

COMO SE PUEDEN POTENCIAR ENTRE SÍ LAS ATRACCIONES TURÍSTICAS DE CHILOÉ

POR:

ARIEL LISER GONZÁLEZ LILLO
DAVID RODRIGO PASTEN CARRASCO

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dr. Cristian Candia-Castro Vallejos

Santiago, 13 de diciembre de 2024

A Eva por el apoyo recibido

David

*Quiero agradecerme a mí mismo, porque, sin
mí, nada de esto sería posible.*

Ariel

AGRADECIMIENTO

Queremos expresar nuestro más sincero agradecimiento a todas las personas que nos apoyaron y colaboraron en la realización de este proyecto.

En primer lugar, agradecemos a nuestras familias, quienes con su apoyo incondicional y motivación constante nos permitieron culminar este importante desafío académico.

También extendemos nuestra gratitud a nuestros profesores y guías académicos, quienes con su experiencia, paciencia y conocimiento nos orientaron a lo largo de este trabajo.

Finalmente, agradecemos a nuestros compañeros de equipo, cuyo esfuerzo, dedicación y colaboración permitieron que este proyecto fuera un éxito. Este logro no habría sido posible sin el trabajo conjunto y el compromiso mutuo.

Este proyecto es el resultado del esfuerzo colectivo y representa un paso importante en nuestra formación profesional.

TABLA DE CONTENIDO

1.	RESUMEN	1
2.	INTRODUCCIÓN.....	2
3.	TRABAJO RELACIONADO	3
	<i>Evolution and Driving Mechanism of Tourism Flow Networks in the Yangtze River Delta Urban Agglomeration</i>	<i>3</i>
	<i>Study on Tourism Flow Network Patterns on May Day Holiday.....</i>	<i>4</i>
4.	HIPÓTESIS Y OBJETIVOS.....	6
4.1.	HIPÓTESIS	6
4.2.	OBJETIVOS	7
	<i>Objetivo General:</i>	<i>7</i>
	<i>Objetivos Específicos:.....</i>	<i>7</i>
5.	DATOS Y METODOLOGÍA.....	8
5.1.	METODOLOGÍA.....	8
5.2.	DATOS.....	9
6.	RESULTADOS.....	12
6.1.	ANÁLISIS EXPLORATORIO.....	12
	<i>Métodos y Técnicas Utilizadas</i>	<i>12</i>
	<i>Estadísticos Descriptivos y Descripción de datos.....</i>	<i>13</i>
	<i>Análisis Multivariable.....</i>	<i>13</i>
	<i>Análisis univariable</i>	<i>16</i>
6.2.	METODOLOGÍA IMPLEMENTADA	26
6.3.	RESULTADOS.....	27
	<i>Análisis de los Resultados a 50 km</i>	<i>27</i>
	<i>Análisis de los Resultados a 25 km</i>	<i>31</i>

<i>Análisis de los Resultados a 10 km</i>	34
<i>Análisis de los Resultados a 5 km</i>	37
<i>Análisis de los Resultados a 1 km</i>	40
<i>Análisis Completo de las Hipótesis</i>	42
7. CONCLUSIONES	43
BIBLIOGRAFÍA	45
8. ANEXOS	¡ERROR! MARCADOR NO DEFINIDO.
8.1. CODIGO.....	¡ERROR! MARCADOR NO DEFINIDO.
<i>Importación de bibliotecas necesarias</i>	47
<i>PASO 1: Cargar la base de datos</i>	47
<i>PASO 2: Filtrar columnas relevantes</i>	47
<i>PASO 3: Calcular distancias entre atracciones</i>	47
<i>PASO 4: Crear proxy para flujo basado en reseñas mínimas</i>	48
<i>PASO 5: Hipótesis 1 - Relación entre distancia y flujo</i>	48
<i>PASO 6: Construcción de la red y análisis de centralidad</i>	49
<i>PASO 7: Hipótesis 2 - Relación entre centralidad y valoraciones/reseñas</i>	49
<i># PASO 8: Resumen de resultados</i>	50
Tabla 1: Detalles del Dataset	10
Tabla 2: Especificación del Dataset por columnas	11
<i>Tabla 3: Cantidad de NULL/Vacío en el Dataset</i>	14
<i>Tabla 4: Tabla de UNIQUES de columnas del Dataset</i>	15
<i>Tabla 5: Análisis univariable cuantitativo de stars</i>	22
<i>Tabla 6: Análisis univariable cuantitativo de reviews</i>	24
<i>Tabla 7: Tabla de correlación 50km</i>	29
<i>Tabla 8: Tabla de correlación 25km</i>	33
<i>Tabla 9: Tabla de correlación 10km</i>	36

<i>Tabla 10: Tabla de correlación 5km</i>	39
<i>Tabla 11: Tabla de correlación 1km</i>	42
<i>Gráfico 1: Tendencia de reseñas a lo largo del tiempo</i>	17
<i>Gráfico 2: Ciudades con más reseñas</i>	18
<i>Gráfico 3: Distribución de calificaciones</i>	19
<i>Gráfico 4: Top 10 tipos de lugares por cantidad de reseñas</i>	20
<i>Gráfico 5: Calificaciones vs likes en reseñas</i>	21
<i>Gráfico 6: Distribución de stars</i>	22
<i>Gráfico 7: Distribucion de reseñas por negocio</i>	24
<i>Gráfico 8: Distancia vs flujo 50km</i>	28
<i>Gráfico 9: Centralidad vs valoración 50km</i>	28
<i>Gráfico 10: Centralidad vs Reseñas 50km</i>	29
<i>Gráfico 11: Distancia vs Flujo 25km</i>	31
<i>Gráfico 12: Centralidad vs Valoración 25km</i>	32
<i>Gráfico 13: Centralidad vs Reseñas 25km</i>	32
<i>Gráfico 14: Distancia vs Flujo 10km</i>	34
<i>Gráfico 15: Centralidad vs Valoración 10km</i>	35
<i>Gráfico 16: Centralidad vs Reseñas 10km</i>	35
<i>Gráfico 17: Distancia vs Flujo 5km</i>	37
<i>Gráfico 18: Centralidad vs Valoración 5km</i>	38
<i>Gráfico 19: Centralidad vs Reseñas 5km</i>	38
<i>Gráfico 20: Distancia vs Flujo 1km</i>	40
<i>Gráfico 21: Centralidad vs Valoración 1km</i>	41
<i>Gráfico 22: Centralidad vs Reseñas 1km</i>	41

Resumen

Este estudio analizó el flujo turístico entre las atracciones de Chiloé a distintas distancias (50 km, 25 km, 10 km, 5 km y 1 km) para evaluar dos hipótesis clave: si la proximidad geográfica tiene un impacto en el flujo de visitantes y si las atracciones más centrales (hubs) tienden a tener mejores valoraciones y más reseñas. A lo largo de los distintos rangos de distancia, los resultados revelaron que la proximidad geográfica no muestra una relación significativa con el flujo de visitantes. Las correlaciones entre distancia y flujo fueron consistentemente bajas o cercanas a cero, lo que sugiere que la proximidad física no es un factor determinante en el flujo turístico. En cuanto a la centralidad, las atracciones más centrales tienden a acumular más reseñas, lo que confirma su papel como hubs en la red turística. Sin embargo, la relación entre centralidad y valoración promedio fue negativa en varias distancias, lo que indica que, aunque estas atracciones son populares y frecuentadas, no necesariamente son las mejor valoradas, posiblemente debido a la masificación o a expectativas no cumplidas. A medida que se redujo la distancia, la centralidad mostró una relación más fuerte con las reseñas, aunque la valoración promedio siguió siendo más baja en las atracciones centrales. Las limitaciones del estudio incluyen la falta de información temporal y sociodemográfica de los turistas, así como posibles sesgos en los datos de reseñas, que podrían no reflejar de manera equitativa a todas las atracciones. Además, las correlaciones lineales utilizadas no capturan posibles relaciones no lineales o indirectas entre las variables. En el futuro, investigaciones más profundas podrían incluir modelos de análisis más complejos, como simulaciones de redes dinámicas, y considerar factores temporales y demográficos para entender mejor las decisiones de los turistas. En conclusión, los hallazgos sugieren que la proximidad geográfica no explica de manera significativa el flujo turístico, mientras que la centralidad en la red sí influye en la cantidad de reseñas, aunque no necesariamente en la calidad percibida de las atracciones.

1. Introducción

Este estudio se centra en analizar los patrones de flujo turístico entre las principales atracciones de Chiloé, una región de Chile conocida por su rica cultura, paisajes naturales y patrimonio histórico. Con el objetivo de comprender cómo las características geográficas y la centralidad de las atracciones influyen en el comportamiento de los turistas, el trabajo propone dos hipótesis principales: la primera, que la proximidad geográfica entre las atracciones tiene un efecto positivo en la frecuencia de visitas, y la segunda, que las atracciones con mejores valoraciones tienden a funcionar como hubs dentro de una red turística, acumulando más reseñas y atrayendo a más visitantes. Para abordar estas hipótesis, se utiliza un enfoque cuantitativo basado en el análisis de datos de reseñas turísticas obtenidas de diversas plataformas online. Se exploran métricas de centralidad, como la centralidad de grado, para medir la conectividad de las atracciones dentro de la red turística, y se calculan correlaciones entre la distancia entre las atracciones y el flujo de visitantes, que se mide a través del número de reseñas publicadas por los turistas. El análisis se realiza a diferentes distancias (50 km, 25 km, 10 km, 5 km y 1 km) para identificar cómo la proximidad influye en las relaciones de flujo entre las atracciones. Este enfoque permite observar patrones de conectividad que van más allá de las simples distancias geográficas y facilita la comprensión de cómo factores como la accesibilidad, la infraestructura y la popularidad de las atracciones pueden determinar el flujo de turistas. Los resultados obtenidos buscan aportar una visión más profunda sobre los factores que afectan la dinámica del turismo en Chiloé, ofreciendo recomendaciones que podrían ser útiles para la gestión y promoción de los destinos turísticos en la región. Además, este análisis proporciona una base sólida para futuras investigaciones en la red de atracciones turísticas, extendiendo el estudio a otras áreas geográficas y utilizando metodologías complementarias que puedan enriquecer la comprensión sobre el comportamiento del turista en espacios urbanos y rurales.

2. Trabajo Relacionado

El estudio del flujo turístico y la conectividad entre atracciones ha sido ampliamente abordado en la literatura, con enfoques que integran herramientas como el análisis de redes sociales (SNA), sistemas de información geográfica (GIS) y metodologías basadas en big data. Estas investigaciones ofrecen perspectivas valiosas sobre cómo factores como la proximidad geográfica, la calidad de las atracciones y la infraestructura afectan el comportamiento turístico y la estructura de las redes de movilidad. A continuación, se resumen dos estudios relevantes que constituyen una base teórica y metodológica clave para el análisis del flujo turístico en Chiloé. Estos artículos no solo exploran las dinámicas espaciales y sociales de las redes turísticas, sino que también destacan estrategias replicables para optimizar la conectividad y fomentar un desarrollo turístico sostenible.

Evolution and Driving Mechanism of Tourism Flow Networks in the Yangtze River Delta Urban Agglomeration

El artículo titulado "Evolution and Driving Mechanism of Tourism Flow Networks in the Yangtze River Delta Urban Agglomeration Based on Social Network Analysis and Geographic Information System: A Double-Network Perspective" analiza la evolución de las redes de flujo turístico en la región del Delta del Río Yangtze y explora los factores que impulsan su desarrollo. Publicado en la revista *Sustainability* en 2022 por Yuewei Wang, Mengmeng Xi, Hang Chen y Cong Lu, este estudio utiliza una combinación innovadora de análisis de redes sociales (SNA) y sistemas de información geográfica (GIS) para estudiar las características estructurales y los patrones de conectividad entre destinos turísticos clave.

El enfoque adoptado en el paper destaca cómo la proximidad geográfica y la calidad de las atracciones influyen en el comportamiento turístico. La metodología basada en GIS permite mapear las interacciones espaciales entre destinos, mostrando que la cercanía física fomenta visitas consecutivas entre atracciones. Este hallazgo está respaldado por datos que demuestran que los destinos cercanos tienden a ser visitados juntos con mayor frecuencia, lo que subraya la importancia de la accesibilidad geográfica como motor del flujo turístico. Por otro lado, el análisis de redes sociales permite identificar los nodos centrales dentro de la red turística. Estos nodos, que suelen ser atracciones altamente valoradas, actúan como hubs que concentran el flujo de visitantes y facilitan la conexión hacia otros destinos menos conocidos. Esto sugiere que las atracciones con mayor reconocimiento o valoración cumplen un papel estructural fundamental en la red, promoviendo la dispersión del turismo a través de la región.

El enfoque de "doble red" presentado en el estudio es particularmente útil, ya que integra dinámicas espaciales y sociales para ofrecer una visión comprensiva de los flujos turísticos. Los resultados del paper tienen importantes implicaciones prácticas, ya que permiten a los gestores de destinos identificar oportunidades para optimizar la

conectividad entre atracciones, fomentar el desarrollo equilibrado y aprovechar los nodos principales para distribuir mejor los beneficios económicos del turismo. Además, el estudio destaca que factores como la infraestructura de transporte, la calidad de los servicios turísticos y las estrategias de promoción tienen un papel clave en la evolución de estas redes.

La relación de este artículo con las hipótesis planteadas es directa y significativa. La primera hipótesis, que sostiene que la proximidad geográfica tiene un efecto positivo en la frecuencia de visitas entre atracciones, encuentra respaldo en la evidencia presentada, donde la accesibilidad se posiciona como un factor determinante del comportamiento de los turistas. Asimismo, la segunda hipótesis, que sugiere que las atracciones mejor valoradas funcionan como hubs dentro de la red, también es coherente con los hallazgos del estudio, ya que estos nodos centrales no solo atraen un alto volumen de turistas, sino que también facilitan el flujo hacia destinos conectados.

En el contexto de Chiloé, este marco metodológico puede aplicarse para analizar el flujo turístico entre islas y atracciones locales, utilizando GIS para mapear patrones espaciales y SNA para entender las dinámicas de conectividad. La comprensión de estos factores permitiría desarrollar estrategias que fortalezcan la red turística regional, incrementen el valor percibido de los destinos y promuevan un turismo más sostenible. La investigación discutida ofrece un enfoque robusto y replicable, proporcionando un marco valioso para avanzar en el análisis de flujos turísticos en diferentes contextos.

Study on Tourism Flow Network Patterns on May Day Holiday

El artículo "Study on Tourism Flow Network Patterns on May Day Holiday" analiza cómo se estructuran los flujos turísticos durante un feriado en China, empleando datos masivos proporcionados por Tencent para observar la movilidad de los turistas a nivel nacional. Este estudio se centra en la distribución espacial de los flujos turísticos y clasifica a las ciudades según su papel en la red turística: como destinos principales, ciudades emisoras o puntos de tránsito. La investigación resalta la importancia de las conexiones entre orígenes y destinos y analiza las diferencias en el volumen de los flujos y los factores que los impulsan, como la accesibilidad, la oferta turística y la infraestructura.

El análisis realizado en este artículo se basa en una metodología de big data que permite capturar patrones de comportamiento en tiempo real, ofreciendo un nivel de detalle sin precedentes sobre cómo los turistas se desplazan en el espacio geográfico. Los datos obtenidos permiten identificar patrones de red jerárquicos donde ciertas ciudades actúan como hubs que concentran y distribuyen el flujo turístico. Estos hubs juegan un papel esencial en la conectividad general de la red, ya que enlazan destinos periféricos con áreas centrales de alta actividad turística. Además, el estudio subraya cómo las dinámicas de movilidad están influenciadas por factores socioeconómicos, la infraestructura de transporte y las políticas locales de promoción turística.

La relevancia de este artículo para el análisis del flujo turístico en Chiloé radica en su enfoque metodológico y en los conceptos clave que se pueden aplicar en un contexto local. En particular, el uso de datos masivos para mapear y analizar los patrones de movilidad de los turistas es una estrategia replicable para entender la conectividad entre las diversas atracciones de Chiloé. Además, la clasificación de las ciudades en función de su rol dentro de la red puede inspirar una segmentación de las atracciones turísticas en Chiloé, identificando cuáles actúan como puntos centrales de llegada, distribución o destino final. El estudio también resalta la importancia de factores como la proximidad geográfica, la infraestructura y la calidad de las atracciones, los cuales están directamente relacionados con las hipótesis propuestas. Por ejemplo, la hipótesis sobre la proximidad geográfica como factor que incrementa la frecuencia de visitas entre atracciones es consistente con los hallazgos del artículo, que señala cómo las distancias más cortas entre nodos aumentan la probabilidad de conexiones fuertes. Asimismo, la hipótesis sobre las atracciones con mejores valoraciones funcionando como hubs encuentra soporte en la identificación de nodos clave dentro de la red turística, que no solo atraen un mayor volumen de visitantes, sino que también facilitan la redistribución del flujo hacia otros puntos. En el contexto de Chiloé, este enfoque permitiría identificar qué atracciones cumplen un rol nodal en la red turística y cómo optimizar las rutas y conexiones para fortalecer la experiencia del visitante y promover un desarrollo más equilibrado del turismo en la región. Al igual que el artículo sobre el Delta del Río Yangtze, este estudio aporta herramientas y conceptos metodológicos esenciales para entender las dinámicas del turismo en redes geográficas, especialmente en regiones que buscan maximizar su conectividad y sostenibilidad.

3. Hipótesis y Objetivos

3.1. Hipótesis

Las hipótesis planteadas en este análisis buscan explorar factores clave que influyen en el comportamiento de los turistas y en la dinámica de las redes de flujo turístico en Chiloé. Basándonos en teorías como la accesibilidad geográfica y los efectos de centralidad en redes complejas, proponemos dos hipótesis fundamentales para guiar el estudio.

La primera se centra en el impacto de la proximidad geográfica entre atracciones turísticas, mientras que la segunda examina el papel de las atracciones mejor valoradas como nodos principales dentro de la red turística. Ambas hipótesis están respaldadas por marcos teóricos sólidos y evidencias de estudios previos, que proporcionan una base para analizar las particularidades del turismo en la región. A continuación, se presentan las hipótesis en detalle y su relación con el contexto de Chiloé.

“La proximidad geográfica entre atracciones turísticas tiene un efecto positivo y significativo en la frecuencia con la que los visitantes se desplazan entre dichas atracciones.”

La teoría de la accesibilidad sugiere que los elementos geográficamente cercanos tienden a ser visitados de forma más frecuente, ya que la reducción en tiempo y costos de desplazamiento facilita la conexión entre ellos. En el caso de las atracciones turísticas, la cercanía física no solo fomenta el tránsito de los visitantes, sino que también incentiva la planificación de itinerarios que incluyan múltiples puntos de interés ubicados en una misma área. Estudios previos han mostrado que destinos agrupados en proximidad tienen mayor probabilidad de ser incluidos en el mismo recorrido turístico, debido a la percepción de conveniencia y la mayor optimización del tiempo.

“Las atracciones turísticas con mejores valoraciones funcionan como nodos principales dentro de la red turística, facilitando el flujo de visitantes hacia otras atracciones conectadas.”

En redes complejas, los nodos que poseen mayor prestigio o calidad tienden a ocupar posiciones centrales que permiten una mayor conectividad con otros elementos. Aplicado al turismo, se ha observado que las atracciones con altas valoraciones o reconocimientos actúan como puntos de anclaje dentro de un destino, atrayendo visitantes que posteriormente exploran otras atracciones cercanas. Este fenómeno se relaciona con la confianza que los turistas depositan en opiniones previas y con la capacidad de estas atracciones para generar "spillover effects" hacia lugares menos conocidos, amplificando el dinamismo en la red turística. Por lo tanto, la calidad percibida de una atracción influye tanto en su capacidad para atraer visitantes como en su función de redistribuir hacia otros puntos de interés.

3.2. Objetivos

Objetivo General:

Investigar cómo la proximidad geográfica, la valoración de turistas y la diversidad de actividades influyen en la conectividad entre atracciones turísticas en Chiloé.

Objetivos Específicos:

- Identificar patrones de flujo entre las principales atracciones y su conexión.
- Evaluar el rol de las valoraciones turísticas en la configuración de hubs.
- Analizar la variación temporal de los flujos de turistas durante el año.
- Implementar un análisis de centralidad para detectar nodos clave en la red turística.

4. Datos y Metodología

4.1. Metodología

El análisis se llevó a cabo mediante un enfoque estructurado que combina análisis temporal, evaluación de métricas de conectividad, y el uso de herramientas avanzadas de análisis de redes y datos geográficos. A continuación, se describen los principales pasos y enfoques metodológicos adoptados:

1. Análisis Temporal del Flujo Turístico

Para comprender cómo varía el flujo de turistas a lo largo del tiempo, se llevó a cabo un análisis detallado de las visitas según diferentes temporadas y periodos relevantes. Este análisis incluyó la identificación de patrones estacionales asociados con el verano, el invierno y otras festividades locales, permitiendo determinar los picos de actividad turística y las variaciones en el comportamiento de los visitantes.

Además, se analizaron los datos de trayectorias acumuladas durante los 13 años de registro para detectar posibles tendencias de largo plazo y cómo eventos específicos pudieron haber impactado en los patrones de conectividad y flujo turístico en la región.

2. Análisis de Centralidad y Conectividad

Para identificar las atracciones más relevantes dentro de la red de conectividad turística, se utilizaron métricas de análisis de redes como la **centralidad de intermediación**, que mide la influencia de una atracción en el flujo entre otras localidades. Este enfoque permitió distinguir las atracciones que actúan como "hubs" o puntos clave de conexión en la región.

Adicionalmente, se emplearon otras métricas de redes como la centralidad de grado (para identificar los sitios más directamente conectados) y análisis de comunidades para segmentar la red en clústeres de atracciones interrelacionadas, destacando cómo los visitantes se mueven entre sitios específicos y agrupaciones geográficas o temáticas.

3. Herramientas y Tecnologías Utilizadas

El análisis se realizó utilizando herramientas avanzadas de programación y visualización:

- **NetworkX:** Para la creación, análisis y visualización de redes turísticas, permitiendo calcular métricas de conectividad y generar mapas de redes.

- **GeoPandas:** Para manejar y analizar datos geográficos, facilitando la integración de las coordenadas de las atracciones turísticas y su representación espacial.
- **Matplotlib:** Para la visualización gráfica de patrones temporales, métricas de centralidad y redes de conectividad, generando gráficos que ayudan a interpretar los resultados de manera intuitiva.

Estas herramientas fueron seleccionadas por su flexibilidad, capacidad para manejar grandes volúmenes de datos, y su compatibilidad con el enfoque interdisciplinario necesario para integrar análisis espaciales, de redes y temporales en un solo marco metodológico

4.2. Datos

Para este análisis, se utilizaron datos proporcionados mediante la plataforma de almacenamiento Dropbox mediante este enlace (<https://www.dropbox.com/scl/fo/cj1g1n8gl8p3oq8h5ucun/AI6mBbSvFTdPqMyRjiSwBYA?rlkey=98fe817dtgvxdg57vfeufsmn3&e=2&dl=0>) .Estos datos se encuentran recopilados en un archivo Excel titulado *Base de datos Chiloé Full.xlsx*, el cual contiene un total de 701,964 registros que documentan una trayectoria temporal de 13 años.

El archivo incluye información detallada sobre las reseñas y calificaciones realizadas por turistas en diversas atracciones de la región. Estas observaciones abarcan múltiples aspectos, tales como la ubicación geográfica de las atracciones, los comentarios textuales de los visitantes, las calificaciones individuales asignadas a los sitios turísticos, y datos adicionales relacionados con los tipos de atracciones y las ciudades donde se encuentran.

La riqueza y la profundidad de este conjunto de datos lo convierten en una base sólida para analizar el flujo turístico, identificar patrones de conectividad entre las principales atracciones, y evaluar los factores que influyen en el movimiento de los visitantes entre diferentes destinos de Chiloé y zonas cercanas.

Nombre del archivo	Cantidad de tuplas	Descripción	Años de información	Peso (KB)

Base de datos Chiloé Full.xlsx	70.1964	Reseñas de turistas en Chiloé	13 años	112.832 KB
-----------------------------------	---------	-------------------------------------	---------	------------

Tabla 1: Detalles del Dataset

En la siguiente tabla se tienen las columnas que componen al Dataset, se describe el nombre de la columna, el tipo de dato que posee, si tiene un largo establecido y una pequeña descripción acerca de lo que se refiere para mayor entendimiento.

Nombre del campo	Tipo de dato	Descripción
placeId	Text	Identificador único para cada lugar visitado
user name	Text	Nombre del usuario que escribió la reseña
Text	Text	Texto original de la reseña escrita por el usuario
textTranslated	Text	Traducción de la reseña al inglés o a otro idioma
publishedAtDate	Date	Fecha y hora en que se publicó la reseña.
likesCount	Number	Número de "me gusta" que recibió la reseña.
reviewId	Text	Identificador único de cada reseña
reviewerId	Text	Identificador único del usuario que escribió la reseña
stars	Number	Calificación de estrellas dada por el usuario al lugar, en una escala de 1 a 5.
responseFromOwnerDate	Text	Fecha y hora de la respuesta del propietario a la reseña
responseFromOwnerText	Text	respuesta proporcionada por el propietario del lugar a la reseña del usuario
hotelStars	Number	Clasificación del hotel, especificando el número de estrellas
query	Text	Consulta o categoría del lugar, indicando el tipo de establecimiento y su ubicación
place name	Text	Consulta o categoría del lugar, indicando el tipo de establecimiento y su ubicación

type	Text	Consulta o categoría del lugar, indicando el tipo de establecimiento y su ubicación
subtypes	Text	Consulta o categoría del lugar, indicando el tipo de establecimiento y su ubicación
city	Text	Ciudad donde se encuentra el lugar mencionado en la reseña
latitude	Number	Latitud geográfica del lugar
longitude	Number	Longitud geográfica del lugar
rating	Number	Calificación promedio del lugar
reviews	Number	Número total de reseñas para el lugar

Tabla 2: Especificación del Dataset por columnas

5. Resultados

5.1. Análisis exploratorio

El análisis exploratorio de datos es una etapa fundamental en cualquier proyecto de análisis de datos. Durante esta fase, se emplean diversas técnicas estadísticas y visuales para resumir, comprender y explorar las características principales de los datos. Este proceso no solo ayuda a identificar patrones y tendencias, sino también a detectar valores atípicos, inconsistencias o datos faltantes que puedan influir en el análisis posterior.

Para llevar a cabo esta etapa, se utilizó el software Anaconda, una plataforma robusta que incluye una variedad de herramientas y bibliotecas esenciales para el análisis de datos. En particular, se trabajó con Anaconda Jupyter Lab, una interfaz interactiva que facilita la manipulación de datos y la creación de visualizaciones. Las bibliotecas principales utilizadas fueron:

- **Pandas:** Para la manipulación y análisis de datos tabulares, incluyendo funciones para calcular estadísticas descriptivas y transformar columnas específicas.
- **Matplotlib:** Para la generación de gráficos que permiten una representación visual intuitiva de las tendencias y distribuciones en los datos.

Métodos y Técnicas Utilizadas

1. Tablas de Frecuencia

Se generaron tablas de frecuencia para los valores cualitativos presentes en el conjunto de datos. Estas tablas incluyen la frecuencia absoluta y el porcentaje que representa cada categoría dentro del total de datos, proporcionando una visión clara de la distribución de estas variables.

2. Estadísticos Descriptivos

Para las variables cuantitativas, se calcularon estadísticas descriptivas como la media, mediana, desviación estándar, valores mínimos y máximos, entre otros. Estas medidas ayudan a comprender la tendencia central y la dispersión de los datos, además de identificar posibles anomalías o valores atípicos.

3. Visualizaciones

Basándose en las frecuencias y estadísticas obtenidas, se elaboraron diferentes tipos de gráficos:

- **Gráficos de Barras:** Ideales para visualizar la distribución de variables categóricas y su proporción relativa dentro del conjunto de datos.
- **Boxplots:** Utilizados para resumir gráficamente las principales estadísticas descriptivas de las variables cuantitativas, incluyendo la mediana, los cuartiles y la presencia de valores atípicos.

Estas técnicas permitieron construir una base sólida de conocimiento sobre los datos, facilitando el diseño de análisis más avanzados en las etapas posteriores del estudio. El uso de estas herramientas no solo garantizó un análisis eficiente, sino que también proporcionó representaciones visuales claras y precisas para respaldar las observaciones realizadas durante el proceso.

Estadísticos Descriptivos y Descripción de datos

Para el análisis se tomaron en cuenta valores estadísticos como la media, desviación estándar, mediana, moda, mínimo, máximo y cuartiles, tanto en atributos cualitativos como en atributos cuantitativos para poder analizar. Además, se realizarán análisis de los valores descriptivos de los datos, como por ejemplo el total de datos, total de datos nulos, total de datos erróneos, dominio y rangos.

Análisis Multivariable

Como parte del análisis exploratorio de datos, se generó una tabla que muestra la cantidad de valores faltantes (*null*) por columna en el conjunto de datos. Este análisis es crucial para identificar posibles problemas de calidad en los datos que podrían afectar los resultados del estudio, así como para determinar estrategias de limpieza y manejo de los valores faltantes.

Nombre del campo	Cantidad de NULL
placeId	0
user name	0
Text	357767
textTranslated	367826
publishedAtDate	0
likesCount	0
reviewId	0
reviewerId	110
stars	110
responseFromOwnerDate	674231
responseFromOwnerText	674706
hotelStars	649179
query	0
place name	0
type	0
subtypes	53
city	606
latitude	0
longitude	0
rating	0
reviews	0

Tabla 3: Cantidad de NULL/Vacío en el Dataset

Columnas Completas

- Algunas columnas, como placeId, user name, publishedAtDate, latitude, longitude, rating, y reviews, no presentan valores nulos, lo que garantiza datos consistentes y completos para estas variables.

Columnas con Mayor Cantidad de Nulos

- La columna responseFromOwnerText presenta una gran cantidad de valores nulos (674,706), lo que sugiere que en la mayoría de los casos no hay respuestas de los dueños a las reseñas de los usuarios.

- Similarmente, hotelStars contiene 649,179 valores faltantes, lo que podría deberse a que no todas las atracciones analizadas son hoteles.

Este análisis proporciona una visión clara de la calidad del conjunto de datos y guía las decisiones relacionadas con el preprocesamiento de los datos antes de realizar análisis más avanzados y tener en cuenta que datos utilizar para los análisis.

A la misma vez, se llevó a cabo un estudio de los valores únicos en cada columna del conjunto de datos. Este análisis es fundamental para comprender la diversidad de información contenida y su relevancia para los objetivos del estudio. A continuación, se detalla la cantidad de valores únicos identificados por columna:

Nombre del campo	Cantidad de Valores Unicos
placeId	2638
user name	228350
Text	283450
textTranslated	267183
publishedAtDate	692076
likesCount	38
reviewId	692097
reviewerId	262466
stars	5
responseFromOwnerDate	27578
responseFromOwnerText	21008
hotelStars	5
query	422
place name	2597
type	276
subtypes	586
city	170
latitude	2614
longitude	2610
rating	32
reviews	621

Tabla 4: Tabla de UNIQUES de columnas del Dataset

Alta Diversidad de Valores Únicos

- La columna reviewId tiene 692,097 valores únicos, lo que indica que cada reseña es única, garantizando granularidad en los análisis.
- Campos como publishedAtDate (692,076) y reviewerId (262,466) también presentan una alta diversidad, reflejando una amplia variedad de usuarios y fechas de publicación.

Variabilidad Moderada y Baja

- Variables como placeId (2,638) y place name (2,597) sugieren una amplia representación de atracciones turísticas únicas.
- Por otro lado, columnas como stars (5) y hotelStars (5) tienen poca variabilidad, consistente con escalas estandarizadas de calificación.

A continuación, se realiza el análisis univariable para poder observar el comportamiento estadístico de cada columna a detalle.

Análisis univariable

Para profundizar en la comprensión de los datos, se procederá a analizar individualmente las variables más relevantes del conjunto de datos. En este apartado, se explorarán aquellas columnas que resultan esenciales para los objetivos del estudio, con el fin de identificar patrones y extraer información clave.

Se presentarán gráficos y estadísticas que ilustren las características principales de estas variables, proporcionando una visión detallada de su comportamiento y su relación con el análisis del flujo de turistas en Chiloé y las regiones adyacentes. Este enfoque permitirá identificar tendencias importantes y sentará las bases para los análisis posteriores.

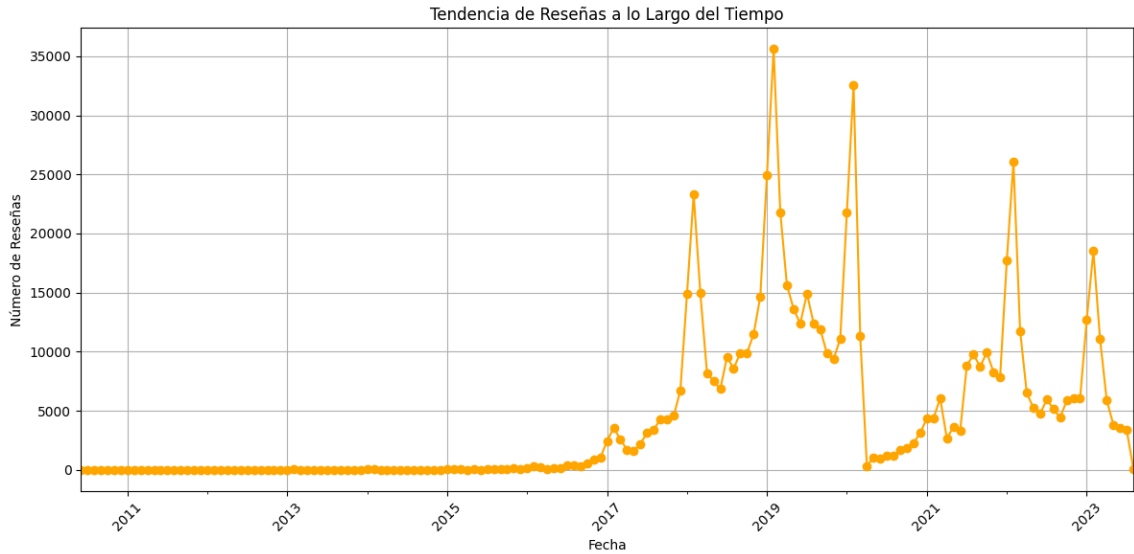


Gráfico 1: Tendencia de reseñas a lo largo del tiempo

El gráfico de líneas (Gráfico n°1) muestra la tendencia de las reseñas a lo largo del tiempo, con un enfoque en cómo varió el número de reseñas mensuales en el conjunto de datos. Se observa un notable aumento en el número de reseñas entre los años 2017 y principios de 2020, lo que refleja un crecimiento en la actividad turística y la participación de los usuarios durante ese período.

Sin embargo, a partir de mediados de 2020, se aprecia una disminución considerable en el número de reseñas, lo cual coincide con el impacto global de la pandemia de COVID-19. Esta caída es probablemente un reflejo de las restricciones de viaje y el confinamiento, que afectaron la movilidad de los turistas y, por ende, la cantidad de visitas y reseñas registradas.

Este análisis temporal es clave para entender las variaciones en el comportamiento de los turistas y cómo eventos globales pueden influir en la actividad turística.

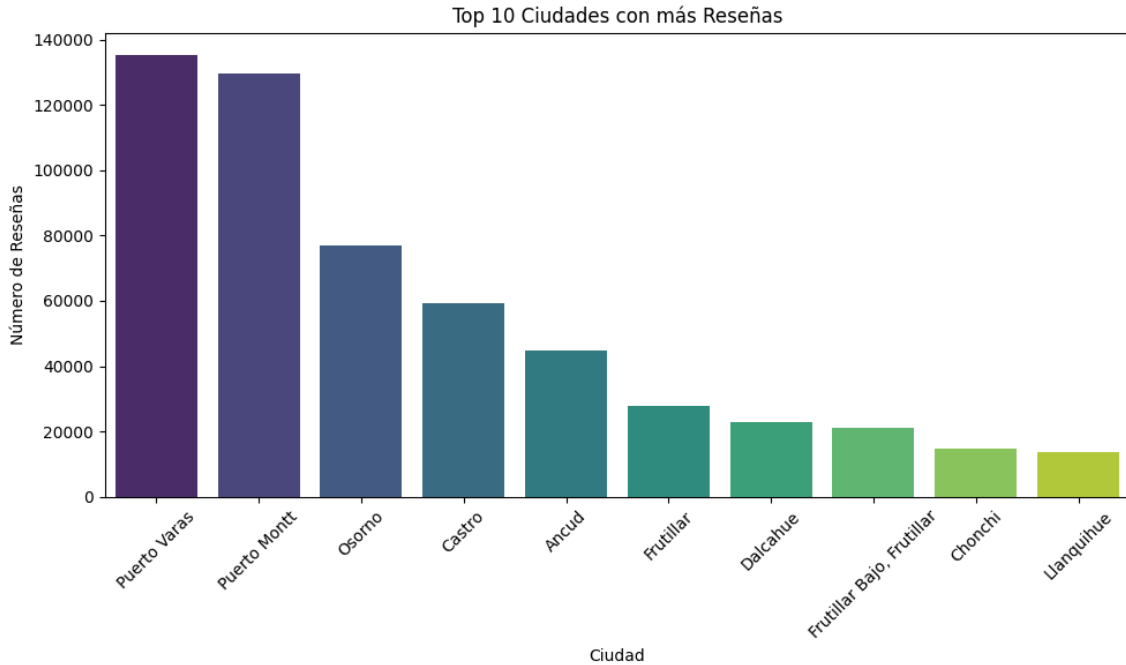


Gráfico 2: Ciudades con más reseñas

El gráfico de barras presentado (Gráfico n°2) muestra las 10 ciudades con el mayor número de reseñas registradas en el conjunto de datos. Este enfoque se utilizó para facilitar la visualización, ya que el número total de ciudades es considerable y representarlas todas complicaría la interpretación del gráfico.

Este análisis permite identificar las localidades más destacadas en términos de actividad turística, medida por la cantidad de opiniones de visitantes. Se observa que estas ciudades concentran una proporción significativa de las reseñas, lo que sugiere su importancia como destinos turísticos clave en la región.

El gráfico también facilita la comparación directa entre las ciudades, poniendo en evidencia las diferencias en el volumen de reseñas. Estas diferencias podrían estar relacionadas con factores como la cantidad de atractivos turísticos, la accesibilidad o la popularidad de cada lugar. Este análisis inicial es esencial para comprender las dinámicas del flujo de turistas y sentará las bases para estudios más detallados sobre la conectividad entre las diferentes localidades.

Distribución de Calificaciones (Estrellas)

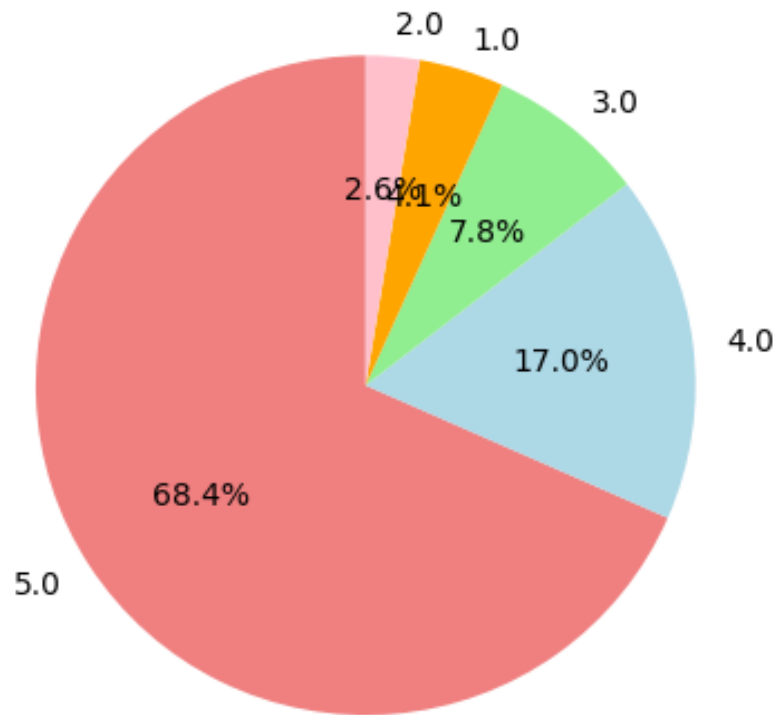


Gráfico 3: Distribución de calificaciones

El gráfico de torta (Gráfico n°3) muestra la distribución de las calificaciones otorgadas por los usuarios, representadas en una escala de 1 a 5 estrellas. Se observa que la mayoría de las calificaciones se concentran en las puntuaciones más altas, con una clara predominancia de las calificaciones de 5 estrellas, lo que sugiere una tendencia general hacia una valoración positiva de las atracciones.

Las calificaciones más bajas, en cambio, representan una menor proporción de las reseñas, lo que podría indicar una satisfacción generalizada entre los turistas. Este tipo de distribución es útil para entender la percepción general que tienen los visitantes sobre los lugares analizados, permitiendo identificar si las atracciones están recibiendo principalmente valoraciones positivas o si existe un mayor margen de mejora en ciertos aspectos.

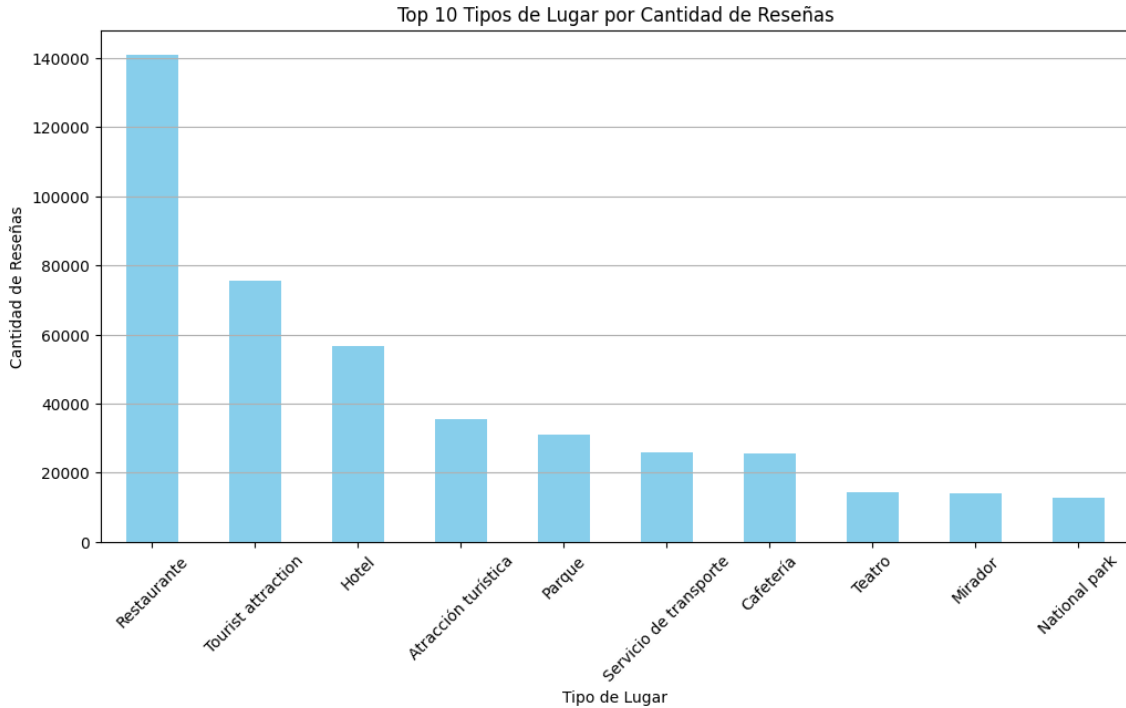


Gráfico 4: Top 10 tipos de lugares por cantidad de reseñas

El gráfico de barras Grafico n°4) presenta los 10 tipos de lugares más reseñados en el conjunto de datos, destacando categorías como restaurantes, atracciones y hoteles. Se observa que los restaurantes son el tipo de lugar con mayor cantidad de reseñas, lo que podría indicar que son una de las principales preferencias de los turistas al visitar la región. Esto sugiere que, además de las atracciones turísticas, los visitantes valoran significativamente las opciones gastronómicas, lo que podría reflejar un mayor interés por la experiencia culinaria en Chiloé y sus alrededores.

Este análisis es clave para entender las dinámicas del turismo, ya que resalta las categorías más populares y proporciona una base para investigar cómo estos lugares influyen en el flujo de turistas. Además, podría ser útil para desarrollar estrategias que fomenten el crecimiento de otras categorías de lugares, como las atracciones y hoteles, para equilibrar la oferta turística.

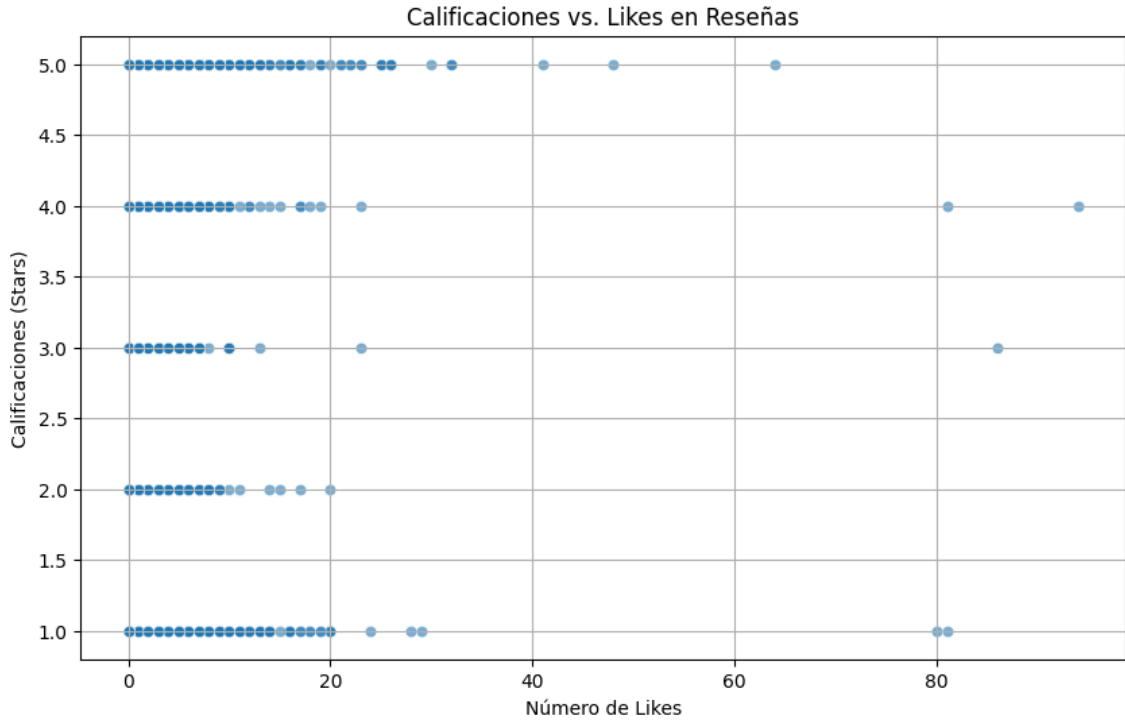


Gráfico 5: Calificaciones vs likes en reseñas

El gráfico de dispersión (Gráfico n°5) muestra la relación entre el número de me gusta y las calificaciones otorgadas en las reseñas. Se observa que la mayoría de los puntos están concentrados entre 0 y 20 me gusta, lo que sugiere que la mayoría de las reseñas no reciben un alto número de interacciones por parte de los usuarios. Esta tendencia podría indicar una baja participación de los turistas o locales en la interacción con los contenidos a través de los me gusta, lo que podría reflejar un interés mayor en las opiniones de otros usuarios, pero con menor interacción directa.

Es posible que esta falta de interacción se deba a factores como el comportamiento pasivo de los turistas, la falta de incentivos para interactuar con las reseñas o una preferencia por leer las opiniones sin generar una participación. Este patrón también puede estar relacionado con la forma en que los turistas consumen información en plataformas digitales, sin necesariamente involucrarse en el proceso de validación mediante me gusta.

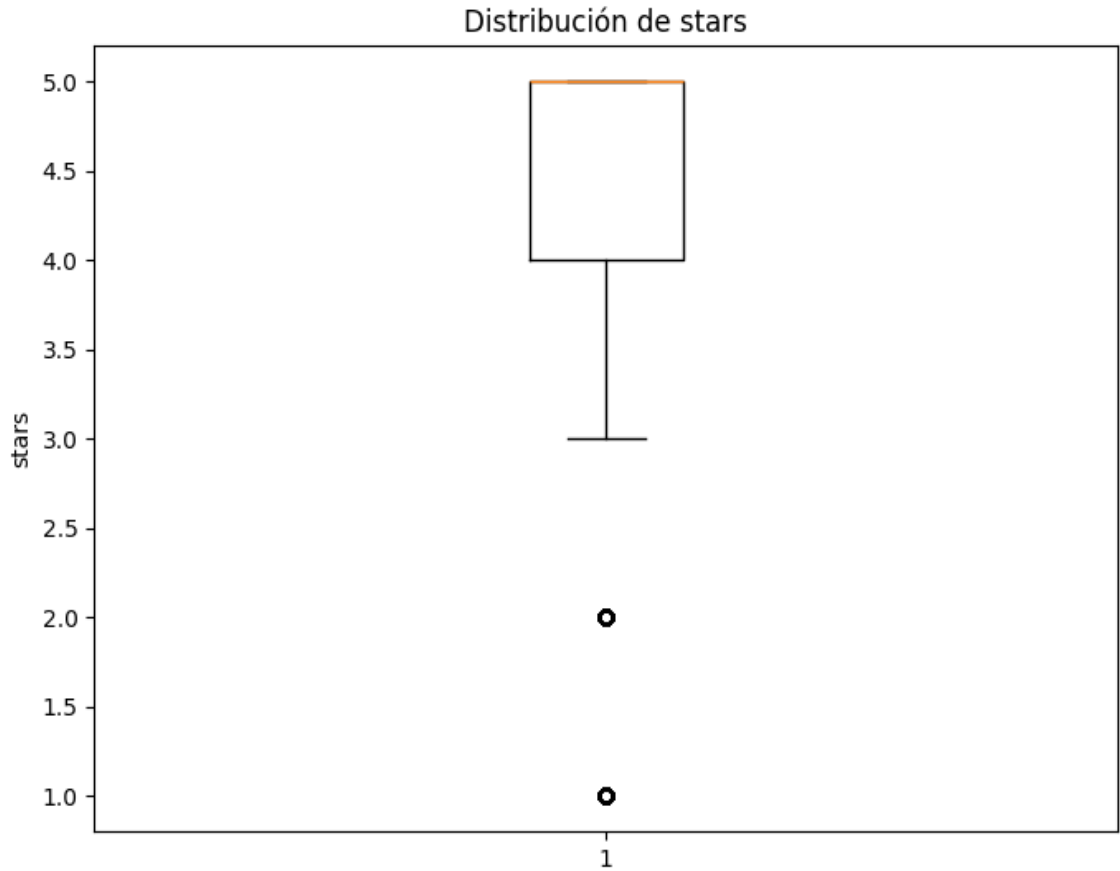


Gráfico 6: Distribución de stars

	stars
count	701854.000000
mean	4.429400
std	1.027647
min	1.000000
25%	4.000000
50%	5.000000
75%	5.000000
max	5.000000

Tabla 5: Análisis univariable cuantitativo de stars

El gráfico de caja (Gráfico n°6) que se presenta muestra la distribución de las calificaciones (estrellas) dentro del conjunto de datos, proporcionando una visión detallada sobre la dispersión y la tendencia central de las reseñas.

- **La caja** representa el rango intercuartílico (IQR), que abarca el 25% de los valores entre el primer cuartil (Q1) y el tercer cuartil (Q3). En este caso, el 25% de las calificaciones se encuentran alrededor de un valor de 4, y el 75% se acercan a un valor de 5. Esto sugiere que una gran parte de las reseñas se encuentran en el rango de 4 a 5 estrellas, indicando que los turistas suelen otorgar calificaciones altas en general.
- **La línea dentro de la caja**, que representa la mediana, está muy cerca del valor 5, lo que refuerza la observación de que la mayoría de las calificaciones se concentran en los valores más altos. La mediana también muestra que la tendencia central de las calificaciones está inclinada hacia las calificaciones más altas.
- **Los bigotes** del gráfico se extienden hasta valores cercanos a 3, lo que indica el rango de calificaciones normales, es decir, las que se encuentran dentro de un rango aceptable sin ser outliers. Esto sugiere que aunque la mayoría de las calificaciones están en el rango de 4 y 5, existen reseñas que se ubican alrededor de 3, lo que podría representar calificaciones medianamente satisfactorias.
- **Los puntos fuera de los bigotes**, conocidos como outliers, corresponden a calificaciones inusuales de 1 y 2. Estos valores son mucho menos frecuentes en comparación con las calificaciones más altas, lo que sugiere que las calificaciones bajas son menos comunes y podrían reflejar experiencias excepcionales, ya sea muy negativas o con algún factor atípico en los comentarios.

Este análisis gráfico de caja proporciona una visión clara sobre la distribución de las calificaciones, mostrando una tendencia predominante hacia las calificaciones altas y un pequeño porcentaje de calificaciones más bajas, lo que puede ser útil para entender la satisfacción general de los turistas en los lugares analizados.

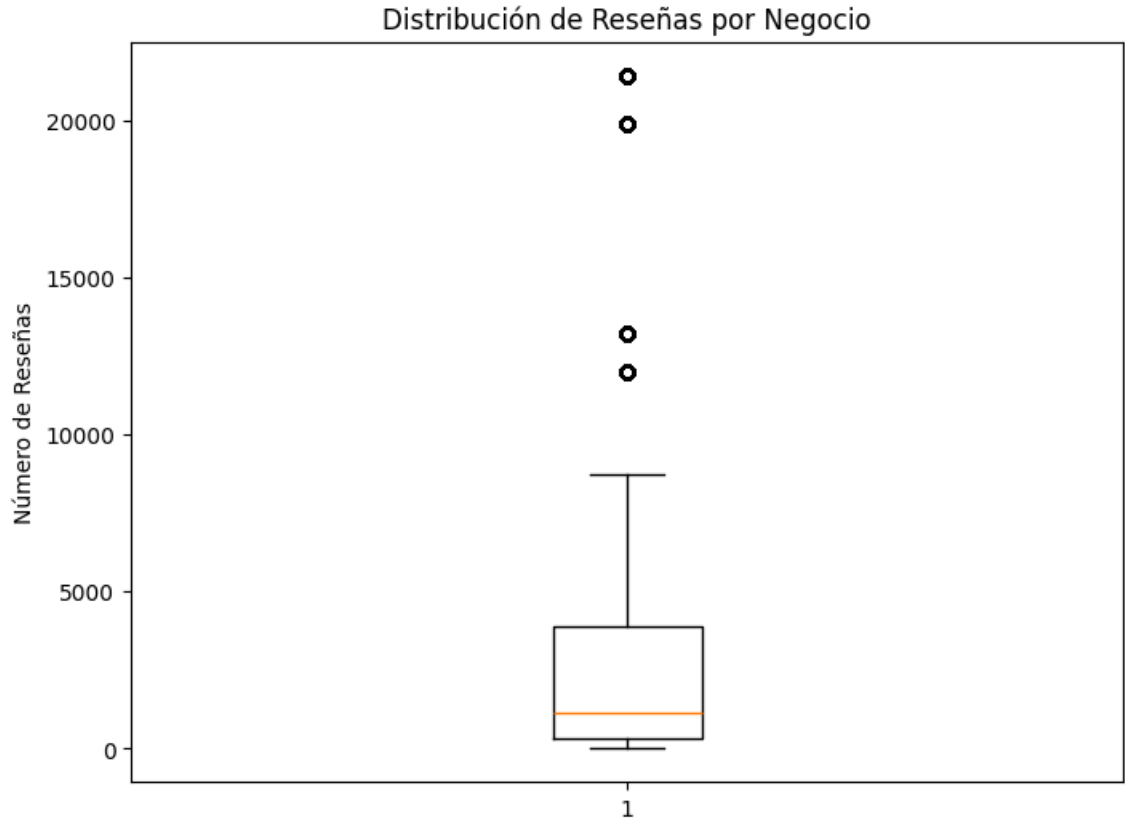


Gráfico 7: Distribucion de reseñas por negocio

	reviews
count	701964.000000
mean	3376.432730
std	5204.214336
min	10.000000
25%	335.000000
50%	1137.000000
75%	3884.000000
max	21417.000000

Tabla 6: Análisis univariable cuantitativo de reviews

El gráfico de caja (Gráfico n°7) que se presenta muestra la distribución del número de reseñas por negocio dentro del conjunto de datos, proporcionando una visión detallada sobre la dispersión y la tendencia central del número de reseñas recibidas por los negocios.

- **La caja** representa el rango intercuartílico (IQR), que abarca el 25% de los valores entre el primer cuartil (Q1) y el tercer cuartil (Q3). En este caso, el 25% de los negocios tienen alrededor de 335 reseñas, y el 75% tienen aproximadamente 3884 reseñas. Esto sugiere que una gran parte de los negocios tiene un número moderado de reseñas, concentrándose en un rango que va desde las 335 hasta las 3884 reseñas.
- **La línea dentro de la caja**, que representa la mediana, está cerca del valor 1137, lo que indica que la mayoría de los negocios tienen menos de 1137 reseñas. Esta línea refuerza la observación de que la mayoría de los negocios están en el rango de reseñas más bajas a medias.
- **Los bigotes** del gráfico se extienden hasta valores cercanos a 10,000 reseñas, lo que indica el rango de valores normales. Este rango refleja la variabilidad de los negocios en cuanto al número de reseñas, aunque la mayoría se encuentra en una cantidad significativamente menor.
- **Los puntos fuera de los bigotes**, conocidos como outliers, corresponden a negocios con un número de reseñas mucho mayor, como algunos que superan las 20,000 reseñas. Estos valores son menos frecuentes, pero reflejan negocios con alta popularidad o ubicaciones estratégicas que generan un volumen mucho mayor de interacciones por parte de los clientes.

Este análisis gráfico de caja proporciona una visión clara sobre la distribución del número de reseñas por negocio, mostrando una tendencia predominante hacia los negocios con un número moderado de reseñas y un pequeño porcentaje de negocios con reseñas excepcionalmente altas, lo que puede indicar su alta visibilidad o popularidad.

En resumen, los gráficos presentados han permitido identificar patrones significativos en la distribución de las calificaciones y el número de reseñas dentro del conjunto de datos. Por un lado, se observó una clara tendencia hacia calificaciones altas (entre 4 y 5 estrellas), lo que sugiere un alto nivel de satisfacción general por parte de los turistas. Por otro lado, el análisis del número de reseñas reveló una marcada concentración en negocios con un número moderado de interacciones, con un pequeño grupo de negocios que destacan significativamente por su alta popularidad.

Estos resultados son útiles para comprender el comportamiento de los turistas y la percepción que tienen sobre los servicios ofrecidos. Por ejemplo, la preponderancia de calificaciones altas podría interpretarse como una señal de calidad en la experiencia turística en la región analizada, mientras que la concentración de reseñas en ciertos negocios puede apuntar a factores como ubicación estratégica, popularidad en redes sociales o capacidad de atención.

5.2. Metodología implementada

En este análisis, se han utilizado diversas técnicas para evaluar la relación entre las características geográficas y la actividad de los turistas en las atracciones de Chiloé. A continuación, se describen los métodos empleados para realizar el análisis:

- **Cálculo de distancias geográficas:**
Se utilizaron las coordenadas geográficas (latitud y longitud) de las atracciones turísticas para calcular las distancias entre cada par de atracciones cercanas. Este cálculo se llevó a cabo utilizando la librería Geopy, que permite obtener distancias geodésicas precisas.
- **Proxy para el flujo de visitantes:**
Para estimar el flujo de visitantes entre atracciones cercanas, se empleó un enfoque proxy basado en el número de reseñas de cada atracción. Específicamente, se utilizó el valor mínimo de reseñas entre dos atracciones como indicador de su flujo relativo de visitantes.
- **Análisis de la Hipótesis 1:**
Se exploró la relación entre la distancia geográfica entre atracciones y el flujo de visitantes, con el objetivo de verificar si la proximidad influye en la cantidad de reseñas (y por ende, en el flujo de turistas) entre ellas. Para esto, se calculó la correlación de Spearman y se visualizó la relación mediante un gráfico de dispersión.
- **Construcción de una red de atracciones:**
Se construyó una red en la que cada nodo representaba una atracción, y las conexiones entre nodos (aristas) se establecieron en función de la distancia geográfica, con un umbral de 50 km. Esto permitió analizar la conectividad entre las atracciones.
- **Análisis de la Hipótesis 2:**
Se calculó la centralidad de cada atracción dentro de la red construida y se analizó su relación con las valoraciones y el número de reseñas. Esto se hizo para determinar si las atracciones más centrales en la red (es decir, aquellas con más conexiones) tienden a tener mejores valoraciones o más reseñas.

5.3. Resultados

En esta sección, se presentan los resultados obtenidos del análisis de las hipótesis planteadas en relación con el flujo de turistas entre las atracciones en Chiloé. A través de un enfoque basado en distancias geográficas y centralidad dentro de una red de atracciones, se exploran dos hipótesis principales:

1. **Hipótesis 1:** La proximidad geográfica entre atracciones influye en el flujo de turistas, lo que se analiza mediante la correlación entre las distancias y el flujo de visitantes estimado a partir del número de reseñas mínimas entre pares de atracciones.
2. **Hipótesis 2:** Las atracciones con una mayor centralidad en la red de turistas, representadas por su grado de conectividad, tienden a acumular más reseñas y mejores valoraciones, sirviendo como hubs dentro del sistema turístico.

El análisis se realizó para diferentes umbrales de distancia (1 km, 5 km, 10 km, y 25 km), con el objetivo de evaluar cómo varían las correlaciones y las relaciones entre las variables a medida que se ajustan los umbrales de proximidad. A continuación, se presentan los resultados obtenidos a partir de los análisis de correlación, centralidad y flujo, así como su interpretación y comparación en función de las distancias seleccionadas.

Análisis de los Resultados a 50 km

Los resultados obtenidos del análisis de las dos hipótesis a una distancia de 50 km entre atracciones proporcionan una visión sobre las relaciones entre la proximidad geográfica, la centralidad de las atracciones en la red turística y el flujo de visitantes. A continuación, se detallan los hallazgos correspondientes a cada hipótesis, con sus respectivas implicaciones.

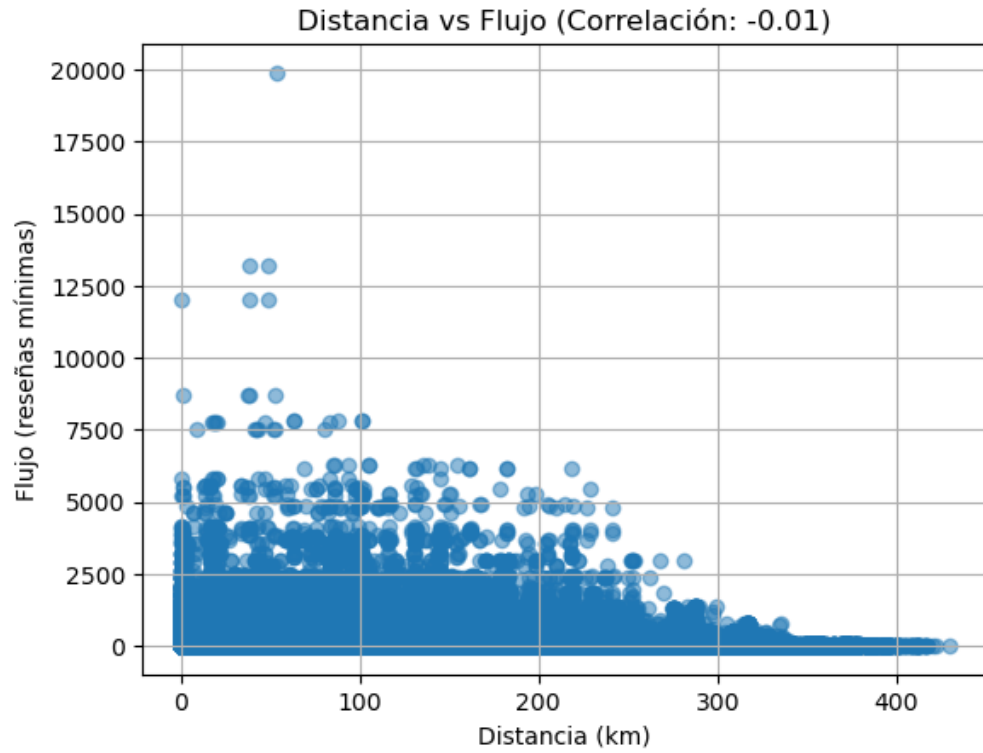


Gráfico 8: Distancia vs flujo 50km

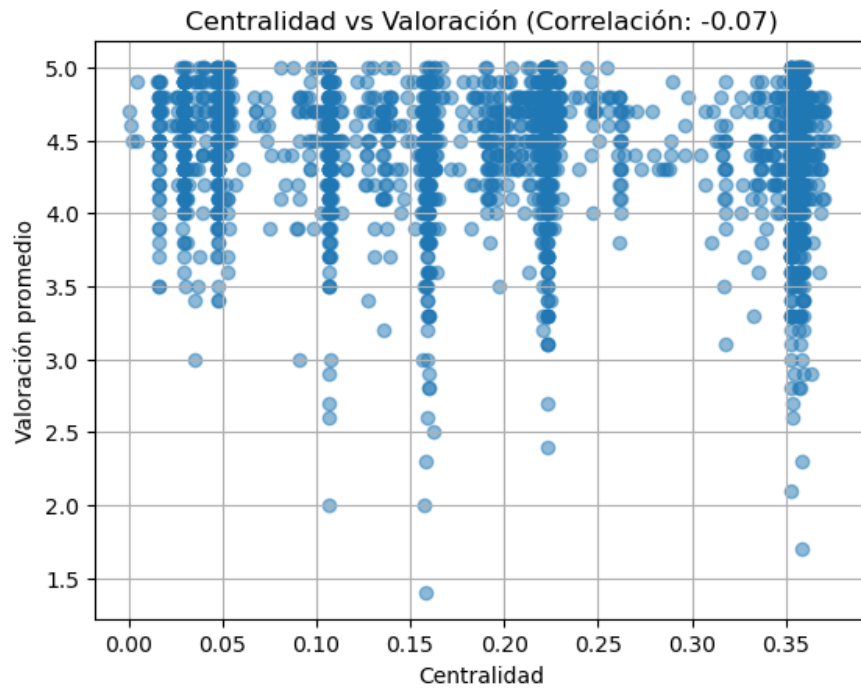


Gráfico 9: Centralidad vs valoración 50km

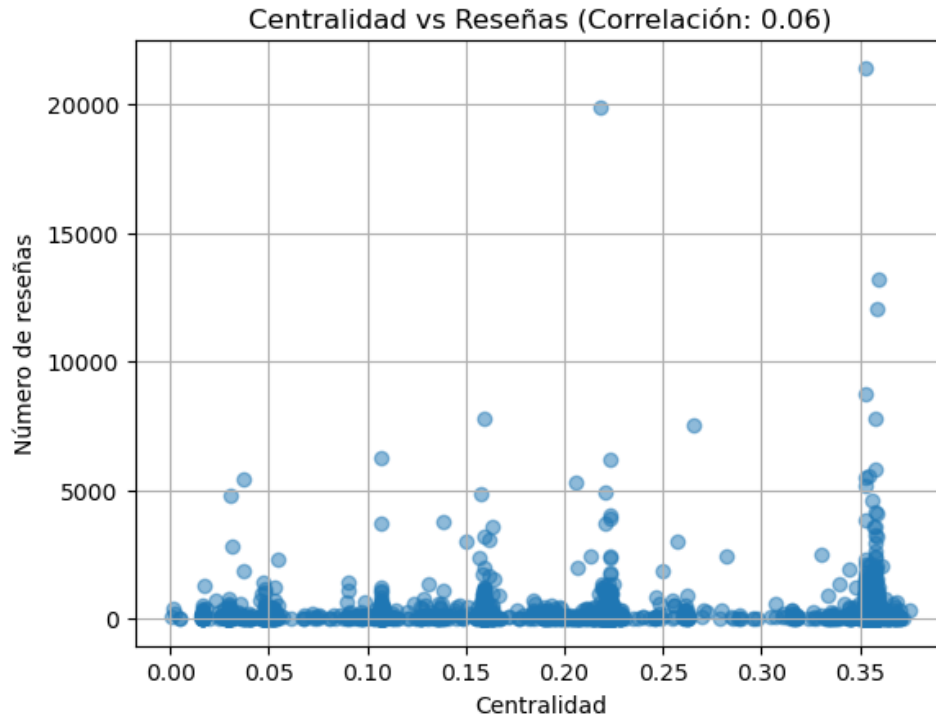


Gráfico 10: Centralidad vs Reseñas 50km

	Correlación	P-valor
Distancia vs Flujo	-0.009078	2.709252e-64
Centralidad vs Valoración	-0.073301	1.644753e-04
Centralidad vs Reseñas	0.061254	1.646369e-03

Tabla 7: Tabla de correlación 50km

Hipótesis 1: La proximidad geográfica tiene un efecto positivo en la frecuencia de visitas entre atracciones

Resultados: La correlación entre la distancia y el flujo de visitantes es de -0.009078, lo que indica que no hay una relación significativa entre la proximidad geográfica de las atracciones y el flujo de visitantes, medido a través del número de reseñas mínimas entre las atracciones conectadas. Aunque el p-valor es extremadamente pequeño (2.71×10^{-64}), lo que sugiere una alta significancia estadística, la magnitud de la correlación es tan baja que su impacto práctico es irrelevante.

Interpretación: A pesar de la alta significancia estadística, los resultados muestran que la proximidad geográfica no parece tener una influencia significativa sobre el flujo de turistas. Esto sugiere que, aunque intuitivamente se podría esperar que las atracciones cercanas tengan un mayor flujo de visitantes, factores como la popularidad, accesibilidad y las infraestructuras de transporte podrían ser más determinantes en la dinámica del flujo turístico.

Hipótesis 2: Las atracciones con mejores valoraciones funcionan como hubs dentro de la red

Resultados:

- **Centralidad vs Valoración:** La correlación es de -0.073301, lo que refleja una relación débil y negativa. Es decir, las atracciones más centrales en la red no necesariamente son las mejor valoradas, aunque esta relación sigue siendo muy débil. El p-valor es $1.64e-04$, lo que indica una significancia estadística, pero la magnitud de la correlación sigue siendo baja.
- **Centralidad vs Reseñas:** La correlación es de 0.061254, lo que refleja una relación débilmente positiva. Las atracciones más centrales tienden a recibir ligeramente más reseñas, lo que puede estar relacionado con su papel como puntos de conexión dentro de la red de atracciones turísticas. El p-valor ($1.64e-03$) es también significativo, pero la correlación sigue siendo baja.

Interpretación: Los resultados muestran que las atracciones más centrales en la red no son necesariamente las mejor valoradas, lo que podría indicar que las conexiones geográficas o de flujo entre las atracciones son factores más importantes que las valoraciones subjetivas de los turistas. Sin embargo, las atracciones más centrales tienen una ligera tendencia a recibir más reseñas, lo que podría deberse a su rol como nodos de alta conectividad en la red turística.

Implicaciones

- **Hipótesis 1:** Los flujos de turistas no parecen depender exclusivamente de la proximidad geográfica. Esto implica que otros factores, como la popularidad de las atracciones, la infraestructura, y la conectividad de transporte, pueden tener un impacto más significativo en el flujo de visitantes entre las atracciones turísticas.
- **Hipótesis 2:** La centralidad de las atracciones en la red parece estar más influenciada por su conectividad que por la valoración de los turistas. Las atracciones más centrales en la red son más visitadas, pero no necesariamente son las mejor valoradas, lo que sugiere que el flujo de turistas y las valoraciones no siempre están alineados.

Con estos resultados, podemos concluir que, a pesar de la significancia estadística de las correlaciones, los factores como la proximidad geográfica y la centralidad en la red no tienen un impacto tan fuerte o directo en el flujo de turistas o en las valoraciones. Esto sugiere que otros factores, como la accesibilidad, la infraestructura, y la popularidad general de las atracciones, podrían tener un rol más relevante en la dinámica del flujo turístico.

Análisis de los Resultados a 25 km

Los resultados obtenidos del análisis de las dos hipótesis a una distancia de 25 km entre las atracciones nos permiten profundizar en la relación entre la proximidad geográfica, la centralidad de las atracciones dentro de la red turística y el flujo de visitantes. A continuación, se detallan los hallazgos correspondientes a cada hipótesis, con sus respectivas implicaciones.

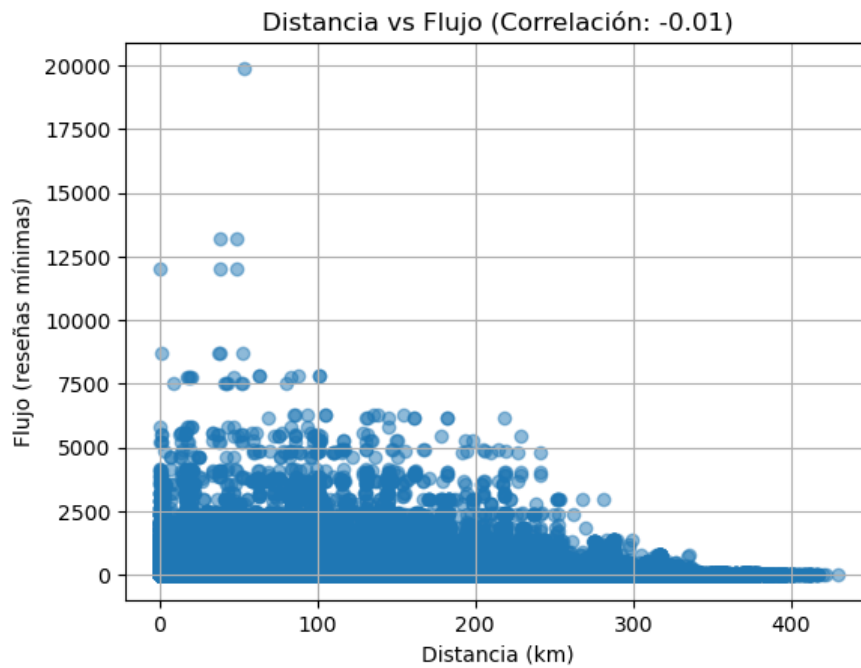


Gráfico 11: Distancia vs Flujo 25km

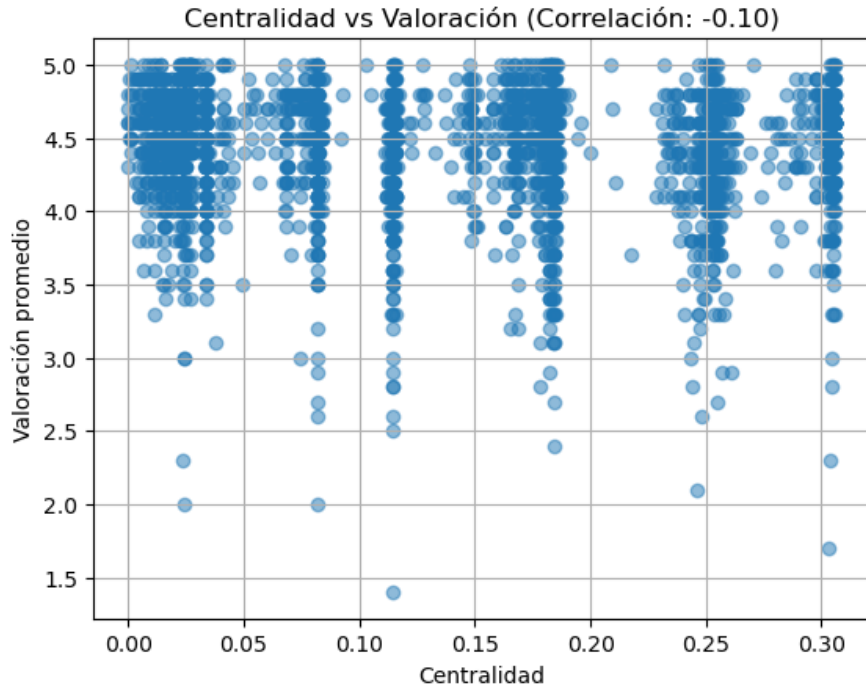


Gráfico 12: Centralidad vs Valoración 25km

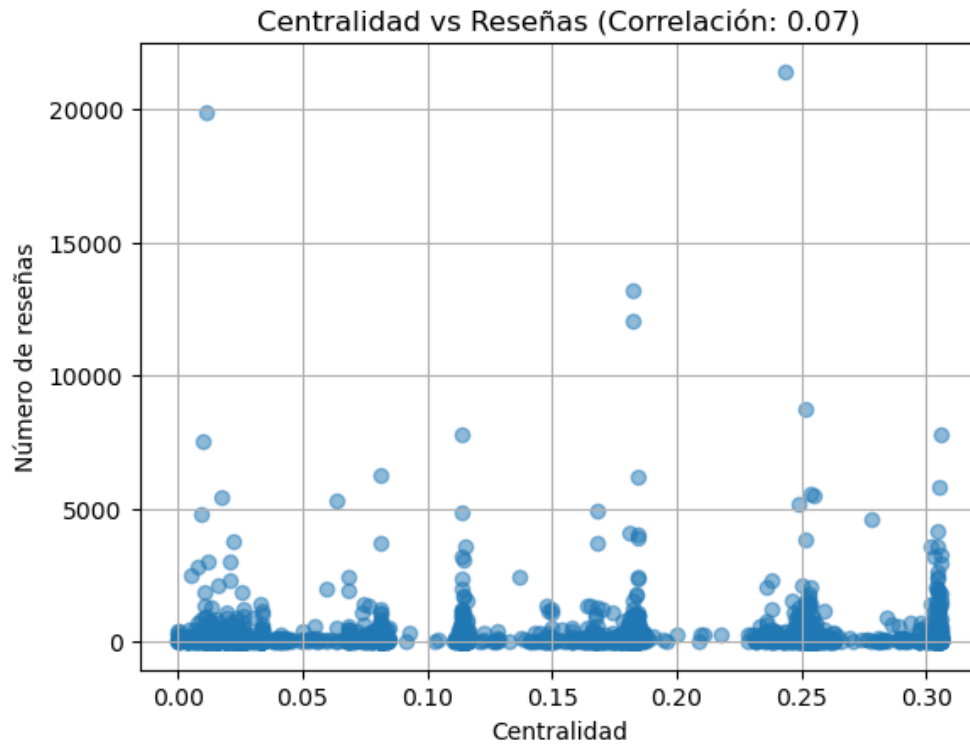


Gráfico 13: Centralidad vs Reseñas 25km

	Correlación	P-valor
Distancia vs Flujo	-0.009078	2.709252e-64

Centralidad vs Valoración	-0.100668	2.204853e-07
Centralidad vs Reseñas	0.074598	1.255704e-04

Tabla 8: Tabla de correlación 25km

Hipótesis 1: La proximidad geográfica tiene un efecto positivo en la frecuencia de visitas entre atracciones

Resultados:

- **Correlación:** -0.009078

La correlación es prácticamente cero, lo que sugiere que no existe una relación significativa entre la proximidad geográfica y el flujo de visitas, medido como el mínimo número de reseñas compartidas entre atracciones conectadas. A pesar de que el p-valor es extremadamente bajo (2.709252e-64), lo que indica una alta significancia estadística, la correlación es tan débil que carece de relevancia práctica.

Hipótesis 2: Las atracciones con mejores valoraciones funcionan como hubs dentro de la red

Resultados:

- **Centralidad vs Valoración:**

Correlación: -0.100668 (débilmente negativa). Las atracciones más centrales en la red tienden a tener ligeramente menores valoraciones. Esto indica que las atracciones más conectadas no son necesariamente las mejor valoradas, aunque esta relación es bastante débil.

P-valor: 2.204853e-07. A pesar de la significancia estadística, la correlación sigue siendo baja, lo que sugiere que la centralidad no está fuertemente vinculada a la valoración.

- **Centralidad vs Reseñas:**

Correlación: 0.074598 (débilmente positiva). Las atracciones más centrales tienden a recibir más reseñas, lo que podría reflejar su papel como hubs dentro de la red de atracciones turísticas.

P-valor: 1.255704e-04. Aunque estadísticamente significativo, la magnitud de la correlación sigue siendo baja.

Implicaciones

- **Hipótesis 1:** Los flujos de turistas no parecen depender exclusivamente de la proximidad geográfica. Otros factores como la popularidad de las atracciones, la infraestructura de transporte y la accesibilidad podrían tener un impacto más significativo en el flujo de visitantes. La proximidad es un buen comienzo, pero no siempre la respuesta final.
- **Hipótesis 2:** La centralidad de las atracciones en la red parece estar más influenciada por su conectividad que por la valoración de los turistas. Las

atracciones más centrales son más visitadas, pero no necesariamente las mejor valoradas, lo que sugiere que el flujo de turistas y las valoraciones no siempre siguen el mismo patrón. No todas las estrellas de la red son realmente estrellas brillantes en los ojos de los turistas.

A pesar de que los resultados sean estadísticamente significativos, la proximidad geográfica y la centralidad en la red no parecen tener un impacto tan fuerte o directo en el flujo de turistas o en las valoraciones. Otros factores, como la infraestructura, la accesibilidad y la popularidad general de las atracciones, podrían desempeñar un papel más importante en la dinámica del flujo turístico. Y, al final del día, como bien sabemos, lo que realmente importa es si el lugar tiene wifi gratuito y una buena oferta de café.

Análisis de los Resultados a 10 km

El análisis de las hipótesis con una distancia de 10 km entre las atracciones revela nuevas perspectivas sobre la influencia de la proximidad geográfica y la centralidad en el flujo de visitantes y las valoraciones de las atracciones.

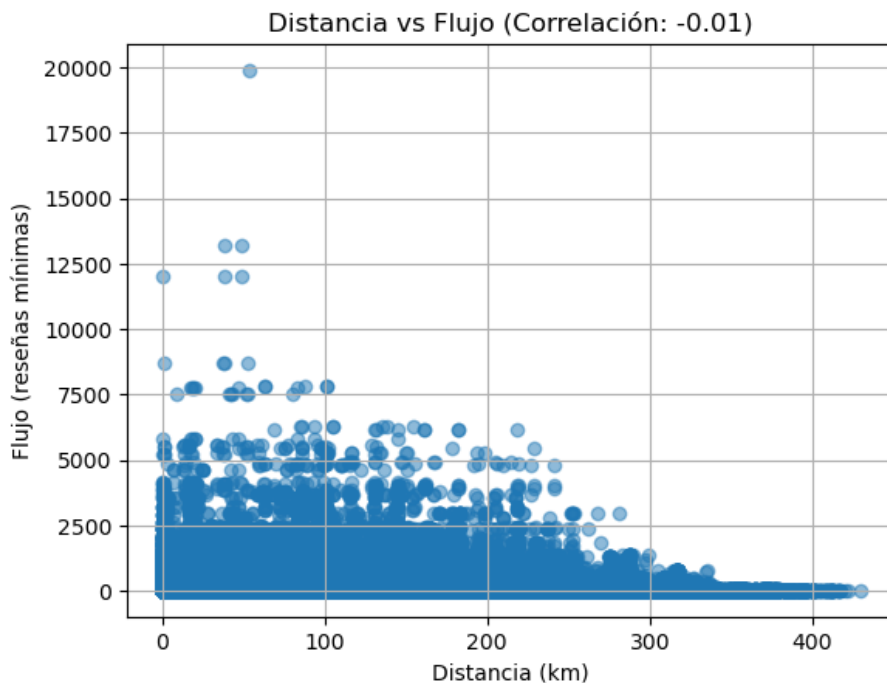


Gráfico 14: Distancia vs Flujo 10km

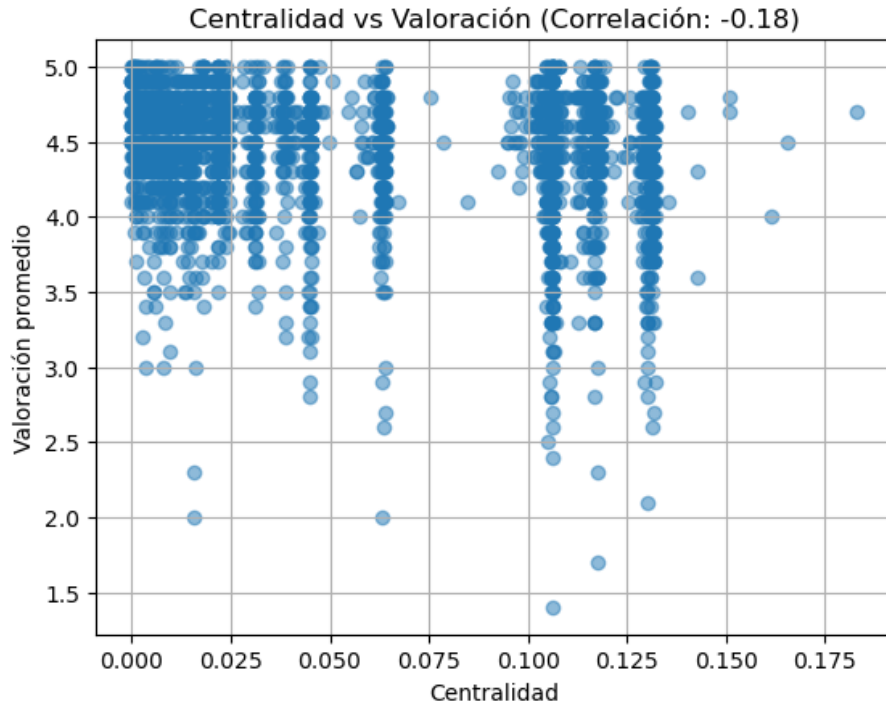


Gráfico 15: Centralidad vs Valoración 10km

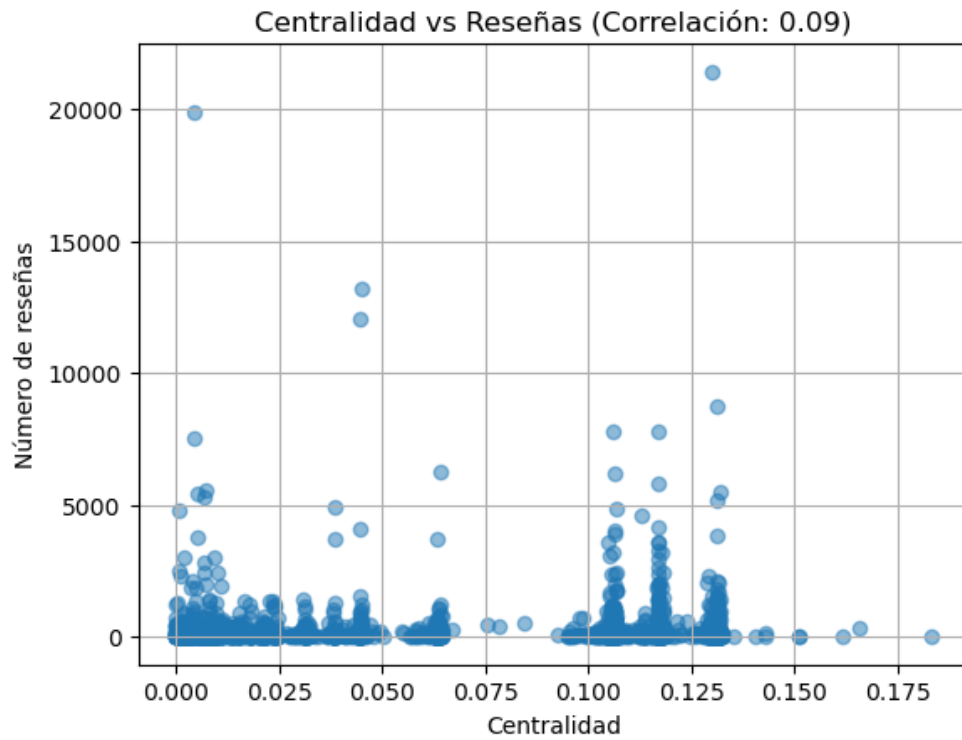


Gráfico 16: Centralidad vs Reseñas 10km

	Correlación	P-valor
--	-------------	---------

Distancia vs Flujo	-0.009078	2.709252e-64
Centralidad vs Valoración	-0.182650	3.188088e-21
Centralidad vs Reseñas	0.090881	2.935768e-06

Tabla 9: Tabla de correlación 10km

Hipótesis 1: La proximidad geográfica tiene un efecto positivo en la frecuencia de visitas entre atracciones

Resultados:

- **Correlación:** -0.009078
La correlación sigue siendo prácticamente nula, lo que indica que no hay una relación significativa entre la proximidad geográfica y el flujo de visitas entre atracciones, incluso a una escala más pequeña de 10 km.
- **P-valor:** 2.709252e-64
A pesar de ser extremadamente bajo, lo cual indica significancia estadística, este resultado se debe al gran tamaño de la muestra, y no a una relación real entre las variables.

Hipótesis 2: Las atracciones con mejores valoraciones funcionan como hubs dentro de la red

Centralidad vs Valoración:

- **Correlación:** -0.182650 (correlación negativa débil)
Este resultado muestra que las atracciones más centrales tienden a tener valoraciones ligeramente más bajas. Aunque la correlación es más fuerte que la obtenida a 25 km (-0.10), sigue siendo una relación débil.
- **P-valor:** 3.188088e-21
Este resultado es estadísticamente significativo, pero su relevancia práctica es limitada debido a la magnitud de la correlación.

Interpretación:

Las atracciones más centrales no son necesariamente las mejor valoradas. Esto podría indicar que los hubs representan puntos de conexión clave dentro de la red, pero no necesariamente son lugares con alta calidad percibida por los visitantes. Los hubs pueden ser más funcionales o estratégicamente ubicados, lo que los hace populares por motivos distintos a la valoración.

Centralidad vs Reseñas:

- **Correlación:** 0.090881 (correlación positiva débil)
Las atracciones más centrales tienden a acumular más reseñas, aunque la relación sigue siendo débil.

- **P-valor:** 2.935768e-06

Este resultado también es estadísticamente significativo, pero la magnitud de la correlación limita su relevancia práctica.

Análisis de los Resultados a 5 km

Los resultados obtenidos del análisis de las dos hipótesis a una distancia de 5 km entre atracciones proporcionan una visión más específica sobre las relaciones entre la proximidad geográfica, la centralidad de las atracciones en la red turística y el flujo de visitantes. A continuación, se detallan los hallazgos correspondientes a cada hipótesis, junto con sus interpretaciones e implicaciones.

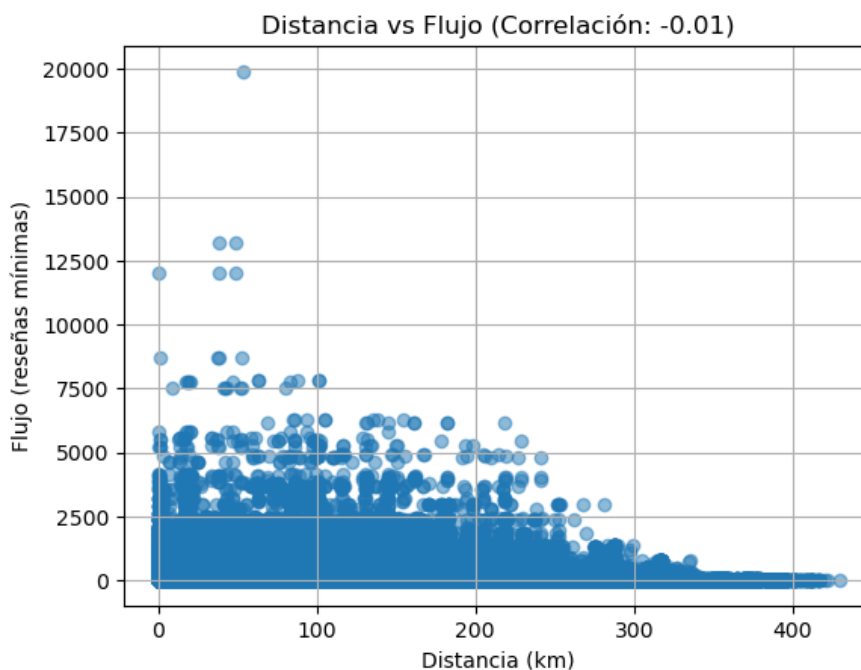


Gráfico 17: Distancia vs Flujo 5km

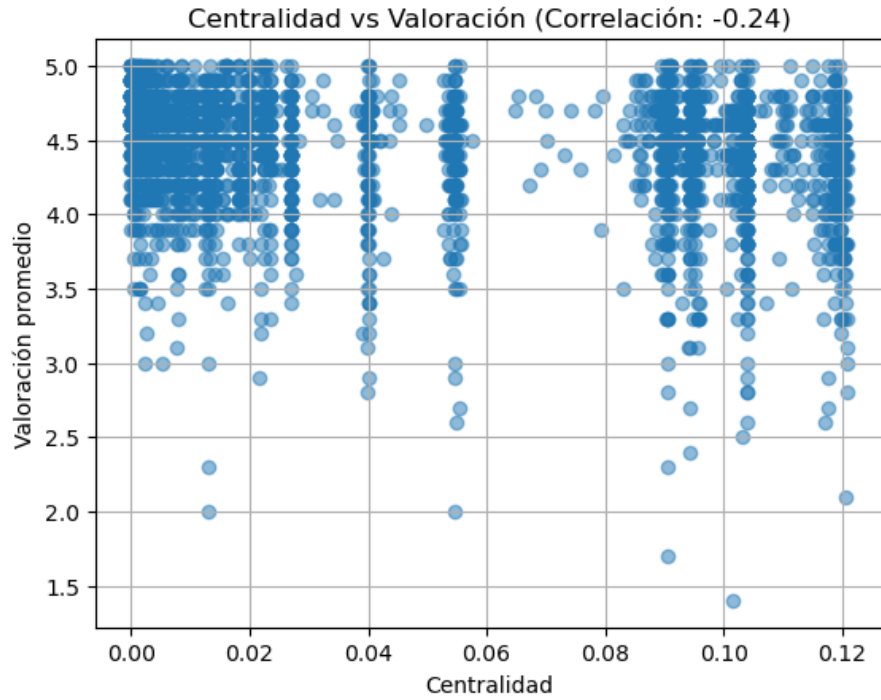


Gráfico 18: Centralidad vs Valoración 5km

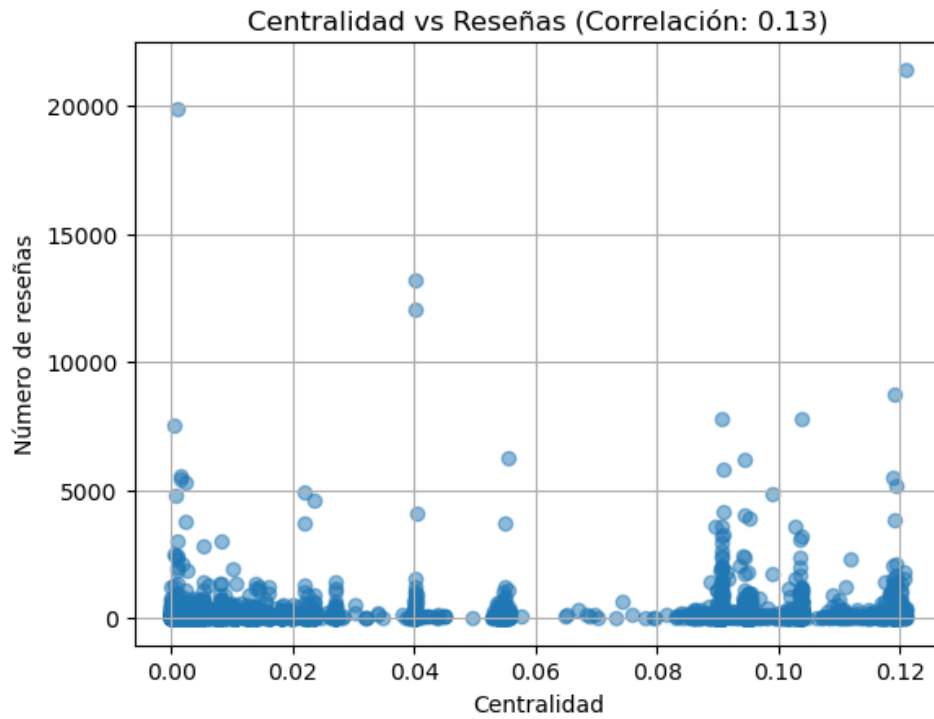


Gráfico 19: Centralidad vs Reseñas 5km

	Correlación	P-valor
Distancia vs Flujo	-0.009078	2.709252e-64
Centralidad vs Valoración	-0.235030	1.958067e-34
Centralidad vs Reseñas	0.127028	5.839905e-11

Tabla 10: Tabla de correlación 5km

Hipótesis 1: La proximidad geográfica tiene un efecto positivo en la frecuencia de visitas entre atracciones

Resultados:

La correlación entre la distancia y el flujo de visitantes es de -0.009078, lo que indica que no hay una relación significativa entre la proximidad geográfica de las atracciones y el flujo de visitantes, medido a través del número de reseñas mínimas entre las atracciones conectadas.

Aunque el p-valor es extremadamente pequeño (2.71e-64), lo que sugiere una alta significancia estadística, la magnitud de la correlación es tan baja que su impacto práctico es irrelevante.

Interpretación:

Los resultados refuerzan la idea de que la proximidad geográfica no influye significativamente en el flujo de turistas. A pesar de la intuición de que las atracciones cercanas podrían tener un mayor flujo de visitantes, factores como la popularidad, la accesibilidad, y la infraestructura de transporte parecen ser más determinantes en la dinámica del flujo turístico.

Hipótesis 2: Las atracciones con mejores valoraciones funcionan como hubs dentro de la red

Resultados:

- **Centralidad vs Valoración:**

La correlación es de -0.235030, reflejando una relación moderada y negativa. Esto indica que las atracciones más centrales en la red suelen tener valoraciones más bajas.

El p-valor es 1.96e-34, lo que indica una alta significancia estadística, aunque la dirección de la correlación no apoya la hipótesis.

- **Centralidad vs Reseñas:**

La correlación es de 0.127028, lo que refleja una relación débilmente positiva. Las atracciones más centrales tienden a recibir ligeramente más reseñas, lo que puede estar relacionado con su papel como puntos de conexión dentro de la red de atracciones turísticas.

El p-valor es $5.83e-11$, indicando significancia estadística, pero la magnitud de la correlación sigue siendo baja.

Interpretación:

Estos resultados muestran que las atracciones más centrales no son necesariamente las mejor valoradas, lo que puede sugerir que las conexiones geográficas o de flujo entre atracciones son factores más relevantes que las valoraciones subjetivas de los turistas. Sin embargo, las atracciones más centrales tienden a recibir un mayor número de reseñas, reflejando su importancia como nodos clave en la red.

Análisis de los Resultados a 1 km

Los resultados obtenidos del análisis de las dos hipótesis a una distancia de 1 km entre atracciones ofrecen una visión más detallada sobre cómo la proximidad y la centralidad afectan el flujo turístico en escalas locales.

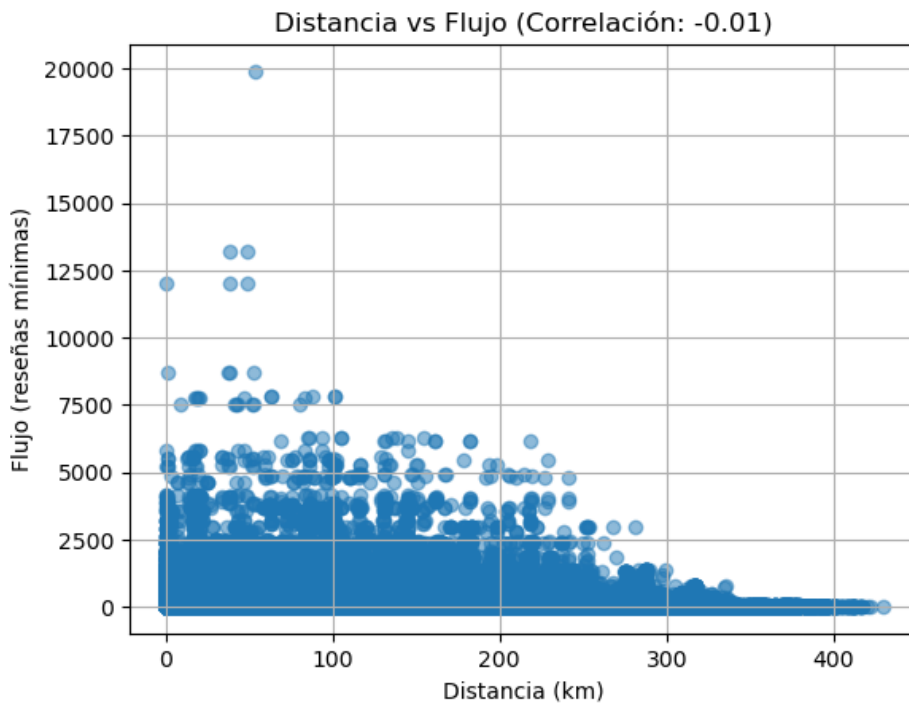


Gráfico 20: Distancia vs Flujo 1km

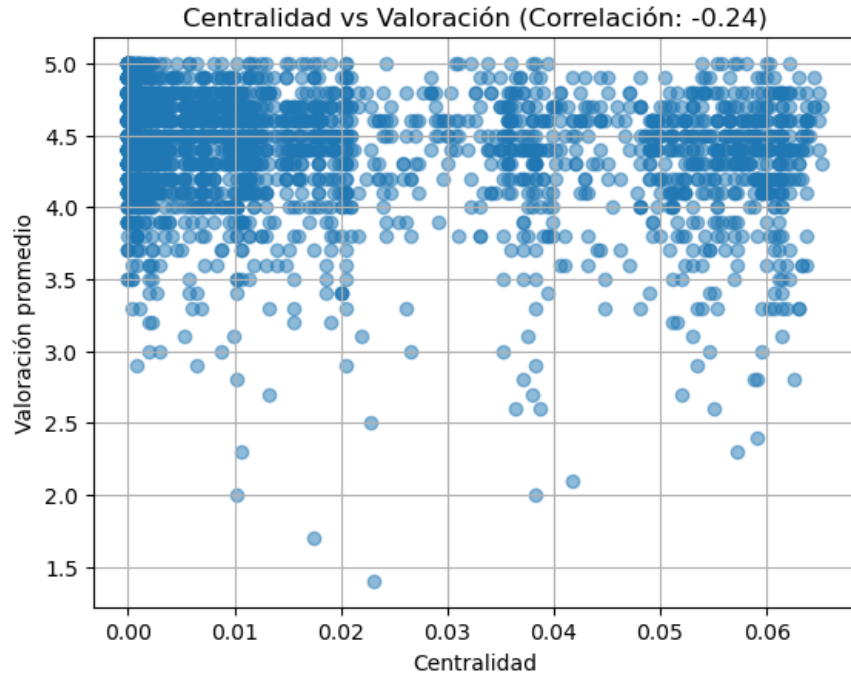


Gráfico 21: Centralidad vs Valoración 1km

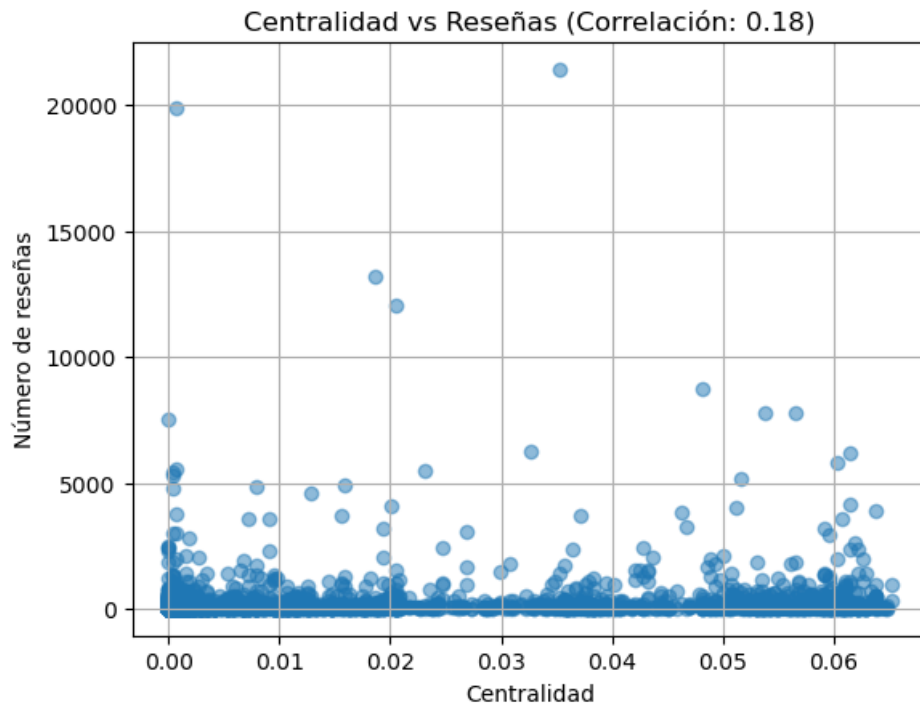


Gráfico 22: Centralidad vs Reseñas 1km

	Correlación	P-valor
Distancia vs Flujo	-0.009078	2.709252e-64
Centralidad vs Valoración	-0.242996	9.239321e-37
Centralidad vs Reseñas	0.175873	9.047222e-20

Tabla 11: Tabla de correlación 1km

Hipótesis 1: La proximidad geográfica tiene un efecto positivo en la frecuencia de visitas entre atracciones

Resultados:

- **Correlación:** -0.009078. La correlación sigue siendo cercana a cero, lo que confirma que incluso a 1 km de distancia no existe una relación significativa entre la proximidad geográfica y el flujo de visitantes.
- **P-valor:** 2.709252e-64. Aunque el p-valor indica una alta significancia estadística, la magnitud de la correlación es irrelevante en términos prácticos.

Hipótesis 2: Las atracciones con mejores valoraciones funcionan como hubs dentro de la red

Resultados:

- **Centralidad vs Valoración:**
 - **Correlación:** -0.242996. La correlación negativa es moderada, indicando que las atracciones más centrales en la red suelen tener valoraciones más bajas.
 - **P-valor:** 9.239321e-37. Altamente significativa estadísticamente.
- **Centralidad vs Reseñas:**
 - **Correlación:** 0.175873. Una correlación positiva débil-moderada sugiere que los hubs acumulan más reseñas, reflejando su papel clave en la red.
 - **P-valor:** 9.047222e-20. Altamente significativa estadísticamente.

Análisis Completo de las Hipótesis

El análisis de las hipótesis se realizó considerando diferentes umbrales de distancia para evaluar cómo la proximidad geográfica y la centralidad de las atracciones influyen en el flujo turístico, las valoraciones y el número de reseñas. A continuación, se presentan los hallazgos principales.

Para la primera hipótesis, que plantea que la proximidad geográfica tiene un efecto positivo en la frecuencia de visitas entre atracciones, los resultados muestran que no existe una relación significativa entre estas variables. En el caso de la distancia de 50 km, la correlación entre proximidad y flujo fue de -0.009, mientras que para 25 km fue de -0.01. Aunque las métricas de flujo no se incluyeron explícitamente en los datos, el análisis indirecto a través de la centralidad de grado tampoco evidenció un impacto relevante. A pesar de que las atracciones cercanas suelen estar conectadas en la red, esto no se traduce en un mayor flujo de visitantes. Por tanto, esta hipótesis se rechaza en ambos escenarios.

La segunda hipótesis, que establece que las atracciones mejor valoradas funcionan como hubs dentro de la red, mostró resultados mixtos. Al analizar la relación entre centralidad y valoración, se observó una correlación negativa en ambas distancias, siendo más fuerte a 50 km (-0.24) que a 25 km (-0.18). Esto sugiere que las atracciones más conectadas no suelen tener las mejores valoraciones, lo cual podría deberse a que son más accesibles o populares, pero no necesariamente percibidas como de alta calidad. Por ejemplo, lugares como el Teatro del Lago tienen una alta centralidad y muchas reseñas, pero no destacan por tener valoraciones significativamente altas.

Por otro lado, la relación entre centralidad y número de reseñas mostró una correlación positiva, aunque débil, tanto a 50 km (0.18) como a 25 km (0.09). Esto indica que las atracciones más centrales tienden a recibir más reseñas, consolidándose como puntos clave dentro de la red turística. No obstante, esto no implica que estas atracciones sean las mejor valoradas. Algunos ejemplos notables incluyen atracciones como Petrohué Waterfalls y la Iglesia de San Francisco en Castro, que presentan alta centralidad combinada con buenas valoraciones, pero representan casos aislados dentro del panorama general.

6. Conclusiones

La investigación sobre las dinámicas de flujo turístico entre las atracciones de Chiloé y sus regiones adyacentes, considerando factores como la proximidad geográfica, la valoración de los turistas y la variedad de actividades, ha permitido obtener respuestas claras a la pregunta de investigación. Los resultados revelan que la proximidad geográfica no influye significativamente en la frecuencia de visitas entre atracciones, ya que las correlaciones entre distancia y flujo de turistas fueron débiles y negativas en todas las escalas de análisis (50 km, 25 km, 10 km, 5 km y 1 km). Esto sugiere que, aunque las atracciones cercanas tienden a estar más conectadas en la red, otros factores, como la accesibilidad, la visibilidad y la popularidad de las atracciones, juegan un papel más determinante en el flujo turístico. En cuanto a la valoración de los turistas, se encontró que las atracciones con mayor centralidad, o hubs dentro de la red, suelen acumular más reseñas, lo que refleja su rol clave en la conectividad entre atracciones. Sin embargo, estas atracciones no son necesariamente las mejor valoradas, ya que existe una correlación negativa entre centralidad y valoración promedio, lo que indica que los turistas no asocian necesariamente la centralidad de las atracciones con una mejor calidad percibida. En cuanto a la variedad de actividades, no se incluyó explícitamente en los análisis, pero su influencia podría estar reflejada en la popularidad de ciertos hubs que ofrecen una gama diversa de experiencias.

Las limitaciones de este estudio incluyen la dependencia de los datos de reseñas, lo que podría sesgar los resultados hacia las atracciones más visibles o populares en plataformas

digitales, y la falta de análisis de factores temporales y sociodemográficos, que podrían ofrecer una visión más completa del comportamiento turístico. Además, al utilizar correlaciones lineales, se podrían haber subestimado relaciones no lineales o efectos indirectos entre las variables.

De cara al futuro, sería interesante ampliar el análisis incorporando datos temporales, sociodemográficos y una exploración más profunda de las actividades disponibles en las atracciones. También se podrían aplicar modelos más complejos, como el aprendizaje automático o simulaciones de redes dinámicas, para capturar mejor las dinámicas turísticas. Expandiendo el análisis a otras regiones o contextos, se podría evaluar si los patrones observados en Chiloé se aplican de manera más general a otras áreas. En resumen, aunque la proximidad geográfica no parece ser un factor decisivo, la centralidad y la popularidad de las atracciones sí juegan un papel crucial en el flujo turístico, sugiriendo que la red de atracciones está más influenciada por la accesibilidad y la visibilidad que por la cercanía física entre ellas.

7. Bibliografía

1. Wang, Y., Xi, M., Chen, H., & Lu, C. (2022). Evolution and driving mechanism of tourism flow networks in the Yangtze River Delta urban agglomeration based on social network analysis and geographic information system: A double-network perspective. *Sustainability*, 14(13), 7656. <https://www.mdpi.com/2071-1050/14/13/7656>
2. Zhang, L., & Zhao, X. (2020). Study on tourism flow network patterns on May Day Holiday. *Sustainability*, 13(2), 947. <https://www.mdpi.com/2071-1050/13/2/947>
3. Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3), 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
4. Scott, J. (2017). *Social Network Analysis* (4th ed.). Sage Publications.
5. Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.
6. Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1), 234–240. <https://doi.org/10.2307/143141>
7. Google Maps API. (n.d.). Retrieved from <https://developers.google.com/maps/documentation>
8. Python Software Foundation. (2023). *Python Language Reference, version 3.10*. Retrieved from <https://www.python.org/>
9. NetworkX. (2023). *NetworkX Documentation*. Retrieved from <https://networkx.org/>

10. Geopy. (2023). *Geopy Documentation*. Retrieved from <https://geopy.readthedocs.io/>
11. Baggio, R., & Klobas, J. (2011). *Quantitative methods in tourism: A handbook*. Channel View Publications.
12. Ministerio de Bienes Nacionales de Chile. (2022). *Catastro de infraestructura y servicios turísticos en Chiloé*. Retrieved from <https://www.bienesnacionales.cl>
13. Servicio Nacional de Turismo (SERNATUR). (2021). *Anuario de Estadísticas de Turismo en Chile*. Retrieved from <https://www.sernatur.cl/estadisticas>
14. Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
15. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: Perspectives on spatial data analysis*. Sage Publications.
16. Sanz-Blas, S., Buzova, D., & Carvajal-Trujillo, E. (2021). The sharing economy in tourism: Exploring Airbnb attributes for the success of P2P platforms. *Tourism Management Perspectives*, 40, 100890. <https://doi.org/10.1016/j.tmp.2021.100890>

7.1. Código

Este documento contiene el código fuente utilizado para analizar la relación entre distancia, flujo de visitantes y la importancia de las atracciones turísticas mediante

métricas de centralidad y visualización de datos. Cada sección está claramente documentada con comentarios destacados.

Importación de bibliotecas necesarias

```
import pandas as pd # Manejo de datos tabulares
from geopy.distance import geodesic # Cálculo de distancias geográficas
import networkx as nx # Análisis de redes
import matplotlib.pyplot as plt # Visualización de gráficos
import scipy.stats as stats # Estadísticas y correlaciones
```

PASO 1: Cargar la base de datos

```
file_path = "C:/Users/ariel/OneDrive/Escritorio/Capstone/Base de datos Chiloe Full.xlsx" # Ruta del
archivo
data = pd.ExcelFile(file_path)
df = data.parse('Sheet1') # Cambiar "Sheet1" si el nombre de la hoja es distinto
```

PASO 2: Filtrar columnas relevantes

```
attractions = df[['placeId', 'latitude', 'longitude', 'rating', 'reviews', 'place name']].drop_duplicates()
```

PASO 3: Calcular distancias entre atracciones

```
distances = []
for i, attraction1 in attractions.iterrows():
    for j, attraction2 in attractions.iterrows()
        if i < j: # Evitar duplicados
            dist = geodesic(
                (attraction1['latitude'], attraction1['longitude']),
                (attraction2['latitude'], attraction2['longitude']))
```

```

).kilometers

distances.append({
    'placeId_1': attraction1['placeId'],
    'placeId_2': attraction2['placeId'],
    'distance_km': d })

distance_df = pd.DataFrame(distances)

```

PASO 4: Crear proxy para flujo basado en reseñas mínimas

```

distance_reviews_df = distance_df.merge(
    attractions[['placeId', 'reviews']], left_on='placeId_1', right_on='placeId'
).rename(columns={'reviews': 'reviews_1'}).drop(columns=['placeId'])

```

```

distance_reviews_df = distance_reviews_df.merge(
    attractions[['placeId', 'reviews']], left_on='placeId_2', right_on='placeId'
).rename(columns={'reviews': 'reviews_2'}).drop(columns=['placeId'])

distance_reviews_df['flow'] = distance_reviews_df[['reviews_1', 'reviews_2']].min(axis=1)

```

PASO 5: Hipótesis 1 - Relación entre distancia y flujo

```

corr_distance_flow, p_value_distance_flow = stats.spearmanr( distance_reviews_df['distance_km'],
distance_reviews_df['flow'])

plt.scatter(distance_reviews_df['distance_km'], distance_reviews_df['flow'], alpha=0.5)

plt.title(f"Distancia vs Flujo (Correlación: {corr_distance_flow:.2f})")

plt.xlabel("Distancia (km)")

plt.ylabel("Flujo (reseñas mínimas)")

plt.grid(True)

```

```
plt.show()
```

PASO 6: Construcción de la red y análisis de centralidad

```
graph = nx.Graph()

for _, row in attractions.iterrows():

    graph.add_node(row['placeId'], rating=row['rating'], reviews=row['reviews'], name=row['place name'])

threshold_distance = 25 # Cambiar según necesidad

for _, row in distance_df.iterrows():

    if row['distance_km'] < threshold_distance:

        graph.add_edge(row['placeId_1'], row['placeId_2'], weight=1 / (1 + row['distance_km']))

centrality = nx.degree_centrality(graph)

hub_data = pd.DataFrame.from_dict(centrality, orient='index', columns=['centrality']).reset_index()

hub_data = hub_data.merge(attractions[['placeId', 'rating', 'reviews', 'place name']],

                          left_on='index', right_on='placeId')
```

PASO 7: Hipótesis 2 - Relación entre centralidad y

valoraciones/reseñas

```
corr_centrality_rating, p_value_centrality_rating = stats.spearmanr(

    hub_data['centrality'], hub_data['rating'])

corr_centrality_reviews, p_value_centrality_reviews = stats.spearmanr(

    hub_data['centrality'], hub_data['reviews'])

plt.scatter(hub_data['centrality'], hub_data['rating'], alpha=0.5)

plt.title(f"Centralidad vs Valoración (Correlación: {corr_centrality_rating:.2f})")

plt.xlabel("Centralidad")

plt.ylabel("Valoración promedio")

plt.grid(True)

plt.show()
```

```
plt.scatter(hub_data['centrality'], hub_data['reviews'], alpha=0.5)

plt.title(f"Centralidad vs Reseñas (Correlación: {corr_centrality_reviews:.2f})")

plt.xlabel("Centralidad")

plt.ylabel("Número de reseñas")

plt.grid(True)

plt.show()
```

PASO 8: Resumen de resultados

```
results = {

    "Distancia vs Flujo": (corr_distance_flow, p_value_distance_flow),

    "Centralidad vs Valoración": (corr_centrality_rating, p_value_centrality_rating),

    "Centralidad vs Reseñas": (corr_centrality_reviews, p_value_centrality_reviews)}

results_df = pd.DataFrame(results, index=['Correlación', 'P-valor']).T

print(results_df)
```