



**OBSERVATORIO
DE TURISMO DE BOGOTÁ**
197

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

Instituto Distrital de Turismo



ALCALDÍA MAYOR
DE BOGOTÁ D.C.





CARLOS FERNANDO GALÁN PACHÓN
Alcalde Mayor de Bogotá

MARÍA DEL PILAR LÓPEZ URIBE
Secretaria de Desarrollo Económico

ANDRÉS SANTAMARÍA GARRIDO
Director General
Instituto Distrital de Turismo

PAOLA ANDREA SÁNCHEZ PRIETO
Subdirectora
Inteligencia y Gestión de Tecnologías de
la Información

DIEGO ALEJANDRO RODRÍGUEZ
Investigación

Contacto
Instituto Distrital de Turismo
www.idt.gov.co
Bogotá D.C., 2024

CONTENIDO

1. INTRODUCCIÓN	3
2. ANTECEDENTES DE CIENCIA DE DATOS APLICADA AL SECTOR TURISMO.....	4
3. METODOLOGÍA DE RECOLECCIÓN DE LOS DATOS	8
4. RESULTADOS.....	11
4.1 Resultados a nivel de Localidad.....	22
4.2 Resultados a nivel de Atractivos	¡Error! Marcador no definido.
5. CONCLUSIONES.....	36
INDICE DE TABLAS, FIGURAS Y GRÁFICOS	38
BIBLIOGRAFIA	40

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. Para el año 2021 buscando aprovechar los datos no convencionales para generar análisis holísticos del sector, se publicó una primera versión del: "[ESTUDIO DE CIENCIA DE DATOS APLICADA AL SECTOR TURISMO, PERCEPCIÓN DE LOS ATRACTIVOS TURÍSTICOS EN BOGOTÁ](#)".

Para este año 2024, utilizando nuevamente "Web Scraping" (también llamada "extracción de datos web" o "minería de datos web"), Procesamiento de Lenguajes Naturales (NLP por sus siglas en inglés) y Análisis de Sentimiento, a marzo de 2024 se logró extraer la información de cerca de 51.000 reseñas para 185 atractivos de la ciudad de Bogotá (mismos atractivos usados en el estudio del 2021), con el fin de conocer la percepción de los turistas sobre la estructura turística de la ciudad, determinar las características, condiciones de la oferta, falencias, fortalezas e incluso descubrir los patrones de movilidad de los turistas en la ciudad como igualmente identificar falencias y fortalezas, que a su vez ayuden a determinar acciones de mejora de los atractivos turísticos de la ciudad. juegan un papel importante.

En este estudio, el cual se utiliza la Ciencia de Datos aplicada al sector turismo, se espera poder entender de mejor manera los cambios en la percepción que tienen los distintos turistas visitantes de los atractivos turísticos en Bogotá, especialmente para los últimos 3 años (2021-2023). Finalmente, desde el OTB se espera seguir utilizando estas 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

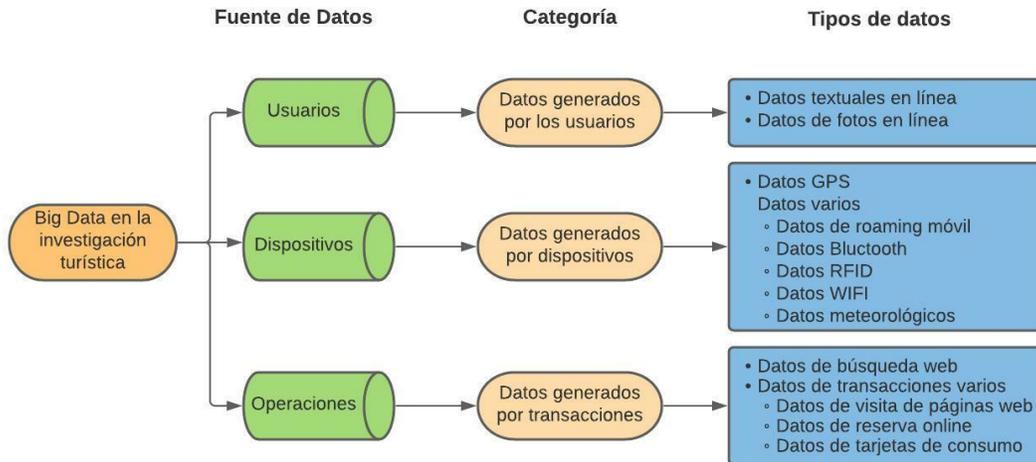
Teniendo como precedente la literatura existente sobre el uso de Big Data en la investigación Turística y sobre Ciencia de Datos aplicada al Sector Turismo, se evidencia que estos temas no son nuevos, ya que se puede encontrar una gran cantidad de trabajos y artículos científicos en los cuales se ha 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 RFID¹, 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.

¹ 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 los 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.

Estudio	Fuente de los Datos	Área Geográfica	Método Extracción de los Datos	Tipo de reseñas	Número Total establecimientos o de Reseñas ²
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 ³	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

² 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.

³ La Alhambra, Sagrada Familia y La mezquita de Córdoba.

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
Aggarwal & Gour (2020)	TripAdvisor	Goa, La India	Web scraping	Atractivos Turísticos	31.295 reseñas
Kim et al. (2021)	Google y TripAdvisor	Incheon, Korea	Web scraping	Siete Atractivos Turísticos ⁴	6.822 reseñas
Safarian & Konyshev (2022)	TripAdvisor	Región Perm, Rusia	No Disponible	Atractivos Turísticos	5.668 reseñas
Sánchez-Vargas et al. (2024)	TripAdvisor	Comarca Tajo-Salor-Almonte (Cáceres, España)	Web scraping	Hoteles, Restaurantes, Atractivos Turísticos	2.442 reseñas

Fuente: Elaboración propia.

⁴ La Barrio Chino de Incheon, la ciudad de cuento de hadas, la isla de Wolmido, la sala conmemorativa de la operación de desembarco de Incheon, el parque Jayu en Incheon y el mercado internacional de Sinpo.

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 HTML⁵ 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 TripAdvisor⁶ la URL para cada una de las páginas web de los 430 atractivos de Bogotá, donde se encontraron 185 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.

⁵ HyperText Markup Language (Lenguaje de Marcas de Hipertexto en español), es el componente más básico de la Web.

⁶ <https://www.tripadvisor.co/>

Para cada uno de estos 185 atractivos con cuenta en TripAdvisor, se analizó la estructura HTML del sitio web (diagramación y navegación) y luego se extrae la información de todas las reseñas. Finalmente, la información recolectada para cada atractivo se registra en una base de datos. Todas las reseñas descargadas corresponden a aquellas que se escribieron en español, conformando finalmente una base de datos con 50.732⁷ reseñas escritas desde diciembre de 2006 hasta marzo de 2024.

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 la extracción 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 legalidad⁸.

Para el total de 50.732 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 calificaciones), Cantidad de Reseñas⁹ 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,

⁷ En el estudio del 2021 se extrajeron 49.955 reseñas, por lo cual para esta actualización del estudio se incremento el total de reseñas en 777 las cuales fueron escritas entre septiembre de 2021 y marzo de 2024.

⁸ 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.

⁹ 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.

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

Para el atractivo	
Variable	Explicación
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
Para las Reseñas	
Variable	Explicación
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 reportó el visitante
Fecha de reseña	Fecha en la cual el visitante escribe la reseña
Fecha de experiencia	Fecha en la cual el visitante visitó el atractivo
Ciudad	Ciudad y país de residencia del visitante
Grupo de Viaje	Grupo de viaje o personas con las que visitó al atractivo el visitante

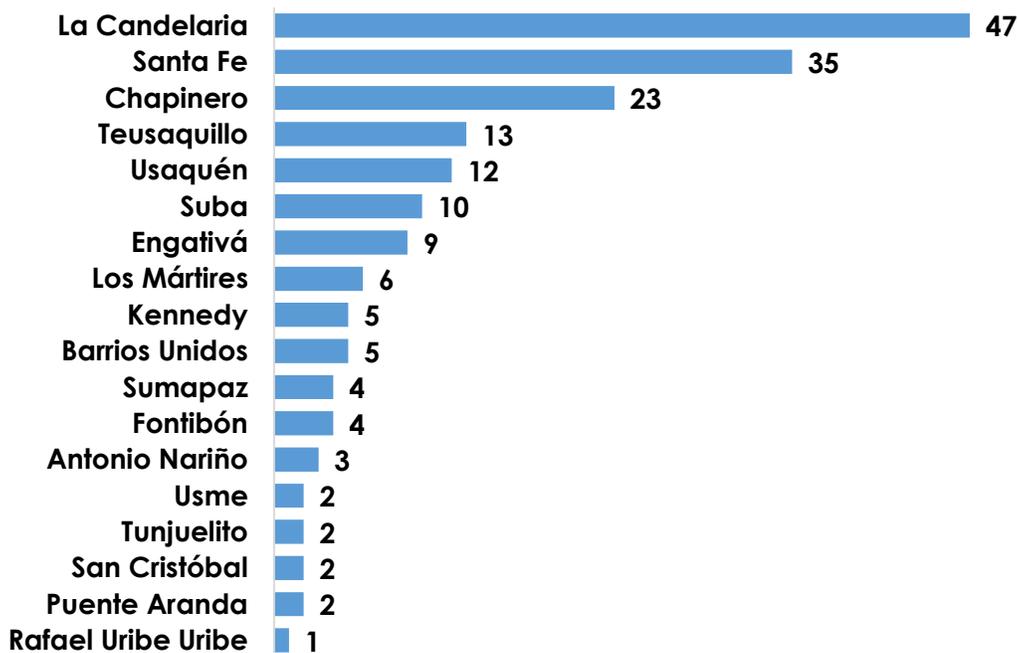
Fuente: Elaboración propia.

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.

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 (2006-2024-I)

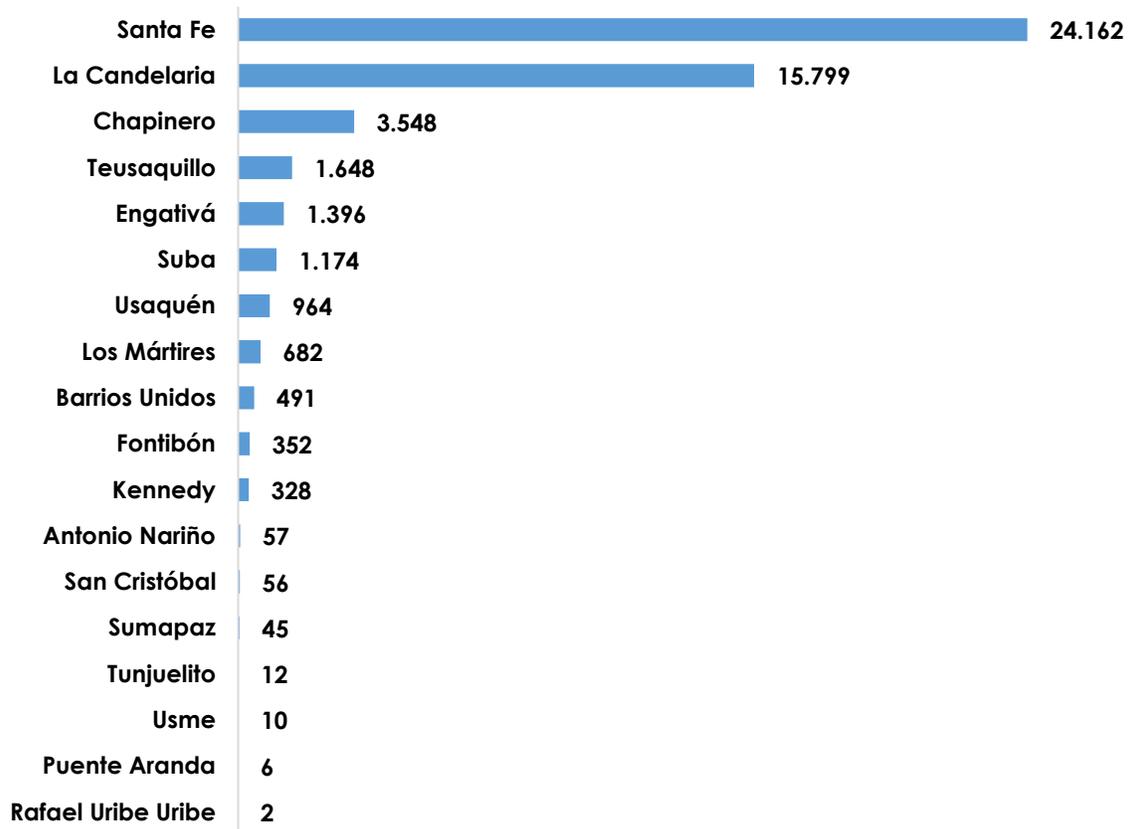


Fuente: TripAdvisor 2024.

Al igual que en el estudio de 2021, La Candelaria es la localidad con más atractivos, sin embargo, no es la localidad que agrega más reseñas, siendo la localidad de Santa Fe la que mayor número de reseñas posee (24.162),

principalmente debido a que los dos atractivos más reseñados están ubicados en esta localidad (El Cerro de Monserrate y El Museo del Oro), por lo cual la localidad de Santa Fe representa el 47,6% de las reseñas recopiladas para el estudio.

Gráfica 2. Número de reseñas por localidad (2006-2024-I)

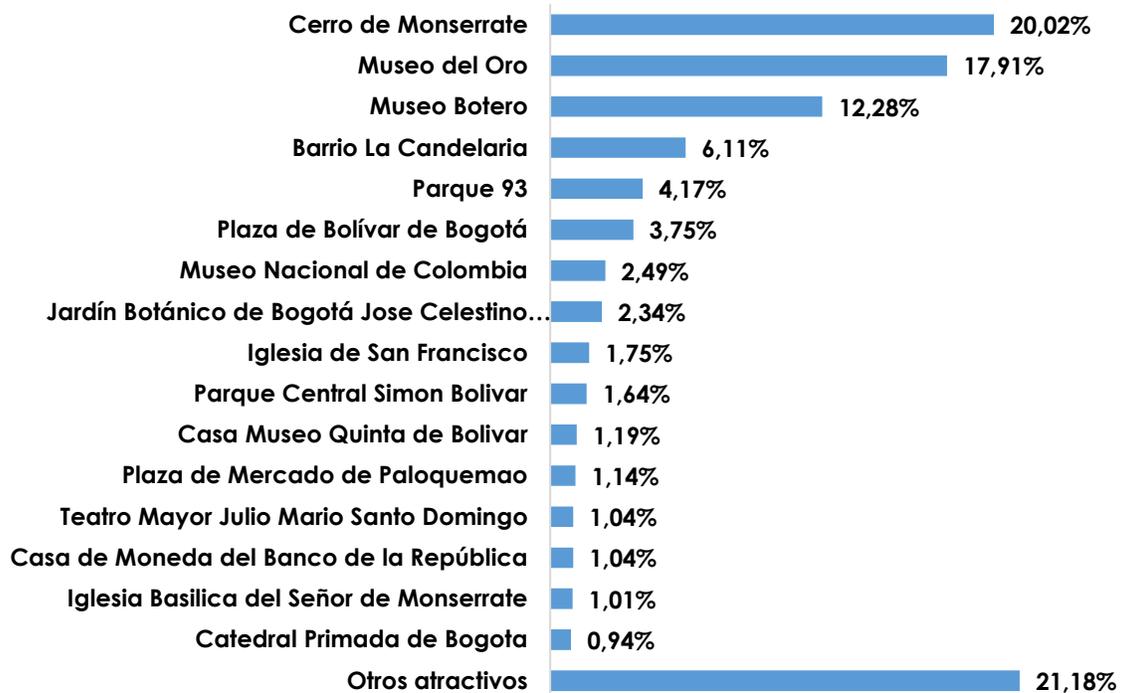


Fuente: TripAdvisor 2024.

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 25.468 y representan el 50,2% del total de reseñas recopiladas en este estudio, también se pudo identificar que 170 atractivos que tienen menos de 500 reseñas representan el 22,1%. 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, tendencia que no ha presentado un cambio significativo en el transcurso de los dos últimos años.

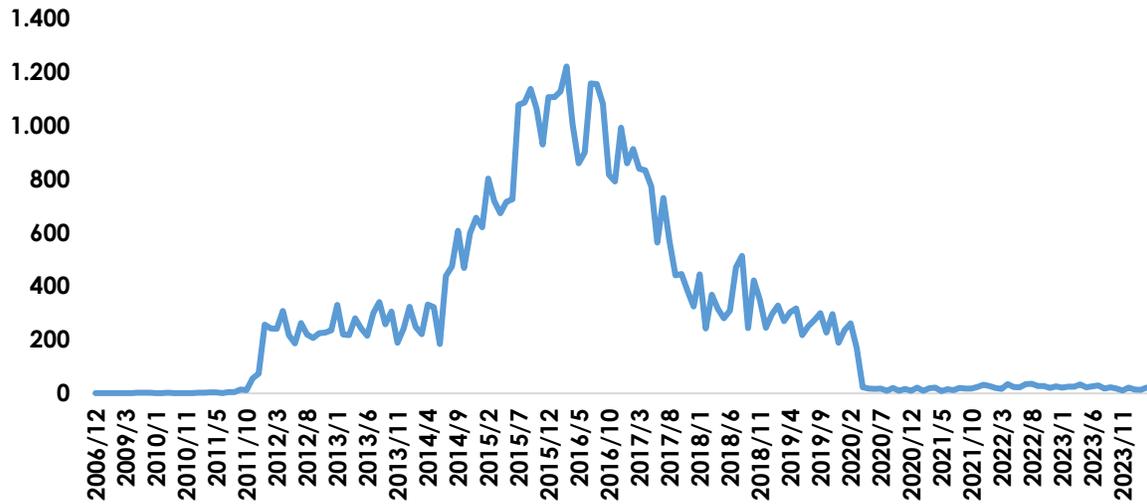
Gráfica 3. Atractivos con el mayor porcentaje de reseñas (2006-2024-I)



Fuente: TripAdvisor 2024.

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 donde la plataforma alcanzó mayor popularidad y se identificaron más de 1000 reseñas por mes. Sin embargo, desde mitad de 2017 hasta inicios del 2020 cae el número promedio de reseñas por mes a 350.

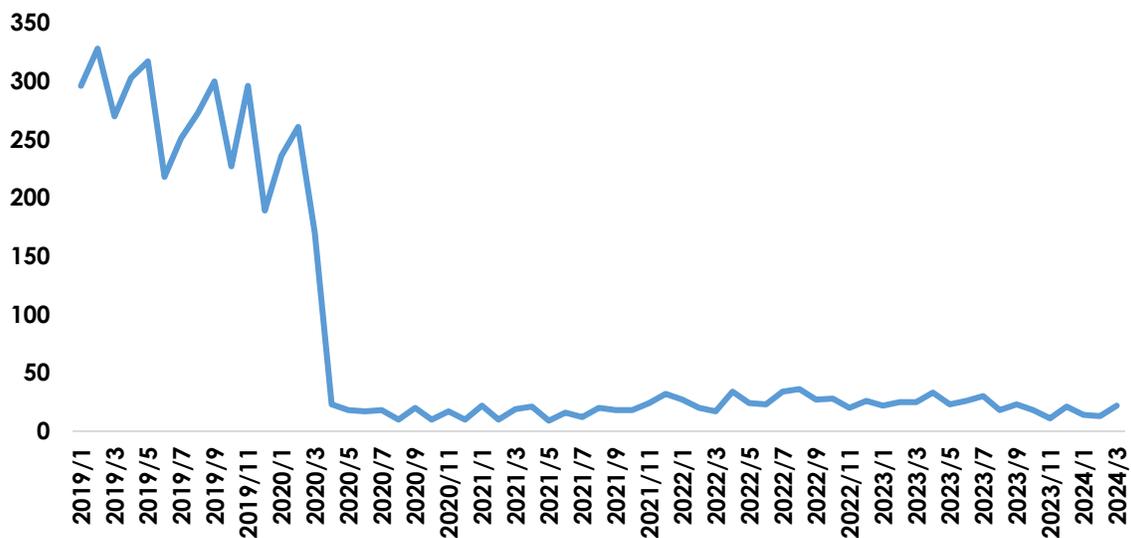
Gráfica 4. Número de reseñas por mes (2006-2024-I)



Fuente: TripAdvisor 2024.

De igual manera, se identifica el efecto de la pandemia, ya que, desde el mes de marzo del 2020 con la caída del turismo internacional y doméstico debido a las restricciones de movilidad, cae el número promedio de reseñas por mes a 20, para estos 185 atractivos turísticos de Bogotá, la tendencia producto de esta caída a 2024 aún no ha cambiado.

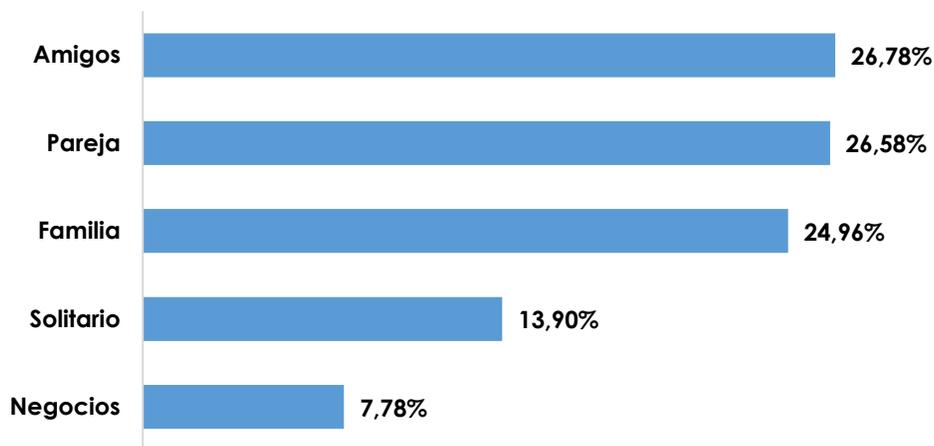
Gráfica 5. Número de reseñas por mes (2019-2024-I)



Fuente: TripAdvisor 2024.

Respecto el grupo de viaje o personas con las que visita el atractivo el turista o visitante, principalmente se encuentran las visitas con amigos en 26,78% de las reseñas y con la pareja 26,58%. Esto seguido de los visitantes que asisten con la familia 24,96%. 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 6. Grupo de Viaje de los visitantes a los atractivos turísticos (2006-2024-I)



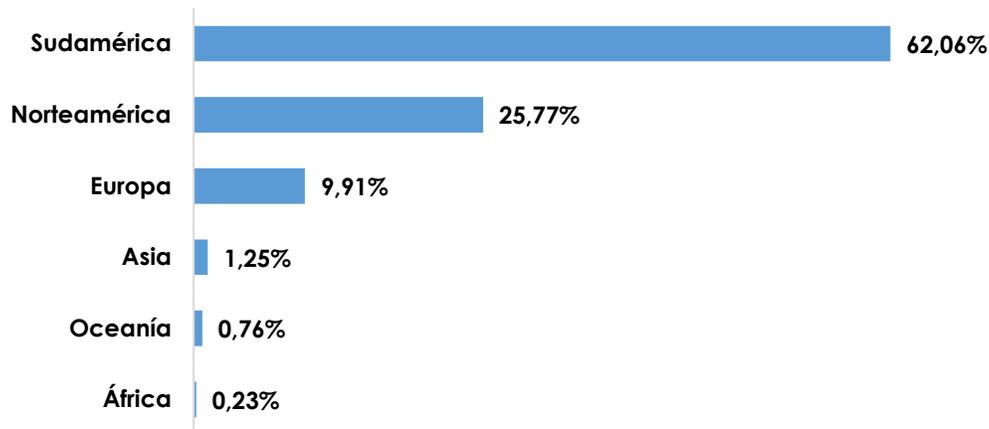
Fuente: TripAdvisor 2024.

Al observar el continente de origen de los turistas y visitantes que escriben las reseñas, en su mayoría se identifican que son de Sudamérica (62,06%), 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 Norteamérica¹⁰ con un 25,77%, principalmente turistas de Estados Unidos o de México, países con los cuales Bogotá tiene varias rutas, frecuencia alta

¹⁰ Esta categoría Incluye países de Norteamérica, Centro América y el Caribe.

de vuelos y hacen parte de los países mayores emisores de turistas hacia Bogotá según la Investigación viajeros en Bogotá.

Gráfica 7. Continente de origen de los visitantes a los atractivos turísticos (2006-2024-I)

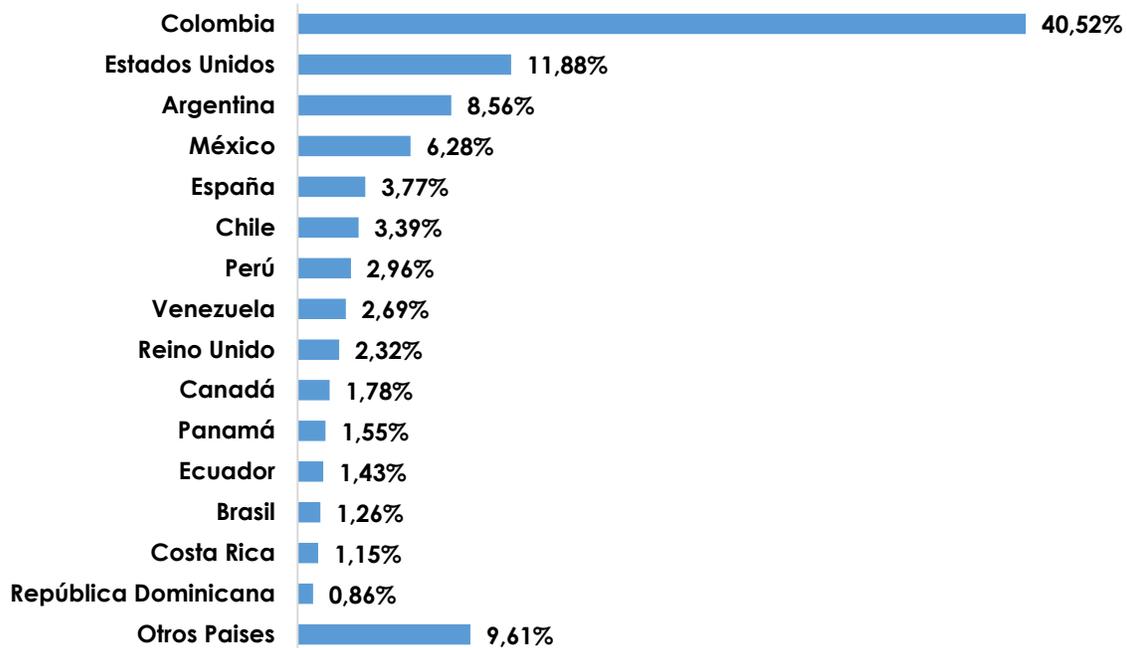


Fuente: TripAdvisor 2024.

Continuando con el análisis sobre el origen de los turistas y visitantes de los atractivos de Bogotá, la gráfica 8 presenta los principales 15 orígenes de los visitantes a los atractivos turísticos, donde se identificó que en primer lugar un 40,52% son de origen nacional (provenientes de distintas ciudades y/o departamentos de Colombia, como también residentes de la ciudad), seguido de visitantes provenientes de Estados Unidos 11,88% y Argentina 8,56%. Parte del continente europeo solo España y el Reino Unido son los países que aparecen en este listado.

También se observa que estos quince primeros países de origen de los visitantes (incluido Colombia) recogen el 90,39% del total de reseñas recolectadas para el estudio y tan solo un 9,61% corresponde al resto de países de origen de los visitantes que escriben reseñas sobre los atractivos de Bogotá.

Gráfica 8. Principales países de origen de los visitantes a los atractivos turísticos (2006-2024-I)

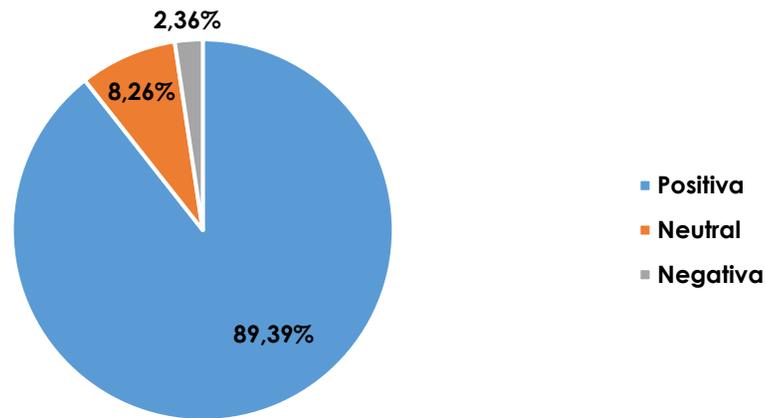


Fuente: TripAdvisor 2024.

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 **neutrales** y valores mayores a tres como reseñas **positivas**.

Al evaluar cómo se distribuyen las reseñas según esta clasificación se encuentra que el 89,39% son reseñas positivas, 8,26% neutras y solamente 2,36% son 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 9. Clasificación de las reseñas según su tipo (2006-2024-I)



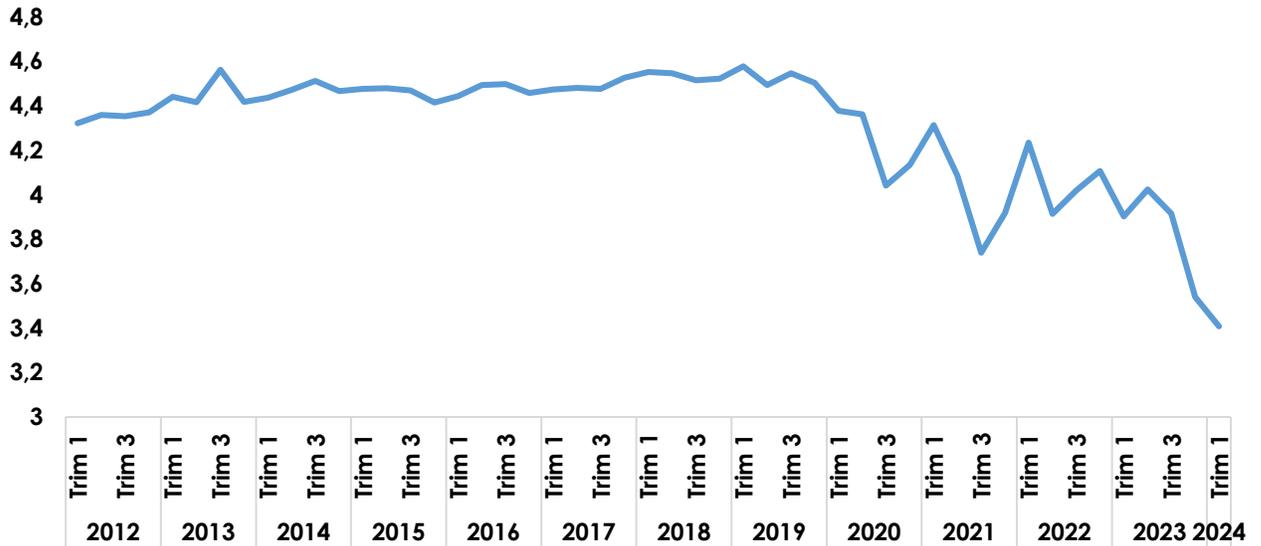
Fuente: TripAdvisor 2024.

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 tercer trimestre de 2023 hasta el primer trimestre de 2024, la calificación promedio de las reseñas cae nuevamente por debajo de 4 (igual que durante la pandemia), presentando una tendencia decreciente para estos tres periodos y alcanzando el menor valor registrado (3,4 promedio trimestre) en el primer trimestre de 2024.

Gráfica 10. Calificación promedio de las reseñas por trimestre (2012-2024-I)



Fuente: TripAdvisor 2024.

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 2006¹¹ hasta 2021 y el segundo para el 2022, 2023 y el primer trimestre de 2024.

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 “*actividades*”, “*ambiente*” y “*agradable*” tiene relevancia para el periodo 2006 y 2021, principalmente relacionado con la oferta con la que cuenta la ciudad. Mientras que para el periodo 2022 y 2024-I, tienen relevancia palabras como “*museo*”, “*vista*”, “*excelente*” e “*experiencia*”, igualmente asociadas a la oferta de Bogotá y mostrando la importancia de los dos museos con más reseñas (El Museo del Oro y el Museo Botero).

¹¹ Año donde se identificó la primera reseña a un atractivo.

Por el contrario, dentro de las reseñas negativas las palabras “*entrada*” y “*fila*”, “*centro*” y “*gente*” tienen protagonismo hasta el 2021, asociado a la manera de acceder a los atractivos turísticos y a la percepción negativa de turistas y visitantes sobre algunas condiciones del centro de la ciudad, mientras que las palabras “*servicio*”, “*atención*”, “*experiencia*” y “*centro*” presentan una mayor relevancia desde el 2022 en adelante, esto asociado a las nuevas problemáticas que encuentran los visitantes durante su visita a los atractivos.

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: Chapinero, Engativá, La Candelaria, Santa Fe, Suba, Teusaquillo y Usaquén. Estas 7 localidades agregan un total de 48.691 reseñas, equivalente al 96% del total de reseñas recolectadas del ejercicio de "Web Scraping" que se ejecutó para los atractivos turísticos de la ciudad.

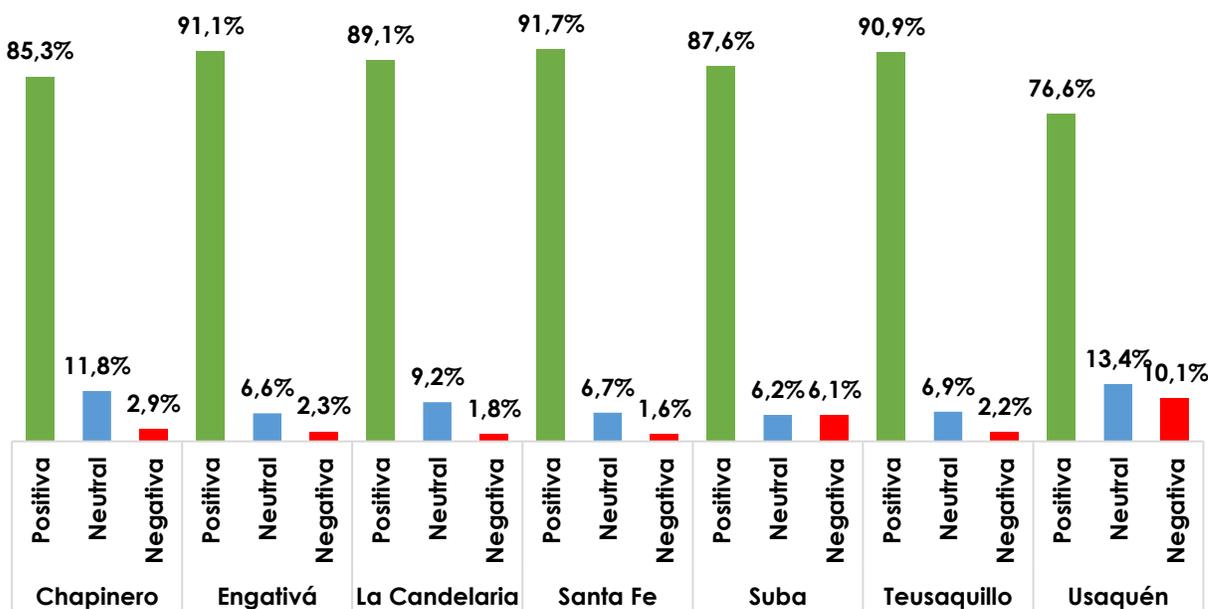
Como se había mencionado anteriormente, se categorizaron las reseñas de acuerdo con la calificación otorgada por el usuario, donde los puntajes menores o iguales a dos se clasifican como reseñas negativas, valores de tres corresponden con reseñas neutrales y puntajes mayores a 3 como reseñas positivas.

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 Chapinero, La Candelaria, y Suba para las cuales el porcentaje de reseñas positivas se encuentra entre el 85% y 90%.

Por último, se encuentra la localidad de Usaquén para la cual se observa un porcentaje de reseñas positivas cercana al 77% y el porcentaje más alto de reseñas negativas 10,1%, seguido de la localidad de Suba con un 6,1% de reseñas negativas.

Gráfica 11. Clasificación de las reseñas según su tipo desagregado por localidad (2006-2024-I)



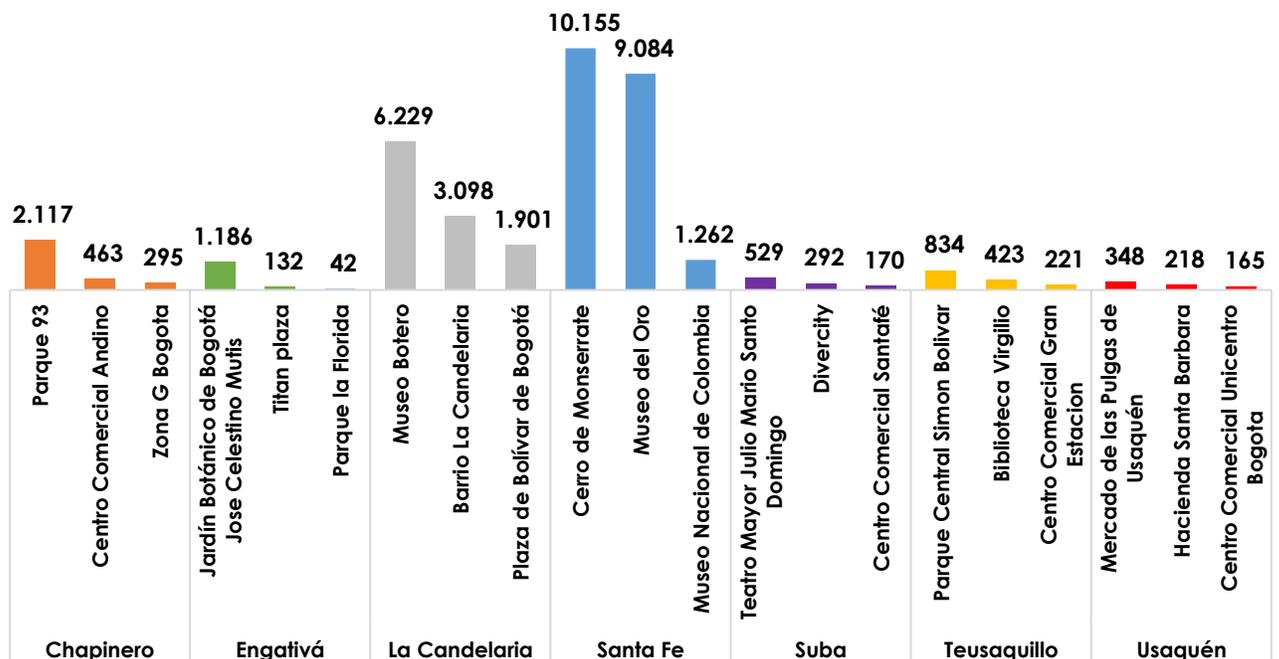
Fuente: TripAdvisor 2024.

Continuando con el análisis, la gráfica 12 presenta los 3 atractivos con el mayor número de reseñas para cada una de las siete localidades analizadas, donde se observa como el Cerro de Monserrate y el Museo del Oro, ambos atractivos en la Localidad de Santa Fe, son los más reseñados (10.155 y 9.084 respectivamente). Al mirar para cada localidad, en Chapinero es el Parque de la 93 (2.117 reseñas), en Engativá el atractivo con más reseñas es el Jardín Botánico de Bogotá (1.186 reseñas), para la Candelaria es el Museo Botero (6.229), en Suba es el Teatro Mayor Mario Julio Santo Domingo (529 reseñas), para Teusaquillo es el Parque Simón Bolívar (834 reseñas) y Finalmente para Usaquén es el Mercado de las Pulgas con 348 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 Andino (Chapinero), el Centro

Comercial Titan Plaza (Engativá), 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.

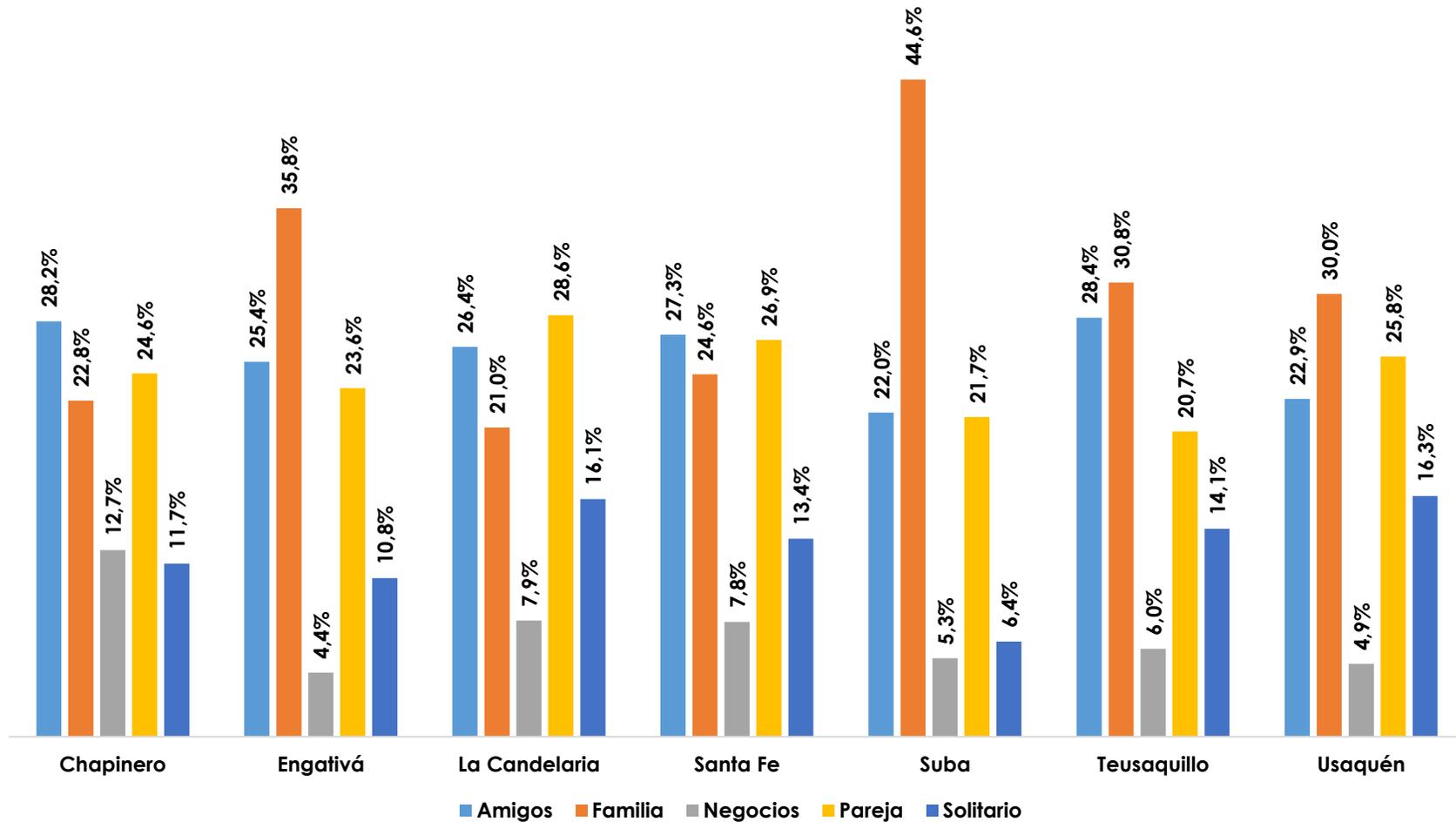
Gráfica 12. Atractivos con el mayor número de reseñas desagregado por localidad (2006-2024-I)



Fuente: TripAdvisor 2024.

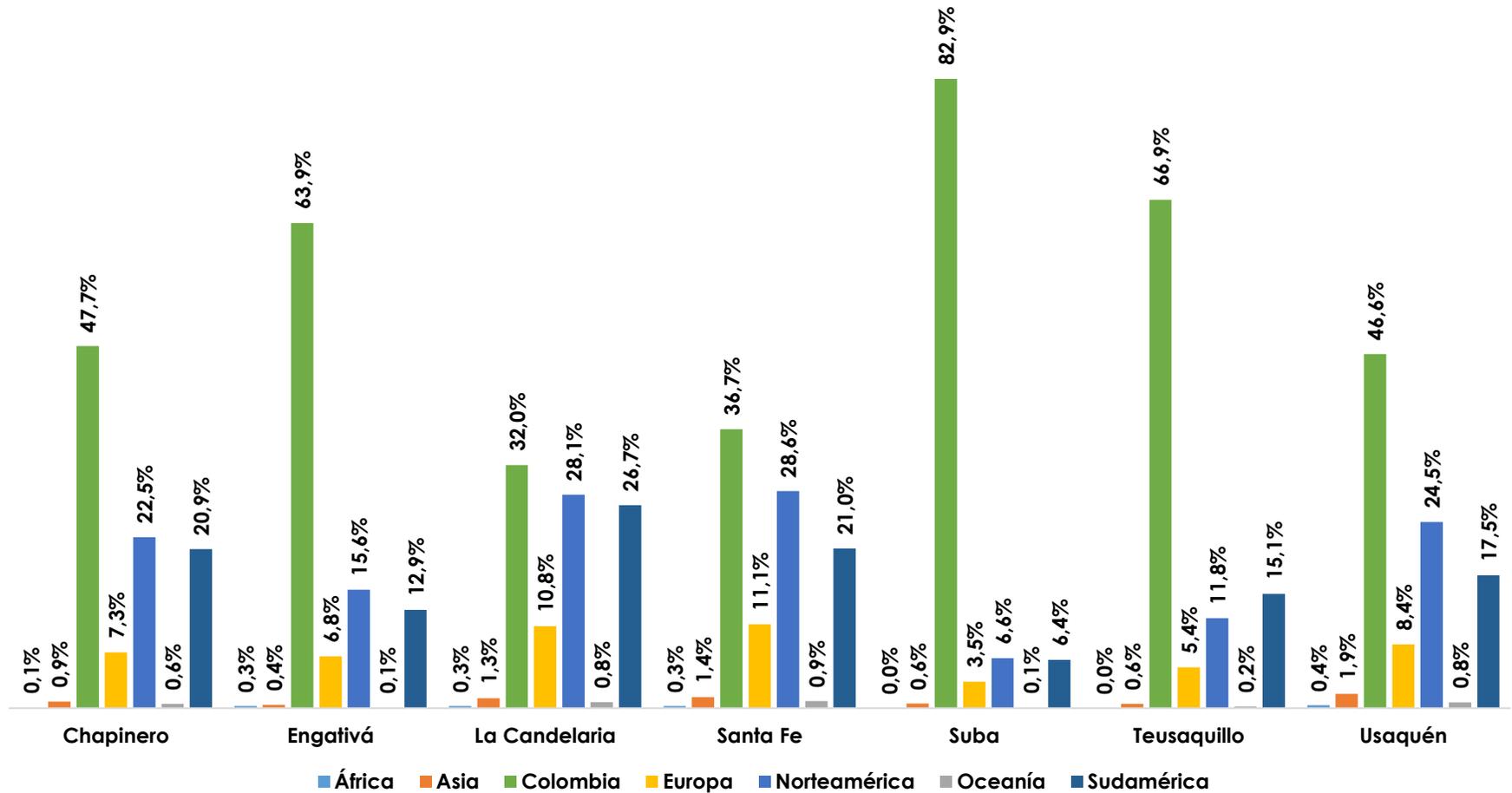
Respecto al motivo de viaje en las localidades de Engativá, Suba, Teusaquillo y Usaquén los visitan principalmente las personas que asisten con la familia, mientras que para las localidades de Chapinero y Santa Fe son principalmente los grupos de amigos los que visitan y reseñan los atractivos, por último, se observa que para La Candelaria son las personas que visitan con la pareja el grupo de viaje predominante, esta tendencia para las 7 localidades no ha cambiado en el transcurso de los dos últimos años.

Gráfica 13. Grupo de Viaje de los visitantes a los atractivos turísticos desagregado por localidad (2006-2024-I)



Fuente: TripAdvisor 2024.

Gráfica 14. Continente de origen de los visitantes a los atractivos turísticos desagregado por localidad (2006-2024-I)



Fuente: TripAdvisor 2024.

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

También para todas las localidades, menos para Teusaquillo el segundo continente de origen de los visitantes es Norteamérica¹² y en algunos casos como en las localidades de Chapinero, La Candelaria, Santa Fe y Usaquén representan más del 20% de los visitantes que reseñaron los atractivos.

De igual manera la gráfica 15 muestra el origen de los visitantes para cada localidad seleccionada, pero de acuerdo con el país de origen, encontrando que los atractivos de las localidades de Engativá, suba y Teusaquillo son visitados y reseñados en una mayor medida por visitantes de origen nacional, alcanzando estos una representatividad mayor al 60%.

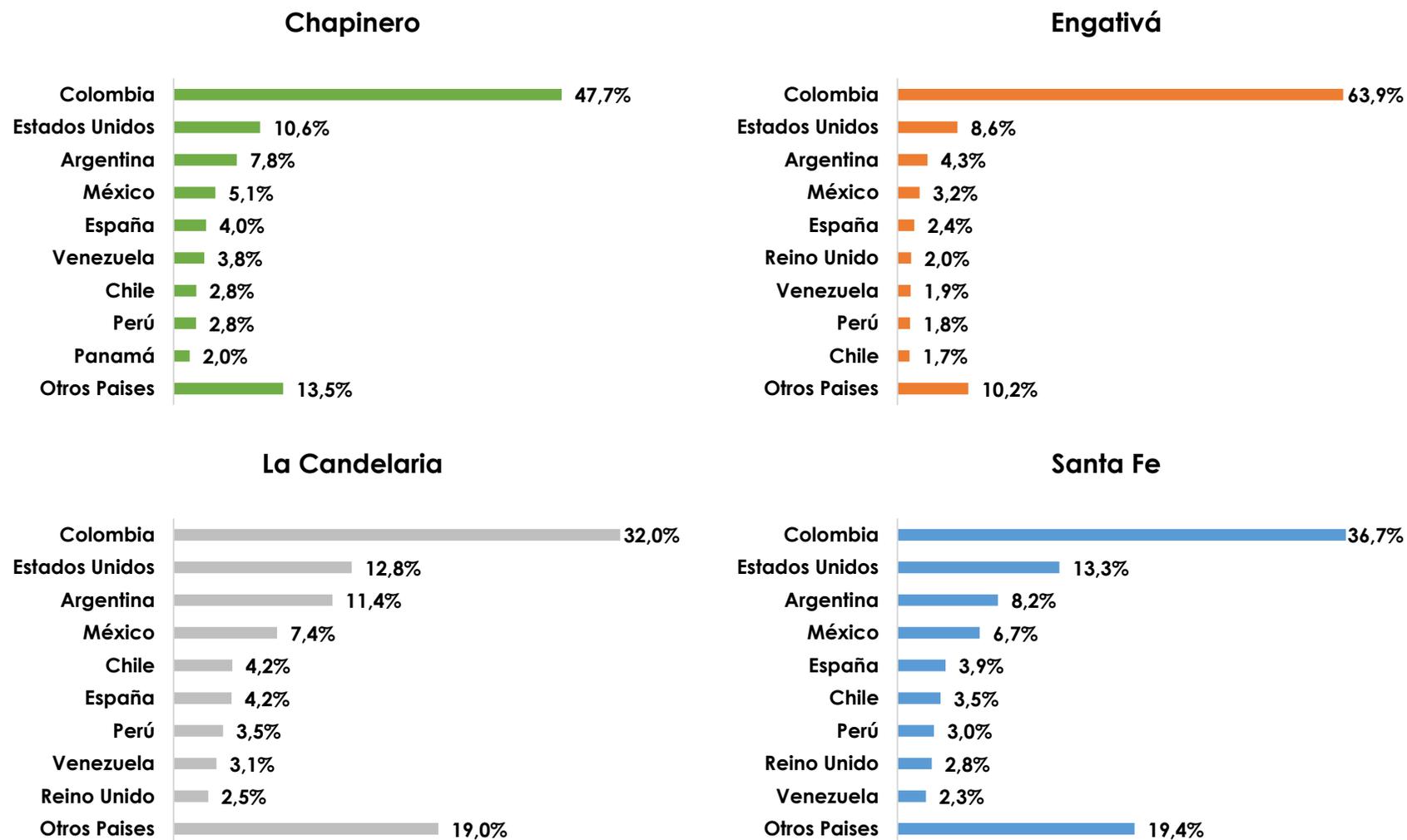
Por otro lado, para localidades como La Candelaria y Santa Fe los visitantes de origen nacional que reseñan los atractivos no representan un porcentaje tan elevado respecto al total de visitantes (menos del 40%), lo que podría indicar que tienen atractivos turísticos más interesantes para los turistas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

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

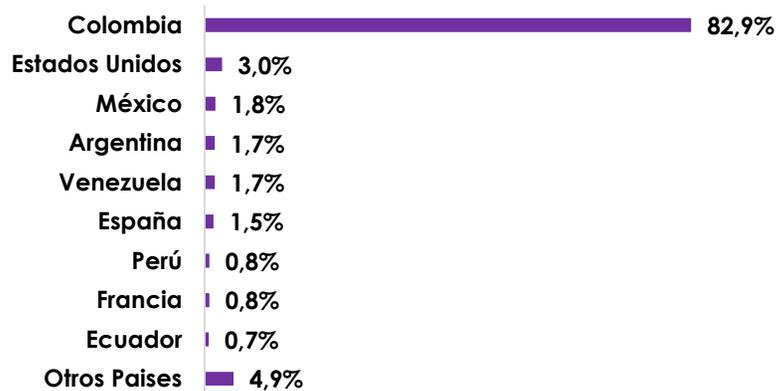
Además, se observa que para todas las localidades excepto para Teusaquillo, el segundo país de origen de donde provienen más reseñas es Estados Unidos, donde para algunas localidades representa más del 10% del total reseñas.

¹² Esta categoría incluye países de Norteamérica, Centro América y el Caribe.

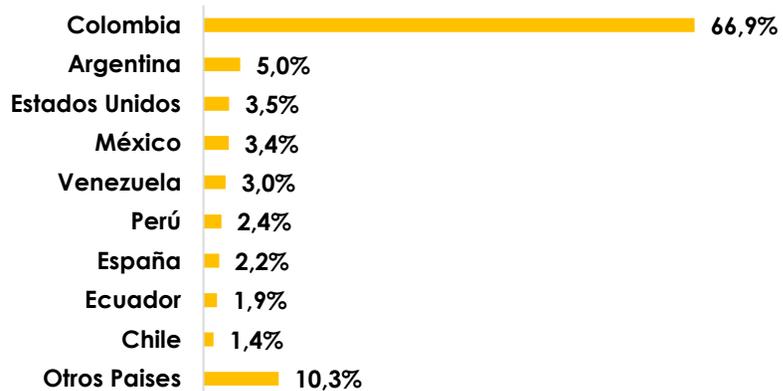
Gráfica 15. Principales países de origen de los visitantes a los atractivos turísticos desagregado por localidad (2006-2024-I)



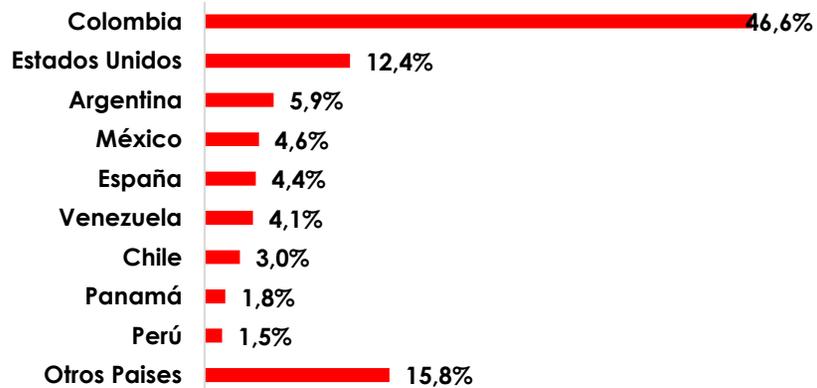
Suba



Teusaquillo



Usaquén



Fuente: TripAdvisor 2024.

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 seleccionados (el primero desde 2006 hasta 2021 y el segundo para el 2022, 2023 y el primer trimestre de 2024.). 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.

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 “restaurantes”, “comida”, “bares”, “ambiente” y “agradables” como factores que los visitantes resaltan de este tipo de establecimientos.

Para Engativá cuyo atractivo con más reseñas es el Jardín Botánico con el 85% del total de reseñas, se identificaron palabras como “jardines”, “variedad”, “especies”, “plantas” y “flores” las cuales están relacionadas con elementos que se pueden encontrar en este atractivo. Por otro lado, para las localidades de La Candelaria y Santa Fe se identificaron palabras como “arte”, “artistas”, “antiguo”, “arquitectura” y “ambiente”, evidenciando la importancia que tienen la oferta cultural y de museos (Museo del Oro, Museo Botero, Museo Nacional entre otros) en ambas localidades.

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

Respecto a las reseñas negativas de las localidades seleccionadas se encuentran palabras como "Seguridad", "servicio", "experiencia" problemáticas en las cuales los administradores de los atractivos y el distrito podría intervenir, con la finalidad de mejorar la percepción de los visitantes.

Las localidades de la Candelaria y Santa Fe se caracterizan también por sus reseñas negativas con palabras asociadas "fila", "vendedores" y "mendigos", que puede asociarse a los tiempos de espera para ingresar a los atractivos como con los vendedores ambulantes, artistas de la calle y personas que piden dinero presentes en la zona.

Por otro lado las localidades de Chapinero, y Usaquén se identificaron palabras como "centro", "comercial", "tiendas" y "restaurantes" conceptos que nuevamente están vinculados con el sector comercial y gastronómico de estas localidades, infiriendo problemáticas en las experiencias de los visitantes en estos tipos de atractivos.

5. CONCLUSIONES

- En este segundo acercamiento el uso de 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.
- Se identifica una caída en el número de reseñas escritas hacia atractivos ubicados en la ciudad de Bogotá ya que en el estudio del 2021 se extrajeron 49.955 reseñas (periodo de diciembre de 2006 a agosto de 2021), mientras que para esta segunda versión del estudio solo se cuenta con un incrementó de 777 en el total de reseñas las cuales fueron escritas entre septiembre de 2021 y marzo de 2024.
- Desde el mes de marzo del 2020 con la caída del turismo internacional y doméstico debido a las restricciones de movilidad, efecto de la pandemia, cae el número promedio de reseñas por mes a 20, para los atractivos turísticos de Bogotá seleccionados, la tendencia producto de esta caída a 2024 aún no ha cambiado.
- Se observa que desde el tercer trimestre de 2023 hasta el primer trimestre de 2024, la calificación promedio de las reseñas escritas para los atractivos de Bogotá seleccionados cae por debajo de 4, presentando una tendencia decreciente para estos tres trimestre y alcanzando el menor valor registrado (3,4 promedio trimestre) en el primer trimestre de 2024.
- 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.

INDICE DE TABLAS, FIGURAS Y GRÁFICOS

Índice de Tablas

Tabla 1. Trabajos en el que se utilizan las reseñas de los usuarios para el análisis los destinos Turísticos.	6
Tabla 2. Estructura de los datos que se descargaron	10

Índice de Figuras

Figura 1. Big Data en la investigación turística	5
Figura 2. Nube de palabras asociadas a las reseñas positivas y negativas (2006-2021) y (2022-2024-I)	21
Figura 3. Nube de palabras asociadas a las reseñas positivas y negativas por localidad (2006-2024-I)	32

Índice de Gráficos

Gráfica 1. Número de atractivos turísticos por localidad (2006-2024-I).....	11
Gráfica 2. Número de reseñas por localidad (2006-2024-I)	12
Gráfica 3. Atractivos con el mayor porcentaje de reseñas (2006-2024-I) ...	13
Gráfica 4. Número de reseñas por mes (2006-2024-I).....	14
Gráfica 5. Número de reseñas por mes (2019-2024-I).....	14
Gráfica 6. Grupo de Viaje de los visitantes a los atractivos turísticos (2006-2024-I)	15
Gráfica 7. Continente de origen de los visitantes a los atractivos turísticos (2006-2024-I)	16
Gráfica 8. Principales países de origen de los visitantes a los atractivos turísticos (2006-2024-I)	17
Gráfica 9. Clasificación de las reseñas según su tipo (2006-2024-I)	18
Gráfica 10. Calificación promedio de las reseñas por trimestre (2012-2024-I)	19
Gráfica 11. Clasificación de las reseñas según su tipo desagregado por localidad (2006-2024-I).....	23
Gráfica 12. Atractivos con el mayor número de reseñas desagregado por localidad (2006-2024-I).....	24
Gráfica 13. Grupo de Viaje de los visitantes a los atractivos turísticos desagregado por localidad (2006-2024-I)	25
Gráfica 14. Continente de origen de los visitantes a los atractivos turísticos desagregado por localidad (2006-2024-I)	26
Gráfica 15. Principales países de origen de los visitantes a los atractivos turísticos desagregado por localidad (2006-2024-I)	28

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.

Aggarwal, S. & Gour, A. (2020). Peeking inside the minds of tourists using a novel web analytics approach. *Journal of Hospitality and Tourism Management*, 45, 580-591.

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.

Kim, W., Kim, S.B. y Park, E. (2021). Mapping tourists' destination (dis) satisfaction attributes with user-generated content. *Sustainability*, 13, 12650.

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.

Sánchez Vargas, E., Campón Cerro, A. M., Di Clemente, E, & Prado Recio, E. (2024). Evaluación de la imagen de un destino de turismo rural a través del UGC. *Investigaciones Turísticas* (27), 105-128.

Safarian, A.A. & Konyshv, E. V. (2022). Geographical and Component Analysis of the Perception of the Tourism and Recreational Space of the Perm Region. *Bulletin of the Serbian Geographical Society*, 102(2), 141-158.

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.



OBSERVATORIO
DE TURISMO DE BOGOTÁ
IST



ALCALDÍA MAYOR
DE BOGOTÁ D.C.

INSTITUTO DISTRIAL DE
TURISMO

