

# Avance Tesis MDS

Nombre Estudiante	<u>Rodrigo Molina Ávila</u>
Profesor/a Guía	<u>Sebastián Ferrada Aliaga</u>
Profesor/a Co-guía	<u>Roxana Chiappa Baros</u>
Nombre Tesis	<u>Representación mediante grafos de conocimiento a trayectorias académicas y producción científica de becarios/as doctorales BECAS CHILE y CONICYT (2014-2016) en disciplinas STEM y Ciencias Sociales</u>

## I. Principales Avances a la fecha

- Registro de la historia de cada proceso de extracción, análisis, o cualquier cambio en los datos en un repositorio GitHub.
- Extracción exitosa enriquecedoras de los datos que ayudan a responder la pregunta de investigación:
  - o Datos de educación secundaria adquiridos desde el ministerio de educación a través de webscraping.
  - o Datos de calidad de posgrado adquirido desde sus páginas a través de webscraping (QS y Shanghái Rank).
  - o Datos de publicaciones académicas adquiridos a través de api de programación de scopus.
  - o Datos de educación de pregrado adquirido a través del portal del investigador e identificación de perfiles en LinkedIn.
- Construcción de una ruta argumentativa desde la emergencia del problema, como ha evolucionado, como se ha estudiado y donde se inserta el problema de investigación; a través de marco teórico que valida las variables a utilizar.
- Preprocesamiento, depuración y modelaje estrella de las tablas dadas las extracciones de datos.
- EDA completo del total de los datos en Power BI
- Diseño, estructura, y montura del grafo en la plataforma neo4j.
- Expansión de la tesis con el desarrollo de una memoria para el Departamento de Ciencias de la Computación: [DCC | Introducción al Trabajo de Título](#)



**MDS**

Master of  
Data Science  
Universidad de Chile

# Avance Tesis MDS

- Construcción de consultas que visualizan y segmentan la muestra en función de los objetivos de la investigación.

II. De la redacción del manuscrito: ¿Qué capítulos están desarrollados? Indicar un número 1 si está completo el capítulo y un 0 si no.

1. Introducción: 0
2. Revisión de la literatura: 1
3. Metodología: 0
4. Resultados: 0
5. Discusión: 0
6. Conclusiones: 0

Estas son las secciones base, en caso de tener completas otras secciones indicar cuales:

En las respuestas anteriores, el 0 no significa que no tenga nada, actualmente, la investigación va tomando rumbos que no me permiten cerrar capítulos definitivamente para decir que si está “completo”

III. Principales dificultades a la fecha

- Detección y verificación de la integridad de los datos: Por la naturaleza de los datos, a veces, pequeños detalles en caracteres se escapan en la muestra. Se han realizado varias etapas de depuraciones para solucionar este problema.
- Vincular el uso de Knowledge Graphs para responder problemas desde la sociología, es decir, como los métodos que nos entrega la ciencia de datos, nos permiten/nos ayudan/nos facilitan (o como sea) responder preguntas complejas como la desigualdad dentro de campos tan específicos de la sociedad. Es una dificultad más bien filosófica que me he encontrado durante el desarrollo de la tesis.
- La decisión de como abordar la pregunta de investigación y entender qué preguntas hay que realizar a los datos.
- Estas dificultades actualmente están resueltas.

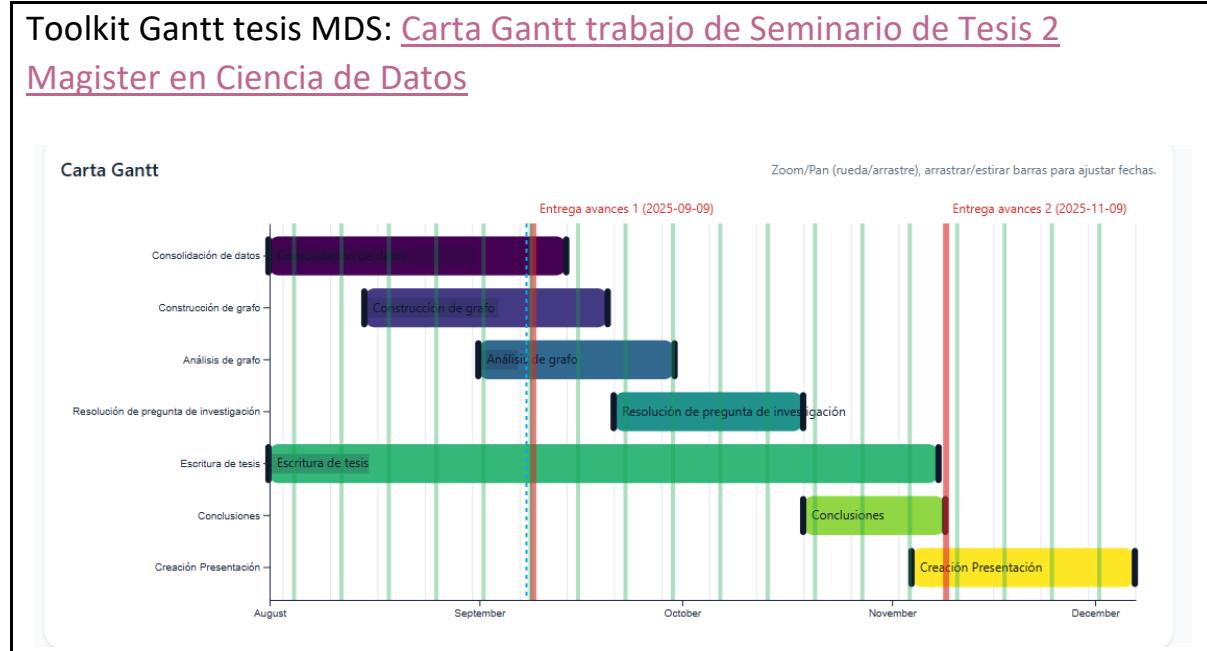


**MDS**

Master of  
Data Science  
Universidad de Chile

# Avance Tesis MDS

## IV. Próximos Pasos (Carta Gantt)



**Firma Profesor/a Guía**



**MDS** Master of  
Data Science  
Universidad de Chile