





FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE



INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE



TÉCNICAS DE DATA MINING SOBRE DATOS DE TELEFONÍA CELULAR PARA CIUDADES INTELIGENTES

Sebastián Ríos - sebastian@rios.tv

2 de Nov. 2016



PRESENTACIÓN

- Sebastián A. Ríos - sebastian@rios.tv
 - *Prof. Asistente, Departamento de Ing. Industrial, U. de Chile*
 - *Ph.D. On Information engineering, University of Tokyo (2007)*
 - *Ing. Civil Industrial, U. de Chile (2003)*
 - *Lic. en Ing. Industrial, U. de Chile*
 - *Lic. en q de la Computación, U de Chile.*
 - *Director del Centro de Investigación en Inteligencia de Negocios*
 - *Director del Magister en Ingeniería de Negocios con TICs*



FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE



INGENIERÍA INDUSTRIAL
UNIVERSIDAD DE CHILE



TÉCNICAS DE DATA MINING SOBRE DATOS DE TELEFONÍA CELULAR PARA CIUDADES INTELIGENTES

Sebastián Ríos - sebastian@rios.tv

2 de Nov. 2016



OUTLINE

.....

1. Presentación y Motivación
2. Data Science (Business Intell.)
3. Descubrimiento de Patrones de uso de Suelo



1. PRESENTACIÓN Y MOTIVACION



CENTRO DE INVESTIGACIÓN EN INTELIGENCIA DE NEGOCIOS





MISION

.....

- Ser un puente entre las empresas y la Universidad de Chile
- Ser un puente con otras Universidades y Centros de Investigación
- Mejorar el capital humano en el área de Inteligencia de Negocios
- Realizar proyectos con impacto (social/tecnológico/industrial)

FINANCIAMIENTO

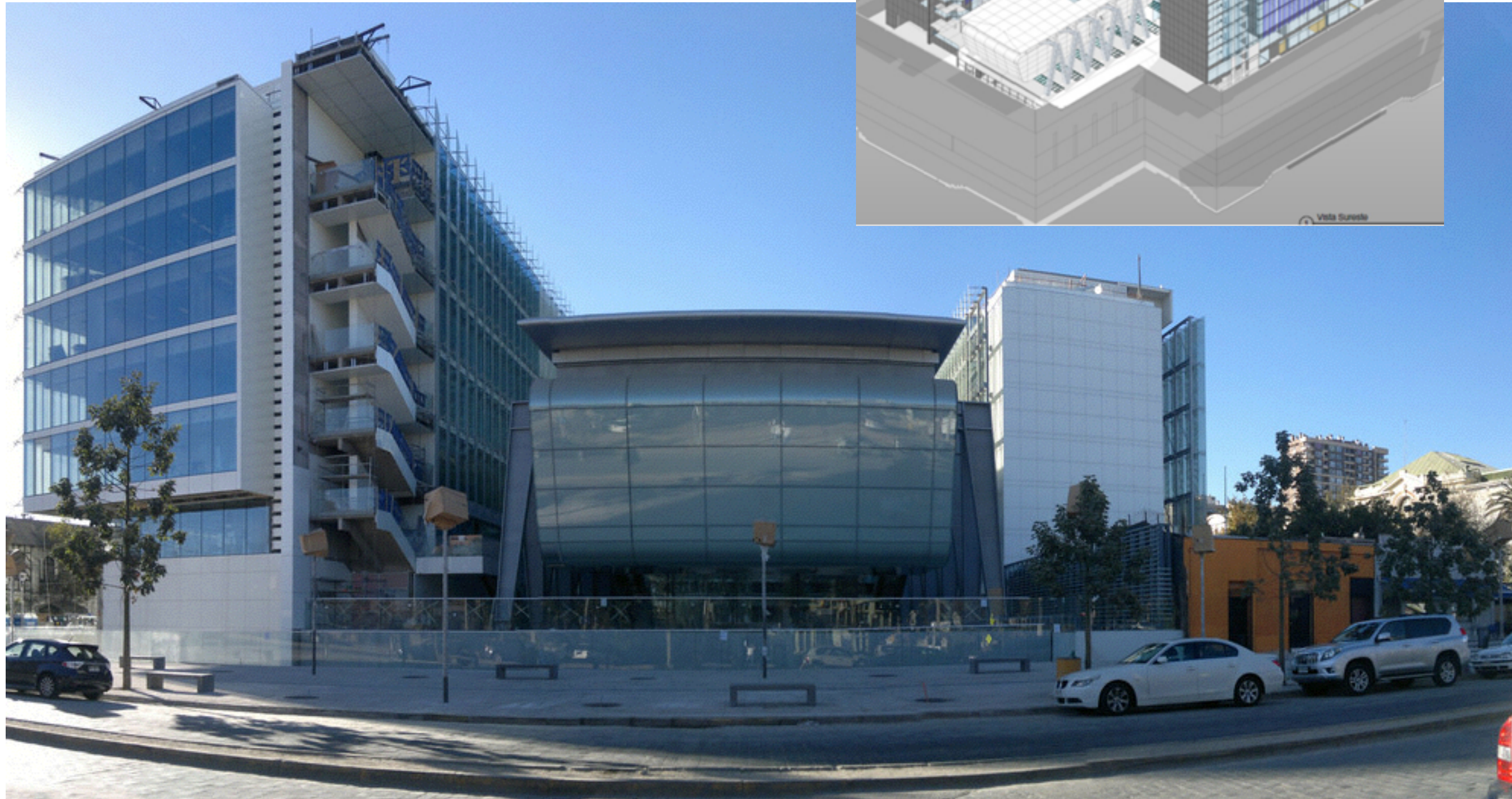
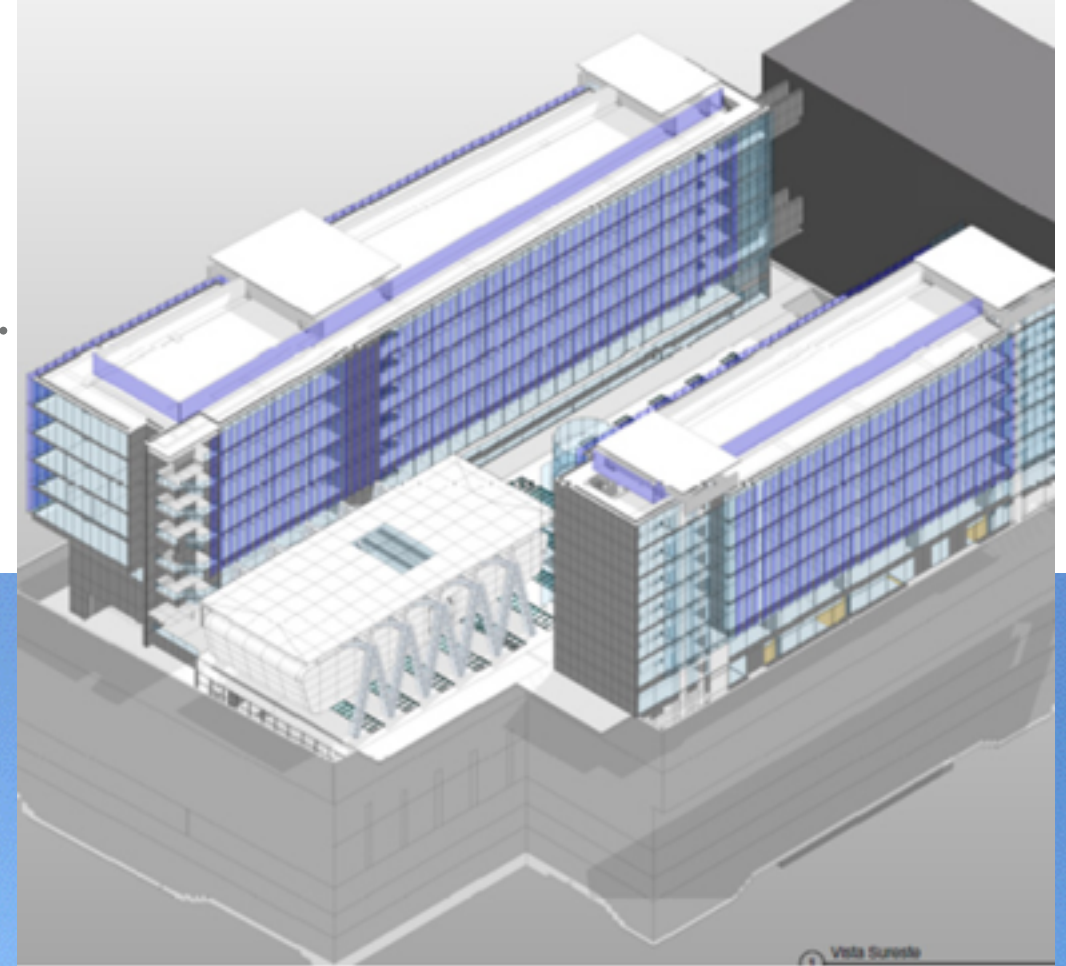
.....

- El CEINE se crea gracias a un convenio de cooperación a 5 años entre la Universidad de Chile y Telefónica Chile
- Actualmente estamos generando convenios de cooperación con otras empresas y organizaciones
- Fondos concursables en Chile y en el exterior como Iberoeka (es), newtons found (uk), ECOS-chile (fr).



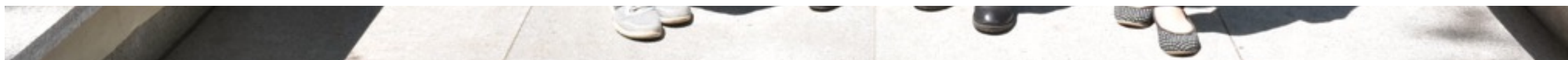
NUEVAS OFICINAS

.....





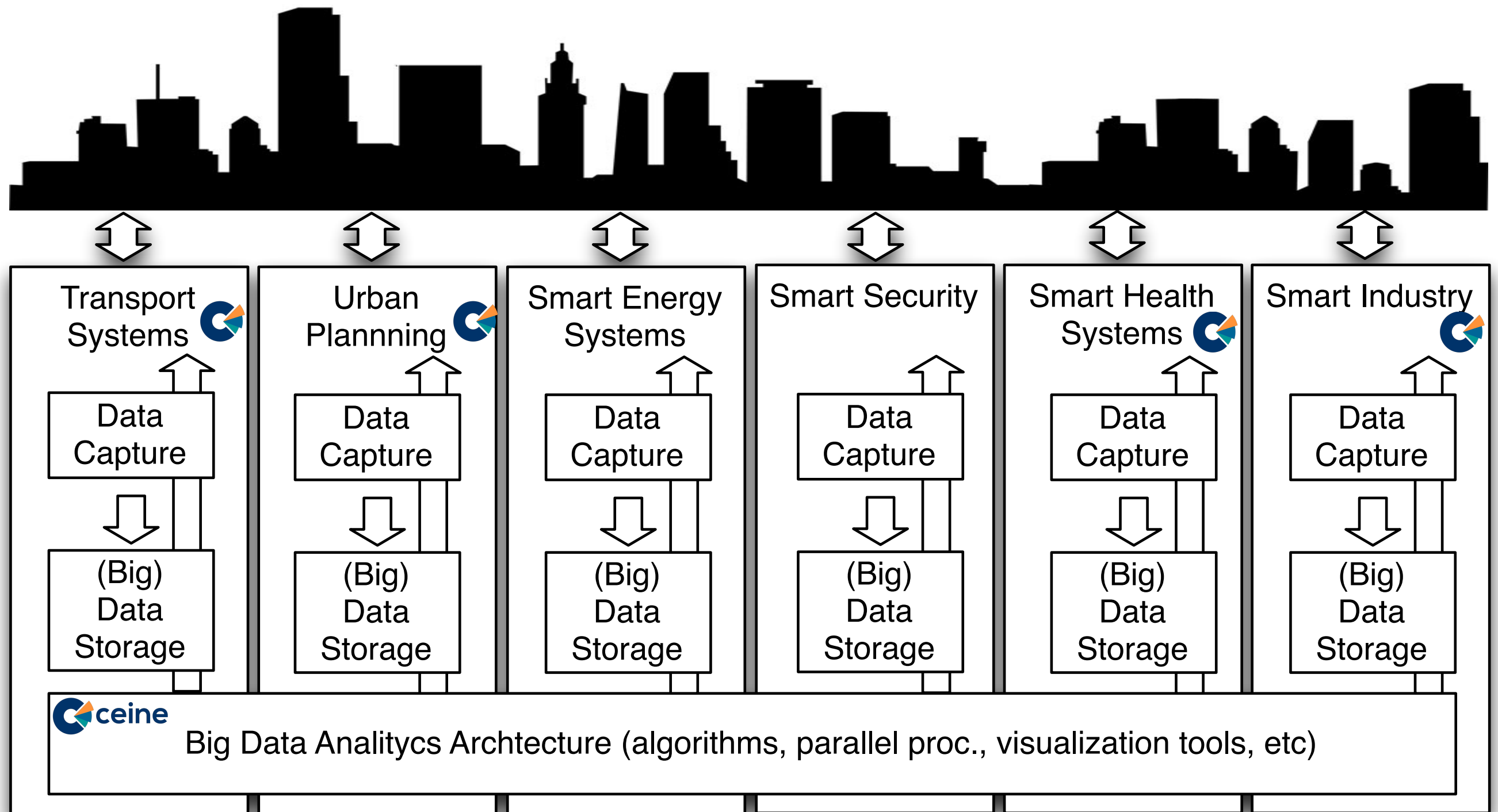
QUIENES SOMOS





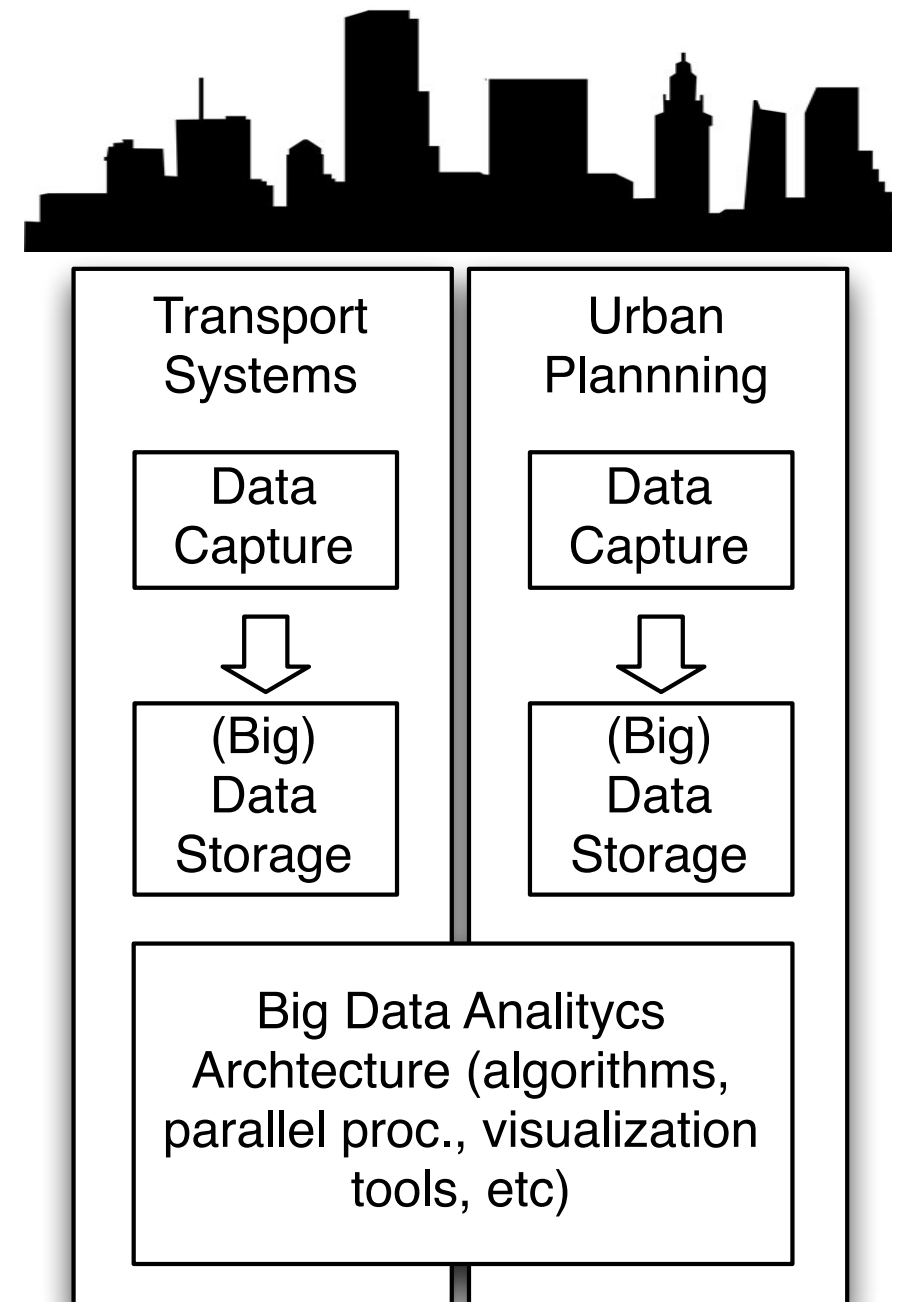
SMART CITIES

SMART CITY & (BIG) DATA ANALITICS



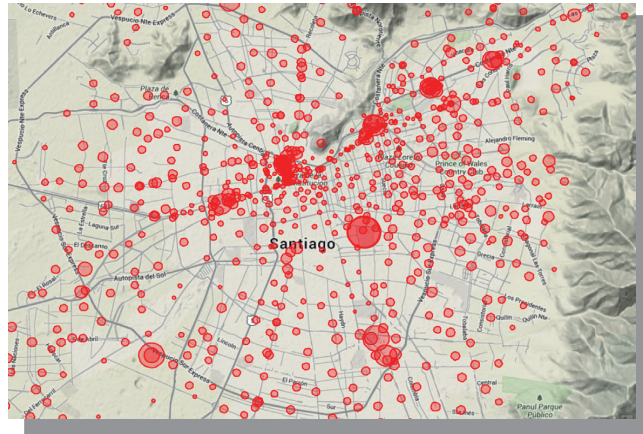
A) TRANSPORTATION SYSTEMS & URBAN PLANNING

- To develop any solution in this area we need to **track how people moves** around a city.
- We have several data sets which allow to obtain the geolocation of a person.
 - Twitter with geotagging
 - Public transportation buses GPS
 - Cell phone network antennas logs (location by triangulation)

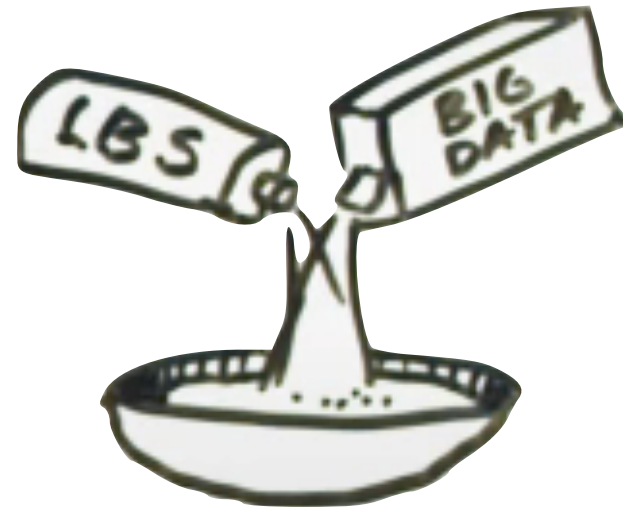


MOBILE MARKETING APPLICATIONS

.....



Patrón de llamados Sector de Comercio-Diversión



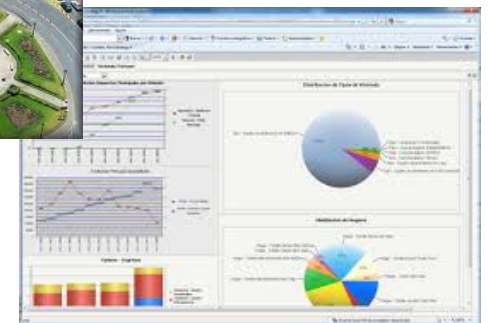
- ☒ RIGHT TARGET AUDIENCE
- ☒ RIGHT PLACE
- ☒ RIGHT TIME

Esta es una tendencia muy fuerte en los grandes de la publicidad (google) y las telcos desde el año 2009.

Algoritmos e Infraestructura



Inducir el uso de nuevos servicios, a usuarios que circulan por zonas cercanas de donde se ofrece el servicio



Construir estudios de perfil de ubicación y consumo de usuarios, para diseñar estrategias de marketing

PREDICTIVE MODELS FOR MARKETING CAMPAIGNS



Casos de uso típicos:

Lugares Frecuentes

Lugares
Esporádicos

Comportamientos
Masivos

PREDICTIVE MODELS FOR MARKETING CAMPAIGNS

.....



Lugares Frecuentes

Lugares
Esporádicos

Comportamientos
Masivos



¿Cuánto tiempo permanecen en el mall?

Sebastián A. Ríos, Ricardo Muñoz, **Land Use detection with cell phone data using topic models: Case Santiago, Chile**, Computers, Environment and Urban Systems, Volume 61, Part A, January 2017, Pages 39–48

TODOS LOS GRANDES ESTÁN INTERESADOS EN ESTO...

Mall Plaza Vespucio

vespucio.mallplaza.cl/ ▼

Ven a Mall Plaza, Encuentra lo más variado en productos y servicios, los mejores panoramas de entretención y cultura con salas de cine ademas de las ...

Catálogo De Tiendas

Panoramas - Como Llegar -
Horarios - Cine - 100% Fútbol

Horarios

Paris, Falabella y Ripley. Lunes a
Domingo; 11:00 a 21:00 hrs.

Como Llegar

Taxis colectivos: Desde Estación
Central, La Florida, La Cisterna ...

[Más resultados de mallplaza.cl »](#)

Panoramas

En diciembre, ¡Vive la cultura en
Biblioteca Viva de Mallplaza ...

Cine

... y cultura con salas de cine ademas
de ... cine. < Volver ...

Contacto

Contacto. Si tienes alguna consulta
que hacemos ...

Mall Plaza

www.mallplaza.cl/ ▼

Ven a Mall Plaza, Encuentra lo más variado en productos y servicios, los mejores panoramas de entretención y cultura con salas de cine ademas de las ...

Mall Plaza Vespucio - Cinemark Chile

www.cinemark.cl/theatres/vespucio ▼

Horarios. 30-Nov:22:50; 1-Dic:14:2017:0019:5022:30; 2-Dic:14:2017:0019:5022: 30; 3-Dic:00:2011:2014:2017:0019:5022:30 ...

Mall Plaza Vespucio - Tiendas, ofertas y descuentos | Tiendeo

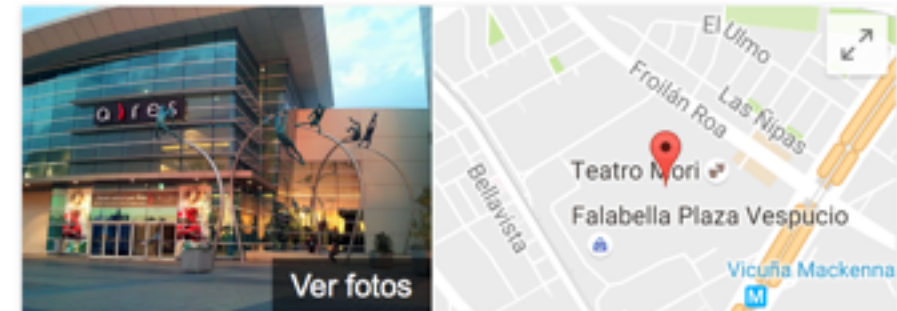
www.tiendeo.cl > [Tiendeo en La Florida](#) > [Malls](#) ▼

Descubre las nuevas promociones de Mall Plaza Vespucio y las mejores ofertas y descuentos de tus malls favoritos en La Florida.

Mall Plaza Vespucio - Cory

www.cory.cl > [Locales](#) ▼

Local: Mall Plaza Vespucio. Av. Vicuña Mackenna 7110 Local 327 / La Florida / Teléfonos: (02) - 2 - 221 1617. Ver Cory en Mall Plaza Vespucio en un mapa ...



Mall Plaza Vespucio

[Sitio web](#)

[Indicaciones](#)

3,9 ★★★★★ 475 comentarios de Google

Centro comercial

Dirección: Av. Vicuña Mackenna Ote. 7110, La Florida, Región Metropolitana, Chile

Horario: Hoy abierto · 10-21:30 ▼

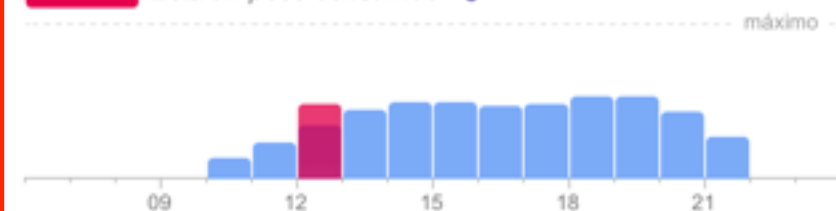
Teléfono: +56 2 2585 7000

[Sugerir una edición](#) · [¿Eres propietario de esta empresa?](#)

Horarios populares ?

Jueves ↕

EN VIVO Está un poco concurrido · [¿Es correcto?](#)



Planifica tu visita: La gente suele pasar De 25 minutos a 1,5 horas aquí.

Opiniones

475 comentarios de Google

[Escribe una opinión](#)

[Agregar una foto](#)

TODOS LOS GRANDES ESTÁN INTERESADOS EN ESTO...

Jueves 1 de Dic. 2016 - 12.10pm

Horarios populares ?

Jueves ↕

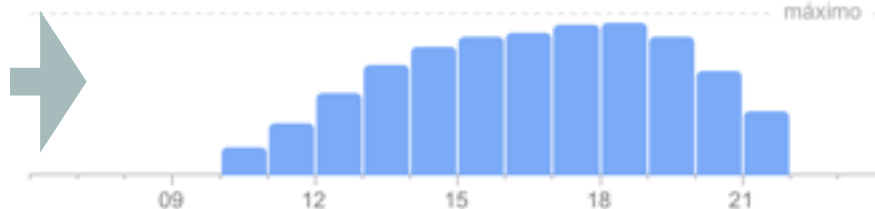
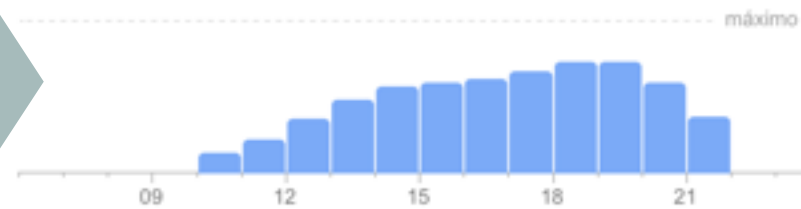
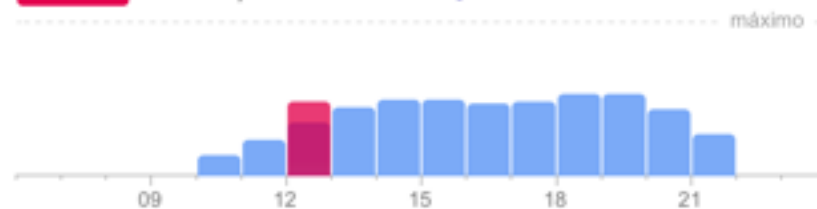
Horarios populares ?

Viernes ↕

Horarios populares ?

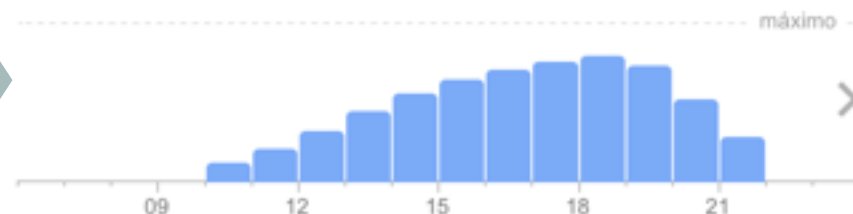
Sábados ↕

EN VIVO Está un poco concurrido - ¿Es correcto?



Horarios populares ?

Domingos ↕



Horarios populares ?

Lunes ↕

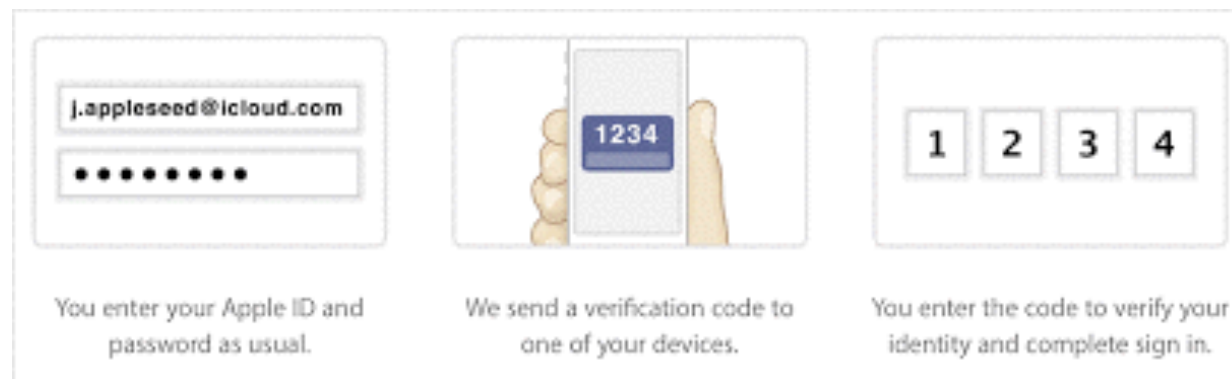


No hay que sorprenderse mucho, es un caso de uso de negocio y por lo tanto, todas las compañías trataran de acceder a los datos de su celular

SE INVENTAN MECANISMOS PARA SABER “QUIEN ES QUIEN” TRAS UN DISPOSITIVO

.....

Two-step verification for iOS



*Actualización de Apps o
Sistemas Operativos que
activan el GPS por defecto*

OPTIMIZACIÓN DEL PROCESO DE ASIGNACIÓN DE CARRERAS DE TAXIS



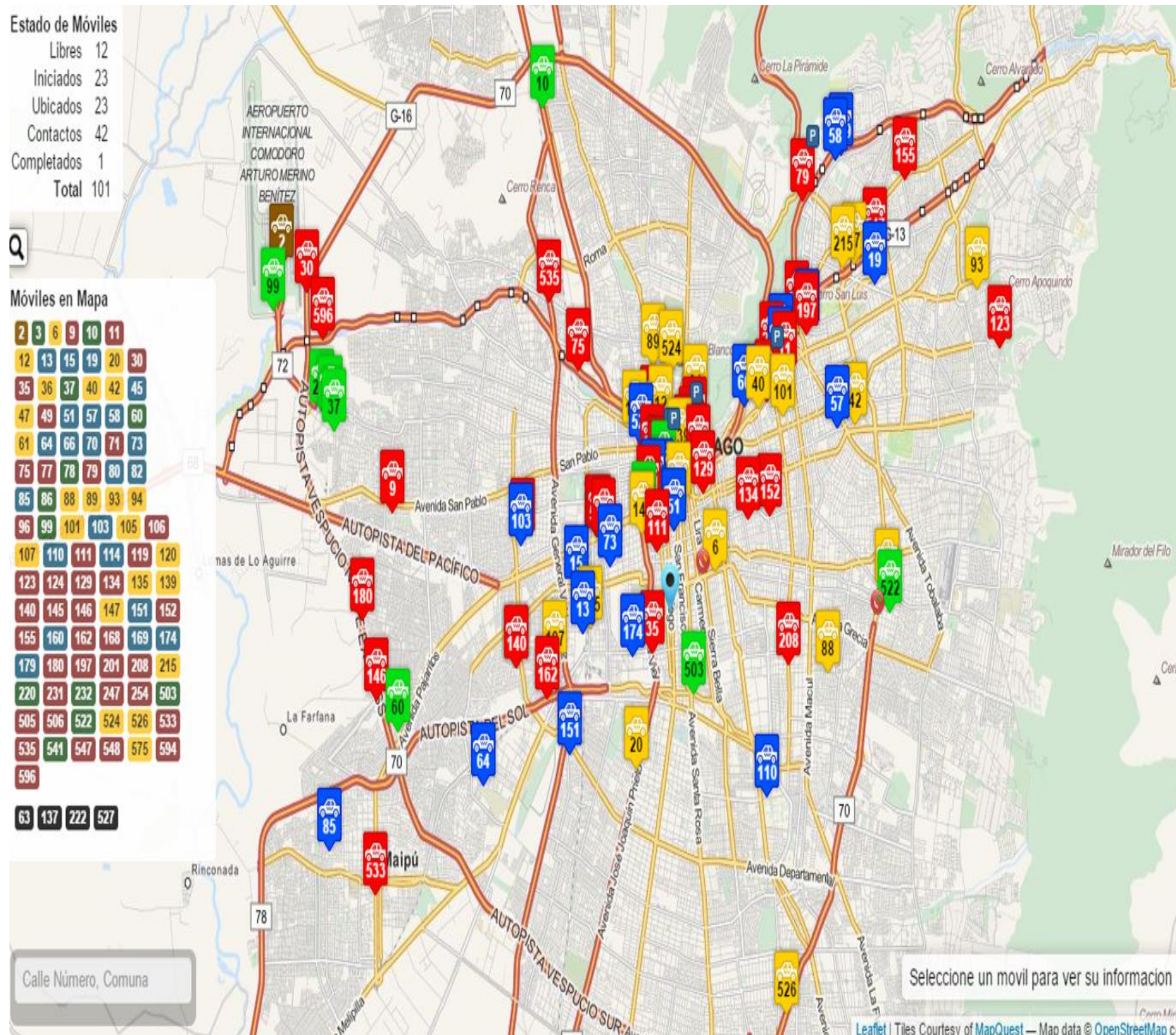
ASIGNACIÓN OPTIMA DE CARRERAS

Estado de Móviles

Libres	12
Iniciados	23
Ubicados	23
Contactos	42
Completados	1
Total	101

Móviles en Mapa

2	3	6	9	10	11
12	13	15	19	20	30
35	36	37	40	42	45
47	49	51	57	58	60
61	64	66	70	71	73
75	77	78	79	80	82
85	86	88	89	93	94
96	99	101	103	105	106
107	110	111	114	119	120
123	124	129	134	135	139
140	145	146	147	151	152
155	160	162	168	169	174
179	180	197	201	208	215
220	231	232	247	254	503
505	506	522	524	526	533
535	541	547	548	575	594
596					
63	137	222	527		



186 móviles de una empresa de radio taxis

926 carreras al día y 17.633 Kms recorridos promedio al día.

Se logró una reducción del 30% en los Kms recorridos por la flota para satisfacer el mismo número de carreras

Reducción de Costos Operacionales: €1,6 Millones de Euros al año

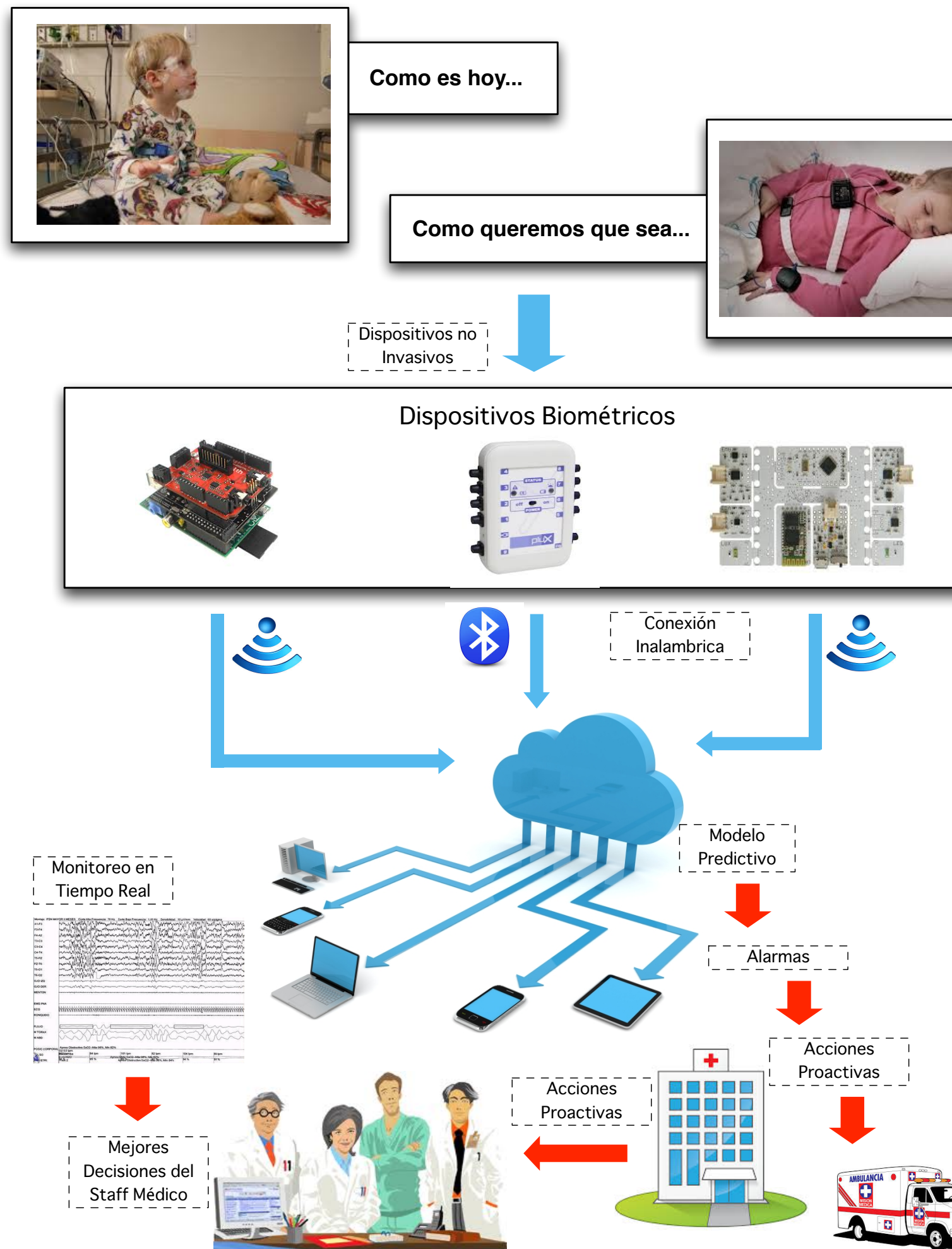


SMART MEDICAL SYSTEMS

.....

MONITOREO DE PACIENTES DOMICILIARIOS

- Plataformas de sensores de bajo costo



1. Plataforma de Sensores: Raspberry Pi



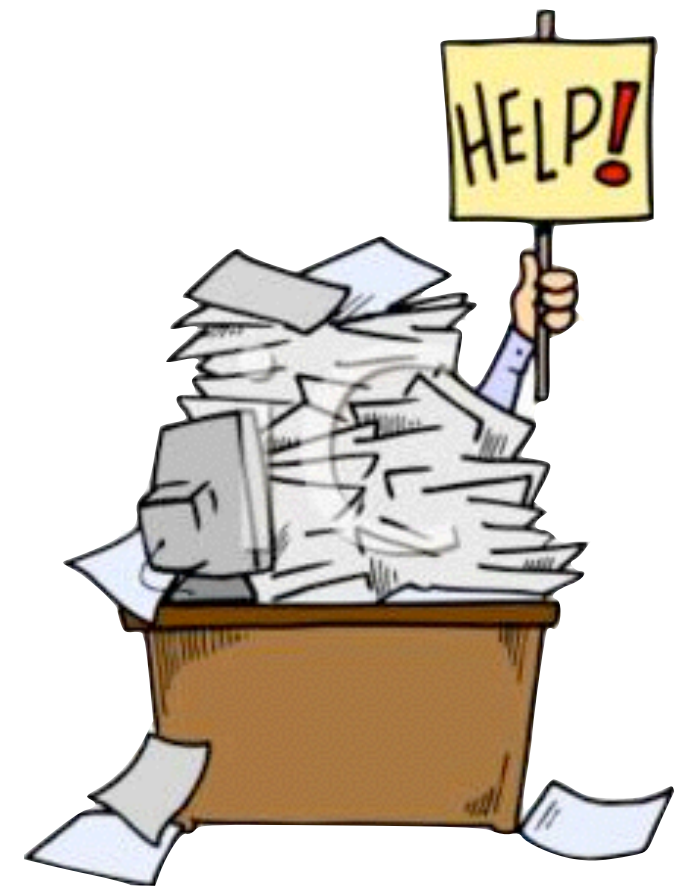
- Puede emular un Polisomnograma
- 21 canales (ECG, EEG, Sat O2, etc.)



2. KNOWLEDGE DISCOVERY IN DATA BASES = ... = DATA SCIENCE

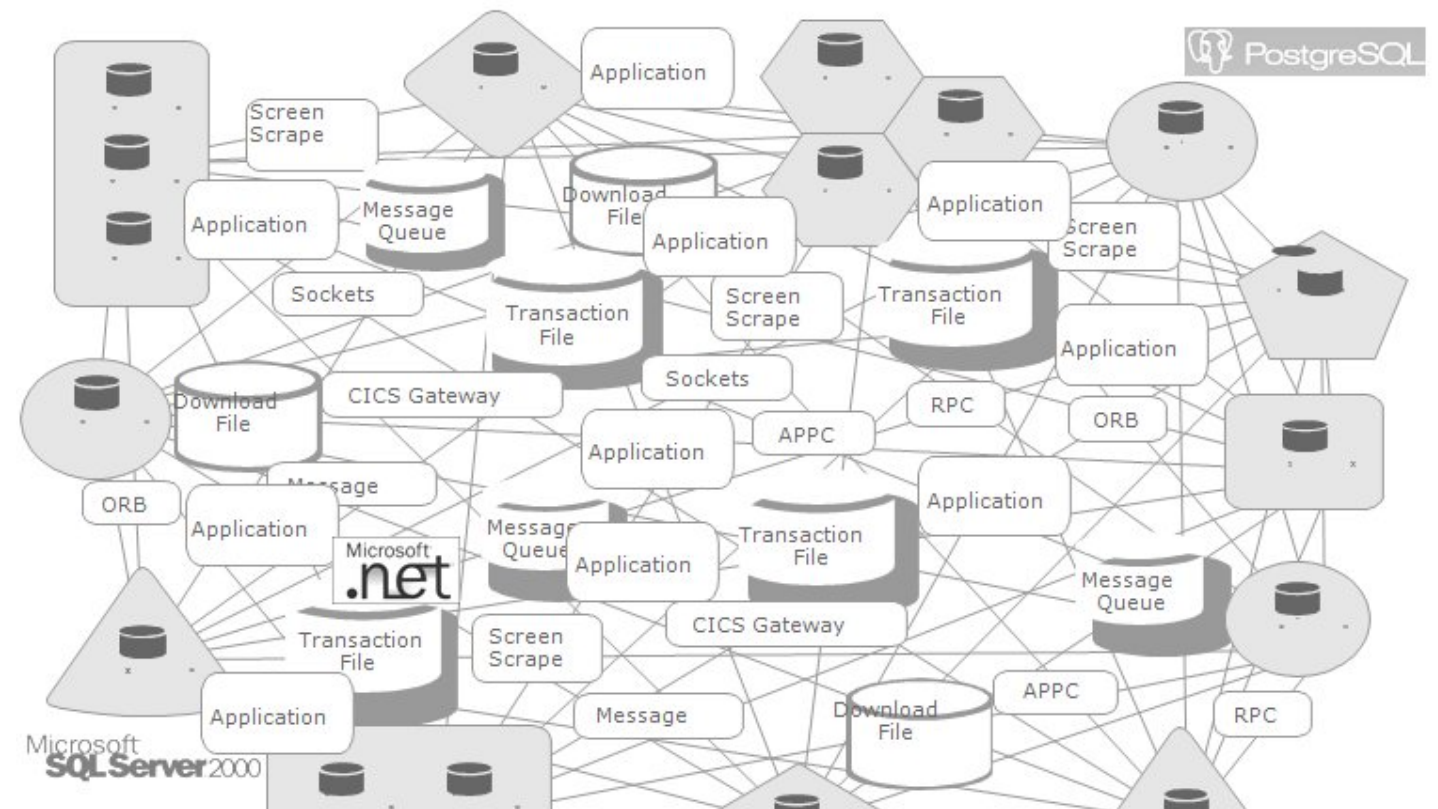
¿CUAL ES LA PRINCIPAL MOTIVACIÓN?

- Para cualquier negocio u organización la toma de decisiones debe ser **ágil, informada y oportuna** para generar una ventaja competitiva.
- Sin embargo, hoy en día la **cantidad de datos** con los que contamos en la empresa/organización es tan **grande** que muchas veces no es posible cumplir con lo anterior



ENORME CANTIDAD DE DATOS EN MULTIPLES PLATAFORMAS

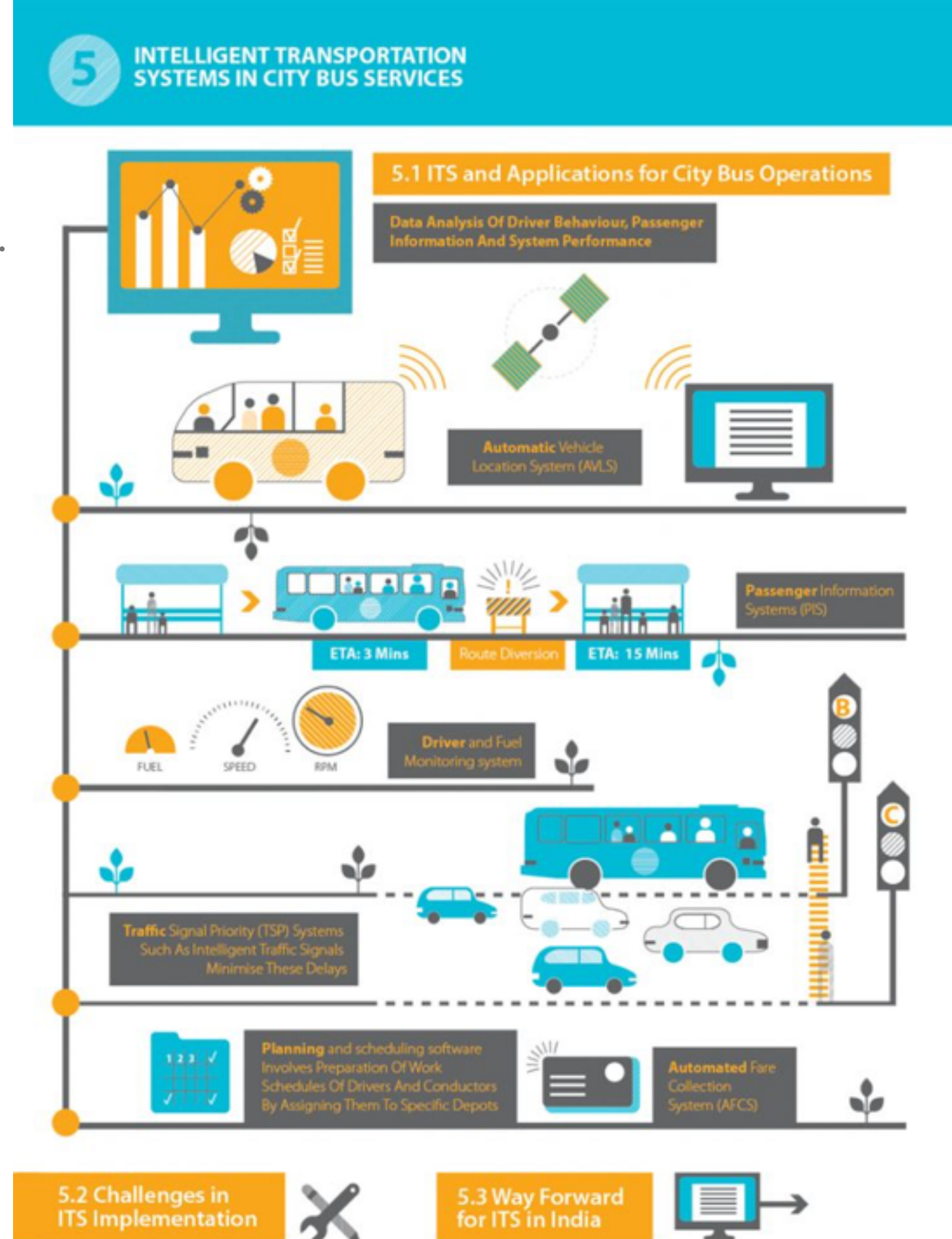
- Multitud de Plataformas
 - SAP, excel, e-mail, etc
- Qué problemas con los Datos?
 - están distribuidos
 - en formatos distintos
 - duplicados
 - desactualizados



Problema de Calidad de Datos

LO MISMO OCURRE CON LOS DATOS DE UNA CIUDAD

- Para que una ciudad pueda contar con algoritmos inteligentes que la hagan más eficiente se requieren datos.
- Los que muchas veces están dispersos en varias aplicaciones no están comunicadas





2.1 TÉCNICAS DE DM

¿QUÉ TAREAS SE PUEDEN REALIZAR CON BI?

- Business Intel. es utilizado comúnmente para realizar una de las siguientes tareas:
 1. Descripción
 2. Estimación
 3. Clasificación
 4. Predicción
 5. Asociación
 6. Clustering (o agrupamiento)

TODOS LOS ALGORITMOS SE ENTRENAN (PARA APRENDER)

Modelos Supervisados

Tareas predictivas

Hay medidas de error o de capacidad de ajuste

Existe una variable objetivo a modelar

Ej: predecir la demanda por tickets de la prox semana; probabilidad de fuga de un cliente

Modelos No Supervisados

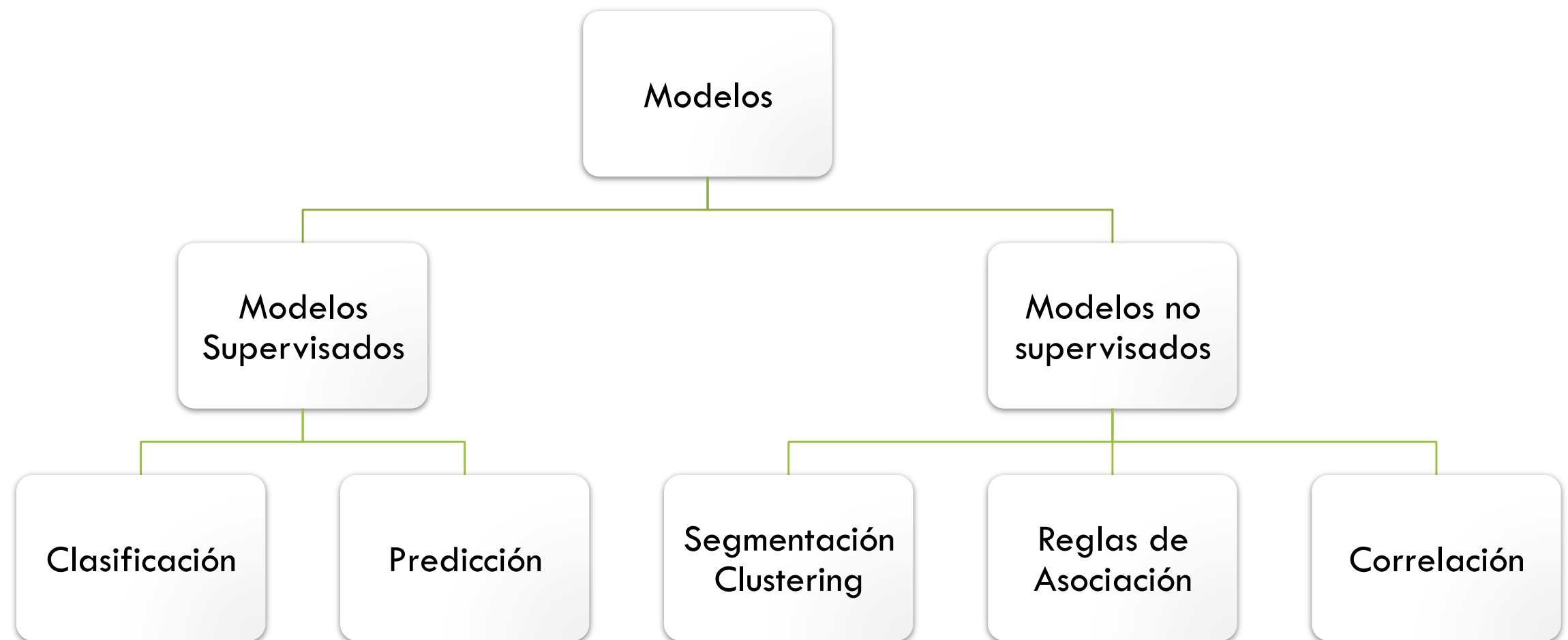
Tareas descriptivas o de resumen de un comportamiento global

Medidas de error más difusas relativas a la capacidad de interpretación de los datos

Existen muchas variables que describen distintos objetos

Ej: Asociación entre productos: Si hace upgrade entonces comprará duty free

TAXONOMÍA DE MODELOS



DESCRIPCIÓN

- Los analistas solo se centran en como encontrar los patrones y tendencias que existen en los datos.
 - Cuales son las variables correlacionadas con un fenómeno y cuales no
 - Que variables explican una cierta tendencia, etc.
 - Descubrir un patrón (cluster) y utilizarlo para formarnos una descripción de los elementos del cluster (generalización).
- En el ejemplos en datos de movilidad:
 - Poder mostrar las comunas desde las cuales llegan los clientes de un centro comercial durante la semana.
 - Podríamos describir desde que sectores de la ciudad se mueven más bicicletas al centro de la ciudad

ESTIMACIÓN

- En base a variables explicativas se construye un modelo matemático que permite estimar una variable objetivo.
- Ej. Podríamos estar interesados en la estimación del porcentaje de accidentes en una carretera por hora.
- La estimación es similar a la clasificación, salvo que la variable de destino es numérica y no categórica



7 Septiembre, 2016

Crean sistema que predice hasta un 80% de los accidentes de tránsito en autopistas.

Utilizando la información que los pódicos de cobro de las carreteras entregan, un grupo de ingenieros apoyados por el ISCI diseñaron un proyecto único en el mundo.

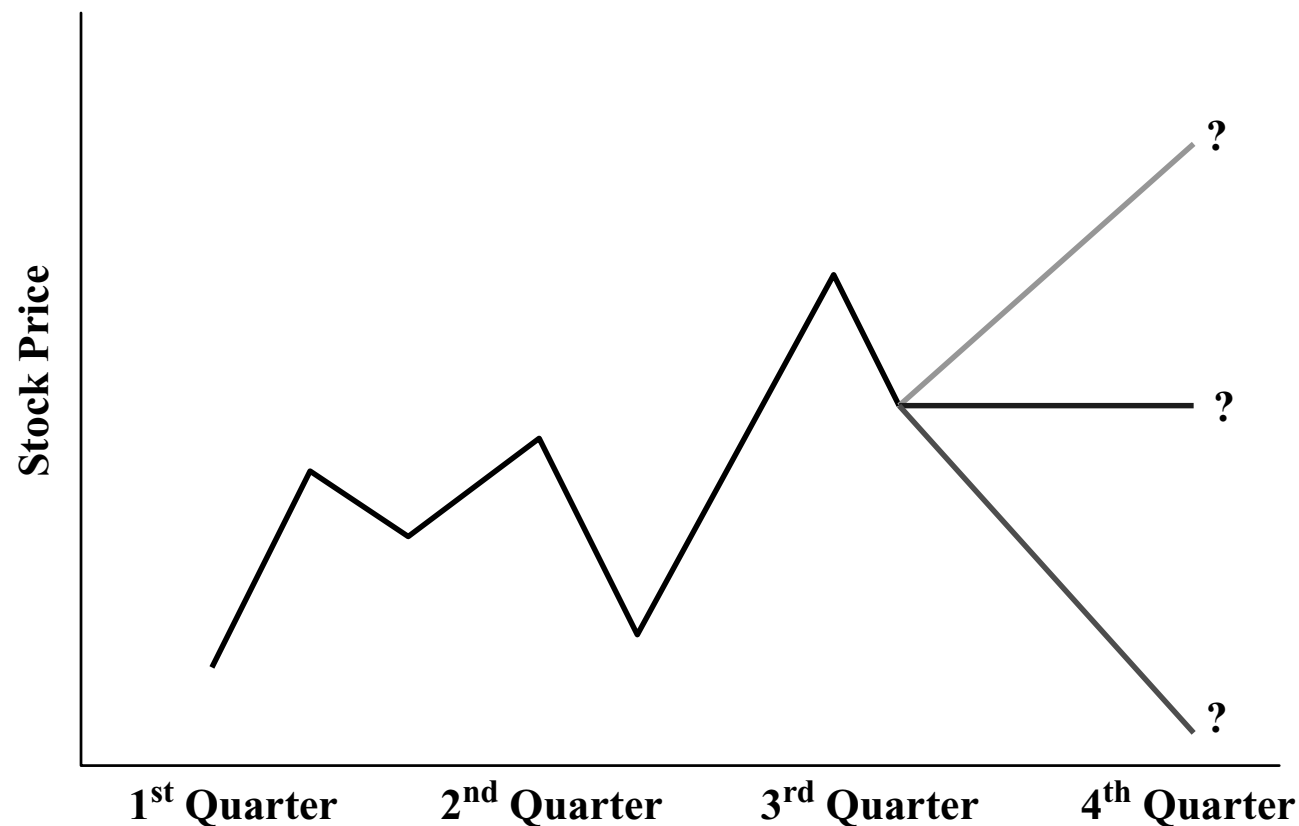


CLASIFICACIÓN

- En este caso, tenemos una variable objetivo que no es numérica, si no que es una categoría o de una clase.
- En el ejemplo anterior:
 - La variable objetivo Accidentes, podría subdividirse en: No hay accidentes, puede haber entre 1 y 3 accidentes; y hay muchos accidentes (sobre 3).
 - También es sistemas de salud inteligentes, una variable es el riesgo de crisis respiratoria. La variable puede ser: Sin riesgo, con riesgo bajo, con riesgo moderado, y con alto riesgo.

PREDICCIÓN

- La predicción es similar a la estimación, excepto que para la predicción, los resultados se encuentran en el futuro.
- O sea, que no son valores conocidos al momento de construir el modelo
 - Ej. Predecir el precio de una acción tres meses en el futuro
 - Ej2. Predecir cual es el riesgo de una crisis respiratoria en un niño durante las próximas 4 horas.



Nota: Análisis de regresiones es la metodología estadística más utilizada para predicción numérica, por lo cual, los dos términos se ocupan como sinónimos usualmente.

ASOCIACIÓN

- La tarea de asociación en minería de datos consiste en encontrar los atributos que "van juntos."
- Las reglas de asociación tienen la siguiente forma:
 - **Si** <antecedente> **entonces** <consecuente>

Rule: $X \Rightarrow Y$

$$\text{Support} = \frac{\text{frq}(X, Y)}{N}$$
$$\text{Confidence} = \frac{\text{frq}(X, Y)}{\text{frq}(X)}$$
$$\text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}$$

N: Total de transacciones analizados
(típicamente compras -> boletas)

Frq(X): Numero de veces (frecuencia) que
aparece el elemento X en las transacciones

Frq(X,Y): Numero de veces (frecuencia) que
aparecen los elementos X e Y en las
transacciones

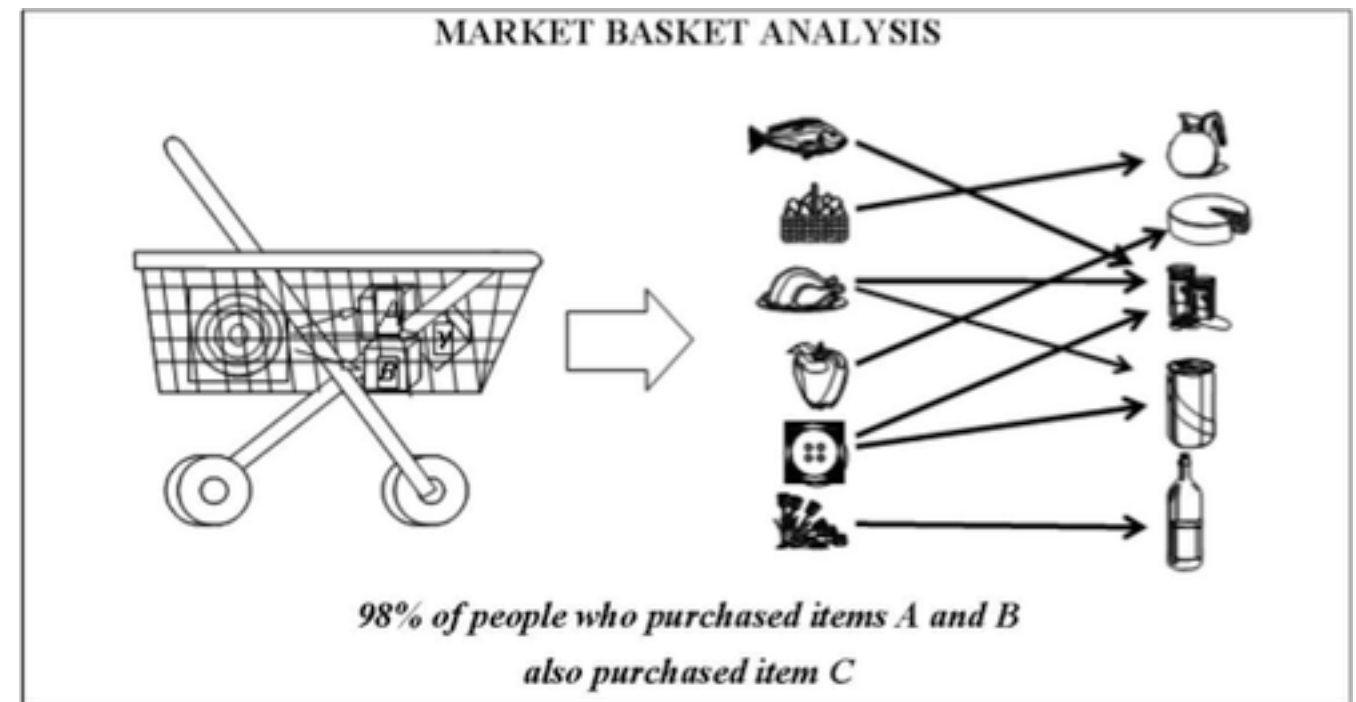
EJ. REGLAS DE ASOCIACIÓN

- Para una compañía de telecomunicaciones una regla de asociación puede ser:
- Si (Cliente, ADSL 1MB) & (Cliente, Servicios de Telefonía) entonces (Cliente, Televisión Digital)
 - Soporte: 60%
 - En el 60% de todas las transacciones, los clientes contratan ADSL 1MB y Servicios de Telefonía
 - Confianza: 67%
 - Significa que cuando el Cliente compra ADSL 1MB y Servicios de Telefonía Digital, en un 67% de los casos el Cliente además contrata televisión digital.

MARKET BASKET ANALYSIS

.....

- Esta es el área más utilizada en la industria de las reglas de asociación.
- Consiste en entender cuales son las diversas canastas de productos que los clientes compran con frecuencia
- Esta información es utilizada para poder dar una recomendación a los clientes cuando vuelven al negocio a comprar.



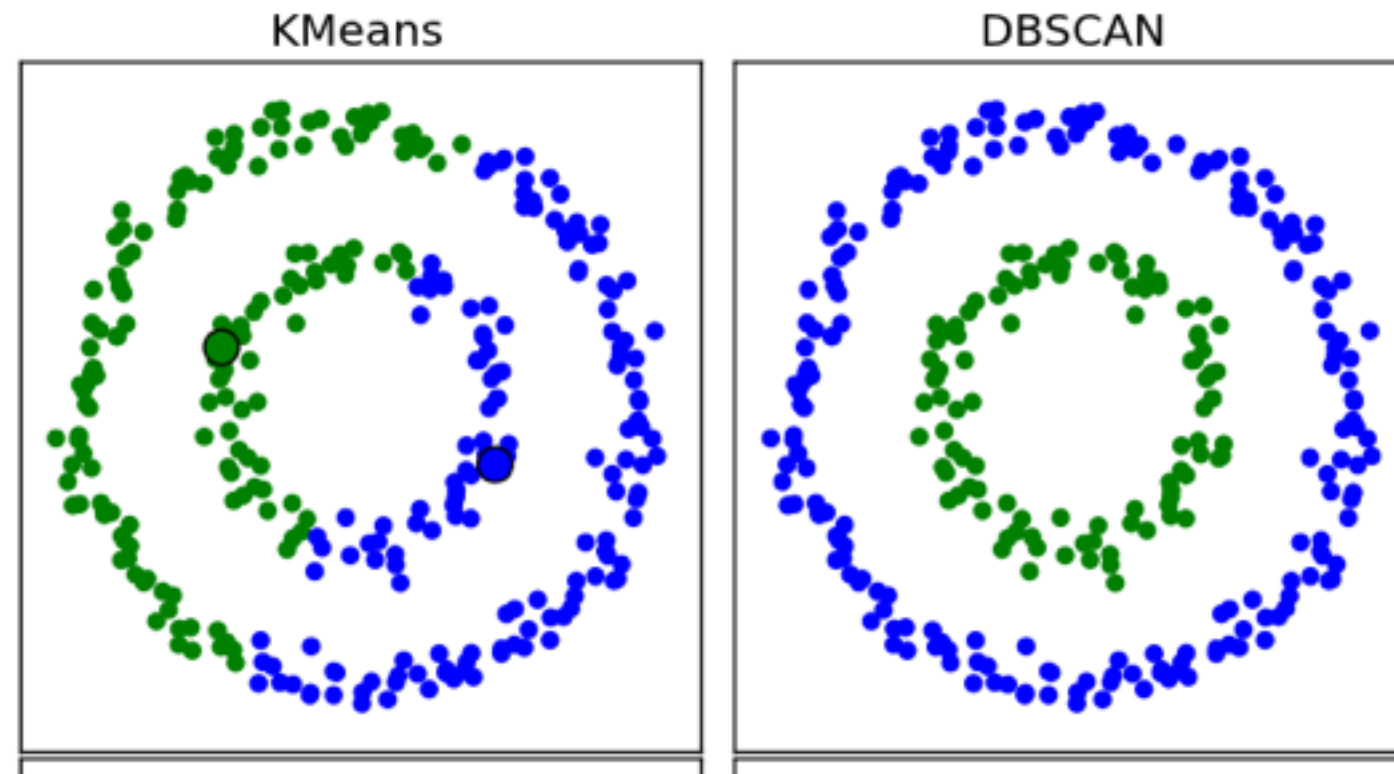
☀ Referencias:

- ☀ Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, in G. Piatetsky-Shapiro & W. J. Frawley, eds, 'Knowledge Discovery in Databases', AAAI/MIT Press, Cambridge, MA.
- ☀ R. Agrawal; T. Imielinski; A. Swami: Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference 1993: 207-216

CLUSTERING

- Consiste en generar grupos de objetos
- A diferencia de las técnicas de estimación, predicción y clasificación donde se conocen las etiquetas o el valor de la variable que se desea modelar; en este caso **NO** se requiere tener conocimiento previo respecto de los datos.
 - Las técnicas se encargan de analizar y aprender desde los mismos datos, para luego agruparlos en clusters.
- Por este motivo, los algoritmos de clustering corresponden a técnicas de **aprendizaje no supervisado**.

EJ. CLUSTERING

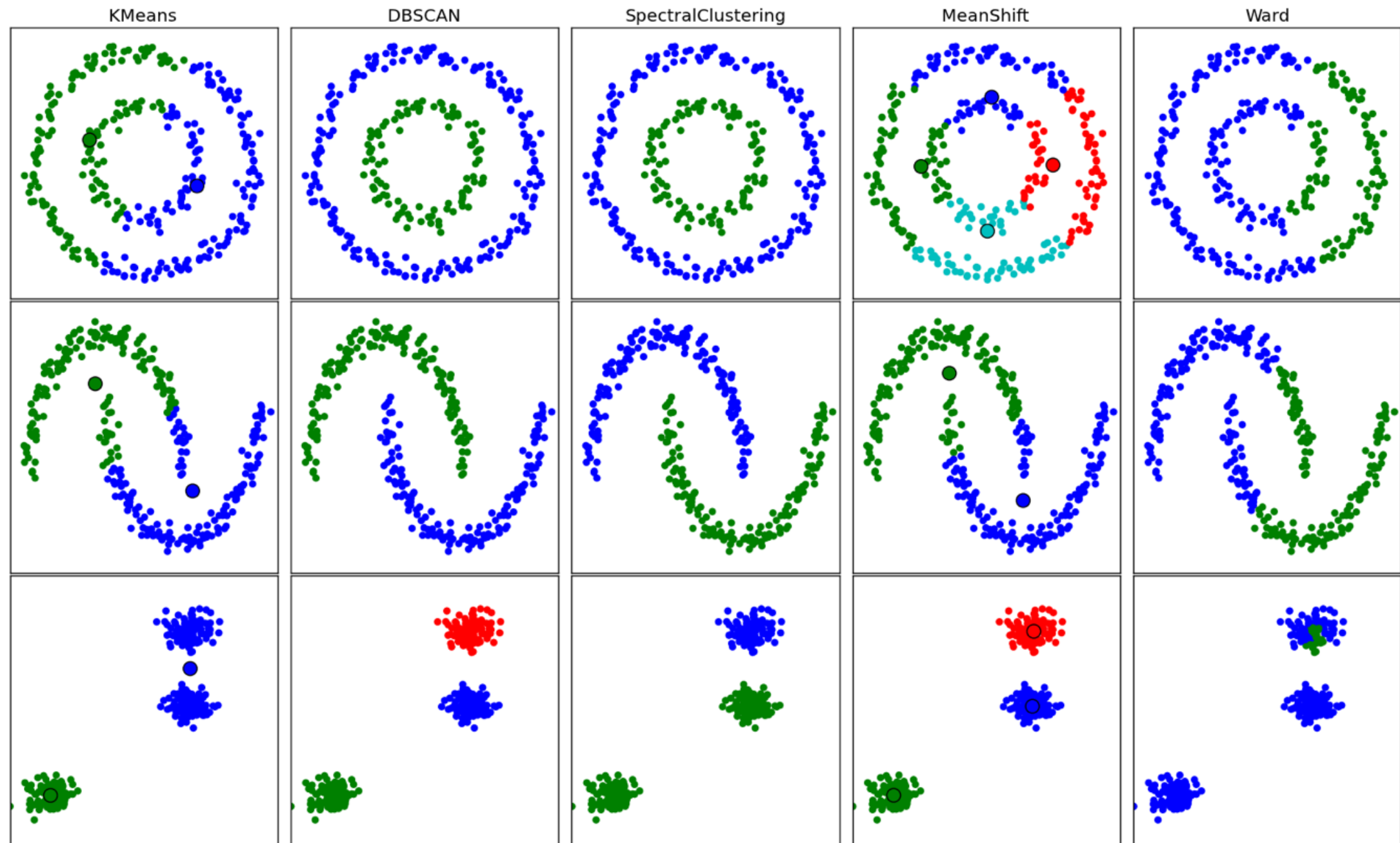


SUPONGAMOS QUE QUEREMOS CLUSTERIZAR LOS PUNTOS DE LA FIGURA EN DOS GRUPOS, LOS PUNTOS EN EL ANILLO EXTERNO Y LOS EN ANILLO INTERNO.

- SIN EMBARGO, LOS DATOS SON SOLO PUNTOS EN EL PLANO
- ES DECIR, NO DICEN SI UN PUNTO (X,Y) PERTENECE AL ANILLO EXTERNO O AL INTERNO.
 - ESTO SIGNIFICA QUE NO ESTA ETIQUETADO (O QUE SON DATOS NO SUPERVISADOS DADO EL PROBLEMA)

PARA RESOLVER ESTE PROBLEMA UTILIZAMOS ALGORITMOS DE CLUSTERING

PROBLEMAS DE ESTOS ALGORITMOS



*Tradicionales: Support Vector Machines, ANN, Naive Bayes,
Particionamiento Jerárquico, Aglomertivos, entre otros.*

¿CÓMO MEDIMOS CALIDAD?

► Normalmente es posible utilizar los índices internos:

$$Intra - C = \frac{\sum_{i=1}^K \left(\sum_{z,t \in C_i} d(z,t) / |C_i| \right)}{K}$$

Cohesión de los
elementos del
cluster (intra-cluster)

$$Inter - C = \frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^K d(c_i, c_j)}{\sum_{i=1}^{K-1} i}$$

Separación de
los clusters
(inter-cluster)

Donde:

d : es la distancia Euclidiana

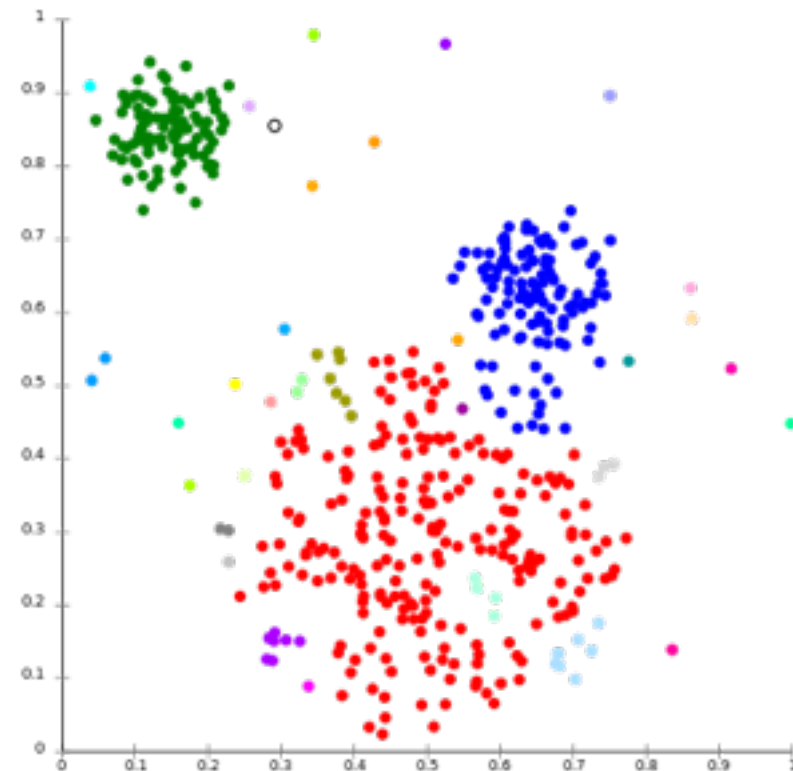
n : es la cantidad de elementos del conjunto de datos

z : representa un elemento del conjunto de datos

K : es la cantidad de clusters

C_i : es el cluster i

c_i : es el centroide del cluster i



Razón muy útil:

$$\frac{Intra-Cluster}{Inter-Cluster}$$



Pequeña (tiende a 0)
Es mejor



2.2 EL CONCEPTO DE PATRÓN



CONCEPTO DE PATRÓN

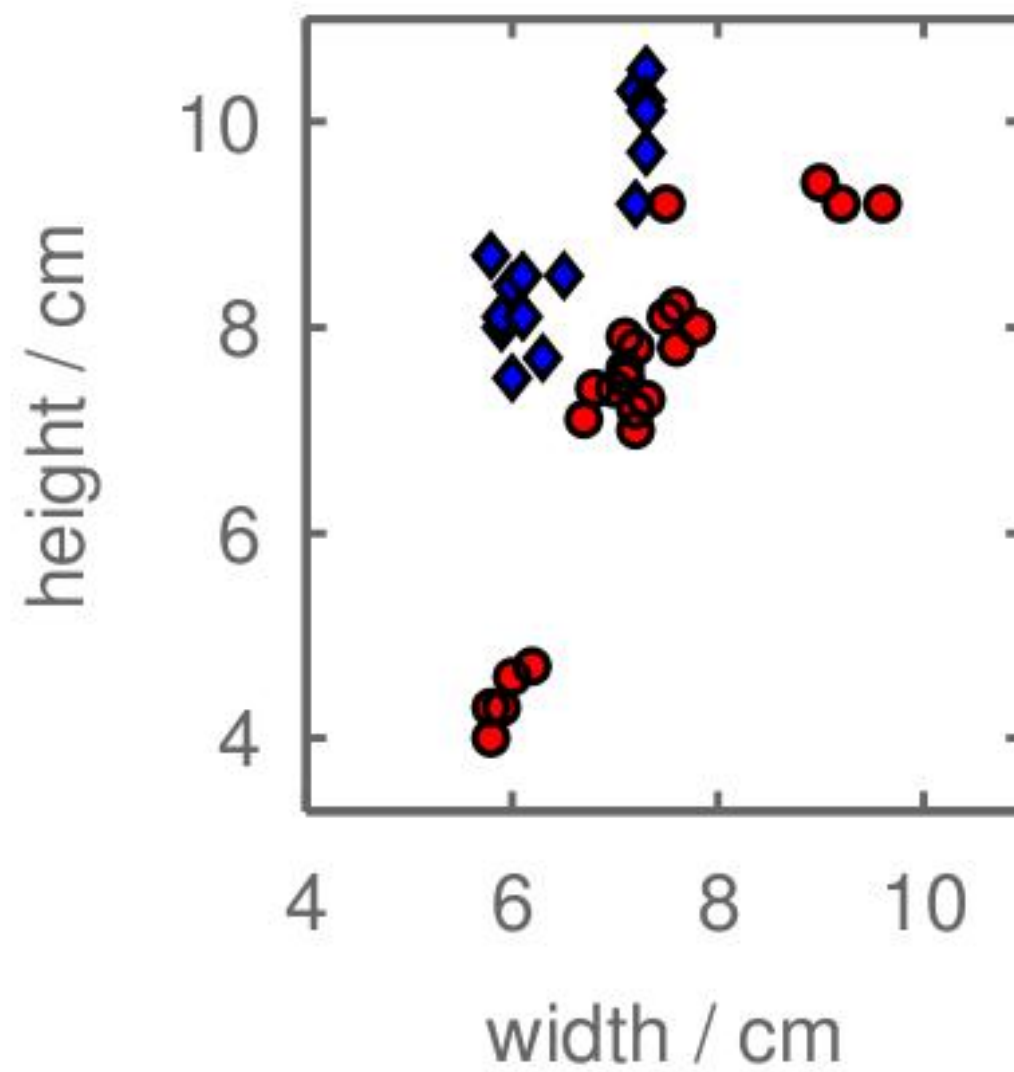


**AQUI PODEMOS APRECIAR
VARIOS PATRONES:**

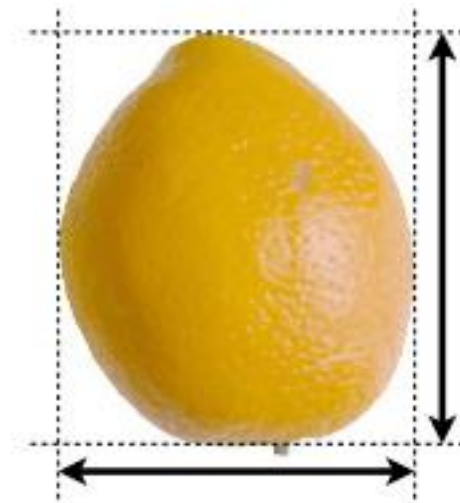
- DE COLOR**
- FORMA**
- SABOR**
- PESO**
- TEXTURA**

**¿QUE VARIABLES PODRÍAMOS
UTILIZAR PARA DESCRIBIR LA
FORMA?**

PATRONES Y REPRESENTACIÓN CARTESIANA

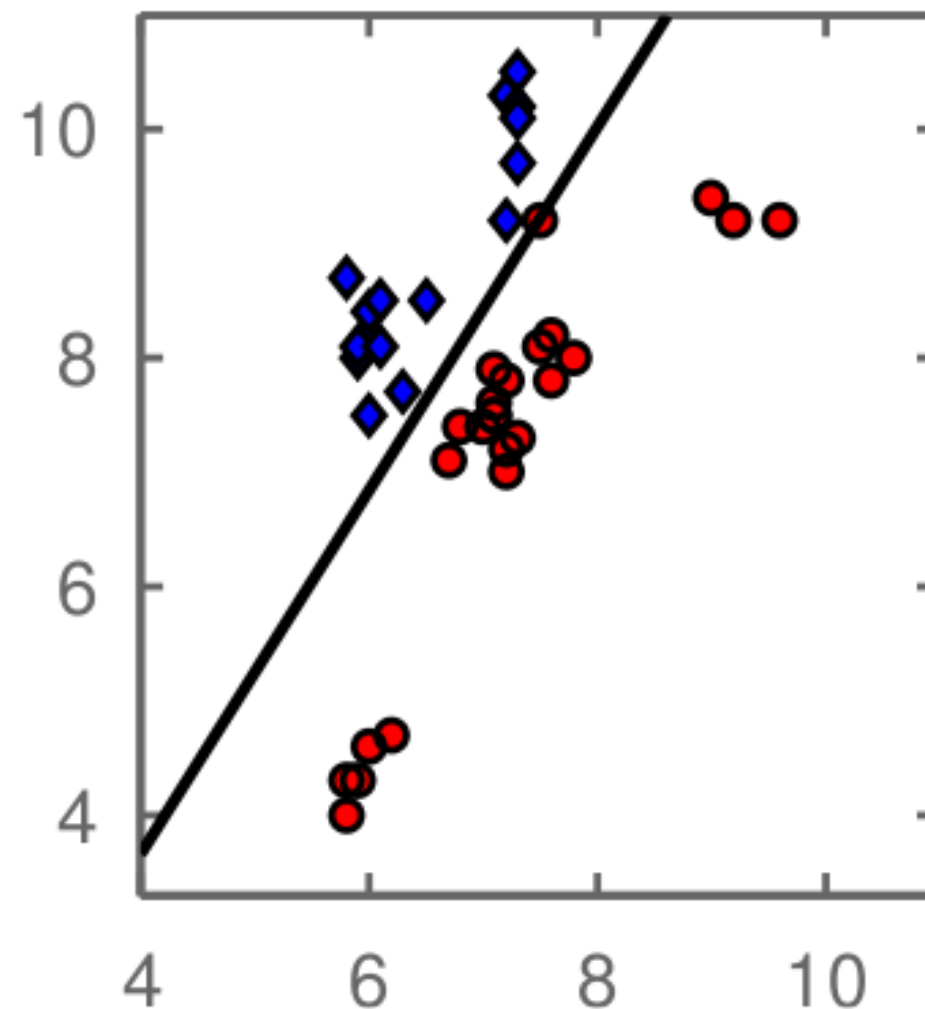


● Oranges
◆ Lemons



USO DE PATRONES

Clasificador Lineal (Aprendizaje Supervisado)



EN LUGAR DE NARANJAS Y LIMONES PODRIAMOS USAR:

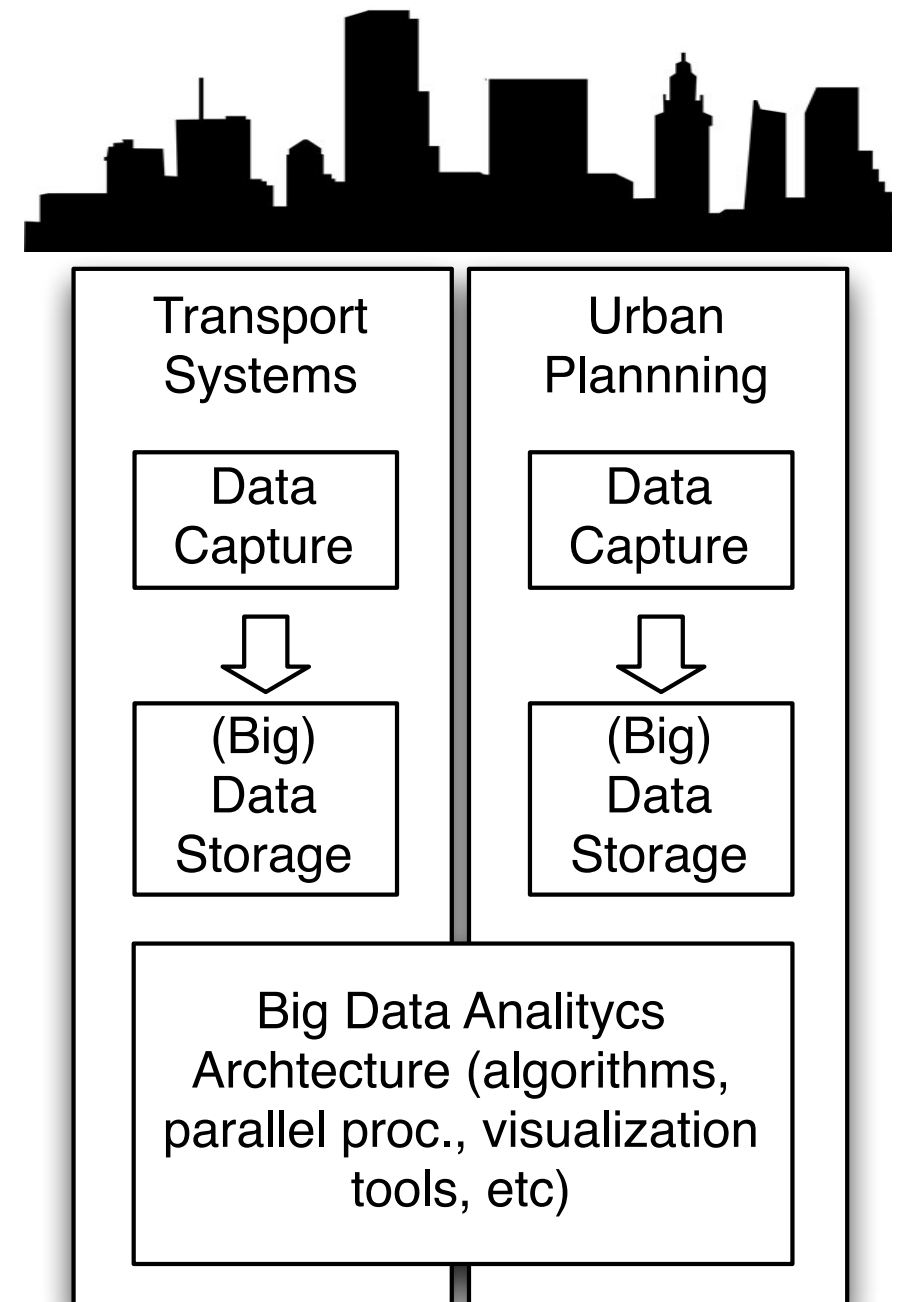
- CLIENTES DE BAJO RIESGO**
- CLIENTES DE ALTO RIESGO**



3. PATRONES EN DATOS DE TELEFONÍA CELULAR

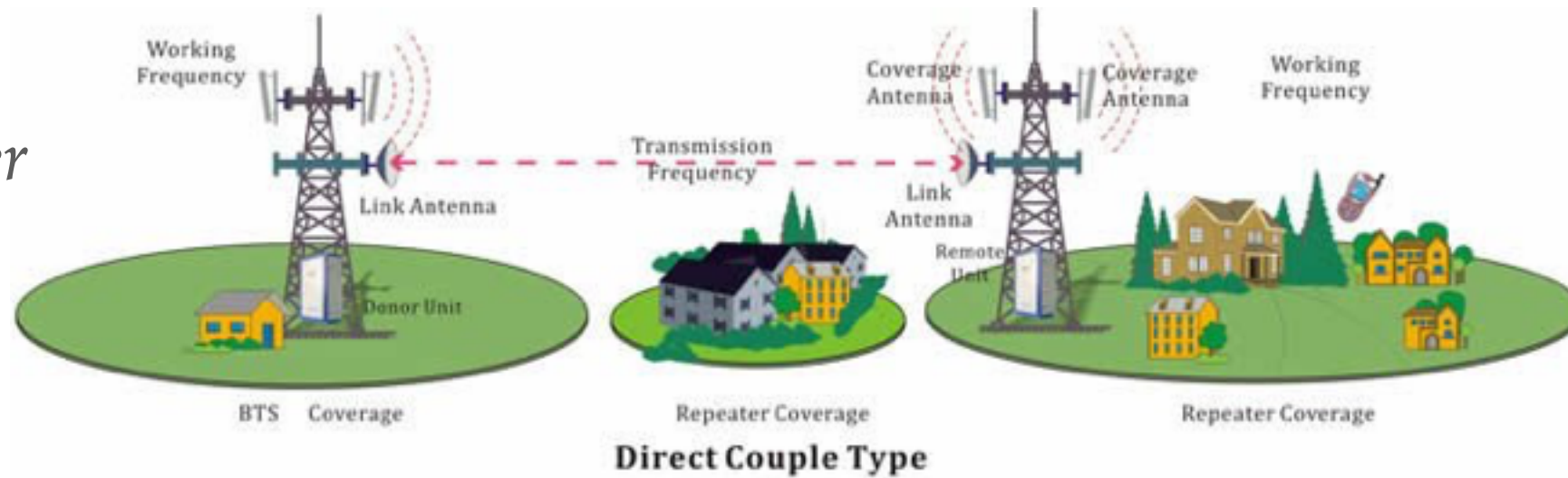
A) TRANSPORTATION SYSTEMS & URBAN PLANNING

- To develop any solution in this area we need to **track how people moves** around a city.
- We have several data sets which allow to obtain the geolocation of a person.
 - Twitter with geotagging
 - Public transportation buses GPS
 - Cell phone network antennas logs (location by triangulation)

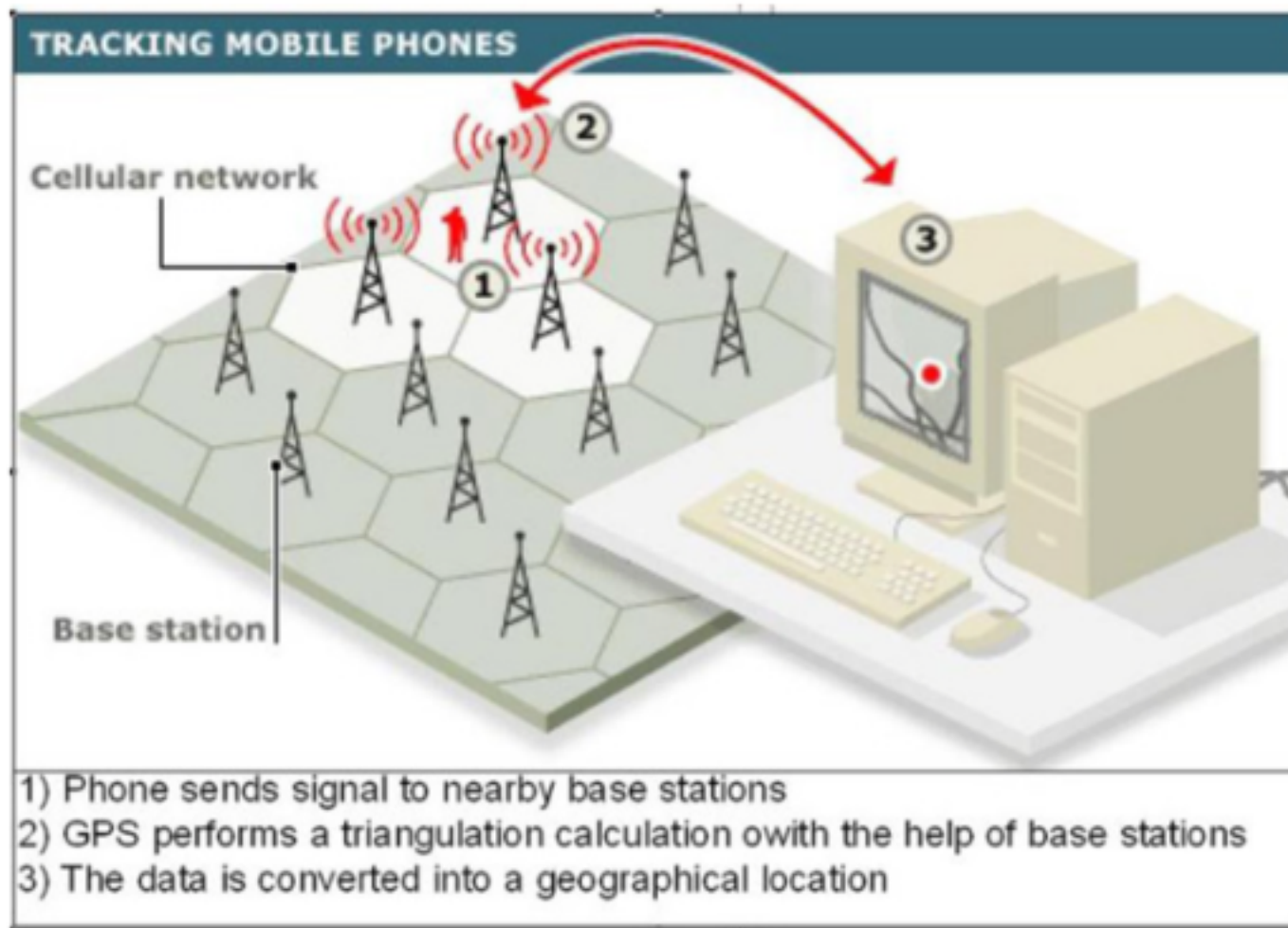


COMO FUNCIONA LA RED CELULAR

*Una torre
puede tener
varias
antenas*



CALL DETAIL RECORD – CDR



Cuando una persona llama, se encuentra la antena más cercana con capacidad libre para atender el llamado.

CDR

Se registra la antena (pos x, y), hora, id del dispositivo, #Tel. Origen, #Tel. de Destino, duración del llamado, tipo (voz, sms, etc.) etc.

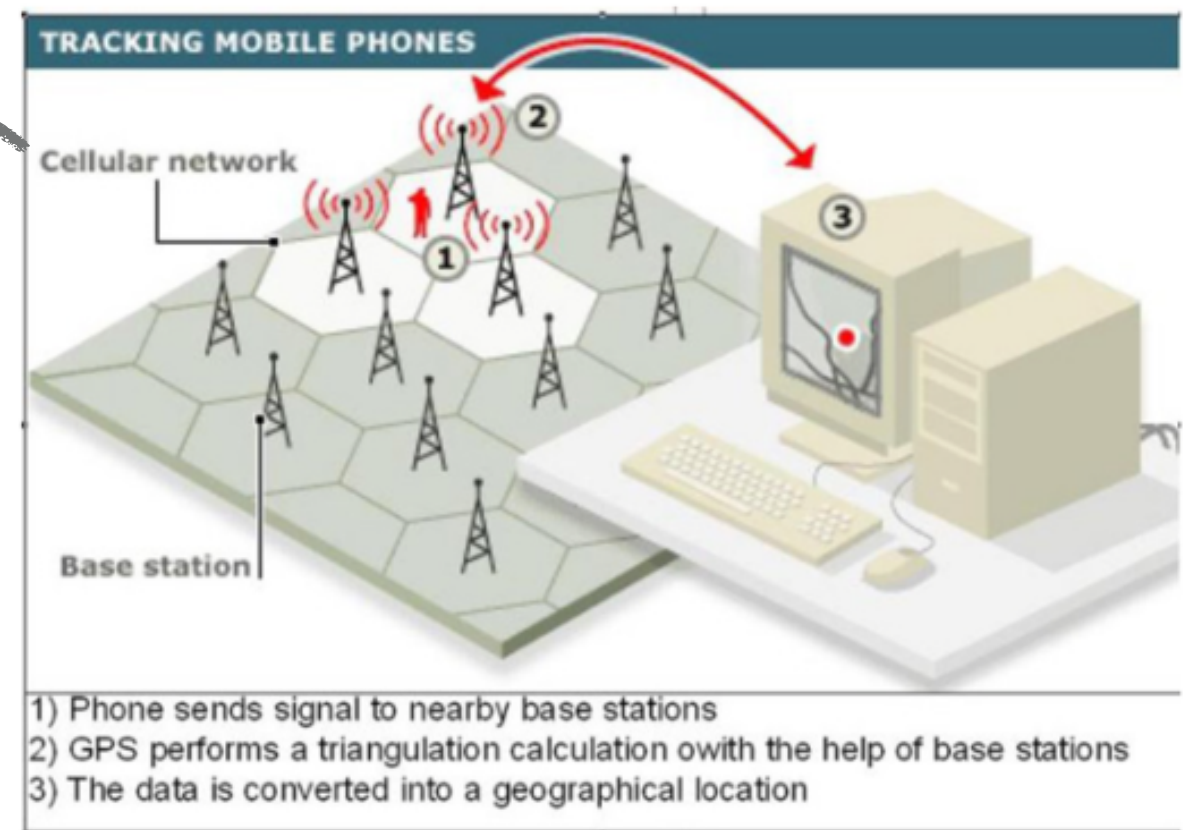
We have data from all cell phone antennas in Chile 77 days

We selected only antennas from Santiago, Chile

We processed 880.000.000 millions of Voice CDRs

from near 3.000.000 customers

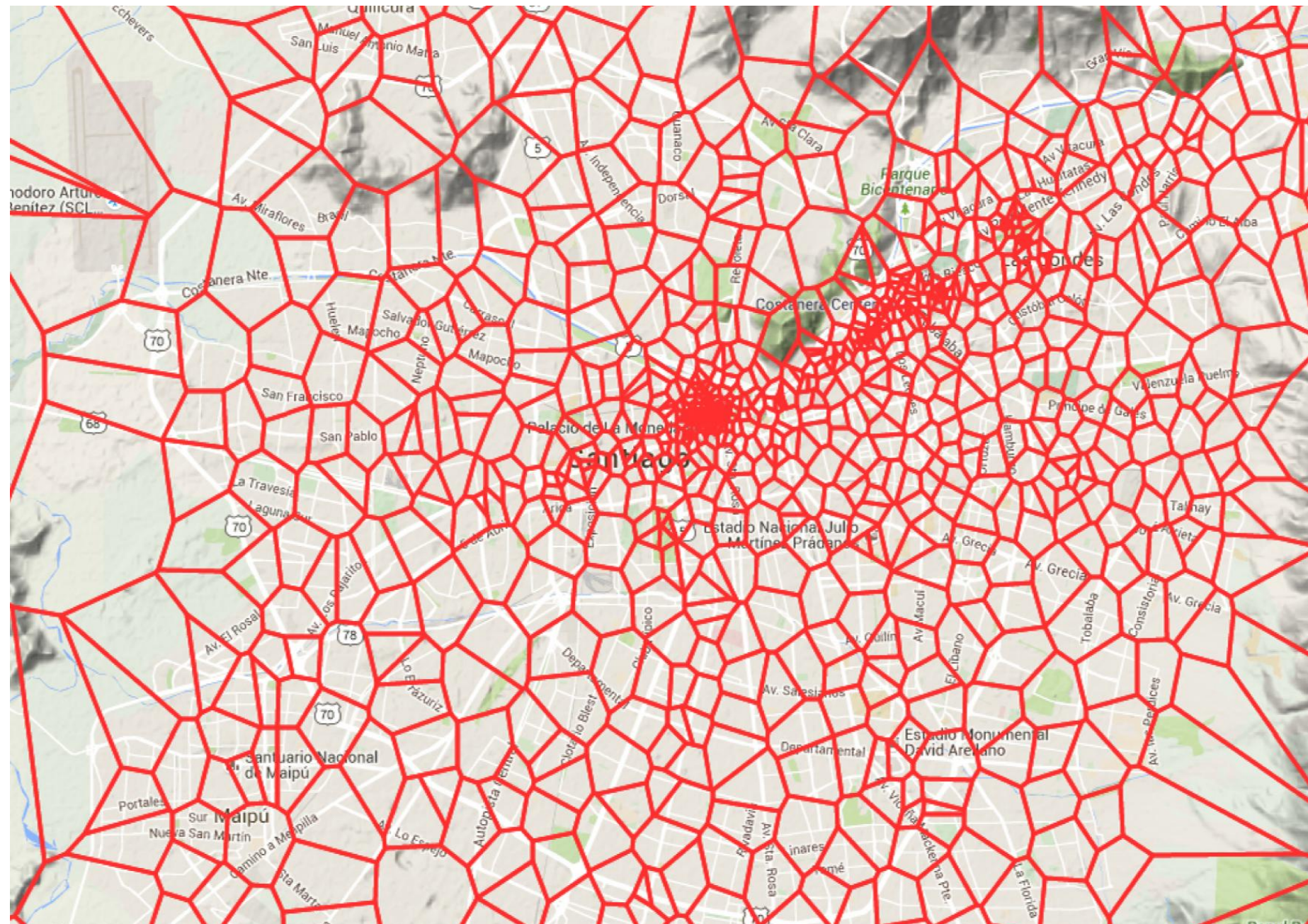
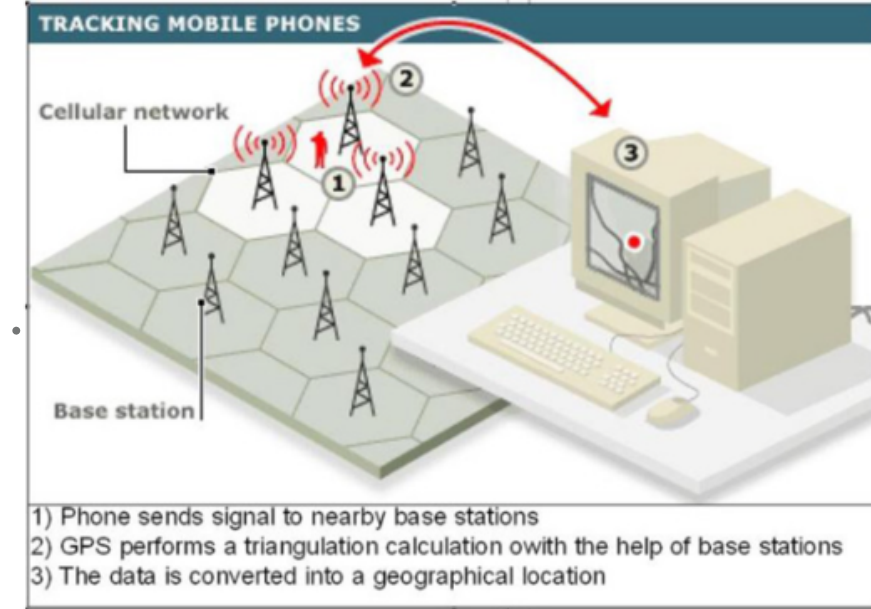
Experiments from 3 Teras of data



Today, we have access to a platform of almost real time CDR - 15 min delay.

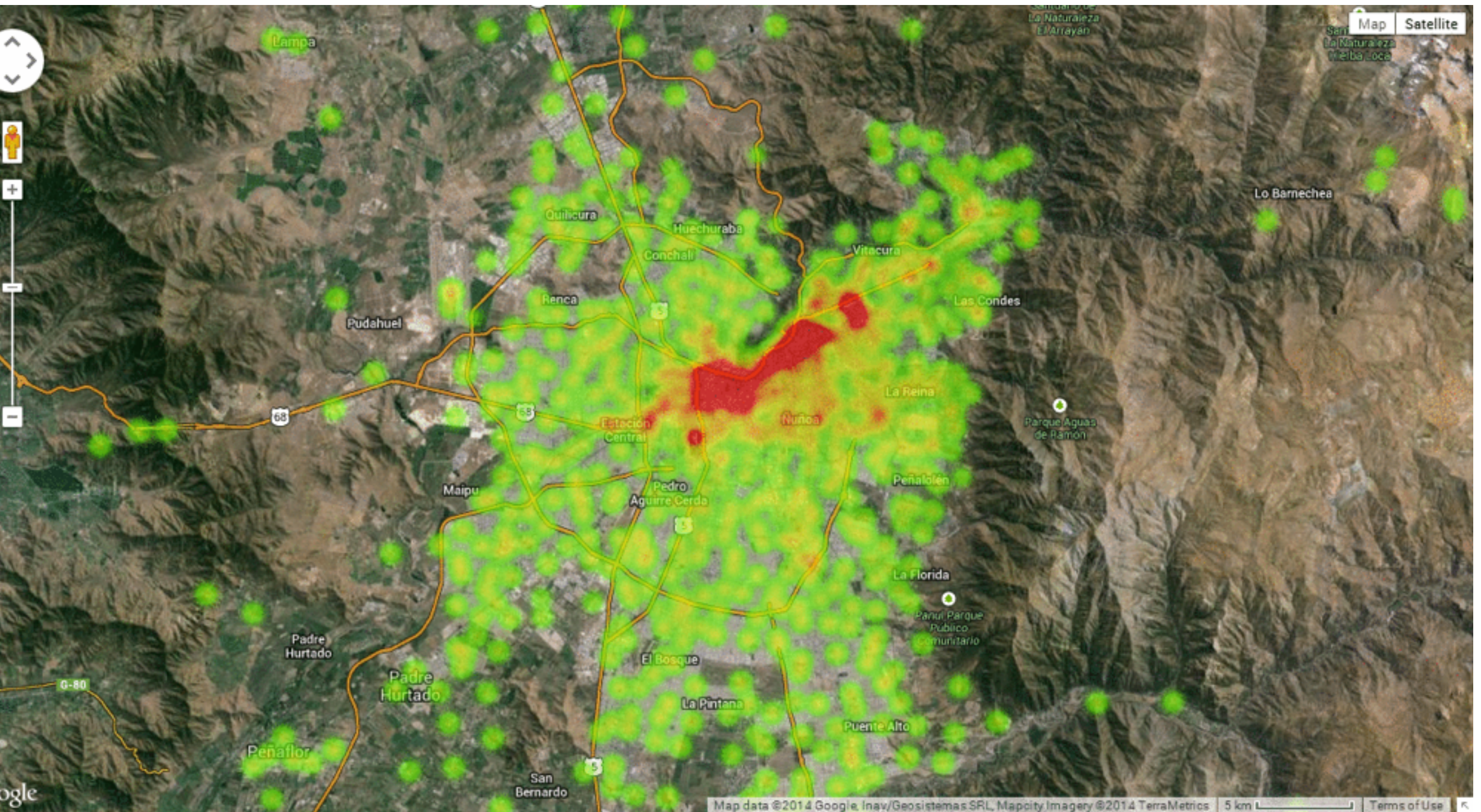
CDRs from Internet Traffic. More massive and allow to have a better tracking.

ZONAS DE VORONOI



VISUALIZACIONES: MAPAS DE CALOR

.....



SE PUEDE ESTUDIAR SECTORES MUY EN DETALLE



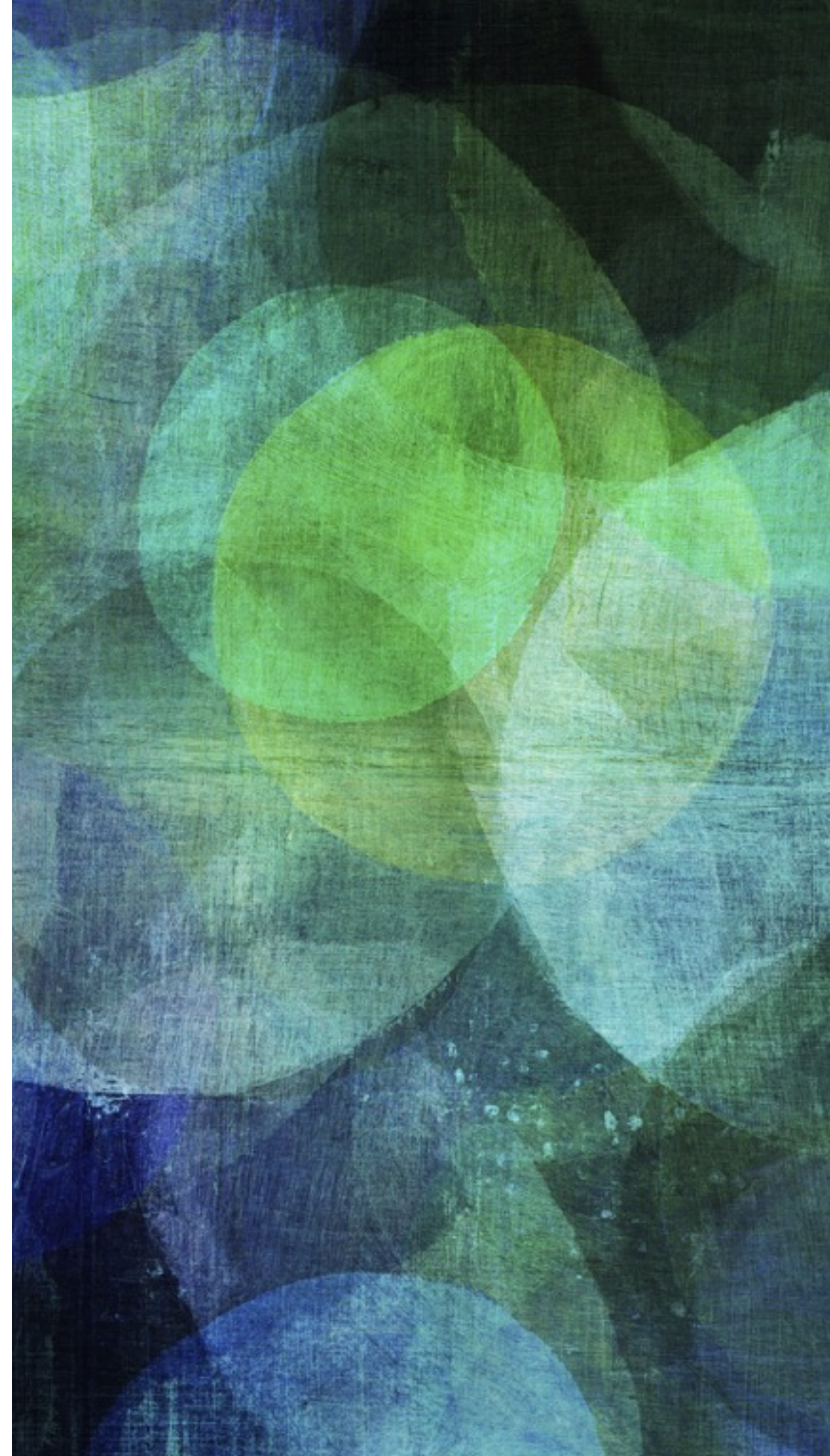
Hora: 10am

Estación
Central

Centro de
Santiago

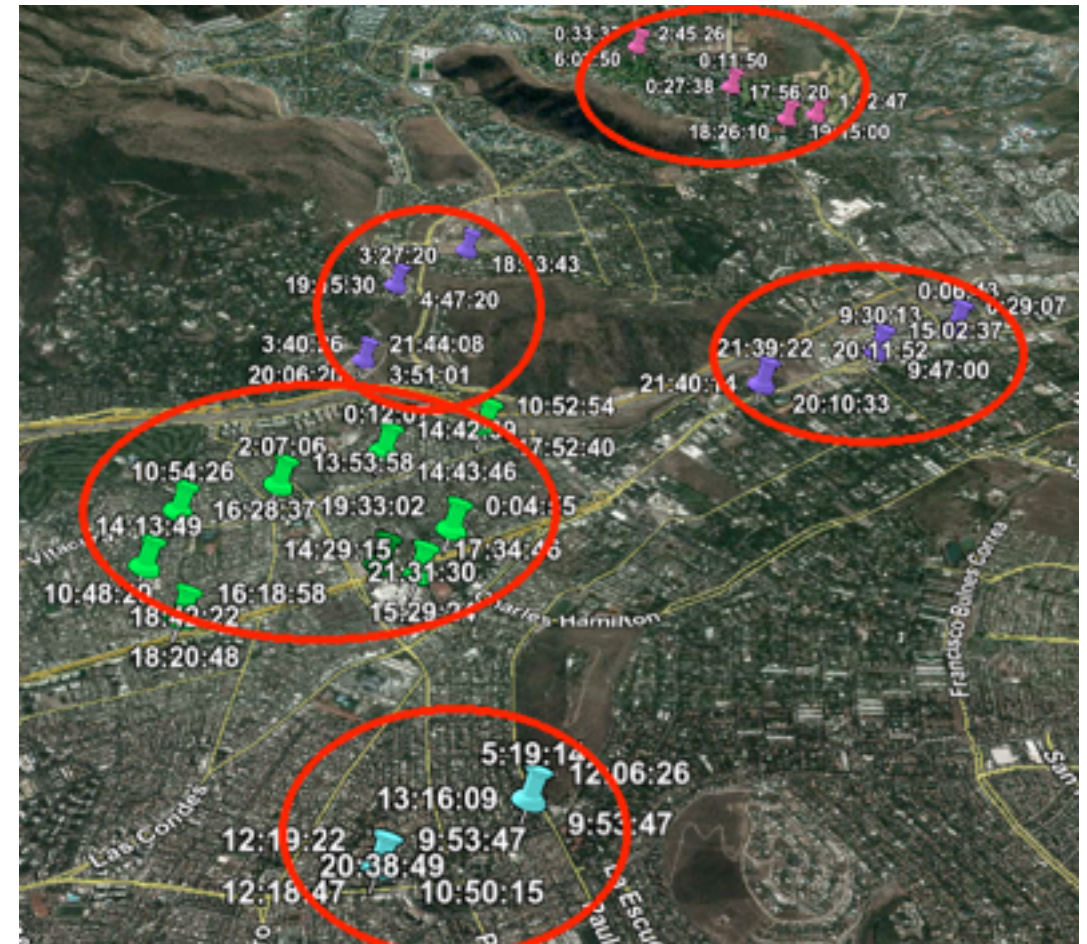
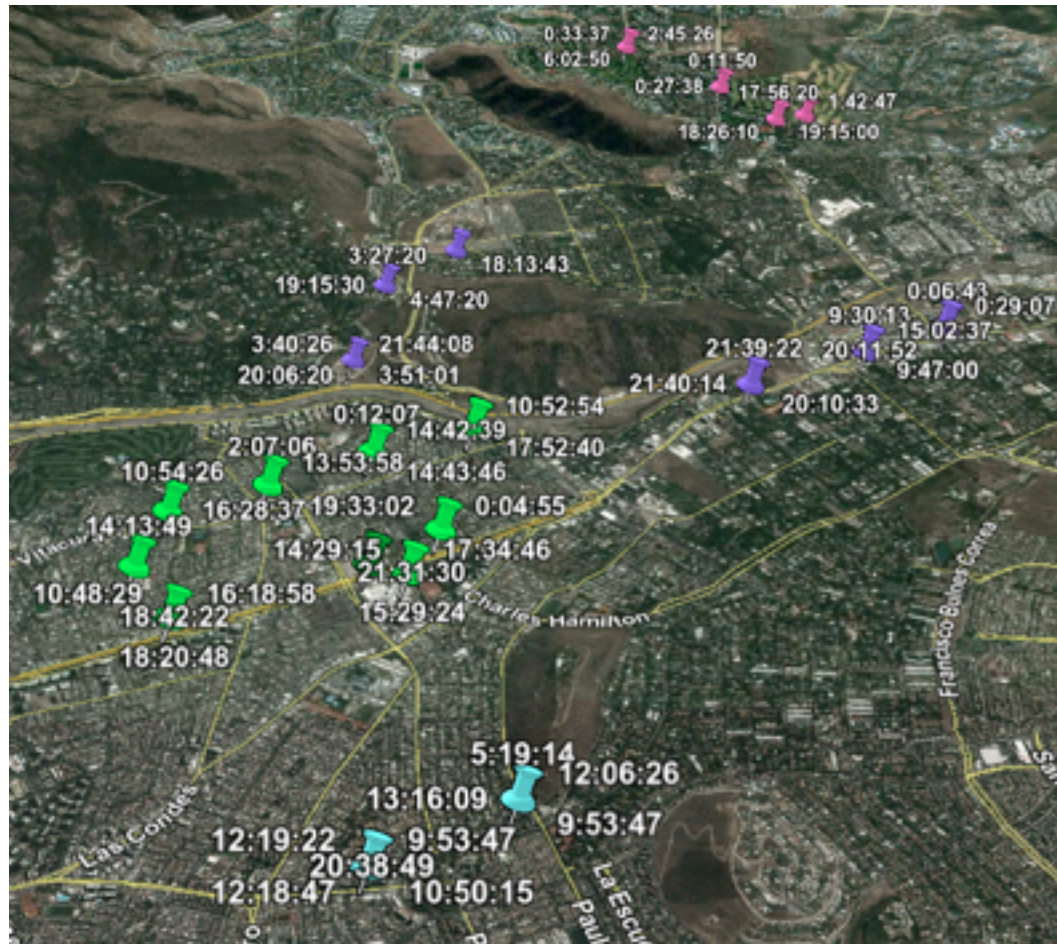
3.1 PATRONES SIMPLES

.....
(sólo contando)



¿QUE SE PUEDE OBTENER DE ESTOS DATOS?

.....



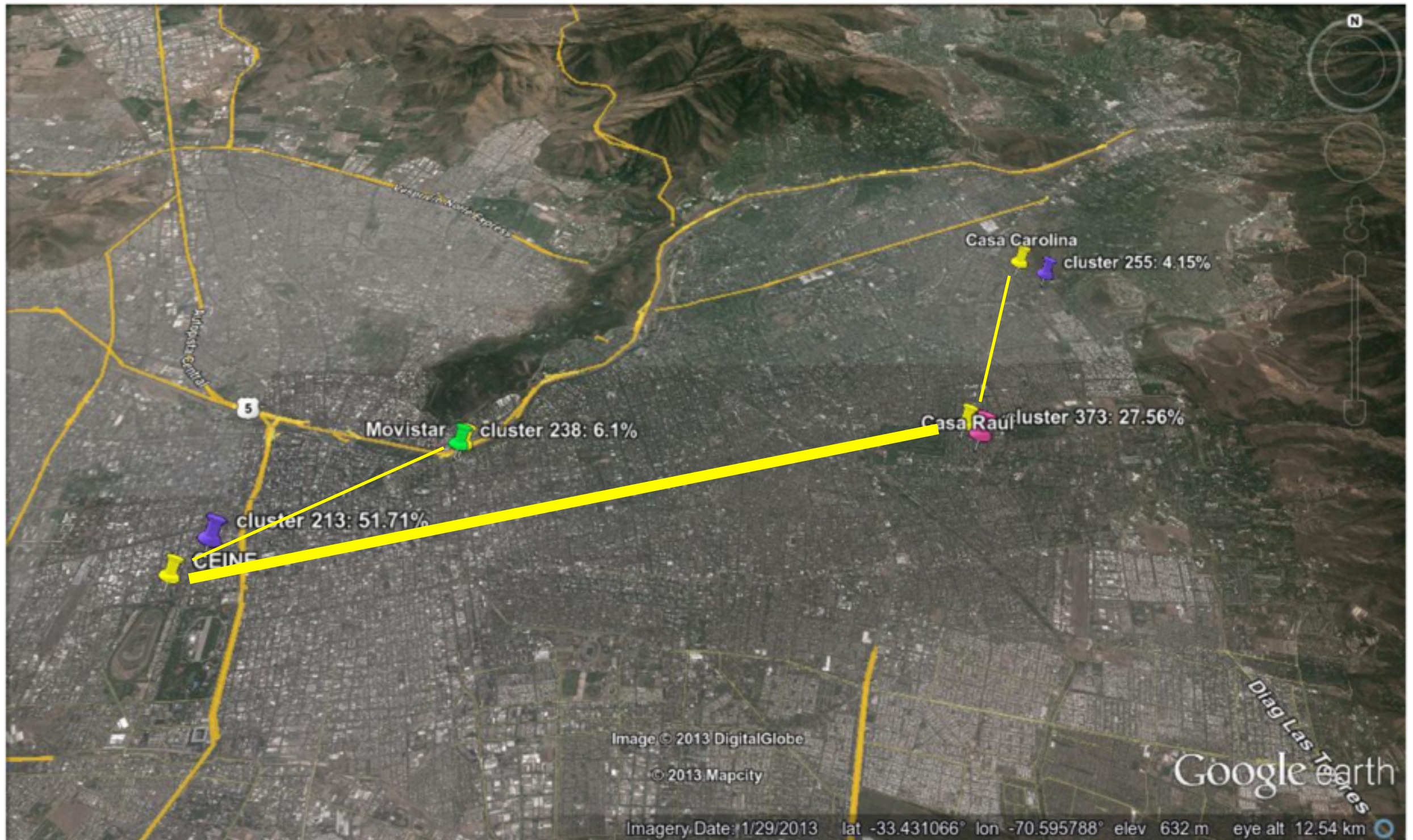
Lugares Frecuentes

Lugares
Esporádicos

Comportamientos
Masivos

LUGARES FRECUENTES

.....

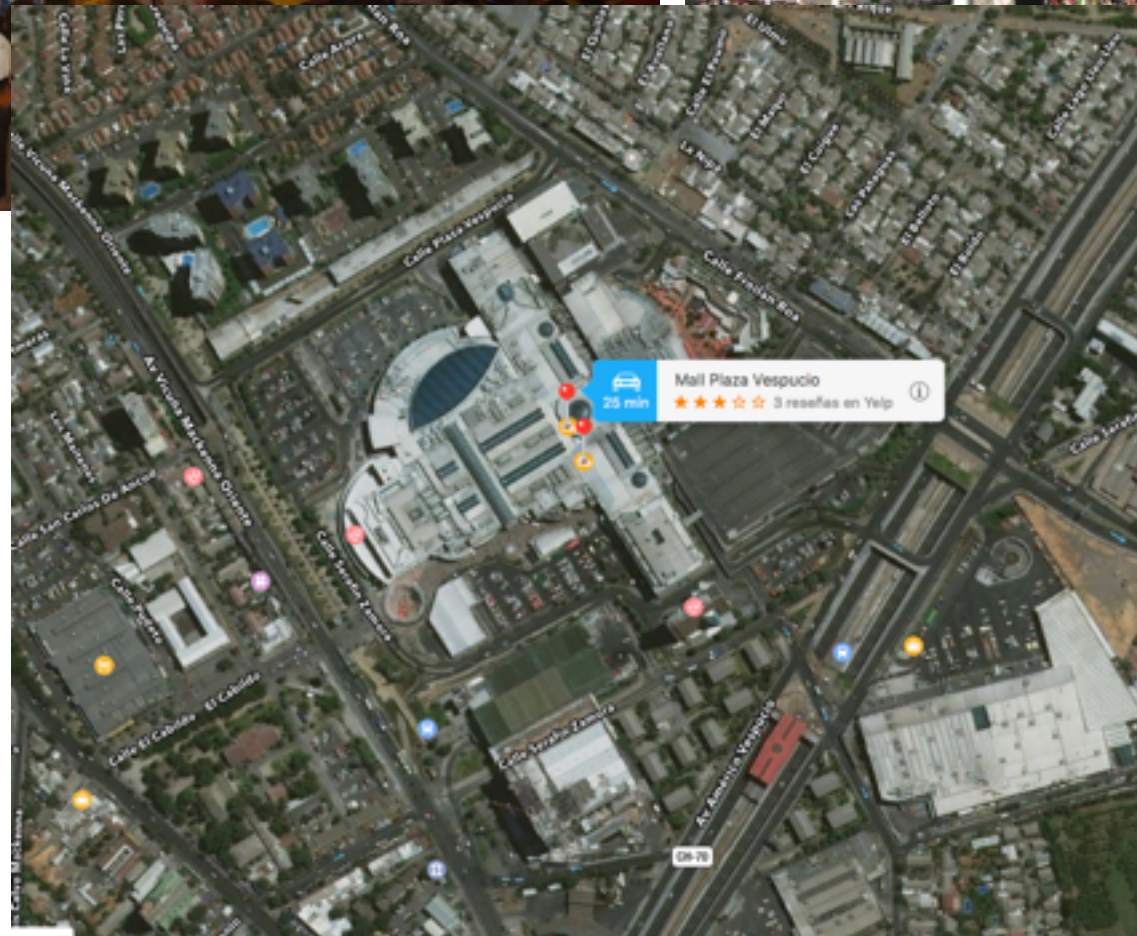


LUGARES FRECUENTES



LUGARES ESPORÁDICOS (MASIVOS)

.....

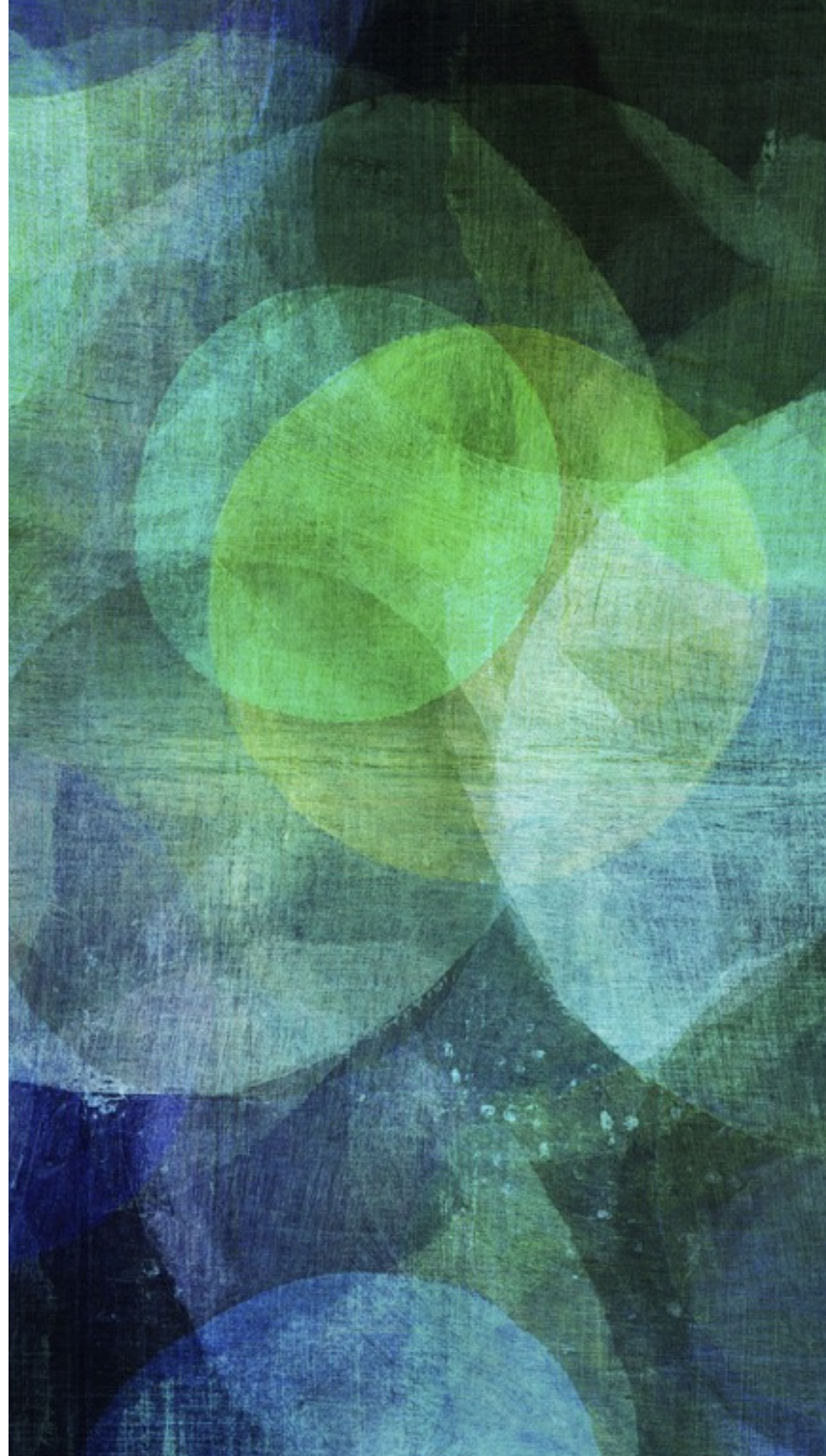


*a. Puntos de
Acumulación*

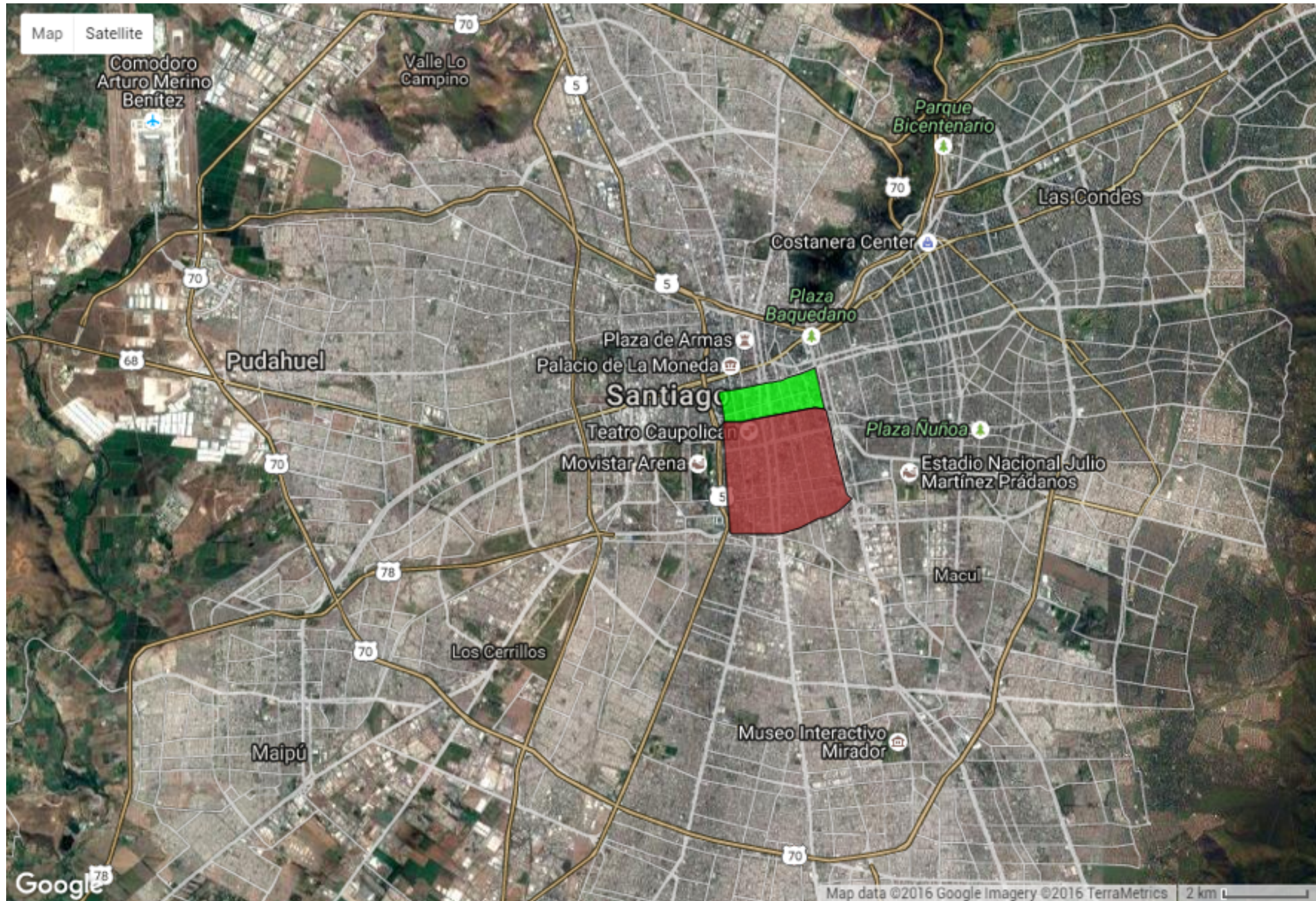
*b. No hay regularidad
Horaria*

*c. Ni regularidad de
participantes*

GENTRIFICACION Y SEGREGACION



GENTRIFICACIÓN Y SEGREGACIÓN



ANÁLISIS DESCRIPTIVO

Base de Datos	CliTestSum
# Registros	40.855.988
# Total Llamadas	160.766.091
# Teléfonos	349.042



ANÁLISIS DESCRIPTIVO

Al cruzar con datos demográficos de la compañía se obtiene un sub conjunto de números para el estudio.

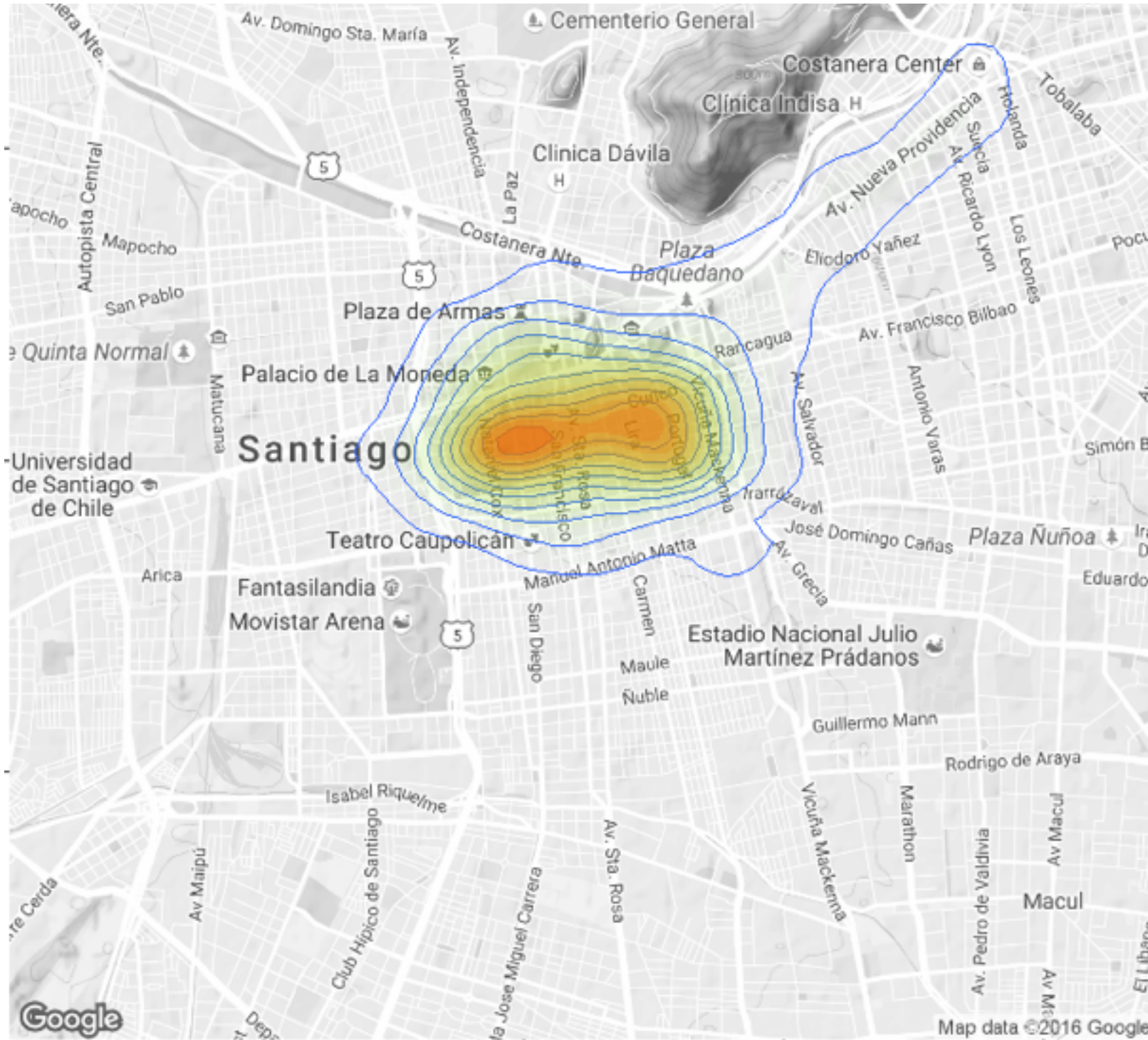
Post aplicar la heurística para determinar hogares, se encuentra que 4285 personas* viven en las zonas de interés.

Santa Isabel	1.704
Bogotá	2.581

Post aplicar la heurística para determinar oficinas, se encuentra que 5923 personas* trabajan en las zonas de interés.

Santa Isabel	2.034
Bogotá	3.889

HOGARES Y OFICINAS – SANTA ISABEL



Densidad de personas que trabajan en Santa Isabel y que viven ahí



HOGARES Y OFICINA – BOGOTA



Densidad de personas que trabajan en Bogotá y que viven ahí



¿DONDE TRABAJA LA GENTE DE ESTOS SECTORES?

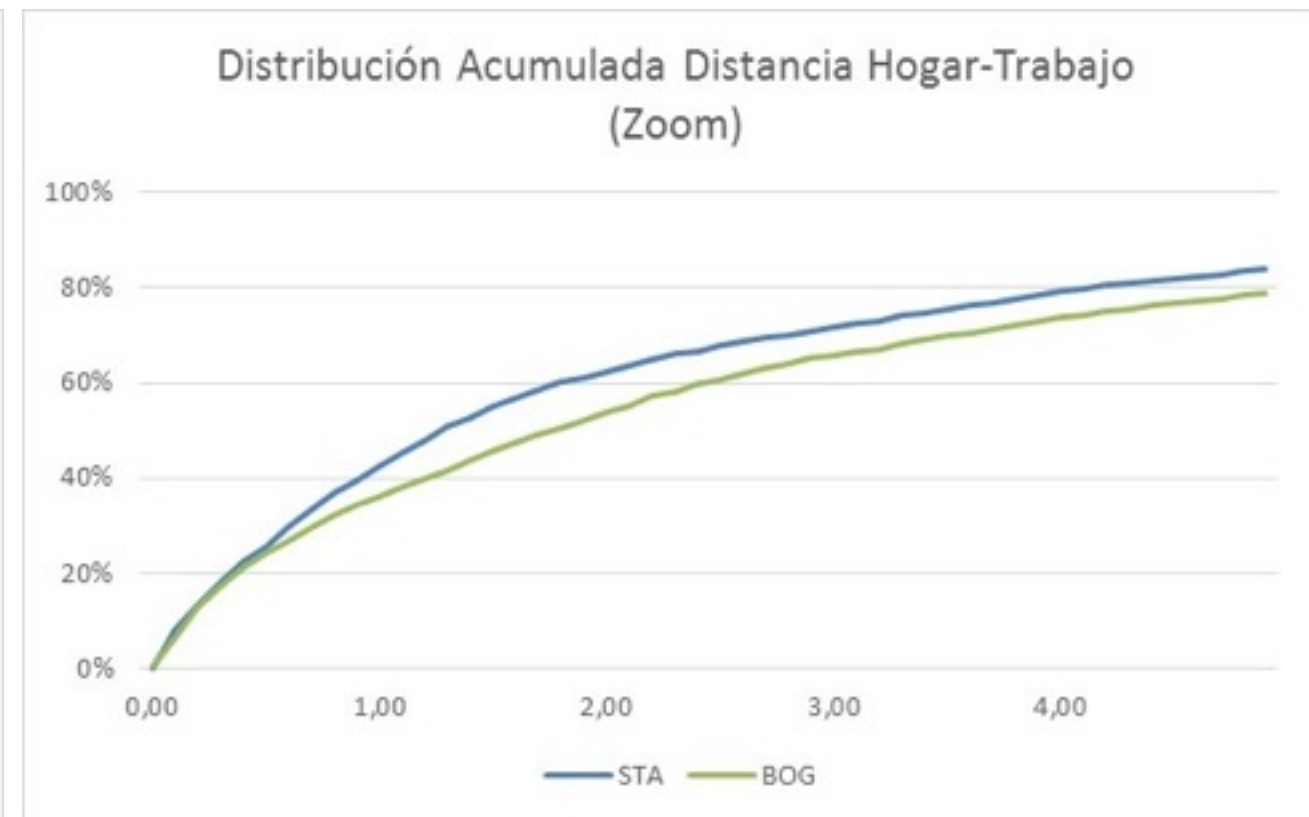
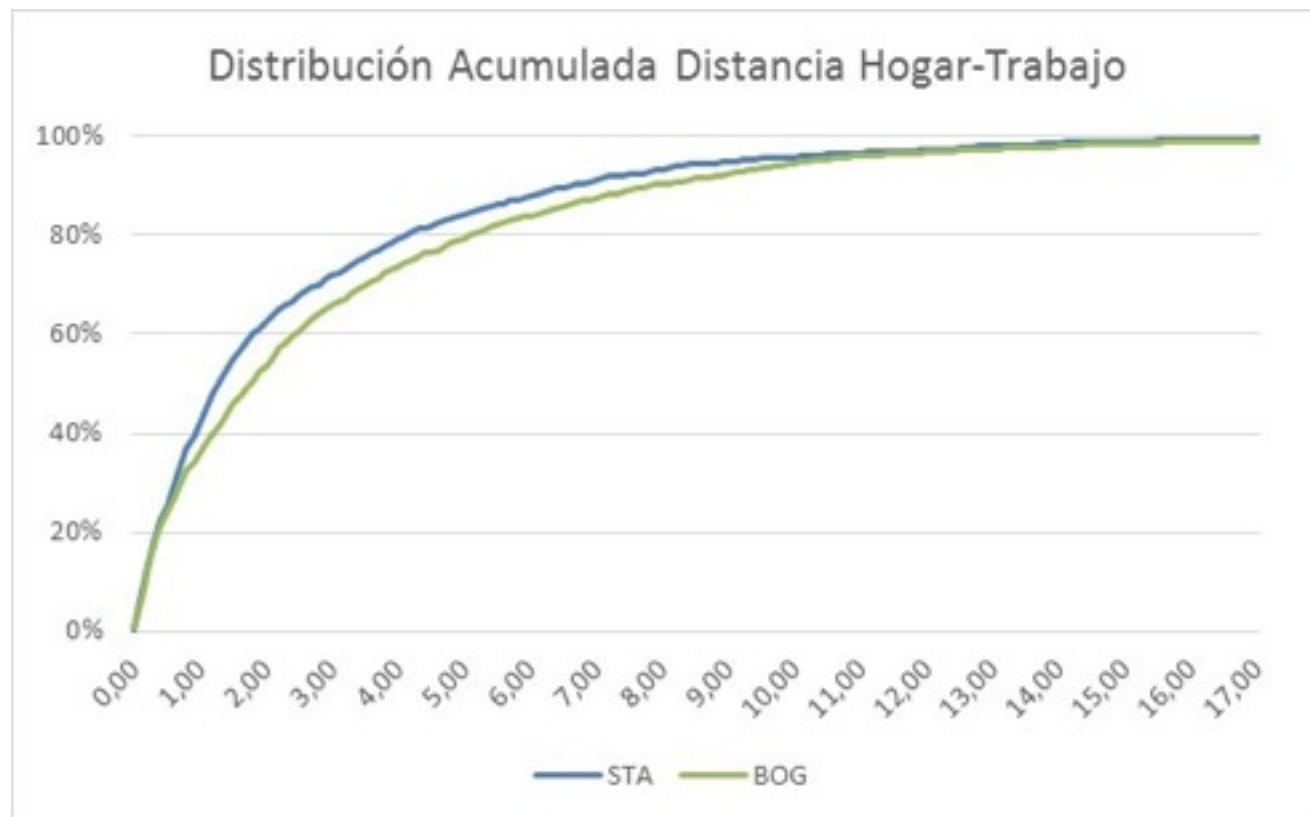
Se presentan las comunas donde se encuentran las oficinas de aquellas personas cuyo hogar se encuentra en las zonas de interés

Comunas	Santa Isabel	%Oficinas	%ACUM
Santiago	993	58,27%	58,27%
Providencia	256	15,02%	73,30%
Las Condes	88	5,16%	78,46%
Ñuñoa	77	4,52%	82,98%
Recoleta	72	4,23%	87,21%
Macul	20	1,17%	88,38%
Independencia	19	1,12%	89,50%
Estacion Central	16	0,94%	90,43%
San Joaquin	15	0,88%	91,31%
Quinta Normal	13	0,76%	92,08%
Vitacura	10	0,59%	92,66%
Huechuraba	9	0,53%	93,19%
La Cisterna	8	0,47%	93,66%
Pedro Aguirre Cerda	8	0,47%	94,13%
Quilicura	8	0,47%	94,60%

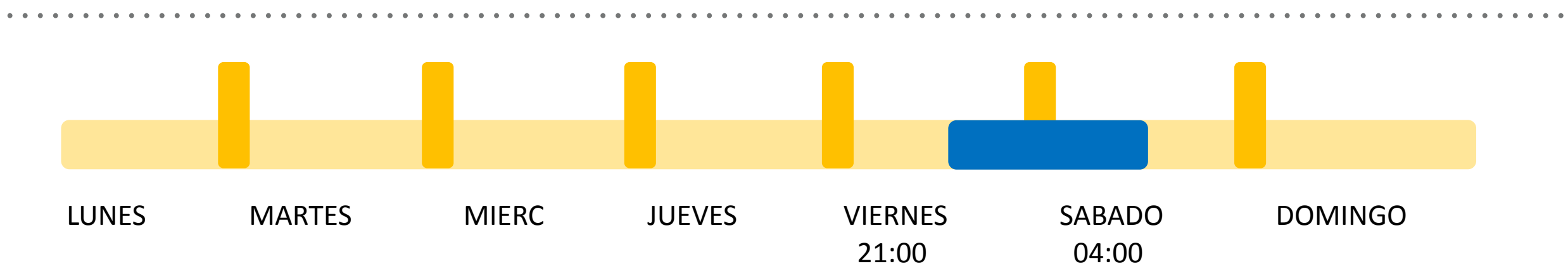
Comunas	BOGOTA	%Oficinas	%ACUM
Santiago	1411	54,7%	54,7%
Providencia	213	8,3%	62,9%
Ñuñoa	193	7,5%	70,4%
Las Condes	94	3,6%	74,0%
San Joaquin	86	3,3%	77,4%
San Miguel	79	3,1%	80,4%
Recoleta	66	2,6%	83,0%
Macul	53	2,1%	85,0%
Estacion Central	43	1,7%	86,7%
Cerillos	26	1,0%	87,7%
La Florida	26	1,0%	88,7%
Pedro Aguirre Cerda	22	0,9%	89,6%
Vitacura	22	0,9%	90,4%
Independencia	20	0,8%	91,2%
La Reina	17	0,7%	91,9%

¿CUANTO SE DESPLAZAN AL TRABAJO?

	Dist. (Kms)	N
Sta. Isabel	2,5839	1.704
Bogotá	3,0676	2.581



CASO: CARRETE = FIESTA



Análisis Ponderando por el numero de llamados en cada antena

	Casa	Oficina
Sta. Isabel	2,42	3.95
Bogotá	3,48	5.278

Análisis de alcance

Casa	Oficina
3,57	4,92
4,61	6.29



Se consideran todos los llamados de personas que viven en zonas de interés

¿DONDE VAN DE CARRETE?

Santa Isabel					
NOMBRE	N	Frec	%N	%FREC	%FREC_ACUM
Santiago	2923	10146	55,20%	65,82%	65,82%
Providencia	519	1141	9,80%	7,40%	73,22%
Ñuñoa	348	1067	6,57%	6,92%	80,14%
Recoleta	245	647	4,63%	4,20%	84,34%
Estacion Central	168	378	3,17%	2,45%	86,79%
Las Condes	176	330	3,32%	2,14%	88,93%
La Florida	86	167	1,62%	1,08%	90,02%
Maipu	72	162	1,36%	1,05%	91,07%
Macul	64	120	1,21%	0,78%	91,85%
Vitacura	47	104	0,89%	0,67%	92,52%
Pudahuel	49	99	0,93%	0,64%	93,16%
San Miguel	47	96	0,89%	0,62%	93,79%
La Reina	46	87	0,87%	0,56%	94,35%
Peñalolen	38	78	0,72%	0,51%	94,86%
Independencia	35	66	0,66%	0,43%	95,28%
Quinta Normal	38	57	0,72%	0,37%	95,65%

BOGOTA					
NOMBRE	N	Frec	%N	%FREC	%FREC_ACUM
Santiago	3213	11809	44,59%	54,03%	54,03%
Ñuñoa	582	1699	8,08%	7,77%	61,80%
San Miguel	330	1045	4,58%	4,78%	66,58%
Providencia	396	890	5,50%	4,07%	70,65%
San Joaquin	159	591	2,21%	2,70%	73,36%
Recoleta	234	576	3,25%	2,64%	75,99%
La Florida	266	558	3,69%	2,55%	78,55%
Macul	203	545	2,82%	2,49%	81,04%
Estacion Central	256	529	3,55%	2,42%	83,46%
Las Condes	163	315	2,26%	1,44%	84,90%
Pedro Aguirre Cerda	102	312	1,42%	1,43%	86,33%
Maipu	131	290	1,82%	1,33%	87,66%
Puente Alto	110	277	1,53%	1,27%	88,92%
La Cisterna	100	249	1,39%	1,14%	90,06%
Peñalolen	82	218	1,14%	1,00%	91,06%
La Reina	63	175	0,87%	0,80%	91,86%

¿DONDE CARRETEAN: SANTA ISABEL?

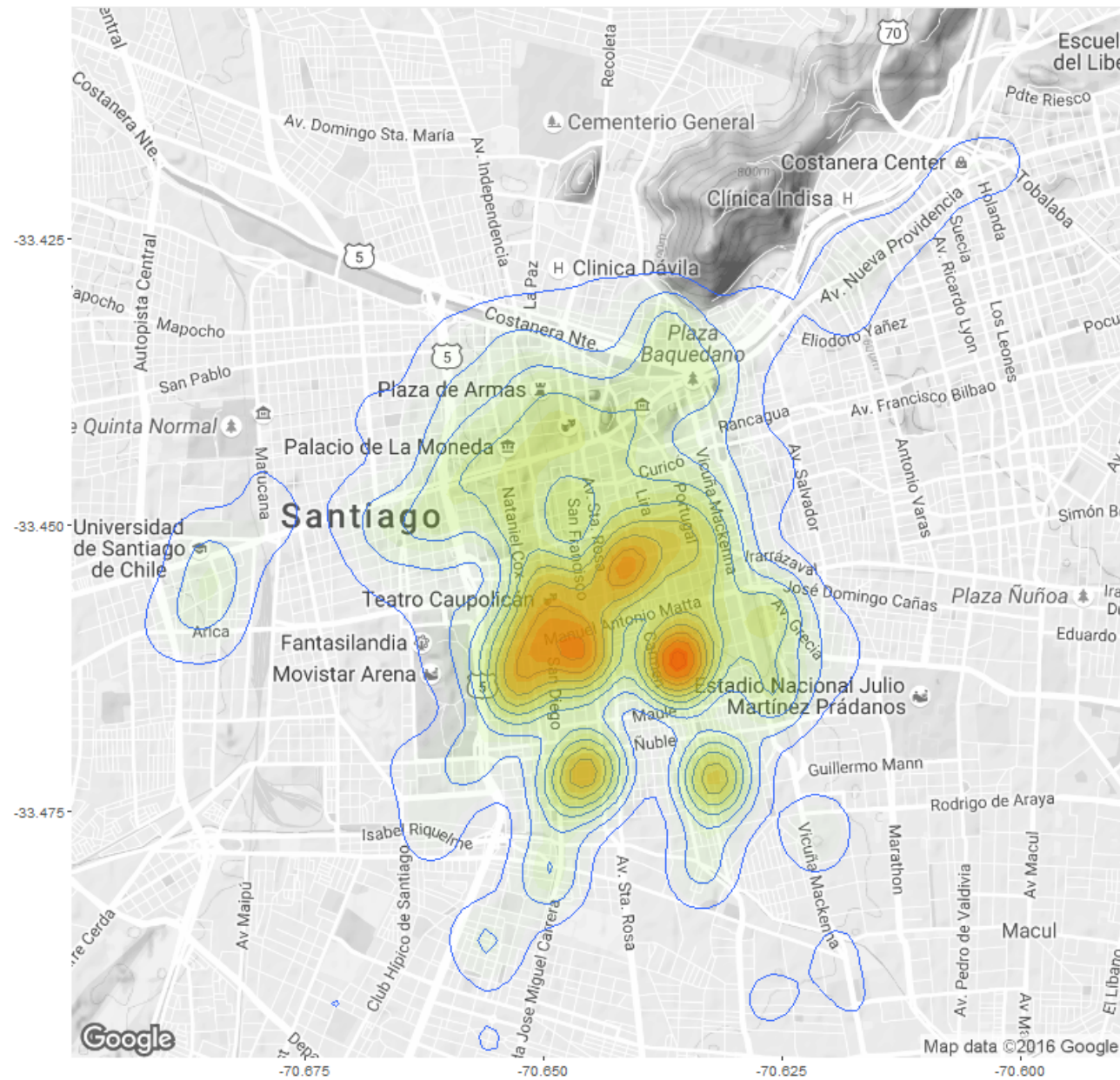


Todo

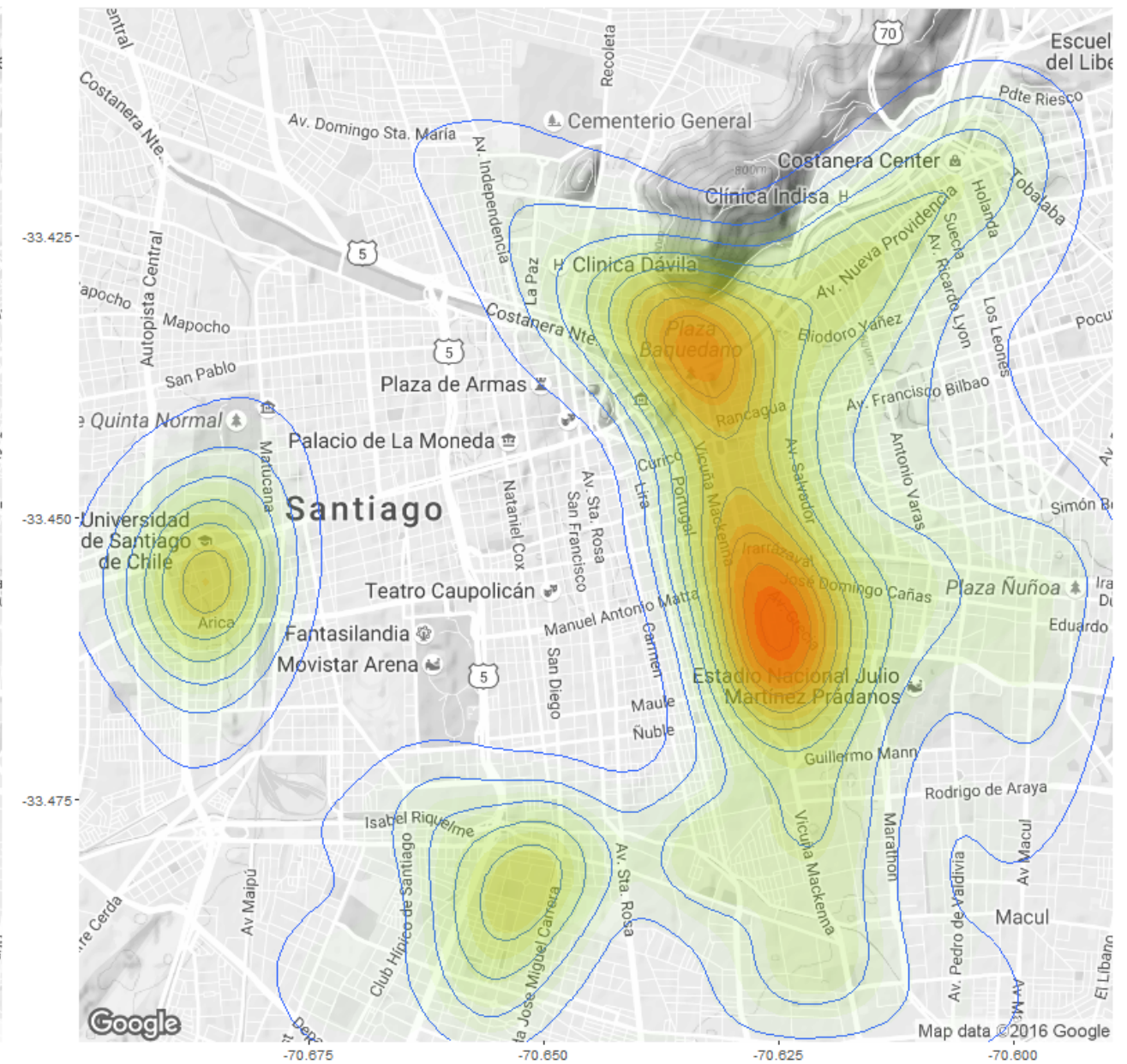


Sin Santiago

¿DONDE CARRETEAN: BOGOTA?



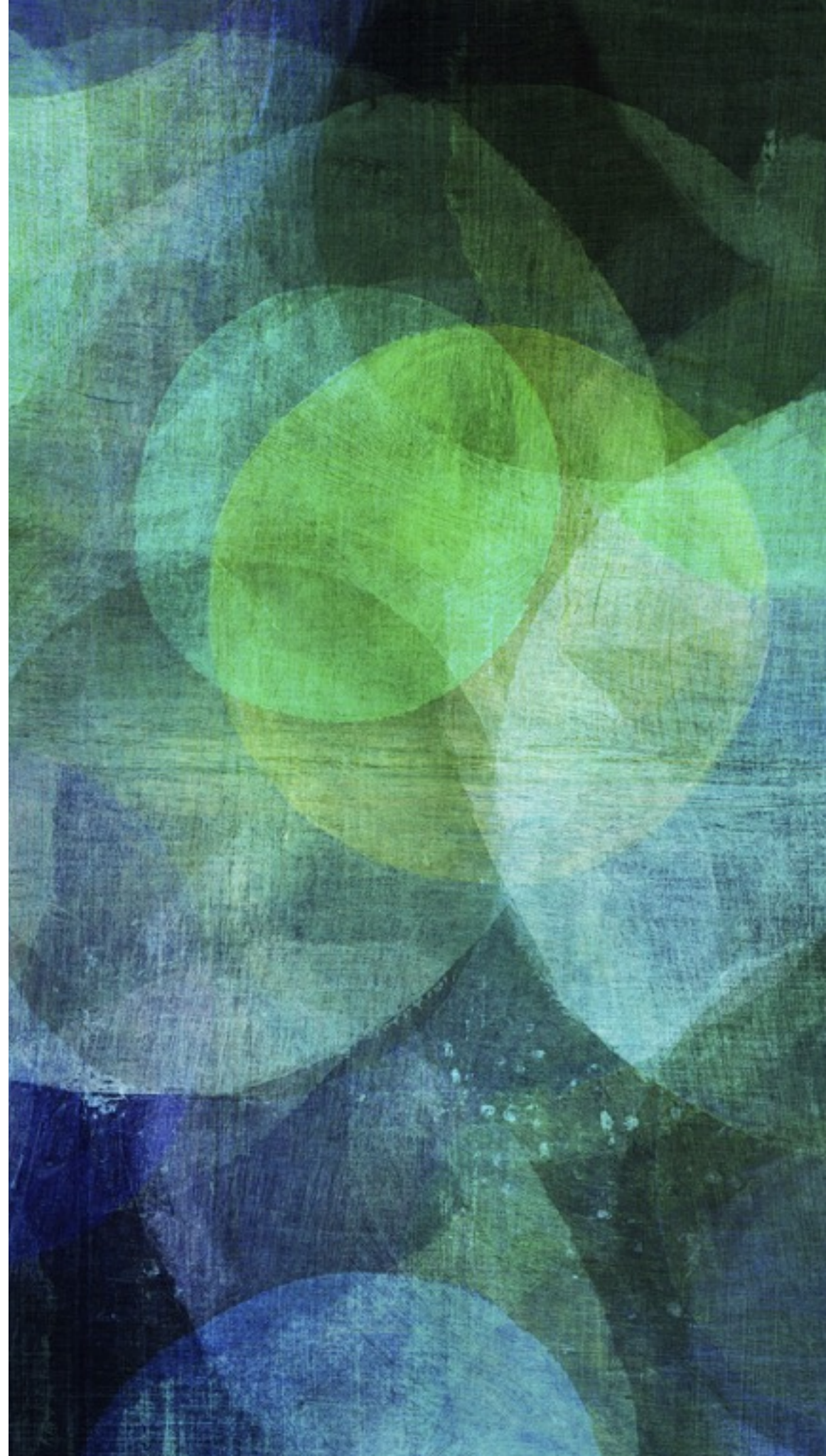
Todo



Sin Santiago

3.2 PATRONES COMPLEJOS

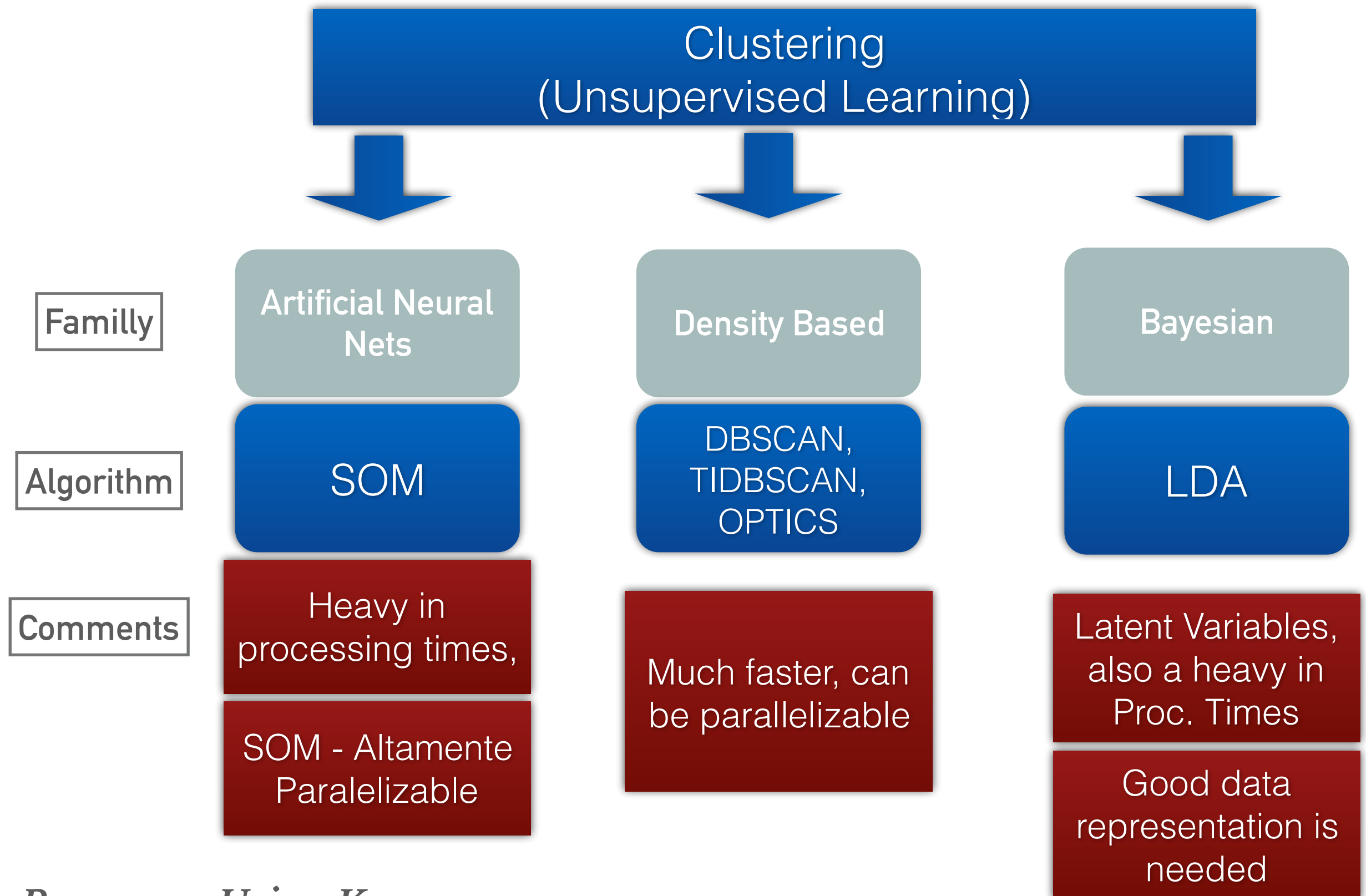
.....
(con algoritmos)



¿PATRONES DE COMPORTAMIENTO MASIVOS?

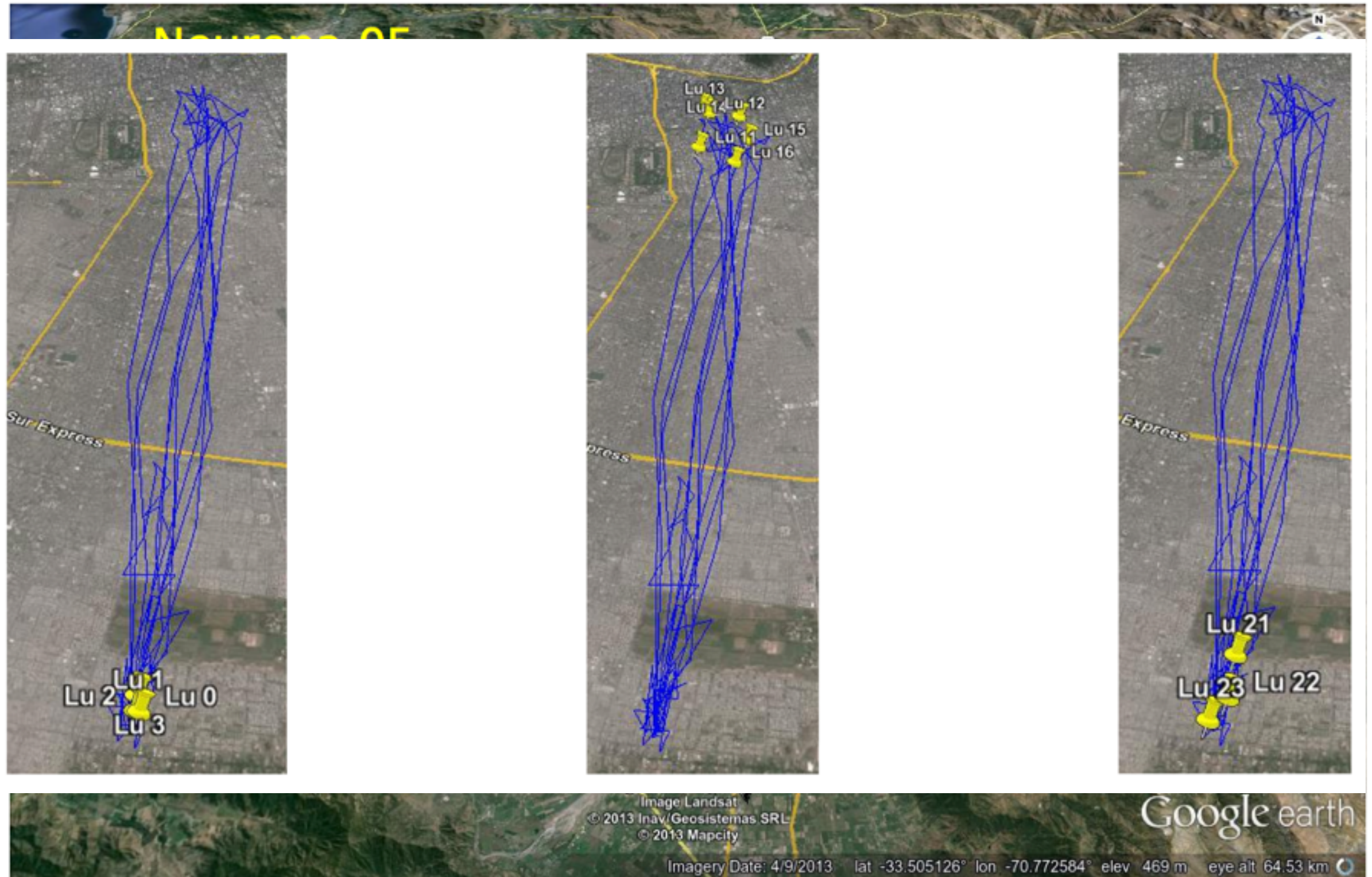
- ¿existirán patrones de uso de suelo masivos - además de los reportados por la literatura? (Oficinas y Casas)
- ¿Cuan generalizables son esos patrones si es que los hay?
- ¿Es posible obtener algunos comportamientos humanos a nivel macro?

SEVERAL TECHNIQUES TO PROCESS MASSIVE DATA



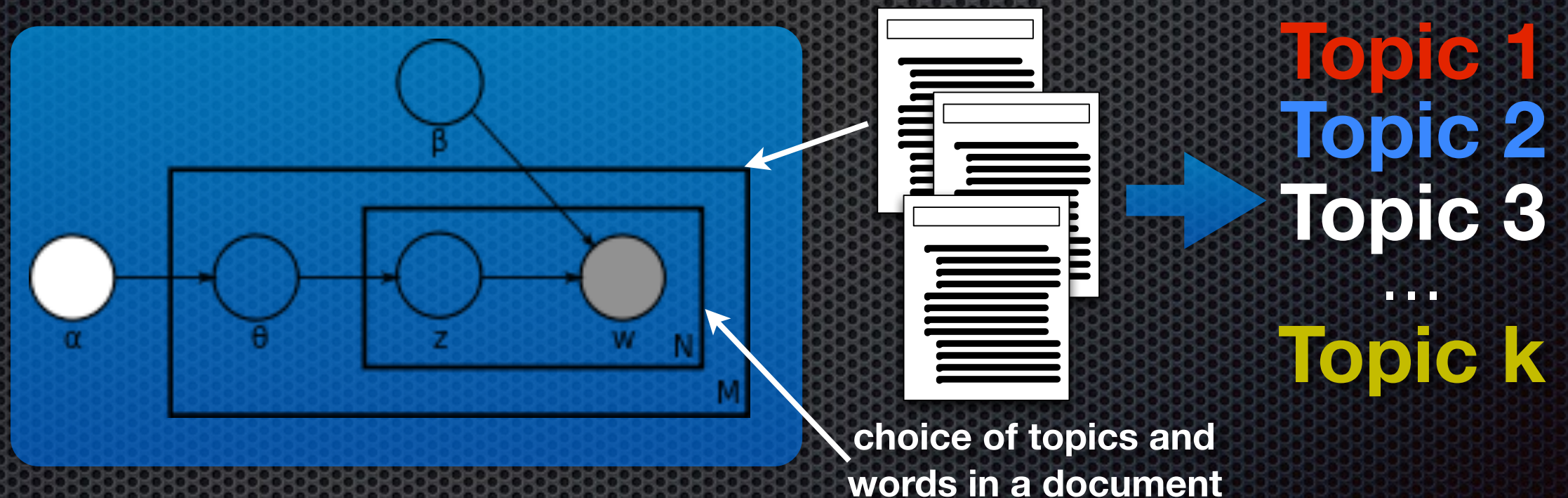
Base case: Using K-means

PATRONES DE MOVIMIENTOS MASIVOS EN LA CIUDAD



Latent Semantics with LDA

- Latent Dirichlet Allocation is a generative Topic Model, which try to mimic the way a document is written.



- The latent Dirichlet allocation allows to find out a set of words in the corpus, that can be called a **topic**.

D. Blei, Latent Dirichlet Allocation, Journal of Machine Learning Research 3 (2003) 993-1022

“Arts” “Budgets” “Children” “Education”

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HATTI

Problem:

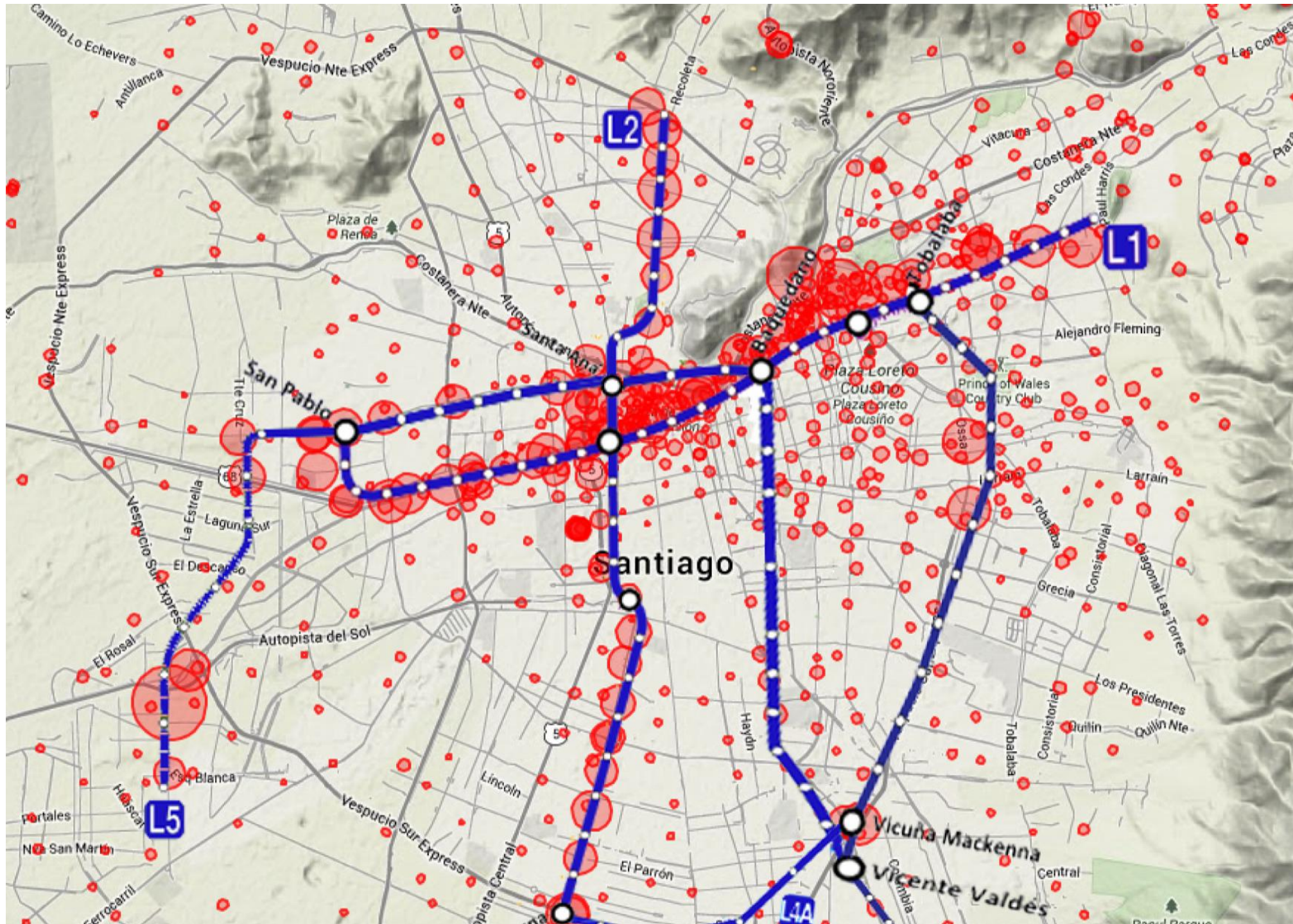
We don't know the Topics,
neither the words that represent
each topic.

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. “Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services,” Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center’s share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

We want the Topics

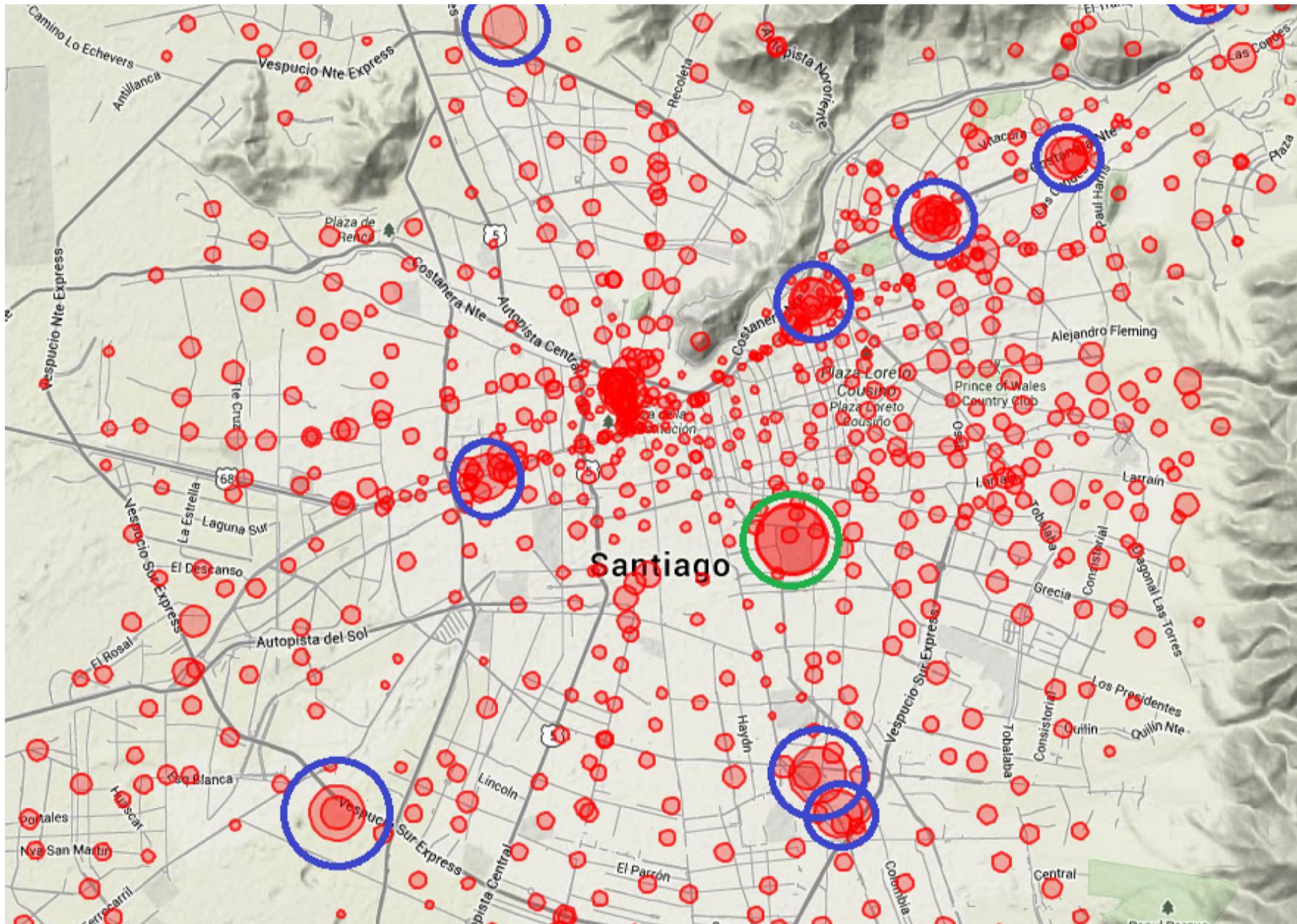
Topic 19 “War”	Topic 15 “Terrorist Attacks (General)”	Topic 33 “Political/Terrorist Trials”	Topic 54 “Religion (Allah)”
government (0.0140)	kill (0.0255)	court (0.0149)	allah (0.0141)
war (0.0113)	police (0.0180)	guantanamo (0.0099)	prophet (0.0163)
military (0.0104)	soldier (0.0168)	trail (0.0080)	messenger (0.0141)
country (0.0091)	attack (0.0139)	prison (0.0073)	peace (0.0118)
united (0.0089)	force (0.0137)	judge (0.0066)	people (0.0113)
security (0.0083)	military (0.0109)	torture (0.0065)	lord (0.0107)
force (0.0079)	official (0.0105)	rights (0.0061)	day (0.0099)
international (0.0059)	security (0.0089)	charges (0.0057)	believer (0.0096)
official (0.0059)	report (0.0086)	government (0.0056)	bless (0.0084)
american (0.0056)	army (0.0075)	arrested (0.0056)	quran (0.0077)

DESCUBRIENDO PATRONES DE USO DE SUELO: HORA PUNTA

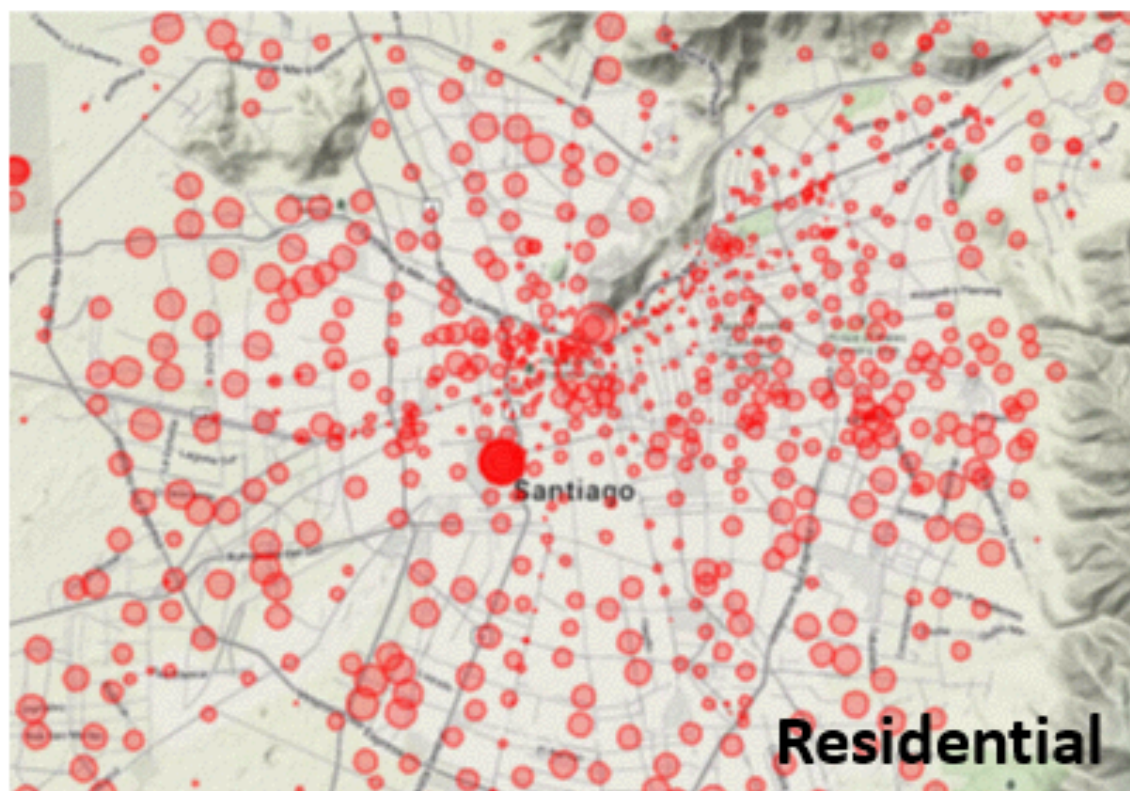
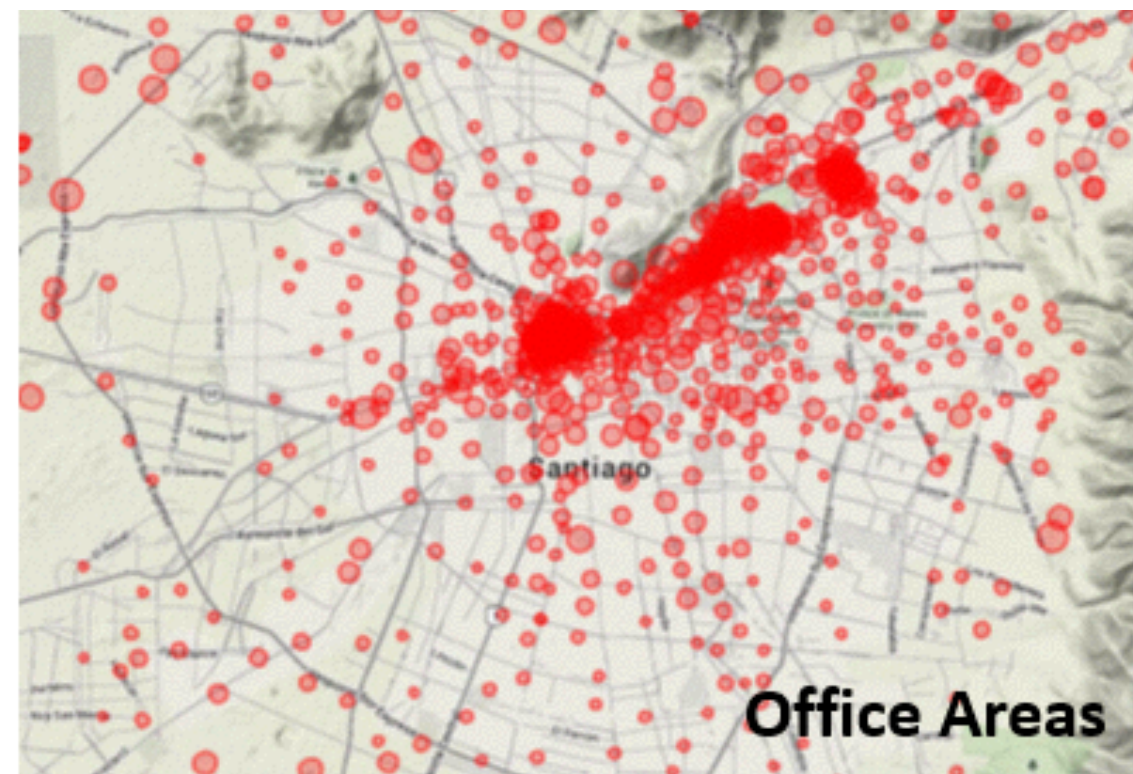


Sebastián A. Ríos, Ricardo Muñoz, **Land Use detection with cell phone data using topic models: Case Santiago, Chile**, Computers, Environment and Urban Systems, Volume 61, Part A, January 2017, Pages 39–48

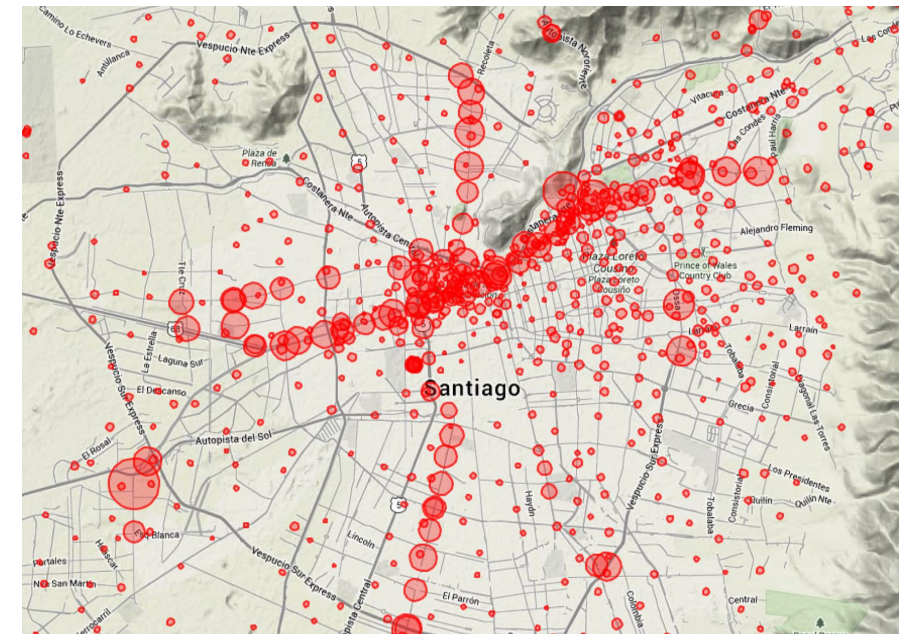
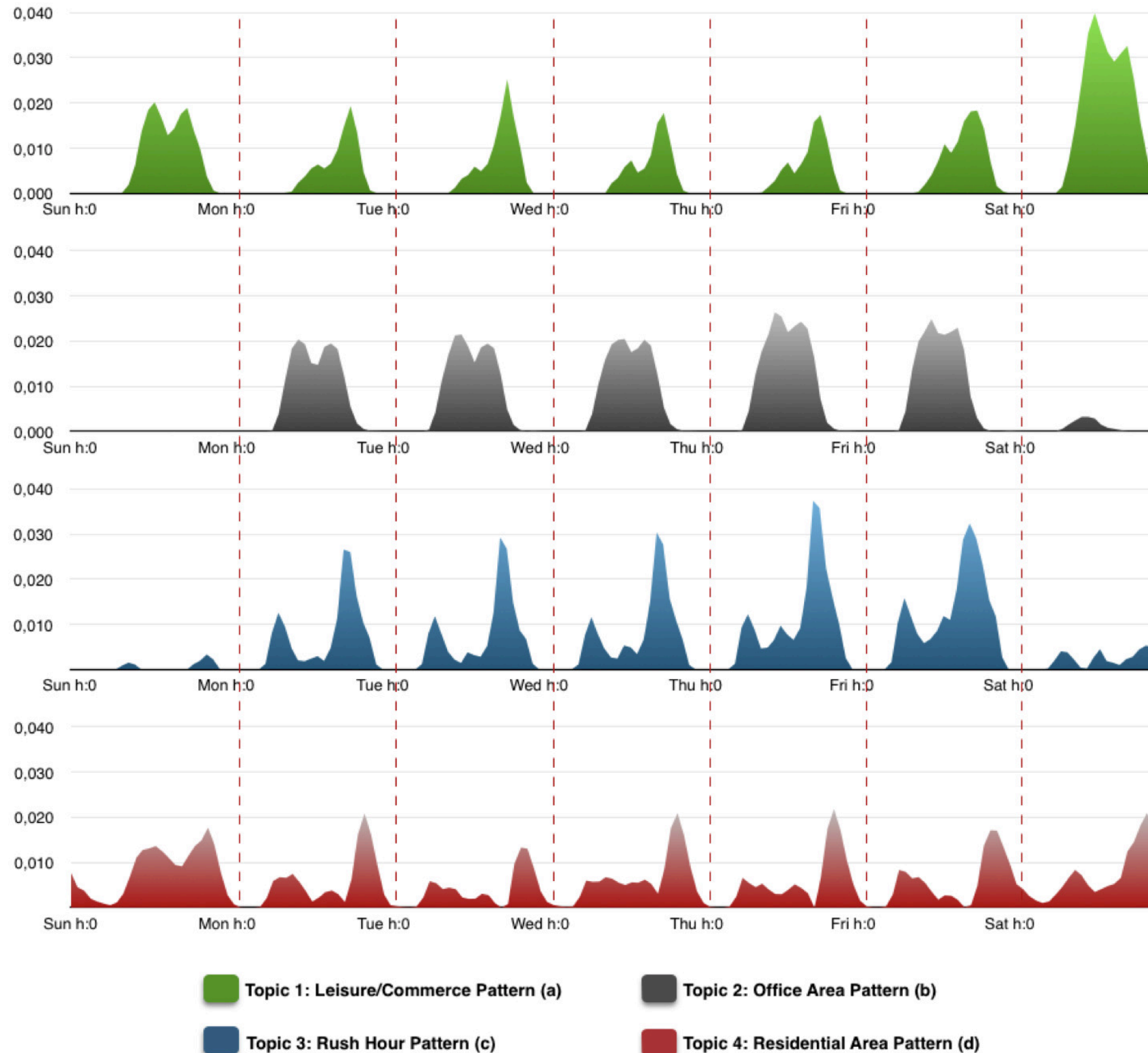
DESCUBRIENDO PATRONES DE USO DE SUELO: COMERCIO & RECREACION



Todos los puntos de acumulación corresponden a Centros Comerciales o Estadios donde es posible realizar conciertos (o actividades similares).



PLACES FINGERPRINTS OF USAGE



Ir al sitio

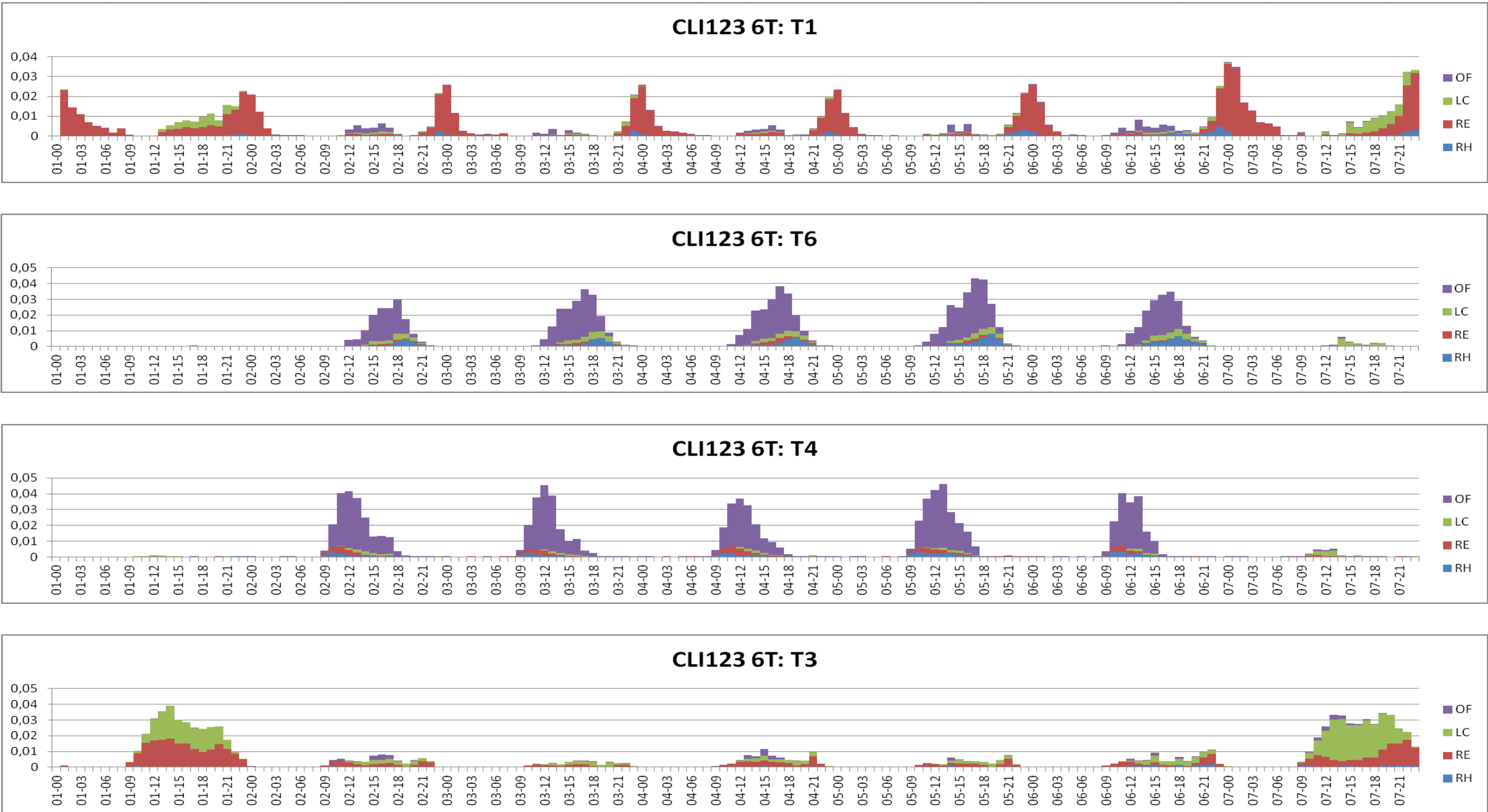
*Frecuencia de eventos por cada
torre de antenas (en todo Chile).*

Para cada hora del día.

Para todo el año 2015.

*Precisiones Menores a 350 mts. en
Santiago*

HUMAN ACTIVITY PATTERNS



Preguntas

