ciencia de datos y gestión municipal

garcy valenzuela peña

21/12/2020

# Información general

## Antecedentes y descripción del proyecto

## Antecedentes

En el contexto del curso Ciencia de Datos y Gestión Municipal, dictado por la Universidad Alberto Hurtado (UAH), surge la necesidad de realizar un proyecto individual, que permita evidenciar lo aprendido en éste.

En este curso se utilizó la aplicación R Studio para la enseñanza del análisis de datos en el área o ámbito de los Datos Públicos para la gestión municipal.

La elaboración del proyecto seguirá la metodología estándar de un proyecto en ciencia de datos compuesta por las siguientes etapas:

## Comprensión y formulación del problema: se plantea el problema de ciencia de datos a abordar durante el proyecto.

## Adquisición de datos: se identifican las fuentes de datos y se procede a extraer, limpiar y transformar dichas fuentes para su posterior análisis.

## Análisis y modelamiento de datos: se usan técnicas estadísticas, de minería de datos y de Machine Learning (ML), para extraer valor a partir de los datos con el fin de resolver el problema inicial.

## Comunicación del resultado: se comunican los resultados usando técnicas de visualización de datos.

## Despliegue: se pone en producción el modelo construido y validado.

## Descripción del proyecto

El proyecto se denomina: MEDICIÓN DE GESTIÓN MUNICIPAL

¿Es posible medir la gestión municipal basada en un modelo manejado o dirigido por datos? ¿Cómo de mide la gestión municipal?

Por interés particular se pretende medir la gestión municipal de la comuna de Lanco. Ésta comuna se encuentra en Chile en la denominada Región de Los Ríos (o región 14).

Para responder a las preguntas se revisó literatura con el objeto de encontrar un modelo que permita cuantificar la gestión de un municipio.

Resultado de la búsqueda se halló que la Técnica de Análisis Envolvente de Datos (AED, o DEA por sus siglas en inglés) es ampliamente utilizada para la estimación de eficiencia en la gestión de gobiernos locales, producto de su simplicidad conceptual y versatilidad como herramienta de trabajo (Australia, España, Finlandia, Inglaterra, a, Finlandia, Inglaterra, EE.UU., entre otros)

La Técnica DEA es un modelo que tiene como propósito evaluar diferentes tipos de eficiencia a un grupo de empresas o unidades de toma de decisiones (municipalidades), que tienen un propósito común. Para nuestro caso se medirá la Eficiencia Técnica, es decir si para un nivel dado de servicios municipales (outputs), se utiliza la menor cantidad de insumos (inputs)

El nivel de eficiencia estimado, corresponde a la situación de un municipio en relación a los demás, no se mide en términos absolutos, sino que en términos relativos.

## Alcance del proyecto

Las actividades a realizar serán las siguientes:

1. Desarrollar una pregunta de investigación que deba ser abordada para un problema en el ámbito local para una comuna en específico.

La pregunta a contestar es **¿Qué comuna es técnicamente eficiente?**

1. Seleccionar conjuntos de datos desagregados a nivel comunal/municipal que se relacionen con su pregunta de investigación. Se describirá el origen, estructura y formato de las distintas fuentes a considerar.

Las comunas que formarán parte de éste análisis serán las de la región de Los Ríos y la región de Los lagos. A priori se consideran éstas dado que su localización geográfica y actividad económica las hace que tengan un perfil similar.

Producto de una investigación y análisis se identificarán cuáles son los datos que se requieren para un análisis DEA y se determinará su disponibilidad o no para ser parte del modelo.

De forma preliminar se ha identificado como fuente de datos los del área de Administración y Finanzas Municipales disponibles en el SINIM: <http://datos.sinim.gov.cl/datos_municipales.php>



En esta fuente existen 9 áreas con datos posibles a utilizar.

1. Desarrollar los procedimientos de limpieza, normalización y formateo de las fuentes en un solo conjunto de datos para el análisis. Se describirán el formato y estructura de los datos finales.

Una vez identificado los datos requeridos y su fuente, haciendo uso de la herramienta R Studio se cargarán éstos y se filtrarán para utilizar los que sean necesarios para modelar los inputs y outputs de una municipalidad.

No está de más explicitar que la geolocalización de las comunas son datos básicos a cargar, para poder luego generar una visualización en un mapa.

1. Generar un Plan de análisis.

Como modelo de análisis se ha definido que se utilizará la técnica DEA para determinar la eficiencia técnica de un municipio.

El documento siguiente describe la técnica o método DEA:

<https://bibing.us.es/proyectos/abreproy/4559/fichero/Cap%C3%ADtulo+II.pdf>

(copiar y pegar enlace en navegador)

1. Generar tablas, gráficos y mapas descriptivos en 3 niveles:
2. Interregional: métricas de resumen para variables seleccionadas según región, con foco en aquella donde se ubica la comuna seleccionada

Se consideran 2 regiones: Los Ríos y Los Lagos

1. Interprovincial: métricas de resumen para variables seleccionadas según provincia, al interior de la región donde se ubica la comuna seleccionada.

Se consideran 2 provincias



Se consideran 4 provincias para la Región de Los Lagos:



1. Intercomunal: métricas de resumen para variables selecciones según comuna, al interior de la provincia donde se ubica la comuna seleccionada

La región de Los Ríos contiene a 12 comunas

La región de Los Lagos contiene a 30 comunas

1. Analizar la situación de la comuna en cada uno de los tres niveles:
2. **Interregional:** situación de la región donde se ubica la comuna, respecto del resto de las regiones del país.
   * Se señalarán qué regiones tienen un perfil similar al de la región en la que se localiza la comuna seleccionada.
   * Se describirá si la visualización geográfica permite identificar situaciones que el resto de las visualizaciones no.
3. **Interprovincial:** situación de la provincia donde se ubica la comuna, respecto del resto de las regiones del país.
   * Se señalarán las provincias que tienen un perfil similar al de la región en la que se localiza la comuna seleccionada.
   * Se describirá si la visualización geográfica permite identificar situaciones que el resto de las visualizaciones no.
4. **Intercomunal:** situación de la comuna seleccionada respecto del resto de las que componen la provincia.
   * Se señalarán qué comunas tienen un perfil similar al de la región en la que se localiza la comuna seleccionada.
   * Se describirá si la visualización geográfica permite identificar situaciones que el resto de las visualizaciones no.

## Requisitos a alto nivel

Es necesario que el nuevo sistema incluya lo siguiente:

* Implementación en R Studio
* Utilizar datos de geolocalización
* Entrega, plazo máximo 11 de enero de 2021

## Entregables

* Datos originales utilizados
* Código desarrollado en R Studio
* Informe en formato Word.

## Partes afectadas

Este proyecto es externo y aislado, por lo cual no existen partes afectadas

## Procesos o sistemas empresariales afectados

Este proyecto es externo y aislado, por lo cual no existen interacciones con sistemas que puedan verse afectados

## Exclusiones específicas fuera del ámbito de acción

* No existirán entregas parciales, el desarrollo será continuo.
* Se considera realizar al menos una presentación de avance y muestra de lo implementado

## Plan de implementación

La implementación se realizará generando scripts en R Studio, haciendo uso de las librerías que sean necesarias para la correcta lectura y despliegue gráfico. En la sección siguiente se describe un cronograma a seguir.

## Escala de tiempo/Programación a alto nivel

Se propone la siguiente programación y candelarizacion:

**Semana 1.** 21-25 diciembre de 2020

* Desarrollar descripción a nivel de detalle.
* Identificar y documentar las fuentes de datos y los datos a utilizar.
* Implementar la lectura de datos, limpieza y transformación pertinentes.

**Semana 2**. 28-29 diciembre de 2020

* Implementar técnica DEA en R Studio
* Desarrollar visualizaciones en los 3 niveles descritos en sección II letra e.

**Semana 3**.- 04-08 enero de 2021

* Documento con Datos originales utilizados (desarrollado en semana 1)
* Documento con Código desarrollado en R Studio
* Generar Informe en formato Word.

# Aprobación y autoridad para proseguir

Aprobamos el proyecto según se describe más arriba, y autorizamos al equipo a que siga adelante.

|  |  |  |
| --- | --- | --- |
| Nombre | Cargo | Fecha |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
| Aprobado por |  |  | Fecha |  | Aprobado por |  |  | Fecha |