

Business Understanding

Determinar Objetivos

1. Background

El uso de información confiable y actualizada es de gran valor para la toma de decisiones dentro de la planificación y desarrollo de un Sistema de Transporte Urbano el cual sea eficiente para la población del lugar en cuestión. Por otro lado, debido a que se busca que la información se actualice lo más frecuentemente posible, el realizar Encuestas de Movilidad las cuales nos den información para poder tomar decisiones es sumamente costoso e inconveniente, ya que estas toman mucho tiempo y recursos. Por ello, para poder tener información de valor de la movilidad dentro de la ciudad de Santiago Chile, se hace uso de la información almacenada por Movistar Chile, con la cual es posible estimar y analizar las diferentes rutas donde se trasladan gran parte de la población de la ciudad de Santiago. Con estas estimaciones podemos realizar diferentes análisis que nos ayuden a tomar decisiones dentro del Sistema de Transporte Urbano. También es importante recalcar que como se está trabajando con datos personales de la población se debe de tener cierto nivel de discreción.

2. Objetivos de negocio

Objetivo Primario:

- Generar conocimiento para identificar el impacto que tienen las universidades dentro de la movilidad.
- Generar una "Matriz de Origen Destino" la cual nos almacene la cantidad de viajes de un punto i a un punto j .

- ¿Los días festivos afectan a la movilidad?
- ¿Los fines de semana afectan a la movilidad?
- ¿Qué métodos de viaje se utilizan más frecuentemente?
- ¿La cantidad de estudiantes por universidad tiene un impacto?

Universidades a analizar

Top 5

-Pontificia Universidad Católica de Chile

Ubicación: -33.441565078548514, -70.63994367134049

-Universidad de Chile

Ubicación: -33.4443503051102, -70.65092770035139

-Universidad Diego Portales

Ubicación: -33.449219560635335, -70.66159598465921

-Universidad de Santiago de Chile

Ubicación: -33.449466250548134, -70.67952809075585

-Universidad del Desarrollo

Ubicación: -33.39090888775374, -70.50097712022519

3. Criterios de éxito

Criterio Subjetivo

- ¿Es de interés y/o utilidad ser consciente de este impacto?
- Hacer uso de una metodología seleccionada definida, justificada y correctamente implementada. (indicar persona que hace el juicio)

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 incluídos 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 abarcará 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

Categoría	Descripción	Esfuerzo
Análisis del Proyecto y entendimiento de los datos	En esta parte se consideran las juntas con la socio formadora, así como con otros interesados. También se contempla el análisis y diseño general de la solución a desarrollar. Obtención de requerimientos y estimación inicial	1 Semanas
Documentación	La gran parte de la documentación se realizará a la par del desarrollo; debido a que el modelo CRISP-DM nos pide que realicemos diferentes documentos donde	4 Semanas (a la par de lo demás)

	se describen muchos aspectos del proyecto como un resumen y entendimiento del negocio.	
Procesamiento de datos, Generación del modelo, Evaluación	Esta parte del proyecto es bastante iterativa, ya que se crean diferentes modelos y dependiendo de su desempeño dentro de la evaluación este puede ser modificado.	5 Semanas
Entrega	El producto final de todo el desarrollo	1 Semana
	Tiempo Total Estimado	10 Semanas

Determinar los objetivos de minería de datos

1. Objetivos de minería de datos


- Construcción de la Matriz de Origen Destino con los datos disponibles → proporcionar información y análisis del impacto que tienen las universidades en la movilidad de la ciudad de Santiago.

2. Criterios de éxito de minería de datos

- El modelo de predicción tiene un error menor al 15%.

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)	Costo
AWS	Proveedor de servicios en la nube	Medio-bajo
Tableau	Herramienta de análisis de datos	Medio-bajo
Spark	Herramienta de análisis y procesamiento de datos	Medio