.00      1: 34
## 1st Qu.: 330      1st Qu.: 24.00      2: 1
## Median : 899      Median : 61.00      3: 1
## Mean :1239      Mean : 79.54      P:221976
## 3rd Qu.:1750      3rd Qu.: 114.00
## Max. :8078      Max. :1603.00
## NA's :1096      NA's :1096
```

Ejecutando el comando “summary” observamos lo siguiente:

- Varias variables únicamente contienen valores NAs, por lo tanto deberan ser eliminadas.
- El codigo de aerolinea que más se utiliza es XX. Podemos entender no existía un código y se rellene este campo con XX.
- La mayoría de vuelos tiene el valor **Departed** en su código de estado (variable general\_status\_code).
- La mayoría de los vuelos despegan y aterrizan en aeropuertos Españoles.
- La máxima distancia de un vuelo son 9275 kil?metros y la mínima son 123 kilómetros.
- Existen vuelos que llegan a su destino con 59 minutos de adelanto así como existen vuelos que llegan a su destino con 59 minutos de retraso.
- Existen vuelos que despegan de su origen con 59 minutos de adelanto así como existen vuelos que despegan de su origen con 59 minutos de retraso.
- La media de piezas de equipajes en los aviones es de 61.
- Información sobre las fechas de los vuelos:

La siguiente gráfica muestra una relacion de los vuelos por año.

```
barplot(table(year(ymd_hms(vuelos$actual_time_of_departure))))
```



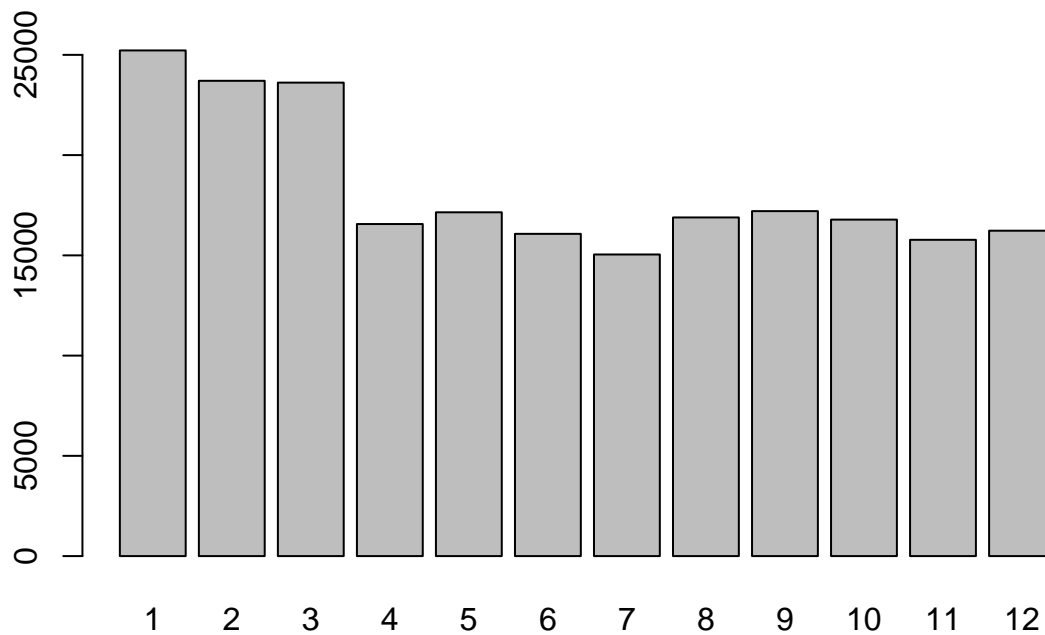
De la gráfica anterior se puede observar que los vuelos están concentrados en el año 2016. A continuación se indica el número de vuelos que se ha realizado cada año.

```
table(year(ymd_hms(vuelos$actual_time_of_departure)))
```

```
##
##  2011    2012    2014    2015    2016    2017
##    13     19    568  43547 151201  24899
```

La siguiente gráfica muestra la relación entre los vuelos y los meses del año:

```
barplot(table(month(ymd_hms(vuelos$actual_time_of_departure))))
```

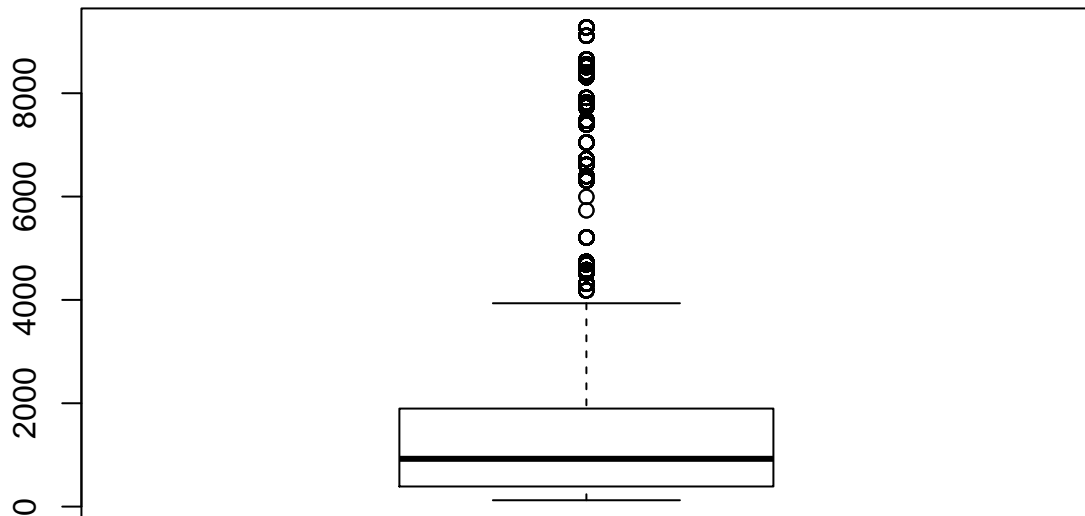


De la gráfica anterior determinamos que los meses en los que más vuelos se hacen son Enero, Febrero y Marzo.

El siguiente boxplot muestra información de la distancia de los vuelos:

```
boxplot(vuelos$distance, main="Distancia vuelos")
```

## Distancia vuelos



```
summary(vuelos$distance)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      123    389     925    1742    1897    9275
```

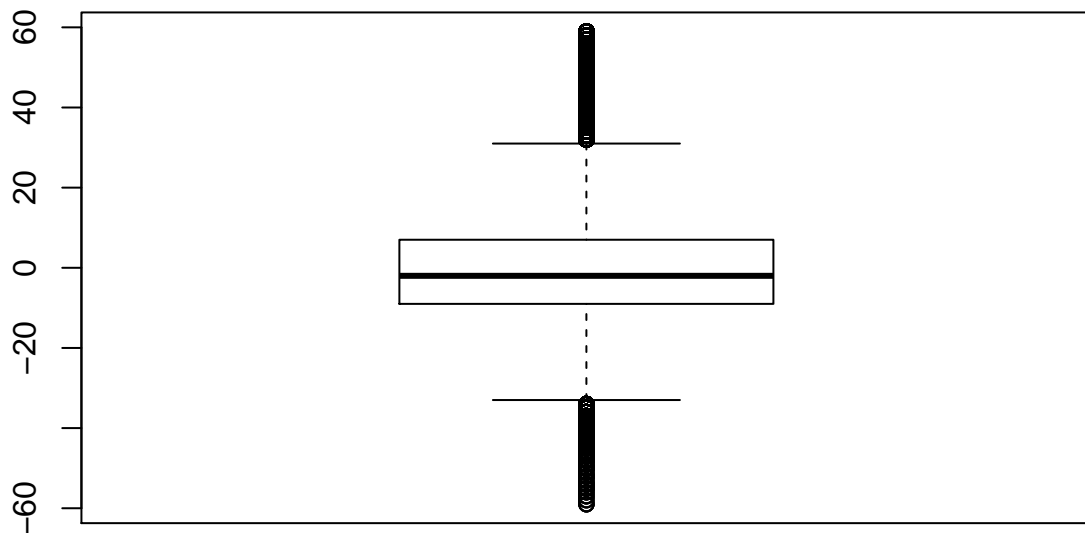
Del boxplot anterior determinamos que la distancia media realizada en los vuelos es de 925 km. El vuelo que menos distancia recorrió fueron 123 km y el vuelo que más distancia recorrió fueron 9275 km.

La siguiente gráfica muestra información sobre el retraso de llegada para los vuelos.

```
boxplot(vuelos$arrival_delay, main="Retraso llegada")
```



## Retraso Llegada



```
summary(vuelos$arrival_delay)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
## -59.000  -9.000   -2.000   -0.171   7.000   59.000   5078
```

Observando la gráfica anterior vemos que los retrasos de llegada de los vuelos se concentran entre -9 y 7 minutos.

Teniendo en cuenta esta primera observación de los datos en bruto, analizaremos cada variable con el fin de obtener el máximo de información posible antes de aplicar cualquier modelo.

### 3. Metodología a utilizar

3.1. La primera acción a tomar será obtener el conjunto de datos a analizar. Para ello, nos fijamos en la variable **general status code**. Obtendremos aquellos registros cuyo valor para esta variable sea **Departed**.

3.2. El siguiente paso será limpiar los datos, analizar las variables de entrada, eliminar los NAs en el caso de ser posible, comprobar que los datos son correctos y crear nuevas variables en base a los datos existentes (siempre que sea posible y tenga sentido).

3.3 Se añadirán datos climáticos al dataset de entrada en función de las coordenadas geográficas.

Los pasos seguidos en orden para el tratamiento de datos y para añadir datos climáticos se incluyen en los archivos adjuntos:

- 1 -DescripciónDatosEntrada\_01.r
- 2 -GráficasDatos.r
- 3 -ScriptWeather.ipynb -> se debe abrir con la herramienta jupyter notebook (python)

- 4 -DescripcionDatosEntrada\_02

Para visualizar como se ha sido el proceso de agregación de datos atmosféricos, se puede ver el anexo adjuntado en el repositorio llamado: **ScriptWeather.md**

### 3.4 Normalización.

Para realizar un mejor estudio, hemos decidido normalizar los datos que tenemos. Para ello se ejecuta la primera parte del archivo en R:

- 5 -Normalizacion\_modelizacion.r

3.5. Se utilizaran diferentes algoritmos para determinar las variables que mas influyen en el retraso de los vuelos. Aplicaremos diferentes modelos y formas para analizar los datos que tenemos.

- Regresion Lineal, PCA, regresión logística y random forest. Por ejemplo mostramos los resultados del análisis con regresión logística:

Para la regresión logística se han realizado los siguientes comandos:

```
## Warning: non-integer #successes in a binomial glm!

##
## Call:
## glm(formula = departure_delay ~ ., family = binomial(link = "logit"),
##      data = Train)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47841  -0.06976  -0.00847   0.05905   1.30477
##
## Coefficients: (6 not defined because of singularities)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.029e+00  1.898e-01  -5.418 6.03e-08 ***
## distance       4.776e-01  2.059e-01   2.319  0.02037 *
## arrival_delay  2.309e+00  5.050e-02  45.721 < 2e-16 ***
## act_blocktime  -1.169e+00  1.849e-01  -6.324 2.55e-10 ***
## cabin_1_fitted_configuration  5.492e-01  7.214e-01   0.761  0.44652
## cabin_1_saleable  -6.635e-01  7.299e-01  -0.909  0.36332
## cabin_1_pax_boarded  1.015e-01  1.835e-01   0.553  0.58009
## cabin_1_rpk      -7.031e+03  3.641e+04  -0.193  0.84690
## cabin_1_ask      -6.122e+03  3.127e+04  -0.196  0.84477
## cabin_2_codeJ     1.828e-01  1.666e-01   1.097  0.27261
## cabin_2_fitted_configuration  1.986e-01  1.590e-01   1.249  0.21166
## cabin_2_saleable  -2.008e-01  3.020e-01  -0.665  0.50606
## cabin_2_pax_boarded  1.290e-01  1.283e-01   1.005  0.31499
## cabin_2_rpk      -1.189e+03  6.160e+03  -0.193  0.84689
## cabin_2_ask      -1.042e+03  5.322e+03  -0.196  0.84476
## total_rpk        7.876e+03  4.079e+04   0.193  0.84690
## total_ask        6.863e+03  3.505e+04   0.196  0.84477
## load_factor     -7.770e-03  7.214e-02  -0.108  0.91422
## total_pax              NA              NA              NA              NA
## total_no_shows     4.703e-01  4.470e-01   1.052  0.29275
## total_cabin_crew    7.365e-02  3.004e-01   0.245  0.80629
## total_technical_crew -5.020e-02  1.486e-01  -0.338  0.73557
## total_baggage_weight  2.941e-01  2.533e-01   1.161  0.24569
## number_of_baggage_pieces -3.214e-01  8.690e-01  -0.370  0.71153
## flightNumberCode    1.978e-02  2.327e-02   0.850  0.39527
```

```

## TMIN_o          -2.442e-04  2.794e-04  -0.874  0.38207
## TMIN_d          8.500e-05  2.770e-04   0.307  0.75894
## TMAX_o         -3.741e-04  3.727e-04  -1.004  0.31560
## TMAX_d          1.664e-05  3.721e-04   0.045  0.96434
## TAVG_o          1.048e-03  6.109e-04   1.715  0.08633 .
## TAVG_d          4.028e-04  5.323e-04   0.757  0.44926
## SNOW_o          NA          NA          NA          NA
## SNOW_d          NA          NA          NA          NA
## PRCP_o          5.647e-02  3.188e-01   0.177  0.85941
## PRCP_d         -5.530e-02  3.359e-01  -0.165  0.86924
## SNWD_o         -4.472e-02  2.890e-01  -0.155  0.87702
## SNWD_d         -1.713e-01  2.952e-01  -0.580  0.56177
## pesosFligthNumber 8.394e-02  4.797e-02   1.750  0.08011 .
## pesosBoardPoint  2.331e-01  1.190e-01   1.959  0.05010 .
## pesosBoardLat     NA          NA          NA          NA
## pesosBoardLon     NA          NA          NA          NA
## pesosBoardCountryCode -1.184e-01  7.084e-02  -1.671  0.09473 .
## pesosOffPoint     -2.508e+00  4.651e+00  -0.539  0.58970
## pesoOffLat        NA          NA          NA          NA
## pesoOffLon        2.577e+00  4.632e+00   0.556  0.57800
## pesoOffCountryCode -4.417e-02  5.651e-02  -0.782  0.43446
## pesoAircraftType  1.051e-01  5.506e-02   1.909  0.05632 .
## pesoAircraftRegNumber 1.717e-02  5.185e-02   0.331  0.74059
## pesoRouting       -3.260e-01  1.030e-01  -3.165  0.00155 **
## pesoMesSalida      5.932e-02  2.412e-01   0.246  0.80571
## pesoDiaSalida      1.299e-02  5.626e-02   0.231  0.81739
## pesoHoraSalida     1.158e-01  2.383e-02   4.861  1.17e-06 ***
## pesoMesLlegada     -2.278e-02  2.412e-01  -0.094  0.92475
## pesoDiaLlegada     1.170e-03  5.617e-02   0.021  0.98338
## pesoHoraLlegada    -1.598e-02  2.243e-02  -0.712  0.47625
## diaSemanaVuelojueves 5.562e-03  2.000e-02   0.278  0.78096
## diaSemanaVuelolunes 6.634e-03  1.993e-02   0.333  0.73920
## diaSemanaVuelomartes -8.582e-03  2.039e-02  -0.421  0.67378
## diaSemanaVuelomiércoles -6.313e-04  2.025e-02  -0.031  0.97513
## diaSemanaVuelosábado 1.005e-02  2.169e-02   0.463  0.64324
## diaSemanaVueloviernes 2.224e-02  1.994e-02   1.115  0.26472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 5851.3 on 148873 degrees of freedom
## Residual deviance: 2793.3 on 148819 degrees of freedom
## (15 observations deleted due to missingness)
## AIC: 186178
##
## Number of Fisher Scoring iterations: 3

```

#### 4. Conclusiones y resumen de resultados

Como conclusiones se puede indicar que las variables que más influyen en el retraso de llegada de los vuelos analizados son:

- El retraso de salida (departure\_delay)

- El tiempo de vuelo. (act\_blocktime)
- El avión. (flight\_number)
- La ruta del vuelo. (routing)
- Y la hora de llegada.

En cuanto a las variables atmosféricas se puede deducir que no son demasiado concluyentes. Al menos los datos que hemos analizado. Se puede ver que quizás la temperatura media es más importante en el retraso del vuelo que otras variables climáticas pero tampoco parece ser importante en el retraso de los vuelos.

Si hubiesemos analizado datos climáticos por hora se observaría mejor si afecta o no al retraso.