

Analítica aplicada al Estudio y Detección de la Criminalidad en Nueva York

Junio 2017

Francisco J. Rodríguez Aragón Ph. D. in Statistics

Senior Data Scientist

Teradata Company Background

Corporate Vision

Enabling data-driven business

Mission

Providing the world's best analytic data solutions to drive competitive advantage for our customers



2,600+ CUSTOMERS in **77** COUNTRIES

10000+ EMPLOYEES

TOP 10 public U.S. software company
✓ Member of
S&P 500



Financially **STRONG** and **GROWING** (Revenue of \$2,732M)

1

End-to-End Solutions and Services

- Data warehousing
- Big data analytics
- Marketing applications

2

Industry Expertise and Experience

- Financial Services
- Communications
- Retail
- Manufacturing
- Healthcare
- Energy/Utilities
- Healthcare
- Government
- Travel/Transportation
- Media/Entertainment

3

Data Analytics Leadership

- Deep expertise
- Analytic engines
- Advanced algorithms
- Industry acclaimed

TERADATA

Un Bootcamp en Milán



https://fjira01.shinyapps.io/shinyapp_v2/



TERADATA.

Agenda



-Introducción

- Un poco de shiny: Análisis visual de la criminalidad
- Previsión de la Criminalidad con Series Temporales
- Deducción de Relaciones entre Crímenes
- Conclusiones

Introducción

- Este diálogo nos cuesta dinero a todos. Falta



FJRA: Vengo a sacar el DNI para mi hija

POLICÍA: ¿Tiene la partida de nacimiento y el volante de empadronamiento?

FJRA: Aquí tengo la partida de nacimiento, ...

POLICÍA: Uy, pero esta partida no tiene apostillado que es para el DNI

FJRA: Ya pero es que no ví que era necesario, ...

POLICÍA: Pues está muy clarito en la web y además no veo el volante, por tanto no puedo hacer el trámite, necesita esos documentos ...



Suponiendo que hubiera leído bien, y que hubiera recopilado la información correctamente, el coste de ir con los 2 papelitos implica:

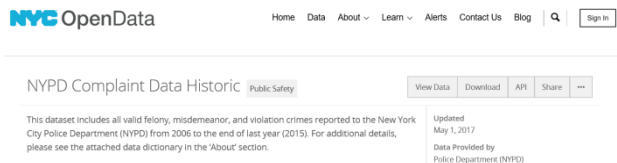
- 2 viajes en metro que si lo imputamos a coste billete sería aprox. 2€
- El tiempo de los 2 funcionarios que si son unos 10min costaría aprox 3€ (suponientos 30€) la hora
- Mi tiempo que son 2h y que podría cobrar, pero que no lo voy a hacer
- El tiempo de espera de las otras personas mientras el funcionario me atiende y que no lo voy a imputar

En fin, el coste total, por hacerlo 2 veces sería de al menos 5€, 3 de los cuales al menos es de dinero puramente público sin sumar la atención policial y el uso de elementos públicos adicionales que no he tenido en cuenta.

¿Por qué los entes públicos no están interrelacionados? ¿Por qué a partir de la identificación por un profesional policial no se puede hacer una query simple a los sistemas de empadronamiento y de partida de nacimiento?

Introducción

- El Departamento de Policía de la Ciudad de Nueva York, publica datos a libre disposición referenciados por tipología, tiempo y lugar de ocurrencia junto con un rico conjunto de características adicionales
- Existen históricos de ocurrencias desde el 2006
- Los enlaces de los datos en bruto estarían en:
 - <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-YTD/5uac-w243>
 - <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i>



The screenshot shows the NYC OpenData website interface. At the top, there's a navigation bar with links: Home, Data, About, Learn, Alerts, Contact Us, Blog, a search icon, and a Sign In button. Below the navigation bar, the main heading is "NYPD Complaint Data Historic" with a "Public Safety" tag. To the right of the heading are buttons for "View Data", "Download", "API", "Share", and a dropdown arrow. Below the heading, there's a descriptive text: "This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to the end of last year (2015). For additional details, please see the attached data dictionary in the 'About' section." To the right of this text, there's a small box containing the text: "Updated May 1, 2017" and "Data Provided by Police Department (NYPD)".

Introducción

- **Se plantean varios tipos de análisis:**

- Análisis simples de ocurrencias de criminalidad
- Análisis predictivos en términos a futuro de la criminalidad en global y por tipología
- Análisis de interrelaciones entre crímenes

- **Todo lo anterior se hará dentro de un mismo entorno analítico: R y un desarrollo shiny**

- **Principales librerías R:**

- shiny shinythemes leaflet rhandsontable
- geosphere insol data.table

Agenda



-Introducción

-**Un poco de shiny: Análisis visual de la criminalidad**

-Previsión de la Criminalidad con Series Temporales

-Deducción de Relaciones entre Crímenes

-Conclusiones

Un poco de shiny: Análisis visual de la Criminalidad

- La primera idea era hacer una aplicación responsive capaz de representar los crímenes ocurridos en Nueva York desde Enero de 2017 hasta final de Marzo de 2017:

- Había que decidir si se representaban todos o parte de los crímenes registrados. Se eligieron crímenes de carácter violento
- Se mapearon los crímenes a unas categorías genéricas

```
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "HARRASSMENT 2", "OTHERS_HAR", DATOS_NYC_ANIO$FNS_DESC)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "ROBBERY", "ROBBERY", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "DANGEROUS DRUGS", "OTHERS_DRU", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "OFFENSES INVOLVING FRAUD", "OTHERS_FRA", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "MISCELLANEOUS PENAL LAW", "OTHERS", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "CRIMINAL MISCHIEF & RELATED OF", "OTHERS_CHI", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "OFF. AGNST PUB ORD SENSBLTY &", "OTHERS", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "ALCOHOLIC BEVERAGE CONTROL LAW", "OTHERS_ALC", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "SEX CRIMES", "RAPE", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "CRIMINAL TRESPASS", "OTHERS_TRE", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "FORGERY", "OTHERS_FOR", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "INTOXICATED & IMPAIRED DRIVING", "OTHERS_ALC", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "OFFENSES AGAINST THE PERSON", "OTHERS_OFE", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "FRAUDS", "OTHERS_FRA", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "ADMINISTRATIVE CODE", "OTHERS", DATOS_NYC_ANIO$Clasificacion)
DATOS_NYC_ANIO$Clasificacion <- ifelse(DATOS_NYC_ANIO$FNS_DESC == "THEFT-FRAUD", "OTHERS_FRA", DATOS_NYC_ANIO$Clasificacion)
```

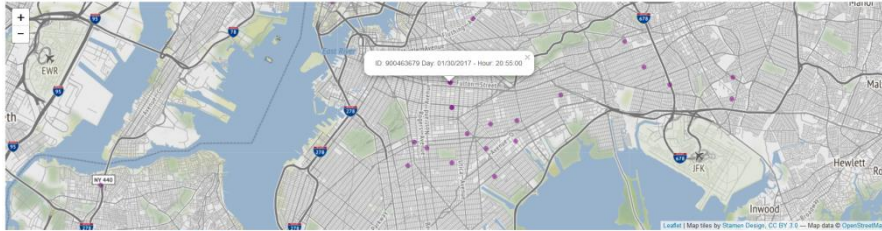
Type of Crime

- ☒ Murder
- ☐ Vehicle Theft
- ☐ Assault
- ☐ Larceny Theft
- ☐ Robbery
- ☐ Burglary

- Se eliminan (en este estudio) los crímenes que no son de la clasificación anterior

Un poco de shiny: Análisis visual de la Criminalidad

- La primera idea era hacer una aplicación responsive capaz de representar los crímenes ocurridos en Nueva York desde Enero de 2017 hasta final de Marzo de 2017:
 - Cada categoría de crímenes es representada por un punto en un mapa. Este punto ofrece cierta interactividad cuando se hace click sobre él



- Se puede hacer zoom sobre a la ocurrencia de un determinado suceso

Un poco de shiny: Análisis visual de la Criminalidad

- La primera idea era hacer una aplicación responsive capaz de representar los crímenes ocurridos en Nueva York desde Enero de 2017 hasta final de Marzo de 2017:
 - Versión móvil o en github: Se ofrece una versión anterior con la mayoría de las funcionalidades que se muestran en <https://github.com/FJROAR> o se puede visualizar en cualquier dispositivo móvil (móvil, table, pc) si se va a la url <https://fjra01.shinyapps.io/MAPAS4/>
 - Algunos elementos clave de codificación: Selección del Tema o de la hoja de estilos

```
library(shinythemes)
library(leaflet)

shinytheme <- function(theme = NULL) {
  # Check that theme exists
  if (is.null(theme) || !theme %in% allThemes()) {
    stop(theme, " is not an available theme. Valid themes are: ",
         paste(allThemes(), collapse = ", ", ".")
    )
  }

  paste0("shinythemes/css/", theme, ".min.css")
}

allThemes <- function() {
  themes <- dir(system.file("shinythemes/css", package = "shinythemes"),
               "*.min.css")
  sub(".min.css", "", themes)
}
```

Un poco de shiny: Análisis visual de la Criminalidad

- La primera idea era hacer una aplicación responsive capaz de representar los crímenes ocurridos en Nueva York desde Enero de 2017 hasta final de Marzo de 2017:
 - Versión móvil o en github: Se ofrece una versión anterior con la mayoría de las funcionalidades que se muestran en <https://github.com/FJROAR> o se puede visualizar en cualquier dispositivo móvil (móvil, table, pc) si se va a la url <https://fjra01.shinyapps.io/MAPAS4/>
 - Algunos elementos clave de codificación: Selección del Crimen por Tipología

```
br(),
br(),
h3("Choose which Crime Category want to be shown: "),
p(),
fluidRow(
  column(3,
    radioButtons("radio", label = h4("Type of Crime"),
      choices = list("Murder" = "purple", #"Rape" = 3,
        "Vehicle Theft" = "black",
        "Assault" = "orange",
        "Larceny Theft" = "grey",
        "Robbery" = "cyan",
        "Burglary" = "brown"), selected = "purple")),
  p(),
```

```
RUTA = "DATA/"
Datos <- read.csv(paste0(RUTA, "DatosMap.csv"), sep = ",")
```

```
points <- reactive({
  Datos[which(Datos$Color == input$radio),]
})
```

```
observe({
  leafletProxy("mymap", data = points()) %>%
    clearShapes() %>%
    clearMarkers() %>%
    addCircleMarkers( color = ~paste(Color),
      popup = ~paste(Label), radius = 2)
})
```


Un poco de shiny: Análisis visual de la Criminalidad

- La primera idea era hacer una aplicación responsive capaz de representar los crímenes ocurridos en Nueva York desde Enero de 2017 hasta final de Marzo de 2017:
 - Versión móvil o en github: Se ofrece una versión anterior con la mayoría de las funcionalidades que se muestran en <https://github.com/FJROAR> o se puede visualizar en cualquier dispositivo móvil (móvil, table, pc) si se va a la url <https://fjra01.shinyapps.io/MAPAS4/>
 - Algunos elementos clave de codificación: Tabla Básica

```
DatosMap.csv
1 "ID", "lat", "long", "Color", "Label"
2 845348933, 40.638018389, -73.898491201, "yellow", "ID: 845348933 Day: 03/31/2017 - Hour: 23:30:00"
3 886921338, 40.759172699, -73.988392793, "orange", "ID: 886921338 Day: 03/31/2017 - Hour: 23:25:00"
4 893265998, 40.84024096, -73.905125257, "cyan", "ID: 893265998 Day: 03/31/2017 - Hour: 23:15:00"
5 518511851, 40.861894559, -73.85766248, "yellow", "ID: 518511851 Day: 03/31/2017 - Hour: 23:00:00"
```

Agenda



-Introducción

-Un poco de shiny: Análisis visual de la criminalidad

-Previsión de la Criminalidad con Series Temporales

-Deducción de Relaciones entre Crímenes

-Conclusiones

Previsión de la Criminalidad con Series Temporales

- **El objetivo del análisis fue una previsión por semana que:**

- Distinguiera por día de la semana y hora
- Distinguiera por tipo de Crimen

- **Se usan:**

- Modelos ARIMA
- Matrices de Probabilidad por Tipología de Crimen

Previsión de la Criminalidad con Series Temporales

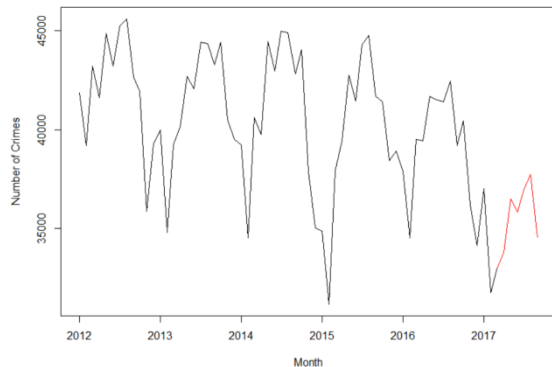
- **Modelo ARIMA**

- A nivel mensual agregado se observan ciertos comportamientos cíclicos y estacionales de los crímenes que se cometen en la ciudad de Nueva York
- Se construye una serie temporal con datos desde el 2012
- El usuario puede elegir el tipo de modelo ARMA (en el modelo a la serie se le aplica una diferencia estacional y otra no estacional)
- Se calcula a 3 meses el ECM y el EAM

Previsión de la Criminalidad con Series Temporales

- Modelo ARIMA. Función R básica de estimación

	SERIE_CRIMEN.csv
1	"Year", "Month", "Crimes", "YearMonth"
2	2012,1,41835,2012-01-01
3	2012,2,39179,2012-02-01
4	2012,3,43187,2012-03-01
5	2012,4,41597,2012-04-01
6	2012,5,44859,2012-05-01
7	2012,6,43189,2012-06-01
8	2012,7,45224,2012-07-01
9	2012,8,45582,2012-08-01
10	2012,9,42627,2012-09-01
11	2012,10,41919,2012-10-01
12	2012,11,35874,2012-11-01
13	2012,12,39262,2012-12-01



```
Modelo_con_Backtesting <- function (SERIE_CRIMEN, p, d, q, P, D, Q, periodo, mesback)
{
  SERIE_back <- SERIE_CRIMEN [1:(nrow(SERIE_CRIMEN) - mesback + 1), ]

  M_back <- arima(SERIE_back$Crimes, order = c(p, d, q),
                  seasonal = list(order = c(P, D, Q), period = periodo),
                  include.mean = FALSE)

  Back <- as.vector(predict(M_back, n.ahead = mesback)$pred)
  Real <- SERIE_CRIMEN[(nrow(SERIE_CRIMEN) - mesback + 1):(nrow(SERIE_CRIMEN)),3]

  Fiabilidad_Agregada = abs(sum(Back - Real)/sum(Real))
  Fiabilidad_EAMR = mean(abs(Back - t(Real))) / mean(t(Real))
  Fiabilidad_ECMR = (mean(abs(Back - t(Real))**2)**0.5) / mean(t(Real))

  Modelo <- arima(SERIE_CRIMEN$Crimes, order = c(p, d, q),
                  seasonal = list(order = c(P, D, Q), period = periodo),
                  include.mean = FALSE)

  Prediccion <- as.vector(predict(Modelo, n.ahead = mesback)$pred)

  Lista_final <- list(Prediccion, Fiabilidad_Agregada, Fiabilidad_EAMR, Fiabilidad_ECMR)

  return(Lista_final)
}
```


Previsión de la Criminalidad con Series Temporales

- **Matrices de Probabilidad por Tipología de Crimen**

- Se estiman probabilidades con datos recogidos a lo largo de un año (que puede ser una ventana móvil) por tipo de crimen distinguiendo por día de la semana y hora

Hora	D1	D2	D3	D4	D5	D6	D7
0	0.00600152	0.00571881	0.00618367	0.00626147	0.00684777	0.00782683	0.00736007
1	0.00381380	0.00342483	0.00395801	0.00416672	0.00469610	0.00684967	0.00671874
2	0.00310986	0.00266968	0.00311366	0.00337550	0.00377016	0.00617419	0.00601480
3	0.00257100	0.00217254	0.00239659	0.00254823	0.00295238	0.00527291	0.00581558
4	0.00240023	0.00181013	0.00207957	0.00213080	0.00240023	0.00487635	0.00531466
5	0.00194485	0.00162039	0.00167352	0.00169439	0.00198280	0.00302068	0.00332806
6	0.00215167	0.00208336	0.002224085	0.00220290	0.00230726	0.00244007	0.00241920
7	0.00317058	0.00332237	0.00360888	0.00346468	0.00338119	0.00231674	0.00233002
8	0.00531276	0.00525015	0.00573968	0.00538296	0.00536209	0.00343811	0.00321612
9	0.00579850	0.00583075	0.00588198	0.00585922	0.00575296	0.00427108	0.00366011
10	0.00605844	0.00597875	0.00641136	0.00593891	0.00591045	0.00505092	0.00453292
11	0.00609829	0.00616280	0.00634685	0.00618557	0.00643034	0.00531086	0.00458605
12	0.00623857	0.00612093	0.00652129	0.00626514	0.00612383	0.00699956	0.00631649
13	0.00684967	0.00732971	0.00753463	0.00731074	0.00716084	0.00601101	0.00548732
14	0.00796535	0.00807919	0.00863513	0.00812663	0.00819873	0.00636013	0.00597686
15	0.00870344	0.00897667	0.00946620	0.00908482	0.00919487	0.00714566	0.00674910
16	0.00842262	0.00901651	0.00936184	0.00900703	0.00906964	0.00739612	0.00701854
17	0.00852319	0.00913415	0.00929733	0.00890457	0.00939030	0.00757827	0.00706028
18	0.00841883	0.00899754	0.00929923	0.00904497	0.00921384	0.00767504	0.00745494
19	0.00760104	0.00837519	0.00873949	0.00834483	0.00880400	0.00747771	0.00709823
20	0.00767125	0.00812283	0.00836001	0.00810386	0.00824427	0.00765796	0.00726520
21	0.00621783	0.00698248	0.00745494	0.00718551	0.00786289	0.00718551	0.00626716
22	0.00568086	0.00645500	0.00619885	0.00668761	0.00757827	0.00686105	0.00603568
23	0.00505851	0.00564291	0.00556132	0.00592183	0.00714187	0.00691418	0.00560306

Hora	D1	D2	D3	D4	D5	D6	D7
0	0.00015749	0.00015938	0.00017836	0.00011005	0.00014041	0.00014420	0.00014800
1	0.00013851	0.00012333	0.00010056	0.00014041	0.00015179	0.00015749	0.00014041
2	0.00012333	0.00015559	0.00012902	0.00014610	0.00013472	0.00014041	0.00013384
3	0.00012333	0.00013851	0.00016128	0.00011954	0.00015369	0.00016128	0.00013472
4	0.00014800	0.00012333	0.00012143	0.00012333	0.00012713	0.00016697	0.00013092
5	0.00010246	0.00010626	0.00008538	0.00010436	0.00010246	0.00011005	0.00009677
6	0.00011384	0.00009867	0.00009867	0.00009108	0.00010626	0.00008159	0.00006261
7	0.00018025	0.00019354	0.00016508	0.00022579	0.00018405	0.00008159	0.00008728
8	0.00025615	0.00024856	0.00028082	0.00024287	0.00027323	0.00009487	0.00008728
9	0.00021820	0.00019733	0.00023907	0.00028651	0.00024856	0.00012523	0.00007590
10	0.00023528	0.00019923	0.00021631	0.00017077	0.00023148	0.00013092	0.00010626
11	0.00019733	0.00015559	0.00020492	0.00023148	0.00021820	0.00015749	0.00012523
12	0.00019543	0.00023718	0.00034097	0.00022200	0.00035046	0.00016128	0.00012523
13	0.00016128	0.00018784	0.00022579	0.00020113	0.00018784	0.00012713	0.00010815
14	0.00019543	0.00015559	0.00023148	0.00017646	0.00022769	0.00013472	0.00011574
15	0.00021441	0.00020113	0.00021061	0.00020113	0.00026374	0.00020113	0.00014231
16	0.00022579	0.00018974	0.00018784	0.00019923	0.00029789	0.00021061	0.0001815
17	0.00021441	0.00018595	0.00023528	0.00021251	0.00028461	0.00023388	0.00018215
18	0.00020492	0.00018215	0.00024666	0.00023528	0.00024287	0.00020492	0.00014800
19	0.00016697	0.00017836	0.00021820	0.00014041	0.00026184	0.00018215	0.00011954
20	0.00016508	0.00014800	0.00015749	0.00014800	0.00020113	0.00020682	0.00013661
21	0.00015369	0.00016128	0.00013282	0.00016508	0.00016887	0.00015749	0.00011195
22	0.00012143	0.00014041	0.00013661	0.00012143	0.00015179	0.00016508	0.00014041
23	0.00013851	0.00016508	0.00013282	0.00014990	0.00013661	0.00015179	0.00014231

Previsión de la Criminalidad con Series Temporales

- Interfaz de usuario

- El diseño shiny implementa las ideas anteriores del siguiente modo

Choose which Crime Category want to be shown:

Type of Crime

☐ Murder

☐ Vehicle Theft

☐ Assault

☐ Larceny Theft

☐ Robbery

☒ Burglary

Crime Prediction:

Number of Weekly Crimes:

Hour of Day:

Week-Days:

AR Parameter:

MA Parameter:

SAR Parameter:

SAM Parameter:

Apply Category in computations:

Crime aggregation by:

☒ None

☐ Day

☐ Hour

Crime Probability = 0.6002%

Expected Criminality = 39.01

EAM ARIMA Model = 0.1767

ECM ARIMA Model = 0.1827

Si no se coloca dato en la ventana crímenes semanales, se hace uso de un modelo ARMA

El usuario puede parametrizar modelos ARMA de una diferencia estacional y otra no estacional

Se puede agregar por hora de la semana y por día de la semana

La Criminalidad Esperada es igual al número de crímenes predichos por la probabilidad de ocurrencia

Se puede predecir probabilidades y crímenes por tipología

Se controla la fiabilidad esperada del modelo

Previsión de la Criminalidad con Series Temporales

- Parte de la Arquitectura del módulo de Previsión

```
h4("Crime Prediction:")
),
fluidRow(
  #Number
  column(4,
    numericInput("n1",
      label = h6("Number of Weekly Crimes: "),
      value = 6500)),
  # Hour
  column(2,
    numericInput("h1",
      label = h6("Hour of Day: "),
      value = 0)),
  # Day
  column(2,
    numericInput("d1",
      label = h6("Week-Days: "),
      value = 1)),
  column(1,
    numericInput("ar",
      label = h6("AR Parameter: "),
      value = 1)),
  column(1,
    numericInput("ma",
      label = h6("MA Parameter: "),
      value = 0)),

```

```
column(1,
  numericInput("sar",
    label = h6("SAR Parameter: "),
    value = 1)),
column(1,
  numericInput("sma",
    label = h6("SAM Parameter: "),
    value = 0)),
column(8,
  sidebarPanel(
    selectInput("Seg", "Apply Category in computations:",
      choices=c("No", "Yes"))),
column(3,
  radioButtons("radio2", label = h4("Crime aggregation by:"),
    choices = list("None"= 1,
      "Day"= 2,
      "Hour" = 3
    ),selected = 1))
),

```

```
output$text1 <- renderText({
  Hour = input$h1
  Day = input$d1
  Crimes = input$n1

  if(is.nan(Hour) == 1 || is.infinite(Hour) == 1 || is.na(Hour) == 1) {Hour = 0}
  if(Hour > 23) {Hour = 23}
  if(Hour < 0) {Hour = 0}
  if(is.nan(Day) == 1 || is.infinite(Day) == 1 || is.na(Day) == 1) {Day = 1}
  if(Day > 7) {Day = 7}
  if(Day < 1) {Day = 1}

  Cat = input$radio

```

```
Agg = input$radio2

if (input$Seg == 'No'){
  cathourFunction <- function (Hour, Day, Cat, Agg){
    RUTA ="DATA/"

    ProbMatrix <- read.csv(paste0(RUTA,"DayHour_Matrix.csv"), sep = ",")
    CrimeProb <- ProbMatrix[which(ProbMatrix$Hora == Hour), Day + 1]

    if (Agg == 2)
    {
      CrimeProb <- colSums(ProbMatrix)[c(2:8)]
      CrimeProb <- CrimeProb[Day]
    }

    if (Agg == 3)
    {
      CrimeProb <- rowSums(ProbMatrix[,c(2:8)])
      CrimeProb <- CrimeProb[(Hour + 1)]
    }

    return(CrimeProb)
  }
}

```

Agenda



-Introducción

-Un poco de shiny: Análisis visual de la criminalidad

-Previsión de la Criminalidad con Series Temporales

-Deducción de Relaciones entre Crímenes

-Conclusiones

Deducción de Relaciones entre Crímenes

- Selección de crímenes cercanos:

- Cercanía por diferencia de días

- Cercanía por diferencia de horas

- Cercanía por distancia

```
CrimeRelatedto <- function(ID, difDay, difHour, Color, dist)
{
```

```
  library(ISO1)
  library(data.table)
  library(geosphere)

  Datos <- Datos[which(Datos$Color != "yellow"),]
  Hora <- as.numeric(substr(Datos$Label, 39, 40))
  Dia <- as.POSIXct(as.Date(substr(Datos$Label, 20, 29), "%m/%d/%Y"))
  NumDia <- JD(Dia)
  VCrime1 <- data.frame(Datos$ID, NumDia, Hora, Datos$lat, Datos$long, Datos$Color, Datos$Label)
  names(VCrime1) <- c("ID", "NumDia", "Hora", "lat", "long", "Color", "Label")

  IndexCrime <- which(VCrime1$ID == ID)
  Datos <- Datos[which(Datos$Color != "yellow"),]

  #Proximidad por hora. Diferencia de 2h
  Hora <- as.numeric(substr(Datos$Label, 39, 40))

  #Proximidad por Día menos de 2 días
  Dia <- as.POSIXct(as.Date(substr(Datos$Label, 20, 29), "%m/%d/%Y"))
  NumDia <- JD(Dia)

  VCrime1 <- data.frame(Datos$ID, NumDia, Hora, Datos$lat, Datos$long, Datos$Color, Datos$Label)
  names(VCrime1) <- c("ID", "NumDia", "Hora", "lat", "long", "Color", "Label")

  IndexHour <- which(abs(VCrime1$NumDia[IndexCrime] - VCrime1[,2]) <= difDay)
  VCrime2 <- VCrime1[IndexHour,]

  IndexHour <- which(abs(VCrime1$Hora[IndexCrime] - VCrime2[,3]) <= difHour)
  VCrime2 <- VCrime2[IndexHour,]
```

```
  if (Color != "All"){
    VCrime2 = VCrime2[which(VCrime2$Color == Color),]
  }
  VCrime2 <- VCrime2[,c(4,5,7)]
  CrimesGeo <- VCrime2[,c(2, 1)]
  if(nrow(CrimesGeo) == 0){
    long = 0
    lat = 0
    Label = "Not valid"
    Color = "None"
    VCrime2 = data.frame(long, lat, Label, Color)
    return(VCrime2)
  }
}
```

```
  GeoBase <- VCrime1[which(VCrime1$ID == ID),][c(5, 4)]
  distancia <- distGeo(GeoBase, CrimesGeo)

  VCrime2 <- data.frame(VCrime2, distancia)

  VCrime2 <- VCrime2[which(VCrime2$distancia <= dist),]

  return(VCrime2)
}
```


Análisis de Inter-relaciones

- Interfaz de usuario:

Input ID Crime to be analyzed:

Se elije un ID válido y automáticamente todo se actualiza acorde a las restricciones

Analysis of crime relationships

- ☒ All
- ☐ Murder
- ☐ Vehicle Theft
- ☐ Assault
- ☐ Larceny Theft
- ☐ Robbery
- ☐ Burglary

Se puede seleccionar por tipología de crimen

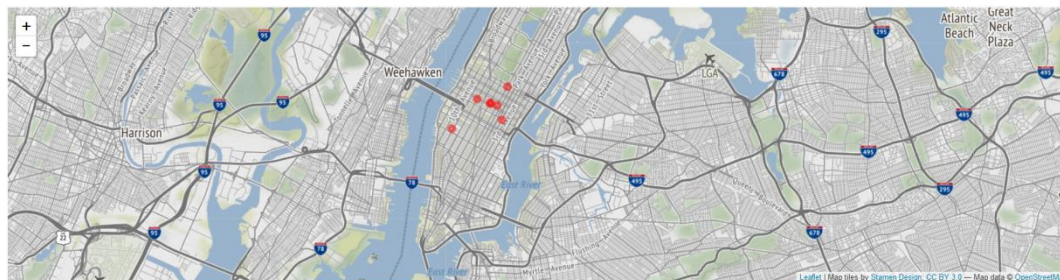
Difference in days

Difference in hours

Difference in meter

Se analiza la relación por lejanía en número de días, horas, y distancia en metros respecto al crimen de referencia

lat	long	Label	distancia
40.7592	-73.9884	ID: 886921338 Day: 03/31/2017 - Hour: 23:25:00	0
40.757	-73.9783	ID: 618408777 Day: 03/31/2017 - Hour: 21:00:00	883.117
40.7514	-73.9761	ID: 667664115 Day: 03/31/2017 - Hour: 21:00:00	1349.7375
40.7577	-73.982	ID: 820494889 Day: 03/30/2017 - Hour: 22:30:00	561.5902
40.7577	-73.982	ID: 227343174 Day: 03/30/2017 - Hour: 22:30:00	561.5902
40.7481	-74.0012	ID: 158827266 Day: 03/30/2017 - Hour: 21:00:00	1636.8087
40.764	-73.9732	ID: 870353647 Day: 03/30/2017 - Hour: 23:08:00	1386.1217



Análisis de Inter-relaciones

- Parte de la Arquitectura del módulo de Inter-relaciones

```
fluidRow(
  br(),
  br(),
  column(3,
    numericInput("ID",
      label = h2("Input ID Crime to be analyzed: "),
      value = 886921338)),
  column(3,
    radioButtons("radio5", label = h2("Analysis of crime relationships"),
      choices = list("All" = "All", "Murder" = "purple", "#Rape" = 3,
        "Vehicle Theft" = "black",
        "Assault" = "orange",
        "Larceny Theft" = "grey",
        "Robbery" = "cyan",
        "Burglary" = "brown"), selected = "All")),
  column(6,
    (
      numericInput("ddif", "Difference in days",
        value = 1 )),
    column(6,
      (
        numericInput("hdif", "Difference in hours",
          value = 2 )),
        column(6,
          (
            numericInput("mdif", "Difference in meter",
              value = 2000 )),
            )
  ),
  )
```

```
leafletOutput("mymap1",
  img(src = "TD.jpg"),  img(src = "Big.jpg"),
  fluidRow(
    column(8,
      rHandsontableOutput("tab1")
    ),
  ),
  )
```

```
points1 <- reactive({
  CrimeRelatedto(input$ID, input$ddif, input$hdif, input$radio5, input$mdif)
})
output$mymap1 <- renderLeaflet({
  leaflet() %>%
    setView(lat = 40.6643, lng = -73.9385, zoom = 12) %>%
    addProviderTiles("Stamen.Terrain",
      options = providerTileOptions(nowrap = TRUE)
    )
})
observe({
  leafletProxy("mymap1", data = points1()) %>%
    clearShapes() %>%
    clearMarkers() %>%
    addCircleMarkers( color = "red",
      #color = ~paste(Color),
      popup = ~paste(Label), radius = 4)
})
```

```
output$tab1 = renderRHandsontable({
  rhandsontable(points1(), readOnly = TRUE, selectCallback = TRUE, rowHeaders = FALSE) %>%
  hot_cols(renderer = "function (instance, td, row, col, prop, value, cellProperties) {
    Handsontable.renderers.TextRenderer.apply(this, arguments);

    td.style.background = 'orange';
  }")
})
```

Agenda



-Introducción

-Un poco de shiny: Análisis visual de la criminalidad

-Previsión de la Criminalidad con Series Temporales

-Deducción de Relaciones entre Crímenes

-Conclusiones

Conclusiones Finales

- El publicar datos interesantes, permitiría que muchos investigadores puedan aportar soluciones analíticas a problemas reales a coste muy bajo para nuestras instituciones
- Mayor información entre instituciones, el que no estén debidamente conectadas es síntoma de debilidad informática
- Con muy poco se puede hacer mucho y con cierto cuidado se puede hacer cosas respetando además los derechos que se nos garantizan desde la LOPD
- Es el momento que las instituciones abran más datos, se repercutirán a éstas beneficios que a medio-largo plazo van a ser siempre superiores a los costes incurridos
- Hacer una app con shiny con visualización en móvil es muy sencillo si sólo se sabe de R



Francisco J. Rodríguez Aragón, PhD
Senior Data Scientist

Teradata Iberia

María de Portugal, 1-3-5
28050 MADRID
+34 682277976

francisco.rodriguezaragon@Thinkbiganalytics.com

francisco.rodriguezaragon@teradata.com

teradata.com

Advanced Analytics Data Platforms | Applications | Services