

Programa de Ciudad inclusivas, sostenibles e inteligentes (CISI)

Taller de capacitación virtual

Aspectos de medición de la movilidad urbana para contribuir al Gran Impulso para la Sostenibilidad

 4, 5 y 6 de mayo, 2022



¿Cómo darle forma a la ciudad a partir de los datos de Telefonía Móvil?



Eduardo Graells-Garrido | Instituto de Data Science, Universidad del Desarrollo

dataScience UDD



Comisión Económica para América
Latina y el Caribe
(CEPAL)



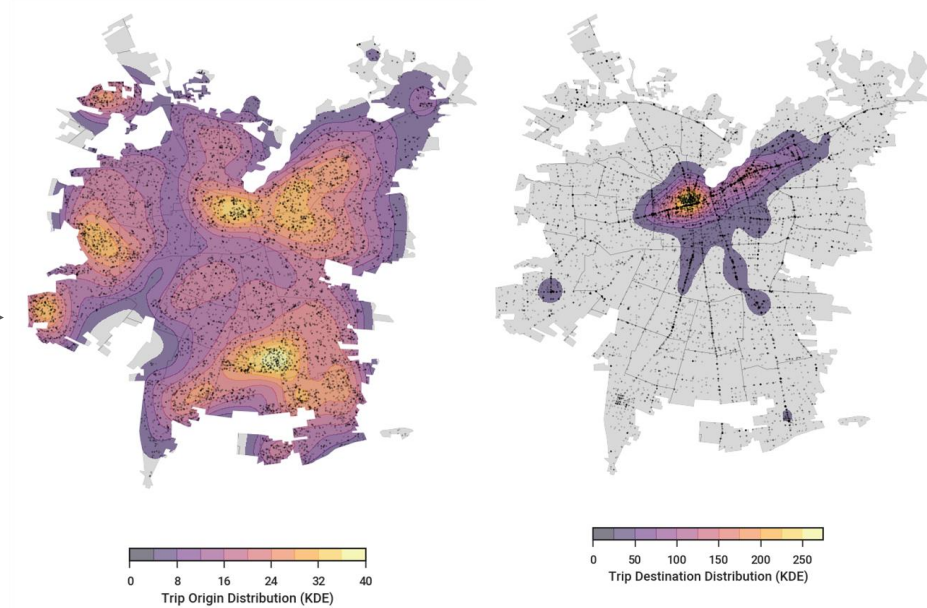
*Programa de Cooperación CEPAL-BMZ/giz
2020-2022: Ciudades Inclusivas, sostenibles e
inteligentes en el marco de la Agenda 2030*.

Motivación/Situación



Encuesta Origen Destino de Santiago 2012

20K Hogares, 100K Encuestados/as
(7M Habitantes)



Distribución de la Población (IZQ) y Polos de Atracción de Trabajo (DER).

Complicación

La ciudad cambia a un ritmo mayor al de captura y análisis de datos tradicional.

Más autos, nuevos modos de transporte, más población, ¡una pandemia!



Mercado automotriz chileno proyecta récord histórico en ventas este 2018

A juicio de la Cámara Nacional de Comercio Automotriz de Chile (Cavem) este año se venderán alrededor de 400 mil unidades debido a un alza en la inversión que se espera para este periodo.

10 de Enero de 2018 | 13:35 | Patricio Gutiérrez, Emol



ENCUENTRA TU PROXIMO VEHICULO

Palabras clave

Región
Todo Chile

Categoría
Vehículos

Marca
Selección Marca

< 2000 Rango de años 2018



Oportunidad

Hoy hay más dispositivos generando trazas digitales que personas.



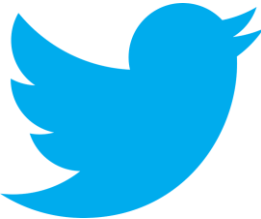
Trazas Digitales



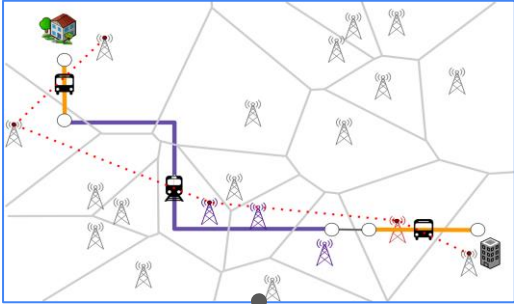
transbank.
APOYANDO NEGOCIOS



Cornershop



Propuesta desde la Ciencia de Datos



Telefonía Redes Sociales Apps de Movilidad

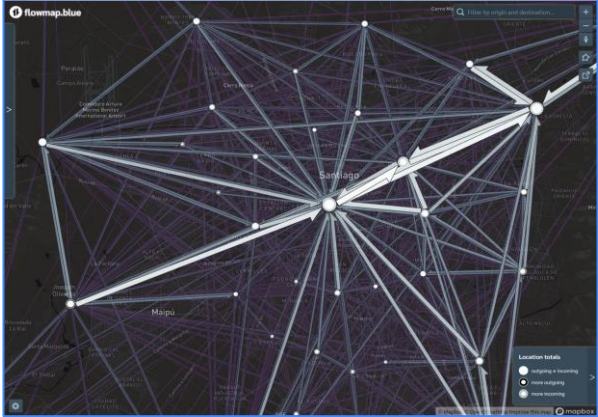
Datos

Detección de Viajes
Inferencia de Modo de Transporte

Patrones de Movilidad Actuales

Objetivos

Impacto



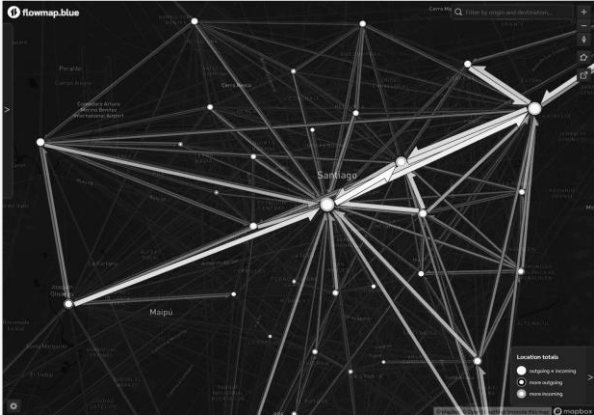
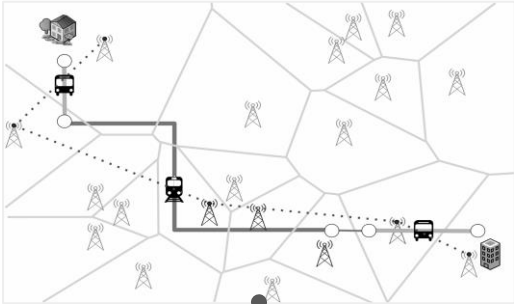
Created by Amethyst Studio from Noun Project

Data Scientist

Machine Learning Visualización

Métodos

Situación Actual



Created by Amethyst Studio from Noun Project

Data Scientist

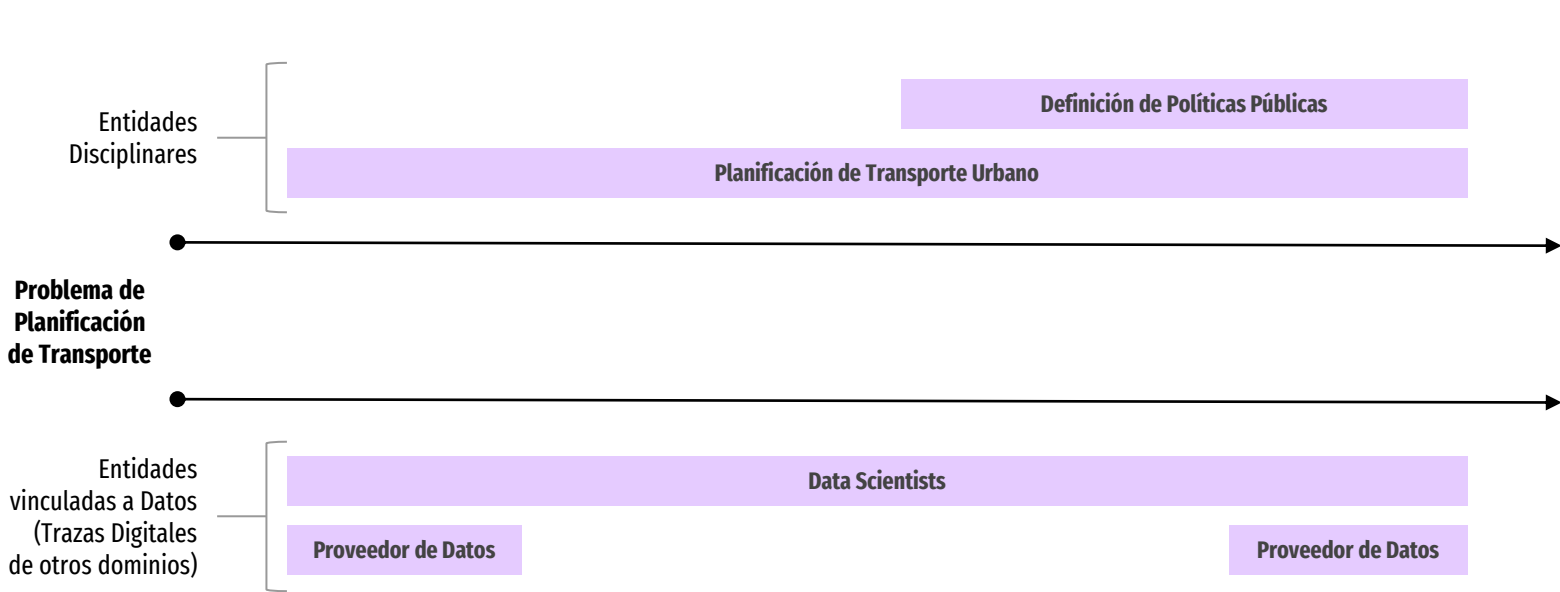


Created by Amethyst Studio from Noun Project

Planificador(a)

- Modelos de Transporte
- Fuentes Tradicionales
- Simuladores

Líneas Paralelas no se Intersectan



Created by AnantYat Studio
from Noun Project

**Solución
Basada en
Datos**

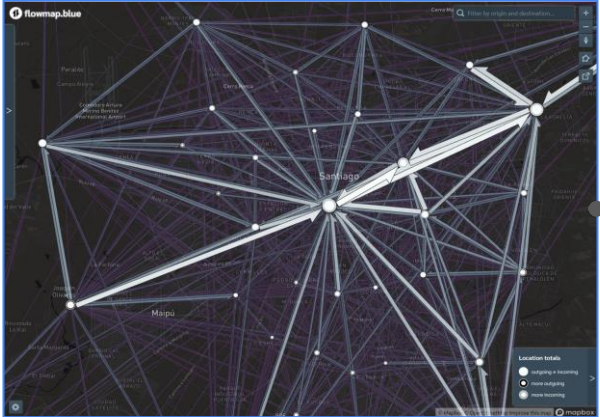
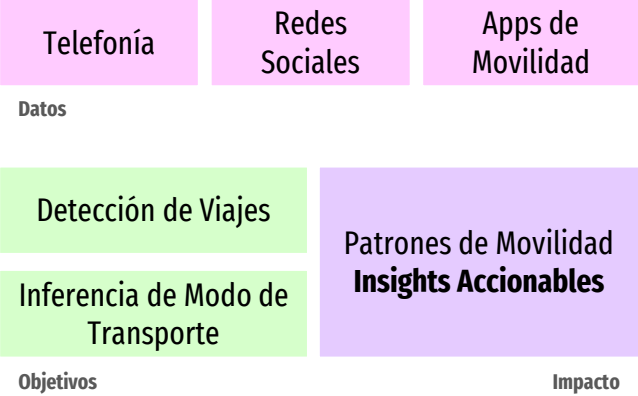
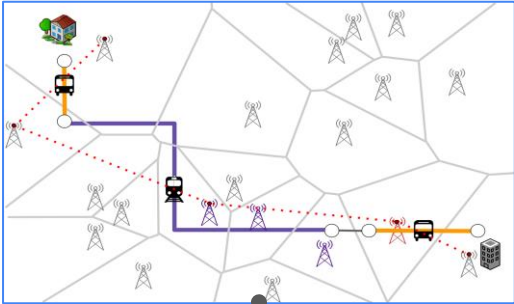
**Solución
Basada en
Datos**



Created by AnantYat Studio
from Noun Project

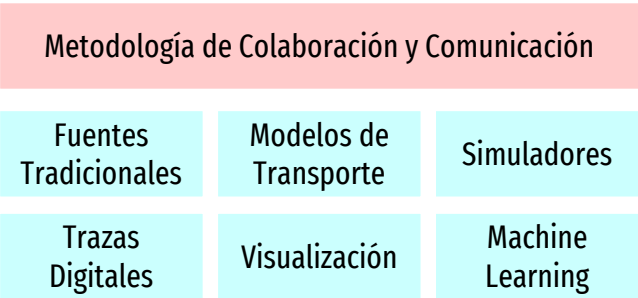
¿Cómo lograr la colaboración?

Situación Ideal



Created by Amethyst Studio from Noun Project

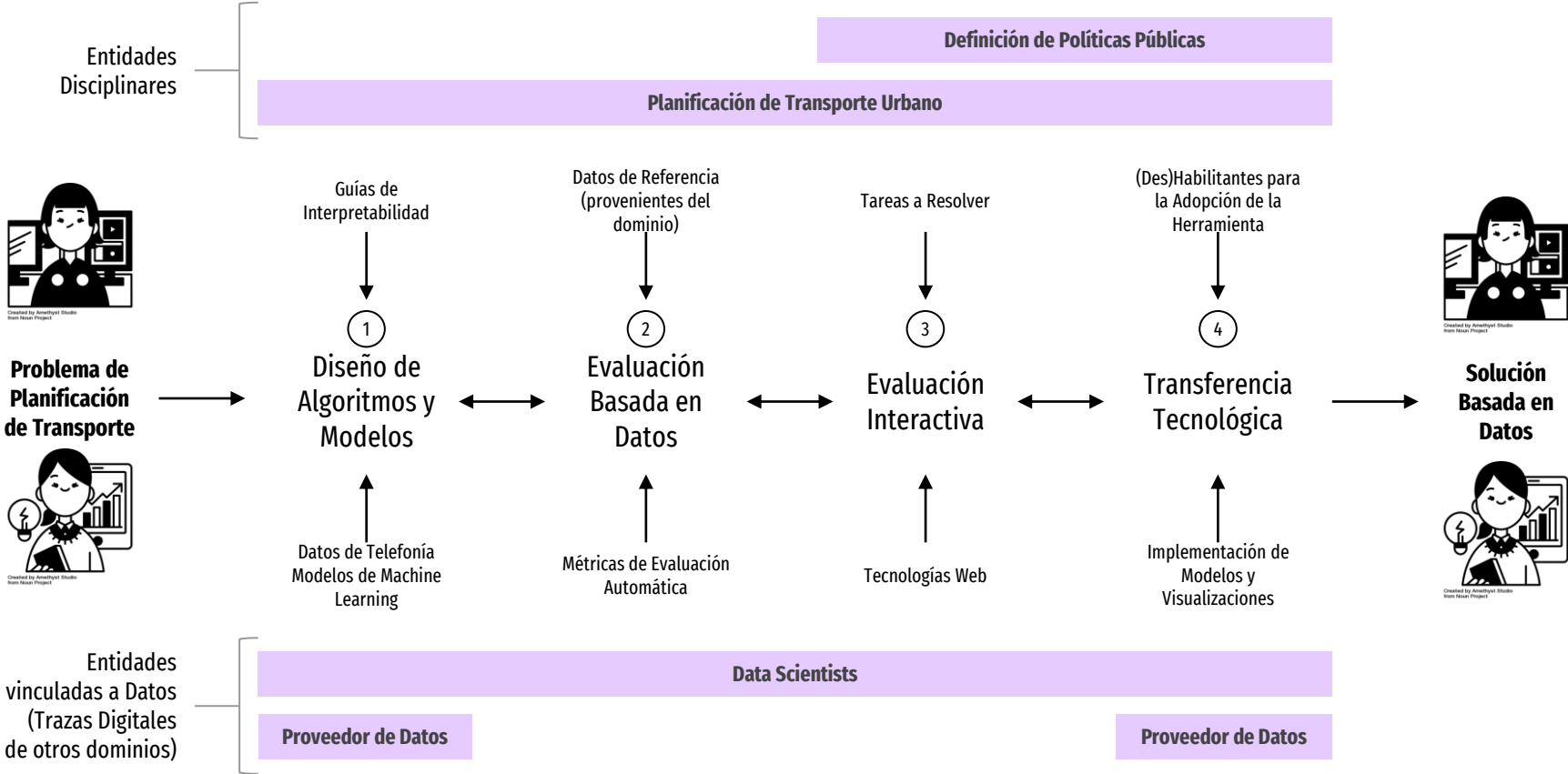
Data Scientist



Created by Amethyst Studio from Noun Project

Planificador(a)

Metodología



Caso de Estudio: Partición Modal

Problema de Planificación

¿Cuánto se usa cada modo de transporte?

¿Cuál es su distribución geográfica?

Determinar la **Partición Modal** (o *Modal Split*) actualizada es un problema que no está resuelto.

Sin una medición de los movimientos de las personas, no podemos planificarla para el futuro ni mejorarla en el presente.

Fotografías por Ricardo Hurtubia @rhurtubia



Nacional

“Súper Lunes”: Marzo llegó con menos movilidad vial de lo esperado

Por: El Desconcierto | Publicado: 01.03.2021



Súper Lunes en el Metro | Agencia Uno



Este 1 de marzo se esperaba que fuese un intenso “Súper Lunes” con mucha congestión vehicular, protestas y masivo regreso de veraneantes a sus funciones habituales, pero según el balance de la ministra de Transportes, Gloria Hutt, hubo menos movilidad de lo esperado, que en comparativa a marzo 2020 previo a la pandemia, es mucho más baja.

Etapa 1

Algoritmos y Modelos

Registros de Telefonía Móvil

Son **trayectorias** de conexiones a torres de transmisión.

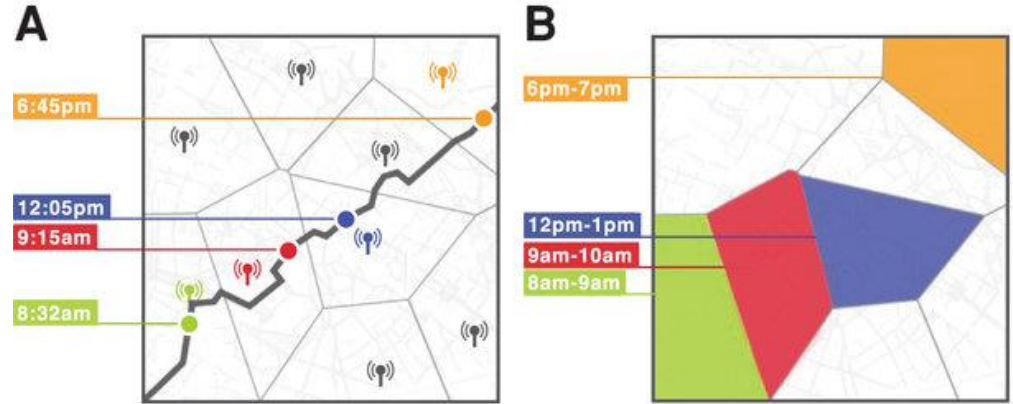
Un registro contiene:

- **Identificador anonimizado de dispositivo** (ej: *ABCDEFGH12345*).
- **Fecha y hora del registro.**
- **Torre** a la que estaba conectado.

La torre no es una ubicación exacta, sino que es un **área de cobertura**.

En promedio pasan **15 minutos entre registros** de un dispositivo.

Trabajamos con registros de más de **un millón de dispositivos en Santiago.**



Telefonica

¿Cómo se ve la ciudad a través de los datos?

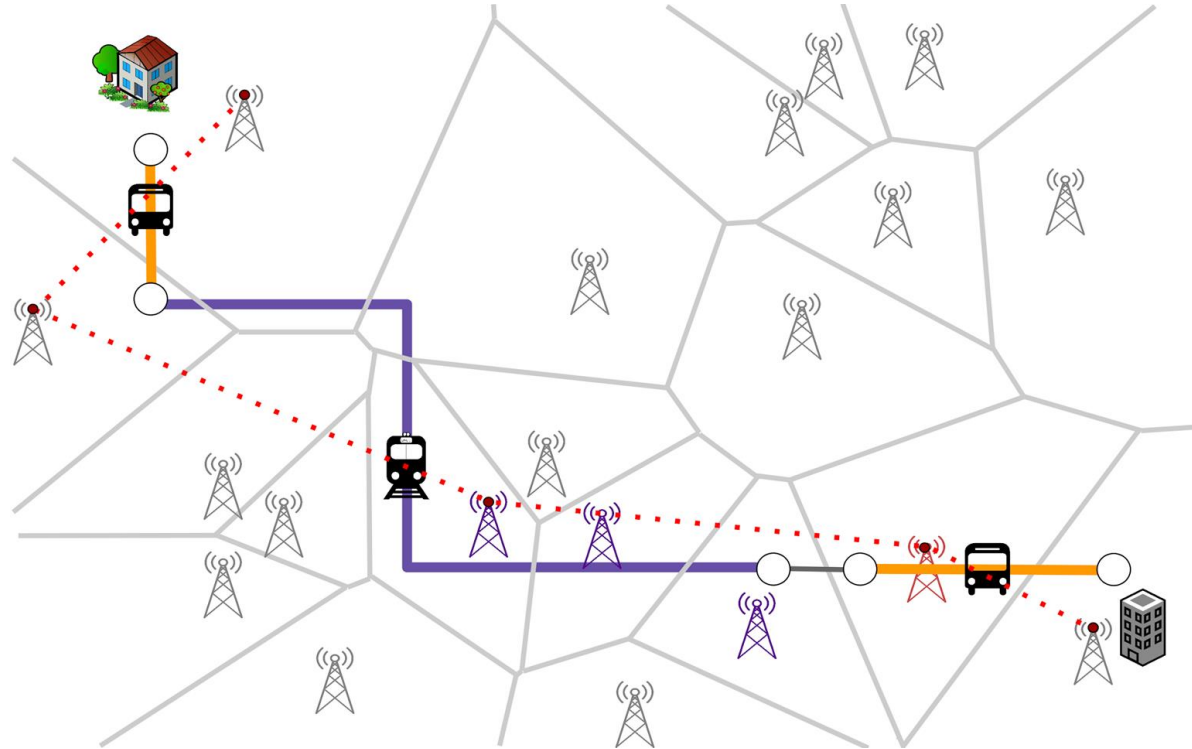
- 1 Diseño de Algoritmos y Modelos
- 2 Evaluación Basada en Datos
- 3 Evaluación Interactiva
- 4 Transferencia Tecnológica

Un viaje de tres etapas (bus-metro-bus) en Santiago está **plenamente identificado en una fuente de datos tradicional**.

En cambio, las conexiones a las torres presentan una **aproximación del viaje**.

Debemos **diseñar un algoritmo** para inferir viajes.

Y debemos **ajustar un modelo** para determinar los modos de transporte utilizados.

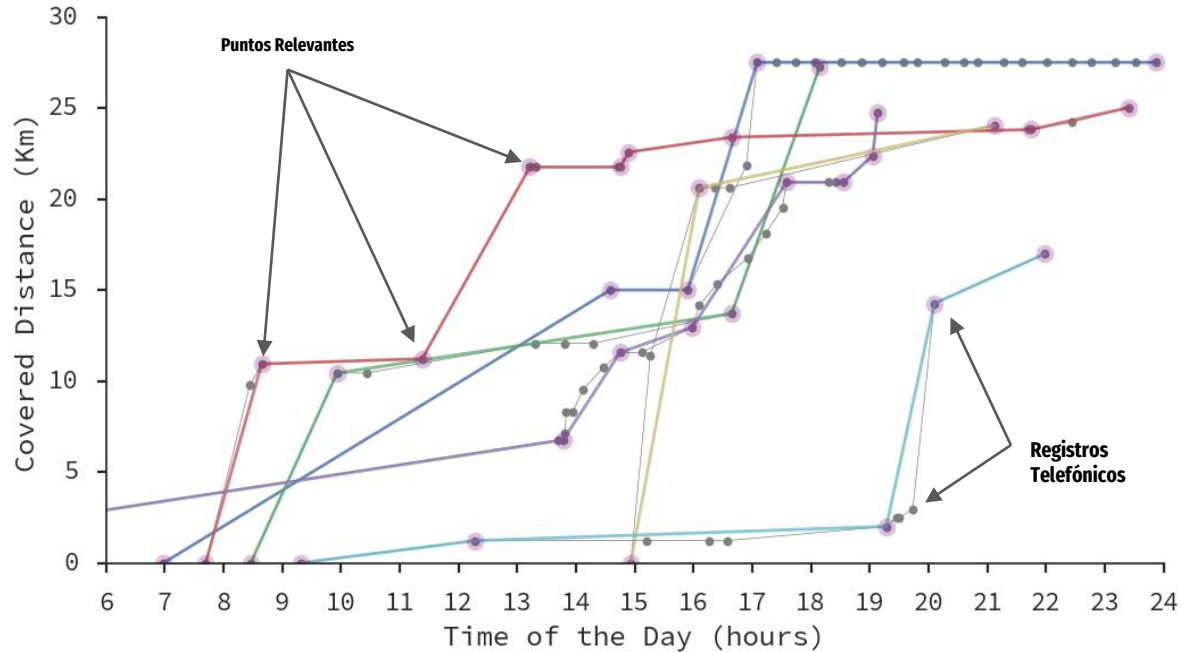
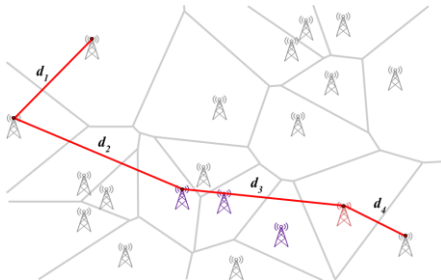


La Geometría de los Viajes

Creamos un **plano** con el **tiempo** y la **distancia acumulada** recorrida por un dispositivo.

Cada dispositivo es una polilínea, definida por sus registros.

Aplicamos un **criterio geométrico** (por ej., ángulos) para identificar los **puntos relevantes de cada trayectoria**.

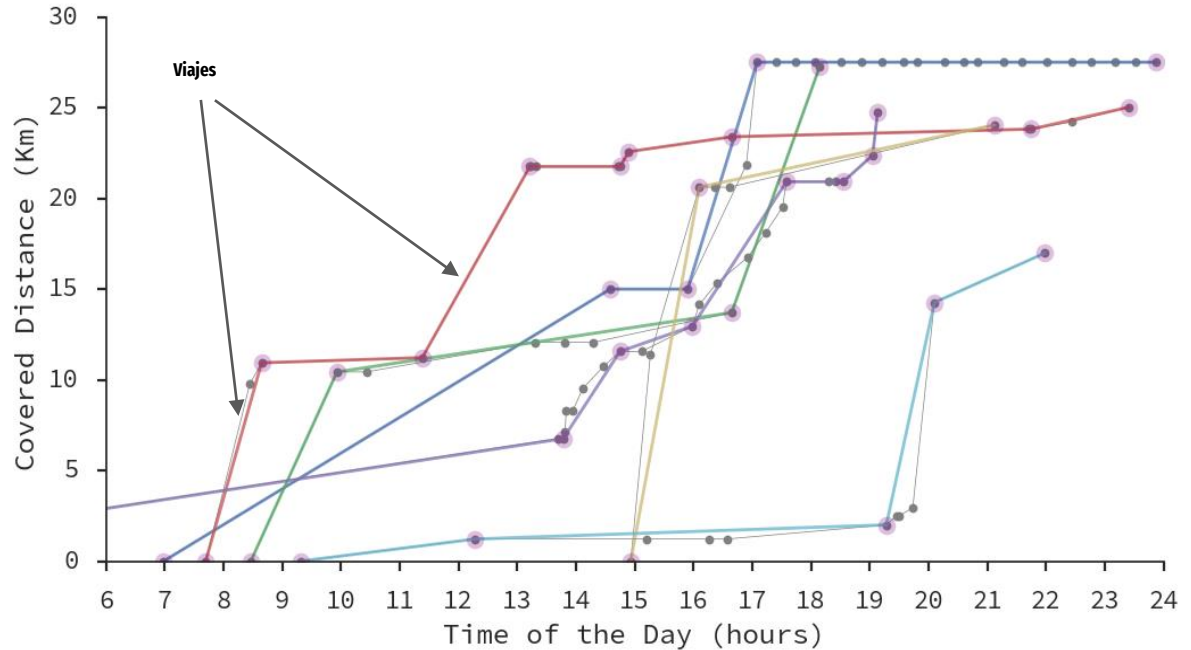


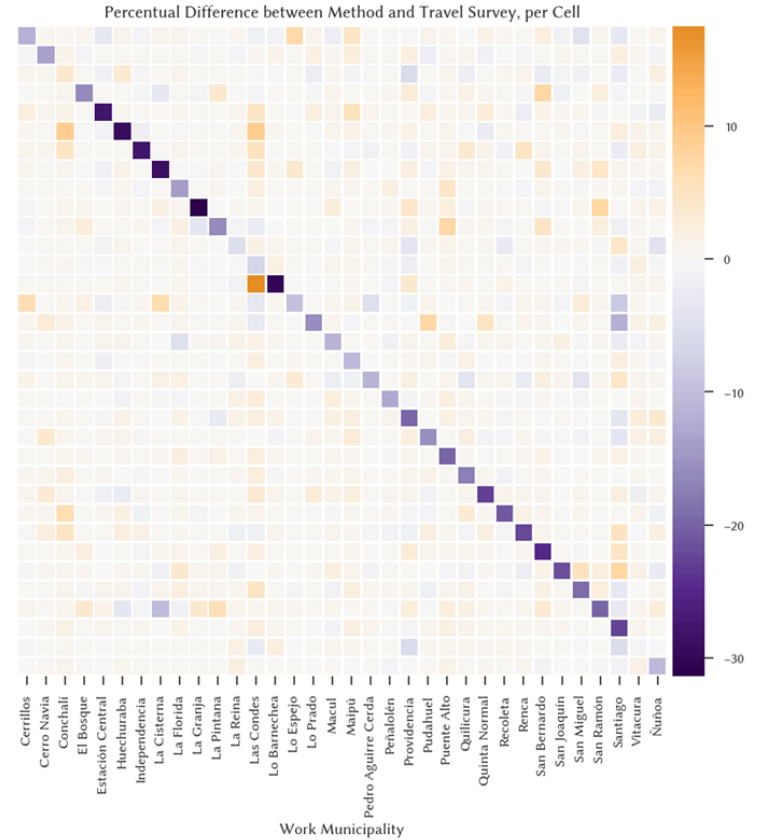
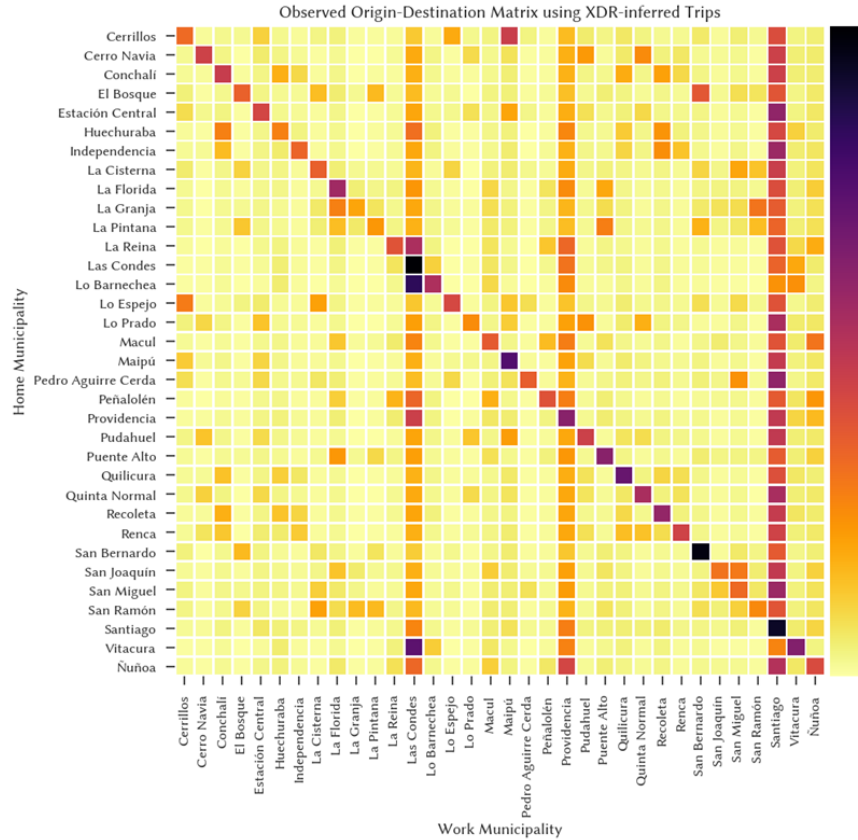
Identificación de Viajes

Los puntos relevantes nos permiten **identificar segmentos en la trayectoria.**

Nuevamente un **criterio geométrico** nos puede decir **cuáles de estos segmentos representan desplazamientos.**

Segmentos que sean **desplazamientos contiguos se pueden agrupar en viajes.**





Agrupación de Viajes por Modo de Transporte

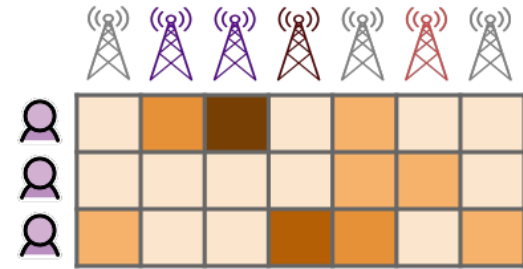
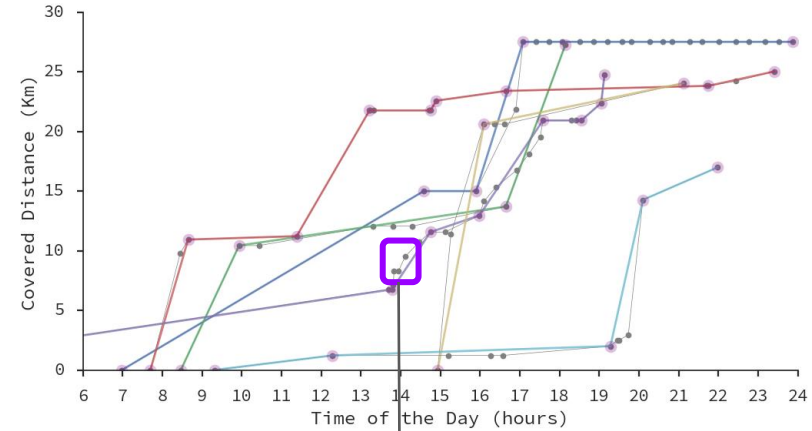
Una manera de identificar el modo de transporte es **agrupar los dispositivos que comparten modo**.

Unificamos los viajes al trabajo (son recurrentes) y construimos una matriz que llamamos **Waypoint Matrix**.

Esta matriz analiza los puntos intermedios de las rutas.

Objetivo: agrupar filas de la matrix de acuerdo a su contenido, es decir, dispositivos de acuerdo a sus rutas.

- 1 Diseño de Algoritmos y Modelos
- 2 Evaluación Basada en Datos
- 3 Evaluación Interactiva
- 4 Transferencia Tecnológica



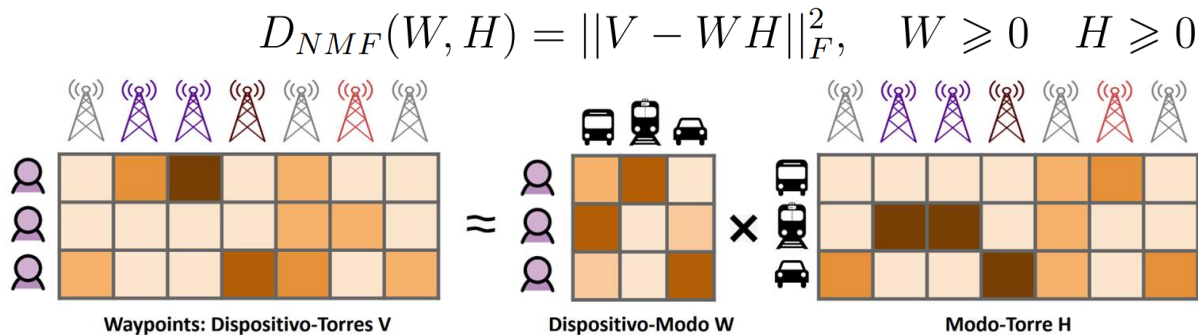
Waypoint Matrix W

Topic-Supervised Non-Negative Matrix Factorization

NMF es una **técnica de descomposición de matrices** que agrupa filas y columnas de una matriz en **dimensiones latentes**.

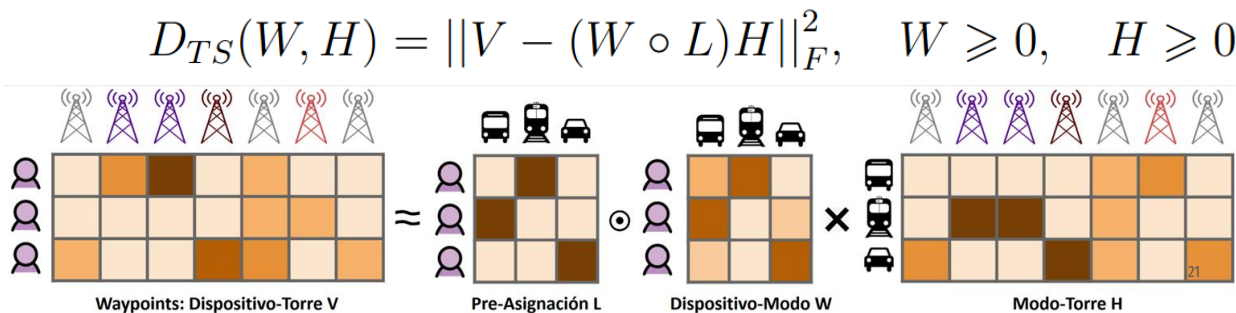
Es un problema **NP-Hard**. Se resuelve con Expectation-Maximization y Maximum Likelihood, entre otras técnicas.

NMF



TS-NMF es una versión **semi-supervisada** que permite adjuntar información al modelo a través de una matriz adicional. Así, **las dimensiones se vuelven interpretables**.

TS-NMF

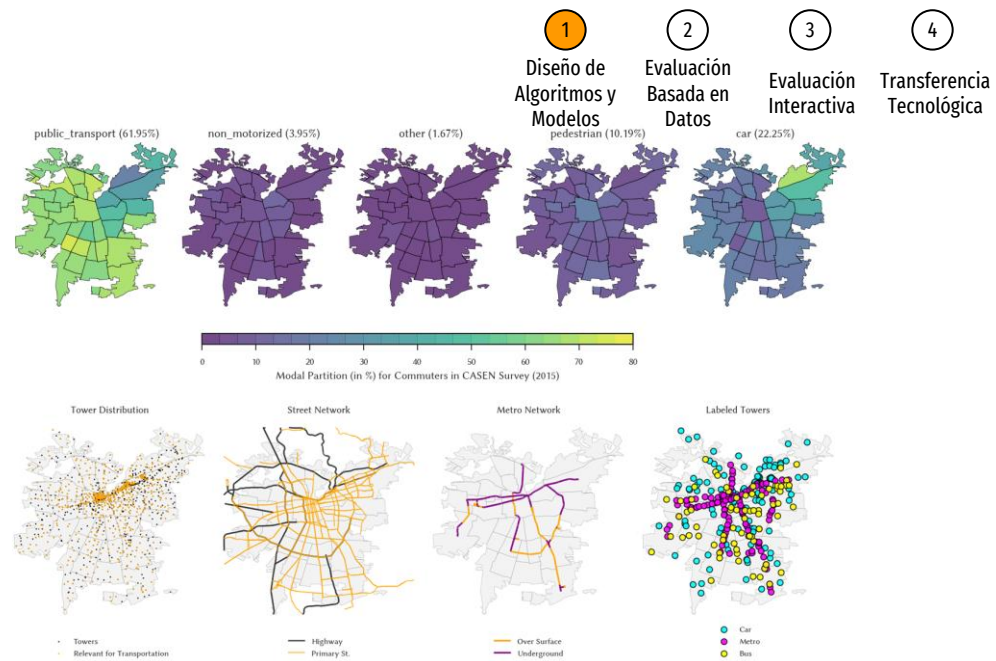


¿Cómo construir la matriz L?

En L podemos indicar el modo de transporte más probable para algunos dispositivos:

- Si transita por torres cercanas/internas a infraestructura de modos de transporte (autopistas, metro).
- Si vive en una comuna donde exista un modo de transporte principal (dada la segregación de Santiago, ésto es frecuente).

Después de construir la matriz L se resuelve el problema de optimización.



Etapa 2

Evaluación con Datos

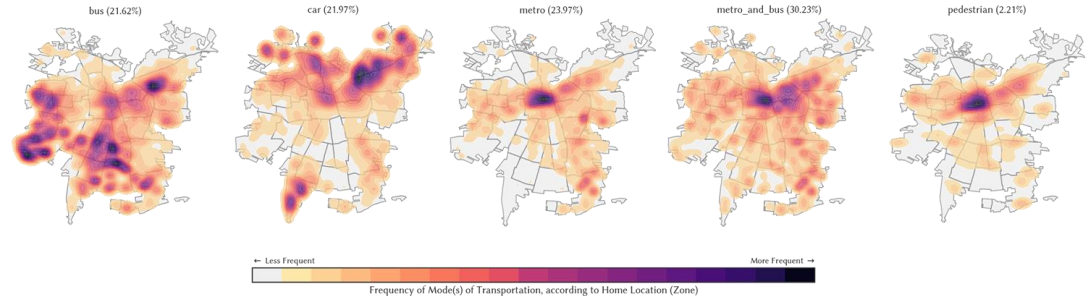
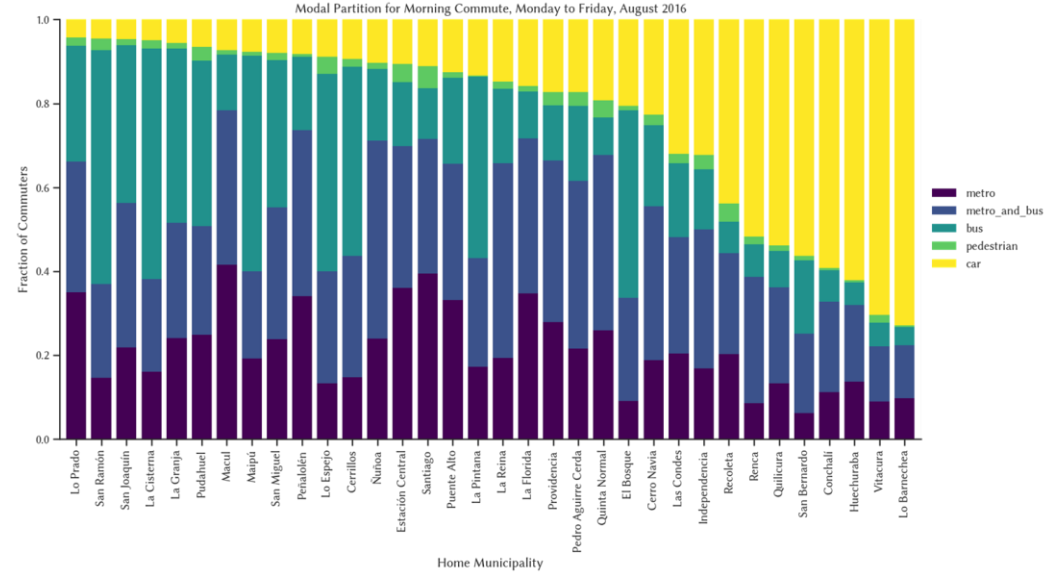
La Particion Modal

Con el hogar y trabajo asociado a los dispositivos podemos calcular la partición modal.

Hacemos un post-proceso: si un usuario está asociado parcialmente a dos modos de transporte, entonces usa transporte **multimodal**.

Agregamos el modo “peatón” para quienes el trabajo está a menos 1 Km de casa.

- 1 Diseño de Algoritmos y Modelos
- 2 Evaluación Basada en Datos
- 3 Evaluación Interactiva
- 4 Transferencia Tecnológica

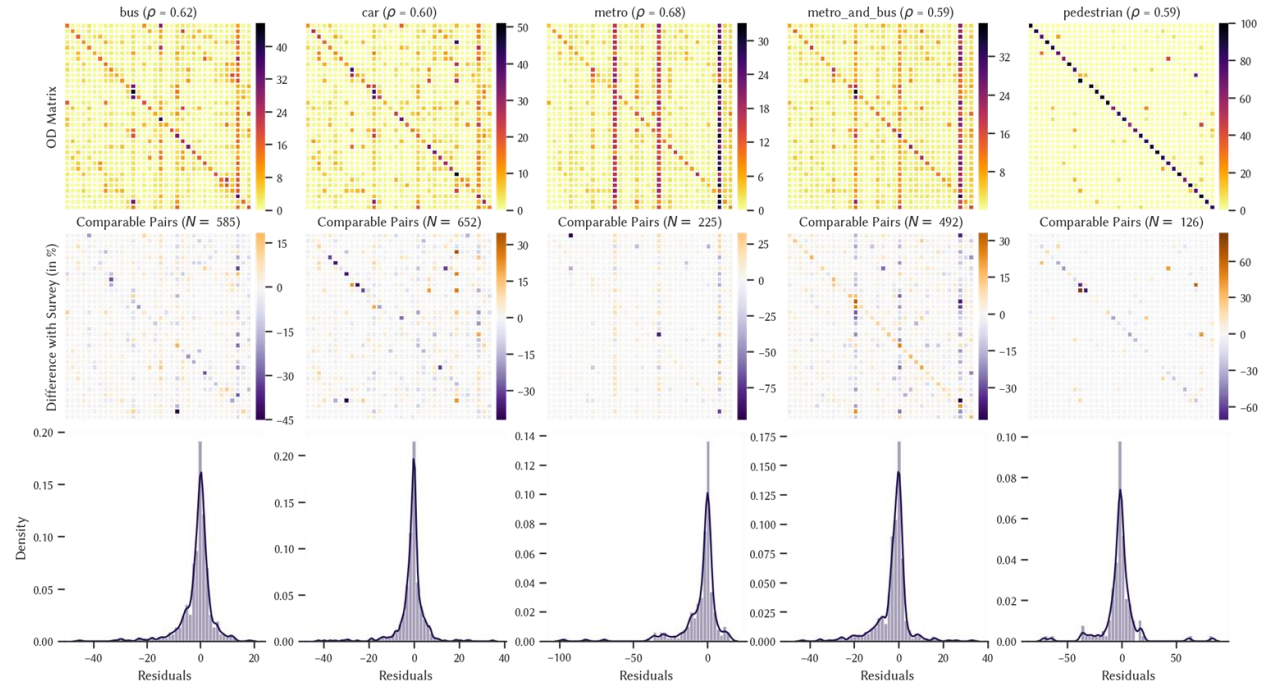


Evaluación de los Datos

No hay datos oficiales actualizados para las fechas en las que se calcula la partición modal actualizada (en este caso, 2017).

Para cada modo (e intermodalidad) calculamos una **matriz Origen-Destino** y su correlación de Spearman con la encuesta Origen-Destino de Santiago. Exploramos las **diferencias**.

Validamos correlaciones y diferencias **en conjunto con personas expertas del área**.



Etapa 3

Evaluación Interactiva

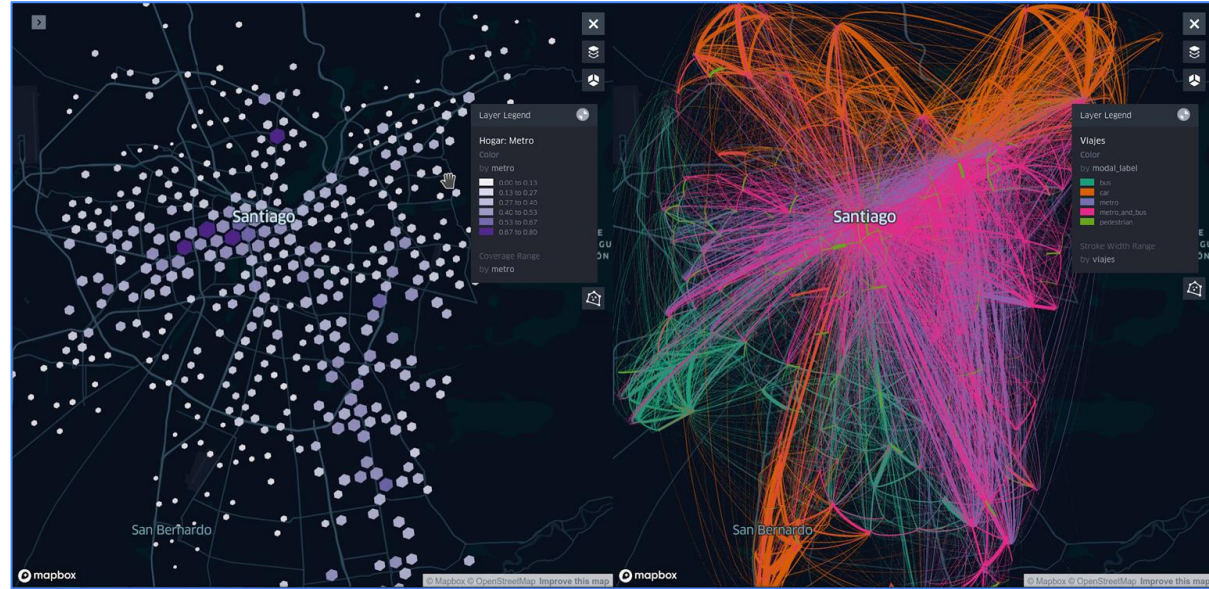
De Modelos a Herramientas

- 1 Diseño de Algoritmos y Modelos
- 2 Evaluación Basada en Datos
- 3 Evaluación Interactiva
- 4 Transferencia Tecnológica

Si no hay datos oficiales, **la única forma de asegurar que nuestra propuesta aporta valor es poniéndola a prueba.**

Para ello creamos **prototipos** que hacen uso de **Visualización de Información** para desplegar los resultados del modelo de manera interactiva.

De este modo, **personas que harían uso del sistema pueden ejecutar tareas concretas.**



Prototipo implementado en herramienta **kepler.gl**

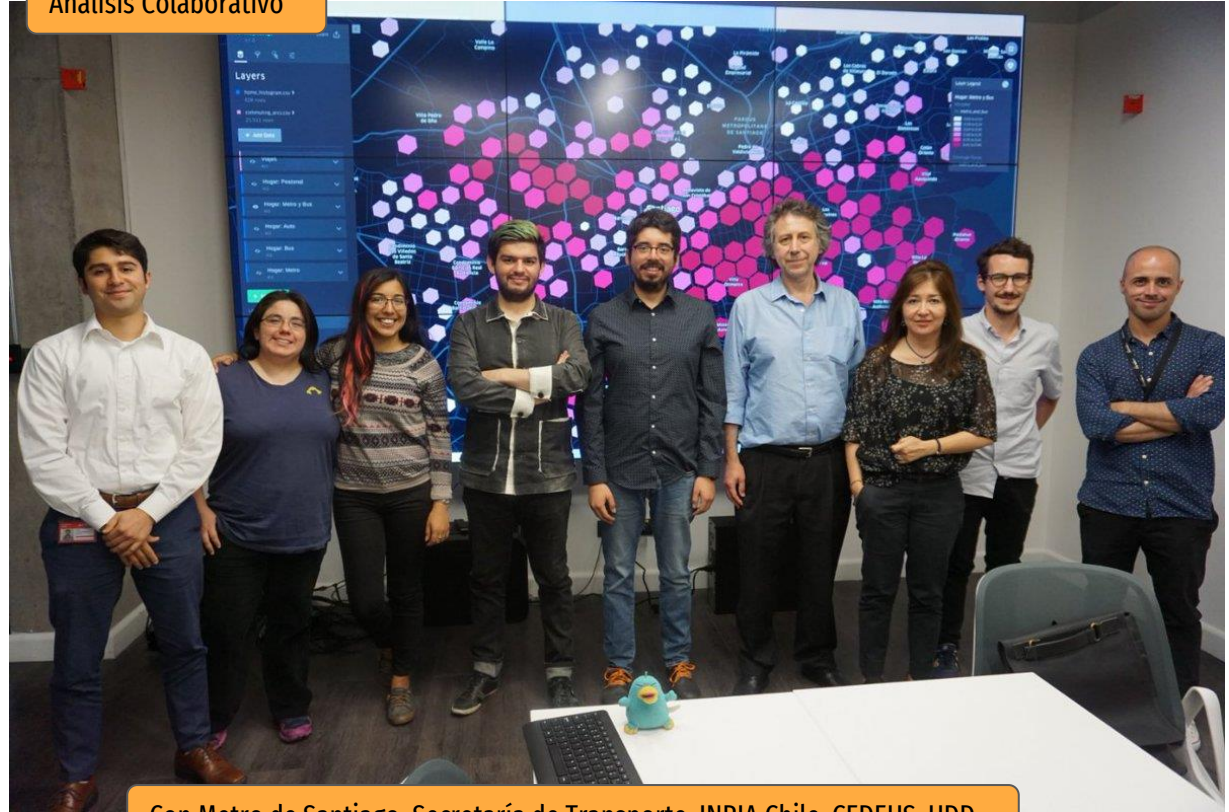
Pruebas de Concepto



Test de Interactividad



Análisis Colaborativo



Con Metro de Santiago, Secretaría de Transporte, INRIA Chile, CEDEUS, UDD



Graells-Garrido et al. *EPJ Data Science* (2018) 7:49
<https://doi.org/10.1140/epjds/s13688-018-0177-1>



REGULAR ARTICLE

Open Access



Inferring modes of transportation using mobile phone data

Eduardo Graells-Garrido^{1,2*}, Diego Caro^{1,2} and Denis Parra³

*Correspondence: egraells@udd.cl
¹Data Science Institute, Faculty of Engineering, Universidad del Desarrollo, Santiago, Chile
²Telefonica R&D, Santiago, Chile
 Full list of author information is available at the end of the article

Abstract

Cities are growing at a fast rate, and transportation networks need to adapt accordingly. To design, plan, and manage transportation networks, domain experts need data that reflect how people move from one place to another, at what times, for what purpose, and in what mode(s) of transportation. However, traditional data collection methods are not cost-effective or timely. For instance, travel surveys are very expensive, collected every ten years, a period of time that does not cope with quick city changes, and using a relatively small sample of people. In this paper, we propose an algorithmic pipeline to infer the distribution of mode of transportation usage in a city, using mobile phone network data. Our pipeline is based on a Topic-Supervised Non-Negative Matrix Factorization model, using a Weak-Labeling strategy on user trajectories with data obtained from open datasets, such as GTFS and OpenStreetMap. As a case study, we show results for the city of Santiago, Chile, which has a sophisticated Intermodal public transportation system. Importantly, our pipeline delivers coherent results that are explainable, with interpretable parameters at each step. Finally, we discuss the potential applications and implications of such a system in transportation and urban planning.

Keywords: Mobile phone networks; Urban informatics; Commuting; Non-negative matrix factorization; Mode of transportation

EPJ Data Science
 a SpringerOpen Journal

Open Access Article

Adoption-Driven Data Science for Transportation Planning: Methodology, Case Study, and Lessons Learned[†]

by Eduardo Graells-Garrido^{1,2,*} Vanessa Peña-Araya³ and Loreto Bravo²

¹ Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain

² Data Science Institute, Faculty of Engineering, Universidad del Desarrollo, Santiago 7610658, Chile

³ LRJ, CNRS, Inria, Université Paris-Saclay, 91190 Paris, France

* Author to whom correspondence should be addressed.

[†] This paper is an extended version of our paper published in Eduardo Graells-Garrido and Vanessa Peña-Araya. 2020. Toward An Interdisciplinary Methodology to Solve New (Old) Transportation Problems. In Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), Taipei, Taiwan, 20–24 April 2020.

Sustainability 2020, 12(15), 6001; <https://doi.org/10.3390/su12156001>

Received: 29 June 2020 / Revised: 21 July 2020 / Accepted: 22 July 2020 / Published: 26 July 2020

(This article belongs to the Special Issue [Smart Cities, City Dashboards, Planning and Evaluation of Urban Performances](#))

[View Full-Text](#)

[Download PDF](#)

[Browse Figures](#)

[Review Reports](#)

[Citation Export](#)

Etapas 4

Transferencia Tecnológica

Propiedad Intelectual

Queremos que las herramientas que hemos diseñado, desarrollado y validado sean adoptadas.

Para que la colaboración sea sustentable debe ser rentable.

Un primer paso para ello es establecer acuerdos de Propiedad Intelectual con las partes involucradas.

Juntos identificamos cómo nuestro proyecto cumple con los requisitos de patentabilidad:

Originalidad / Aplicación Industrial / Nivel Inventivo.

- 1 Diseño de Algoritmos y Modelos
- 2 Evaluación Basada en Datos
- 3 Evaluación Interactiva
- 4 Transferencia Tecnológica



UDD recibe nueva patente de invención

Publicado en: Destacado, Noticias, Noticias Home | 17 noviembre 2021

La Universidad del Desarrollo (UDD) -junto a Telefónica-, obtuvo su quinta patente de invención, concedida por el Instituto Nacional de Propiedad Intelectual (Inapi) por la tecnología denominada "**Método para determinar el Modo de Transporte de una población a partir de datos de facturación de servicios móviles (MOD)**".

La tecnología del Instituto Data Science (IDS) de la Facultad de Ingeniería UDD, cuyos inventores son Eduardo Graells y Diego Caro, permite -a través de un algoritmo- conocer la movilidad de las personas entre y dentro de las comunas a partir de la ubicación de las antenas de sus dispositivos móviles y la inteligencia tras la ciencia de datos. Para esto, se utilizan registros agregados y anonimizados de telefonía celular que permiten estimar los desplazamientos realizados.



Disponibilidad de Software

Usamos software libre y generamos software libre.

SUPER-FLY utiliza deck.gl para implementar la funcionalidad que resultó ser valiosa en las sesiones interactivas, y también los diseños nuevos resultantes de la investigación.

TS-NMF-sparse es una implementación eficiente del algoritmo TS-NMF que funciona con matrices dispersas de gran tamaño.

Vokturz/tsnmf-sparse

Topic supervised non-negative matrix factorization with sparse matrices

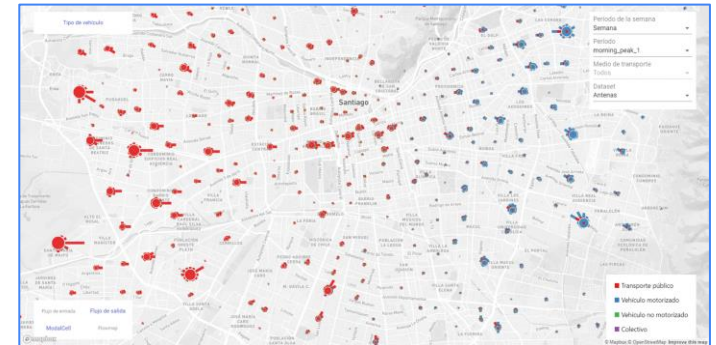
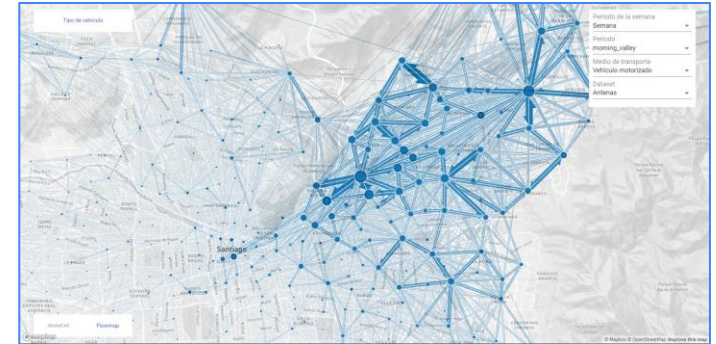
2 Contributors 0 Issues 8 Stars 1 Fork



zorzalerrante/super-fly

0 Contributors 0 Issues 0 Stars 0 Forks

- 1 Diseño de Algoritmos y Modelos
- 2 Evaluación Basada en Datos
- 3 Evaluación Interactiva
- 4 Transferencia Tecnológica



Solución Basada en Datos

Solución Transferida y Adoptada

El proyecto fue utilizado para la **primera Encuesta Origen-Destino híbrida**: encuesta tradicional + trazas digitales en **Rancagua/Machalí**.

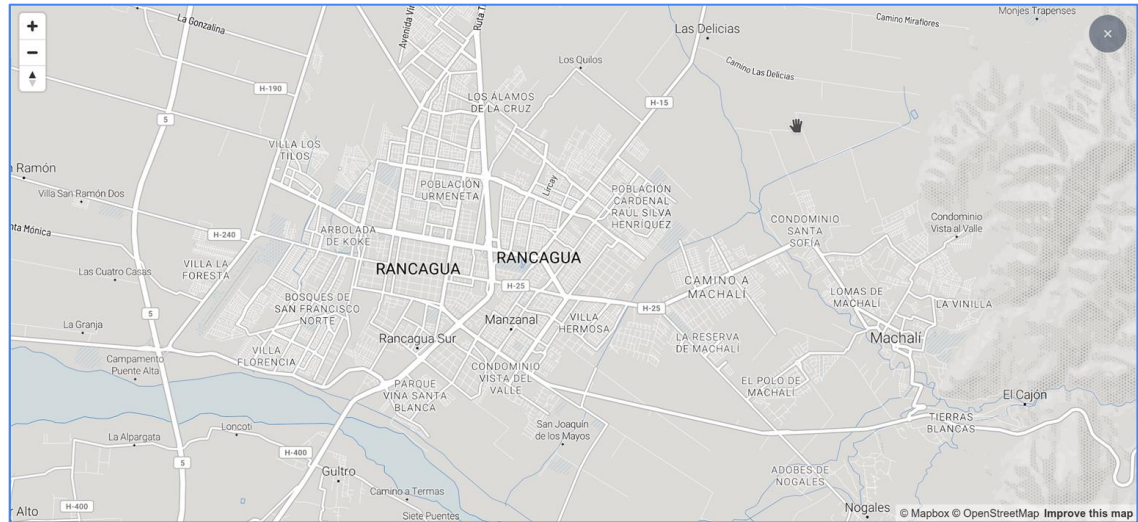
Se ganó una licitación con un proyecto conjunto entre Telefónica Movistar / CityPlanning / UDD.



cityPlanning

Telefónica

dataScience UDD



Conclusiones

Se requieren herramientas que **modelen las ciudades** para enfrentar **circunstancias actuales y futuras**.

En Data Science existen fuentes de datos y herramientas que permiten estudiar la ciudad y **complementar las herramientas tradicionales**.

Para que una herramienta de DS sea transferida y adoptada **es necesario trabajar de manera colaborativa**.

¡Existen metodologías para ello!



Imagen generada por computadora ante la descripción “Santiago, Ciudad Feliz y Sustentable, ilustrada por Estudio Ghibli” (MidJourney)



¡Gracias! Colaboremos: egraells@udd.cl

¿Preguntas?

Formalización en Fusión de Datos

Estamos trabajando en un nuevo modelo que formaliza la incorporación de conocimiento experto o del dominio en el modelo, a través de un framework de Fusión de Datos. Esta representación formal permite **generar** datos y **controlar sesgos**.

Con ese modelo estamos generando la nueva partición modal de Santiago.

