

# Data Science Project Scoping Worksheet

This worksheet is designed for social good organizations (government agencies, non-profits, social enterprises, and others) to scope actionable data science projects.

## 1. Project Name:

Llamado oportuno

## 2. Organization Name:

Secretaría de seguridad ciudadana

## 3. Problem Description:

Se tienen 20 ambulancias para atender emergencias, por lo que se busca saber si una llamada es potencialmente legítima o no, para ser socorrida y dar la atención de manera correcta.

### 3.1 What is the problem you are facing?

Dado el número de unidades limitadas se necesita saber si una llamada es legítima para ser socorrida y no desperdiciar recursos esenciales de auxilio.

### 3.2 Who/what is affected by this problem? (people of certain type, organizations, neighborhoods, environment)

La ciudadanía en general ya que pueden sufrir accidentes viales.

### 3.3 How many people/organizations/places/etc and how much are they affected? (e.g. mean wait time for surgery, number of students dropping out of school, cost due to tax fraud, etc.)

Tiempo en ser atendido una emergencia vial.

Gente lastimada en accidente vial sin recibir servicio de ambulancia.

### 3.4 Why is solving this problem a priority for your organization

Es una prioridad que la ciudadanía que se encuentra en esta situación de emergencia pueda ser atendida de manera oportuna para salvaguardar su salud o vida.

## 4. Goals: What are the business/policy goals that will be accomplished by solving this problem and what constraints do you have? (in order of priority)

- The technical solution that will be built (e.g. predictive model or dashboard or map) is not the business/policy goal - that is the tool that will achieve your goal
- The goal should be specific and measurable
- Achieving the goal should help solve the problem you're tackling
- Typical goals include improving/maximizing/increasing or decreasing/mitigating/reducing some outcome or metric (such as school graduation rates or unemployment rates) .
- Typical constraints include budget, lack of human capital, legal restrictions, political will and social license.
- Consider tradeoffs between conflicting goals.

	Goal	Constraints
1	Minimizar la tasa de atenciones de llamadas de emergencia falsas	El número limitado de unidades de emergencia (20 ambulancias)
2		
3		

## 5. Actions

- Actions are activities or programs that institutions are doing/will do to address a problem. Actions can involve allocating resources, such as inspecting facilities, providing preventive services, outreach, etc.
- Actions should improve when the institution has the information that is generated in the project.
- Ideal actions should help you achieve the goals defined above.

	Action 1	Action 2	Action 3
<b>Action:</b> <i>eg. inspection for compliance with fishing quotas for boats in ports</i>	Identificar llamadas falsas de emergencia vial		
<b>Who is executing the action?</b> <i>eg. Inspector, Department of Inspections</i>	C5		
<b>Who/what is the action being taken on?</b> <i>eg. fishing boats</i>	Llamadas		
<b>How often is the decision to take this action made?</b> <i>eg. Daily</i>	Todos los días las 24 hrs		

<b>What channels are/can be used to take this action</b> <i>Eg. In person</i>	Análisis de voz en llamadas, revisión de cámaras, evasión policiaca de manera presencial		
<b>Other useful information about the action</b>			

## 6. Data

- The data has to connect to the actions its informing so the organization can achieve its goal
- Typical data science projects use administrative data as the primary data source, and enhance it with publicly available data sources (Census, other open data). Partnering with the private sector or non-profits could be a way to obtain data you might be missing internally.

### A. What data sources do you have internally?

	Data Source 1	Data Source 2	Data Source 3
<b>Name</b> e.g. Hospital Admissions database	Datos de accidentes viales en la CDMX reportados por C5 y Sistema de Datos Personales del Servicio de Atención de Llamadas de Emergencia 9-1-1 en la Ciudad de México		
<b>What does it contain?</b> Describe the attributes included in the data source. <i>eg. admission and discharge records for hospitals nationwide, including patient sociodemographic data, insurance type, medical doctor information, etc.</i>	Folio, fecha de creación del reporte, hora de creación del reporte, día de la semana de creación del reporte, fecha de cierre de reporte, hora de cierre de reporte, motivo del incidente dependiendo del tipo de emergencia, alcaldía donde sucedió el incidente, latitud y longitud del incidente, código de cierre del incidente reportado, clasificación del		

	incidente, origen del incidente por tipo, alcaldía en que se dio resolución al incidente o emergencia.		
<b>What level of granularity?</b> <i>eg. transaction, person, organization, location</i>	Llamada, folio		
<b>How frequently is it collected/updated after it's captured?</b> <i>eg. real time, daily, weekly, monthly, yearly, one off</i>	Con un día de vencimiento, pues se realiza una recarga diaria		

<b>Does it have reliable and unique identifiers that can be linked to other data sources?</b> <i>eg. SSN, National identifier, patient identifier, insurance number, etc</i>	ID Unico, no es seguro que se pueda unir con otras bases		
<b>Who's the internal owner of the data?</b> <i>eg. Hospitals</i>	el(a) Titular de la Dirección General de Operaciones		
<b>How is it stored?</b> <i>eg. in a database, pdfs, excel, spss</i>	Database		
<b>Additional comments</b>			

**B. What data can you get from external, private or public sources?**

	Data Source 1	Data Source 2	Data Source 3
<b>Name</b> <i>eg. Air Quality database</i>			
<b>What does it contain?</b> Describe the attributes included in the data source. <i>eg. distinct pollution's particle concentration</i>			
<b>What level of granularity?</b> <i>eg. geolocalized hourly sensor data</i>			
<b>How frequently is it collected/updated after it's captured?</b> <i>eg. daily</i>			
<b>Does it have unique identifiers that can be linked to other data sources?</b> <i>eg. sensor identifier</i>			

<b>Who's the internal owner of the data?</b> <i>eg. NOAA</i>			
<b>How is it stored?</b> <i>eg. API endpoint from an open data portal</i>			
<b>Additional comments</b>			

C. In an ideal world, is there additional data would you want to get/gather that would be relevant to his problem? (surveys, CCTV, phone records, DNA, different frequency or granularity for currently available data, etc)

Historial de llamadas de la persona solicitando apoyo, localización del que realiza la llamada.

## 7. Analysis

- Typical data science projects include a combination of analysis, typically including description, detection, prediction, optimization, and/or behavior change.
- Again, the analysis is not the goal of the project - the **analysis** helps you use the **data** you have to inform the **actions** you have access to in order to achieve your **goals**.
- Choose the right set of analysis for each problem
- You must validate the analysis and use a validation process that matches how your analysis will be used in practice

	<b>Analysis 1:</b>	<b>Analysis 2:</b>	<b>Analysis 3:</b>
<b>Analysis type</b> <i>e.g. Description, Prediction, Detection, Behavior Change</i>	Clasificación		
<b>Purpose of the analysis</b> <i>eg. understand historical behavior of</i>	Definir si una llamada es o no falsa para priorizar correctamente los recursos de ambulancias en accidentes viales de la CDMX.		

<i>individuals, estimate risk of disease of patient, identify which actions will diminish overfishing in the region</i>			
<b>Which action will this analysis inform?</b> <i>eg. inspections of compliance regarding fishing quotas</i>	Identificación de llamada falsa de accidente vial		
<b>How will you validate this analysis using existing data? What methodology and what metrics will you use?</b> <i>eg. using historical data, running an RCT</i>	Se usará eficacia.		

## 8. Ethical considerations

<b>Privacy</b> Are you working with personal and/or sensitive data that is individually identifiable? Mention them.	Sí se trabaja con información personal, en la app 911 incluso puede venir ficha de salud y se debe “firmar” La Leyenda de Protección de Datos Personales.
<b>Transparency</b> Which stakeholders should know about which parts of the project? <i>Stakeholders typically include policymakers, frontline workers, people who will be affected by the actions, etc</i>	Operaciones de C5 y Hospitales de la ciudad, deben conocer el contexto de la mejora en llamadas de accidentes viales.
<b>Discrimination/Equity</b> Are there any specific groups for whom you want to ensure equity of outcomes?	Equidad a cualquier persona que pida ayuda verdadera.

<i>eg. groups of interest defined by gender, age, localization, social class, educational level, urban/rural, ethnicity</i>	
<b>Social Licence</b> If the entire population of the country finds out about your project, will they be ok with it? Why?	Estarían de acuerdo con priorizar el recurso a quién más lo necesite, siempre que no sean ellos los afectados que no reciben la atención. En cuyo caso el descontento iría ligado al número de recursos destinados a los incidentes viales.
<b>Accountability</b> Who are the people responsible for all the things above?	C5, Equipo Ciencia de Datos, Secretaría de seguridad ciudadana.
<b>Any other considerations such as consent, legal, etc</b>	

9. **What field trial or randomized controlled trial can you design to validate the project in the field?** The outcomes you will measure should match your goals. Define the population in which the model will be tested. Define the duration of the trial. Specify the baseline. You should measure the impact in different population subgroups (see section 8)

Seleccionar aleatoriamente solicitudes de ayuda a las que se enviará la ambulancia con base en la clasificación del modelo contra las que no se mandará.

10. **Who are the external organizations and internal departments that will need to be involved?**

(Typically, data science projects need involvement from data owners, IT infrastructure owners, problem owner, analytics people)

<b>Organization/Department</b>	<b>Description of desired involvement</b>	<b>Name/role of counterpart</b>
<i>Sistemas</i>	<i>Infraestructura y acceso base de datos</i>	<i>Director del area de TI</i>
<i>Operaciones</i>	<i>Comunicación del impacto de la propuesta en el negocio</i>	<i>Equipo de operaciones</i>



This worksheet was originally developed by the Center for Data Science and Public Policy at the University of Chicago. For more information about our programs and work, please visit <http://datasciencepublicpolicy.org> or email us at [info@datascienceforsocialgood.org](mailto:info@datascienceforsocialgood.org)

This version of the worksheet has been extended through a collaboration between GobLab UAI and Carnegie Mellon University.

GobLab UAI is the innovation lab of the School of Government at Adolfo Ibáñez University. Its mission is to promote the use of data science in the public sector in order to improve public management and have more evidence-based public policies. For more information visit. It trains public servants, does applied research and projects in partnership with government agencies. <https://gobierno.uai.cl/centro-investigacion/goblab-uai/> or email [goblab@uai.cl](mailto:goblab@uai.cl)