AGGRESSIVE BEHAVIOR PREDICTION IN BOGOTÁ CITY, USING TIME SERIES

Por: Jorge Victorino

Grupo de modelado computacional de la facultad de Ciencias Universidad Nacional de Colombia

- Francisco Gómez
- Jorge Rudas
- Ana María Reyes
- Jorge Victorino
- Miguel Barrero
- Cristian Pulido
- Luisa Chaparro
- Luz Angela Narvaez

AGGRESSIVE BEHAVIORS PREDICTION

Predicting and identifying crime patterns is an essential task in the design of policies for the safety of citizens [Baraniul, 2018]

Fights and aggressive behaviors correspond to violent actions
 to harm physical or psychological pain [Kim, 2017].

In Bogotá, fights leading cause of crimes, with 70% such as homicides and personal injuries [Greenberg 2018, Secretaria, 2018]

CLASSIFICATION OF FIGHTS

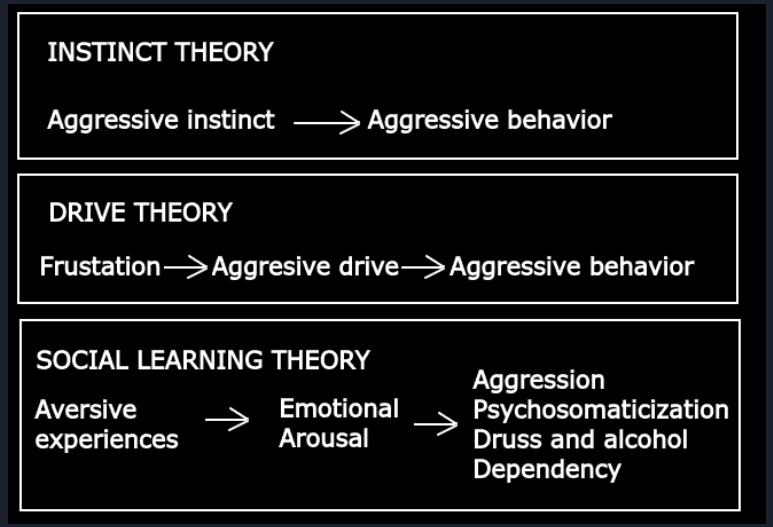
Drive theory [Davitz 1952]: fights over drivers in traffic, fights between relatives

Social learning theory [Bandura 1978]: fights between fan groups

Deindividualization theory [Zimbardo 1969]: fights between fan groups

Instinct theory [Bandura 1978]: defend against an aggressor

CLASSIFICATION OF FIGHTS



Theoretical formulations couched in terms of frustrating instigators and injurious aims have limited explanatory power [Bandura, 1973]"

DATA SOURCE

Source	NUSE	SIEDCO	RNMC
Rows	2.254.000	141.270	912.556
Time	6 Years	10 Years	10 Years
Origin	Citizen report	Complaints	Infraction
Advantage	Quantity	Confirmed	Confirmed
Disavantage	Repeated	Underreporting	Underreporting

Theoretical formulations couched in terms of frustrating instigators and injurious aims have limited explanatory power [Bandura, 1973]"

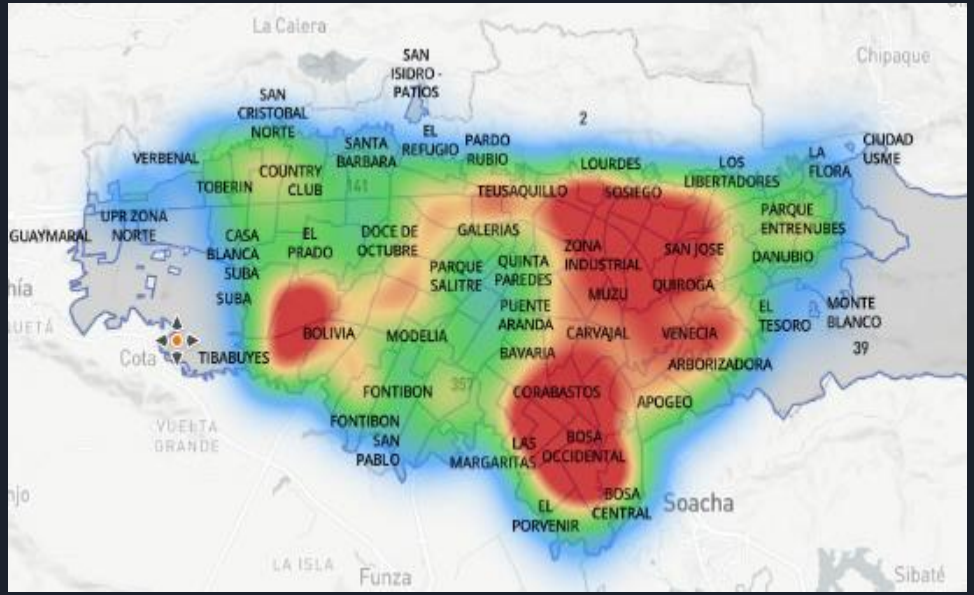
METHODS FOR CRIME PREDICTION

Based on [Wheeler & Steenbeek, 2020], the methods for crime prediction belong to one of three classes

- Hot spot maps
- Point processes
- Regression based methods

Hot spots maps

- Identify areas with elevated crime rates
- It dividing space into cells: (grid or political boundaries)
- Hot spots in the past are the same location in the future
- Spatio temporal events are sparse
- Problem: counts are sensitive to scale and offset
- Problem: The measures widely vary in the selected area unit



Spatio-Temporal Point Processes (STPP)

- STPP estimates an intensity function that predicts the rate of incidents
- Poisson point processes:
- Hawkes processes, Self-Exciting Point Process (**SEPP**):

$$\lambda(s,t|H_t) = \mu(s) + \sum_{i:t_i<t} g(s-s_i, t-t_i)$$

- Extended model with covariates:

$$\lambda(s,t) = exp(\beta X_{C(s)}) + \sum_{i:t_i<t} g(s-s_i, t-t_i, M_i)$$

- Examples: retaliation homicides, vandalism damage, looting or pillage
- Problem: Self-excitation assumption is not always true

Regression based methods

- Provide a simple and interpretable method
  Allowing multiple regression with covariates and regularization
- generally, a linear relationship is assumed between the variables
- Allowing longer-term forecasts
- Problem: counts are sensitive to scale and offset
- Problem: high sensitivity to setting parameters and outliers

Methods for crime prediction, problems

- Assumptions and conditions that are not true
  Biased sensing of incidents
- Sparse data and non-homogeneous representation
- Low interpretability of the results and low usability of the system
- Expert knowledge and experience are not used
- Do not combine computing power with the analysis capacity of expert humans

COVARIATES

- Static covariates help to understand the dynamics of the phenomenon
- Time-varying covariates improve the model prediction
- The dynamics of the covariates is latent in the data
- Problem: An unknown set of covariates affect the phenomenon and its respective weights are unknown
- Problem: The availability of reliable Time-varying covariates are hard to come by and maintain

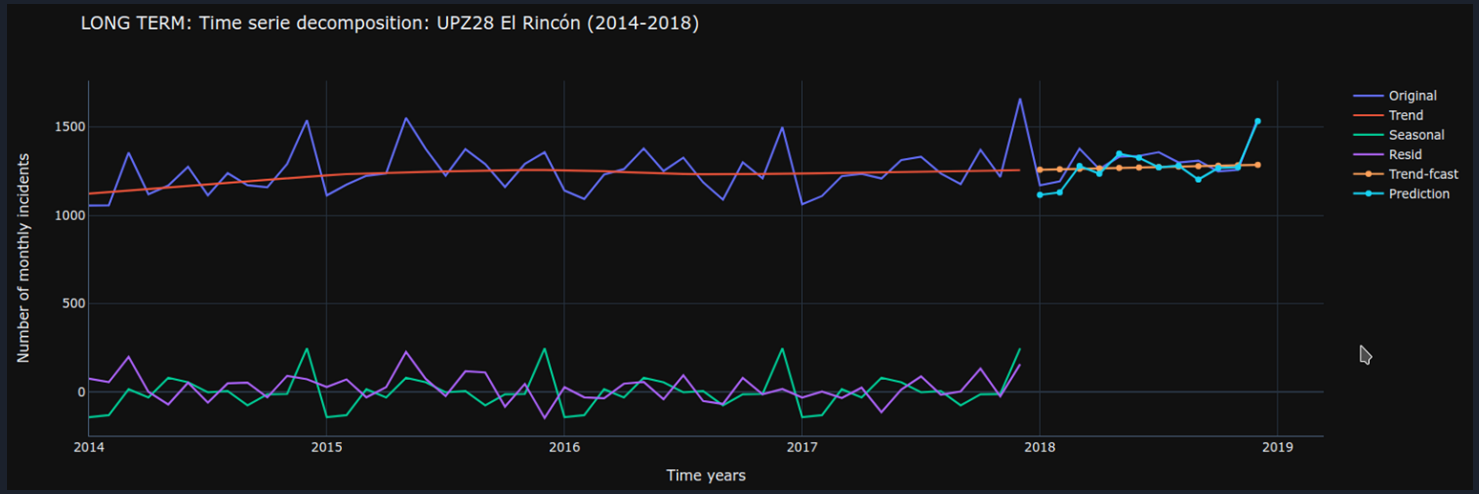
THEORY OF ROUTINE ACTIVITY

- Aggressive behaviors is related to the social-environmental features
- When certain circumstances converge in space and time, the probability of incidents increases
- The theory suggests three elements to characterize the occurrence of incidents:
    - **The rhythm**, **The tempo and The timing**



PREDICTION OF AGGRESSIVE BEHAVIORS

- Quantitative analysis of the reported incidents of aggressive behaviors using Time Series
- Time serie decompostion in: trend, seasonal and residual
- Characterization of rhythm through the seasonal component of the time series
- Characterization of tempo through the trend component of the time series
- Advantage: Simple and interpretable model, allows interactive visual analysis
- Advantage: Long term and short term predictions



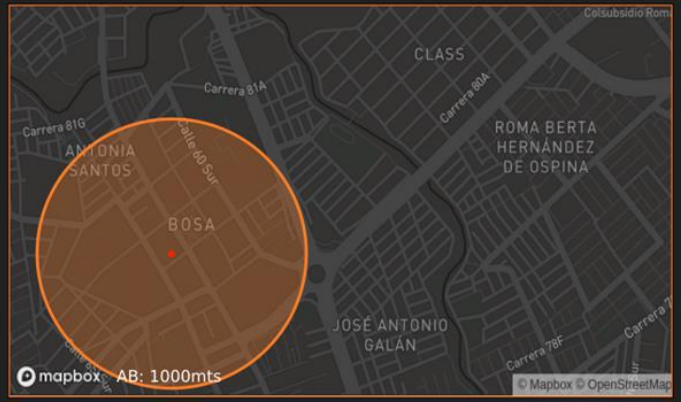
TIME SERIES CHALLENGES

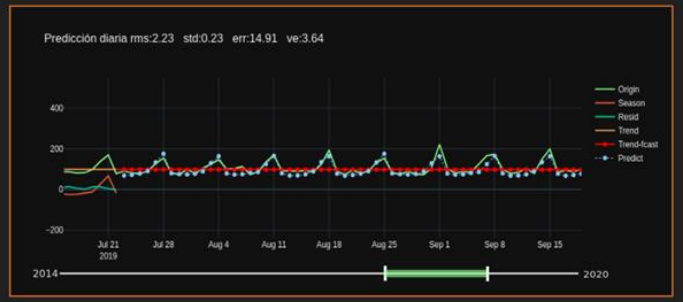
The model is highly sensitive to:

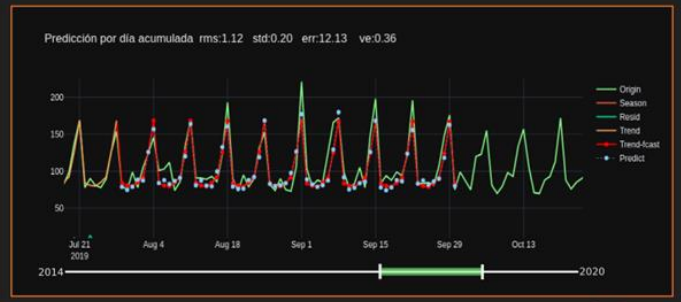
- Discretization in space and time
- Scale and Shift
- Data discontinuity
- Setting parameters
- Time window selection
- Outliers
- Data selection

PROPOSED APPROACH

- Model based in Visual Analytics
- Interactive parameter setting
- Select long, middle, and short term prediction
- Generator of multiple time series
- Interactive time serie decomposition
- Outlier enhancement for markup and annotations
- Expert user knowledge learning module (interactions and setting)
- Orientation process and guidance [Ceneda et al, 2019]
- Assistance in performance evaluation
- High confidence and interpretability of results







Configuración de la predicción

Tipo de ventana de tiempo

Corto plazo | Mediano plazo | Largo plazo

Ventana de tiempo de los datos para la predicción

Fecha de inicio: 2015-01-22
Fecha de fin: 2019-07-22

2014 — 2020

Influencia de los datos en la predicción

- [ ] Lineal
- [ ] Exponencial creciente
- [ ] Exponencial decreciente
- [ ] Potencia

Función: e^x

Ventana de tiempo de predicción

Semanas: 10

2014 — 2020

Etiquetado de valores atípicos

Etiquetar

- [ ] Manifestaciones
- [ ] Festividades
- [ ] Eventos deportivos
- [ ] Eventos culturales

Predecir

Resultados

Serie semanal
Índice de confianza

Serie diaria
Índice de confianza

Serie diaria acumulada
Índice de confianza

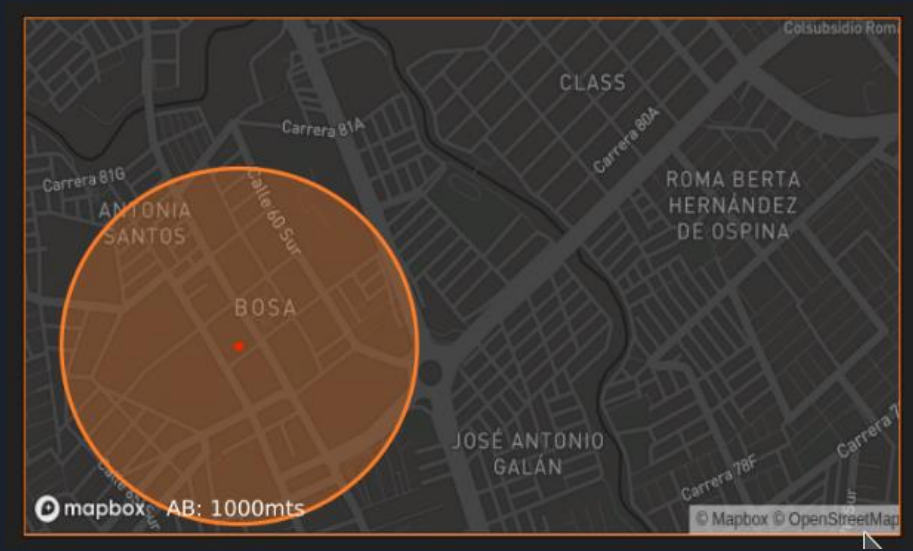
Orientation Procedure

- T0 Select User profile
- T1 Assitance in select point prediction
- T2 Select Time window of incidents and prediction
- T3 Exploration of time series
- T4 Assistence in the treatment of outliers
- T5 Interactive selection of model parameters
- T6 Selection of prediction

Based en guidance methodology by [Ceneda et al., 2019]

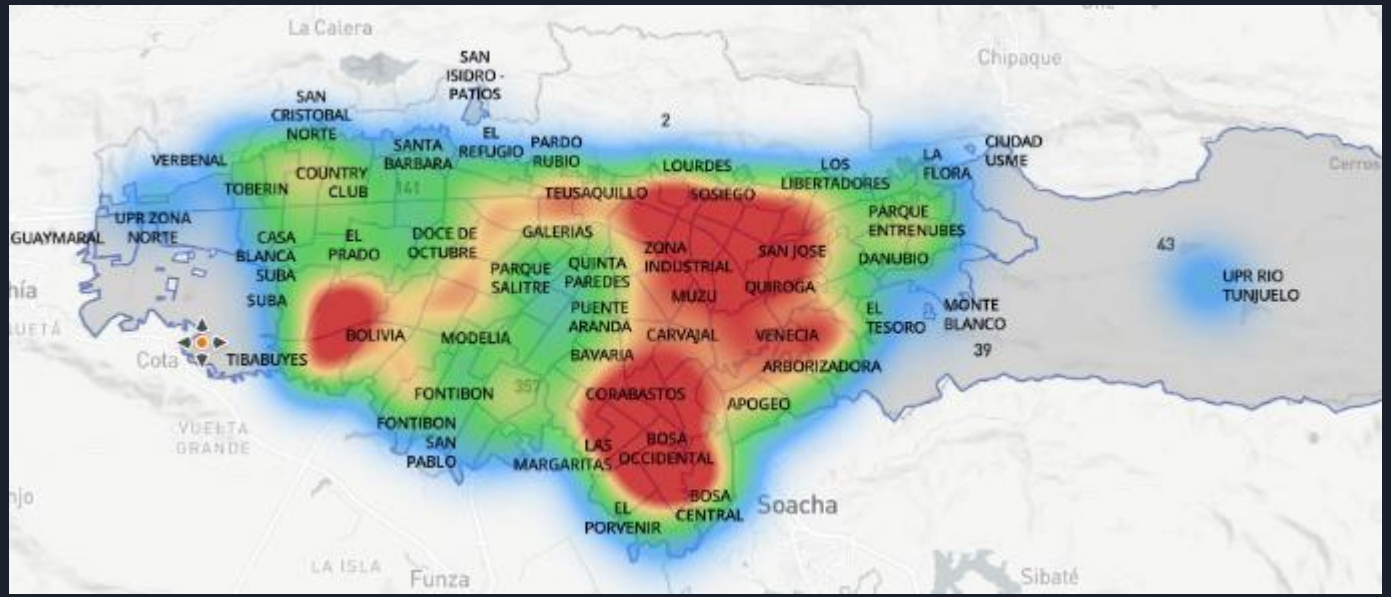
T0 SELECT USER PROFILE

- The system has been designed for unique profile
- Taking in to account a professional of the "Secretaría de seguridad"



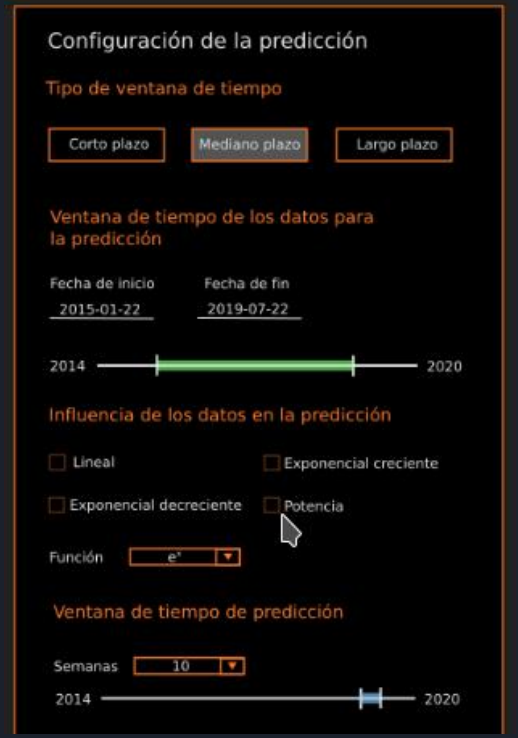
T1 ASSITANCE IN SELECT POINT PREDICTION

- Select point in the interactive map (latitude, longitude)
- Assistance: map of hotspots in Bogotá
- knowledge gap: user with visual methods



T2 Select Time window of incidents and prediction

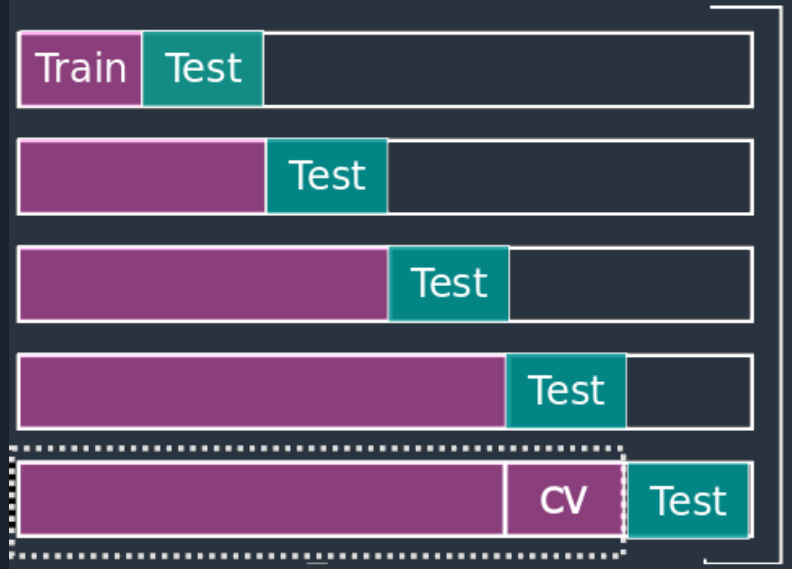
- Buttons to select: long, middle or short term prediction
- Interactive selection of incidents window
- Interactive selection of prediction window
- Assistance: calendar visualization
- Knowledge gap: user with visual methods



T3 Exploration of time series

- knowledge gap: system -> relevant information
- Performance measures

Nested Cross-Validation

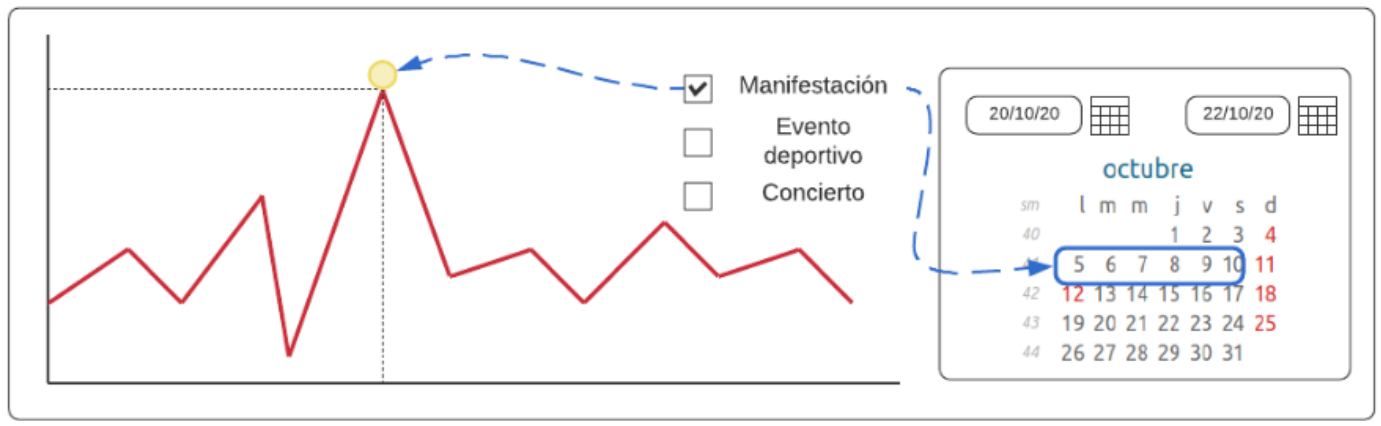


ANOVA

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares (MS)	F
Within	$SSW = \sum_{j=1}^{k}\sum_{j=1}^{l}(X-\bar{X}_j)^2$	$df_w = k-1$	$MSW = \frac{SSW}{df_w}$	$F = \frac{MSB}{MSW}$
Between	$SSB = \sum_{j=1}^{k}(\bar{X}_j-\bar{X})^2$	$df_b = n-k$	$MSB = \frac{SSB}{df_b}$	
Total	$SST = \sum_{j=1}^{n}(\bar{X}_j-\bar{X})^2$	$df_t = n-1$		

T4 Assistance in the treatment of outliers

- Show outliers on the different time series
- Allows logging or labeling of the outlier trigger event
- Highlight previously marked outliers
- Decide whether or not to leave the outlier
- Knowledge gap: system -> manage the outliers
- Knowledge gap: system -> learn causes of outliers

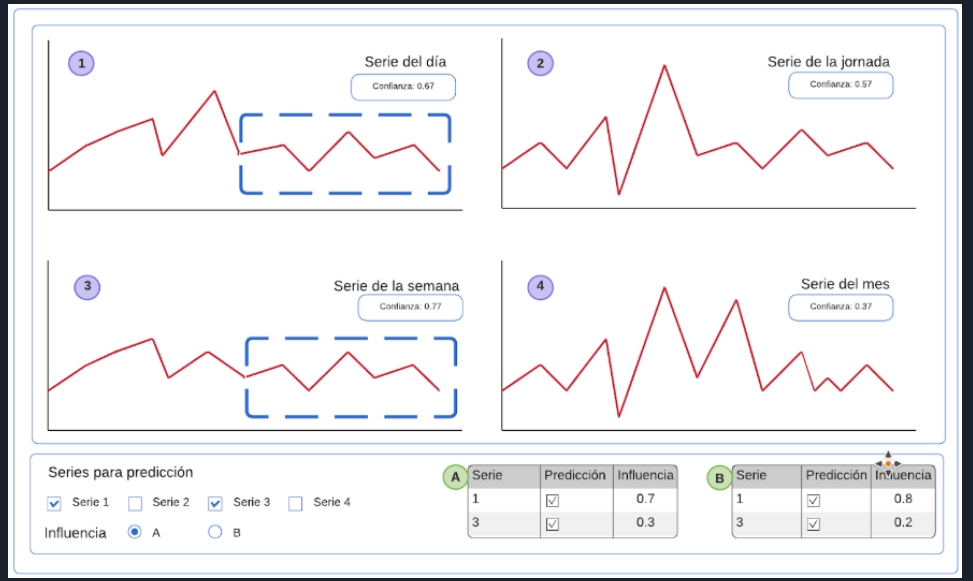


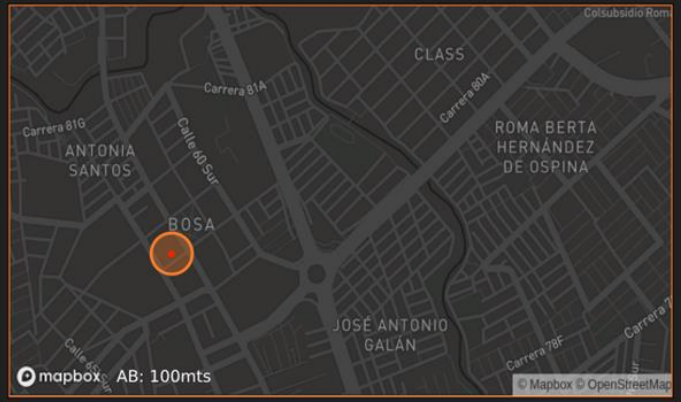
T5 Interactive selection of model parameters

- Select function to modify the degree of influence of data serie
- View statistics as feedback to make decisions
- Knowledge gap: system -> setting parameters

T6 Selection of prediction

- Reduce the time series using customizable weights
- Labeling events in prediction time window
- Setting weights to combine de time serie for prediction
- Knowledge gap: system -> Combine the time-series





Configuración de la predicción

Tipo de ventana de tiempo

Corto plazo | Mediano plazo | Largo plazo

Ventana de tiempo de los datos para la predicción

Fecha de inicio: 2015-01-22
Fecha de fin: 2019-07-22

2014 — 2020

Influencia de los datos en la predicción

- [ ] Lineal
- [ ] Exponencial creciente
- [ ] Exponencial decreciente
- [ ] Potencia

Función: e^x

Ventana de tiempo de predicción

Semanas: 10

2014 — 2020

Etiquetado de valores atípicos

Etiquetar

- [ ] Manifestaciones
- [ ] Festividades
- [ ] Eventos deportivos
- [ ] Eventos culturales

Predecir

Resultados

Serie semanal
Índice de confianza

Serie diaria
Índice de confianza

Serie diaria acumulada
Índice de confianza

Questions?