

# Taxis Inteligentes

Fabian Camilo Heredia Rodríguez y Manuel Alejandro Alvarado Cobo

**Abstract**—La movilidad es un tema de gran importancia para la ciudad de Bogotá, con el paso de los años varias administraciones se han puesto en la tarea de mejorar la movilidad; otro tema de interés que se ha dado en los últimos años, es el manejo de los taxis y las compañías informales que se han creado alrededor de la prestación del servicio de movilizar a las personas, compañías como Uber y Cabify, no han legalizado su operación, pero prestan un servicio un poco más ordenado y seguro para conductores y pasajeros. La Secretaría de Movilidad ha lanzado un proyecto denominado Taxis Inteligentes, el cual consta de tres componentes: plataformas tecnológicas (apps para taxistas y usuarios), registro único de conductores y modernización de la flota. En este proyecto se trabaja con los datos generados por el componente tecnológico, datos que permitirán analizar el comportamiento del negocio y crear valor agregado.

**Index Terms**—taxis, Bogotá, movilidad, transporte, rutas, análisis de transporte, visualizaciones, transporte terrestre.

## INTRODUCCIÓN

Desde la Secretaría Distrital de Movilidad (SDM) de la Alcaldía Mayor de Bogotá se está promoviendo el uso de las tecnologías con el proyecto de Taxi Inteligente que busca “impulsar la economía y los sectores de transporte y tecnología al activar la innovación y la capacidad de producción de las empresas colombianas, por una parte, y la mejor prestación del servicio público de transporte individual, por otra” [1]. Este proyecto implica el uso de aplicaciones móviles por parte de los conductores de taxi, donde estas aplicaciones son desarrolladas por terceros y validadas por la SDM de forma tal que la información recabada y generada por éstas es enviada también a la SDM.

Con la paulatina implementación del proyecto mencionado, la SDM está interesada en evidenciar los comportamientos habituales de los taxis que hacen parte del programa en los momentos en que usan las aplicaciones para buscar y transportar pasajeros así mismo en los momentos en que no se usan.

Para esto la SDM ha suministrado un conjunto estático de datos anonimizados de trayectos realizados por taxis vinculados al programa distribuidos en varios archivos en formato Json.

A continuación, se hace la abstracción y caracterización de los datos según el framework de Tamara Munzner [2].

## 1 ESTADO DEL ARTE

La SDM, pensando en la movilidad de la ciudad además de la seguridad de pasajeros y prestadores de servicio de transporte, evaluó la movilidad de otras ciudades que tuvieran una población similar, de estas ciudades revisaron sus soluciones en temas de transporte por taxi, ciudades como Washinton D.C., Ciudad de México, Mumbai, Querétaro y Meneapolis. Entendiendo las alternativas de solución de estas ciudades se procedió a evaluar las ciudades más grandes en Colombia, viendo las tarifas que hoy se cobran y la cantidad de habitantes de las ciudades, esto llevo a la conclusión de determinar tarifas por tiempo y por distancia que determinan los valores por carrera.



Fig. 1: Solución SDM, para el análisis de indicadores

### 1.1 Alternativa de Solución de SDM

La secretaria tiene como herramienta de solución una visualización con Tableau, esta solución pretende mostrar indicadores del proyecto de taxis inteligente; la tarea principal de esta visualización es la de presentar tendencias en el servicio de taxis en la ciudad, existen otras tareas las cuales se denominan secundarias y es la de presentar la distribución de carreras de taxi a través del tiempo entre otras. Es un poco difícil interpretar la información de algunos de los gráficos, dado que la distribución, empleo de la información, marcas y canales no permite analizar de forma sencilla lo que está ocurriendo. Ver Fig. 1.

## 2 WHAT

Existen dos tipos de dataset, un dataset contiene información de algunos vehículos y el estado de estos, un estado hace referencia a un vehículo que se encuentra ocupado o disponible. Un segundo dataset, describe las solicitudes de servicio y el detalle de este si el servicio fue ejecutado completamente o si fue interrumpido antes de llegar al punto de destino.

### 2.1 Transformación de datos

Los datos originales entregados por el cliente corresponden a archivos json no válidos debido a que en realidad es una lista de objetos json, por lo tanto, se procede a reparar los archivos de forma que se puedan preprocesar para extraer los datos con los que se pretende trabajar. En el repositorio se encuentran bajo la carpeta /py los scripts utilizados para preprocesar los archivos originales y transformarlos en los datos insumo de las visualizaciones entregadas.

- Heredia R. Fabián, Estudiante de la Universidad de los Andes. E-mail: fc.heredia10@uniandes.edu.co.
- Alvarado C. Manuel, Estudiante de la Universidad de los Andes. E-mail: ma.alvarado@uniandes.edu.co
- Nemocón F. Camilo, Cliente del Proyecto. Secretaría Distrital de Movilidad. E-mail: cnemocon@movilidadbogota.gov.co.
- Guerra G. John, Docente de la Universidad de los Andes. E-mail: ja.guerrag@uniandes.edu.co.

Fecha de Publicación 4 dic. 2018

## 2.2 DataSet 1

Dataset Availability: Static

Data Type: Items, Attributes

Dataset Type: Temporal

Tabla 1: Tipos de atributo DataSet 1

Attribute	Type	Description
tarjetaControl	Categorical	Identifica el taxi
fechaHora	Temporal	fecha con hora a la que se valida el estado del vehículo
latitud	Ordered Quantitative Diverging	representa la latitud del vehículo
longitud	Ordered Quantitative Diverging	representa la longitud del vehículo
estado	Categorical	estado del vehículo al momento de la adquisición del dato: D,O; disponible, ocupado
idCarrera	Categorical	Identifica el servicio
rango_hora*	Ordered Ordinal	Clasificación en rangos de una hora de la hora de la medición del estado
origen*	Categorical	Dato binario que identifica si la medición es el punto de origen de un servicio

\* Derived fields.

## 2.3 DataSet 1.1

Dataset Availability: Dinamic (depends on filters)

Data Type: Items Attributes

Dataset Type: Table

Tabla 2: Tipos de atributo DataSet 1.1

Attribute	Type	Description
estado	Categorical	Estado del vehículo: disponible, ocupado
minutos*	Ordered Quantitative Sequential	Cantidad de minutos en un estado determinado

\* Derived fields.

## 2.4 DataSet 2

Dataset Availability: Static.

Data Type: Table -> Items -> Attributes -> links

Data and Dataset Types: Network

Dataset Types: Networks.

Tabla 3: descripción atributos DataSet 2

Attribute	type	Description
name *(node)	Categorical	nombres de las localidades de Bogotá
Node* (node)	Categorical	identificador único del nodo
region *(node)	Categorical	zona geográfica del set de datos
Source* (link)	Categorical	identificador de la localidad para el origen en la red

Target* (link)	Categorical	identificador de la localidad para el destino en la red
Value* (link)	Ordered -> Quantitative -> Sequential	cantidad de servicios del origen al destino

\* Derived fields.

## 3 WHY

Se identifican las siguientes tareas, en orden de relevancia, que se deben poder ejecutar con la visualización propuesta:

### 3.1 Tarea Principal 1

#### 3.1.1 Query - Identify - Trends

Evidenciar los momentos en los que los taxis pertenecientes al proyecto de Taxi Inteligente permanecen ocupados y disponibles. Para esto se entiende que el taxi se encuentra *ocupado* cuando se está transportando pasajeros y las aplicaciones móviles están en uso, el estado *disponible* significa que el conductor está a la espera de recibir una solicitud de servicio en la aplicación móvil o recoger un pasajero en la calle.

### 3.2 Tarea Principal 2

#### 3.2.1 Search – Browse – Outliers

Identificar las localidades de la ciudad de Bogotá, donde mas se reciben servicios y donde mas se termina los servicios de taxis.

### 3.3 Tareas Secundarias

- Identificar la cantidad de tiempo en minutos que un taxi pasa en estado activo u ocupado. (Summarize -> Features)
- Se debe procesar la ubicación de los puntos para identificar el lugar donde tenían estado ocupado o disponible (Summarize -> Spatial Data).
- Identificar las zonas donde se reciben mas servicios de taxi. (Query -> Compare -> Spatial Data)
- Comparar las zonas donde se reciben y de terminan servicios de taxi en la ciudad de Bogotá (Query -> Compare -> Spatial Data)
- Identificar en el mapa los lugares donde se produjeron los servicios. (Query -> Compare -> Spatial Data)
- Se deben procesar los datos de los dos sets para cumplir con las tareas principales y secundarias. (Analyze -> Produce -> Derive -> Features)

## 4 How

Para esta tarea siguiendo la convención del framework de Tamara Munzner.

### 4.1 Tarea Principal 1

Debido a que los datos a usar son de carácter categórico (estado), temporal, ordenado, cuantitativo y cíclico (hora) y un derivado ordenado, cuantitativo y secuencial (cantidad), se usó lo siguiente:

- Encode:
  - Estado: Map – Color, hue.
  - Hora: Arrange – Express.
  - Cantidad: Arrange – Express.
- Manipulate: Select.
- Facet: Juxtapose.
- Mark: Area.
- Channels:
  - Estado: Color, hue.
  - Hora: Position.
  - Cantidad: Size.

## 4.2 Tarea Principal 2

Debido a que los datos a usar son de carácter categórico en su mayoría resultado de los procesos de transformación del set de datos original, se usó lo siguiente:

- Encode: Separate, Order , Aling
- Manipulate: Select.
- Mark: Area.
- Channels:
  - Región: Node Color, hue.
  - Origen/Destino: Node Horizontal Position.
  - Cantidad de servicios: Node Vertical Position.
  - Cantidad de servicios: Link Size.

## 4.3 Definición de Modismos

Por lo tanto, siguiendo los principios de expresividad y efectividad, también de acuerdo a las pruebas de usuario realizadas Ver 5; se definieron los siguientes modismos:

- *En la Tarea Principal 1: el modismo elegido para esta tarea es un diagrama de barras. Para facilitar la comparación entre barras debido al uso de juxtapose se implementó una interacción que por medio de anotaciones muestra los valores exactos de todas las barras correspondientes a la misma hora.*
- *En la Tarea Principal 2: el modismo a utilizar es un sankey diagram, ya que este me permite visualizar de forma clara la distribución de los servicios y las relaciones que se tienen entre los diferentes nodos.*

## 4.4 Tareas Secundarias Asociadas

Para mostrar la cantidad de tiempo que pasan los taxis en cada estado, se implementó otro diagrama de barras, horizontales, en este caso, de forma que los datos correspondieran con el estado mostrado al mismo nivel horizontal y la codificación de los datos categórico (estado) y el derivado cuantitativo, secuencial (minutos), como sigue:

- Encode:
  - Estado: Map – Color, hue.
  - Minutos: Arrange – Express.
- Mark: Area.
- Channels:
  - Estado: Color, hue.
  - Minutos: Size.

Para establecer el lugar geografico donde un taxi se encuentra en un estado específico, implementa un mapa para georreferenciar las posiciones y se codifica de la siguiente forma:

- Encode:
  - Estado: Map – Color, hue.
  - Latitud, Longitude: Arrange – Use.
- Mark: Shape (point).
- Channels:
  - Estado: Color, hue.
  - Latitud, Longitude: Position.

Para identificar los lugares donde mas se reciben o se terminan servicios de taxi, además poder comparar las zonas de estos servicios

- Encode:
  - Origen/Destino: Map – Color, hue.
  - Latitud, Longitude: Arrange – Use.
- Mark: Shape (point or path).
- Channels:
  - Origen/Destino: Color, hue.
  - Latitud, Longitude: Position.

Para realizar el derive de los datos:

- En el caso de la sección de Actividad de Taxis se ejecuta primero el script **preprocess-bars.py** y su resultado se ordena con las columnas **tid** y **hora** (usando Excel debido a que la cantidad de datos dificulta su ordenamiento en el mismo script de preprocesamiento) para luego agregar los puntos de origen con el script **add-origen.py**.
- Para la sección de Uso de Taxis una vez corregidos los archivos json originales se aplica el script **uso.py** para generar el archivo insumo del sankey diagram.
- En ambos casos se genera un servicio de ArcGIS para el set de datos de los mapas.

## 5 EVALUACIÓN Y RESULTADOS

Una vez definidos los modismos y gracias a la caracterización de los datos, se genera una primera versión mockup y se somete a evaluación con los compañeros del curso de Visual Analytics; gracias a la retroalimentación recibida se genera un segundo mockup (Fig. 2, Fig. 3) y se diseña una prueba de usabilidad (Anexo 3) que se aplica a 10 usuarios incluido el cliente del proyecto; la recopilación de los problemas encontrados a continuación:

### 5.1 Actividad de Taxis

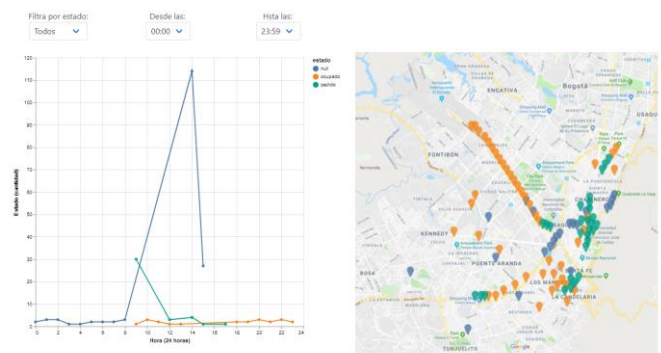


Fig. 2 - Mockup Actividad de Taxis

#### 5.1.1 Modismo 1

- Los usuarios presentan confusión respecto del número de taxis de la muestra.
- A los usuarios se les dificulta identificar que el eje x corresponde a las horas del día.

#### 5.1.2 Modismo 2

- Se dificulta la relación de los datos entre las dos gráficas.
- A los usuarios les gustaría conocer el punto de origen del servicio.

### 5.2 Uso de Taxis

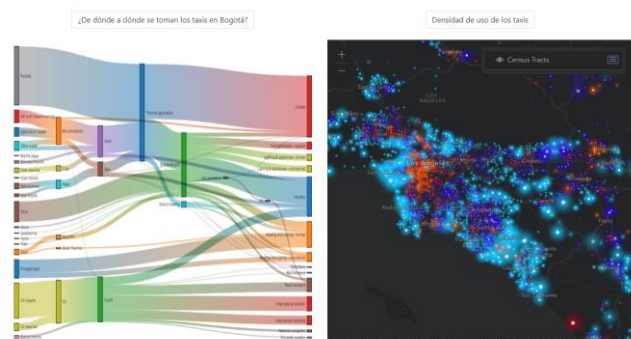


Fig. 3 - Mockup Uso de Taxis

### 5.2.1 Modismo 1

- A los usuarios se les dificulta identificar el origen y el destino.
- A los usuarios les preocupa desconocer a qué corresponde el color.

### 5.2.2 Modismo 2

- Los usuarios no identifican qué puntos corresponden a origen y a destino.
- Se considera que el uso de un heat map sea más eficiente.

Aunque los resultados de las pruebas realizadas fueron satisfactorios se obtiene retroalimentación importante por parte de la academia, los usuarios y el cliente y se realizan ajustes para finalmente generar los modismos e interacciones finales (Fig. 4, Fig. 5 y Fig. 6).

## 6 CONCLUSIONES

Gracias a la retroalimentación recibida durante la fase de pruebas se decide cambiar el modismo 1 de un diagrama de líneas a un diagrama de barras así mismo como implementar un nuevo modismo que representa la cantidad de tiempo que pasan los taxis en cada estado debido a que en los mockups iniciales no estaba contemplado.

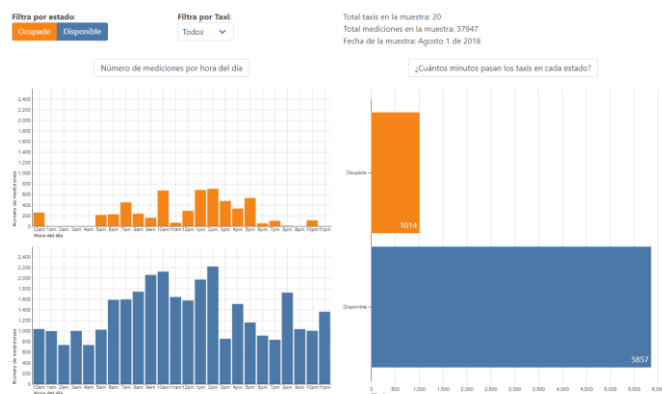


Fig. 4 - Modismos Finales 1 y 2 (Actividad de Taxis)

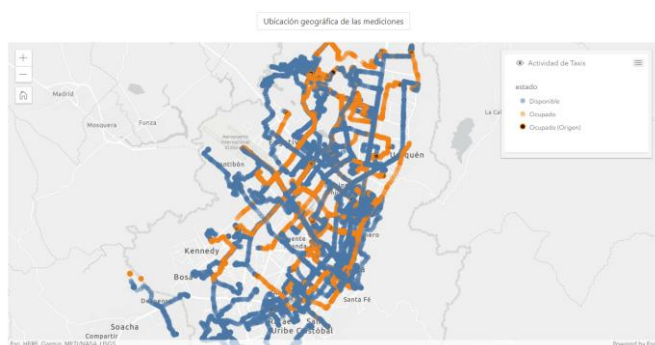


Fig. 5 - Modismo Final 3 (Actividad de Taxis)

Gracias a los modismos creados para la sección de Actividad de Taxis es posible decir que:

- Debido a la naturaleza de los datos se esperaba que el estado dominante fuera el Ocupado, al menos en ciertas horas del día, pero el estado Disponible domina por completo las mediciones.
- Los picos más altos (donde había más cantidad de taxis activos y por ende mayor número de mediciones) es entre

las 10 y 11 de la mañana y la 1 y 3 de la tarde, siendo específicamente el más alto entre las 2 y 3 de la tarde.

- A pesar de que hay taxis disponibles no se encontraron servicios en el extremo sur de la ciudad (localidades de Ciudad Bolívar, San Cristóbal y Rafael Uribe Uribe).
- Muy posiblemente algunos taxis perdieron conexión o transitaron con los sistemas fuera de línea debido a que al filtrar por taxi en algunos casos se puede notar que hay tramos de ruta que no están conectados.



Fig. 6 - Modismos Finales 1 y 2 (Uso de Taxis)

En el caso de Uso de Taxis, los modismos permiten afirmar que:

- Para el mes de abril de 2018 y mayo de 2018 la mayor cantidad de servicios generados fue en las localidades de Puente Aranda y Suba respectivamente.
- Los destinos más servicios presentaron en los meses de abril y mayo de 2018 fueron Usaquén, Fontibón y suba.
- Las zonas centro y norte son las que más servicios de taxi solicitan y donde más servicios de taxi finalizan
- Con el mapa se puede notar puntos calientes donde se podrían instalar zonas amarillas para que los taxis estén atentos a prestar el servicio.

La implementación final se encuentra disponible en [https://fabianheredia.github.io/Taxis\\_Inteligentes/](https://fabianheredia.github.io/Taxis_Inteligentes/) Github Pages y su código fuente en [https://github.com/fabianheredia/Taxis\\_Inteligentes](https://github.com/fabianheredia/Taxis_Inteligentes).

## REFERENCES

- [1] Secretaría Distrital de Movilidad. “Toda la innovación al servicio de los taxis inteligentes de Bogotá”, disponible en <http://www.movilidadbogota.gov.co/web/TODA%20LA%20INNOVACION%20C3%93N%2C%20AL%20SERVICIO%20DE%20LOS%20TAXIS%20INTELIGENTES%20DE%20BOGOT%C3%81>. 2018.
- [2] T. Munzner. Visualization Analysis and Design. 2014.

## ANEXOS

1. Mockup: [mockup.pdf](#)
2. Bitácora y cronograma: [bitácora-cronograma.xlsx](#)
3. Pruebas de usabilidad: [Pruebas Proyecto Taxis Inteligentes.pdf](#)