Data Science Project Scoping Worksheet[1]

1. **Project Name:** Commander[2]

2. **Organization Name:** Santiago Fire Department

3. Problem Description:

3.1 What is the problem you are facing?

In Chile, there is a special number for emergency calls to the Fire Department: 132. All calls are received by a central dispatch, which then sends the dispatch message to the relevant station according to data on the address and type of emergency. The Fire Department responds to fire emergencies, vehicle and urban rescues, gas leaks and handling of hazardous materials.

There are currently no measurements for the average travel times of the fire trucks. Additionally, no information about street closures, traffic congestion or others is used when responding to an emergency call; a linear route is simply traced. Arriving to an incident on a timely manner is of great importance. For instance, it can prevent the risk of a fire spreading and can even save lives.

[1] Created by the Center for Data Science and Public Policy at the University of Chicago and GobLab at Universidad Adolfo Ibáñez. This project description was created by GobLab at Universidad Adolfo Ibáñez as part of a curriculum that is available here. You may use it quoting its creators and adapt it following the Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) License. The terms and conditions are available here.

[2] This worksheet was created by GobLab UAI for teaching purposes. It is a retrospective exercise conducted using public information that doesn't necessarily represent the project's initial formulation.

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

People who call the 132 number. Firefighters who lose valuable time traveling to the incident.

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

At a national level, there were 38,094 volunteers in 2009, and 107,826 service requests[3] were responded. In 2009, the Santiago Fire Station had 2,140 volunteer firefighters and 126 officers[4], and it attended 4,775[5] events.

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

The purpose of the Chilean Fire Department is to develop an efficient voluntary firefighting service that guarantees the safety of the lives and assets of citizens.[6] Solving the issue of arrival times to incidents is not only a benefit for the people who require the firefighters' services, but it also aligns with the objectives and history of the Chilean Fire Department. The Santiago Fire Department is among those that face the greatest challenges, as almost 50% of the Chilean population lives in the Metropolitan Region, and traffic is highly variable during the day, with peaks in the morning and afternoon, and a high level of different routes to get to the same place.

4. Goals (in order of priority)

[3] Memoria anual Bomberos Chile, 2009
[4] Memoria CBS, 2009
[5] Un nuevo sistema de despacho para bomberos, UChile
[6] Bomberos.cl/historia

- The technical solution (e.g. predictive model) is not the goal.
- The goal must be measurable.
- Achieving the goal should help solve the problem.
- Typical goals include improving/maximizing/increasing or decreasing/mitigating/reducing some outcome or metric.
- Typical constraints include budget, lack of human capital, legal restrictions, political will and social license.
- Consider tradeoffs between conflicting goals.

	Goal	**Constraints**
1	To decrease the response time between the dispatch referral and the arrival of the truck at the incident by 30% within 1 year	There is a limited number of stations located at fixed points in the city

5. Actions

- Actions are what institutions can do to address a problem by allocating resources, for instance, inspect facilities, provide preventive services, outreach, etc.
- Ideal actions should help you achieve the goal defined above.

	Action 1	Action 2	Action 3
Action: *e.g. inform the owner of a vehicle on how to renew their permit*	Dispatch of fire trucks	Selection of a route to get to the incident	
Who is executing the action? *e.g. IT department (sends email)*	Santiago Fire Department central dispatch	Each station (the driver of the fire truck)	
Who/what is the action being taken on? e.g. vehicle owner	On each incident call	On each incident call (on the fire truck)	

How often is the decision to take this action made? e.g. annually	Every time a call is received	Every time a fire truck is dispatched	
What channels are/can be used to take this action? e.g. email	The decision is made verbally	On-site and immediately	
Other useful information about the action	The dispatch decision is made based on the shortest distance according to predetermined quadrants and the characteristics of each fire that may require special equipment (e.g. chemical, altitude, etc.). This may be modified during the trip given the circumstances of the moment.	The central dispatch sends the dispatch order, but the truck can take the route it deems most pertinent.	

6. Data

- The data has to connect to the actions it informs 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?

(add columns for more sources if applicable)

	Station Characteristics	**Data Source 2**	**Data Source 3**
What does it contain? e.g. hospital admission and discharge records at a national level	Location and characteristics of the stations that are part of the Santiago Fire Department		
What level of granularity? e.g. transaction, person, organization, location	At a station level		
How frequently is it collected/updated once it is captured? e.g. in real time, daily, weekly, monthly, yearly, one-off	Once a year		

Does it have reliable and unique identifiers that can be linked to other data sources? e.g. national identifier	Yes, the number and name of the station		
Who is the internal owner of the data? e.g. hospitals	Santiago Fire Department (CBS, Cuerpo de Bomberos de Santiago)		
How is it stored? e.g. in a database, PDF, Excel	Database		
Additional comments	It contains the location of the station, types of fire trucks, and the staff and volunteers		

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

	Map of Santiago	**Santiago Traffic Information**	**Land Use Distribution**
What does it contain?	Streets of Santiago	Trip duration of Transantiago routes	Information on the location of hospitals, schools, parks, etc.
What level of granularity?	It has information on georeferenced coordinates	At a route level, by hour, minutes and seconds	At an establishment level
How frequently is it collected/updated once it is captured?	Several times a year	It is collected in real time, but updated daily	Once a month
Does it have reliable and unique identifiers that can be linked to other data sources?	Yes, georeferenced coordinates	Yes, georeferencing of the route	Yes, the data is georeferenced
Who is the internal owner of the data?	Google	Transantiago	MINVU (Ministry of Housing and Urbanism)
How is it stored?	Database	Database	Database

Additional comments	It delivers a database of the map of Santiago	It allows obtaining the traffic jam levels of some streets in Santiago (those with Transantiago routes)	

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

Online information about the condition of the streets.

7. Analysis

- Typical data science projects include a combination of analyses.
- The analysis is not the goal of the project.
- Choose the right analysis for the right problem.
- You must validate the analysis, and the validation process must match your goal.

	Analysis 1:	**Analysis 2:**	**Analysis 3:**
Type of analysis (e.g. description, prediction, detection, behavior change)	Prediction	Optimization	
Purpose of the analysis (e.g. understand historical behavior of individuals, estimate a patient's risk of disease)	To predict the estimated times of routes to the incident based on traffic and geographic variables, and with this, to obtain the fastest route from the station to the place of the incident	Based on the prediction of estimated route times and characteristics of the incident, it is possible to find out which are the optimal stations for dispatching one or more trucks	
Which action will this analysis inform?	Action 2: Selection of a route to get to the incident	Action 1: Dispatch of fire trucks	
How will you validate this analysis using existing data? (e.g. using historical data, running an RCT)	A randomized trial can be conducted, measuring the response time of stations that do not use the system versus others that do use it.	It is possible to compare with historical data of the area in which the system starts to be used. It is also possible to conduct a randomized trial with incidents that use the system versus the traditional system.	

1. Ethical Considerations

Privacy Are you working with personal and/or sensitive data that is individually identifiable?	No
Transparency Which stakeholders should know about which parts of the project? (Stakeholders typically include policymakers, frontline workers, people who will be affected by the actions, etc.)	Volunteers and fire truck drivers. It is important that they know how the system works so they can follow the recommendations and not seek alternative routes during the trucks' dispatch. Coordinator of the Central Dispatch
Discrimination/Equity Are there any specific groups for whom you want to ensure equity of outcomes?	No
Social License If the entire population of the country finds out about your project, will they be ok with it?	Yes, because it would allow reducing arrival times to accidents and fires and helping people, something that is valued by the population
Accountability Who are the people responsible for all the things above?	Commander of the Santiago Fire Department
Other considerations such as consent, legal, etc.	

8. **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.

A randomized trial can be implemented in which some stations will have this new tool. Then, the response times before/after the implementation can be compared, and also between stations.

9. 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, the problem owner, analytics people)

Organization/Department	**Description of desired involvement**	**Name/role of counterpart**
Universidad de Chile	Creation of the tool	Data analysts
Santiago Fire Department	Implementation of the tool in its pilot format	Superintendent
Various fire stations	Measurement of results (times)	Commander