



Dirección Nacional de Datos e Información Pública

Subsecretaría de Innovación Pública y Gobierno Abierto
Ministerio de Modernización

Dr. Martin Elias Costa
Analista de datos
@meliascosta

Lic. Natalia Sampietro
Dir. de Datos públicos
@nasampietro



Datos 3.0: *Data-driven government*



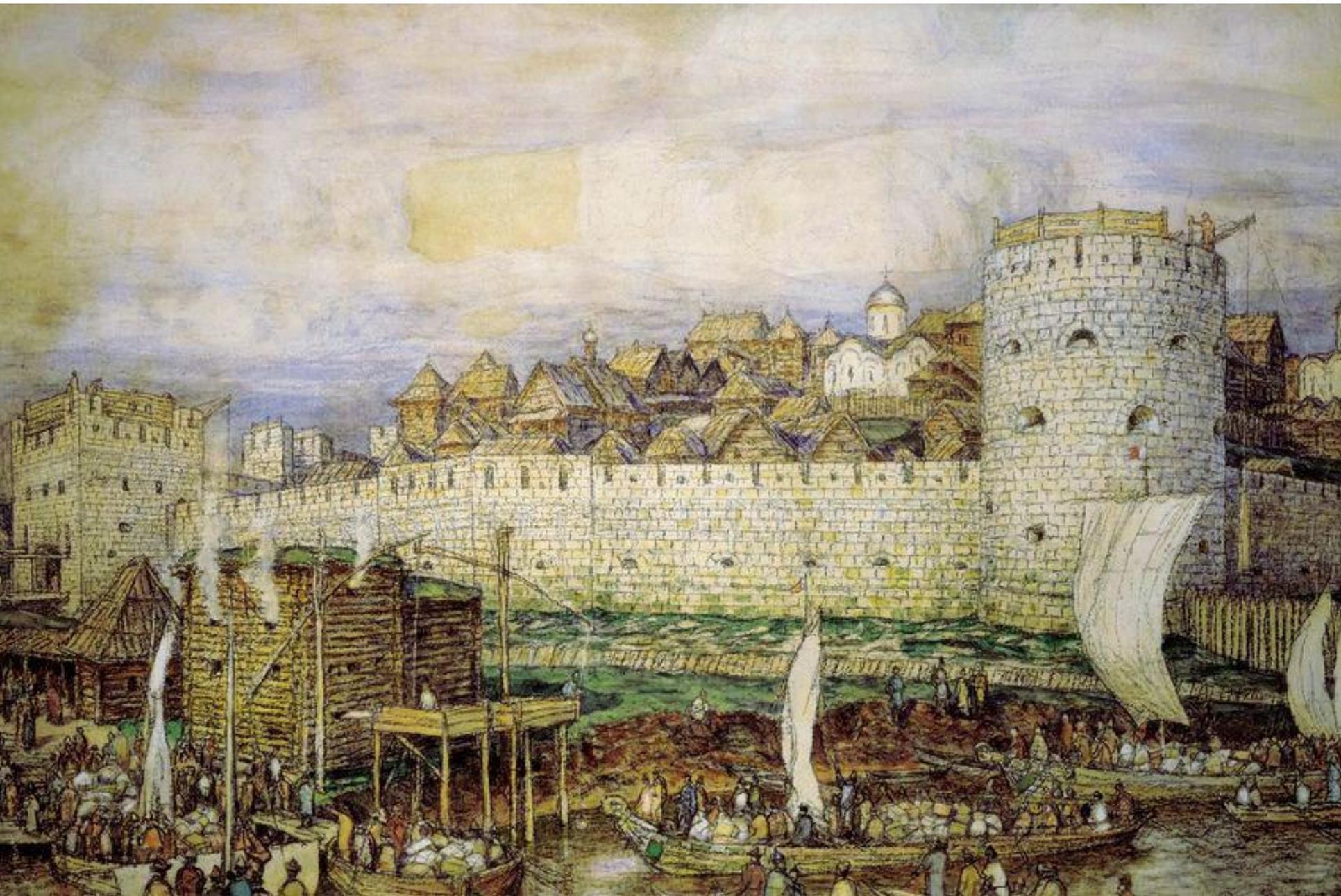
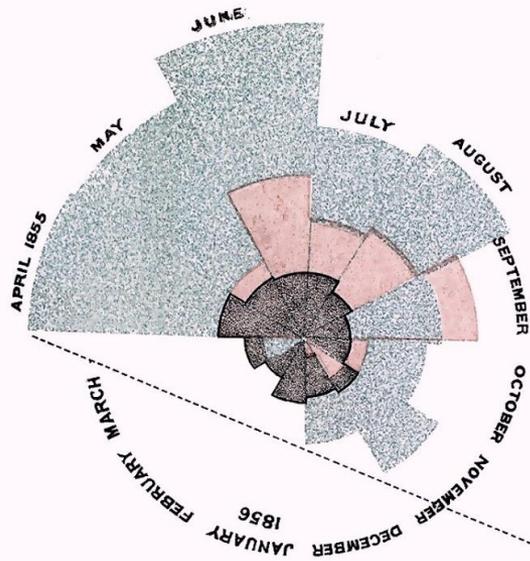
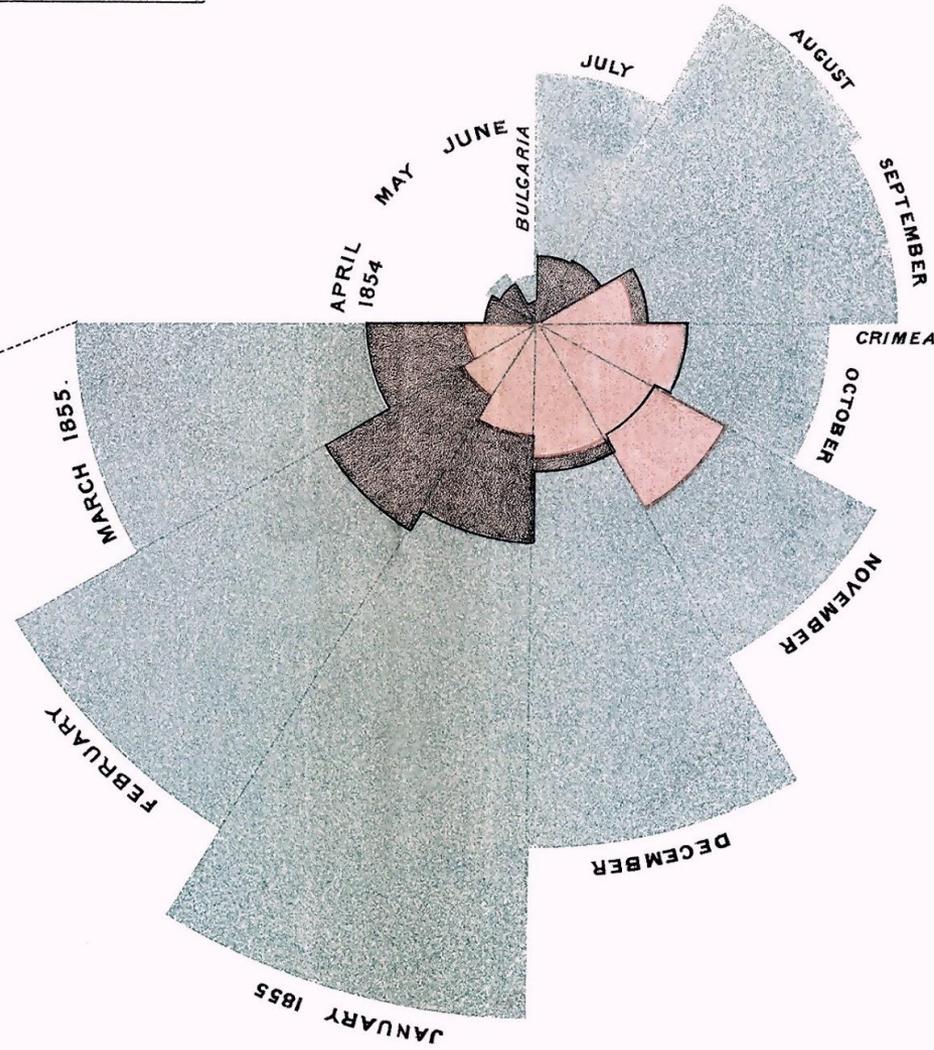


DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855; the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Un estudio de IBM del 2013 revelaba que **solo el 50% de los decisores hacen al menos la mitad de sus decisiones basados en datos y minería de datos.**

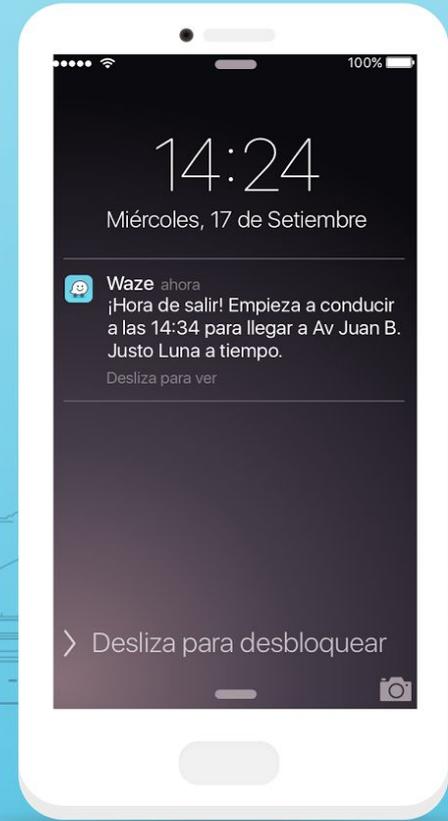
La razón... **no disponen de los datos y la capacidad para analizarlos.**

Data-driven government: Challenges and a path forward, IBM Analytics. White Paper October 2015



NETFLIX







Analytics para auditar edificios en New York City

<http://blog.datalook.io/using-data-analytics-to-make-bad-buildings-better-in-new-york-city/>

FUTURE

PAST

Enter an address

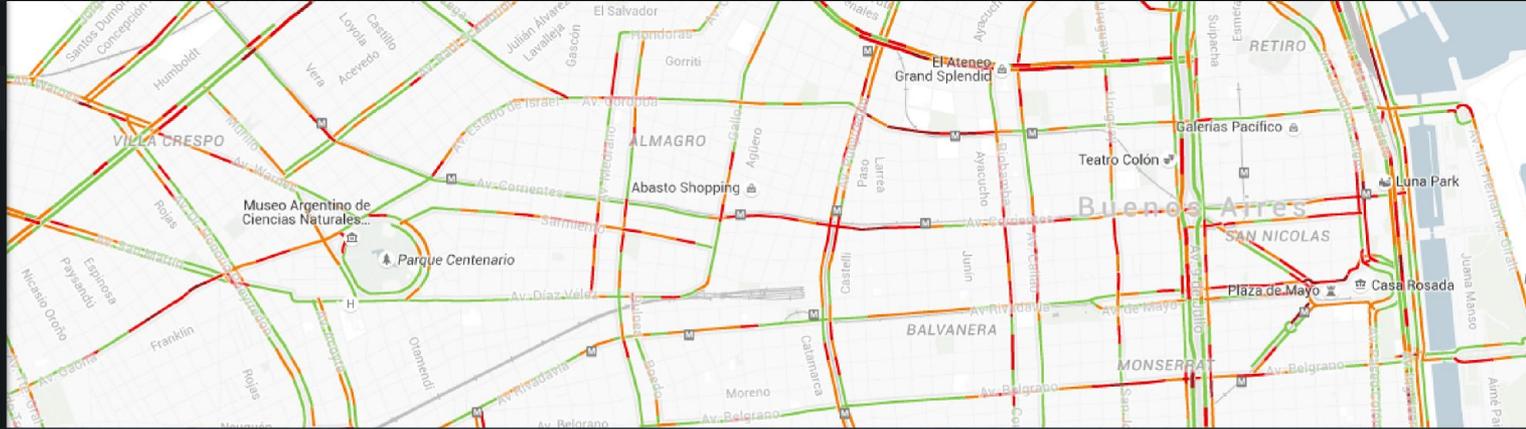
Machine learning para predecir crimen en Río.



CRIME FORECAST

NOW	TODAY 10/10	TUESDAY 11/10	WEDNESDAY 12/10	THURSDAY 13/10	FRIDAY 14/10	SATURDAY 15/10
16:44	EVE	NIGHT MRNG AFTN EVE	NIGHT MRNG AFTN EVE	NIGHT MRNG AFTN EVE	NIGHT MRNG AFTN EVE	NIGHT MRNG AFTN EVE
HIGH	MED	LOW LOW HIGH MED	LOW MED HIGH MED	LOW LOW HIGH MED	LOW LOW HIGH MED	LOW LOW HIGH MED

- Alcorta ✓
- Libertador !
- 9 de Julio ✓
- Illia ✓
- Córdoba ✓
- Pueyrredón ✓
- Av. de Mayo ✓
- Rivadavia ✓
- Independencia !
- Paseo Colón ✓
- Alem !
- Corrientes ✓
- JuanB. Justo !
- Cabildo !
- San Martín ✓



Análisis en tiempo real de los accesos y egresos a la Ciudad

Corredor Libertador

← A CAPITAL

Aviso de anomalía Hace 15 min. Causa [?]

Aviso de anomalía Hace 15 min. Causa [?]

Aviso de anomalía Hace 15 min. Causa [?]

Reportar estado

	CERRITO	SARMIENTO	LA PAMPA	GENERAL PAZ
Aviso de anomalía	---	Aviso de anomalía	---	---
Tiempo del trayecto	20 min.	Tiempo del trayecto	20 min.	Tiempo del trayecto 20 min.
Demora	---	Demora	---	---
Causa	[?]	Causa	[?]	Causa [?]

→ A PROVINCIA

Estado

TRAYECTO CERRITO - SARMIENTO

SENTIDO A Capital

ANOMALÍA Crítica

DEMORA +10% (+5')

CAUSA Seleccionar causa...

DESCRIPCIÓN

Escriba aquí...

REPORTAR MODIFICAR

Para que los datos y las capacidades estén...

```
42     email: @el.find('.email-input').val().trim()
43     telephone: @el.find('.telephone-input').val().trim()
44     country: country
45     id_type: if person.country == 'Argentina' then @el.find('.id-type-input').val() else country
46   }
47   applicantAttr = {
48     occupation: @el.find('.position-input').val().trim()
49     absent: @el.find('applicant-didnt-participated').is('checked')
50     person: personAttr
51   }
52
53   valid = true
54
55   personNameValid = @validateName(personAttr.name)
56   valid = valid and personNameValid
57   @el.find('.name-input').toggleClass('invalid', !personNameValid)
58
59   personSurnameValid = @validateName(personAttr.surname)
60   valid = valid and personSurnameValid
61   @el.find('.surname-input').toggleClass('invalid', !personSurnameValid)
62
63   personIdValid = @validatePersonId(personAttr.person_id, personAttr.country)
64   valid = valid and personIdValid
65   @el.find('.person-id-input').toggleClass('invalid', !personIdValid)
66
67   personEmailValid = @validateEmail(personAttr.email)
68   valid = valid and personEmailValid
69   @el.find('.email-input').toggleClass('invalid', !personEmailValid)
70
71   personCountryValid = @validateCountry(personAttr.country)
72   valid = valid and personCountryValid
73   @el.find('.countries-select').toggleClass('invalid', !personCountryValid)
74
75   applicantOccupationValid = @validateName(applicantAttr.occupation)
76   valid = valid and applicantOccupationValid
77   @el.find('.position-input').toggleClass('invalid', !applicantOccupationValid)
78
79   if valid
80     @updateApplicant(applicantAttr)
81
82   updateApplicant (applicantData) =>
83     data = {
84       audience: {
85         id: @audience.get('id'),
86         new: !!@audience.get('new')
87       },
88       applicant: applicantData
89     }
90
91     if data.audience.new
92       data.audience.obligee_id = audiencias.globals.obliges.currentObligee().get('id')
93       data.audience.author_id = audiencias.globals.users.currentUser().get('id')
94
95     $.ajax(
96       url: '/intranet/editor_audencia'
97       method: 'POST'
98       data: data
99     )
100     success: (response) =>
101       if response.success and response.audience
102         response.audience.editingApplicant = false
103         response.audience.new = false
104         @audience_forceUpdate(response.audience)
```

Gestión de Datos e Información



Infraestructura. Generación del contexto propicio para una gestión eficiente de datos e información. (contexto humano, normativo, tecnológico, etc.)

Gestión de Datos e Información



Hub de Datos. Identificación e integración de distintas fuentes de datos, favoreciendo la interoperabilidad entre las mismas.



Datos Abiertos



Coordinación de la política de datos abiertos y el **portal de datos públicos** de la Administración Pública Nacional (datos.gob.ar)



Datos Argentina

Portal de datos abiertos del Gobierno de la República Argentina. Acá encontrarás información pública, herramientas y recursos para desarrollar aplicaciones y visualizaciones y crear historias con datos.

009
DATASETS

014
ORGANIZACIONES

Buscar Dataset



Administración Pública

Cultura y Recreación

Economía y Finanzas

Educación

Estadísticas

Infraestructura y Obras Públicas

Medio Ambiente

Movilidad y Transporte

Salud y Servicios Sociales

Seguridad y Justicia

Urbanismo y Territorio

Datasets destacados

Declaraciones Juradas Patrimoniales Integrales

Ministerio de Justicia y Derechos Humanos. Oficina Anticorrupción

Datos correspondiente a las Declaraciones Juradas Patrimoniales Integrales presentadas por los sujetos obligados conforme lo normado

Ejecución Presupuestaria de la Administración Pública Nacional

Ministerio de Hacienda. Secretaría de Hacienda. Subsecretaría de Presupuesto.



Datos MINEM

Iniciativa de apertura de datos del Ministerio de Energía y Minería.

Buscar Dataset

257 Datasets disponibles

 Exploración producción hidrocarburos (12)	 Transporte hidrocarburos (10)	 Refinación de hidrocarburos (2)	 Comercialización de hidrocarburos (39)	 Generación energía eléctrica (4)	 Transporte energía eléctrica (3)	 Distribución energía eléctrica (39)	 Energía renovables (24)
 Minería (9)	 Otros (15)	 Socioeconómico (97)	 Obras (3)				

Etiquetas (178) ▾

Formatos (4) ▾

Ordenar ▾

Balanza comercial minera

Balanza comercial de sustancias minerales y productos derivados totales. Años 2001-2015 (en millones de US\$ FOB). Esta serie refleja la evolución del aporte neto de divisas del sector minero argentino.

Barrios Ciudad Autónoma de Buenos Aires

Límites y ubicación geográfica de los barrios de la Ciudad Autónoma de Buenos Aires. ISO 19126: SettlementsAssociated, Este subgrupo es un grupo de conceptos que...

Cantidad de usuarios de gas de red por segmento de calle

La presente capa muestra la cantidad de usuarios de gas de red que existe por vereda (par e impar) en todo el país. Es...

Cantidad de Usuarios por Provincia de Redes de Gas Natural

Número total de usuarios, por provincia - Datos de ENARGAS.



Datos Abiertos Agroindustria

En el Portal de Datos Abiertos del Ministerio de Agroindustria podrás obtener datos numéricos, estadísticos y referenciados geográficamente del sector agropecuario. Ingresá periódicamente y descubrí la información nueva y actualizada.

Podrás buscar y descargar los datos que prefieras.



Categorías



Producción

Actividad primaria



Agroindustria

Industria
agropecuaria



Comercio

Intercambio
comercial



Inversión

Financiamiento e
inversión



Precios

Cultivos y
productos



Padrones

Establecimientos y
productores



IDE

Información
Geo-Referenciada

Datos

Encontrá en esta plataforma las últimas actualizaciones de los datos abiertos de PAMI.

019
DATASETS

Grupos



Organizaciones

PAMI (19)

Etiquetas

Servicios (11)

Ubicación (8)

Gestión (8)

Médicos (7)

Dirección (7)

Salud (6)

Prestadores (6)

Buscar Dataset



Ordenar:

Ultima modificación



Organigrama

Información sobre la estructura y autoridades del Instituto.



pdf

Consejos Participativos

Información sobre los consejos participativos realizados con el objetivo de profundizar el diálogo y la integración entre los afiliados y el Instituto.



pdf

Campaña de Vacunación Antigripal

Información de la campaña de vacunación antigripal.



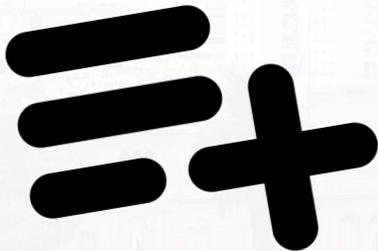
csv

xlsx

Prestadores Médicos



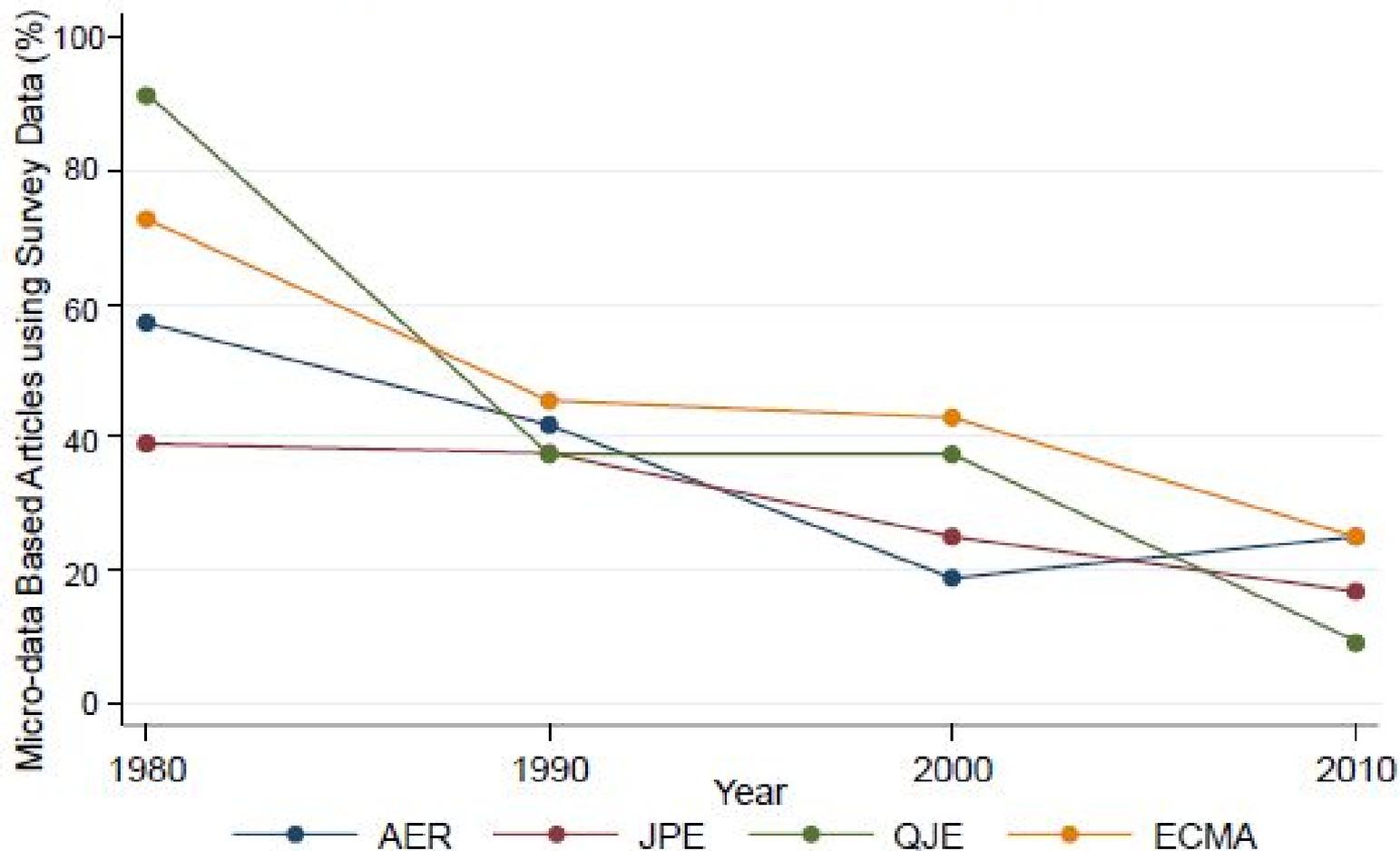
Valor agregado sobre los Datos



Exploración y análisis de los datos para la **generación de producto** interno y externo. Promoviendo la participación de reutilizadores internos y externos.



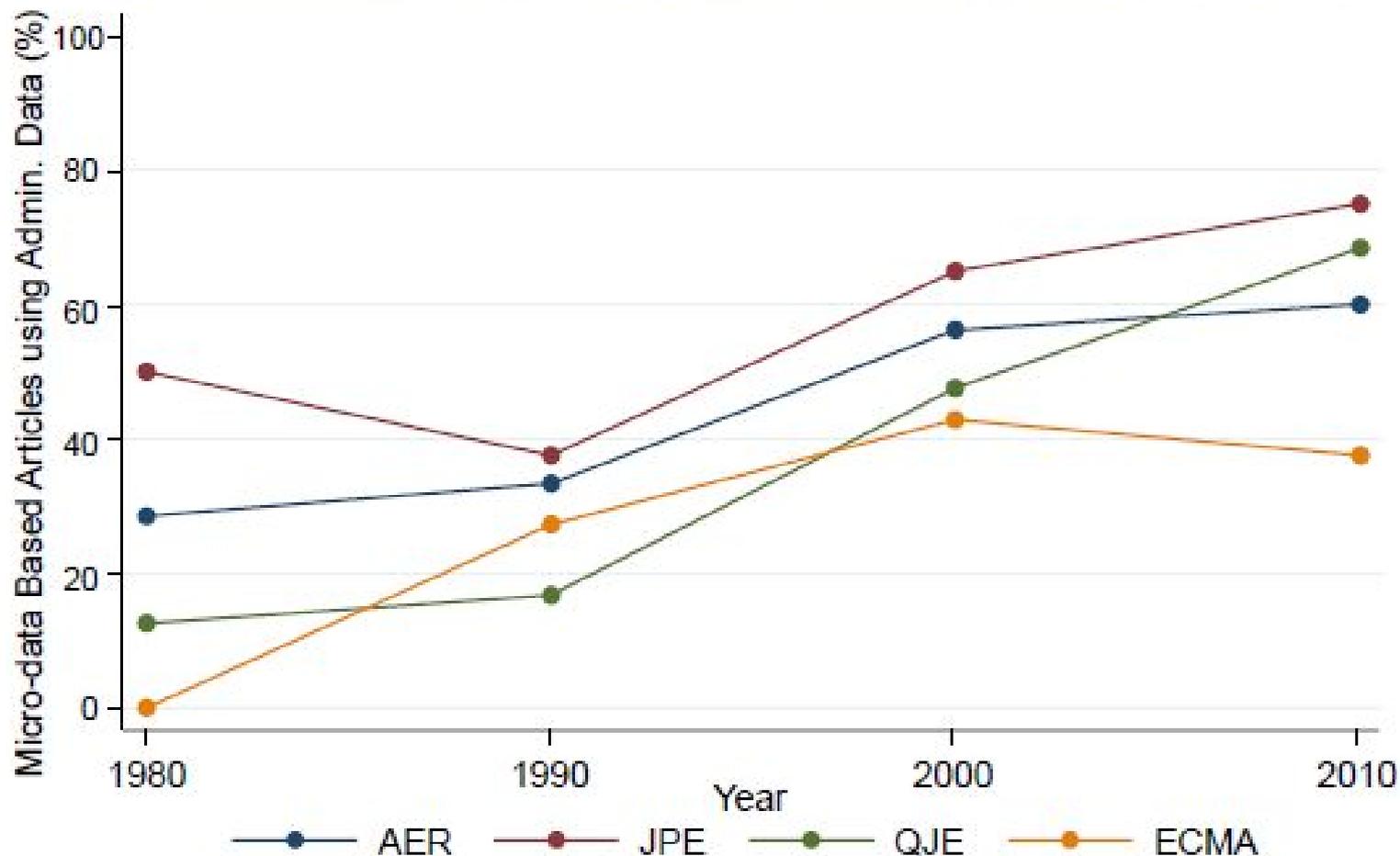
Use of Pre-Existing Survey Data in Publications in Leading Journals, 1980-2010



Note: "Pre-existing survey" datasets refer to micro surveys such as the CPS or SIPP and do not include surveys designed by researchers for their study. Sample excludes studies whose primary data source is from developing countries.

Fuente: Time Trends in the Use of Administrative Data for Empirical Research. Raj Chetty, Harvard and NBER (July 2012)

Use of Administrative Data in Publications in Leading Journals, 1980-2010



Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records).

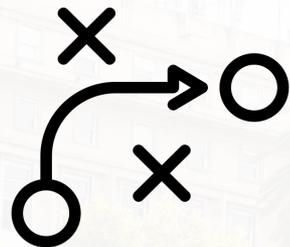
Sample excludes studies whose primary data source is from developing countries.

Fuente: Time Trends in the Use of Administrative Data for Empirical Research. Raj Chetty, Harvard and NBER (July 2012)

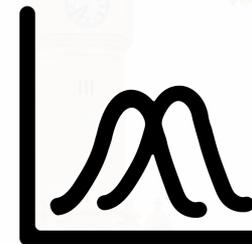
¿Cómo lo hacemos?

```
42     email: @el.find('.email-input').val().trim()
43     telephone: @el.find('.telephone-input').val().trim()
44     country: country
45     id_type: if person.country == 'Argentina' then @el.find('.id-type-input').val() else country
46   }
47   applicantAttr = {
48     occupation: @el.find('.position-input').val().trim()
49     absent: @el.find('applicant-didnt-participated').is(':checked')
50     person: personAttr
51   }
52
53   valid = true
54
55   personNameValid = @validateName(personAttr.name)
56   valid = valid and personNameValid
57   @el.find('.name-input').toggleClass('invalid', !personNameValid)
58
59   personSurnameValid = @validateName(personAttr.surname)
60   valid = valid and personSurnameValid
61   @el.find('.surname-input').toggleClass('invalid', !personSurnameValid)
62
63   personIdValid = @validatePersonId(personAttr.person_id, personAttr.country)
64   valid = valid and personIdValid
65   @el.find('.person-id-input').toggleClass('invalid', !personIdValid)
66
67   personEmailValid = @validateEmail(personAttr.email)
68   valid = valid and personEmailValid
69   @el.find('.email-input').toggleClass('invalid', !personEmailValid)
70
71   personCountryValid = @validateCountry(personAttr.country)
72   valid = valid and personCountryValid
73   @el.find('.countries-select').toggleClass('invalid', !personCountryValid)
74
75   applicantOccupationValid = @validateName(applicantAttr.occupation)
76   valid = valid and applicantOccupationValid
77   @el.find('.position-input').toggleClass('invalid', !applicantOccupationValid)
78
79   if valid
80     @updateApplicant(applicantAttr)
81
82   updateApplicant (applicantData) =>
83     data = {
84       audience: {
85         id: @audience.get('id'),
86         new: !!@audience.get('new')
87       },
88       applicant: applicantData
89     }
90
91   if data.audience.new
92     data.audience.obligee_id = audiencias.globals.obliges.currentObligee().get('id')
93     data.audience.author_id = audiencias.globals.users.currentUser().get('id')
94
95   $.ajax(
96     url: '/intranet/editor_audencia'
97     method: 'POST'
98     data: data
99     success: (response) =>
100       if response.success and response.audience
101         response.audience.editingApplicant = false
102         response.audience.new = false
103         @audience_forceUpdate(response.audience)
```

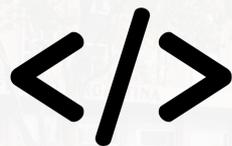
DN de Datos e Información Pública



Hacedores de
Políticas



Data Scientist

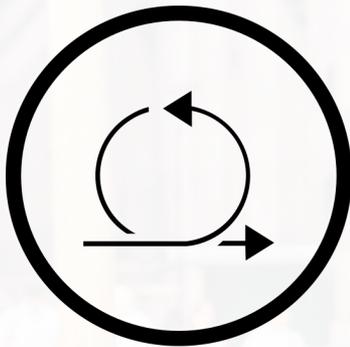


Desarrolladores



Diseñadores

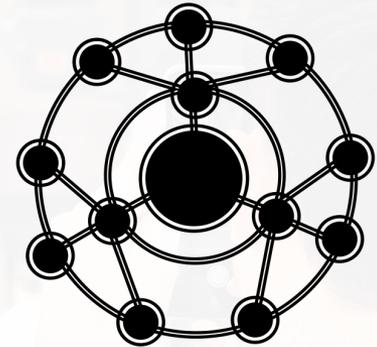
DN de Datos e Información Pública



Ágil

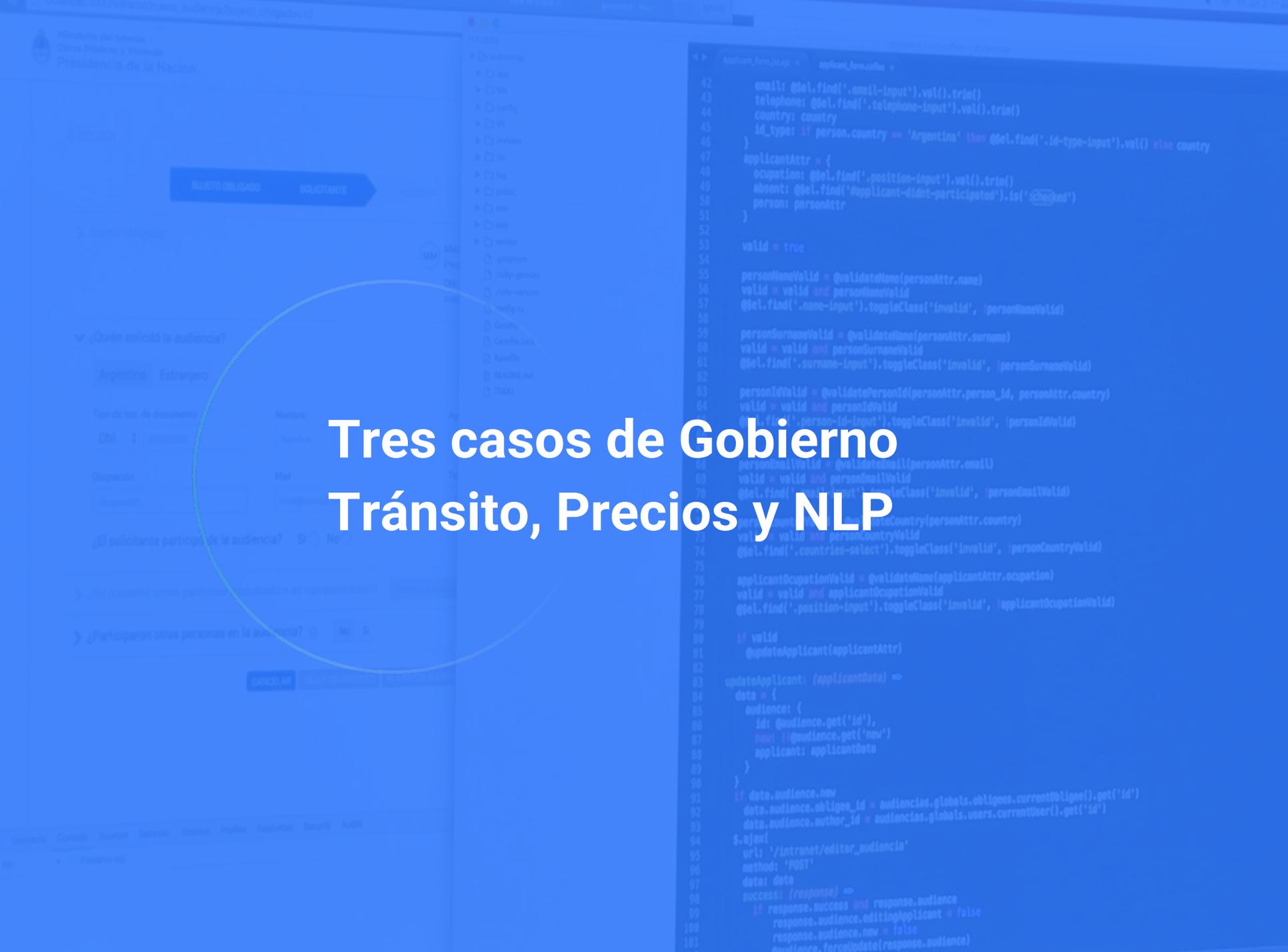


Abierto



Conectado

Tres casos de Gobierno Tránsito, Precios y NLP

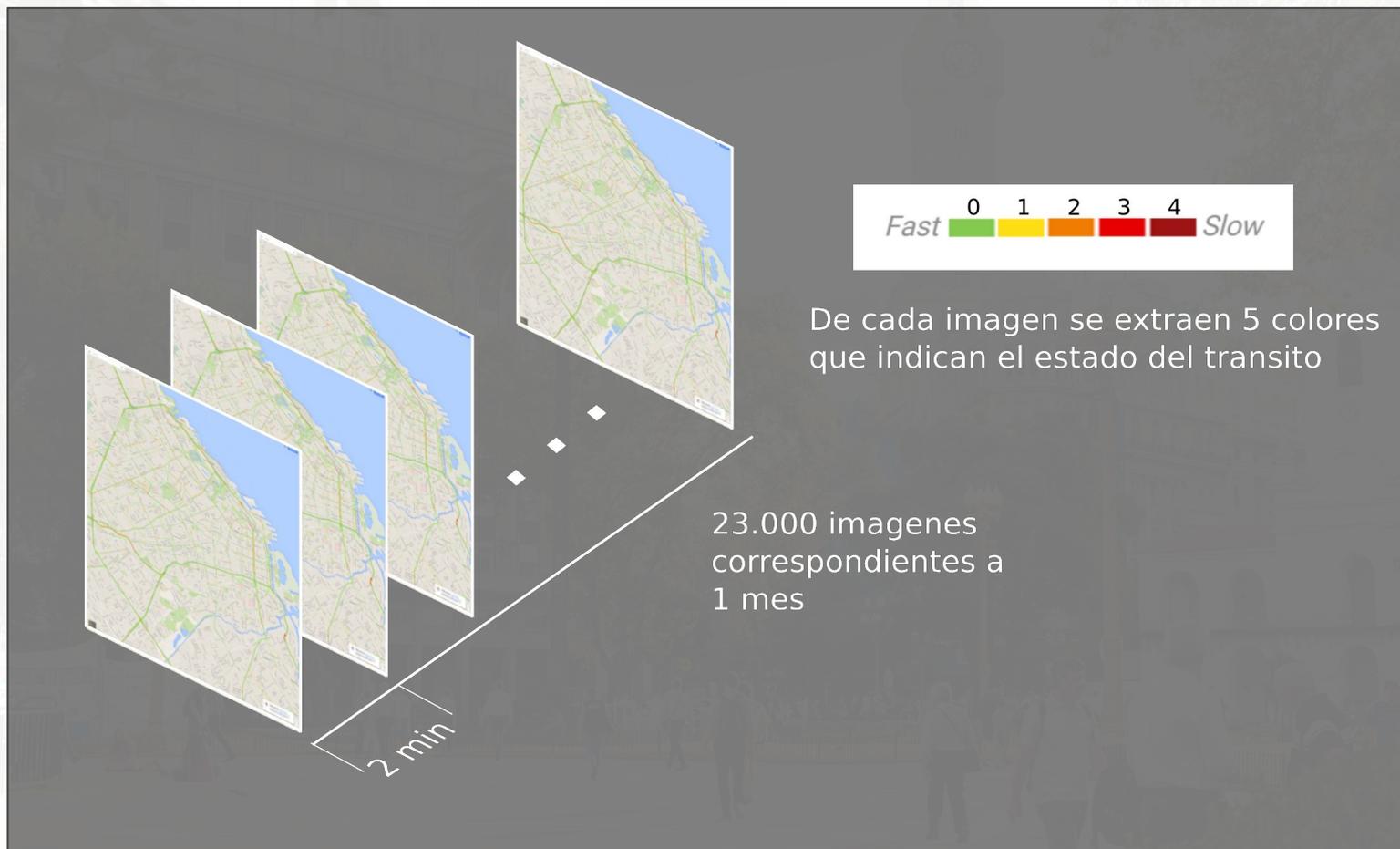


Monitoreo de Tránsito en CABA



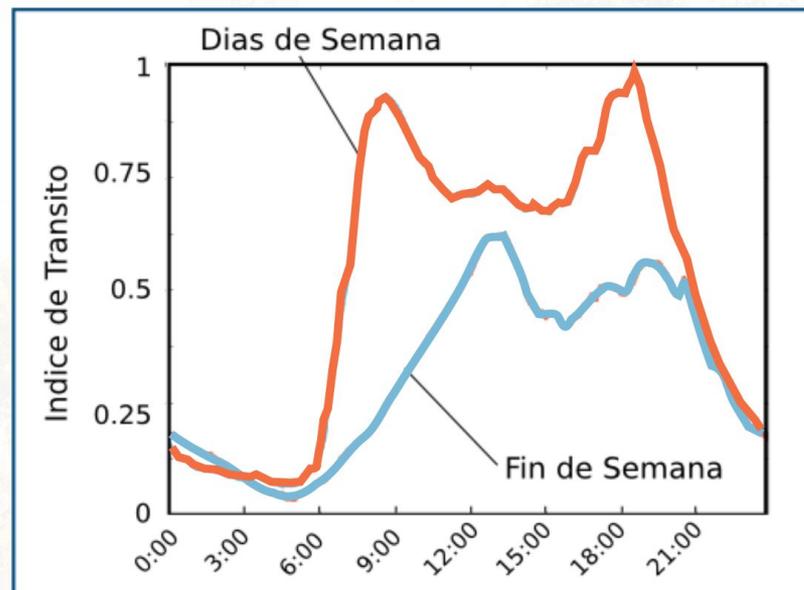
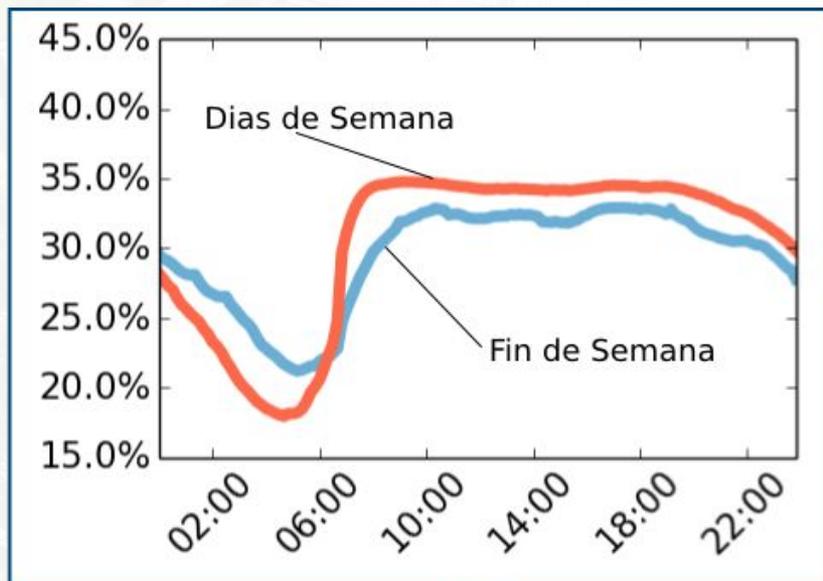
Tránsito

Monitoreo con Google Maps



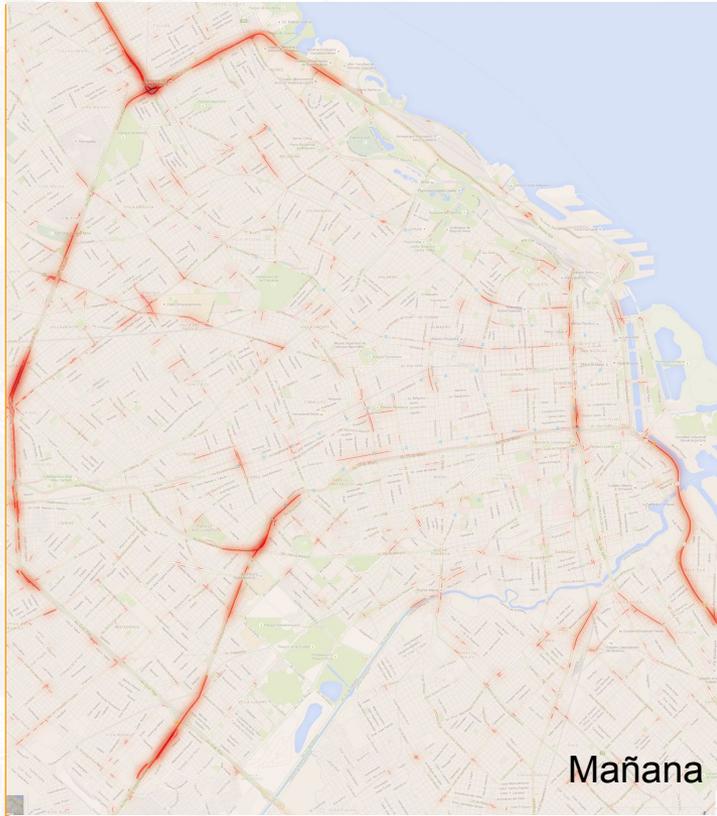
Tránsito

Monitoreo con Google Maps



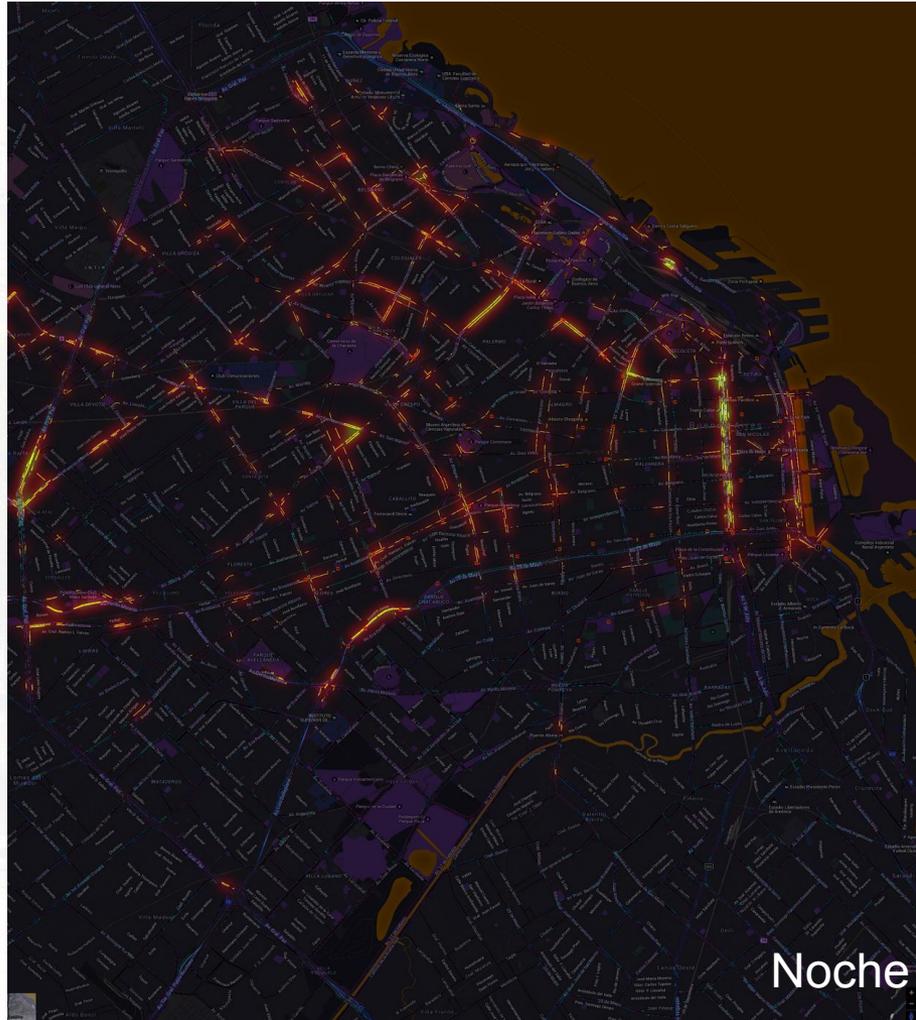
Tránsito

Monitoreo con Google Maps



Tránsito

Monitoreo con Google Maps

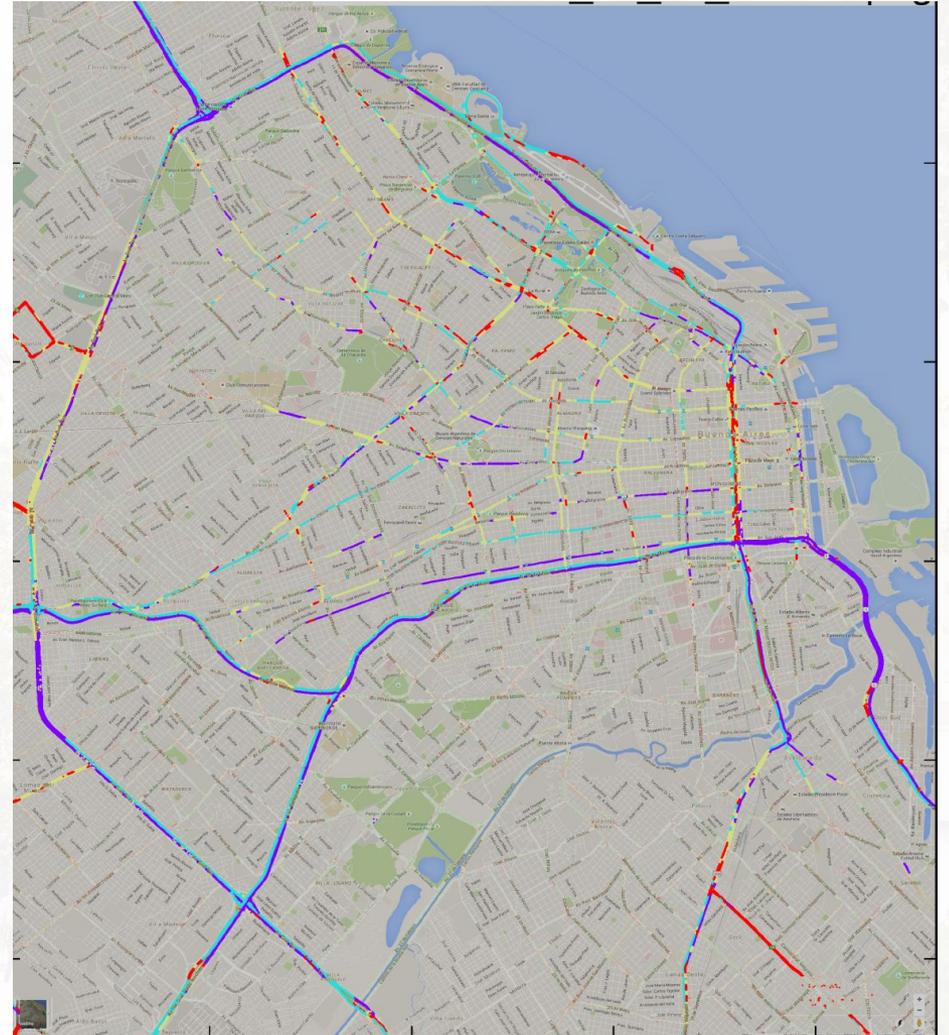
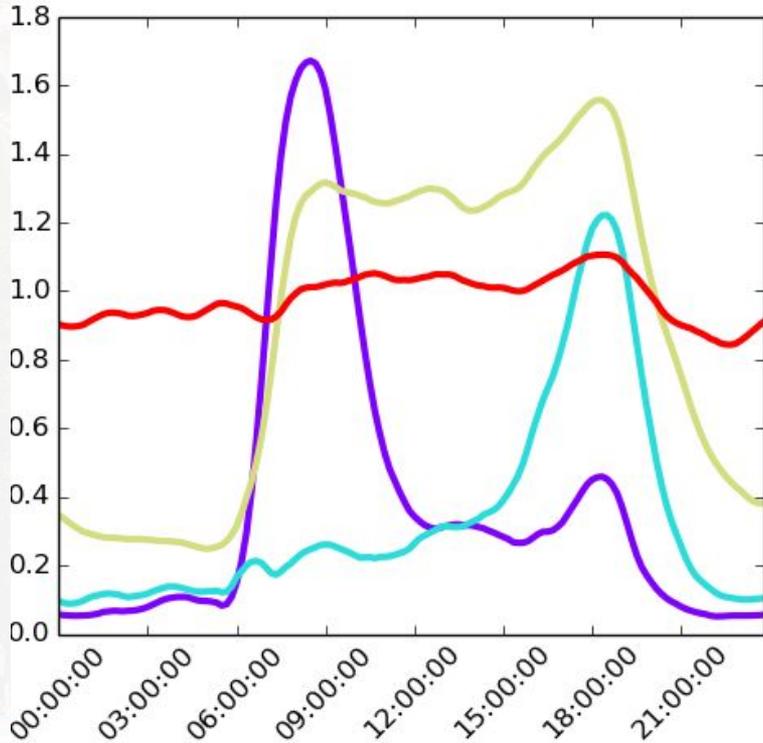


Clasificación de corredores

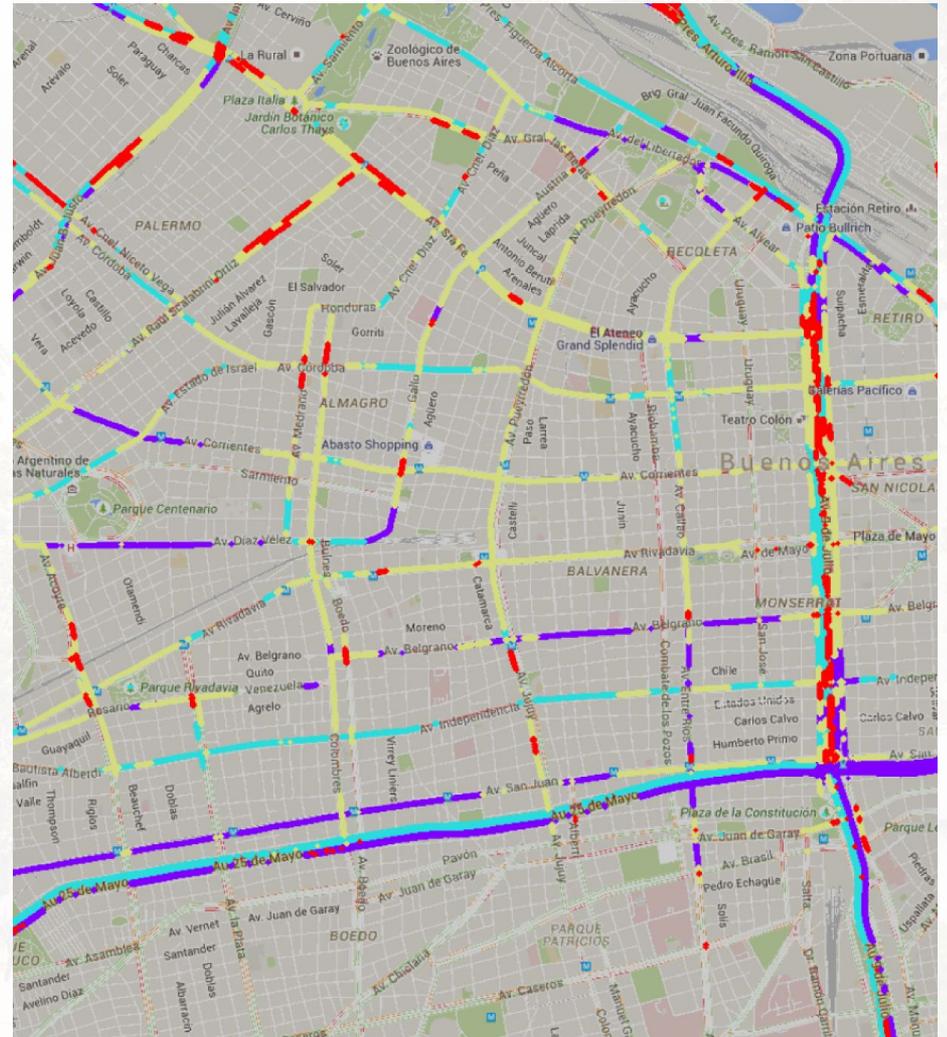
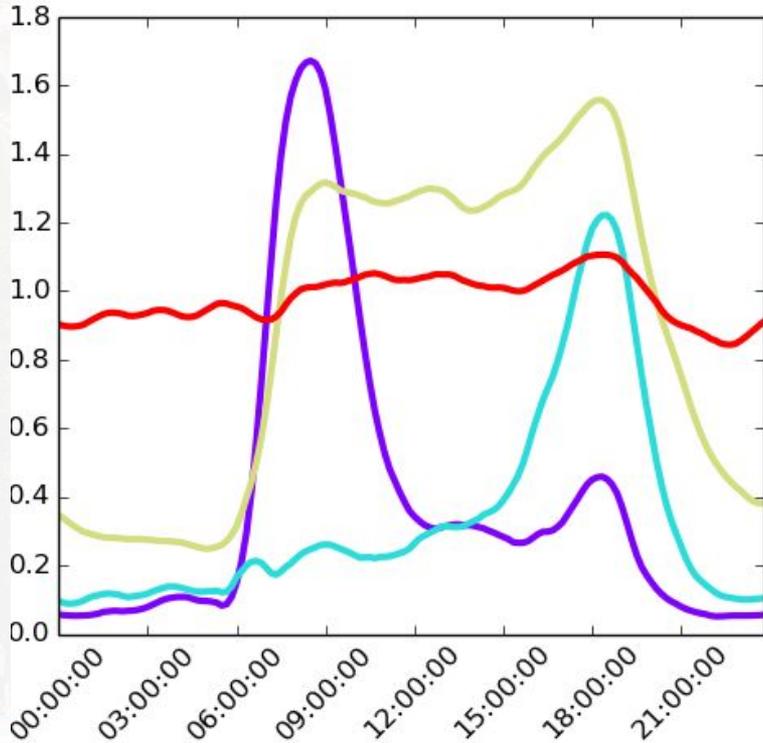
Clasificación de corredores por patrones similares



Clusters



Clusters



Procesamiento del Lenguaje Natural en el Congreso



NLP

Preguntas en el Congreso



¿Cómo organizar esta información no estructurada?

Jefatura de Gabinete



Informes al Congreso

Preguntas de legisladores

NLP

Preguntas en el Congreso

¿Cómo puedo categorizar las preguntas que recibo eficientemente?

Etiquetado automático

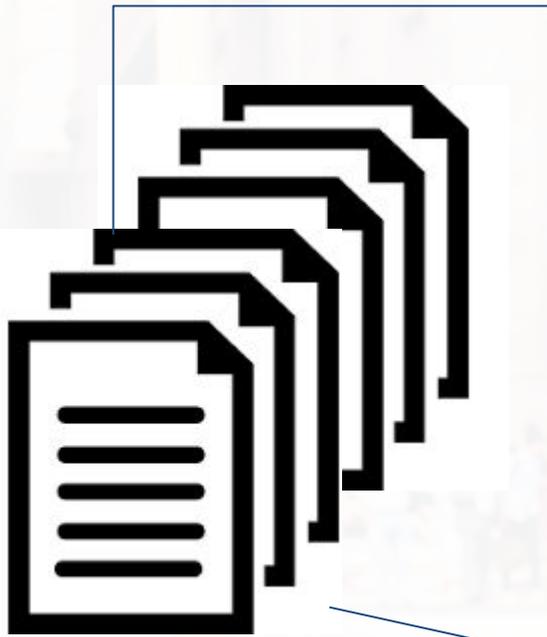
¿Cómo puedo saber si ya respondí preguntas similares?

Búsqueda de textos similares



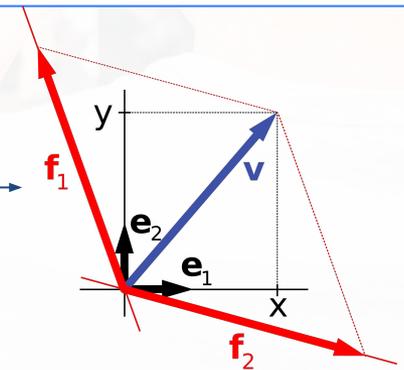
NLP

Vectorización de textos



En cuanto al relevamiento de la recuperación de armas de fuego y municiones, indique como fue realizada esa recuperación teniendo en cuenta: -por cada provincia -por tipos de armas de fuego por municiones considerando cada tipo aclarando quienes han sido las personas que hicieron la entrega de las mismas; si eran "legales" y cuántas provenían del "mercado gris" (armas no registradas ni usadas para delinquir") y del negro; las recibidas por el programa de entrega voluntaria; como las incautadas por cuestiones administrativas y aquellas miles enviadas por los poderes judiciales con orden de destrucción.

TF - IDF



NLP

Vectorización de textos

$$\text{TF-IDF}_{(n,d)} = \text{TF}_{(n,d)} \times \text{IDF}_{(n)}$$

The diagram illustrates the TF-IDF formula. The term $\text{TF-IDF}_{(n,d)}$ is bracketed in red and linked to a box labeled "Peso de un término (n) en un documento (d)". The term $\text{TF}_{(n,d)}$ is bracketed in blue and linked to a box labeled "Frecuencia de aparición de un término (n) en un documento (d)". The term $\text{IDF}_{(n)}$ is bracketed in yellow and linked to a box labeled "Factor IDF de un término (n)".

En cuanto al relevamiento de la **recuperación** de **armas** de **fuego** y **municiones**, indique como fue realizada esa **recuperación** teniendo en cuenta: -por cada provincia -por **tipos** de **armas** de **fuego** por **municiones** considerando cada tipo aclarando quienes han sido las personas que hicieron la **entrega** de las mismas; si eran "**legales**" y cuántas provenían del "mercado gris" (**armas** no **registradas** ni usadas para delinquir") y del negro; las recibidas por el programa de **entrega** voluntaria; como las incautadas por cuestiones administrativas y aquellas **miles** enviadas por los poderes judiciales con orden de **destrucción**.

Palabras con mayor TFIDF

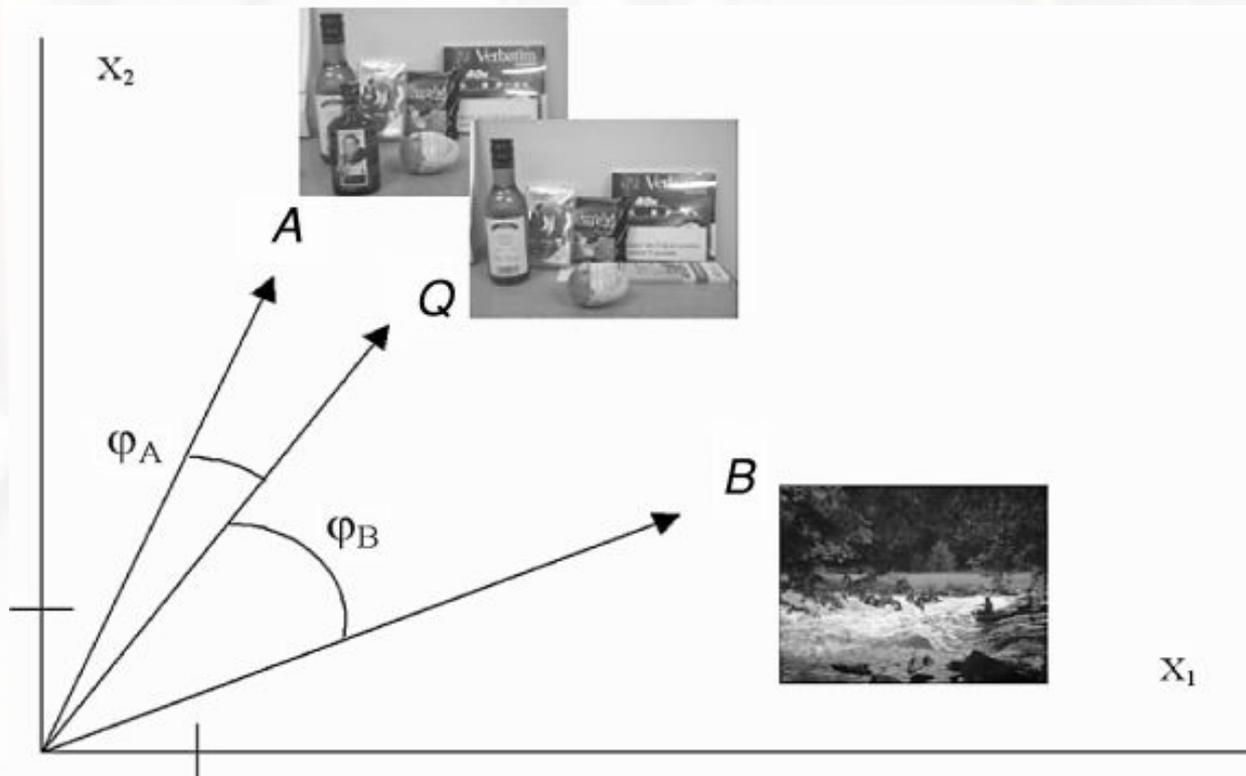
armas
municiones
recuperación
entrega
fuego
destrucción
miles
legales
tipos
registradas



NLP

Búsqueda de textos similares

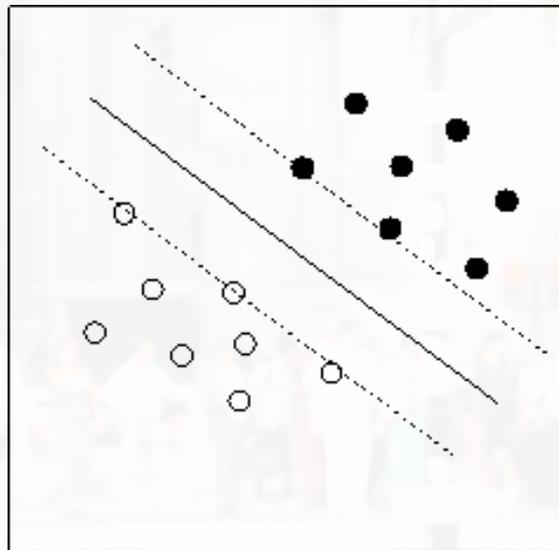
Cosine similarity



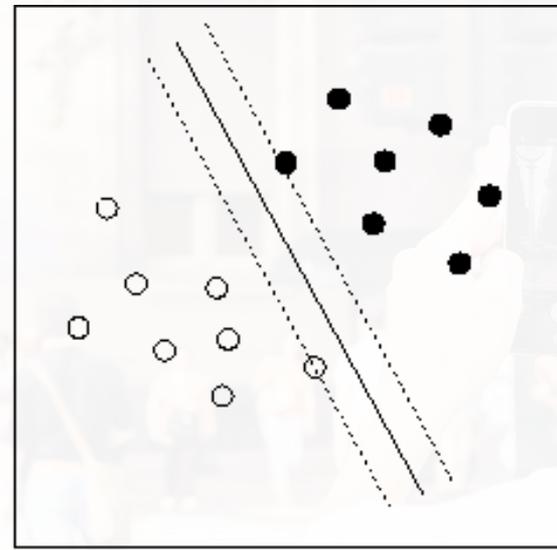
NLP

Etiquetado automático

Support Vector Machines



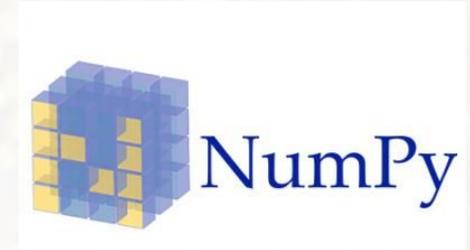
(a) Larger margin



(b) Smaller margin

Con 700 casos de entrenamiento se logra una performance del 80% de etiquetas correctas entre las primeras 5 de 20 opciones (50% para la primera opción).

Herramientas usadas:



Seaborn



<http://datos.gob.ar>



<https://twitter.com/datosgobar>



<https://datosgobar.github.io/>

