

UNIVERSIDAD DE LOS ANDES

Documento del resultado de la investigación para la convocatoria  
706 Colciencias de Jóvenes Investigadores 2015

## **Exploración de datos espacio - temporales para la analítica visual en sistemas urbanos.**

Joven Investigador:

**CAMILO ANDRÉS NEMOCÓN FARFÁN**

Director del proyecto:

**JOSE TIBERIO HERNANDEZ. PhD.**

## **1. OBJETIVOS:**

### **1.1 Objetivo General:**

Desarrollar una herramienta web que facilite la forma de expresión de las consultas Espacio Temporales de forma visual e interactiva, a partir de un modelo que administra la información en tres aspectos de espacio, tiempo, y datos, así como de sus niveles de granularidad, para visualizar, navegar y explorar los datos en sistemas urbanos de Bogotá (uso del suelo y movilidad) apoyando la tarea de análisis de planeación urbana.

### **1.2 Objetivos Específicos:**

- Crear un modelo de datos que permita la exploración y navegación en datos espacio - temporales.
- Diseñar y desarrollar una herramienta de interacción web de consultas espacio temporales que apoye la tarea de análisis de la exploración de datos en sistemas urbanos de Bogotá.

## **2. INTRODUCCIÓN:**

El crecimiento acelerado de la población y de la ciudad de Bogotá, requiere de una planeación urbana que satisfaga las nuevas necesidades de servicios e infraestructura de la ciudad, teniendo en cuenta los planes de gobierno, los cuales cambian constantemente y dependen de las políticas elaboradas por cada mandato.

El crecimiento de Bogotá se evidencia con las proyecciones de población realizadas por el Censo General del 2005, en donde se estima que para el año 2020 habitarán en la ciudad 8'118.782 personas y esto se argumenta con base en que en el año 2010 las localidades de Usaquén, Suba y Usme han sido las zonas con mayor expansión y en ese año se llegó a una densidad urbana de 191 personas / ha<sup>1</sup> [1].

Por tanto para un buen desarrollo y planeación urbana en torno al crecimiento de la ciudad es necesario ver y evaluar el comportamiento actual de la ciudad para así tomar las mejores decisiones, planeaciones y políticas en torno a los sistemas urbanos. Para ello los expertos necesitan visualizar simultáneamente gran cantidad de datos y explorarlos para descubrir patrones, relaciones, tendencias u otras características que apoyen su posterior proceso de análisis, planteamiento de hipótesis y toma de decisiones.

Los datos que usan los expertos en planeación urbana, son datos espacio – temporales, es decir datos que varían o existen dependiendo de un determinado momento en el tiempo y de una específica ubicación [2], ya que a partir del plan de desarrollo realizado en el tiempo de mandato de un gobierno se sacan datos que evidencian las actividades realizadas, las cuales pueden ser visualizados por medio de sistemas de información geográficos interactivos en plataforma web para explorar estos datos.

Para el proceso de exploración de datos espacio – temporales, se necesita una herramienta de visualización interactiva web que le permita al experto la selección de los factores que son importantes para él y evidenciar su comportamiento en el espacio y en el tiempo, lo cual, le permite tener una visión general de la situación, con el propósito de poder ver el estado actual y la evolución de la ciudad en términos de sistemas urbanos.

Por tanto es importante generar una herramienta que permita la integración de los datos espacio - temporales dentro de un sistema de información geográfica, siguiendo un protocolo que permita determinar la calidad de los datos, para su posterior uso dentro de un sistema de modelado y visualización de datos [3], ya que a partir de la validación de los datos se pueden tomar decisiones y realizar análisis por medio de aplicación web que vinculen la analítica visual para generar hipótesis y conclusiones que apoyen la tarea de planeación urbana.

### 3. ESTADO DEL ARTE: APLICACIONES Y HERRAMIENTAS QUE APOYAN LA EXPLORACIÓN DE DATOS DE SISTEMAS URBANOS.

En el área de sistemas urbanos y analítica visual encontramos trabajos de investigación para la integración de sistemas de información con fines de simulación y planeación, como es el trabajo de Wang cuya aplicación llamada CAL3QHC [4], integra los sistemas de información geográfica como ArcGis, con simuladores de Trafico y simuladores de calidad de aire, generando visualizaciones pertinentes para los expertos, sin embargo el usuario debe generar los datos en cada simulador para ser integrados en la aplicación, lo cual genera un aumento en el tiempo para la preparación de los datos para la visualización de cada tema a analizar y no permite la visualización web centrando y limitando la información y la toma de decisiones.

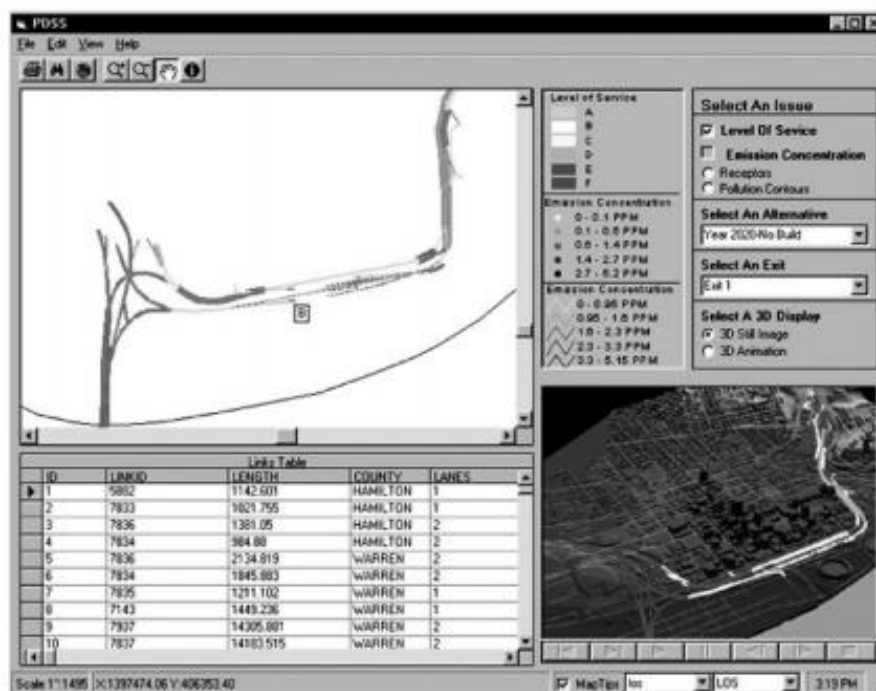


Fig 1. CAL3QHC, Vista gráfica del sistema de análisis del tráfico.

Otro proyecto es DHTML [5], que aunque no se enfoca en sistemas urbanos, si permite una exploración de datos a partir de las interfaces denominadas consultas dinámicas, para acelerar los procesos de consulta y proporcionar a los usuarios las presentaciones visuales de grandes conjuntos de resultados, permitiéndoles de la misma forma ver fácilmente adiciones o supresiones de las consultas a partir del diseño de un solo control el cual es el slider, sin embargo este proyecto se queda en el desarrollo de un solo control de interacción para el manejo de atributos sin diseñar el resto de elementos que ayudan a construir un mejor análisis de los datos.

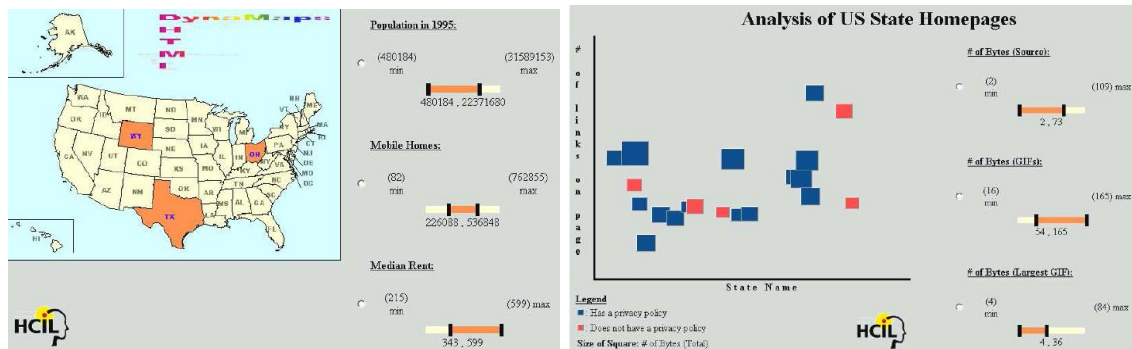


Fig 2. DHTML, densidad poblacional visualizada por medio de un mapa y de un Scattergram.

Otro proyecto de vinculación de analítica visual, Big data e interfaces de consulta es VisGets [6], en donde se centran en el desarrollo de una interfaz web para formular consultas sobre los datos, donde muestra un listado con las respuestas de la consultas utilizando widgets interactivos que permite ver la granularidad temporal, sin embargo la visualización de las respuestas no es clara y es limitado para la visualización de datos heterogéneos, ya que no hay una normalización comprensiva de los datos para el usuario.

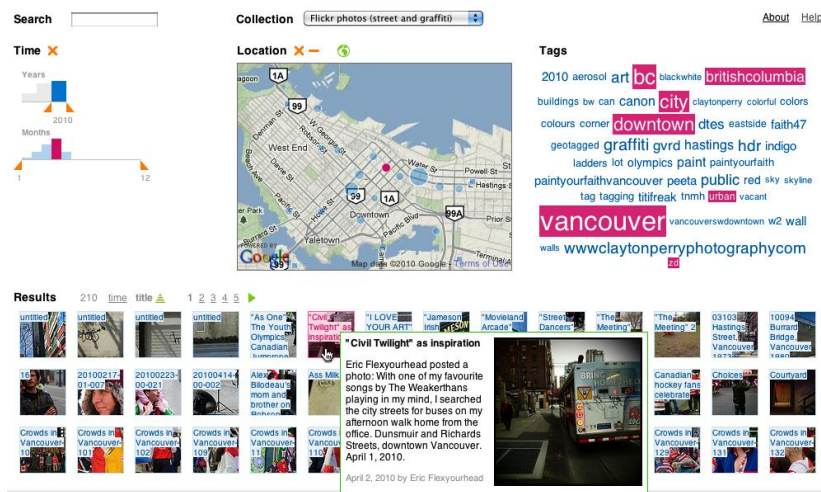


Fig 3. VISGETS, Fotos de Flickr del tema Graffiti mostrando los resultados de esta consulta.

También encontramos otra aplicación llamada EuropeanaConnect [7], en donde se puede obtener información relevante frente a tópicos históricos dependiendo de la ubicación y el momento en el tiempo consultado, permitiendo visualizar espacial y temporalmente la información de un tópico específico, sin embargo no permite la relación de temas que podrían ser importantes para el usuario, ya que solo muestra los datos de cada punto de forma individual sin poderlos relacionar.

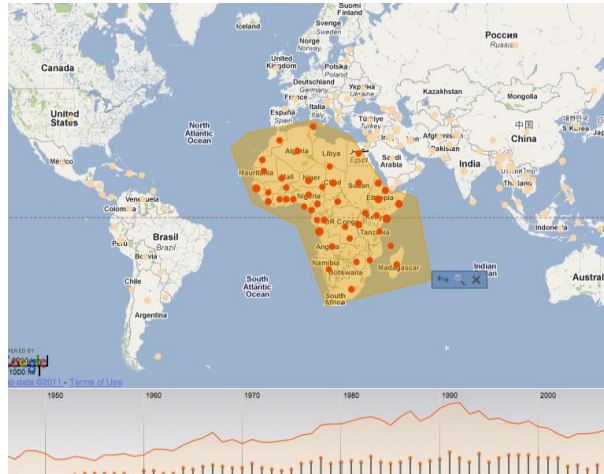


Fig 4. EUROPEANACONNECT, Despliegue de datos individuales a partir de un punto y un tiempo.

Existen otros proyectos en donde introducen los conceptos que subyacen a la correcta visualización de información y que permiten la interrelación de los datos, lo cual ayuda al descubrimiento de ciertas tendencias espaciales y temporales, permitiendo generar clústeres, umbrales, correlaciones, relaciones jerárquicas y así sucesivamente, y esto se evidencia en el sistema Panorama SOLAP [8] en donde involucran vistas múltiples, las cuales permiten ver los datos desde diferentes perspectivas, donde cada representación destaca diferentes aspectos de los datos; sin embargo no hay diseño centrado en el usuario, lo cual no permite un mejor diseño, cognición y percepción (aumento de la percepción humana), lo cual genera dificultad en el descubrimiento del conocimiento, en aprender a manejar la herramienta y en las interfaces para realizar las consultas ya que no existen controles claros para el filtro temporal, espacial y de atributos.

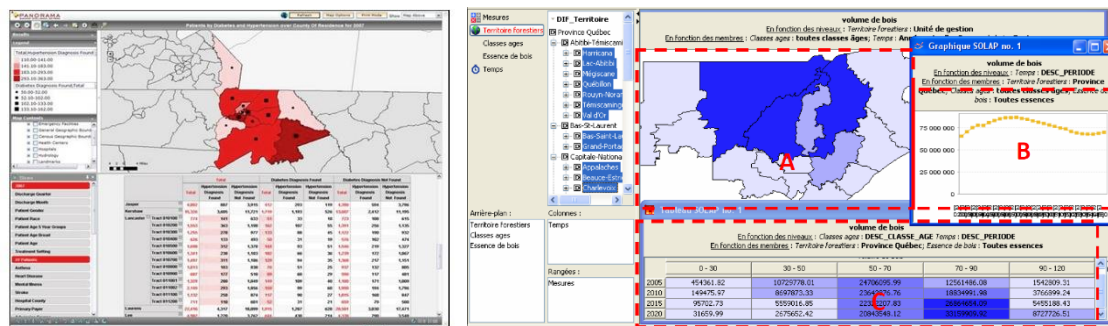


Fig 5. Interfaz Panorama SOLAP.



Volviendo a enfocarnos hacia sistema urbanos existe un proyecto centrado en la movilidad en New York [9], construye su propio modelo base de datos y proponen el uso de controles de restricciones temporales, restricciones de atributos, restricciones espaciales para generar consultas atómicas como input retornando un conjunto de elementos (vehículos) que satisface una restricción (restricción de tipo de vehículo, rango de tiempo, ubicación). Así mismo permite utilizar múltiples puntos de vista para generar y así comprobar distintas hipótesis de los expertos.

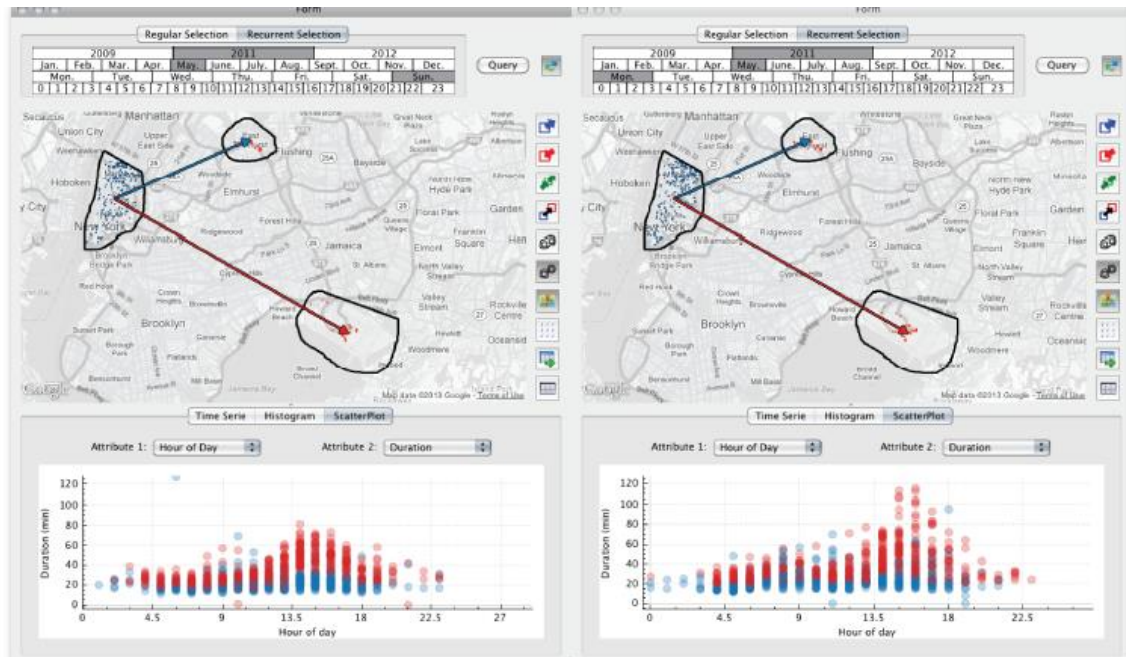


Fig 6. Recorridos realizados por los Taxis en Nueva York.

Otro proyecto con planteamiento similar frente al comportamiento del tráfico en una ciudad, es el realizado por Alberto Debiassi [10], quien utiliza la analítica visual, la interacción sobre las visualizaciones y el ambiente en 3D para establecer los parámetros de comportamiento y así los usuarios pueden detectar cuales son las vías más congestionadas dado un tiempo y número de vehículos en la vía, sin embargo este proyecto está enfocado hacia la simulación.

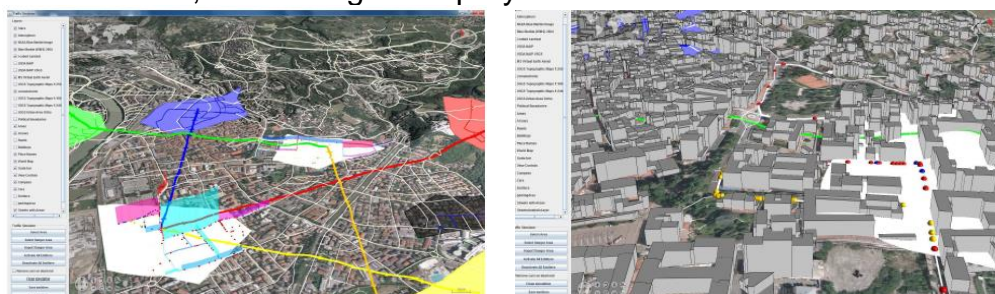


Fig 7. Simulación del tráfico.

Así mismo encontramos un proyecto de visualización del recorrido de vehículos [11], esta aplicación muestra un histograma temporal y un diagrama de barras los intervalos de tiempo, donde las alturas de las barras representan el número de lugares visitados o la distancia recorrida y las barras tienen una orientación o sentido y un color el cual refleja la dirección de los carros (analogía rosa de los vientos). Por otro lado cada punto posee un radio que representa el número de

vehículos con una velocidad determinada por el usuario. Sin embargo la interfaz no es intuitiva y necesita de una interpretación previa para su entendimiento, lo cual no permite una buena exploración de los datos más allá de averiguar la cantidad y dirección de los vehículos.

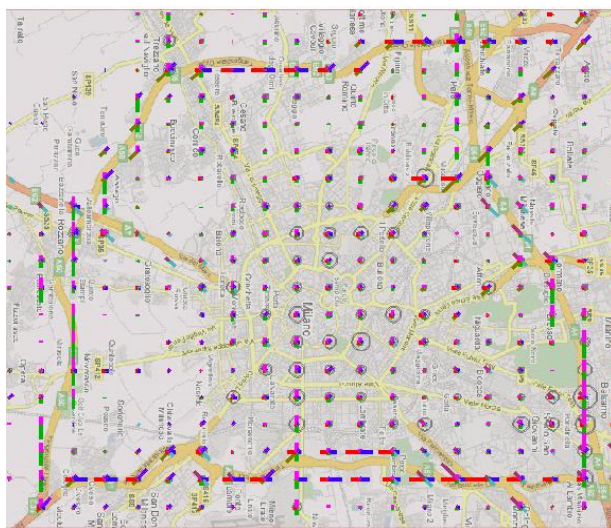


Fig 8. Recorrido de vehículos.

#### 4. METODOLOGÍA:

Este proyecto tiene como base el diseño de un modelo de manejo de datos espacio – temporales y la implementación del mismo sobre una aplicación web que permita generar consultas sobre estos. Esta aplicación es una herramienta interactiva de análisis visual, la cual se desarrolla en torno al diseño centrado en el usuario generando un proceso iterativo donde a partir de las pruebas se va implementando funcionalidades dentro de la aplicación.

- **Definición de datos de estudio:**

En este proyecto se trabajó un conjunto de datos del transporte masivo de la ciudad de Bogotá, Transmilenio, los cuales permiten su manejo enfocado hacia los sistemas urbanos y específicamente hacia el área de movilidad; estos datos son espacio – temporales, los cuales permiten a los usuarios identificar y ubicar los factores importantes en torno a una o varias ubicaciones y un tiempo o rango de tiempos, de esta forma el experto puede ver el comportamiento de las estaciones de Transmilenio a partir de la segmentación por periodos de tiempo (año, mes, día y hora), permitiendo el análisis y visualización de la información, probando el modelo de exploración y manejo de datos desarrollado, con el cual se puede realizar de forma fácil y rápida las consultas y visualizar los datos dentro de un ambiente web, apoyando las tareas de análisis de exploración de datos.

Los datos obtenidos del sistemas de transporte Transmilenio, corresponden a las entradas y salidas por torniquete de las estaciones de la fase 1, 2 y 3, de los años 2011, 2012, 2013, 2014, cuyo tipo de archivo está en formato .xls y la fuente de información es de Transmilenio.

Estos datos se normalizaron con base al modelo de manejo de datos desarrollado, generando el siguiente orden:

<b>Dato</b>	<b>ESCALA CUANDO</b>	<b>CUANDO</b>							
Entradas y salidas por torniquete . Fases I y II	AÑO	2011		2012		2013		2014	
	MES	MAYO	SEPTIEMBRE	MAYO	SEPTIEMBRE	MAYO	SEPTIEMBRE	MAYO	SEPTIEMBRE
	DIA	1 AL 31	1 AL 30	1 AL 31	1 AL 30	1 AL 31	1 AL 30	1 AL 31	1 AL 30
	HORA	24	24	24	24	24	24	24	24
Entradas y salidas por torniquete . Fase III	AÑO			2012		2013		2014	
	MES			SEPTIEMBRE		MAYO	SEPTIEMBRE	MAYO	SEPTIEMBRE
	DIA			1 AL 30		1 AL 31	1 AL 30	1 AL 31	1 AL 30
	HORA			24		24	24	24	24

<b>Dato</b>	<b>ESCALA DONDE</b>	<b>DONDE</b>	
Entradas y salidas por torniquete. Fases I y II	CIUDAD	BOGOTÁ	
	FASE	1 Y 2	
	TRONCAL		
	ESTACIÓN	1 A LA 4	
Entradas y salidas por torniquete. Fase III	CIUDAD	BOGOTÁ	
	FASE	3	
	TRONCAL	zona k	zona L
	ESTACIÓN	Portal Eldorado/ Modelia /Normandía/Av. Rojas/El Tiempo/ CAN/Gobernación/Quinta Paredes/Corferias/Ciudad Universitaria/Plaza de la Democracia/Centro Memoria/Universidades/	Portal 20 de Julio/Country Sur/ Av. Primero de Mayo/Ciudad Jardín/Policarpa/San Victorino/Las Nieves/San Diego/Hospitales

<b>Dato</b>	<b>ESCALA QUE</b>	<b>QUE</b>			
Entradas y salidas por torniquete . Fases I y II	TEMA	MOVILIDAD			
	SUBTEMA	ESTACIONES TRANSMILENIO			
	VARIABLES	DESCRIPCIÓN TIPO DE VIAJE	CANTIDAD POR MOLINETE	NÚMERO MOLINETE	CODIGO ESTACIÓN
Entradas y salidas por torniquete . Fase III	TEMA	MOVILIDAD			
	SUBTEMA	ESTACIONES TRANSMILENIO			
	VARIABLES	ACCESO PEATONAL	NUMERO DE SALIDAS	NUMERO DE ENTRADAS	

En resumen, estos son datos espacio – temporales, donde la ubicación de las estaciones se normalizaron dentro de un formato expresado en longitud y latitud,



teniendo de esta forma una información espacial las estaciones de Transmilenio de la ciudad de Bogotá. Cada estación muestra la cantidad de personas que entran y salen de la estación de acuerdo a la orientación de la estación, es decir muestra los accesos norte, sur, oriente y occidente.

A nivel de tiempo, tenemos que de cada año hay datos de los meses de mayo y septiembre, donde cada mes se tienen datos de todos los días correspondientes a cada mes y así mismo se obtienen las 24 horas de cada día, por tanto se tiene en todo nivel granular temporal la cantidad de personas que acceden al servicio de Transmilenio en las fases estaciones correspondientes a la fase I, II y III.

- **Tarea de análisis:**

Se realizó la implementación de un modelo de tareas de análisis que utilizan los expertos dentro del área de sistemas urbanos para el análisis y toma de decisiones, donde a partir de una serie de pasos le permiten entender una situación o un estado en un sistema mediante la exploración de tiempos, ubicaciones y atributos.

Modelo de Tarea de Análisis [12]:

1. Exploración: observación de datos de un periodo de tiempo a otro, a partir de la selección de una ubicación en donde se visualizan únicamente las variables de contexto.
2. Evaluación: proceso subjetivo donde se miden los indicadores de impacto a partir de la variable de decisión y la variable de contexto. El indicador de impacto determina cuales de las variables o atributos es mejor que otra, a partir de los valores de los atributos seleccionados por el usuario y con este resultado se escoge y se plantea una alternativa.
3. Generación de alternativas: creación de nuevos estados sobre el sistema, donde estos estados corresponden a los cambios que realiza el usuario seleccionando nuevas ubicaciones, tiempos o cambios en las variables de decisión y/o en nuevos datos agregados, que generan un nuevo indicador de impacto.
4. Valoración de alternativas: se determina la importancia de las variables para los expertos (factor de relevancia de los atributos), y se generan comparaciones entre las múltiples vistas en donde se evidencian los cambios dados por el factor de relevancia.

Este proyecto se centró en la primera tarea de análisis, la cual es: exploración de datos espacio – temporales, ya que estos datos y esta tarea es la base para la modelación, planeación y análisis del sistema urbano, debido a que son datos que varían en el tiempo refiriéndose a una ubicación espacial. Estos datos se pueden estimar y visualizar, permitiendo la exploración de comportamientos evidenciando valores de atributos en lugares específicos en tiempos determinados. De esta forma se puede observar y correlacionar diferentes eventos en el tiempo y el lugar para descubrir relaciones entre ellos, e identificar y ubicar diferentes patrones en los datos [13].

- **Antecedentes y contexto del modelo de datos espacio -temporales:**

Los datos espacio - temporales se pueden representar conceptualmente mediante un cubo, esta representación la realiza A. Koussoulakou, donde cada celda en el cubo representa un objeto espacial, un atributo y un paso en el tiempo [14]; la navegación por el cubo se puede realizar: sobre el eje “X” para la exploración de atributos, sobre el eje “Y” para los lugares y sobre el “Z” podemos encontrar el tiempo, esto permite la búsqueda de los datos a través de la navegación por cada uno de sus ejes.

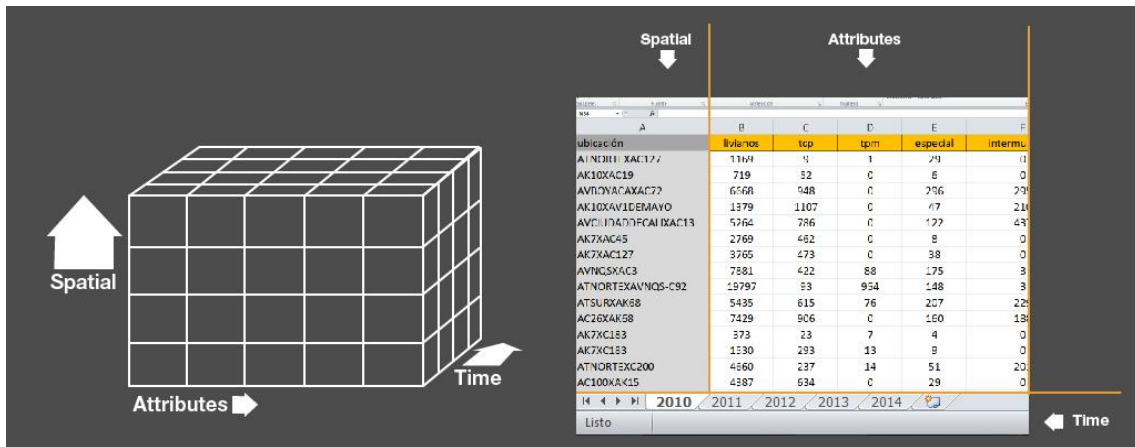


Fig 9. Cubo espacio -temporal.

El eje del tiempo en datos espacio–temporales, se define como una dimensión para comprender un evento o un determinado estado del mundo [15] pero esta dimensión está compuesta por un sistema jerárquico de granularidades, incluyendo segundos, minutos, horas, días, semanas, meses, años, décadas, siglos; sin embargo en este proyecto se consideró trabajar con variables temporales agrupándolas por los niveles de granularidad que utilizan los expertos en sistemas urbanos en el área de movilidad, por lo que utilizamos la granularidad de año, mes, día y horas.

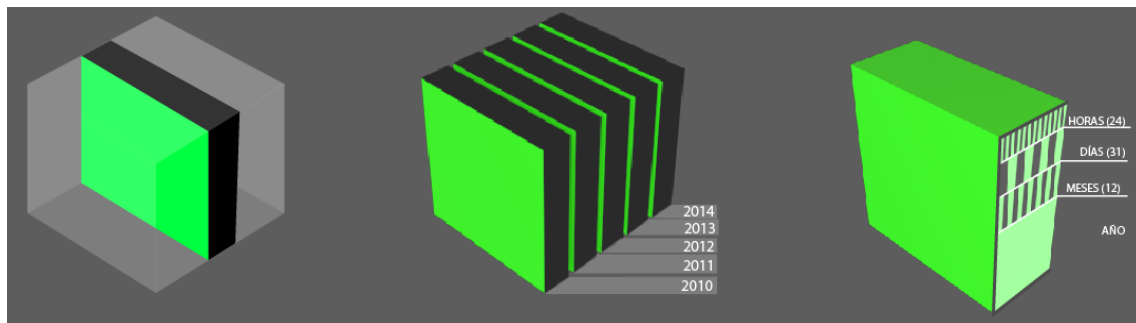


Fig 10. Cubo temporal.

El eje del espacio en datos espacio–temporales, se define como la ubicación constituida por la latitud y longitud ocupada por un objeto y describe las acciones en el tiempo de un conjunto de objetos a partir de una ubicación. Los datos espaciales pueden experimentar cambios existenciales a través de una referencia espacial, la cual se muestra por medio de un mapa que transporta la información de los datos espaciales referenciados a una ubicación espacial [16], sin embargo el entorno espacial está constituido por la heterogeneidad del espacio geográfico la cual puede estar categorizada por niveles de granularidad

dada por divisiones políticas, geográficas, etc. Por tanto, este trabajo se basó se determinó su granularidad como Ciudad, Fase, Troncal, Estaciones.

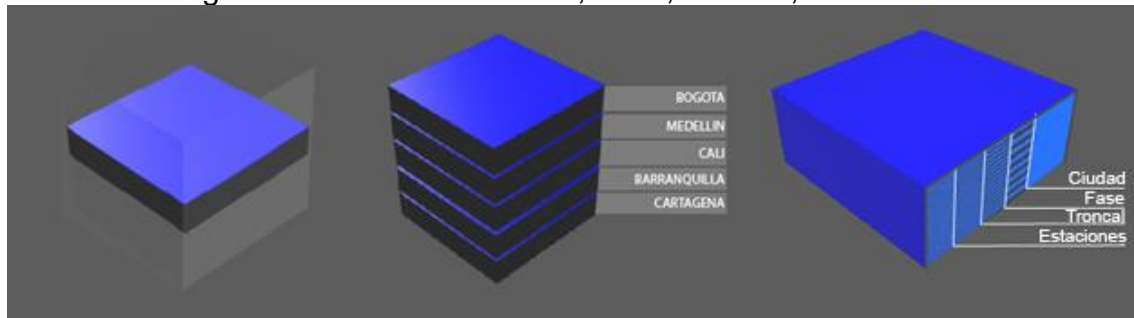


Fig 11. Cubo espacial.

El eje de los atributos en datos espacio–temporales se define como un número limitado de valores de datos heterogéneos debido a que los datos pueden ser cualitativos permitiendo condiciones igualitarias entre sus categorías y cuantitativos los cuales permiten condiciones numéricas y por intervalos [17].

Los atributos se pueden agrupar usando la taxonomía clásica de clasificación de tipos de variable de la estadística descriptiva, generando una granularidad dada por un tema, subtema y las variables que posee cada subtema. El orden de los atributos en la herramienta corresponderá al orden como se presentan los datos, así mismo nos enfocaremos en el uso de variables cuantitativas y cualitativas nominales ya que son las más usadas y las más frecuentes en el dominio de Sistemas Urbanos.

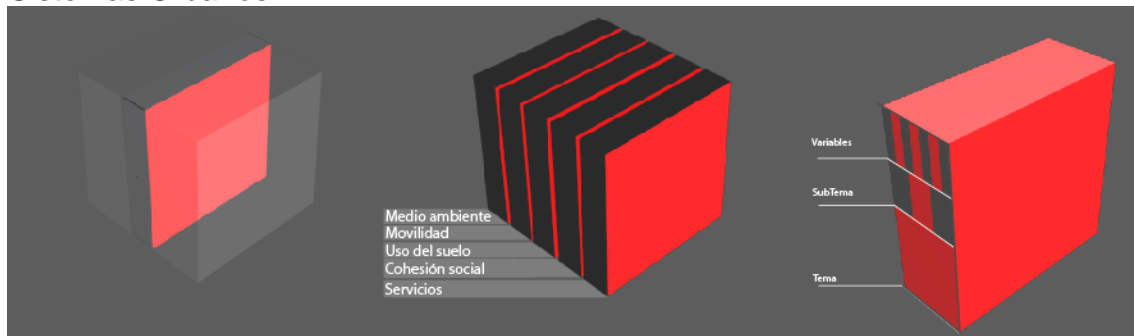


Fig 12. Cubo atributos.

Para realizar la navegación por los datos espacio–temporales existe un modelo de consultas propuesto por D. Peuquet, en donde permite considerar todos los componentes de los datos y ayuda a expresar restricciones de consulta espacio temporales, convirtiendo los ejes de estos datos en adverbios interrogativos, es decir, tiempo (¿cuándo?), espacio (¿dónde?) y atributo (¿qué?), para así generar filtros y búsquedas sobre los datos [18].

En la figura 13 encontramos características de cada una de las preguntas que se pueden realizar con este modelo de consultas y su representación sobre el cubo espacio temporal de Koussoulakou [14].

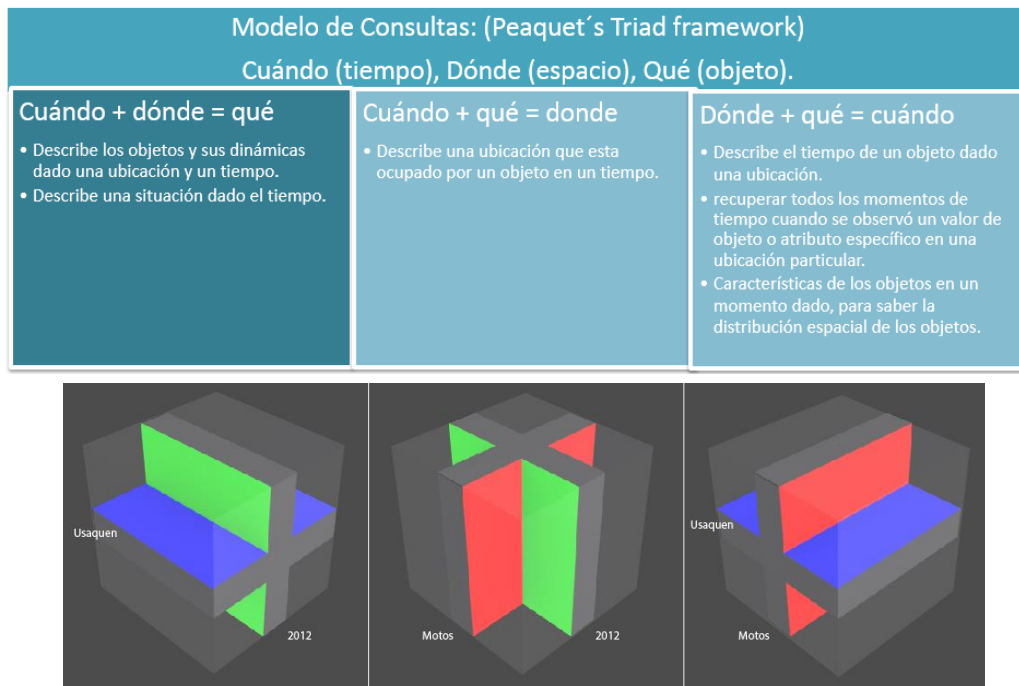


Fig 13. Modelo D. Piuquet.

## 5. PROPUESTA:

- **Modelo de manejo de datos “Modelo Exploratorio”:**

Se utilizó como base y referencia el modelo de consultas de D. Peuquet [18], en donde permite considerar todos los componentes de los datos y ayuda a expresar restricciones de consulta espacio - temporales por medio de las preguntas que facilitan la expresión de las consultas. A partir de éste, se desarrolló un modelo que maneja los datos permitiendo la navegación, filtrado y exploración de datos manejando cada uno de sus componentes de forma independiente, con lo cual el experto en el área de sistemas urbanos puede seleccionar tres tipos de componentes de interés (espacio, tiempo y atributos) para observar los datos de un periodo de tiempo a otro, a partir de una selección de una o varias ubicaciones en donde se visualizan únicamente las variables de contexto de interés.

El modelo Exploratorio es en una estructura que genera el desacoplamiento del espacio, tiempo y atributos y sus niveles de granularidad, permitiendo la visualización y exploración para la lectura, búsqueda y filtrado de los datos por medio del semantic zoom y la desagregación de los datos permitiendo acceder a éstos desde sus diferentes niveles.

El modelo apoya la tarea de análisis de exploración de datos mostrando la información solicitada y eliminando de la vista del usuario los datos que no se ajusten a la restricciones de consulta, permitiendo además la combinación entre los niveles de granularidad que poseen los datos temporales y espaciales ya que cada celda del cubo representa un objeto espacial asociado a un atributo en un paso en el tiempo y dentro de cada uno se puede mostrar sus niveles de granularidad.

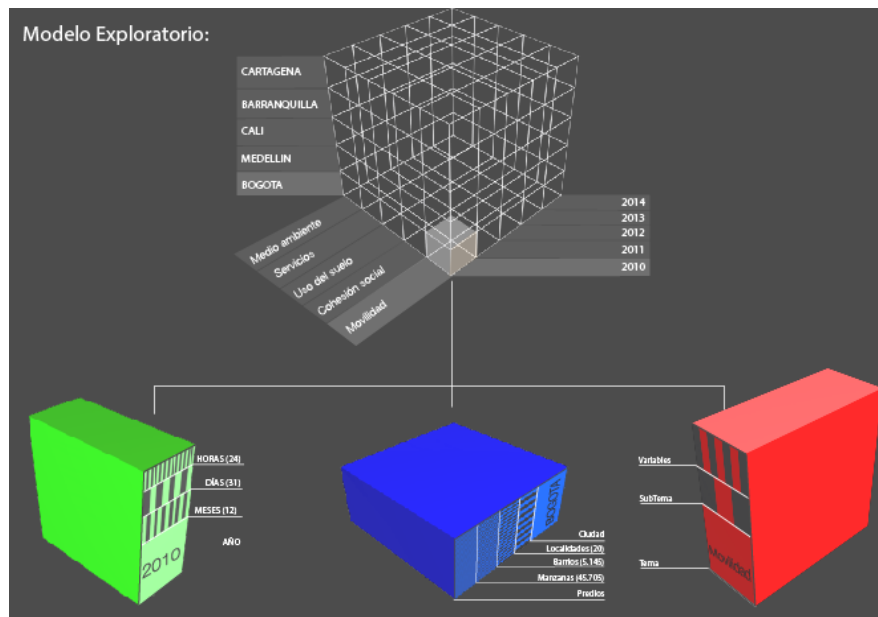


Fig 14. Modelo Exploratorio.

El diseño de este modelo permite ver los distintos niveles de granularidad de uno o varios componentes de cada uno de los ejes, y se les denominó como modos exploratorios.

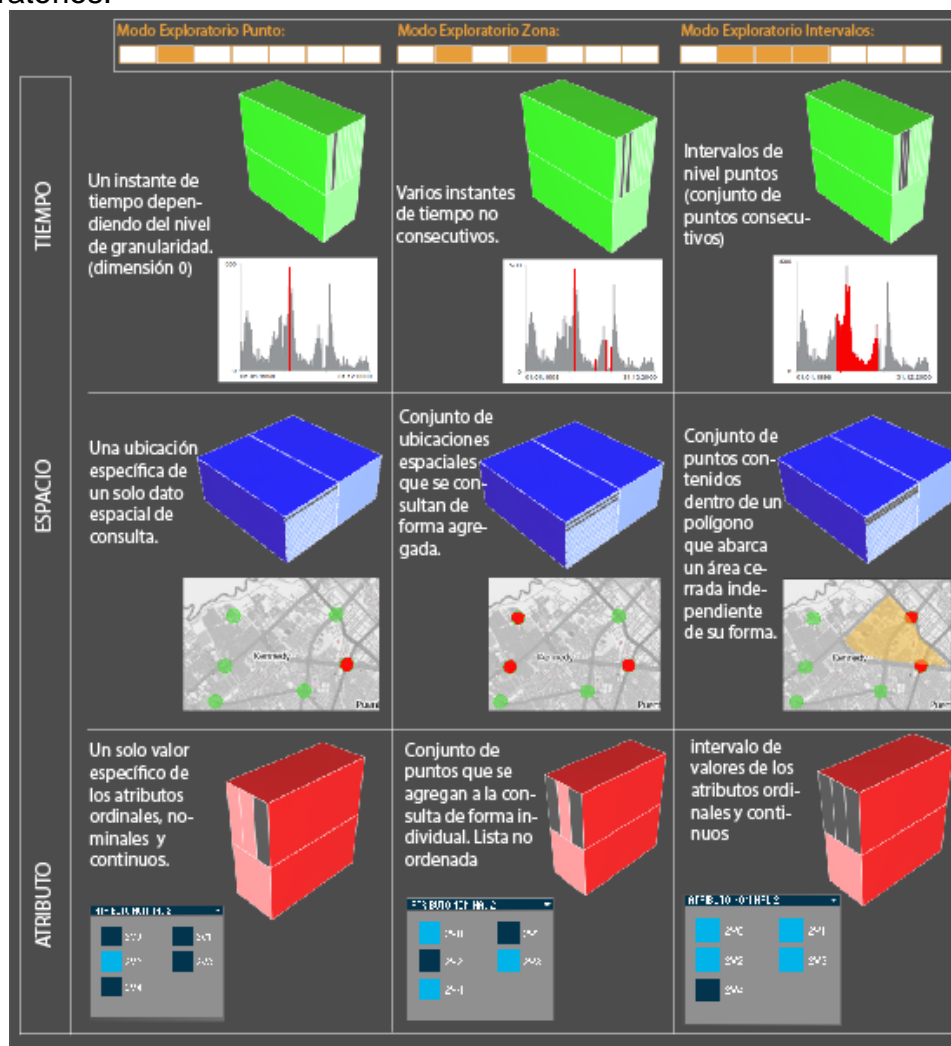


Fig 15. Modos Exploratorios.

- **Grafo del modelo de manejo de datos “Modelo Exploratorio”:**

El modelo se representa mediante un grafo, en donde su núcleo está compuesto por los componentes de Espacio, Tiempo y Atributos, los cuales poseen sus propio niveles de granularidad, generando una estructura basada en un árbol por cada componente, representando de esta forma la jerarquía de las granularidades de cada componente, con un sistema no direccional, que permite la relación libre del núcleo del grafo con cualquier grado de granularidad de alguno de sus componentes para así realizar cualquier tipo de consulta sobre datos espacio – temporales.

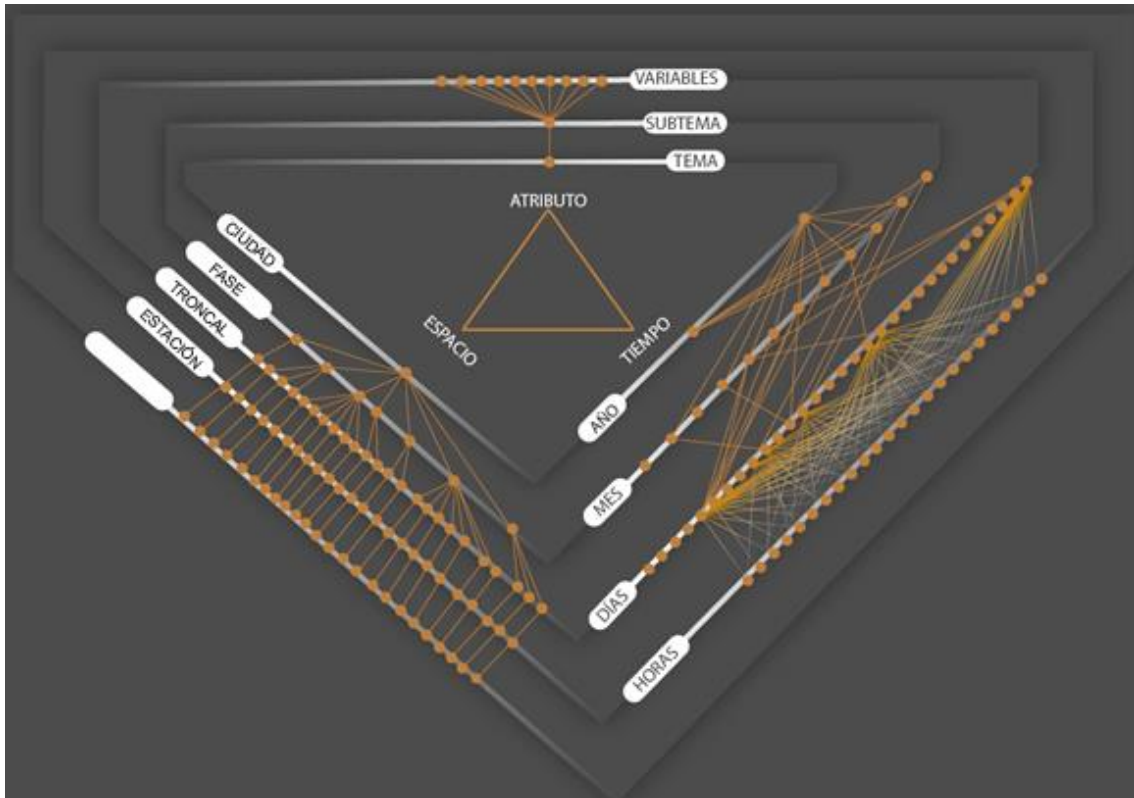


Fig 15. Modos Exploratorios.

- **Implementación del modelo y aplicativo Web:**

La herramienta interactiva web permite la implementación del modelo para resolver la tarea de análisis de la exploración de los datos dando resultados visuales e interactivos de los datos. La exploración de los datos se realiza a través de consultas mediante el uso de preguntas, las cuales se generan a partir del filtrado de los ejes temporales, espaciales o de atributos y de la selección de uno o varios datos en sus distintas granularidades.

El desarrollo de la aplicación web se realizó con base de datos SQLite utilizando procesos que permiten una descarga y búsqueda rápida de las consultas que pueda realizar el usuario sobre una base de datos que inicialmente tenía un peso de 98MB y se optimizo a 18MB, esta optimización se realiza mediante el navegador, el cual descarga la versión comprimida, permitiendo una carga de la página rápida y luego se descomprime con Javascript, mostrándole al usuario un “overlay” o capa de inicio en donde se muestra un temporizador, que evidencia



el progreso de la descarga y descompresión de los datos, y la configuración de la base de datos para posteriormente empezar a interactuar con la página.

Las tecnologías utilizadas fueron Javascript y SQL.js, que se implementaron junto a SQLite, donde éste sistema es finalmente un manejador de base de datos en memoria donde todo el proceso corre en memoria por el lado del cliente, de ésta forma todas las consultas y procesos se realizan en el navegador del cliente y el servidor solo funciona como un repositorio de archivos e información, ya que toda la lógica se implementa en el index.html de la página web.

El sistema utiliza los índices de las tablas de la base de datos para realizar las consultas permitiendo una velocidad de respuesta entre consulta y consulta de 200 milisegundos o menos, mediante los id de espacio, tiempo y atributos, y utilizando como componente un time out que se resetea cuando se interactúa con cualquier elemento de la interfaz. La interfaz se realizó con Bootstrap, el cual permite darle la estructura a la página para que sea responsive y me permita ubicar los elementos de consulta y respuesta. Así mismo, se utilizó la librería de JQuery para el manejo de los componentes gráficos e interactivos de la interfaz.

#### - Visualización:

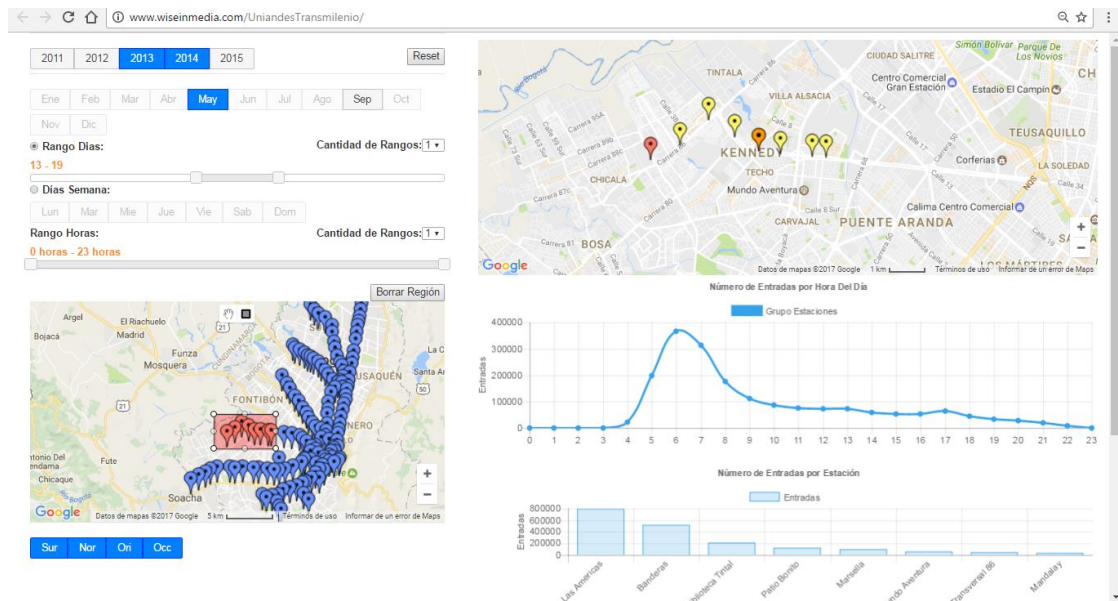


Fig 16. Aplicativo web ([www.wiseinmedia.com/UniandesTransmilenio/index.html](http://www.wiseinmedia.com/UniandesTransmilenio/index.html))

La visualización web está compuesta por dos áreas, al costado izquierdo encontramos el área de consultas y al costado derecho está el área respuestas, donde en el área de consultas el usuario se centra en la selección de ítems de espacio, tiempo y atributos para filtrar de esa forma la información que es mostrada en el área de resultados, en donde se evidencia puntualmente los lugares de interés para el usuario y muestra la información individual de estos en la parte superior de esa área y en la parte inferior se muestran los resultados en conjunto, para que el usuario pueda comparar los resultados de su selección y

exploración de forma conjunta vs. el resultado individual de cada componente espacial seleccionado.

- Área de consultas:

En el área de consultas se ubican los tres tipos de componentes de interés (espacio, tiempo y atributos) con sus respectivas granularidades, permitiendo al usuario tener una vista general de los datos que posee y sobre estos poder realizar las consultas en los tres componentes, donde cada componente tiene una forma de visualización e interacción diferente para realizar consultas y filtrados sobre los datos.

Temporal:

The screenshot displays the temporal query interface. At the top, there are buttons for years from 2011 to 2015, with 2014 selected. Below this, there are buttons for months from January (Ene) to December (Dic), with May (May) selected. Under the month selection, there is a radio button for 'Rango Dias' (Day Range) and a dropdown menu for 'Cantidad de Rangos' (Number of Ranges) set to 3. Below this, there are three horizontal sliders for day ranges: '13 - 15', '4 - 9', and '22 - 27'. Below the sliders, there is a radio button for 'Días Semana' (Days of the Week) and a dropdown menu for 'Cantidad de Rangos' set to 2. Below this, there are buttons for days of the week from Monday (Lun) to Sunday (Dom), with Wednesday (Mie) selected. At the bottom, there is a section for 'Rango Horas' (Hour Range) with a dropdown menu for 'Cantidad de Rangos' set to 2. Below this, there are two horizontal sliders for hour ranges: '6 horas - 8 horas' and '16 horas - 19 horas'.

Fig 17. Área de consulta temporal

En la parte superior de esta área, se encuentran los botones correspondientes a los años, donde el sistema coloca automáticamente la cantidad de botones de acuerdo a la cantidad de años que hay en la base de datos, de esta forma le muestra al usuario los años de los cuales posee información para consultar. Este mismo sistema aplica para la parte inferior que le sigue a este componente, en donde se muestran todos los meses del año, sin embargo aparecen en opacidad los botones de los meses en los que no hay datos.

Luego encontramos la granularidad de los días, que permite seleccionar un día a la semana, un día del mes o un rango de días del mes, así mismo permite generar varios slider para poder seleccionar varios rangos de días en el mes, permitiendo la mayor flexibilidad en la selección temporal y utilizando los modos exploratorios del modelo (Modo Punto, Modo Zona y Modo Intervalo).

Bajo este mismo concepto se realizó la selección de las horas permitiendo la selección de una hora, un rango de horas o varios rangos de horas.

Se determinó utilizar el sistema de slider ya que la cantidad de días y horas son numerosas, y es más fácil para el usuario seleccionar deslizar un slider que oprimir “n” número de botones. Con este componente el usuario puede realizar consultas temporales desde cualquier nivel granular, es decir desde el nivel más general como es el año en donde va a obtener todos los datos de ese año, como a nivel específico en donde puede obtener todos los datos de uno o varios años, meses, días y horas.

Espacial:



Fig 18. Área de consulta espacial

Aquí nos permite ver los puntos espaciales de las cuales poseemos datos, ubicándolos sobre un mapa, a partir de la latitud y la longitud. Los puntos rojas corresponden a los puntos seleccionados por el usuario y los azules los que no han sido seleccionados.

En la parte central y superior del mapa encontramos dos botones, un cuadrado, que nos permite hacer una o varias áreas rectangulares sobre el mapa, lo cual genera automáticamente la selección de aquellas estaciones que se encuentran dentro de estas áreas. Estas áreas se pueden borrar seleccionando cada área y oprimiendo el botón que se encuentra en el costado derecho superior del mapa.

El otro botón que está en la parte central corresponde a una mano la cual nos permite modificar las áreas rectangulares realizadas y también permite el paneo y la navegación sobre el mapa. Así mismo esta mano permite la selección individual de una estación o la selección puntual de varias estaciones.

Al costado lateral derecho encontramos los botones de zoom in y zoom out para ver la granularidad espacial de donde se encuentran la información de los datos.

Atributos:



Fig 19. Área de consulta de atributos

Aquí encontramos los botones correspondientes a los atributos de los datos, donde, en nuestro caso de estudio se encuentran los accesos a las estaciones. Mediante estos botones el usuario puede seleccionar uno o más accesos para observar la cantidad de personas que ingresaron por cada acceso.

- Área de Resultados:

En el área de resultados esta compuesta por dos elementos un mapa focalizado y un sector de gráfico de utilidad que se encuentran abajo del mapa. Ambos sectores se activan simultáneamente cuando el usuario realiza las consultas de espacio, tiempo y atributos, evidenciando los resultados generados por cada selección que se realiza en el área de consulta.

Mapa focalizado:

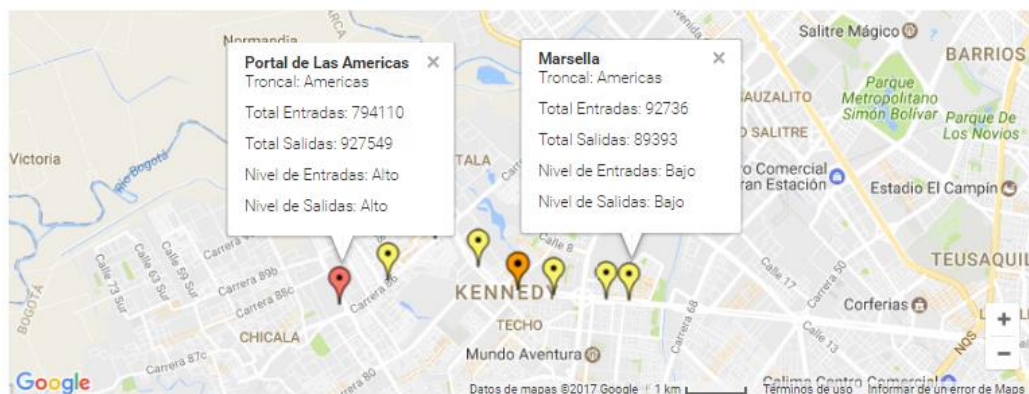


Fig 20. Mapa focalizado

En la parte superior del área de resultados se encuentra un mapa en donde se visualizan las ubicaciones espaciales de interés del usuario, es decir, las estaciones seleccionadas por el usuario en el área de consulta espacial. De esa forma aparecen las estaciones con la información de interés para el usuario de forma individual, por tanto el usuario puede seleccionar cada uno de estas estaciones y le mostrará su ubicación, el nombre, la troncal, la cantidad de personas que entraron y salieron en esa estación seleccionada.

Así mismo cada punto espacial visualizada tiene un color que determina la cantidad de un atributo ordenando de mayor a menor, el valor del atributo en cada punto espacial, de esta forma el atributo con mayor valor generará un color rojo, el siguiente atributo con un valor menor tendrá el color naranja, amarillo y blanco el atributo con menor valor entre los puntos espaciales seleccionados y comparados.

Por otro lado el usuario tiene los controles de navegación sobre el mapa para navegar, como los botones de zoom out, zoom in y paneo.

### Gráficos de utilidad:

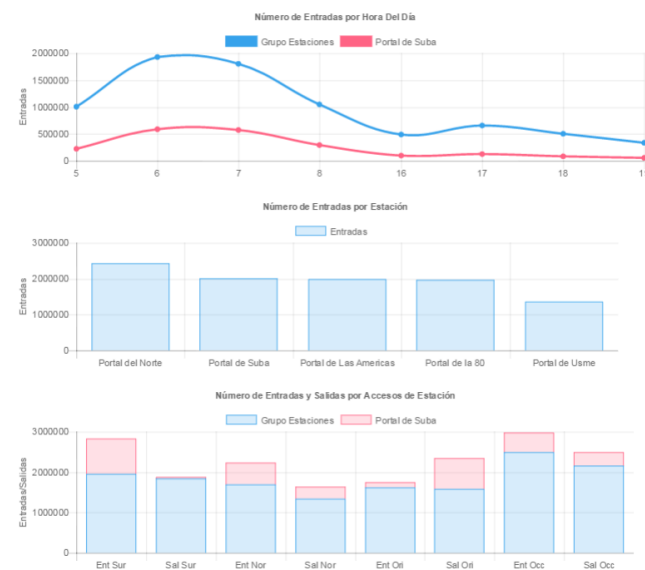
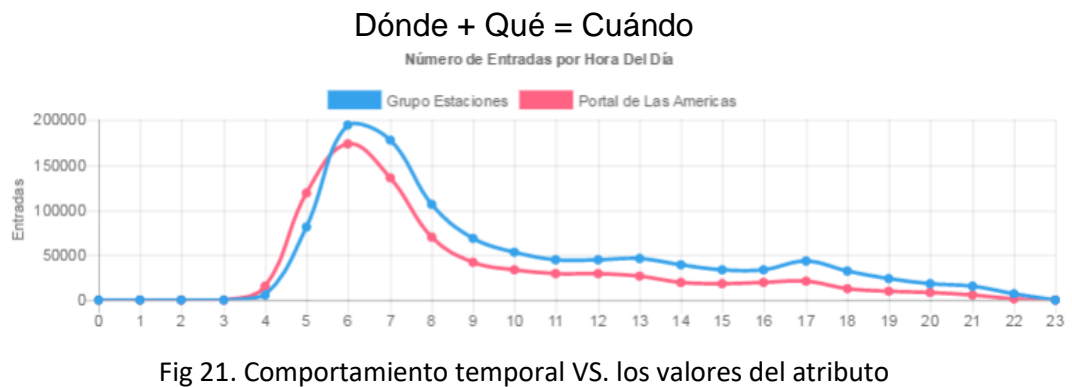


Fig 21. Gráficos de utilidad

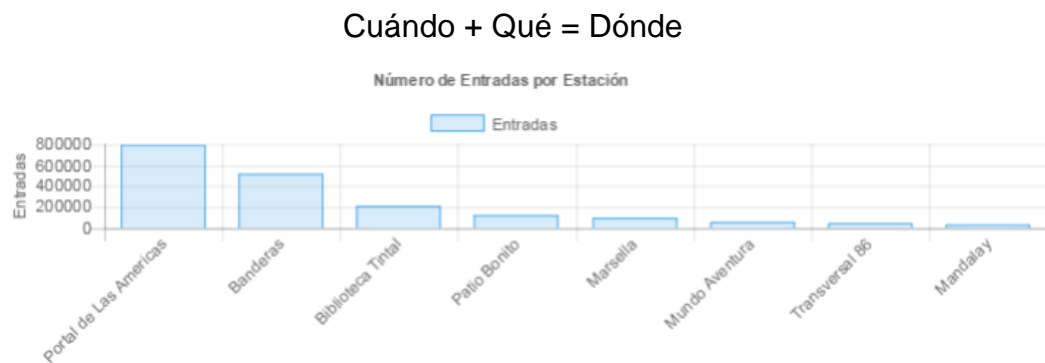
Las gráficas de utilidad son los diagramas que se encuentran abajo del mapa de resultados, donde estas gráficas permiten distintos análisis sobre los mismos datos, lo cual permite ver diferentes enfoques de la información y su retroalimentación automática de acuerdo a la interacción que se tenga sobre el área de consulta y el mapa focalizado.

Las gráficas de utilidad representan el resultado colectivo o total de las ubicaciones seleccionadas por el usuario versus los atributos seleccionados por el usuario, sin embargo estas cambian dinámicamente, cuando el experto toca una estación en el mapa focalizado, ya que en el área de utilidad se va a ver reflejado los resultados de esa estación en particular vs. los datos acumulados de todas las estaciones, para así evaluar el comportamiento de la estación en comparación con las demás.

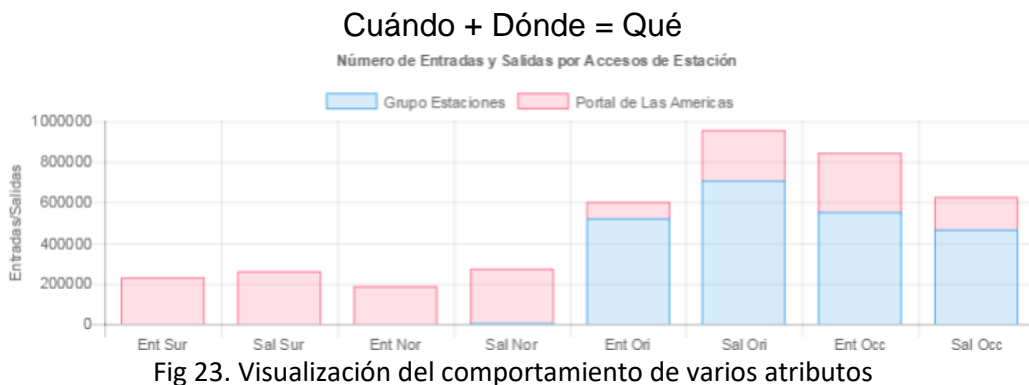
Haciendo uso del Modelo Exploratorio, se presentan tres tipos de gráficas, cada una corresponde al tipo de pregunta que se puede realizar, de esta forma se presentan las siguientes consultas:



Este tipo de gráfica permite ver los valores de los atributos seleccionados vs. la menor granularidad del tiempo que se tenga de acuerdo a los puntos espaciales seleccionados por el usuario, permitiendo ver el comportamiento completo de todos los puntos espaciales en un cada momento comparándolo contra cada punto espacial que seleccione el usuario desde el mapa focalizado.



A partir de la selección de un tiempo específico se puede visualizar como los valores del atributo influyen en cada punto espacial sobre el cual se tienen datos, permitiendo ver la relación entre cada punto espacial y el atributo de acuerdo al filtrado del tiempo que se establezca.



Esta consulta permite ver y comparar el comportamiento de cada atributo de forma individual den cada punto espacial de acuerdo a un tiempo



filtrado por el usuario, por tanto el usuario puede ver el valor de cada uno de los atributos.

## **5. PRUEBAS Y CASO DE ESTUDIO:**

Para nuestro caso de estudio se utilizaron datos espacio temporales enfocados en el área de movilidad, específicamente al acceso de las estaciones del transporte masivo de la ciudad de Bogotá, Transmilenio. Los datos corresponden a los años 2011, 2012, 2013 y 2014, donde de cada año encontramos un data set de # dimensiones compuesto por un componente espacial representado por la longitud y latitud de la estación, un componente temporal dado por un intervalo de tiempo de 24 horas, con los días del mes de mayo y septiembre, de cuatro años correspondientes a cada estación de Transmilenio. Finalmente se encuentran los atributos compuestos por cuatro variables, norte, sur, oriente y occidente, donde estas variables hacen referencia a los accesos de entradas y salidas en cada estación.

Por tanto se generó una base de datos compuesta por 3 tablas, obedeciendo al modelo propuesto, donde una tabla es la del tiempo, compuesta por el “Tiempo” con sus granularidades (id, año, meses, días, horas) con cerca de 6.500 registros, otra tabla denominada “Espacio” (id, código de la estación, nombre de la estación, latitud, longitud) constituida por 150 registros, y una última tabla que es “Atributos” (id, accesos, códigos de los torniquetes) la cual posee 1.200.000 registros.

Con los datos obtenidos y cargados se desarrollaron pruebas con la participación de quince (15) personas. Antes del experimento a los participantes se les explicó el funcionamiento de la herramienta a partir del concepto del modelo de consultas y exploración de datos. Seguidamente de la explicación, se realizó una tarea práctica con el fin de familiarizarse con la tarea y la interfaz, el ejercicio se centraba en encontrar la hora máxima de demanda de las estaciones. Esta tarea de análisis se desarrolló conjuntamente entre el participante y el expositor encontrando las horas pico en un (1) minuto. Para resolver esta tarea de análisis nos enfocamos en la formulación de la consulta por medio del (dónde + qué = cuándo), para ello se seleccionaron todas las estaciones de Transmilenio en el mapa y se escogieron todos los accesos de los atributos, generándose una gráfica de cantidad de personas VS. Las 24 horas del día, y se determinó que las horas pico son de 5:00am a 8:00am, y de 4:00pm a 7:00pm. Esta tarea de análisis también nos permite comparar la sumatoria de entradas de personas en todas las estaciones seleccionadas VS. Cada una de las estaciones que el usuario seleccione en el mapa focalizado, de esta forma se puede evidenciar que tanto influye cada estación frente sobre el total de estaciones.

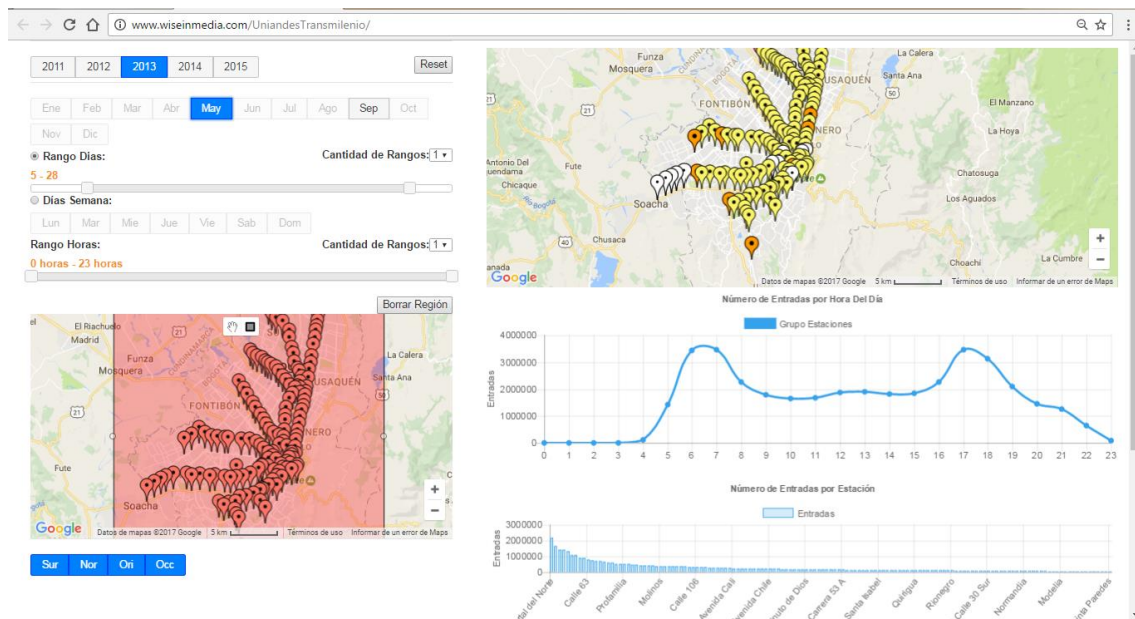


Fig 24. Visualización del acceso a las estaciones durante las 24 horas del día.

Posteriormente se les pidió que encontrarán las estaciones con mayor congestión en las horas pico, y debido al resultado anterior en donde ya habían determinado las horas pico el 73,3% de los usuarios escogieron correctamente la formulación de la consulta, la cual era (qué + cuándo = dónde), por tanto, el usuario selecciono la horas pico en el área temporal y posteriormente selecciono todos los accesos de entrada, lo cual permitió ver una gráfica de la cantidad de personas VS. Las todas las estaciones y se pudo observar varios picos, correspondientes a las estaciones Portal Norte, Portal 80, Portal Suba, Portal Américas, Portal Sur. Esta tarea de análisis apoya al proceso de evaluación de recorridos, ya que permite ver el comportamiento del ingreso de los usuarios en todas las estaciones en las horas pico, con lo cual gráficamente se evidencia que estaciones tiene mayor congestión y permitiría determinar rutas y frecuencias de las rutas. Esta tarea de análisis la lograron terminar el 73,3% de usuarios en un tiempo aproximado de 3 minutos.

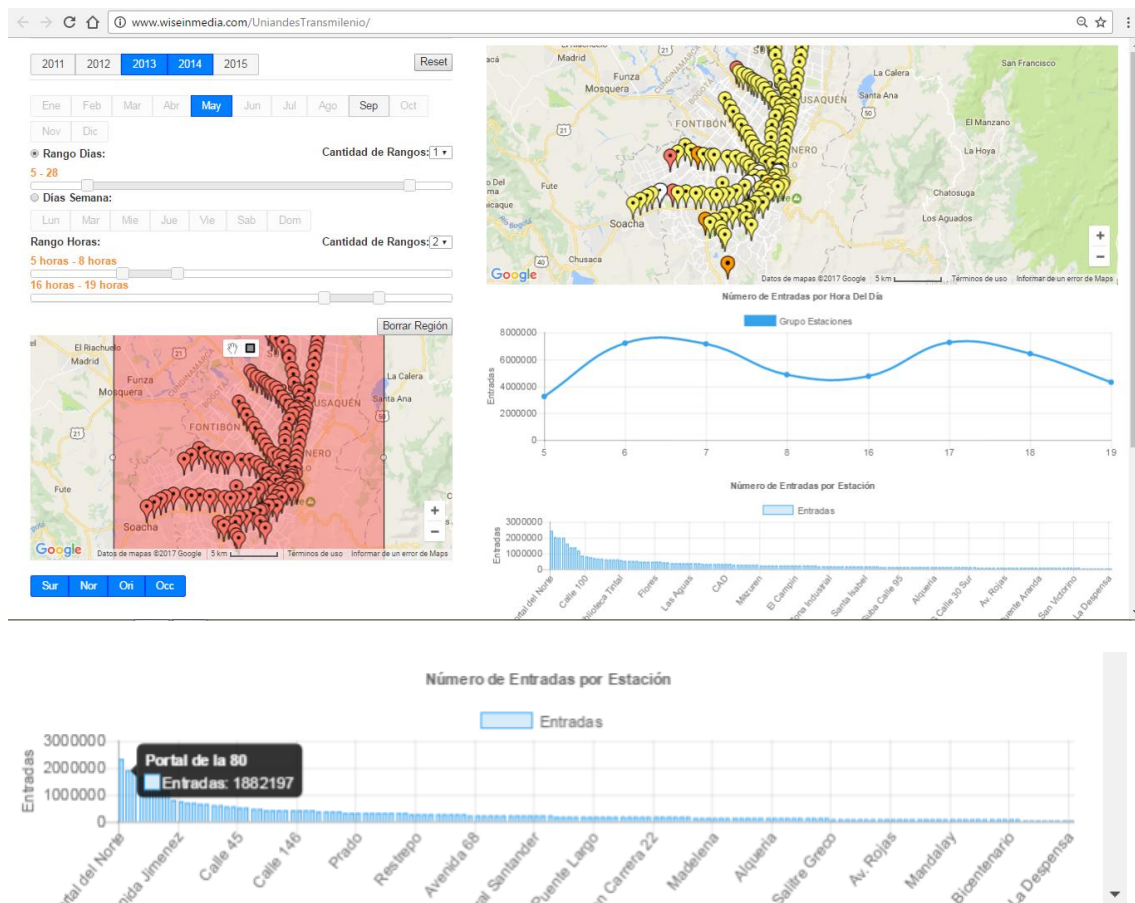


Fig 25. Visualización de los accesos a todas las estaciones de Transmilenio en horas pico.

La última tarea de análisis que se les pidió fue determinar cuántas rutas debe haber en el sistema. El objetivo de esta pregunta es evidenciar si la herramienta le permite al usuario llegar a determinar patrones o tendencias en la circulación del sistema, para así conocer la frecuencia optima de las rutas. Para ello solo el 46,6% de los usuarios escogieron correctamente la formulación de la consulta, la cual es (cuándo + dónde = qué), en donde la gráfica permite ver la cantidad de usuario que entran y salen de las estaciones seleccionadas por el usuario, en los tiempos de hora pico. Estos 7 usuarios pudieron dar hipótesis de posibles circulaciones, ya que al saber las entradas de las estaciones en una hora específica, se puede saber la salida de las personas en la siguiente hora y así empezar a determinar los patrones de circulación en el sistema.

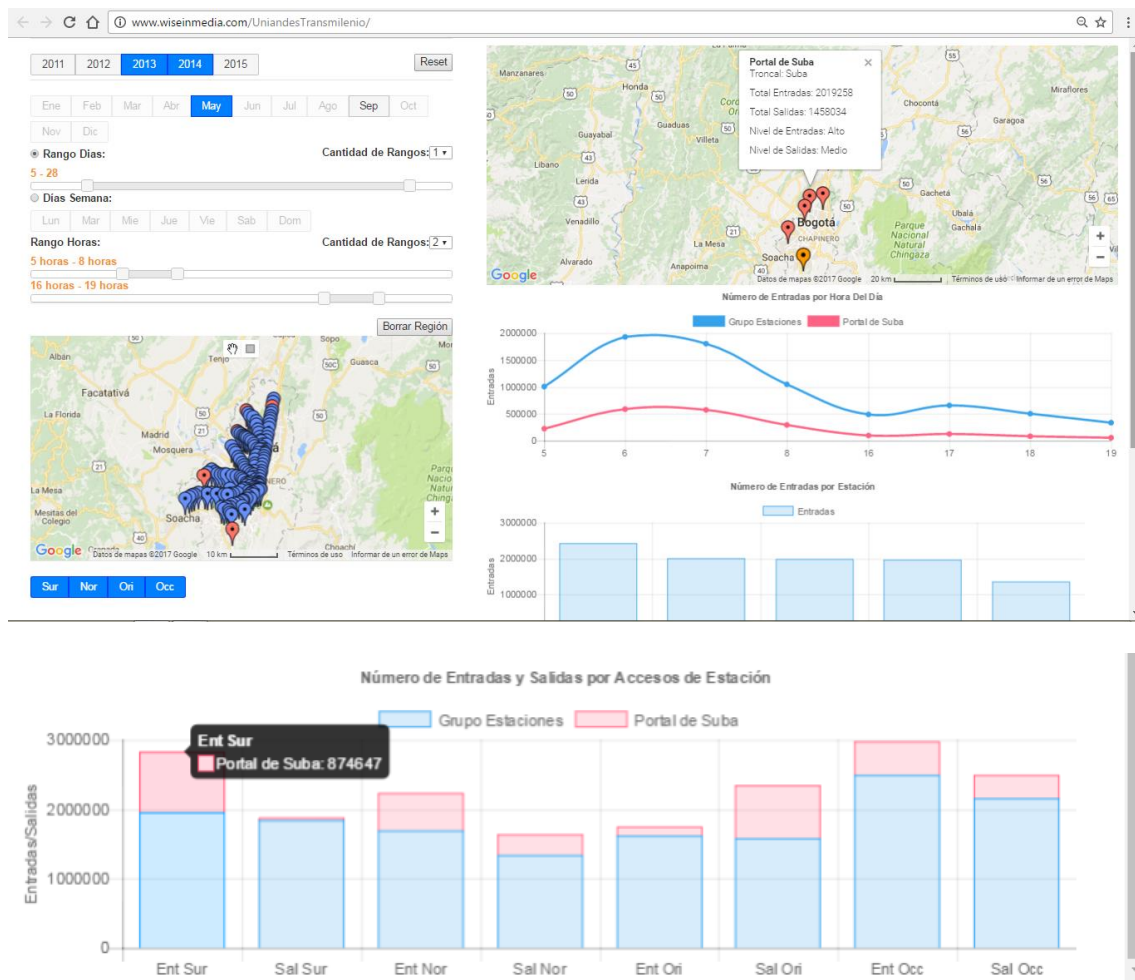


Fig 26. Visualización de los accesos a todas las estaciones más congestionadas en horas pico.

## 6. CONCLUSIONES Y TRABAJO FUTURO:

En conclusión el modelo de manejo de datos permite realizar consultas rápidas y definidas de una gran cantidad de datos, y permitió su implementación en una herramienta web que tiene un área de consultas y un área de respuestas que va acorde con las funciones planteadas por el modelo.

La herramienta web cumplió con el objetivo de implementar el modelo y permitir al usuario realizar consultas complejas y detalladas mediante una expresividad visual, a través del diseño de la visualización y el uso de controles intuitivos y fáciles de utilizar.

Así mismo la herramienta permite ver una vista general de los datos que se encuentran en la base de datos, permitiéndole al usuario evaluar cuales son los datos que posee en cada una de sus granularidades, para posteriormente realizar análisis coherentes frente a los datos que se poseen. También se confirmó que la herramienta y el modelo permiten el manejo consulta y visualización de datos espacio temporales enfocados al área de sistemas

urbanos, evidenciando su uso en el caso de estudio en donde el usuario pudo explorar, analizar y resolver tareas de análisis específicas.

Dentro del caso de estudio se pudo concluir que el tipo de consulta (Dónde + Qué = Cuándo) permite ver la cantidad de personas que entran a las estaciones seleccionadas por el usuario según la hora, para así poder determinar las horas pico y valle del sistema y poder analizar el comportamiento de las estaciones.

Por otro lado el tipo de consulta (Cuándo + Qué = Dónde) permite determinar el acceso de las personas por cada estación, a partir de la selección de fechas y horas, lo cual ayuda a evidenciar cuales son las estaciones con mayor congestión y permitiría analizar las rutas y las frecuencias de acuerdo a la cantidad de gente que ingresa por estación de acuerdo a las horas pico en las que se presentan mayor congestión.

Finalmente el tipo de consulta (Cuándo + Dónde = Qué) permite deducir la circulación determinando las entradas por estación en horas específicas y contrarrestándolas contra las horas posteriores.

Como trabajo a futuro se propone generar una interacción consecutiva entre las respuestas y preguntas posteriores, lo cual permitiría una consecución del análisis sin necesidad de estar utilizando el área de consulta todo el tiempo para realizar el filtrado de los datos. Así mismo esto permitiría generar un historial en donde se puede evidenciar todas las decisiones de consulta y sus respectivas respuestas para que el usuario pueda realizar nuevas consultas a partir de las consultas realizadas por otro experto.

## Bibliografía:

- [1] Instituto de estudios urbanos, crecimiento urbano boletín No. 23
- [2] Space, time and visual analytics Space, Gennady Andrienko\*, Natalia Andrienko, 2010.
- [3] Wolfgang Aigner, S. M. (2011). *Visualization of Time-Oriented Data*. Springer.
- [4] Integrating GIS, simulation models, and visualization in traffic impact analysis. *Computers, Environment and Urban Systems* 29:471–496, 2005.
- [5] Dynamic query visualization on World Wide Web Clients: DHTML, *International Journal of Web Engineering and Technology*, 2002.
- [6] VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery, 2012.
- [7] Max Kaiser, “EuropeanaConnect,” eContentplus programme, a multiannual Community program to make digital content in Europe more accessible, pp. 65, 1994.
- [8] Hans-Jörg Schulz, “Interactive Exploration of Multi-granularity Spatial and Temporal Datacubes: Providing Computer-Assisted Geovisualization Support”, *ACM New York, NY, USA*, Sep. 2013.
- [9] Nivan Ferreira, “Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips”, *IEEE Transactions on Visualization and Computer Graphics*. Dic 2013.
- [10] Alberto Debiasi, “Visual Analytics Tool For Urban Traffic Simulation”, *Proceedings of the 6th International ICST Conference on Simulation Tools and Techniques*, p. 51-56. Dic 2013.
- [11] G. Andrienko, “Spatio-temporal Aggregation for Visual Analysis of Movements”, *VAST '08. IEEE Symposium on*, p. 51-58. Oct 2013.
- [12] Andrienko & Andrienko, “CommonGIS visual analytics toolkit”, *ERCIM News No.41* - April 2000.
- [13] Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis, “Mastering the Information Age Solving Problems with Visual Analytics”, *Eurographics Association*. 2010.
- [14] Sara Johansson, “GeoAnalytics Visual Inquiry and Filtering Tools in Parallel Coordinates Plots”, *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. Nov 2007.
- [15] G. Andrienko, “Space, time and visual analytics”, *International Journal of Geographical Information Science*, vol. 14, Jul 2010.



[16] G. Andrienko, "Exploratory spatio-temporal visualization: an analytical review", *Journal of Visual Languages & Computing*. Dic 2003.

[17] Manuel Córdova Zamora, "Estadística Descriptiva", Dic 2014.

[18] D. Peuquet, "It's about time: Framework for the representation of temporal dynamics in geographic information systems", *Annals of the Association of American Geographers*, Vol. 84, No. 3, Sep. 1994.