

# **Reporte Etapa 1: Business Understanding**

## Determinar Objetivos

### 1. Background

Se busca tomar decisiones dentro del Sistema de Transporte Urbano, por medio del uso de información confiable y actualizada. Los datos fueron obtenidos por medio de Movistar Chile con la cuál es posible estimar diferentes rutas de movilidad de la población, dichas rutas permitirán realizar un análisis que permita la toma de decisiones acertadas. Cabe mencionar que se está trabajando con datos personales por lo cual se debe tener cierto nivel de discrecionalidad.

Universidades a analizar - Top 5 Universidades en Santiago de Chile de acuerdo al ranking realizado por University Guru (<https://www.universityguru.com/es/universidades-santiago>)

Considerando escuelas públicas y privadas, se encuentran:

-Universidad de Chile

Ubicación: -33.4443503051102, -70.65092770035139

Cantidad de estudiantes: 40494

-Pontificia Universidad Católica de Chile

Ubicación: -33.441565078548514, -70.63994367134049

Cantidad de estudiantes: 33769

-Universidad de Santiago de Chile

Ubicación: -33.449466250548134, -70.67952809075585

Cantidad de estudiantes: 24396

-Universidad del Desarrollo

Ubicación: -33.39090888775374, -70.50097712022519

Cantidad de estudiantes: 18801

-Universidad Diego Portales

Ubicación: -33.449219560635335, -70.66159598465921

Cantidad de estudiantes: 18031

### 2. Objetivos de negocio

Objetivo Primario:

- Identificar el impacto que tienen las top 5 universidades dentro de la movilidad.
  - ¿Cuáles son las rutas más frecuentes?
  - ¿De qué comunas vienen más personas? y ¿Cómo influye la comuna de origen en la universidad destino?
  - ¿Cuáles son las horas en las que hay más desplazamiento de llegada y salida de las universidades?

### 3. Criterios de éxito

- Determinar las rutas más comunes hacia cada universidad
- Identificar la comuna que genera más viajes hacia cada universidad
- Definir los horarios en los que se realizan más viajes hacia cada universidad

El socio formador define si se cumple con el criterio de éxito del objetivo

## **Evaluar la Situación**

### 1. Inventario de recursos

- Expertos
  - Loreto Bravo y su equipo
- Asesores
  - Dr. José Antonio Cantoral - Área de deep learning y técnicas de lenguaje natural
  - Dr. Alberto Dorantes - Área de estadística
  - Dr. Benjamín Valdés Aguirre - Área de deep learning
  - MTI. Eduardo Daniel Juárez Pineda - Área de team management
  - Dr. Ismael Solis Moreno - Área de big data y cloud computing
- Dataset
  - Los datos se componen de 5 columnas las cuales son:
    - phoneid - El ID del usuario.
    - timestamp - Fecha de la conexión
    - bts\_id - ID de la antena telefónica
    - lat - Ubicación en latitud de la antena telefonica
    - lon - Ubicación en longitud de la antena telefónica
- Recursos de cómputo
  - Ordenadores personales
- Software
  - AWS
  - Tableau
  - Python

### 2. Requerimientos, supuestos y restricciones

- Lista de requerimientos:
  - Calendario de entrega

Fase	Fecha de Entrega
Business understanding	30 de Septiembre
Data understanding	12 de Octubre
Data preparation	
Modeling	4 de Noviembre
Evaluation	15 de Noviembre
Deployment	
Entrega preliminar	18 de noviembre
Entrega final	30 de noviembre

- Comprensibilidad y calidad de los datos

Sabemos que los datos provienen de Movistar Chile, por lo tanto son de una fuente confiable, ya que estos se obtuvieron del socio formador y estos ya pasaron por un proceso de limpieza y procesamiento para respetar la privacidad de los usuarios.

Los datos ya son bastante digeribles pero se podrían hacer mucho más procesamiento para hacerlos mucho más intuitivos de leer.

- Seguridad

Los datos cuentan con un procedimiento de seudonimización para garantizar la anonimidad de los sujetos. Por otro lado, si se tiene conocimiento de las localizaciones de las torres telefónicas y con estas se pueden hacer supuestos de la ubicación de algún usuario, lo cual puede afectar la privacidad y por lo tanto la vida de los mismos, de manera negativa.

- Aspectos legales

Contamos con el permiso de la empresa para el uso y manejo del dataset

- Lista de supuestos del proyecto sobre:

- Datos del proyecto

Los datos incluidos contienen información de ubicaciones de torres telefónicas así como las conexiones entre los teléfonos y las antenas, generando información acerca de las llamadas, duración de las mismas y el uso de datos móviles.

- El negocio

## Optimizar el Sistema de Transporte Urbano con información relevante

- Lista de restricciones del proyecto sobre:
  - Disponibilidad de los recursos  
Debido a que tenemos información limitada acerca de los usuarios no es posible saber quiénes son estudiantes, es por esto que el análisis abarca al público en general.
  - Aspectos técnicos
    - Considerando el número de antenas y la forma de detectar la ubicación de las personas, existe la posibilidad de tener un error alto durante el análisis.
    - Inicialmente no contamos con todos los conocimientos, pero estos se estarán obteniendo durante el transcurso del bloque.

### 3. Riesgos y contingencia

#### Matriz de riesgos

- Lista de riesgos que podrían ocasionar retrasos o fallos en el proyecto.
- Evaluación de los riesgos
- Plan de mitigación de los riesgos de mayor impacto
- Plan de contingencia para cada riesgo

### 4. Terminología

- Glosario de la terminología relevante para el proyecto. Debe incluir:
  - Terminología relevante del negocio
    - Comunas: f. Am. municipio (II entidad local).
    - Sistema de Transporte Urbano: Medio colectivo de desplazamiento de personas de un punto a otro.
    - Matriz de Origen Destino: Cantidad de viajes de un punto a otro.
  - Terminología técnica relevante ilustrada con ejemplos como minería de datos, IA, estadística, entre otros
    - Minería de datos es una de las técnicas principales que se utilizarán, la cual es el proceso que consiste en la detección de anomalías, patrones y correlaciones dentro de conjuntos de datos para la predicción de resultados.
    - Redes neuronales convolutivas: hacen uso de Kernel o filtros para la detección de ciertos criterios.
    - Kernel o filtro: método para el análisis de patrones.


5. Costos y beneficios

Dado que este proyecto se está desarrollando dentro de un contexto académico no es necesario hacer un registro de costos.

Determinar los objetivos de minería de datos

- 1. Objetivos de minería de datos
  - Identificar patrones de movilidad de las top 5 universidades.
- 2. Criterios de éxito de minería de datos
  - Identificar el 20% de rutas más frecuentes hacia cada universidad.
  - Identificar la comuna que registre el 35% o más de los viajes totales a cada universidad.
  - Determinar los rangos de horarios en los que el desplazamiento es 8% mayor que la media del día.

Producir el plan de proyecto

- 1. Plan de proyecto
  -  Plan de trabajo.xlsx
- 2. Evaluación inicial de herramientas y técnicas

Herramienta/Tecnología	Descripción (Considerar capacitaciones)
AWS	Proveedor de servicios en la nube
Tableau	Herramienta de análisis de datos
Spark	Herramienta de análisis y procesamiento de datos