

# CIENCIA DE DATOS APLICADA AL SECTOR TURISMO, PERCEPCIÓN DE LOS ATRACTIVOS TURÍSTICOS EN BOGOTÁ





CLAUDIA NAYIBE LÓPEZ HERNÁNDEZ  
**Alcaldesa Mayor de Bogotá**

CAROLINA DURÁN PEÑA  
**Secretaria de Desarrollo Económico**

KAROL FAJARDO MARIÑO  
**Directora General  
Instituto Distrital de Turismo**

DANIEL VALENCIA VALENCIA  
**Asesor  
Observatorio de Turismo**

**Investigación**  
LUIS FERNANDO PINEDA GONZALEZ  
DIEGO ALEJANDRO RODRÍGUEZ GUERRERO

**Contacto**  
Instituto Distrital de Turismo  
<http://www.idt.gov.co>  
Bogotá D.C., 2021

## CONTENIDO

1. INTRODUCCIÓN .....	3
2. ANTECEDENTES DE CIENCIA DE DATOS APLICADA AL SECTOR TURISMO .....	5
3. METODOLOGÍA DE RECOLECCIÓN DE LOS DATOS .....	8
4. RESULTADOS .....	11
4.1 Resultados a nivel de Localidad .....	21
4.2 Resultados a nivel de Atractivos .....	35
5. CONCLUSIONES.....	42
INDICE DE TABLAS, FIGURAS Y GRÁFICOS .....	44
BIBLIOGRAFIA .....	46

## 1. INTRODUCCIÓN

El Instituto Distrital de Turismo (IDT), a través del Observatorio de Turismo de Bogotá (OTB), consolida, sistematiza, analiza, investiga y divulga la información del sector turístico de Bogotá. A través de un equipo de encuestadores se captura la información primaria relacionada tanto con la oferta como con la demanda y adicionalmente se complementa la información sectorial a través de alianzas con aliados estratégicos como OPAIN, Terminal de Transportes, La ANI, las agremiaciones del sector, entre otros. Sin embargo, en algunos casos esta información es limitada, por lo anterior, se busca aprovechar los datos no convencionales para generar análisis holísticos del sector.

Es así que, técnicas como: "Web Scraping" (también llamada "extracción de datos web" o "minería de datos web"), el Procesamiento de Lenguajes Naturales (NPL por sus siglas en inglés) y Análisis de Sentimiento juegan un papel importante. Estas técnicas se pueden aplicar, a las distintas atracciones turísticas y/o establecimientos de alojamiento de la ciudad, con el fin de determinar las características, condiciones de la oferta, falencias, fortalezas e incluso descubrir los patrones de movilidad de los turistas en la ciudad.

Utilizando la técnica de "Web Scraping", desde el OTB, se logró extraer la información de 50.000 reseñas para 185 atractivos de la ciudad de Bogotá, con el fin de conocer la percepción de los turistas sobre la estructura turística de la ciudad, a modo de identificar falencias y fortalezas, que a su vez ayuden a determinar acciones de mejora de los atractivos turísticos de la ciudad.

Este estudio, en el cual se utiliza la Ciencia de Datos aplicada al sector turismo, corresponde a una primera aproximación utilizando los desarrollos

tecnológicos que se han generado en las últimas décadas, con el cual se espera poder entender mejor la percepción que tienen los distintos turistas visitantes de los atractivos turísticos en Bogotá.

Se espera seguir utilizando nuevas metodologías de Big data y Ciencia de Datos para lograr obtener más información y de mejor calidad, lo cual se verá reflejado en mejores tomas de decisiones.

## 2. ANTECEDENTES DE CIENCIA DE DATOS APLICADA AL SECTOR TURISMO

### **Antecedentes de Ciencia de Datos aplicada al Sector Turismo**

Teniendo como precedente la literatura existente sobre el uso de Big Data en la investigación Turística y sobre Ciencia de Datos aplicada al Sector Turismo, se evidencia que estos temas no son nuevos, ya que se puede encontrar una gran cantidad de trabajos y artículos científicos en los cuales se ha aplicado una variedad de técnicas a la investigación turística.

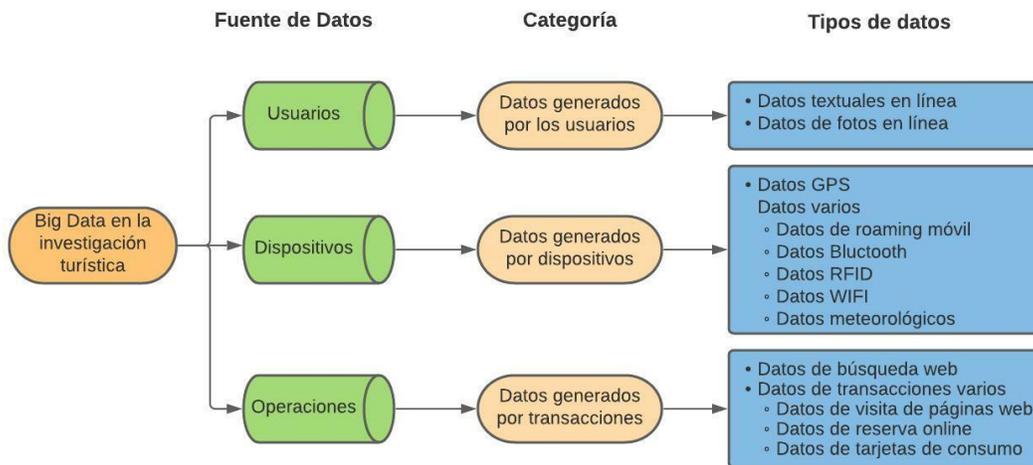
Li et al. (2018) presenta una revisión exhaustiva de la literatura sobre los diferentes tipos de Big Data en la investigación turística, los cuales generan un marco analítico (figura 1) de referencia que se clasifica en tres categorías de acuerdo con el origen de los datos, a saber: (1) Datos generados por usuarios: datos textuales y datos fotográficos; (2) Datos generados por dispositivos electrónicos los cuales son recopilados de forma pasiva a través de datos de GPS, datos de itinerancia móvil, datos de Bluetooth, datos de RFID<sup>1</sup>, datos de WIFI, entre otras: y (3) Datos generados por transacciones que registran las operaciones relacionadas por el usuario: datos de búsqueda web, datos de visita a la página web, datos de reserva en línea y datos de las transacciones por tarjetas débito y crédito.

Como se ilustra en la Figura 1 y de acuerdo con el origen de los datos usados en el presente documento, correspondiente a datos textuales en línea (reseñas en la plataforma de TripAdvisor), los cuales son generados por los turistas y/o visitantes (usuarios) que recorren al atractivo Turístico, se procede a formular el marco analítico de este estudio.

---

<sup>1</sup> Identificación por radiofrecuencia.

**Figura 1. Big Data en la investigación turística**



Fuente: Adaptado al español de Li et al. (2018)

En la revisión de literatura respecto a trabajos y artículos donde se emplearon datos textuales en línea (reseñas), se pudo identificar que TripAdvisor es la principal plataforma de donde se recopilan los datos para los estudios seleccionados (ver Tabla 1). Esto debido a que a nivel mundial esta plataforma es uno de los mayores repositorios de reseñas relacionadas con viajes y turismo.

También se pudo determinar que la mayoría de estudios en el sector turístico utilizan reseñas asociadas principalmente a hoteles, restaurantes y atractivos turísticos para distintas ciudades, regiones o países completos. De igual manera se encontró que la metodología predominante en la recopilación de los datos es el "Web Scraping", por encima de la adquisición directa de la base de datos de proveedores especializados. Esto debido a la facilidad que ofrecen este tipo de técnicas para recopilar grandes volúmenes de información de las páginas web.

**Tabla 1. Trabajos en el que se utilizan las reseñas de los usuarios para el análisis los destinos Turísticos.**

<b>Estudio</b>	<b>Fuente de los Datos</b>	<b>Área Geográfica</b>	<b>Método Extracción de los Datos</b>	<b>Tipo de reseñas</b>	<b>Número Total establecimientos o de Reseñas<sup>2</sup></b>
Flôres Limberger et al. (2014)	TripAdvisor	Brasil	No Disponible	Hoteles	660 reseñas
Berezina et al. (2015)	TripAdvisor	Saratoga (Estados Unidos)	Online robot developed	Hoteles	2.510 reseñas
Fang et al. (2015)	TripAdvisor	New Orleans (Estados Unidos)	No Disponible	Atractivos Turísticos	41.061 reseñas para 106 Atractivos
Pearce & Wu. (2015)	TripAdvisor	Yangshuo (China)	No Disponible	Atractivo Turístico (Impression Sanjie Liu)	573 reseñas
Dhiratara et al. (2016)	Instagram	Paris (Francia)	Web scraping	Atractivos Turísticos	82.381 publicaciones
Valdivia et al. (2017)	TripAdvisor	España	No Disponible	Tres Atractivos Turísticos <sup>3</sup>	45.303 reseñas
Moro et al. (2018)	TripAdvisor	Portugal	Web scraping	Hoteles	967 establecimientos
Almeida & Abrantes (2018)	TripAdvisor	Minas Gerais (Brasil)	Web scraping	Atractivos Turísticos	1.384 atractivos
Agüero-Torales et al. (2019)	TripAdvisor	Provincia de Granada (España)	Web scraping	Restaurantes	33.594 reseñas
Adhinugroho et al. (2020)	Agoda y Pegipegi	Indonesia	Web scraping	Hoteles	19.834 establecimientos

Fuente: Elaboración propia.

<sup>2</sup> Se maneja esta clasificación por número de establecimientos o número de reseñas, ya que no se cuenta con una información homogénea en este campo entre todos los artículos seleccionados.

<sup>3</sup> La Alhambra, Sagrada Familia y La mezquita de Córdoba.

### 3. METODOLOGÍA DE RECOLECCIÓN DE LOS DATOS

La metodología utilizada para la recolección de los datos es de la “Web Scraping” o “raspado de la web”, también llamado “recolección web”, “extracción de datos web” o incluso “minería de datos web”), la cual se puede definir como “la construcción de un agente para descargar, analizar y organizar datos de la web de manera automatizada”. (Vanden Broucke & Baesens B., 2018).

Para poder extraer los datos utilizando la técnica de “Web Scraping”, es necesario conocer la estructura de la página web, que normalmente está compuesta por código HTML<sup>4</sup> y que varía de acuerdo con el tipo de los establecimientos de interés. Por ello, inicialmente se conformó un listado de atractivos turísticos que posee Bogotá de acuerdo con el inventario entregado por el Viceministerio de Turismo en junio 2020, identificando 430 atractivos que se pueden clasificar dentro de una de las siguientes categorías: Aguas lénticas, Aguas lólicas, Áreas protegidas, Formaciones cársicas, Montañas, Patrimonio cultural inmaterial, Patrimonio cultural material inmueble y Patrimonio Cultural Mueble.

Posterior a la recolección del marco de atractivos, se buscó en TripAdvisor<sup>5</sup> la URL para cada una de las páginas web de los 430 atractivos de Bogotá, donde se encontraron 195 atractivos de los 430 del marco inicial con un sitio web asociado en la plataforma y con reseñas web disponibles, adicionalmente, en este ejercicio se identificó que son muy pocos los atractivos de tipo Patrimonio cultural inmaterial y Patrimonio Cultural Mueble que cuentan con un espacio web dentro de TripAdvisor.

---

<sup>4</sup> HyperText Markup Language (Lenguaje de Marcas de Hipertexto en español), es el componente más básico de la Web.

<sup>5</sup> <https://www.tripadvisor.co/>

Para cada uno de estos 195 atractivos con cuenta en TripAdvisor, se analizó la estructura HTML del sitio web (diagramación y navegación) y luego se raspa la información de todas las reseñas. Finalmente, la información recolectada para cada atractivo se registra en una base de datos.

En este proceso se identificaron 10 atractivos que tenían páginas web en la plataforma de TripAdvisor pero al momento del “Web Scraping” no contaban con ninguna reseña, adicionalmente, todas las reseñas descargadas corresponden a aquellas que se escribieron en español, conformando finalmente una base de datos con 49.955 reseñas para 185 atractivos hasta agosto de 2021.

Dados los intereses de análisis, para este estudio se consideran las calificaciones de los usuarios con un valor de uno y dos como negativas, tres como neutrales y cuatro y cinco como positivas.

Como la naturaleza de la metodología es programar para que la interacción y el raspado de la información se realice de manera automática, una de las preguntas más frecuentes sobre la metodología de “Web Scraping” es respecto a las cuestiones de legalidad<sup>6</sup>.

Para el total de 49.955 reseñas recopiladas de los 185 atractivos se generó una base de datos que manejó una estructura que se presenta en la Tabla 2, donde se tienen dos grupos de información: 1. La que corresponde al atractivo: Nombre, Calificación General (promedio del total de

---

<sup>6</sup> Como resultado de la investigación, se encuentra que esta acción es completamente legal ya que recuperar datos públicos de algún sitio web específico de Internet sin vulnerar los filtros de seguridad de la página web, por lo que se considera una acción legal. En el caso de TripAdvisor, los datos disponibles en el sitio web son datos públicos y no contiene información privada o sensible de los usuarios.

calificaciones), Cantidad de Reseñas<sup>7</sup> y las Palabras clave asociadas al atractivo. 2. La información de la reseña que deja un usuario: título de la reseña, texto de la reseña, calificación de la reseña, fecha de la reseña, fecha de la experiencia (día de la visita al atractivo), ciudad y país de residencia del visitante y el grupo de viaje.

**Tabla 2. Estructura de los datos que se descargaron**

<b>Para el atractivo</b>	
<b>Variable</b>	<b>Explicación</b>
Nombre Atractivo	Contiene el Nombre del Atractivo
Calificación General	Muestra la calificación general del atractivo
Cantidad de Reseñas	Muestra la cantidad de reseñas que tiene el atractivo
Palabras clave	Muestra las palabras claves asociadas al atractivo
<b>Para las Reseñas</b>	
<b>Variable</b>	<b>Explicación</b>
Título Reseña	El título que el visitante del atractivo escribió para la reseña
Texto Reseña	El texto completo de la reseña que escribió el visitante del atractivo
Calificación Reseña	La calificación del atractivo que reporto el visitante
Fecha de reseña	Fecha en la cual el visitante escribe la reseña
Fecha de experiencia	Fecha en la cual el visitante visito el atractivo
Ciudad	Ciudad y país de residencia del visitante
Grupo de Viaje	Grupo de viaje o personas con las que visito al atractivo el visitante

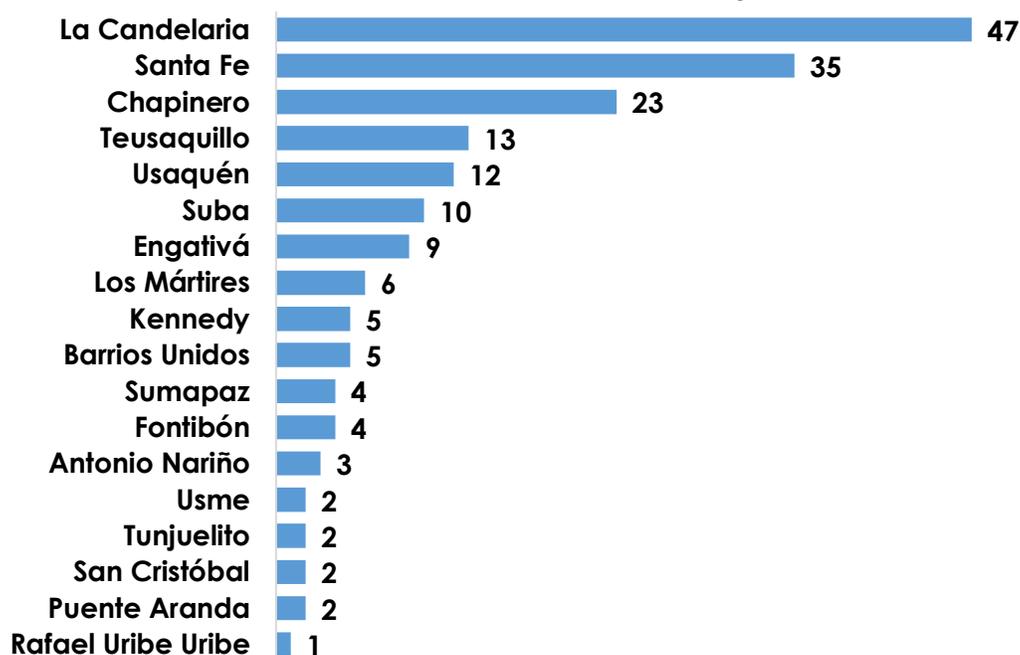
<sup>7</sup> La cantidad de reseñas que se reportan para esta variable tienen en cuenta las reseñas para el atractivo en todos los idiomas, pero para el ejercicio de este estudio solo se recolectaron las reseña en español.

## 4. RESULTADOS

En esta sección de resultados solo se tendrá en cuenta la información de los 185 atractivos turísticos de Bogotá (que se encuentran en el inventario del Viceministerio de Turismo – Ministerio de Comercio, Industria y Turismo) que tienen reseñas en la plataforma de Tripadvisor<sup>8</sup>.

La ciudad de Bogotá está conformada por 20 localidades y como se observa en la gráfica 1, en 18 de ellas se encuentran los atractivos estudiados, también se identifica que la localidad de La Candelaria es la que mayor cantidad de atractivos tiene con 47, seguida de la localidad de Santa Fe con 35 y Chapinero con 23. Entre estas tres localidades se encuentran el 56,7% de los atractivos usados en el estudio.

**Gráfica 1. Número de atractivos turísticos por localidad**

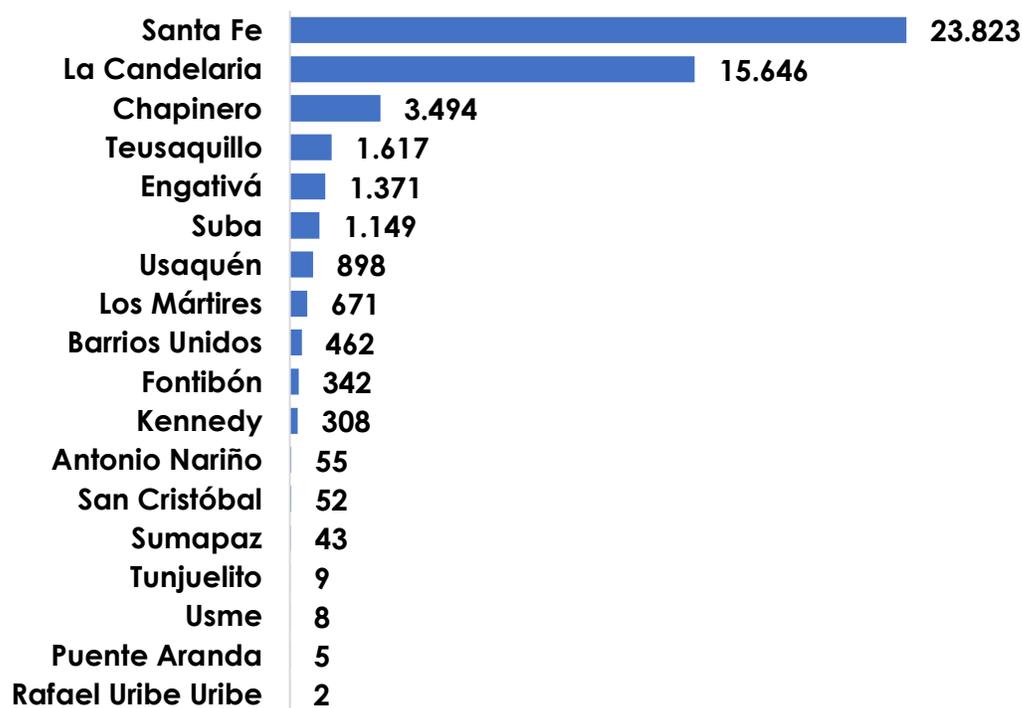


Fuente: TripAdvisor 2021

<sup>8</sup> El sitio web estadounidense de reseñas Trip Advisor fue fundado en el año 2000, pero es hasta diciembre del 2006 que se realiza la primera reseña para uno de los atractivos del estudio. Sin embargo, solo hasta 2012 la plataforma se vuelve popular entre los turistas y visitantes de los atractivos en Bogotá y se comienza a ver más de 200 reseñas por mes.

Aunque La Candelaria es la localidad con más atractivos, no es la localidad que agrega más reseñas, siendo la localidad de Santa Fe la que mayor número de reseñas posee, principalmente debido a que los dos atractivos más reseñados están ubicados en esta localidad (El Cerro de Monserrate y El Museo del Oro). Es así como, la localidad de Santa Fe representa el 47,6% de las reseñas recopiladas para el estudio.

**Gráfica 2. Número de reseñas por localidad**

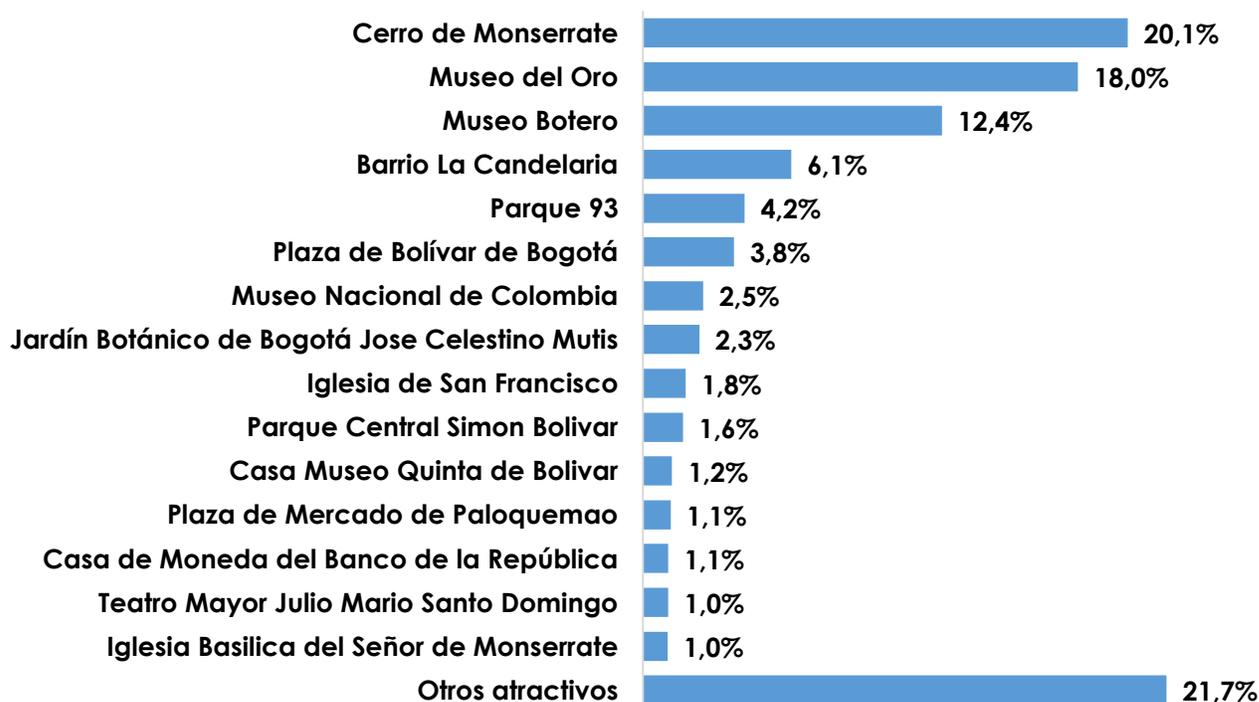


Fuente: TripAdvisor 2021

Como se mencionó anteriormente los dos atractivos con más reseñas en la plataforma de Trip Advisor para Bogotá son: El Cerro de Monserrate y El Museo del Oro, los cuales junto con el Museo Botero suman un total de 25.210 y representan el 50,4% del total de reseñas recopiladas en este estudio, también se pudo identificar que 170 atractivos que tienen menos de 500 reseñas representan el 21,7%. En esta primera caracterización de la base de datos por localidad y por atractivo se encuentra que 3 localidades y 3

atractivos concentran aproximadamente el 50% de las reseñas, lo cual da indicios sobre las preferencias de los turistas y visitantes en la ciudad.

**Gráfica 3. Atractivos con el mayor porcentaje de reseñas**

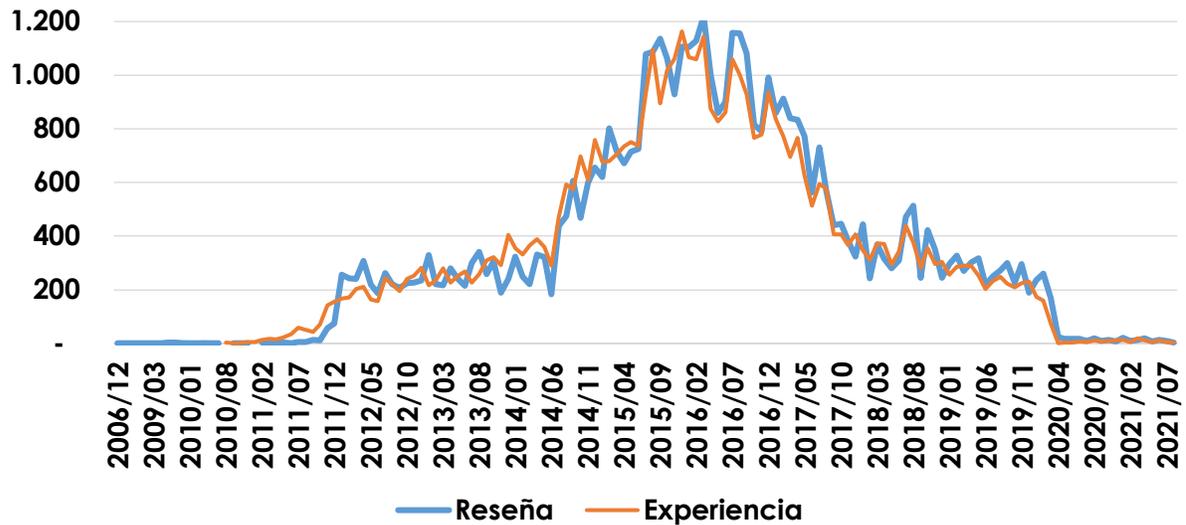


Fuente: TripAdvisor 2021

Al observar cómo ha sido el comportamiento de las reseñas de los atractivos de Bogotá durante el tiempo, se encuentra que para el periodo entre el 2015 y el 2016 es en donde la plataforma alcanza mayor popularidad y se identificaron más de 1000 reseñas por mes.

También, se identifica el efecto de la pandemia, ya que, para el mes de marzo del 2020 con la caída del turismo internacional y doméstico debido a las restricciones de movilidad, caen por completo las reseñas sobre los atractivos de Bogotá, y de esta caída aún no se han podido recuperar.

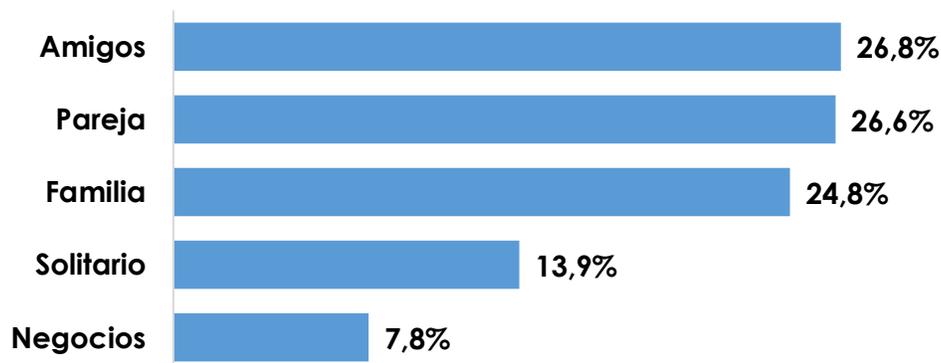
**Gráfica 4. Número de reseñas por mes (2006-2021)**



Fuente: TripAdvisor 2021

Al mirar el grupo de viaje o personas con las que visita el atractivo el turista o visitante, principalmente se encuentran las visitas con amigos en 26,8% de las reseñas y con la pareja 26,6%. Esto seguido de los visitantes que asisten con la familia 24,8%. Es importante resaltar que en la plataforma de Trip Advisor la información sobre el motivo por negocios se encuentra dentro de la información de grupo de viaje.

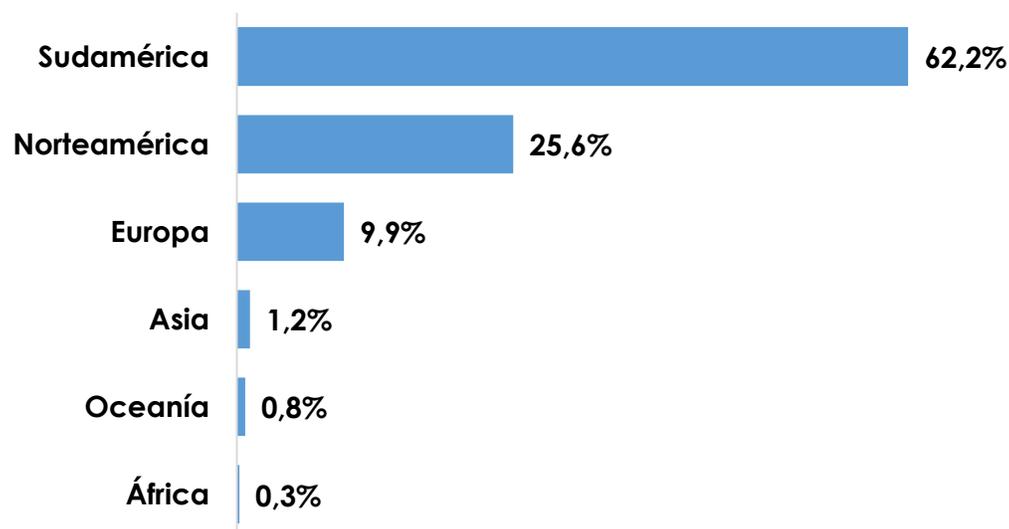
**Gráfica 5. Grupo de Viaje de los visitantes a los atractivos turísticos**



Fuente: TripAdvisor 2021

Respecto al continente de origen se observa que los turistas y visitantes que escriben las reseñas en su mayoría son de Sudamérica (62,2%), esto debido principalmente a los visitantes de los atractivos que residen en distintas ciudades y/o departamentos de Colombia, incluso también de residentes capitalinos, seguido por los turistas que provienen de Norteamérica<sup>9</sup> con un 25,6%, principalmente turistas de Estados Unidos o de México, países con los cuales Bogotá tiene varias rutas, frecuencia alta de vuelos y hacen parte de los países mayores emisores de turistas hacia Bogotá según la Investigación viajeros en Bogotá.

**Gráfica 6. Continente de origen de los visitantes a los atractivos turísticos**



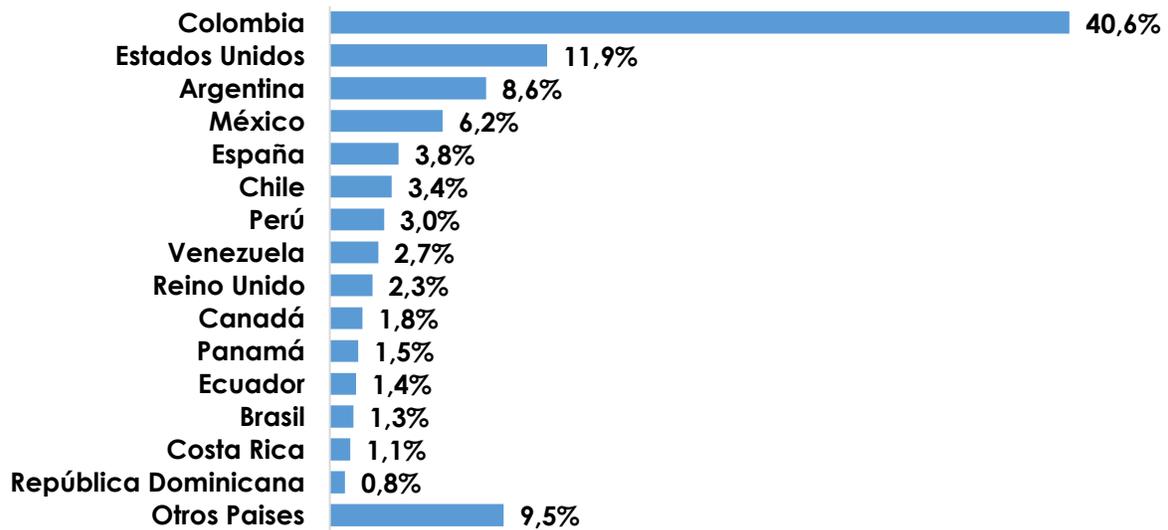
Fuente: TripAdvisor 2021

Siguiendo con el análisis sobre el origen de los turistas y visitantes de los atractivos de Bogotá, se identificó que un 40,6% son de origen nacional, seguido de turistas de Estados Unidos 11,9% y Argentina 8,6%. También se pudo observar que los quince primeros países de origen de los visitantes (incluido Colombia) recogen el 90,5% del total de reseñas recolectadas

<sup>9</sup> Esta categoría Incluye países de Norteamérica y Centro América.

para el estudio, siendo España y el Reino Unido los únicos países que aparecen en este listado por parte del continente Europeo.

**Gráfica 7. Principales países de origen de los visitantes a los atractivos turísticos**

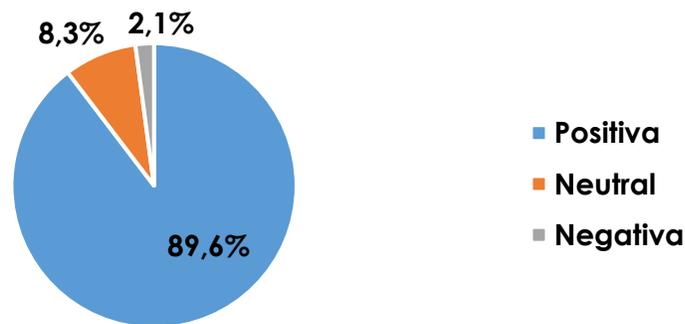


Fuente: TripAdvisor 2021

En el presente estudio se consideran las calificaciones de los usuarios a cada atractivo, las cuales asociadas a las reseñas permiten categorizar los comentarios de la siguiente manera: valores menores o iguales a dos como reseñas negativas, valores de tres como reseñas neutras y valores mayores como reseñas positivas.

Al evaluar cómo se distribuyen las reseñas según esta clasificación se encuentra que el 89,6% son reseñas positivas, 8,3% neutras y solamente 2,1% negativas, lo cual indica que la base de datos se encuentra desbalanceada, sin embargo, esta conclusión es congruente, en el sentido que TripAdvisor fue creada con el objetivo de resaltar la popularidad de los sitios que tienen presencia dentro de la plataforma.

**Gráfica 8. Clasificación de las reseñas según su tipo**



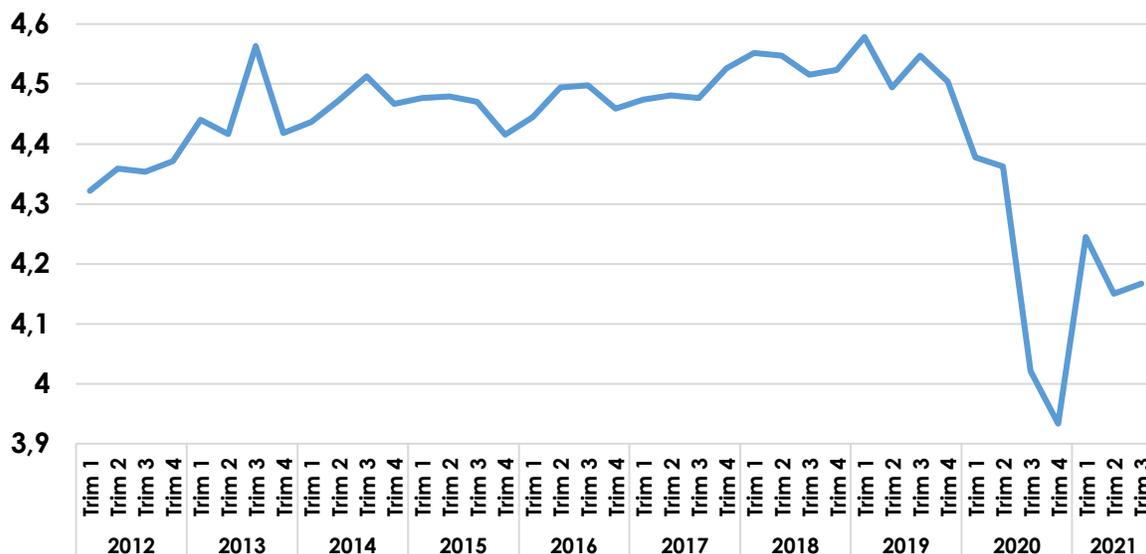
Fuente: TripAdvisor 2021

Como se había mencionado anteriormente, el año 2012 fue el momento donde la plataforma de TripAdvisor se volvió popular entre los turistas y visitantes de los atractivos en Bogotá, razón por la cual se observó un aumento en el número de reseñas por mes. Teniendo en cuenta esta información y calculando la calificación promedio por trimestre se genera la gráfica 9.

Se puede observar que desde 2012 hasta el primer trimestre del 2020 la calificación promedio de las reseñas siempre estuvo por encima de 4,3, mientras que, para el segundo trimestre del 2020, por efectos de la pandemia como confinamientos y el cierre de oferta turística y cultural en la ciudad, la calificación promedio de las reseñas cae ubicándose por debajo de 4.

Por otro lado, se puede observar que desde el 2021, como consecuencia de los distintos programas de reactivación económica y del sector turismo en la ciudad, la calificación promedio de las reseñas se ha estado recuperando, aun así, aún no se recupera los niveles alcanzados pre-pandemia.

**Gráfica 9. Calificación promedio de las reseñas por trimestre (2012-2021)**



Fuente: TripAdvisor 2021

Finalmente, a nivel de Bogotá se realizó un análisis del texto de las reseñas, utilizando técnicas de Procesamiento de Lenguaje Natural (NLP), donde se identifican las palabras asociadas a las reseñas positivas y negativas para dos periodos de tiempo, el primero desde 2006<sup>10</sup> hasta 2019 y el segundo para el 2020 y 2021.

Este conjunto de palabras identificadas para ambos periodos y para los dos tipos de reseñas se ilustran en 4 nubes de palabras (Figura 2). Desde las reseñas positivas se logró identificar que las palabras “restaurante”, “comer” y “café” pierden relevancia para el 2020 y 2021, principalmente por los métodos de comercialización y restricciones de aforo producto de la pandemia, además de la reorganización de los gastos de los hogares. También palabras como “vestir”, “pasear”, “colección” y “conoce” perdieron protagonismo.

<sup>10</sup> Año donde se identificó la primera reseña a un atractivo.

De otro modo, se observa como las palabras “museo”, “plan”, “opción” e “historia” pasan a tener un mayor protagonismo dentro de las reseñas positivas, lo cual tiene bastante sentido ya que los museos se reabrieron rápidamente y mucho antes que otros atractivos de la ciudad, así como la búsqueda de nuevas opciones y planes que paso a ser un tema de discusión y que corresponden con las nuevas tendencias en turismo.

Por el contrario, dentro de las reseñas negativas las palabras “atención” y “servicio” pierden protagonismo en los años de la pandemia, posiblemente asociado a las medidas de aforo reducido que permite ofrecer una mejor atención a los turistas y visitantes, mientras que las palabras “experiencia”, “fila” e “ingresar” presentan una mayor relevancia, esto asociado a las nuevas problemáticas que encuentran los visitantes en su ingreso de los atractivos.

Así mismo, aparecen nuevas palabras como “experiencia”, “administración”, “alcanzar”, “planificar”, “organizar”, “natural” y “asador” que inducen a pensar que en el cambio hacia nuevas experiencias más naturales se generaron problemáticas por falta de planificación.



## 4.1 Resultados a nivel de Localidad

En esta sección se presenta un análisis más detallado a nivel de localidad, donde se contemplan aquellas localidades con suficientes reseñas para hacer análisis, es decir, Engativá, La Candelaria, Chapinero, Santa Fe, Suba, Teusaquillo y Usaquén. Estas 7 localidades agregan un total de 47.998 reseñas, equivalente al 96% del total de reseñas recolectadas del ejercicio de "Web Scraping" que se ejecutó para los atractivos turísticos de la ciudad.

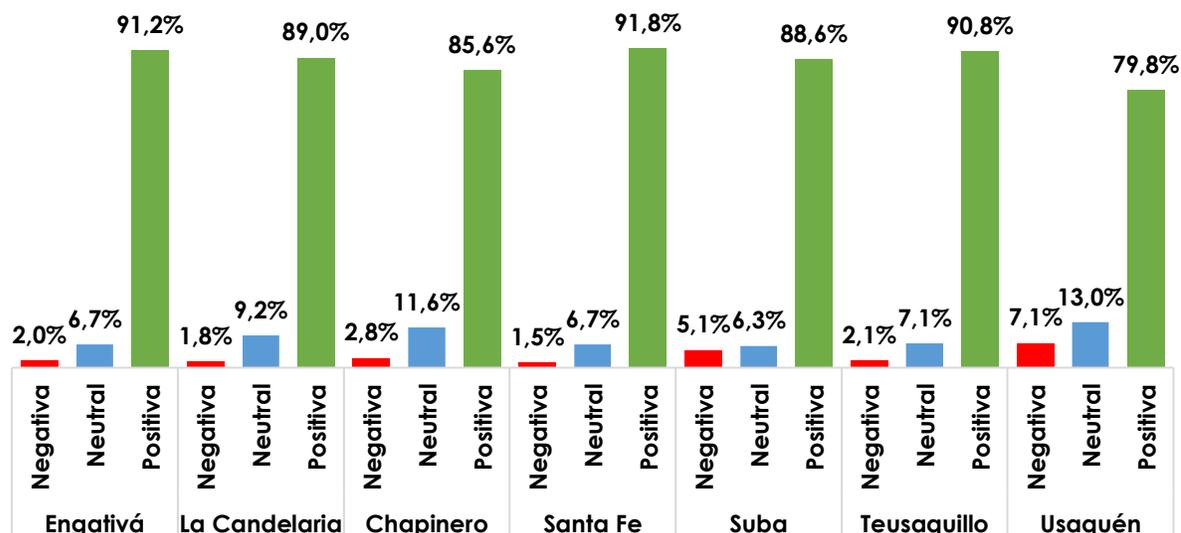
Como se había mencionado anteriormente, se categorizaron las reseñas de acuerdo con la calificación otorgada por el usuario, donde los puntajes menores o iguales a dos se clasifican como reseñas negativas, valores de tres corresponden con reseñas neutrales y puntajes mayores a 3 como reseñas positivas.

Al revisar la relación entre reseñas positivas y negativas para las localidades seleccionadas se observa que, para Engativá, Santa Fe y Teusaquillo el porcentaje de reseñas positivas representan más del 90% del total de reseñas que tiene en cada localidad.

En el segundo grupo de localidades se encuentra La Candelaria, Chapinero y Suba para las cuales el porcentaje de reseñas positivas se encuentra entre el 80% y 90%.

Por último se encuentra la localidad de Usaquén para la cual se observa un porcentaje de reseñas positivas cercana al 80% y el porcentaje más alto de reseñas negativas 7,1%, seguido de la localidad de Suba con un 5,1% de reseñas negativas.

**Gráfica 10. Clasificación de las reseñas según su tipo desagregado por localidad**



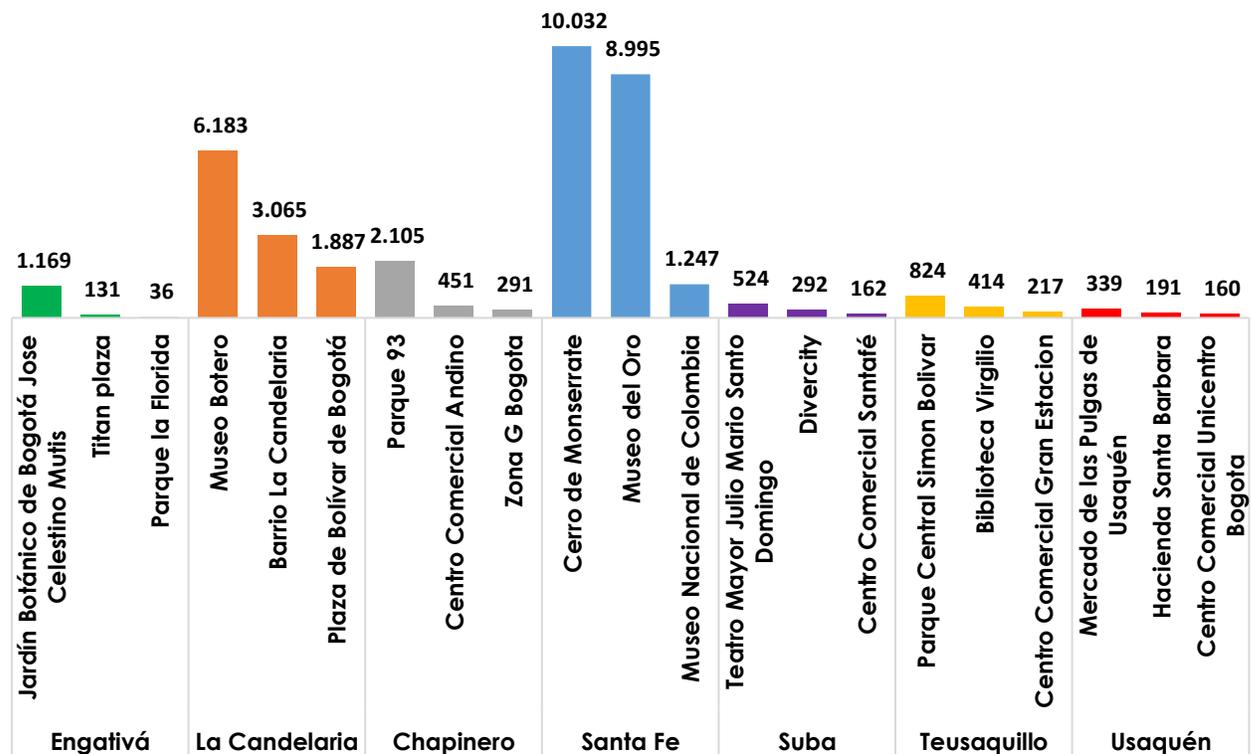
Fuente: TripAdvisor 2021

Continuando con el análisis, la gráfica 11 presenta los 3 atractivos con el mayor número de reseñas para cada una de las siete localidades analizadas, observando como el Cerro de Monserrate y el Museo del Oro, ambos atractivos en la Localidad de Santa Fe, son los más reseñados. Al mirar para cada localidad, en Engativá el atractivo con más reseñas es el Jardín Botánico de Bogotá (1.169 reseñas), para la Candelaria es el Museo Botero (6.183), en Chapinero es el Parque de la 93 (2.105 reseñas), Suba es el Teatro Mayor Mario Julio Santo Domingo (1.247 reseñas), para Teusaquillo es el Parque Simón Bolívar (824 reseñas) y Finalmente para Usaquén es el Mercado de las Pulgas con 339 reseñas y que funciona hace más de 30 años.

También es importante resaltar que en 5 de las 7 localidades, dentro de los tres atractivos con más reseñas se encuentra al menos un centro comercial, particularmente el Centro Comercial Titan Plaza (Engativá), el Centro Comercial Andino (Chapinero), el Centro Comercial Santafé (Suba), Centro

Comercial Gran Estación (Teusaquillo) y en Usaquén se tienen el Centro Comercial Hacienda Santa Bárbara y el Centro Comercial Unicentro Bogotá, mostrando la importancia que tienen este tipo de atractivos dentro de los visitantes de origen local, nacional e internacional que escriben reseñas en Trip Advisor.

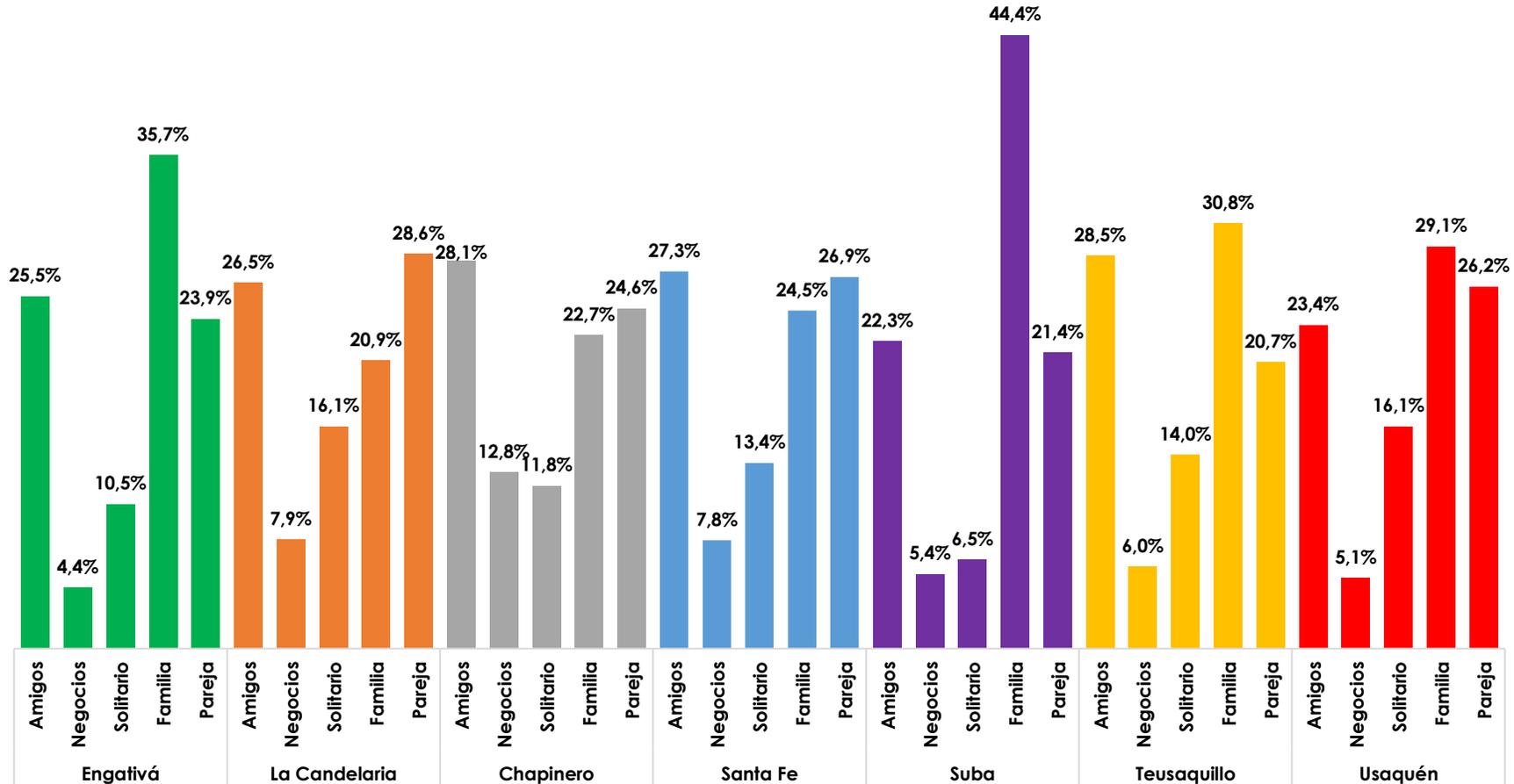
**Gráfica 11. Atractivos con el mayor número de reseñas desagregado por localidad**



Fuente: TripAdvisor 2021

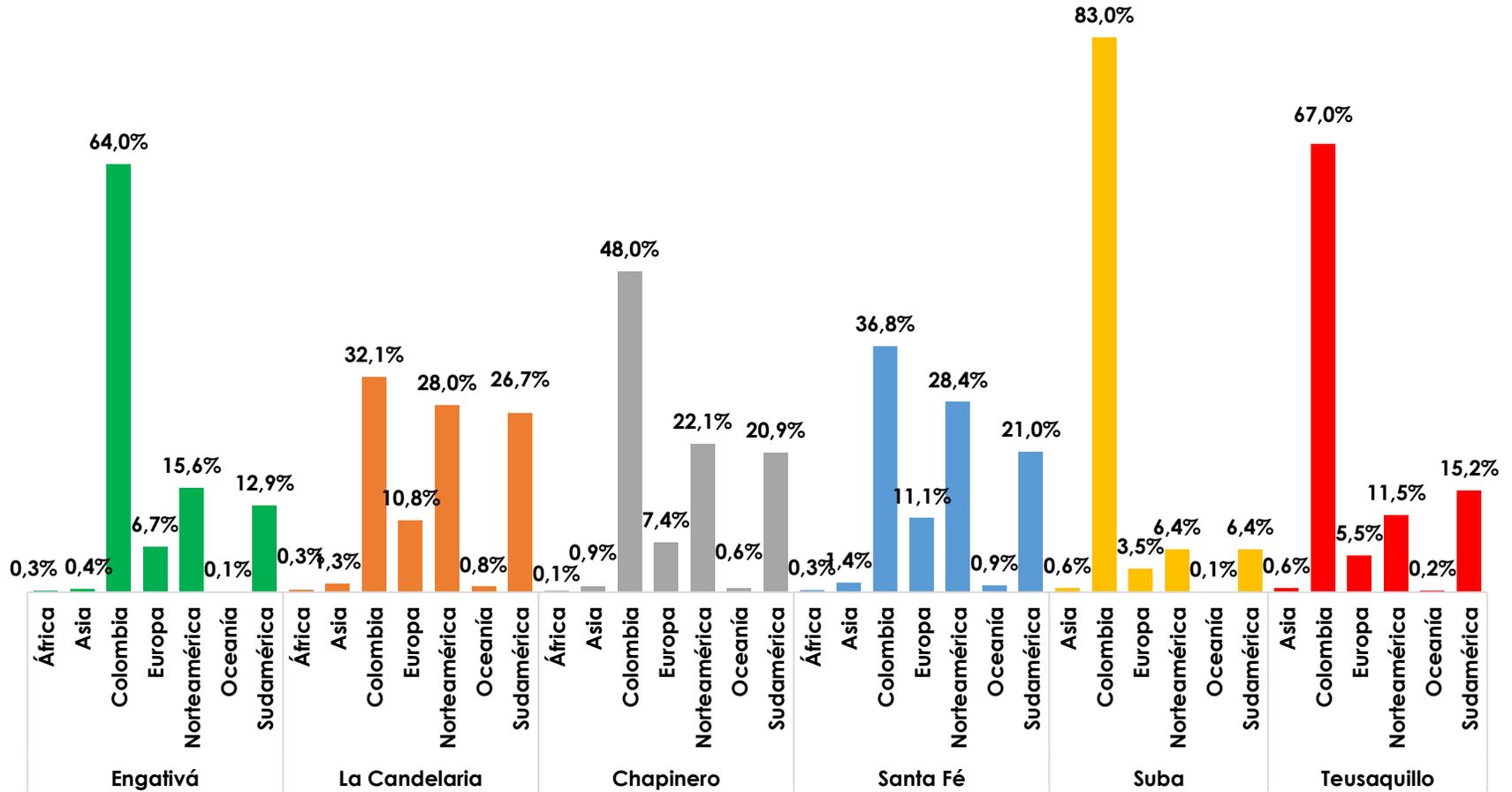
La grafica 12 muestra que a los atractivos de las localidades de Engativá, Suba, Teusaquillo y Usaquén los visitan principalmente las personas que asisten con la familia, mientras que para las localidades de Chapinero y Santa Fe son principalmente los grupos de amigos los que visitan y reseñan los atractivos, por último, se observa que para La Candelaria son las personas que visitan con la pareja el grupo de viaje predominante.

**Gráfica 12. Grupo de Viaje de los visitantes a los atractivos turísticos desagregado por localidad**



Fuente: TripAdvisor 2021

Gráfica 13. Continente de origen de los visitantes a los atractivos turísticos desagregado por localidad



Fuente: TripAdvisor 2021

Respecto al continente y teniendo en cuenta que los visitantes de origen nacional se cuentan en la gráfica como Colombia, se observa esta es la categoría que en todas las localidades representa el mayor porcentaje de visitantes.

También para todas las localidades, menos para Teusaquillo el segundo continente de origen de los visitantes es Norteamérica<sup>11</sup> y en algunos casos como en las localidades de La Candelaria, Chapinero, Santa Fe y Usaquén representan más del 20% de los visitantes que reseñaron los atractivos.

De igual manera la gráfica 14 muestra el origen de los visitantes para cada localidad seleccionada, pero de acuerdo con el país de origen, en donde se encuentra que las localidades de Engativá, suba y Teusaquillo son visitadas en una mayor medida por visitantes de origen nacional, alcanzando representatividades mayores al 60%.

Por otro lado, para localidades como La Candelaria y Santa Fe los visitantes de origen nacional no representan un porcentaje tan elevado respecto al total de visitantes (menos del 40%), lo que podría indicar que tienen atractivos turísticos más interesantes para los turistas de origen internacional.

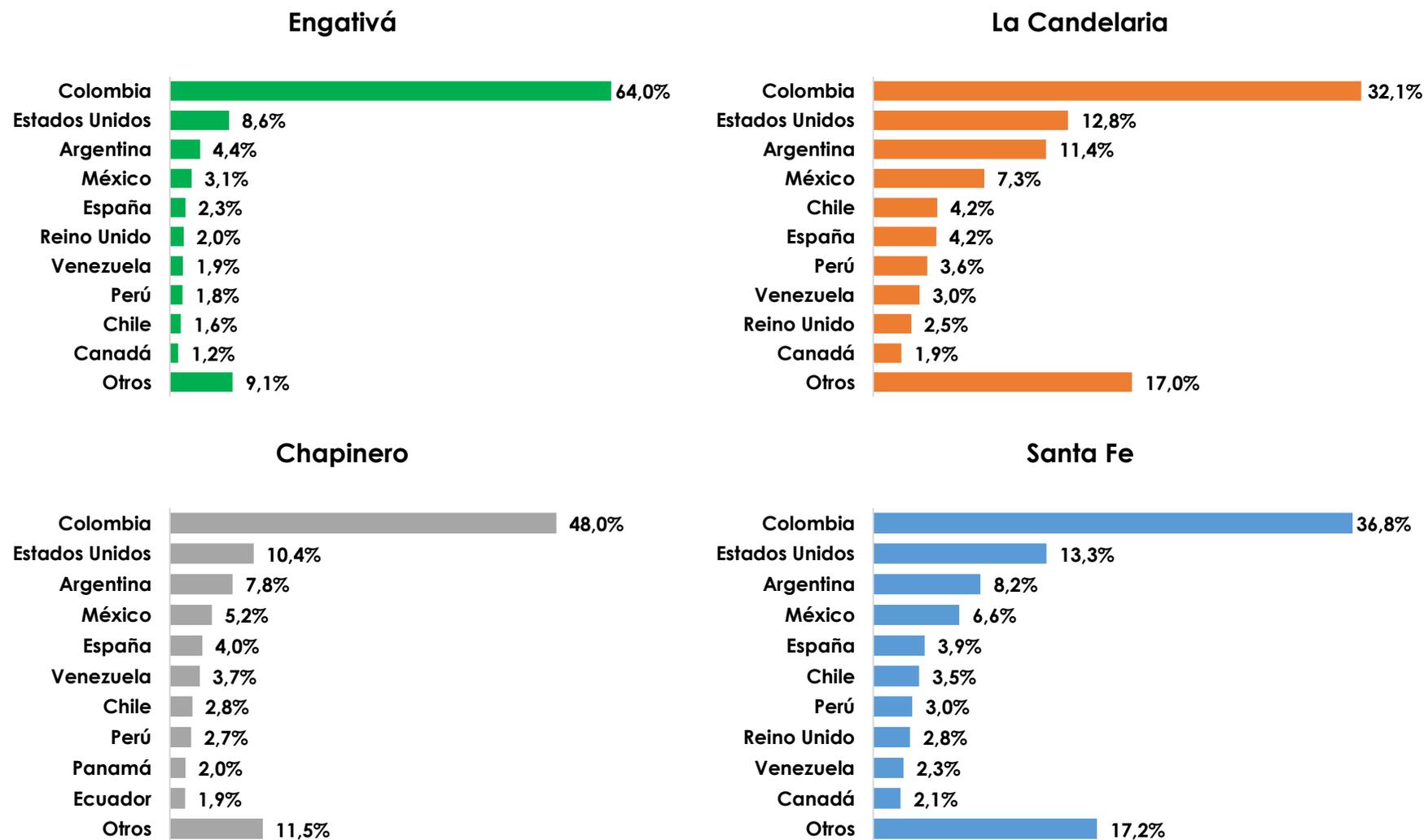
En cuanto a Usaquén y chapinero, existe mayor equilibrio entre los visitantes nacionales e internacionales, debido a que el porcentaje de reseñas de residentes colombianos es más cercano al 50%.

Además, se observa que para todas las localidades excepto para Teusaquillo, el segundo país de origen de donde provienen más reseñas es Estados Unidos, en donde para algunas localidades representa más del 10% del total reseñas.

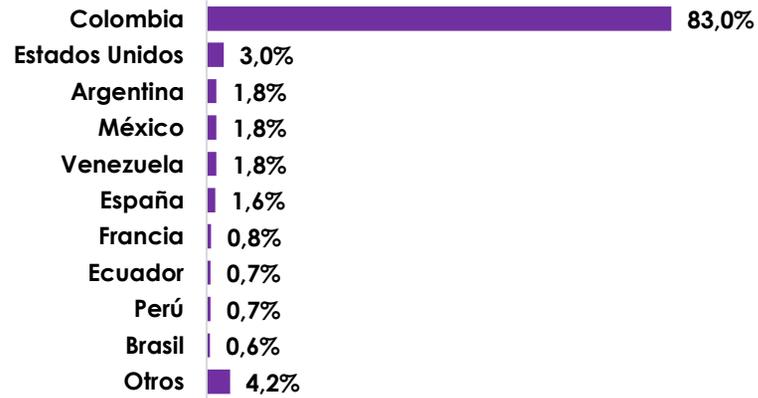
---

<sup>11</sup> Esta categoría Incluye países de Norteamérica y Centro América.

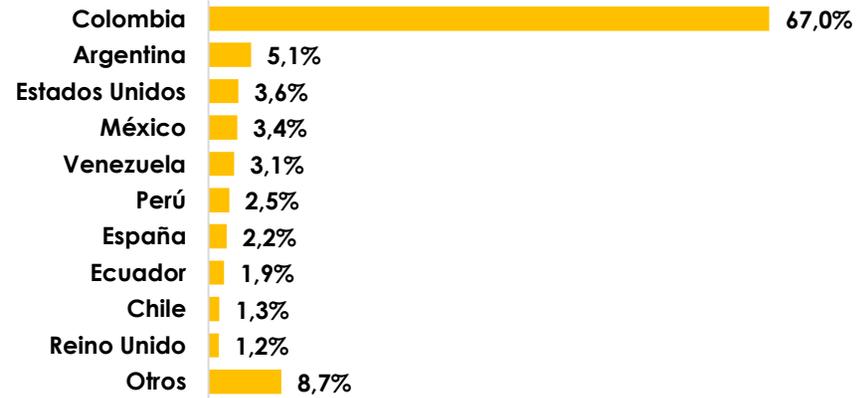
**Gráfica 14. Principales países de origen de los visitantes a los atractivos turísticos desagregado por localidad**



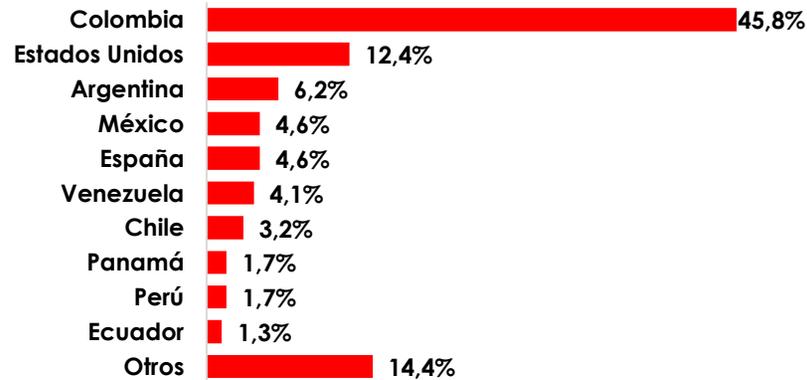
### Suba



### Teusaquillo



### Usaquén



Fuente: TripAdvisor 2021

Análogamente a los resultados a nivel de Bogotá, para las siete localidades seleccionadas se analizaron los textos de las reseñas, utilizando técnicas de Procesamiento de Lenguaje Natural (NLP), aunque en esta ocasión no se hace la comparación para los periodos de tiempo antes y durante la pandemia. El ejercicio que se realizó identifica las palabras asociadas a las reseñas positivas y negativas.

La Figura 3 muestra las nubes de palabras asociadas a cada localidad, donde las reseñas positivas se caracterizan por contener palabras principalmente asociadas con el tipo de atractivos que contiene cada localidad.

Para Engativá cuyo atractivo con más reseñas es el Jardín Botánico con el 85% del total de reseñas, se identificaron palabras como “*caminar*”, “*pasear*”, “*plan*”, “*disfrutar*”, las cuales están relacionadas con las actividades que se pueden realizar en este atractivo. Por otro lado, para las localidades de La Candelaria y Santa Fe se identificaron palabras como “*colección*”, “*museo*” y “*restaurante*”, evidenciando la importancia que tienen la oferta de museos (Museo del Oro, Museo Botero, Museo Nacional entre otros) y también sector gastronómico en ambas localidades.

Los sectores gastronómicos del Parque de la 93, la Zona T y la Zona G en Chapinero y el sector gastronómico de Usaquén se identifican como los más importantes para estas localidades, ya que las nubes de palabras de las reseñas positivas muestran “*restaurante*”, “*comer*”, “*calidad*” como factores que los visitantes resaltan.

En la localidad de Suba en donde el atractivo con mayor número de reseñas es el Teatro Mayor Julio Mario Santo Domingo, se identifican palabras positivas como “*acústica*”, “*programación*”, “*sonido*”, “*espectáculo*”, todas relacionadas con las actividades que se desarrollan en el teatro.

Respecto a las reseñas negativas para las localidades de Chapinero, Santa Fe y Usaquén se identificaron palabras como “atención”, “atender”, “pagar” conceptos que nuevamente están vinculados con el sector gastronómico de estas localidades.

Otras palabras que se encuentran en las reseñas negativas de las localidades seleccionadas son “olor”, “suciedad”, “basura” problemáticas en las cuales el distrito podría intervenir.

Adicionalmente, también resaltan palabras tales como “robar”, “amenazar” o “marihuana”, que son indicios de problemas de seguridad en algunas localidades. Así mismo, destacan también las palabras “parquear”, “parqueadero” y “carro” que evidencian la oferta de plazas de estacionamiento no sean suficientes para suplir la demanda.

La localidad de la Candelaria se caracteriza también por sus reseñas negativas con palabras asociadas “desordenar” o “invadir”, que puede asociarse con los vendedores ambulantes y artistas de la calle presentes en la zona.



### Chapinero - Reseñas Positivas



### Chapinero - Reseñas Negativas



### Santa Fé - Reseñas Positivas



### Santa Fe - Reseñas Negativas





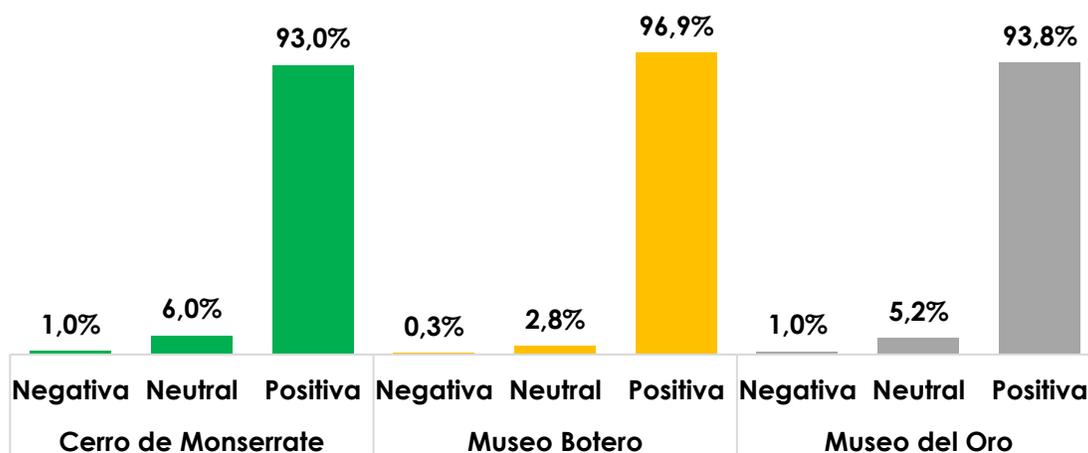


## 4.2 Resultados a nivel de Atractivos

Continuando con el análisis en esta sección se presenta un análisis desagregado para los tres atractivos con más reseñas, los cuales son: el Cerro de Monserrate, el Museo Botero y el Museo del Oro, estos tres atractivos suman un total de 25.210 comentarios, representando el 50,4% del total de reseñas recolectadas.

Al mirar la relación entre reseñas positivas y negativas para estos tres atractivos, se observa que están desbalanceadas ya que los tres atractivos tienen un porcentaje de reseñas positivas superior al 93%, mientras que las reseñas negativas solo alcanzan el 1%, de este resultado se puede concluir que estos tres atractivos cuentan con una percepción positiva entre los distintos visitantes que los reseñan<sup>12</sup>.

**Gráfica 15. Clasificación de las reseñas según su tipo desagregado por atractivo turístico**

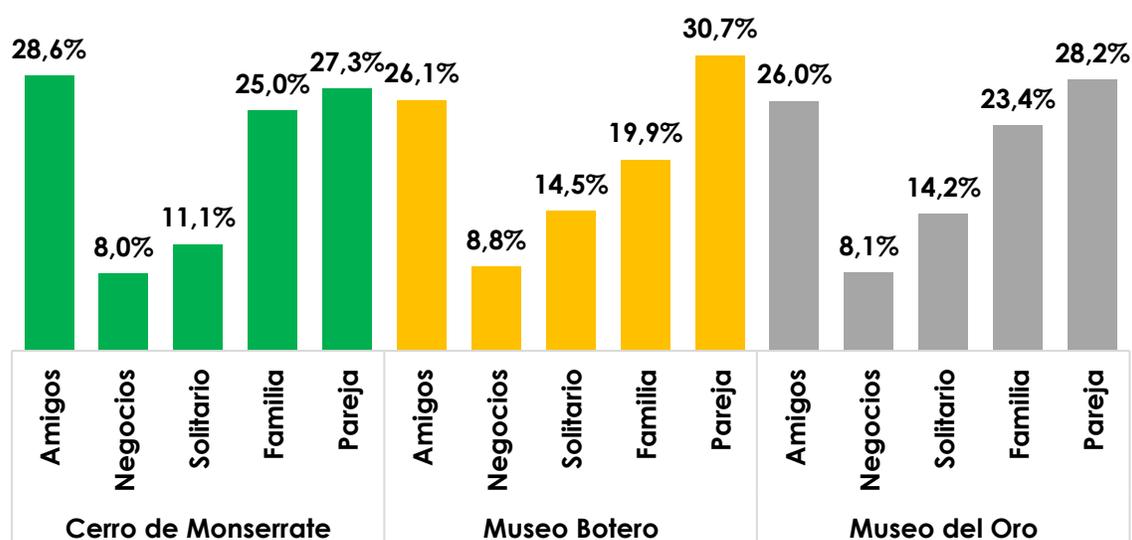


Fuente: TripAdvisor 2021

<sup>12</sup> Aunque este resultado es de esperarse dado que la filosofía de la página de trip-advisor en principio fue la de potenciar o identificar lugares que fueran populares para las personas, es decir, no se esperaban según está lógica muchos comentarios negativos.

Respecto al grupo de viaje que indicaron de los visitantes de estos tres atractivos, visitar con la pareja es el más común para el Museo Botero y el Museo del Oro, seguido de visitar con Amigos. El caso contrario ocurre con el Cerro de Monserrate en donde el 28,6% de los visitantes indican que fueron con amigos, seguido por quienes realizaron la visita con su pareja (27,3%). También se puede observar como para los tres atractivos la visita con la familia se ubica en tercer lugar.

**Gráfica 16. Grupo de Viaje de los visitantes desagregado por atractivo turístico**



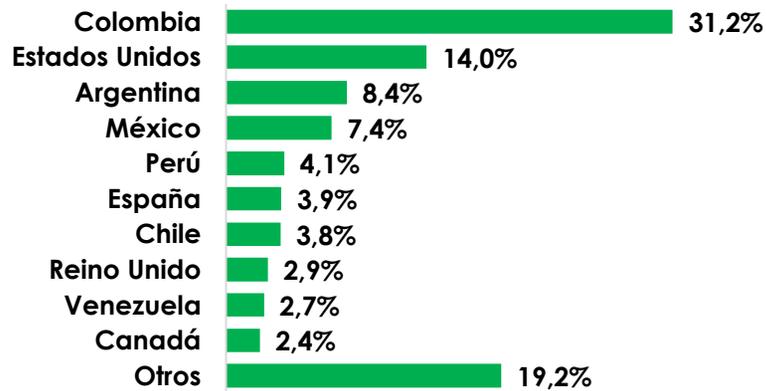
Fuente: TripAdvisor 2021

Para los tres atractivos el top 4 de país de origen de los visitantes se mantiene, siendo Colombia el principal origen seguido de Estados Unidos, Argentina y México; al llegar al quinto país de origen surgen diferencias en el orden de los países donde Perú, Chile y España se alternan esta posición.

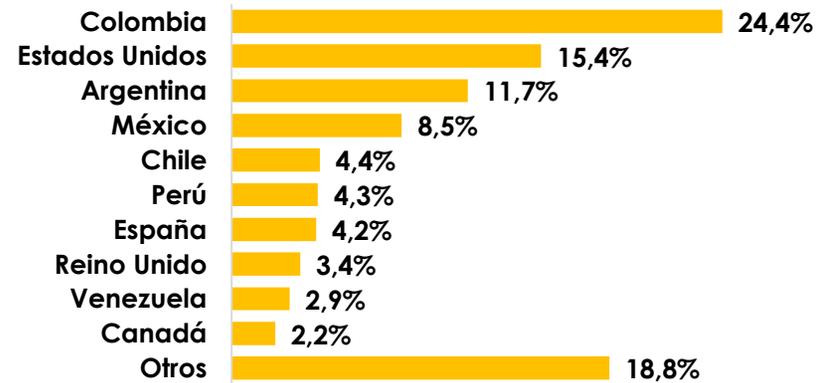
Estados Unidos es el primer origen internacional, que en los tres atractivos llega a representar porcentajes entre el 14% y 15% de las reseñas de los atractivos, mostrando la importancia de este mercado dentro de los visitantes y turistas a los atractivos de Bogotá.

**Gráfica 17. Principales países de origen de los visitantes desagregado por atractivo turístico**

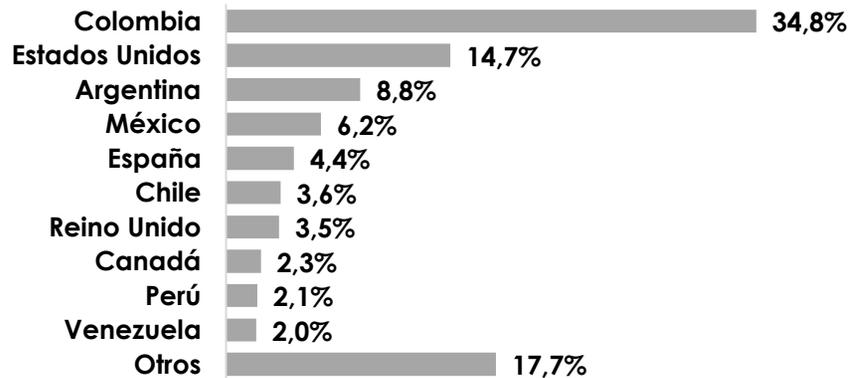
**Cerro de Monserrate**



**Museo Botero**



**Museo del Oro**



Fuente: TripAdvisor 2021

Análogo a las secciones anteriores, para estos tres atractivos también se realizó un análisis de los textos de las reseñas, utilizando técnicas de Procesamiento de Lenguaje Natural, identificando las palabras asociadas a las reseñas positivas y negativas.

Para cada uno de estos tres atractivos en la Figura 4 se puede observar las nubes de palabras asociadas, nuevamente evidenciando el desbalanceo entre las reseñas positivas y negativas, ya que estos tres atractivos al contar con muy pocas reseñas negativas, sus nubes de palabras también cuentan con pocas palabras asociadas.

Para el Cerro de Monserrate, se identificaron palabras como “*vestir*”, “*vista*”, “*experiencia*”, “*comer*”, “*restaurante*”, “*ejercicio*” que se asocian con las diferentes actividades que se pueden realizar en este atractivo. Por un lado subir utilizando alguno de los distintos medios con ropa adecuada para el clima frío de la capital, incluyendo la caminata como actividad física, adicionalmente, desde la cima poder apreciar la vista de la ciudad, así como, disfrutar de la oferta gastronómica que se encuentra en este atractivo.

Dentro de las palabras asociadas a las reseñas negativas de este atractivo fueron “*olor*”, “*suciedad*” y “*organización*” las cuales pueden asociarse a la cantidad de visitantes que recibe especialmente los fines de semana generando así mismo efectos negativos ambientales y de logística para la subida. También se identificaron palabras como “*cobrar*”, “*servicio*” y “*calidad*” las cuales están relacionadas con los restaurantes ubicados al interior del atractivo.

Por otro lado, en las reseñas de los atractivos Museo Botero y Museo del Oro se identificaron palabras como “*colección*”, “*exposición*”, “*visitar*”, evidenciando como las temáticas y contenidos de estos dos museos tienen

una gran importancia entre los visitantes que posteriormente escriben una reseña.

Respecto a las reseñas negativas para estos dos museos se identificaron muy pocas palabras, pero dentro de las que se identificaron se encuentran "atención", "entender" conceptos que están relacionados con el servicio ofrecido y en los cuales se podrían mejorar.

Figura 4. Nube de palabras asociadas a las reseñas positivas y negativas desagregado por atractivo turístico

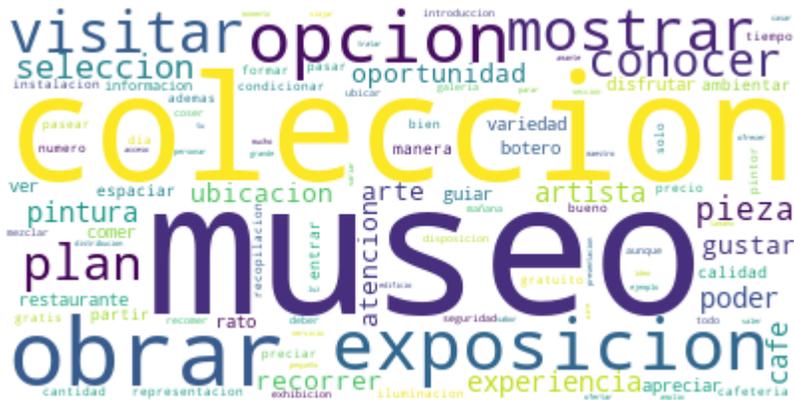
**Cerro de Monserrate - Reseñas Positivas**



**Cerro de Monserrate - Reseñas Negativas**



**Museo Botero - Reseñas Positivas**



**Museo Botero - Reseñas Negativas**





## 5. CONCLUSIONES

- El uso de nuevas herramientas como el Big Data y la ciencia de datos en la investigación turística de Bogotá representan un papel importante, ya que su aplicación en distintos estudios produce información importante que no puede ser recolectada a través de procesos de aplicación de formularios, por tanto, se genera información complementaria a la ya producida por el Observatorio de Turismo de Bogotá.
- En este primer acercamiento hacia las nuevas tecnologías y metodologías, aplicado en este caso a las reseñas de distintos atractivos turísticos que conforman parte de la oferta turística de la ciudad (pero no toda), se logró determinar la percepción de los turistas y visitantes asociadas a características y condiciones de la oferta haciendo uso de técnicas de Procesamiento de Lenguajes Naturales (PLN), además de características (origen, grupo de viaje y percepción) propias de los turistas y visitantes.
- Se identificó un grupo de 3 localidades: La Candelaria, Santa Fe y Chapinero, las cuales contienen cerca del 47% de los atractivos utilizados en el estudio y agregan más del 50% de las reseñas recolectadas.
- Más del 40% de los usuarios (turistas o visitantes) que escriben y publican reseñas sobre los atractivos de Bogotá son de Origen Nacional, seguidos de los turistas de Estados Unidos.
- Para Bogotá, al igual que para la mayoría de los destinos turísticos urbanos a nivel mundial, se presenta que durante el periodo de la pandemia la percepción de los turistas que los visitaron alcanzó los niveles más bajos en la historia de reseñas de TripAdvisor, así mismo, se evidencia que para Bogotá las reseñas asociadas con el sector

gastronómico perdieron relevancia, mientras que las filas, museos y organización de nuevos planes emergieron.

## INDICE DE TABLAS, FIGURAS Y GRÁFICOS

### Índice de Tablas

<b>Tabla 1. Trabajos en el que se utilizan las reseñas de los usuarios para el análisis los destinos Turísticos. ....</b>	<b>7</b>
<b>Tabla 2. Estructura de los datos que se descargaron .....</b>	<b>10</b>

### Índice de Figuras

<b>Figura 1. Big Data en la investigación turística .....</b>	<b>6</b>
<b>Figura 2. Nube de palabras asociadas a las reseñas positivas y negativas antes y después de la pandemia .....</b>	<b>20</b>
<b>Figura 3. Nube de palabras asociadas a las reseñas positivas y negativas desagregado por localidad .....</b>	<b>31</b>
<b>Figura 4. Nube de palabras asociadas a las reseñas positivas y negativas desagregado por atractivo turístico .....</b>	<b>40</b>

## Índice de Gráficos

<b>Gráfica 1. Número de atractivos turísticos por localidad .....</b>	<b>11</b>
<b>Gráfica 2. Número de reseñas por localidad .....</b>	<b>12</b>
<b>Gráfica 3. Atractivos con el mayor porcentaje de reseñas .....</b>	<b>13</b>
<b>Gráfica 4. Número de reseñas por mes (2006-2021) .....</b>	<b>14</b>
<b>Gráfica 5. Grupo de Viaje de los visitantes a los atractivos turísticos .....</b>	<b>14</b>
<b>Gráfica 6. Continente de origen de los visitantes a los atractivos turísticos</b>	<b>15</b>
<b>Gráfica 7. Principales países de origen de los visitantes a los atractivos turísticos.....</b>	<b>16</b>
<b>Gráfica 8. Clasificación de las reseñas según su tipo.....</b>	<b>17</b>
<b>Gráfica 9. Calificación promedio de las reseñas por trimestre (2012-2021)</b>	<b>18</b>
<b>Gráfica 10. Clasificación de las reseñas según su tipo desagregado por localidad .....</b>	<b>22</b>
<b>Gráfica 11. Atractivos con el mayor número de reseñas desagregado por localidad .....</b>	<b>23</b>
<b>Gráfica 12. Grupo de Viaje de los visitantes a los atractivos turísticos desagregado por localidad .....</b>	<b>24</b>
<b>Gráfica 13. Continente de origen de los visitantes a los atractivos turísticos desagregado por localidad .....</b>	<b>25</b>
<b>Gráfica 14. Principales países de origen de los visitantes a los atractivos turísticos desagregado por localidad .....</b>	<b>27</b>
<b>Gráfica 15. Clasificación de las reseñas según su tipo desagregado por atractivo turístico .....</b>	<b>35</b>
<b>Gráfica 16. Grupo de Viaje de los visitantes desagregado por atractivo turístico .....</b>	<b>36</b>
<b>Gráfica 17. Principales países de origen de los visitantes desagregado por atractivo turístico .....</b>	<b>37</b>

## BIBLIOGRAFIA

Adhinugroho, Y., Putra, A. P., Luqman, M., Ermawan, G. Y., Takdir, Mariyah, Sifi, & Pramana, Setia. (2020). Development of online travel “Web Scraping” for tourism statistics in Indonesia. *Information Research*, 25(4), paper 885.

Agüero-Torales, M. M., Cobo, M. J., Herrera-Viedma, E., & López-Herrera, A. G. (2019). A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor. *Procedia Computer Science*, 162, 392-399.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191.

Almeida, R. de & Abrantes, R.M. (2018). The development of tourism indicators through the use of social media data: the case of Minas Gerais, Brazil *Information Research*, 23(4), paper 805.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2015). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1-24.

vanden Broucke, S., & Baesens, B. (2018). Practical “Web Scraping” for Data Science: Best Practices and Examples with Python. Apress.

Caballero, R., Martín, E., & Riesco, A. (2018). Big Data con Python. Recolección, almacenamiento y proceso. Alfaomega Grupo Editor.

Choong, W. J. (2019). An automated “Web Scraping” tool for Malaysia tourism (Degree dissertation, UTAR).

Dhiratara, A., Yang, J., Bozzon, A., & Houben, G. J. P. M. (2016). Social media data analytics for tourism: A preliminary study. KDWEB 2016.

Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52, 498-506.

Flôres Limberger, P., dos Anjos, F. A., de Souza Meira, J. V., & dos Anjos, S. J. G. (2014). Satisfaction in hospitality on TripAdvisor. com: An analysis of the correlation between evaluation criteria and overall satisfaction. *Tourism & Management Studies*, 10(1).

Heydt, M. (2018). Python "Web Scraping" Cookbook: Over 90 proven recipes to get you scraping with Python, microservices, Docker, and AWS. Packt Publishing Ltd.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301-323.

Mitchell, R. (2018). "Web Scraping" with Python: Collecting more data from the modern web. " O'Reilly Media, Inc."

Moro, S., Rita, P., Oliveira, C., Batista, F., & Ribeiro, R. (2018). Leveraging national tourist offices through data analytics. *International Journal of Culture, Tourism and Hospitality Research*.

Pearce, P. L., & Wu, M. Y. (2015). Entertaining international tourists: An empirical study of an iconic site in China. *Journal of Hospitality & Tourism Research*, 42(5), 772-792.

Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment analysis on tripadvisor: Are there inconsistencies in user reviews?. In *International Conference on Hybrid Artificial Intelligence Systems* (pp. 15-25). Springer, Cham.



ALCALDÍA MAYOR  
DE BOGOTÁ D.C.

INSTITUTO DISTRITAL DE  
**TURISMO**



9.39

731.42

341.79

325.13

774.72

341.79  
154.18